

**UTILISATION DES COPULES POUR ANALYSER L'IMPACT DES
DEPENDANCES SUR UN PORTEFEUILLE DE CREDITS**

RAPPORT DE STAGE D'INGENIEUR



CONFIDENTIEL

5 MAI – 22 NOVEMBRE 2002

MAITRES DE STAGES : Mikael BOUVIER
Viven BRUNEL

REMERCIEMENTS

Je tiens à remercier en premier lieu messieurs Kim Ly et Richard Dalaud pour m'avoir accueilli dans leur service et m'avoir offert la possibilité de mener cette étude.

Il m'est difficile de trouver les mots pour exprimer mon énorme gratitude vis-à-vis de Mikael Bouvier et Vivien Brunel qui m'ont encadré, conseillé et dirigé pendant ces six mois. J'ai particulièrement apprécié leur grande expérience du métier et leur professionnalisme à toute épreuve.

Je terminerai en remerciant tous les membres de l'équipe RISQ/RCE et les stagiaires de l'été 2002. J'ai été très sensible à leur haut degré de compétence et de manière générale à leur constante disponibilité.

SOMMAIRE

<u>INTRODUCTION</u>	4
<u>I. STRUCTURE DE DÉPENDANCE ET GESTION DE PORTEFEUILLES DE CRÉDITS</u>	6
<u>Résumé</u> :	6
<u>DÉPENDANCE</u> :	9
<u>CORRÉLATION</u>	10
<u>Tau de Kendall (corrélation de rang)</u> :	10
<u>LES COPULES</u>	14
<u>Définition</u>	14
<u>Copules Archimédiens bivariés</u>	16
<u>La notion de “tail dependence” ou “dépendance de queue”</u>	20
<u>Choix du bon copule</u>	22
<u>Copules multidimensionnels</u>	25
<u>Copules de la famille H_n</u>	25
<u>Copule Gaussien</u>	27
<u>Copule t de Student</u>	27
<u>ALGORITHMES</u>	31
<u>CADRE FINANCIER</u>	36
<u>Portefeuille homogène – modèle KMV</u>	36
<u>Le modèle de Merton (publié en 1974)</u>	36
<u>Modèle de la firme à un facteur</u>	37
<u>Capital économique</u>	40
<u>Capital marginal</u>	41
<u>Contribution</u>	41
<u>SIMULATION</u>	44
<u>II. CALCUL DES PROBABILITÉS DE DEFAULT JOINT</u>	49
<u>Résumé</u> :	49
<u>PROBABILITE DE DEFAULT JOINT</u>	51
<u>Barrière affine</u> :	51
<u>Barrière affine par morceaux</u>	52
<u>Calcul des probabilités de défaut joint</u> :	55
<u>DEROULEMENT DU STAGE</u>	56
<u>BILAN ET PERSPECTIVES</u>	59
<u>BIBLIOGRAPHIE</u>	60
<u>ANNEXES</u>	62
<u>Copules archimédiens à un paramètre et leur générateur</u>	63
<u>Tau de kendall pour les copules archimédiens à un paramètre</u>	65
<u>Calcul de Andersen</u>	65
<u>Calcul de Andersen</u>	66
<u>Le modèle APT</u>	67

INTRODUCTION

Dans le domaine des activités de crédits, la concurrence interbancaire est aujourd'hui telle que la recherche d'une meilleure rentabilité est devenue un objectif prioritaire pour les banques.

Pour ce faire, depuis 1997, la Société Générale met en place un projet RAROC – Risk Adjusted Return On Capital – dont l'objectif est de mesurer le capital apporté par l'actionnaire en couverture du risque lié aux activités de crédit et d'en estimer la rentabilité. Le département RISQ/RCE (Rentabilité Capital Economique) est en charge de ce projet, ainsi que de l'estimation du capital économique associé aux autres types de risque (marché, opérationnel).

RISQ/RCE a développé des modèles et outils informatiques pour atteindre ces objectifs. Il s'agit essentiellement, au final :

- d'outils de notation ("rating"), qui permettent de mesurer la qualité financière d'un (éventuel) client,
- d'outils d'aide à la structuration, qui permettent d'ajuster la structure d'un crédit dans le souci de la rentabilité pour l'actionnaire,
- d'un outil de gestion de portefeuilles : estimation de la distribution des pertes futures, et en particulier de la perte moyenne et du capital économique, pour l'ensemble du portefeuille de crédit de la banque ou pour tout sous-portefeuille. L'outil permet également l'estimation de la rentabilité et de la valeur créée.

La modélisation d'un portefeuille de crédit passe inévitablement par la dépendance qui existent entre les différentes firmes du portefeuille. En effet, le risque lié à un crédit est d'autant plus élevé qu'il est fortement corrélé avec l'ensemble de ceux déjà présents en portefeuille.

On évalue souvent la dépendance entre deux firmes à l'aide de leur corrélation. Or cet indicateur ne permet pas de modéliser la dépendance d'une part et dépend des marges d'autre part.

Les copules, qui sont une « fonction de dépendance », permettent de séparer la structure de dépendance et les marges. En effet le risque d'un portefeuille de crédit ne dépend que du copule et pas des loi marginales.

Tout comme ce rapport, mon stage s'est divisé en deux missions.

La première mission, consistait à étudier les copules puis à les utiliser dans la gestion d'un portefeuille de crédit.

La deuxième mission consiste à calculer les probabilités de défaut joint de deux entreprises, en utilisant une approche type « marché » qui s'intéresse aux cours des actions.

Les copules étant un sujet complexe, nous n'avons pas voulu faire de ce rapport un cours théorique sur les copules mais plutôt un outil facilement utilisable par l'entreprise et compréhensible par un non initié.

Dans cette optique, nous ne démontrerons pas les théorèmes que nous utiliserons, mais les résultats utiles pour une application directe des copules.

Le lecteur souhaitant avoir une étude plus théorique sur le sujet pourra se reporter aux différentes références présentes dans le rapport.

I. STRUCTURE DE DEPENDANCE ET GESTION DE PORTEFEUILLES DE CREDITS

Résumé :

En finance, la dépendance est stratégique dans de nombreux domaines comme le risque de crédit, le risque pays, les dérivées de crédit, etc.

Il est courant de représenter la dépendance en n'utilisant que le coefficient de corrélation linéaire, or cet usage introduit forcément une erreur car celui-ci ne reflète pas la structure de dépendance de deux variables aléatoires, mais seulement un résumé potentiellement dangereux.

Pour résoudre ce problème, nous allons donc utiliser les copules, qui sont un outil prometteur pour évaluer la dépendance. Le copule est une écriture de la fonction de répartition jointe qui permet de disposer de toute l'information sur la structure de dépendance.

Durant mon stage, nous avons étudié et utilisé les copules pour analyser l'impact de la dépendance sur un portefeuille de crédit.

En effet lorsqu'une banque détient un portefeuille de crédit, elle cherche à évaluer l'influence que peut avoir ce portefeuille sur sa santé financière. Pour cela, elle calcule le capital économique du portefeuille, qui n'est d'autre qu'un quantile des pertes.

Pour simplifier notre étude, nous considérerons que ce portefeuille est homogène, c'est à dire que tous les crédits qui le composent sont de même exposition, ont la même probabilité de faire défaut et sont de même corrélation. Nous disposons d'une formule analytique pour calculer le capital économique de ce portefeuille, dans ce cas.

Lorsque l'on ajoute un deuxième portefeuille (qui sera lui aussi considéré comme homogène) au premier, le calcul du capital économique du portefeuille total devient beaucoup plus complexe. Il faut en effet tenir compte lors du calcul du capital du fait que les deux portefeuilles ne soient pas indépendants. Ce calcul est le plus souvent réalisé à l'aide de simulations de Monte Carlo.

L'une des méthodes couramment utilisée pour obtenir le capital économique de deux portefeuilles consiste à calculer dans un premier temps le coefficient de corrélation linéaire

des deux portefeuilles, puis à générer par simulation dans un second temps deux variables aléatoires ayant cette même corrélation et représentant les distributions de pertes des deux portefeuilles.

L'enjeu financier de cette étude pour l'entreprise est énorme : en effet le portefeuille de crédit de la Société Générale se compte en dizaines de milliards d'euro et le capital économique en milliards d'euro, un changement dans la structure de dépendance peut donc avoir un impact conséquent sur le bilan de la banque.

En plus de l'enjeu quantitatif, cette étude présente également des intérêts qualitatifs pour la banque : elle permettra d'améliorer sa maîtrise des risques, et pourra s'étendre avec de nouvelles applications. En effet, bien que nous ayons étudié les copules dans le cadre d'un sujet précis, nos travaux restent néanmoins utilisables dans une multitude d'autres domaines financiers utilisant comme outil la dépendance (banque, assurance).

Ma mission consiste donc à :

- dans un premier temps, étudier les différents copules bivariés et plus précisément les archimédiens présents dans la littérature, notamment concevoir des algorithmes pour les simuler et les identifier
- réaliser la même étude pour les copules multivariés
- obtenir un encadrement de la VaR (Value at Risk) en fonction de la structure de dépendance

Application à un cas réel :

- identifier le copule multivarié sur les 8 sous portefeuilles du portefeuille de crédits de la Société Générale et l'utiliser pour calculer le capital économique ainsi que les contributions
- optimiser un portefeuille de crédits à l'aide des copules : à partir d'un portefeuille initial, déterminer la part d'un autre portefeuille à acquérir pour optimiser l'allocation de capital

Résultat :

- élaboration d'un rapport sur les copules avec les principaux algorithmes pour simuler et identifier les copules
- identification du copule bivarié sur des distributions de pertes provenant du simulateur de la Société Générale
- optimisation d'un portefeuille de crédits composé de deux sous portefeuilles homogènes à partir de leur structure de dépendance

Mon rôle était donc d'étudier les copules et de les rendre compréhensibles et utilisables par le reste de l'équipe dans le but d'une possible implémentation dans le simulateur de la Société Générale.

Difficultés et perspectives :

Il est pratiquement impossible de déterminer la structure exacte de dépendance d'une série de donnée, on cherche à approcher le plus possible cette structure à l'aide d'une structure connue tels que les copules archimédiens, Gaussien ou bien Student. C'est ce que l'on fait lorsqu'on utilise un test d'identification : on ne détermine pas le copule empirique, mais le copule connu qui s'en rapproche le plus.

Les principales difficultés résident dans le manque de copules multivariés (à ce jour, seulement le Gaussien et le Student sont utilisables), ainsi que le manque de données disponible pour identifier les copules. Sur certains crédits par exemple, on ne dispose que d'une centaine d'observations, ce qui est insuffisant pour utiliser les différents tests.

Le copule est un outil récent, les promesses de ce domaine est donc énorme, autant sur les types de copules que sur les tests. En effet le test utilisé pour discriminer le copule Gaussien et identifier le Student ne date que de Juin 2002, ce qui laisse imaginer les progrès possibles à venir.

Le passage du copule Gaussien vers le copule de Student (qui tend vers le gaussien quand le degré de liberté tend vers l'infini) serait déjà une grande avancée en finance. En effet, J.P. Morgan et Lehmann Brothers sont les seuls organismes financiers suspectés d'utiliser les copules (et plus particulièrement le copule de Student) pour pricer leurs dérivés de crédit.

Je pense donc que les copules seront largement utilisés par les banques et qu'ils constituent un domaine d'étude à ne pas négliger. Les copules sont donc susceptibles d'avoir un impact sur la finance à long terme.

DEPENDANCE :

Lorsque l'on évoque le concept de dépendance, le premier mot qui vient généralement à l'esprit est celui de coefficient de corrélation linéaire de Pearson. Ce coefficient a pourtant ces limites car il sous entend que le monde est elliptique (distributions Normales ou Student multivariées).

Il s'agit d'une mesure symétrique de dépendance linéaire alors que certaines dépendances ne le sont pas.

Exemple

Les variables aléatoires X de lois normales centrées réduites, et X^2 , bien que dépendantes l'une de l'autre (« comonotomes »), présentent une corrélation nulle au sens du coefficient de Pearson.

Nous présenterons un autre coefficient de corrélation, à savoir le coefficient de Kendall, qui présente les avantages de demeurer inchangé sous l'hypothèse d'une transformation strictement croissante des variables aléatoires et d'être un estimateur robuste du coefficient de corrélation linéaire de Pearson pour des distributions elliptiques.

Une mesure de dépendance doit en fait nous permettre de nous faire une idée de la structure de dépendance entre deux variables aléatoires.

En toute rigueur, la dépendance entre deux variables est fonctionnelle, via, au choix :

- La densité jointe
- La fonction de répartition jointe
- La transformée de Laplace ou de Fourier
- Les 2 lois conditionnelles

Le copule est une écriture « uniformisée », c'est à dire indépendante des marges, de la fonction de répartition jointe de X et Y . Elle est d'une importance cruciale en finance, car ce sont les dépendances qui nous intéressent dans l'étude des risques.

Nous examinerons également les notions de “tail dependence” ou de dépendance de queue, très importantes dans l'étude des dépendances entre les valeurs extrêmes et directement associées aux copules.

CORRELATION

Le tau de Kendall est une mesure de la dépendance entre deux variables aléatoires, mais contrairement au coefficient de corrélation linéaire, il ne dépend pas des lois des deux variables étudiés mais que de leur ordre. Il est d'une importance fondamentale dans l'étude des copules, car il existe une relation analytique entre le taux de Kendall et le copule.

Tau de Kendall (corrélacion de rang) :

Définition : le Tau de Kendall entre les variables aléatoires X et Y est égal à la différence entre la probabilité de concordance et la probabilité de non concordance.

Pour les variables (X, Y) , le Tau de Kendall empirique est donné par :

$$\hat{\tau} = \frac{c-d}{c+d}$$

où c est le nombre de paires concordantes et d le nombre de paires non concordantes.

Pour un échantillon de n points (x_i, y_i) , il y a $n(n-1)/2$ paires de points $((x_i, y_i), (x_j, y_j))$ avec $i < j$. La paire $((x_i, y_i), (x_j, y_j))$ est concordante si $(x_i - x_j)(y_i - y_j) > 0$ et non concordante si $(x_i - x_j)(y_i - y_j) < 0$.

$$\hat{\tau} = \frac{\#\{(i, j) : (x_i - x_j)(y_i - y_j) > 0, i < j\} - \#\{(i, j) : (x_i - x_j)(y_i - y_j) < 0, i < j\}}{n(n-1)/2}$$

propriétés :

- Pour toutes fonctions T_1 et T_2 strictement croissantes de $R \rightarrow R$ on a :

$$\tau(X, Y) = \tau(T_1(X), T_2(Y))$$

- Soient les variables aléatoires X, Y de loi normale (ou elliptique) de coefficient de corrélation linéaire de Pearson ρ , alors on a la relation suivante :

$$\tau = \frac{2}{\pi} \arcsin \rho \quad (\text{cf. [7]})$$

La première propriété sera de grande importance dans l'utilisation des copules, et la deuxième permettra d'utiliser le Tau de Kendall comme estimateur du coefficient de corrélation linéaire ρ . En effet le Tau de Kendall est un estimateur plus robuste que l'estimateur standard.

Pour illustrer cela, nous allons simuler 3000 échantillons indépendants de tailles 90, de loi Student-3 bivarée de coefficient de corrélation linéaire 0.5. Puis nous allons estimer ρ avec les deux différents estimateurs.

