COMISEF WORKING PAPERS SERIES

WPS-038 17/05/2010

Exact Maximum Likelihood Estimation for Copula Models

J. Zhang
W. L. Ng

Exact Maximum Likelihood Estimation for Copula Models

Jin Zhang*   Wing Long Ng [†]

Abstract

In recent years, copulas have become very popular in financial research and actuarial science as they are more flexible in modelling the co-movements and relationships of risk factors as compared to the conventional linear correlation coefficient by Pearson. However, a precise estimation of the copula parameters is vital in order to correctly capture the (possibly nonlinear) dependence structure and joint tail events. In this study, we employ two optimization heuristics, namely Differential Evolution and Threshold Accepting to tackle the parameter estimation of multivariate t distribution models in the EML approach. Since the evolutionary optimizer does not rely on gradient search, the EML approach can be applied to estimation of more complicated copula models such as high-dimensional copulas. Our experimental study shows that the proposed method provides more robust and more accurate estimates as compared to the IFM approach.

Key words. Copula Models, Parameter Inference, Exactly Maximum Likelihood, Differential Evolution, Threshold Accepting.

1   Introduction

Nowadays, copulae have been widely applied by practitioners to model the dependence structure of financial risk factors, such as equities and exchange rates. The popularity of copulae is mainly due to their flexibility as they can be used to model both the linear and non-linear dependence structure of a multivariate distribution. The linear correlation by Pearson is not only insufficient in describing the dependence of risk factors which moving away from elliptical distributions, but

---

*Centre for Computational Finance and Economic Agents, University of Essex, United Kingdom. `jzhangf@essex.ac.uk`

[†]Centre for Computational Finance and Economic Agents, University of Essex, United Kingdom. `wlng@essex.ac.uk`

also inconsistent under nonlinear strictly increasing transformations of risk factors (see McNeil et al. [2005]). Therefore, using copula-based dependence measures will be more accurate in capturing the dependence structure than calculating the linear correlation.

However, a precise estimation of parameters in copula models is crucial to dependence modelling. In the literature, several ways based on the statistical inference theory were developed to estimate the parametric and non-parametric copula models (see 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. Furthermore, since 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 with that from the EML estimation in general. Due to this reason, the former set of estimates are usually used as a starting guess for the latter, leading a cumbersome procedure, i.e. a 'two-step' maximum likelihood.

Two evolutionary methods, namely Differential Evolution (DE) and Threshold Accepting (TA), are used to tackle the parameter inference problem of the multivariate t copula models under the EML framework in this paper. By resorting to the evolutionary approaches, the parameter inference can be realized in a one-step estimation procedure for the EML estimation, and the approaches do not require any starting guess of decision variables. The proposed approach is particularly suitable for the inference of complicated copula models by using the EML estimation, traditional optimization procedures tend to stop at local optima in such cases.

The structure of this paper is organized as follows. Section 2 introduces the copula model and the parameter inference problem. Section 3 presents the optimization problem and the evolutionary methods for solving the problem. Section 4 reports the experiment results and Section 5 summarizes the paper.

2 Copula Theory

Copula has become an important tool in finance with various applications, e.g., risk management, derivatives pricing, portfolio management, etc. In fact, copulae 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

$$H(x_1, ...x_d) = C(F_1(x_1), ..., F_d(x_d)), \tag{1}$$

if $F_1, ..., F_d$ are continuous. The copula model interprets multivariate distributions by coupling the marginal distribution function $F_{x_1}(x_1), ..., F_{x_d}(x_d)$ with the dependence structure C (Nelsen [1998]). In other words, the joint distribution can be expressed by combining the marginal distributions with the dependence structure, yielding

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

with $\boldsymbol{u} \in [0, 1]^d$, 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 behavior, which can be applied to capture financial extreme events (Bollerslev [1987]). 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

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

where $\Gamma(\cdot)$ is the Gamma function, η denotes the marginal degrees of freedom (DoF), μ and σ represent location and dispersion of the marginal distribution respectively (see Meucci [2005]).

