MODELLING CLAIM SIZE IN TIME VIA COPULAS

Pettere, Gaida
Riga Technical University,
Kalku Street 1, Riga, LV 1658, Latvia,
Phone: 371 9463236, fax: 371 7089490
gaida@latnet.lv

Kollo, Tonu
University of Tartu
J. Liivi Street 2, Tartu, 50409, Estonia,
Phone: 372 7375486, fax: 372 7375898
tonu.kollo@ut.ee

Abstract

In the paper claims of a Latvian Insurance company are studied. The two investigated variables are claim size and time from the moment when claim occurred to the moment when the payment has been reported. Approximations to the distributions of both variables have been found as well as the bivariate distribution modelled using Archimedean copulas. IBNR claim reserves have been calculated using the bivariate Clayton copula model. The results are illustrated by graphs and tables.

Keywords: Archimedean copula, claim distribution, Clayton copula, Frank copula, Gumbel copula, IBNR reserves

1. INTRODUCTION

In the paper claims of a Latvian insurance company from the year 2003 are studied. There has been growing interest to the modelling of financial data by copulas (see Cherubini et al (2004), for example). In insurance mathematics Archimedean copulas have become the most common tool since Frees, Valdez (1998). Usually claim size and expenses allocated to each claim have been the two variables of interest in copula models. We accentuate to the claim size and its development factor. The second variable "development factor" measures time from the moment when claim occurred to the moment when claim has reported. These two variables form the basis for calculation IBNR claim reserves. As they are correlated we need to take into account both variables in the modelling process. In part 3.1 we describe univariate marginal distributions of the claim size and the development factor. Also correlation coefficients between the variables are found and basic characteristics of the marginal distributions given. In part 3.2 we give a short summary of necessary notions and notation from the copula theory. Archimedean copulas are of the main interest. In part 3.3 several models via copulas are examined and goodness-of-fit of the models tested. Part 3.4 deals with estimating IBNR reserves on the basis of the found models. Finally the results are summarized and conclusions made.

2. METHODOLOGY

When finding approximations to the distributions of the random variables of interest: claim size and development factor, the least squares method was used to find out the best models. Copula methodology was used to model joint bivariate distribution of the variables. Several classes of Archimedean copulas: Clayton, Frank, Gumbel were used and once again the best model was found by the least squares method. Estimation of IBNR reserves was based on simulation from the obtained bivariate Clayton copula model. All the calculations have been made using Mathcad and Microsoft Excel software.

3. DISCUSSION

3.1. Modelling claim size and development factor.

The data under consideration consists of 1237 claims. Basic characteristics of the claim size are presented in the following Table 1.

Table 1. Characteristics of claim size

Mean	298.57
Standard Error	17.85
Median	160.25
Mode	0
Standard Deviation	627.94
Sample Variance	394312.19
Kurtosis	90.06
Skewness	8.31
Range	9000
Minimum	0
Maximum	9000
Sum	369332.52
Count	1237
Largest (1)	9000
Smallest (1)	0
Confidence Level (95,0%)	35.03

As one can see from Table 1, the distribution is heavily skewed. Skewness measure β_1 is approximately 8.3 and the sample mean is almost twice bigger than sample median: while the sample mean equals 298.57 LVL, the sample median is only 160,25 LVL. The distribution was modelled by four different distributions: Pareto, *t*-distribution, lognormal, and Wald (or inverse Gaussian) distribution. From these four classes of distributions only lognormal and Wald distributions gave satisfactory models. The lognormal distribution was used in the following parameterisation: random variable X is lognormally distributed with parameters μ and σ, if $\ln X = Y \sim N(\mu, \sigma)$, where $EY = \mu$ and $DY = \sigma^2$. The best fitting lognormal density is sketched in Figure 1.

