STOCHASTIC MODELLING OF INSURANCE

Konstantins Didenko[1], Vladimirs Jansons[2], Vitalijs Jurenoks[3], Gaida Pettere[4]

[1,2,3]Technical University of Latvia, Faculty of Engineering Economics
1658 Riga, Kalku 1, Latvia
[4]Technical University of Latvia, Faculty of Applied Mathematics
1650 Riga, Meza 1, Latvia

vladjans@latnet.lv (Vladimirs Jansons)

Abstract

From ancient time one of the most important problem in non-life insurance companies has been how to calculate incurred but not reported claim provisions (IBNR). When accounting period ends, premiums have received, but there is always situation that claims have incurred but still not reported because it is not always possible to report them in the same day of occurrence. To have right accounts money from received premiums have to be reserved for such claims. There are many classical methods how to do that. All these methods are based on different coefficient calculations and deal with classical development triangle. Nowadays new stochastic methods actual have become to calculate the mentioned reserves (Charpentier, 2004 [1]). Every claim relates with loss adjusted expenses which are needed for insurance company to be able to pay out claim. We have gone further to look on each claim in non-life insurance like event consists from three important characteristics: the claim size, the allocated loss adjusted expense and the development time (time from the moment of claim occurrence until its settlement). Our interest is concentrated to the joint study of all three random variables what is very important for IBNR claim provision calculations. Using multivariate distributions with different dependence structures, we statistically evaluated IBNR claim provision.

Keywords: Non-Life Insurance, Insurance, Copula, Modelling, Algorithm.

Presenting Author's biography

Vladimirs Jansons was born in Daugavpils, Latvia and is a graduate of the University of Latvia, where he studied mathematical science and obtained his degree in 1970. For eight years he has worked in the University of Latvia Computing Centre. Since 1978, he has been lecturing at Riga Technical University, in 1983 was awarded the doctoral degree in the mathematical science. The main field of research pursued is statistical modelling and optimization of technical and economic systems.

General

Understanding and quantifying dependence is at the core of all modelling efforts in financial econometrics. The linear correlation coefficient, which is the far most used measure to test dependence in the finance and also elsewhere, is only a measure of linear dependence. This means that it is a meaningful measure of dependence if asset returns are well represented by an elliptical distribution. Outside the world of elliptical distributions, however, using the linear correlation coefficient as a measure of dependence may lead to misleading conclusions. Hence, alternative methods for capturing co-dependency should be considered. One class of alternatives is copula-based dependence measures. Copula like a tool for modelling different dependence structures more and more widely have been used in different fields of research: finance, insurance, risk theory. Copulas were introduced in 1959 by Sklar [2] but only in 1997 Wang introduced copula models in insurance. Many conferences, seminars have been about that topic from that time [3,4,8]. Most commonly copulas in non-life insurance are used to model claim sizes and allocated loss adjusted expenses (Frees, E. W., Valdez, E. A. (1998) [7]), evaluate economic capital (Tang, Valdez (2006) [8], for combining different risks (Clemen, Reilly (1998) [5].

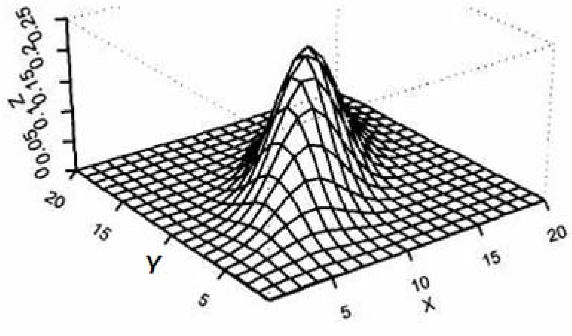

Fig. 1 The density of the bivariate Gaussian distribution with correlation cor = 0.7 and standard normal marginals