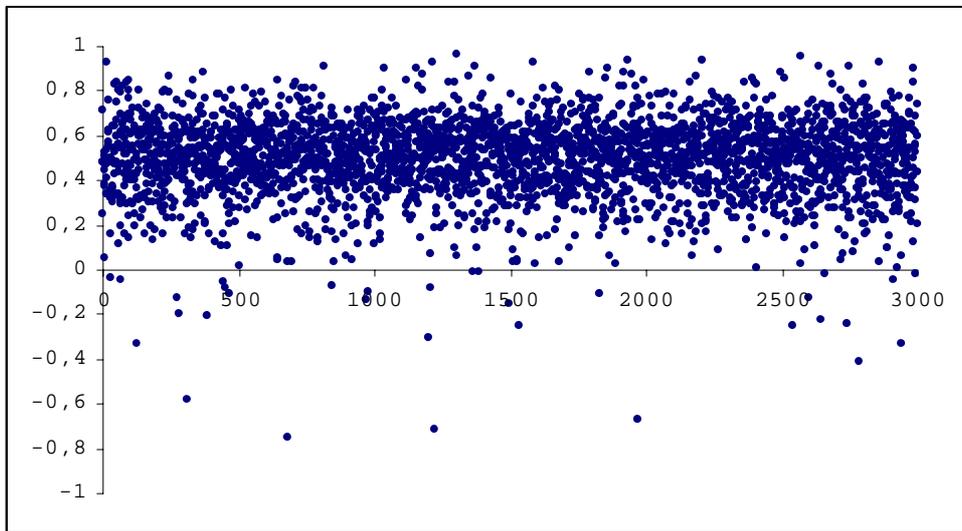
Estimateur standard :

$$\hat{\rho} = \frac{n \sum_i X_i Y_i - \sum_i X_i \sum_i Y_i}{\sqrt{n \sum_i X_i^2 - (\sum_i X_i)^2} \cdot \sqrt{n \sum_i Y_i^2 - (\sum_i Y_i)^2}}$$

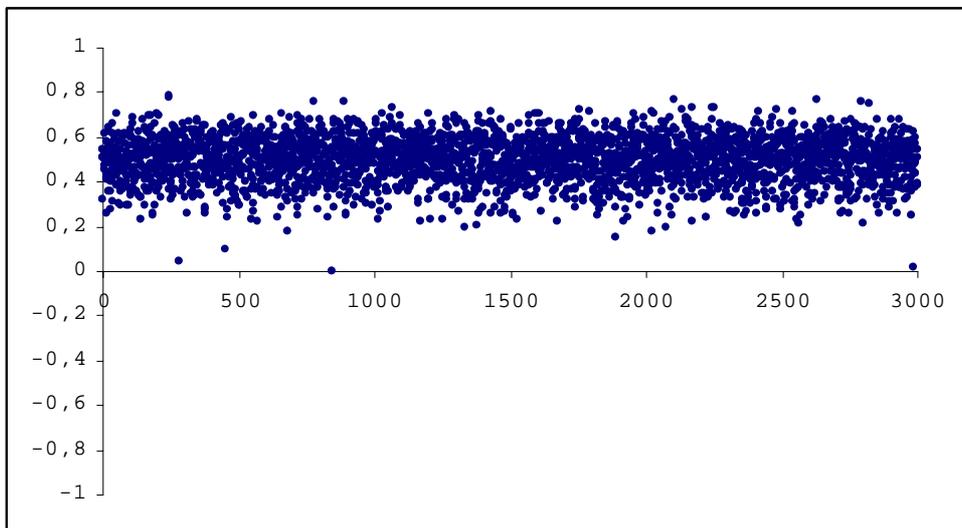
Estimateur de Kendall :

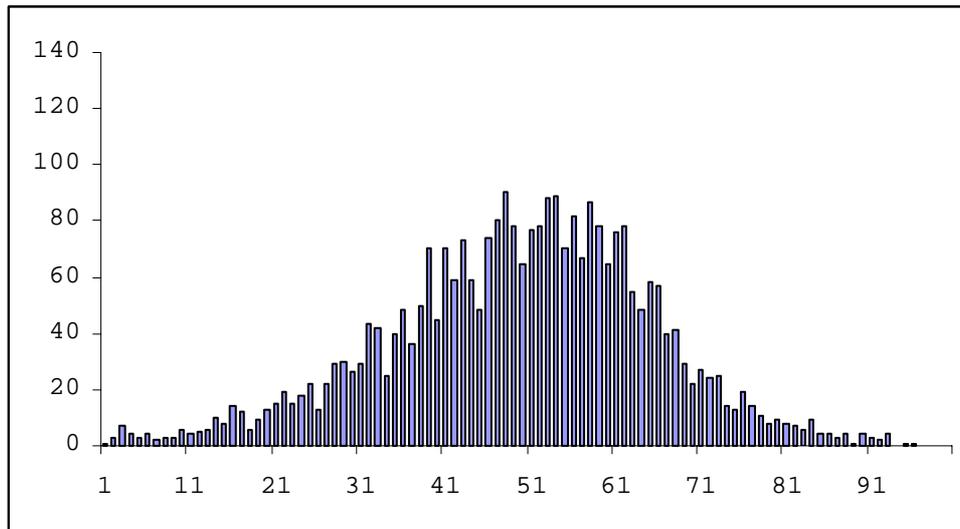
$$\hat{\rho} = \sin\left(\frac{\pi \cdot \hat{\tau}}{2}\right)$$

Estimateur standard



Estimateur de Kendall

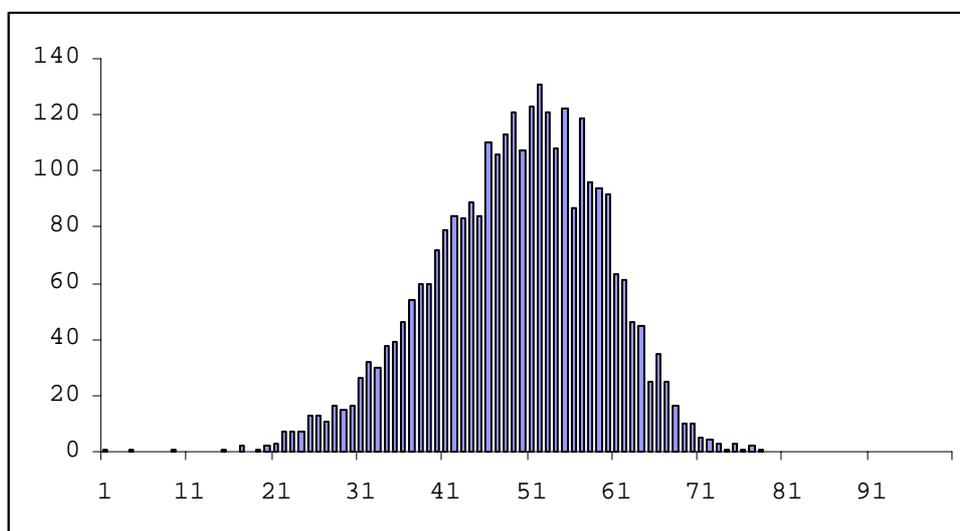




Histogramme 1 : estimateur standard, en abscisse le coefficient de corrélation linéaire en % et en ordonnée le nombre de réalisations

Moyenne : 0.4944
 Ecart type : 0.17

Estimateur de Kendall



Histogramme 2 : estimateur de Kendall, en abscisse le coefficient de corrélation linéaire en % et en ordonnée le nombre de réalisations

Moyenne : 0.4971
 Ecart type : 0.10

Conclusion :

Il est donc évident au vu de ces résultats que l'estimateur de Kendall est plus précis que l'estimateur standard.

Il sera donc utilisé comme estimateur du coefficient de corrélation de Pearson pour les lois elliptiques, et pour évaluer la dépendance gaussienne pour les autres types de corrélations.

Mes recherches ont donc permis de trouver une utilisation nouvelle du tau de Kendall pour l'équipe RISK/RCE.

Maîtrisant le tau de Kendall, nous allons maintenant exposer les copules, qui sont un outil mathématique qui permet de modéliser la dépendance et qui utilise ce coefficient de corrélation de rang.

LES COPULES

(cf. [3] [14])

La dépendance entre les variables aléatoires réelles X_1, \dots, X_n est totalement décrite par leur fonction de répartition jointe

$$F(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n]$$

Considérons les variables aléatoires X_1, \dots, X_n de fonctions de répartition continue F_1, \dots, F_n .

Ceci permet d'introduire le copule

$$\begin{aligned} F(x_1, \dots, x_n) &= P[F_1(X_1) \leq F_1(x_1), \dots, F_n(X_n) \leq F_n(x_n)] \\ &= C(F_1(x_1), \dots, F_n(x_n)) \end{aligned}$$

Le copule est donc une écriture de la fonction de répartition jointe des variables aléatoires X_1, \dots, X_n indépendante de leurs distributions (où marges)

En effet, on rappelle le résultat fondamental suivant : $F_i(X_i)$ est de loi uniforme.

Définition

Le copule C est une application de $[0,1]^n \rightarrow [0,1]$ qui vérifie ces trois propriétés :

1. $C(x_1, \dots, x_n)$ est une application croissante pour toutes les composantes x_i
2. $C(1, \dots, 1, x_i, 1, \dots, 1) = x_i \quad \forall i \in \{1, \dots, n\}$, $x_i \in [0,1]$
3. Pour tout $(a_1, \dots, a_n), (b_1, \dots, b_n) \in [0,1]^n$ avec $a_i \leq b_i$ on a :

$$\sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1 + \dots + i_n} C(x_{1i_1}, \dots, x_{ni_n}) \geq 0$$

$$\text{où } x_{j1} = a_j \text{ et } x_{j2} = b_j \quad \forall j \in \{1, \dots, n\}$$

En deux dimensions, on a :

Un copule C de dimension 2 est une application de $[0,1]^2 \rightarrow [0,1]$ qui vérifie

1. $C(u,0) = C(0,v) = 0 \quad \forall u, v \in [0,1]$
2. $C(u,1) = u$ et $C(1,v) = v \quad \forall u, v \in [0,1]$
3. Pour tout $(u_1, v_1), (u_2, v_2) \in [0,1]^2$ avec $u_1 \leq u_2$ et $v_1 \leq v_2$ on a :

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

Il existe un lien entre la densité jointe et le copule :

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \cdot c_{X,Y}(F_X(x), F_Y(y))$$

où
$$c_{X,Y}(u,v) = \frac{\partial^2 C_{X,Y}(u,v)}{\partial u \partial v}$$

Théorème:

Soient deux variables aléatoires continues X_1 et X_2 de copule $C(X_1, X_2)$. Si h_1 et h_2 sont deux fonctions strictement croissantes, alors

$$C(h_1(X_1), h_2(X_2)) = C(X_1, X_2)$$

Donc, l'application de transformations croissantes ne modifie pas le copule mais seulement les marges.

- Copules C^- , C^\perp , C^+ :

C^- et C^+ sont les copules extrémaux de Fréchet :

$$\left\{ \begin{array}{l} C^-(u_1, u_2) = \max(u_1 + u_2 - 1, 0) \\ C^+(u_1, u_2) = \min(u_1, u_2) \end{array} \right.$$

C^\perp est le copule indépendant :

$$C^\perp(u_1, u_2) = u_1 u_2$$

Définition : Soient C_1 et C_2 deux copules. On dit que C_1 est plus petit que C_2 et on note $C_1 \prec C_2$ si et seulement si $C_1(u_1, u_2) \leq C_2(u_1, u_2)$ pour tout $(u_1, u_2) \in [0,1]^2$

Quel que soit le copule C , nous avons toujours la relation

$$C^- \prec C \prec C^+$$

Nous étudierons d'abord les copules archimédiens à 2 dimensions et ensuite nous généraliserons à ceux de n dimensions (avec $n > 2$)

Copules Archimédiens bivariés

Soit Ψ la classe des applications définies sur $[0,1]$ à valeurs dans $[0, +\infty[$, continues, strictement décroissantes et convexes, pour lesquelles $\varphi(1) = 0$

On désigne par φ^{-1} la fonction inverse de φ

Alors la fonction

$$C_\varphi(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \quad \text{pour } u, v \in]0,1] \quad (1)$$

est le copule archimédien de générateur φ .

Famille	Générateur $\varphi(t)$	Espace de Définition de α	Copule bivarié $C_\varphi(u, v)$
Indépendant	$-\ln t$	Pas applicable	uv
Clayton (1978), Cook-Johnson (1981), Oakes (1982)	$t^{-\alpha} - 1$	$\alpha > 1$	$(u^{-\alpha} + v^{-\alpha} - 1)^{-1/\alpha}$
Gumbel (1960), Hougaard (1986)	$(-\ln t)^\alpha$	$\alpha \geq 1$	$\exp\left\{-\left[(-\ln u)^\alpha + (-\ln v)^\alpha\right]^{1/\alpha}\right\}$
Frank (1979)	$\ln \frac{e^{\alpha t} - 1}{e^\alpha - 1}$	$-\infty < \alpha < \infty$	$\frac{1}{\alpha} \ln \left(1 + \frac{(e^{\alpha u} - 1)(e^{\alpha v} - 1)}{e^\alpha - 1}\right)$

Tableau 1 : Quelques copules archimédiens à un paramètre et leur générateur

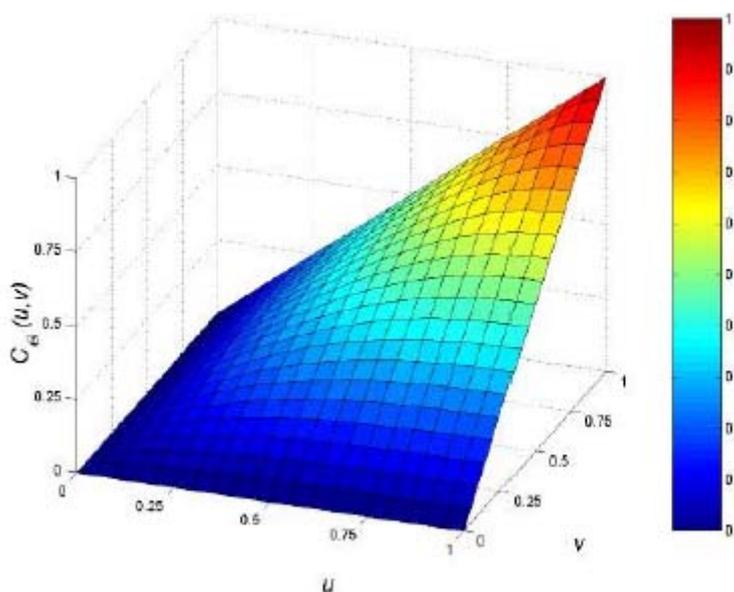


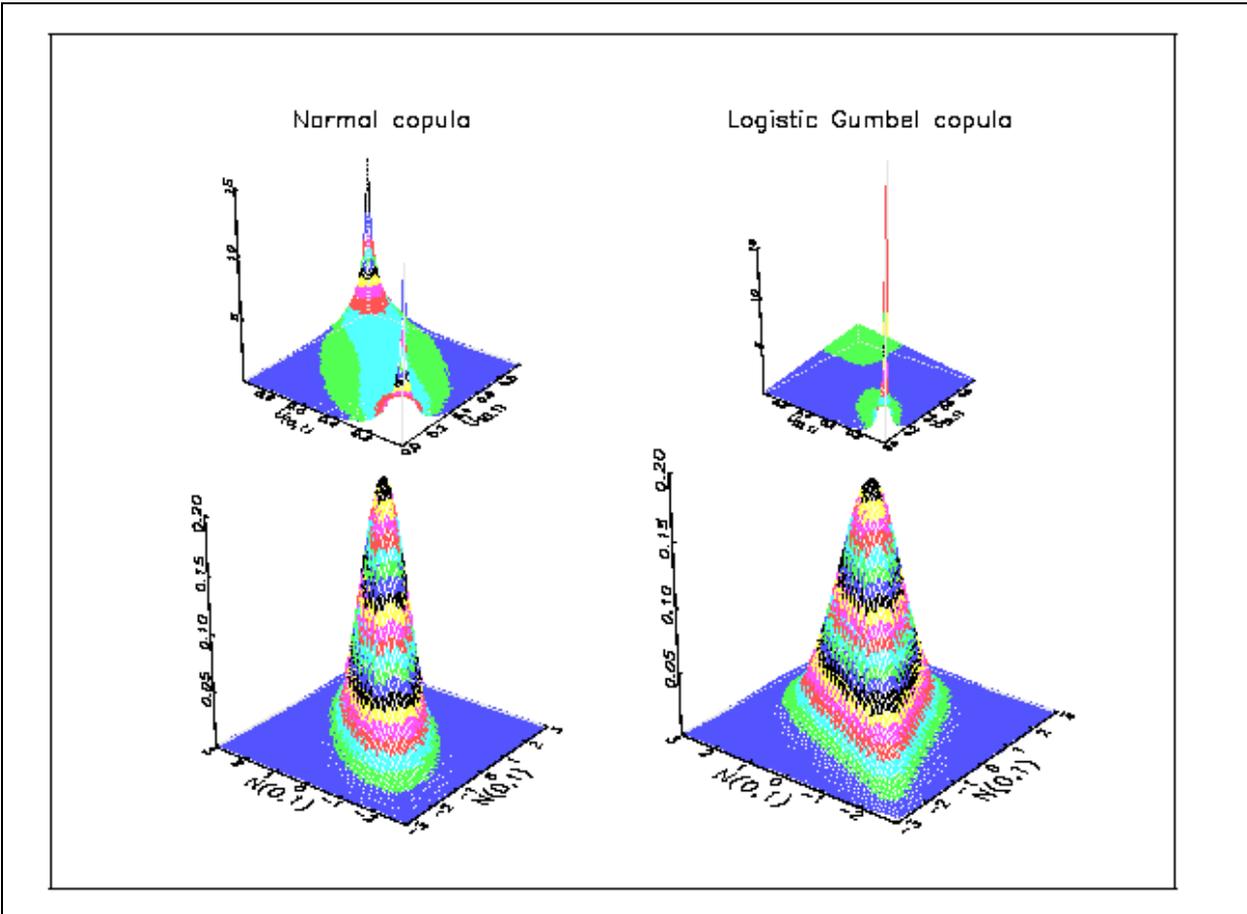
Figure1 : Copule de Gumbel, $C_{Gumbel}(u, v)$ pour $\alpha = 2$

La représentation des copules en 3D ne présente pas de grand intérêt, en effet elle est illisible. Les courbes iso-densités quand à elles, fournissent l'informations sur la structure de dépendance. De plus on utilisera ces courbes dans un plan Normale-Normal (cf. figure 3) pour mieux visualiser leur valeurs extrêmes ce qui est plus difficile dans le plan Uniforme-Uniforme.

