

Analysis of bivariate tail dependence using extreme value copulas: An application to the SOA medical large claims database

Ana C. Cebrián,¹ Michel Denuit² and Philippe Lambert³

Abstract. The aim of this work is to analyze the dependence structure between losses and ALAE's relating to large claims using extreme value copulas. A procedure to select and estimate the copula based on a parametric estimation of the dependence function is proposed. An application to the evaluation of reinsurance premiums is performed in group medical insurance. This work clearly enhances the relevance of the copula-based approach to model claim amounts and their associated ALAE's.

Keywords: extreme value copulas, bivariate extreme value distributions, dependence function, excess-of-loss reinsurance premiums.

1 Introduction and Motivation

1.1 The situation under study

In this work we are going to deal with the Loss-ALAE data recorded in a database about medical insurance large claims, available from the website of the Society of Actuaries (<http://www.soa.org>).

ALAE's are type of insurance company expenses that are specifically attributable to the settlement of individual claims such as lawyers' fees and claims investigation expenses. The possible dependence between losses and ALAE's has to be accounted for (e.g. to price an excess-of-loss reinsurance treaty when the reinsurer shares the claim settlement costs; see Section 1.3 below).

1.2 Data description

The dataset we are going to work with is part of a larger database that records medical claim amounts exceeding \$25,000. The study by Grazier *et al.* (1997) collects information from 26 insurers. The 171,000 claims recorded are part of a database including about 3,000,000 claims over the years 1991-92. The total amount paid by the insurer for each claim (coded as Total), which behaviour was studied in Cebrián *et al.* (2001), is the sum of hospital charges (LOSS) and other

expenses (ALAE). In this study only the 1991 record relating to Plan type 4 is used. The size of this sample is 3,901. Univariate descriptive statistics for the the three variables are provided in Table 1.

	Mean	Q ₂₅	Q ₅₀	Q ₇₅	Min.	Max.	St. dev.
LOSS	63,450	31,360	42,310	66,130	25,000	1,430,000	74,665
ALAE	18,920	8,088	13,940	22,850	23	568,000	22,073
Total	82,370	43,040	58,560	88,500	25,320	1,530,000	85,523

Table 1. Claims summary statistics.

The relationship between LOSS and ALAE is illustrated in Figure 1, where the scatterplots of the two variables, in original and log-scale are shown. Although there is a clear dependence, it is not simple to model. Tests for independence (Pearson $\hat{r} = 0.380$, Spearman $\hat{\rho} = 0.437$ and Kendall $\hat{\tau} = 0.305$) all clearly yield to reject independence (p -values less than 10^{-2}).

1.3 Aim of the study

Our purpose is to price an excess-of-loss reinsurance treaty where the settlement costs are shared between the reinsurer and insurer on a proportional basis. Specifically, given a realization of (LOSS,ALAE), the reinsurance indemnity amounts to

$$g(\text{LOSS}, \text{ALAE}) = \tag{1}$$

$$\begin{cases} 0 & \text{if } \text{LOSS} \leq R, \\ \text{LOSS} - R + \frac{\text{LOSS} - R}{\text{LOSS}} \text{ALAE} & \text{if } \text{LOSS} > R. \end{cases}$$

The pure premium relating to this reinsurance treaty is expressed in terms of

$$\pi = E[g(\text{LOSS}, \text{ALAE})].$$

The pure reinsurance premium is obtained by multiplying π by the expected number of claims generated by the portfolio. Clearly, the joint distribution of the pair (LOSS,ALAE) has to be modelled to compute π . This is precisely the aim of the next section.

¹ Dpto. Métodos Estadísticos. Universidad de Zaragoza. Zaragoza 50009, Spain. Email: acebrian@unizar.es

² Institut des Sciences Actuarielles. Université Catholique de Louvain. B-1348 Louvain-la-Neuve, Belgium. Email: denuit@stat.ucl.ac.be

³ Institut de Statistique. Université Catholique de Louvain. B-1348 Louvain-la-Neuve, Belgium. E-mail: lambert@stat.ucl.ac.be

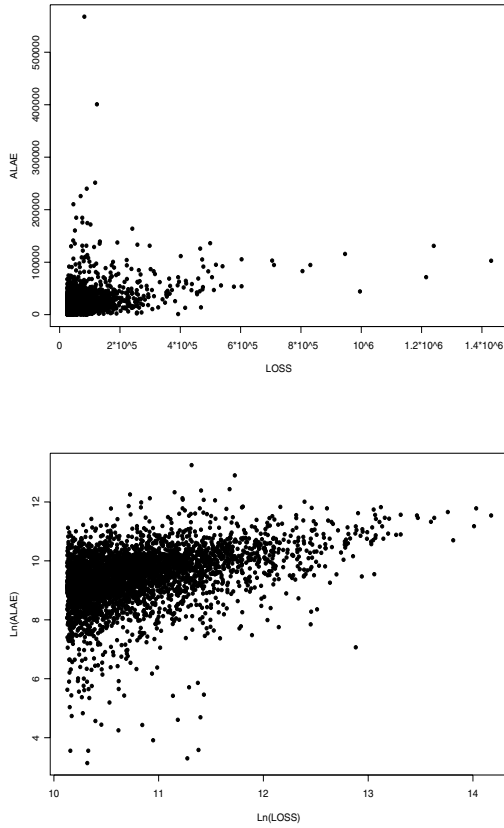


Figure 1. Bivariate plots of LOSS and ALAE on original and log-scale.

