Variable Dependency 'D' Distributions: A General Framework to Generate Skew Elliptical Multivariate Copulas with Polytonal Dependency

Working paper presented to Institute of Actuaries Finance and Risk conference, Dublin June 2007

Martyn Dorey and Konstantina Papastamati
Quantitative Solutions
P-Solve ASSET SOLUTIONS

Abstract

A simple procedure generating a multivariate density function that satisfies high asymmetry and polytonal dependency is defined and studied. Traditional approaches to multivariate distributions develop functions that first establish joint density, with dependency inferred, and often poorly understood, as an adjunct at the end. The result is that separate copula designs, such as the Clayton or skew-T, are sought to capture particular dependency structures. The alternative approach presented here generates a joint density function after the dependency is explicitly predefined. This approach, which we have termed the *D* distribution, has produced a variety of satisfactory results relating to different univariate and multivariate distributions through the choice of an appropriate dependency transfer function. Here we show a new application of the transfer function to construct the flexible and hence highly applicable *D* distribution and *D* copula. This concept is connected to copulas, neural networks, skew normal distributions and conditioning on hidden variables.

Key Words: Transfer functions, neural networks, polytonal dependency, local dependence function, copulas, multivariate copulas, skew multivariate distributions, skew-normal, skew-Cauchy, skew-T.

1. Introduction

Motivation and aims

The motivation for this paper was the need for a tool for 'defensive portfolio engineering':

- System breaks are a feature of the markets, we need an econometric model able to identify thresholds of various variables of interest above or below which some kind of impact will occur (e.g. when oil price will have an impact on share prices).
- Asset models that better capture downside dependency and deepening dependency in tails enable robust asset allocation against liabilities. In addition, assets that become jointly negatively correlated in upside events and jointly positively correlated in downside events have been observed in the capital markets.

Asymmetric dependency features of capital markets complicate traditional measures such as the 'hedge ratio', which relies heavily on a parameter driven by correlation. Sensitivity testing and hedging of a portfolio requires a methodology where the beta may only exist in the downside tail, or where the beta may reverse on deep upside and downside tails. So if we consider a joint distribution shaped like a banana, traditional techniques will not be able to capture this type of joint probability distribution. The aim of this study was to identify a method that can control and generate banana-shaped distributions, where correlation switches across the distribution.

This paper aims to challenge the following paradigms:

- The sum of normal distributions is normal.

- Beta and portfolio hedging ratios are symmetric.
- Different dependency structures require different copulas.
- Dependency falls out of a probability density function.

The paper also introduces the idea that a single correlation number is actually made up of two partial correlations multiplied together; in the same way that variance is two standard deviations multiplied together.

2. Background

2.1 Methods of introducing skewness

There is a large body of literature on parametric families of non-normal multivariate distributions. The study of the skew normal distribution by Azzalini and Dalla Valle (1996) is strongly connected to this work.

Some of the methods used to introduce skewness are briefly described below. These broadly fall into the same category of multiplication by some constant.

2.1.1 A skewing variable

Conic moments for modelling jump diffusion processes (Barndorff-Nielsen, 1998. Also discussed by the economist Smith, 2003). The Normal Inverse Gaussian distribution is an example of applying a skewing variable

2.1.2 Dividing and multiplying by distributions

This is an extension of 2.1. Some such examples involving various distributions are the work by Azzalini and Dalla Valle (1996) (skew normal distribution) and the work by Tan and Peng (2005) (skew-slash distributions). For additional references and a review on related literature, see Arnold and Beaver (2002) and Genton (2004) for a collection of papers on the subject. Distributions such as the t-distribution can be constructed independently using the normal distribution and multiplying it by an inverse gamma distribution, which has the effect of stretching the tails.

2.1.3 Conditioning

Samples can be taken from a variable conditioned on hidden variables. The latter have cut off points, which can determine how they influence the seen variables. In simple terms, a conditioning variable is forced to take a specific value so that we can investigate its effect on the rest of the variables; for example, US equity returns can be set to -10%. The correlation between markets leads to an impact on other international equity returns. Hidden variables and truncated sampling can generate highly flexible joint distributions, and arguably any joint distribution can be generated from sufficient numbers of truncated samples.

2.1.4 Local Density functions

Holland and Wang (1987) defined the local dependence function as the change of the natural logarithm of a joint distribution for a change in the corresponding marginals. They also proved that a local dependence function and two marginals uniquely determine the joint bivariate normal density.

2.1.5 Neural Networks

Neural networks attempt to approximate a variety of functions and systems from simple algebraic functions to systems of pattern recognition, signal processing, etc. They use a sequence of stages, in which transfer functions together with appropriate parameter values, are used to manipulate inputs and transform them into a smaller set of outputs with discontinuities and rotations. A transfer function is a function that takes an input between [$-\infty$ ∞] and forces it to a particular range [0 1], or [-1 1]. An inverse normal function is an example of a transfer function, as is the logistic function $\{1/[1-\mathbf{exp}(-x)]\}$. A summation of logistic transfer functions and continuous nesting of transfer functions could in theory fit any surface, and in the literature it is assumed that this process broadly mimics the operation of biological neurons. (Hagan *et al.*, 1996).

2.1.6 Dependency driven distributions

It would be appropriate to contrast these approaches with the proposed *D* distribution. The *D* distribution connects together approaches 2.1.3 to 2.1.5 by considering that variables could have been generated by conditioning them on an unseen uniformly distributed variable, *U*. A univariate dependency structure on the unseen variable is defined as a function of *U*. The joint dependency is simply the multiplication of the univariate dependencies and the uniform variable is essentially the height of a joint copula. The conventional approach for the generation of distributions defines the marginal and joint distributions first, while the dependency structure is solved for afterwards. In this new approach, the univariate dependency is defined first and the marginal distribution follows.

2.2 Preliminaries

2.2.1 Elliptical distributions

We start with the definition of the spherical distribution. A vector $Y = (Y_1, Y_2,..., Y_n)^T$ has a spherical distribution, if for every orthogonal map $U \in R^{n \times n}$ ($UU^T = U^T U = I_{n \times n}$), UY has the same distribution as Y [2], which using the symbolism from Embrechts *et al.*, (2002), is expressed as

$$UY =_d Y. \quad \text{Eq. 2.1}$$

The stochastic representation of a spherical distribution is given as $Y \sim S_n(\varphi)$ if and only if

$$Y =_{d} R \cdot U \qquad \text{Eq. 2.2}$$

where U is a random vector uniformly distributed on the unit hypersphere $S_{n-1} = \{y \in R^n \mid y^T y = 1\}$ in $R^n \geq 0$ and R is a positive random variable independent of U and known as the generating variable. As a consequence, spherical distributions can be regarded as mixtures of uniform distributions defined on spheres of various radii in R^n.

An n-dimensional continuous random variable X is said to have an elliptical distribution with parameters μ (n x 1) and Σ (n x n) if Y has the same distribution as $\mu + A^T Y$, where Y follows a k-dimensional spherical distribution with characteristic generator φ (characteristic function $\psi_Y(t) = \varphi(t^T t)$) and A is a (k x n) matrix such that $A^T A = \Sigma$ with rank(Σ) = k. The elliptical distribution is an extension of the n-dimensional normal distribution $N_n(\mu, \Sigma)$. By setting $\varphi(u) = \exp(-u/2)$, a vector variable X (n x 1) with n-dimensional distribution $N_n(\mu, \Sigma)$ is also said to follow a n-dimensional elliptical distribution, $EC_n(\mu, \Sigma, \varphi)$ (source: *Wikipedia*). Thus, mathematically elliptical distributions are affine maps of spherical distributions in R^n.

2.2.2 The geometric parallel of correlation

In practice, different radii of spherical and elliptical distributions correspond to different correlations. In the elliptical case we deal with unequal correlations. The concepts of absolute and relative variance and of relative correlation can be easily understood through a geometrical parallel, as illustrated below.

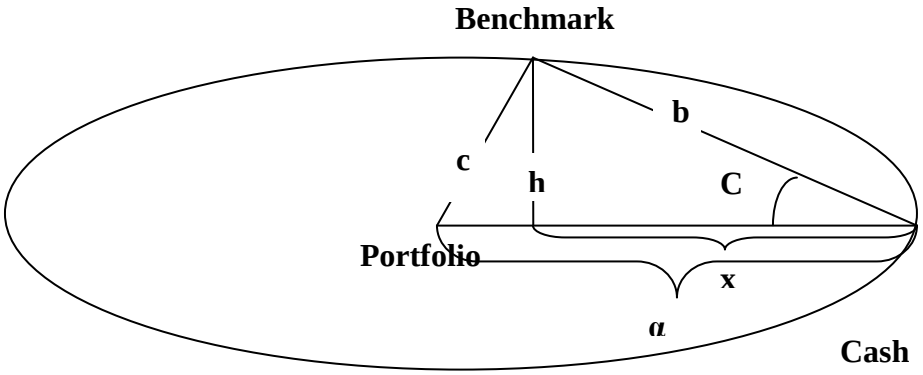


Fig. 2.1. Illustration of the correspondence between the cosine rule and the concept of relative correlation.

Applying now the Pythagorian theorem we get the following relationships

$$c^2 = (a-x)^2 + h^2 \Rightarrow c^2 = a^2 + x^2 - 2ax + h^2$$
$$b^2 = x^2 + h^2 \qquad \text{Eq. 2.3}$$
$$\cos(C) = \frac{x}{b} \Rightarrow x = b\cos(C)$$

Therefore,

$$c^2 = a^2 + b^2 - 2ab\cos(C) \Rightarrow \cos(C) = \frac{a^2 + b^2 - c^2}{2ab} \qquad \text{Eq. 2.4}$$

If we replace α, b and c with the relative variances or absolute covariance relating to the portfolio, P_1, and cash, P_2, the expression for the relative correlation coefficient is a mirror image of the cosine rule.

$$\text{Rel.corr.coeff.} = \frac{\text{Rel.}Var(P_1) + \text{Rel.}Var(P_2) - \text{Abs.}Cov(P_1, P_2)}{2\sqrt{\text{Rel.}Var(P_1)}\sqrt{\text{Rel.}Var(P_2)}} \Rightarrow$$

$$\rho_{(P_1-B,P_2-B)} = \frac{\sigma^2_{(P_1-B)} + \sigma^2_{(P_2-B)} - \sigma^2_{(P_1-P_2)}}{2\sqrt{\sigma^2_{(P_1-B)}}\sqrt{\sigma^2_{(P_2-B)}}} \qquad \text{Eq. 2.5}$$

where B is the benchmark. *If* we have a portfolio with asset weights W and corresponding covariance matrix, C, and a portfolio of the same assets with weights relative to some benchmark B and corresponding relative covariance matrix C_R, then

$$(W - B)' C (W - B) = W' C_R W \qquad \text{Eq. 2.6}$$

Dorey's lemma

We have named Eq. 2.6 Dorey's lemma, and the solution of C_R is a suitable exercise to set to quantitative job candidates.

