

COPULA INFERENCE FOR MULTIPLE LIVES ANALYSIS–PRELIMINARIES*

Yogo Purwono

Department of Management
Faculty of Economics, University of Indonesia
Depok, INDONESIA

Abstract. Copula models are becoming increasingly popular tool for modeling dependencies between random variables, especially in such fields as biostatistics, actuarial science, and finance. The purpose of the present paper is to provide preliminary discussion about the project aimed to investigate models and inference methods for multiple lives–based insurance data by means of copula. In the life contingency models for several individuals, the life time of insured individuals have been assumed to be mutually independent. This means that for combination of lives, the life time probability of one live and life time of another lives do not directly impact each other. Independence assumption in multiple lives contingency models is often considered because this is more mathematically tractable to compute straightforwardly actuarial present value or benefit premiums and reserves. However, this may be unrealistic because intuitively, life times of several associated individuals, such as married couples, can exhibit “dependence” because of such conditions as common disaster, common life style, or the broken–heart syndrome. Using actual data on mortality of spouses that hold a last–survivor annuity policies in Indonesia, this research project will apply the basic concepts of copula inference in empirically investigating the presence of dependence in multiple lives–based insurance contract. There is a growing number of papers that explore the issue of dependencies on joint life times contracts, but no paper that has provided a detail lists of inference methods in copula modeling for multiple lives theory.

Key-words: copula function, multiple lives model, joint survival function, conditional copula, IFM method, Bayesian estimation.

1 Introduction

One important issue in actuarial practices is to model, in a much efficient way, the dependence of random life time of several individuals. Developing and finding the good model for this problem become absolutely necessary in insurance. The theory and the markets urged the economic agents to combine individual components. Insurance or annuities products covering several lives, like last–survivor annuities for a married couple for instance are increasingly demanded and have to be fairly priced. Similarly, finance instruments based on more than one asset, like multi–name credit risk–bearing derivatives, are widely used and have become standard. As part of a whole, random times–to–event behaviour are seldom independent and can even have rather intricate dependence structures. Therefore, flexible multivariate models and good inference methods are called for.

* Submitted to *the 13th East Asian Actuarial Conference*

Formerly, in finance (mainly) and in insurance (to some extent) there were two main approaches to the dependence modelling question: independence or multivariate normality. Either the several risks were admitted to be independent like in a portfolio of insurance policies, or multivariate normality was assumed as in the classical analysis of financial time series. In each case, risk aggregation is straightforward but often too far from providing realistic models. Similarly, when we are dealing with insurance policies insuring a group of lives, the premiums or reserved that we calculate under life time independencies assumption may be too far from the fair one.

A multivariate model has two components: the univariate (marginal) one, which characterizes each of the variables and the dependence structure between these marginal variables. This obvious separation can be specified in mathematical terms and it is exploited for modelling in this proposed study. The dependence structure of the random variables is known as the copula. Through the combination of copula with specific univariate distributions it is possible to construct an infinite number of multivariate distributions which we will refer to as copula-based models.

The notion of a function characterizing the dependence structure between several random variables comes from the work of Hoeffding in the early forties. Independently, other authors introduced related notions after-wards but it was in 1959 that Sklar [36] defined and named as copula a functional that gives the multivariate distribution as a function of the univariate marginal distributions. Allowing for the study of the dependence structure apart from the univariate behaviour of each variable, copula became very useful. The literature kept growing as the interest in copula increased, for instance in environmental data modelling. In the late nineties this notion was brought into finance and insurance problems. In 1997, Wang [39] proposes copula for modelling aggregate loss distributions of correlated insurance policies. Frees and Valdez [10] in 1998 and Klugman and Parsa [22] in 1999 use copula to model bivariate insurance claim data. Embrechts et al. [7] give a comprehensive overview of the application of copula to finance in a preprint already available on-line in 1999.

By now, the theory of copula is well established. On the other hand, the literature on inference questions is still scarce. References like Genest et al. [12], Genest and Rivest [13], Joe [20] and Oakes [31] are important in this context. We can find copula models fitted to multiple lives-based insurance in Anderson et al. [1], Frees et al. [9], Youn and Shemyakin [40], and Youn et al. [41]. Our interest lies on parsimonious multivariate copula-models of joint survival analysis, able to capture the relevant facts of insurance data and on the inference methods for those models. Such fitted models are essential to accurately estimate functionals of dependent life times, pricing insurance contracts based on more than one live and evaluate risk measures.

The primary purpose of this research project is to investigate copula models together with the proper inference methods for a broad context of multiple lives analysis and dwell on one particular problem of two associated lives. Straightforward copula analysis based on mixing marginal distribution functions as suggested in [1] and [9] might work successfully or not work at all depending on particular structure of association between two lives.

Section 2 discusses the basic definitions, properties, and examples of several types of copula functions. Here we will discuss in detail the types of copulas, which found their applications, for instance, in the analysis of extremes in financial asset

returns, and also the types of copulas, which seem to better fit the problems of life time analysis.

Section 3 is dedicated to the formulation of the range of multiple lives analysis problems, which we are going to consider. A special attention is paid to the effects of left truncation and right censoring on data. Joint first–life and joint last–survivor functions, and their relationship to the bivariate survival function are discussed.

In section 4 we explain approaches we will consider in our project, which allows for the estimation of joint first–life and joint last–survivor functions with the help of joint bivariate survival function. We briefly discuss the estimation methods we use to empirically estimate and infer copula models for general multivariate survival functions. We also explain the properties of these estimators in brief.