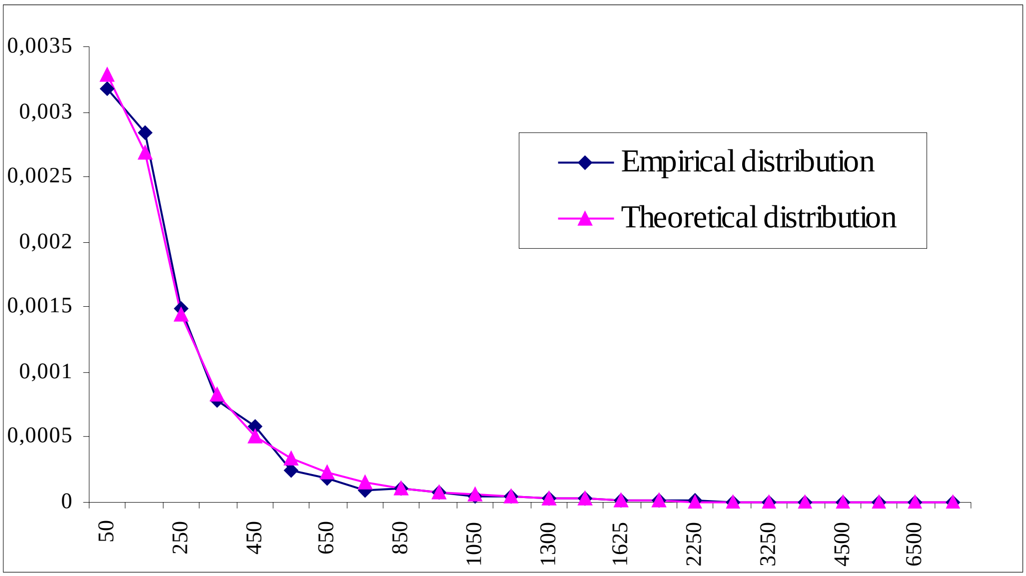

Figure 1. Lognormal approximation of claim size

The best fit was obtained with the lognormal distribution with parameters $\mu = 5.053$, $\sigma = 1.007$. The Kolmogorov-Smirnov two sided test-statistic $K_n = 0.012$ which allows us to accept the model at the significance level $\alpha = 0.05$ (critical value equals 0.038668).
Wald distribution gave slightly worse fit with the data (the value of the Kolmogorov-Smirnov test-statistic $K_n = 0.0299$ for the distribution with parameters $\mu = 282$ and $\lambda = 146$). Random variable X has Wald distribution with parameters μ and λ, if the density is of the form (Balakrishnan, Nevzorov, 2003):

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

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

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

where $\Phi(x)$ is the distribution function of $N(0,1)$.

Consider now modelling of the development factor. Range on values of this variable runs from 0 to 1100 days (about three years). Basic characteristics of the variable are presented in Table 2.

Table 2. Characteristics of the development factor

Mean	25.21
Standard Error	2.27
Median	2
Mode	1
Standard Deviation	79.86
Sample Variance	6377.25
Kurtosis	38.81
Skewness	5.79
Range	740.41
Minimum	0
Maximum	740.41
Sum	31189.51
Count	1237
Largest (1)	740.41
Smallest (1)	0
Confidence Level (95,0%)	4.45

Again we have extremely skewed distribution with the sample mean being much bigger than median. The same time skewness measure β_1 equals 5.79 approximately.

Lognormal distribution gives us again quite a good model. The two-sided Kolmogorov-Smirnov statistic is $K_n = 0.0306$ for the distribution with parameters $\mu = 0.487$ and $\sigma = 2.487$ (critical value equals 0.038668 at the 5% significance level).

