ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 13, 2009

Stochastic modelling of insurance liabilities

Gaida Pettere

Abstract. Our aim is to present a method for estimating incurred but not reported (IBNR) claim reserves. Each claim is described by three characteristics: the claim size, the allocated loss adjusted expense and the development time. We concentrate on the joint study of all three random variables. First, the marginal univariate distributions are estimated using families of lognormal, Pareto, Wald and Gamma distributions. Next, the matrix of dependence characteristics is found between the three variables and then different multivariate copulas are used to model the joint distribution. The obtained models are fitted to the real data of motor liability insurance of a Latvian insurance company. By simulation the average claim size and allocated loss adjustment expenses in each development day have been estimated. Finally, outstanding claim reserve has been estimated.

1. Introduction

Copulas have been widely used as a tool for modelling different dependence structures in different fields: finance, insurance, risk theory. Copulas were introduced in 1959 by Sklar [22]. Most commonly copulas in non-life insurance are used to model claim sizes and allocated loss adjusted expenses (Frees, Valdez [12]), to evaluate economic capital (Tang, Valdez [23], Jackie [14], Bagarry [1]), for combining different risks (Clemen, Reilly [9], Cherubini et al. [8], Embrechts et al. [11], Klugman, Parsa [16]) or in ruin theory (Bregman, Kluppelberg [5]).

For a long time one of the most important problems for non-life insurance companies has been how to calculate incurred but not reported (IBNR) claim provisions. There are many classical methods how to do that (see, for example, Benjamin [3], Hossack et al. [13], Bornhuetter, Ferguson [4]). All these methods are based on different coefficient calculations and deal with the classical development triangle. A new approach was introduced in

Received October 12, 2009.

2000 *Mathematics Subject Classification.* 47N30.

Key words and phrases. Claim size, dependence characteristics, development time, multivariate copulas, loss adjusted expense, IBNR reserve.

IBNR calculations by using credibility theory (see, for example, De Vylder [10], Mack [19]). These methods are based on Bühlmann-Straub model described in Bühlmann, Gisler [6] and deal with estimation of outstanding claim sizes. Nowadays new stochastic methods have become actual to calculate the mentioned reserves (see Charpentier [7]). Every claim is related with loss adjusted expenses which are needed for an insurance company to be able to pay out claims. Therefore insurers always put provisions not only for not known claims but additionally for expenses which are needed to evaluate claims. Frees and Valdez [12] have studied how to evaluate claim sizes and loss adjusted expenses allocated with each claim. They considered every claim characterized by two random variables: claim amount and loss adjusted expenses. A similar result can be reached by using multidimensional credibility approach (see, for example, Bühlmann, Gisler [6]). In our article we have gone further and considered each claim in non-life insurance as event given by 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 on the joint study of all three random variables what is very important for IBNR claim provision calculations. Two-dimensional case, when every claim is given by two random variables: claim size and development time, is investigated in Kollo, Pettere [17], and detailed IBNR calculation algorithm in that case is given in Pettere [21].

2. Methodology

We consider every claim as a joint realization of three random variables $(X_i, LAE_i, , T_i)$ where X_i is amount of claim incurred at time i, LAE_i is the loss adjusted expenses of corresponding claim and T_i is the time difference between reported and incurred time units of corresponding claim. Our main assumptions are

1) the accident units are independent of each other,
2) $Cov(X_i, LAE_i/i) \neq 0$, $Cov(X_i, T_i/i) \neq 0$ and $Cov(T_i, LAE_i/i) \neq 0$ for every i,
3) distributions of X_i, LAE_i and T_i are independent of i,
4) common distribution function is independent of i and can be expressed through copula

$$H(x_i, l_i, t_i) = C(F_{X_i}(x_i), F_{LAE_i}(l_i), F_{T_i}(t_i))$$

where x_i, l_i and t_i is realizations of corresponding random variables X_i , LAE_i and T_i at accident unit i.

Therefore, if appropriate copula is found, it is possible to estimate by simulation distribution of claim amounts, corresponding to LAE and development time, and calculate average values of claim amounts and LAE in each development unit.