According to Sklar [1959], the Student t copula of the random vector $\mathbf{u}$ can be expressed as

$$C_{\nu,\boldsymbol{\rho}}^t(\mathbf{u}) = t_{\nu,\boldsymbol{\rho}}(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d)), \tag{4}$$

where $\rho_{i,j} = \Sigma_{i,j}/\sqrt{\Sigma_{i,i}\Sigma_{j,j}}$, with $i, j \in 1, ..., d$. Σ is the variance-covariance matrix, $t_{\nu,\rho}(\cdot)$ denotes the distribution function $H(\cdot)$, and $t_\nu^{-1}(\cdot)$ represents the inverse of the marginal t distribution function $F_i^{-1}(\cdot)$. The corresponding Student t copula density $c(u_1, ..., u_d) = \frac{\partial^d C(u_1, ..., u_d)}{\partial u_1 ... \partial u_d}$ can be written as

$$c_{\nu,\boldsymbol{\rho}}^t(u_1, ...u_d) = \frac{1}{\sqrt{|\boldsymbol{\rho}|}} \frac{\Gamma(\frac{\nu+d}{2})\Gamma(\frac{\nu}{2})^{d-1}}{\Gamma(\frac{\nu+1}{2})^d} \frac{\prod_{j=1}^d (1 + \frac{\boldsymbol{y}_j^2}{\nu})^{\frac{\nu+1}{2}}}{(1 + \frac{\boldsymbol{y}'\boldsymbol{\rho}^{-1}\boldsymbol{y}}{\nu})^{\frac{\nu+d}{2}}}. \tag{5}$$

One should note that, if the DoF η of the marginal distribution of Eq. (3) is consistent with the DoF ν in the copula function in Eq. (5), the multivariate distribution is referred to as a multivariate t distribution (see McNeil et al. [2005]).

The complete copula model has two parts, the marginal cumulative distributions $F_j(\cdot)$ and a joint cumulative distribution $H(\cdot)$. Ideally, 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

$$\ell_j^m = -\mathfrak{n}_o \cdot \left[\log(\sigma_j) + \log(\sqrt{\eta_j}) + \log(\sqrt{\pi}) + \log\left(\Gamma\left(\frac{\eta_j}{2}\right) \right) + \log\left(\Gamma\left(\frac{1+\eta_j}{2}\right) \right) \right]$$
$$- \left(\frac{\eta_j + 1}{2}\right) \cdot \sum_{i=1}^{\mathfrak{n}_o} \log\left(1 + \frac{(x_{j,i} - \mu_j)^2}{\sigma_j^2 \cdot \eta_j} \right), \tag{6}$$

where $\mathfrak{n}_o$ is the observation number; and μ_j, σ_j, η_j denote location, dispersion and DoF of the j-th marginal distribution, respectively. The log-likelihood function ℓ^C of the Student t copula density in Eq. (5) can be written as

$$\ell^C = \mathfrak{n}_o \cdot \left[-\frac{1}{2} \cdot \log(|\boldsymbol{\rho}|) - 2 \cdot \log\left(\Gamma\left(\frac{\nu+1}{2}\right) \right) + \log\left(\Gamma\left(\frac{\nu+2}{2}\right) \right) + \log\left(\Gamma\left(\frac{\nu}{2}\right) \right) \right]$$
$$+ \sum_{j=1}^{d} \sum_{i=1}^{\mathfrak{n}_o} \frac{\nu+1}{2} \cdot \log\left(1 + \frac{y_{j,i}^2}{\nu} \right) - \frac{\nu+2}{2} \cdot \sum_{i=1}^{\mathfrak{n}_o} \log\left[1 + \frac{1}{\nu} \boldsymbol{y_i}' \cdot \boldsymbol{\rho}^{-1} \cdot \boldsymbol{y_i} \right], \tag{7}$$

where d denotes the dimension of the risk factors; $y_{j,i}$ represents the inverse transform of Student t with ν DoF for the i-th observation of the j-th risk factor after a strictly increasing transform (i.e. the Student t cumulative distribution function) of the original observation $x_{j,i}$.