Exemple, pour le copule de Gumbel, les courbes de niveaux reflètent l'asymétrie de ce copule (cf. figure 3 : copule de Gumbel). En effet, les petites valeurs sont fortement corrélées (forme en « pique » du copule de Gumbel dans les petites valeurs) tandis que les grandes ne le sont pas.

Il existe des paramètres qui permettent d'estimer la dépendance des valeurs extrêmes suivant le copule.

Exemple de courbes iso-densité de copule.



Le tau de Kendall ne dépendant pas des marges mais que de la structure de dépendance, il existe donc un lien entre le paramètre α et τ pour les copules archimédiens (cf. annexe A).

Exemple : Simulation de 2000 variables aléatoires de même tau de Kendall et de marges gaussienne ($\rho = 60\%$)

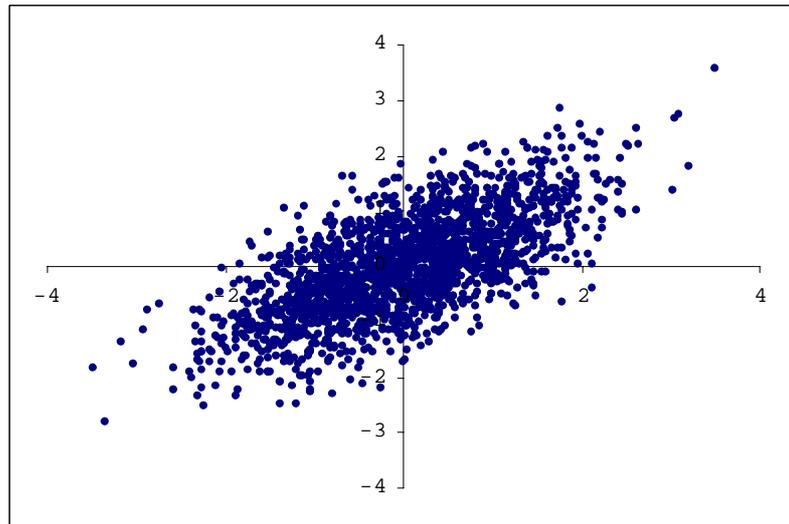


Figure 4 : copule gaussien

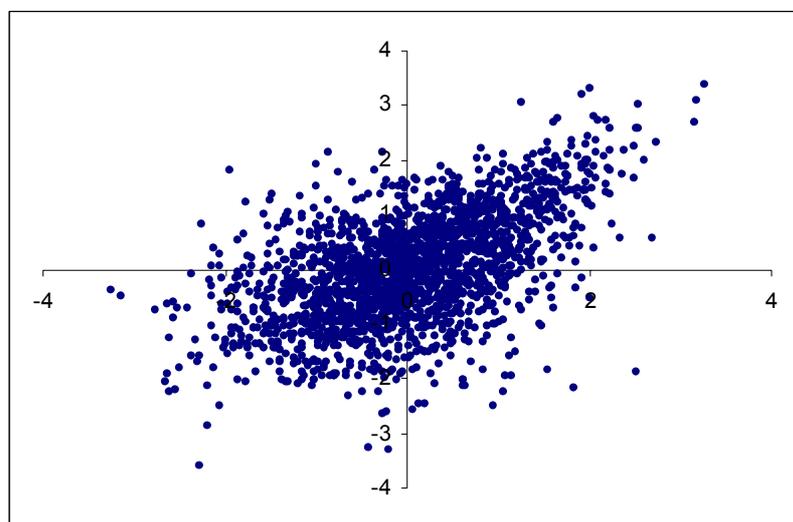


Figure 5 : copule Gumbel

La notion de “tail dependence” ou “dépendance de queue”

Cette notion est très importante dans l'étude de la dépendance asymptotique entre deux variables aléatoires. Cela va nous permettre de voir le niveau de dépendance dans les valeurs extrêmes (upper tail dependence) et dans les valeurs petites (lower tail dependence). Ce concept sera totalement basé sur celui des copules.

Upper tail dependence.

L'objet est donc l'étude de la dépendance dans la queue commune de la distribution bivariée. Prenons donc deux variables aléatoires continues X et Y ayant pour fonction de distribution respectives F_X et F_Y .

Le coefficient d' "upper tail dependence" de X et Y est défini

$$\lambda_u = \lim_{u \rightarrow 1^-} P(Y > F_Y^{-1}(u) | X > F_X^{-1}(u)) = \lim_{u \rightarrow 1^-} \frac{C(u, u) - 2u + 1}{1 - u}$$

si toutefois cette limite $\lambda_u \in [0, 1]$ existe.

La quantité λ est une fonction du copule et est donc invariante par transformation croissante.

Si $\lambda \in (0, 1]$, il existe alors une dépendance asymptotique.

Si $\lambda = 0$, on dit qu'il y a indépendance asymptotique.

Le copule la plus fréquemment donnée en exemple dans la littérature pour illustrer ce concept est celui de Gumbel

Exemple pour le copule de Gumbel

$$C_\alpha^{Gu}(u, u) = u^{2^{1/\alpha}} \quad (\text{Rappel : } \alpha \geq 1)$$

donc

$$\lambda_u = \lim_{u \rightarrow 1} \frac{1 - 2u + u^{2^{1/\alpha}}}{1 - u} = 2 - \lim_{u \rightarrow 1} \frac{u^{2^{1/\alpha}} - 1}{1 - u}$$

On obtient donc :

$$\lambda_u = 2 - \lim_{u \rightarrow 1} \frac{\partial \left(u^{2^{1/\alpha}} \right)}{\partial u} = 2 - 2^{1/\alpha}$$

La dépendance augmente vers 1 au fur et à mesure que α tend vers l'infini.

A l'inverse, le copule Gaussien par exemple ne présente aucune dépendance asymptotique.

Exemple pour le copule Student :

Pour le copule de Student, nous avons une relation entre la dépendance de queue ("upper tail dependence") et le coefficient de corrélation linéaire de Pearson.

$$\lambda_U = 2 - 2t_{\nu+1} \left(\left(\frac{(\nu+1)(1-\rho)}{(1+\rho)} \right)^{1/2} \right)$$

Nous en déduisons que

$$\lambda_U : \begin{cases} > 0 & \text{si } \rho > -1 \\ 0 & \text{si } \rho = -1 \end{cases}$$

Lower tail dependence

En gardant, les mêmes notations qu'auparavant, le coefficient de "lower tail dependence" de X et Y est défini comme étant

$$\lambda_l = \lim_{u \rightarrow 0^+} P(Y \leq F_Y^{-1}(u) | X \leq F_X^{-1}(u)) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}$$

si cette limite λ_l existe.

Exemple pour le copule de Clayton

$$C_\alpha^{cl}(u, u) = \left(2u^{-1/\alpha} - 1 \right)^{-\alpha}$$

donc

$$\lambda_l = \lim_{u \rightarrow 0} \frac{1}{u \left(2u^{-1/\alpha} - 1 \right)^\alpha} = \lim_{u \rightarrow 0} \frac{1}{\left(2 - u^{1/\alpha} \right)^\alpha} = \frac{1}{2^\alpha}$$

Le copule de Clayton possède donc un coefficient de "lower tail dependence".

$\lambda_l = 0$ pour Gumbel, ce qui confirme l'interprétation des iso densités.

Choix du bon copule

Après avoir défini le concept de copule, on peut se demander comment déterminer la structure de dépendance d'une série de données. Il existe plusieurs méthodes qui permettent de faire le tri entre les différents copules archimédiens à un paramètre.

Dans cette partie, nous allons exposer les deux méthodes qui nous paraissent être les meilleures. Il s'agit de la méthode utilisant la fonction de répartition du copule et de la méthode du maximum de vraisemblance.

1. Fonction de répartition du copule (cas archimédien)

Après avoir calculé le Tau de Kendall de notre série de données, on estime le paramètre du copule archimédien que l'on souhaite tester (cf. annexe A : relation entre le tau de Kendall et paramètre des copules archimédiens)

La fonction de répartition d'un copule archimédien bivarié est unique et est donnée par :

$$K_c(t) = P[C(U, V) \leq t] = t - \frac{\varphi(t)}{\varphi'(t)} \quad U, V \text{ v.a. de loi uniforme}$$

On calcule la fonction de répartition empirique du copule et on la compare graphiquement (ou bien avec un test de Kolmogorov – Smirnov) avec la fonction de répartition théorique K_c

Fonction de répartition empirique :

$$z_i = \frac{\text{nombre de paires } \{x_j, y_j\} \text{ telles que } x_j < x_i \text{ et } y_j < y_i}{n-1}$$

$$K_n(z) = \frac{\{\text{nombre de } z_i \leq z\}}{n}$$

illustration :

On simule 2000 variables aléatoires uniformes de copule de Gumbel de paramètre 1.7, et on teste cette méthode avec le copule de Gumbel et le copule de Clayton.

On calcule le tau de Kendall empirique ($\bar{\tau} = 0.356$) et on en déduit les paramètres du copule de Gumbel ($\alpha = 1.6$) et du copule de Clayton ($\alpha = 1.2$).

Copule de Gumbel : $K_{Gumbel}(t) = t(1 - \frac{\ln t}{\alpha})$

Copule de Clayton : $K_{Clayton}(t) = t + \alpha t(1 - t^{1/\alpha})$

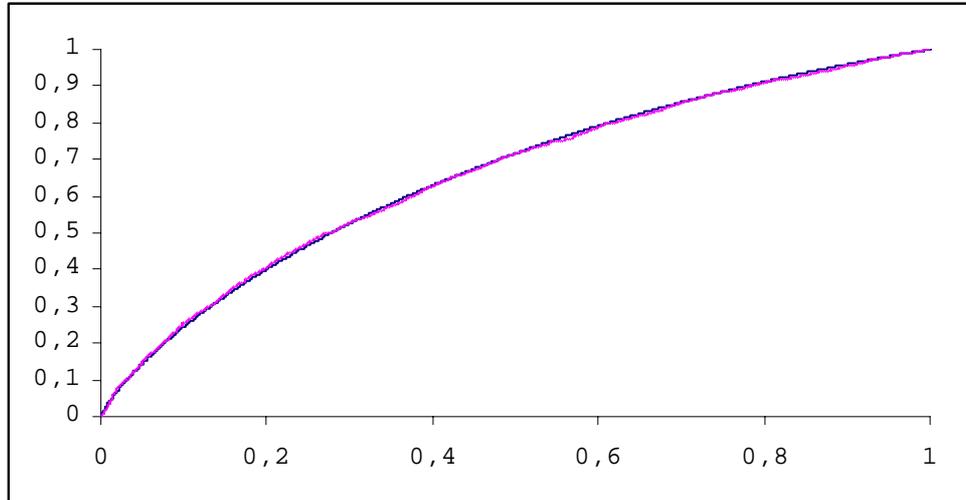


Figure 6 : fonction de répartition empirique et fonction de répartition théorique pour le copule de Gumbel

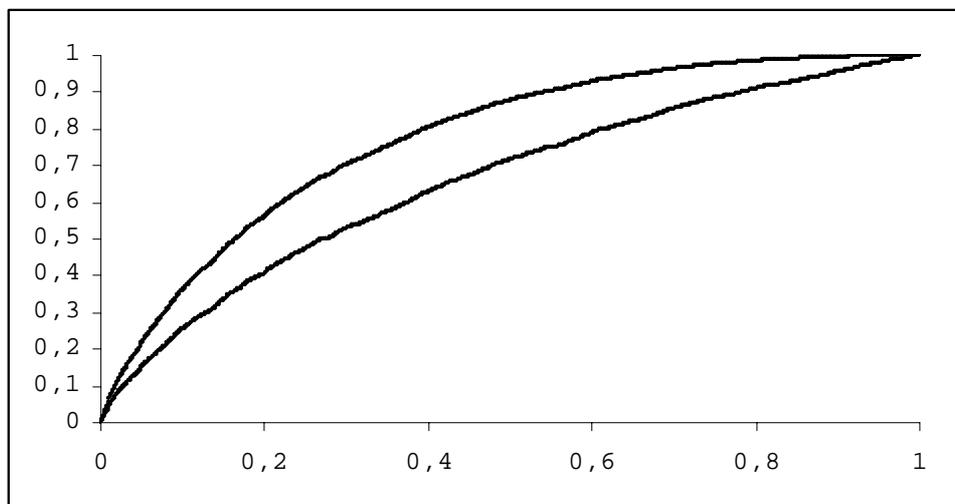


Figure 7 : fonction de répartition empirique et fonction de répartition théorique pour le copule de Clayton

Dans cette exemple, il est donc évident que le copule de Gumbel est plus proche que le copule de Clayton.

Cette méthode est graphique, elle présente donc l'avantage d'être visuelle mais ayant estimé le paramètre du copule elle est moins précise que la méthode utilisant le maximum de vraisemblance, qui est par ailleurs plus facile programmer.

2. Maximum de vraisemblance

Cette méthode consiste, étant donné un échantillon de valeurs x_1, x_2, \dots, x_n à prendre comme estimation de α la valeur de α qui rend maximale la log-vraisemblance (cf. Algorithme 5):

$$\log L(x_1, x_2, \dots, x_n; \alpha)$$

Justification mathématique : cf. [17]

On utilise la même simulation que pour la 1^{ère} méthode, et on la teste avec la 2^{ème}.

Copule	Estimation de α	log-vraisemblance
Gumbel	1.62	426
Clayton	0.77	183
Ali	1	151
Frank	4.14	336
Gaussien	0.55	352

Tableau 2 : maximum de vraisemblance pour 5 différents copules archimédiens, (Cf. algorithme 5 page 27)

C'est donc le copule de Gumbel, qui « fitte » le mieux nos données.

Remarque : on trouve une valeur de α plus proche pour la 2^{ème} méthode.

Conclusion :

Après plusieurs tests, on en déduit que la méthode du maximum de vraisemblance est plus simple à mettre en oeuvre et donne de meilleurs résultats pour les échantillons de grande taille. De plus cette méthode n'estime pas de paramètre, cela lui confère une plus grande robustesse. En effet le lien entre le paramètre du copule et le tau de Kendall n'est pas toujours simple. Elle présente aussi l'avantage d'être utilisable pour les copules multidimensionnels, tandis que la première ne l'est que pour les copules archimédiens bivariés.

On utilisera donc cette méthode dans la suite du rapport pour fitter un copule.

Copules multidimensionnels

Après avoir introduit les copules bivariés nous allons généraliser les copules archimédiens à n dimensions, puis présenter les copules multidimensionnels les plus courants qui sont les copules Gaussien et Student, le but du stage étant de fitter un copule multivarié sur les 8 sous-portefeuilles de crédit de la Société Générale, chaque sous portefeuille étant composé de crédits de même qualité (8 ratings)

Copules de la famille H_n

Pour simplifier, nous étudierons les lois multidimensionnelles dont les marges sont uniformes sur le carré unité $[0,1]$.

Une première généralisation intuitive des copules archimédiens de la forme (1) à l'ordre $n \geq 3$ pourrait être de la forme :

$$C_\varphi(x_1, \dots, x_n) = \varphi^{-1}(\varphi(x_1) + \dots + \varphi(x_n)), \quad 0 \leq x_1, \dots, x_n \leq 1 \quad (2)$$

D'après le théorème de Schweizer et Sklar, l'application de la forme (2) est une fonction de répartition pour tout $n \geq 3$ si et seulement si φ^{-1} est complètement monotone, c'est à dire :

$$(-1)^k \frac{\partial^k}{\partial t^k} \varphi^{-1}(x) \geq 0, \forall k = 1, 2, \dots \quad (3)$$

En pratique, une modélisation de cette forme n'a pas grand intérêt puisque les dépendances entre les sous-ensembles de variables aléatoires de $\{X_1, \dots, X_n\}$ ne sont pas mises en valeurs, seule une dépendance globale est prise en compte.

Pour cette raison nous allons donc étudier une famille, notée H , de copules archimédiens à l'ordre $n \geq 3$, proposée pour la première fois par Joe (1993), qui tient compte de toutes les dépendances entre les variables aléatoires X_1, \dots, X_n .

Si on note H_n le copule archimédien de dimension n du vecteur aléatoire (X_1, \dots, X_n) , alors cette famille est définie par :

$$H_n(x_1, \dots, x_n) = \varphi_{n-1}^{-1}(\varphi_{n-1}(H_{n-1}(x_1, \dots, x_{n-1})) + \varphi_{n-1}(x_n)) \quad (4)$$

où $H_{n-1}(x_1, \dots, x_{n-1}) = \varphi_{n-2}^{-1}(\varphi_{n-2}(H_{n-2}(x_1, \dots, x_{n-2})) + \varphi_{n-2}(x_{n-1}))$ avec $0 \leq x_1, \dots, x_n \leq 1$ et l'application φ_i pour $1 \leq i \leq n-1$ appartient à la classe Ψ

Proposition : Une application H_n pour $n \geq 3$ de type (4) est une fonction de répartition si et seulement si les applications $(-\varphi_{n-1} \circ \varphi_{n-2}^{-1}), \dots, (-\varphi_2 \circ \varphi_1^{-1})$ sont complètement monotones.

