# Empirical Estimation of Dependence in a Portfolio of Insurance Claims - Preliminaries*

Emiliano A. Valdez† and Andrew Chernih
School of Actuarial Studies
Faculty of Commerce & Economics
University of New South Wales
Sydney, AUSTRALIA

## Abstract

The purpose of this paper is to provide preliminary discussion about a project aimed to investigate the presence of "dependencies" in individual claims in a portfolio of insurance policies. In traditional risk theory, the individual claims in a portfolio of insurance have been assumed to be mutually independent. This means that for combination of claims, the probabilities of claim occurrence and the claim sizes do not directly impact each other. The traditional approach is to assume claims are independent because this is more mathematically tractable, that is, it becomes straightforward to compute for surplus requirements or probabilities of ruin. However, this may be unrealistic because intuitively, claims can exhibit "dependence" and they may be subject to the same physical or economic environment that might be driving the dependency. Using actual policy and claims data for a period of 9 years, this project aims to empirically investigate and validate the presence of dependencies in a portfolio of insurance policies. There is a growing number of papers on the literature addressing the issue of dependencies on claims, but no paper has provided empirical evidence to completely reject or accept the traditional framework of independence.

*Acknowledgment*: This project is being funded by the Actuarial Education and Research Fund (AERF) and the Casualty Actuarial Society (CAS), and the authors thank both organizations. The authors also thank Dr. Michel Verschuere of K.U. Leuven, Belgium and Vienna University of Technology, Austria for his helpful comments on earlier version of this draft.

*This paper is being prepared for presentation at the First Brazilian Conference on Statistical Modelling in Insurance & Finance, Ubatuba, Brazil, September 2003.

†Please do not quote without permission from authors. Email: e.valdez@unsw.edu.au

## 1 Introduction

Consider a portfolio of $n$ insurance risks $X_1$, $X_2$,..., $X_n$. The aggregate claims is defined to be the sum of these risks:

$$S = X_1 + X_2 + \cdots + X_n = \sum_{k=1}^{n} X_k,$$

where generally the risks are non-negative random variables, i.e. $X_k \geq 0$. The distribution of this sum is of considerable importance to the insurer because for example, it is needed as an input in premium and reserve determination for one. In classical risk theory, the risks are typically assumed to be independent and computation of the aggregate claims become more tractable. For certain types of individual claim distribution, one can even determine the exact distribution for this aggregate claims. Approximations using the central limit theorem can also be justified for a large number of insurance risks. See Bowers, et al. (1998), Buhlmann (1970), and Klugman, et al. (1998).

The primary purpose of this research project is to estimate the presence of dependence in claims in a portfolio of insurance risks. Using the methods of copulas, we can easily express the joint distribution of claims and thereby allowing us to construct the appropriate likelihood for estimation purposes. The copula will contain a parameter of dependence which will then provide a measure of the level of the presence of dependence in the insurance portfolio. This measure can easily be re-expressed in terms of more familiar measures of dependence such as Spearman's correlation which will allow us to have a better understanding of the magnitude of the dependence. Section 3 of this paper provides a quick review about copulas.

The results of this project can have far-reaching implications to the practicing actuary who may be concerned about the financial consequences of assuming independent claims when in fact there may be forces driving dependencies. The actuary can be provided with better information to make more informed decisions. Not only will this project provide possible empirical evidence of dependencies, it will also offer an innovative procedure for examining the presence of such dependencies.

# 2 Ruin Probabilities with Dependent Claims - Some Result

In classical risk theory, the surplus process is a very important model for understanding how capital or surplus of an insurance company evolves over time. It is defined to be

$$U_t = u_0 + P_t - S_t$$

where $U_t$ represents the company surplus at time $t$, $u_0$ is the initial company surplus, $P_t$ is the amount of premium collected up until time $t$ and $S_t$ is the accumulated amount of claims paid up until time $t$. The claims process $\{S_t\}$ is usually modelled as the sum of individual risks as

$$S_t = X_1 + X_2 + \cdots + X_{N(t)},$$

where the $X$'s are individual claim amounts and $N(t)$ is the number of claims up until time $t$. One important quantity considered in risk theory is the probability of ruin defined as

$$\psi(u) = \text{Prob}\left(U_t < 0, \text{ for some time } t > 0 \,|u_0 = u\right)$$

which represents the probability that the surplus will ever reach below zero given a level of initial surplus $u$. For mathematical simplicity, the individual risks $X_k$'s are typically assumed to be independent random variables, although there may be several situations for which this is no longer true. For example, the policyholders in an insurance portfolio may share common characteristics, or may be subject to the same physical or economic environment, for which they no longer are considered independent risks. Consider the case of insuring contiguous risks for fire or earthquake protection. It is very possible that claims occur in multiplicity when they do occur.

