COPULA BASED VaR APPROACH FOR EUROPEAN STOCKS PORTFOLIO

Mária Bohdalová, Michal Greguš, Comenius University in Bratislava,
maria.bohdalova@fm.uniba.sk, michal.gregus@fm.uniba.sk

The paper gives stochastic assessments of the financial crisis and discusses the Value at Risk European stocks from the point of view of copula based approach. Copula techniques can be based on the connection between rank correlation and certain one–parameter bivariate copulas. This relation allows easy calibration of the parameters. We use more general numerical calibration techniques that are based on maximum likelihood estimation (MLE). Using this approach we want to estimate VaR of the EU stocks portfolio using Monte Carlo simulation. The focus will be on modelling the interdependence between two risk factor returns. We suppose that the risk factor returns have some assumed marginal distributions, which need not be identical, and their dependency is modelled with copulas. We find that standard parametric copula functions (such as Gaussian) are not able to provide a good fit to the data. This is especially true when one or more of the marginal distributions has fat tails. We overcome this problem by fitting a t–copula with different marginal which can approximate any possible shape for the joint density.

Introduction

Understanding the joint distribution of risk factors is fundamental for investigating and computing the Value–at–Risk (VaR) of portfolios. The accuracy of this measure has crucial importance in determining the capital requirement for financial institutions. If we analyse the distribution of losses, one verifies that large losses are influenced by simultaneous losses in risk factors. Therefore, the distribution of the losses depends on joint distribution of risk factors. Modelling dependence has the key importance to all economic fields in which uncertainty plays a large role. It is a crucial element of decision making under uncertainty and risk analysis. Consequently, an inappropriate model for dependence can lead to suboptimal decisions and inaccurate assessments of risk exposures. Traditionally, correlation is used to describe dependence between random variables, but recent studies have ascertained the superiority of copulas to model dependence.

In this paper we use multivariate data and attempt to model the joint distribution of daily log returns of two European market indices DAX and EURO STOXX 50. We find that standard parametric copula functions (such as Gaussian) are not able to provide a good fit to the data. This is especially true when one or more of the marginal distributions has fat tails. We overcome this problem by fitting a t–copula with different marginal which can approximate any possible shape for the joint density.

Literature review

Kole (2005) has shown the superiority of copulas to model dependence, as they offer much more flexibility than the correlation approach. Frees and Valdez (1998) shown the use of copulas in actuarial risk analyses and Embrechts, McNeil and Straumann (1999) introduced the copulas into finance. Li (2000) has studied the problem of default correlation in credit risk models and has shown that "the current Credit Metrics approach to default correlation through asset correlation is equivalent to using a normal copula function". In the Risk special report of November 1999 on Operational Risk, Ceske and Hernández (1999) explain that copulas may be used in conjunction with Monte Carlo methods to aggregate correlated losses.

An important reason to consider other copulas than the correlation–implied Gaussian copula is its failure to capture dependence of extreme events. Especially in the financial crisis, because sometimes extreme losses occur, the financial data are approximated over skewed distribution. Due to financial crisis in 2008, U.S. stock markets have suffered their worst volatile trading days in memory, and various stock indices have fallen dramatically. The Dow Jones and S&P 500 are on course to record their worst yearly returns since the Great Depression. Meanwhile, U.S. stock index futures fluctuated a day after the Dow Jones snapped a seven–day losing streak. There is a number of empirical evidence that the dependence between many important asset returns is non–normal in crisis. Pearson correlation coefficient is not an appropriate dependence measure for very fat–tailed risks when extreme losses occurred. This inadequacy of correlation requires an appropriate dependence measure. Copula method may be the right tool for the job, which is applied to research on non–normal dependence of financial time series (Chai & Guo, 2011). However, still no consensus that copula is more suitable for use in specific applications, or which methods for estimating the copula parameters are better.

