MODELLING DEPENDENCY BETWEEN DIFFERENT LINES OF BUSINESS WITH COPULAS

LI, Jackie BCom, MApplFin, FIAA

Centre for Actuarial Studies, University of Melbourne, Australia

Email: k.li3@pgrad.unimelb.edu.au

ABSTRACT

In this paper we select various practically tractable copulas and demonstrate their use in practical circumstances under the current Australian regulatory framework. The copulas under discussion include Gaussian copula, *t* copula, Cook-Johnson copula, and a few Archimedean copulas. We also examine the feasibility of the simulation procedures of the copulas in practice. We set up two hypothetical examples, which are based on real life claims features. In particular, we propose the incorporation of a copula into the traditional collective risk model. We demonstrate that copulas are a set of flexible mathematical tools for modelling dependency, in which the extreme percentiles of the aggregate portfolio value vary considerably for different choices of model copulas. We also show that some measures of association have better properties than the correlation coefficient, which is a common measure in practice.

Relevant legislation and preliminary information are introduced in Sections 1 and 2. The general definition of a copula is set out in Section 3. Some elementary measures of association between pairwise random variables, for computing the parameters of the copulas, are described in Section 4. Different types of copula and simulation techniques, with some examples for illustration, are introduced in Section 5. Examples of practical applications on assessing the uncertainty of some general insurance liabilities are provided in Sections 6 and 7. Final discussion is set forth in Section 8.

Keyword: Copula, dependency, multivariate, simulation, line of business, liability, uncertainty, percentile, margin

1. INTRODUCTION

Australian Prudential Standard GPS 210 requires an allowance for diversification benefits, which arise from the underlying dependency structures between different lines of business for an insurer. The liabilities of different lines can be conceived as multivariate random variables. Furthermore, the Australian solvency benchmark for probability of ruin is 0.5% on a one-year time horizon. Appropriate dependency modelling is hence essential for determining the mandatory solvency level of reserves.

Copulas have been known and studied for more than 45 years. These mathematical tools have remained largely in theoretical development and only until recently experimental applications have been tested in some practical financial areas. An example is the VaR (Value at Risk) calculation by Micocci and Masala (2004), in which the multinormal assumption is dropped and the dependency between the extreme returns of different stocks in a portfolio is more properly accommodated. The application in general insurance is still in its infancy, despite the general familiarity with DFA (Dynamic Financial Analysis) in dealing with risk concentration. As the types of copula are many and varied, this paper identifies expedient copulas and applicable simulation techniques for common general insurance practice under the current legislative environment in Australia.

We broadly differentiate two sources of dependency between the liabilities of different lines of business (refer to Taylor and McGuire (2005) for a similar discussion):

Inherent dependency – This component is inherent in the claims run-off process. Driven by similar factors such as inflation, interest rate, exchange rate, economic cycles, and weather patterns, the number of unreported claims or the claim sizes of different lines may move in the same direction to some extent. In practice, the volume of the past claims data is usually insufficient for quantifying this dependency with satisfactory statistical confidence. Due judgement is integral.

Other dependency – This component relates to the practices implemented by the insurer across its various lines of business. A few examples are: similar reserving methods and judgement are applied to different lines; the claims department adopts a particular mechanism of setting up the case estimates for all lines; the claims management procedures are similar for some lines; and there is a consistent data system in the company. This dependency cannot be readily quantified and again appropriate judgement is necessary.

All the calculations were carried out through Excel spreadsheets with VBA (Visual Basic for Applications) coding and the software Mathematica. Some details regarding the use of Excel functions and Mathematica coding are provided in the appendices.

The main references of the contents of this chapter are Joe (1997), Nelsen (1999), and Embrechts et al (2001), which provide elaborate descriptions and proofs about the dependency concepts.

2. NOTATION, TERMINOLOGY, AND PRELIMINARIES

2.1 Notation

For the notation used in this paper, the superscripts generally refer to the lines of business concerned, e.g. $\rho^{X,Y}$ represents Pearson's correlation coefficient between X and Y, which are say the total outstanding claims liabilities of two different lines. The bracket in the subscript also indicates a particular line of business, e.g. $X_{(i)}$ may represent the total outstanding claims liability of line *i*.

2.2 Terminology

The terms 'marginal univariate probability distribution' and 'marginal' are used interchangeably. The terms 'dependent', 'related', 'associated', and 'multivariate' have the same meaning in general. The increasing dimensions are described as 'univariate', 'bivariate' (or 'pairwise'), and 'multivariate' (or 'm-variate' for *m* dimensions). The abbreviations 'pdf' and 'cdf' stand for 'probability density function' and 'cumulative distribution function' respectively. 'Multinormal' is the same as 'multivariate normal', so too 'binormal' and 'bivariate normal'.

2.3 Preliminary Information

Suppose X and Y are two associated random variables that are continuous. Their joint bivariate cdf is $F_{X,Y}(x,y) = \Pr(X \le x, Y \le y)$. Their marginal univariate cdfs are $F_X(x) = \Pr(X \le x)$ and $F_Y(y) = \Pr(Y \le y)$ respectively, in which $F_X(x) = \lim_{y\to\infty} F_{X,Y}(x,y)$, $F_Y(y) = \lim_{x\to\infty} F_{X,Y}(x,y)$, $0 \le F_X(x) \le 1$, and $0 \le F_Y(y) \le 1$. (Only strictly increasing univariate cdfs are considered in this chapter.) These properties can be extended similarly to the multivariate case with more than two dimensions. In addition, a sample of X can be generated by simulating a sample of $U \sim \mathrm{U}(0,1)$, where $\mathrm{U}(0,1)$ is standard uniform, and then using the inverse $F_X^{-1}(U)$.

3. DEFINITION OF COPULA

A copula is a mathematical function that joins univariate probability distributions to form a multivariate probability distribution. Sklar (1959) proved that for a multivariate distribution of m dimensions with continuous marginals, there exists a unique copula function C such that:

$$F_{X_{(1)},X_{(2)},\ldots,X_{(m)}}\left(x_{(1)},x_{(2)},\ldots,x_{(m)}\right)=C\left(F_{X_{(1)}}\left(x_{(1)}\right),F_{X_{(2)}}\left(x_{(2)}\right),\ldots,F_{X_{(m)}}\left(x_{(m)}\right)\right),$$

where $X_{(i)}$'s are related random variables that are continuous, $F_{X_{(1)},X_{(2)},\ldots,X_{(m)}}$ is the joint multivariate cdf, $F_{X_{(i)}}$'s are the marginal univariate cdfs, and C is the unique copula function.

From Sklar's theorem above, for every set of continuous multivariate random variables with a particular dependency structure, there is a unique copula function that links the marginal univariate cdfs to form the joint multivariate cdf. This feature allows separate consideration between selecting the marginals and choosing the dependency structure, and so offers tremendous flexibility in modelling multivariate random variables. This flexibility contrasts with the traditional use of the multinormal distribution, in which the dependency structure is restricted as the linear correlation matrix and the marginals as normally distributed.

There is a host of types of copula. They describe both the shape and strength of the relationships between dependent random variables. They possess varying characteristics, such as different tail dependence (refer to Subsection 4.4), and positive or negative association. The choice of copula can be tailored to a particular situation, e.g. if two lines of business are not related except for the right tails of their underlying liability distributions (say, for catastrophic events), some copulas with tail dependence can be used to model these two lines properly, while simple linear correlation cannot capture this feature.

4. MEASURES OF ASSOCIATION

This section sets out several fundamental measures of association between pairwise random variables. These measures are required to compute certain parameters of the copulas discussed in the next section. For the m-variate case, a $m \times m$ matrix of pairwise measures can be constructed unless otherwise specified. Suppose X and Y are two associated random variables that are continuous, $u = F_X(x)$, $v = F_Y(y)$, and (X_i, Y_i) represent the i[th] pair of (independent) observations of (X, Y). There are a total of n pairs of observations. Let C be the copula between X and Y. Suppose C is also the copula between two related random variables A and B, in which $A \sim \mathrm{U}(0,1)$ and $B \sim \mathrm{U}(0,1)$.

4.1 Kendall's Tau (τ)

According to Nelsen (1999), Kendall's Tau between X and Y is defined as:

$$\tau^{X,Y} = \Pr\left(\left(X_1 - X_2\right)\left(Y_1 - Y_2\right) > 0\right) - \Pr\left(\left(X_1 - X_2\right)\left(Y_1 - Y_2\right) < 0\right) = 4\mathrm{E}\left(C\left(A, B\right)\right) - 1 \ ,$$

which is estimated by $\sum_{i<j} \mathrm{sign}\left(\left(x_i - x_j\right)\left(y_i - y_j\right)\right) \Big/ \binom{n}{2}$.

Nelsen (1999) states that τ $(-1 \le \tau \le 1)$ is invariant under strictly increasing linear or non-linear transformations of X and Y, i.e. $\tau^{f(X),g(Y)} = \tau^{X,Y}$, where f and g are strictly increasing functions. As such, this measure is unaffected by changing the scale of the marginals. The overall association between X and Y is positive if $\tau > 0$ and negative if $\tau < 0$.

Nelsen (1999) also states that (X_i, Y_i) and (X_j, Y_j) (for $i \neq j$) are described as concordant if $X_i < X_j$ when $Y_i < Y_j$ (or $X_i > X_j$ when $Y_i > Y_j$) and as discordant if $X_i < X_j$ when $Y_i > Y_j$ (or $X_i > X_j$ when $Y_i < Y_j$). Thus, τ is equal to the probability of concordance minus the probability of discordance for (X_i, Y_i) and (X_j, Y_j).

Kendall and Gibbons (1990) state that when the number of pairs of observations exceeds 30 (i.e. n > 30), the hypothesis $H_0 : \tau = 0$ can be tested by using the test statistic $3\hat{\tau}\sqrt{n(n-1)/(2(2n+5))} \sim N(0,1)$, where $\hat{\tau}$ is the estimator of τ (a two-sided test).

4.2 Spearman's Rho (ρ_s)

According to Nelsen (1999), Spearman's Rho between X and Y is defined as:

$$\rho_s^{X,Y} = 3\left(\Pr\left((X_1 - X_2)(Y_1 - Y_3) > 0\right) - \Pr\left((X_1 - X_2)(Y_1 - Y_3) < 0\right)\right) = \frac{\mathrm{E}(AB) - \mathrm{E}(A)\mathrm{E}(B)}{\sqrt{\mathrm{Var}(A)\mathrm{Var}(B)}} ,$$

which is estimated by Pearson's correlation coefficient (refer to Subsection 4.3) between $\mathrm{Rank}(x_i)$ and $\mathrm{Rank}(y_i)$.

Nelsen (1999) states that ρ_s ($-1 \le \rho_s \le 1$) is invariant under strictly increasing linear or non-linear transformations of X and Y, i.e. $\rho_s^{f(X),g(Y)} = \rho_s^{X,Y}$, where f and g are strictly increasing functions. This measure is thus unaffected by changing the scale of the marginals. The overall association between X and Y is positive if $\rho_s > 0$ and negative if $\rho_s < 0$.

In effect, ρ_s is equal to the probability of concordance minus the probability of discordance for (X_i, Y_i) and (X_j, Y_k) (for unequal i, j, and *k*). Furthermore, as stated in Nelsen (1999), some relationships between τ and ρ_s are $-1 \le 3\tau - 2\rho_s \le 1$, $(1+\tau)^2 \le 2(1+\rho_s)$, and $(1-\tau)^2 \le 2(1-\rho_s)$. Kendall and Gibbons (1990) mention that in practice, when neither $|\tau|$ nor $|\rho_s|$ is too close to unity, ρ_s is about 50% greater than τ in magnitude.

Kendall and Gibbons (1990) also note that for $n > 35$, the hypothesis $\mathrm{H}_0 : \rho_s = 0$ can be tested by using the test statistic $\hat{\rho}_s \sqrt{n-1} \sim \mathrm{N}(0,1)$, where $\hat{\rho}_s$ is the estimator of ρ_s (a two-sided test).

4.3 Pearson's Correlation Coefficient (ρ)

Pearson's correlation coefficient is a common statistical measure. Between X and Y, it is defined as:

$$\rho^{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} = \frac{\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \ ,$$

which is estimated by $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \Big/ \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}$,

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

The two associated random variables X and Y are said to be positively correlated if $\rho > 0$, negatively correlated if $\rho < 0$, and uncorrelated if $\rho = 0$. It is well-known that independence implies zero correlation, but the reverse is not true unless the probability distribution is bivariate normal. As discussed in Watson (1983), the two hypotheses $\mathrm{H}_0 : \rho = 0$ and $\mathrm{H}_0 : \rho = k$ (for $-1 < k < 1$ and $k \neq 0$) can be tested by computing the test statistics $(n-2)\hat{\rho}^2 / (1 - \hat{\rho}^2) \sim \mathrm{F}_{1,n-2}$ (a one-sided test) and $\ln((1+\hat{\rho})/(1-\hat{\rho}))/2$ $\sim \mathrm{N}(\ \ln((1+k)/(1-k))/2\ ,\ 1/(n-3)\ )$ (a two-sided test) respectively, where $\hat{\rho}$ is the estimator of ρ.

This linear correlation measure ($-1 \le \rho \le 1$) is the most popular and well-known measure between pairwise random variables. It is frequently adopted in assessing the relationships between different lines of business, such as the calculations in Bateup and Reed (2001) and in Collings and White (2001). Despite its simplicity and plain rationale, Embrechts et al (2001) note that ρ is simply a measure of the dependency of elliptical distributions, such as the binormal distribution (i.e. the marginals are normally distributed, linked by the Gaussian copula (refer to Subsection 5.3)). Only computing ρ between related random variables and ignoring other aspects (e.g. tail dependence as discussed in Subsection 4.4) of the dependency structure are equivalent to implicitly assuming all the marginals are elliptically distributed and of the same type. It is proverbial that many lines of business have right-skewed liability distributions whereas elliptical distributions are symmetric. Different lines of an insurer would also tend to have different liability distributions. Moreover, ρ measures a linear relationship itself and does not capture a non-linear one on its own, as noted in Priest (2003). Consider a hypothetical case where $Y = X^2$ and $X \sim \mathrm{N}(0,1)$. In this example, even though X and Y are perfectly associated, it can be shown that $\mathrm{Cov}(X,Y) = \mathrm{E}(X^3) - \mathrm{E}(X)\mathrm{E}(X^2) = 0$ and so $\rho^{X,Y} = 0$, i.e. no correlation is detected. These properties constitute obvious limitations for modelling the dependency structure.

In contrast to τ and ρ_s, Melchiori (2003) comments that ρ is invariant only under strictly increasing linear transformations of X and Y, but is variant under non-linear transformations. Any non-linear change of the scale of the marginals thus impinges on this measure. This feature adds another layer of inflexibility.

Embrechts et al (2001) pinpoint that τ and ρ_s are more desirable than ρ as a measure of association for non-elliptical distributions, in which ρ is often misleading and inappropriate. Many general insurance liabilities have non-elliptical distributions that are right-skewed and so care is needed when ρ is used due to simplicity in practice.

4.4 Tail Dependence (λ_{upper} and λ_{lower})

As discussed in Joe (1997), upper tail dependence and lower tail dependence between X and Y are defined as follows:

$$\lambda_{\text{upper}} = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1-u} \text{ and } \lambda_{\text{lower}} = \lim_{u \to 0} \frac{C(u,u)}{u} \ ,$$

in which $0 \le \lambda_{\text{upper}} \le 1$ and $0 \le \lambda_{\text{lower}} \le 1$.

The variables X and Y have upper tail dependence if $\lambda_{\text{upper}} > 0$ and have lower tail dependence if $\lambda_{\text{lower}} > 0$. Upper tail dependence represents the association in the upper-right-quadrant tail and lower tail dependence represents the association in the lower-left-quadrant tail. There is no upper tail dependence if $\lambda_{\text{upper}} = 0$ and there is no lower tail dependence if $\lambda_{\text{lower}} = 0$. The two measures λ_{upper} and λ_{lower} are computed from the selected copula function with its parameters estimated.

