A Study on the Dependence Structure of Aggregate Loss Distributions Based on Frequency Dependence

Woohwan Kim* and Seungbeom Bang**

*The dependence structure among cell by cell in operational risk matrix is a critical issue in the determination of bank-wide operational risk capital under Advanced Measurement Approach (AMA). Especially, Loss Distribution Approach (LDA), which is one of the main methods in AMA, involves statistical inference on the frequency and severity distribution, and then loss distribution is generated via Monte-Carlo simulation. This procedure is conducted based on a specific cell in the operational risk matrix whose dimension is 7 * 8 in AMA. To generate bank-wide operational loss distribution, dependence structure among cell by cell loss distribution has to be modeled in a consistent manner. This paper proposes a methodology for generating firm-wide loss distribution based on frequency dependence, which is a primary source of the dependence structure. It especially highlights how to model the frequency dependence of loss distribution via copula, which is a function that links marginal and joint distributions. The main contribution of this paper is that it provides a very flexible method for modeling the dependence of individual loss distribution, and it is confirmed that the proposed method is satisfied in terms of conservatism. The study finds that bank-wide capital increases at about 15% as the dependence becomes stronger, and this empirical result is consistent with changes in the severity distribution of marginal loss distributions.*

Introduction

Operational risk is defined as the risk of loss resulting from inadequate or failed internal processes, people or systems, or from external events. The New Basel Capital Accord (BCBS, 2004) emphasizes the quantification and management of operational risk because of its pervasive nature. The proposed methods for operational risk quantification are: Basic Indicator Approach, Standardized Approach and Advanced Measurement Approach (AMA),[1] which includes Loss Distribution Approach (LDA) and Scenario-Based Approach[2]. Among these methods, LDA measures operational risk capital based on bank's own historical loss data of least five years. LDA involves two types of distribution fitting and analysis, which are severity and frequency distribution, and then loss distribution is generated employing mathematical

* Research Professor, Department of Economics, Yonsei University, Seoul, Korea. E-mail: jumnjump@yonsei.ac.kr

** Team Manager, Consulting Division, F1 Consulting Korea Co. Ltd., Seoul 100-170, Korea. E-mail: sbbang@f1ckorea.com

[1] For more details on the operational risk capital, see BCBS (2004).

[2] Scenario-Based Approach uses managerial judgment, supplemented by data analysis, to identify and quantify a wide range of potential operational losses for material risk based on consideration of potential causes and effects.

convolution, usually conducted by Monte-Carlo simulation in practice. There is a 7 * 8 operational loss data matrix for LDA, where 7 denotes types of loss events and 8 denotes business lines. The atomic unit for the calculation of operational risk capital is a specific cell of the 7 * 8 matrix. To quantify bank-wide operational risk capital, we need to model the dependence among cell by cell loss distributions employing statistical methodology.[3] For this purpose, we should consider the source of dependence of bank-wide aggregate loss distribution. In this paper, we focus on the dependence among loss frequencies of the cells of the 7 * 8 matrix and propose how to generate bank-wide loss distribution considering the frequency dependence.

Bank-wide operational loss distribution is believed to have strong dependence on the cell by cell loss distribution. However, there is no proper method for modeling dependence, so the bank simply sums up the cells' VaR to calculate the bank-wide operational capital. In this paper, we propose a conservative method for dependence modeling for the determination of bank-wide operational risk capital. The main idea of the proposed method is to model the dependence of cell by cell frequency distributions via copula. Copula is a function that joins or couples multivariate distribution function and marginal distribution function. We can separate marginal and joint distributions using copula function with appropriate dependence structure. It makes copula very useful to model multivariate distribution in practical views. Especially, with a copula, it is possible to capture the comovements among random variables, which is the well-known feature of risk management. However, the numerous studies on copula have focused on continuous marginal distributions. In this paper, we provide the modeling scheme of copula among frequency distributions, since we believe that dependence among loss frequencies is the primary source of aggregate loss distribution. Our main argument is that this is particularly true for high severity events, as severity-independence likely dominates the frequency correlation. Denuit and Lambert (2005) propose a continuousation method to transform the discrete random variable to a continuous one. The important thing is that this continuousation does not alter the concordance between the pairs of random variables. We employ this continuousation to model correlated Poisson distribution and then we compare our proposed method with the existing method.