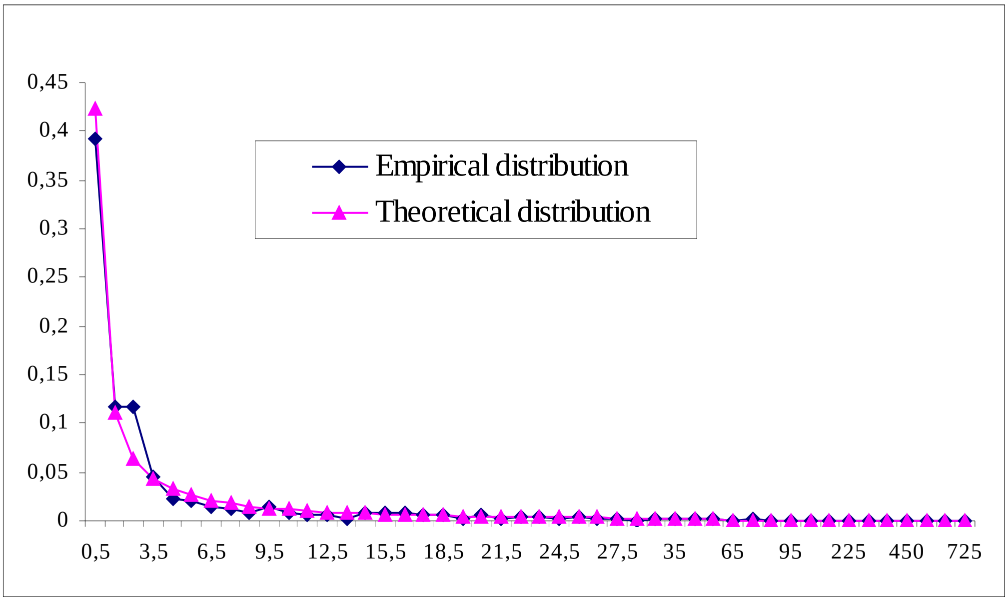

Figure 2. Approximation of development factor with lognormal distribution

Correlation between these two variables is not high and varies for different samples up to 0.3. In our case the sample linear correlation is 0.10. For the sample size 1237 this correlation appears to be significant.

3.2. Notions and results from the copula theory.

Copulas have become an intensively use method in modelling of financial data in recent years. Frees, Valdez (1998), Durrleman et al (2000), Lindskog (2000), Rank (2000), Cherubini, Luciano (2001), Jouanin et al (2001), Roncalli (2001), de Matteis (2001), Embrechts et al (2003), Marena (2003) are referred to here from the long list of existing studies. For additional references Cherubini et al (2004) can be recommended. Our intension is to use bivariate copulas to model the bivariate random vector under consideration.

First we try to find the best Archimedean copula. Classes of bivariate Archimedean copulas are attractive in two reasons. Firstly, Archimedean copulas are characterized by a scalar parameter, which can be related to a rank correlation coefficient (Spearman or Kendall) between the two variables. Secondly, due to Genest, Rivest (1993) an algorithm is available to check the fit of the obtained model with the data via univariate Kolmogorov-Smirnov test-statistic (see also Genest et al, 1995). Necessary notions and notation are given below. Bivariate copula is the joint distribution function of two uniformly distributed random variables:

$$C(u,v) = P(U \le u, V \le v),$$

where $U, V \sim U(0,1)$. Let X and Y be two random variables with distribution functions F and G respectively. Because $F(X)$ and $G(X)$ are uniformly distributed, we may take $F(X) = U$, $G(Y) = V$, and get the copula in the following form

$$C(F(x), G(y)) = P(F(X) \le F(x), G(Y) \le G(y)) = P(X \le x, Y \le y) = F_{X,Y}(x, y).$$

DEFINITION 1. If the copula $C(u,v)$ can be presented of the form

$$C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \tag{1}$$

for a continuous strictly decreasing function φ from $[0,1]$ to $[0,\infty)$ with $\varphi(1) = 0$, we call the copula Archimedean. The function φ in (1) is called the generator of the copula.

Detailed list of different classes of Archimedean copulas can be found in Nelsen (1999), pp. 94-97, their properties and fitting problems are carefully studied in the thesis de Matteis (2001).

There exist several classes of Archimedean copulas, which are of particular interest in financial applications. Some important classes are presented in the following Table 3.