2 Modelling the dependence between LOSS and ALAE

2.1 Copulas

2.1.1 Definition

The stochastic behaviour of two random variables X_1 and X_2 with respective marginal cdf's F_1 and F_2 (assumed to be continuous) is appropriately described by their joint cdf

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

Since $F_1(X_1)$ and $F_2(X_2)$ are uniformly distributed between 0 and 1, their joint distribution for every $(u_1, u_2) \in [0, 1]^2$ can be expressed as

$$\begin{aligned} C(u_1, u_2) &= P[F_1(X_1) \leq u_1, F_2(X_2) \leq u_2] \\ &= P[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2)] \quad (2) \\ &= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \end{aligned}$$

where the F_i^{-1} 's, $i = 1, 2$, are the quantile functions associated with the F_i 's, i.e.

$$F_i^{-1}(p) = \inf\{x \in \mathbb{R} | F_i(x) \geq p\}, \quad p \in (0, 1).$$

Under this construction C is a distribution function of two random variables with unit uniform values. Such function is called *bidimensional copula*. As it can be seen from (2) the copula can be recovered from the knowledge of the joint distribution F and of its margins F_j , $j = 1, 2$. For a more exhaustive introduction to copulas we refer the interested reader to Nelsen (1999), Joe (1997) and for copulas with applications in actuarial sciences to Frees and Valdez (1998) and Klugman and Parsa (1999). Of course, the notion of copula extends into higher dimensions: n -dimensional copulas are joint distribution functions of n random variables with unit uniform marginals.

The interest of the copulas as an statistical tool originates from the result stated in the following section, referred to as Sklar's theorem.

2.1.2 Sklar's construction

Given a bidimensional distribution function F , with univariate margins F_1 and F_2 , there exists a copula C such that for all $(x_1, x_2) \in \mathbb{R}^2$,

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \quad (3)$$

Conversely, if C is a copula and F_1 and F_2 are distribution functions then the function F defined by (3) is a bidimensional distribution with margins F_1 and F_2 . Moreover, if F_1 and F_2 are both continuous, then C is uniquely defined.

2.1.3 More properties about copulas

It is not difficult to prove that C is a two-dimensional copula if and only if it is a function $C(u_1, u_2) : [0, 1]^2 \rightarrow [0, 1]$ that satisfies the following two conditions

- For every $u_1, u_2 \in [0, 1]$, $C(u_1, 0) = C(0, u_2) = 0$, $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$.
- For every $u_{11} \leq u_{12}$, $u_{21} \leq u_{22}$ with $u_{11}, u_{12}, u_{21}, u_{22} \in [0, 1]$, the following inequality holds:

$$C(u_{12}, u_{22}) - C(u_{12}, u_{21}) - C(u_{11}, u_{22}) + C(u_{11}, u_{21}) \geq 0$$

If C is considered to be a distribution function of two random variables U_1 and U_2 , the first condition ensures that U_1 and U_2 have uniform marginal distributions. The second condition, known as the rectangular inequality, simply requires that C is a valid distribution function, that is $P(u_{11} \leq U_1 \leq u_{12}, u_{21} \leq U_2 \leq u_{22}) \geq 0$.

Another important property is that copulas capture the dependence between two variables, regardless of the scale in which each variable is measured. More precisely, given two arbitrary functions g_1 and g_2 strictly increasing over the range of X_1 and X_2 , then the transformed variables $g_1(X_1)$ and $g_2(X_2)$ have the same copula than X_1 and X_2 .

Probably the most frequently used copulas are the Archimedean ones. Copulas in this family can be written in the following way

$$C(u_1, u_2) = \Phi^{-1}(\Phi(u_1) + \Phi(u_2))$$

where Φ is a decreasing convex function called the Archimedean generator, for reference see Nelsen (1999), chapter 4. Archimedean copulas are widely used in actuarial science; see e.g. Frees and Valdez (1998).

2.1.4 Copulas and dependence measures

The traditional measure of dependence, the Pearson correlation coefficient, presents some drawbacks as a measure for bivariate distributions (see Joe (1997), page 32, as well as Denuit and Dhaene in this volume). Measures with better properties include Kendall τ and Spearman ρ . The latter can be expressed uniquely in terms of the copula. Specifically, the following representations hold true:

$$\begin{aligned}\rho &= 12 \int_0^1 \int_0^1 \{C(u, v) - uv\} dudv \\ \tau &= 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.\end{aligned}$$

2.2 Extreme value copulas

In the situation of interest in this paper, the combination of the copula construction with extreme value theory seems promising. This is precisely the aim of this section.