The paper briefly explains the basic concept of the dependence of aggregate loss distribution and introduces a correlated Poisson distribution in general dimension. Subsequently, it provides a copula with discrete marginal distributions, especially focusing on continuousation and sampling methods for correlated Poisson distribution. Then, it summarizes the empirical results and simulation-based analysis, and thus concludes.

Aggregate Loss Distribution

The aggregate loss L is defined as a random sum of individual losses occurring in a given time horizon, typically, one year.

[3] The Basel Committee has not clearly provided how to model the dependence structure among cell by cell loss distribution. The proposed principle for dependence modeling is summarized as the statistical soundness and conservativeness, especially in case we do not have sufficient data for dependence modeling.

$$L = \sum_{i=1}^{N} X_i \qquad \text{...(1)}$$

where X_i denotes individual loss amount, which is also known as severity, and N is annual number of loss events. To generate aggregate loss distribution, it is needed to conduct mathematical convolution of frequency and severity distributions such as Monte-Carlo simulation as well as fast fourier transformation. In this paper, we use Monte-Carlo simulation to generate aggregate loss distribution.

The regulatory capital requirement K is the 99.9% of aggregate loss distribution G, i.e.,

$$K = G^{-1}(0.999) \qquad \text{...(2)}$$

Let L_1 and L_2 denote the two aggregate losses, and L is the firm-wide aggregate loss which is expressed as the sum of the two aggregate losses, i.e.,

$$L = L_1 + L_2 = \sum_{i=1}^{N_1} X_i + \sum_{i=1}^{N_2} Y_i \qquad \text{...(3)}$$

Under the classical assumption,[4] the covariance of L_1 and L_2 are given as:

$$\begin{aligned} Cov(L_1, L_2) &= E(L_1 L_2) - E(L_1)E(L_2) \\ &= E\left(\sum_{i=1}^{N_1} X_i \sum_{i=1}^{N_2} Y_i\right) - E\left(\sum_{i=1}^{N_1} X_i\right) E\left(\sum_{i=1}^{N_2} Y_i\right) \\ &= [E(N_1 N_2) - E(N_1)E(N_2)]E(X)E(Y) \end{aligned} \qquad \text{...(4)}$$

Additionally, the correlation of L_1 and L_2 is calculated as:

$$corr(L_1, L_2) = corr(N_1, N_2)\frac{E(X)}{\sqrt{E(X^2)}}\frac{E(Y)}{\sqrt{E(Y^2)}} \qquad \text{...(5)}$$

Since X is a positive random variable, we conclude that aggregate loss correlation is always lower than the frequency correlation, that is,

$$0 \le corr(L_1, L_2) \le corr(N_1, N_2) \le 1 \qquad \text{...(6)}$$

The above formula intimates the upper limit of the firm-wide aggregate loss distribution. Especially, financial institutions should make conservative risk estimates in the absence of sufficient data to model the dependence structure. Additionally, the dependence among loss frequency is the primary source of dependence of firm-wide loss distribution. We introduce the method for generating correlated aggregate distribution with modeling frequency dependence via copula in the next section.

[4] Classical assumption via Monte-Carlo simulation for generating an aggregate loss distribution is characterized by the independence of frequency and individual loss, also called severity.

Copula with Discrete Marginal Distribution

The most prominent method to model multivariate distribution is to introduce a copula which connects marginal distribution and joint distribution in a consistent manner. Employing copula theory, we can describe the joint distribution of asset returns with proper choice of copula.

Copula (Sklar, 1959): A d-dimensional copula is a function C: $[0, 1]^d \mapsto [0, 1]$ which has the following properties:

- $C(u)$ is increasing in each component u_k with $k \in \{1, ..., d\}$.
- For every vector $u \in [0, 1]^d$, $C(u) = 0$, if at least one coordinate of the vector u is 0 and $C(u) = u_k$, if all the coordinates of u are equal to 1 except the k^{th} one.
- For every $a, b \in [0, 1]^d$ with $a \leq b$, given a hypercube.

  $B = [a, b] = [a_1, b_1] \times [a_2 \times b_2] \times ... \times [a_n \times b_n]$ whose vertices lie in the domain of C, its volume $V_C(B) \geq 0$.

This definition states that copula is a multivariate uniform distribution function with uniformly distributed margins.

The following theorem is one of the most important theorems, that makes the copula used to construct the multivariate distribution. A proof of Sklar's theorem can be found in Nelsen (1999).

