

## **Copules et dépendances : application pratique à la détermination du besoin en fonds propres d'un assureur non vie**

**David Cadoux**

**Institut des Actuaires (IA)**

GE Insurance Solutions

107 rue Saint-Lazare, 75009 Paris FRANCE

Tél. : +331 55 07 13 91

Fax : +331 55 07 14 02

E-mail : david.cadoux@ge.com

**Jean-Marc Loizeau**

**Institut des Actuaires (IA)**

MAAF Assurances

Chauray, 79036 Niort Cedex 9 FRANCE

Tél. : +335 49 34 43 00

Fax : +335 49 17 55 83

E-mail : jean-marc.loizeau@maaf.fr

### **Résumé**

La modélisation des dépendances entre lignes d'affaire d'un portefeuille d'assurance non vie a souvent fait l'objet d'hypothèses simplificatrices. Les copules sont un outil flexible qu'il est désormais possible d'utiliser pour représenter la dépendance entre risques de façon réaliste. Cet article présente un cas pratique d'application de la théorie des copules à la détermination du besoin en fonds propres d'un assureur non vie du marché français tenant compte du seul risque de fluctuation des sinistres. Une méthodologie spécifique de recherche des dépendances au sein du portefeuille d'assurance est mise en place. Les copules représentant le mieux ces interactions sont ensuite sélectionnées à l'aide d'un test d'adéquation du chi-deux. Des simulations de Monte-Carlo permettent enfin de mesurer l'impact des dépendances sur le besoin en fonds propres de la compagnie, ce dernier étant évalué à l'aide de plusieurs mesures de risque.

**Mots-clefs** : besoin en fonds propres, mesures de risque, dépendance, copule, tau de Kendall, statistique du chi-deux, Monte-Carlo

### **Abstract**

The modeling of dependent business lines of a non life insurance portfolio has often been based on a set of simplified assumptions. Copulae are a versatile tool which can now be used to represent dependence between risks in a realistic way. This study presents a practical case where copulae theory is applied to assess the capital adequacy of a French non life insurer only adjusted for the risk of volatility in insurance losses. Firstly, a specific methodology is developed to detect dependencies embedded in the portfolio. Secondly, copulae that better reflect those links are selected through a chi-square goodness-of-fit test. Finally, Monte-Carlo simulations are used to measure the impact of dependencies on the capital adequacy of the insurer, the latter being assessed thanks to several risk measures.

**Keywords** : capital adequacy, risk measures, dependency, copula, Kendall's Tau, Chi-square statistic, Monte-Carlo

## **INTRODUCTION**

L'activité d'assurance requiert de disposer d'un niveau minimum de fonds propres pour absorber les mouvements défavorables de résultats non anticipés. La détermination de ce montant minimal est devenue une problématique majeure pour les assureurs, notamment depuis les deux dernières décennies. En effet, auparavant les marchés étaient régulés, moins volatils et les rendements exigés par les actionnaires étaient de fait moins élevés. Progressivement, la compétition s'est accrue, la sinistralité a augmenté avec une plus grande concentration des risques assurés, l'environnement juridique est devenu de plus en plus incertain, les mouvements capitalistiques et la notion de création de valeur pour l'actionnaire sont apparus. Dans ce nouveau contexte, les assureurs sont désormais fortement incités à développer une gestion optimale de leurs fonds propres qui doit satisfaire des intérêts divergents. D'un côté, les autorités de contrôle, les agences de notation, et dans une mesure croissante les assurés, recherchent une solvabilité maximale de la société et donc plus de fonds propres. De l'autre, les actionnaires espèrent maximiser le rendement de leur investissement en limitant le capital injecté dans la société.

La recherche d'une optimalité du niveau des fonds propres soulève une question importante : l'optimum souhaité reflète-t-il des dépendances potentielles entre les risques supportés par la compagnie ? Récemment encore les réponses actuarielles apportées à cette question étaient basées sur des hypothèses fortes : soit les risques étaient supposés indépendants, soit on admettait leur dépendance mais en leur imposant de suivre la même loi de distribution. Désormais, une modélisation plus réaliste des liaisons entre risques est possible grâce à la théorie des copules, objet d'un vif regain d'intérêt depuis quelques années.

L'impact de la dépendance entre risques sur la détermination du besoin en fonds propres d'un assureur n'a été analysé à l'aide des copules que dans un cadre théorique. Cet article présente une mise en œuvre pratique de cette problématique appliquée au portefeuille non vie d'un assureur du marché français, en se concentrant sur la dépendance entre charges sinistres de plusieurs garanties.

La première partie sera consacrée aux mesures de risque. On discutera de leurs propriétés et on présentera celles qui peuvent être retenues pour déterminer le besoin en fonds propres. Dans la deuxième partie, on rappellera les principaux résultats autour de la dépendance et des copules, et on proposera notamment une méthode de sélection de la meilleure copule basée sur la statistique du chi-deux. La troisième partie présentera les données du portefeuille d'assurance analysé, décrira une méthodologie permettant d'exhiber des dépendances intra-branche et inter-branches au sein de ce portefeuille et mettra en exergue les copules décrivant le mieux ces dépendances. Au cours de la dernière partie, on modélisera le besoin en fonds propres via diverses mesures de risque, et on évaluera via des simulations de Monte-Carlo comment il est affecté par la prise en compte de dépendances modélisées par diverses copules.

## **Partie 1. MESURES DE RISQUE**

Le principal outil théorique pour calculer le besoin en fonds propres est défini sous le vocable « mesures de risque ». Certaines de ces mesures de risque sont manipulées depuis fort longtemps par les actuaires, plus particulièrement dans le domaine de la tarification. Les mesures de risque utilisées pour le besoin en fonds propres ont donné lieu à de nombreux travaux actuariels ces dernières années. De manière générale, elles visent à fixer un niveau de capital pour un portefeuille de risques donné, et mesurent le risque en un ou plusieurs nombres. Dans ce qui suit on donnera la définition d'une mesure de risque et les propriétés associées qui peuvent être recherchées pour évaluer le besoin en fonds propres, puis on présentera les principales mesures de risque permettant de fixer un niveau optimal de fonds propres.

### **1.1. Définition**

Soit  $\Omega$  l'ensemble fini des états de nature possibles, on appelle variable aléatoire réelle une fonction  $X$  qui à un état de la nature  $\omega$  associe le réel  $X(\omega)$ . En assurance, le résultat d'une compagnie sur une période donnée ou la charge agrégée de sinistres (d'une ou plusieurs garanties) peuvent être formalisés de cette manière.  $X$  pourra être caractérisée par sa fonction de répartition  $F_X: F_X(x) = P\{\omega / X(\omega) \leq x\} = P\{X \leq x\}$ .

On appelle mesure de risque une fonction  $\rho$  associant à un risque  $X$  un réel positif  $\rho(X)$ . Le besoin en capital d'une compagnie d'assurance non vie peut être formalisé en utilisant une telle fonction.  $X$  représentera dans ce qui suit le résultat<sup>1</sup> de la société, le besoin en capital étant fonction de cette variable.

### **1.2. Propriétés**

Une mesure de risque doit pouvoir vérifier un certain nombre de propriétés élémentaires. A ce jour, il n'existe pas de consensus dans la littérature actuarielle sur les propriétés que doit nécessairement respecter une mesure de risque. Nous rappelons ci-après le corps des propriétés pouvant s'appliquer aux mesures de risque utilisées pour estimer le besoin en capital et présentons une traduction simple de ces propriétés en nous inspirant des travaux de De La Foata et Odjo [2001]. Il existe d'autres propriétés qui ne sont pas présentées car elles sont plus spécifiques au calcul du chargement de la cotisation. Pour deux risques quelconques  $X$  et  $Y$ , les propriétés suivantes peuvent être formulées :

#### **Invariance par translation**

$$\rho(X + c) = \rho(X) + c, \text{ pour toute constante } c.$$

Si on ajoute (respectivement on retranche) un montant certain  $c$  au résultat d'un centre de profit<sup>2</sup>, le besoin en capital décroît (respectivement augmente) du même montant.

#### **Sous-additivité**

$$\rho(X + Y) \leq \rho(X) + \rho(Y)$$

---

<sup>1</sup> Volontairement le résultat n'est pas défini ici car il dépend du modèle d'évaluation du besoin en capital mis en œuvre.

<sup>2</sup> Le centre de profit peut être soit la compagnie dans son ensemble soit limité à une branche d'activité donnée.