2.2.3 Conditioning an elliptical distribution

When (Y_1, Y_2), with $Y_1 \in R^p$, $Y_2 \in R^q$ and $p+q = n$, follow an elliptical distribution $EC_n(\mu, \Sigma, \varphi)$, then the conditional distribution of Y_1 given Y_2 $(Y_1|Y_2)$ is also elliptical, even though with a different generator $\tilde{\varphi}$. So, $Y_1|Y_2 \sim EC_p(\mu_{1.2}, \Sigma_{11.2}, \tilde{\varphi})$, where

$$\mu_{1..2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Y_2 - \mu_2)$$
$$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \qquad \text{Eq. 2.7}$$

When a multivariate distribution of dimension n is conditioned, the resulting conditional distribution is of dimension (n-1) because the conditioning variable is set to a specific value (hence it behaves as a constant and not a variable). In Fig. 2.2, the 3-D plot of the bivariate normal distribution is sliced where vector *vecX2* is equal to -1. This slice is the conditional distribution N(*vecX1, vecX2*|*vecX2* = -1) (Fig. 2.2 below).

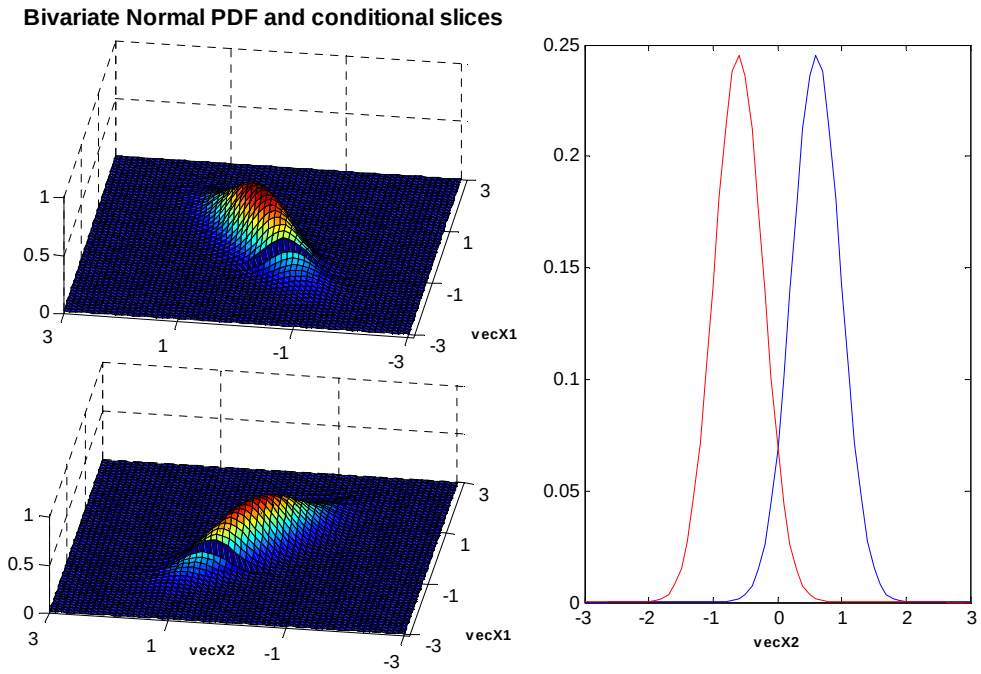


Fig. 2.2. The bivariate Normal distribution N(vecX1, vecX2; Mu, Sigma) and the conditional slices N(vecX1, vecX2| vecX1 = -1; Mu, Sigma) when the correlation coefficient is 0.8 (top-left and red line, respectively) or -0.8 (bottom-left and blue line, respectively).

2.2.4 Pearson distributions

Another generalisation of the normal distribution is the family of the Pearson distributions. These distributions are used in financial applications because of the ability to parameterise them in a way meaningful for market traders. Particularly the Pearson VII distribution function has a wide range of applications due to the ability to vary its width and decay rate of its tails, depending on the value of parameters *K* and *M*, respectively (Prevéy, 1986):

$$f(x) = \left[1 + K^2 (x - x_0)^2 / M\right]^{-M} \qquad \text{Eq. 2.8}$$

For $M = 1$, we get the Cauchy distribution, $M = 2$ the Lorentzian and for M approaching infinity we get the Gaussian distribution. Furthermore, in some areas of research including finance and social sciences, it is important to know the distribution of the product of two variables X and Y, which could be correlated or not. Nadarajah and Kotz (2006) studied this problem when X and Y are jointly distributed with a Pearson type VII distribution. The latter is given by the following expression:

$$f(x,y) = \frac{N-1}{\pi m\sqrt{1-\rho^2}}\left(1 + \frac{x^2 + y^2 - 2\rho xy}{m(1-\rho^2)}\right)^{-N} \qquad \text{Eq. 2.9}$$

for $-\infty < x < +\infty$, $-\infty < y < +\infty$, $N > 1$, $m > 0$ and $-1 < \rho < 1$. The bivariate t distribution and the bivariate Cauchy distribution are special cases of the Pearson type VII bivariate

distribution for $N = (m+2)/2$ (m is the degrees of freedom) and $m = 1$, $N = 3/2$, respectively. The bivariate Pearson type VII distribution has become very popular because it is a generalization of the univariate t distribution and has more realistic tails in view of analysis of real data, as opposed to the more limited multivariate Normal distribution. Overall, Pearson proposed a continuous probability density proportional to

$$(1+x/\alpha_1)^{\nu\alpha_1}\ (1-x/\alpha_2)^{\nu\alpha_2} \qquad \text{Eq. 2.10}$$

for $-\alpha_1 < x < \alpha_2$, which generalises the hypergeometric distribution. Taking different limits of this expression, he derived the types I-III and V. Type II, is characterised by a limited range in both directions and by symmetry. Olano and North (1997) mention the type II Pearson distribution as part of a class of distributions with elliptical contours. The density function for each member of this class of distributions is given as a function of the following distance function

$$d(x,\mu,\Sigma) = (x-\mu)\Sigma^{-1}(x-\mu)^T \qquad \text{Eq. 2.11}$$

Then the general expression of the density function is given as

$$f(x,\mu,\Sigma) = \frac{k}{|\Sigma|^{1/2}}\, g\left[d(x,\mu,\Sigma)\right] \qquad \text{Eq. 2.12}$$

It follows that if

$$g\left[d(x,\mu,\Sigma)\right] = g(t) \text{ then}$$
$$\text{if } g(t) = e^{-t/2} \Rightarrow \text{Gaussian} \qquad \text{Eq. 2.13}$$
$$\text{if } g(t) = \left(1-\frac{t}{2^m}\right)^m \Rightarrow \text{Pearson type II distribution}$$

The parameter m in the Pearson type II distribution is a constant whose value controls how close it gets to the Gaussian distribution. Considering the symmetry of the type II Pearson distribution, a d-dimensional point is said to be centrally symmetric around a point ξ, if $X - \xi = \xi - X$. This requirement implies that the corresponding density function f satisfies $f(y - \xi) = f(\xi - y)$.

3. Copulas

3.1 Introduction

Copulas are tools for understanding the relationship within multivariate events. Simplistically copulas explain relationships by working with percentiles. The copula is a surface that describes the dependency structure between percentiles.

Regression analysis is a widely used statistical technique for the investigation of relationships between variables. However, regression always separates one variable (the

dependent) from the rest (the independent or explanatory), while copulas look at all variables comprising a multivariate event and explain the joint multivariate distribution of these variables (Frees and Valdez, 1997). While the marginals of ordinary multivariate distributions share the same characteristics, copulas have the advantage that different assumptions can be made for the marginal distribution of each component of a multivariate event. Then the joining of these different assumptions through the use of a copula, leads to a better understanding of the underlying relationships within multivariate events and consequently better modelling of such events and their components. Copulas are multidimensional distributions with uniform marginals on [0, 1]. A copula C of dimension n is defined as follows

$$C(u_1, u_2, ..., u_n) = \Pr ob(U_1 \le u_1, U_2 \le u_2, ..., U_n \le u_n) \qquad \text{Eq. 3.1}$$

If we choose some marginal distributions $F_i(x_i) = u_i$, $i = 1, 2, \ldots, n$, so that

$$C\left[F_1(x_1), F_2(x_2), ..., F_n(x_n)\right] = F(x_1, x_2, ..., x_n) \qquad \text{Eq. 3.2}$$

we get

$$C(u_1, u_2, ..., u_n) = F\left[F_1^{-1}(u_1), F_2^{-1}(u_2), ..., F_n^{-1}(u_n)\right] \qquad \text{Eq. 3.3}$$

which is the copula of the joint distribution of the n-dimensional set of variables (X_1, X_2,…, X_n)

3.2 Fréchet-Hoeffding bounds for joint distribution functions.

For every copula C and vector of percentiles $u = (u_1, u_2, \ldots, u_n)$ in I^n, the following inequality is always true

$$W^n(u_1, u_2, ..., u_n) \le C(u_1, u_2, ..., u_n) \le M^n(u_1, u_2, ..., u_n) \qquad \text{Eq. 3.4}$$

where $M^n(u_1, u_2, \ldots, u_n) = \min(u_1, u_2, \ldots, u_n)$ and $W^n(u_1, u_2, \ldots, u_n) = \max(u_1 + u_2 + \ldots + u_n - n + 1, 0)$ and known as the Fréchet-Hoeffding upper and lower bound, respectively (see Fig. 3.1 below).

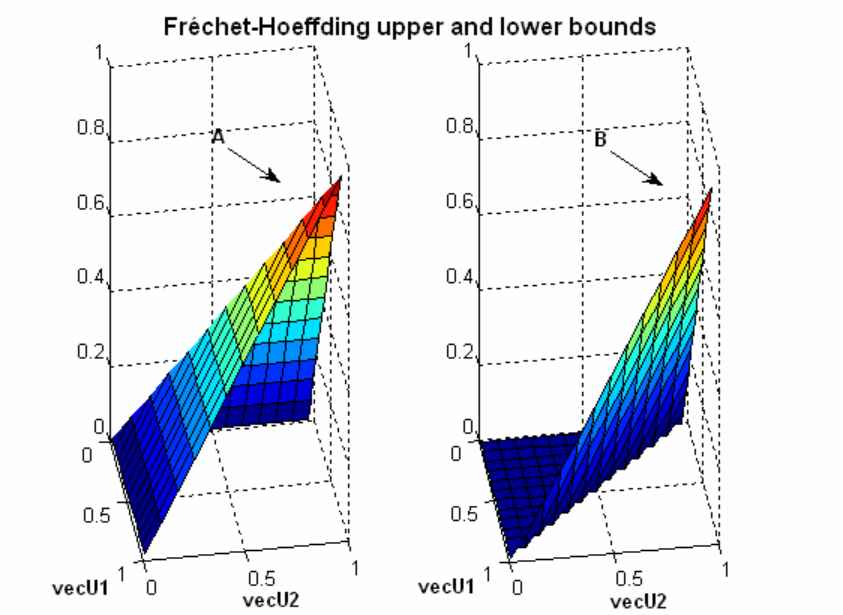


Fig. 3.1. The Fréchet-Hoeffding upper (left) and lower (right) bounds, which apply to all copulas.