Sklar's Theorem (1959): *Let H be a d-dimensional distribution function with margins F_1, ... F_d. Then there exists a d-dimensional copula C such that, for $x \in R^d$ we have,*

$H(x_1, ... x_d) = C\ (F_1(x_1), ..., F_d(x_d))$

Moreover, if F_1, ..., F_d are continuous, then C is unique.[5]

Sklar's theorem shows that joint distribution can be separated into marginal distribution and dependence structure. This means that the basic idea of dependence modeling via copula is that the univariate margins and the dependence structure can be separated, with the latter completely described by a copula function. For more details on copula, see Joe (1997) and Nelsen (1999).

Another important property of copula is that it is invariant under strictly increasing transformations, i.e.,

$$C_{X_1,...,X_d} = C_{h_1(X_1),...,h_d(X_d)}, \text{ if } \partial_x h_n(x) > 0 \quad ...(7)$$

The dependence structure between random variables does not change under increasing and continuous transformations of the margins. This invariance property implies that the copulas are the natural framework used to study dependence properties which are invariant under increasing transformation of the margins.

[5] In discrete case, C is uniquely defined on Range F_X x Range F_Y.

One more advantage of using copula to model joint density is that with copula it is possible to model comovements among variables with tail dependence index. Generally, correlation is considered as the measure of dependence between two random variables; however, it only captures the linear relationship, i.e., it does not describe fully the dependence structure between the variables. To overcome this problem, two additional dependence measures are commonly used in empirical analysis. One is Spearman's rank correlation and the other is Kendall's tau. Both measures are expressed as functions of copula, which make them more attractive in practical application. Kendall's tau and the Spearman's rho are defined, respectively, as follows:

$$\tau = 4\iint_{I^2} C(u_1, u_2) dC(u_1, u_2) - 1 \qquad ...(8)$$

$$\rho_S = 12\iint_{I^2} u_1 . u_2 \, dC(u_1, u_2) - 3 \qquad ...(9)$$

There are broadly two classes of copula: one is named as elliptical copula and the other is Archimedean copula. Elliptical copula is derived from multivariate distribution function. Gaussian copula and *t* copula are included in this class. Archimedean copula is derived via generator and includes various copulas such as Gumbel, Frank and Clayton. This class is very useful since it can be characterized by a simple generator function and it can model extreme comovements in tail regions.

Gaussian Copula:

$$C^G(u_1, ..., u_d, \Sigma) = \Phi_\Sigma^d\left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\right)$$

$$= \int_{-\infty}^{\Phi^{-1}(u)} \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2} y^T \Sigma^{-1} y\right\} dy \qquad ...(10)$$

where $\Phi^{-1}(u) \equiv \left(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_d)\right)$ and $|\Sigma|$ denote matrix determinant and T is the transpose operator. The integration represented by the above formula is component-wise integration.

The parameter of Gaussian copula is Σ, the usual linear correlation matrix which characterizes the dependence of random variables.

***t* Copula:**

$$C(u_1, ..., u_d; \nu, \Sigma) = t_{\nu,\Sigma}^d\left(t_\nu^{-1}(u_1), ..., t_\nu^{-1}(u_d)\right)$$

$$= \int_{-\infty}^{t_\nu^{-1}(u)} \frac{\Gamma((\nu + d)/2)}{\Gamma(\nu/2)(\nu\pi)^{d/2}|\Sigma|^{1/2}} \left(1 + y^T \Sigma^{-1} y/\nu\right)^{-(\nu+d)/2} dy \qquad ...(11)$$

where ν denotes the degree of freedom.

The parameters of t copula are linear correlation with the same role as Gaussian copula and the degree of freedom which characterizes the comovement of random variables. A lower degree of freedom corresponds to strong comovement in the tail area.

Archimedean copula is generally written in the following form:

$$C(u,v)=\phi^{-1}[\phi(u)+\phi(v)] \quad ...(12)$$

where ϕ is a generator of Archimedean copula and ϕ^{-1} is the inverse function of ϕ. The generator is a function $\phi : I \rightarrow R^{+}$ that is continuous, strictly decreasing, convex and satisfies $\phi(1) = 0$.