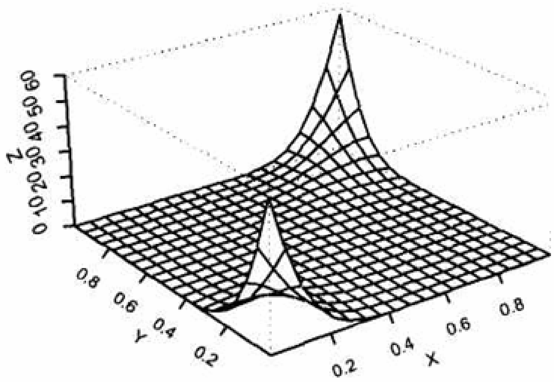

Fig. 2 The density of the bivariate Gaussian copula with parameter ρ = 0.7

1 Copula Technique

1.1 Definition of copula

In the real world, there is often a non-linear dependence between different variables and correlation cannot be an appropriate measure of co-dependency. Therefore linear Spearmen's correlation coefficient is a limited measure of dependence. It is not surprising that alternative methods (the copula method) for capturing co-dependency have been considered. The concept of copulas comes from Sklar [2] in 1959.

In rough terms, a copula is a function:

$$C : [0,1]^n \to [0,1], \tag{1}$$

with certain special properties.

1.2 Gaussian copula

The Gaussian copula is given by

$$C_\rho(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} a \cdot \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dxdy$$

where ρ is the parameter of the copula, and Φ^{-1} is the inverse of the standard univariate Gaussian distribution function and $a = 1/(2\pi(1-\rho^2)^{1/2}$.

1.3 Student's t-copula

The Student's t-copula allows for joint fat tails and an increased probability of joint extreme events compared with the Gaussian copula. This copula can be written as

$$C_{\rho,v}(u,v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} a \cdot \left(1 + \frac{s^2 - 2\rho st + t^2}{v(1-\rho^2)}\right)^{-\frac{v+2}{2}} dsdt$$

where ρ and *v* is the parameter of the copula, and t_v^{-1} is the inverse of the standard univariate t-distribution function with v degrees of freedom, expectation 0 and variance v/(v-2).

1.4 Sklar's theorem

The most useful results of copula theory is Sklar's theorem. Let F be a joint multivariate distribution with marginals F_1 and F_2. Then, for any x_1, x_2 there exists a copula C such that

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{2}$$

Furthermore, if marginals F_1 and F_2 are continuous, the copula C is unique. Conversely, if F_1 and F_2 are marginal distributions and C is a copula, then the function F defined by $C(F_1(x_1), F_2(x_2))$ is a joint distribution function with marginals F_1 and F_2. If we have a random vector $X = (X_1, X_2)$ the copula of their joint distribution function may be extracted from equation (1):

$$C(u_1,u_2)=F(F_1^{-1}(u_1),F_2^{-1}(u_2)), \qquad (3)$$

where the F_1^{-1}, F_2^{-1} are the quantile functions of the margins.

1.5 Definition of a two-dimensional copula

A two-dimensional copula is a two-dimensional distribution function C with uniformly distributed marginals U(0,1) on [0,1]. Thus a copula is a function C:

$$C:[0,1]^2 \rightarrow [0,1], \qquad (4)$$

satisfying the following three properties:

1. For every u, v from [0, 1]:

C(u, 0) = C(0, v), C(u, 1) = u and C(1, v) = v.

2. C(u, v) is increasing in u and v.

3. For every u_1, u_1, v_1, v_2 from [0, 1] with $u_1 \le u_2$ and $v_1 \le v_2$ we have:

$$C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1)\ge 0$$

Condition 1 provides the restriction for the support of the variables and the marginal uniform distribution. Conditions 2 and 3 correspond to the existence of a nonnegative "density" function.
Sklar's theorem provides a decomposition of the joint distribution into marginal features (that are F_1 and F_2) and dependence features (represented by copula C). The two variables X and Y are independent if and only if F(X) and G(Y) are independent. The independence condition can be written in terms of copula as C(u, v) = uv. When C(u,v) ≠ uv, the variables X and Y (or F(X), G(Y)) are dependent and the dependence summarized in the copula depends on the variables up to (nonlinear) increasing transformation of the variables. It is important to see if they are more or less dependent, and the "sign" of the dependence. The comparison of dependence can be based on the usual first order dominance stochastic ordering applied to copula. In most financial cases we can effectively use Archimedean copulas. The Archimedean copulas provide analytical tractability and a large spectrum of different dependence measure. These copulas can be used in a wide range of applications for the following reasons:

a) the simplicity with which they can be constructed;

b) the many parametric families of copulas belonging to this class;

the great variety of different dependence structures.