In plain language, this means that any copula always lies between *A* and *B* (Fig. 3.1). When $n = 2$, both limits in Eq. 3.4 are copulas and *A* and *B* are the bivariate distribution functions of $(U, 1\text{-}U)^T$ and $(U, U)^T$, respectively, with $U \sim U(0, 1)$. It is then said that *A* represents perfect negative dependence, while *B* describes perfect positive dependence (Embrechts *et al.*, 2002). The transfer function in the *D* distribution aims to directly control whether we are considering the minimum or maximum by controlling the dependency up the univariate dependency structure.

In terms of our objective, we considered whether a weighted sum of bounds *A* and *B* (Fig. 3.1 above) with turning points could create polytonal copulas. However, this proved difficult to fit without an underlying assumption of a p.d.f. This led us to the *D* distribution described in the next section.

4. The *D* distribution

4.1 Definition

The *D* distribution is a conceptually straight forward idea. Generate a variable *U* that describes whether variables are at their lowest or highest percentile. Now include a dependence function that generates simulations where the dependency structure is dependent on *U*.

We shall define a positive random variable *X* as distributed with a *D* distribution if its distribution is expressed as a function of variable *U* with random generator distribution, *R*, dependency transfer function, *T*, and transfer function parameters, *B*. Then

$$\overset{d_1}{X} \sim D\left(\overset{k}{U}, \overset{d_2}{R}\left(X, Y(U) \middle| Y(U)\right), \overset{d_1}{T}(U), \overset{d_1}{B}\right) \qquad \text{Eq. 4.1}$$

where:

$\boldsymbol{d_1}$ is the number of dimensions modelled with $d_1 \geq 1$ and $d_2 = d_1+1$.

$\boldsymbol{X}$ is a conditioned vector of random variables, each directly independent of each other but directly dependent on a hidden variable U.

$\boldsymbol{U}$ is a continuous uniform univariate variable , i.e. U~U(0,1). U can be described as a percentile generating variable or the height of the joint copula surface and must have an inverse. Correlation is a function of U, i.e. $T(U)$ (see below). There must be at least one U variable, i.e. $1 \leq k \leq d_2$.

$\boldsymbol{R}$ is best described as a random generator that can be conditioned. The purpose of R is to carry the effects of the percentile generating function to the conditional distribution function it represents, the carrier function. R is a vector of multivariate random number generators or density generators of dimension d. A requirement of the random number generator is that it must be conditioned on k variables U through its parameters which are functions of the U variable/s. In order to construct a univariate distribution of a variable X with values x, we use a bivariate distribution and X represents a sample from the corresponding conditional distribution of $(X|Y = y)$, where variable Y is the univariate inverse of variable U. If, for example, the carrier distribution function is the Normal distribution and we want to generate a univariate D distribution, then y = NormalInverse(U) and the conditional distribution for $(X|Y = y)$ is derived from the bivariate normal distribution.

$\boldsymbol{T}$ is a vector of transformation functions that define some dependency pattern of dimension d_1. These functions take as input the values u of the variable U, which is uniform in the range [0 1], and convert them into measures of real dependency in the range [-1,1]. Examples of possible transfer functions are the logistic and the sine functions.

$\boldsymbol{B}$ is the vector (univariate case, β_i) or matrix (multivariate case, β_{ij}) of parameters determining the dependency transfer function, T, with the number of row vectors equal to the dimension parameter, d_1 (d_1 vector parameters). By changing the vector parameters the dependency structure, i.e. the correlation, changes (i refers to the transfer function and j to the parameter).

For simplicity, it is often helpful to have the same random number generator and transfer function. We will adopt the convention that where the same random number generator and transfer function are used, there is only a single row argument in the D distribution function definition. Furthermore, when there is only one variable U, we omit it from the D descriptor. The number of generated X variables is then indicated by the number of rows in the parameter matrix, B. Initially, we shall ignore location and scale parameters and assume we are modelling normalised distributions.

4.2 Example

4.2.1Definition

A variable *X* with values *x*, is *D* distributed as follows:

$$x \sim D\left(Normal, Logistic, \begin{bmatrix} -1 & 1 & 0 \end{bmatrix}\right) \quad \text{Eq. 4.2}$$

The random generator is a bivariate normal distribution and the dependency transfer function of the uniform variable *U* is the logistic function with row parameter vector, *B*, with elements β_1 = -1, β_2 = 1 and β_3 = 0. There is only one row of parameters because here we describe a univariate distribution

4.2.2 Generating samples from this D distribution

The stochastic form for this univariate distribution is a three step process:

1. Sample values u ~U(0,1) to represent percentiles of the variable *X*.
2. For each sample of *U* calculate a corresponding correlation ρ using the dependency transfer function, $T(u)$. In this example, $T(u)$ (and hence ρ) is a logistic function with parameters -1, 1 and 0.
3. Generate a sample from a conditioned bivariate normal distribution using the dependency transfer function $T(u)$, with mean, μ, and covariance, Σ.

So, for the stochastic representation we have:

$u \sim U(0,1)$ and is randomly generated Eq. 4.3

y = norminv(u), is the conditioning variable Eq. 4.4

$$T(u) = \frac{2}{1+e^{-n}} - 1, \text{ where } n = \beta_1 + \beta_2 u + \beta_3 u^2 \quad \text{Eq. 4.5}$$

$$\mu = \begin{bmatrix} \mu_1 & \mu_2 \end{bmatrix} = \begin{bmatrix} T(u) y & y \end{bmatrix}, \text{ refers to } (x|y, y) \quad \text{Eq. 4.6}$$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 - T^2(u) & 0 \\ 0 & 0 \end{bmatrix}, \text{ refers to } (x|y, y) \quad \text{Eq. 4.7}$$

$$x = \mu + (\Sigma)^{1/2} z \text{ , where } z \sim N(0,1) \quad \text{Eq. 4.8}$$

Even though *Y* is normally distributed, *X* will follow a skewed path determined by the βi parameters in the transfer function. In Table 1 we summarise the parameters and functions involved for the construction of the univariate standard *D* distribution, which we define as a univariate *D* distribution with carrier function the Normal distribution.

Table 1. Parameters and functions involved in the generation of the standard *D* distribution, i.e. the carrier distribution function is the normal distribution.

Parameters/Functions	Formula
Percentile generating function	U~U(0, 1)
Conditioning variable	y = norminv(u)
Transfer function	$T(u) = \frac{2}{1+e^{-n}} - 1$,
B parameter vector function	$n = \beta_1 + \beta_2 u + \beta_3 u^2$
Mean vector of (x|y, y)	$\mu = [\mu_1 \quad \mu_2] = [T(u) y \quad y]$
Covariance matrix of (x|y, y)	$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 - T^2(u) & 0 \\ 0 & 0 \end{bmatrix}$
Values of bivariate carrier distribution	$x = \mu + (\Sigma)^{1/2} z$, z ~N(0, 1)

In Table 2 we give a numeric example of applying Eq. 4.4-4.8 when the parameter vector is B = [-1 1 0] and in Fig. 4.1 we illustrate the full results for the same parameters.

Table 2. Worked example

Values of y, n, $T(u)$, mean, μ, variance, Σ_{11} and x as described by Eq. 4.4-4.8 for a sample value of U, u, when the parameter vector is B = [-1 1 0].

U = u	***y=normal inverse(u)* (Eq. 4.4)**	***n= -1 +1*u+0*u^2* (Eq. 4.5)**	**T(u) (Eq. 4.5)**	***μ* (Eq. 4.6)**	**Σ_{11} (Eq. 4.7)**	***x***
0.371	-0.329	-0.629	-0.305	0.1	0.907	-0.213

So the random value u = 0.371 leads directly to y as its normal inverse. The mean is the product of -0.329 and -0.305 and the standard deviation is $\left[1-(-0.305)^2\right]$ from which x is sampled.

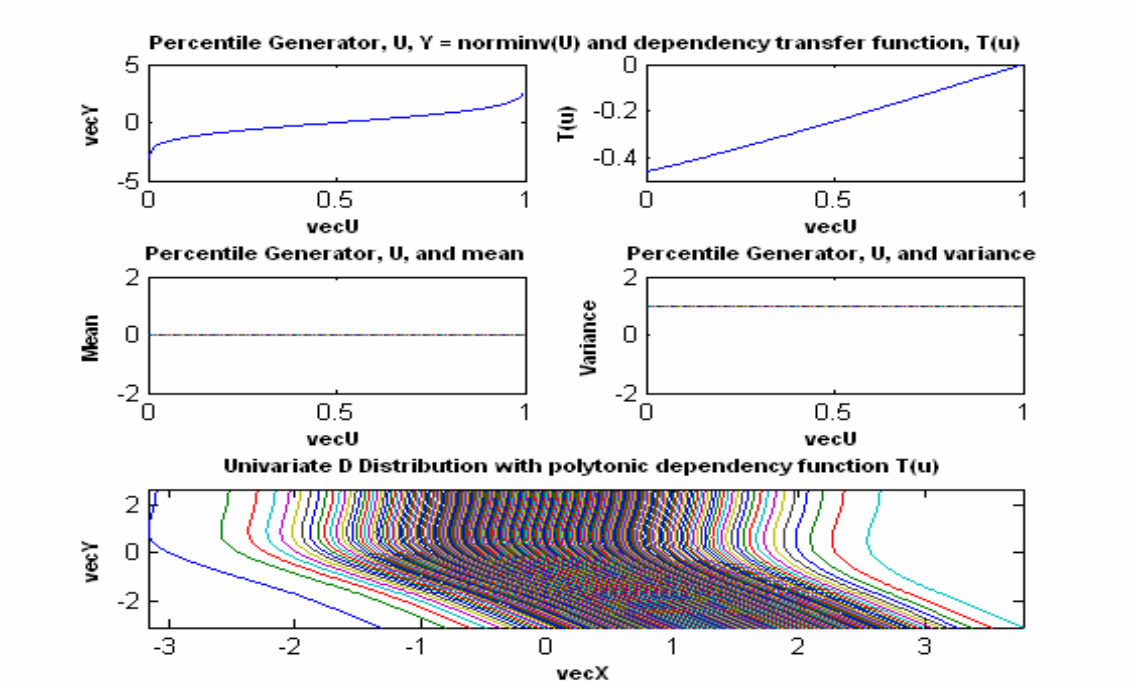


Fig. 4.1. The components of the univariate D distribution described by Eq. 4.3-4.8 in the text when the carrier distribution is Normal and the parameter vector of the dependency transfer function is $B = [-1 \quad 1 \quad 0]$.