La fusion de deux centres de profit ne crée pas de risque supplémentaire. Au contraire, la diversification tend à réduire le risque global. Cette propriété permet ainsi une gestion décentralisée du besoin en capital dans les différents centres de profit sans courir le risque d'un besoin global supérieur à la somme des besoins individuels de chacun des centres. Si cette propriété n'était pas respectée, une société ne respectant pas un certain niveau requis de capital pourrait être incitée à se scinder artificiellement en deux entités afin de réduire son besoin en capital.

#### **Homogénéité positive**

$$\rho(\lambda X) = \lambda \rho(X), \quad \forall \lambda \geq 0$$

De même qu'une fusion ne crée pas de risque supplémentaire ( $\rho(\lambda X) \leq \lambda \rho(X)$ ), une fusion sans diversification ne réduit pas le besoin global en capital.

#### **Monotonie**

$$X \leq Y \Rightarrow \rho(X) \geq \rho(Y)$$

Si les pertes encourues avec le risque  $X$  sont toujours supérieures à celles obtenues avec  $Y$ , le besoin en capital pour  $X$  doit être supérieur à celui pour  $Y$ .

#### **Propriété de borne supérieure**

$\rho(X) \leq \max(-X)$ , le besoin en capital est borné par la perte maximale possible.

#### **Conservatisme**

$\rho(X) = \rho(X^-)$  où  $X^-$  désigne la partie négative de  $X$ , ( $X^- = \min(X, 0)$ ).

Elle est définie par Artzner *et al.* [1999] notamment pour la détermination du besoin en capital, et conduit à ne prendre en compte que les valeurs négatives de  $X$ , traduisant ainsi une forte aversion au risque.

#### **Propriété de sévérité**

Le besoin en capital doit tenir compte des amplitudes possibles des pertes.

Selon Artzner *et al.* [1999], une mesure de risque vérifiant les quatre premières propriétés est dite cohérente. Selon ces auteurs, une mesure de risque appropriée doit respecter le critère de cohérence. Ce corps de propriétés définissant une mesure cohérente ne fait pas l'unanimité dans la communauté actuarielle. On peut montrer en effet qu'il existe des situations où le non respect d'une des propriétés est préférable. Le principe d'homogénéité positive est remis en question lorsqu'une variation d'échelle du risque conduit à un effet plus que proportionnel sur le besoin en capital, lié, par exemple, à une difficulté pour se réassurer. Un autre exemple construit par Perronnet *et al.* [2001] montre que la diversification, recherchée à travers l'agrégation des risques, limite bien la dispersion du risque mais n'augmente pas nécessairement la probabilité d'être solvable. Dans l'exemple mis en œuvre, un agent basant son critère de souscription sur l'écart-type, en raison de sa sous-additivité, réduit effectivement la dispersion du risque mais augmente sensiblement la probabilité d'être insolvable.

### **1.3. Principales mesures de risque pour évaluer le besoin en fonds propres**

Cette section décrit les caractéristiques, avantages et inconvénients des principales mesures de risque susceptibles d'être mises en œuvre pour déterminer le besoin en fonds propres.

## Probabilité de ruine

La probabilité de ruine dans sa version primaire correspond au percentile du point au-delà duquel le capital initial est totalement épuisé sur une période donnée suite à un résultat déficitaire. Autrement dit, si on note  $C$  le capital initial donné, la probabilité de ruine est égale à  $F_X(-C) = P(X < -C)$ . Assez logiquement, elle a aussi été utilisée dans les modèles assurance en fixant un niveau minimum de probabilité de ruine,  $\alpha$ , jugé acceptable selon l'aversion au risque de l'agent et en déterminant ensuite le niveau de capital qui lui correspond. Dans ce cas, on a alors  $F_X(-C) = \alpha$  et  $C$  est cette fois déterminé en fonction de  $\alpha$ . La probabilité de ruine issue des modèles d'assurance est à l'origine de la Value at Risk (VaR), bien connue du monde bancaire.

## Value at Risk (VaR)

La VaR est une mesure de risque commune aux organismes financiers qui est totalement équivalente au concept de Sinistre Maximum Probable (SMP), très familier des assureurs non vie. Elle est égale à la perte maximale que peut subir une organisation, dans des conditions normales de marché, sur une période de temps donnée pour un certain niveau de probabilité  $\alpha$ . On a donc  $VaR_\alpha(X) = -\inf\{x : F_X(x) \geq \alpha\}$ .

De manière pratique, la mise en œuvre de la VaR consiste à faire varier la probabilité  $\alpha$  et à déterminer le capital minimal associé en début de période qui permette de faire face à la perte maximale en fin de période. On obtient donc rétrospectivement le niveau de capital en deçà duquel l'organisation est en ruine avec une probabilité supérieure à  $\alpha$ . La VaR utilisée dans ce contexte n'est autre que la probabilité de ruine évoquée plus haut.

L'avantage de la VaR est d'être un concept simple et facile à calculer. Elle vérifie les propriétés d'invariance par translation, d'homogénéité positive, de monotonie, de borne supérieure et de conservatisme. En revanche, elle n'est pas sous-additive et ne tient pas compte de la sévérité de la ruine, ce qui constitue la critique la plus souvent formulée à son égard. Cependant cette mesure de risque peut faire sens dans un objectif de solvabilité.

## TailVaR (TVaR)

La TailVaR (TVaR) aussi appelée Expected Shortfall, Conditional Tail Expectation ou bien encore Conditional VaR, est une mesure de risque définie de la manière suivante :

$$TVaR_\alpha(X) = VaR_\alpha(X) + \frac{P(X < -VaR_\alpha(X))}{\alpha} \cdot E[-X - VaR_\alpha(X) | X < -VaR_\alpha(X)]$$

Cette expression s'explique par le fait que pour une distribution discrète, on peut avoir  $P(X < -VaR_\alpha(X)) < \alpha$ . Dans la pratique, on se place généralement dans le cas où  $F_X(x)$  est continue<sup>3</sup> et on a alors

$$\begin{aligned} TVaR_\alpha(X) &= VaR_\alpha(X) + E[-X - VaR_\alpha(X) | X < -VaR_\alpha(X)] \\ &= VaR_\alpha(X) + E[-X | X < -VaR_\alpha(X)] - VaR_\alpha(X) \\ &= E[-X | X < -VaR_\alpha(X)] \end{aligned}$$

Par construction, cette mesure de risque est très sensible à la queue de distribution et est de fait plus conservatrice que la VaR car pour un seuil  $\alpha$  on a toujours  $TVaR_\alpha(X) > VaR_\alpha(X)$ .

---

<sup>3</sup> Dans la pratique,  $F_X(x)$  est le plus souvent inconnue et approchée par simulations. Grâce aux capacités informatiques actuelles, on peut disposer aujourd'hui d'un très grand nombre de simulations qui permet de considérer que l'on a un continuum d'observations.

Elle est également facile à expliquer et présente la particularité de respecter un très grand nombre de propriétés puisqu'elle est non seulement cohérente au sens d'Artzner *et al.* [1999] mais elle vérifie aussi les propriétés de borne supérieure, de conservatisme et de sévérité. Ceci explique qu'elle soit la mesure de risque la plus souvent recommandée.

### **Excess TailVaR (XTVaR)**

La XTVaR n'est autre que la ruine moyenne au-delà d'un certain seuil :

$$XTVaR_\alpha(X) = E[-X - VaR_\alpha(X) \mid X < -VaR_\alpha(X)]$$

Cette mesure de risque, aussi appelée Mean Excess Loss, est équivalente à l'espérance de vie résiduelle utilisée dans les modèles d'actuariat vie. Ce n'est autre que la différence entre la TVaR et la VaR :  $TVaR_\alpha(X) = VaR_\alpha(X) + XTVaR_\alpha(X)$ .

Si l'on pose  $\beta = VaR_\alpha(X)$ , alors on a  $XTVaR_\alpha(X) = TVaR_\alpha(X) - \beta$  et l'on voit que par construction la XTVaR ne peut vérifier les propriétés de sous additivité, d'homogénéité positive et de monotonie.

## **Partie 2. DEPENDANCE ET COPULES**

Supposer l'indépendance entre les risques d'un portefeuille d'assurance est une hypothèse forte qui doit être testée. Les copules présentent de nombreux avantages pour modéliser la dépendance entre risques. D'une part, elles permettent de décrire le comportement individuel de chaque risque et « couplent » les lois marginales pour obtenir la loi jointe. D'autre part, elles offrent une représentation fonctionnelle de la dépendance qui donne une description très complète de la forme de cette dernière. Dans cette partie sont présentés les résultats fondamentaux de la théorie des copules.