1.6 Definition of an Archimedean copula

Definition of Archimedean copula's generator. Let us consider a function φ: [0;1]→[0;1] which is continuous, strictly decreasing, convex and for which φ(0) = ∞ and φ(1)=0. We then define the pseudo inverse of $\varphi^{[-1]}$: [0; ∞] → [0;1] such that:

$$\varphi^{[-1]}(t)=\begin{cases}\varphi^{-1}(t) & 0\le t\le\varphi(0)\\ 0 & \varphi(0)\le t\le\infty\end{cases}.$$

As φ is convex, the function C: $[0; 1]^2 \rightarrow [0; 1]$ defined as $C(u_1, u_2)= \varphi^{-1}[\varphi(u_1) + \varphi(u_2)]$ is an Archimedean copula and φ is called the generator of the copula [2,7]. In case of the multivariate extension for all n ≥ 2, the function C: $[0; 1]^n \rightarrow [0; 1]$ defined as

$$C(u_1, ..., u_n) = \varphi^{-1}[\varphi(u_1) +. . . + \varphi(u_n)], \qquad (5)$$

is an n-dimensional Archimedean copula if and only if φ^{-1} is completely monotone on [0, ∞).

Alternatively we can say that it is a multivariate distribution function defined on the unit cube $[0; 1]^n$. Copula functions are well studied object in the statistical literature. These functions have been introduced to model a joint distribution once the marginal distributions are known. When multivariate normal distribution is rejected by data, the copula may be used as an important alternative to represent the dependence in joint distributions.

1.7 Copula-based dependence measures

Since the copula of a multivariate distribution describes its dependence structure, it might be appropriate to use measures of dependence which are copula-based. The bivariate concordance measures Kendall's tau and Spearman's rho, as well as the coefficient of tail dependence, can, as opposed to the linear correlation coefficient, be expressed in terms of the underlying copula alone.

Kendall's tau. Kendall's tau of two variables X and Y is

$$\rho_\tau(X,Y)=4\int_0^1\int_0^1 C(u,v)dudv-1,$$

where C(u, v) is the copula of the bivariate distribution function of X and Y. For the Gaussian and Student's t-copulas and also all other elliptical copulas, the relationship between the linear correlation coefficient and Kendall's tau is given by

$$cor(X,Y)=\sin\left(\frac{\pi}{2}\cdot\rho_\tau\right).$$

Spearman's rho of two variables X and Y is given by

$$\rho_S(X,Y)=12\int_0^1\int_0^1 C(u,v)dudv-3.$$

For the Gaussian and Student's t-copulas, we have that the relationship between the linear correlation coefficient and Spearman's rho is

$$cor(X,Y) = 2\sin\left(\frac{\pi}{6}\rho_S\right).$$

Both $\rho_\tau(X, Y)$ and $\rho_S(X, Y)$ may be considered as measures of the degree of monotonic dependence between X and Y, where as linear correlation measures the degree of linear dependence only. These measures are invariant under monotone transformations, while the linear correlation generally isn't. Hence, according to Embrechts (2003) [6] it is better to use these measures than the linear correlation coefficient. Therefore copulas have become a powerful tool for modeling dependence between random variables. Also copula methodology is effective for modeling joint distributions with fat tails. Fat tails in financial return data have been documented in numerous real cases. Joint distribution on financial data returns is very important issue in risk management.

1.8 Simulating algorithm from the copula

If in addition to equation (5) function ϕ equals the inverse of Laplase transform of a distribution function G on R+ satisfying G(0) = 0, the following algorithm can be used for simulating from the copula (Marshall and Olkin, 1988) [9]:

I. Simulate a variate X with distribution function G such that the Laplace transform of G is the inverse of the generator.