In Table 3 D_1 and D_2 denote the Debye functions defined as

$$D_k(x) = \frac{k}{x^k} \cdot \int_0^x \frac{t^k}{e^t - 1} dt.$$

For negative arguments we have (Frees, Valdez, 1998):

$$D_k(-x) = D_k(x) + \frac{k \cdot x}{k+1}.$$

The procedure suggested by Genest, Rivest (1993) to identify an Archimedean copula defined by a generator φ, is based on the following argument. From bivariate data (x_i, y_i), $i = 1,\dots,n$ we construct values of the univariate pseudovariable Z :

$$z_i = \frac{No\{x_j, y_j : x_j < x_i, y_j < y_i\}}{n-1}, \; i = 1,..,n. \quad (2)$$

Then $F_Z(x) = K_\varphi(x)$,

where

$$K_\varphi(x) = x - \frac{\varphi(x)}{\varphi'(x)} \quad (3)$$

with $\varphi'(x)$ being the derivative of $\varphi(x)$.

Now the standard procedure of Kolmogorov-Smirnov criteria can be used by comparing empirical distribution function $F_n(x)$ of z_i and $K_\varphi(x)$.

Table 3. Archimedean copulas

Family	Generator φ	Parameter area	Copula $C(u,v)$	Kendall's τ	Spearman's ρ
Independence	$-\ln t$	-	$u \cdot v$	0	0
Clayton	$t^{-\theta}$	$\theta > 1$	$(u^{-\theta} + v^{-\theta} - 1)^{\frac{1}{\theta}}$	$\frac{\theta}{\theta+2}$	complicated
Gumbel	$(-\ln t)^\theta$	$\theta \ge 1$	$\exp\{-[(-\ln u)^\theta +$ $+(-\ln v)\theta]^{\frac{1}{\theta}}\}$	$1-\theta^{-1}$	no explicit form
Frank	$\ln\frac{e^{\theta t}-1}{e^\theta - 1}$	$-\infty < \theta < \infty$	$\frac{1}{\theta}\ln(1+$ $+\frac{(e^{\theta u}-1)(e^{\theta v}-1)}{e^\theta - 1}$	$1-\frac{4}{\theta}\cdot[D_1(-\theta)-$ $-1]$	$1-$ $-\frac{12}{\theta}[D_2(-\theta)-$ $-D_1(-\theta)]$

3.3. Modelling claims in time.

In this section we are going to present results of modelling of the bivariate distribution of the two variables of interest: the claim size and the development factor. The variables are both considered lognormally distributed, with $\mu = 5.053$ and $\sigma = 1.007$ for claim size and with $\mu = 0.487$ and $\sigma = 2.487$ for development factor.

After time-consuming studies of many possibilities (independence, Gaussian copula, HRT copula and many others) we examined more carefully three classes of Archimedean copulas with function $K_\varphi(x,\theta)$ presented in Table 4 and found parameters of them by the least squares method by minimizing the function

$$\Psi(\theta) = \sum_{i=1}^{n} (K_\varphi(x,\theta) - F_Z(x))^2,$$

where $n = 100$.

Table 4. Distribution function $K_\varphi(x,\theta)$ for used copulas

Copula	Clayton	Gumbel	Frank
$K_\varphi(x,\theta)$	$x - \frac{x^{\theta+1} - x}{\theta}$	$x - x\cdot\frac{\ln x}{\theta}$	$x - \frac{\ln(\frac{e^{-\theta x}-1}{e^{-\theta}-1})}{\theta}\cdot(e^{\theta x}-1)$

At the next step we compared the values of the function $\Psi(\theta)$ and values of the function

$$LM(\theta) = \left|K_{\varphi}(x,\theta) - F_Z(x)\right|$$

for chosen copulas with estimated value of parameter θ. The results are shown in Table 5.

Table 5. Values of parameters and functions used for fitting Archimedean copulas

Copula	Clayton	Gumbel	Frank
Values of θ	-0.3	0.85	–1.3
$\Psi(\theta)$	0.0522	0.213	0.13973
$LM(\theta)$	0.134	0.179	0.18

Comparison of theoretical functions $K_{\varphi}(x,\theta)$ and empirical distribution function $F_Z(x)$ are shown in Figures 3, 4 and 5.