A precise estimation of parameters in copula models is crucial to dependence modelling (Zhang & Ng, 2010). In the literature several ways, based on the statistical inference theory, were developed to estimate the parametric and non–parametric copula models (Joe, 1997). These approaches can be mainly classified into three types: parametric approaches (e.g. the maximum likelihood estimation), semi–parametric estimation and non–parametric methods. The maximum likelihood estimations usually include the exact maximum likelihood method (EML) and the inference for margins method (IFM). The EML method for the parameter estimation of complex copula models, such as a high–dimensional copula model, could be computationally intensive while using traditional numerical methods. Because, the EML jointly estimates the marginal distribution parameters and the dependence structure (copula) parameters, the solutions from traditional optimization approaches tend to stuck in local optima. Joe (1997) proposed the IFM approach, a computationally simpler approach that first estimates the marginal distribution parameters and then the copula parameters. However, the estimators from the IFM method do not hold equality with the EML estimation in general. Due to this reason, the former set of estimates is usually used as a starting guess for the „two–step "maximum likelihood method (Zhang & Ng, 2010).
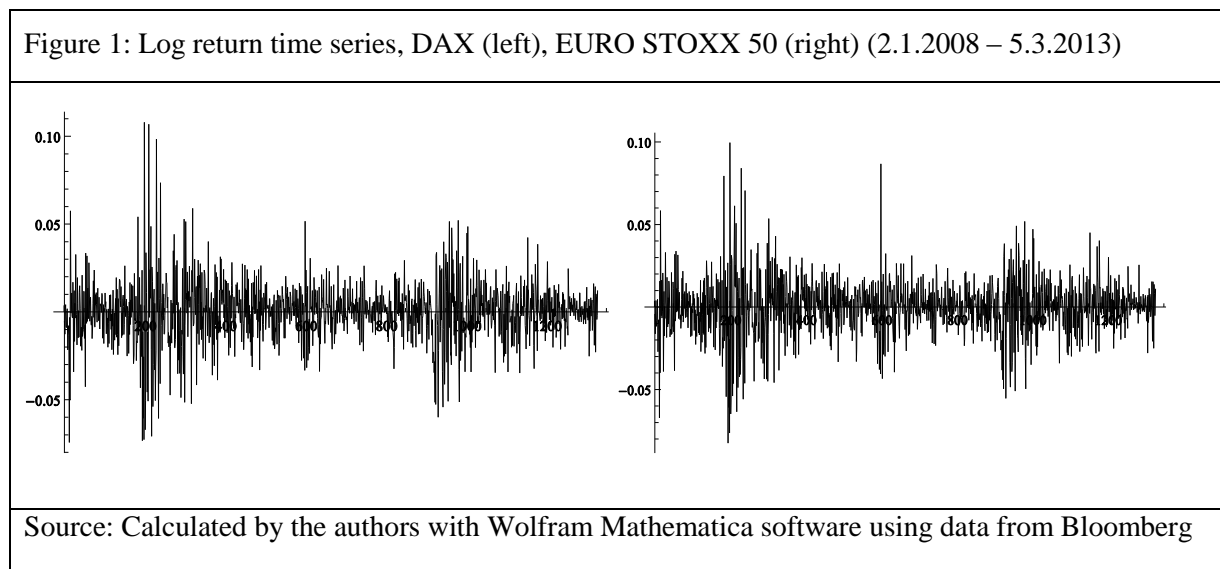
The primary motivation for this paper is as follows. Copula models for financial time series are used to extract the dependence between two market stock indices and its underlying asset when financial crisis breaks out. In this study, we empirically examine the log return time series of DAX and EURO STOXX 50 market indices. It is concluded that, these stock indices are not normally distributed in financial crisis and that the *t*–copula function can provide a better fit to the empirical data. In this paper we compare EML with IFM estimation results for estimation of copula parameters.

The paper is structured as follows. Next section gives data and methodology of this paper. A briefly describes copula functions as an intuitive and practical way to simplify the problem of modelling multivariate distributions and the maximum likelihood methods for estimation of the copula parameters. Section Results and Discussion reports our results for DAX and EURO STOXX 50 market indices. Section Conclusion concludes and discusses several possible avenues of future research.

Data and methodology

Assume that we should determine the VaR of a market indices portfolio equally proportioned from DAX and EURO STOXX 50, for next two weeks (10 trading days) for confidence level α=0.01 and α=0.05. The sample period for historical data is January 2, 2008 to March 5, 2013 (denoted as 1…1328). Data has been downloaded from Bloomberg. DAX (Deutscher Aktien IndeX) is a stock market index consisting of 30 major German companies trading on the Frankfurt Stock Exchange, the most important stock index of DTB (Deutsche boerse) in Germany. EURO STOXX 50 Index is leading Blue-chip index for Eurozone, provides a Blue-chip representation of supersector leaders in the Eurozone. The test data is the natural logarithm return of the closing price. All the estimation processes are carried out in Wolfram Mathematica v. 9.

