Copulas

Statistical Methods in Finance

Copula Methods (See Chapter 8)

- **Copula**

  Financial data often display a great deal of dependency. Classical statistical methods for multivariate analysis typically assume joint normality. Such an assumption could be too restrictive as financial data are typically "non-normal". Copula methods provide flexible and general ways to generate multivariate distributions with specified (univariate) marginal distributions. Because of the marginal specification, the joint distribution can be transformed to the corresponding one on the unit cube $[0, 1]^n$ via the (specified) marginal distribution functions. An n-variate copula is a multivariate distribution on the unit cube $[0, 1]^n$, with uniform(0, 1) marginals. To fix ideas, we start with bivariate copulas.

Bivariate Copula: Definition

Definition 1 A bivariate function G on $[0,1] \times [0,1]$ is said to be grounded if $G(u,0) = G(0,v) = 0$ for all u, v in $[0,1]$.

Definition 2 A bivariate function G is said to be 2-increasing if, for all $u_1 \leq u_2$, $v_1 \leq v_2$,

$$G(u_2, v_2) - G(u_1, v_2) - G(u_2, v_1) + G(u_1, v_1) \geq 0.$$

When G is twice differentiable, 2-increasing is equivalent to

$$\frac{\partial^2 G(u,v)}{\partial u \partial v} \geq 0.$$

Bivariate Copula: Definition

Definition 3 A bivariate copula C is a function on $[0,1] \times [0,1]$ satisfying

(a) C is grounded;

(b) C is 2-increasing;

(c) $C(u,1) = u$ and $C(1,v) = v$ for all u, v in $[0,1]$.

Bivariate Copula

Corollary 1 A bivariate function is a copula if and only if it is a distribution function on $[0, 1] \times [0, 1]$ with uniform marginals.

Corollary 2 The following three functions are copulas:

(a) $C^-(u, v) = \max(u + v - 1, 0)$

(b) $C^+(u, v) = \min(u, v)$

(c) $C^\perp = uv$

Definition 4 We say that C_1 is smaller than C_2, denoted by $C_1 \prec C_2$, if

$$C_1(u, v) \leq C_2(u, v), \quad \text{for all } u, v.$$

Corollary 3 For any copula C, we have $C^- \prec C \prec C^+$.

Sklar Theorem

Theorem (Sklar, 1959) Let F_1 and F_2 be two univariate distribution functions. Then the following results hold.
(a) If C is a copula, then $C(F_1(x), F_2(y))$ is a bivariate distribution function.
(b) If $F(x, y)$ is a bivariate distribution function with marginals F_1 and F_2, then there exists a copula C such that $F(x, y) = C(F_1(x), F_2(y))$.

Note that in (b), the copula can be expressed as

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)). \tag{1}$$

This expression also allows us to simulate bivariate random variables with C as their joint distribution as long as we can simulate from F. We first simulate (X, Y) from $F(x, y)$ and then transform them to $(U, V) = (F_1(X), F_2(Y))$, which follows C.

Parametric Family: Gaussian

Gaussian Copulas For a bivariate normal (Gaussian) distribution with marginals $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ and correlation coefficient ρ, we can use equation (1) to obtain a copula C. Because of the standardization (to uniform (0,1)) on the marginals, the resulting copula C does not depend on the means and variances of the marginals but only the correlation ρ. Using Φ for standard normal and Φ_ρ for bivariate normal with standard normal as the marginals and correlation coefficient ρ, we can write

$$C_\rho^G(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)). \tag{2}$$

We call $\{C_\rho^G : \rho \in [-1, 1]\}$ the family of Gaussian copulas.

Parametric Families

Archimedean Copulas Generator g is a strictly decreasing function on $[0, 1]$ such that $g(1) = 0$. Given generator g, the corresponding Archimedean copula is

$$C_g^A(u, v) = g^{-1}(g(u) + g(v)).$$

When $g(0) < \infty$, we define $g^{-1}(x) = 0$ for $x \geq g(0)$.

1. Gumbel (1960): $g_\alpha(u) = (-\log u)^\alpha$, $\alpha \in [1, \infty)$. In this case,

$$C_\alpha(u, v) = \exp\{-[(-\log u)^\alpha + (-\log v)^\alpha]^{\frac{1}{\alpha}}\}.$$

For $\alpha = 1$, $g_\alpha(u) = -\log u$ and, therefore, $C(u, v) = uv = C^\perp(u, v)$, the independent case. On the other hand, as α approaches ∞, $C(u, v)$ approaches $\min(u, v) = C^+(u, v)$.

