The XIII International Conference
"Applied Stochastic Models and Data Analysis"
(ASMDA-2009)
June 30-July 3, 2009, Vilnius, LITHUANIA

ISBN 978-9955-28-463-5
L. Sakalauskas, C. Skiadas and
E. K. Zavadskas (Eds.): ASMDA-2009
Selected papers. Vilnius, 2009, pp. 51–55

RISK MODELING FOR FUTURE CASH FLOW USING SKEW *T*-COPULA

Gaida Pettere[1], Tõnu Kollo[2]

[1]*Dept of Engeneering Mathematics, Riga Technical University, Kalku 1, Riga, Latvia*
[2]*Institute of Mathematical Statistics, University of Tartu, J. Liivi 2, Tartu, Estonia*
E-mail: [1]gaida@latnet.lv; [2]tonu.kollo@ut.ee

Abstract: In the paper we estimate future annual cash flow of an insurance company. Data of three main different business lines of an insurance company are used to predict the next year cash flow. The individual business lines are modelled by Gamma, Pareto and lognormal distributions. The analysis showed that daily payments were correlated. Therefore the joint distribution of the business lines had to be found. As the model the skew *t*-copula introduced in Kollo, Pettere (2008) was used. Simulation from the joint distribution was carried out to predict the next year cash flow. The obtained results were compared with the case when the dependencies between the business lines were not taken into account. It came out that the last model gives a too optimistic model for future payments.

Keywords: loss distributions, method of moments, simulation, skew *t*-copula, skew *t*-distribution, *VaR*

1. Introduction

From 2006, when International Financial Reporting Standard IFRS 4 became in force, an insurer has to prepare Liability adequasy test in the end of each calendar year which has to show that company's insurance liabilities are adequate using current estimates of future cash flows under company's insurance contracts. It is very common to forecast future cash flows independently for each business line in nonlife insurance. But it comes out from practice that different business lines can be highly correlated. The aim of the present paper is to model future cash flows in the situation when different business lines are correlated. We form a joint distribution of the business lines using skew *t*-copula. Then by simulation from the joint distribution we get an empirical distribution of the annual loss which is used to find the risk measure *VaR*. As information we used paid claims in current year and the history of the loss ratio.

2. Skew *t*-copula

We are going to model the joint distribution of different business lines via skew *t*-copula. The copula is introduced in Kollo, Pettere (2008). As marginal distributions of the business lines are skewed, a skewed copula will be a natural model to give a good fit with the data.
The construction of the skew *t*-copula is based on the multivariate skew *t*-distribution introduced in Azzalini ja Capitanio (2003). To define a skew *t*-distribution the multivariate *t*-distribution is needed. The density function of the p-variate *t*-distribution with ν degrees of freedom is of the form

$$t_{p,\nu}(\mathbf{x},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{(\pi\nu)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\left[1+\frac{(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu}\right]^{-\frac{\nu+p}{2}}$$

where $\boldsymbol{\Sigma}$ is a positive definite $p\times p-$matrix and $\boldsymbol{\mu}$ is a $p-$vector. The multivariate skew *t*-distribution is defined as follows.

DEFINITION 1. A random p-vector $\mathbf{X}=(X_1,...,X_p)^T$ *has p-variate skew* t*-distribution with parameters* $\boldsymbol{\mu}$, $\boldsymbol{\alpha}$ *and* $\boldsymbol{\Sigma}$, *if its density function is of the form*

$$g_{p,\nu}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha})=2\cdot t_{p,\nu}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\cdot T_{1,\nu+p}\left[\boldsymbol{\alpha}^T\mathbf{W}^{-1}(\mathbf{x}-\boldsymbol{\mu})\left(\frac{\nu+p}{Q+\nu}\right)^{\frac{1}{2}}\right], \quad (2.1)$$

where Q *denotes the quadratic form*

$Q=(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})$.

$T_{1,\nu+p}(\cdot)$ *denotes the distribution function of the central univariate* t*- distribution with* $\nu+p$ *degrees of freedom.*

The parameter $\boldsymbol{\alpha}$ is called the shape parameter and it regulates both, shape and location and $\boldsymbol{\mu}$ is considered as the location or shift parameter.