The summary statistics is given in Table 1 and gives, for example, the annualized mean, annualized standard deviation, median, max, min, skewness, kurtosis, Jarque–Bera test of normality. In both cases, the null hypothesis of normality is rejected at any level of significance, and there is evidence of significant excess kurtosis of the return series. This indicates that the distributions of these return series are non–Gaussian. Daily log returns series and their histograms are shown in Figure 1 and Figure 2 respectively.

Figure 1: Log return time series, DAX (left), EURO STOXX 50 (right) (2.1.2008 – 5.3.2013)

Source: Calculated by the authors with Wolfram Mathematica software using data from Bloomberg

Maximum likelihood for copula parameters estimation

Copulas were initially introduced by (Sklar, 1959). Let H denote a joint distribution of function with margins $F_1, …, F_d$, then there exists an unique copula C

$$H(x_1; …, x_d) = C(F_1(x_1); …; F_d(x_d));$$ (1)

if $F_1; \dots; F_d$ are continuous functions. The copula model interprets multivariate distributions by coupling the marginal distribution function $F_{x_1}(x_1), \dots, F_{x_d}(x_d)$ with the dependence structure C (Nelsen, 1998, referenced by Zhang & Ng, 2010).

Table 1: Summary statistics of daily log returns of European market stocks indices. (2.1.2008 – 5.3.2013)

Descriptive statistics	DAX	EURO STOXX 50
Mean	-0.00001	-0.00032
Mean Annualized	-0.00389	-0.08004
Median	0.000681	-0.00018
Max	0.107975	0.099599
Min	-0.07433	-0.0825
Stand. Dev.	0.01693	0.016771
Stand. Dev. Ann	0.267786	0.265178
Skewness	0.120292	0.009396
Kurtosis	8.263859	7.120973
Jarque–Bera t stat	1556.472	952.4647
Jarque–Bera p value	0.0000	0.0000
Sample size	1328	1328

Source: Calculated by the authors with Wolfram Mathematica software using data from Bloomberg

Figure 2: Histograms of log return time series, DAX (left), EURO STOXX 50 (right), (2.1.2008 – 5.3.2013)
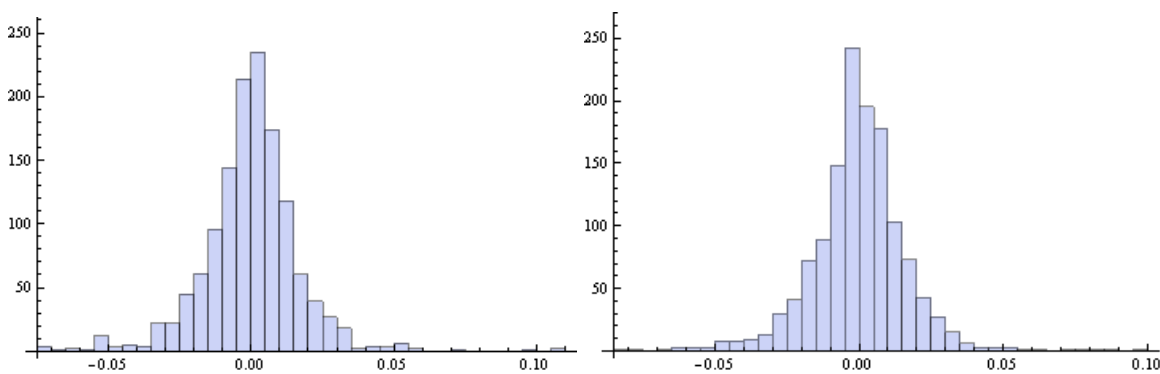
Source: Calculated by the authors with Wolfram Mathematica software using data from Bloomberg

In other words, the joint distribution can be expressed by combining the marginal distributions with the dependence structure, yielding

$$C(u_1, \dots, u_d) = H(F_1^{-1}(u_1), \dots, F_d^{-1}(u_d)); \tag{2}$$

with $u \in [0; 1]^d$, and $F_i^{-1}(\cdot)$ denoting the inverse of the marginal distribution $F_i(\cdot)$. In this paper, the general Student t distribution and Student t copula are used to model the marginal distribution $F_i(\cdot)$ and the dependence structure $C(\cdot)$, respectively.