Valdez and Mo (2002) investigate the impact of dependence on ruin probabilities. By employing simulation procedures and imposing a copula function on the dependence structure, they found larger ruin probabilities for portfolios assuming the presence of dependence and that the Lundberg upper bound is violated. See Figure 1. By additionally examining for evidence of dependence in a portfolio of insurance risks, we can further extend this work. Another interesting work on estimating ruin probabilities with dependent claims is that of Albrecher and Kantor (2002).

# 3 Modelling Dependence with Copulas

## 3.1 Definition

Suppose that an $n$-dimensional random vector $\mathbf{X} = (X_1, ..., X_n)$ has the cumulative distribution function

$$F(x_1, ..., x_n) = \text{Prob}\left(X_1 \le x_1, ..., X_n \le x_n\right).$$

We can decompose this c.d.f. $F$ into the univariate marginals of $X_k$ for $k = 1, 2, ..., n$ and another distribution function called a copula. Before formally defining a copula function, let us examine the properties of a multivariate distribution function. Following Joe (1997), a function $F$ with support $\mathbf{R}^n$ and range $[0, 1]$ is a multivariate cumulative distribution function if it satisfies the following:

1. it is right-continuous;
2. $\lim_{x_k \to -\infty} F(x_1, ..., x_n) = 0$, for $k = 1, 2, ..., n$;
3. $\lim_{x_k \to \infty,\ \forall k} F(x_1, ..., x_n) = 1$; and
4. the following rectangle inequality holds: for all $(a_1, ..., a_n)$ and $(b_1, ..., b_n)$ with $a_k \le b_k$ for $k = 1, 2, ..., n$ ,we have
$$\sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} F(x_{1i_1}, ..., x_{ni_n}) \ge 0,$$
where $x_{k1} = a_k$ and $x_{k2} = b_k$.

Suppose $\mathbf{u} = (u_1, ..., u_n)$ belong to the $n$-cube $[0, 1]^n$. A copula, $C(\mathbf{u})$, is a function, with support $[0, 1]^n$ and range $[0, 1]$, that is a multivariate cumulative distribution function whose univariate marginals are $U(0, 1)$. As a consequence of this definition, we see that

$$C(u_1, ..., u_{k-1}, 0, u_{k+1}, ..., u_n) = 0$$

and

$$C(1, ..., 1, u_k, 1, ..., 1) = u_k$$

for all $k = 1, 2, ..., n$. Any copula function $C$ is therefore the distribution of a multivariate uniform random vector. From the definition of a multivariate distribution function, the rectangle inequality leads us to

$$\begin{aligned} & \text{Prob}\left(a_1 \leq U_1 \leq b_1, ..., a_n \leq U_n \leq b_n\right) \\ = & \sum_{i_1=1}^{2} \cdots \sum_{i_n=1}^{2} (-1)^{i_1+\cdots+i_n} C\left(u_{1i_1}, ..., u_{ni_n}\right) \geq 0, \end{aligned}$$

for all $u_k \in [0,1]$, $(a_1, a_2, ..., a_n)$ and $(b_1, b_2, ..., b_n)$ with $a_k \leq b_k$ for $k = 1, 2, ..., n$, and $u_{k1} = a_k$ and $u_{k2} = b_k$.

The significance of copulas in examining the dependence structure of $X_1, X_2, ..., X_n$ comes from a result which first appeared in Sklar (1959). Known as Sklar's theorem, it relates the marginal distribution functions to copulas. Suppose $\mathbf{X}$ is a random vector with joint distribution function $F$. According to Sklar (1959), there exists a copula function $C$ such that

$$F\left(x_1, ..., x_n\right) = C\left(F_1\left(x_1\right), ..., F_n\left(x_n\right)\right)$$

where $F_k$ is the $k$th univariate marginal, for $k = 1, 2, ..., n$. The function $C$ need not be unique, but it is unique if the univariate marginals are absolutely continuous. For absolutely continuous univariate marginals, the unique copula function is clearly

$$C\left(u_1, ..., u_n\right) = F\left(F_1^{-1}\left(x_1\right), ..., F_n^{-1}\left(x_n\right)\right)$$

where $F_1^{-1}, ..., F_n^{-1}$ denote the quantile functions of the univariate marginals $F_1, ..., F_n$. It is apparent that the copula is a function which "couples,", "links," or "connects" the joint distribution to its marginals.

