IEOR E4602: Quantitative Risk Management

Introduction to Copulas

Martin Haugh

Department of Industrial Engineering and Operations Research
Columbia University
Email: martin.b.haugh@gmail.com

References: Chapter 7 of 2^{nd} ed. of QRM by McNeil, Frey and Embrechts, Chapter 8 of SDAFA by Ruppert and Matteson, and the book chapter "Coping With Copulas" by Thorsten Schmidt.

Outline

Why Study Copulas?

- Copulas separate the marginal distributions from the dependency structure in a given multivariate distribution.
- Copulas help expose the various fallacies associated with correlation.
- Copulas play an important role in pricing securities that depend on many underlying securities
  - e.g. equity basket options, collateralized debt obligations (CDO's), n^{th}-to-default options.
- They provide a source of examples regarding model risk!
  - e.g. the (in)famous Gaussian copula model that was used for pricing CDO's.
- But there are problems with copulas as well!
  1. Not always applied properly.
  2. Generally **static** in nature.
  3. Not easy to estimate in general.
- Nevertheless, an understanding of copulas is important in QRM.

The Definition of a Copula

Definition: A d-dimensional copula, $C : [0,1]^d \ :\rightarrow\ [0,1]$ is a cumulative distribution function (CDF) with uniform marginals.

We write $C(\mathbf{u}) = C(u_1, \ldots, u_d)$ for a generic copula and immediately have (why?) the following properties:

1. $C(u_1, \ldots, u_d)$ is non-decreasing in each component, u_i.
2. The i^{th} marginal distribution is obtained by setting $u_j = 1$ for $j \neq i$ and since it it is uniformly distributed

$$C(1, \ldots, 1, u_i, 1, \ldots, 1) \ = \ u_i.$$

3. For $a_i \leq b_i$, $P(U_1 \in [a_1, b_1], \ldots, U_d \in [a_d, b_d])$ must be non-negative. This implies the rectangle inequality

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d}\ C(u_{1,i_1}, \ldots, u_{d,i_d}) \ \geq \ 0$$

where $u_{j,1} = a_j$ and $u_{j,2} = b_j$.

Properties of a Copula

The reverse is also true: any function that satisfies properties 1 to 3 is a copula.

Easy then to confirm that $C(1, u_1, \ldots, u_{d-1})$ is a $(d-1)$-dimensional copula

- more generally, all k-dimensional marginals with $2 \leq k \leq d$ are copulas.

Recall the definition of the quantile function or generalized inverse: for a CDF, F, the generalized inverse, $F^{\leftarrow}$, is defined as

$$F^{\leftarrow}(x) \ := \ \inf\{v \ : \ F(v) \geq x\}.$$

We now recall the following well-known result:

Proposition: If $U \sim U[0,1]$ and F_X is a CDF, then

$$P\left(F^{\leftarrow}(U) \leq x\right) \ = \ F_X(x).$$

In the opposite direction, if X has a continuous CDF, F_X, then

$$F_X(X) \ \sim \ U[0,1].$$

Sklar's Theorem (1959)

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a multivariate random vector with CDF F_X and with continuous marginals.

Then (why?) the joint distribution of $F_{X_1}(X_1), \ldots, F_{X_d}(X_d)$ is a copula, C_X.

Can we find an expression for C_X? Yes! We have

$$\begin{aligned} C_X(u_1, \ldots, u_d) &= P\left(F_{X_1}(X_1) \leq u_1, \ldots, F_{X_d}(X_d) \leq u_d\right) \\ &= P\left(X_1 \leq F_{X_1}^{-1}(u_1), \ldots, X_d \leq F_{X_d}^{-1}(u_d)\right) \\ &= F_{\mathbf{X}}\left(F_{X_1}^{-1}(u_1), \ldots, F_{X_d}^{-1}(u_d)\right). \end{aligned} \tag{1}$$

Now let $u_j := F_{X_j}(x_j)$ so that (1) yields

$$F_{\mathbf{X}}(x_1, \ldots, x_d) = C_X(F_{X_1}(x_1), \ldots, F_{X_d}(x_d))$$

– this is one side of Sklar's Theorem!

Sklar's Theorem (1959)

Consider a d-dimensional CDF, F, with marginals F_1, ..., F_d. Then there exists a copula, C, such that

$$F(x_1, \ldots, x_d) \;=\; C\left(F_1(x_1), \ldots, F_d(x_d)\right) \tag{2}$$

for all $x_i \in [-\infty,\ \infty]$ and $i = 1, \ldots, d$.

If F_i is continuous for all $i = 1, \ldots, d$, then C is unique; otherwise C is uniquely determined only on $\text{Ran}(F_1) \times \cdots \times \text{Ran}(F_d)$ where $\text{Ran}(F_i)$ denotes the range of the CDF, F_i.

In the opposite direction, consider a copula, C, and univariate CDF's, $F_1, \ldots, F_d$. Then F as defined in (2) is a multivariate CDF with marginals $F_1, \ldots, F_d$.

An Example

- Let Y and Z be two IID random variables each with CDF, $F(\cdot)$.
- Let $X_1 := \min(Y, Z)$ and $X_2 := \max(Y, Z)$.

We then have

$$P(X_1 \leq x_1,\ X_2 \leq x_2) \ = \ 2\ F\left(\min\{x_1, x_2\}\right)\ F(x_2) \ - \ F\left(\min\{x_1, x_2\}\right)^2$$

- can show this by considering separately the two cases

(i) $x_2 \leq x_1$

and

(ii) $x_2 > x_1$.

Would like to compute the copula, $C(u_1, u_2)$, of (X_1, X_2)!

An Example

First note the two marginals satisfy

$$\begin{aligned} F_1(x) &= 2F(x) - F(x)^2 \\ F_2(x) &= F(x)^2. \end{aligned}$$

Sklar's Theorem states that $C(\cdot, \cdot)$ satisfies

$$C(F_1(x_1),\ F_2(x_2)) = F(x_1, x_2).$$

So just need to connect the pieces(!) to obtain

$$C(u_1, u_2) = 2 \min\{1 - \sqrt{1 - u_1},\ \sqrt{u_2}\}\sqrt{u_2} - \min\{1 - \sqrt{1 - u_1},\ \sqrt{u_2}\}^2.$$

When the Marginals Are Continuous

Suppose the marginal distributions, $F_1, \ldots, F_n$, are continuous. Then can show

$$F_i\left(F_i^{\leftarrow}(y)\right) \;=\; y. \tag{3}$$

Now evaluate (2) at $x_i = F_i^{\leftarrow}(u_i)$ and use (3) to obtain

$$C(\mathbf{u}) \;=\; F\left(F_1^{\leftarrow}(u_1), \ldots, F_d^{\leftarrow}(u_d)\right) \tag{4}$$

- a very useful characterization!