Since the EML estimation for complex copula models could be computationally burdensome, the literature suggests 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 Eq. (6). 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 Eq. (7) 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, such as the Newton-Raphson algorithm. However, the IFM approach cannot guarantee the parameter η_j in Eq. (6) and the ν in Eq. (7) 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

$$\ell = \ell^C + \sum_{j=1}^{d} \ell_j^m,$$ (8)

which has been discussed in the work of Zhang and Ng [2010].

3 Maximum Likelihood for Parameter Estimation

3.1 Optimization Problem

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 Eq. (8). 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 $\mathfrak{O}$ depends on μ_j, σ_j, ρ and ν, thus the optimization problem can be simply formulated as

$$\max_{\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{\rho},\nu} \mathfrak{O} = \ell$$ (9)

subject to

$$1 > \rho > -1, \quad \nu > 3.$$

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 (see Fantazzini [2009]). When ν is smaller than 3, the third and fourth moments of the distribution are not defined. Therefore, the minimum value of ν is constrained as grater 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.

3.2 Differential Evolution

Heuristic methods provide ways of tackling combinatorial optimization problems. Differential Evolution (DE) which was originally proposed by Storn and Price Storn and Price [1997], is a population based heuristic method for solving the optimization problems with continuous space. The approach generates new solutions by linear combination and cross-over based on current solutions. The resulting solution would replace the current best solution if the new solution has a higher

---

Algorithm 1 Differential Evolution.

1: randomly initialize population of vectors $\boldsymbol{\imath_p}$, $p= 1,...,P$
2: **while** the halting criterion is not met **do**
3:    **for** all current solutions $\boldsymbol{\imath_p}$, $p=1,...,P$ **do**
4:       randomly pick three different solutions, i.e. $p_1 \neq p_2 \neq p_3 \neq p$
5:       $\imath_c[i] \leftarrow \imath_{p_1}[i]+(K+z_1[i])(\imath_{p_2}[i]-\imath_{p_3}[i]+z_2[i])$ with probability π_1, or $\imath_c[i] \leftarrow \imath_p[i]$ otherwise
6:       compute the fitness value of $\boldsymbol{\imath_p}$, i.e. the sum of log-likelihood value of the marginal and copula density functions
7:    **end for**
8:    **for** the current solution $\boldsymbol{\imath_p}$, $p = 1,...,P$ **do**
9:       **if** Fitness($\boldsymbol{\imath_c}$) > Fitness($\boldsymbol{\imath_p}$) **then** $\boldsymbol{\imath_p} \leftarrow \boldsymbol{\imath_c}$ **end if**
10:    **end for**
11: **end while**

---

fitness value. For each current solution $\boldsymbol{\imath_p}$, a new solution $\boldsymbol{\imath_c}$ is generated from the following procedure: randomly selecting three different members from the current population ($p_1 \neq p_2 \neq p_3 \neq p$); linearly combining the solution vectors at probability π_1, or inheriting the original p-th solution otherwise. We use the 'Ditter' and 'Jitter' version of the standard DE (see Price *et al.* Price et al. [1998]), which considers adding normally distributed random numbers to the weighting factor K, and the difference of two solution vectors respectively. Vectors $\boldsymbol{z_1}$ and $\boldsymbol{z_2}$ represent the extra noise in the algorithm; they contain random numbers being zero with probability $\boldsymbol{\pi_2}$ and $\boldsymbol{\pi_3}$ respectively, or independently follow the normal distributions $N(0, \mathfrak{d}_1^2)$ and $N(0, \mathfrak{d}_2^2)$.