Contrary to the method of generation of multivariate copulas in Tang, Valdez [23], Jackie [14], Bagarry [1] we have used the simulation algorithm developed by Lee [18] and described in Frees, Valdez [12] for simulation of multivariate Archimedean copulas. Copula parameter and copula for further using were found by using the fitting measure

$$SEE = \frac{\sum_{k=1}^{H} \sum_{j=1}^{H} \sum_{i=1}^{H} (O_{i,j,k} - S_{i,j,k})^2}{H^3},$$

where values of each variable are divided into H subsets and so totally H^3 rectangles are created, $O_{i,j,k}$ is the observed frequency in each rectangle and $S_{i,j,k}$ is the simulated frequency in each rectangle under the condition that the size of simulated data coincides with the sample size of the data.

3. Characteristics of marginals

First the marginal univariate distributions are examined. We 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 (Balakrishnan, Nevzorov [2]):

$$f_X(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \cdot exp\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\},\ \mu, \lambda > 0,\ x > 0.$$

The distribution function of the Wald distribution can be presented through the standard normal distribution:

$$F_X(x) = \Phi(\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu} - 1)) + e^{\frac{2\lambda}{\mu}} \cdot \Phi(-\sqrt{\frac{\lambda}{x}}(\frac{x}{\mu} + 1)), x > 0,$$

where $\Phi(x)$ is the distribution function of $N(0,1)$. In the paper property claims of the third party liability insurance of a Latvian insurance company from the first quarter of the year 2004 are studied. Our experience shows that during two years until the end of the first quarter 2006 almost all property claims have been reported. The data under consideration consist 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 1.

Table 1. Characteristics of defined random variables.

Random variable	Development factor (days)	Claim size (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
Sample size	1657	1657	1657
Largest(1)	707	9000	106.20
Smallest(1)	1	0	0

Table 2. Kolmogorov test statistics for all three random variables.

	Development factor	Claim size	Loss adjusted expenses
Lognormal distribution	0.019955	0.008259	0.045872
Pareto distribution		0.028969	
Wald distribution			0.036157

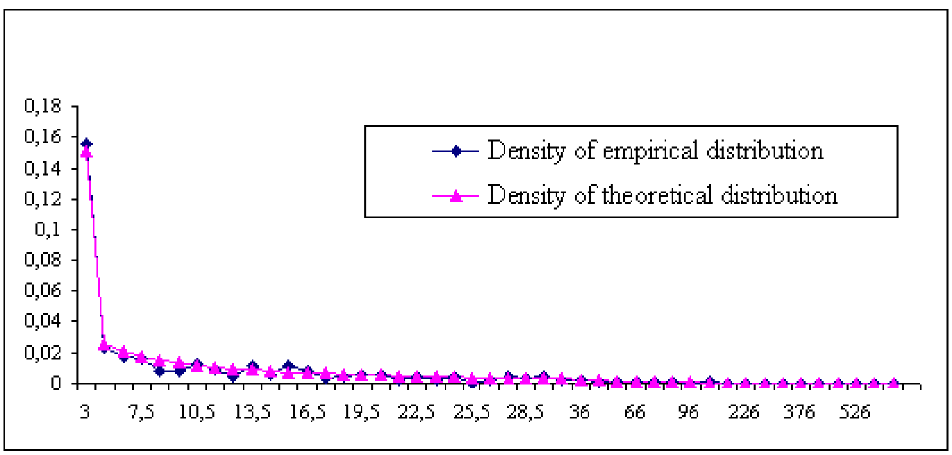

Figure 1. Approximation of development factor by lognormal distribution.

As one can see from Table 1, distributions of all three random variables are skewed and kurtosis is far from zero for all the random variables. This creates large difficulties to find marginal distributions. We have used Kolmogorov goodness-of-fit test to find the best approximation for all three random variables from lognormal, Pareto and Wald distributions. The test statistics for comparison have been calculated at 5% confidence level. Values of Kolmogorov test (0.03341) are shown in Table 2. Graphs of both, theoretical and sample densities, are shown in Figures 1, 2 and 3. Sample estimates of matrices of linear correlation and Kendall's tau are shown in Table 3.