Inconvénients de la famille H_n :

Prenons le cas de dimension 3.

$$H_3(x_1, x_2, x_3) = \varphi_2^{-1}(\varphi_2[\varphi_1^{-1}(\varphi_1(x_1) + \varphi_1(x_2))] + \varphi_2(x_3)) = H_2(H_1(x_1, x_2), x_3)$$

$$\text{où } H_1(u, v) = \varphi_1^{-1}(\varphi_1(u) + \varphi_1(v)) \text{ et } H_2(u, v) = \varphi_2^{-1}(\varphi_2(u) + \varphi_2(v))$$

donc :

$$H_3(x_1, x_2, 1) = H_2(H_1(x_1, x_2), 1) = H_1(x_1, x_2)$$

$$H_3(x_1, 1, x_3) = H_2(H_1(x_1, 1), x_3) = H_2(x_1, x_3)$$

$$H_3(1, x_2, x_3) = H_2(H_1(1, x_2), x_3) = H_2(x_2, x_3)$$

On en déduit donc :

Le couple (x_1, x_2) a comme structure de dépendance un copule de générateur φ_1 .

Les couples (x_1, x_3) et (x_2, x_3) ont comme structure de dépendance un copule de générateur φ_2 .

En généralisant au degré n on obtient donc :

Structure de dépendance : Copule de générateur	couples
φ_1	(x_1, x_2)
\vdots	
φ_i	$(x_1, x_{i-1}), (x_2, x_{i-1}) \dots (x_{i-2}, x_{i-1})$
\vdots	
φ_{n-1}	$(x_1, x_{n-1}), (x_2, x_{n-1}) \dots (x_{n-2}, x_{n-1})$

Il vient immédiatement que lorsque les marges sont gaussiennes, la matrice de corrélation est de la forme

$$\begin{pmatrix} 1 & \rho_1 & \rho_2 & \rho_3 & \cdot & \cdots & \rho_{n-2} & \rho_{n-1} \\ \rho_1 & 1 & \rho_2 & \rho_3 & \cdot & \cdot & \cdot & \cdot \\ \rho_2 & \rho_2 & 1 & \rho_3 & \cdot & \cdot & \cdot & \cdot \\ \rho_3 & \rho_3 & \rho_3 & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot & \cdot & \ddots & \rho_{n-2} & \rho_{n-1} \\ \rho_{n-2} & \cdot & \cdot & \cdot & \cdot & \rho_{n-2} & 1 & \rho_{n-1} \\ \rho_{n-1} & \cdot & \cdot & \cdot & \cdot & \rho_{n-1} & \rho_{n-1} & 1 \end{pmatrix}$$

Les copules multidimensionnels appartenant à la famille H_n sont donc difficilement utilisables compte tenu des énormes contraintes qui portent sur la structure de dépendance des couples (x_i, x_j) .

Dans le cas multidimensionnel, mis à part la famille H_n , il n'existe dans la littérature que deux autres types de copules : le Gaussien et le Student

Le copule de Student a deux paramètres, la matrice de corrélation Σ et le degré de liberté ν , tandis que le Gaussien n'a que Σ .

Ces deux copules sont donc très proches, ils ont le paramètre Σ en commun et le copule de Student converge vers le copule Gaussien quand le degré de liberté tend vers l'infini.

Copule Gaussien

Soit Σ une matrice diagonale définie positive avec $\text{diag} \Sigma = 1$ et Φ_ρ la distribution normale multivariée standard de matrice de corrélation Σ . Le copule Normale est alors définie de la façon suivante :

$$C(u_1, \dots, u_n; \Sigma) = \Phi_\rho(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n))$$

La densité du copule Normale est :

$$c(u_1, \dots, u_n; \Sigma) = \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \zeta^T (\Sigma^{-1} - I) \zeta\right)$$

avec $\zeta_i = \Phi^{-1}(u_i)$ et I la matrice identité de dimension $(n \times n)$

En deux dimensions, le copule Gaussien s'exprime de la façon suivante

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\left\{-\frac{s^2 - 2\rho \cdot st + t^2}{2(1-\rho^2)}\right\} ds dt$$

Copule t de Student

¹ Σ se calcul à l'aide du tau de Kendal des v.a. X_i et X_j . On a la relation $\Sigma_{ij} = \sin\left(\frac{\pi\tau_{ij}}{2}\right)$

Comme pour le copule Gaussien, le copule t (ou copule de Student) est la fonction de dépendance associée à la distribution t multidimensionnelle.

Soit Σ une matrice diagonale définie positive avec $\text{diag}\Sigma = 1$ et $t_{\rho, \nu}$ la distribution de Student multivariée standard à ν degrés de liberté et de matrice de corrélation ρ . Le copule Student est alors définie de la façon suivante :

$$C(u_1, \dots, u_n; \rho) = t_{\rho, \nu}(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_n))$$

La densité du copule t-copule de dimension d et de degré de liberté ν est :

$$c(u_1, \dots, u_d; \nu, \Sigma) = \frac{\Gamma((\nu + d)/2) \cdot [\Gamma(\nu/2)]^{d-1} (1 + y^T \Sigma y / \nu)^{-(\nu+d)/2}}{|\Sigma|^{1/2} [\Gamma((\nu + 1)/2)]^d \prod_{i=1}^d (1 + y_i^2 / \nu)^{-(\nu+1)/2}}$$

où, $y = (y_1, \dots, y_d)$ et $y_i = t_\nu^{-1}(u_i)$

Σ est la matrice de corrélation linéaire

$$\Gamma(x) = \int_0^\infty \exp(-t) t^{x-1} dt$$

$$\Gamma(n+1) = n!$$

$$\Gamma(k + \frac{1}{2}) = (k - \frac{1}{2}) \Gamma(k - \frac{1}{2})$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

En deux dimensions, le copule de Student s'exprime de la façons suivante

$$C_{\nu, \rho}(u, \nu) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \left\{ 1 + \frac{s^2 - 2\rho \cdot st + t^2}{\nu(1-\rho^2)} \right\}^{-(\nu+2)/2} ds dt$$

exemple :

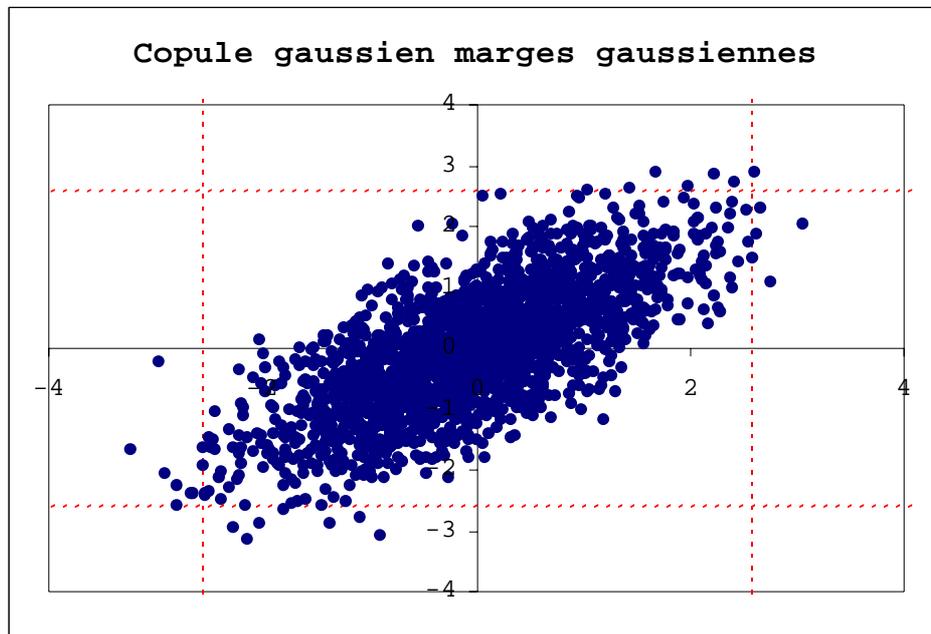


Figure 8 : simulation de 2 000 v.a. de copule gaussien ($\rho = 70\%$) et de marges gaussiennes

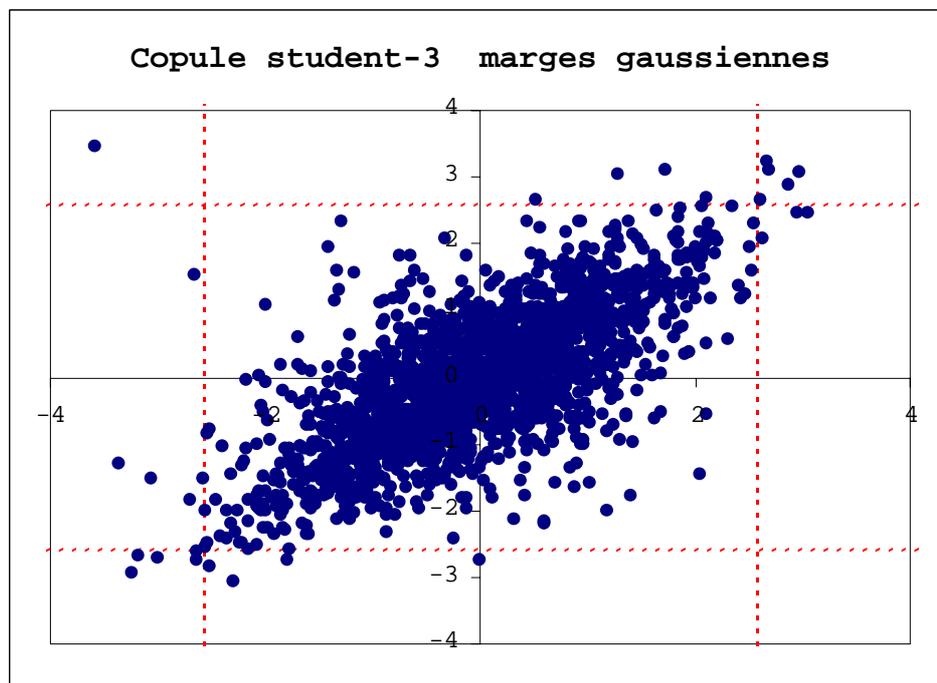


Figure 9 : simulation de 2 000 v.a. de copule Student ($\nu = 3$; $\rho = 70\%$) et de marges gaussiennes

Note : Les pointillés représentent les quantiles à 0.5%

Dans la gestion de portefeuille ce sont les « catastrophes » qui nous intéressent, c'est à dire la fréquence des évènements joints extrêmes. Ces catastrophes dépendent de la dépendance de queue du copule (cf. chapitre « Tail dépendance »)

Les copules Gaussien et Student sont très différents du point de vue de cette dépendance. En effet il y a indépendance des valeurs extrêmes pour le copule Normale tandis que le Student présente une forte dépendance de queue.

	Nombre de « Catastrophes » à	
	99%	99.5%
Copule Normale	10	2
Copule Student-3	19	11

Tableau 3 : Nombre de valeurs extrêmes pour une simulation de 2000 v.a. de copule Normale et Student de degré de liberté 3 (cf. Figure 1 et 2.)

Dans cet exemple, on réalise l'impact que peut avoir le copule sur les valeurs extrêmes. Le nombre de « catastrophes » à 99% (id. 99.5%) est multiplié par 2 (id. 5) lorsque l'on passe d'une structure de dépendance Gaussienne à une Student de degré de liberté 3.

ALGORITHMES

L'un des but du stage fut de réaliser ces algorithmes qui permettent :

- de simuler les copules archimédiens bivariés
- de simuler les copules de la famille H_n
- de simuler les copules Gaussien et Student multivariés
- de fitter un copule archimédien
- d'identifier le copule de Student

Algorithme 0 : Algorithme permettant de générer deux variables aléatoires uniformes (u, v) dont la fonction de répartition jointe est un copule C .

- (1) On génère une variable aléatoire u de loi uniforme.
- (2) On génère une variable aléatoire v de fonction de répartition

$$P(V \leq v / U = u) = \frac{\partial C(u, v)}{\partial u}$$

Remarque : En pratique il est difficile d'utiliser cet algorithme, en effet on ne connaît que rarement la fonction de répartition conditionnelle.

Algorithme 1 : Algorithme permettant de générer deux variables aléatoires uniformes (u, v) dont la fonction de répartition jointe est un copule archimédien C de générateur φ .

- (1) On génère 2 variables aléatoires indépendantes, s et q de lois uniformes.
- (2) On calcule $t = K_C^{(-1)}(q)$, ou $K_C^{(-1)}(q)$ est l'inverse de la fonction de répartition K_C
- (3) On obtient donc la pair (u, v) avec

$$u = \varphi^{[-1]}(s\varphi(t)) \text{ et } v = \varphi^{[-1]}((1-s)\varphi(t))$$

exemple pour un copule Gumbel de paramètre θ :

- (1) On génère 2 variables aléatoires indépendantes, s et q de lois uniformes.
- (2) On résout $q = t(1 - \frac{\ln t}{\theta})$
- (3) Et donc :

$$u = t^{s^{(1/\theta)}} \quad \text{et} \quad v = t^{(1-s)^{(1/\theta)}}$$

Algorithme 2 : Algorithme permettant de générer des variables aléatoires uniformes (u_1, u_2, \dots, u_n) dont la fonction de répartition jointe est un copule archimédien multidimensionnel de la famille H_n .

- On génère une variable aléatoire q de loi uniforme.

- On calcule $t = K_{C_{\theta_1}}^{(-1)}(q)$, ou $K_{C_{\theta_1}}^{(-1)}(q)$ est l'inverse de la fonction de répartition K_C
- On génère une variable aléatoire de loi uniforme s_1 indépendante de q .
- On obtient donc $a_1 = \varphi_{\theta_1}^{[-1]}(s_1 \varphi_{\theta_1}(t_1))$ et $u_n = \varphi_{\theta_1}^{[-1]}((1-s_1) \varphi_{\theta_1}(t_1))$
-
- On calcule $t_i = K_{C_{\theta_i}}^{(-1)}(a_{i-1})$
- On génère une variable aléatoire de loi uniforme s_i indépendante de q et de s_1, \dots, s_{i-1} .
- On obtient donc $a_i = \varphi_{\theta_i}^{[-1]}(s_i \varphi_{\theta_i}(t_i))$ et $u_{n-i+1} = \varphi_{\theta_i}^{[-1]}((1-s_i) \varphi_{\theta_i}(t_i))$
-
- On calcule $t_{n-1} = K_{C_{\theta_{n-1}}}^{(-1)}(a_{n-2})$
- On génère une variable aléatoire de loi uniforme s_{n-1} indépendante de q et de s_1, \dots, s_{n-2} .
- On obtient donc $u_1 = \varphi_{\theta_{n-1}}^{[-1]}(s_{n-1} \varphi_{\theta_{n-1}}(t_{n-1}))$ et $u_2 = \varphi_{\theta_{n-1}}^{[-1]}((1-s_{n-1}) \varphi_{\theta_{n-1}}(t_{n-1}))$
- On obtient donc les variables (u_1, u_2, \dots, u_n) de fonction de répartition jointe $C^n(u_1, u_2, \dots, u_n; \theta_1, \theta_2, \dots, \theta_{n-1})$

Algorithme 3 : Algorithme permettant de générer des variables aléatoires uniformes (u_1, \dots, u_n) dont la fonction de répartition jointe est le copule Gaussien de matrice de corrélation Σ .

- (1) Appliquer la décomposition de Cholesky à la matrice Σ . ($A.A^T = \Sigma$, A triangulaire inférieur)
- (2) Générer des variables aléatoires (z_1, \dots, z_n) de loi Normale centrée réduite
- (3) Calculer $x = Az$
- (4) Calculer $u_i = \Phi(x_i)$ pour $i = 1, \dots, n$: Φ la distribution normale
- (5) On obtient donc les variables (u_1, \dots, u_n) de fonction de répartition jointe le copule Normale de matrice de corrélation Σ .

Algorithme 4 : Algorithme permettant de générer des variables aléatoires uniformes (u_1, \dots, u_n) dont la fonction de répartition jointe est le copule de Student de degré de liberté ν et de matrice de corrélation Σ .

- (1) Appliquer la décomposition de Cholesky à la matrice Σ . ($AA^T = \Sigma$, A triangulaire inférieure)
- (2) Générer des variables aléatoires (z_1, \dots, z_n) de loi Normale centrée réduite
- (3) Générer une variable s de loi χ_v^2 indépendante de (z_1, \dots, z_n)
- (4) Calculer $y = Az$
- (5) Calculer $x = \frac{\sqrt{v}}{\sqrt{s}} y$
- (6) Calculer $u_i = t_v(x_i)$ pour $i = 1, \dots, n$
- (7) On obtient donc les variables (u_1, \dots, u_n) de fonction de répartition jointe le copule Student de degré de liberté v de matrice de corrélation Σ .

Fitter un copule :

A. Copules bivariés

Algorithme 5 : Algorithme permettant de fitter un copule archimédien C bivarié de paramètre α sur une série de données.