In section 5 we will discuss the conditional bayesian copula model suggested first in [21]. Its construction is based on the application of copula mixing to conditional rather than to marginal survival functions. It allows for an explicit utilization of the prior information available on the conditional survival functions. The discussion of why this approach should be sufficient for the estimation of the first–life context, but may fail for the last–survivor context, will be briefly given.

In section 6 a numerical data example for the application of copula analysis, which was previously discussed in section 2, is explained. It will be discussed copula models that will be considered to analyse the data. In the end part of this section, a potential application of join survival analysis to actuarial science is discussed and some suggestions for that issue are given. Section 7 concludes.

The results of this research can have far–reaching implications to the practicing actuary who may be concerned about the financial consequences of assuming independent life length of a pair lives when in fact there may be forces driving dependencies. The actuary can be equipped with better methods to make more informed decisions.

Whilts the literature on copula is devoted either to probabilistic theory, to inference methods or to applications, moreover, this research will combine of these three modeling aspects. Therefore, the results of this research will contribute considerably to the scientific related literatures and to making the new developed tools known to a wider audience of practitioners and researchers alike.

2 Modelling Dependence with Copulas

Suppose that a d –dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)$ has the distribution function

$$F(\mathbf{x}_1, \dots, \mathbf{x}_d) = \Pr(X_1 \leq \mathbf{x}_1, \dots, X_d \leq \mathbf{x}_d).$$

One can decompose F into the univariate marginals of X_k for $k = 1, 2, \dots, d$ and another distribution function called a copula. Referring to Joe in [20], a function F with support \mathbf{R}^d and range $[0, 1]$ is a multivariate distribution function if it satisfies the following:

- i) It is right–continuous;
- ii) $\lim_{x_k \rightarrow -\infty} F(\mathbf{x}_1, \dots, \mathbf{x}_d) = 0$, for $k = 1, 2, \dots, d$;
- iii) $\lim_{x_k \rightarrow -\infty, \forall k} F(\mathbf{x}_1, \dots, \mathbf{x}_d) = 1$; and

- iv) For all (a_1, \dots, a_d) and (b_1, \dots, b_d) with $a_k \leq b_k$ for $k = 1, 2, \dots, d$, we have

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} F(x_{1i_1}, \dots, x_{di_d}) \geq 0,$$

where $x_{k1} = a_k$ and $x_{k2} = b_k$.

A copula is a multivariate distribution function with standard uniform $(0, 1)$ margins. An equivalent definition that provides some copula properties follows. Suppose $\mathbf{u} = (u_1, \dots, u_d)$ belong to the d -cube $[0, 1]^d$. A copula, $C(\mathbf{u})$, is a function $C: [0, 1]^d \rightarrow [0, 1]$ satisfying the conditions:

- i) For all (u_1, \dots, u_d) in $[0, 1]^d$, if at least one component u_i is zero, $C(u_1, \dots, u_d) = 0$;
- ii) For $u_i \in [0, 1]$, $C(1, \dots, 1, u_i, 1, \dots, 1) = u_i$ for all $i \in \{1, 2, \dots, d\}$;
- iii) For all $[u_{11}, u_{12}] \times \dots \times [u_{d1}, u_{d2}]$ d -dimensional rectangles in $[0, 1]^d$,

$$\sum_{i_1=1}^2 \dots \sum_{i_d=1}^2 (-1)^{i_1+\dots+i_d} C(u_{1i_1}, \dots, u_{di_d}) \geq 0.$$

The ability of a copula to separate the dependence structure from the marginal behaviour in a multivariate distribution comes from Sklar's theorem in [36]. According to Sklar's theorem, when we have a d -dimensional distribution function F with univariate margins F_1, F_2, \dots, F_d and the ranges of F_i are \mathbf{R}_i , for $i = 1, 2, \dots, d$, then there exists a unique function H , defined on $\mathbf{R}_1 \times \mathbf{R}_2 \times \dots \times \mathbf{R}_d$ such that

$$F(\mathbf{x}_1, \dots, \mathbf{x}_d) = H(F_1(\mathbf{x}_1), \dots, F_d(\mathbf{x}_d)).$$

The extension of the function H to $[0, 1]^d$ is a copula C . For C such an extension of H , we have that

$$F(\mathbf{x}_1, \dots, \mathbf{x}_d) = C(F_1(\mathbf{x}_1), \dots, F_d(\mathbf{x}_d)).$$

The verifying of the theorem can be seen in [20, page 41]. If F_1, \dots, F_d in Sklar's theorem are continuous, the function H coincides with the copula C which is then unique.

Besides the ability of isolating the dependence structure of a multivariate distributions, the copula provides a way of constructing distributions from given margins. Sklar then stated that given univariate distribution functions F_1, \dots, F_d and a d -dimensional copula C , the function defined by

$$F(\mathbf{x}_1, \dots, \mathbf{x}_d) = C(F_1(\mathbf{x}_1), \dots, F_d(\mathbf{x}_d)) \tag{2.1}$$

is a d -dimensional distribution function with univariate margins F_1, \dots, F_d . For the proof, see [20, page 41].

So we are able to write the joint distribution as a function of the copula and the univariate marginal distributions. As the copula does not depend on the univariate marginal distributions, we can say that it capture completely all the information about the dependence between the variables. For this exposition, from here on we will refer to a copula and a dependence structure indifferently. We will refer to a statistical model given by a multivariate distribution written like (2.1) as a copula-based model.

Some important concepts in the construction of special types of copula-based statistical models are quantile functions, probability-integral and quantile transformations. Suppose that random variables X, X_1, \dots, X_d have distribution functions F, F_1, \dots, F_d respectively.