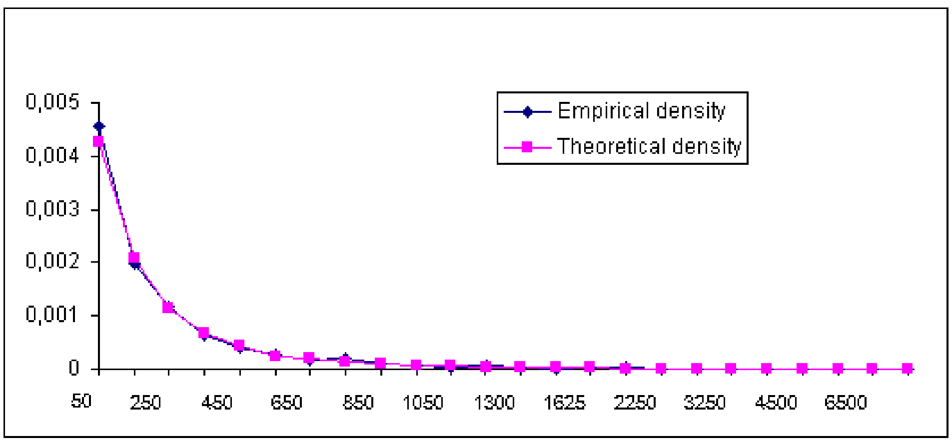

Figure 2. Approximation of claim sizes by Pareto distribution.

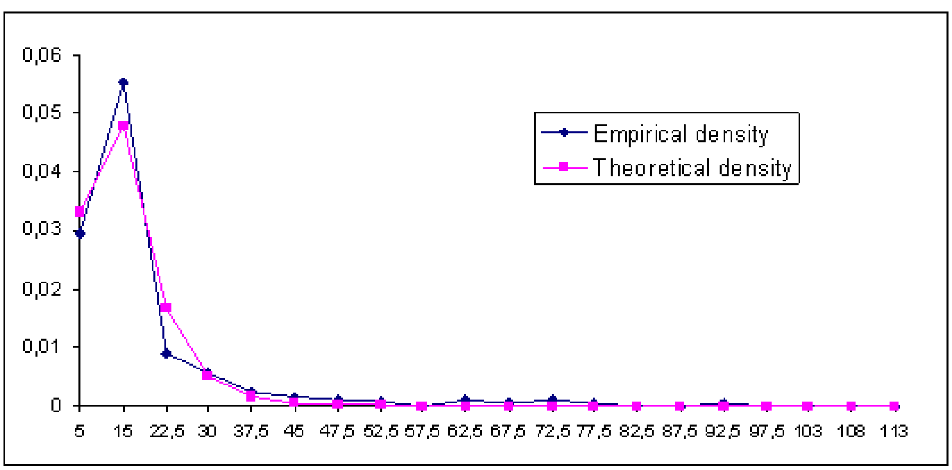

Figure 3. Approximation of loss adjusted expenses by Wald distribution.

Table 3. Relationship between variables.

	Linear correlation coefficient			Kendall's tau		
	Development factor	Claim size	LAE	Development factor	Claim size	LAE
Development factor	1	-0.1352	-0.3532	1	-0.274	-0.350
Claim size	-0.1352	1	0.200	-0.274	1	0.276
LAE	-0.3532	0.200	1	-0.350	0.276	1

4. Finding appropriate copula model

Appropriate copula model is possible to find using classical n-dimensional copulas like Clayton n-copula, Gumbel n-copula and Frank n-copula for which simulation algorithms are given in Cherubini et al. [8]. We have tried to concentrate not only on these widely known copulas but have used several other copulas mentioned in Nelsen [20] as examples and derived several new multidimensional copulas from two dimensional copulas. Due to the arguments in Kimberling [15], in the case when generator φ of an Archimedean copula is non-strictly decreasing and its pseudo-inverse $\varphi^{[-1]}$ is *m-monotonic* on $[0,\infty)$, the function given by

$$C^n(\bar{u}) = \varphi^{[-1]}[\varphi(u_1) + \varphi(u_2) + \ldots + \varphi(u_n)],$$