4.5 Empirical Measurement

In practice, there are often insufficient data for estimating the measures of association, e.g. computing τ for inherent dependency between the outstanding claims liabilities of two lines of business. The assessment hence remains largely judgemental, and can become more feasible only when more data are collected in the future or when an insurer has a long history and keeps a good track of relevant claims data records. On the other hand, determination of other dependency is likely to continue to rely heavily on practical judgement.

5. TYPES OF COPULA

This section presents our selection of a few types of serviceable copula that possess varying modelling flexibility and reasonable tractability in practical situations. These copulas include Gaussian copula, *t* copula, Cook-Johnson copula, and a group of flexible copulas called Archimedean copulas. Some general properties of a copula are explained at the outset. The definition, the copula fitting technique, and the simulation procedure of each type of copula are set forth, in which different inferences can be drawn from the simulated samples. Examples of application to two hypothetical lines of business are given as each type of copula is introduced.

Suppose $X_{(i)}$'s are continuous m-variate random variables, $F_{X_{(1)},X_{(2)},\dots,X_{(m)}}$ is the joint m-variate cdf, $F_{X_{(i)}}$'s are the marginal univariate cdfs, and $u_{(i)} = F_{X_{(i)}}\left(x_{(i)}\right)$. Let C be the underlying m-variate copula. There are in total $m(m-1)/2$ pairs of bivariate random variables. Some further details regarding $F_{X_{(1)},X_{(2)},\dots,X_{(m)}}$ and $F_{X_{(i)}}$ are provided in Appendix I.

5.1 General Properties

The m-variate copula C can be expressed in the following forms:

$$C\left(F_{X_{(1)}}\left(x_{(1)}\right) , F_{X_{(2)}}\left(x_{(2)}\right) , \dots , F_{X_{(m)}}\left(x_{(m)}\right) \right) = C\left(u_{(1)} , u_{(2)} , \dots , u_{(m)} \right)$$

$$= F_{X_{(1)},X_{(2)},\dots,X_{(m)}}\left(x_{(1)} , x_{(2)} , \dots , x_{(m)} \right) = \Pr\left(X_{(1)} \le x_{(1)} , X_{(2)} \le x_{(2)} , \dots , X_{(m)} \le x_{(m)} \right),$$

where, as discussed in De Matteis (2001), $C\left(u_{(1)}, u_{(2)}, \dots, u_{(m)}\right)$ is a strictly increasing function of each $u_{(i)}$, with $0 \le C\left(u_{(1)}, u_{(2)}, \dots, u_{(m)}\right) \le 1$ and $C\left(1, \dots, 1, u_{(i)}, 1, \dots, 1\right) = u_{(i)}$.

Nelsen (1999) states that the copula becomes $C\left(u_{(1)}, u_{(2)}, \dots, u_{(m)}\right) = u_{(1)} u_{(2)} \dots u_{(m)}$ if and only if the $X_{(i)}$'s are independent. De Matteis (2001) states that C is invariant under strictly increasing transformations of $X_{(i)}$'s. For example, let $f_{(i)}$'s be strictly increasing functions of $X_{(i)}$'s, then the copula between $X_{(i)}$'s is the same as the copula between $f_{(i)}\left(X_{(i)}\right)$'s.

Moreover, according to Nelsen (1999), upper and lower bounds for C are derived as $\max\left(u_{(1)}+u_{(2)}+...+u_{(m)}-m+1\,,0\right)\leq C\left(u_{(1)},u_{(2)},...,u_{(m)}\right)\leq \min\left(u_{(1)},u_{(2)},...,u_{(m)}\right)$. The bounds are called Fréchet-Hoeffding bounds. The bounds are copulas themselves, except for the lower bound when the dimension is greater than two. According to Embrechts et al (2001), for the bivariate case with two random variables $X_{(1)}$ and $X_{(2)}$, the copula between them is the upper bound $\min\left(u_{(1)},u_{(2)}\right)$ if and only if $\tau^{X_{(1)},X_{(2)}}=\rho_s^{X_{(1)},X_{(2)}}=1$. In this case, $X_{(1)}$ and $X_{(2)}$ are said to be comonotonic. On the other hand, the copula between them is the lower bound $\max\left(u_{(1)}+u_{(2)}-1\,,0\right)$ if and only if $\tau^{X_{(1)},X_{(2)}}=\rho_s^{X_{(1)},X_{(2)}}=-1$, and $X_{(1)}$ and $X_{(2)}$ are then countermonotonic. Comonotonicity and countermonotonicity are extreme cases of concordance and discordance respectively.

In addition, as noted in Clemen and Reilly (1999), the copula density function is defined as $c\left(u_{(1)},u_{(2)},...,u_{(m)}\right)=\dfrac{\partial^m C\left(u_{(1)},u_{(2)},...,u_{(m)}\right)}{\partial u_{(1)}\partial u_{(2)}...\partial u_{(m)}}$. The joint multivariate pdf is then derived as $f_{X_{(1)},X_{(2)},...,X_{(m)}}\left(x_{(1)},x_{(2)},...,x_{(m)}\right)=c\left(u_{(1)},u_{(2)},...,u_{(m)}\right)f_{X_{(1)}}\left(x_{(1)}\right)f_{X_{(2)}}\left(x_{(2)}\right)...f_{X_{(m)}}\left(x_{(m)}\right)$, in which $f_{X_{(i)}}\left(x_{(i)}\right)$'s are the marginal univariate pdfs.

5.2 Simulation

Embrechts et al (2001) and De Matteis (2001) describe the general procedure of simulating multivariate random variables as follows, which is inefficient in many cases:

(1) Define $C_{(i)}\left(u_{(1)},u_{(2)},...,u_{(i)}\right)=C\left(u_{(1)},u_{(2)},...,u_{(i)},1,...,1\right)$ for i = 1, 2, … , m.

(2) Derive $C_{(i)}\left(u_{(i)}\middle|u_{(1)},u_{(2)},...,u_{(i-1)}\right)=\dfrac{\dfrac{\partial^{i-1}C_{(i)}\left(u_{(1)},u_{(2)},...,u_{(i)}\right)}{\partial u_{(1)}\partial u_{(2)}...\partial u_{(i-1)}}}{\dfrac{\partial^{i-1}C_{(i-1)}\left(u_{(1)},u_{(2)},...,u_{(i-1)}\right)}{\partial u_{(1)}\partial u_{(2)}...\partial u_{(i-1)}}}$ for i = 1, 2, … , m.

(3) Simulate $Q_{(1)} \sim \mathrm{U}(0,1)$. Take $V_{(1)} = Q_{(1)}$.

(4) Simulate $Q_{(2)} \sim \mathrm{U}(0,1)$. Compute $V_{(2)} = C_{(2)}^{-1}\left(Q_{(2)} \middle| V_{(1)}\right)$, using $V_{(1)}$ from the previous step.

(5) Simulate $Q_{(3)} \sim \mathrm{U}(0,1)$. Compute $V_{(3)} = C_{(3)}^{-1}\left(Q_{(3)} \middle| V_{(1)}, V_{(2)}\right)$, using $V_{(1)}$ and $V_{(2)}$ from the previous steps.

(6) Effectively, the computation is $V_{(i)} = C_{(i)}^{-1}\left(Q_{(i)} \middle| V_{(1)}, V_{(2)}, \ldots, V_{(i-1)}\right)$.

(7) Repeat (6) until a set of $V_{(i)}$'s is generated for $i = 1, 2, \ldots, m$.

(8) Compute $X_{(i)} = F_{X_{(i)}}^{-1}\left(V_{(i)}\right)$ for $i = 1, 2, \ldots, m$. Go to (3) to repeat the simulation.

If the function inverses do not have a closed form, numerical root finding is necessary. For some of the copulas introduced in the following subsections, the simulation procedures are simpler and more efficient. Excel and Mathematica can be used to carry out all the simulations (refer to Appendix II for details).

5.3 Gaussian Copula

The Gaussian copula is the copula embedded in the standard multinormal distribution. Instead of being mixed with some normally distributed marginals to come up naturally with a multinormal distribution, it can be amalgamated with other marginal distributions.

As stated in Embrechts et al (2001), the Gaussian copula is of the following form:

$$C\left(u_{(1)}, u_{(2)}, \ldots, u_{(m)}\right) = \Phi_m\left(\Phi^{-1}\left(u_{(1)}\right), \Phi^{-1}\left(u_{(2)}\right), \ldots, \Phi^{-1}\left(u_{(m)}\right)\right),$$

where Φ_m is the m-variate standard normal cdf and Φ is the univariate standard normal cdf with the inverse Φ^{-1}.

Embrechts et al (2001) describe the following copula fitting technique. To fit the Gaussian copula, τ $(-1 \le \tau \le 1)$ is estimated for each pair of random variables (refer to Appendix II for computing the sample τ). Then r $(-1 \le r \le 1)$ is computed for each pair of random variables as $r = \sin(\pi\tau/2)$, and all the r's calculated are used as the components of the dispersion matrix of the Gaussian copula. In effect, this dispersion matrix contains the parameters for the underlying dependency. The correlation coefficient ρ is not estimated directly between each pair of random variables and then used as r because it only measures the linear relationship and the marginals may be non-elliptical. This point is discussed in Subsection 4.3. Clemen and Reilly (1999) note that the dispersion matrix must be positive definite and symmetric and adjustment is needed if the estimated matrix is not positive definite. Clemen and Reilly (1999) also note that ρ_s ($-1 \le \rho_s \le 1$) can also be used and r is then computed as $r = 2\sin(\pi\rho_s/6)$.

After forming the dispersion matrix as stated above, m-variate random variables with the Gaussian copula can be simulated by the following steps, as detailed in Embrechts et al (2001):

(1) Sample $Z_{(1)}$, $Z_{(2)}$, … , and $Z_{(m)}$ from the m-variate standard normal distribution with the estimated dispersion matrix as the linear correlation matrix, via the Cholesky decomposition process (refer to Appendix II).

(2) Compute $V_{(i)} = \Phi(Z_{(i)})$ for i = 1, 2, … , m.

(3) Compute $X_{(i)} = F_{X_{(i)}}^{-1}(V_{(i)})$ for i = 1, 2, … , m.

In some practical situations, there is more concern about the relationship between severe events than between normal events. In this regard, the Gaussian copula does not have upper tail dependence or lower tail dependence, i.e. $\lambda_{\text{upper}} = \lambda_{\text{lower}} = 0$. To illustrate this property, we set up an example of two lines of business. Suppose bivariate $(X_{(1)}, X_{(2)})$ represent the outstanding claims liabilities of two lines of business. Their probability distributions are known to be $X_{(1)} \sim \text{LN}(15\ ,\ 0.2)$ ($\text{E}(X_{(1)}) \approx 3{,}335{,}056$; coefficient of variation ≈ 20%) and $X_{(2)} \sim \gamma(44\ ,\ 7.3 \times 10^{-6})$ ($\text{E}(X_{(2)}) \approx 6{,}027{,}397$; coefficient of variation ≈ 15%). Suppose the dependency structure is unknown and modelled arbitrarily by the Gaussian copula, with estimated $\tau^{X_{(1)},X_{(2)}} = 0.35$. Figure 1 at the end of this section shows a scatter plot of $\left(X_{(1)}, X_{(2)}\right)$ with 30,000 samples. The relationship weakens when moving to the northeast direction. Table 3 shows the sample $\tau^{X_{(1)},X_{(2)}}$ in the upper-right-quadrant tail. The sample $\tau^{X_{(1)},X_{(2)}}$ decreases as the percentile above which the figure is calculated increases, and is much smaller than 0.35 in general. Furthermore, Figure 1 also exhibits the bivariate pdf of $\left(X_{(1)}, X_{(2)}\right)$, in which the upper-right-quadrant tail does not have a concentrated density. The corresponding contours are also shown.

Accordingly, the high-end percentiles of the total outstanding claims liabilities of the two lines and so the necessary safety margin allowance may be underestimated if the Gaussian copula is assumed and there is no allowance for potential upper tail dependence. Despite the tractable mathematics of the Gaussian copula, its upper tail independence imposes a serious constraint on modelling any dependent tail events.

5.4 *t* copula

The *t* copula is the copula embedded in the multivariate central *t* distribution (with zero mean vector). This copula can be combined with marginal distributions other than Student's *t*. As stated in Embrechts et al (2001), the *t* copula has the following form:

$$C\left(u_{(1)}, u_{(2)}, \ldots, u_{(m)}\right) = t_{v:m}\left(t_v^{-1}\left(u_{(1)}\right), t_v^{-1}\left(u_{(2)}\right), \ldots, t_v^{-1}\left(u_{(m)}\right)\right),$$

where $t_{v:m}$ is the m-variate central *t* cdf and t_v is the univariate Student's *t* cdf with the inverse t_v^{-1} (v is the degrees of freedom).

According to Embrechts et al (2001), fitting the *t* copula is similar to fitting the Gaussian copula since both the multivariate *t* and multinormal distributions are elliptical distributions. (As noted in Lindskog (2000), the *t* copula converges to the Gaussian copula as v increases to infinity.) Accordingly, τ $(-1 \le \tau \le 1)$ is estimated for each pair of random variables and r $(-1 \le r \le 1)$ is computed as $r = \sin(\pi\tau/2)$, which is used to construct the dispersion matrix of the *t* copula. Again, the dispersion matrix must be positive definite and symmetric. The parameter v is estimated by carrying out the formal tests described in Appendix III on different trial values of v, or it is determined by general judgement on the tail dependence property.

After forming the dispersion matrix and deciding the value of v, m-variate random variables with the t copula can be simulated by the following steps, as detailed in Embrechts et al (2001):

(1) Sample $Z_{(1)}$, $Z_{(2)}$, … , and $Z_{(m)}$ from the m-variate standard normal distribution with the estimated dispersion matrix as the linear correlation matrix, via the Cholesky decomposition process (refer to Appendix II).

(2) Sample $S \sim \chi_v^2$, where v is the degrees of freedom.

(3) Compute $V_{(i)} = t_v\left(Z_{(i)}\sqrt{v/S}\right)$ for i = 1, 2, … , m (the same S for all i).

(4) Compute $X_{(i)} = F_{X_{(i)}}^{-1}\left(V_{(i)}\right)$ for i = 1, 2, … , m.

The t copula has the same upper and lower tail dependence for each pair of random variables, as derived in Embrechts et al (2001):

$$\lambda_{\text{upper}} = \lambda_{\text{lower}} = 2 - 2t_{v+1}\left(\sqrt{(v+1)(1-r)/(1+r)}\right),$$

which tends to zero when v goes to infinity.

In contrast to the Gaussian copula, the t copula has upper tail dependence for modelling dependent severe events. Figures 2 to 8 exhibit some scatter plots of bivariate $\left(X_{(1)}, X_{(2)}\right)$ for the same marginal distributions and the same estimated $\tau^{X_{(1)},X_{(2)}}$ as in the previous example, but modelled with the t copula and various values of v. For small values of v, the relationship remains significant even when moving to the northeast direction. Moreover, Table 3 shows that for small values of v the sample $\tau^{X_{(1)},X_{(2)}}$, not much smaller than 0.35 in general, only decreases slightly as the percentile considered increases. As shown in Figures 2 to 8, the concentration of the density of $\left(X_{(1)}, X_{(2)}\right)$ at their common upper tails decreases from a higher level towards the extent of the Gaussian copula as v increases. Figures 2 to 8 also present the contours, which show declining tail association as v increases. Accordingly, upper tail dependence can be accommodated by selecting an appropriate value of v, apart from setting the value of τ.

Table 4 presents our calculation of λ_{upper} (and so λ_{lower}) for different values of τ and v. As v increases to infinity, the t copula converges to the Gaussian copula and so λ_{upper} tends to zero. Furthermore, even for zero (or negative) τ, there is still upper tail dependence for small values of v. This property is suitable for the situation when only the most severe events, not normal events, of different lines are related.

There are two drawbacks of applying the t copula in modelling general insurance liabilities. First, like the Gaussian copula, the t copula is symmetric – as noted in Demarta and McNeil (2004), for the multivariate case, tail dependence of any corner of the t copula is the same as that of the opposite corner – and so $\lambda_{\text{upper}} = \lambda_{\text{lower}}$ for each pair of random variables. There are many situations where asymmetry is more likely, e.g. large claims are related to some extent but small claims are not related or not as equally related. If the focus is only on severe large claims, however, the symmetric tail dependence at the lower end can be ignored as an approximation. Second, the t copula involves only one parameter v for all pairs of random variables, despite its allowance for a particular r for each pair. As noted in Venter (2003), the parameter v is involved in determining tail dependence for all pairs of random variables, and thus there is less modelling flexibility.