2.2.1 Definition

A random vector (X_1, X_2) is said to conform to an Extreme Value Distribution (EVD, in short) with unit exponential margins if,

$$P[X_1 > y_1] = \exp(-y_1) \text{ and } P[X_2 > y_2] = \exp(-y_2)$$

for $y_1, y_2 > 0$ and the joint survival function $\bar{F}(x_1, x_2) = P(X_1 > x_1, X_2 > x_2)$ verifies the scaling property,

$$\bar{F}^n(y_1, y_2) = \bar{F}(ny_1, ny_2)$$

for any y_1, y_2 and integer $n \geq 1$. For more details, see e.g. Tawn (1988). A random couple obeys to some EVD if and only if, the survival function can be expressed as

$$\bar{F}(y_1, y_2) = \exp \left[-(y_1 + y_2) A \left(\frac{y_2}{y_1 + y_2} \right) \right]$$

for any $y_1, y_2 > 0$, where

$$A(w) = \int_0^1 \max[(1-w)q, w(1-q)] dL(q)$$

for some positive finite measure L on $[0, 1]$. The function A is called the *dependence function*.

Expressing this relationship in terms of uniform margins, we obtain

$$C(u_1, u_2) = \exp \left\{ [\ln(u_1) + \ln(u_2)] A \left[\frac{\ln(u_2)}{\ln(u_1) + \ln(u_2)} \right] \right\}$$

for $0 \leq u_1, u_2 \leq 1$. This function is known as the *extreme value copula* (EV copula, in short) and it allows to model dependence between the components of a random couple which represent the largest values of two characteristics observed over the same period of time.

One important property of EV copulas is that they are max-stable, meaning that, if $(X_{11}, X_{21}), (X_{12}, X_{22}), \dots, (X_{1n}, X_{2n})$ are iid random pairs from an EV copula C and $M_n = \max(X_{11}, X_{12}, \dots, X_{1n})$ and $N_n = \max(X_{21}, X_{22}, \dots, X_{2n})$, the copula associated with the random pair (M_n, N_n) is also C .

2.2.2 Properties of the dependence function

The dependence function A involved in an EV copula C must satisfy the following properties:

- $A(0) = A(1) = 1$.
- $\max(w, 1-w) \leq A(w) \leq 1$ for $0 \leq w \leq 1$. Moreover, if $A(w) = 1$ then (X_1, X_2) are independent and if $A(w) = \max(w, 1-w)$, then (X_1, X_2) are perfectly dependent (or comonotonic).
- $A(w)$ is a convex function in the region $0 \leq w \leq 1$.

2.2.3 Some examples

Gumbel copula The dependence function is

$$A(t) = (t^r + (1-t)^r)^{1/r}$$

with $r \geq 1$. Gumbel copula is the only one that belongs to both Archimedean and Extreme Value families.

Logistic model The dependence function is

$$A(w) = [(1-w)^r + w^r]^{1/r}$$

with $r \geq 1$. The corresponding copula function is given by

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

Independence and complete dependence correspond to $r = 1$ and $r = \infty$ respectively. In this model the variables are exchangeable.

Asymmetric logistic model The dependence function is

$$A(w) = [\theta^r (1-w)^r + \phi^r w^r]^{1/r} + (\theta - \phi)w + 1 - \theta$$

with $\theta \geq 0, \phi \leq 1$, and $r \geq 1$ and the copula function

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

This model is very flexible and contains several of the existing models such as the logistic ($\phi = \theta = 1$), the biextremal, the dual of the biextremal, the Gumbel, as well as a mixture of logistic and independence models. Complete dependence corresponds to $\theta = \phi = 1$ and $r = \infty$ whereas independence corresponds to $\theta = 0$ or $\phi = 0$ or $r = 1$.

Mixed model The dependence function is

$$A(w) = \theta w^2 - \theta w + 1$$

and the corresponding copula function is

$$C(u_1, u_2) = u_1 u_2 \exp \left\{ -\theta \frac{\ln u_1 \ln u_2}{\ln(u_1 u_2)} \right\}$$

with $0 \leq \theta \leq 1$. Independence corresponds to $\theta = 0$ but we cannot have complete dependence. In this model the variables are exchangeable.

2.2.4 Kendall's tau

Some important properties of the variables can be formulated in terms of the dependence function A . The latter plays a role comparable to the generator of Archimedean families. In particular Kendall's τ is given by

$$\tau = \int_0^1 \frac{t(1-t)}{A(t)} A''(t) dt. \quad (4)$$

For the models presented above, τ_A is given by

1. Logistic model:

$$\tau_A = 1 - 1/r.$$

2. Mixed model:

$$\tau_A = \frac{8 \arctan \sqrt{\theta/(4-\theta)}}{\sqrt{\theta(4-\theta)}} - 2.$$

3. Asymmetric Logistic model:

$$\tau_A = \int_0^1 \frac{w(1-w)(r-1)\theta^r \phi^r (w-w^2)^{r-2}}{[\theta^r(1-w)^r + \phi^r w^r]^2} dw.$$

For the asymmetric logistic model, a closed expression for τ_A does not exist. Therefore, one needs to resort to numerical methods to evaluate τ_A .