where $\bar{u} = (u_1, u_2, \ldots, u_n)'$, is an n-dimensional copula. Nelsen has shown in Example 4.25 (see Nelsen [20], p. 124) that generator

$$\varphi(t) = 1 - t^{\frac{1}{\delta}} \quad \text{for} \quad \delta > 1,$$

despite that it is non-strictly decreasing and its pseudo-inverse is *m-monotonic*, where $m \leq \delta < m+1$, generates an n-dimensional copula

$$C^n(\bar{u}) = max([u_1^{\frac{1}{\delta}} + u_2^{\frac{1}{\delta}} + \ldots + u_n^{\frac{1}{\delta}} - n + 1]^{\delta}, 0)$$

for $n \leq m$. Below it will be called Copula 4.25.

Additionally we derived some new multidimensional copulas from two-dimensional copulas with non-strictly decreasing generators but the best result was obtained with the generator used for two-dimensional Copula 4.2.7 (see Nelsen [20], p. 94):

$$\varphi(t) = -ln[\delta t + (1-\delta)] \quad \text{for} \quad \delta \in (0,1].$$

It is non-strictly decreasing because of $\varphi(0) = -ln(1-\delta)$. But its pseudo-inverse

$$\varphi^{[-1]}(t) = \begin{cases} \frac{1}{\delta}[e^{-t} - (1-\delta)], & t \in [0, -ln(1-\delta)], \\ 0, & t \geq -ln(1-\delta), \end{cases}$$

is completely monotonic and therefore it generates the n-dimensional copula

$$C^n(\bar{u}) = max([\delta u_1 u_2 \cdot \ldots \cdot u_n + (1-\delta)(u_1 + u_2 + \ldots + u_n - n + 1)], 0).$$

Exception is the case $\delta = 1$. Then the generator is a strictly decreasing function and the copula is the product copula.

We have used for simulation an algorithm described in Frees, Valdez [12]:

1. Generate $U_1, U_2, \ldots, U_p$ realizations from a uniform random variable on $(0, 1)$.
2. Set $X_1 = F_1^{-1}(U_1)$.
3. For $k = 2, \ldots, p$, 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 an Archimedean copula, $(\varphi^{-1})^{(k-1)}$ is the $k-1$ order derivative of the inverse of generator, F is the marginal distribution function and

$$c_k = \varphi[F_1(x_1)] + \varphi[F_2(x_2)] + \ldots + \varphi[F_k(x_k)].$$

The explicit expressions for simulation using mentioned algorithm for Copula 4.2.7 and Copula 4.25 (see Nelsen [20], p. 94 and p. 124) are shown in Table 4. Explicit expressions for Clayton n-copula, Gumbel n-copula and Frank n-copula are given in Cherubini et al. [8].

We have checked the goodness-of-fit by using our fitting measure for each of the described copulas and different sets of marginals: a) lognormal, lognormal, lognormal, b) lognormal, Pareto, lognormal, c) lognormal, lognormal, Wald, and d) lognormal, Pareto, Wald, to find the best copula parameter δ and finally to find the best copula with appropriate marginals. The average values of the fitting measure and its standard deviation for the best copulas are shown in Table 5.

Finally, we have chosen Clayton copula with lognormal marginals to describe the average claim amount and loss adjusted expense in each development day. Slightly worse results were obtained with Clayton copula with lognormal-Pareto-lognormal marginals, with Copula 4.2.7 with lognormal-Pareto-lognormal marginals and with Copula 4.25 with lognormal-Pareto-Wald marginals. More detailed analysis is needed there and we have left this for further research.

Table 4. Generators of Copula 4.2.7 and Copula 4.25 and functions needed for simulation algorithm.