Invariance of the Copula Under Monotonic Transformations

Proposition: Suppose the random variables $X_1, \ldots, X_d$ have continuous marginals and copula, C_X. Let $T_i : \mathbb{R} \to \mathbb{R}$, for $i = 1, \ldots, d$ be strictly increasing functions.

Then the dependence structure of the random variables

$$Y_1 := T_1(X_1),\ \ldots,\ Y_d := T_d(X_d)$$

is also given by the copula C_X.

Sketch of proof when T_j's are continuous and $F_{X_j}^{-1}$'s exist:

First note that

$$\begin{aligned} F_Y(y_1, \ldots, y_d) &= P(T_1(X_1) \leq y_1,\ \ldots\ , T_d(X_d) \leq y_d) \\ &= P(X_1 \leq T_1^{-1}(y_1),\ \ldots\ , X_d \leq T_d^{-1}(y_d)) \\ &= F_X(T_1^{-1}(y_1),\ \ldots\ , T_d^{-1}(y_d)) \end{aligned} \tag{5}$$

so that (why?) $F_{Y_j}(y_j) = F_{X_j}(T_j^{-1}(y_j))$.

This in turn implies

$$F_{Y_j}^{-1}(y_j) = T_j(F_{X_j}^{-1}(y_j)). \tag{6}$$

Invariance of the Copula Under Monotonic Transformations

Now to the proof:

$$
\begin{aligned}
C_Y(u_1, \ldots, u_d) &= F_Y\left(F_{Y_1}^{-1}(u_1),\ \ldots\ , F_{Y_d}^{-1}(u_d)\right) \text{ by } (4) \\
&= F_X\left(T_1^{-1}\left(F_{Y_1}^{-1}(u_1)\right),\ \ldots\ , T_d^{-1}\left(F_{Y_d}^{-1}(u_d)\right)\right) \text{ by } (5) \\
&= F_X\left(F_{X_1}^{-1}(u_1),\ \ldots\ , F_{X_d}^{-1}(u_d)\right) \text{ by } (6) \\
&= C_X(u_1, \ldots, u_d)
\end{aligned}
$$

and so $C_X = C_Y$.

The Fréchet-Hoeffding Bounds

Theorem: Consider a copula $C(\mathbf{u}) = C(u_1, \ldots, u_d)$. Then

$$\max\left\{1 - d + \sum_{i=1}^{d} u_i,\ 0\right\} \ \le\ C(\mathbf{u})\ \le\ \min\{u_1, \ldots, u_d\}.$$

Sketch of Proof: The first inequality follows from the observation

$$\begin{aligned} C(\mathbf{u}) &= P\left(\bigcap_{1\le i\le d} \{U_i \le u_i\}\right) \\ &= 1 - P\left(\bigcup_{1\le i\le d} \{U_i > u_i\}\right) \\ &\ge 1 - \sum_{i=1}^{d} P(U_i > u_i) \ =\ 1 - d + \sum_{i=1}^{d} u_i. \end{aligned}$$

The second inequality follows since $\bigcap_{1\le i\le d}\{U_i \le u_i\} \subseteq \{U_i \le u_i\}$ for all i. □

Tightness of the Fréchet-Hoeffding Bounds

The upper Fréchet-Hoeffding bound is tight for all d.

The lower Fréchet-Hoeffding bound is tight only when $d = 2$.

Fréchet and Hoeffding showed independently that copulas always lie between these bounds

- corresponding to cases of **extreme** of dependency, i.e. comonotonicity and countermonotonicity.

Comonotonicity and The Perfect Dependence Copula

The comonotonic copula is given by

$$M(\mathbf{u}) \ := \ \min\{u_1, \ldots, u_d\}$$

- the Fréchet-Hoeffding upper bound.

It corresponds to the case of extreme positive dependence.

Proposition: Let $X_1, \ldots, X_d$ be random variables with continuous marginals and suppose $X_i = T_i(X_1)$ for $i = 2, \ldots, d$ where $T_2, \ldots, T_d$ are strictly increasing transformations. Then $X_1, \ldots, X_d$ have the comonotonic copula.

Proof: Apply the *invariance under monotonic transformations* proposition and observe that the copula of $(X_1, X_1, \ldots, X_1)$ is the comonotonic copula.

Countermonotonic Random Variables

The countermonotonic copula is the 2-dimensional copula that is the Fréchet-Hoeffding lower bound.

It satisfies

$$W(u_1, u_2) \; = \; \max\{u_1 + u_2 - 1, \; 0\} \tag{7}$$

and corresponds to the case of perfect negative dependence.

Can check that (7) is the joint distribution of $(U, 1 - U)$ where $U \sim U(0, 1)$.

Question: Why is the Fréchet-Hoeffding lower bound **not** a copula for $d > 2$?

The Independence Copula

The independence copula satisfies

$$\Pi(\mathbf{u}) \;=\; \prod_{i=1}^{d} u_i.$$

And random variables are independent if and only if their copula is the independence copula

- follows immediately from Sklar's Theorem.

The Gaussian Copula

Recall when marginals are continuous we have from (4)

$$C(\mathbf{u}) \ = \ F\left(F_1^{\leftarrow}(u_1), \ldots, F_1^{\leftarrow}(u_d)\right).$$

Now let $\mathbf{X} \sim \mathsf{MN}_d(\mathbf{0}, \mathbf{P})$, where P is the correlation matrix of $\mathbf{X}$.

Then the corresponding Gaussian copula is given by

$$C_P^{Gauss}(\mathbf{u}) \ := \ \boldsymbol{\Phi}_{\boldsymbol{P}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right) \tag{8}$$

- where $\Phi(\cdot)$ is the standard univariate normal CDF
- and $\boldsymbol{\Phi}_{\boldsymbol{P}}(\cdot)$ denotes the joint CDF of $\mathbf{X}$.

If $\mathbf{Y} \sim \mathsf{MN}_d(\mu, \Sigma)$ with $\mathsf{Corr}(\mathbf{Y}) = \mathbf{P}$, then $\mathbf{Y}$ has (why?) the same copula as $\mathbf{X}$

- hence **a Gaussian copula is fully specified by a correlation matrix, $\mathbf{P}$.**

For $d = 2$, obtain countermonotonic, independence and comonotonic copulas when $\rho = -1,\ 0$, and 1, respectively.

The t Copula

Recall $\mathbf{X} = (X_1, \ldots, X_d)$ has a multivariate t distribution with ν dof if

$$\mathbf{X} \;=\; \frac{\mathbf{Z}}{\sqrt{\xi/\nu}}$$

- where $\mathbf{Z} \sim \mathsf{MN}_d(\mathbf{0}, \mathbf{\Sigma})$
- and $\xi \sim \chi^2_\nu$ independently of $\mathbf{Z}$.