3 Estimation of the copula C and its probability integral transform K

3.1 Copula probability integral transform

Genest and Rivest (2000) suggest a general formula for computing the distribution function K of the random variable

$$V = F(X_1, X_2) = C(F_1(X_1), F_2(X_2)).$$

In contrast to the univariate case, it is not true that V is uniformly distributed over $[0, 1]$. In fact, any distribution function K for $F(X_1, X_2)$ has to fulfill the inequalities

$$v \leq K(v) \leq 1$$

for $0 \leq v \leq 1$. Moreover, $K(v) = v$ and $K(v) = 1$ correspond to the two extreme cases where the random variables X_1 and X_2 are in perfect positive dependence (comonotonicity) and in perfect negative dependence (countermonotonicity), respectively.

3.2 Computing K for bivariate extreme value copulas

It can be proved that

$$K(v) = v - \lambda(v)$$

with

$$\lambda(v) = - \int_v^1 \frac{\partial C(x_1, x_2^{x_1 v})}{\partial x_1} dx_1$$

and $x_2^{x_1 v}$ the root of the equation

$$C(x_1, x_2^{x_1 v}) = v.$$

For the EV copulas, we obtain

$$\lambda(v) = -(1 - \tau_A)v \log(v)$$

where τ_A is given by (4) so that

$$K(v) = v - (1 - \tau_A)v \log(v). \quad (5)$$

Obviously, in order to estimate K via (5), only τ_A needs to be estimated from the data.

3.3 Parameter estimation

Given a random sample of size n , $(X_{11}, X_{21}), \dots, (X_{1n}, X_{2n})$ say with copula function C describing the dependence between each pair of variables and marginal density functions f_1 and f_2 , the corresponding joint probability density function is

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)C''(F_1(x_1), F_2(x_2); p)$$

where p is a vector of parameters and

$$C''(u_1, u_2; p) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2; p).$$

Assuming known margins, only C'' depends on the parameters to be estimated and the function

$$L(p) = \sum_{i=1}^n \log C''(u_{1i}, u_{2i}; p) \quad (6)$$

(where $u_{ji} = F_j(x_{ji})$ for $j = 1, 2$) must be maximized, using for example algorithms based on quasi-Newton methods.

3.4 Nonparametric estimation of C

To check the validity of the parametric models, it would be desirable to compare them with nonparametric estimators. In particular, we make use of the one proposed by Scaillet (2000). This author developed a kernel based approach without putting any particular parametric a priori on the dependence structure between margins.

To estimate the copula function at given points, $\{u_i = (u_{1i}, u_{2i}), i = 1, \dots, n\}$, with $u_{ji} \in (0, 1)$, a simple plug-in method in expression (2) is used, resulting in,

$$\hat{C}(u_1, u_2) = \hat{F}(\hat{F}_1^{-1}(u_1), \hat{F}_2^{-1}(u_2))$$

where \hat{F} , \hat{F}_1 and \hat{F}_2 represent estimators of the bivariate distribution function F and of the corresponding marginal distributions F_1 and F_2 . In this work, the empirical distribution function will be used to estimate each of the marginal functions F_1 and F_2 , but kernel or parametric estimators (if proper distributions are found to fit the margins) could also be used.

Note that a kernel estimator, based on the pseudo-observations

$$c_i = \hat{C}(\hat{F}(x_{1i}), \hat{F}(x_{2i})),$$

can be used to estimate the function K .

4 LOSS-ALAE analysis

4.1 Estimation results

Table 2 summarizes estimations for the three parametric models discussed in Section 2.3. Not only the value of the objective function (6) is given, but also an AIC goodness-of-fit measure,

$$AIC = -2L + 2n_p/n$$

where L has been defined in (6) and n_p is the number of parameters.

Model	Parameters	$\hat{\tau}_A$	(6)	AIC
Logistic	$\hat{\tau} = 1.406$	0.289	-7,321.3	14,643
Asym. Log.	$\hat{\theta} = 0.983, \hat{\phi} = 0.825,$ $\hat{\tau} = 1.486$	0.288	-7,315.7	14,631
Mixed	$\hat{\theta} = 0.738$	0.289	-7,327.3	14,654

Table 2. Summary of dependence function estimation.

In Figure 2 the three estimations of the dependence function are represented. It can be seen that the three models provide similar results and it seems all of them are able to represent the dependence structure. Also the estimated Kendall τ is almost the same for the three models.

In plots of Figures 3 and 4 we represent for each model the differences between the nonparametric and the constrained estimations of the copula function. From the upper left graph, the difference between the nonparametric estimation of the copula and its independent counterpart is always positive, indicating a positive dependence between LOSS and ALAE. The results from all the parametric models are rather similar, but the mixed model looks slightly better than the others (since the differences are smaller for this model).

In Figure 5 we compare the parametric estimations of K with a kernel estimation of this function calculated from the pseudo-sample $\hat{c}_i = \hat{C}(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}))$. Again the independent model is rather bad, producing large departures from the nonparametric estimation. No significant differences are observed between the three parametric models.