5.5 Cook-Johnson Copula

Studied by Cook and Johnson (1981), the Cook-Johnson copula has the following form:

$$C\left(u_{(1)}, u_{(2)}, \ldots, u_{(m)}\right) = \max\left(\left(u_{(1)}^{-\theta} + u_{(2)}^{-\theta} + \ldots + u_{(m)}^{-\theta} - m + 1\right)^{-1/\theta},\ 0\right).$$

The term θ is the sole parameter of the Cook-Johnson copula. Since there is only one parameter, the same extent of association has to be assumed for all pairs of random variables. This property is very restrictive. The copula is only useful for certain lines of business that have very similar levels of association. Wang (1998) states that the association between each pair of random variables has to be positive, i.e. $\tau > 0$. Cook and Johnson (1981) note that negative association can be allowed for by replacing $u_{(i)}$ with $\left(1-u_{(i)}\right)$ for some (but not all) i.

To fit the copula, τ ($0 < \tau \le 1$) is estimated for each pair of random variables. Assuming all the estimated τ 's are close, an average value of τ is determined and used to obtain θ ($\theta > 0$), using the formula $\theta = 2\tau/(1-\tau)$ as stated in Wang (1998).

After obtaining θ as above, m-variate random variables with the Cook-Johnson copula can be simulated by the following steps, as detailed in Cook and Johnson (1981):

(1) Sample $A_{(i)} \sim \exp(1)$ for $i = 1, 2, \ldots, m$.

(2) Sample $B \sim \gamma(1/\theta, 1)$.

(3) Compute $V_{(i)} = \left(1 + A_{(i)}/B\right)^{-1/\theta}$ for $i = 1, 2, \ldots, m$ (the same B for all i).

(4) Compute $X_{(i)} = F_{X_{(i)}}^{-1}\left(V_{(i)}\right)$ for $i = 1, 2, \ldots, m$.

We find that the Cook-Johnson copula does not have upper tail dependence as $\lambda_{\text{upper}} = 0$, but it has lower tail dependence as $\lambda_{\text{lower}} = 2^{-1/\theta}$. This upper tail independence is inflexible for modelling the relationship between severe events. Figure 9 corresponds to Figure 1 with the Gaussian copula replaced by the Cook-Johnson copula, whose lower tail dependence and upper tail independence are clearly reflected in the graphs. From Table 3, the sample $\tau^{X_{(1)},X_{(2)}}$ is even lower than that of the Gaussian copula across different percentiles.

For the bivariate case, the Cook-Johnson copula is equivalent to the so-called Clayton copula, which is a type of the Archimedean copulas described in the next subsection. For the multivariate case, the Cook-Johnson copula is equivalent to one possible multivariate extension of the Clayton copula.

5.6 Archimedean Copulas

The Archimedean copulas are a group of copulas that possess a number of similar characteristics. Many of them are flexible and tractable and furnish numerous structures for modelling different dependency properties. Starting from the bivariate case, the Archimedean copulas have the following basic form, as stated in Nelsen (1999):

$$C\left(u_{(1)},u_{(2)}\right)=\varphi^{-1}\left(\varphi\left(u_{(1)}\right)+\varphi\left(u_{(2)}\right)\right),$$

where φ is a function called generator and φ^{-1} is the corresponding inverse.

As noted in Nelsen (1999), each type of Archimedean copula has a unique form of φ. In general, φ is continuous and strictly decreasing with $\varphi(1)=0$. For practicality, we only consider strict generators with $\varphi(0)=\infty$, and the copulas are then called strict Archimedean copulas. Moreover, only the one-parameter family (for pairwise random variables) is discussed here.

To fit the copula, $\tau^{X_{(1)},X_{(2)}}$ is estimated between bivariate $X_{(1)}$ and $X_{(2)}$. The parameter θ of the selected copula is then computed by exploiting the formula $\tau=1+4\int_0^1\varphi(t)/\varphi'(t)dt$ as derived in Nelsen (1999), where $\varphi'(t)$ is the first derivative of $\varphi(t)$ with respect to *t*.

Using the formula $K_C(t)=t-\varphi(t)/\varphi'(t)$ as derived in Nelsen (1999), bivariate random variables with the selected Archimedean copula can be simulated by the following procedure, as detailed in Embrechts et al (2001):

(1) Sample $S\sim \mathrm{U}(0,1)$ and $Q\sim \mathrm{U}(0,1)$.

(2) Compute $T=K_C^{-1}(Q)$, in which K_C^{-1} is the inverse of K_C. This step may require numerical root finding using $\frac{d}{dt}K_c(t)$ (refer to Appendix II).

(3) Compute $V_{(1)}=\varphi^{-1}\left(S\varphi(T)\right)$ and $V_{(2)}=\varphi^{-1}\left((1-S)\varphi(T)\right)$.

(4) Compute $X_{(1)} = F_{X_{(1)}}^{-1}\left(V_{(1)}\right)$ and $X_{(2)} = F_{X_{(2)}}^{-1}\left(V_{(2)}\right)$.

Furthermore, according to Embrechts et al (2001), upper and lower tail dependence are derived as:

$$\lambda_{\text{upper}} = 2 - 2\lim_{s\to 0}\frac{\varphi^{-1\prime}(2s)}{\varphi^{-1\prime}(s)} \text{ and } \lambda_{\text{lower}} = 2\lim_{s\to\infty}\frac{\varphi^{-1\prime}(2s)}{\varphi^{-1\prime}(s)},$$

where $\varphi^{-1\prime}(s)$ is the first derivative of $\varphi^{-1}(s)$ with respect to s.

Tables 1 and 2 in the following present the characteristics of four one-parameter strict Archimedean copulas, including the Clayton copula, the Gumbel-Hougaard copula, the Frank copula, and the Archimedean copula no. 12 in Nelsen (1999) (referred to as Nelsen no. 12 copula here). The contents in the tables emanate from Joe (1997), Nelsen (1999), Embrechts et al (2001), Melchiori (2003), and our derivation ($\varphi'(t)$ and $\varphi^{-1}(s)$ of the Clayton copula, $\varphi'(t)$, $K_C(t)$ and its first derivative, and tail dependence of the Nelsen no. 12 copula, and $\varphi^{-1\prime}(s)$ of the four copulas). This information is required for simulating multivariate random variables, computing tail dependence, and understanding the possible ranges of association.

These four copulas demonstrate some of the flexibility of the Archimedean copulas. The Clayton copula, proposed by Clayton (1978), allows for lower tail dependence and positive association (negative association is excluded here, but is indeed possible with the result that the copula is then not strict). The Gumbel-Hougaard copula, proposed by Gumbel (1960) and discussed by Hougaard (1986), accommodates upper tail dependence and non-negative association. While the Frank copula, first studied by Frank (1979), has no tail dependence behaviour, it allows for negative association readily (but only positive association is feasible if C is extended to more than two dimensions and then φ^{-1} needs to be completely monotonic, as discussed later). In addition, the Debye function $D_1(\theta) = \int_0^\theta t/(\exp(t)-1)\,dt \big/ \theta$ for the Frank copula can be solved by using a Riemann sum or Mathematica (refer to Appendix II). The last copula involves both upper and lower tail dependence and accommodates positive association only in the range of $1/3 \le \tau \le 1$.

Table 1 Clayton Copula and Gumbel-Hougaard Copula

	Clayton Copula	Gumbel-Hougaard Copula
$C(u_{(1)}, u_{(2)})$	$\max\left(\left(u_{(1)}^{-\theta} + u_{(2)}^{-\theta} - 1\right)^{-\frac{1}{\theta}}, 0\right)$	$\exp\left(-\left((-\ln u_{(1)})^{\theta} + (-\ln u_{(2)})^{\theta}\right)^{\frac{1}{\theta}}\right)$
θ	$\frac{2\tau}{1-\tau}$ $(\theta > 0\ ;\ 0 < \tau \le 1)$	$\frac{1}{1-\tau}$ $(\theta \ge 1\ ;\ 0 \le \tau \le 1)$
$\varphi(t)$	$\frac{t^{-\theta} - 1}{\theta}$	$(-\ln t)^{\theta}$
$\frac{d}{dt}\varphi(t)$	$-t^{-\theta-1}$	$-\frac{\theta(-\ln t)^{\theta-1}}{t}$
$\varphi^{-1}(s)$	$(1+\theta s)^{-\frac{1}{\theta}}$	$\exp\left(-s^{\frac{1}{\theta}}\right)$
$\frac{d}{ds}\varphi^{-1}(s)$	$-(1+\theta s)^{-\frac{1}{\theta}-1}$	$-\frac{1}{\theta}\exp\left(-s^{\frac{1}{\theta}}\right)s^{\frac{1}{\theta}-1}$
$K_C(t)$	$t - \frac{t^{\theta+1} - t}{\theta}$	$t - \frac{t\ln t}{\theta}$
$\frac{d}{dt}K_C(t)$	$1 - \frac{(\theta+1)t^{\theta} - 1}{\theta}$	$1 - \frac{1+\ln t}{\theta}$
λ_{upper}	0	$2 - 2^{\frac{1}{\theta}}$
λ_{lower}	$2^{-\frac{1}{\theta}}$	0

Table 2 Frank Copula and Nelsen No. 12 Copula

	Frank Copula	Nelsen No. 12 Copula
$C(u_{(1)},u_{(2)})$	$\frac{-1}{\theta}\ln\left(1+\frac{(\exp(-\theta u_{(1)})-1)(\exp(-\theta u_{(2)})-1)}{\exp(-\theta)-1}\right)$	$\left(1+\left((u_{(1)}^{-1}-1)^{\theta}+(u_{(2)}^{-1}-1)^{\theta}\right)^{\frac{1}{\theta}}\right)^{-1}$
θ	$\tau=1-\frac{4(1-D_1(\theta))}{\theta}\begin{pmatrix}\theta\neq0;\ \tau\neq0;\\ -1\leq\tau\leq1\end{pmatrix}$	$\frac{2}{3(1-\tau)}$ $(\theta\geq1;\ \frac{1}{3}\leq\tau\leq1)$
$\varphi(t)$	$-\ln\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$\left(\frac{1}{t}-1\right)^{\theta}$
$\frac{d}{dt}\varphi(t)$	$\frac{\theta}{1-\exp(\theta t)}$	$-\frac{\theta}{t^2}\left(\frac{1}{t}-1\right)^{\theta-1}$
$\varphi^{-1}(s)$	$-\frac{1}{\theta}\ln\left(1+(\exp(-\theta)-1)\exp(-s)\right)$	$\left(1+s^{\frac{1}{\theta}}\right)^{-1}$
$\frac{d}{ds}\varphi^{-1}(s)$	$\frac{(\exp(-\theta)-1)\exp(-s)}{\theta\left(1+(\exp(-\theta)-1)\exp(-s)\right)}$	$-\frac{1}{\theta}s^{\frac{1}{\theta}-1}\left(1+s^{\frac{1}{\theta}}\right)^{-2}$
$K_C(t)$	$t-\frac{1}{\theta}(\exp(\theta t)-1)\ln\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$t-\frac{t(t-1)}{\theta}$
$\frac{d}{dt}K_C(t)$	$-\exp(\theta t)\ln\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$1-\frac{2t-1}{\theta}$
λ_{upper}	0	$2-2^{\frac{1}{\theta}}$
λ_{lower}	0	$2^{-\frac{1}{\theta}}$

For more than two dimensions with all pairs of random variables having the same φ and the same θ (i.e. exactly the same type of Archimedean copula with the same parameter), C can be extended as follows, as detailed in Embrechts et al (2001):

$$C(u_{(1)},u_{(2)},\ldots,u_{(m)})=\varphi^{-1}(\varphi(u_{(1)})+\varphi(u_{(2)})+\ldots+\varphi(u_{(m)})),$$

provided that φ^{-1} is completely monotonic, i.e. φ^{-1} belongs to the class of Laplace transforms $\mathcal{L}_{\infty}$ (refer to Appendix IV).

According to Joe (1997) and Embrechts et al (2001), φ^{-1} of each of the four copulas described previously is completely monotonic and so the multivariate extension above is feasible. For other Archimedean copulas with φ^{-1} not being completely monotonic, however, this extension does not provide a multivariate copula.

We derive an example of the extension of the m-variate Clayton copula as follows, which is equivalent to the m-variate Cook-Johnson copula:

$$C\left(u_{(1)}, u_{(2)}, \ldots, u_{(m)}\right) = \max\left(\left(u_{(1)}^{-\theta} + u_{(2)}^{-\theta} + \ldots + u_{(m)}^{-\theta} - m + 1\right)^{-1/\theta},\ 0\right),$$

and the simulation (as described in Subsection 5.2) involves the following formulae:

$$C_{(i)}\left(u_{(i)} \middle| u_{(1)}, u_{(2)}, \ldots, u_{(i-1)}\right) = \left(\frac{u_{(1)}^{-\theta} + u_{(2)}^{-\theta} + \ldots + u_{(i)}^{-\theta} - i + 1}{u_{(1)}^{-\theta} + u_{(2)}^{-\theta} + \ldots + u_{(i-1)}^{-\theta} - i + 2} \right)^{-1/\theta - i + 1}, \text{ and}$$

$$V_{(i)} = \left(\left(Q_{(i)}^{1/(-1/\theta - i + 1)} - 1\right)\left(V_{(1)}^{-\theta} + V_{(2)}^{-\theta} + \ldots + V_{(i-1)}^{-\theta} - i + 2\right) + 1 \right)^{-1/\theta}.$$

On the other hand, if φ is the same but θ varies for different pairs of random variables (i.e. the same type of Archimedean copula but with different parameters), Joe (1997) states that there are several ways to extend C to more than two dimensions. As noted in Embrechts et al (2001), one possible extension is shown in the following, say $m = 5$:

$$C\left(u_{(1)}, u_{(2)}\right) = \varphi_4^{-1}\left(\varphi_4\left(u_{(1)}\right) + \varphi_4\left(u_{(2)}\right)\right),$$

$$C\left(u_{(1)}, u_{(2)}, u_{(3)}\right) = \varphi_3^{-1}\left(\varphi_3\left(C\left(u_{(1)}, u_{(2)}\right)\right) + \varphi_3\left(u_{(3)}\right)\right),$$

$$C\left(u_{(1)}, u_{(2)}, u_{(3)}, u_{(4)}\right) = \varphi_2^{-1}\left(\varphi_2\left(C\left(u_{(1)}, u_{(2)}, u_{(3)}\right)\right) + \varphi_2\left(u_{(4)}\right)\right), \text{ and}$$

$$C\left(u_{(1)}, u_{(2)}, u_{(3)}, u_{(4)}, u_{(5)}\right) = \varphi_1^{-1}\left(\varphi_1\left(C\left(u_{(1)}, u_{(2)}, u_{(3)}, u_{(4)}\right)\right) + \varphi_1\left(u_{(5)}\right)\right),$$

provided that φ_i^{-1} (for i = 1 to 4 in which some or all θ_i's are different) is completely monotonic and $\varphi_i \circ \varphi_{i+1}^{-1}$ (for i = 1 to 3) belongs to the class $\mathcal{L}_\infty^*$ (refer to Appendix IV).

According to Joe (1997) and Embrechts et al (2001), $\varphi_i \circ \varphi_{i+1}^{-1}$ of the four types of copula discussed belong to the class $\mathcal{L}_\infty^*$ if $\theta_1 < \theta_2 < \theta_3 < \theta_4$.

Effectively, the relationship between $X_{(1)}$ and $X_{(2)}$ is determined by θ_4, the relationship between $X_{(1)}$ (or $X_{(2)}$) and $X_{(3)}$ is determined by θ_3, the relationship between $X_{(1)}$ (or $X_{(2)}$ or $X_{(3)}$) and $X_{(4)}$ is determined by θ_2, and the relationship between $X_{(1)}$ (or $X_{(2)}$ or $X_{(3)}$ or $X_{(4)}$) and $X_{(5)}$ is determined by θ_1. Hence there are only m – 1 (= 4) distinct parameters for m (= 5) random variables, compared to $m(m-1)/2$ (= 10) distinct parameters if the copula was Gaussian. This property is called partial exchangeability and it imposes a constraint on setting the parameters.