The d-dimensional t-copula is defined as

$$C^t_{\nu,P}(\mathbf{u}) \;:=\; \mathbf{t}_{\nu,\mathbf{P}}\left(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d)\right) \tag{9}$$

- where again $\mathbf{P}$ is a correlation matrix
- $\mathbf{t}_{\nu,\mathbf{P}}$ is the joint CDF of $\mathbf{X} \sim \mathbf{t}_d(\nu, \mathbf{0}, \mathbf{P})$
- and t_ν is the standard univariate CDF of a t-distribution with ν dof.

The Bivariate Gumbel Copula

The bivariate Gumbel copula is defined as

$$C_\theta^{Gu}(u_1, u_2) \;:=\; \exp\left(-\left((-\ln u_1)^\theta \;+\; (-\ln u_2)^\theta\right)^{\frac{1}{\theta}}\right)$$

where $\theta \in [1, \infty)$.

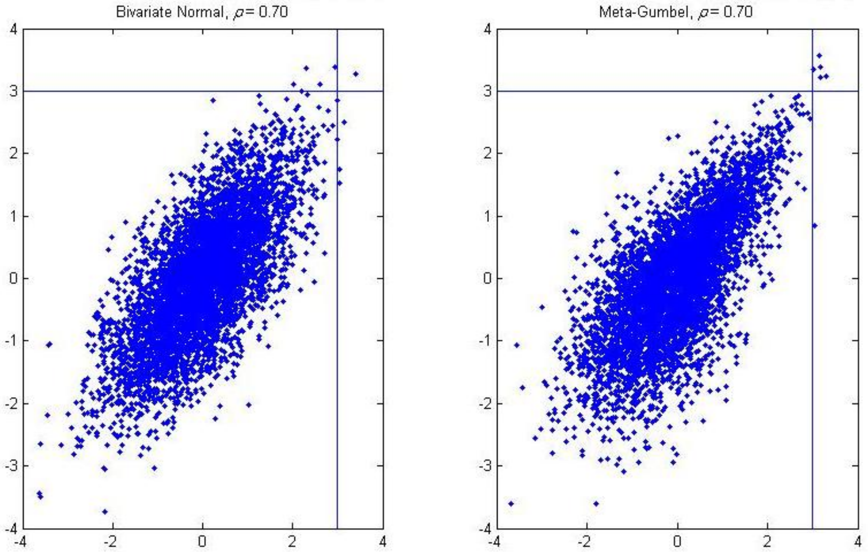
When $\theta = 1$ obtain the independence copula.

As $\theta \to \infty$ the Gumbel copula $\to$ the comonotonicity copula

- an example of a copula with tail dependence in just one corner.

e.g. Consider bivariate Normal and **meta-Gumbel** distributions on next slide:

- $5,000$ points simulated from each distribution
- **marginal** distributions in each case are standard normal
- correlation is $\approx .7$ in both cases
- but meta-Gumbel is much more likely to see large **joint** moves.



Bivariate Normal, ρ= 0.70
Meta-Gumbel, ρ= 0.70

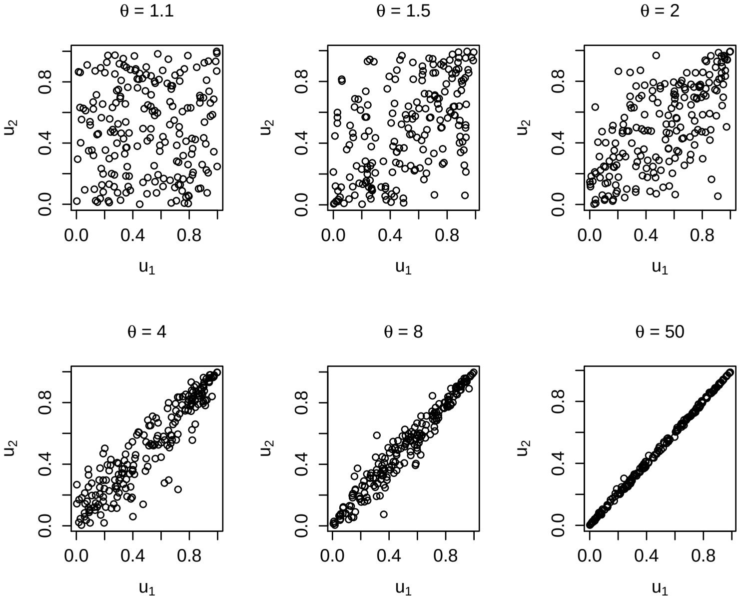


Figure 8.4 from Ruppert and Matteson: Bivariate random samples of size 200 from various Gumbel copulas.

The Bivariate Clayton Copula

The bivariate Clayton copula is defined as

$$C_\theta^{Cl}(u_1, u_2) \ := \ \left(u_1^{-\theta} \ + \ u_2^{-\theta} - 1\right)^{-1/\theta}$$

where $\theta \in [-1, \infty) \backslash \{0\}$.

- As $\theta \to 0$ obtain the independence copula.
- As $\theta \to \infty$ the Clayton copula $\to$ the comonotonic copula.
- For $\theta = -1$ obtain the Fréchet-Hoeffding lower bound
  - so Clayton moves from countermonotonic to independence to comonotonic copulas.

- Clayton and Gumbel copulas belong to the Archimedean family of copulas
  - they can be be generalized to d dimensions
  - but their d-dimensional versions are exchangeable – so $C(u_1, \ldots, u_d)$ unchanged if we permute $u_1, \ldots, u_d$
  - implies all pairs have same dependence structure – has implications for modeling with Archimedean copulas!

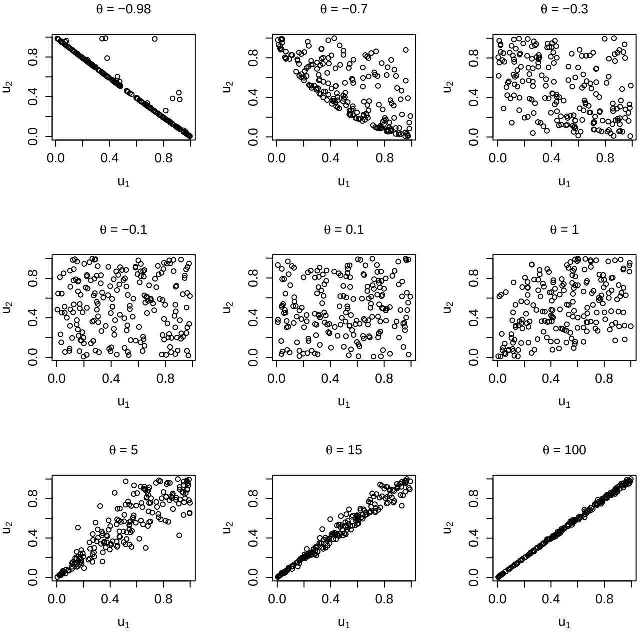


Figure 8.3 from Ruppert and Matteson: Bivariate random samples of size 200 from various Clayton copulas.