Particularly in finance and risk management, the Student t distribution has been used instead of the normal distribution, because of its fat tail behaviour, which can be applied to capture financial extreme events (Bollerslev, 1987, referenced by Zhang & Ng, 2010). The marginal distributions of a multivariate t distribution are univariate Student t distributions. The probability density function $f(\cdot)$ of general Student t distributions can be written as

$$f_{\eta,\mu,\sigma}^t(x) = \frac{\Gamma(\frac{\eta+1}{2})}{\Gamma(\frac{\eta}{2})} \frac{1}{\sqrt{\eta\pi\sigma^2}} \left(1 + \frac{1}{\eta}\frac{(x-\mu)^2}{\sigma^2}\right)^{-\frac{\eta+1}{2}},$$ (3)

where $\Gamma(\cdot)$ is the Gamma function, η denotes the marginal degrees of freedom, μ and σ represent location and dispersion of the marginal distribution respectively (Meucci, 2005, referenced by Zhang & Ng, 2010).

According to (Sklar, 1959, referenced by Zhang & Ng, 2010), the Student t copula of the random vector u can be expressed as

$$C_{\eta,\rho}^t(u) = t_{v,\rho}\left(t_v^{-1}(u_1),\ldots,t_v^{-1}(u_d)\right),$$ (4)

where $\rho_{i,j}$ is a correlation coefficient, with $i, j \in 1, \ldots, d$, composed into covariance matrix Σ, $t_{v,\rho}(\cdot)$ denotes the distribution function $H(\cdot)$ and $t_v^{-1}(\cdot)$ represents the inverse od the marginal t distribution function $F_i^{-1}(\cdot)$. The corresponding Student t copula density

$$c(u_1,\ldots,u_d) = \frac{\partial^d C(u_1,\ldots,u_d)}{\partial u_1 \ldots \partial u_d}$$ (5)

can be written as

$$c_{v,\rho}^t(u_1,\ldots,u_d) = \frac{1}{\sqrt{|\rho|}} \frac{\Gamma\left(\frac{v+d}{2}\right)\Gamma\left(\frac{v}{2}\right)^{d-1}}{\Gamma\left(\frac{v+1}{2}\right)^d} \frac{\prod_{j=1}^d \left(1 + \frac{y_j^2}{v}\right)^{\frac{v+1}{2}}}{\left(1 + \frac{y'\rho^{-1}y}{v}\right)^{\frac{v+d}{2}}}.$$ (6)

We ought to note that, if the degree of freedom η of the marginal distribution of equation (3) is consistent with the degree of freedom v in copula function in (6), the multivariate distribution is referred to as a multivariate t distribution (McNeil et al., 2005).

The complete copula model has two parts, the marginal cumulative distribution $F_j(\cdot)$ and a joint cumulative distribution $H(\cdot)$. The distribution parameters of the complete copula models should be estimated jointly according to the exact maximum likelihood (EML) method.

The log-likelihood function ℓ_j^m of the j-th Student t marginal distribution can be written as

$$\ell_j^m = -n_0\left[\log(\sigma_j)+\log\left(\sqrt{\eta_j}\right)+\log\left(\sqrt{\pi}\right)+\log\left(\Gamma\left(\frac{\eta_j}{2}\right)\right)+\log\left(\Gamma\left(\frac{1+\eta_j}{2}\right)\right)\right]$$
$$-\left(\frac{\eta_j+1}{2}\right)\cdot\sum_{i=1}^{n_0}\log\left(1+\frac{(x_{j,i}-\mu_j)^2}{\sigma_j^2\eta_j}\right),\tag{7}$$

where n_0 is the observation number, μ_j, σ_j, η_j denote location, dispersion and degree of freedom of the j-th marginal distribution, respectively. The log-likelihood function ℓ^C of the Student t copula density in equation (6) can be written as

$$\ell^C = n_0\cdot\left[-\frac{1}{2}\cdot\log(|\rho|)-2\cdot\log\left(\Gamma\left(\frac{\nu+1}{2}\right)\right)+\log\left(\Gamma\left(\frac{\nu+2}{2}\right)\right)+\log\left(\Gamma\left(\frac{\nu}{2}\right)\right)\right]$$
$$+\sum_{j=1}^{d}\sum_{i=1}^{n_0}\frac{\nu+1}{2}\cdot\log\left(1+\frac{y_{j,i}^2}{\nu}\right)-\frac{\nu+2}{2}\cdot\sum_{i=1}^{n_0}\log\left[1+\frac{1}{\nu}y_i^{'}\cdot\rho^{-1}y_i\right],\tag{8}$$

where d denotes the dimension of the risk factors, $y_{j,i}$ represents the inverse transform of Student t distribution with ν degree of freedom for the i-th observation of the j-th risk factor after a strictly increasing transformation of the original observation $x_{j,i}$.