- i) The quantile function of F is defined for all u in $(0, 1)$ by the generalised inverse of F : $F^{\leftarrow}(u) = \inf \{x \in \mathbf{R} : F(x) \geq u\}$.
- ii) The probability integral transformation is the mapping $T : \mathbf{R}^d \rightarrow [0, 1]^d$, $(x_1, \dots, x_d) \rightarrow (F_1(x_1), \dots, F_d(x_d))$.
- iii) The quantile transformation is the operation $T^{\leftarrow} : [0, 1]^d \rightarrow \mathbf{R}^d$, and $(u_1, \dots, u_d) \rightarrow (F_1^{\leftarrow}(u_1), \dots, F_d^{\leftarrow}(u_d))$.

F^{-1} denotes the usual inverse function of F and is a particular case of F^{\leftarrow} when F is continuous and strictly increasing.

It is well-known that, if U is a uniform random variable on $(0, 1)$, then $F^{\leftarrow}(U)$ has distribution function F . On the other hand, the distribution function F of a random variable X is continuous if and only if $F(x)$ is uniformly distributed on $(0, 1)$.

Suppose that F_i is continuous and write $F_i(x_i) = u_i$ for $i = 1, \dots, d$. Making these substitution in (2.1), we obtain

$$C(u_1, \dots, u_d) = F(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)) \quad (2.2)$$

for $(u_1, \dots, u_d) \in [0, 1]^d$, which is an explicit expression for the copula as a function of the joint and the univariate marginal distribution functions.

Consider one example as follows. Suppose that we have the bivariate logistic distribution function given by

$$F(x, y) = \exp\left(-\left(x^{-\theta} + y^{-\theta}\right)^{1/\theta}\right), \quad x > 0, y > 0, \theta \geq 1.$$

We can rewrite this function in a more convenient form as follows:

$$F(x, y) = \exp\left(-\left(\left(-\log e^{-1/x}\right)^\theta + \left(-\log e^{-1/y}\right)^\theta\right)^{1/\theta}\right) \quad (2.3).$$

The margins of a bivariate logistic are standard Fréchet. It means that they have distribution function $F(z) = e^{-1/z}$ for $z > 0$. Therefore, one can readily obtain the copula function C of a pair of random variables with bivariate logistic distribution from (2.3) using (2.2), leading to

$$C(u_1, u_2) = \exp\left(-\left(\left(-\log u_1\right)^\theta + \left(-\log u_2\right)^\theta\right)^{1/\theta}\right), \quad (2.4)$$

where $\theta \geq 1$ and $(u_1, u_2) \in [0, 1]^2$.

The family of copula function, like in the previous example, is known as Gumbel, Gumbel–Hougaard or logistic copula. This copula family has the characteristics that are frequently suitable for modelling financial or insurance data.

Copula has a very interesting feature, that is invariance property. The formal explanation of invariance property is as follows. Suppose that C be the copula of the random vector $\mathbf{X} = (X_1, \dots, X_d)$ and \mathbf{R}_i the range of X_i , for $i = 1, \dots, d$. If the

function $g_i : \mathbf{R}_i \rightarrow \mathbf{R}$ for $i = 1, \dots, d$ are continuous and strictly increasing a.s., C is still the copula of $(g_1(X_1), \dots, g_d(X_d))$. If the univariate marginal distributions are continuous then the functions g_i only have to be increasing a.s. in order to keep the invariance of C .

The invariance property means that copula are the framework to study dependence properties which are invariant under increasing transformations of the individual random variables. Consider a set of dependent insurance claims and we want to come up with a multivariate distribution as a model for the losses. The joint distribution for the losses will have the same copula as the one for the logarithm of the losses or for the integral-transforms of the losses. So, in terms of copula modelling, we are indifferent with which of the three marginal scales we work (see 15).

In many applications, the interesting model is sometimes defined via the so-called survival family rather than the copula family itself. This modeling approach is usually used, in many cases, in modeling dependencies between life times of several objects. The survival copula appears as the function which relates the joint survival function of a multivariate distribution with the survival functions of the univariate margins. The formal proposition (for the proof see [27, page 28]) is as follows.

Suppose we have the distribution function F of the random vector (X_1, \dots, X_d) , with marginal distribution functions F_1, \dots, F_d respectively. Then, there exists a copula \bar{C} such that

$$\bar{F}(x_1, \dots, x_d) = \bar{C}(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)),$$

where $\bar{F}(x_1, \dots, x_m) = \Pr(X_1 > x_1, \dots, X_m > x_m)$ for $m = 1, \dots, d$. Furthermore, in the bivariate case if C is the copula of (X_1, X_2) then

$$\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2).$$

The proof of the above proposition we can see in [27, page 28].

In the final part of this section, we give some examples of copula important for life times dependencies modeling in insurance and finance. First we give an example of so-called explicit copula and then two examples of implicit copula. An implicit copula is obtained from a multivariate distribution function thorough (2.2). The copula is said to be explicit, when the expression which defines the copula is not written as a function of the joint and marginal distribution functions.