### **2.1. Dépendance ou corrélation ?**

Il est important de rappeler que la dépendance et la corrélation sont des notions différentes. En effet, on a  $X$  et  $Y$  indépendantes  $\Rightarrow X$  et  $Y$  non corrélées ou  $\rho(X, Y) = 0$  mais la réciproque est fautive sauf dans le cas où les variables sont gaussiennes car la dépendance est alors entièrement caractérisée par le coefficient de corrélation. Le contre-exemple le plus connu dans la littérature est le suivant : soit  $X \sim N(0,1)$  et  $Y = X^2$ , alors  $cov(X, Y) = E(X^3) = 0$ .

Bien qu'il soit facile à calculer et fréquemment présent dans les travaux actuariels, en assurance comme en finance, le coefficient de corrélation doit être utilisé avec précaution car il n'est pertinent qu'en présence de distributions elliptiques (distribution multivariée Normale ou de Student) ou de dépendance linéaire. Les erreurs d'interprétation et limites liées à son utilisation sont discutées dans Embrechts *et al.* [1999].

### **2.2. Qu'est-ce qu'une copule ?**

#### **Définition 1 : copule**

Une copule est une fonction de répartition multivariée  $C$  définie sur l'hypercube  $[0;1]^n$  et dont les marginales sont uniformes sur  $[0;1]$ .

### **Théorème 1 - Théorème de Sklar [1959]**

Soit  $F$  une fonction de répartition  $n$ -dimensionnelle avec des marginales  $F_1, \dots, F_n$ , alors il existe une  $n$ -copule  $C$  telle que pour tout  $x$  de  $R^n$ ,  $F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$ .

Le théorème de Sklar appliqué à des variables aléatoires continues conduit au théorème suivant.

### **Théorème 2**

Soit  $(X_1, \dots, X_n)$  un vecteur de variables aléatoires continues admettant  $F_1, \dots, F_n$  comme fonctions de répartition marginales et  $F$  comme fonction de répartition jointe, alors il existe une copule  $C$  qui vérifie la relation du théorème 1. Si les marginales  $F_1, \dots, F_n$  sont continues, alors  $C$  est unique, autrement  $C$  est uniquement déterminée sur  $\text{Im}(F_1) \times \dots \times \text{Im}(F_n)$ . ( $\text{Im}(X)$  représentant l'ensemble des valeurs prises par  $X$ )

A partir du théorème 2, on voit qu'une copule permet d'exprimer une fonction de répartition multivariée selon ses marginales et que cette copule résume toute la structure de dépendance. Une copule n'est autre qu'une fonction de répartition dont les lois marginales sont uniformes, ce qui conduit à une expression probabiliste simple de la copule :

$$C(u_1, \dots, u_n) = \Pr(U_1 \leq u_1, \dots, U_n \leq u_n).$$

Avec les copules, on ne travaille plus en nombre ou montant mais en rang. Autrement dit, dans la pratique on transforme linéairement les réalisations  $x_1, \dots, x_n$  en uniformes empiriques

$$u_1, \dots, u_n \text{ où } u_i = \frac{\text{Rang}(x_i)}{n+1} \text{ pour tout } i \text{ de } 1 \text{ à } n.$$

### **Définition 2 : densité d'une copule**

La densité  $c$  d'une copule  $C$ , si elle existe, est définie comme suit

$$c(u_1, \dots, u_n) = \frac{\partial^n C}{\partial u_1 \dots \partial u_n}(u_1, \dots, u_n).$$

## **2.3. Mesures de concordance**

Il convient tout d'abord de rappeler la notion de concordance.

Soient  $(x, y)$  et  $(\tilde{x}, \tilde{y})$  deux réalisations d'un vecteur aléatoire continu  $(X, Y)$ , alors  $(x, y)$  et  $(\tilde{x}, \tilde{y})$  sont dites concordantes si  $(x - \tilde{x})(y - \tilde{y}) > 0$  et discordantes si  $(x - \tilde{x})(y - \tilde{y}) < 0$ .

### **Théorème 3**

Soit  $\kappa$  une mesure de concordance pour des variables aléatoires continues  $X$  et  $Y$ .

1. si  $Y$  est une fonction croissante de  $X$ , alors  $\kappa\langle X, Y \rangle = 1$ ;
2. si  $Y$  est une fonction décroissante de  $X$ , alors  $\kappa\langle X, Y \rangle = -1$ ;
3. si  $\alpha$  et  $\beta$  sont des fonctions strictement croissantes, alors  $\kappa\langle \alpha(X), \beta(Y) \rangle = \kappa\langle X, Y \rangle$ .

Il est facile de construire des exemples où le coefficient de corrélation linéaire de Pearson n'est pas invariant par transformation strictement croissante (voir Embrechts *et al.* [2001]). Par conséquent, la corrélation linéaire n'est pas une mesure de concordance. En revanche, le

tau de Kendall et le rho de Spearman sont deux mesures de concordance bien connues en statistique. Elles donnent une mesure de la corrélation entre les rangs des observations, à la différence du coefficient de corrélation linéaire qui lui apprécie la corrélation entre les valeurs des observations. Elles offrent par ailleurs l'avantage de s'exprimer simplement en fonction de la copule associée au couple de variables aléatoires.

**Définition 3 : tau de Kendall**

Soient  $(X, Y)$  un couple de vecteurs aléatoires et  $(X', Y')$  une copie de  $(X, Y)$ , c'est-à-dire un couple de vecteurs en tout point identique à  $(X, Y)$ , le tau de Kendall s'écrit alors :

$$\tau(X, Y) = \Pr\{(X - X')(Y - Y') > 0\} - \Pr\{(X - X')(Y - Y') < 0\}.$$

Le tau de Kendall n'est autre que la différence entre la probabilité de concordance et celle de discordance. Son expression en terme de copule est la suivante :

**Théorème 4**

Soit  $(X, Y)$  un couple de variables aléatoires continues de copule  $C$ , alors

$$\tau(X, Y) = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1 = 1 - 4 \iint_{[0,1]^2} \partial_u C(u, v) \partial_v C(u, v) dudv.$$

ce qui peut s'écrire encore  $\tau(X, Y) = 4E(C(U, V)) - 1$ , avec  $U, V \sim U(0,1)$ .

Un estimateur du tau de Kendall se construit à partir d'un échantillon  $\{(x_1, y_1), \dots, (x_T, y_T)\}$  de  $(X, Y)$  de la façon suivante :

$$\hat{\tau}(X, Y) = \frac{2}{T(T-1)} \sum_{j=2}^T \sum_{i=1}^{j-1} \text{sign}\{(x_j - x_i)(y_j - y_i)\} \quad \text{où } \text{sign}(z) = \begin{cases} 1 & \text{si } z \geq 0 \\ -1 & \text{si } z < 0 \end{cases}$$

**Définition 4 : rho de Spearman**

Soient  $(\tilde{X}, \tilde{Y})$  et  $(X', Y')$  deux couples de vecteurs aléatoires copies d'un vecteur aléatoire  $(X, Y)$ , alors le rho de Spearman est égal à :

$$\rho_s(X, Y) = 3 \left[ \Pr\{(X - \tilde{X})(Y - Y') > 0\} - \Pr\{(X - \tilde{X})(Y - Y') < 0\} \right]$$

Le rho de Spearman s'écrit aussi en fonction du coefficient  $\rho$  de corrélation linéaire de Pearson :  $\rho_s(X, Y) = \rho(F_X(X), F_Y(Y))$  où  $F_X$  et  $F_Y$  sont les fonctions de répartition respectives de  $X$  et  $Y$ . Son expression en terme de copule est la suivante :

**Théorème 5**

Soit  $(X, Y)$  un couple de variables aléatoires continues de copule  $C$ , alors

$$\rho_s(X, Y) = 12 \iint_{[0,1]^2} uv dC(u, v) - 3 = 12 \iint_{[0,1]^2} C(u, v) dudv - 3.$$

Un estimateur du rho de Spearman se construit à partir d'un échantillon  $\{(x_1, y_1), \dots, (x_T, y_T)\}$  de  $(X, Y)$  de la façon suivante :

$$\hat{\rho}_s(X, Y) = \frac{\sum_{i=1}^T (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^T (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^T (S_i - \bar{S})^2}} \quad \text{où } R_i \text{ est le rang de } x_i, S_i \text{ celui de } y_i \text{ et } \bar{Z} = \frac{1}{T} \sum_{i=1}^T Z_i.$$



## 2.4. Dépendance de queue

Le concept de dépendance de queue fournit une description de la dépendance au niveau des queues de distribution, très intéressante pour étudier la survenance simultanée de valeurs extrêmes. C'est une mesure locale contrairement au tau de Kendall et au rho de Spearman qui mesurent la dépendance sur l'ensemble de la distribution.