The skew *t*-copula we define through its density function (Kollo & Pettere, 2008):

DEFINITION 2. A copula is called skew t*-copula, if its density function is*

$$c_{p,\nu}(\mathbf{u};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha})=\frac{g_{p,\nu}[\{G_{1,\nu}^{-1}(u_1;0,\sigma_{11},\alpha_1),...,G_{1,\nu}^{-1}(u_p;0,\sigma_{pp},\alpha_p)\};\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha}]}{\prod_{i=1}^{p}g_{1,\nu}[G_{1,\nu}^{-1}(u_i;\mu_i,\sigma_{ii},\alpha_i);\mu_i,\sigma_{ii},\alpha_i]} \quad (2.2)$$

where *the density function* $g_{p,\nu}(\cdot;\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\alpha}):R^p\to R$ *is defined by (2.1) and function* $G_{1,\nu}^{-1}(\cdot;\mu_i,\sigma_{ii},\alpha_i)$ *:* $R^1\to I$, $i\in\{1,...,p\}$ *denotes the inverse of the univariate* $t_{1,\nu}$*-distribution function.*

We are going to apply the skew *t*-copula in a special case when the shift parameter $\boldsymbol{\mu}=\mathbf{0}$. To find a model for our data we have to estimate the parameters $\boldsymbol{\Sigma},\boldsymbol{\alpha}$ and ν. For that we shall apply the method of moments. Parameters $\boldsymbol{\Sigma}$ and $\boldsymbol{\alpha}$ are estimated from the first two moments

(Kollo & Pettere, 2008). Let $\bar{\mathbf{X}}$ and $\mathbf{S_X}$ denote the sample mean and the sample covariance matrix, respectively. Then the estimates are

$$\hat{\boldsymbol{\Sigma}}=\frac{\nu-2}{\nu}(\mathbf{S_X}+\bar{\mathbf{X}}\bar{\mathbf{X}}^T) \quad (2.3)$$

$$\hat{\boldsymbol{\alpha}}=\frac{b(\nu)\cdot\boldsymbol{\beta}}{\sqrt{b^2(\nu)-\bar{\mathbf{X}}^T\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{X}}}}, \quad (2.4)$$

where

$$\boldsymbol{\beta}=\frac{1}{b(\nu)}\hat{\mathbf{W}}\hat{\boldsymbol{\Sigma}}^{-1}\bar{\mathbf{X}}, \quad (2.5)$$

with $\hat{\mathbf{W}}=(\delta_{ij}\sqrt{\hat{\sigma}_{ij}})$, $i,j=1,...,p$, *where* δ_{ij} *is the Kronecker delta and*

$$b(\nu)=\left[\frac{\nu}{\pi}\right]^{\frac{1}{2}}\cdot\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}.$$

In formula (2.3) we have to assume that $\nu>2$.

3. Description of the model and data

The total claim amount paid in one year is the sum of the individual claims which have occurred over the selected time periods. We have taken two-days periods from 250 working days in a year. The total claim amount is found as a sum of total amounts of claims over period *j* for the business line *l*:

$$C_{j,l} = \sum_{k=1}^{K} c_{j,l,k}$$

where

K – is the number of individual claims over period j for the business line l,

$C_{j,l}$ - total amount of claims over period j for the business line l,

$c_{j,l,k}$ - individual claim over period j for the business line l.

Total amount of claims in period j over the year is affected by total amount of written premiums in years $0,-1,-2,...,-s_l$ and by inflation. To calculate cash flow for the next year one has to take into account additionally previous years premiums' impact to future claims and the inflation. Therefore we have to project claims for next year by correcting the sums $C_{j,l}$ of the previous year using formula

$$CG_{j,l} = C_{j,l}(1+i_y)^{\frac{1}{2}}(1+i_{y+1})^{\frac{1}{2}} \cdot \frac{r_{s_l} P_{y-s_l+1,l} + ... + r_1 P_{y,l}}{r_{s_l} P_{y-s_l,l} + ... + r_1 P_{y-1,l} + r_0 P_{y,l}}, \tag{3.1}$$

where i_t is inflation in year t, $P_{t,l}$ denotes written premiums in the year t in line l and loss ratios r_{n_l} are calculated in the following way

$$r_{n_l} = \frac{C_{y+n,l}}{P_{y,l}}$$

where $C_{y+n,l}$ denotes claims paid in the period $y+n$ for the contracts started in period y.

We have used data (paid claims in periods of one year) from the three most important business lines. The first and third business line are highly correlated ($r_{1,3}$ *=0.6*) while the other correlations are relatively small ($r_{1,2}$ *=0.14 and* $r_{2,3}$ *=0.114)*. Descriptive statistics of the marginal distributions of business lines are presented in Table 1.

Table 1. Descriptive statistics of used data

Business lines	1	2	3
Size	125	125	125
Mean	43 769.01	6 794.74	62 390.55
Median	41 274.28	3 285.15	59 154.16
Standard deviation	21 622.92	8 425.98	28 879.03
Skewness	0.96	2.15	0.46
Kurtosis	1.42	5.00	0.25

Marginal distributions were approximated by Gamma, Pareto and lognormal distributions. The best fit for the business line 1 was obtained by Gamma distribution, for business line 2 Pareto distribution gave the best model and business line 3 was modeled by lognormal distribution. The goodness-of-fit was measured by the Kolmogorov test (the 5% critical value equals 0,1216). Results of testing are shown in Table 2.