One example of explicit copula is Clayton copula. The Clayton copula with parameter $\theta \in (0, +\infty)$ is given by:

$$C(u_1, \dots, u_d) = \left(\sum_{i=1}^d (u_i^{-\theta} - 1) + 1 \right)^{-1/\theta}, \quad (u_1, \dots, u_d) \in [0, 1]^d.$$

The first example of implicit copula is Gaussian copula. For a d -dimensional random vector \mathbf{X} that has a multivariate normal distribution with mean vector zero and correlation matrix Σ , the Gaussian copula is defined as

$$C(u_1, \dots, u_d) = P(\Phi_1(X_1) \leq u_1, \dots, \Phi_1(X_d) \leq u_d)$$

$$\begin{aligned}
 &= P \left(X_1 \leq \Phi_1^{-1}(X_1), \dots, X_d \leq \Phi_1^{-1}(X_d) \right) \\
 &= \Phi_d \left(\Phi_1^{-1}(X_1), \dots, \Phi_1^{-1}(X_d) \right)
 \end{aligned}$$

t-copula is a second example of implicit copula. Suppose that \mathbf{X} has a zero mean d-dimensional multivariate t-distribution with density function

$$f(x_1, \dots, x_d) = \frac{\Gamma\left(\frac{\nu + d}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{(\pi\nu)^d |\Sigma|}} \left(1 + \frac{x' \Sigma^{-1} x}{\nu}\right)^{-\left(\frac{\nu+d}{2}\right)}$$

where Σ is the correlation matrix and ν are the degrees of freedom. Analogously to the Gaussian copula, the t-copula is given by

$$C(u_1, \dots, u_d) = t_d \left(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_d) \right) \tag{2.5}$$

where $(u_1, \dots, u_d) \in [0, 1]^d$, t_d represents the distribution function of a d-dimensional random vector with density (2.5) and t_ν denotes the distribution function of a standard univariate student-t random variable with ν degrees of freedom.

3 Multiple Lives Analysis

The specific problem of dependence analysis considered in the project deals with studying associations between several lives. For lives L_j define X_j as “age at death” of life L_j random variables. Assume that associated pairs of d-dimensional lives (L_1, \dots, L_d) are observed during a certain limited period of time T . An observation begins simultaneously at random entry age e_1 for live L_1 , e_2 for live L_2 , \dots , and e_d for live L_d . The observation like this is applied to each observation. This condition represents the “left truncation” application to the data. Life L_j is observed until L_j dies before the age $e_j + T$ or L_j is still alive at age $e_j + T$, in which case the death time will not be observed. This condition represents the “right censoring”.

Therefore, in a sample of $\mathbf{Y} = (Y_1, \dots, Y_n)$, each i-th observation y_i of an associated pair of lives (L_{i1}, \dots, L_{id}) may be represented as a vector

$$Y_i = \{(e_{i1}, \dots, e_{id}), (T_{i1}, \dots, T_{id}), (\delta_{i1}, \dots, \delta_{id})\} \tag{3.1}$$

where t_{ij} is the termination time length for life L_{ij} , and δ_{ij} is the right censoring indicator:

$$\delta_{ij} = 0, \text{ when } T_{ij} = T \text{ and } = 1 \text{ when } T_{ij} < T.$$

The range of application of such structure of observations is frequently encountered in health, epidemiological, or demographical studies. In these area of study, it is usually considered two lives $j = 1, 2$ where T_1 and T_2 may represent times to failure of associated organs (eyes, kidneys), see, e.g., [32], [35], or lives of twins [1], [17] or lives of married couples. The latter example, playing important role in insurance pricing and first analyzed by copula methods in [9], will be the numerical example in our project.

One of the important tasks in many applications like in the multiple lives-based insurance policies, is to estimate the future lifelength probabilities for given entry ages. As an example, consider an insurance covering benefits for a married couple, like last survivor life annuity. In this case, the interest are usually the joint first-life survival function

$$S_{FL}(t; e_1, e_2) = \Pr\{\min(X_1 - e_1, X_2 - e_2) > t \mid \min(X_1 - e_1, X_2 - e_2) > 0\} \quad (3.2)$$

and the joint last-survivor function

$$S_{LS}(t; e_1, e_2) = \Pr\{\max(X_1 - e_1, X_2 - e_2) > t \mid \min(X_1 - e_1, X_2 - e_2) > 0\} \quad (3.3)$$

A very natural approach to the estimation of $S_{FL}(t; e_1, e_2)$ and $S_{LS}(t; e_1, e_2)$ is to first estimate the bivariate survival function $S(t_1, t_2)$ as defined in section 2, and then use the formulas

$$S_{FL}(t; e_1, e_2) = \frac{S(e_1 + t, e_2 + t)}{S(e_1, e_2)}$$

and

$$S_{LS}(t; e_1, e_2) = \frac{S(e_1, e_2 + t) + S(e_1 + t, e_2) - S(e_1 + t, e_2 + t)}{S(e_1, e_2)} \quad (3.4)$$

We will consider this approach for explaining the numerical data example in the final section.

4 Estimation for Copula Models

Estimation of $S(t_1, t_2, \dots, t_d)$ may be carried out in two stages: at the first stage, the marginal survival functions $S_j(t_j)$ for $j = 1, \dots, d$ are estimated non-parametrically (via product-limit estimation method or its modification) or with the help of a parametric techniques by proposing a number of parametric distributions (popular choices could be Gompertz or Weibull distributions). At the second stage, the estimated survival functions are mixed according to some copula models (see section 2), and the association parameter is estimated via a statistical estimation method we will consider in our study.

An alternative approach consists of estimating simultaneously parameters of the marginal distributions and the association parameters as well. If there is a certain information available regarding the marginals, e.g., there exist additional data available on the individual lives, it is natural to consider Bayesian estimation methods with informative priors on the marginal parameters and non-informative prior on the parameter of association.