Measures of Dependence

There are three principal measures of dependence:

1. The usual Pearson, i.e. linear, correlation coefficient
   - invariant under positive linear transformations, but not under general strictly increasing transformations
   - there are many fallacies associated with the Pearson correlation
   - not defined unless second moments exist.
2. Rank correlations
   - only depend on the unique copula of the joint distribution
   - therefore (why?) invariant to strictly increasing transformations
   - also very useful for calibrating copulas to data.
3. Coefficients of tail dependence
   - a measure of dependence in the extremes of the distributions.

Fallacies of The Correlation Coefficient

Each of the following statements is false!

1. The marginal distributions and correlation matrix are enough to determine the joint distribution
   - how would we find a counterexample?

2. For given univariate distributions, F_1 and F_2, and any correlation value $\rho \in [-1, 1]$, it is always possible to construct a joint distribution F with margins F_1 and F_2 and correlation ρ.

3. The VaR of the sum of two risks is largest when the two risks have maximal correlation.

Definition: We say two random variables, X_1 and X_2, are of the same type if there exist constants $a > 0$ and $b \in \mathbb{R}$ such that

$$X_1 \sim aX_2 + b.$$

On Fallacy $\#2$

Theorem: Let (X_1, X_2) be a random vector with finite-variance marginal CDF's F_1 and F_2, respectively, and an unspecified joint CDF.

Assuming $\text{Var}(X_1) > 0$ and $\text{Var}(X_2) > 0$, then the following statements hold:

1. The attainable correlations form a closed interval $[\rho_{\text{min}}, \rho_{\text{max}}]$ with $\rho_{\text{min}} < 0 < \rho_{\text{max}}$.

2. The minimum correlation $\rho = \rho_{\text{min}}$ is attained if and only if X_1 and X_2 are countermonotonic. The maximum correlation $\rho = \rho_{\text{max}}$ is attained if and only if X_1 and X_2 are comonotonic.

3. $\rho_{\text{min}} = -1$ if and only if X_1 and $-X_2$ are of the same type. $\rho_{\text{max}} = 1$ if and only if X_1 and X_2 are of the same type.

The proof is not very difficult; see Section 7.2 of *MFE* for details.

Spearman's Rho

Definition: For random variables X_1 and X_2, Spearman's rho is defined as

$$\rho_s(X_1, X_2) \; := \; \rho(F_1(X_1), F_2(X_2)).$$

So Spearman's rho is simply the linear correlation of the probability-transformed random variables.

The Spearman's rho matrix is simply the matrix of pairwise Spearman's rho correlations, $\rho(F_i(X_i), F_j(X_j))$ – a positive-definite matrix. Why?

If X_1 and X_2 have continuous marginals then can show

$$\rho_s(X_1, X_2) \; = \; 12 \int_0^1 \int_0^1 \; (C(u_1, u_2) \; - \; u_1 u_2) \; du_1 \; du_2.$$

Can show that for a bivariate Gaussian copula

$$\rho_s(X_1, X_2) \; = \frac{6}{\pi} \; \arcsin \frac{\rho}{2} \; \simeq \; \rho$$

where ρ is the Pearson, i.e., linear, correlation coefficient.

Kendall's Tau

Definition: For random variables X_1 and X_2, Kendall's tau is defined as

$$\rho_\tau(X_1, X_2) \;:=\; \mathsf{E}\left[\text{sign}\left((X_1 - \tilde{X}_1)\,(X_2 - \tilde{X}_2)\right)\right]$$

where $(\tilde{X}_1, \tilde{X}_2)$ is independent of (X_1, X_2) but has same joint distribution as (X_1, X_2).

Note that Kendall's tau can be written as

$$\rho_\tau(X_1, X_2) \;=\; P\left((X_1 - \tilde{X}_1)\,(X_2 - \tilde{X}_2)\; > 0\right) - \;P\left((X_1 - \tilde{X}_1)\,(X_2 - \tilde{X}_2)\; < 0\right)$$

- so if both probabilities are equal then $\rho_\tau(X_1, X_2) = 0$.

Kendall's Tau

If X_1 and X_2 have continuous marginals then can show

$$\rho_\tau(X_1, X_2) \;=\; 4 \int_0^1 \int_0^1 \; C(u_1, u_2) \; dC(u_1, u_2) \;-\; 1.$$

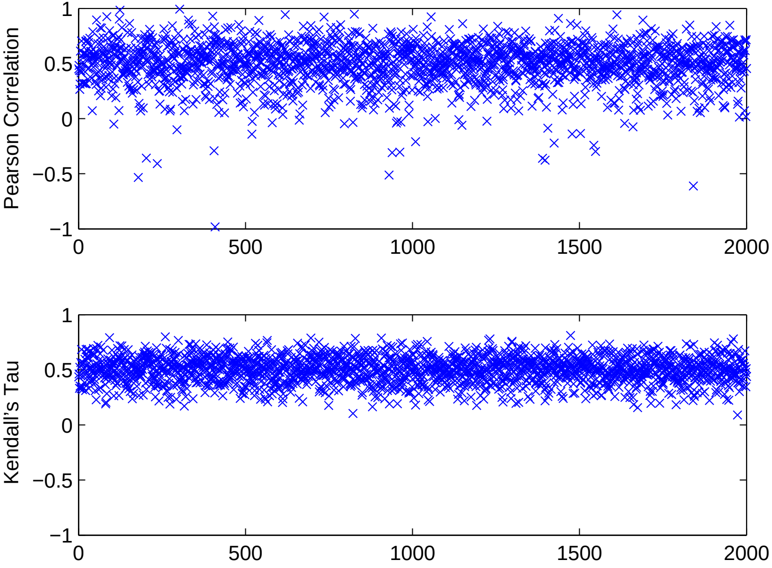
Can also show that for a bivariate Gaussian copula, or more generally, if $\mathbf{X} \sim \mathbf{E}_2(\mu, \mathbf{P}, \psi)$ and $P(\mathbf{X} = \mu) = 0$

$$\rho_\tau(X_1, X_2) \;=\frac{2}{\pi}\; \arcsin \rho \tag{10}$$

where $\rho = \mathbf{P}_{12} = \mathbf{P}_{21}$ is the Pearson correlation coefficient.