### Définition 5 : coefficients de dépendance de queue

Le coefficient de dépendance de queue inférieure ou *lower tail dependence coefficient* de deux variables aléatoires  $X$  et  $Y$ , de fonctions de répartition respectives  $F_X$  et  $F_Y$ , est défini par  $\lambda_L(X, Y) = \lim_{\alpha \rightarrow 0^+} \Pr[X \leq F_X^{-1}(\alpha) \mid Y \leq F_Y^{-1}(\alpha)]$  (si cette limite existe).

Le coefficient de dépendance de queue supérieure ou *upper tail dependence coefficient* de deux variables aléatoires  $X$  et  $Y$ , de fonctions de répartition respectives  $F_X$  et  $F_Y$ , est défini par  $\lambda_U(X, Y) = \lim_{\alpha \rightarrow 1^-} \Pr[X > F_X^{-1}(\alpha) \mid Y > F_Y^{-1}(\alpha)]$  (si cette limite existe).

On dit que  $X$  et  $Y$  sont asymptotiquement dépendantes au niveau supérieur de la queue de distribution si  $\lambda_U \in (0; 1]$  et asymptotiquement indépendantes au niveau supérieur de la queue de distribution si  $\lambda_U = 0$ . Lorsque  $\lambda_U > 0$ , une interprétation simple peut être donnée lorsqu'on étudie la sinistralité extrême concomitante sur deux branches d'assurance : sachant qu'un sinistre extrême est survenu dans une branche, il existe une probabilité non nulle qu'un sinistre d'une intensité relative comparable survienne concomitamment dans l'autre branche.

La notion de copule de survie est très utile pour l'étude de la dépendance de queue. La définition en est la suivante.

### Définition 6 : copule de survie

Soit  $\tilde{C}(u_1, \dots, u_n)$  la fonction définie par

$$\tilde{C}(u_1, \dots, u_n) = \bar{C}(1 - u_1, \dots, 1 - u_n) \quad \text{où} \quad \bar{C}(u_1, \dots, u_n) = \Pr[U_1 > u_1, \dots, U_n > u_n].$$

Alors  $\tilde{C}(u_1, \dots, u_n)$  est appelée copule de survie (*survival copula*) de la copule  $C$ .

### Définition 7

Soient  $X$  et  $Y$  deux variables aléatoires de copule  $C$ , alors on a

$$\lambda_L(X, Y) = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u} \quad \text{et} \quad \lambda_U(X, Y) = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

On montre facilement que le coefficient de dépendance de queue inférieure (resp. supérieure) de  $C$  est le coefficient de dépendance de queue supérieure (resp. inférieure) de  $\tilde{C}$ . Autrement dit, à partir d'une copule donnée, il est possible de créer une autre copule présentant une structure de dépendance de queue inversée.

## 2.5. Copules archimédiennes

Volontairement les copules gaussienne et de Student, dites copules elliptiques, ne seront pas présentées dans cet article. Elles sont en effet moins bien adaptées en assurance car elles s'appliquent à des distributions symétriques. Le lecteur pourra néanmoins en trouver une présentation dans Cadoux *et al* [2004]. Les copules archimédiennes ont le grand avantage de décrire des structures de dépendance très diverses dont notamment les dépendances dites asymétriques, où les coefficients de queue inférieure et de queue supérieure diffèrent.

### Définition 8 : copules archimédiennes

Soit  $\varphi$  une fonction convexe, continue, strictement décroissante de  $[0; 1]$  dans  $[0; \infty[$  telle que  $\varphi(1) = 0$  et  $\varphi(0) = \infty$  alors  $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$  est une copule archimédienne stricte et  $\varphi$  est appelé générateur strict de  $C$ .

Notons que cette méthode de génération de copules peut être facilement étendue en dimension  $n$ . Elle présente un double intérêt. D'une part, elle permet de construire une grande variété de familles de copules. D'autre part, les copules ainsi générées ont des formes analytiques closes. Le tableau suivant regroupe les caractéristiques des familles les plus connues. On rappelle que le paramètre  $a$  mesure le degré de dépendance entre les risques. Plus il est élevé plus la dépendance est forte et une valeur positive de  $a$  indique une dépendance positive.

| Copule         | $\varphi(u)$  | $C(u, v)$  | $\lambda_L$        | $\lambda_U$   | $\tau_a$                      |
|----------------|---|--|--------------------|---------------|-------------------------------|
| <b>Gumbel</b>  | $(-\ln u)^a, a \geq 1$                                      | $\exp\left(-\left[(-\ln u)^a + (-\ln v)^a\right]^{1/a}\right)$                     | 0                  | $2 - 2^{1/a}$ | $1 - 1/a$                     |
| <b>Franck</b>  | $-\ln\left(\frac{e^{-au} - 1}{e^{-a} - 1}\right), a \neq 0$ | $-\frac{1}{a} \ln\left[1 + \frac{(e^{-au} - 1)(e^{-av} - 1)}{(e^{-a} - 1)}\right]$ | 0                  | 0             | $1 - \frac{4(1 - D_1(a))}{a}$ |
| <b>Clayton</b> | $\frac{(u^{-a} - 1)}{a}, a > 0$                             | $(u^{-a} + v^{-a} - 1)^{-\frac{1}{a}}$   | $2^{-\frac{1}{a}}$ | 0             | $\frac{a}{a + 2}$             |

La copule de Gumbel n'appréhende que des dépendances positives et possède la caractéristique de pouvoir représenter des risques dont la structure de dépendance est plus accentuée sur la queue supérieure. Elle est à ce titre particulièrement adaptée en assurance et en finance pour étudier l'impact de la survenance d'événements de forte intensité sur la dépendance entre branches d'assurance ou actifs financiers.

La copule de Franck permet de modéliser les dépendances aussi bien positives que négatives. On note qu'il n'existe pas de dépendance de queue pour cette copule. Le tau de Kendall s'exprime en fonction de  $a$  grâce à la fonction Debye définie par

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt.$$

Comme la copule de Gumbel, la copule de Clayton ne permet de modéliser que les dépendances positives. A l'inverse de la copule de Gumbel, elle vise à rendre compte d'une dépendance sur les événements de faible intensité.

## 2.6. Copule HRT

La copule HRT n'appartient pas à la famille des copules archimédiennes. Elle a été introduite par Venter [2001] pour modéliser la dépendance sur des événements extrêmes de forte intensité. Elle a été construite comme étant la copule de survie de la copule de Clayton et présente donc une structure de dépendance inversée.

$$C(u, v) = u + v - 1 + \left( (1-u)^{-a} + (1-v)^{-a} - 1 \right)^{-\frac{1}{a}}, \text{ et } \lambda_L = 0, \lambda_U = 2^{\frac{-1}{a}}, \tau_a = \frac{a}{a+2}.$$

## 2.7. Copule empirique

**Définition 9**, Deheuvels [1979] a introduit la notion de copule empirique :

Soit  $\{r'_1, \dots, r'_n\}$  la statistique de rang associée à l'échantillon multivarié  $\{x'_1, \dots, x'_n\}$  dont chaque vecteur possède  $J$  observations ( $\forall i \in [1; n]$ ,  $r'_i$  est le rang de  $x'_i$  parmi  $(x'_i)_{l=1 \dots J}$ ) alors toute

copule  $\hat{C}$  définie sur le treillis  $\left\{ \left( \frac{t_1}{T}, \dots, \frac{t_n}{T} \right) \mid 1 \leq k \leq n, 0 \leq t_k \leq T \right\}$  par

$$\hat{C} \left( \frac{t_1}{T}, \dots, \frac{t_n}{T} \right) = \frac{1}{T} \sum_{l=1}^T \prod_{k=1}^n 1_{\{r'_k \leq t_k\}} \text{ est une copule empirique.}$$

Un exemple simple de calcul de copule empirique en dimension 2 avec  $J=T=3$  est donné dans Cadoux *et al.* [2004]. Il est important de noter que la copule empirique n'est pas unique mais que toutes les copules empiriques prennent les mêmes valeurs sur le treillis. Par ailleurs, si l'on note  $\hat{C}_T$  la copule empirique d'ordre  $T$ , où  $T$  représente la taille de l'échantillon utilisé pour sa construction, alors  $\hat{C}_T$  converge asymptotiquement vers  $C$  (voir Faivre [2002]).  $\hat{C}$  offre l'avantage de s'affranchir des fonctions de répartition marginales. En revanche, l'estimateur empirique naturel de la copule fait lui appel aux lois marginales et s'écrit  $\tilde{C}(u_1, \dots, u_n) = \tilde{F}(\tilde{F}_1^{-1}(u_1), \dots, \tilde{F}_n^{-1}(u_n))$  où  $\tilde{F}$  est la fonction de répartition empirique jointe et  $\tilde{F}_i$  la fonction de répartition empirique marginale.