- (1) Transformer les données (X, Y) en rang (U, V) , en utilisant la fonction de répartition empirique.

$$U_i(\cdot) = \frac{1}{n+1} \sum_{i=1}^n 1\{X_i \leq \cdot\} \quad V_i(\cdot) = \frac{1}{n+1} \sum_{i=1}^n 1\{Y_i \leq \cdot\}$$

où $1\{\cdot\}$ est la fonction indicatrice.

- (2) Calculer le maximum de vraisemblance :

$$\hat{\alpha} = \arg \max_{\alpha \in R} L(\alpha; U, V) = \arg \max_{\alpha \in R} \sum_{i=1}^n \log C_{12}(U_i, V_i)$$

où R représente les valeurs possibles de α (voir en annexe)

$$\text{et } C_{12}(u, v) = \frac{\partial}{\partial u} \frac{\partial}{\partial v} C(u, v)$$

- (3) On estimera chaque copule candidat par maximum de vraisemblance. On retiendra comme « meilleur » copule celui qui à la plus grande vraisemblance (critère AIC - Akaike's Information Criterion).

La difficulté réside dans le calcul de la densité du copule. En effet pour certain copules archimédiens on ne peut calculer une formule explicite de leur densité.

Exemple : densité du copule gaussien

En utilisant la relation entre la densité jointe et le copule

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)C_{12}[F_1(x_1), F_2(x_2)]$$

dans le cas d'une loi Gaussienne bivariée centrée réduite de coefficient de corrélation linéaire ρ , on obtient :

$$\frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho x_1x_2}{2(1-\rho^2)}\right) = C_{12}[F_1(x_1), F_2(x_2)] \cdot \frac{1}{2\pi} e^{-\frac{x_1^2}{2}} \cdot \frac{1}{2\pi} e^{-\frac{x_2^2}{2}}$$

La densité est donc :

$$C_{12}(u_1, u_2, \rho) = \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\rho^2 x_1^2 + \rho^2 x_2^2 - 2\rho x_1x_2}{2(1-\rho^2)}\right)$$

$$\text{où } \begin{cases} x_1 = F_1^{-1}(u_1) \\ x_2 = F_2^{-1}(u_2) \end{cases}$$

B. Copule multidimensionnel

Identification du t-copule de Student :

Algorithme 6 : Algorithme permettant d'identifier un copule de Student de dimension d sur une série de données.

- (1) Transformer les données $X_d = (X_1, \dots, X_d)$ avec $X_i = (X_{i1}, \dots, X_{in})$ en rang U_d , en utilisant la fonction de répartition empirique.

$$\hat{F}_j(\cdot) = \frac{1}{n+1} \sum_{i=1}^n 1\{X_{ij} \leq \cdot\} \text{ pour } j=1, \dots, d.$$

où $1\{\cdot\}$ est la fonction indicatrice.

On obtient donc :

$$\hat{U}_i = (\hat{F}_1(X_{i1}), \dots, \hat{F}_d(X_{id}))$$

- (2) Estimer la matrice de corrélation $\hat{\Sigma}$ en utilisant l'estimateur de Kendall :

$$\hat{\Sigma}_{ij} = \sin\left(\frac{\pi}{2} \hat{\tau}_{ij}\right).$$

où, $\hat{\tau}_{ij} = \hat{\tau}(\hat{U}_i, \hat{U}_j)$

(3) Utiliser un calcul numérique pour trouver $\hat{\nu}$ tel que :

$$\hat{\nu} = \arg \max_{\nu \in (2, \infty]} \left[\sum_{i=1}^n \log(c(\hat{U}_i, \nu, \hat{\Sigma})) \right]$$

On a la statistique :

$$\Lambda_n(\hat{\nu}/\nu_0) = -2 \log \frac{\prod_{i=1}^n c(\hat{U}_i, \nu_0, \hat{\Sigma})}{\prod_{i=1}^n c(\hat{U}_i, \hat{\nu}, \hat{\Sigma})}$$

où $\hat{\nu}$ est déterminé à l'aide du maximum de vraisemblance en (3) et où ν_0 représente l'hypothèse que l'on souhaite tester. On prendra $\nu_0 = 10^5$ comme approximation du copule gaussien.

Quand n tend vers l'infini, on a l'approximation :

$$\Lambda_n(\hat{\nu}/\nu_0) \approx 2 \chi_1^2$$

CADRE FINANCIER

On s'intéresse au risque associé au portefeuille de crédit de la Société Générale.

Pour cela nous allons définir le modèle de Merton à un facteur et considérer que nous sommes dans le cas du portefeuille homogène.

Puis nous allons introduire le concept de Capital Economique, qui représente le « risque » du portefeuille.

Et enfin nous allons utiliser plusieurs types de structures de dépendance (ou copules) pour évaluer le risque lié à un portefeuille composé de deux sous portefeuilles homogènes.

Portefeuille homogène – modèle KMV

Le modèle de Merton (publié en 1974)

Dans la théorie de la firme, Merton interprète les actions d'une société en terme d'options. Supposons que l'entreprise n'ait que deux classes de passif : les fonds propres et la dette, considérés comme homogènes (notamment, de même maturité). La société promet le remboursement de la dette aux créanciers à maturité. A cette unique date, ou bien la valeur de la firme est supérieure à la valeur faciale de la dette, qui peut alors être remboursée – les actionnaires reçoivent donc la différence entre cette valeur et la dette – ou bien la firme fait défaut, les créanciers s'emparent des actifs de la société et les actionnaires n'ont rien. En termes financiers, les actionnaires détiennent un call (une option d'achat) sur l'entreprise de strike (ou prix d'exercice) la dette et de maturité celle de la dette.

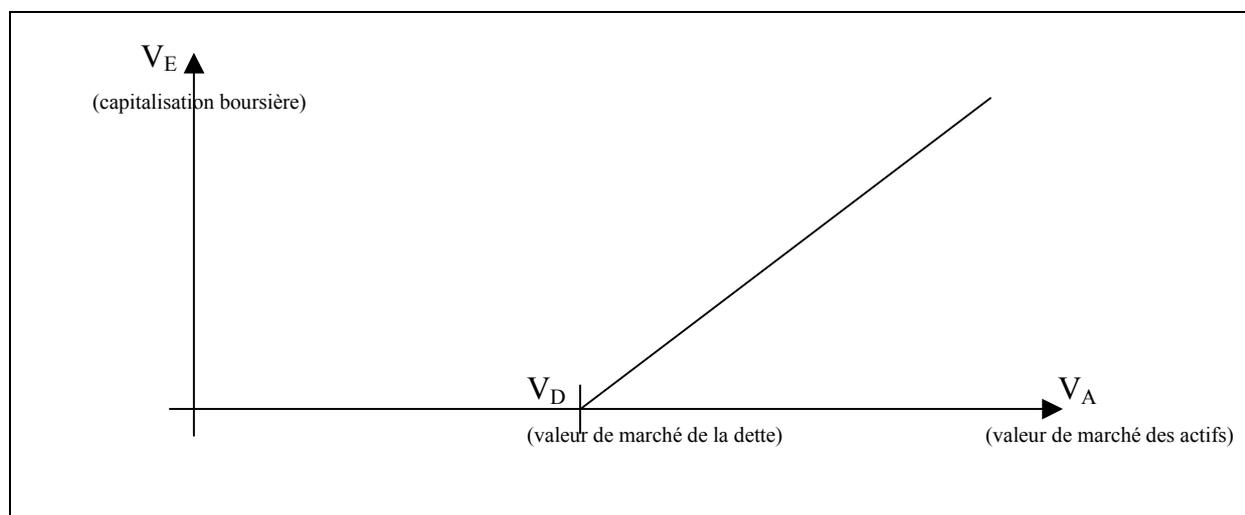


Figure 10 : capitalisation boursière en fonction de la valeur de marché des actifs

Modèle de la firme à un facteur :

Soit un portefeuille de crédits composé de N obligors. Le défaut de chaque contrepartie dépend de la valeur de sa firme à échéance $V_i(T)$, la contrepartie i fait défaut si la valeur de sa firme est inférieure à un seuil s_i donc si : $V_i(T) \leq s_i$.

Dans le modèle de la firme à un facteur, la valeur de la chaque contrepartie dépend d'un facteur Y et est égale à :

$$V_i(T) = Y\sqrt{\rho} + \varepsilon_i\sqrt{1-\rho} \quad \forall i \leq N$$

où Y et $\varepsilon_i, i \leq N$ suivent une loi normale centrée réduite.

Remarques :

Les valeurs des actifs de deux obligors i et j avec $i \neq j$ ont pour coefficient de corrélation linéaire ρ .

Pour une réalisation de Y donnée, les valeurs des actifs sont indépendantes.

Définition : Un portefeuille est homogène si les N contreparties qui le composent ont

- même probabilité de défaut
- même exposition
- des corrélation identiques

C'est donc le cas particulier du modèle de la firme à un facteur où tous les obligors ont même seuil s et même exposition E .

Pour simplifier nous allons considérer que $E = 1$

Le taux de défaillance du portefeuille sachant Y est donc égale à :

$$L(Y) = \frac{\sum_{i=1}^N 1_{\{V_i(T) \leq s | Y\}}}{N}$$

$1_{\{V_i(T) \leq s | Y\}}$ pour $i \leq N$ sont des variables aléatoires indépendantes de même loi, donc en utilisant la loi des grands nombres on a :

$$L(Y) \xrightarrow{\infty} E(1_{\{V_i(T) \leq s|Y\}})$$

Pour un portefeuille homogène de taille importante on a donc :

$$\begin{aligned} L(Y) &= E(1_{\{V_i(T) \leq s|Y\}}) \\ &= P(V_i(T) \leq s|Y) \\ &= P(Y\sqrt{\rho} + \varepsilon_i\sqrt{1-\rho} \leq s) \\ &= P(\varepsilon_i \leq \frac{s-Y\sqrt{\rho}}{\sqrt{1-\rho}}) \end{aligned}$$

$$\text{donc : } L(Y) = N\left(\frac{s-Y\sqrt{\rho}}{\sqrt{1-\rho}}\right)$$

où N est la fonction de répartition de la loi normale centrée réduite

Calculons la fonction de répartition de L

$$\begin{aligned} P(L \leq x) &= P\left(N\left(\frac{s-Y\sqrt{\rho}}{\sqrt{1-\rho}}\right) \leq x\right) \\ &= P\left(Y \geq \frac{s - N^{-1}(x)\sqrt{1-\rho}}{\sqrt{\rho}}\right) \\ &= P\left(Y \leq \frac{N^{-1}(x)\sqrt{1-\rho} - s}{\sqrt{\rho}}\right) \end{aligned}$$

$$\text{donc : } P(L \leq x) = N\left(\frac{N^{-1}(x)\sqrt{1-\rho} - s}{\sqrt{\rho}}\right)$$

on en déduit que L a pour densité :

$$f(x) = \sqrt{\frac{1-\rho}{\rho}} \exp\left\{\frac{1}{2}(N^{-1}(x)) - \frac{1}{2\rho}(s - N^{-1}(x)\sqrt{1-\rho})^2\right\}$$

KMV (Kealhofer McQuown Vasicek) est le nom le plus souvent donné à ce type de loi.

Algorithme 1 : Algorithme permettant de générer une variable aléatoire de loi KMV de seuil s et de corrélation ρ

(1) Générer f variable aléatoire de loi Normale centrée réduite

$$(2) \text{ Calculer } L = \mathbf{N}\left(\frac{s - f\sqrt{\rho}}{\sqrt{1-\rho}}\right)$$

(3) On obtient donc la variable aléatoire L de loi KMV de seuil s et de corrélation ρ

remarque : f est appelé facteur de la loi KMV

Propriété : La structure de dépendance de deux lois KMV est la même que la structure de dépendance de leurs deux facteurs f_1 et f_2 .

Démonstration :

Soit deux variables aléatoires L_1 et L_2 de loi KMV.

$$L_i = \mathbf{N}\left(\frac{s_i - f_i\sqrt{\rho_i}}{\sqrt{1-\rho_i}}\right) \quad \text{avec } i \in \{1;2\}$$

f_1 et f_2 étant des variables aléatoires symétrique, en introduisant les fonctions h_i on a donc :

$$L_i = h_i(f_i) = \mathbf{N}\left(\frac{s_i + f_i\sqrt{\rho_i}}{\sqrt{1-\rho_i}}\right) \quad \text{avec } i \in \{1;2\}$$

h_1 et h_2 sont des fonctions strictement croissantes, donc d'après le théorème principal sur les copules, on a donc :

$$C(h_1(L_1), h_2(L_2)) = C(L_1, L_2)$$

De cette propriété on en déduit l'algorithme 2.

Algorithme 2 : Algorithme permettant de générer deux variables aléatoires de loi KMV de seuils s_1, s_2 de corrélation ρ_1, ρ_2 et de structure de dépendance un copule C

- (1) Générer u_1 et u_2 deux variables aléatoires uniformes de structure de dépendance le copule C (cf. simulation de copules algorithme 1)
- (2) Calculer $f_1 = \mathbf{N}^{-1}(u_1)$ et $f_2 = \mathbf{N}^{-1}(u_2)$
- (3) Calculer $L_1 = \mathbf{N}\left(\frac{s_1 - f_1\sqrt{\rho_1}}{\sqrt{1-\rho_1}}\right)$ et $L_2 = \mathbf{N}\left(\frac{s_2 - f_2\sqrt{\rho_2}}{\sqrt{1-\rho_2}}\right)$
- (3) L_1 et L_2 sont donc les deux variables aléatoires recherchées.

Capital économique :

La définition la plus répandue du capital économique d'un portefeuille de crédits utilise la notion de quantile de la distribution des pertes.

Pour un portefeuille de taille 1, le capital économique à $\alpha\%$ est égal à la différence entre le α -quantile de la distribution des pertes et l'espérance des pertes, plus précisément :

$$EC(\alpha) = q_\alpha(L) - E(L)$$

$$\text{où : } q_\alpha(L) = \inf \left\{ y / P[L > y] > 1 - \frac{\alpha}{100} \right\}$$

C'est donc le capital dont doit disposer la banque pour couvrir $\alpha\%$ de ses pertes.

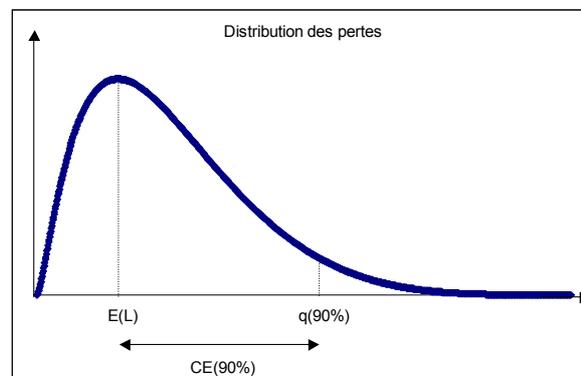
Remarque : la notion de α -quantile est équivalente à celle de la VaR (Value at Risk) à $\alpha\%$

$$q_\alpha(L) = VaR_{\alpha\%}(L)$$

Propriété : Soit f une fonction croissante et L une variable aléatoire, alors

$$VaR_{\alpha\%}(f(L)) = f(VaR_{\alpha\%}(L))$$

exemple : Capital économique à 90% pour un portefeuille de crédit dont la distribution des pertes suit une loi KMV.



En risk management, on retiendra une valeur de 99.98 pour α , ce qui représente la probabilité de défaut à 3 ans d'une entreprise notée AA^- .

L'espérance des pertes au dessus d'un seuil K peut être une autre définition du capital économique :

$$EC_K = E[L / L > K]$$

Capital marginal :

Soit deux portefeuilles de crédit P_1 et P_2 , alors le capital marginal du portefeuille P_2 est égal à la différence entre le capital économique du portefeuille composé de P_1 et P_2 et le capital économique du portefeuille P_1 .

$$CM(P_2) = EC(P_1 + P_2) - EC(P_1)$$

Le capital marginal représente donc l'incrément de capital lié à l'acquisition de P_2 , partant de P_1 .

Contribution :

Nous allons définir la notion de contribution dans le cadre de loi Normale, mais pour cela nous allons exposer la propriété de la VaR dans ce cadre.

Soit N_σ , une variable aléatoire de loi Normale centrée d'écart type σ , on a le résultat suivant

$$VaR_\alpha(N_\sigma) = k_\alpha \cdot \sigma$$

Donc si la distribution des pertes d'un portefeuille P de montant E suit une loi $N(m, \sigma^2)$ alors :

$$CE_\alpha(P) = VaR_\alpha(N_\sigma) * E = E \cdot k_\alpha \cdot \sigma$$

Soit un portefeuille P composé de 2 sous portefeuilles P_1 et P_2 de montants E_1 et E_2 .

Le capital économique étant homogène en fonction de E_1 et E_2 , on a donc d'après le théorème d'Euler :

$$CE = E_1 \frac{\partial CE}{\partial E_1} + E_2 \frac{\partial CE}{\partial E_2}$$

Soit σ l'écart type du portefeuille P , on a donc la relation :

$$\sigma = (\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)^{1/2}$$

où σ_1 et σ_2 sont les écarts types respectifs des portefeuilles P_1 et P_2 .

$$\text{En posant } \begin{cases} \sigma_1 = E_1 \cdot s_1 \\ \sigma_2 = E_2 \cdot s_2 \end{cases}$$

On a donc : $E_i \frac{\partial CE}{\partial E_i} = \frac{E_i^2 s_i^2 + \rho E_1 E_2 s_1 s_2}{\sigma^2} EC = \frac{\text{cov}(L_i, L)}{\text{Var}(L)} EC$ avec $i \in \{1;2\}$

où L et L_i sont les lois des pertes des portefeuilles P et P_i .