II. Simulate n independent variates $V_1,\dots, V_n$.

III. Return U = $(\phi^{-1}(-\log(V_1)/X),\dots, (\phi^{-1}(-\log(V_n)/X))$.

Frank, Clayton and the Gumbel copula can be simulated using this procedure. For example for the Clayton copula simulation algorithm becomes

I. Simulate a Gamma variate X~Gamma(1/θ, 1).

II. Simulate n independent standart uniforms variates V1, ..., Vn.

III. Return U = $((1-\log(V_1)/X)^{-1/\theta},\dots, (1-\log(V_n)/X)^{-1/\theta})$.

Functions needed for simulating algorithm are shown in Tables 1, 2.

Tab. 1.

Tab. 1 Frank copula

Functions	Frank copula
$K_C(t)$	$t - \frac{\ln\left(\frac{g(t)}{g1}\right)}{\theta}\left(e^{\theta t} - 1\right)$

$\varphi^{-1}(s\cdot\varphi(t))$	$-\frac{\ln\left(e^{-\theta - s\varphi(t)} - e^{-s\varphi(t)} + 1\right)}{\theta}$
$\varphi^{-1}((1-s)\cdot\varphi(t))$	$-\frac{\ln\left(e^{-\theta-(1-s)\varphi(t)} - e^{-(1-s)\varphi(t)} + 1\right)}{\theta}$

Tab. 2 Gumbel and Clayton copula

Functions	Gumbel copula	Clayton copula
$K_C(t)$	$t - \frac{t\ln(t)}{\theta}$	$t - \frac{t^{\theta+1} - t}{\theta}$
$\varphi^{-1}(s\cdot\varphi(t))$	$t^{s^{\frac{1}{\theta}}}$	$\left(1 + s\left(t^{-\theta} - 1\right)\right)^{-\frac{1}{\theta}}$
$\varphi^{-1}((1-s)\cdot\varphi(t))$	$t^{(1-s)^{\frac{1}{\theta}}}$	$\left(1 + (1-s)\left(t^{-\theta} - 1\right)\right)^{-\frac{1}{\theta}}$

2 IBNR claim provision calculations

First the marginal univariate distributions are examined to use families of lognormal, Pareto and Wald distributions. Random variable X has Wald distribution with parameters μ and λ, if the density is of the form:

$$f_X(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\cdot\exp\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\},$$

where μ, λ and x >0. The distribution function of the Wald distribution can be presented through the standard normal distribution:

$$F_X(x) = \Phi(\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu} - 1)) + e^{\frac{2\lambda}{\mu}}\Phi(-\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu} + 1))$$

where x > 0 and $\Phi(x)$ is the distribution function of N(0,1). The parameters of the marginal distributions are first estimated, and then each parametric marginal distribution function is plugged into the copula likelihood, and this full likelihood is maximized with respect to the copula parameters. [4].

In the paper claims of a Latvian insurance company from the first quarter of the year 2004 are studied. The data under consideration consists of 1657 claims characterized by development factor, claim size and loss adjusted expenses. Basic characteristics of all three random variables are presented in the following Table 3.

Tab. 3.

Tab. 3 Characteristics of defined random variables

Random variable	Development factor (days)	Claim sizes (LVL)	Loss adjusted expanses (LVL)
Used notation	R1	R2	R3
Mean	36.74	284.36	15.81
Median	3	121.49	19.47
Mode	1	0	19.47
Standard Deviation	94.19	668.41	13.65
Sample Variance	8872.25	446774.40	186.29
Kurtosis	16.93	70.42	6.03
Skewness	3.92	7.17	1.55
Range	706	9000	106.20
Count	1657	1657	1657
Largest (1)	707	9000	106.20
Smallest (1)	1	0	0

As one can see from Table 3, distributions of all three random variables are skewed and kurtosis is far from zero for all random variables. That creates large difficulties to define marginal distributions. We have used Kolmogorov-Smirnov goodness of fit test to find the best approximation for all three random variables and lognormal, Pareto and Wald distributions. The test statistics for comparison with 5% critical value of Kolmogorov-Smirnov test (0.03341) are shown in Table 4.