In the purely discrete case, denote the $k$th univariate distribution function by $F_k\left(i_k\right) = \text{Prob}(X_k \leq i_k)$ together with its probability mass function

$$\begin{aligned} p_k\left(i_k\right) &= \text{Prob}\left(X_k = i_k\right) \\ &= \text{Prob}\left(X_k \leq i_k\right) - \text{Prob}\left(X_k \leq i_k^-\right) \\ &= F_k\left(i_k\right) - F_k\left(i_k^-\right), \end{aligned}$$

where $i_k$ belongs to its set of support, say $D_k$. A copula function $C$ then that is associated with the joint distribution of $X_1, ..., X_n$ will satisfy

$$P\left(i_1, ..., i_n\right) = C\left(F_1\left(i_1\right), ..., F_n\left(i_n\right)\right)$$

for all $i_k \in D_k$, where $P(\cdot, ..., \cdot)$ denotes the cumulative distribution of the discrete random vector. Here the copula $C$ although it exists, need not be unique.

An example of a copula is the independence copula which is given by

$$C\left(u_1, ..., u_n\right) = u_1 \cdots u_n$$

and is the copula associated with the joint distribution of independent random variables $X_1, X_2, ..., X_n$. This copula is often denoted simply by $\mathbf{\Pi}\left(u_1, ..., u_n\right)$. Another very important copula is the normal copula. Denote the density and cumulative distribution functions of a univariate standard normal by $\phi(\cdot)$ and $\Phi(\cdot)$, respectively, so that

$$\Phi(z) = \int_{-\infty}^{z} \phi(w)\, dw, \text{ where } \phi(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2}.$$

Consider an $n$-variate normal random vector $\mathbf{Z} = (Z_1, ..., Z_n)$ with standard normal marginals, i.e. $Z_k \sim N(0,1)$ for $k = 1, 2, ..., n$ and positive-definite, symmetric variance-covariance matrix $\mathbf{V} = (v_{ij})$. Clearly, the elements of $\mathbf{V}$ satisfy

$$v_{ij} = \begin{cases} 1, & \text{if } i = j \\ \text{corr}\left(Z_i, Z_j\right), & \text{if } i \neq j \end{cases}.$$

The joint density of $\mathbf{Z}$ can be expressed as

$$f\left(z_1, ..., z_n\right) = \frac{1}{\sqrt{(2\pi)^n}\, |\mathbf{V}|} \exp\left(-\frac{1}{2}\mathbf{z}^T \mathbf{V} \mathbf{z}\right),$$

with $\mathbf{z} = (z_1, ..., z_n)$. Now denote the joint distribution function by

$$H\left(z_1, ..., z_n\right) = \int_{-\infty}^{z_n} \cdots \int_{-\infty}^{z_1} f\left(z_1, ..., z_n\right) dz_1 \cdots dz_n.$$

The copula defined by

$$C\left(u_1, ..., u_n\right) = H\left(\Phi^{-1}\left(u_1\right), ..., \Phi^{-1}\left(u_n\right)\right)$$

is called the normal copula and is easily seen to define a multivariate uniform cumulative distribution function. Although the normal copula does not appear to be simple in form, it generally leads to simple simulation procedures. For those interested, we refer them to the paper by Wang (1998).

## 3.2 Copula Construction

The use of Laplace transforms can lead us to construct special types of copulas. Suppose $H$ is the cumulative distribution function of a positive random variable $X$, that is, $H(x)$ =Prob$(X \leq x)$. We shall denote the Laplace transform of $X$ by $\phi$ and it is defined by

$$\phi(t) = E\left(e^{-tX}\right) = \int_0^{\infty} \exp(-tx)\, dH(x), \text{ for } t \geq 0.$$