In the present project, we will restrict our shelves to only consider the second approach as the main interest of our study. Specifically, we will pay attention to the case of estimating directly the bivariate survival function as a copula

$$S(t_1, t_2; \theta_1, \theta_2, \alpha) = C(S_1(t_1; \theta_1), S_2(t_2; \theta_2), \alpha)$$

where θ_j for $j = 1, 2$ are parameters vectors of the marginals and α is the parameter of association. In the presence of right censoring, the likelihood function for the vector parameter $\theta = (\theta_1, \theta_2, \alpha)$ with data \mathbf{y} structured as described in (3.1).

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 $\mathbf{X}_i = (X_{i1}, \dots, X_{id})$, $i = 1, \dots, n$ are given with their corresponding marginal distribution functions $F_k(\cdot; \theta_k)$ and density functions $f_k(\cdot; \theta_k)$, for $k = 1, \dots, d$. Consider the random vector $\mathbf{X} = (X_1, \dots, X_d)^t$. Suppose we want to estimate the parametric copula-based model for \mathbf{X} given by

$$F(\mathbf{x}; \alpha_1, \dots, \alpha_d, \theta) = C(F_1(x_1; \alpha_1), \dots, F_d(x_d; \alpha_d)) \quad (4.1)$$

where $F_i(x_i; \alpha_i)$ is the distribution function of X_i with parameter vector $\alpha_i \in \mathbf{R}^{p_i}$ with $p_i \in \mathbf{N}$ for all univariate margins $i = 1, \dots, d$ and C is a copula family parameterised by the vector $\theta \in \mathbf{R}^q$ with $q \in \mathbf{N}$. Assume that C has a density function c given by

$$c(u_1, \dots, u_d; \theta) = \frac{\partial^d C(u_1, \dots, u_d; \theta)}{\partial u_1 \dots \partial u_d} \quad (4.2)$$

with $(u_1, \dots, u_d) \in [0, 1]^d$ and that F_i has a density f_i for all $i = 1, \dots, d$. For the case where the margins are discrete, we denote by f_i the probability mass function of X_i . As for our applications, margins are absolutely continuous, this will not be an issue. The density of the copula-based model (4.1) for \mathbf{X} is

$$f(\mathbf{x}; \alpha_1, \dots, \alpha_d, \theta) = c(F_1(x_1; \alpha_1), \dots, F_d(x_d; \alpha_d); \theta) \prod_{i=1}^d f_i(x_i; \alpha_i) \quad (4.3)$$

Suppose that we have n iid d -dimensional vectors of observations $(\mathbf{x}_1, \dots, \mathbf{x}_n)$. We assume that all the necessary regularity conditions (see [24]) on c and f_i for $i = 1, \dots, d$ are met. The log-likelihood function for the univariate margins of model (4.1) takes the form

$$L_i(\alpha_i; \mathbf{x}) = \sum_{j=1}^n \log f_i(x_{ij}; \alpha_i), \text{ for } i = 1, \dots, d, \quad (4.4)$$

while the log-likelihood for the copula model F is

$$L(\alpha_1, \dots, \alpha_d, \theta; \mathbf{x}) = \sum_{j=1}^n \log f(x_j; \alpha_1, \dots, \alpha_d, \theta) \quad (4.5)$$

Assuming that the usual regularity conditions are fulfilled, there is a vector solution $\underline{\alpha}_i$ to each one of the d systems of equations

$$\left(\frac{\partial L_i(\alpha_i; \mathbf{x})}{\partial \alpha_{i1}}, \dots, \frac{\partial L_i(\alpha_i; \mathbf{x})}{\partial \alpha_{ip_i}} \right) = 0, \text{ for } i = 1, \dots, d \quad (4.6)$$

which is the maximum likelihood estimator (MLE) for the marginal parameters. Note that these estimators are obtained independently for each margin.

The process of finding estimators from the previous procedure will be complex when the function in (4.5) is to be complicated. Instead of using the full maximum likelihood, we can use the so-called Inference Function for Margins (IFM) Method. This terminology comes from McLeish and Small in [26] and it has been followed by authors like Joe in [20] for copula-based model statistical inference. The method consists of estimating the model parameters by finding the roots of a conveniently defined set of inference functions. In the case of maximum likelihood estimation, the inference functions are the partial derivatives of the log-likelihood function. In the IFM method, the score functions of the margins and of the copula constitute the set of estimating equations.

The IFM method consists first of obtaining the MLE vectors $\bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_d$ for the marginal parameters solving (4.6) then substitute these marginal estimates in (4.5) to maximise (in most cases numerically)

$$L(\boldsymbol{\theta}; \mathbf{x}, \bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_d) = \sum_{j=1}^n \log f(x_j, \bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_d; \boldsymbol{\theta}), \quad (4.7)$$

in order to estimate $\boldsymbol{\theta}$, or to use the score function of (4.5) and estimate the dependence parameter vector $\bar{\boldsymbol{\theta}}$, solving the following system of equations

$$\left(\frac{\partial L(\boldsymbol{\theta}; \mathbf{x}, \bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_d)}{\partial \theta_1}, \dots, \frac{\partial L(\boldsymbol{\theta}; \mathbf{x}, \bar{\boldsymbol{\alpha}}_1, \dots, \bar{\boldsymbol{\alpha}}_d)}{\partial \theta_q} \right) = 0. \quad (4.8)$$

This procedure is far more easy numerically and less computationally intensive than a direct optimisation of (4.5). However, in case it is feasible to obtain the MLE for the full vector of model parameters from (4.5), we can use the IFM estimates as good starting values for the optimisation routine.