La contribution du portefeuille P_i dans le portefeuille total est donc :

$$Cb_i = \frac{\text{cov}(L_i, L)}{\text{Var}(L)} * CE(SG + M)$$

Identification de la structure de dépendance entre le portefeuille SG et une tranche d'ABS²

Il s'agit d'identifier la structure de dépendance (copule) entre le portefeuille de la Société Générale et un autre portefeuille représenté par une tranche d'ABS.

Ainsi on pourra calculer le capital marginal et comparer les résultats obtenus avec le cas Gaussien.

Pour cela on dispose de 60 000 couples de données représentant les pertes des deux portefeuilles.

Après avoir maximisé la vraisemblance pour les différents copules archimédiens (algorithme 5, page 28), on constate que le copule qui satisfait le critère AIC est le Gaussien.

Nous avons ensuite utilisé le test d'identification du copule de Student (algorithme 6, page 28). Ce test nous permet de déterminer, avec une erreur fixée, l'intervalle auquel appartient le degré de liberté du copule de Student ainsi que celui qui maximise la pseudo vraisemblance.

	Erreur de 1%	Erreur de 0.05%
$\nu \in$	[18 ;30]	[16 ;35]

Tableau 4 : intervalle de ν , maximum atteint pour $\nu = 22$ (cf. figure 12),

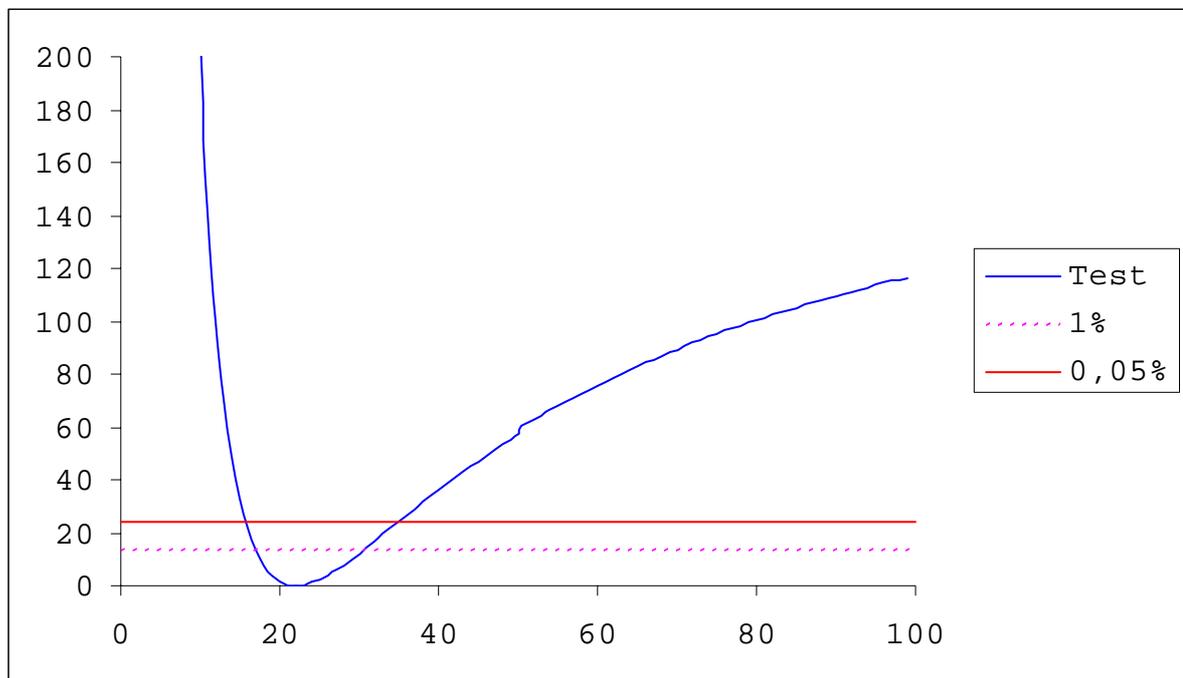


Figure 11 : test d'identification du copule de Student entre le portefeuille SG et une tranche d'ABS, en abscisse le degré de liberté.

²Dans notre étude, nous prendrons comme tranche d'ABS (Asset Back Securitie) un portefeuille homogène

SIMULATION

Pour des raisons de confidentialité, nous allons supposer que les deux portefeuilles homogènes suivants, représentent le portefeuille de la société générale ainsi que celui de la tranche d'ABS :

	Probabilité de défaut	Corrélation entre les actifs	Montant
P(SG)	1.2%	30%	1000
P(ABS)	2.3%	40%	100

On prendra une corrélation de 25% entre les facteurs.

Pour calculer le capital économique et le capital marginal, nous prendrons pour structure de dépendance des 2 facteurs un copule de Student de degré ν .

Pour alléger le temps de calcul, nous prendrons un capital économique à 99% :

$$EC(99\%) = (VaR_{99\%} - P_m) * E$$

où P_m et E sont les pertes moyennes et le montant total du portefeuille

- Calcul du capital marginal et de la contribution de l'ABS en capital à 99% pour 5 000 000 de simulations

$$CM(99\%) = EC_{99\%}(SG + ABS) - EC_{99\%}(SG)$$

$$Cb_{ABS}(99\%) = \frac{\text{cov}(L_{ABS}, L_{SG+ABS})}{\text{Var}(L_{SG+ABS})} * EC_{99\%}(SG + ABS)$$

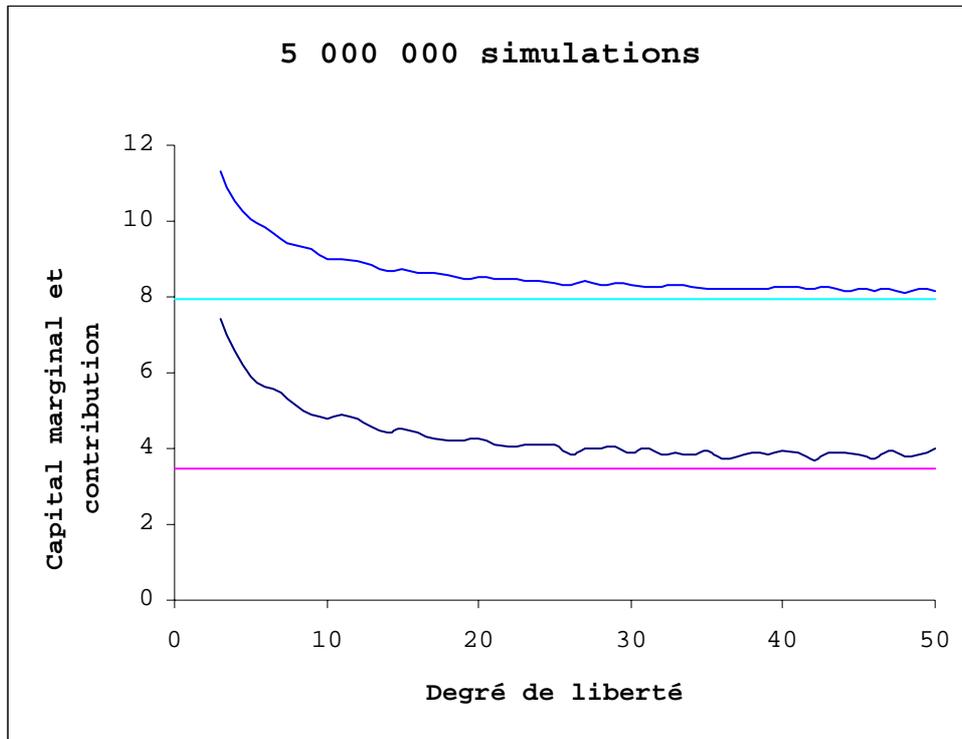
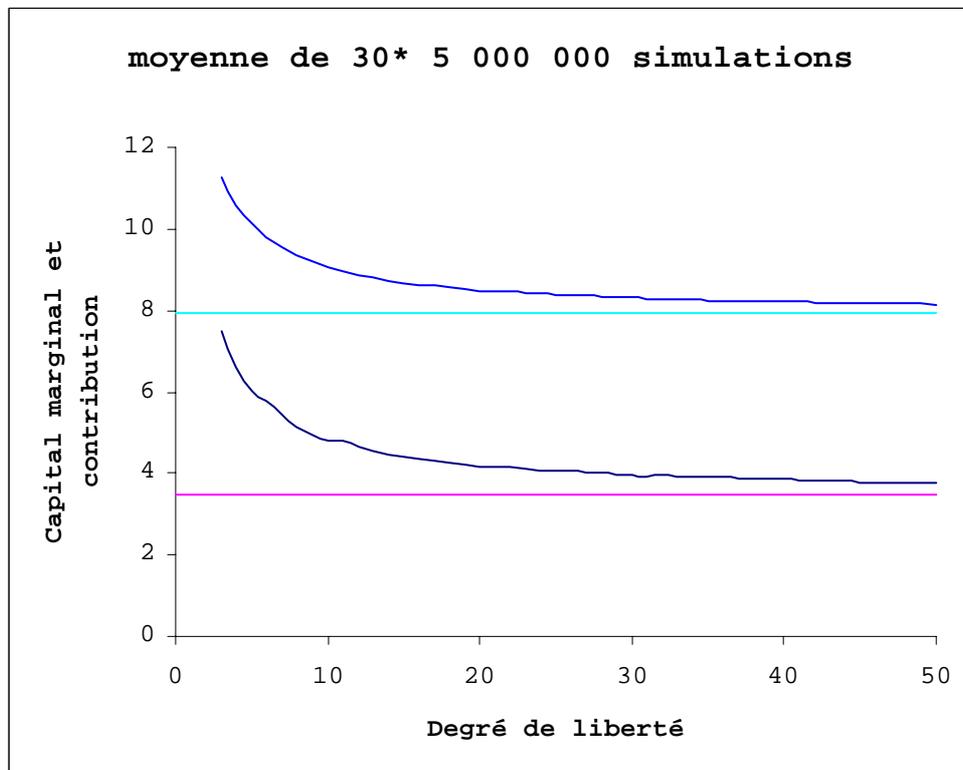


Figure 12 : la courbe du haut représente les contributions et celle du bas le capital marginal
 Les deux droites, représentent le cas où la structure de dépendance des facteurs est le copule gaussien.

Bien que les courbes ne soient pas trop lisses, on remarque que lorsque le degré de liberté tend vers l'infini, on converge vers le copule gaussien.
 Néanmoins il est possible de lisser les courbes en faisant la moyenne de plusieurs simulations.
 Voici les résultats pour la moyenne de 30 simulations de 5 000 000 de simulations chacune.



Sur ce portefeuille, si l'on prenait comme structure de dépendance un copule de Student de degré de liberté 5, on multiplierait le capital marginal (resp. la contribution) par 1.7 (resp. 1.3)

A partir de ces résultats, on peut recalculer le capital marginal et comparer avec le cas gaussien :

degré de liberté	variation du capital marginal	variation de la contribution
16	22,51%	9,02%
17	21,45%	8,44%
18	20,54%	7,98%
19	19,66%	7,58%
20	18,71%	7,11%
21	18,17%	6,86%
22	17,49%	6,56%
23	16,69%	6,24%
24	16,10%	5,98%
25	15,69%	5,78%
26	15,11%	5,52%
27	14,47%	5,32%
28	14,11%	5,08%
29	13,66%	4,92%
30	13,23%	4,84%
31	12,79%	4,65%
32	12,55%	4,49%
33	12,13%	4,33%
34	11,76%	4,24%
35	11,40%	4,06%

Bien que les degrés de liberté du copule de Student soient élevés, on constate que les variations sont importantes : entre 11% et 22% d'augmentation du capital marginal. Les variations des contributions sont elles, beaucoup moins importantes : entre 4% et 9%.

Cependant, les variations du capital marginal dépendent de la corrélation entre les facteurs. Dans cet exemple, on a fixé une corrélation de 25%, nous allons maintenant calculer le capital marginal en la faisant varier.

Calcul du capital marginal en fonction du coefficient de corrélation linéaire :

Soit 2 portefeuilles homogènes de caractéristiques :

	Probabilité de défaut	Corrélation entre les actifs	Montant
P(SG)	1.2%	30%	150
P(ABS)	1%	50%	20

Pour nos calculs, le nombre de simulations est de 30 fois 1.000.000 de simulations

Figure 10 : capital marginal en fonction du coefficient de corrélation linéaire (en%)

- Courbe du haut : copule Student de degré de liberté 5 de paramètre ρ
- Courbe du milieu : copule Gaussien de paramètre ρ
- Courbe du bas : copule de Gumbel de paramètre $\frac{1}{1 - \frac{2}{\pi} \arcsin \rho}$

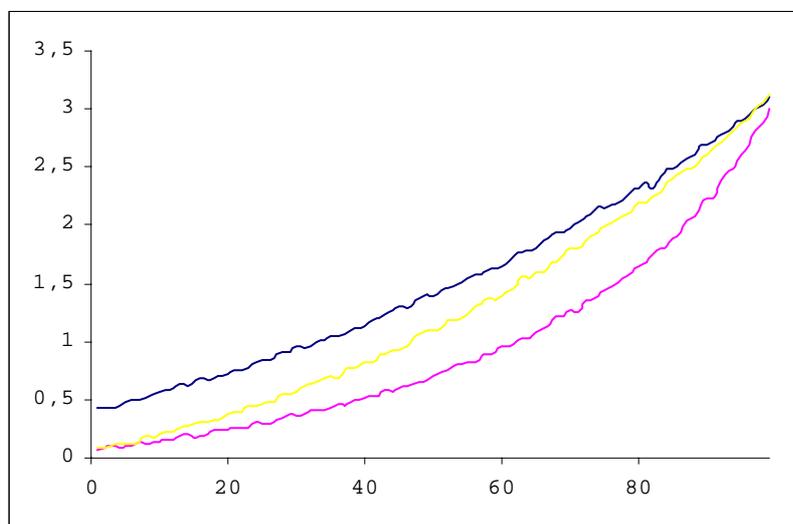


Figure 10 : Capital marginal en fonction du coefficient de corrélation linéaire (en%)

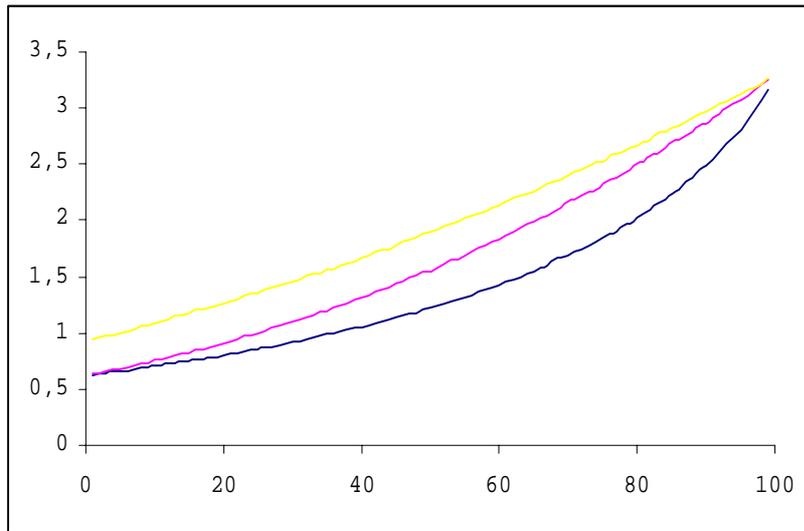


Figure 11 : Contribution en fonction du coefficient de corrélation linéaire (en%)

Remarques :

- le copule de Gumbel est à titre indicatif
- le copule de Student est susceptible d'être utilisé par la Société Générale, nous nous intéressons donc aux différences de celui-ci avec le Gaussien

Conclusion : les différences entre les copules Gaussien et Student sont importantes lorsque la corrélation est faible. En effet, lorsque la corrélation est nulle, les variables aléatoires sont indépendantes dans le cas du copule Gaussien, ce qui n'est pas le cas avec le copule de Student. Dans le cas où les corrélations sont élevées, les différences sont très faibles. Les contributions sont elles moins sensibles au copule lorsque les corrélations sont faibles.

Nous avons donc constaté que seuls les copules Gaussien et Student sont utilisables dans le cas multidimensionnel. Une deuxième étude m'a donc été confiée à la mi septembre. Elle consiste à calculer la probabilité de défaut jointe de deux entreprises en utilisant une approche de type « marché »

II. CALCUL DES PROBABILITES DE DEFAUT JOINT

Résumé :

Lorsqu'une banque accorde un crédit garanti ou bien lors de différents autres types de contrats, elle doit tenir compte de la probabilité de défaut joint, c'est à dire la probabilité pour que l'obligor et le garant fassent défaut a la même période. Dans le cas où ils sont fortement corrélés (par exemple une maison mère et sa fille), cette probabilité n'est pas négligeable.