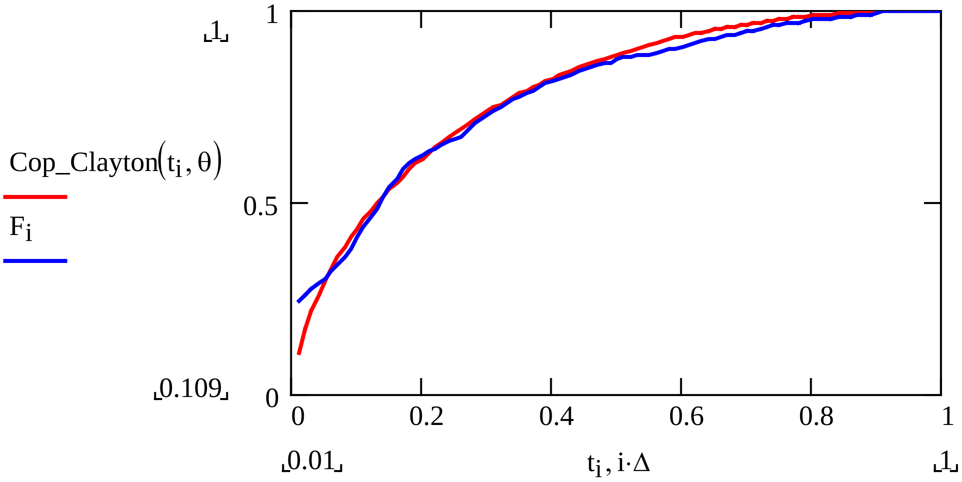

Figure 3. Genest-Rivest approximation by Clayton copula

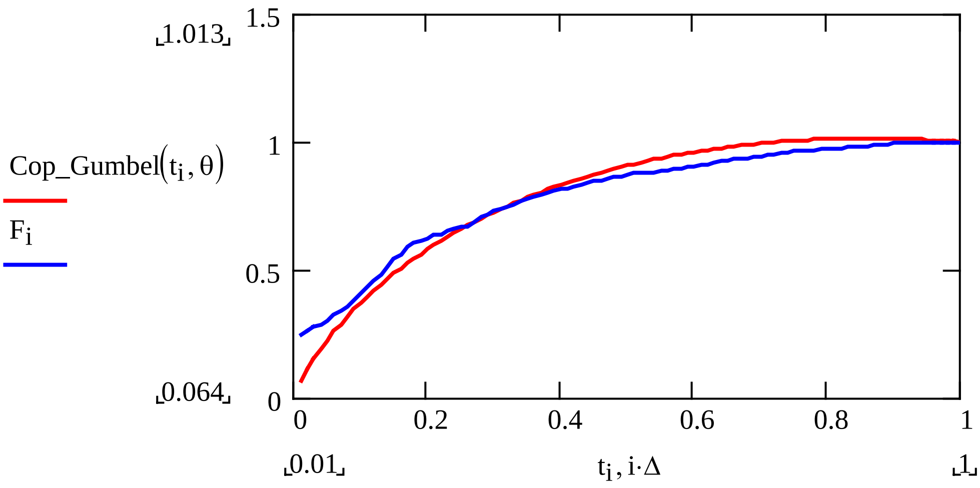

Figure 4. Genest-Rivest approximation by Gumbel copula

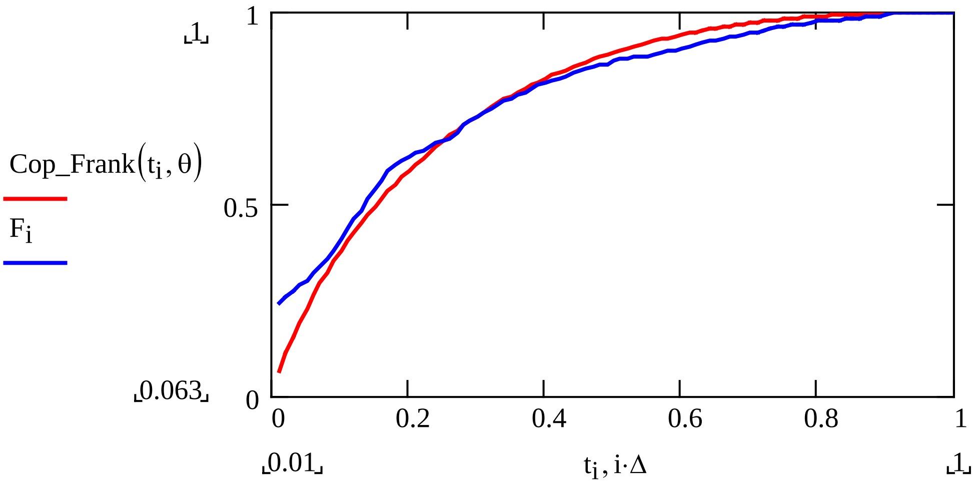

Figure 5. Genest-Rivest approximation by Frank copula

Fitting of the obtained Archimedean copulas was tested by the Kolmogorov-Smirnov test again which was applied to pseudovalues z_i, defined by (2) in part 3.2. From Table 5 and Figures 3, 4 and 5 one can see that the best fit is obtained by Clayton copula. Closer examination of the difference between empirical and theoretical distribution functions of pseudovalues z_i shows that the only value above the 5% critical one occurs in the case of zero development values, which does not influence reserve calculation because claims with zero development are known.

3.4. Estimation IBNR reserves