Joe in [20] proves that the IFM estimator verifies, under regular conditions (the research project will confirm these conditions), the property of asymptotic normality, and we have:

$$\sqrt{T} (\boldsymbol{\theta}_{IFM} - \boldsymbol{\theta}_0) \rightarrow N(0, G^{-1}(\boldsymbol{\theta}_0)) \quad (4.9)$$

with $G(\boldsymbol{\theta}_0)$ is the Godambe information matrix. Thus, if we define a score function

$$s(\boldsymbol{\theta}) = \left(\frac{\partial l_1}{\partial \theta_{11}}, \dots, \frac{\partial l_d}{\partial \theta_{1d}}, \frac{\partial l_c}{\partial \theta_2} \right)', \quad (4.10)$$

where l_j denotes the log-likelihood of the j th marginal, and l_c the log-likelihood for the copula itself, splitting the log-likelihood in two parts l_1, \dots, l_d for each margin and l_c for the copula, the Godambe information matrix takes the following form:

$$G(\boldsymbol{\theta}_0) = D^{-1}V(D^{-1})' \quad (4.11)$$

with

$$D = E \left[\frac{\partial s(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \text{ and } V = E [s(\boldsymbol{\theta})s(\boldsymbol{\theta})']$$

The estimation of this covariance matrix requires us to compute many derivatives. Joe in [20] then suggests the use of the jackknife method or other bootstrap methods to estimate

it. Joe [20] points out that the IFM method is highly efficient compared with the ML method (we also confirm this in the research project). It is worth noting that the IFM method may be viewed as a special case of the generalized method of moments (GMM) with an identity weight matrix (see [4]).

In this project, we will use SAS to code the maximum likelihood procedure and in particular, use the Interactive Matrix Language (IML) procedure with optimization routines. These routines normally give estimates of the second derivatives of the likelihood function so that an estimated asymptotic covariance in (4.11) is a standard output. Using these asymptotic results, we can then construct confidence intervals for our parameters of interest 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 copulas.

5 Conditional Copula for Joint Survival

Shemyakin et al in [33] explained that using (unconditional) copula model to directly estimate joint survival function $S(t_1, t_2)$ sometimes proved to be insufficient for some paired lives. The main reason for this is that the problem of dimensionality will appear when the entry ages of the paired lives do not coincide, which is usually happened in the case of the spouses' lives. Therefore, it seems that no single copula model for multivariate survival function is going to provide a good tool for estimation of joint first–life and last–survivor probabilities in (3.2) and in (3.3) respectively, for paired lives with different ages and association tied to real time rather than to the age scale. For more detailed discussion of this problem and its consequences, one can find in [34].

A solution to this controversy, consists in creating separate copula models with common structure and common value of the association parameter for each pair of entry ages. Instead of mixing two marginal survival functions into a joint bivariate function, one can mix two conditional survival functions conditioned to the survival of paired lives to their entry ages. This approach appears to be efficient for estimation of joint first–life, but faces serious problems when applied to estimation of last–survivor probabilities. An alternative but, probably, inferior solution of introducing the “age difference” as an additional variable was discussed in [33].

Consider the problem of estimating the joint first–life survival function like in (3.2). such estimation requires observing of a pair of associated lives until first death. A possible problem with the data structure as described in (3.1) in actuarial applications consists in the underpinning of the time of the first death in a pair, if only the second death brings about a change in the insurance payout. We will not discuss this issue because it is not clear how serious this problem where it can be depending on the origin of data.

The idea of conditional copulas is to apply a copula function to conditional survival functions instead of the marginals so that

$$P_{FL}(t; e_1, e_2) = C \left(\frac{S(e_1 + t)}{S(e_1)}, \frac{S(e_2 + t)}{S(e_2)} ; \alpha \right) \quad (5.1),$$

where the type of copula used could be the same as discussed above. The dimensionality problem, which was pointed out in the above, does not arise in this situation, because the joint-life probability is clearly conditioned to the entry ages. If the underlying survival distributions of individual lives allow for closed-form conditional distributions, this approach does not lead to a substantial increase of complexity of the likelihood. For example see [42].

Bayesian models based on the conditional copula approach with Weibull marginals and stable copulas were applied to maritalties of married couples in [34]. The choice of priors and hyperparameters for the shape and scale distributions is based on the fitness to the existing male and female mortality tables. This choice reflected the belief that a substantial amount of prior information on individual male and female mortality can be incorporated in a copula model.

However, this approach does not promise an equally easy success while dealing with the joint last-survivor probabilities described in (3.3) for the same application to mortalities of the married couples. For more detail for this interesting phenomena, one can see [40].

For last-survivor problem, Shemyakin et al proposed a reasonable approximation for the conditional copula of last survival probabilities (for detailed reasoning, see [34], page 10):

$$P_{LS}(t; e_1, e_2) \approx p_1(t; e_1 : 0) + p_2(t; e_2 : 0) - P_{FL}(t; e_1, e_2), \quad (5.2)$$

where $p_1(t; e_1 : 0)$ can be interpreted as the probability of a woman *alive* and *married* at entry age e_1 to survive for t more years, and p_2 defined similarly.

This would dictate a different choice of priors. A model, which will make use of approximation (5.2), may look like this:

$$P_{LS}(t; e_1, e_2) \approx p_1(t; e_1 : 0) + p_2(t; e_2 : 0) - P_{FL}(t; e_1, e_2)$$

where

$$p_j(t; e_j : 0) = \exp \left\{ \left(\frac{\alpha_j}{\beta_j} \right)^{\gamma_j} - \left(\frac{e_j + t}{\beta_j} \right)^{\gamma_j} \right\}, \quad (5.3)$$

and the prior distributions for parameters β_j and γ_j are set to reflect that they are related not to entire male and female populations, but to the populations of married men and women. As it is evident from Table 10 in [8], married men and women tend to have lower mortality comparatively to men and women in general. When we introduce Weibull priors based on mortality tables, we have to take into account this effect, leading to increased values in both scale and shape parameters. This modeling paradigm will be considered as one of analysis tools we want to apply to the data as an illustration of the application of copula models inference as describe in the previous section.