Note (10) very useful for estimating ρ with fat-tailed elliptical distributions

- generally provides much more robust estimates of ρ than usual Pearson estimator
- see figure on next slide where each estimate was constructed from a sample of $n = 60$ (simulated) data-points.



Estimating Pearson's correlation using the usual Pearson estimator versus using Kendall's τ. Underlying distribution was bivariate t with $\nu = 3$ degrees-of-freedom and true Pearson correlation $\rho = 0.5$.

Properties of Spearman's Rho and Kendall's Tau

Spearman's rho and Kendall's tau are examples of rank correlations in that, when the marginals are continuous, they depend only on the bivariate copula and not on the marginals

- they are invariant (why?) in this case under strictly increasing transformations.

They both take values in $[-1, 1]$:

- they equal 0 for independent random variables; but possible for dependent variables to also have a rank correlation of 0.
- they take the value 1 when X_1 and X_2 are comonotonic
- they take the value -1 when X_1 and X_2 are countermonotonic.

They are very useful for calibrating copulas via **method-of-moments** type algorithms.

Fallacy $\#2$ is no longer an issue when we work with rank correlations.

Tail Dependence

Definition: Let X_1 and X_2 denote two random variables with CDF's F_1 and F_2, respectively. Then the coefficient of upper tail dependence, λ_u, is given by

$$\lambda_u \;:=\; \lim_{q \nearrow 1} \; P\left(X_2 \;>\; F_2^{\leftarrow}(q) \mid X_1 \;>\; F_1^{\leftarrow}(q)\right)$$

provided that the limit exists.

Similarly, the coefficient of lower tail dependence, λ_l, is given by

$$\lambda_l \;:=\; \lim_{q \searrow 0} \; P\left(X_2 \;\leq\; F_2^{\leftarrow}(q) \mid X_1 \;\leq\; F_1^{\leftarrow}(q)\right)$$

provided again that the limit exists.

- If $\lambda_u > 0$, then we say that X_1 and X_2 have upper tail dependence while if $\lambda_u = 0$ we say they are asymptotically independent in the upper tail.
- Lower tail dependence and asymptotically independent in the lower tail are similarly defined using λ_l.

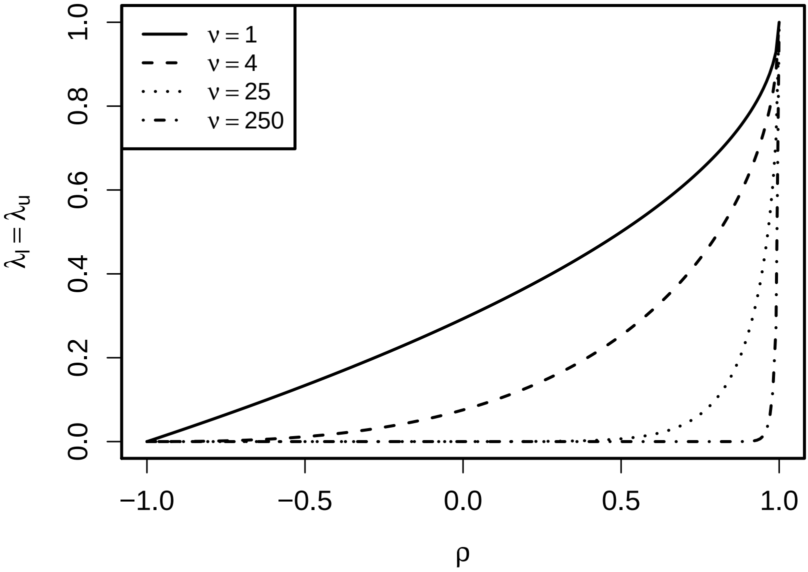


Figure 8.6 from Ruppert and Matteson: Coefficients of tail dependence for bivariate t-copulas as functions of ρ for $\nu = 1$, 4, 25 and 250.

Simulating the Gaussian Copula

1. For an arbitrary covariance matrix, $\mathbf{\Sigma}$, let $\mathbf{P}$ be its corresponding correlation matrix.
2. Compute the Cholesky decomposition, $\mathbf{A}$, of $\mathbf{P}$ so that $\mathbf{P} = \mathbf{A}^T\mathbf{A}$.
3. Generate $\mathbf{Z} \sim \mathsf{MN}_d(\mathbf{0}, \mathbf{I}_d)$.
4. Set $\mathbf{X} = \mathbf{A}^T\,\mathbf{Z}$.
5. Return $\mathbf{U} = (\Phi(X_1), \ldots, \Phi(X_d))$.

The distribution of $\mathbf{U}$ is the Gaussian copula $C_P^{Gauss}(\mathbf{u})$ so that

$$\mathsf{Prob}(U_1 \leq u_1, \ldots, U_d \leq u_d) \;=\; \mathbf{\Phi}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right)$$

- this is also (why?) the copula of $\mathbf{X}$.

Desired marginal inverses can now be applied to each component of $\mathbf{U}$ in order to generate the desired multivariate distribution with the given Gaussian copula

- follows again by invariance of copula to monotonic transformations.

Simulating the t Copula

1. For an arbitrary covariance matrix, $\mathbf{\Sigma}$, let $\mathbf{P}$ be its corresponding correlation matrix.
2. Generate $\mathbf{X} \sim \mathsf{MN}_d(\mathbf{0}, \mathbf{P})$.
3. Generate $\xi \sim \chi^2_\nu$ independent of $\mathbf{X}$.
4. Return $\mathbf{U} = \left(t_\nu(X_1/\sqrt{\xi/\nu}), \ldots, t_\nu(X_d/\sqrt{\xi/\nu})\right)$ where t_ν is the CDF of a univariate t distribution with ν degrees-of-freedom.

The distribution of $\mathbf{U}$ is the t copula $C^t_{\nu,P}(\mathbf{u})$

- this is also (why?) the copula of $\mathbf{X}$.

Desired marginal inverses can now be applied to each component of $\mathbf{U}$ in order to generate the desired multivariate distribution with the given t copula.

Estimating / Calibrating Copulas

There are several related methods that can be used for estimating copulas:

1. Maximum likelihood estimation (MLE)
2. Pseudo-MLE of which there are two types:
   - parametric pseudo-MLE
   - semiparametric pseudo-MLE
3. Moment-matching methods are also sometimes used
   - they can also be used for finding starting points for (pseudo) MLE.

MLE is often considered too difficult to apply

- too many parameters to estimate.

Pseudo-MLE seems to be used most often in practice

- marginals are estimated via their empirical CDFs
- then the copula can be estimated via MLE.

Maximum Likelihood Estimation

Let $\mathbf{Y} = (Y_1 \ldots Y_d)^\top$ be a random vector and suppose we have parametric models $F_{Y_1}(\cdot \,|\, \boldsymbol{\theta}_1), \ldots, F_{Y_d}(\cdot \,|\, \boldsymbol{\theta}_d)$ for the marginal CDFs.