4.2.3 Results

To illustrate the role of the transfer function, we show the transfer function and the resulting joint distribution generated from Eq. 4.1 (Fig. 4.2-4.7 below).

If the transfer function generates a constant correlation, then the output will be noise around a straight line with gradient equal to β_1 (Fig. 4.2). If the transfer function graduates from perfect correlation to zero without changing sign, the distribution will become increasingly skew (Fig. 4.3). If the transfer function changes sign, such as illustrated in Fig. 4.4-4.5, the distribution will become skew and resemble a bifurcation pattern. If the dependency transfer function does not change sign but is concave, then the distribution will have tighter dependency in the centre and be more scattered at the tails (Fig. 4.6-4.7).

Fig 4.2 Constant T(u)

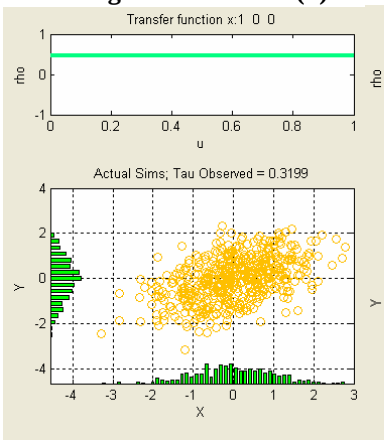


Fig 4.3 Monotonic T(u)

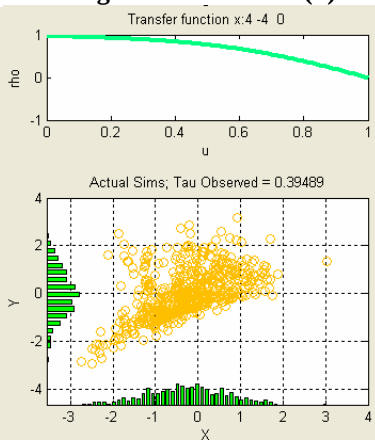


Fig 4.4 Polytonic T(u)

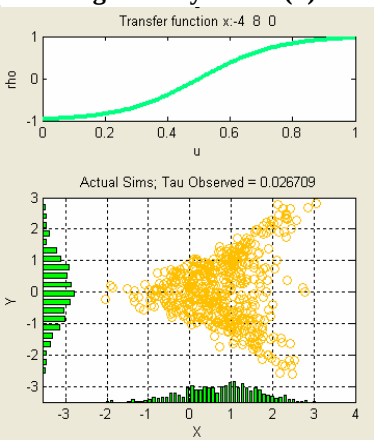


Fig 4.5 Hard limit, discontinuous

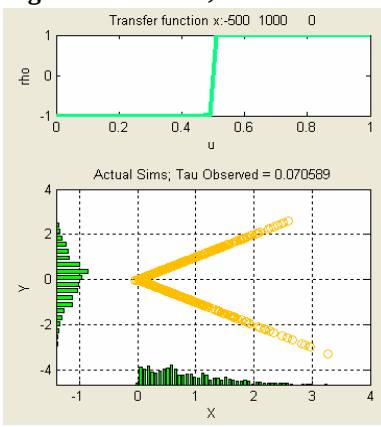


Fig 4.6 Return 1

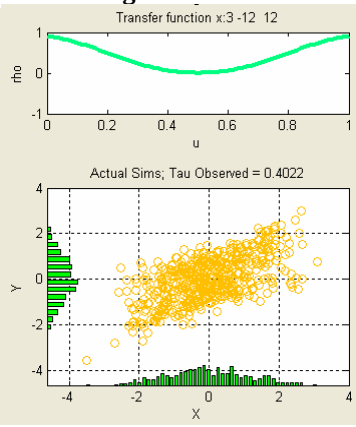
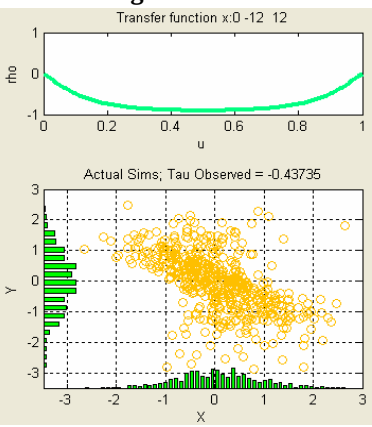


Fig 4.7 Return 2



In each example illustrated in the previous figures, *Y* is normally distributed. However, the transfer parameters in Fig 4.3-4.5 generate asymmetric univariate distributions for variable *X*. In particular, in Fig. 4.5 the distribution of variable *X* is a half normal distribution. This is equivalent to max(*X*, *Y*) when *X* and *Y* are perfectly negatively correlated normal variables. This is another way to arrive at the skew normal distribution described by Azzalini and De Valle (1996).

5. An alternative carrier function

5.1 The Parallelepid distribution

The integration of exponential forms, as involved in the normal distribution, can be complicated. Our next question was whether some other distribution or geometrical construction might be easier to tackle mathematically. If a variable has a conditional uniform distribution, then the integration over the range of the conditioning variable should result in a four sided plane. Depending on the degree of inter-dependency (for simplicity we will be using the term' dependency') between the two variables, their joint scatter plot can vary from a perfect square or rectangle, to a parallelepid of non acute angles, all the way to a straight diagonal line when dependency is perfect (angle of data spread is 0^o). The cosine of the angle of data spread is equal to the dependency between the variables. The sign of the dependency parameter changes the direction of the pattern so that when dependency is negative, the pattern is the mirror image of the one obtained when dependency is positive. Below are examples of when the dependency parameter is zero, 0.4 and -0.4, 0.8 and -0.8 and 1 and -1. In each of these cases the angle of data spread is 90^o, arccos(0.4), arccos(-0.4), arccos(0.8), arccos(-0.8), 0^o and 0^o, respectively.

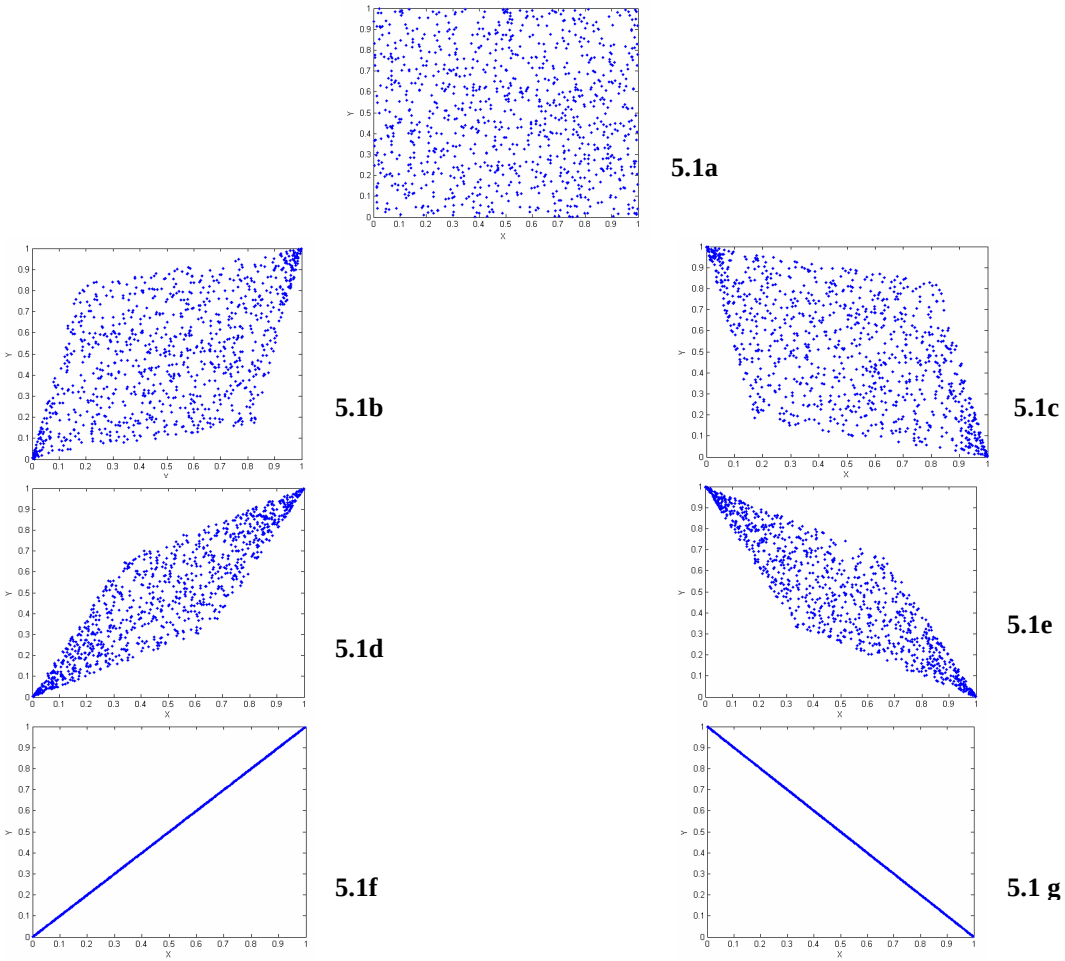


Fig. 5.1. Joint parallelepid distribution when the dependency parameter is 0 (**5.1a**), 0.4 (**5.1b**), -0.4 (**5.1c**), 0.8 (**5.1d**), -0.8 (**5.1e**), 1 (**5.1f**) and -1 (**5.1g**), respectively.

If the general diamond shaped pattern is divided into two halves along the median value of the *X* variable, the two parts are a reversed mirror image of each other and therefore, it suffices to find the functional forms for one half of the diamond shape (the left part). In the case of positive correlation, in order to produce the other half of the pattern, i.e. for values of $x \geq median(X)$, the same functions need to be applied on the converted values $[max(X)-x]$. In the case of negative correlation, all is needed is to convert the *x* values into $[max(X)-x]$ and then proceed as if the correlation was positive. This is so because if we examine the two shapes obtained for the same absolute correlation but of opposite sign along their correlation diagonal (Fig. 5.2), then the two halves above the diagonal and the two halves below the diagonal are mirror images of each other. As shown in the diagram below, part abc on the right (negative correlation) is treated as part *123* on the left (positive correlation) and part *adc* on the right is treated like part 143 on the left, in order to get the functions producing either shape.