It is clearly true that

$$M_X(-t) = E\left(e^{-tX}\right)$$

Table 1: Some Examples of Laplace Transforms

| Family of Distributions | Laplace Transform $\phi(t)$ | Parameter Constraint |
|---|---|---|
| Gamma | $(1+t)^{-1/\theta}$ | $\theta \geq 1$ |
| Positive Stable | $\exp\left(-t^{1/\theta}\right)$ | $\theta \geq 0$ |
| Logarithmic Series | $-\frac{1}{\theta}\log\left[1-\left(1-e^{-\theta}\right)e^{-t}\right]$ | $\theta > 0$ |
| Power Series | $1-\left(1-e^{-t}\right)^{1/\theta}$ | $\theta \geq 1$ |

where $M$ denotes the moment generating function of $X$. The Laplace transform $\phi$ is a continuous function of $t$ and is strictly decreasing since

$$\phi'(t) = \int_0^\infty -x\exp(-tx)\,dH(x) \leq 0.$$

Also, $\phi(0) = 1$. We shall denote by $\phi^{-1}$ to be the inverse function of $\phi$. It can be verified that $\phi^{-1}$ is also a strictly decreasing function with

$$\phi^{-1}(0) = \infty \text{ and } \phi^{-1}(1) = 0.$$

See Feller (1971) for verification of these properties and to learn of other properties of the Laplace transform. Table 1 gives examples of random variables with their Laplace transforms.

Now, consider an arbitrary distribution function $F$ of a positive random variable $X$. It is well-known that we can construct another $G$ from the Laplace transform of $X$ as follows. There is a unique distribution function $G$ for which

$$F(x) = \int_0^\infty G^z(x)\,dH(z) = \phi(-\log G(x))$$

so that

$$G(x) = \exp\left[-\phi^{-1}(F(x))\right].$$

To extend this to the bivariate case, let

$$G_k(x_k) = \exp\left[-\phi^{-1}(F_k(x_k))\right] \text{ for } k = 1, 2.$$

Then the joint distribution function becomes

$$\begin{aligned} & \int_0^\infty [G_1(x)\,G_2(x)]^z\,dH(z) \\ = \;& \phi[-\log(G_1(x)\,G_2(x))] \\ = \;& \phi[-\log G_1(x) - \log G_2(x)] \\ = \;& \phi\left[\phi^{-1}(F_1(x_1)) + \phi^{-1}(F_2(x_2))\right] \end{aligned}$$

so that the copula function can be defined as

$$C(u_1, u_2) = \phi\left[\phi^{-1}(u_1) + \phi^{-1}(u_2)\right].$$

We can extend this procedure to higher than two dimensions by first defining $G_k(x_k) = \exp\left[-\phi^{-1}(F_k(x_k))\right]$ for $k = 1, 2, ..., n$ and then the following is a joint distribution function

$$\begin{aligned} & \int_0^\infty [G_1(x)\cdots G_n(x)]^z\,dH(z) \\ = \;& \phi\left[\phi^{-1}(F_1(x_1)) + \cdots + \phi^{-1}(F_n(x_n))\right]. \end{aligned}$$

Any copula satisfying this form belongs to the class of Archimedean copulas. More formally, we say that a copula function $C$ is *Archimedean* if it can be written in the form

$$C(u_1, ..., u_n) = \psi^{-1}[\psi(u_1) + \cdots + \psi(u_n)]$$

for all $0 \leq u_1, ..., u_n \leq 1$ and for some function $\psi$ (often called the generator) satisfying:

(i) $\psi(1) = 0$; and

(ii) $\psi$ is decreasing and convex, i.e. for all $t \in (0, 1)$, $\psi'(t) < 0$ and $\psi''(t) \geq 0$.

This class of copulas has been extensively studied by Genest and Mackay (1986) and Genest and Rivest (1993). As demonstrated in these articles, these types of copulas possess several desirable properties that make them attractive for statistical inference and simulation. They are also useful for extending copulas to higher dimensions. The independent copula belongs to this class with generator

$$\psi(t) = -k \cdot \log(t)$$

for some arbitrary positive constant $k$. Although the Frechet lower bound is not a copula for dimensions

$n>2$, its bivariate form is Archimedean with generator

$$\psi(t) = 1 - t.$$

However, the Frechet upper bound can be proven by contradiction that it does not belong to the class of Archimedean copulas. To do this, suppose there exists a generator $\phi$, then it must be that

$$M(u_1, ..., u_n) = \psi^{-1}[\psi(u_1) + \cdots + \psi(u_n)]$$

so that

$$\psi(M) = \psi(u_1) + \cdots + \psi(u_n).$$

Now, take the case where

$$u_1 = \cdots = u_n = u,$$

then we have

$$\psi(u) = n \cdot \psi(u)$$

which is impossible since the only solution to this is the case where $\psi(u) = 0$ for all $u$. Other generators and their corresponding copulas are given in Table 2 below.