A final possibility to check the model is to compare the quantiles of the parametric and the nonparametric estimation of the copula using qq-plots. The plots corresponding to the

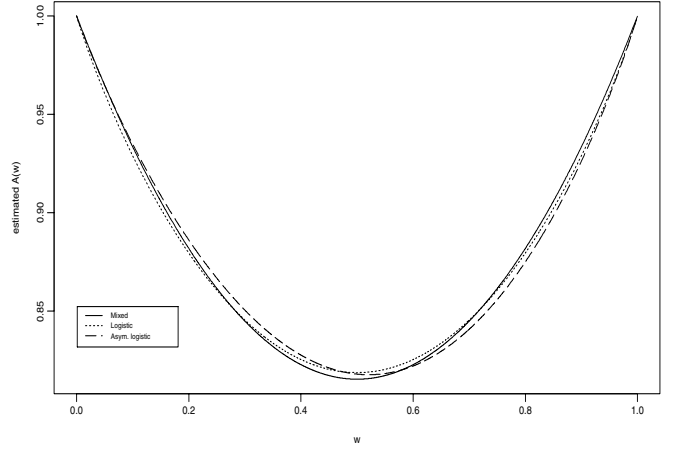


Figure 2. Parametric estimations of the dependence function A .

independent and the parametric models are shown in Figure 6. Again, independence excepted, the three parametric models behave well on the data set under study.

4.2 Goodness-of-fit tests

In addition to the graphical tools, some more objective measures and tests are desirable to complete the goodness-of-fit analysis. First, we performed a general test proposed by Ghoudi et al. (1998) to verify whether the copula belongs to the EV family without specifying a particular dependence function. It is based on the distribution function of the copula, K ; more precisely, the hypothesis $H_0 : K(v) = K_\tau(v)$ with K_τ of the form (5) for some $\tau \in (0, 1)$ is checked using the test statistic

$$S_n = \frac{8}{n(n-1)} \sum_{i \neq j}^n \delta_{ij} - \frac{9}{n(n-1)(n-2)} \sum_{i \neq j \neq k}^n \delta_{ij} \delta_{ik} - 1,$$

where

$$\delta_{ij} = \delta[(X_{1i}, x_{2i}), (X_{1j}, x_{2j})]$$

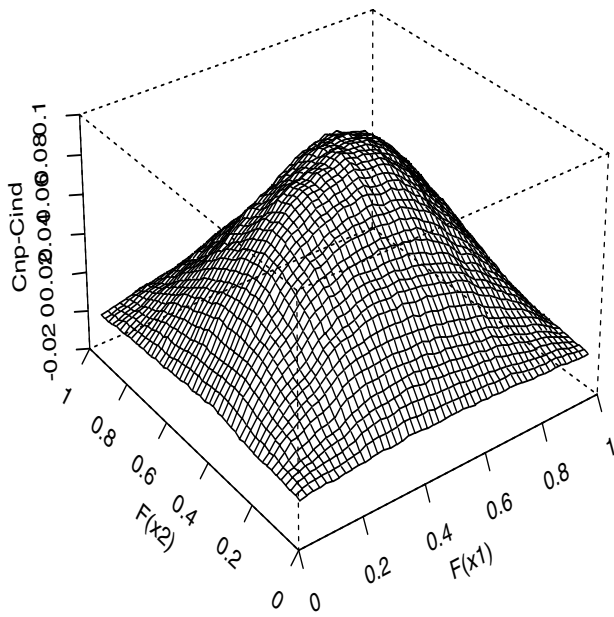
and

$$\delta[(x_1, y_1), (x_2, y_2)] = \begin{cases} 1 & \text{if } x_1 \geq x_2 \text{ and } y_1 \geq y_2 \\ 0 & \text{otherwise.} \end{cases}$$

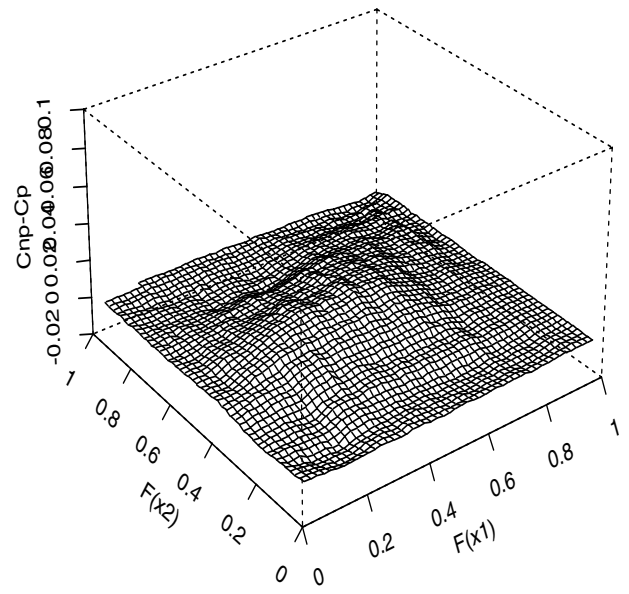
The standardized test statistic $|S_n|/\hat{V}ar(S_n)$ is asymptotically normal and a jackknife method is suggested to estimate the variance,