Il est également possible de définir une densité empirique  $\hat{c}$  (équivalente à la densité de Radon-Nikodym) pour la copule empirique  $\hat{C}$  :

$$\hat{c} \left( \frac{t_1}{T}, \dots, \frac{t_n}{T} \right) = \sum_{i_1=1}^2 \dots \sum_{i_n=1}^2 (-1)^{i_1 + \dots + i_n} \hat{C} \left( \frac{t_1 - i_1 + 1}{T}, \dots, \frac{t_n - i_n + 1}{T} \right).$$

Elle est parfois appelée *empirical copula frequency* dans les travaux anglo-saxons et représente la probabilité d'appartenance à un hypercube de  $[0; 1]^n$  de longueur d'arrête  $\frac{1}{T}$ .

Il est possible d'exprimer directement  $\hat{C}$  en fonction de  $\hat{c}$  de la manière suivante :

$$\hat{C} \left( \frac{t_1}{T}, \dots, \frac{t_n}{T} \right) = \sum_{i_1=1}^{t_1} \dots \sum_{i_n=1}^{t_n} \hat{c} \left( \frac{i_1}{T}, \dots, \frac{i_n}{T} \right).$$

$\hat{C}$  et  $\hat{c}$  sont discontinues par construction et leur discontinuité s'accroît lorsque  $T$  diminue. Dans la pratique, leur temps de calcul, croissant avec  $T$ , peut être assez long. Il est conseillé de ne pas choisir une valeur trop grande de  $T$  afin de limiter les temps de traitement informatique tout en veillant à ne pas retenir une valeur trop faible qui conduirait à des résultats trop discontinus, engendrant alors une perte d'information sur la forme de la distribution. Ces deux fonctions empiriques peuvent être utilisées graphiquement mais elles ne permettent pas par simple visualisation de choisir la famille paramétrique de copules la mieux adaptée à la structure de dépendance des données.

## 2.8. Inférence statistique

Cette partie présente les principales méthodes permettant d'estimer le paramètre de la copule.

### Méthode des moments

Cette méthode revient à se donner une mesure de concordance  $\kappa\langle C \rangle$  et à considérer que la valeur du paramètre  $a$  de la copule  $C$  est celle qui égalise la valeur théorique  $\kappa\langle C(a) \rangle$  à  $\hat{\kappa}\langle C \rangle$ . Elle n'assure aucune robustesse de l'estimateur. Dans la pratique, l'estimateur empirique du tau de Kendall est le plus souvent utilisé comme mesure de concordance en raison de sa simplicité de calcul.

### Méthode du maximum de vraisemblance

A partir du théorème 1 et de la définition 2, sous des conditions de continuité, la densité jointe de la distribution  $F$  s'écrit :

$$f(x_1, \dots, x_n) = c(F_1(x_1), \dots, F_n(x_n)) \prod_{i=1}^n f_i(x_i).$$

Soient  $\{(x'_1, \dots, x'_n)\}_{t=1}^T$  l'échantillon d'observations,  $\theta$  le vecteur  $K \times 1$  des paramètres à estimer et  $\Theta$  l'espace dans lequel  $\theta$  prend ses valeurs.

La log-vraisemblance s'exprime alors  $l(\theta) = \sum_{t=1}^T \ln c(F_1(x'_t), \dots, F_n(x'_t)) + \sum_{t=1}^T \sum_{k=1}^n \ln f_k(x'_k)$ .

L'estimateur de  $\theta$ , noté  $\hat{\theta}_{ML}$ , est tel que  $l(\hat{\theta}_{ML}) \geq l(\theta)$ ,  $\forall \theta \in \Theta$ . Il vérifie la propriété de normalité asymptotique :

$$\sqrt{T}(\hat{\theta}_{ML} - \theta_0) \xrightarrow{L} N(0, I^{-1}(\theta_0)), \text{ avec } I(\theta_0) \text{ la matrice d'information de Fisher.}$$

Le problème avec cette méthode est qu'elle peut engendrer des temps de calcul très longs dans le cas d'une grande dimension car elle nécessite d'estimer conjointement les paramètres des lois marginales et les paramètres de la structure de dépendance. De plus, l'estimation de la copule est sensible à une éventuelle erreur d'estimation des marginales car elles interviennent directement dans le calcul de la log-vraisemblance.

### Méthode IFM (Inference Functions for Margins)

La méthode IFM ou méthode des fonctions d'inférence des marginales a été proposée par Joe *et al.* [1996]. Elle repose sur le fait que la représentation en copule permet de séparer les paramètres spécifiques des distributions marginales de ceux de la structure de dépendance. Ainsi, la log-vraisemblance peut s'écrire de la manière suivante :

$$l(\theta) = \sum_{t=1}^T \ln c(F_1(x'_t, \theta_1), \dots, F_n(x'_t, \theta_n); a) + \sum_{t=1}^T \sum_{k=1}^n \ln f_k(x'_t; \theta_k)$$

où  $\theta = (\theta_1, \dots, \theta_n, a)$  le vecteur des paramètres contient les paramètres  $\theta_k$  de chaque marginale et les paramètres  $a$  de la copule. La maximisation s'effectue en deux étapes :

on estime les paramètres de chaque marginale

$$\begin{aligned} \hat{\theta}_k &= \arg \max l_k(\theta_k) \\ &= \arg \max \sum_{t=1}^T \ln f_k(x'_t; \theta_k) \end{aligned}$$

on estime  $a$  à partir des estimateurs précédents

$$\begin{aligned} \hat{a} &= \arg \max l_c(a) \\ &= \arg \max \sum_{t=1}^T \ln c(F_1(x'_t, \hat{\theta}_1), \dots, F_n(x'_t, \hat{\theta}_n); a). \end{aligned}$$

Là encore, l'estimateur  $\hat{\theta}_{IFM} = (\hat{\theta}_1, \dots, \hat{\theta}_n, \hat{a})$  vérifie la propriété de normalité asymptotique et on a  $\sqrt{T}(\hat{\theta}_{IFM} - \theta_0) \rightarrow N(0, \nu^{-1}(\theta_0))$

avec  $\nu(\theta_0)$  la matrice d'information de Godambe. Si on définit  $g(\theta) = \left( \frac{\partial}{\partial \theta_1} l_1, \dots, \frac{\partial}{\partial \theta_n} l_n \right)$  alors

la matrice de Godambe s'écrit (Joe [1997]) :

$$\nu(\theta_0) = D^{-1} M (D^{-1})^T \text{ où } D = E \left[ \frac{\partial}{\partial \theta} (g(\theta)^T) \right] \text{ et } M = E [g(\theta)^T g(\theta)]$$

Cette méthode pourrait présenter l'avantage de reposer sur des calculs plus légers que ceux générés par la méthode du maximum de vraisemblance mais la détermination de la matrice de Godambe peut s'avérer très complexe en raison des multiples calculs de dérivées. Là encore, la méthode est sensible à une éventuelle erreur de spécification des marginales pour la même raison que celle évoquée plus haut.

### Méthode CML (Canonical Maximum Likelihood)

Cette méthode, recommandée par Bouye *et al.* [2000], est voisine de la méthode IFM à la différence qu'elle ne nécessite pas d'avoir recours à l'estimation des marginales. Pour cela, les observations  $\{(x'_1, \dots, x'_n)\}_{t=1}^T$  sont transformées en uniformes  $\{(\hat{u}'_1, \dots, \hat{u}'_n)\}_{t=1}^T$  en utilisant les fonctions de répartition empirique univariées et en estimant le paramètre comme suit :

$$\hat{a} = \arg \max \sum_{t=1}^T \ln c(\hat{u}'_1, \dots, \hat{u}'_n; a).$$

Cette méthode présente le grand avantage de procéder à une estimation paramétrique de la copule totalement indépendante de la forme paramétrique des lois marginales. En outre, elle génère des temps de calcul limités<sup>4</sup>. Ce sont deux atouts majeurs qui la rendent très attractive. C'est cette méthode qui a été choisie dans notre application.