Tab. 4.

Tab. 4 Kolmogorov-Smirnov test statistics for all three random variables

	Development factor	Claim size	Loss adjusted expenses
Lognormal distribution	0.01995	0.00826	0.04587
Pareto distribution		0.02897	
Wald distribution			0.03616

Graph of both theoretical and sample densities for some of them are shown in Figure 3, 4 and 5. The success of the fully-parametric method obviously depends upon finding appropriate parametric models for the margins, which may not always be so straightforward if they show evidence of heavy tails and/or skewness. Hence, it would be better to have a procedure that avoids marginal risk as much as possible. Several authors, e.g. Jansons and Jurenoks [3], have therefore proposed a semi-parametric approach, for which one do not have any parametric assumptions for the margins. Instead, using semi-parametric method the univariate empirical cumulative distribution functions are plugged.

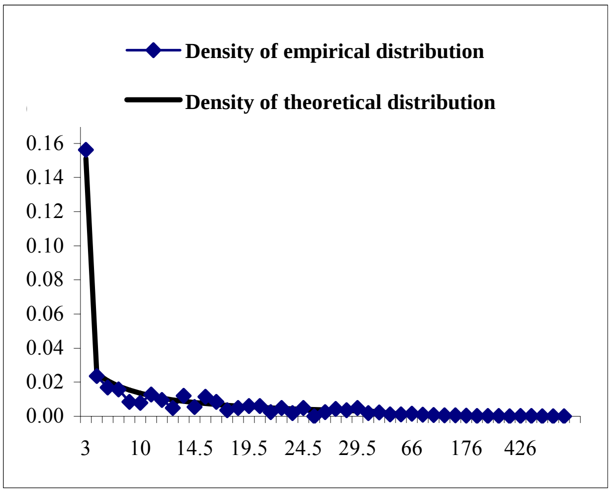

Fig. 3 Approximation of development factor by lognormal distribution

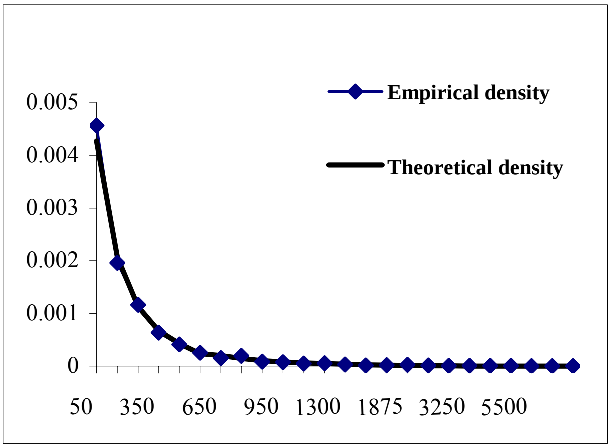

Fig. 4 Approximation of claim sizes by Pareto distribution

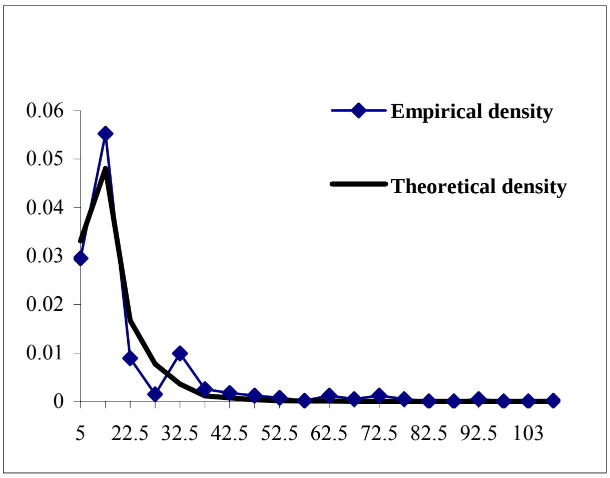

Fig. 5 Approximation of loss adjusted expenses by Wald distribution Figures, tables, and equations

Matrices of linear correlation and Kendall's tau between samples are shown in Table 5, 6.