6 Data and Analysis

The most important piece of the project is illustrating the application of copula modeling of multiple lives data to the empirical estimation for the presence of dependence in the mortality of married couples. We will use the data coming from a huge number of joint last-survivor annuity contracts of a large Indonesian insurer in payoff status over 5-year observation period beginning in 1998 and ending in 2002. The data structure follows situation in section 3, where we

considered right censoring and left truncation to the data. In elicitation of the hyperparameters of the Bayesian models we also used Table 10 of [8] providing the group mortality rates by age, sex, and marital status.

We will analyse the data in five ways of modelling. For Model I to IV, we assume that marginal survival functions for men and women described by two-parameter Weibull distributions with scales β_j and shapes γ_j , where $j = 1$ corresponds to female mortality, and $j = 2$ to male mortality. For the copula model, we consider stable (Gumbel–Hougaard) copula with the association parameter α .

In Model I, we assume that the independence of male and female mortalities may be treated as a degenerate case of stable copula with $\alpha = 1$. In this model the estimates for scale and shape parameters were obtained by maximum likelihood method. In Model II, the simultaneous IFM estimation was performed for all five parameters (scales, shapes, and association) follows the same likelihood constructing paradigm done in [42]. Model III was first introduced in [34] and requires Bayesian estimation of five parameters of the conditional copula function described in section 5 with the following priors distributions: $\beta_j \sim N(\phi_j, \eta_j)$, $\gamma_j \sim N(\mu_j, \sigma_j)$, $\pi(\alpha) \propto 1/\alpha$, where hyperparameters were determined from US male and female mortality tables. The computation of last-survivor probability can be carried out with the help of approximate formula (7) in [42].

Model IV uses the conditional copula function described in section 5 with a choice of priors motivated by the discussion in section 6 of [34]. The general type of the prior distributions is similar to Model III: $\beta_j \sim N(\phi_j, \eta_j)$, $\gamma_j \sim N(\mu_j, \sigma_j)$, $\alpha \sim U(1, 1)$, but the choice of hyperparameters is not related to general male and female populations, but to married men and women. A possible way is to utilize the estimates of the marginal parameters obtained in Model I. This seems consistent with the results obtained by fitting parametric models to group mortality rates in Table 10 of [8] and comparing scale and shape parameters of Weibull mortality models for married men and women with those for the general population.

Model V is the one suggested in the present research project. It uses the conditional copula functions described in section 5. The choice of priors distributions and the choice of hyperparameters like the one used in Model IV. In this Model, we will not directly assume that marginal survival functions are Weibull distributed, but here we infer the marginal survival functions directly from the best fitted copula model which we select based on an approach borrowing from a statistical inference method. The choice of copula in the Model V is one of the issues to resolve in the research project. Joe in [20] and Nelsen in [27] list 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. in [5] 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 goodness-of-fit arguments may then be applied.

7 Concluding Remark

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) How to empirically estimate and test the presence of dependencies in multiple life times of

insured individuals?; (2) How to statistically select the appropriate copula model that can be used to approximate the true behaviour of dependence structure of spouses life times?; and (3) If there is such presence, what are its implications in computing insurance and/or annuity premiums, insurance reserves and/or pension liabilities?

Reference

- [1] Anderson, J.E., Louis, T.A., Holm, N.V., and Harvald, B. (1992) Time dependent association measures for bivariate survival distributions, *J. of the Amer. Statist. Assoc.*, 87, 419, 641-650.
- [2] Bowers, N., Gerber, H., Hickman, J.C., Jones, D.A., and Nesbitt, C. (1997) *Actuarial Mathematics*, Schaumburg, IL: Society of Actuaries.
- [3] Clayton, D.G. (1978) A model for association bivariate tables and its application in epidemiological studies of family tendency in chronic disease incidence, *Biometrika*, 65, 141-151.
- [4] Davidson, R. and MacKinnon, J. (1993) *Estimation and inference in econometrics*. Oxford University Press, Oxford.
- [5] Durrleman, V., Nikeghali, A. And Roncalli, T. (2000) "Which Copula is the Right One?," working paper, Groupe de Recherche Operationelle, Credit Lyonnais, France.
- [6] Embrechts, P., McNeil, A. J., and Straumann, D. (2002) Correlation and dependence in risk management: Properties and pitfalls. In: *Risk Management: Value at Risk and Beyond* (Ed. M. Dempster), Cambridge University Press, Cambridge, 176-223.
- [7] Embrechts, P., Lindskog, F., McNeil, A. (2003) Modelling dependence with copulas and applications to risk management In: *Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, Chapter 8, 329-384.
- [8] England and Wales. Death rates. (2002) In: *Mortality Statistics. Series DH1 no.35*, Office of National Statistics, London, <http://www.statistics.gov.uk/downloads/>
- [9] Frees, E., Carriere, J. and Valdez, E. (1996) Annuity valuation with dependent mortality, *Journal of Risk and Insurance*, Vol. 63, 229.
- [10] Frees, E. and Valdez, E. (1998) Understanding relationships using copulas, *Actuarial Research Clearing House, Proceedings*, 32nd Actuarial Research Conference, August 6-8, 1997, 5.
- [11] Genest, C. (1987) Frank's family of bivariate distributions. *Biometrika* 74, 3, 549-555.
- [12] Genest, C., Ghoudi, K., and Rivest, L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82, 3, 543-552.
- [13] Genest, C., and Rivest, L.-P. (1993) Statistical inference procedures for bivariate Archimedean copulas. *J. Amer. Statist. Assoc.* 88, 423, 1034-1043.
- [14] Genest, C., and Werker, B. J. M. (2002) Conditions for the asymptotic semiparametric efficiency of an omnibus estimator of dependence