Also have a parametric model $c_{\mathbf{Y}}(\cdot \,|\, \boldsymbol{\theta}_C)$ for the copula density of $\mathbf{Y}$.

By differentiating (2) we see that the density of $\mathbf{Y}$ is given by

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{Y}}(y_1, \ldots, y_d) = c_{\mathbf{Y}}\left(F_{Y_1}(y_1), \ldots, F_{Y_d}(y_d)\right) \prod_{j=1}^{d} f_{Y_j}(y_j). \tag{11}$$

Given an IID sample $\mathbf{Y}_{1:n} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$, we obtain the log-likelihood as

$$\begin{aligned}
\log L(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \boldsymbol{\theta}_C) &= \log \prod_{i=1}^{n} f_{\mathbf{Y}}(\mathbf{y}_i) \\
&= \sum_{i=1}^{n} \Big(\log[c_{\mathbf{Y}}\left(F_{Y_1}(y_{i,1} \,|\, \boldsymbol{\theta}_1), \ldots, F_{Y_d}(y_{i,d} \,|\, \boldsymbol{\theta}_d) \,|\, \boldsymbol{\theta}_C\right)] \\
&\quad + \log\left(f_{Y_1}(y_{i,1} \,|\, \boldsymbol{\theta}_1)\right) + \cdots + \log\left(f_{Y_d}(y_{i,d} \,|\, \boldsymbol{\theta}_d)\right) \Big)
\end{aligned} \tag{12}$$

Maximum Likelihood Estimation

The ML estimators $\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_d, \hat{\boldsymbol{\theta}}_C$ are obtained by maximizing (12).

But there are problems with this:

1. Many parameters – especially for large d – so optimization can be difficult.
2. If any of the parametric univariate distributions $F_{Y_i}(\cdot \mid \boldsymbol{\theta}_i)$ are misspecified then this can cause biases in estimation of both univariate distributions and the copula.

The pseudo-MLE approach helps to resolve these problems.

Pseudo-Maximum Likelihood Estimation

The pseudo-MLE approach has two steps:

1. First estimate the marginal CDFs to obtain $\hat{F}_{Y_j}$ for $j = 1, \ldots, d$. Can do this using either:
   - The empirical CDF of $y_{1,j}, \ldots, y_{n,j}$ so that
   $$\hat{F}_{Y_j}(y) = \frac{\sum_{i=1}^{n} 1_{\{y_{i,j} \leq y\}}}{n+1}$$
   - A parametric model with $\hat{\boldsymbol{\theta}}_j$ obtained using usual MLE approach.
2. Then estimate the copula parameters $\boldsymbol{\theta}_C$ by maximizing
$$\sum_{i=1}^{n} \log\left[c_{\mathbf{Y}}\left(\hat{F}_{Y_1}(y_{i,1}), \ldots, \hat{F}_{Y_d}(y_{i,d}) \,|\, \boldsymbol{\theta}_C\right)\right] \tag{13}$$
   - Note relation between (12) and (13)!

Even (13) may be difficult to maximize if d large

- important then to have good starting point for the optimization or to impose **additional structure** on $\boldsymbol{\theta}_C$.

Fitting Gaussian and t Copulas

Proposition (Results 8.1 from Ruppert and Matteson):
Let $\mathbf{Y} = (Y_1 \ldots Y_d)^\top$ have a meta-Gaussian distribution with continuous univariate marginal distributions and copula $C_{\mathbf{\Omega}}^{Gauss}$ and let $\Omega_{i,j} = [\mathbf{\Omega}]_{i,j}$. Then

$$\rho_\tau(Y_i, Y_j) = \frac{2}{\pi} \arcsin(\Omega_{i,j}) \tag{14}$$

and

$$\rho_S(Y_i, Y_j) = \frac{6}{\pi} \arcsin(\Omega_{i,j}/2) \approx \Omega_{i,j} \tag{15}$$

If instead $\mathbf{Y}$ has a meta-t distribution with continuous univariate marginal distributions and copula $C_{\nu,\mathbf{\Omega}}^t$ then (14) still holds but (15) does not. □

Question: There are several ways to use this Proposition to fit meta Gaussian and t copulas. What are some of them?

Collateralized Debt Obligations (CDO's)

Want to find the expected losses in a simple 1-period CDO with the following characteristics:

- The maturity is 1 year.
- There are $N = 125$ bonds in the reference portfolio.
- Each bond pays a coupon of one unit after 1 year if it has not defaulted.
- The recovery rate on each defaulted bond is zero.
- There are 3 tranches of interest:

1. The equity tranche with attachment points: 0-3 defaults
2. The mezzanine tranche with attachment points: 4-6 defaults
3. The senior tranche with attachment points: 7-125 defaults.

Assume probability, q, of defaulting within 1 year is **identical** across all bonds.

Collateralized Debt Obligations (CDO's)

X_i is the normalized asset value of the i^{th} credit and we assume

$$X_i = \sqrt{\rho} M + \sqrt{1-\rho}\, Z_i \tag{16}$$

where $M, Z_1, \ldots, Z_N$ are IID normal random variables

- note correlation between each pair of asset values is identical.

We assume also that i^{th} credit defaults if $X_i \leq \bar{x}_i$.

Since probability, q, of default is identical across all bonds must therefore have

$$\bar{x}_1 = \cdots \bar{x}_N = \Phi^{-1}(q). \tag{17}$$

It now follows from (16) and (17) that

$$\begin{aligned} \mathsf{P}(i \text{ defaults} \,|\, M) &= \mathsf{P}(X_i \leq \bar{x}_i \,|\, M) \\ &= \mathsf{P}(\sqrt{\rho} M + \sqrt{1-\rho}\, Z_i \leq \Phi^{-1}(q) \,|\, M) \\ &= \mathsf{P}\left(Z_i \leq \frac{\Phi^{-1}(q) - \sqrt{\rho} M}{\sqrt{1-\rho}} \,\middle|\, M \right). \end{aligned}$$

Collateralized Debt Obligations (CDO's)

Therefore conditional on M, the total number of defaults is **Bin**(N, q_M) where

$$q_M := \Phi\left(\frac{\Phi^{-1}(q) - \sqrt{\rho}M}{\sqrt{1-\rho}}\right).$$

That is,

$$p(k \,|\, M) = \binom{N}{k} q_M^k (1 - q_M)^{N-k}.$$

Unconditional probabilities computed by numerically integrating the binomial probabilities with respect to M so that

$$\mathsf{P}(k \text{ defaults}) = \int_{-\infty}^{\infty} p(k \,|\, M)\phi(M)\, dM.$$