This means of extension can be applied to other dimensions similarly. Accordingly, the procedure in Subsection 5.2 is required to simulate m-variate random variables. Some information of exploiting Excel or Mathematica to carry out the simulation is provided in Appendix II. Other means of extension are set forth in Joe (1997).

Figures 10 to 13 correspond to Figure 1 with the Gaussian copula replaced by the four Archimedean copulas (the Clayton copula is just the Cook-Johnson copula in this bivariate example). The Gumbel-Hougaard copula shows very strong upper tail dependence with $\lambda_{\text{upper}} = 0.4308$ (larger than that of the t copula with v = 3). The Clayton copula and the Nelsen no. 12 copula have very strong lower tail dependence with $\lambda_{\text{lower}} = 0.5254$ and $\lambda_{\text{lower}} = 0.5087$ respectively. The Nelsen no. 12 copula does not show obvious upper tail dependence in the graphs because its λ_{upper} is only equal to 0.0344. The graphs of the Frank copula look quite similar to those of the Gaussian copula.

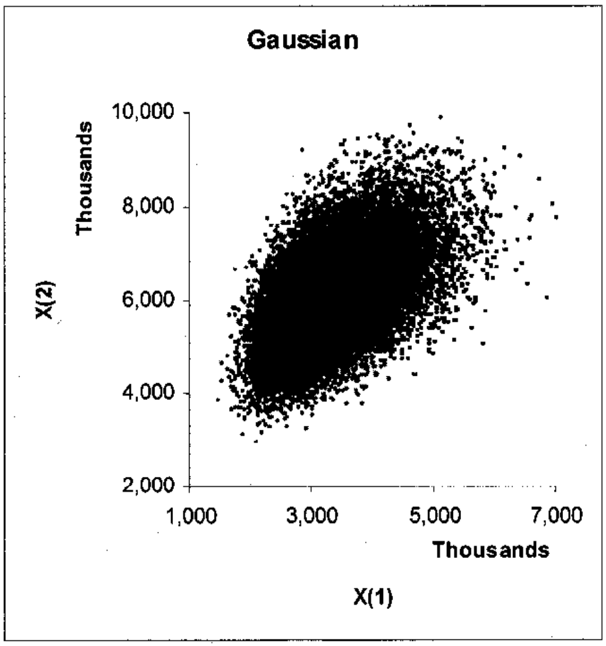
Table 3 exhibits the sample $\tau^{X_{(1)},X_{(2)}}$ in the upper-right-quadrant tail. For the Gumbel-Hougaard copula, the sample $\tau^{X_{(1)},X_{(2)}}$ remains at about 0.3 for all percentiles. For the other three copulas, the sample $\tau^{X_{(1)},X_{(2)}}$ is very low. In addition, the results of the Clayton copula are very similar to those of the Cook-Johnson copula, as the two copulas are identical in this bivariate example.

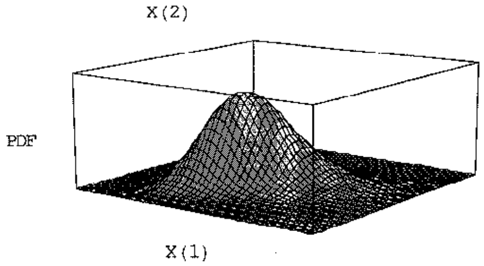
5.7 Copula Selection

As discussed in the previous subsections, r or θ is computed from the estimated τ once the type of copula is chosen. Some formal tests to justify the selection of copula, based on the observations of two bivariate random variables, are provided in Appendix III. These tests include goodness of fit test, cdf test, $K_C(t)$ test, and binomial test. They can also be used to warrant the estimated value of r or θ (or v for the t copula).

In many practical circumstances, appropriate judgement regarding the suitability of the selected copula and the reasonableness of the estimated values of the parameters is vital. A judgemental approach is of particular importance if the data are scarce, which is common in general insurance practice. Assumptions have to be set up carefully via thorough investigation of the portfolios. Experience analysis is essential for responding to the possibility of setting erroneous assumptions in the first instance and to any changes of the dependency structures over time.

Figure 1 Scatter Plot, PDF, and Contours – Gaussian Copula





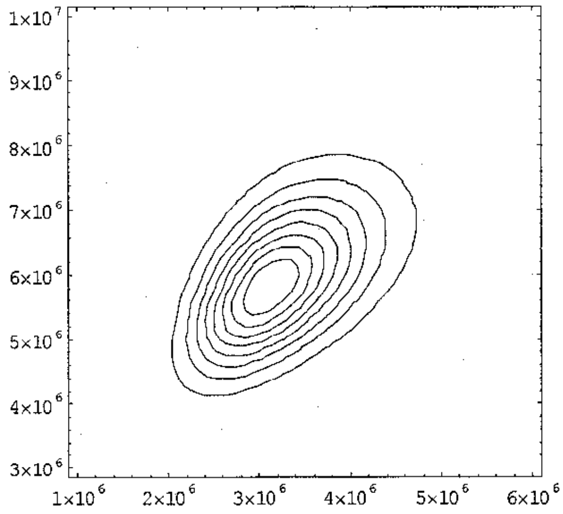


Figure 2 Scatter Plot, PDF, and Contours – t_3 Copula

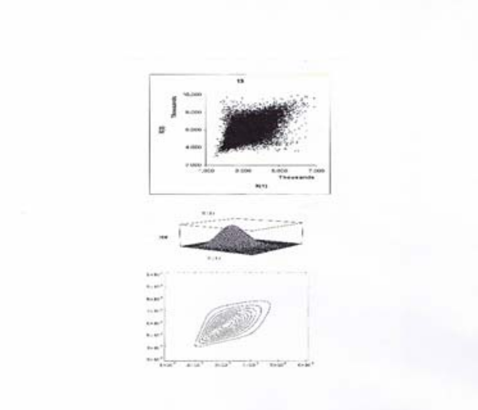


Figure 3 Scatter Plot, PDF, and Contours – t_5 Copula

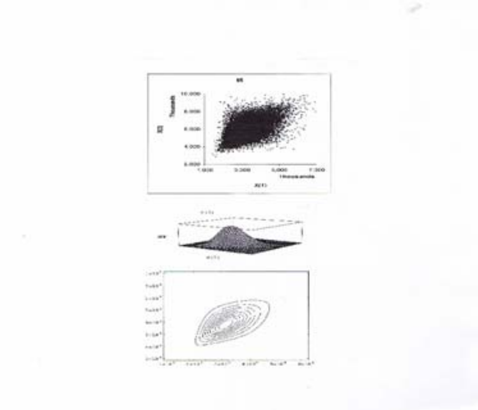


Figure 4 Scatter Plot, PDF, and Contours – t_8 Copula

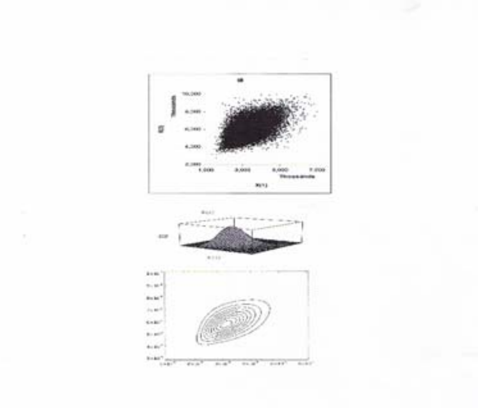


Figure 5 Scatter Plot, PDF, and Contours – t_{10} Copula

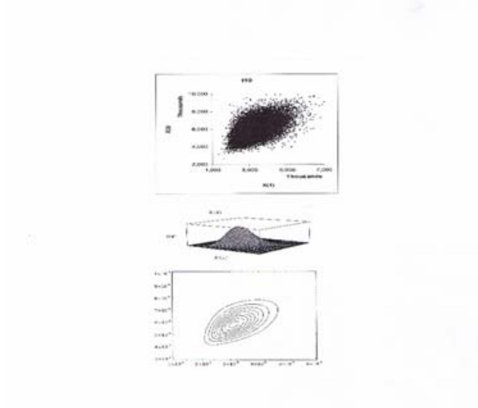


Figure 6 Scatter Plot, PDF, and Contours – t_{50} Copula

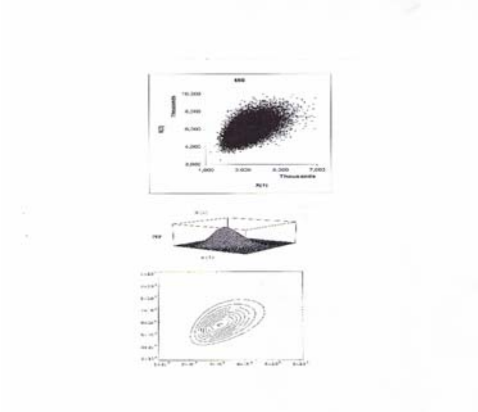


Figure 7 Scatter Plot, PDF, and Contours – t_{100} Copula

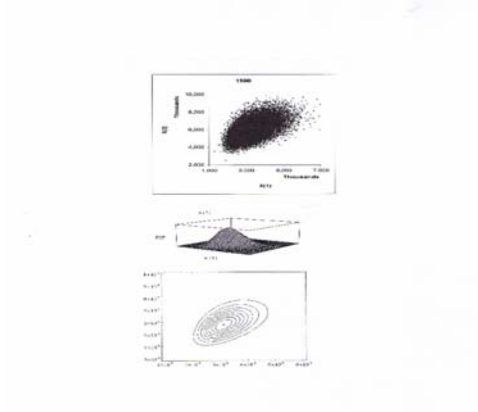


Figure 8 Scatter Plot, PDF, and Contours – t_{200} Copula

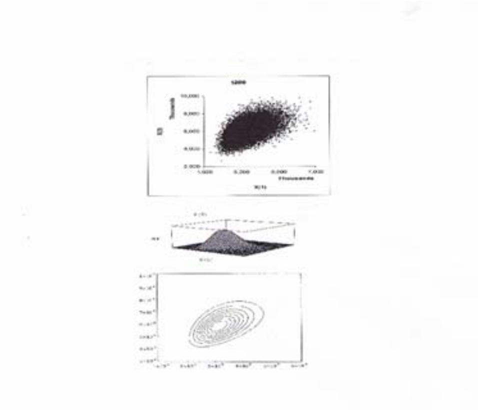


Figure 9 Scatter Plot, PDF, and Contours – Cook-Johnson Copula

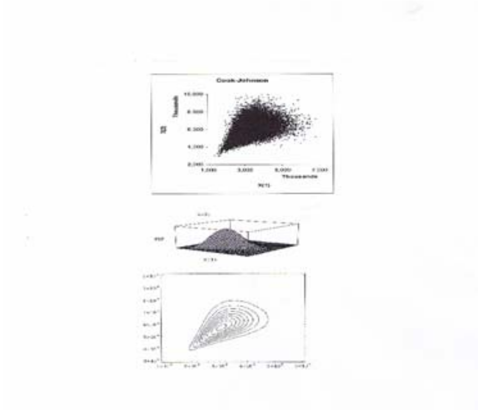


Figure 10 Scatter Plot, PDF, and Contours – Clayton Copula

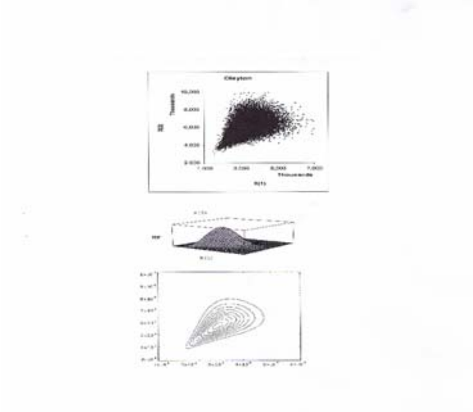


Figure 11 Scatter Plot, PDF, and Contours – Gumbel-Hougaard Copula

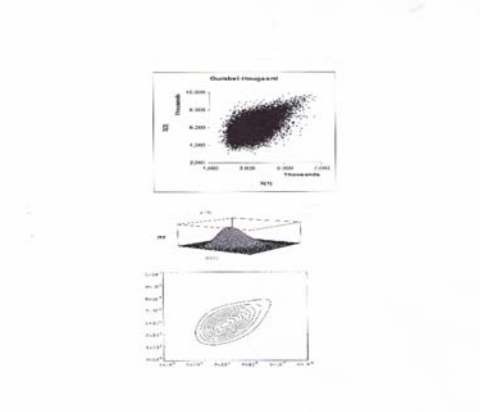


Figure 12 Scatter Plot, PDF, and Contours – Frank Copula

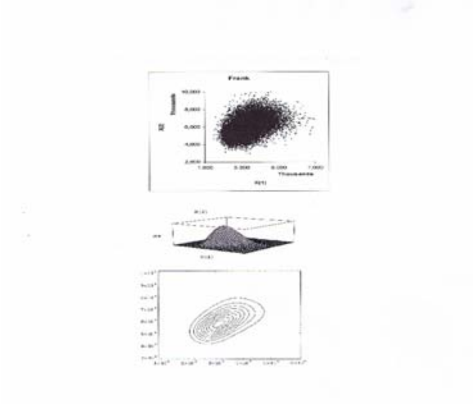


Figure 13 Scatter Plot, PDF, and Contours – Nelsen No. 12 Copula

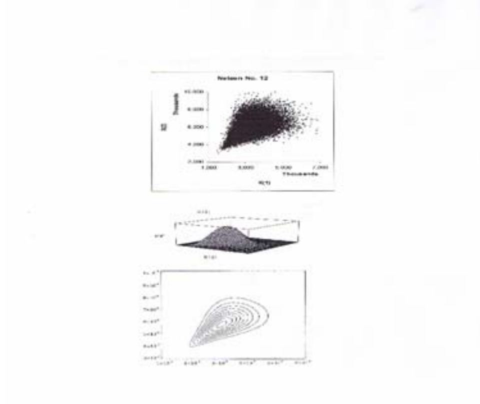


Table 3 Sample Kendall's Tau between Bivariate Outstanding Claims Liabilities of Two Lines of Business

Percentile	Gaussian	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	Cook-Johnson	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
95th	0.09	0.27	0.20	0.16	0.15	0.07	0.07	0.09	0.00	0.02	0.35	0.01	0.00
90th	0.07	0.27	0.19	0.18	0.17	0.10	0.09	0.09	0.01	0.03	0.31	0.06	0.09
85th	0.10	0.29	0.22	0.17	0.16	0.10	0.09	0.10	0.03	0.02	0.31	0.03	0.06
80th	0.11	0.29	0.22	0.18	0.16	0.11	0.10	0.10	0.01	0.02	0.29	0.04	0.05
75th	0.12	0.29	0.23	0.19	0.18	0.13	0.12	0.12	0.01	0.02	0.30	0.04	0.04
70th	0.14	0.29	0.24	0.20	0.19	0.16	0.15	0.14	0.02	0.02	0.30	0.06	0.04
65th	0.15	0.29	0.24	0.22	0.20	0.17	0.16	0.16	0.02	0.02	0.29	0.08	0.04

Table 4 Tail Dependence of *t* Copula

$v \setminus \tau$	-0.8	-0.6	-0.4	-0.2	0	0.2	0.35	0.4	0.6	0.8	1
3	0.0002	0.0035	0.0172	0.0512	0.1161	0.2199	0.3254	0.3658	0.5512	0.7673	1.0000
5	0.0000	0.0003	0.0030	0.0150	0.0498	0.1254	0.2192	0.2585	0.4564	0.7114	1.0000
8	0.0000	0.0000	0.0002	0.0026	0.0150	0.0572	0.1272	0.1607	0.3551	0.6460	1.0000
10	0.0000	0.0000	0.0000	0.0008	0.0069	0.0346	0.0902	0.1192	0.3043	0.6098	1.0000
50	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006	0.0244	0.2633	1.0000
100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0015	0.1146	1.0000
200	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0258	1.0000

6. PRACTICAL APPLICATION

In the following, we set up a hypothetical example of practical application of some copulas on measuring the uncertainty of outstanding claims liabilities. The assumptions are set out at the start and the simulation results are presented thereafter. We exploit the simulation procedures of the copulas as discussed in the previous section.