ELM estimation for complex copula models could be computationally demanding and therefore we can use the inference for margins (IFM) approach, which can obtain the estimates more simply, but at the cost of a higher bias. The IFM approach first estimates the parameters of marginal distributions, such as the one in equation (8) (Zhang & Ng, 2010). Then the variables $x_{j,i}$ are transferred into $y_{j,i}$ based on the estimated parameters of the marginal distribution. After that, the inference of the copula parameters in equation (8) is performed while taking the $y_{j,i}$ as input observations. The IFM approach is a two-step procedure and it can be implemented by using traditional numerical approaches. However, the IFM approach cannot guarantee the parameter η_j in equation (7) and the ν in equation (8) being consistent. In contrast to the IFM approach, the EML estimation overcomes the barrier since it estimates the marginal distributions and the copula density jointly. The objective function used in the EML approach is simply defined as

$$\ell = \ell^C + \sum_{j=1}^{d}\ell_j^m,\tag{9}$$

which has been discussed in the work of (Zhang & Ng, 2010b).

Estimation of the copula parameters is based on the maximization of the objective function, i.e. the log–likelihood functions from the complete copula model defined in equation (9). The fitness of the final objective function is defined as the sum of log–likelihood values of both the marginal and copula density functions. The fitness value of the objective function U depends on μ_j, σ_j, ρ and ν, thus the optimization problem can be simply formulated as

$$\max_{\mu,\sigma,\rho,\nu} U = l\tag{10}$$

subject to $1 > \rho > -1$, $\nu > 3$.

In practice, when ν is greater than 30, the Student t copula can be approximated by using the Gaussian copula, which does not consider any tail dependence (Fantazzini, 2009, referenced by Zhang & Ng, 2010). When ν is smaller than 3, the third and fourth moments of the distribution are not defined.

Therefore, the minimum value of v is constrained as greater than 3 in the maximum likelihood estimation. In order to solve the optimization problem, two population based evolutionary methods are utilized to search optimal solutions for the copula model while taking the marginal distributions and the dependence structure into account simultaneously.

Correspondence between Copulas and Rank Correlation

McNeil (2005) referenced by Alexander (2008) showed that Kendall's tau (τ) has a direct relationship with a bivariate copula function $C(u_1,u_2)$ as follows:

$$\tau = 4\int_0^1\int_0^1 C(u_1,u_2)\,dC(u_1,u_2)-1 \tag{11}$$

If the copula depends on one parameter then (11) provides a means of calibrating this parameter using a sample estimate of the rank correlation. Equation (11) has simple solution for instance for the bivariate normal copula. For the correlation ρ from the identity (11) yields

$$\rho = \sin\left(\frac{\pi}{2}\tau\right). \tag{12}$$

Note that equation (12) can be applied to the Student t copula and to the copula based in an elliptical distribution (Lindskog et al., 2003, referenced by Alexander, 2008). These authors have showed that for the normal copula there is a relationship between the correlation parameter and Spearman's rho (ρ_S):

$$\rho = 2\sin\left(\frac{\pi}{6}\rho_S\right) \tag{13}$$

Selection the best copula

A straightforward way to determine which copula provides the best fit to the data is to compare the values of the optimized likelihood function. But the more parameters in the copula, the higher the likelihood tends to be. So to reward parsimony in the copula specification the *Akaike information criterion* (*AIC*) or the *Bayesian information criterion* (*BIC*) can be applied. The *AIC* is defined as

$$AIC = 2k - 2\ln\ell \tag{14}$$

where $\ln\ell$ is the optimized value of the log likelihood function and k is the number of parameters to be estimated; and the *BIC* is defined as

$$BIC = T^{-1}\left(k\ln T - 2\ln\ell\right) \tag{15}$$

where T is the number of data points. Then the copula that yields the lowest value of the *AIC* or the *BIC* is considered to be the best fit (Alexander, 2008).