Gumbel Copula: Gumbel (Gumbel, 1960) copula with parameter δ has the following form:

$$C(u,v)=\exp\left\{-\left[\left(-\ln u^{\theta}+\left(-\ln v^{\theta}\right)\right)\right]^{1/\theta}\right\},\theta \geq 1 \quad ...(13)$$

The generator of Gumbel copula is $\phi(t)=(-\ln t)^{\theta}$ and the parameter θ of Gumbel copula controls the strength of dependence; when $\theta = 1$, there is no dependence and when θ goes to infinity, there is perfect dependence.

To implement copula for empirical analysis, it is needed to estimate copula parameters and select the best copula according to the given data. Estimation is conducted via maximization of likelihood function given a specific copula. Copula selection is conducted to minimize the sum square deviance between empirical copula and theoretical copula. Goodness of fit test and likelihood-based selection are also employed to choose the best fitted copula (Nelsen, 1999).

Copula with discrete margin is a developing area, compared to copula with continuous margin. In this paper, we show how to transform a discrete random variable into a continuous one, maintaining concordance among the variables. The probability integral transformation states that if the random variable Y is a continuous variable with cumulative distribution F, then:

$$Z = F(Y) \quad ...(14)$$

is uniformly distributed on [0, 1]. However, the probability integral transformation with discrete random variable does not yield satisfactory result, which implies that Z is not uniformly distributed on [0, 1] in discrete case.

Denuit and Lambert (2005) proposed a continuousation method to overcome this problem. The important thing is that this continuousation does not alter the concordance between the pairs of random variables. The main idea of continuousation is to create a new random variable X^{*}, by adding to a discrete variable X a continuous variable U valued in [0; 1], independent of Y, with a strictly increasing cdf, sharing no parameter with the distribution of X, such as the Uniform [0, 1], i.e.,

$$X^{*} = X + (U - 1) \quad ...(15)$$

If X is a discrete random variable with probability mass function $f_x = P(X = x)$, and the probability density function of continuouused variable Z^*, given as

$$Z^* = F^*\left(X^*\right) = F^*\left(X + (U-1)\right) = F\left(\left[X^*\right]\right) + f_{\left[X^*\right]+1} U = F(X-1) + f_x U \qquad \text{...(16)}$$

is uniformly distributed on [0; 1], then [X] denotes the integer part of X. In the case of discrete data, we use Z^* as an argument in the copula, if the marginal model is well-specified. This will ensure that the conditions for use of a copula are met.

Suppose Y is another discrete random variable valued and let $Y^* = Y + V$ denote its continuous version. We assume that the random vector (U, V) is independent of (X, Y). It is then easily seen that the joint distribution function of (X^*, Y^*) can be written in terms of the copula C of (U, V) as:

$$H^*(x,y) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} C(x-i, y-j)\, p_{ij},\ (x,y) \in R \times R \qquad \text{...(17)}$$

where p_{ij} denotes the joint probability function of (X, Y).

Empirical Analysis

In this paper, our primary concern is to quantify the impact of frequency dependence on the bank-wide operational loss distribution. For this purpose, we conduct a simulation-based analysis with various frequency dependence structures under lognormal severity and Poisson frequency. The frequency of each cell is set at Poisson (20), and severity represented by the lognormal distribution is set at (10, 1.5). We employ Gaussian, *t*,[6] and Gumbel copula for frequency dependence modeling with various Kendall's tau. Table 1 shows the firm-wide VaR with 99.9% confidence level, which is obtained by taking the average of 100 replications, with the number of simulation for generating loss distribution being 100,000.

From Table 1, we confirm that bank-wide operational capital cumulatively increases at about 12%, as the frequency dependence gets stronger from 0.05 (almost independent case) to 0.95 (almost perfect dependent case). Specifically, Gaussian copula increases at 13.46%, *t* copula at 12.07% and Gumbel copula at 11.51% (Figure 1).

To investigate the impact of severity distribution on bank-wide operational capital charge, we conduct the same simulation with Gamma distribution (0.52, 0.000015) which is almost similar to the lognormal (10, 1). From Table 2, we confirm that bank-wide operational capital cumulatively increases at about 15%, as the frequency dependence gets stronger from 0.05 (almost independent case) to 0.95 (almost perfect dependent case). Specifically, Gaussian copula increases at 15.95%, *t* copula at 16.27% and Gumbel copula at 14.89% (Figure 2).

We may conclude that the bank-wide operational capital is not affected by the changing marginal severity distributions as long as their shapes are similar.