6.1 Assumptions and Parameter Estimation

Table 5 states our assumptions of the probability distributions (refer to Appendix I) of the total outstanding claims liabilities of eight lines of business. We assume all the lines have the same expected total outstanding claims liability of 80 million but with different variability levels (including parameter error and process error as noted in Li (2006)). The value of the expected total outstanding claims liabilities of the whole portfolio is then 640 million. We compute the parameters for each line to ensure the magnitude of one half of the coefficient of variation corresponds to the industry levels as stated in Bateup and Reed (2001) and Collings and White (2001). According to Wang (1998), the gamma, Weibull, inverse Gaussian, and lognormal distributions have an increasing order of right-tail heaviness. Since the last three lines in the table are generally regarded as having high uncertainty and heavy right tails in their liability distributions, we arbitrarily make use of the lognormal distribution for these three lines. In practice, the lognormal distribution is popular for modelling general insurance liabilities conceived to possess heavy tails. We assume the other lines have liability distributions with lighter tails.

In reality, the mean and the standard deviation can be estimated by say the Mack model (refer to Mack (1993)) and the form of liability distribution can be determined by prior knowledge regarding the nature of the business or by analysis of the past claims data.

Table 5 Assumptions of Probability Distributions

Line	Probability Distribution	Mean	Standard Deviation	$\frac{1}{2}$ Coefficient of Variation
Motor	$\gamma\left(25\ ,\ 3.1250\times10^{-7}\right)$	80,000,000	16,000,000	10%
Home	$\gamma\left(25\ ,\ 3.1250\times10^{-7}\right)$	80,000,000	16,000,000	10%
Fire	$\mathrm{W}\left(4.1782\times10^{-30}\ ,\ 3.6965\right)$	80,000,000	24,000,000	15%
Marine	$\mathrm{W}\left(3.4402\times10^{-22}\ ,\ 2.6984\right)$	80,000,000	32,000,000	20%
Other	$\mathrm{IG}\left(80,000,000\ ,\ 3.3541\times10^{-5}\right)$	80,000,000	24,000,000	15%
Workers' Compensation	$\mathrm{LN}\left(18.1233\ ,\ 0.3853\right)$	80,000,000	32,000,000	20%
Liability	$\mathrm{LN}\left(18.1233\ ,\ 0.3853\right)$	80,000,000	32,000,000	20%
Professional Indemnity	$\mathrm{LN}\left(18.0860\ ,\ 0.4724\right)$	80,000,000	40,000,000	25%

Table 6 states our assumptions of an 8×8 matrix of pairwise τ between the total outstanding claims liabilities of the eight lines, referring to the linear correlation matrices as shown in Bateup and Reed (2001) and Collings and White (2001). We assume the two sources of dependency (refer to Section 1) are approximately accommodated in the assumptions. Since the Institute of Actuaries of Australia (IAAust) Guidance Note GN 353 states that caution should be taken when low correlation is assumed, negative τ is not considered here. The first five lines of short-tailed business (with a relatively short period of time to settle all claims) are assumed to be independent of the rest that are long-tailed (with a relatively long period of time to settle all claims). Thus, τ is zero between a short-tailed line and a long-tailed line. In effect, there are two independent sets of multivariate random variables, in which one has five dimensions and the other has three. In practice, pairwise measures of association can be estimated as discussed in Section 4 or determined by judgement. Moreover, we assume the underlying copula is unknown and we attempt to model the liabilities with different copulas.

Table 6 Assumptions of Kendall's Tau

τ	**Motor**	**Home**	**Fire**	**Marine**	**Other**	**WC**	**Liab**	**PI**
Motor	1	0.15	0.1	0.05	0.15	0	0	0
Home	0.15	1	0.15	0.05	0.1	0	0	0
Fire	0.1	0.15	1	0.05	0.1	0	0	0
Marine	0.05	0.05	0.05	1	0.05	0	0	0
Other	0.15	0.1	0.1	0.05	1	0	0	0
WC	0	0	0	0	0	1	0.2	0.15
Liab	0	0	0	0	0	0.2	1	0.2
PI	0	0	0	0	0	0.15	0.2	1

Several copulas discussed previously are applied in this example. Based on the assumed matrix of τ , pairwise *r* is computed for the Gaussian copula and the *t* copula referring to Subsections 5.3 and 5.4. Pairwise θ is also computed for the Cook-Johnson copula and the Archimedean copulas with regard to Subsections 5.5 and 5.6. These figures are shown in Tables 7 and 8. For the Gaussian copula and the *t* copula, pairwise *r* is computed by using $r = \sin(\pi\tau/2)$. The dispersion matrix computed is positive definite and symmetric.

Table 7 *r* of Gaussian Copula and *t* Copula

r	**Motor**	**Home**	**Fire**	**Marine**	**Other**	**WC**	**Liab**	**PI**
Motor	1	0.23	0.16	0.08	0.23	0	0	0
Home	0.23	1	0.23	0.08	0.16	0	0	0
Fire	0.16	0.23	1	0.08	0.16	0	0	0
Marine	0.08	0.08	0.08	1	0.08	0	0	0
Other	0.23	0.16	0.16	0.08	1	0	0	0
WC	0	0	0	0	0	1	0.31	0.23
Liab	0	0	0	0	0	0.31	1	0.31
PI	0	0	0	0	0	0.23	0.31	1

Table 8 θ of Archimedean Copulas

Short-Tailed	θ_1	θ_2	θ_3	θ_4
Clayton	0.1053	0.2642	0.2857	0.3529
Gumbel-Hougaard	1.0526	1.1321	1.1429	1.1765
Frank	0.4509	1.0617	1.1395	1.3752

Long-Tailed	θ_1	θ_2
Clayton	0.4242	0.5000
Gumbel-Hougaard	1.2121	1.2500
Frank	1.6154	1.8609

For the Cook-Johnson copula, the average τ of the short-tailed lines is computed as 0.0950 and the average τ of the long-tailed lines is about 0.1833. These are used as approximations since there is only one parameter for either the short-tailed lines or long-tailed lines. Making use of $\theta = 2\tau/(1-\tau)$, pairwise θ is computed as 0.2099 for the short-tailed lines and as 0.4490 for the long-tailed lines.

For the Archimedean copulas, the five-dimension extension example given in Subsection 5.6 is used here for the short-tailed lines. The order of extension from $X_{(1)}$ to $X_{(5)}$ is selected as Motor, Home, Fire, Other, and Marine, since Marine has the same value of τ with all the other short-tailed lines and this value of τ (and so the corresponding θ) is the smallest. When τ is not the same, an average value of τ is computed, e.g. $\tau^{\text{Motor,Fire}} = 0.1$, $\tau^{\text{Home,Fire}} = 0.15$, and average τ = 0.125. This approximation is required due to partial exchangeability. The way θ is computed from τ depends on whether the copula is Clayton, Gumbel-Hougaard, or Frank, as described in Subsection 5.6. The Nelsen no. 12 copula is not examined in this example as the copula does not allow τ lower than one-third. For the long-tailed lines, a similar three-dimension extension is made with the selected order Workers' Compensation, Liability, and Professional Indemnity.

6.2 Results

Based on the assumptions in the previous subsection, 10,000 samples of the total outstanding claims liability are simulated for each line of business, and are then summed across the eight lines to obtain 10,000 samples of the aggregate portfolio value. The simulation process is carried out by using the selected copula to link the marginals to form a joint multivariate distribution. This process is repeated for each copula in turn.

The outstanding claims liabilities must be assessed at the 75th percentile (the risk margin, when expressed as a percentage of the mean, is subject to a minimum of one half of the coefficient of variation) according to GPS 210. Moreover, the Australian solvency benchmark for probability of ruin is 0.5% on a one-year time horizon. Accordingly, the following figures are estimated from the 10,000 simulated samples of the aggregate portfolio value: one half of the coefficient of variation, the difference (as a percentage of the mean) between the 75th percentile and the mean, and the difference (as a percentage of the mean) between the 99.5th percentile and the mean. Table 10 at the end of this section presents these figures.

Overall, one-half the coefficient of variation and the 75th percentile margin are more or less the same across different copula selections. The main disparities lie in the 99.5th percentile margin. This phenomenon reflects that the effect of exploiting a particular copula becomes obvious when the tail dependence behaviour is concerned.

The analysis of the results regarding the 99.5th percentile margin of the aggregate portfolio value is as follows. The Gumbel-Hougaard copula and the t_3 copula lead to the highest margins (54% and 52%). These two copulas have very strong upper tail dependence, which explains the results. (For the short-tailed lines only, the t_3 copula produces a higher margin than the Gumbel-Hougaard copula, and vice versa for the long-tailed lines.) As v increases from 3 to 200, the results of the t copula converge to those of the Gaussian copula. The Cook-Johnson copula and the Clayton copula are not equivalent for more than two dimensions in this case, but they still produce similarly low margins

(43% and 42%) due to some similarity between their structures. Having no upper tail dependence, the Frank copula and the Gaussian copula lead to similar results (both around 47%). (Their results are very similar for the short-tailed lines.) The largest difference is around 12% between the Gumbel-Hougaard copula and the Clayton copula.

Figures 14 to 18 at the end of this section show the extreme percentiles of the aggregate portfolio value for different copula selections. The lighter lines in these figures are for the Gaussian copula and the heavier lines are for the other copulas. In all the figures, the cdf graph of the Gaussian copula is plotted against those of the other copulas as a comparison. As shown in the figures, the t_3 copula and the Gumbel-Hougaard copula lead to larger percentiles (above around the 93rd percentile) than the Gaussian copula. On the contrary, the Cook-Johnson copula and the Clayton copula lead to smaller percentiles than the Gaussian copula. The cdf graphs of the Gaussian copula and the Frank copula are close to each other. All these features correspond to the previous analysis of the 99.5th percentile margin of the aggregate portfolio value.

These varying results highlight the importance of choosing an appropriate copula when an assessment, such as probability of ruin, involves the computation of an extreme percentile. An understatement of the underlying upper tail dependence due to an unsuitable copula choice would lead to an underestimation of the high-percentile margins.

6.3 Further Results

To stress-test the copulas in the last example, we now enhance each pairwise τ by 100%. The following figures are computed after the increase in τ.

Table 9 Parameters after τ Increase

Cook-Johnson θ	Short-Tailed : 0.4691	Long-Tailed : 1.1579

τ	Motor	Home	Fire	Marine	Other	WC	Liab	PI
Motor	1	0.3	0.2	0.1	0.3	0	0	0
Home	0.3	1	0.3	0.1	0.2	0	0	0
Fire	0.2	0.3	1	0.1	0.2	0	0	0
Marine	0.1	0.1	0.1	1	0.1	0	0	0
Other	0.3	0.2	0.2	0.1	1	0	0	0
WC	0	0	0	0	0	1	0.4	0.3
Liab	0	0	0	0	0	0.4	1	0.4
PI	0	0	0	0	0	0.3	0.4	1

r	Motor	Home	Fire	Marine	Other	WC	Liab	PI
Motor	1	0.45	0.31	0.16	0.45	0	0	0
Home	0.45	1	0.45	0.16	0.31	0	0	0
Fire	0.31	0.45	1	0.16	0.31	0	0	0
Marine	0.16	0.16	0.16	1	0.16	0	0	0
Other	0.45	0.31	0.31	0.16	1	0	0	0
WC	0	0	0	0	0	1	0.59	0.45
Liab	0	0	0	0	0	0.59	1	0.59
PI	0	0	0	0	0	0.45	0.59	1

Short-Tailed	θ_1	θ_2	θ_3	θ_4
Clayton	0.2222	0.6087	0.6667	0.8571
Gumbel-Hougaard	1.1111	1.3043	1.3333	1.4286
Frank	0.9074	2.1982	2.3719	2.9174

Long-Tailed	θ_1	θ_2
Clayton	1.0769	1.3333
Gumbel-Hougaard	1.5385	1.6667
Frank	3.5088	4.1611

Table 11 shows the margins after the increase in τ. These margins are generally higher than those in Table 10, as the association between each pair of lines of business is 100% higher in terms of τ. The increase in the 99.5th percentile margin of the aggregate portfolio value ranges from 3% for the Clayton copula to around 8% for the Gaussian copula and the *t* copula. The effect of the increase in τ is more prominent for the long-tailed lines, in which the increase in the 99.5th percentile margin ranges from 9% for the Cook-Johnson copula to around 19% for the *t* copula. It can be seen that the margin levels are fairly sensitive to the τ assumptions, particularly for the long-tailed lines with the copulas that possess upper tail dependence. As such, apart from choosing a suitable copula structure, proper measurement of τ (and also proper determination of v for the *t* copula) is also essential in computing the extreme percentiles.

Furthermore, the patterns across different copula selections are similar to those before the increase in τ. The t_3 copula and the Gumbel-Hougaard copula lead to the highest 99.5th percentile margins of the aggregate portfolio value, the Cook-Johnson copula and the Clayton copula lead to the lowest margins, and the Gaussian copula and the Frank copula have their margins in between. In particular, it appears that the Gumbel-Hougaard copula leads to a very skewed aggregate liability distribution for the long-tailed lines.

Figures 19 to 23 show the extreme percentiles of the aggregate portfolio value for different copula selections after the increase in τ. As shown in the figures, the t_3 copula and the Gumbel-Hougaard copula lead to larger percentiles (above around the 95th percentile) than the Gaussian copula. On the contrary, the Cook-Johnson copula, the Clayton copula, and the Frank copula lead to smaller percentiles than the Gaussian copula. All these features correspond to the results presented in Table 11.

Table 10 One Half of Coefficient of Variation, 75th Percentile Margin, and 99.5th Percentile Margin

Whole Portfolio	**Gaussian**	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	**Cook-Johnson**	**Clayton**	**Gumbel-Hougaard**	**Frank**
0.5CV	8%	8%	8%	8%	8%	8%	8%	8%	8%	7%	8%	8%
75% Margin	10%	9%	9%	9%	9%	10%	10%	10%	10%	10%	9%	9%
99.5% Margin	47%	52%	50%	50%	49%	48%	49%	48%	43%	42%	54%	47%

Short-Tailed Lines	**Gaussian**	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	**Cook-Johnson**	**Clayton**	**Gumbel-Hougaard**	**Frank**
0.5CV	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%	8%
75% Margin	11%	9%	10%	10%	10%	11%	11%	11%	11%	10%	10%	10%
99.5% Margin	43%	52%	49%	47%	46%	44%	43%	43%	40%	39%	48%	43%

Long-Tailed Lines	**Gaussian**	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	**Cook-Johnson**	**Clayton**	**Gumbel-Hougaard**	**Frank**
0.5CV	16%	16%	16%	16%	16%	16%	16%	16%	15%	15%	16%	15%
75% Margin	17%	16%	17%	17%	17%	17%	17%	17%	18%	18%	15%	18%
99.5% Margin	114%	120%	117%	118%	118%	115%	115%	115%	95%	88%	129%	100%

Table 11 One Half of Coefficient of Variation, 75th Percentile Margin, and 99.5th Percentile Margin after τ Increase

Whole Portfolio	**Gaussian**	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	**Cook-Johnson**	**Clayton**	**Gumbel-Hougaard**	**Frank**
0.5CV	9%	9%	9%	9%	9%	9%	9%	9%	9%	8%	9%	9%
75% Margin	11%	10%	11%	11%	11%	11%	11%	11%	11%	10%	10%	11%
99.5% Margin	55%	59%	57%	57%	57%	57%	56%	55%	48%	45%	58%	51%

Short-Tailed Lines	**Gaussian**	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	**Cook-Johnson**	**Clayton**	**Gumbel-Hougaard**	**Frank**
0.5CV	9%	9%	9%	9%	9%	9%	9%	9%	9%	8%	9%	9%
75% Margin	12%	11%	11%	11%	12%	12%	12%	12%	13%	12%	11%	12%
99.5% Margin	49%	57%	56%	54%	52%	50%	50%	49%	44%	39%	55%	46%