Can now compute expected (risk-neutral) loss on each of the three tranches:

Collateralized Debt Obligations (CDO's)

$$\begin{aligned}
\mathsf{E}_0^{\mathbb{Q}}\,[\text{Equity tranche loss}] &= 3 \times \mathsf{P}(3 \text{ or more defaults}) + \sum_{k=1}^{2} k\,\mathsf{P}(k \text{ defaults}) \\
\mathsf{E}_0^{\mathbb{Q}}\,[\text{Mezz tranche loss}] &= 3 \times \mathsf{P}(6 \text{ or more defaults}) + \sum_{k=1}^{2} k\,\mathsf{P}(k+3 \text{ defaults}) \\
\mathsf{E}_0^{\mathbb{Q}}\,[\text{Senior tranche loss}] &= \sum_{k=1}^{119} k\,\mathsf{P}(k+6 \text{ defaults}).
\end{aligned}$$

Collateralized Debt Obligations (CDO's)

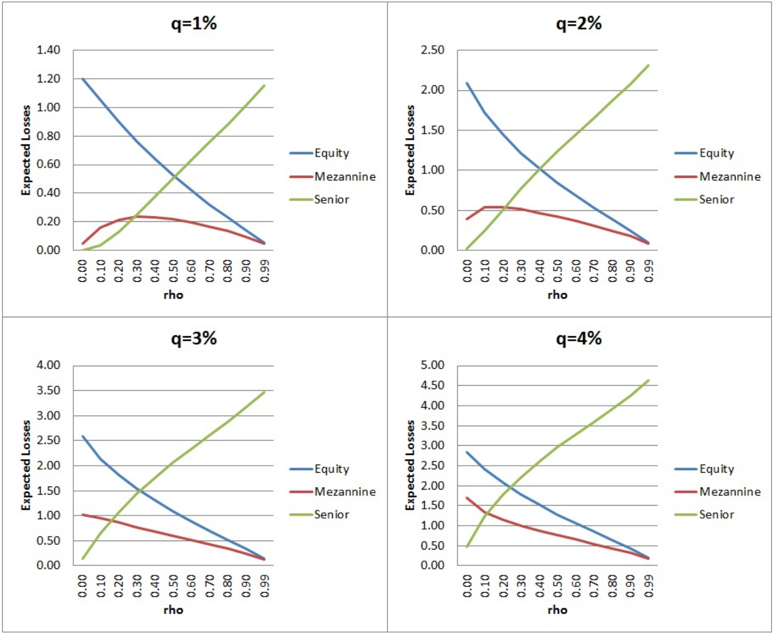
Regardless of the individual default probability, q, and correlation, ρ, we have:

$$\mathsf{E}_0^{\mathbb{Q}}\left[\% \text{ Equity tranche loss}\right] \geq \mathsf{E}_0^{\mathbb{Q}}\left[\% \text{ Mezz tranche loss}\right] \geq \mathsf{E}_0^{\mathbb{Q}}\left[\% \text{ Senior tranche loss}\right].$$

Also note that expected equity tranche loss always **decreasing** in ρ.

Expected mezzanine tranche loss often relatively insensitive to ρ.

Expected senior tranche loss (with upper attachment point of 100%) always **increasing** in ρ.



Expected Tranche Losses As a Function of ρ

Collateralized Debt Obligations (CDO's)

Question: How does the total expected loss in the portfolio vary with ρ?

The **dependence structure** we used in (**??**) to link the default events of the various bonds is the famous Gaussian-copula model.

In practice CDO's are **multi-period** securities and can be **cash** or **synthetic** CDO's.

Multi-period CDO's

Will now assume:

- There are N credits in the reference portfolio.
- Each credit has a notional amount of A_i.
- If the i^{th} credit defaults, then the portfolio incurs a loss of $A_i \times (1 - R_i)$
  - R_i is the recovery rate
  - we assume R_i is fixed and known.
- The default time of i^{th} credit is $\text{Exp}(\lambda_i)$ with CDF F_i
  - λ_i easily estimated from either CDS spreads or prices of corporate bonds
  - so can compute $F_i(t)$, the risk-neutral probability that i^{th} credit defaults before time t.

We can easily relax the exponential distribution assumption

- we simply need to be able to estimate $F_i(t)$.

Multi-period CDO's

Again X_i is the normalized asset value of the i^{th} credit and we assume

$$X_i \;=\; a_i M \;+\; \sqrt{1-a_i^2}\, Z_i \tag{18}$$

where $M, Z_1, \ldots, Z_N$ are IID normal random variables.

The factor loadings, a_i, are assumed to lie in interval $[0, 1]$.

Clear that $\text{Corr}(X_i, X_j) \;=\; a_i a_j$ with covariance matrix equal to correlation matrix, $\mathbf{P}$, where $\mathbf{P}_{i,j} = a_i a_j$ for $i \neq j$.

Let $F(t_1, \ldots, t_n)$ denote joint distribution of the default times of the N credits.

Then **assume**

$$F(t_1, \ldots, t_n) \;\;=\;\; \mathbf{\Phi}_P\left(\Phi^{-1}(F_1(t_1)), \ldots, \Phi^{-1}(F_n(t_n))\right) \tag{19}$$

where $\mathbf{\Phi}_P(\cdot)$ is the multivariate normal CDF with mean $\mathbf{0}$ and correlation $\mathbf{P}$

- so distribution of default times has a Gaussian copula!

Computing the Portfolio Loss Distribution

In order to price credit derivatives, we need to compute the portfolio loss distribution.

We fix $t = t_1 = \cdots = t_N$ and set $q_i := F_i(t)$.

As before, the N default events are independent, conditional on M. They are given by

$$q_i(t|M) \ = \ \Phi\left(\frac{\Phi^{-1}(q_i) - a_i M}{\sqrt{1 - a_i^2}}\right).$$

Now let $p^N(l, t) =$ risk-neutral probability that there are a total of l portfolio defaults before time t.

Then may write

$$p(l, t) \ = \ \int_{-\infty}^{\infty} p^N(l, t|M) \ \phi(M) \ dM. \tag{20}$$

Computing the Portfolio Loss Distribution

Straightforward to calculate $p^N(l, t|M)$ using a simple iterative procedure.

Can then perform a numerical integration on the right-hand-side of (20) to calculate $p(l, t)$.

If we assume that the notional, A_i, and the recovery rate, R_i, are constant across all credits, then the loss on any given credit will be either 0 or $A(1 - R)$.

This then implies that knowing probability distribution of the number of defaults is equivalent to knowing probability distribution of the total loss in the reference portfolio.

The Mechanics and Pricing of a CDO Tranche

A tranche is defined by the lower and upper attachment points, L and U, respectively.