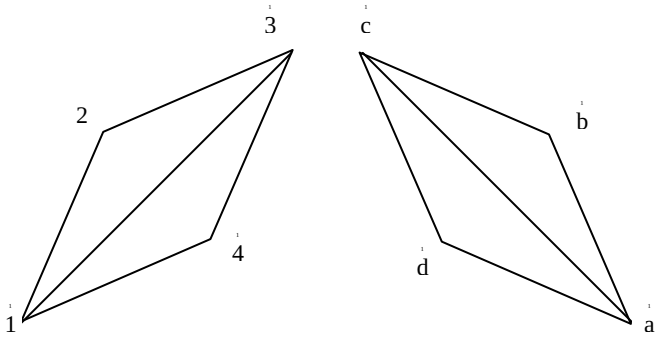


Fig. 5.2. General parallelepid distribution shape when dependency is positive (left) or negative (right).

In the case of the parallelepid distribution being the carrier distribution, we get

$$f(x|y) = \frac{1}{b-a}\Big|U = u, \text{ with } \begin{cases} a = k_2 x \\ b = k_1 x, x \leq x_* \\ b = k_2(x - x_*) + y_*, x > x_* \end{cases} \qquad \text{Eq. 5.1}$$

where

$$x_* = \frac{1}{k_1 + 1},\ y_* = k_1 x_* \text{ and} \qquad \text{Eq. 5.2}$$

$$k_1 = \mathbf{tan}\left\{\frac{\pi}{4} + \frac{[\mathbf{arccos}(T(u))]}{2}\right\}$$

$$k_2 = \mathbf{tan}\left\{\frac{\pi}{4} - \frac{[\mathbf{arccos}(T(u))]}{2}\right\} \qquad \text{Eq. 5.3}$$

and $T(u)$ is the transfer function described in section 4.2.2. After substituting Eq. 5.2 and Eq. 5.3 into 5.1 and using some trigonometric inequalities, Eq. 5.1 becomes

$$f(x|y) = \frac{1}{b-a}\bigg|u = \begin{cases} \dfrac{1}{2\tan\{\arccos[T(u)]\}x}, x \le x_* \\ \dfrac{1}{\tan\{\arccos[T(u)]\}\left\{1-\tan\left\{\dfrac{\arccos[T(u)]}{2}\right\}\right\}}, x > x_* \end{cases} \qquad \text{Eq. 5.4}$$

6. The probability density function: univariate *D* distribution

6.1 Normal carrier function

The probability distribution function is unique to each transfer function and random generating function. If we study the example described in 4.2 (i.e. the carrier distribution is Normal and there is a single percentile generating function *U*) in more detail, we derive the following expression for the p.d.f. of variable *X*|*Y*:

$$f(x|y) = \frac{1}{|\Sigma|^{\frac{1}{2}}(2\pi)^{\frac{d}{2}}} e^{-\frac{1}{2}\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)}\bigg|U = u \qquad \text{Eq. 6.1}$$

6.2 Parallelepid transfer function

Using some trigonometric relationships, Eq. 5.4 can be further simplified and written as

$$f(x|y) = \frac{1}{b-a}\bigg|u = \begin{cases} \dfrac{T(u)}{2\sqrt{[1-T^2(u)]}x}, x \le x_* \\ \dfrac{T(u)}{2\sqrt{[1-T^2(u)]}\left\{1-\dfrac{\sqrt{[1-T^2(u)]}}{1+T(u)}\right\}}, x > x_* \end{cases} \qquad \text{Eq. 6.2}$$

6.3 Difficulties of integration

In order to derive the formula for the *D* distribution, with either the normal or the parallelepid as the carrier distribution function, the formula described by Eq. 6.1 or Eq. 6.2, respectively, needs to be integrated across the range of the *U* variable, i.e. between 0 and 1. However, this is not a simple calculation and it has not been possible to reach a result of a closed form formula for the p.d.f. for either case of carrier function. This is because of the dependency of the correlation coefficient, and hence the variance of variable *X*, on the hidden variable *U*. The integration was attempted using integration by parts, i.e.

$$\int_a^b f_1'(y)\, f_2(y)\, dy = \left[f_1(y)\ f_2(y)\right]_a^b - \int_a^b f_1(y)\ f_2'(y)\, dy \qquad \text{Eq. 6.3}$$

where *f* and *g* are functions of variable *Y* and the prime sign denotes a first order derivative. In the case of the normal distribution, i.e. Eq. 6.1, the two parts representing

$f_1'(y)$ and $f_2(y)$, are $\frac{1}{|\Sigma|^{\frac{1}{2}}}$ and $e^{-\frac{1}{2}\left((x-\mu)'\Sigma^{-1}(x-\mu)\right)}$, respectively. Furthermore, in order to try and simplify the calculation without ignoring the dependency on U, $(1-T^2(u))$ was represented as general function of U, $g(u)$. In the case of the uniform carrier distribution function, i.e. Eq. 6.2, when $x \le x_*$ the two parts representing $f_1'(y)$ and $f_2(y)$ are $\frac{T(u)}{\sqrt{[1-T^2(u)]}}$ and $\frac{1}{2x}$, respectively, while when $x > x_*$ $f_1'(y)$ and $f_2(y)$, are $\frac{T(u)}{2\sqrt{[1-T^2(u)]}}$ and $\frac{1}{\left\{1-\frac{\sqrt{[1-T^2(u)]}}{1+T(u)}\right\}}$, respectively. The purpose of integrating by parts is to try and group terms together into recognisable forms of functions that are easy to integrate. However, in this case the full derivative of either expression is quite complicated because of the dependency on U and not fully present in either Eq. 6.1 or Eq. 6.2.

7. Multivariate *D* Distribution

7.1 Partial correlations

To construct a sample from a multivariate *D* distribution, a set of transfer functions with corresponding β_{ij} parameters are required. Each transfer function leads to a precise dependency structure between $X_1, \ldots, X_n$ given that they are conditional on variable U. In the example presented in *4.2.2*, there is no direct correlation between X and Y because correlation applies indirectly through variable U. The equivalent bivariate case is

$$(X_1, X_2) \sim D\left(Normal, Logistic, \begin{bmatrix} -1 & +1 & 0 \\ -1 & +1 & 0 \end{bmatrix}\right) \qquad \text{Eq. 7.1}$$

Even though X_1 and X_2 are correlated across a range of U, they are not correlated given U, i.e. X_1 and X_2 are correlated indirectly through some values of U. This property means that the correlation between X_1 and X_2 given U is equal to the product of the 'marginal' or partial correlations, i.e.

$$\rho(X_1, X_2 \mid U) = \rho(X_1 \mid U)\,\rho(X_2 \mid U) \qquad \text{Eq. 7.2}$$

This means that the covariance matrix for variables Y [=FunctionInverse(U)], X_1 and X_2 given U is

$$Cov(Y, X_1, X_2 \mid U) = \begin{bmatrix} 1 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 \\ \rho_{1,2}\sigma_1\sigma_2 & 1 & 0 \\ \rho_{1,3}\sigma_1\sigma_3 & 0 & 1 \end{bmatrix} \qquad \text{Eq. 7.3}$$

where the subscripts 1, 2 and 3 refer to Y, X_1 and X_2, respectively.

A worked example of a bivariate *D* distribution has been submitted to the conference web site as an excel sheet.

7.2 Mixture probability density

Since by definition of the *D* distribution E(E(*X*|*Y*)) = E(*X*) = *T*(*U*)*Y*, *U*~U(0 1), the expected value is a scale mixture of normal variables. Because the percentile generating function refers to the whole of the *D* distribution, while the two sets of percentiles refer to each of the marginals for *X* and *Y*, the *D* distribution consists of a set of slices, which can be overlapping (Fig. 7.1), and it is a weighted sum of bivariate distributions (Fig. 7.2) (*X* = X_2 and *Y* = X_1).

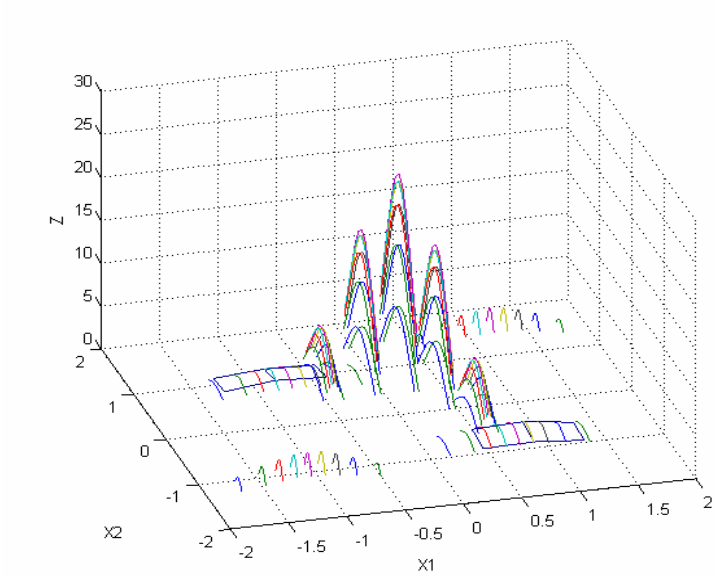


Fig. 7.1. Components of the bivariate *D* distribution when the carrier distribution function is Normal.

Each percentile of the full distribution is a weighted mixture of probability densities, so that the sum of the weights used is unity. Being a mixture probability density, each data point has a degree of membership in the distribution. This is also true for the means E(*X*|*Y*) and variances Var(*X*|*Y*), where the degrees of membership are the scaling weights and their squares, respectively. The standard deviation of the *D* distribution is the square root of the weighted sum of the component variances, i.e. $SD_D = \sqrt{\sum_{j=1}^{n} component\ weight_j^2\ Variance_j}$.

In order to build the p.d.f., first a grid is created using the ranges of the modelled variables, X_1 and X_2. Then we calculate a weighted sum of the p.d.f. value at interpolated pairs of values for X_1 and X_2 from the grid surface. The c.d.f is calculated in exactly the same way by replacing the p.d.f. An example of the bivariate *D* distribution is shown in Fig. 7.2 below.