## 2.9. Sélection d'une copule à partir du test d'adéquation du chi-deux

La sélection de la meilleure copule doit s'appuyer sur un test statistique. Le test du chi-deux utilisé dans le cadre de l'ajustement d'une distribution paramétrique à une distribution empirique constitue à cet égard un outil intéressant pour le choix d'une copule. Habituellement mis en œuvre pour ajuster une distribution univariée, il offre l'avantage de facilement s'adapter au cas des distributions multivariées. Il est vrai qu'il nécessite de procéder à un découpage en classes qui réduit sa puissance et que la manière de choisir ces classes peut être critiquée. Néanmoins, sa facilité d'utilisation en fait un outil de décision très intéressant pour sélectionner la meilleure copule.

Nous proposons ci-après une méthode statistique de sélection de copule basée sur un ajustement bivarié entre la copule paramétrique et la copule empirique effectué à partir de la statistique du chi-deux. Notons que l'extension de cette méthodologie au cas trivarié est immédiate en veillant à bien adapter le calcul des probabilités trivariées (voir Cadoux *et al.* [2004]). Notre approche s'inspire d'un article de Hurlimann [2001a] dans lequel le test du chi-deux est utilisé dans une version modifiée de la méthode IFM. Le processus de sélection d'une copule est décrit ci-après.

### 1. Estimation du paramètre de la copule

On estime le paramètre de la copule bivariée à partir de la méthode CML afin de s'affranchir d'une éventuelle erreur de spécification des lois marginales. On note  $C_{\hat{a}_{cml}}(u_1, u_2)$  la copule paramétrique obtenue.

### 2. Calcul de la copule empirique

On définit un treillis d'ordre  $T$  adapté au nombre d'observations et aux contraintes de temps de calcul puis on calcule la copule empirique bivariée sur ce treillis :

$$\hat{C}\left(\frac{t_1}{T}, \frac{t_2}{T}\right) = \frac{1}{T} \sum_{i=1}^T \prod_{k=1}^2 1_{\{t_k^i \leq t_k\}}$$

### 3. Choix d'un nombre d'intervalles pour découper les uniformes univariées

On se donne un nombre  $k$  afin de découper chacune des uniformes en  $k$  intervalles et de construire un tableau de contingence bivarié des effectifs.  $k$  et  $T$  doivent être choisis de concert afin d'obtenir aisément les effectifs empiriques à partir de la copule empirique. Sans a priori sur la distribution empirique bivariée, on découpe chaque uniforme sur des intervalles identiques. On note  $(b_0, b_1], (b_1, b_2], \dots, (b_{k-1}, b_k]$  et  $(c_0, c_1], (c_1, c_2], \dots, (c_{k-1}, c_k]$  les  $k$  intervalles pour découper respectivement  $u_1$  et  $u_2$ .

### 4. Calcul des effectifs et construction du tableau de contingence bivarié

L'objet de cette étape est de calculer les effectifs empiriques et théoriques sur chacun des  $k^2$  intervalles bidimensionnels  $(b_{i-1}, b_i] \times (c_{j-1}, c_j]$ ,  $i, j = 1, \dots, k$ .

<sup>4</sup> Pour accélérer la maximisation du programme, il est conseillé de partir d'une valeur initiale de  $a$  égale à celle qui découle de l'estimateur des moments.

Pour ce faire, il convient de calculer les probabilités bivariées  $p_{i,j}$  d'appartenance à un intervalle. La formule de calcul est la suivante :

$$p_{i,j} = C(b_i, c_j) - C(b_{i-1}, c_j) - C(b_i, c_{j-1}) + C(b_{i-1}, c_{j-1})$$

où  $i, j = 1, \dots, k$  et  $C(x, y) = \hat{C}(x, y)$  ou  $C_{\hat{a}_{cml}}(x, y)$ .

Les effectifs  $f_{i,j}$  sont calculés en multipliant les probabilités bivariées par le nombre total d'observations. On obtient alors un tableau de contingence bivarié donnant les effectifs empiriques issus de la copule empirique et les effectifs théoriques issus de la copule paramétrique. Il a la forme suivante :

| $\mathbf{u}_1 \setminus \mathbf{u}_2$ | $(c_0, c_1]$ | $(c_1, c_2]$ | ... | $(c_{k-1}, c_k]$ |
|---------------------------------------|--------------|--------------|-----|------------------|
| $(b_0, b_1]$                          | $f_{1,1}$    | $f_{1,2}$    | ... | $f_{1,k}$        |
| $(b_1, b_2]$                          | $f_{2,1}$    | $f_{2,2}$    | ... | $f_{2,k}$        |
| ...                                   | ...          | ...          | ... | ...              |
| $(b_{k-1}, b_k]$                      | $f_{k,1}$    | $f_{k,2}$    | ... | $f_{k,k}$        |

### 5. Application du critère de Cochran et regroupement en classes

Dans la pratique, les effectifs théoriques de certaines cases peuvent être très faibles voir proches de zéro. Il faut alors procéder à un regroupement des  $k^2$  intervalles initiaux en  $n$  classes permettant de respecter le critère de Cochran qui recommande d'avoir des effectifs théoriques au moins supérieurs à 1% du nombre total d'observations dans chaque classe, et supérieurs à 5% du nombre total d'observations dans au moins 80% des classes. Ce regroupement est aussi à effectuer sur le tableau bivarié des effectifs empiriques. Selon les zones de dépendance recherchées, on pourra notamment isoler les cellules des queues de distribution ou retenir des directions particulières comme la diagonale principale par exemple.

### 6. Calcul de la statistique bivariée du chi-deux et sélection de la copule

Après avoir défini le regroupement en  $n$  classes vérifiant le critère de Cochran, on calcule la statistique bivariée du chi-deux observée comme suit :

$$\chi_{obs}^2 = \sum_{j=1}^n \frac{(O_j - E_j)^2}{E_j}$$

où  $O_j$  et  $E_j$  représentent respectivement les effectifs observés et attendus dans chaque classe  $j$ ,  $j = 1, \dots, n$ . Cette statistique suit une loi du chi-deux à  $(n-r-1)$  degrés de liberté avec  $r$  le nombre de paramètres estimés de la copule. On se donne ensuite un seuil critique  $\alpha$  pour la zone de rejet et on calcule la p-valeur  $\hat{\alpha} = P(\chi_{n-r-1}^2 > \chi_{obs}^2)$ . On rejette l'adéquation de la copule paramétrique si  $\hat{\alpha} < \alpha$ . Il faut préciser pour les praticiens que la p-valeur exacte  $\hat{\alpha}$  se situe en fait dans l'intervalle suivant :

$$P(\chi_{n-1}^2 > \chi_{obs}^2) \leq \hat{\alpha} \leq P(\chi_{n-r-1}^2 > \chi_{obs}^2).$$

Cet encadrement étant de faible amplitude lorsque  $r$  est petit vis à vis de  $n$ , le seuil critique  $\alpha$  sera le plus souvent en dehors de cet encadrement et on conclura sans ambiguïté (Besson *et al* [2004]). Parmi les copules acceptées, on sélectionne alors celle qui maximise la p-valeur.

### **Partie 3. PORTEFEUILLE ETUDIE ET RECHERCHE DES DEPENDANCES**

Cette partie aborde l'aspect pratique de nos travaux. On y présentera le portefeuille d'assurance non vie étudié et les risques retenus pour l'application. On mettra ensuite en évidence un certain nombre de dépendances et on sélectionnera les copules les représentant le mieux.

#### **3.1. Portefeuille étudié**

Les données en notre possession sont celles d'une société d'assurance importante du marché français des particuliers et des petits professionnels. L'application se limite au seul risque de fluctuation de la charge sinistres du portefeuille, qui constitue un aléa majeur parmi ceux susceptibles d'affecter la solvabilité d'un assureur non vie. Trois branches ont été retenues au sens de la vision produit qu'en a la compagnie : Automobile (AUTO), Multirisque Habitation des Particuliers (MR PART), et Multirisque des Professionnels (MR PROF). Ce périmètre d'analyse représente plus de 70% du chiffre d'affaire non-vie de la compagnie. La base de données exploitée contient tous les sinistres survenus de 1996 à 2001, soit plus de 7,5 millions de sinistres. Les branches de la compagnie se déclinent ensuite en garanties. Au total, on en dénombre 23 comme entre autres les garanties Responsabilité Civile (RC), Tempête, Vol, Catastrophes Naturelles (Cat Nat).

Afin de disposer d'un échantillon suffisamment grand pour détecter les dépendances entre charge sinistres, on agrège les coûts individuels des sinistres par mois de survenance, ce qui conduit à 72 mois de charge sinistres pour chaque garantie. Il est important de souligner que la dépendance est évaluée sans analyser d'éventuels phénomènes de saisonnalité en raison de l'historique limité à six ans. Les mois constituant l'échantillon sont donc considérés comme indépendants, et on fait l'hypothèse qu'on a la même loi d'un mois à l'autre pour une garantie donnée.