The tranche loss function, $TL^{L,U}(l)$, for a fixed time, t, is a function of the number of defaults, l, and is given by

$$TL_t^{L,U}(l) \;:=\; \max\left\{\min\{lA(1-R),\; U\} - L,\; 0\right\}.$$

For a given number of defaults it tells us the loss suffered by the tranche.

For example:

- Suppose $L = 3\%$ and $U = 7\%$
- Suppose also that total portfolio loss is $lA(1-R) = 5\%$

Then tranche loss is 2% of total portfolio notional

- or 50% of tranche notional.

The Mechanics and Pricing of a CDO Tranche

When an investor sells protection on the tranche she is guaranteeing to reimburse any realized losses on the tranche to the protection buyer.

In return, the protection seller is paid a premium at regular intervals

- typically every three months
- though in some cases protection buyer may also pay an upfront amount in addition to, or instead of, a regular premium
- an upfront typically occurs for equity tranches which have a lower attachment point of zero.

The fair value of the CDO tranche is that value of the premium for which the expected value of the premium leg equals the expected value of the default leg.

The Mechanics and Pricing of a CDO Tranche

Clearly then the fair value of the CDO tranche depends on the expected value of the tranche loss function.

Indeed, for a fixed time, t, the expected tranche loss is given by

$$\mathsf{E}\left[TL_t^{L,U}\right] = \sum_{l=0}^{N} TL_t^{L,U}(l)\ p(l,t)$$

– which we can compute using (20).

We now compute the fair value of the premium and default legs ...

Fair Value of the Premium Leg

Premium leg represents the premium payments that are paid periodically by the protection buyer to the protection seller.

They are paid at the end of each time interval and they are based upon the remaining notional in the tranche.

Formally, the time $t = 0$ value of the premium leg, $P_0^{L,U}$, satisfies

$$PL_0^{L,U} \;=\; s \sum_{t=1}^{n} d_t \Delta_t \left((U - L) - \mathsf{E}_0 \left[TL_t^{L,U} \right] \right) \tag{21}$$

- n is the number of periods in the contract
- d_t is the risk-free discount factor for payment date t
- s is the annualized spread or premium paid to the protection seller
- Δ_t is the accrual factor for payment date t, e.g. $\Delta_t = 1/4$ if payments take place quarterly.

Note that (21) is consistent with the statement that the premium paid at any time t is based only on the remaining notional in the tranche.

Fair Value of the Default Leg

Default leg represents the cash flows paid to the protection buyer upon losses occurring in the tranche.

Formally, the time $t = 0$ value of the default leg, $DL_0^{L,U}$, satisfies

$$DL_0^{L,U} \;=\; \sum_{t=1}^{n} d_t \left(\mathsf{E}_0 \left[TL_t^{L,U} \right] - \mathsf{E}_0 \left[TL_{t-1}^{L,U} \right] \right). \tag{22}$$

Fair Value of the Tranche

The fair premium, s^* say, is the value of s that equates the value of the default leg with the value of the premium leg:

$$s^* \;:=\; \frac{DL_0^{L,U}}{\sum_{t=1}^{n} d_t \Delta_t \left((U-L) - \mathsf{E}_0\left[TL_t^{L,U} \right] \right)}.$$

As is the case with swaps and forwards, the fair value of the tranche to the protection buyer and seller at initiation is therefore zero.

Easy to incorporate any possible upfront payments that the protection buyer must pay at time $t = 0$ in addition to the regular premium payments.

Can also incorporate recovery values and notional values that vary with each credit in the portfolio.

Cash CDO's

First CDOs to be traded were all cash CDOs

- the reference portfolio actually existed and consisted of corporate bonds that the CDO issuer usually kept on its balance sheet.

Capital requirements meant that these bonds required a substantial amount of capital to be set aside to cover any potential losses.

To reduce these capital requirements, banks converted the portfolio into a series of tranches and sold most of these tranches to investors.

Banks usually kept the equity tranche for themselves. This meant:

- They kept most of the economic risk and rewards of the portfolio
- But they also succeeded in dramatically reducing the amount of capital they needed to set aside
- Hence first CDO deals were motivated by regulatory arbitrage considerations.

Synthetic CDO's

Soon become clear there was an appetite in the market-place for these products.

e.g. Hedge funds were keen to buy the riskier tranches while insurance companies and others sought the AAA-rated senior and super-senior tranches.

This appetite and explosion in the CDS market gave rise to synthetic tranches where:

- the underlying reference portfolio is no longer a physical portfolio of corporate bonds or loans
- it is instead a fictitious portfolio consisting of a number of credits with an associated notional amount for each credit.

Synthetic CDO's

Mechanics of a synthetic tranche are precisely as described earlier.

But they have at least two features that distinguish them from cash CDOs:

(i) With a synthetic CDO it is no longer necessary to tranche the entire portfolio and sell the entire "deal"

e.g. A bank could sell protection on a 3%-7% tranche and never have to worry about selling the other pieces of the reference portfolio. This is not the case with cash CDOs.

(ii) Because the issuer no longer owns the underlying bond portfolio, it is no longer hedged against adverse price movements

- it therefore needs to dynamically hedge its synthetic tranche position and would have typically done so using the CDS markets.

Calibrating the Gaussian Copula Model

In practice, very common to calibrate synthetic tranches as follows:

1. Assume all pairwise correlations, $\text{Corr}(X_i, X_j)$, are identical
   - equivalent to taking $a_1 = \cdots = a_N = a$ in (18) so that $\text{Corr}(X_i, X_j) = a^2 := \rho$ for all i, j.

2. In the case of the liquid CDO tranches whose prices are observable in the market-place, we then choose ρ so that the fair tranche spread in the model is equal to the quoted spread in the market place.

We refer to this calibrated correlation, ρ_{imp} say, as the tranche implied correlation.

Calibrating the Gaussian Copula Model

If the model is correct, then every tranche should have the same ρ_{imp}.

Unfortunately, this does not occur in practice.

Indeed, in the case of mezzanine tranches it is possible that there is no value of ρ that fits the market price!
It is also possible that there are **multiple** solutions.

The market responded to this problem by introducing the concept of base correlations

- they are the implied correlations of equity tranches with increasing upper attachment points.

Implied base correlations can always be computed and then bootstrapping techniques are employed to price the mezzanine tranches.

Just as equity derivatives markets have an implied volatility surface, the CDO market has implied base correlation curves.

Calibrating the Gaussian Copula Model

The implied base correlation curve is generally an increasing function of the upper attachment point.

A distinguishing feature of CDOs and other credit derivatives such as n^{th}-to-default options is that they can be very sensitive to correlation assumptions.

As a risk manager or investor in structured credit, it is very important to understand why equity, mezzanine and super senior tranches react as they do to changes in implied correlation.