Results and Discussion

Figure 3 shows a scatter plot of daily log return on DAX index on the horizontal axis and the daily log return on EURO STOXX 50 index on the vertical axis. We use these data to calibrate the Student t copula. We assume the marginals are Student t distributed.

Figure 3: Scatter plot on DAX and EURO STOXX 50 indices log return time series, (2.1.2008 – 5.3.2013)
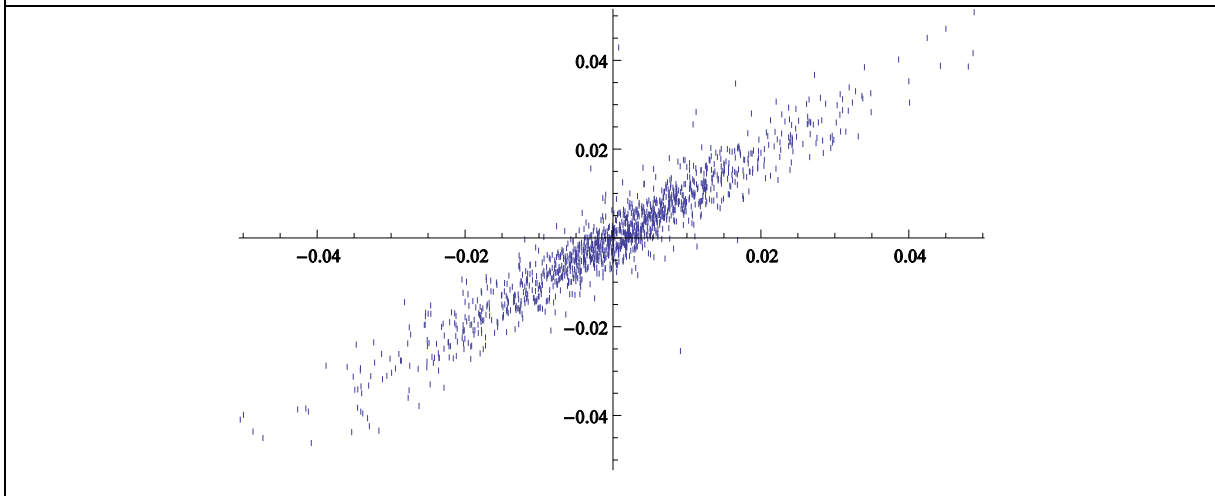
Source: Calculated by the authors with Wolfram Mathematica software using data from Bloomberg

We found the sample mean and standard deviation for both log returns series. The degree of freedom was obtained for standardized returns (with zero mean and unit variance) using maximum likelihood to fit a standardized t distribution to both standardized log returns series. These calibrated parameters are shown in Table 2.

Table 2: Calibrated parameters for Student t marginals (2.1.2008 – 5.3.2013)

Series	Mean	Standard Deviation	Degree of freedom
DAX	-0.00001	0.01693	3.4475
EURO STOXX 50	-0.00032	0.01677	3.7635

Source: Calculated by the authors with Wolfram Mathematica software unsing data from Bloomberg

Now we calibrate the copula parameters. The bivariate Student t copula has two parameters: the correlation ρ and the degrees of freedom v. In this paper we compare the calibration of the copula parameters under two different approaches:

1. EML approach: we have calibrated both ρ and v simultaneously using MLE.

2. IFM approach: we calibrate ρ first using the relationship (12) with Spearman`s rho and then we use MLE to calibrate v.

The results are shown in Table 3. The lowest value of the *AIC* or the *BIC* is obtained for EML approach.

Table 4 shows VaR results for 1% and 5% 10–days VaR. We compare the Normal copula with normal or t –marginals with Student t –copula with t marginals. Normal copula VaR with normal marginals gives the lowest estimate. Student t-copula with t–marginals gives the highest estimate of VaR thus taking into account fat tails.