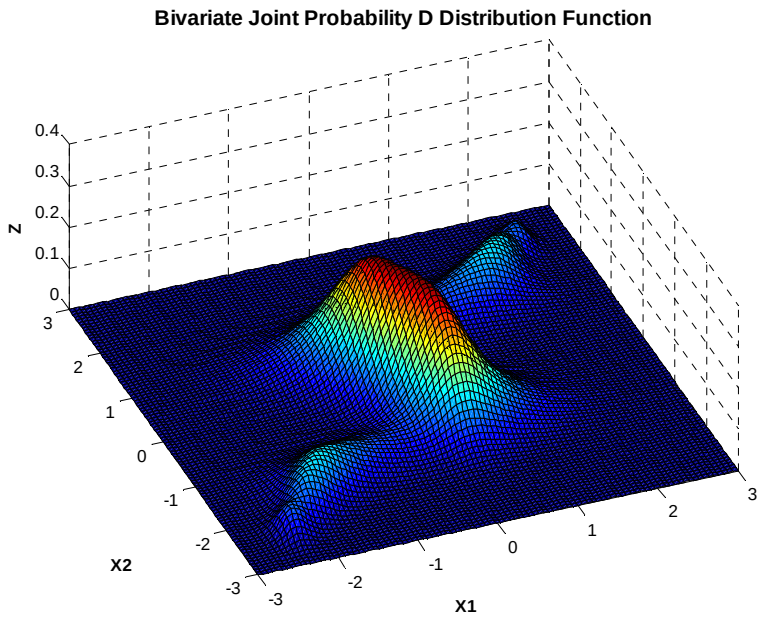


Fig. 7.2. Joint probability distribution function for the *D* distribution as a weighted sum of bivariate distributions when the carrier dependency function is normal and the parameters of the dependency transfer function are $B = \begin{bmatrix} -10 & 80 & -80 \\ -8 & 20 & -20 \end{bmatrix}$. The stochastic counterpart is shown in Fig. 7.7 below.

7.3 Examples: same transfer function

7.3.1 Examples of dependency structures with symmetry and normal carrier function

Below are some examples of the stochastic version of the bivariate *D* distribution for various cases of symmetry and tail dependency when the carrier distribution function is normal and the transfer function is the same for both modelled variables.

If the dependency changes from perfect correlation to zero without changing sign, then the joint distribution becomes increasingly skew (Fig. 7.3-7.4). If neither of the transfer functions changes sign and both are convex, the distribution is tighter towards the centre and more scattered towards the tails (Fig. 7.5).

Fig. 7.3 **Fig. 7.4** **Fig. 7.5** **Fig. 7.6**

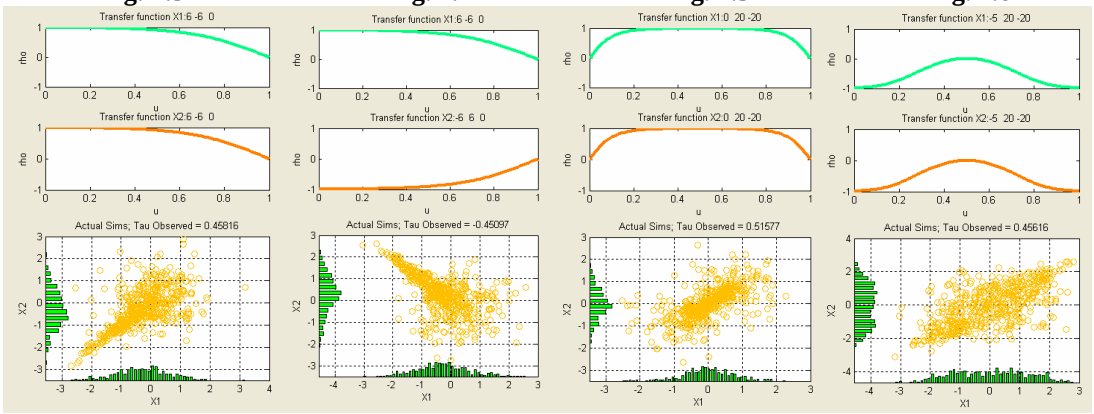
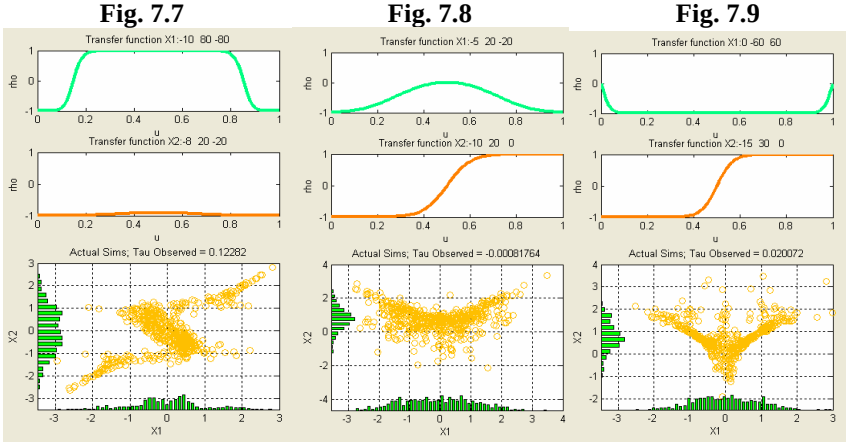


Fig. 7.7 **Fig. 7.8** **Fig. 7.9**



If there are multiple changes in sign in the transfer function, then the distribution will be characterised by turning points in correlation (Fig. 7.7-7.9).

The marginal distributions are generally pretty close to normal. However, when the distribution is highly skewed (Fig. 7.3-7.4) or has a pronounced change of correlation the marginal distributions tend to become more asymmetric with longer tails (Fig. 7.7-7.9).

7.3.2 Examples with Parallelepid carrier function

Below are some examples of the stochastic version of the bivariate *D* distribution for various cases of symmetry and tail dependency when the carrier distribution function is parallelepid, i.e.

$$(X,Y) \sim D\left(Parallelepid, Logistic, \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix}\right) \qquad \text{Eq. 7.4}$$

with a logistic dependency transfer function with parameters $B_1 = [\beta_{11}\ \beta_{12}\ \beta_{13}]$ for variable X_1 and parameters $B_2 = [\beta_{21}\ \beta_{22}\ \beta_{23}]$ for variable X_2.

Fig. 7.10 **Fig. 7.11** **Fig. 7.12** **Fig. 7.13**

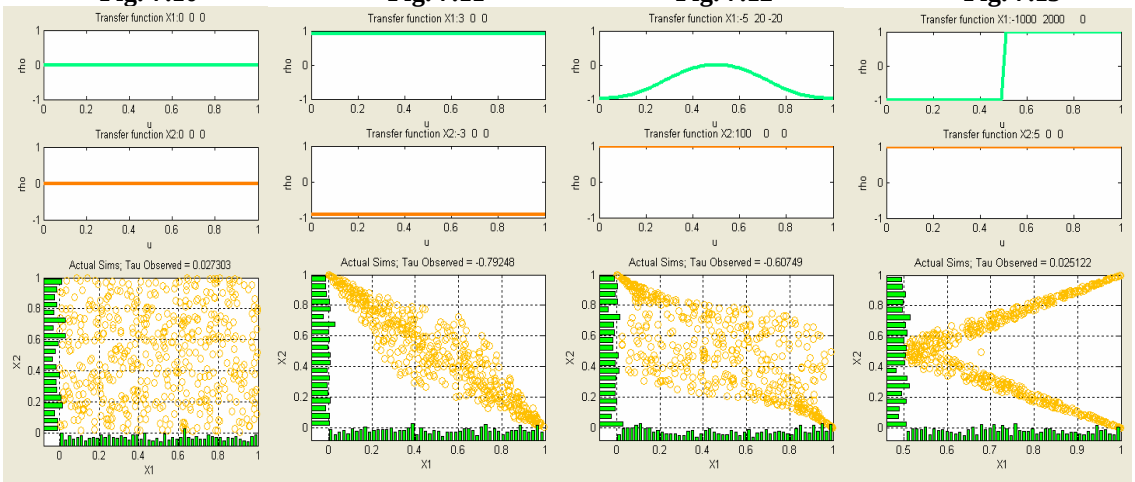
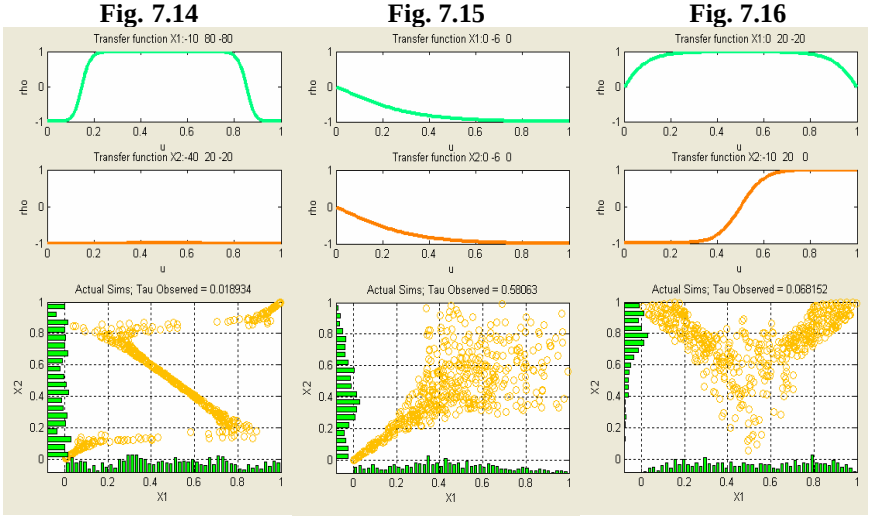


Fig. 7.14 **Fig. 7.15** **Fig. 7.16**



When there is no dependency between the variables (Fig. 7.10) then the distribution is perfectly uniform. When there is constant dependency other than zero, the resulting distribution is of a diamond shape with spread depending on the absolute value of dependency, i.e. the higher the absolute dependency, the more the distribution tends to one of the diagonals. The direction of the distribution depends on whether the transfer functions are of the same or opposite signs (Fig. 7.11). If one of the transfer functions is concave without changing sign, while the other remains constant, the distribution will be tight at the tails with noise in the centre (Fig. 7.12). If one of the transfer functions is constant and the other one changes sign abruptly the distribution will resemble a bifurcation pattern, while if the variable transfer function changes sign twice, the distribution will be of the 'Zoro' shape with more concentration at the diagonal and the end points (Fig. 7.13-7.14, respectively). When both transfer functions change from perfect correlation to zero, the distribution will become increasingly skew (Fig. 7.15). If one of the transfer functions gradually changes sign and the other one is concave and does not change sign the distribution will have a clear turning point and change of correlation (Fig. 7.16).

The marginal distributions are pretty uniform when the corresponding dependency transfer functions are constant (Fig. 7.10-7.11, variable X_2 in Fig. 7.12-7.14), while becoming more skew as the transfer functions deviate from a constant value with or without a change of sign (Fig. 7.15-7.16).

Comparing the distributions derived from using a normal dependency transfer function (Fig. 7.3-7.9) and a parallelepid dependency transfer function (Fig. 7.10-7.16), it is evident that the spread of the data differs even when looking at virtually the same distribution patterns (maybe with the exception of Fig. 7.15).

Note that in the distribution there is an alternation of parts of data with perfect correlation (tight clusters of points) and parts of data with greater variability (scattered groups of points). This indicates a distribution able to model processes with both deterministic and stochastic responses. This must have application in wider industry.