Table 2. Test results for the marginal distributions

Business line	Used distribution	Parameters	
1	Gamma	α	4
		β	10 680
		Test value	0.0478
2	Pareto	α	5.7
		λ	32 000
		Test value	0.0740
3	Lognormal	μ	10,9
		σ	0,44
		Test value	0,0957

4. Results

The found marginal distributions were joined into a three-dimensional distribution by the skew *t*-copula using copula density (2.2). The shape parameter $\boldsymbol{\alpha}$ and the scale parameter $\boldsymbol{\Sigma}$ were estimated from data using formulas (2.3)-(2.5). As the result we had the following estimates for our model:

$$\hat{\boldsymbol{\alpha}}^T = (0,300448\; 0,115422 - 0,059621)\,;\; \hat{\boldsymbol{\Sigma}} = \begin{pmatrix} 1,60610^6 & 9,50110^3 & 1,27910^4 \\ 9,50110^3 & 2,38810^5 & 9,88010^3 \\ 1,27910^4 & 9,88010^3 & 2,79910^4 \end{pmatrix}.$$

The number of degrees of freedom ν was taken 3 to be able to use the multivariate *t*-distribution with maximally heavy tail area. In Table 4 we have presented simulation results. Simulation is based on the representation of a random vector $\mathbf{X}$ with the multivariate skew *t*-distribution (see Kotz & Nadarajah, 2004, p. 103, for instance):

$$\mathbf{X} = \boldsymbol{\mu} + V^{1/2}\mathbf{Y},$$

where $\boldsymbol{\mu}$ is the shift parameter, νV is chi-square distributed with ν degrees of freedom and $\mathbf{y}$ has multivariate skew-normal distribution with parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\Sigma}$ (for the skew-normal distribution see Kotz & Nadarajah (2004), p. 103 or Azzalini & Capitanio (2003)). The simulation rule can be found in Cherubini, Luciano & Vecchiato (2004), p. 181.

In the simulation experiment triples were simulated from the joint 3-variate distribution of business lines. The number of replication was 150. Results of simulation are collected in Table 4. On the first line `Real values´ we have data from the current year with the total annual claim size of the portfolio. On the next lines characteristics of the simulated portfolios are given. From the obtained empirical distributions we get estimates of *VaR* for all business lines as well as for the portfolio on the last line of Table 3.

Table 3. Simulation results

	Business lines			Portfolio
	1	2	3	
Real values	10 942 253	1 698 685	15 597 638	28 238 577
Mean	10 558 135	1 725 715	16 538 033	28 821 884
Median	10 549 088	1 716 799	16 494 779	28 806 834
St. deviation	305 722	129 409	433 715	702 708
95% confidence interval for mean	10 509 209	1 705 005	16 468 624	28 709 427
	10 607 061	1 746 425	16 607 442	28 934 341
Skewness	0.04	0.24	0.23	0.19
Kurtosis	0.60	-0.10	0.01	-0.13
VaR(99.5%)	11 494 016	2 050 740	17 698 194	30 667 410

For prediction of the cash flow for the next year we have to calculate corrected claim amounts which take into account written premiums and inflation rate using formula (3.1). From these values the coefficients for all three business lines are found. The results are summarized in Table 4.

Table 4. Calculated coefficients for the next year using formula (3.1)

Business line	1	2	3
Coefficients	0.88	1.10	0.81

Now we are able to find the predicted cash flow as a linear combination of the simulated mean values of marginal total claim amounts for the next year combining data from Tables 4 and 5. In the same way we get prediction of 99.5% *VaR* for the next year. For comparison we present in the third line of Table 5 the predicted cash flow in the case when the business lines are considered as independent random variables. As one can see we underestimate the risk when we do not take into account correlations between the business lines and certain additional capital is needed.

Table 5. Projected cash flow for the next year and additional capital

Mean of future cash flow	24 131 823.80
VaR (99,5%) of future cash flow	26 706 086.33
Calculated cash flow without taking account correlation	23 505 195.35
Additional capital	3 200 890.98

5. Conclusions

The study showed that when forecasting next year cash flow for an insurance company, it is important to take into account dependencies between different business lines. Using multivariate skew *t*-copula it was possible to construct a reliable model which enables to model loss distributions with heavy tail area and gives a more adequate model than the traditional approach.

References

Azzalini, A.; Capitanio, A. 2003. Distributions generated by pertubation of symmetry with emphasis on the skew *t*-distribution,. *J. R. Stat. Soc. Ser B, Stat. Methodol.* 65: 367–389.

Cherubini, U.; Luciano, E.; Vecchiato, W. 2004. *Copula Methods in Finance.* Wiley, Chichester.

Kollo, T.; Pettere, G. 2008. Parameter estimation for the multivariate skew *t*-copula. *Test*, 9 pp. (submitted)

Kotz, S.; Nadarajah, S. 2004. *Multivariate t Distributions and Their Applications.* Cambridge University Press, Cambridge.