	Copula 4.2.7 $\delta \in (0,1]$	Copula 4.25 $\delta > 1$
Generator φ	$-\ln(\delta \cdot t + (1-\delta))$	$1 - t^{\frac{1}{\delta}}$
Inverse generator φ^{-1}	$\frac{1}{\delta}(e^{-t} - (1-\delta))$	$(1-t)^{\delta}$
First order derivative of inverse generator	$-\frac{1}{\delta} \cdot e^{-t}$	$-\delta \cdot (1-t)^{\delta-1}$
Second order derivative of inverse generator	$\frac{1}{\delta} \cdot e^{-t}$	$\delta \cdot (\delta-1) \cdot (1-t)^{\delta-2}$
c_1	$-\ln(\delta \cdot F_1(x_1) + (1-\delta))$	$1-(F_1(x_1))^{\frac{1}{\delta}}$
x_2	$F_2^{-1}(\frac{U_2-(1-\delta)}{\delta})$	$F_2^{-1}(U_2 \cdot (1-c_1)^{\delta-1} + c_1)^{\delta}$
c_2	$c_1 - \ln(\delta \cdot F_2(x_2) + (1-\delta))$	$c_1 + 1 - (F_2(x_2))^{\frac{1}{\delta}}$
x_3	$F_3^{-1}(\frac{U_3-(1-\delta)}{\delta})$	$F_3^{-1}(\left(U_3^{\frac{1}{\delta-2}} \cdot (1-c_2) + c_2\right)^{\delta})$

Table 5. The average values of fitting measure and its standard deviation with the best copula parameter.

Marginals	Clayton copula $\delta = 0.1$	Copula 4.2.7 $\delta = 0.96$	Copula 4.25 $\delta = 5$
Lognormal, lognormal, lognormal	$\mu = \mathbf{216.2}$ $\sigma = \mathbf{8.8}$	$\mu = 218.3$ $\sigma = 9.1$	$\mu = 214.2$ $\sigma = 10.3$
Lognormal, Pareto, lognormal	$\mu = 212.4$ $\sigma = 10.4$	$\mu = 216.4$ $\sigma = 9.1$	$\mu = 210.4$ $\sigma = 9.8$
Lognormal, lognormal, Wald	$\mu = 216.8$ $\sigma = 8.9$	$\mu = 219.1$ $\sigma = 10.1$	$\mu = 215.8$ $\sigma = 9.9$
Lognormal, Pareto, Wald	$\mu = 218.4$ $\sigma = 8.9$	$\mu = 220.3$ $\sigma = 10.2$	$\mu = 215.1$ $\sigma = 8.9$

5. IBNR reserve calculation

We did 100 simulations of 10000 data from the Clayton copula with log-normal marginals and calculated the average claim amount and loss adjusted expense in each development day. Results are shown in Table 6.

Table 6. The average value of claim amount and loss adjusted expenses in each development day.

Development days	0	1	2	...	697	698	699	700
Average claim amount (LVL)	255,69	278,86	275,08	...	168,46	149,27	24,98	88,13
Average LAE (LVL)	14,08	14,69	14,67	...	6,72	6,85	2,00	3,76

At the next step the number of claims happening in one day had to be estimated. The sample characteristics of that variable are shown in Table 7.

Table 7. Characteristics of number of claims per development day.

Random variable	Number of claims
Mean	13.82
Median	13
Mode	9
Standard Deviation	5.93
Sample Variance	35.11
Kurtosis	-0.12
Skewness	0.49
Range	29
Sample size	90
Largest	32
Smallest	3

For simplicity we have assumed that this variable follows a normal distribution because of small values of kurtosis and skewness. The hypothesis about normality of number of claims per day was tested with Kolmogorov–Smirnov test with mean 13 and standard deviation 6. The test statistic equals 0.0808, while the 5% critical value is 0.1436.

To obtain outstanding liability of one day (one row of the development triangle) we multiply the average claim amount plus the average loss adjusted expenses by the expected number of claims reported in each development day. It is necessary for further calculations to get not known part of liability triangle to multiply obtained results in each day by the number of developed days (day one repeats once, day two – twice, day three – three times, ...). We have calculated liabilities by using the average number of claims in one day plus one 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. The results are shown in Table 8.

Table 8. 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 (%)
113162.20	19	13.57
148897.60	25	2.28
184633.00	31	0.13

For testing we compared the calculated liabilities with the real amount of money, what was necessary to 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. This is very important because there is a significant correlation between the mentioned variables. For further investigation we have envisaged the problem how copulas could be used in credibility theory.