For the case of the two dimensions $(n=2)$, Nelsen (1999) demonstrated that given a function $\psi$ that is continuous, strictly decreasing with $\psi(1) = 0$ and a function $C$ that satisfies $C(u_1, u_2) = \psi^{-1}[\psi(u_1) + \psi(u_2)]$, then $C$ is a copula if and only if $\psi$ is convex. That proof can be easily extended to multiple dimensions. Thus, we see that Archimedean copulas do in fact satisfy the definitions of a copula provided the generator is at least a convex function. It is easy to verify that the generators in our illustrative examples in Table 2 are convex functions.

Another procedure used to construct families of copulas is to use the concept of frailty models that have been extensively used in survival analysis. Although the usual development is based on the survival function, we illustrate this method using the distribution function. Denote the (frailty) random variable by $Z$ and assume that $Z$ is non-negative with density $g(z)$, distribution $F(z)$, and moment generating function $M_Z(t) = E_Z[\exp(tZ)]$ Now, consider the random vector $\mathbf{X} = (X_1, ..., X_n)$ and assume that these $n$ random variables $X_1, ..., X_n$ are independent given the frailty $Z$. Thus, we have

$$\begin{aligned} & \text{Prob}(X_1 \le x_1, ..., X_n \le x_n | Z = z) \\ = & \prod_{k=1}^{n} \text{Prob}(X_1 \le x_1 | Z = z) \\ = & \prod_{k=1}^{n} F_{k|z}(x_k | Z = z) \end{aligned}$$

and the joint distribution function can be derived using

$$\begin{aligned} & F(x_1, ..., x_n) \\ = & \int_0^\infty \text{Prob}(X_1 \le x_1, ..., X_n \le x_n | Z = z) \, g(z) \, dz \\ = & \int_0^\infty \left[\prod_{k=1}^{n} F_{k|z}(x_k | Z = z)\right] g(z) \, dz. \end{aligned}$$

Let us consider the special case where we assume that

$$F_{k|z}(x_k | Z = z) = [F_k(x_k)]^z, \text{ for all } k = 1, ..., n$$

where $F_{k|z}$ and $F_k$ denote the marginal distribution functions given the frailty and no frailty, respectively. It follows that we can express the joint distribution function as

$$\begin{aligned} & F(x_1, ..., x_n) \\ = & \int_0^\infty \left[\prod_{k=1}^{n} F_{k|z}(x_k | Z = z)\right] g(z) \, dz \\ = & \int_0^\infty \left[\prod_{k=1}^{n} F_k(x_k)\right]^z g(z) \, dz \\ = & E_Z\left\{\left[\prod_{k=1}^{n} F_k(x_k)\right]^Z\right\} \\ = & E_Z\left\{\exp\left[Z \sum_{k=1}^{n} \log(F_k(x_k))\right]\right\} \\ = & M_Z\left[\sum_{k=1}^{n} \log(F_k(x_k))\right]. \end{aligned}$$

Note that we can also express the marginals in terms of the inverses of the moment generating function of the frailty variable as

$$F_k(x_k) = \exp\left[M_Z^{-1}(F_k(x_k))\right].$$

Therefore the copula associated with the random vector is given by

$$C(u_1, ..., u_n) = M_Z\left[\sum_{k=1}^{n} M_Z^{-1}(u_k)\right].$$

This is exactly the form of an Archimedean copula. To illustrate, suppose the frailty random variable $Z$ has a positive stable distributions with m.g.f. given by

$$M_Z(t) = \exp\left[-(-t)^{1/\theta}\right].$$

which leads us to the Gumbel-Hougaard family of copulas

$$C(u_1, ..., u_n) = \exp\left\{-\left[\sum_{k=1}^{n} (-\log u_k)^\theta\right]^{-1/\theta}\right\}.$$