7.3.3 Examples wth Student-t carrier function

Below are examples of the stochastic version of the bivariate *D* distribution for two cases where the carrier distribution function is the Student's-t with stretch parameter, $\nu = 4$, and stretch parameter, $\nu = 1$, in which case the resulting carrier distribution is the Cauchy, i.e.

$$(X,Y) \sim D\left(Student's-t, Logistic, \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \right) \qquad \text{Eq. 7.5}$$

with a logistic dependency transfer function with parameters $B_1 = [\beta_{11}\ \beta_{12}\ \beta_{13}]$ for variable X_1 and parameters $B_2 = [\beta_{21}\ \beta_{22}\ \beta_{23}]$ for variable X_2.

The t-distribution function, provides long tailed results. The resulting distribution resembles a cluster of points, which becomes tighter as the stretch parameter, ν, of the t-distribution carrier function reduces. However when it comes to adjusting the shape of the joint distribution, the t-distribution does not offer any obvious advantages over the use of the ordinary normal distribution or the parallelepid distribution as carrier functions within a copula, where the stretching can be expressed as a long tailed univariate distribution.

Fig. 7.17

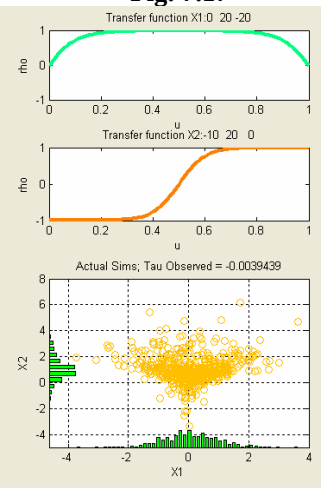
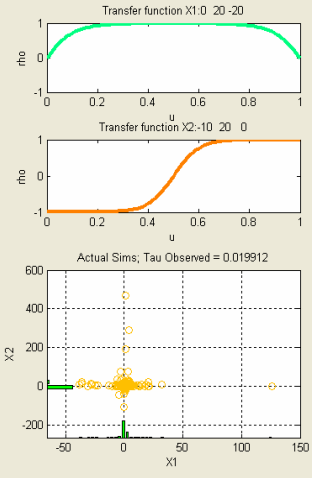


Fig. 7.18



8. Discussion of dependency structures

8.1 Murphy's Law

Major Edward Murphy, an engineer in the U.S. Air Force, was testing the effects of G-force on human beings in the late 1940s with extra diligence against risk in this dangerous field. He had a great understanding of links, dependency and defensive design. To apply Murphy's law to problem solving, every possible outcome, especially anything that can go wrong or against our expectations, has to be considered and engineered against. He law is often misquoted as 'what can go wrong will go wrong'.

We define a Murphy dependency structure as one where dependency increases in downside events, such as Fig. 7.3 or 7.15. Our motivation for the creation of the *D* distribution was the ability to be able to explain as many possible forms of dependency as

possible, which is something that portfolio design can benefit from. In the financial world, portfolio performance can be very different during normal (or more ordinary) situations as opposed to situations of crisis and extreme events. Hence, it is to be expected that various financial behavioural patterns will be better described by a flexible distribution, which is in fact a mixture of probability density functions, as is the case of the *D* distribution.

8.2 Tail independency/dependency

While it is possible to have returns of two assets behaving like the data points on any of the figures shown in section 7.3, e.g. Fig. 7.3, a lot of the modelling on joint asset movements is based on the assumption of normal behaviour and hence, the heavy use of the Gaussian distribution. However, in the bivariate Gaussian distribution correlation approaches zero as we asymptotically move towards the tails and in the case of the data shown in Fig. 7.3 the Gaussian distribution would be a poor choice for explaining their behaviour and dependency structure. The latter is not only variable but it also varies in an unequal and asymmetric way towards either of the tails: on one end (the bottom left corner) there is a tight relationship between the data, i.e. a large degree of dependency, while on the top right corner, where the data points are much more scattered, the degree of dependency is very low and it would not be wrong to say that at the high end of the scale for either of the two assets there is a great deal of independency. The *D* distribution, as is shown in this paper, is able to capture the dependency structure between data even when it changes unevenly and in opposite directions at different low-high data value combinations.

8.3 'Zoro' dependency

In practice, a 'Zoro'-type dependency structure occurs when under normal (ordinary) circumstances the day to day dependency is strong but under extreme circumstances dependency is reversed, Fig. 7.7 or Fig. 7.14.

8.4 'Banana' dependency

A regulator, like the FSA, may require negative dependency between equities and bonds when equities are on an upward trend and positive dependency when equities are on a downward trend for prudence Fig. 7.8. There is some evidence that this has happened over recent time periods. Using the *D* distribution it is possible to examine whether there is banana dependency or whether changes in dependency are spurious calculations on quadrants of a joint distribution.

8.5 'Stag' dependency

In practice, a 'Stag'-type dependency structure occurs when non-extreme upside and downside events are co-dependent, while middle range and extreme events are independent.

9. *D* Copula

The *D* copula models the joint percentiles between variables. It is calculated by considering the joint cumulative distribution function, c.d.f., over a grid formed from two

variables, say X_1 and X_2. A c.d.f. takes as input values of variables, such as X_1 = 5th percentile, X_2 = 50th percentile and gives as a result the cumulative probability.

To build the *D* copula we calculated a weighted sum of joint normal c.d.f.'s for a range of *U* values in equal increments from 0 to 1. For each value of *U* we have a c.d.f. joint distribution for X_1 and X_2. The c.d.f. was calculated on model points on the X_1, X_2 grid ranging from -5 to +5. Expressing X_1 and X_2 as percentiles and using interpolation to convert the grid into linear percentile units leads to a copula with clear straight edges, which characterise a copula. Taking second order differences of the copula surface, i.e. along both X_1 and X_2 axes, gives a density function for percentiles, which is the copula p.d.f. Every row and column on the grid sum up to the same constant, and the total area under the curve is 1.

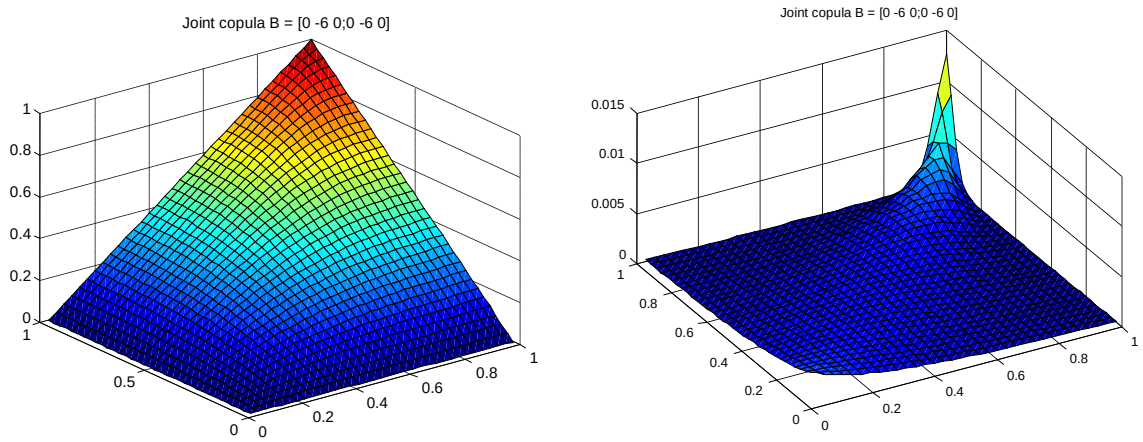


Fig. 9.1 Copula surface of the bivariate *D* distribution when the carrier function is normal and the parameters of the dependency transfer function are $B = \begin{bmatrix} 0 & -6 & 0 \\ 0 & -6 & 0 \end{bmatrix}$

The *D* copula is a useful tool because the modeller can parameterise a dependency structure for any data set using percentiles. Similarly, samples of the copula can be transformed into any univariate distribution.

10. Ongoing research

10.1 Parameter estimation

10.1.1 Maximum likelihood

At the time of writing we had not derived an expression for the p.d.f. of the *D* distribution. So it was not possible to use the method of maximum likelihood for the estimation of model parameters analytically. This remains an open problem.

However, an empirical likelihood maximising function, *L*, can be obtained by using the copula p.d.f. surface from Fig. 9.1. For every pair of empirical observations and a given parameter matrix *B* calculate the probability of occurrence, call it L_k. The sum of all the L_k's is the likelihood *L*. The best fit copula is the set of parameters that maximises *L* for a given set of observed joint percentiles.

10.1.2 Confidence intervals

Once L has been optimised, the confidence interval for the estimated parameters will be calculated using the Cramér-Rao inequality, a lower bound for the variance of an unbiased statistical estimator, say θ, based on the Fisher information $I(\theta)$.

$$\mathbf{var}\left(\hat{\theta}\right) \geq \frac{1}{\mathrm{I}(\theta)} \qquad \text{Eq. 10.1}$$

In statistics, score (or score function), say S, is defined as the partial derivative of the (natural) logarithm of the likelihood function in terms of the function parameters. In our case the function parameters are the elements of the parameter matrix B.

$$S = \frac{\partial}{\partial \beta_{ij}} \mathbf{ln}\left[L\left(\beta_{ij}; x_1, x_2 \ldots\right)\right] = \frac{1}{L\left(\beta_{ij}; x_1, x_2 \ldots\right)} \frac{\partial\left[L\left(\beta_{ij}; x_1, x_2 \ldots\right)\right]}{\partial \beta_{ij}} \qquad \text{Eq. 10.2}$$

The Fisher information is the variance of the score. Because the expected value of the score is zero, its variance is simply the expected value of the squared score. The greater the variance of the score is, the lower the variance in a parameter.

10.1.3 Objective Function

Because the p.d.f. surface is calculated over a percentile grid, computational constraints place a limit on the grid size generated by X_1 and X_2 and the degree of resolution of U, the percentile generating variable that can be used to estimate the likelihood function. In our analysis we achieved a better parameter fit to the empirical copula surface by developing a hybrid function that includes the likelihood maximising function and measures of surface fit. This investigation is still ongoing as we have achieved limited stability in the derivation of optimised parameters. This difficulty was expected because of the lack of mathematical formulas for the p.d.f. as mentioned in 10.1 and therefore the lack of analytical solutions for the distribution parameters.

The objective function is a surface, and where the local optimum on the surface is very pronounced we have stronger confidence in the parameters true value. On the other hand very gentle changes in the surface for a change in the parameter lead to less confidence in the precise positioning of the true underlying parameter value. However, applying Eq. 10.2 to our objective function leads to a lower bound for the variance of each parameter β_{ij} and hence a corresponding approximate confidence interval.