$$\hat{V}ar(S_n) = \frac{n-1}{n} \sum_{i=1}^n [S_n^{-i} - S_n]^2.$$

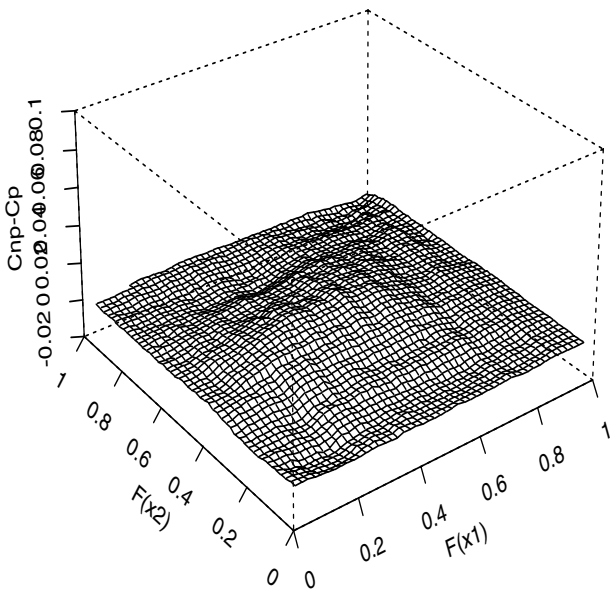
Satisfactory results were found with an observed value of S_n equal to 1.016 and the corresponding p-value equal to 0.31.



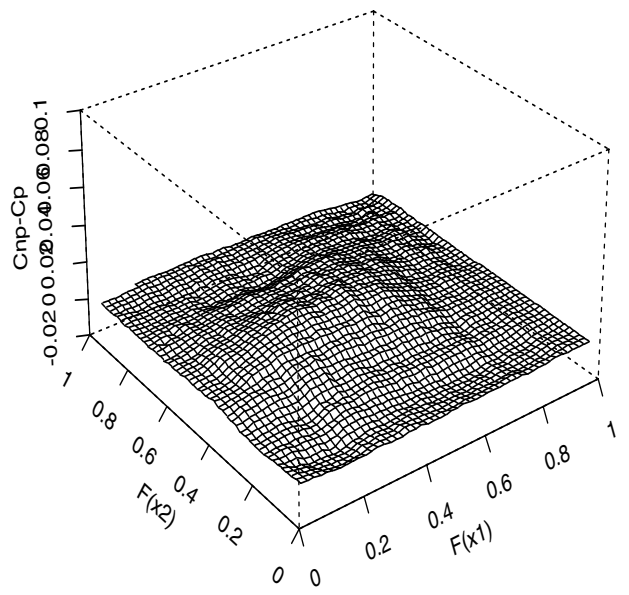
Independent model



Logistic model



Mixed model



Asymmetric Logistic model

Figure 3. Differences between the nonparametric and the Independent and Mixed estimation of the copula.

Figure 4. Differences between the nonparametric and the Logistic and Asymmetric Logistic estimation of the copula.

After that, we checked the goodness of fit of the different dependence functions. We used a Kolmogorov-Smirnov type test to compare the empirical function distribution of the sample $\hat{C}(\hat{F}_1(x_{1i}), \hat{F}_2(x_{2i}))$ to the estimation of the function K estimated under one of the parametric models presented in Section 2.3 and evaluated in the same observations. As we can see in Table 3, according to the K-S test no model is rejected with any usual confidence level.

Test	Logistic	Asym. Logistic.	Mixed
K-S test statistic	0.0121	0.0118	0.0118
p-value	0.61	0.64	0.64

Table 3. Goodness of fit tests.

5 Reinsurance premiums

Let us now analyze the impact of the dependence structure on premium valuation in reinsurance treaties. We consider a reinsurance treaty on a policy with unlimited liability and insurer's retention R . Assuming a prorata sharing of expenses, the reinsurer's payment for a given realization of $(\text{LOSS}, \text{ALAE})$ is described by the function (1).

As it is demonstrated in Joe (1997), any extreme value copula verifies the association property. This means that if g_1 and $g_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ are non-decreasing functions, the inequality

$$E[g_1(X_1, X_2)g_2(X_1, X_2)] \geq E[g_1(X_1, X_2)]E[g_2(X_1, X_2)]$$

holds true. Association is a rather strong dependence property. In particular it implies positive quadrant dependence, that is

$$P(X_1 > x_1, X_2 > x_2) \geq P(X_1 > x_1)P(X_2 > x_2).$$

From Denuit, Dhaene and Ribas (2001), it is known that the premium π computed under the independence hypothesis for $(\text{LOSS}, \text{ALAE})$ will be smaller than the premium computed under an extreme value copula.

The results in Table 4 provide the premiums the reinsurer would have assessed to cover LOSS and ALAE charges according to various insurer's retention. Three situations have been considered:

1. Assuming independence, an estimator of the premium π is

$$\hat{\pi} = \frac{1}{n^2} \sum_{t=1}^n \sum_{t'=1}^n g(\text{LOSS}_t, \text{ALAE}_{t'}).$$

Since g is supermodular, this value is a lower bound to the unknown π under positive quadrant dependence for LOSS and ALAE . It is worth to mention that Figure 3 (upper left panel) suggested such a positive dependence.