Table 2: Some Examples of One-Parameter Archimedean Copulas

| Family | Generator $\psi(t)$ | Copula Form $C(u_1,...,u_n)$ |
|---|---|---|
| Independence | $-\log(t)$ | $u_1 \cdots u_n$ |
| Cook-Johnson | $t^{-\theta}-1$, $\theta > 1$ | $\left(\sum_{k=1}^{n} u_k^{-\theta} - n + 1\right)^{-1/\theta}$ |
| Gumbel-Hougaard | $(-\log t)^{\theta}$, $\theta \geq 1$ | $\exp\left\{-\left[\sum_{k=1}^{n}(-\log u_k)^{\theta}\right]^{-1/\theta}\right\}$ |
| Frank | $\log\left(\frac{\theta^t - 1}{\theta - 1}\right)$, $\theta \geq 0$ | $\frac{1}{\log\theta}\log\left[1+\frac{\prod_{k=1}^{n}(\theta^{u_k}-1)}{(\theta-1)^{n-1}}\right]$ |
| Joe | $-\log\left[1-(1-t)^{\theta}\right]$, $\theta \geq 1$ | $1-\left\{1-\prod_{k=1}^{n}\left[1-(1-u_k)^{\theta}\right]\right\}^{1/\theta}$ |
| Gumbel-Barnett | $\log[1-\theta(1-t)]$, $0 < \theta \leq 1$ | $\exp\left\{-\frac{1}{\theta}\left[1-\prod_{k=1}^{n}(1-\theta\log u_k)\right]\right\}$ |

See Table 2. One of the advantage of the frailty framework is it allows us to interpret the dependencies of the random variables. For example, in an insurance portfolio, it is possible that an external variable, the frailty, may exist that contributes to the dependencies. Given then this external factor, the insurance risks would otherwise be unrelated or independent. Examples of possible external factor is the case where we have a portfolio of homeowner's insurance who may be living in the same neighborhood. The frailty random variable models the risks that could possibly exist among the homeowners within the neighborhood.

# 4 Data and Estimation

The most important piece of the project is empirically estimating for the presence of dependence in a portfolio of insurance claims. An insurance organization has provided us the required data on motor vehicle insurance to pursue this investigation. The types of information requested from the organization included dates of claims, gender, years of driving experience, among other things. Essentially, we have obtained the raw data which included policy exposures and claims experience for an observation period of nine (9) years beginning in 1993 and ending in 2001. The data comes from a portfolio of motor vehicle insurance policies. Figures 2 and 3 below provides graphical display of the monthly number and amount of claims during the observed period. These figures are very preliminary at the moment. We are currently in the process of cleaning up the data files, extracting information from the files, merging data files, and converting them into SAS-readable formats.

In this section, we briefly describe the procedure that may possibly be employed to estimate the model parameters. Here we discuss a general procedure for estimating parameters when a set of independent multivariate observations $\mathbf{X}_t = (X_{1t},...,X_{nt})$, $t = 1,...,T$ are given with their corresponding marginal distribution functions $F_k(\cdot;\boldsymbol{\theta}_k)$ and density functions $f_k(\cdot;\boldsymbol{\theta}_k)$, for $k = 1,...,n$. Here $\boldsymbol{\theta}_k$ denotes a vector of parameters. Consider a copula-based parametric model for the vector $\mathbf{X}$ so that its joint distribution function is expressed as

$$F(x_1,...,x_n;\boldsymbol{\theta},\alpha) = C(F_1(x_1;\boldsymbol{\theta}_1),...,F_n(x_n;\boldsymbol{\theta}_n);\alpha)$$

with $\alpha$ denoting the dependence parameter. Assume the mixed partial derivatives of $C$ exists and denote it by

$$c(u_1,...,u_n) = \frac{\partial^n C(u_1,...,u_n)}{\partial u_1 \cdots \partial u_n}.$$

It can be easily verified that the joint density is given by

$$\begin{aligned} & f(x_1,...,x_n;\boldsymbol{\theta},\alpha) \\ = \; & c(F_1(x_1;\boldsymbol{\theta}_1),...,F_n(x_n;\boldsymbol{\theta}_n);\alpha)\prod_{k=1}^{n} f_k(x_k;\boldsymbol{\theta}_k). \end{aligned}$$

For a sample size of $T$, the log-likelihood function can be expressed as

$$L(\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_n,\alpha) = \sum_{t=1}^{T}\log f(x_{1t},...,x_{nt};\boldsymbol{\theta},\alpha)$$

which can be a complicated function of the parameters. Instead of using the full maximum likelihood, an alternative is to use the so-called Method for Inference Functions for Margins (IFM). See Joe (1997). This method consists of first maximizing the $n$ univariate likelihoods by solving for the set of equations

$$\frac{\partial L_k}{\partial \boldsymbol{\theta}_k} = 0$$

and then using the solutions to these equations $\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_n$ to maximize the full likelihood by solving then

$$\frac{\partial L\left(\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_n, \alpha\right)}{\partial \alpha} = 0$$

to get the m.l.e. of $\alpha$. This method is believed to be less computationally intense.