En se basant sur le modèle de la firme de Merton, nous avons considéré que les actions d'une société étaient assimilables à un call sur l'entreprise de strike la dette et de maturité celle de la dette, et que l'on peut donc modéliser le logarithme de la valeur des actions d'une entreprise par un mouvement Brownien simple.

Toujours d'après la même théorie, le coefficient de corrélation linéaire des deux lois Normales génératrices des deux mouvements Browniens est égale au coefficient de corrélation linéaire du logarithme des rendements des firmes, c'est à dire au logarithme du rendement de leurs valeurs en bourse.

On peut donc, en simulant deux mouvements Browniens « corrélés » représenter la valeur de leurs actifs. Il ne reste plus qu'à déterminer la valeur de leurs dettes (ou barrières) pour déterminer la probabilité de défaut jointe.

Enjeux :

Aujourd'hui les probabilités de défaut joint sont calculées à partir de la loi Normale bivariée, L'inconvénient est qu'elle est réputée sous estimer la fréquences des défauts joints, de plus, ne disposant pas d'autre modèle, on ne peut donc pas évaluer sa validité. L'enjeu est donc de disposer d'une autre approche plus précise pour évaluer les défauts joints.

Ma mission consiste donc à:

- Etudier le modèle de barrière continue de M. Avallaneda et J. Zhu
- A partir de ce modèle, déterminer la barrière continue à dix ans pour chacune des 7 notations internes de la Société Générale (notations équivalentes à celles des agences de notations, Standard & Poors, Moody's)

Mon but sera de calculer la probabilité de défaut joint par année de deux entreprises à partir de leur coefficient de corrélation linéaire (on ne peut dans ce cas utiliser les copules car les

actifs sont représentés par un mouvement brownien). Je pourrais alors comparer ces valeurs avec celles obtenues par la méthode « classique » de la loi normale bivariée.

Résultat :

- j'ai déterminé la barrière affine à 8 ans avec un pas de temps d'un an pour les notations 7 et 4 (cas le plus courant pour un obligor garant) avec une erreur relative de l'ordre de 0.5% pour la notation 7 et de 1% pour la 4.
- j'ai mis en place un programme qui calcule par simulation de Monte Carlo la probabilité de défaut joint à 8 ans, d'un obligor de notation 7 et de son garant de notation 4 pour une corrélation ρ donnée
- j'ai comparé ces résultats avec la méthode « Gaussienne » : contrairement à ce que l'on pourrait penser, les résultats obtenus avec la méthode de la Normale bivariée sont proches de ceux obtenus avec la méthode de la barrière affine.

Difficultés rencontrées :

A l'origine, le but de ma mission était de déterminer les barrières continues pour chacune des notations internes de la Société Générale. Je devais pour cela reproduire la méthode utilisée dans l'article « Distance to default » (cf. [18]).

Après avoir passé plus de deux semaines à tenter de résoudre ce problème, je n'ai réussi, en fin de compte, à ne déterminer que la barrière affine par morceaux, puis on m'a demandé de calculer par simulation les probabilités de défaut joint entre deux entreprises de notation 7 et 4 (équivalentes à des notation CCC et A pour Standard & Poor's)

Ce n'est que plus tard que je me suis aperçu que la barrière continue n'avait pas grand intérêt pour le calcul de ces probabilités et que le modèle affine suffisait amplement. J'ai donc perdu du temps sur ce calcul alors que j'aurais pu dès le début me satisfaire d'une méthode plus simple basée seulement sur des simulations.

Ce fut donc lors simulations que j'ai rencontré la principale difficulté de cette étude. En effet, les probabilités de défaut de la notation 4 étant très faible, il me fallut réaliser un nombre de simulation très important pour obtenir les résultats recherchés, ce qui a énormément augmenté les temps de calcul. Pour palier ce problème, j'ai utilisé deux ordinateurs pour effectuer mes simulations, cela m'a ainsi permis de déterminer les barrières en une semaine seulement.

PROBABILITE DE DEFAUT JOINT

Mon travail consiste à calculer la probabilité de défaut jointe de deux entreprises, mais cette fois-ci nous emprunterons une approche différente de celle utilisée lors de ma mission précédente. En effet, au lieu d'utiliser leur structure de dépendance, nous allons modéliser la valeur des actifs des deux entreprises en question par des mouvements browniens.

Ainsi, nous pourrions dire qu'une entreprise fait défaut lorsque la valeur de ces actifs est inférieure à un seuil ou barrière (cf. chapitre précédent : Modèle de la firme de Merton)

Nous calculerons d'abord la barrière affine d'une entreprise en fonction de sa notation, puis nous déterminerons la probabilité de défaut jointe en fonction du coefficient de corrélation linéaire.

Barrière affine :

En prenant une dette fixe au cours du temps et en représentant les actions d'une société par un mouvement brownien simple, il est possible de calculer la probabilité pour que l'entreprise fasse défaut dans la période $[0, T]$.

A titre d'exemple, on souhaite déterminer la barrière b (ou la dette) pour qu'une entreprise ait une probabilité de faire défaut P_1 à une date T_1 et P_2 à une date T_2 .

La valeur de b pour obtenir une probabilité de défaut P_1 à une date T_1 est donnée par la formule :

$$b = \sqrt{T_1} \cdot N^{-1}\left(\frac{P_1}{2}\right)$$

où N^{-1} est l'inverse de la loi normale

Une fois b déterminé, en inversant la formule on obtient la probabilité de faire défaut à une date T_2

A partir d'une série de probabilités de défaut, on ne peut donc, trouver une barrière fixe y correspondant.

L'une des solutions de ce problème est d'utiliser un modèle dans lequel la barrière est affine par morceaux.

Barrière affine par morceaux

En considérant le cas simple d'un mouvement Brownien sans drift et une barrière affine de la forme $b(t) = -\alpha - \beta.t$ ($\alpha > 0$, $0 < t < t_0$), alors la probabilité $P(t)$ pour que le Brownien traverse la barrière avant une date t ainsi que la dérivée $P'(t)$ de cette probabilité sont données par :

$$P(t) = N\left(\frac{-\alpha - \beta.t}{\sqrt{t}}\right) + e^{-2\alpha\beta} N\left(\frac{-\alpha + \beta.t}{\sqrt{t}}\right) \quad (1)$$

$$P'(t) = \frac{\alpha}{t\sqrt{2\pi t}} e^{-\frac{(\alpha + \beta.t)^2}{2t}} \quad (2)$$

Dans l'article « Distance to default », il est présenté une méthode pour résoudre l'équation aux dérivées partielles que doit vérifier la densité de la probabilité de défaut joint. Cette méthode permet donc d'obtenir la barrière continue à partir des probabilités de défaut de l'entreprise (id. notation).

Je n'ai pas réussi à mettre en œuvre cette méthode, mais j'ai néanmoins pu obtenir une barrière affine par morceaux. De plus, je me suis aperçu, lors des simulations, qu'il est inutile d'avoir une barrière continue car un autre problème plus important se pose lors des simulations.

A barrière continue fixe, on n'obtient pas les mêmes probabilités de « taper » la barrière en fonctions des pas de temps que l'on utilise.

En effet, le mouvement Brownien est simulé suivant un pas temps. On simule une loi Normale centrée d'écart type $\sqrt{t_2 - t_1}$ que l'on ajoute à la valeur du brownien à la date t_1 pour obtenir sa valeur à la date t_2 .

Ainsi, lorsque l'on simule un brownien, il faut tenir compte du fait que c'est un processus continue, et qu'il y a donc une probabilité pour qu'il ait « tapé » la barrière tout en étant au dessus aux temps t_1 et t_2 , et ceci quelque soit le pas de temps $t_2 - t_1$ (cf. annexe 3 : Calcul de Andersen). Cette probabilité dépend du pas de temps d'une part, mais également de la barrière.

Or, lorsque le pas de temps change, la barrière change elle aussi car on discrétise une courbe continue (cf. Figure 1) et donc la probabilité de faire défaut n'est plus la même. C'est pour

cette raison que nous avons d'abord fixé le pas de temps, puis nous avons déterminé la barrière pas à pas.

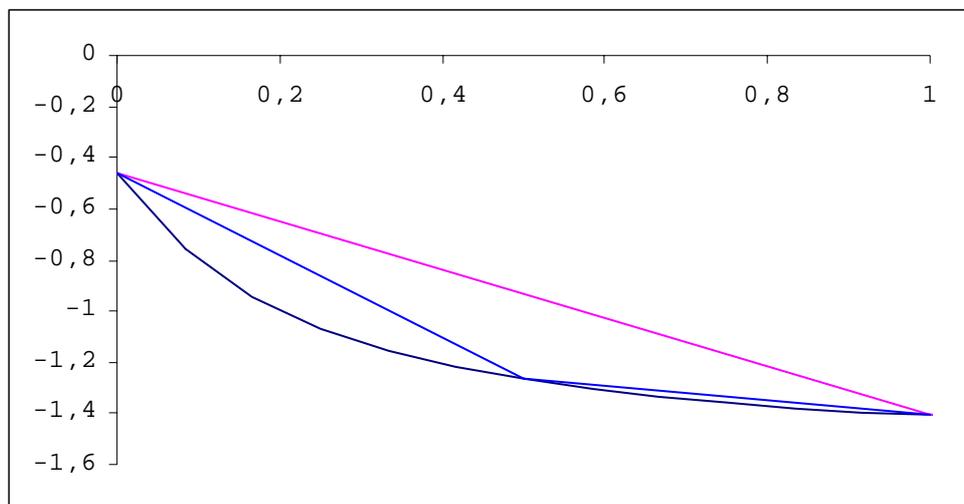


Figure 1 : Barrière continue, Barrière affine avec un pas de temps de 6 mois et de 1 an pour une notation 7 (année en abscisse et valeur de la barrière en ordonnée)

Lorsque le pas de temps passe de 6 mois à 1 an, la barrière est modifiée et donc la probabilité de la « taper » aussi.

Dans le but d'alléger les temps de calcul, nous avons fixé un pas de temps d'une année et déterminé la barrière pour les notations 4 et 7 (cf. figure 2 : probabilités de défaut), qui sont les notations les plus répandues lors de prêts garantis.

année	probabilités de défaut cumulées		probabilités de défaut	
	notation 4	notation 7	notation 4	notation 7
1	0,1680%	19,0263%	0,1680%	19,0263%
2	0,4967%	32,6291%	0,3287%	13,6028%
3	0,9694%	42,5118%	0,4727%	9,8827%
4	1,5685%	49,8222%	0,5991%	7,3103%
5	2,2766%	55,3368%	0,7081%	5,5146%
6	3,0770%	59,5839%	0,8004%	4,2471%
7	3,9542%	62,9249%	0,8772%	3,3410%
8	4,8940%	65,6088%	0,9398%	2,6839%

Tableau 1 : Probabilités de défaut et probabilités de défaut cumulées pour les notations 4 et 7

Détermination de la barrière :

On calcule la valeur de la barrière en 0 et 1 an à partir des formules analytiques (1) et (2). Puis, à l'aide de 20 millions de simulations de browniens et par dichotomie on détermine la valeur de la barrière à 2 ans. On réitère le processus jusqu'à obtenir les valeurs pour les huit premières années (cf. figure 3).

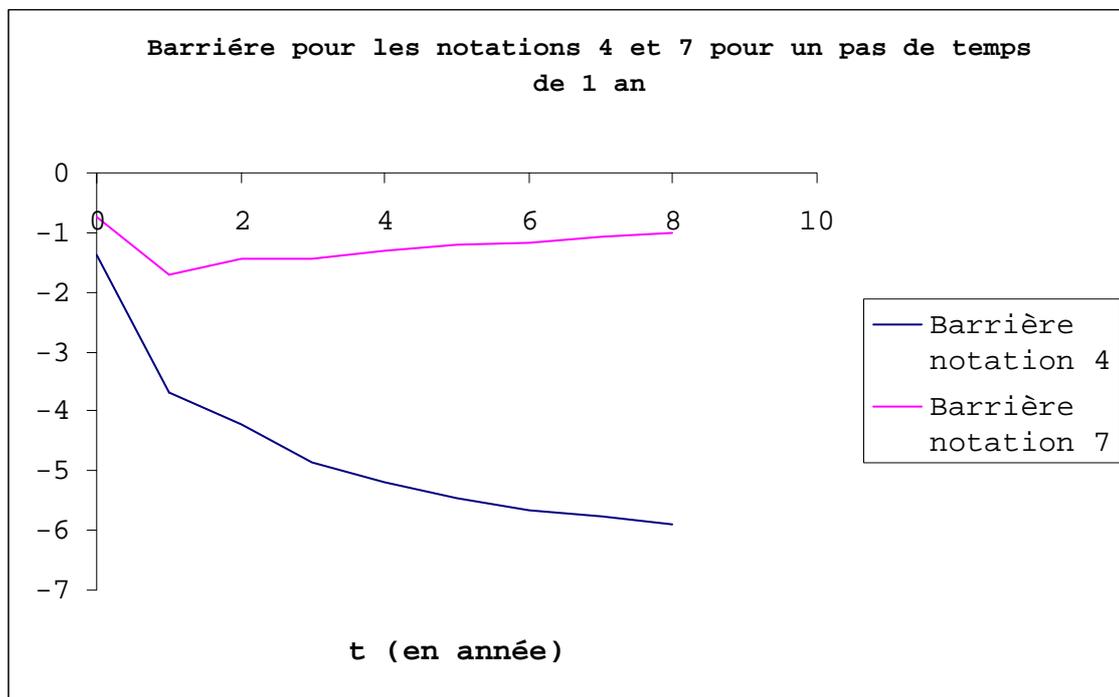


Figure 2 : Barrière affine par morceaux pour un pas de temps de 1 an, pour les notations 4 et 7

Remarque : D'après le modèle de Merton, la barrière représente la dette de l'entreprise, on peut donc en conclure que les entreprises les mieux notées ont tendance à se désendetter (la barrière diminue : cf. figure 3) et que les moins bien notées ont tendance à s'endetter (la barrière remonte : cf. figure 3)

La difficulté réside dans le manque de précision des simulations d'une part, et du temps de calcul d'autre part. C'est pour cette raison que nous avons évalué les erreurs relatives entre les valeurs théoriques et les valeurs empiriques.

Compte tenu des faibles probabilités de défaut de la notation 4 (cf. figure 2), il est donc évident que l'erreur pour cette notation est de loin supérieure à celle de la notation 7.

Néanmoins, les erreurs relatives étant très faibles, on peut donc évaluer par simulation de Monte Carlo les probabilités de défaut joint par année.

	erreur barrière 4			erreur barrière 7		
	Moyenne	Minimum	Maximum	Moyenne	Minimum	Maximum
1	0,551%	0,006%	1,720%	0,0732%	0,0105%	0,1325%
2	0,435%	0,007%	0,890%	0,0311%	0,0036%	0,0754%
3	1,068%	0,503%	1,511%	0,0418%	0,0123%	0,0744%
4	0,215%	0,007%	0,600%	0,0648%	0,0379%	0,0985%
5	0,769%	0,515%	1,132%	0,0473%	0,0238%	0,0687%
6	0,843%	0,636%	1,068%	0,0707%	0,0452%	0,0884%
7	0,354%	0,150%	0,495%	0,0156%	0,0017%	0,0346%
8	0,621%	0,450%	0,816%	0,0126%	0,0009%	0,0307%

Tableau 2 : Moyenne, minimum et maximum des erreurs relatives pour 35 fois 20 000 000 de simulations

Calcul des probabilités de défaut joint :

Le tableau suivant compare les probabilités de défaut joint de deux entreprises de notation 4 et 7 de corrélation 40% obtenues avec deux différentes méthodes.

La première, consiste à les calculer à l'aide des barrières affines par morceaux et à modéliser leur actif à l'aide d'un mouvement brownien. La deuxième méthode considère que la distribution des pertes des deux entreprises est une loi normale. Et donc les probabilités de défaut joint sont calculer à l'aide de la loi Normale bivariée (ou copule Gaussien avec des marges gaussiennes)

t	probabilités de défaut cumulé		probabilités de défaut joint	
	Notation 4	Notation 7	Méthode de la Normale bivariée	Méthode de la barrière
1	0,1680%	19,0263%	0,1129%	0,0748%
2	0,4967%	32,6291%	0,3864%	0,3397%
3	0,9694%	42,5118%	0,8043%	0,7586%
4	1,5685%	49,8222%	1,3486%	1,2870%
5	2,2766%	55,3368%	2,0010%	1,9176%
6	3,0770%	59,5839%	2,7443%	2,6419%
7	3,9542%	62,9249%	3,5626%	3,4851%
8	4,8940%	65,6088%	4,4419%	4,3658%

Tableau 3 : probabilités de défaut joint de deux entreprises de notation 4 et 7 de corrélation 40%

La première colonne représente les années, la deuxième et la troisième les probabilités de défaut cumulé à l'année i pour les deux entreprises. Les deux dernières colonnes, les probabilités de défaut joint entre l'année 0 et l'année i pour chacune des méthodes.