[6] For empirical analysis, the degree of freedom of *t* copula is set to 5 in this paper.

Table 1: The Result of Firm-Wide VaR with Lognormal (10, 1.5) Severity Distribution			
Frequency Dependence	Gaussian Copula	*t* Copula	Gumbel Copula
τ = 0.05	5,054,447	5,127,485	5,215,532
τ = 0.10	5,155,918	5,187,423	5,334,985
τ = 0.20	5,289,368	5,317,207	5,477,000
τ = 0.30	5,343,914	5,436,169	5,608,038
τ = 0.40	5,442,777	5,530,121	5,658,734
τ = 0.50	5,559,256	5,626,612	5,704,291
τ = 0.60	5,672,085	5,689,834	5,748,338
τ = 0.70	5,691,586	5,713,632	5,765,708
τ = 0.80	5,759,981	5,719,665	5,785,652
τ = 0.90	5,764,607	5,777,965	5,792,574
τ = 0.95	5,775,398	5,779,191	5,845,395

Figure 1: The Cumulative Increasing Rate of Bank-Wide Risk Capital with Frequency Dependence

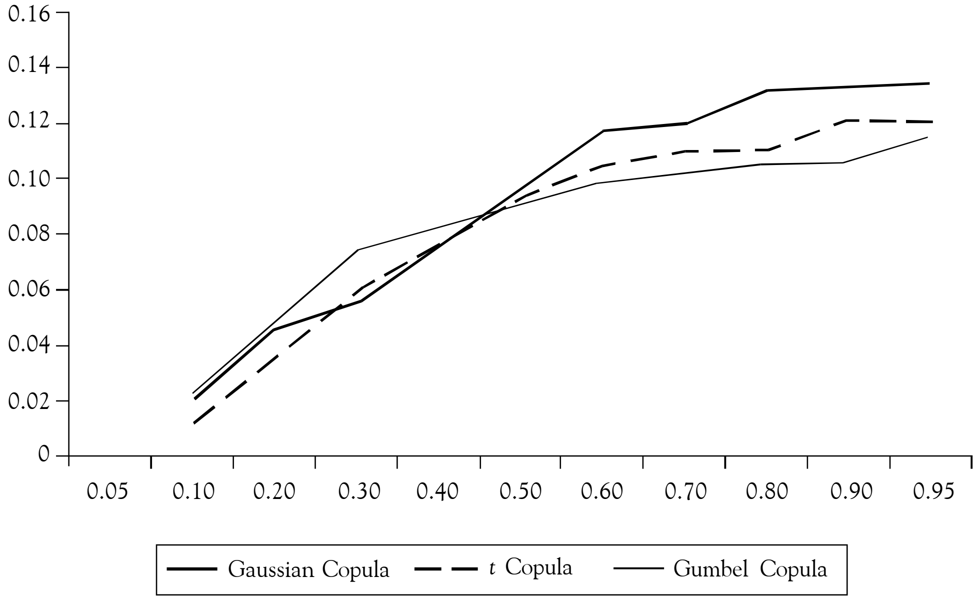
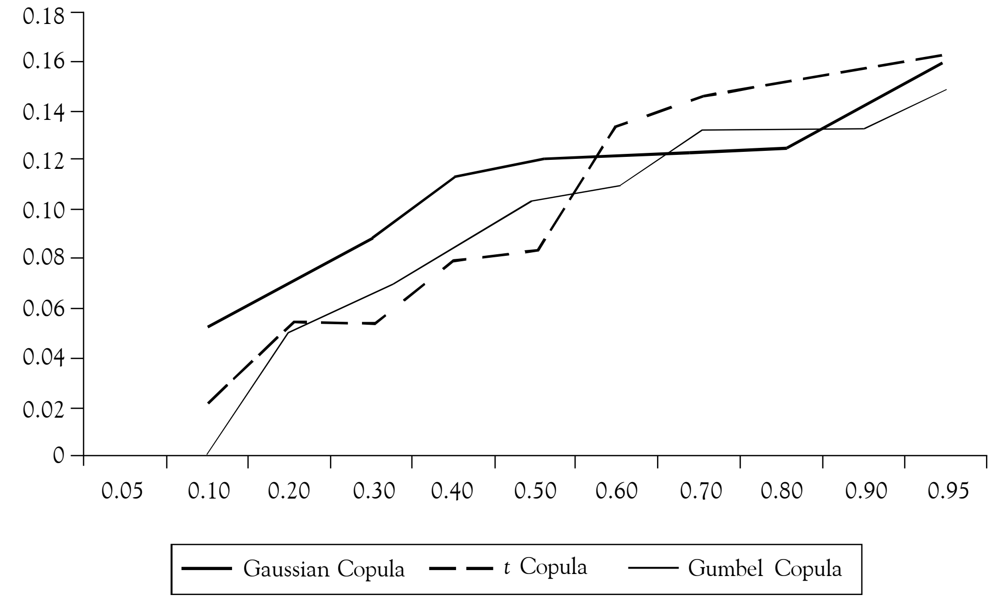