Under certain regularity conditions (the research project will confirm these conditions), the estimated vector of parameters, say

$$\widehat{\boldsymbol{\xi}} = \left(\widehat{\boldsymbol{\theta}}_1, ..., \widehat{\boldsymbol{\theta}}_n, \widehat{\alpha}\right),$$

it can be shown that

$$\sqrt{T}\left(\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}\right)$$

is asymptotically normal with mean $\mathbf{0}$ and covariance matrix $I^{-1}(\boldsymbol{\xi})$ defined to be the inverse matrix of

$$\begin{aligned} I(\boldsymbol{\xi}) &= E\left[-\frac{1}{T}\frac{\partial^2 L(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^2}\right] \\ &= E\left[\frac{1}{T}\left(\frac{\partial L(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\right)\left(\frac{\partial L(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\right)^T\right]. \end{aligned}$$

We will use SAS to code the maximum likelihood procedure and in particular, use the Interactive Matrix Language (IML) procedure in SAS with optimization routines. These routines normally give estimates of the second derivatives of the likelihood function so that an estimated asymptotic covariance $I^{-1}\left(\widehat{\boldsymbol{\xi}}\right)$ is a standard output. Using these asymptotic results, we can then construct confidence intervals for our parameters and develop hypothesis tests. In particular, we will be interested in constructing test for the presence of dependence and this will largely depend on the choice of the parametric copula.

One of the issues to resolve in the research project is the choice of the copula. Joe (1997) and Nelsen (1999) lists several families of parametric copulas, and we described some of the more common ones in this proposal. Choosing the right copula is going to be a challenging aspect of the project because even in the theory of statistics, this area is not well-developed. Durrleman, et al. (2000) chooses a copula based on the construction of the empirical copula, but constructing this in the multivariate case is not straightforward. Our suggested strategy is to fit several families of copulas and choose among these families the best fitting model. An approach similar to a likelihood ratio arguments may then be applied.

We may be able to simplify further the likelihood estimation procedure if instead of directly specifying the copula, we construct the distribution function based on the concept of frailty. The frailty random variable is unobservable, however, given its presence, it gives independence. The distribution of the frailty random variable will contain a parameter (or a vector of parameters) that will describe the presence of dependence. We will explore this as an alternative to the choice of copula. Note that some of the familiar parametric copulas can be constructed using the concept of frailty, so the idea of a frailty is somewhat similar to copula specification.

# 5 Concluding Remarks

This paper lays out the theoretical foundations necessary to develop the solution to the problems with which we wish to address in the research project. In summary, our project aims to address the following fundamental issues: (1) Is there empirical evidence to support the presence of dependencies of claims in an insurance portfolio?; and (2) If there is such presence, what are its implications in computing insurance premiums, portfolio capital/surplus requirements, and/or ruin probabilities?