2. Clayton (1978): $g_\alpha(u) = \alpha^{-1}(u^{-\alpha} - 1)$, $\alpha \in [-1, \infty)$. Note that for $\alpha = 0$, it should be understood as $-\log u$. The corresponding copula

$$C_\alpha(u, v) = \frac{1}{[\max(u^{-\alpha} + v^{-\alpha} - 1, 0)]^{\frac{1}{\alpha}}}.$$

3. Frank (1979): $g_\alpha(u) = -\log[\frac{\exp(-\alpha u)-1}{\exp(-\alpha)-1}]$, $\alpha \in (-\infty, \infty)$. Again for $\alpha = 0$, it should be understood as $-\log u$.

First-to-default (FTD) Swap

Two obligors, with default times τ_1 and τ_2. Pay 1 unit of currency if at least one defaults before T. Assume r the interest rate and present time 0. Then the present value is

$$e^{-rT} I(\tau_1 \leq T \text{ or } \tau_2 \leq T)$$

Price (FTD) should be

$$\begin{aligned} & E[e^{-rT} I(\tau_1 \leq T \text{ or } \tau_2 \leq T)] \\ =& e^{-rT}[P(\tau_1 \leq T) + P(\tau_2 \leq T) - P(\tau_1 \leq T, \tau_2 \leq T)] \end{aligned}$$

First-to-default Swap

Suppose that $\tau_1 \sim F_1$ and $\tau_2 \sim F_2$ with copula C. Then

$$FTD = e^{-rT}[F_1(T) + F_2(T) - C(F_1(T), F_2(T))]$$

(1) $FTD \geq e^{-rT} \max F_1(T), F_2(T)$.

(2) When τ_1 and τ_2 are independent,

$$FTD = e^{-rT}[F_1(T) + F_2(T) - F_1(T)F_2(T)]$$

First-to-default Swap

Example. We have a basket of two obligors, both of which have credit rating of BBB, $r = 4\%$ and $T = 5($ years$)$. Thus, $F_1(T) = F_2(T) = 4.7\%$. Pay \$1 million in case at least one defaults.

$$FTD = 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - C(0.047, 0.047)]$$

(1) $FTD \geq 1000000 \times e^{-0.04\times 5} \times 0.047 = 38562$

(2) When two default times are independent,

$$\begin{aligned} FTD =& 1000000 \times e^{-0.04\times 5} \times (0.047 + 0.047 - 0.047 \times 0.047) \\ =& 75152 \end{aligned}$$

First-to-default Swap

Suppose C is a Gaussian copula with $\rho = 0.3$ and we want to compute $C(u, v)$ in R:

```
library(mvtnorm)

# C(u,v)
pmvnorm(lower = c(-Inf,-Inf), upper = c(qnorm(u),qnorm(v)),
sigma = matrix( c(1,0.3,0.3,1),2,2))
```

Example.

(3) When C is gaussian copula with $\rho = 0.2$,

$$\begin{aligned} FTD =& 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - C(0.047, 0.047)] \\ =& 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - 0.004719249] \\ =& 73096.9 \end{aligned}$$

(4) When C is gaussian copula with $\rho = 0.5$,

$$\begin{aligned} FTD =& 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - C(0.047, 0.047)] \\ =& 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - 0.01117551] \\ =& 67810.96 \end{aligned}$$

(5) When C is gaussian copula with $\rho = 0.9$,

$$\begin{aligned} FTD =& 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - C(0.047, 0.047)] \\ =& 1000000 \times e^{-0.04\times 5}[0.047 + 0.047 - 0.02975759] \\ =& 52597.24 \end{aligned}$$

Second-to-default Swap

Again we have two obligors with default times τ_1 and τ_2. But the triggering event for paying 1 unit of currency is the second default, i.e. if both default before T. Thus the expected present value becomes

$$E[e^{-rT} I(\tau_1 \leq T \text{ and } \tau_2 \leq T)] = e^{-rT} P(\tau_1 \leq T, \tau_2 \leq T)$$

Suppose that $\tau_1 \sim F_1$ and $\tau_2 \sim F_2$ with copula C. Then it becomes

$$C(F_1(T), F_2(T)).$$

Second-to-default Swap

Example. Now the one million is paid when both default. We have the following results under various scenarios.

(1) When C is gaussian copula with $\rho = 0.2$,

$$\begin{aligned} LTD =& 1000000 \times e^{-0.04\times 5} \times C(0.047, 0.047) \\ =& 1000000 \times e^{-0.04\times 5} \times 0.004719249 \\ =& 3863.794 \end{aligned}$$

(2) When C is gaussian copula with $\rho = 0.5$,

$$\begin{aligned} LTD =& 1000000 \times e^{-0.04\times 5} \times C(0.047, 0.047) \\ =& 1000000 \times e^{-0.04\times 5} \times 0.01117551 \\ =& 9149.736 \end{aligned}$$