- parameters in copula models. In: *Distributions with Given Marginals and Statistical Modelling* (Ed. C. M. Cuadras, J. Fortiana and J. A. Rodriguez-Lallena), Kluwer Academic Publishers, Dordrecht, 103-112.
- [15] Hauksson, H. A., Dacorogna, M. M., Domenig, T., MÅuller, U. A., and Samorodnitsky, G. (2001) Multivariate extremes, aggregation and risk estimation. *Quant. Finance* 1, 79-95.
- [16] Hickman, J.C. and Jones, D.A. (2002) Discussion of papers already published, *North Amer. Actuarial J.*, Vol. 6, 4, 113-114.
- [17] Hougaard, P., Harvald, B., and Holm, N.V. (1992) Measuring the similarities between the lifetimes of adult Danish twins born between 1881-1930, *J. of the Amer. Statist. Assoc.*, 87, 417, 17-24.
- [18] Hougaard, P. (1986) A class of multivariate failure time distributions, *Biometrika*, 73, 671-678.
- [19] Hougaard, P. (2000) *Analysis of Multivariate Survival Data*, Springer-Verlag, New York-Berlin-Heidelberg.
- [20] Joe, H. (1997) *Multivariate Models and Dependence Concepts*, Chapman and Hall, London.
- [21] Jouanin, J-F, Riboulet, G., and Roncalli, T. (2004) Financial applications of copula functions, In: *Risk Measures for the 21st Century*, ed. G.Szego, John Wiley.
- [22] Klugman, S. A., and Parsa, R. (1999) Fitting bivariate loss distributions with copulas. *Insurance Math. Econom.* 24, 139-148.
- [23] Lehmann, E. L. (1991) *Theory of Point Estimation*. Wadsworth & Brooks, Pacific Grove, California.
- [24] Lehmann, E. L., and Casella, G. (1998) *Theory of Point Estimation*, second ed. Springer Texts in Statistics. Springer-Verlag, New York.
- [25] Mashal, R., Naldi, M., and Zeevi, A. (2003) Extreme events and multiname credit derivatives, In: *Credit Derivatives: The Definitive Guide*, Ed. J. Gregory, RiskWaters Press.
- [26] McLeish, D. L., and Small, C. G. (1988) *The Theory and Applications of Statistical Inference Functions*, vol. 44 of *Lecture Notes in Statistics*. Springer-Verlag, New York.
- [27] Nelsen R. (1999) *An Introduction to Copulas*, Springer-Verlag, New York.
- [28] Oakes, D. (1982) A model for association in bivariate survival data. *J. Roy. Statist. Soc. Ser. B* 44, 3, 414-422.
- [29] Oakes, D. (1986) Semiparametric inference in a model for association in bivariate survival data. *Biometrika* 73, 2, 353-361.
- [30] Oakes, D. (1989) Bivariate survival models induced by frailties, *J. of the Amer. Statist. Assoc.*, 84, 487-493.
- [31] Oakes, D. (1994) Multivariate survival distributions. *J. Nonparametr. Statist.* 3, 3-4, 343-354.

- [32] Romeo, J.S., Tanaka, N.I., and Pedroso de Lima, A.C. (2004) Bivariate survival modeling: a Bayesian approach based on copulas, *Relatorio Tecnico, Universidade de Sao Paulo, Sao Paulo, Brasil.*, RT-MAE 2004-13.
- [33] Shemyakin, A. and Youn, H. (2000) Statistical aspects of joint-life insurance pricing, *1999 Proceedings of Amer. Stat. Assoc.*, 34-38.
- [34] Shemyakin, A. and Youn, H. (2001) Bayesian estimation of joint survival functions in life insurance, *In: Bayesian Methods with Applications to Science, Policy and Official Statistics*, European Communities, 489-496.
- [35] Shih, J.H. and Louis, T.A. (1985) Inferences on the association parameter in copula models for bivariate survival data, *Biometrika*, 51, 1384-1399.
- [36] Sklar, A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8, 229-231.
- [37] Spiegelhalter, D.J., Thomas, A., and Best, N.G. (2003) *WinBUGs 1.4. Computer Program*, Imperial College and MRC Biostatistics Unit, Cambridge.
- [38] Valdez E. A., and Chernih, A. (2003) Empirical estimation of dependence in a portfolio of insurance claims—preliminaries, working paper, UNSW, Sidney, AU.
- [39] Wang, S. (1997) Aggregation of correlated risk portfolios: Models and algorithms. Preprint, Casualty Actuarial Society (CAS).
- [40] Youn, H. and Shemyakin, A. (2001) Pricing practices for joint last survivor insurance, *Actuarial Research Clearing House*, Vol. 2001.1.
- [41] Youn, H., Shemyakin, A., and Herman, E. (2002) A re-examination of joint mortality functions, *North Amer. Actuarial J.*, Vol. 6, 1, 166-170.
- [42] Youn, H., and Shemyakin, A. (2004) Copula models of joint survival analysis, working Paper, University of St. Thomas, MN, USA.

YOGO PURWONO: Department of Management
Faculty of Economics, University of Indonesia
Depok, INDONESIA