Tab. 5.

Tab. 5 Relationship between samples – linear correlation coefficient

	Linear correlation coefficient		
	Development factor	Claim size	LAE
Development factor	1	-0.135	0.353
Claim size	-0.135	1	0.200
LAE	-0.353	0.200	1

Tab. 6.

Tab. 6 Relationship between samples – Kendall's tau

	Kendall's tau		
	Development factor	Claim size	LAE
Development factor	1	-0.274	0.350
Claim size	-0.274	1	0.276
LAE	-0.350	0.276	1

We have used multivariate Frank, Clayton and Gambel copula and generators for copulas (see Table 1). For modeling of Archimedean copula we have also used algorithm described in Frees, Valdez [7]:

I. Generate U_1, U_2,…, U_p independent standard uniform random numbers.

II. Set $X_1 = F_1^{-1}(U_1)$ and $c_0 = 0$.

III. For k = 2, …, p recursively calculate X_k as the solution of

$$U_k = \frac{\varphi^{-1(k-1)}\{c_{k-1} + \varphi(F_k(x_k))\}}{\varphi^{-1(k-1)}(c_{k-1})},$$

where ϕ is the generator of Archimedean copula, F - marginal distribution function and $c_k = \varphi[F_1(x_1)] + \varphi[F_2(x_2)] + ... + \varphi[F_k(x_k)]$.

3 Conclusions

Finally to obtain total liabilities we multiply average claim amount plus average loss adjusted expenses with expected number of claims reported in each development day and with number of days. We have calculated liabilities by using the average number of claims in one day plus standard deviation, the average number of claims in one day plus two standard deviations and the average number of claims in one day plus three standard deviations. To have back testing we compared calculated liabilities with real money necessary to pay out and cover loss adjusted expenses. It was 128145.32 LVL. Copula theory makes it possible to approximate joint distribution of the claim size, loss adjusted expenses and the development factor. Copula theory makes it possible to approximate joint distribution of the claim size, loss adjusted expenses and the development factor. The results are shown in Table 7.

Tab. 7.

Tab. 7 Calculated IBNR reserves and related probabilities

Calculated reserve (LVL)	Used number of claims happening in one day	Probability that necessary paid out sum will be larger (%)
113162.2	19	13.57
148897.6	25	2.28
184633	31	0.13

On the basis of the obtained copula model it is possible to estimate liabilities of insurance company in relation with probabilities and therefore its expected values what is not possible to do with classical methods.

4 References

[1] Charpentier, A. Advanced Statistical Methods in Non-Life Insurance, Solvency and simulation based methods. Lecture notes based on joint work with Michel Denuit (UCL), 3rd Conference in Actuarial Science&Finance in Samos, September 2004.

[2] Sklar, A. Fonctions de repartition a n dimensions etleurs marges. Publ. Inst. Statist. Univ. Paris 8, 1959, 229-231.

[3] Jansons, V., Didenko, K., Jurenoks, V., Insurance as a tool for steady development of agriculture. VIII International scientific conference, Management and Sustainable Development Bulgarija, 2006.

[4] Pettere, G., Jansons, V., Stochastic analysis of insurance liabilities. 9-th International Vilnius Conference on Probability Theory and Mathematical Statistics. 2006.

[5] Clemen, R. T., Reilly, T. Correlations and copulas for decision and risk analysis. Management Science, 1999, vol. 45, 208-224.

[6] Embrechts, P., Lindskog, F., McNeil, A. Modelling dependence with copulas and applications to risk management. In: Handbook of Heavy Tailed Distributions in Finance, Ed. E. Rachev, 2003. Elsevier, 329-384.

[7] Frees, E. W., Valdez, E. A. Understanding relationships using copulas. North American Actuarial Journal, 1998, vol. 2, 1-25.

[8] Tang, A., Valdez, E., A., Economic Capital and the aggregation of Risks Using Copulas, http://papers.ica2006.com/1A.html, 2006.

[9] Marshall, A.W., Olkin, I., Families of multivariate distributions. The Journal of the American Statistical Association, 83, 1988.