Avant toute estimation statistique sur les charges sinistres mensuelles, les redressements suivants ont été réalisés :

- Les charges de sinistres ne sont pas comparables d'un mois à l'autre. Pour l'être, on doit procéder à une mise « as if » qui corrigera les sinistres des éléments suivants : évolution de la taille du portefeuille pour obtenir un portefeuille de taille constante, évolution de l'inflation pour avoir les sinistres en euros constants, évolution des franchises pour comparer les sinistres avec un même seuil de franchise.
- Les charges de sinistre issues des fichiers ne sont pas toutes stabilisées : des sinistres tardifs peuvent encore être déclarés, des modifications d'évaluations et des règlements complémentaires des sinistres en cours de gestion peuvent intervenir ... Il est donc nécessaire de calculer la charge finale prévisible de chaque exercice et de chaque mois de survenance, c'est-à-dire d'avoir une vision « à l'ultime » des sinistres.

En accord avec les experts de la compagnie, l'hypothèse a été faite que les évolutions de structure de portefeuille avaient un impact limité. De plus, le contenu des garanties et leurs plafonds ainsi que les méthodes de provisionnement n'ont pas subi de changement significatif sur la période observée. Ainsi, ces éléments n'ont conduit à aucune correction des charges sinistres.



Les tempêtes de décembre 1999, Lothar et Martin, représentent plus de 180 millions d'euros de charge sinistres pour la compagnie, dont plus de 90% concernent la garantie Tempête. Leur impact exceptionnel sur le portefeuille, rendant difficile l'ajustement de lois de distribution sur les charges sinistres mensuelles, nous a conduit à les exclure des traitements.

Après redressement, une loi de distribution marginale a pu être ajustée pour chaque charge sinistres des 23 garanties. La loi Log-Normale présente le meilleur ajustement dans 19 cas. Les lois de Weibull et Gamma sont retenues chacune deux fois.

### 3.2. Recherche des dépendances

La recherche porte à la fois sur les dépendances entre garanties à l'intérieur d'une même branche et sur les dépendances entre garanties de branches différentes. L'analyse repose sur :

- Les mesures de concordance du tau de Kendall et du rho de Spearman dont les valeurs absolues sont significatives au-delà de 0,21 et 0,30 respectivement pour un seuil de confiance de 1% et un échantillon de 72 observations (voir Saporta [1990])
- Le test d'indépendance du chi-deux au seuil 1%, qui contrairement aux mesures précédentes, ne donne aucune information sur le sens de la dépendance.

Le processus de sélection peut être décomposé en trois étapes.

#### *Etape 1 : sélection des couples avec une dépendance significative*

L'analyse de la significativité des mesures de concordance et du test d'indépendance du chi-deux conduit à ne retenir que 29 couples de garanties sur 253 combinaisons possibles.

#### *Etape 2 : exclusion des couples avec une dépendance négative*

L'intérêt de la compagnie portant avant tout sur les dépendances susceptibles de générer un besoin additionnel en capital, les dépendances négatives ont été volontairement omises.

#### *Etape 3 : exclusion des couples avec une dépendance difficilement interprétable*

Enfin, seules ont été sélectionnées les dépendances faisant l'objet d'une interprétation possible de la part des experts de la compagnie : relation de cause à effet entre deux garanties, ou relation due à une cause commune.

Le tableau suivant regroupe les 9 couples finalement sélectionnés à l'issue de ce processus, triés par valeurs décroissantes du tau de Kendall (mesure de concordance privilégiée car elle peut être reliée simplement au paramètre de la copule).

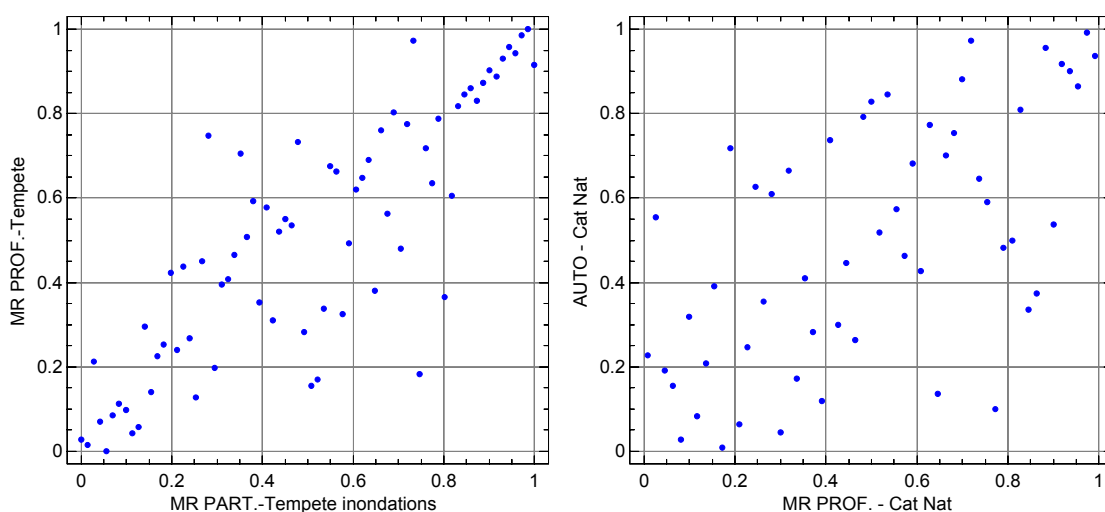
| Couples triés par valeurs décroissantes du tau de Kendall |                              | tau de Kendall | rho de Spearman | Proba. chi-deux 4 classes |
|---|------------------------------|----------------|-----------------|---------------------------|
| MR PART.-Tempete inondations                              | MR PROF.-Tempete             | 0.68           | 0.83            | 0.00%                     |
| AUTO-Tempete  | MR PART.-Tempete inondations | 0.61           | 0.79            | 0.00%                     |
| AUTO-Tempete  | MR PROF.-Tempete             | 0.55           | 0.74            | 0.00%                     |
| AUTO-Cat Nat  | MR PROF.-Cat Nat             | 0.53           | 0.70            | 0.00%                     |
| MR PART.-Cat Nat  | MR PROF.-Cat Nat             | 0.48           | 0.66            | 0.00%                     |
| MR PROF.-Pertes Financières                               | MR PROF.-Incendie            | 0.48           | 0.66            | 0.00%                     |
| AUTO-Cat Nat  | MR PART.-Cat Nat             | 0.42           | 0.59            | 0.02%                     |
| AUTO-Vol  | MR PROF.-Vol                 | 0.34           | 0.48            | 0.13%                     |
| AUTO-RC matérielle  | AUTO-Tous risques            | 0.31           | 0.44            | 0.00%                     |

On note que 7 couples illustrent une dépendance inter-branches et deux couples une dépendance intra-branche. Les dépendances observées s'interprètent comme suit :

- **Couples Tempête et Cat Nat** : ils présentent les plus fortes dépendances. Les événements climatiques impactent les automobiles, les habitations des particuliers et les locaux des professionnels en même temps.
- **AUTO-RC matérielle / AUTO-Tous risques** : l'interprétation de la dépendance observée est qu'un même sinistre en assurance Automobile peut fréquemment mettre en jeu la garantie RC pour les dommages aux tiers, et la garantie Tous risques pour les dommages au véhicule assuré.
- **MR PROF-Pertes Financières / MR PROF-Incendie** : l'incendie de bâtiments professionnels peut fréquemment générer des pertes d'exploitation ultérieures.
- **AUTO-Vol / MR PROF-Vol** : la liaison entre ces deux garanties Vol s'explique par l'évolution de la délinquance qui constitue une variable latente les affectant conjointement.

Les autres garanties non retenues seront considérées comme indépendantes entre elles et aussi vis-à-vis des garanties dépendantes sélectionnées.