Table 3: Copula parameters calibrations, 2.1.2008 – 5.3.2013

	Student t Copula (EML)		Student t Copula (IFM)	
	v	ρ	N	ρ_S
Calibrated parameter	6.6227	0.97722	6.66433	0.95447
AIC	-4495.64		−4300.96	
BIC	-3.37745		−3.23085	

Source: Calculated by the authors with Wolfram Mathematica software using data from Bloomberg

Table 4: VaR of the portfolio 2.1.2008 – 5.3.2013

	Normal copula		Student t- copula
	Normal marginals	t- marginals	t- marginals
1% 10–day VaR	0.81%	1.51%	1.52%
5% 10–day VaR	0.55%	0.77%	0.78%

Source: Calculated by the authors with Wolfram Mathematica software using data from Bloomberg

Conclusion

In this paper we have proposed Student t copula approach to calculate VaR of the portfolio. We have applied two estimation approaches EML and IFM to calibrate copula parameters. Our results indicate that EML approach gives a better estimate of copula parameters. For VaR estimation we compare three approaches – we use normal copula (standard approach), normal copula with Student marginal and Student t copula with Student t marginals (calibrated using EML approach). Student t copula could be used if returns extremely oscillate. Through a simple simulation study, it has been proven that Student t copula with Student t marginals gives reasonably good estimation of VaR.

Acknowledgment

References

Alexander C. (2008). *Market Risk Analysis*. Chichester, England: John Wiley & Sons.

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69, 542-547. http://dx.doi.org/10.2307/1925546

Ceske, R. & Hernandez, J. (1999). Where Theory Meets Practice. Operational Risk Special Report. *Risk*, 11, 17–20.

Chai, S. & Guo, Ch. (2011). Copula-Based Dependence Analysis of U.S. Stock Index and Futures Time Series in Financial Crisis. *International Business and Management*, 3 (1), 82-85.

Cherubini, U., Luciano, E. & Vecchiato, W. (2004). *Copula methods in finance*. London, England: John Wiley & Sons. http://dx.doi.org/10.1002/9781118673331

Embrechts, P., Lindskog, F. & McNeil, A. J. (2011). *Modelling dependence with copulas and application to risk management: Handbook of Heavy Tailed Distributions in Finance*. Amsterdam, Netherlands: Elsevier. PMid:11312524

Embrechts, P., McNeil, A. J. & Straumann, D. (2001). *Correlation and Dependence in Risk Management. Properties and Pitfalls.* Cambridge, England: Cambridge University Press.

Fantazzini, D. (2009). Three-stage semi-parametric estimation of t-copulas: Asymptotics, finite-sample properties and computational aspects. *Computational Statistics and Data Analysis*, http://dx.doi.org/10.1016/j.csda.2009.02.004

Frees, E. W. & Valdez, E. A. (1998). Understanding relationships using copulas. *North American Actuarial Journal*, 2 (1), 1-25. http://dx.doi.org/10.1080/10920277.1998.10595667

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London, England: Chapman & Hall. http://dx.doi.org/10.1201/b13150

Kole, E., Koedijk, K. & Verbeek, M. (2005). *Testing copulas to model financial dependence.* Retrieved March, 10, 2013, from http://www.fbv.kit.edu/symposium/10th/papers/Kole_Koedijk_Verbeek%20-%20Testing%20copulas%20to%20model%20financial%20dependence.pdf

Li, D. X. (2000). *On default correlation: a copula function approach*. (Working Paper of RiskMetrics Group). Retrieved March, 10, 2013, from http://www.stat.ncsu.edu/people/bloomfield/courses/st810j/wrap/defcorr.pdf

Lindskog, F., McNeil, A. J. & Schmock, U. (2003). *Kendall's tau for elliptical distributions*. Heidelberg, Germany: Physica Verlag. PMCid:2747540

McNeil, A. J., Frey, R. & Embrechts, P. (2005). *Quantitative Risk Management.* Princeton, NJ: Princeton University Press.

Meucci, A. (2005). *Risk and Asset Allocation*. Berlin, Germany: Springer Verlag. http://dx.doi.org/10.1007/978-3-540-27904-4

Nelsen, R. B. (1998). *An Introduction to Copulas*. Berlin, Germany: Springer.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges [Distribution functions in *n* dimensions and their margins]. *Publications de l'institut de statistique de l'Université de Paris*, 8, 229-231.

Zhang, J. & Ng, W. L. (2010). EML-estimation of multivariate t copulas with heuristic optimization. In *Proceedings of International Conference on Computer Science and Information Technology 2010*, p. 469-473.

Zhang, J. & Ng, W. L. (2010). Exact Maximum Likelihood Estimation for Copula Models. (Research report of COMISEF). Retrieved March 10, 2013, from http://comisef.eu/files/wps038.pdf