Long-Tailed Lines	**Gaussian**	t_3	t_5	t_8	t_{10}	t_{50}	t_{100}	t_{200}	**Cook-Johnson**	**Clayton**	**Gumbel-Hougaard**	**Frank**
0.5CV	18%	18%	18%	18%	18%	18%	18%	18%	17%	17%	19%	17%
75% Margin	19%	18%	19%	19%	19%	19%	19%	19%	22%	21%	17%	21%
99.5% Margin	127%	137%	138%	135%	136%	135%	134%	134%	104%	98%	147%	115%

Figure 14 **CDF – Gaussian Copula vs t_3 Copula**

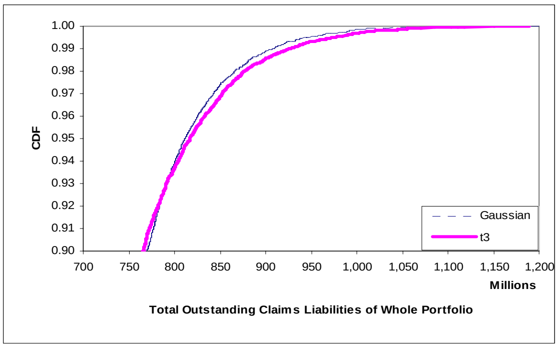


Figure 15 **CDF – Gaussian Copula vs Cook-Johnson Copula**

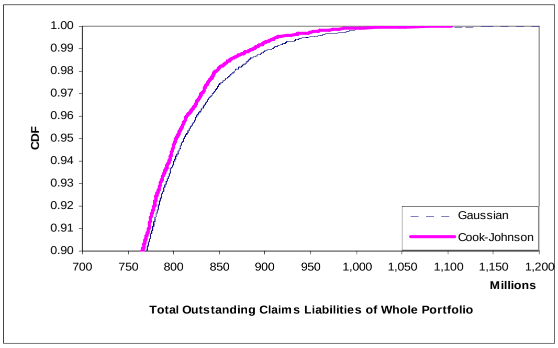


Figure 16 **CDF – Gaussian Copula vs Clayton Copula**

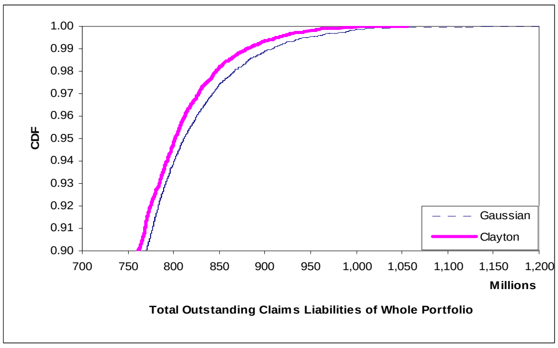


Figure 17 **CDF – Gaussian Copula vs Gumbel-Hougaard Copula**

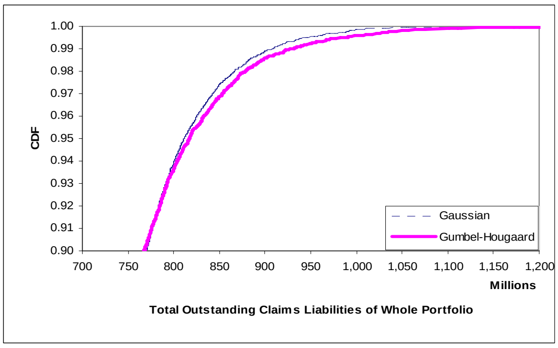


Figure 18 **CDF – Gaussian Copula vs Frank Copula**

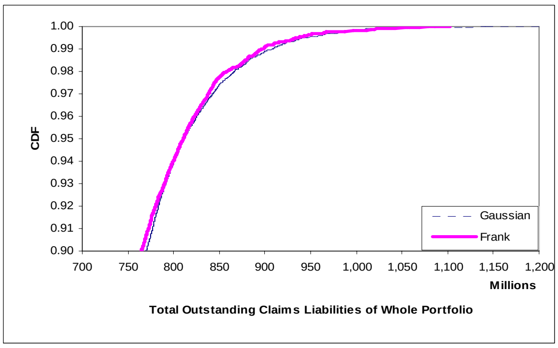


Figure 19 **CDF after Increase – Gaussian Copula vs t_3 Copula**

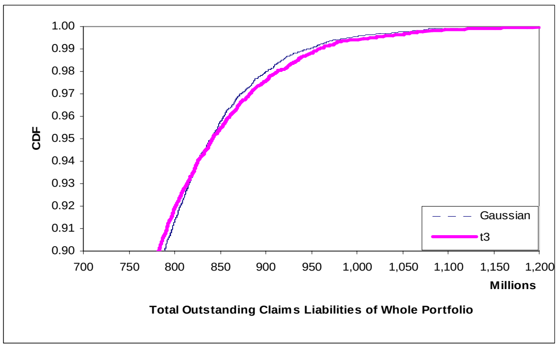


Figure 20 CDF after Increase – Gaussian Copula vs Cook-Johnson Copula

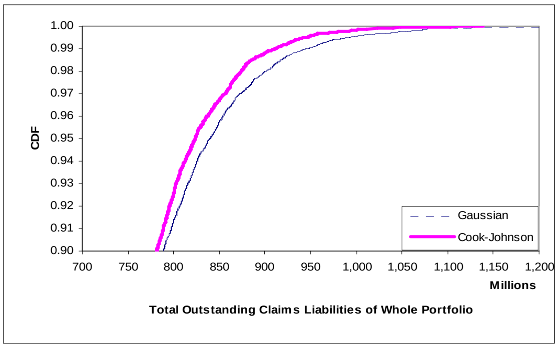


Figure 21 CDF after Increase – Gaussian Copula vs Clayton Copula

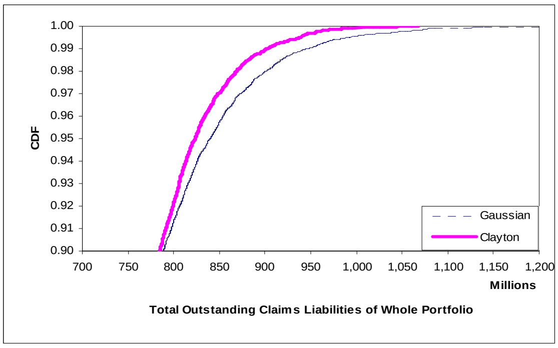


Figure 22 CDF after Increase – Gaussian Copula vs Gumbel-Hougaard Copula

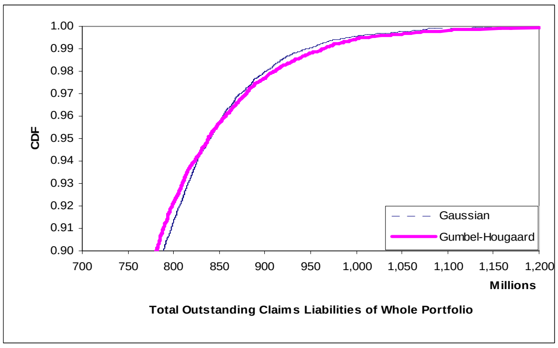
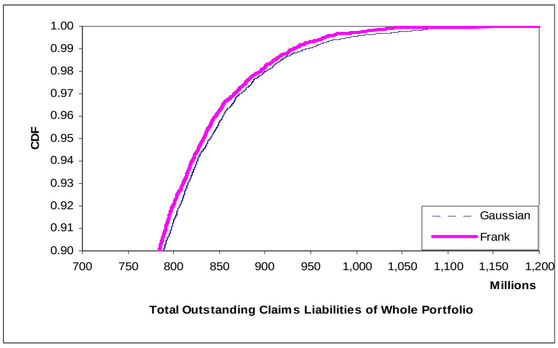


Figure 23 CDF after Increase – Gaussian Copula vs Frank Copula



7. FURTHER APPLICAITON

Copulas can also be used to model smaller components of general insurance liabilities. In this section, we propose a few such examples regarding the number of claims and the claim amounts. We consider the total liability of line i (a total of n lines) as $TL_{(i)} = \sum_{j=1}^{N_{(i)}} X_{(i),j}$, where $N_{(i)}$ is the number of claims and $X_{(i),j}$ is the amount of the jth claim.

This model is the traditional collective risk model, which is useful for short-term contracts and is discussed in Dickson and Waters (1992). The number of claims $N_{(i)}$ can be assumed to follow a distribution of Poisson, binomial, or negative binomial, and $X_{(i),j}$ can be assumed to follow a distribution of exponential, gamma, Weibull, inverse Gaussian, lognormal, or Pareto (refer to Appendix I).

7.1 Linking Claim Counts

Traditionally, the association between $N_{(i)}$'s of different lines is modelled by incorporating a common parameter. The following is an example given by Wang (1998): $N_{(i)} \sim \text{Pn}\left(\lambda_{(i)}\theta\right)$, $\theta \sim \gamma(\alpha, \beta)$, $N_{(i)}$ and $X_{(i),j}$ are independent for any given i, and $X_{(i),j}$'s are independent.

In this example, $\lambda_{(i)}$ is unique for each line of business, while θ represents a common driving force for $N_{(i)}$'s of different lines. When θ is large, there is a higher chance of a larger number of claims for all lines of business, e.g. a windstorm affects the number of claims for both motor and home insurance.

Alternatively, the relationship between $N_{(i)}$'s of different lines can be modelled by fitting a copula. The advantage of copula fitting is its flexibility in dealing with various levels and shapes of association between multivariate random variables. To exemplify, instead of employing a common parameter θ as described above, $\theta_{(i)}$'s of different lines can be modelled to be associated by a particular copula structure, in which $N_{(i)} \sim \text{Pn}\left(\lambda_{(i)}\theta_{(i)}\right)$ and $\theta_{(i)} \sim \gamma\left(\alpha_{(i)}, \beta_{(i)}\right)$, with $\lambda_{(i)}$, $\theta_{(i)}$, $\alpha_{(i)}$, and $\beta_{(i)}$ being unique for each line of business.

7.2 Linking Claim Counts and Claim Amounts

Sometimes, $N_{(i)}$ and $X_{(i),j}$ are linked by a common parameter for a line of business, as shown in the example below:

$$N_{(i)} \sim \text{Pn}\left(\lambda_{(i)}\theta_{(i)}\right),\ X_{(i),j} \sim \text{LN}\left(g_{(i)}\left(\mu_{(i)}, \theta_{(i)}\right),\ \sigma_{(i)}\right),\text{ and } \theta_{(i)} \sim \gamma\left(\alpha_{(i)}, \beta_{(i)}\right).$$

In this example, $\lambda_{(i)}$, $\theta_{(i)}$, $\mu_{(i)}$, $\sigma_{(i)}$, $g_{(i)}$, $\alpha_{(i)}$, and $\beta_{(i)}$ are unique for each line of business and $g_{(i)}$ is an increasing function of $\mu_{(i)}$ and $\theta_{(i)}$. As such, $\theta_{(i)}$ represents a common driving force between $N_{(i)}$ and $X_{(i),j}$. When $\theta_{(i)}$ is large, both $N_{(i)}$ and $X_{(i),j}$ have a higher chance to become large, e.g. a cyclone impinges on both the number of claims and the claim amounts for crop insurance.

Again, the relationship between $\theta_{(i)}$'s of different lines can be modelled by a selected copula structure. In this way, $N_{(i)}$'s and $X_{(i),j}$'s of different lines are linked via the chosen copula between $\theta_{(i)}$'s.

7.3 Application Example

We set up a hypothetical example of two lines of business, using the structure mentioned in the previous subsection. The following assumptions are made arbitrarily:

$$\theta_{(1)} \sim \gamma(50,1),\ \theta_{(2)} \sim \gamma(50,1),\ N_{(1)} \sim \text{Pn}\left(\theta_{(1)}\right),\ N_{(2)} \sim \text{Pn}\left(1.2\theta_{(2)}\right),$$

$$X_{(1),j} \sim \text{LN}\left(6 + 0.02\theta_{(1)}, 0.2\right),\text{ and } X_{(2),j} \sim \text{LN}\left(6 + 0.026\theta_{(2)}, 0.2\right).$$

Suppose the underlying copula is unknown. We associate $\theta_{(1)}$ and $\theta_{(2)}$ with different copulas including the Gaussian copula, the t_3 copula, and the four Archimedean copulas discussed. Accordingly, 30,000 samples of $TL_{(1)} + TL_{(2)} = \sum_{j=1}^{N_{(1)}} X_{(1),j} + \sum_{j=1}^{N_{(2)}} X_{(2),j}$ are simulated for each type of copula considered and for τ equal to 0, 0.2, 0.4, 0.6, 0.8, and 1. For those copulas becoming intractable when τ is equal to 0 or 1, approximation is made as τ equal to say 0.01 or 0.99. The coefficient of variation, margins of various percentiles, and coefficient of skewness are sampled from the simulated figures. The results are shown in Tables 12 to 18 at the end of this section.

Overall, as shown in Tables 12 to 17, one-half the coefficient of variation and those margins below the 99th percentile are more or less the same across different copula selections for different values of τ. The main differences are shown in the 99th percentile margin and the 99.5th percentile margin. In general, the t_3 copula and the Gumbel-Hougaard copula lead to the highest 99th percentile margins and the highest 99.5th percentile margins, as these two copulas possess upper tail dependence. The largest difference is around 19% between the 99.5th percentile margin of the Gumbel-Hougaard copula and that of the Clayton copula at τ equal to 0.4.

In addition, at τ equal to 1, there is not much disparity between the figures of different copulas. As $\theta_{(1)}$ and $\theta_{(2)}$ become comonotonic at this value of τ, no matter which copula is chosen, the overall structure is reduced to the following, which is effectively the traditional way of using a common parameter as mentioned in Subsection 7.1:

$\theta \sim \gamma(50,1)$, $N_{(1)} \sim \text{Pn}(\theta)$, $N_{(2)} \sim \text{Pn}(1.2\theta)$,

$X_{(1),j} \sim \text{LN}(6 + 0.02\theta, 0.2)$, and $X_{(2),j} \sim \text{LN}(6 + 0.026\theta, 0.2)$.

Table 18 shows that the t_3 copula and the Gumbel-Hougaard copula lead to a larger coefficient of skewness than the other copulas in general. Again, this skewness corresponds to the tail dependence property. Moreover, Figures 24 to 29 exhibit the pdfs of $TL_{(1)} + TL_{(2)}$. For all copulas, as τ increases, the association between $\theta_{(1)}$ and $\theta_{(2)}$ increases, and the pdf has a heavier right tail.