Acknowledgement. The author thanks participants of the 9th International Vilnius Conference on Probability Theory and Mathematical Statistics, whose suggestions and comments have improved the presentation of the paper.

References

[1] Bagarry, M. (2006), *Economic Capital: a Plea for the Student Copula.* http://papers.ica2006.com/1A.html

[2] Balakrishnan, N., Nevzorov, V. B. (2003), *A Primer on Statistical Distributions*, Wiley, New York.

[3] Benjamin, B. (1977), *General Insurance*, Heinemann, London.

[4] Bornhuetter, R. L., Ferguson, R. E. (1972), *The actuary and IBNR.* Proc. CAS **59**, 181–195.

[5] Bregman, Y., Kluppelberg, C. (2005), *Ruin estimates in multivariate models with Clayton dependence structure*, Munich University of Technology, unpublished manuscript.

[6] Bühlmann, H., Gisler, A. (2005), *A Course in Credibility Theory and its Applications*, Springer-Verlag, Berlin.

[7] Charpentier A. (2004), *Advanced Statistical Methods in Non-Life Insurance, Solvency and Simulation Based Methods*, Lecture notes based on joint work with Michel Denuit (UCL), 3^{rd} Conference in Actuarial Science & Finance in Samos, September 2004.

[8] Cherubini, U., Luciano, E., Vecchiato, W. (2004), *Copula Methods in Finance*, Wiley, Chichester.

[9] Clemen, R. T., Reilly, T. (1999), *Correlations and copulas for decision and risk analysis*, Manag. Sci. **45**, 208–224.

[10] De Vylder, F. (1982), *Estimation of IBNR claims by credibility theory*, Insurance Math. Econom. **1**, 35–40.

[11] 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. E. Rachev), Elsevier, Amsterdam, 329–384.

[12] Frees, E. W., Valdez, E. A. (1998), *Understanding relationships using copulas*, North Amer. Actuarial J. **2**, 1–25.

[13] Hossack, I. B., Pollard, J. H., Zehnwirth B. (1983), *Introductory Statistics with Applications in General Insurance*, Cambridge University Press, Cambridge.

[14] Jackie, Li (2006), *Modelling Dependency between Different Lines of Business with Copulas*. http://papers.ica2006.com/1A.html

[15] Kimberling, C. H. (1974), *A probabilistic interpretation of complete monotonicity*, Aequationes Math. **10**, 152–164.

[16] Klugman, S. A., Parsa, R. (1999), *Fitting bivariate loss distributions with copulas*, Insurance Math. Econom. **24**, 139–148.

[17] Kollo, T., Pettere, G. (2006), *Copula models for estimating outstanding claim provisions*; In: Festschrift for Tarmo Pukkila on his 60^{th} Birthday (Eds. E.P. Liski, J. Isotalo, J. Niemel, S. Puntanen, G.P.H. Styan), Department of Mathematics, Statistics and Philosophy, University of Tampere, Vammala, 115–125.

[18] Lee, A. J. (1993), *Generating random binary deviates having fixed marginal distributions and specified degrees of association*, Amer. Statist. **47**, 209–215.

[19] Mack, T. (1990), *Improved estimation of IBNR claims by credibility theory*, Insurance Math. Econom. **9**, 51–57.

[20] Nelsen, R. B. (1999), *An Introduction to Copulas*, Springer-Verlag, New York.

[21] Pettere G. (2006), *Modelling incurred but not reported claim reserve using copulas*; In: International Conference on Mathematical and Statistical Modeling in Honor of Enrique Castillo (ICMSM 2006), Program and Abstracts (Eds. R. Minguez, C. Castillo, C. Solares, R. E. Pruneda), June 28 - 30, 2006, University of Castilla - La Mancha, Ciudad Real, Spain, R. No. 06/33297, ISBN 84-689-8577-5.

[22] Sklar, A. (1959), *Fonctions de repartition à n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris **8**, 229–231.

[23] Tang, A., Valdez, E. A. (2006), *Economic Capital and the Aggregation of Risks Using Copulas*. http://papers.ica2006.com/1A.html

Department of Engineering Mathematics, Riga Technical University, Kalku street 1, Riga, Latvia

E-mail address: `gaida@latnet.lv`