The DE algorithm is described by the pseudo code in Algorithm 1. π_1 is the cross-over probability. After the linear combination and cross-over, DE updates the population. More specifically, if the fitness value of $\boldsymbol{\imath_c}$ is higher than the one of $\boldsymbol{\imath_p}$, the solution $\boldsymbol{\imath_p}$ is replaced by $\boldsymbol{\imath_c}$, and the updated $\boldsymbol{\imath_p}$ exists in the current population; otherwise $\boldsymbol{\imath_p}$ survives.

The parameter settings of the DE algorithm used for solving the optimization problem are listed as follows. Population size and iteration number were set at 50 and 500; the value of K was set at a value 0.5; and the crossover probability $\mathfrak{z}_1$ was at 60%. The parameters used for generating the extra noises $\mathbf{z}_1$ and $\mathbf{z}_2$ were : $\mathfrak{z}_2 = 50\%$, $\mathfrak{z}_3 = 10\%$, $\mathfrak{d}_1^2 = 0.1$ and $\mathfrak{d}_2^2 = 0.1$. Repair functions are used to translate these solutions to values which meet the following constraints

$$\nu = 3 + |\dot{\nu}| \tag{10}$$

$$\rho = \begin{cases} \exp \frac{-0.15}{|\dot{\rho}|} & \text{if } \dot{\rho} > 0, \\ -\exp \frac{-0.15}{|\dot{\rho}|} & \text{if } \dot{\rho} < 0, \end{cases} \tag{11}$$

where $\dot{\nu}$ and $\dot{\rho}$ are used to represent the solutions from the DE. The above repair mechanism was utilized to guarantee that the two parameters from DE met the parameter requirement from the copula model: $1 > \rho > -1$, $\nu > 3$. Eq. (11) is analogous to the logistic function in the literature which maps real numbers to values within a specified range.

3.3 Threshold Accepting

Dueck and Scheuer [1990] introduced Threshold Accepting (TA), and Winker [2001] gave a comprehensive discussion of TA and its applications in economics. TA is a refined version of the standard local search procedure, mainly it differs from the standard approaches in its acceptance criterion. Given a minimization problem, let $\dot{\mathbf{s}}^c$ denote an initial solution and $\dot{\mathbf{s}}^n$ represent a candidate solution in the neighborhood of the initial element $\mathfrak{N}(\dot{\mathbf{s}}^c)$, TA will accept $\dot{\mathbf{s}}^n$ as a new solution if and only if the solution is better than $\dot{\mathbf{s}}^c$ in terms of an objective function, i.e. $\mathfrak{O}(\dot{\mathbf{s}}^n) - \mathfrak{O}(\dot{\mathbf{s}}^c) < \mathfrak{T}$ for some pre-assigned non-negative threshold value $\mathfrak{T}$. The threshold $\mathfrak{T}$ is decreased gradually and reaches the value of zero after a given number of steps $\mathfrak{n}_s$. This algorithm was applied to solve the optimization problem by simply changing the sign of the objective function, which turned the original maximization problem to minimization.

To generate a candidate solution in the neighborhood of $\dot{\mathbf{s}}^c$, normally distributed randomness from $\mathcal{N}(0, \mathfrak{d}_3^2)$ was added to each gene in the chromosome of the $\dot{\mathbf{s}}^c$ at a probability of $\mathfrak{z}_4$; otherwise the gene of the new solution $\dot{\mathbf{s}}^n$ inherits the one of $\dot{\mathbf{s}}^c$. The parameter $\mathfrak{d}_3^2$ and $\mathfrak{z}_4$ were assigned with values of 0.1 and 0.5. The sequence of threshold $\mathfrak{F}_i, i = 1, ..., \mathfrak{n}_s$ was decided by using a data-driven approach suggested by Winker [2001] as a standard approach in deciding the sequence. The data-driven approach is briefly described as follows. First, a distance measure is defined as the absolute difference in the fitness values of a solution and a candidate solution from its neighborhood. The empirical distribution of the distance is then constructed on the basis of the distance values between a number of randomly chosen solutions and their neighborhood solutions. After that, the empirically observed distance measures are sorted in decreasing order, and the k-th quantile of the sorted distance is taken as the threshold value for the k-th step. The TA algorithm is described in Algorithm 2.