Table 12 One Half of Coefficient of Variation and Various Margins ($\tau = 0$)

$\tau = 0$	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.5CV	13%	13%	13%	13%	13%	–
75% Margin	14%	14%	15%	14%	15%	–
80% Margin	20%	18%	20%	19%	20%	–
85% Margin	26%	24%	26%	26%	26%	–
90% Margin	34%	33%	34%	34%	34%	–
95% Margin	47%	47%	47%	47%	48%	–
99% Margin	77%	82%	77%	77%	78%	–
99.5% Margin	91%	97%	90%	90%	91%	–

Table 13 One Half of Coefficient of Variation and Various Margins ($\tau = 0.2$)

$\tau = 0.2$	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.5CV	14%	14%	14%	14%	14%	–
75% Margin	16%	15%	16%	15%	16%	–
80% Margin	21%	20%	21%	21%	22%	–
85% Margin	28%	27%	28%	28%	28%	–
90% Margin	38%	36%	37%	37%	38%	–
95% Margin	53%	52%	50%	53%	52%	–
99% Margin	87%	92%	80%	91%	82%	–
99.5% Margin	102%	109%	94%	108%	97%	–

Table 14 One Half of Coefficient of Variation and Various Margins ($\tau = 0.4$)

$\tau = 0.4$	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.5CV	15%	15%	15%	16%	15%	15%
75% Margin	17%	16%	18%	16%	18%	17%
80% Margin	23%	22%	23%	22%	24%	23%
85% Margin	30%	29%	30%	30%	31%	30%
90% Margin	40%	39%	39%	40%	41%	39%
95% Margin	57%	57%	53%	58%	56%	54%
99% Margin	95%	98%	84%	100%	87%	88%
99.5% Margin	111%	115%	98%	117%	102%	104%

Table 15 One Half of Coefficient of Variation and Various Margins ($\tau = 0.6$)

$\tau = 0.6$	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.5CV	16%	16%	16%	16%	16%	16%
75% Margin	17%	17%	19%	17%	18%	17%
80% Margin	24%	24%	25%	23%	25%	23%
85% Margin	32%	31%	32%	31%	33%	31%
90% Margin	43%	42%	42%	43%	43%	42%
95% Margin	60%	60%	57%	60%	59%	59%
99% Margin	101%	102%	89%	103%	92%	99%
99.5% Margin	118%	120%	103%	120%	107%	117%

Table 16 One Half of Coefficient of Variation and Various Margins ($\tau = 0.8$)

$\tau = 0.8$	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.5CV	17%	17%	16%	17%	16%	17%
75% Margin	18%	18%	19%	17%	18%	17%
80% Margin	24%	24%	25%	24%	25%	24%
85% Margin	33%	32%	34%	33%	33%	32%
90% Margin	44%	44%	44%	44%	45%	44%
95% Margin	63%	63%	60%	62%	62%	61%
99% Margin	105%	105%	94%	105%	98%	104%
99.5% Margin	122%	122%	108%	122%	112%	121%

Table 17 One Half of Coefficient of Variation and Various Margins ($\tau = 1$)

$\tau = 1$	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.5CV	17%	17%	17%	17%	17%	17%
75% Margin	18%	18%	18%	18%	18%	18%
80% Margin	25%	25%	24%	24%	24%	24%
85% Margin	33%	33%	33%	33%	33%	33%
90% Margin	44%	44%	44%	44%	45%	44%
95% Margin	63%	63%	63%	62%	63%	62%
99% Margin	106%	105%	104%	105%	103%	105%
99.5% Margin	125%	124%	119%	122%	120%	122%

Table 18 Coefficient of Skewness

τ	Gaussian	t_3	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.0	0.90	1.04	0.88	0.89	0.89	–
0.2	0.94	1.07	0.74	1.16	0.87	–
0.4	1.00	1.09	0.70	1.21	0.87	0.84
0.6	1.06	1.10	0.73	1.17	0.89	1.06
0.8	1.10	1.11	0.85	1.13	0.95	1.11
1.0	1.11	1.10	1.04	1.11	1.04	1.11

Figure 24 PDF – Gaussian Copula

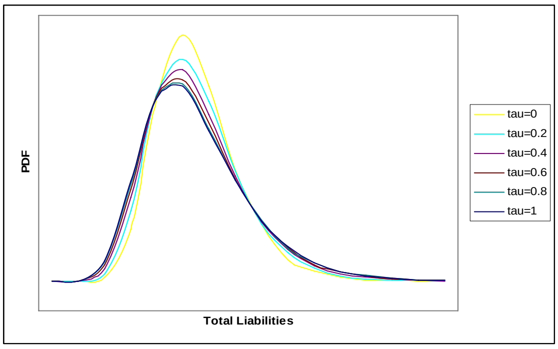


Figure 25 PDF – t_3 Copula

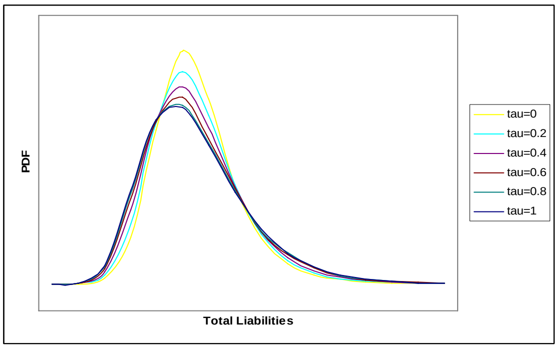


Figure 26 PDF – Clayton Copula

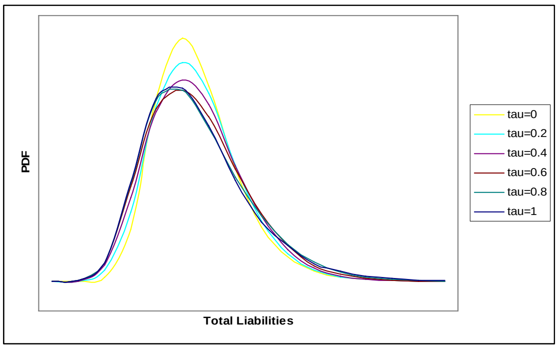


Figure 27 PDF – Gumbel-Hougaard Copula

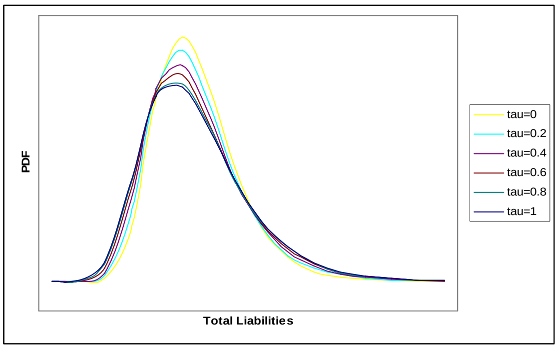


Figure 28 PDF – Frank Copula

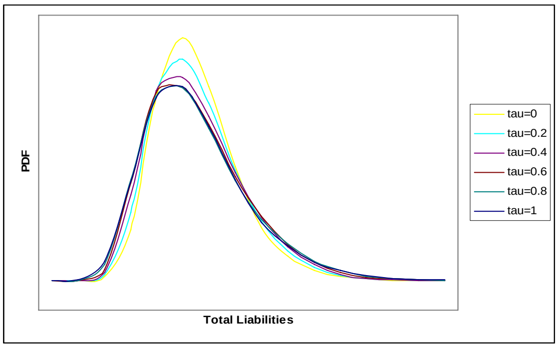
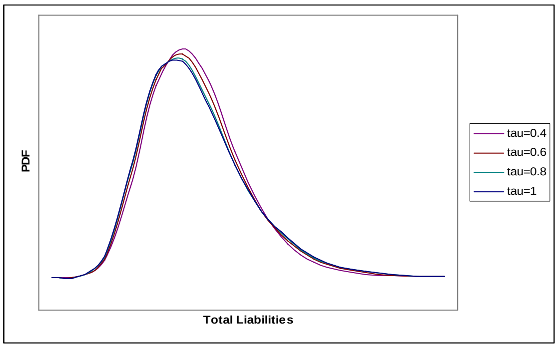


Figure 29 PDF – Nelsen No. 12 Copula



8. DISCUSSION

From the previous sections, we can see that the concept of copula offers a clear view of multivariate distributions. For each set of continuous multivariate random variables, there is a unique copula function that represents a particular dependency structure between the random variables and links the marginal univariate distributions to form a joint multivariate distribution. With the ability to separate the consideration of the marginals and the dependency structure, an actuary can readily blend any right-skewed distributions with the various forms of copula, based on analysis of the past claims data and on subjective judgement. As it is widely accepted that most general insurance liabilities have right-skewed distributions, we find that copula modelling is a versatile mathematical tool to link the liabilities of varying characteristics to form a picture of the aggregate portfolio for an insurer.

Several useful pairwise measures of association such as Kendall's Tau and Spearman's Rho, which have more desirable properties than Pearson's correlation coefficient, can be readily estimated with the help of Excel spreadsheets and VBA coding. Moreover, the Gaussian copula, the *t* copula, and the Cook-Johnson copula (and some Archimedean copulas with a small number of dimensions) can be readily implemented on the Excel platform. Copula modelling is hence highly feasible in practice, especially for DFA and assessment of risk concentration. While only the extreme percentiles of the aggregate portfolio value are affected in an apparent way by the choice of copula, as demonstrated in the hypothetical examples in Sections 6 and 7, it is often a good starting point to model the marginals and the dependency structure separately and appropriately, from which more proper inferences can be drawn.

In addition, the copula modelling techniques discussed in this chapter can be extended to much broader applications. For example, the relationships between the premium liabilities of different lines, the association between the premium liabilities and the outstanding claims liabilities, the dependency structure between individual claim sizes of different lines, and the co-movements between and within the assets and the liabilities can be modelled with the variety of copula functions available. Copulas are particularly useful for modelling dependent severe events in reinsurance.

Nevertheless, there are three problems of using a copula in general insurance practice. First, claims data of an insurer are usually insufficient for computing statistically significant estimates of association measures. Certain industry reports or professional judgement is often required to determine the levels of association. Second, other dependency cannot be readily quantified and its assessment remains largely judgemental and subjective. Finally, the algorithm for simulating multivariate random variables linked by an Archimedean copula (when the dimension is greater than two) is extremely tedious and software such as Mathematica is needed to carry out the simulation. This problem hinders the use of Archimedean copulas, though being versatile and flexible, from becoming more popular amongst practitioners.

9. BIBLIOGRAPHY

Australian Prudential Regulation Authority (ARPA 2001). *Prudential Standard GPS 210 and Guidance Note GGN 210.1*.

Bateup R. and Reed I. (2001). *Research and data analysis relevant to the development of standards and guidelines on liability valuation for general insurance*. The Institute of Actuaries of Australia and Tillinghast – Towers Perrin.

Clayton D. G. (1978). *A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence*. Biometrika 65: 141:151.

Clemen R. T. and Reilly T. (1999). *Correlations and copulas for decision and risk analysis*. Institute for Operations Research and the Management Sciences 45(2).

Collings S. and White G. (2001). *APRA risk margin analysis*. The Institute of Actuaries of Australia XIIIth General Insurance Seminar, Trowbridge Consulting.

Cook R. D. and Johnson M. E. (1981). *A family of distributions for modelling non-elliptically symmetric multivariate data*. Journal of the Royal Statistical Society B43: 210-218.

De Matteis R. (2001). *Fitting copulas to data*. Institute of Mathematics, University of Zurich.

Demarta S. and McNeil A. (2004). *The t copula and related copulas*. Department of Mathematics, ETH Zurich.

Dickson D. C. M. and Waters H. R. (1992). *Risk models*. Department of Actuarial Mathematics and Statistics, Heriot-Watt University.

Embrechts P., Lindskog F., and McNeil A. (2001). *Modelling dependence with copulas and applications to risk management*. Department of Mathematics, ETH Zurich.

Frank M. J. (1979). *On the simultaneous associativity of* $F(x,y)$ *and* $x + y - F(x,y)$. Aequationes Math. 19: 194-226.

Frees E. W. and Valdez E. A. (1998). *Understanding relationships using copulas*. North American Actuarial Journal 2(1).

Guégan D. and Ladoucette S. (2004). *Dependence modelling of the joint extremes in a portfolio using Archimedean copulas: application to MSCI indices*. Institutions et Dynamiques Historiques de l'Economie (IDHE).

Gumbel E. J. (1960). *Distributions des valeurs extremes en plusiers dimensions*. Publ. Inst. Statist. Univ. Paris 9: 171-173.

Hougaard P. (1986). *A class of multivariate failure time distributions*. Biometrika 73: 671-678.

Joe H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall.

Kendall M. and Gibbons J. D. (1990). *Rank correlation methods*. Edward Arnold, Hodder & Stoughton.

Kupiec P. H. (1995). *Techniques for verifying the accuracy of risk measurement models*. Journal of Derivatives 3: 73-84.

Li J. (2006). *Comparison of stochastic reserving methods*. Working Paper No. 131, Centre for Actuarial Studies, The University of Melbourne.

Lindskog F. (2000). *Modelling dependence with copulas and applications to risk management*. RiskLab, ETH Zurich.

Mack T. (1993). *Distribution-free calculation of the standard error of chain ladder reserve estimates*. ASTIN Bulletin 23(2): 213-225.

Melchiori M. R. (2003). *Which Archimedean copula is the right one?* YieldCurve.com e-Journal, Universidad Nacional del Litoral, Argentina.

Micocci M. and Masala G. (2004). *Backtesting value-at-risk estimation with non Gaussian marginals*. Faculty of Economics, University of Cagliari, Italy.

Nelsen R. B. (1999). *An introduction to copulas*. Springer-Verlag.

Priest C. (2003). *Correlations: what they mean and more importantly what they don't mean*. The Institute of Actuaries of Australia Biennial Convention 2003.

Sklar A. (1959). *Fonctions de répartition a n dimensions et leurs marges*. Publ. Inst. Statist. Univ. Paris 8: 229-231.

Taylor G. and McGuire G. (2005). *Bootstrapping of loss reserves*. The Institute of Actuaries of Australia Research Conference – New Developments in the Quantitative Modelling of Insurance and Financial Risk.

The Institute of Actuaries of Australia (IAAust 2002). *Professional Standard PS 300 and Guidance Note GN 353*.

Ventor G. G. (2003). *Quantifying correlated reinsurance exposures with copulas*. Casualty Actuarial Society Forum.

Wang S. S. (1998). *Aggregation of correlated risk portfolios: models and algorithms*. PCAS Volume LXXXV.

Watson R. K. (1983). *Elementary mathematical statistics*. Vaba Publishing.

Wolfram S. (1999). *The Mathematica book*. Cambridge University Press.

APPENDIX I

This appendix lists some common univariate and multivariate probability distributions for modelling general insurance liabilities. Suppose X is a continuous random variable and $\vec{X}$ is a vector of continuous multivariate random variables. Let $f_X(x)$ be the pdf of X and $f_{\vec{X}}(\vec{x})$ be the multivariate pdf of $\vec{X}$.

The following univariate probability distributions are common for modelling individual claim amounts or the total liability of a line of business. Wang (1998) suggests that the order of increasing heaviness of the right tail is gamma, Weibull, inverse Gaussian, lognormal, and Pareto. Currie (1993) depicts some useful distribution fitting techniques including moment matching, percentile matching, and maximum likelihood. Further references include Watson (1983) and Seshadri (1999).

Exponential $X \sim \exp(\lambda)$: $f_X(x) = \lambda \exp(-\lambda x)$

Gamma $X \sim \gamma(\alpha, \lambda)$: $f_X(x) = \lambda^{\alpha} x^{\alpha-1} \exp(-\lambda x) / \Gamma(\alpha)$

Weibull $X \sim \mathrm{W}(c, \gamma)$: $f_X(x) = c\gamma\, x^{\gamma-1} \exp(-cx^{\gamma})$

Inverse Gaussian $X \sim \mathrm{IG}(\mu, \sigma)$: $f_X(x) = \left(\sigma\sqrt{2\pi}\right)^{-1} x^{-3/2} \exp\left(-(x-\mu)^2 / \left(2x\mu^2\sigma^2\right)\right)$

Lognormal $X \sim \mathrm{LN}(\mu, \sigma)$: $f_X(x) = \left(x\sigma\sqrt{2\pi}\right)^{-1} \exp\left(-(\ln x - \mu)^2 / 2\sigma^2\right)$

Pareto $X \sim \mathrm{Pa}(\alpha, \lambda)$: $f_X(x) = \alpha\lambda^{\alpha} / (\lambda + x)^{\alpha+1}$

The following are two multivariate probability distributions. Relevant references include Kotz et al (2000) and Kotz and Nadarajah (2004).

Multinormal $\vec{X} \sim \mathrm{N_m}\left(\vec{\mu}, \vec{D}\right)$: $f_{\vec{X}}(\vec{x}) = \dfrac{1}{(2\pi)^{\frac{m}{2}} \sqrt{\left|\vec{D}\right|}} \exp\left(-\dfrac{1}{2}(\vec{x} - \vec{\mu})' \vec{D}^{-1} (\vec{x} - \vec{\mu})\right)$

Multivariate t $\vec{X} \sim t_{v:m}\left(\vec{\mu}, \vec{R}\right)$: $f_{\vec{X}}\left(\vec{x}\right) = \dfrac{\Gamma\left(\dfrac{v+m}{2}\right)}{\left(\pi v\right)^{\frac{m}{2}} \Gamma\left(\dfrac{v}{2}\right)\sqrt{\left|\vec{R}\right|}} \left(1 + \dfrac{1}{v}\left(\vec{x} - \vec{\mu}\right)' \vec{R}^{-1}\left(\vec{x} - \vec{\mu}\right)\right)^{-\frac{v+m}{2}}$

Now suppose X is a discrete random variable and $f_X\left(x\right)$ represents the pmf (probability mass function) of X. The following three distributions are common for modelling the number of claims, which are discussed in Dickson and Waters (1992), Wang (1998), and Dickson (2005).