We are going to estimate necessary IBNR reserves using the bivariate Clayton copula model obtained above. But firstly we estimate distribution of number of claims reported in k days ($k = 0,1,2,...,1100$). Maximum period for reporting was taken 1100 days, which is fixed by Latvian legislation rules. Once again lognormal distribution gave a good fit.

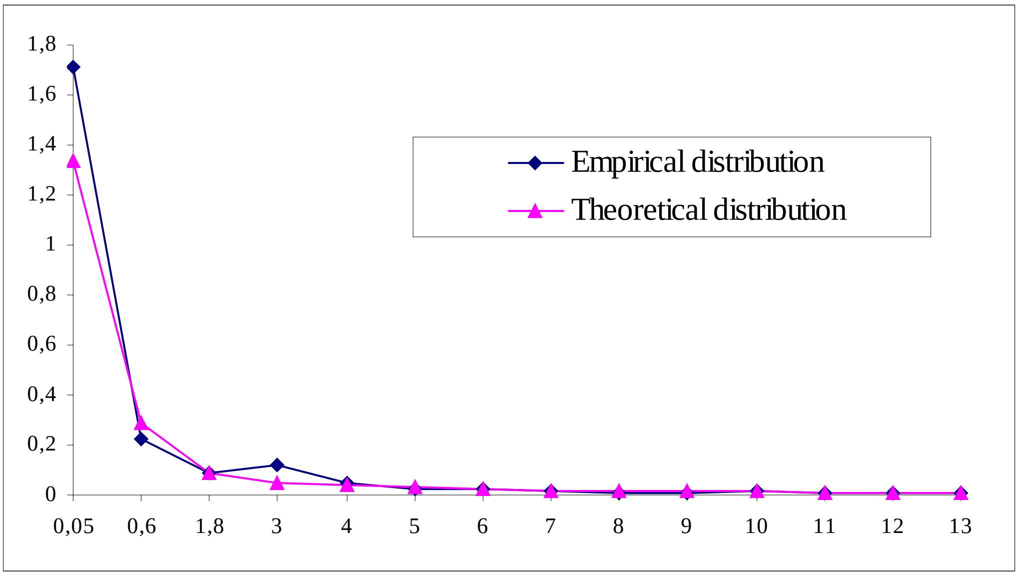

Figure 6. Lognormal approximation to the distribution of number of claims

The best model found by least squares method was with parameters $\mu = 0.6$ and $\sigma = 2.62$. Value of Kolmogorov-Smirnov test statistic equals 0.0379, which is less than the 5% critical value. The obtained lognormal approximation of number of claims is presented in Figure 6.

Secondly we estimate number of claims occurred in one day using data from 90 days in 2003. Descriptive statistics are shown in Table 6.