4 Experimental Results

In order to tell which evolutionary approach is more efficient and stable for solving the optimization problem, the fitness values, i.e. the log-likelihood values, from independent restarts of the DE and TA algorithms were compared with each other.

Algorithm 2 Threshold Accepting.
_____
 1: Initialize threshold sequence $\mathfrak{T}_i$, with $i = 1, ..., \mathfrak{n}_s$ and
 2: Randomly generate an initial solution $\dot{\mathbf{s}}^c$
 3: **for** $i = 1$ to $\mathfrak{n}_s$ **do**
 4:     Choose $\dot{\mathbf{s}}^n \in \mathfrak{N}(\dot{\mathbf{s}}^c)$
 5:     **if** $\mathfrak{O}(\dot{\mathbf{s}}^n) - \mathfrak{O}(\dot{\mathbf{s}}^c) < \mathfrak{T}_i$, **then**
 6:       $\dot{\mathbf{s}}^c \leftarrow \dot{\mathbf{s}}^n$
 7:     **end if**
 8: **end for**
_____

The two algorithms have been restarted 150 times. The fitness values i.e., the log-likelihood values defined in Eq. (8) from restarting the two algorithms, are provided in Figure 1. The figure shows that the DE yields higher and more stable fitness values than that from the TA in each run. The suboptimal performance of the TA is possible due to the simple way used for defining the local neighborhood structure. As Winker [2001] pointed out, the performance of TA highly depends on the construction of local neighborhood structure and the threshold sequence. DE is simpler than TA for implementation as the algorithm only requires a fine tune of the weighting factor K and the crossover probability $\mathfrak{z}_1$. Most of the time, the solutions from DE are insensitive to small changes of its parameter settings. Due to the above reasons, DE was employed as the EML optimizer in the following simulation study.

To assess the accuracy of the estimated parameters from the EML estimation against the IFM approach, a set of 200×2 random variables from the bivariate Student t distribution was generated at a total iteration number of $n_{\mathfrak{S}} = 5,000$. The true distribution parameters used for generating the random numbers were set as $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 0.2548$, $\sigma_2 = 0.2250$, $\rho = 0.43$ and $\nu = 6$. Since the standard hill-climbing algorithm such as the Newton-Raphson approach for the EML estimation did not generate any results but only for the IFM approach, the results from the IFM were compared with that from the DE procedure applied for the EML estimation. Table 1 shows the numerical results with the standard descriptive statistics of the estimated parameters based on the $n_{\mathfrak{S}}$ bootstrap samples. As expected, the EML estimators, obtained by maximizing the log-likelihood function with the DE, are often (i) closer to the true values, (ii) less biased, (iii) less skewed and (iv) less kurtotic as compared to the IFM alternatives.