Poisson $X \sim \mathrm{Pn}\left(\lambda\right)$: $f_X\left(x\right) = \exp\left(-\lambda\right)\lambda^x / x!$

Binomial $X \sim \mathrm{Bi}\left(n, p\right)$: $f_X\left(x\right) = \binom{n}{x} p^x \left(1 - p\right)^{n-x}$

Negative Binomial $X \sim \mathrm{NB}\left(\alpha, p\right)$: $f_X\left(x\right) = \binom{\alpha + x - 1}{x} p^\alpha \left(1 - p\right)^x$

APPENDIX II

This appendix sets out some basic means of exploiting Excel and Mathematica to carry out simulation with the selected copula and probability distributions. In general, the notation used here has the same meanings as stated previously.

Excel Functions

Given the function RAND( ) generates a random variable $U \sim \mathrm{U}(0,1)$, the following table demonstrates some expedient use of spreadsheet functions.

Table A1 Excel Functions

Excel Function	**Use**
NORMSINV(U)	Generate a random variable $X \sim \mathrm{N}(0,1)$.
r * $X_{(1)}$+SQRT(1–r^2)* $X_{(2)}$	Generate a random variable $X \sim \mathrm{N}(0,1)$ that is correlated with $X_{(1)}$, in which $X_{(1)}$ and $X_{(2)}$ are independent $\mathrm{N}(0,1)$ random variables and r is the correlation coefficient.
NORMSDIST(x)	Compute $\Phi(x)$.
CHIINV(1–U , v)	Generate a random variable $X \sim \chi_v^2$.
IF(x < 0,TDIST(–x , v ,1),1–TDIST(x , v ,1))	Compute $t_v(x)$.
–LN(1–U)/λ	Generate a random variable $X \sim \exp(\lambda)$.
GAMMAINV(U , α , 1)/λ	Generate a random variable $X \sim \gamma(\alpha, \lambda)$.
(–LN(1–U)/c)^(1/γ)	Generate a random variable $X \sim \mathrm{W}(c, \gamma)$.

NORMSDIST(1/σ/SQRT(X)*(X/μ–1))+EXP(2/μ/σ^2)*NORMSDIST(–1/σ/SQRT(X)*(X/μ+1))–U	Generate a random variable $X \sim \text{IG}(\mu,\sigma)$ by using Goal Seek to change X to set the function to zero.
LOGINV(U, μ, σ)	Generate a random variable $X \sim \text{LN}(\mu,\sigma)$.
λ/(1–U)^(1/α)–λ	Generate a random variable $X \sim \text{Pa}(\alpha,\lambda)$.
2*TDIST(SQRT((v+1)*(1–r)/(1+r)),v+1,1)	Compute λ_{upper} of the *t* copula.
FINV(1–probability, d_1, d_2)	Compute the inverse of the cdf of F_{d_1,d_2} based on the probability provided.
EXP(GAMMALN(α))	Compute $\Gamma(\alpha)$.
MDETERM(matrix) ; MINVERSE(matrix)	Compute the determinant and the inverse of a square matrix.
FACT(k)	Compute $k!$.

For generating a univariate random variable from one of the continuous probability distributions listed in the previous appendix, U is simply sampled from using RAND( ). For generating bivariate or multivariate random variables linked by a particular copula, as described in Subsections 5.2 to 5.6, the $V_{(i)}$'s required are sampled in a specific way according to the type of copula selected.

Simulation of Bivariate or Multivariate Random Variables

For generating bivariate random variables with an Archimedean copula, K_C^{-1} (the inverse of K_C) can be obtained by numerical root finding if it cannot be solved analytically. To cite the Clayton copula as an example, the following VBA code can be used.

```
t = 0
diff = t – (t ^ (p + 1) – t) / p – u
Do Until Abs(diff) < 0.000000000000001 And t > 0 And t < 1
s = 1 – ((p + 1) * t ^ p – 1) / p
t = t – diff / s
```

```
diff = t – (t ^ (p + 1) – t) / p – u
Loop
```

In the coding above, u is a sample of a random variable $U \sim \mathrm{U}(0,1)$, p is θ, diff is the difference between $K_C(t)$ and u, and s is $\frac{d}{dt}K_C(t)$. The initial value of t is set to zero, and it is updated repeatedly in the loop until the magnitude of diff is less than 10^{-15}. The final value of t is close to $K_C^{-1}(U)$. Alternatively, Goal Seek can be used to find $K_C^{-1}(U)$ by setting diff to zero.

Numerical root finding can also be done on the Mathematica platform, as shown in the code below, in which k represents $K_C(t)$.

```
u = Random[ ] ; k = t – (t ^ (p + 1) – t) / p ; FindRoot[k = = u , {t , 0}]
```

These numerical methods can be applied similarly to the other Archimedean copulas.

The means of extension to more than two dimensions for the Archimedean copulas discussed in Subsection 5.6 can be implemented with Mathematica. The code below is an example of the Clayton copula extended to three dimensions.

```
Array[q , 3] ; Array[p , 3] ; Array[c , 3] ; Array[d , 3] ; Array[u , 3] ; s = “ ” ;
c[1] = u[1] ; c[2] = (c[1] ^ (–p[2]) + u[2] ^ (–p[2]) – 1) ^ (–1 / p[2]) ;
c[3] = (c[2] ^ (–p[1]) + u[3] ^ (–p[1]) – 1) ^ (–1 / p[1]) ;
d[1] = c[1] ; d[2] = D[c[2] , u[1]] / D[c[1] , u[1]] ;
d[3] = D[c[3] , u[1] , u[2]] / D[c[2] , u[1] , u[2]] ;
Do[(q[1] = Random[ ] ; q[2] = Random[ ] ; q[3] = Random[ ] ; u[1] = q[1] ;
temp = u[2] /. FindRoot[d[2] = = q[2] , {u[2] , 0.5}] ; u[2] = temp ;
temp = u[3] /. FindRoot[d[3] = = q[3] , {u[3] , 0.5}] ; u[3] = temp ;
Print[u[1] , s , u[2] , s , u[3]] ; u[1] = . ; u[2] = . ; u[3] = . ;) , {1000}]
```

In the coding above, q[i] is a random variable $Q_{(i)} \sim \mathrm{U}(0,1)$, p[i] represents θ_i, c[i] and d[i] are $C_{(i)}\left(u_{(1)}, u_{(2)}, \ldots, u_{(i)}\right)$ and $C_{(i)}\left(u_{(i)} \middle| u_{(1)}, u_{(2)}, \ldots, u_{(i-1)}\right)$ respectively, and u[i] is a sample of $V_{(i)}$. With the Do function, the simulation is repeated for 1,000 times. This code can be applied in a similar way for different dimensions and for the other Archimedean copulas.

Sample τ

The sample τ between two related random variables X and Y can be computed by the following VBA code.

```
tau = 0
h = 0
For i = 1 To n – 1
For j = i + 1 To n
temp = (x(i) – x(j)) * (y(i) – y(j))
If temp > 0 Then
tau = tau + 1
ElseIf temp < 0 Then
tau = tau – 1
End If
h = h + 1
Next j
Next i
tau = tau / h
```

In the coding above, x(i) and y(i) are the i^{th} pair of observations of X and Y, and n is the total number of pairs of observations.

Debye Function

For the Frank copula, the Debye function $D_1(\theta) = \int_0^\theta t/(\exp(t)-1)dt \Big/ \theta$ can be solved by using a Riemann sum. The following VBA code makes use of a Riemann sum to compute the Debye function, where p is θ. Then, θ can be estimated from the sample τ.

```
Debye = 0
For i = 0.0000001 To p Step 0.0000001
Debye = Debye + i / (Exp(i) – 1) * 0.0000001
Next i
Debye = Debye / p
```

Alternatively, θ can be estimated for the Frank copula by using Mathematica as shown below, with say the sample τ being equal to 0.35.

```
tau = 1 – 4 / p * (1 – 1 / p * Integrate[t / (Exp[t] – 1) , {t , 0 , p}]) ;
FindRoot[tau = = 0.35 , {p , 1}]
```

Values of θ for Archimedean Copulas

Table A2 below shows the values of θ we generate for different values of τ for the four Archimedean copulas discussed.

Table A2 θ of Archimedean Copulas

τ	Clayton	Gumbel-Hougaard	Frank	Nelsen No. 12
0.05	0.1053	1.0526	0.4509	–
0.10	0.2222	1.1111	0.9074	–
0.15	0.3529	1.1765	1.3752	–
0.20	0.5000	1.2500	1.8609	–
0.25	0.6667	1.3333	2.3719	–
0.30	0.8571	1.4286	2.9174	–
0.35	1.0769	1.5385	3.5088	1.0256
0.40	1.3333	1.6667	4.1611	1.1111
0.45	1.6364	1.8182	4.8942	1.2121
0.50	2.0000	2.0000	5.7363	1.3333
0.55	2.4444	2.2222	6.7278	1.4815
0.60	3.0000	2.5000	7.9296	1.6667
0.65	3.7143	2.8571	9.4376	1.9048
0.70	4.6667	3.3333	11.4115	2.2222
0.75	6.0000	4.0000	14.1385	2.6667
0.80	8.0000	5.0000	18.1915	3.3333
0.85	11.3333	6.6667	24.9054	4.4444
0.90	18.0000	10.0000	38.2812	6.6667
0.95	38.0000	20.0000	78.3198	13.3333

Cholesky Decomposition

The Cholesky decomposition process is a convenient way to turn independently drawn normal random variables into multinormal random variables for a particular correlation matrix. The crucial step is to form a Cholesky matrix from the correlation matrix. Golan (2004) demonstrates the algorithm. Suppose $N_{(i)}$'s are independently drawn standard normal random variables, samples of m-variate standard normal random variables $Z_{(i)}$'s are then generated by $\vec{Z} = L_{\text{Cholesky}}\vec{N}$ ($\vec{N}$ is a vector of $N_{(i)}$'s, $\vec{Z}$ is a vector of $Z_{(i)}$'s, and

L_{Cholesky} is the Cholesky matrix). The following VBA code can be used to carry out the Cholesky decomposition process, where m is the number of dimensions, LT represents L'_{Cholesky} (the transpose of L_{Cholesky}), and the correlation matrix is input into LT before running the code.

```
For k = 1 To m – 1
For i = k + 1 To m
g = LT(i , k) / LT(k , k)
For j = k To m
If j = k Then
LT(i , j) = 0
Else
LT(i , j) = LT(i , j) – g * LT(k , j)
End If
Next j
Next i
Next k

For i = 1 To m
h = LT(i , i) ^ 0.5
For j = i To m
LT(i , j) = LT(i , j) / h
Next j
Next i
```

Alternatively, the Cholesky decomposition routines implemented in many mathematical software packages can be used.

APPENDIX III

Four formal tests are set forth in this appendix for justifying copula selection between two bivariate random variables. Suppose X and Y are two associated random variables that are continuous, (X_i, Y_i) are the i$^{\text{th}}$ pair (a total of n pairs) of observations of (X, Y), $u_i = F_X(x_i)$, and $v_i = F_Y(y_i)$. Suppose C is the underlying unknown copula between X and Y and C^* is the selected copula between X and Y with its r or θ estimated from the sample τ .

Goodness of Fit Test

The observations of (X_i, Y_i) are recorded on a scatter plot. The plot is then subdivided into regions. For each region, the actual number of observations is counted, and the expected number of observations is computed from using np^*, where p^* represents the estimated probability of having X and Y lying in that region. For each region, p^* is calculated from C^*, which is the selected joint bivariate cdf.

The hypothesis that C^* represents C between X and Y is tested by using the test statistic $\sum_{i=1}^{k} (\text{observed}_i - \text{expected}_i)^2 / \text{expected}_i \sim \chi^2_{k-1}$. This statistic is actually the sum of the weighted square difference between the actual and expected number of observations of each region, and k is the number of regions. The hypothesis is rejected if the test statistic is larger than the selected critical value (one-sided).

Different types of copula are selected and tested, and the one with the lowest value of the test statistic is preferred over the others. According to Watson (1983), this test works properly when n is large, p^* is not too small for each region, $np^* > 5$ for each region, and k is large.

CDF Test

This test is proposed by Guégan and Ladoucette (2004). For each pair of observations, $\Pr(X < x_i, Y < y_i)$ is estimated by directly counting the observations. The corresponding $C^*(u_i, v_i)$ is also computed. The following numerical criterion is then calculated:

$$\sum_{i=1}^{n}\left(\Pr(X < x_i, Y < y_i) - C^*(u_i, v_i)\right)^2 ,$$

which is effectively the sum of the square difference between the sample cdf and the selected cdf.

Different types of copula are selected and the one with the lowest value of the numerical criterion is preferred over the others.

$K_C(t)$ **Test**

This test is suggested by Frees and Valdez (1998) for Archimedean copulas. For each pair of observations, $\Pr(X < x_i, Y < y_i)$ is estimated by directly counting the observations. This estimate forms a sample of $C(A, B)$, which is a random variable, where $A \sim \mathrm{U}(0,1)$ and $B \sim \mathrm{U}(0,1)$. Over all the observations, all samples of $C(A, B)$ constitute the sample $K_C(t)$, which is the sample cdf of $C(A, B)$. The following integral is then computed:

$$\int_0^1 \left(K_C(t) - K_{C^*}(t)\right)^2 dK_C(t) \;\approx \sum_{t \,\in\, \text{samples of } C(A,B)} \left(\text{sample } K_C(t) - K_{C^*}(t)\right)^2 \Delta K_C(t),$$

in which the summation is carried over increasing values of *t*, $K_{C^*}(t)$ is computed from the selected copula, and $\Delta K_C(t)$ represents the increase in the sample $K_C(t)$ for successive pairs of *t*.

Different types of Archimedean copula are selected and the one with the lowest value of the integral is preferred over the others.

A Q-Q plot between the sample $K_C(t)$ and the computed $K_{C^*}(t)$ can also be used. A particular Archimedean copula fitting is better than the other if the line in its Q-Q plot is straighter and has a slope closer to one.

Binomial Test

Proposed by Kupiec (1995), this test can be exploited to check whether the upper-right-quadrant tail is fitted properly. Consider the upper-right-quadrant at the 99.5$^{\text{th}}$ percentile for X and Y. The probability $\Pr\left(X > F_X^{-1}(0.995), Y > F_Y^{-1}(0.995)\right)$ is computed both from directly counting the observations and from C^*. The former estimate is noted as $\hat{p}$ and the latter as p^*. The parameter p is the true underlying probability, which is unknown.

Suppose B is a random variable representing the number of observations lying in the 99.5$^{\text{th}}$ percentile quadrant, with $B \sim \mathrm{Bi}(n, p)$ and $\hat{p} = B/n$. The hypotheses $\mathrm{H}_0 : p = p^*$ and $\mathrm{H}_1 : p \neq p^*$ are tested by the test statistic:

$$2\left(\ln\left(\hat{p}^B (1-\hat{p})^{n-B}\right) - \ln\left(p^{*B}(1-p^*)^{n-B}\right)\right) \sim \chi_1^2 \ ,$$

in which $H_0 : p = p^*$ is rejected if the test statistic is larger than the selected critical value (one-sided).

In addition, n has to be very large to ensure the functionality of this test.

Parameter Selection

All the four tests described above can also be applied in a way to identify the optimal values of the parameters. With a particular copula selected, the optimality of different values of r, θ, or v can be examined (by trial and error) with the four tests.

APPENDIX IV

The class of Laplace transforms $\mathcal{L}_\infty$ and the class $\mathcal{L}_\infty^*$ are defined in Joe (1997).

$$\mathcal{L}_\infty = \left\{ \phi : [0,\infty) \to [0,1] \mid \phi(0) = 1 \,,\ \phi(\infty) = 0 \,,\ (-1)^j \phi^{(j)} \geq 0 \,,\ j = 1,2,...,\infty \right\} .$$

Any function pertaining to $\mathcal{L}_\infty$ has derivatives that alternate in sign.

$$\mathcal{L}_\infty^* = \left\{ \omega : [0,\infty) \to [0,\infty) \mid \omega(0) = 0 \,,\ \omega(\infty) = \infty \,,\ (-1)^{j-1} \omega^{(j)} \geq 0 \,,\ j = 1,2,...,\infty \right\} .$$

Any function pertaining to $\mathcal{L}_\infty^*$ has derivatives that alternate in sign.