(3) When C is gaussian copula with $\rho = 0.9$,

$$\begin{aligned} LTD =& 1000000 \times e^{-0.04\times 5} \times C(0.047, 0.047) \\ =& 1000000 \times e^{-0.04\times 5} \times 0.02975759 \\ =& 24363.45 \end{aligned}$$

Canonical Maximum Likelihood Estimation

The canonical maximum likelihood (CML) is a useful method for the estimation of copula parameter(s). Let C_θ denote the copula with parameter θ and c_θ the corresponding density. Suppose we have observations (X_t, Y_t) $(t = 1, \ldots, T)$. Let $\hat{F}_k$ $(k = 1, 2)$ be the (marginal) empirical distributions of X and Y, i.e.

$$\hat{F}_1(x) = \frac{1}{n} \sum_{t=1}^{T} I(X_t \leq x)$$

$$\hat{F}_2(y) = \frac{1}{n} \sum_{t=1}^{T} I(Y_t \leq y)$$

Parameter Estimation

We approximate the likelihood function

$$L(\theta) = \prod_{t=1}^{T} c_\theta(F_1(X_t), F_2(Y_t))$$

by

$$\hat{L}(\theta) = \prod_{t=1}^{T} c_\theta(\hat{F}_1(X_t), \hat{F}_2(Y_t)).$$

The CML estimator $\hat{\theta}$ is obtained as the maximizer of $\hat{L}(\theta)$.

Relationship with Rank Correlation

Rank correlation: nonparametric approach to association There are two important quantities in nonparametric statistics that measure the strength of bivariate association or rank correlation.

Kendall's τ Let (X, Y) be a pair of random variables and (X_t, Y_t) $(t = 1, \ldots, T)$ be its i.i.d. copies. The Kendall's τ is defined to be

$$\tau = E\left\{\mathrm{sign}[(X_t - X_s)(Y_t - Y_s)]\right\}.$$

Note that if we denote the marginal distributions of X and Y by F and G, then

$$\tau = E\left\{\mathrm{sign}[(F(X_t) - F(X_s))(G(Y_t) - G(Y_s))]\right\}.$$

Thus, Kendall's τ depends only on the copula, not the marginal distributions.

Relationship with Rank Correlation

Spearman's ρ Let (X, Y) be a pair of random variables with marginal distributions F and G. The Spearman's ρ is defined to be

$$\rho = \mathrm{Corr}[F(X), G(Y)].$$

Note that $F(X)$ and $G(Y)$ both have uniform distributions. Let C be the copula for (X, Y). By definition, C is the joint distribution of $F(X)$ and $G(Y)$. So ρ is also determined by the copula.

Multivariate Case Suppose we have random variables $X_1, \ldots, X_n$ with joint distribution F and marginal distributions $F_1, \ldots, F_n$. The associated multivariate copula C satisfies

$$F(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)).$$

Multivariate Gaussian Copula

Gaussian Copulas $C_M^G(u_1, \ldots, u_n) = \Phi_M(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n))$, where M is an $n \times n$ correlation coefficient matrix.

We may apply the CML as follow. Suppose we have data from T periods: $(X_{1t}, \ldots, X_{nt})$, $t = 1, \ldots, T$. We can use data X_{it}, $t = 1, \ldots, T$ to compute $\hat{F}_i$. Next we compute $\tilde{X}_{it} = \Phi^{-1}(\hat{F}_i(X_{it})$. The CML estimator of M is them the sample correlation from $(\tilde{X}_{1t} \ldots \tilde{X}_{nt})$, $t = 1, \ldots, T$.

Multivariate Copula: Parametric Families

Archimedean Copulas Generator g is a decreasing function on [0, 1] such that $g(0) = +\infty$ and $g(1) = 0$. Given generator g, the corresponding Archimedean copula is

$$C_g^A(u_1, \ldots, u_n) = g^{-1}(g(u_1) + \cdots + g(u_n)).$$

Gumbel (1960): $g_\alpha(u) = (-\log u)^\alpha$, $\alpha \in [1, \infty)$.

Clayton (1978): $g_\alpha(u) = \alpha^{-1}(u^{-\alpha} - 1)$.

Frank (1979): $g_\alpha(u) = -\log[\frac{\exp(-\alpha u)-1}{\exp(-\alpha)-1}]$.

Tail Dependency

Tail dependency In financial markets, extreme events are of great importance. The dependence between two random variables at the extreme values can be measured by the coefficient of lower and upper tail dependence. The coefficient of lower (upper) dependence, denoted by λ_l (λ_u) is defined by

$$\lambda_l = \lim_{u \downarrow 0} \frac{C(u, u)}{u}$$

$$\lambda_u = \lim_{u \uparrow 1} \frac{\bar{C}(u, u)}{1 - u}$$

Additional References

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance.* Wiley.

Joe, H. (2014). *Dependence Modeling with Copulas.* Chapman & Hall/CRC.