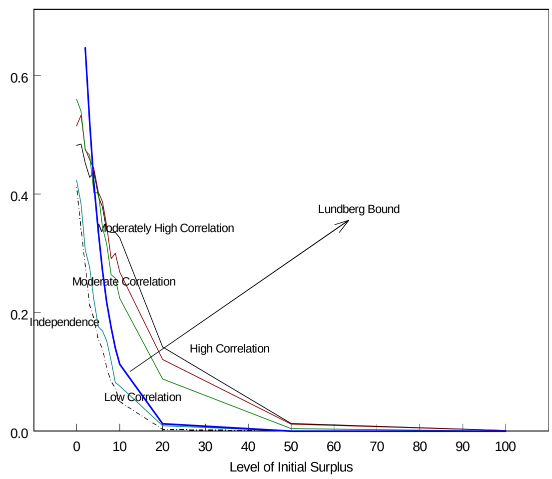


Figure 1: Dependence vs Ruin Probabilities

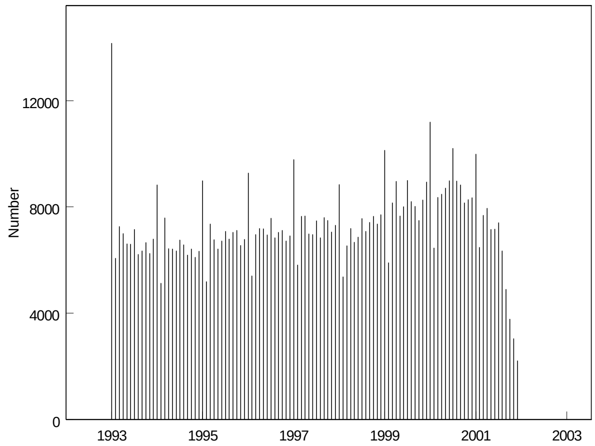


Figure 2: Monthly Number of Claims

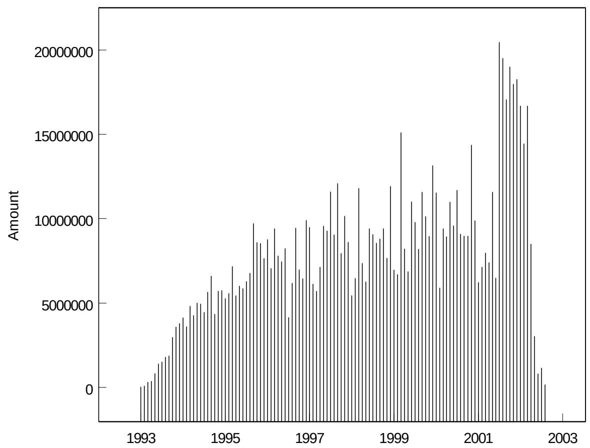


Figure 3: Monthly Amount of Claims

# A Bibliography


Albrecher, H., and Kantor, J. (2002) “Simulation of Ruin Probabilities for Risk Processes of Markovian Type,” *Monte Carlo Methods and Applications*, to appear.

Bowers, N.L., Gerber, H.U., Hickman, J.C., Jones, D.A., and Nesbitt, C.J. (1997) *Actuarial Mathematics* 2nd edition Schaumburg, Illinois: Society of Actuaries.

Buhlmann, H. (1970) *Mathematical Methods in Risk Theory* New York: Springer-Verlag.

Durrleman, V., Nikeghali, A. and Roncalli, T. (2000) “Which Copula is the Right One?,” working paper, Groupe de Recherche Operationelle, Credit Lyonnais, France.

Embrechts, P., McNeil, A., and Straumann, D. (2001) “Correlation and Dependence in Risk Management: Properties and Pitfalls,” *Risk Management: Value at Risk and Beyond,* ed. by Dempster, M. and Moffatt, H.K., Cambridge University Press.

Embrechts, P., Lindskog, F., and McNeil, A. (2001) “Modelling Dependence with Copulas and Applications to Risk Management,” working paper.

Feller, W. (1971) *An Introduction to Probability Theory and its Applications* Vol. 2, New York: John Wiley.

Frank, M.J. (1979) “On the Simultaneous Associativity of F(x,y) and x+y-F(x,y),” *Aequationes Mathematicae* 19: 194-226.

Frees, E.W. and Valdez, E.A. (1998) “Understanding Relationships Using Copulas,” *North American Actuarial Journal* 2: 1-25.

Frees, E.W., Carriere, J.F. and Valdez, E.A. (1996) “Annuity Valuation with Dependent Mortality,” *Journal of Risk and Insurance* 63: 229-261.

Genest, C. (1987) “Frank’s family of Bivariate Distributions,” *Biometrika* 74: 549-555.

Genest, C. and McKay, J. (1986) “The Joy of Copulas: Bivariate Distributions with Uniform Marginals,” *American Statistician* 40: 280-283.

Genest, C. and Rivest, L. (1993) “Statistical Inference Procedures for Bivariate Archimedean Copulas,” *Journal of the American Statistical Association* 88:1034-1043.

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

Klugman, S., Panjer, H. and Willmot, G. (1998) *Loss Models: From Data to Decisions* New York: John Wiley & Sons, Inc.

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

Panjer, H.H. and Willmott, G.E. “Models for the Distribution of Aggregate Claims in Risk Theory,” *Transactions of the Society of Actuaries* 36: 399-452 (including discussion).

Sklar, A. (1959) “Fonctions de repartition a n dimensions et leurs marges,” *Publications de l’Institut de Statistique de l’Universite de Paris* 8: 229-231.

Valdez, E.A. and Mo, K. (2002) “Ruin Probabilities with Dependent Claims,” submitted for publication, *British Actuarial Journal*.

Wang, S. (1998) “Aggregation of Correlated Risk Portfolios: Models and Algorithms,” *Proceedings of the Casualty Actuarial Society,* 85: 848-939.