2. Applying the classical comonotonic approximation for $(\text{LOSS}, \text{ALAE})$ we obtain,

$$\hat{\pi} = \frac{1}{n} \sum_{t=1}^n g\left(\text{LOSS}_t, \hat{F}_{\text{ALAE}}^{-1}\left[\hat{F}_{\text{LOSS}}(\text{LOSS}_t)\right]\right).$$

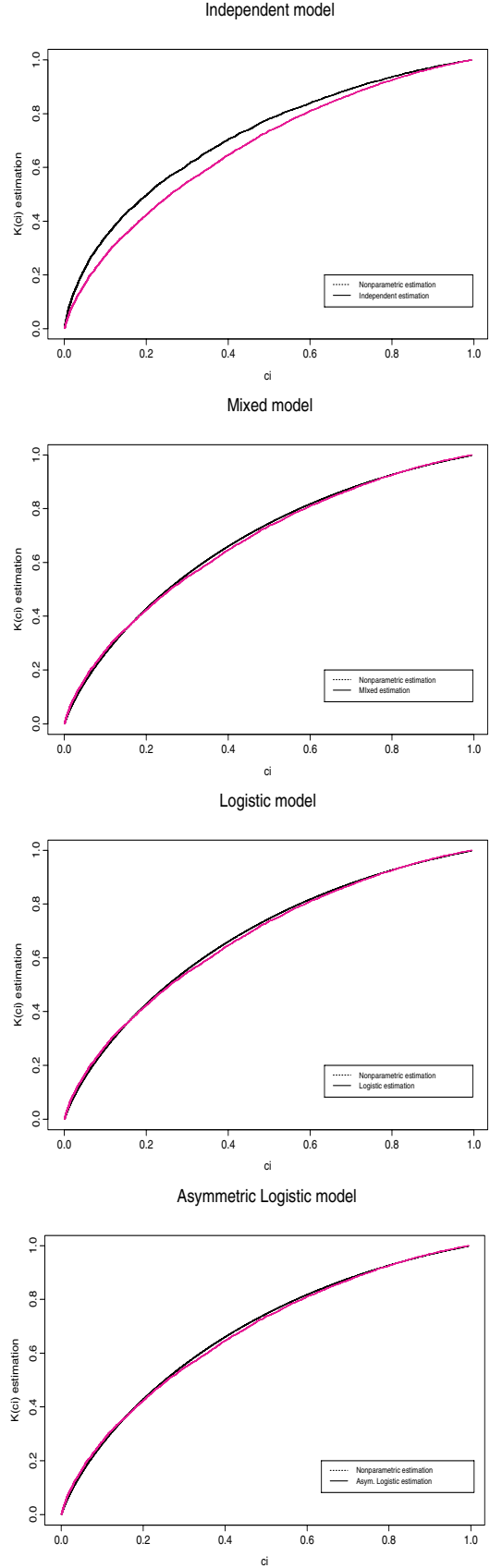


Figure 5. Estimation of K .

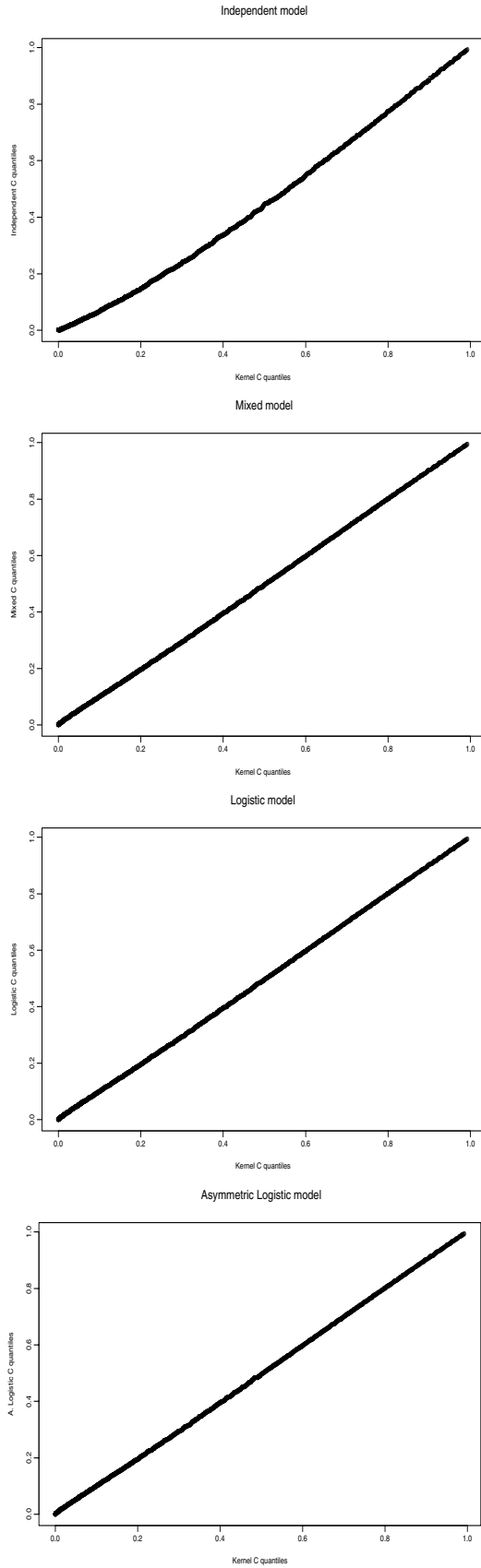


Figure 6. QQ-plots of the copula quantiles.