10.1.4 Fitting a simple symmetrical model

For a symmetrical bivariate D distribution, where dependency increases or decreases symmetrically at the tails, the β_{ij} parameters can be reduced to three variables as shown below.

$$B = \begin{bmatrix} \alpha & -\alpha\times\beta & \alpha\times\beta \\ \alpha\times\rho & -\alpha\times\beta\times\rho & \alpha\times\beta\times\rho \end{bmatrix} \begin{cases} \alpha \geq 0 \\ \beta \geq 0 \\ \rho = +1,-1 \end{cases}, \qquad \text{Eq. 10.3}$$

Parameter ρ can be established by calculating a rank correlation statistic such as Kendall's τ. If τ is positive, ρ is +1, while if τ is negative, ρ is -1. This then leaves only two parameters, the strength of correlation, α, and the inflection between the middle and the tails, β, to be fitted. This reduced model is easier to optimise.

10.2 Measures of dependency

In this paper we have demonstrated the *D* distribution in its simplest form, i.e. we have always used one *U* variable, assumed no correlation between X_1 and X_2 (or *X* and *Y*) and have used the same transfer function in every case. The advantage of introducing a dependency transfer function is the ability to control polytonal dependency.

It is evident from the results presented earlier in this document that the output of the transfer function, which is affected by the choice of parameters, is directly linked to tail dependency. If the transfer function generates dependency values close to +1 or -1 for variables when U is in the bottom 5th percentile or top 95th percentile then there will be strong tail dependency.

Fitting a *D* distribution and subsequent examination of the transfer function parameters appears to offer a new way to calibrate tail dependency in a data set.

10.3 Conditioning

Conditioning is a fundamental building block to the *D* distribution. However the next step is to generate a multivariate *D* distribution conditioned on $X = x$. Forcing *X* to take a particular value leads to a distribution of *U* at that point. Stochastic simulation from the distribution of *U* leads to a conditioned multivariate distribution.

10.4 Deterministic and Probability models

The *D* copula has the ability to model deterministic and probabilistic events within the same structure. A deterministic range is where the joint variation goes to zero, driven by high partial correlations in a range of *U*. We believe this has a lot of applications in the financial markets.

11. Conclusions

At the beginning of this paper we set out to challenge a few long standing statistical concepts. We have demonstrated that a sum of normal distributions where there is dynamic correlation is far from normal and this property allows us to create a new class of distribution functions driven by dependency, the *D* distribution. We have a distribution where there is strong asymmetry of beta and portfolio hedging ratio. We have developed a copula where one copula form, the *D* copula, can bring together vastly different dependency structures. We are confident that with the development of the *D* distribution we have managed to shake these concepts and offer a more realistic insight

into statistical modelling and especially, the modelling of dependency structures in the financial markets.

The most important issue here is dependency and the use of the correlation coefficient to describe this. The standard correlation coefficient is fine as long as any multivariate event under consideration is static, i.e. Variance and correlation between variables is constant.

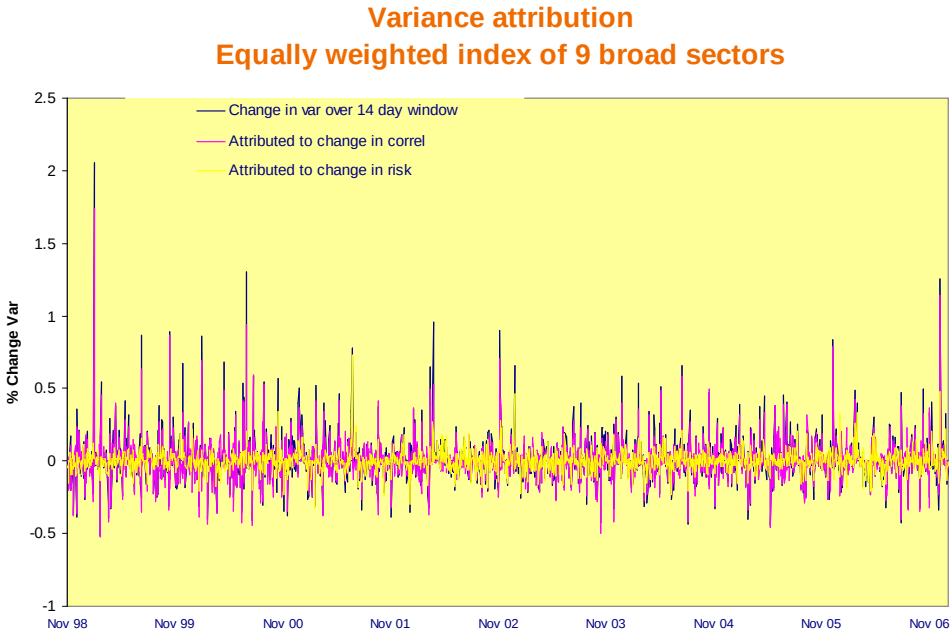


Fig. 11.1. An attribution of variance on an equally weighted index of broad UK FTSE sector indices across a 14 day window. The calculation separates out change due to dynamic correlation between sectors and change due to dynamic variance of the sectors themselves.

Fig 11.1 supports the argument for dynamic correlation. The chart attributes movements at an index level due to changes in dynamic risk and dynamic correlation over a 14 day window in underlying constituent indices. The index was built of equally weighted UK sectors. Around 75% of the movements in the index can be attributed to dynamic correlation in underlying constituents and only 25% due to dynamic variance. Therefore, it is important that dynamic dependency measures are employed to model and understand risk. The use of the normal distributions or sums of various normal distributions can at best be a rough approximation.

Correlation will work for a whole distribution on average. Therefore, it can easily misinterpret and/or neglect data points that lie beyond the average bulk of the data. Likewise, beta and the hedging ratio will only apply correctly if the average market conditions are true but will be inadequate and false otherwise. That is because beta is a function of the correlation coefficient between variables and is defined as the ratio of the standard deviations of the portfolio and the market, scaled by their correlation coefficient, i.e.

$$Beta_p = \rho(\text{portfolio, market})\frac{\sigma_p}{\sigma_m} \qquad \text{Eq. 11.1}$$

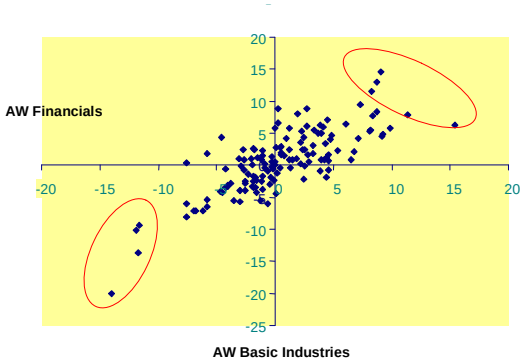


Fig. 11.2. Example of a joint distribution between AW Financials and AW Basic Industries. The blue points are monthly returns January 1994 to March 2005. The red circles highlight the fact that correlation can be very strong at the joint tails: over 90% positive in the joint downside tail (bottom left corner) and negative (about -20%) on the joint upside tail (top right corner). This is a Murphy dependence structure.

In Fig. 11.2 above, it would seem fine to fit a straight line though the data points, which would imply a constant positive hedging coefficient. However, as shown on the bottom left hand quadrant of the chart, there is strong correlation at the joint downside tails. Any hedging based on a constant positive correlation would be valid for the market conditions underlying the central part of this distribution, but if conditions became extreme, the hedge will fail. It is most likely that we want the hedge to work in an extreme upside or downside tail event, and it is during these tail events that the hedge will break down. This is an example of Murphy's law. Murphy structures are common place in financial time series.

The *D* copula is flexible. Altering the parameters involved generates a whole set of completely different dependency surfaces, rather than simple variations of the same oval correlation shape. As long as the proper dependency structure of the modelled variables has been established, the *D* copula can be used as a multivariate copula with very different dependency structures between pairs of variables. The authors are not aware of a technique able to model different positive and negative skew structures in a multivariate copula.

About the authors

Martyn Dorey FIA is head of the P-Solve quantitative solutions team, and an experienced developer of financial models. He is an advisor for pension schemes, fund managers and external consultancies on risk modelling, risk budgeting, and defensive asset allocation.

martyn.dorey@p-solve.com

Dr Konstantina Papastamati has been conducting scientific research since 1993 in the fields of maritime and agricultural biology. Konstantina is a regular contributor to journals on mathematical modelling. Konstantina is providing analytical support and developing investment models to assist in the ongoing development and enhancement of the P-Solve defensive asset allocation and governance business.

konstantina.papastamati@p-solve.com

Bibliography

1. **Arnold, B. C. and Beaver, R. J. 2002.** Skewed multivariate models related to hidden truncation and/or selective reporting. *Test*, **11**, 7-54.
2. **Azzalini, A. and Dalla-Valle, A. 1996.** The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.
3. **Barndorff-Nielsen, O. 1998.** Processes of normal inverse Gaussian type. *Finance and Stochastics*, **2**, 41-68.
4. **Carlin, S. and Smith, A. 2003.** The Importance of Being Normal. *Prepared for*: the Institute and Faculty of Actuaries' Finance and Investment Conference, Edinburgh, pp. 48.
5. **Embrechts, P., McNeil, A. and Straumann, D. 2002.** Correlation and dependence in risk management: properties and pitfalls. *In*: Risk Management, Value at Risk and Beyond. M. A. H. Dempster (Ed.), Cambridge University Press, Cambridge, pp. 176-223.
6. **Embrechts, P., McNeil, A. and Straumann, D. 1999.** Correlation: pitfalls and alternatives. A short non-technical article. *RISK* magazine, May, 69-71.
7. **Frees, E. W. and Valdez, E. A. 1997.** Understanding relationships using copulas. *North American Actuarial Journal*, 2 (1), 1-25.
8. **Genton, M. G., ed., 2004.** Skew elliptical distributions and their applications: a journey beyond normality. Chapman and hall/CRC, Boca Raton.
9. **Hagan, M. T., Demuth, H. B., Beale, M. H., 1996.** Neural Network design. PWS Publishing Company, ISBN: 0-9717321-0-8.
10. **Holland, P. W. and Wang, Y. J. 1987.** Dependence function for continuous bivariate densities. *Communication in Statistics- Theory and Methods*, **16**, 863-876.
11. **Nadarajah, S. and Kotz, S. 2006.** On the product XY for the elliptically symmetric Pearson type VII distribution. *Mathematical proceedings of the Royal Irish Academy*, **106A (2)**, 149-162.
12. **Olano, M. and North, M. 1997.** Normal Distribution Mapping. *UNC Chapel Hill Computer Science Technical Report 97-041*, Department of Computer Science, University of North Carolina, Chapel Hill, North Carolina.
13. **Prevéy, P. S. 1986.** The use of Pearson VII distribution functions in X-ray diffraction residual stress measurement. *Advances in X-Ray Analysis*, **29**, 103-111.
14. **Tan, F. and Peng, H. 2005.** The Slash and Skew-slash Student t Distributions. Preprint, *Elsevier Science*. May 2005