Table 6. Characteristics of the number of claims per day

Mean	13.82
Standard Error	0.62
Median	13
Mode	9
Standard Deviation	5.93
Sample Variance	35.11
Kurtosis	-0.12
Skewness	0.49
Range	29
Minimum	3
Maximum	32
Sum	1244
Count	90
Largest (1)	32
Smallest (1)	3
Confidence Level (95,0%)	1.24

This variable is modelled by the normal law with mean $\mu = 13$ and standard deviation $\sigma = 6$. The maximum difference between empirical and theoretical distribution function is 0.06973 (critical value of the Kolmogorov-Smirnov test statistic equals 0.143357 at the 5% significance level).

By using both random variables we can find number of claims reported in the same day, one day later, two days later, and so on. Then we multiply probability of occurring claims from our lognormal model to the number of expected total claims 35.8 ($\mu + 3.8 \cdot \sigma$). After that 50 times 50000 pairs of claim sizes and development factors were simulated from our Clayton copula model and average amount of claim size corresponding to each development factor from 0 to 1100 was found. Finally IBNR reserve is calculated like there would be present the second part of the development triangle: average claim size reported in the first day was multiplied by number of claims in the first day, average claim size reported in the second day was multiplied by number of claims in the second day and with two, average claim size reported in the third day was multiplied by number of claims in the third day and with three and so on until 1100. Sum of all these products represents IBNR reserve which was equal to 145 809 LVL what is in agreement with real paid out sum in the company. The obtained result shows applicability of the procedure.

4. RESULTS

For the claim size lognormal distribution is the best model but Wald distribution can be used too. Development factor can also be described by the lognormal distribution. When approximating the bivariate distribution of the claim size and the development factor with Archimedean copulas, the Clayton copula gave the best model. Using the bivariate Clayton copula in simulations IBNR reserves were estimated. Applicability of the method of estimation IBNR reserves was checked on data from past of the company. The real empirical values for necessary reserves were in good accordance with the predicted estimate using Clayton copula.

5. CONCLUSIONS AND SUGGESTIONS

Copula theory makes it possible to approximate joint distribution of the claim size and the development factor. On the basis of the obtained copula model it is possible to estimate IBNR reserves. The obtained prognosis is in good agreement with real data.

REFERENCES

1. Balakrishnan, N., Nevzorov, V. B. *A Primer on Statistical Distributions*. Wiley, New York, 2003.
2. Cherubini, U., Luciano, E. Value at risk trade-off and capital allocation with copulas. *Econ. Notes*, 2001, vol.30, 235-256.
3. Cherubini, U., Luciano, E., Vecchiato, W. *Copula Methods in Finance*. Wiley, Chichester, 2004.
4. Clemen, R. T., Reilly, T. Correlations and copulas for decision and risk analysis. *Management Science*, 1999, vol. 45, 208-224.
5. Durrleman, V., Nikeghbali, A., Roncalli, T. *Which Copula is the Right One?* Groupe de Recherche Operationelle, Credit Lyonnais, Working paper, 2000.
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. Genest, C., Rivest, L. P. Statistical inference procedures for bivariate Archimedean copulas. *JASA*, 1993, vol. 88, 1034-1043.
9. Genest, C., Ghoudi, K., Rivest, L. P. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, 1995, vol. 82, 543-552.
10. Jouanin, J. F., Riboulet, G., Roncalli, T. *Modelling Dependence for Credit Derivatives with Copulas*. Groupe de Recherche Operationelle. Credit Lyonnais, Working Paper, 2001.
11. Klugman, S. A., Parsa, R. Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics*, 1999, vol. 24, 139-148.
12. Luciano, E., Marena, M. Copulae as a new tool in financial modelling. *Operational Research*: *An International Journal*, 2003, vol. 2, 139-155.
13. de Matteis R. *Fitting Copulas to Data*. Diploma Thesis, Institute of Mathematics of the University of Zurich, 2001.
14. Nelsen, R. B. *An Introduction to Copulas*. Springer-Verlag, New York, 1999.
15. Rank, J. *Copulas in Financial Risk Management*. Diploma Thesis in Mathematical Finance, University of Oxford, 2000.
16. Roncalli, T. *Copulas: A Tool for Dependence in Finance*. Groupe de Recherche Operationelle. Credit Lyonnais, Working Paper, 2001.