Remarque : nous avons obtenu ces résultats à l'aide de 20.000.000 de simulations.

Les résultats obtenus à l'aide de la Normale bivariée sont proches mais supérieurs à ceux obtenus à l'aide de la méthode de la barrière affine. On peut ajouter qu'ils convergent quand le nombre d'années tend vers huit.

Je n'ai pas pu expliquer ce phénomène, il est en effet encore à l'étude.

DEROULEMENT DU STAGE

Ma mission principale consistait à analyser l'impact de la structure de dépendance sur un portefeuille de crédit. Pour cela, j'ai procédé en trois étapes :

La première étape consistait à découvrir le sujet et le maîtriser. A partir de recherches sur Internet et d'articles fournis par mes maîtres de stages, j'ai découvert les copules et je me suis familiarisé avec leur concept. Ce fut donc un travail de recherche bibliographique dans le but de déterminer des méthodes de simulation et d'identification de copules.

Les articles sur le sujet étant trop nombreux et répétitifs, la première étape fut donc de « résumer » ces articles et de mettre en lumière les résultats essentiels.

Cette étape a duré environ 1 mois et demi, pendant laquelle j'ai passé la majorité du temps à lire des articles et à étudier les théorèmes exposés.

Maîtrisant mieux le sujet et disposant d'outils pour simuler et identifier les copules, la deuxième étape consistait à tester et à valider les modèles. J'ai répondu par exemple à des questions tels que « combien de points sont nécessaire pour fitter un copule ? », ou bien « le tau de Kendall est-il le meilleur estimateur du coefficient de corrélation linéaire ? ». Lors des réunions avec mes responsables, les questions furent plus précises et plus concrètes. Le but était de répondre aux problèmes que peuvent poser les copules lors de leur possible mise en œuvre dans le simulateur de la Société Générale.

Cette étape fut plus courte que la précédente : j'ai passé environ un mois à simuler des copules par Monte Carlo, à les identifier et à valider la méthodologie mise en place.

Après avoir défini et testé les méthodes d'identification et de simulation des copules, la troisième étape a eu pour but d'appliquer les copules aux problèmes qui se posent en risque de crédit. Concrètement, les copules nous ont permis d'analyser l'impact de la dépendance sur un portefeuille de crédit en terme de capital économique, de capital marginal ainsi qu'en terme d'optimisation d'allocation de capital.

Il me fallut environ un mois pour simuler les différentes configurations de portefeuilles possibles et les analyser.

Pour résumer, j'ai donc débuté mon stage par une période de recherche et de familiarisation avec de nouveaux outils, j'ai ensuite testé et validé ces outils, et j'ai terminé en les appliquant à des cas concrets. Cette façon de procéder ne fut pas prédéfinie, mais elle s'est parfaitement inscrite dans la logique du travail de l'ingénieur, depuis la découverte d'un nouveau sujet jusqu'à son application à un cas pratique.

Ma deuxième mission, plus courte que la première, consistait à calculer les probabilités de défaut joint en utilisant le modèle de Merton avec des barrières affines par morceaux.

Il a fallu que je reproduise une méthode de calcul de barrière présentée dans l'article « Distance to default » afin de l'utiliser pour le calcul des défauts joints.

Contrairement à ma première mission, « l'outil dont je me suis servi » m'était déjà familier car il nous avait été présenté en troisième année lors du cours de Monsieur Gabet sur le calcul stochastique. J'ai donc pu me plonger directement dans le sujet avec le but précis de calculer les probabilités de défaut joint.

L'approche fut donc totalement différente de celle que j'avais suivie lors de la première partie de mon stage. En effet, lors de ma mission précédente, le but était de découvrir les copules et tenter de leur trouver une application inédite en finance : cette approche était donc de la recherche, car nous n'étions pas sûrs d'obtenir un résultat. Lors de la deuxième mission, il a fallu que j'arrive à un résultat à partir d'une nouvelle méthodologie, sachant à l'avance que le problème était soluble et disposant d'une méthode pour le résoudre.

Mes réunions hebdomadaires avec mes maîtres de stages m'ont permis d'aborder avec eux les problèmes que j'ai rencontrés au fur et à mesure de l'avancement de mes travaux. Ils m'ont guidé tout au long de mes recherches en envisageant avec moi l'approche la plus adaptée à la résolution de chacun des problèmes que je leur soumettais.

Mes deux maîtres de stage se sont donc beaucoup investis dans mes recherches et, compte tenu de la complexité de mon sujet, leur aide a été cruciale. Ces réunions étaient d'autant plus intéressantes qu'elles permettaient de confronter deux points de vues différents sur la manière d'aborder les problèmes de mon stage. En outre, la très grande disponibilité de mes maîtres de stage m'a permis de leur demander conseil à chaque fois que je le souhaitais et d'augmenter ainsi l'efficacité de mon travail.

La structure professionnelle et hiérarchique de la division RISQ/RCE ne permettait pas au stagiaire de s'intégrer dans le travail d'équipe. Je participais plus de façon indépendante à la préparation en amont d'un projet. Ma mission ne justifiait pas un réel travail d'équipe puisqu'il s'agissait plus d'étudier les copules au préalable pour une potentielle intégration dans le simulateur de crédit de la Société Générale. Néanmoins, si 90% de mon travail se faisait « indépendamment », il s'agissait de pouvoir recouper mon travail avec les recherches effectuées par un autre stagiaire travaillant sur la structure de dépendance des rendements d'actifs des firmes. En effet, il s'agissait pour lui de pouvoir utiliser de façon effective mon travail pour identifier le copule de Student sur les facteurs du modèle APT. Il a donc fallu poursuivre nos travaux en binôme afin de concevoir et tester l'algorithme « identification du t-copule de Student » (cf. algorithme 6)

Compte tenu de la taille réduite de la division, les rapports avec mes collègues ne se limitaient pas à de simples rapports professionnels. En effet, non seulement les membres de la division déjeunaient quotidiennement ensemble de façon non imposée, mais ils le faisaient plus par réel partage d'affinités communes et création de liens amicaux. Il m'a ainsi été facile de pouvoir m'intégrer au sein de la division à laquelle j'ai été affecté.

Néanmoins, la dimension humaine se trouve réduite à un microcosme car rien ne justifiait le développement de relations avec des employés extérieurs au service. Il se trouve en effet que la division RISK/RCE est relativement indépendante et ne recevait des personnes que de façon ponctuelle.

Compte tenu de l'aspect foncièrement technique de mon stage, il ne me semble pas judicieux de m'attarder ici sur les modalités techniques en tant que telles et déjà relatées dans le corps du rapport, mais plus sur la découverte d'un métier, et les débouchés du domaine d'activité. Je recherchais pour mon stage un secteur qui non seulement pouvait me permettre de mettre en

application des connaissances acquises en mathématiques et finance, mais également de pouvoir travailler sur un sujet « technique ».

Mon stage au sein de la division RISQ/RCE m'a donc apporté entière satisfaction en répondant parfaitement à mes attentes sur le plan « technique », mais j'ai également le sentiment d'avoir pu participer de façon active et intéressante à l'élaboration d'un projet à réelle valeur ajoutée pour l'entreprise. En effet, le copule de Student va pouvoir être intégré au simulateur de la Société Générale, et sera mis en service pour une utilisation effective dans le calcul du capital économique. Il est ainsi valorisant pour un stagiaire de voir ses travaux reconnus et intégrés dans une réalité d'entreprise.

BILAN ET PERSPECTIVES

Les copules sont donc un outil conceptuellement simple mais qui présente un certain nombre d'inconvénients. En effet, l'écriture sous la forme de copule d'une structure de dépendance simple peut être parfois très complexe. De plus, seuls les copules Gaussien et Student sont utilisables dans le cas multidimensionnel.

Le simulateur du portefeuille de crédit de la Société Générale se base sur le modèle APT (cf. annexe). Le modèle APT permet d'expliquer les rendements de toutes les actions du marché en fonction de quelques facteurs, qui rassemblent donc toute l'information commune. Mes travaux ont donc permis d'identifier un copule de Student sur les facteurs. Cela a donc apporté des informations nouvelles par rapport à ce qui était supposé jusqu'alors. En effet, il était admis auparavant que les facteurs suivaient un copule Gaussien.

Fort de ces premiers résultats, la prochaine étape consistera à implémenter dans le simulateur un copule de Student sur les facteurs du modèle APT, ce qui entraînera une augmentation d'environ 40% du capital économique du portefeuille de crédit de la Société Générale

La modélisation des rendements des actifs des firmes par des mouvements browniens, et leur utilisation pour le calcul des probabilités de défaut joint avec un modèle à barrière affine, n'apporte pas les résultats attendus. En effet, les probabilités de défaut joint obtenues avec ce modèle sont inférieures à celles obtenues à l'aide de la loi Normale bivariée. Celle-ci étant réputée sous-estimer la fréquence des défauts joints, nous n'avons pas encore pu aboutir à une explication de ces résultats, le sujet étant encore à l'étude. Là encore, c'est peut être l'usage du copule Gaussien, plutôt que la forme de la barrière, qui est à remettre en cause. Néanmoins, l'intérêt de cette méthode et qu'elle nous permet de disposer de l'évolution moyenne de l'endettement d'une firme en fonction de son rating.

BIBLIOGRAPHIE

- [1] Edward W. Frees, Emiliano A. Valdez (1997). **Understanding relationships using copulas**, *University of Wisconsin-Madison*.
- [2] Paul Embrechts (août 1999), Alexander McNeil, Daniel Straumann. **Correlation and dependence in risk management: properties and pitfalls**, *Working paper, Risklab ETH Zurich*
- [3] Filip Lindskog (mars 2000). **Modelling dependence with copulas**, *Master Thesis,, ETH Zurich(Ecole polytechnique fédérale de Zurich)*
- [4] Eric Bouyé (juillet 2000). **Copulas for finance : A reading guide and some applications**, *Financial Econometrics Research Centre, City University Business School, London*
- [5] V. Durrleman, A. Nikeghbali & T. Roncalli (septembre 2000). **How to get bounds for distribution convolutions? A simulation study and an application to risk management**, *Groupe de Recherche Opérationnelle, Crdit Lyonnais*
- [6] Philipp J. Schonbucher (décembre 2000). **Factor models for portfolio credit risk**, *Department of statistics, Bonn University*
- [7] Filip Lindskog (décembre 2000). **Linear correlation estimation**, *Working paper, Risklab ETH Zurich*
- [8] Rudiger Frey, Alexander J. McNeil (mars 2001), **Modelling dependent defaults : Asset correlations are not enough!**, *Institute of Mathematics of the University of Zurich(ETH)*
- [9] Roberto De Matteis (juin 2001). **Fitting copulas to data**, *Diploma thésis, Institute of Mathematics of the University of Zurich*
- [10] Rudiger Frey, Alexander J. McNeil (aout 2001), **Modelling Dependent Defaults**, *Institute of Mathematics of the University of Zurich(ETH)*
- [11] Rudiger Frey, Alexander J. McNeil (octobre 2001), **Copulas and credit models**, *Institute of Mathematics of the University of Zurich(ETH)*
- [12] Eric Bouyé, Nicolas Gaussel, Mark Salmon (décembre 2001). **Investigating dynamic dependence using copulae**, *Financial Econometrics Research Centre, CUBS, London & HSBC Asset Management Europe*
- [13] Eric Bouyé (janvier 2002). **Multivariate Extremes at Work for portfolio Risk**

Measurement, *Financial Econometrics Research Centre, CUBS, London & HSBC Asset Management Europe*

- [14] Thierry Roncalli (février 2002). **Gestion des risques multiples**, *Groupe de Recherche Opérationnelle, Crdit Lyonnais*
- [15] Roy Mashal, Assaf Zeevi (21 juin 2002). **Beyond Correlation: Extreme Co-movements Between Financial Assets**, *Columbia University*
- [16] Nicolas Gaussel, Jérôme Legras, Francois Longin, Roger Rabemananjara. **Au-dela de la VaR**, *Quants n°37*
- [17] G. Saporta, **Probabilités analyse des données et statistique**, *Edition TECHNIP*
- [18] Marco Avellaneda, Jingyi Zhu (décembre 2001). **Distance to default**, *RISK december 2001*

ANNEXES

COPULES ARICHIMEDIENS A UN PARAMETRE ET LEUR GENERATEUR

TAU DE KENDALL POUR LES COPULES ARCHIMEDIENS A UN PARAMETRE

CALCUL DE ANDERSEN

Soit $W_t = \mu t + B_t$ un mouvement Brownien avec dérive ($\mu > 0$ ici). Nous nous proposons de calculer la probabilité conditionnelle $P(\exists t \in [a, b] / W_t = H \mid W_a, W_b)$. H est un réel donné (barrière) ainsi que a et b ($a < b$). On suppose tout d'abord que W_a et W_b sont tous les deux du même côté par rapport au seuil H , car sinon cette probabilité est égale à 1 par continuité des trajectoires. Supposons donc, par exemple, que $W_a > H$ et $W_b > H$. Réécrivons la probabilité cherchée sous la forme :

$$\begin{aligned} P(\exists t \in [a, b] / W_t = H \mid W_a, W_b) &= P(\inf(W_t, t \in [a, b]) \leq H \mid W_a, W_b) \\ &= P(\inf(Z_t, t \in [0, \delta]) \leq H - W_a \mid W_a, W_b), \end{aligned}$$

où Z_t désigne le processus $W_{t+a} - W_a$ et $\delta = b - a$.

On montre facilement que $Z_t - \mu t$ est un mouvement Brownien standard indépendant de la tribu \mathfrak{S}_a . On peut donc considérer W_a comme une constante dans cette dernière écriture, si bien que

$$\begin{aligned} P(\inf(Z_t, t \in [0, \delta]) \leq H - W_a \mid W_a, W_b) &= P(\inf(Z_t, t \in [0, \delta]) \leq d \mid W_b) \text{ avec } d = H - W_a \\ &= P(\inf(Z_t, t \in [0, \delta]) \leq d \mid Z_\delta). \end{aligned}$$

Mis sous cette forme, le calcul est maintenant immédiat. En effet, puisque Z_t est un mouvement Brownien avec dérive, on connaît la densité du couple formé par les deux variables aléatoires $X = Z_\delta$ et $Y = \inf(Z_t, t \in [0, \delta])$:

$$f_{X,Y}(x,y) = \frac{2(x-2y)}{\sqrt{2\pi\delta^3}} \exp\left(-\frac{(x-2y)^2}{2\delta} + \mu x - \mu^2 \frac{\delta}{2}\right) \text{ pour } y < x, y < 0.$$

Comme Z_δ a pour loi $N(\delta\mu, \delta)$, on a :

$$f_X(x) = \frac{1}{\sqrt{2\pi\delta}} \exp\left(-\frac{(x-\delta\mu)^2}{2\delta}\right) \text{ pour } x \in \mathfrak{R}.$$

La probabilité recherchée s'écrit alors

$$P(\exists t \in [a, b] / W_t = H \mid W_a, W_b) = \int_{y=-\infty}^d \frac{f_{X,Y}(x,y)}{f_X(x)} dy \Big|_{x = Z_\delta = W_b - W_a}.$$

Nous ne détaillons pas le calcul qui est direct et conduit à :

$$P(\exists t \in [a, b] / W_t = H \mid W_a, W_b) = \exp\left(-\frac{2}{\delta}(W_b - H)(W_a - H)\right).$$

Précisons que, par symétrie, cette expression est également valide si on suppose $W_a < H$ et $W_b < H$. Les cas intermédiaires $(W_b - H)(W_a - H) < 0$ donnent une probabilité égale à 1. Remarquons également que cette probabilité conditionnelle ne fait pas apparaître le paramètre de dérive μ , alors que la probabilité non conditionnelle en dépend : en fait, la dérive se trouve résumée par la donnée des variables W_a et W_b dont les lois (normales) dépendent effectivement de μ .

LE MODELE APT

Le modèle APT permet d'expliquer les rendements de toutes les actions du marché en fonction de quelques facteurs, qui rassemblent donc toute l'information commune :

$$(0.1) \tilde{R}_j = r_f + \sum_{k=1}^H \beta_{jk} (\tilde{X}_k - r_f) + \tilde{\varepsilon}_j$$

Avec :

X_k	Rentabilité du $k^{\text{ème}}$ facteur.
β_{jk}	Sensibilité de l'action j à ce facteur.
ε_j	Partie spécifique du rendement de la firme, orthogonale aux facteurs.
r_f	Rendement de l'actif sans risque.

L'APT modélise les prix en fonction de facteurs mais n'indique jamais quels sont ces facteurs. Un certain nombre de questions restent ouvertes : combien doit-on prendre de facteurs, comment les identifier, sont-ils uniques ?

En 1986, Chen, Roll et Ross identifient quatre facteurs principaux : l'inflation non anticipée, la croissance du PIB, la pente de la structure par terme des taux d'intérêt, et la prime de risque sur les marchés financiers.

Mais ceux-ci ne donnent pas satisfaction, ce qui justifie que l'on tente d'identifier d'autres facteurs par des méthodes de traitement du signal.