Because of the supermodularity of g , this value is an upper bound π .

3. If we consider the dependence structure, we have two possibilities,
 - a.- either we do not specify any model for the dependence, i.e. we compute the empirical analogue of π given by

$$\hat{\pi} = \frac{1}{n} \sum_{t=1}^n g(LOSS_t, ALAE_t);$$

- b.- or we estimate π under one of the copula models presented in Section 2.3. Under these models, the pure premium can be estimated by

$$\hat{\pi} = \hat{E}[g(LOSS, ALAE)] = \int_{25,000}^{\infty} \int_0^{\infty} g(l, a) \hat{f}_{LOSS, ALAE}(l, a) dl da;$$

where

$$\hat{f}_{LOSS, ALAE}(l, a) = \hat{f}_{LOSS}(l) \hat{f}_{ALAE}(a) \hat{C}''[\hat{F}_{LOSS}(l), \hat{F}_{ALAE}(a)].$$

Since no parametric models have been posed for the margins, the marginal densities are estimated using kernel methods. Having an estimator of the bivariate density function available, the double integral is then calculated using numerical procedures. Since the integrand is a rather complicated function, a powerful software to perform analytical calculations is necessary; in this case MathCad 2000 has been used. Although the procedure is not complicated, it is rather time consuming, mainly due to the evaluations of the kernel estimators of the density and distribution functions. The code is available from Ana Cebrián upon request.

The $\hat{\pi}$ values for $R = 25000, 50000, 100000, 500000, 1000000$ estimated under the different models are shown in Table 4. We see that substantial mispricing could result from the independence hypothesis, while the comonotonic approximation is too conservative. Thus, as expected, independence generates lower premiums than those taking into account dependence in the data, which themselves are smaller than those based on the comonotonic assumption. Two estimators considering dependence are shown, the one based on the empiric dependence of the data and the one based on the extreme value copula with the mixed model, that provided the best fit to the bivariate sample. Comparing both estimators, we observe that they are rather similar for small R values, but differ substantially for higher values of R . It must be remembered that, in fact, the main drawback of nonparametric estimation for π is that for high R values the sample size is not large enough to get reliable estimations.

6 Conclusions

To conclude we can say that the results show the importance to consider the dependence structure of the two variables: substantial mispricing can result from the independence assumption, while the comonotonic approximation seriously overestimates the premium.

R	25,000	50,000	100,000	500,000	1,000,000
Independence	46,365.7	25,642.3	12,785.5	1,247.1	230.3
empirical	48,350.6	27,639.1	13,842.3	1,336.6	244.7
(# obs. LOSS > R)	(3901)	(1516)	(470)	(17)	(3)
Mixed model	48,357.5	27,804.6	14,188.3	1,431.6	269.3
Comonotonicity	50,318.8	29,760.7	15,374.9	1,570.6	302.1

Table 4. Pure premiums for a reinsurance treaty with retention R .

Given the characteristics of the analyzed variables, the use of EV copulas is intellectually satisfying. Moreover, the SOA medical large claims database supported these models. Therefore, EV copulas should be considered for bivariate reinsurance pricing in the future.

ACKNOWLEDGEMENTS

Ana Cebrián thanks the Université catholique de Louvain, Louvain-la-Neuve, Belgium, for financial support through a FSR research grant.

REFERENCES

- [1] Capérea, P., Fougères, A.-L. and Genest. (1997). A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**,3, 567-577.
- [2] Cebrián, A., Denuit, M., & Lambert, Ph. (2003). Generalized Paréto fit to the society of Actuaries' large claims database. *North American Actuarial Journal* **7**, 18-36.
- [3] Denuit, M., Dhaene, J., and Ribas, C. (2001). Does positive dependence between individual risks increase stop-loss premiums? *Insurance: Mathematics and Economics* **28**, 305-308.
- [4] Frees, E.W. and Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* **2**, 1-15.
- [5] Genest, C. and Rivest, L.P. (2000). On the Multivariate Probability Integral Transformation. *Statistics and Probability Letters* **53**, 391-399.
- [6] Ghoudi, K., Khoudraji, A. and Rivest L.P. (1998). Propriétés Statistiques des Copules de Valeurs Extrêmes Bidimensionnelles. *The Canadian Journal of Statistics* **26** No 1, 187-197.
- [7] Grazier, K.L., and G'Sell Associates (1997). *Group Medical Insurance Large Claims Database and Collection*. SOA Monograph M-HB97-1.
- [8] Joe, H. (1997). *Multivariate Models and Dependence Concepts*. Chapman & Hall, London
- [9] Klugman, S. A. and Parsa, R. (1999). Fitting Bivariate Loss Distributions with Copulas. *Insurance: Mathematics and Economics* **24**, 139-148.
- [10] Nelsen, R.B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics **139**, Springer Verlag, New York.
- [11] Tawn, J. (1988). Bivariate Extreme Value Theory: Models and Estimation. *Biometrika* **75**, 397-415.
- [12] Scaillet, O. (2000). Nonparametric Estimation of Copulas for Time Series. Downloadable on <http://www.hec.unige.ch>.