Figure 2 compares the kernel densities of the estimated distribution parameters from EML and IFM. As discernible, the differences between the distribution of estimators for μ_1, μ_2 (top panels) and ρ (bottom left panel) are negligible. More interestingly, the middle panels reveal that the dispersion parameters σ_1 and σ_2 can be more accurately estimated with the EML approach as their kernel densities are higher in the centered region and lower in the tail regions. Finally, the bottom
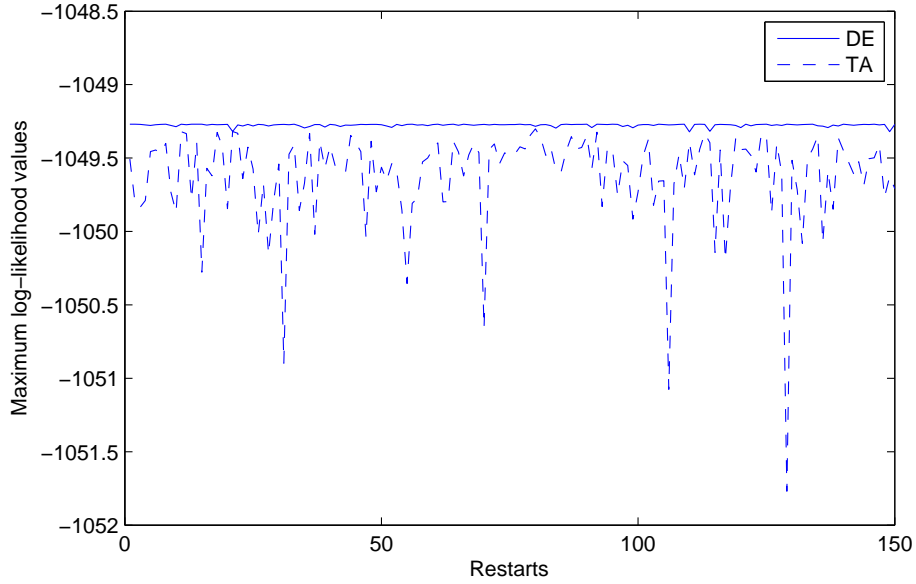
8

Figure 1: Log-likelihood Value from Independent Restarts

right panel shows that the estimate for the degree of freedom ν that controls the probability of tail events in the distribution is even biased from the IFM method as the peak of its kernel density is not localized at the true parameters position. Overall, it can be seen that the parameters responsible centered moments of the distribution, i.e. σ_1, σ_2 and ν can be better estimated with the EML approach. These results are indeed essential as they reveal that the IFM approach often preferred in the financial literature is more likely to provide less reliable estimators of the underlying joint distribution. Hence, the IFM approach based on the gradient search is less able to correctly capture the tail dependence of risk factors (e.g., the extreme losses) than the EML estimation which is powered by the DE algorithm.

9

Table 1: Comparison of Sample Moments of the Parameters for EML and IFM

		EML				
θ	True Value		Mean($\hat{\theta}$)	SD($\hat{\theta}$)	SK($\hat{\theta}$)	KU($\hat{\theta}$)
μ_1	0.0000		0.0000	0.0105	0.0390	2.9275
μ_2	0.0000		0.0000	0.0092	0.0051	3.0494
σ_1	0.2548		0.2551	0.0095	0.0080	3.0637
σ_2	0.2252		0.2255	0.0086	0.0863	3.0170
ρ	0.4300		0.4374	0.0330	-0.0950	3.0309
ν	6.0000		6.2589	1.1123	1.2754	6.2565