Table 2: The Result of Firm-Wide VaR with Poisson (20) and Gamma Distribution (0.52, 0.000015)

Frequency Dependence	Gaussian Copula	*t* Copula	Gumbel Copula
$\tau = 0.05$	4,770,476	4,734,298	4,820,625
$\tau = 0.10$	5,019,228	4,828,523	4,823,517
$\tau = 0.20$	5,117,236	4,989,281	5,069,971
$\tau = 0.30$	5,206,093	4,990,717	5,135,962
$\tau = 0.40$	5,324,967	5,113,906	5,230,200
$\tau = 0.50$	5,369,653	5,140,668	5,335,192
$\tau = 0.60$	5,384,522	5,395,701	5,364,871
$\tau = 0.70$	5,385,833	5,465,099	5,482,064
$\tau = 0.80$	5,394,635	5,492,186	5,490,806
$\tau = 0.90$	5,467,554	5,523,156	5,492,867
$\tau = 0.95$	5,582,696	5,557,665	5,582,727

Figure 2: The Cumulative Increasing Rate of Bank-Wide Risk Capital with Frequency Dependence



Conclusion

This paper shows how to model frequency dependence in aggregate loss distribution and presents empirical analysis with various dependence structures employing copula. The important conclusion is that the strength of frequency dependence does not show strong impact on firm-wide operational capital charge. It also infers that bank-wide operational capital increases at about 15% as the frequency dependence becomes stronger. Additionally, it reveals that bank-wide operational capital is not affected by the changing marginal severity distributions as long as their shapes are similar. ❖

Bibliography

1. Andreas Heinen and Erick Rengifo (2008), "Multivariate Reduced Rank Regression in Non-Gaussian Contexts, Using Copulas", *Computational Statistics & Data Analysis*, Vol. 52, No. 6, pp. 2931-2944.
2. Basel Committee on Banking Supervision (BCBS) (2004), International Convergence of Capital Measurement and Capital Standards: A Revised Framework, June.
3. Denuit M and Lambert P (2005), "Constraints on Concordance Measures in Bivariate Discrete Data", *Journal of Multivariate Analysis,* Vol. 93, pp. 40-57.
4. Embrechts P, McNeil A J and Straumann D (2001), *Correlation and Dependency in Risk Management: Properties and Pitfalls in Value at Risk and Beyond,* Cambridge University Press.
5. Genest C and Rivest L (1993), "Statistical Inference Procedures for Bivariate Archimedean Copulae", *Journal of the American Statistical Association,* Vol. 88, No. 423, pp. 1034-1043.
6. Genest C and Rivest L P (2001), "On the Multivariate Probability Integral Transformation", *Statistics and Probability Letters*, Vol. 53, No. 4, pp. 391-399.
7. Gumbel E J (1960), "Bivariate Exponential Distributions", *Journal of the American Statistical Association*, Vol. 55, pp. 698-707.
8. Joe H (1997), *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.
9. McNeil A J, Frey R and Embrechts P (2006), *Quantitative Risk Management: Concepts, Techniques and Tools*, Princeton University Press, NJ.
10. Mhamed Mesfioui and Abdelouahid Tajar (2005), "On the Properties of Some Nonparametric Concordance Measures in the Discrete Case", *Journal of Nonparametric Statistics,* Vol. 17, No. 5, pp. 541-554.
11. Nelsen R B (1999), *An Introduction to Copulas* (Lecture Notes in Statistics No. 139), Springer, New York.
12. Sklar A (1959), *Fonctions de Repartitions a n dimensions et Leurs Marges*, Publ. Inst. Statist. Univ. Paris, Vol. 8, pp. 229-231.

Reference # 37J-2009-09/12-05-01