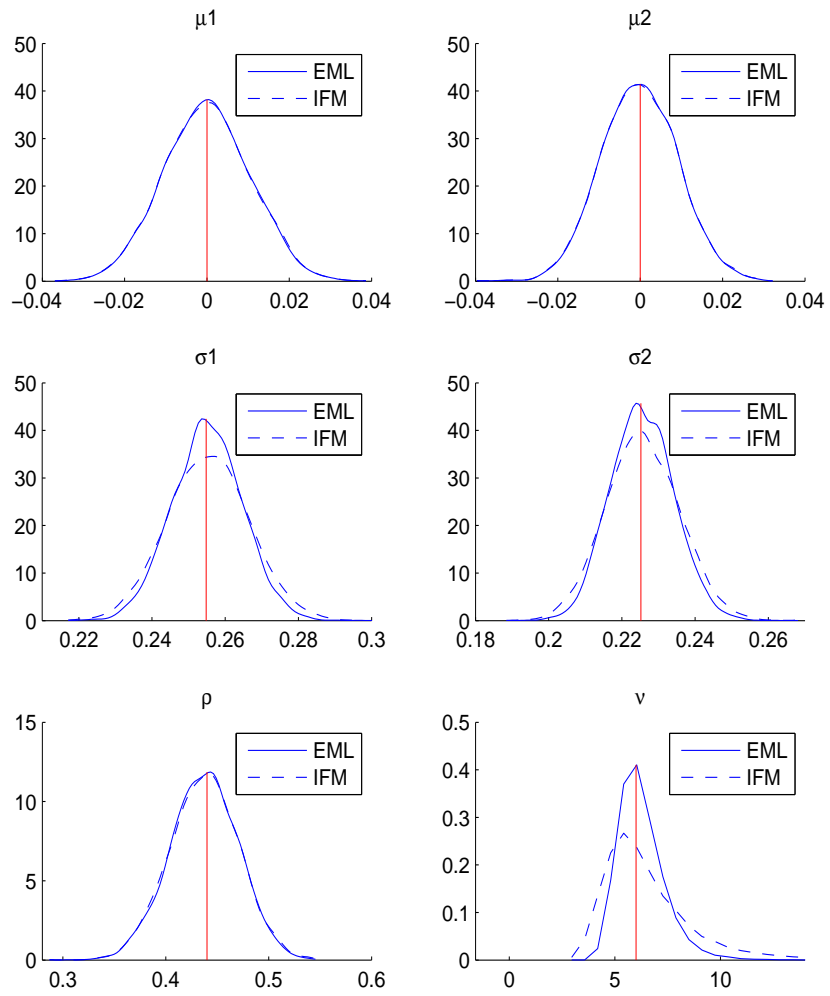|		IMF				
θ	True Value		Mean($\hat{\theta}$)	SD($\hat{\theta}$)	SK($\hat{\theta}$)	KU($\hat{\theta}$)
μ_1	0.0000		0.0000	0.0106	0.0397	2.9311
μ_2	0.0000		0.0000	0.0093	0.0103	3.0495
σ_1	0.2548		0.2554	0.0111	0.0453	3.0506
σ_2	0.2252		0.2257	0.0100	0.0966	3.0502
ρ	0.4300		0.4375	0.0336	-0.0620	3.0299
ν	6.0000		6.6798	2.7531	4.8531	59.7094

Figure 2: Kernel Densities of the Estimated Copula Parameters

5 Comments and Summary

This paper suggests implementing an evolutionary algorithm in the exact maximum likelihood estimation of a multivariate Student t copula model. Usually, the standard Newton-Raphson algorithm fails to solve complex optimization problems such as the EML estimation for the parameter inference of copula models, while a derivative-free optimizer can conquer such problems. Two evolutionary approaches, namely Differential Evolution and Threshold Accepting were employed to tackle the EML estimation problem, and it has been found that the former yielded a better performance. Through a simple simulation study, it has been proven that the proposed methodology for EML estimation already provided reasonably good results in a simple two-dimensional setting with a Student t copula model. As it is expected, the estimates obtained by the EML approach enhanced with Differential Evolution are often closer to the true values as compared to the IFM alternatives. Differential Evolution should be competent for the EML inference of more complicated copula models than the bivariate Student t copula studied in this paper.

References

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69:542–547.

Dueck, G. and Scheuer, T. (1990). Threshold Accepting: A general purpose optimization algorithm appearing superior to simulated annealing. *Journal of Computational Physics*, 90:161–175.

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

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Chapman & Hall, London.

McNeil, A. J., Frey, R., and Embrechts, P. (2005). *Quantitative Risk Management.* Princeton University Press, New Jersey.

Meucci, A. (2005). *Risk and Asset Allocation.* Springer, Berlin.

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

Price, K., Storn, R., and Lampinen, J. (1998). *Differential Evolution: A practical approach to global optimization.* Springer, Berlin.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de 1'institut de statistique de 1'Université de Paris*, 8:229–231.

Storn, R. and Price, K. (1997). Differential Evolution – A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359.

Winker, P. (2001). *Optimization Heuristics in Econometrics: Applications of threshold accepting.* John Wiley & Sons, West Sussex.

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