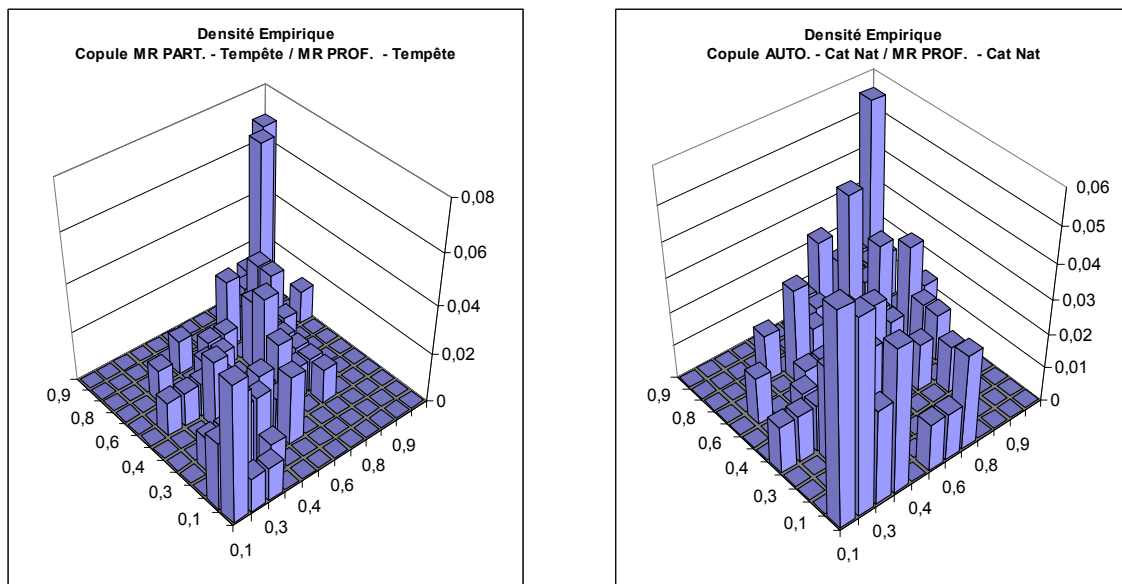
On donne ci-dessous un exemple graphique de deux couples illustrant la dépendance entre uniformes empiriques. A titre de comparaison, en cas d'indépendance parfaite nous devrions trouver dans chaque carré du graphique 72/25 observations, soit environ 3 points.



- **MR PART–Tempête / MR PROF–Tempête** : Le nuage de points est très proche de la 1<sup>ère</sup> bissectrice. Les queues de distribution sont très concentrées, notamment la queue supérieure.
- **MR PROF–Cat Nat / AUTO–Cat Nat** : Le nuage de points est moins proche de la 1<sup>ère</sup> bissectrice et les queues de distribution sont moins concentrées.

Le sens et l'intensité des dépendances entre uniformes empiriques observés graphiquement ont confirmé le classement des couples obtenu à l'issue de la phase de recherche de dépendances. En outre, ces graphiques donnent un premier renseignement sur la dépendance de queue qui peut être complété par une étude graphique des densités empiriques. Dans le cas bivarié, la densité empirique de la copule empirique n'est autre que la probabilité d'appartenance à un carré  $[0; 1]^2$  découpé en  $T$  arrêtes de longueur  $1/T$ .

Les graphiques suivants présentent la densité empirique de la copule empirique calculée avec  $T=12$ . Plus une barre est haute, plus il existe d'observations appartenant à l'intervalle bivarié  $[k/T; (k+1)/T] \times [j/T; (j+1)/T]$ ,  $k, j \in [0; T-1]$ .



De manière générale, la répartition des barres dessine une forme assez proche de celle du nuage de points des uniformes empiriques. Cela signifie que l'ordre  $T$  de construction de la copule empirique n'est pas trop faible car peu d'information est perdue sur la forme de la distribution bivariée. Autrement dit, les barres n'agrègent pas trop de données et l'analyse de la densité empirique fait sens. Pour la copule empirique liée à la garantie Tempête, on note une très forte surpondération dans la queue de distribution supérieure, et à un degré moindre dans la queue de distribution inférieure. Pour la copule empirique liée à la garantie Cat Nat, une surpondération dans les queues de distribution apparaît, plus marquée dans la queue de distribution inférieure que dans la queue de distribution supérieure.

### 3.3. Recherche des meilleures copules : estimation des paramètres, et sélection

Les copules testées sont la copule HRT et les copules archimédiennes de Gumbel, Franck, Clayton. Néanmoins, la présence de dépendances de queue pour certains couples, illustrée notamment par les graphiques ci-dessus, pourrait éventuellement conduire à tester la copule  $t$  de Student.

Dans un premier temps, on calcule l'estimateur des moments du paramètre  $a$  de chaque copule,  $\hat{a}_{moments}$ , en retenant l'estimateur empirique du tau de Kendall comme mesure de concordance. Dans un second temps, on utilise  $\hat{a}_{moments}$  comme valeur initiale permettant d'obtenir après maximisation  $\hat{a}_{CML}$ , estimateur du maximum de la logvraisemblance selon la méthode CML.

Pour les deux couples précédents, les tableaux suivants regroupent  $\hat{a}_{moments}$ ,  $\hat{a}_{CML}$ , le tau de Kendall issu de  $\hat{a}_{CML}$ ,  $\hat{\tau}_{\hat{a}_{CML}}$ , et le maximum de la logvraisemblance calculé à partir de  $\hat{a}_{CML}$ .

**MR PART. – Tempête / MR PROF. – Tempête**

|                              | Gumbel | Clayton | Franck | HRT    |
|------------------------------|--------|---------|--------|--------|
| $\hat{a}_{moments}$          | 3,094  | 4,189   | 10,426 | 4,189  |
| $\hat{a}_{CML}$              | 2,917  | 2,345   | 9,512  | 2,866  |
| $\hat{\tau}_{\hat{a}_{CML}}$ | 0,657  | 0,540   | 0,652  | 0,589  |
| Logvrais.                    | 48,468 | 35,839  | 43,604 | 41,355 |

**AUTO – Cat Nat / MR PROF. – Cat Nat**

|                              | Gumbel | Clayton | Franck | HRT    |
|------------------------------|--------|---------|--------|--------|
| $\hat{a}_{moments}$          | 2,068  | 2,136   | 6,044  | 2,136  |
| $\hat{a}_{CML}$              | 1,724  | 0,394   | 5,126  | 1,398  |
| $\hat{\tau}_{\hat{a}_{CML}}$ | 0,420  | 0,165   | 0,465  | 0,411  |
| Logvrais.                    | 24,676 | 12,721  | 22,910 | 22,593 |

Le test du chi-deux a été mis en œuvre pour sélectionner les copules représentant au mieux la structure de dépendance des différentes associations. Cela nécessite de découper chaque uniforme en  $k$  intervalles. Le nombre élevé d'associations testées a nécessité de développer une méthode standard de découpage en intervalles afin de limiter l'ampleur des travaux et des analyses. Pour ce faire, chacune des uniformes au sein d'une association a été découpée de manière identique : même nombre d'intervalles et même longueur d'intervalles. Plusieurs valeurs de  $k$  ont dû être testées afin de respecter le critère de Cochran.

In fine, les valeurs retenues pour  $k$  sont 4 et 6. A titre d'exemple, si dans le cas bivarié on a  $k = 4$ , alors le découpage de chaque uniforme correspondra aux intervalles  $(0; 0.25]$ ,  $(0.25; 0.50]$ ,  $(0.50; 0.75]$ ,  $(0.75; 1]$ .

Les effectifs empiriques dans chaque intervalle sont calculés à partir de la copule empirique. Le treillis de la copule empirique a été ajusté afin d'obtenir directement les fréquences empiriques sur les  $k$  intervalles. Si l'on reprend par exemple le cas bivarié avec  $k = 4$ , l'ordre  $T$  retenu pour le calcul de la copule empirique est égal à 18.

La valeur du seuil critique  $\alpha$  pour la zone de rejet est unique et égale à 5% et l'encadrement de la p-valeur exacte,  $\hat{\alpha}$ , est systématiquement calculé. Parmi les copules situées dans la zone d'acceptation, on sélectionne celle dont les bornes de l'encadrement de  $\hat{\alpha}$  sont maximales, ce qui est équivalent à sélectionner celle dont la statistique du chi-deux est la plus faible.

Afin d'illustrer la mise en pratique du test du chi-deux, on donne ci-après le détail des résultats pour le couple « MR PART–Tempête / MR PROF–Tempête », avec uniquement les effectifs théoriques issus de la copule de Gumbel.

| Effectifs empiriques                     |          |             |             |          | Ecart au carré pondéré                               |              |             |             |      |  |
|--|----------|-------------|-------------|----------|--|--------------|-------------|-------------|------|--|
|  | 0 - 0,25 | 0,25 - 0,50 | 0,50 - 0,75 | 0,75 - 1 |  | Franck       | Clayton     | Gumbel      | HRT  |  |
| 0 - 0,25                                 | 13.00    | 5.00        | 0.00        | 0.00     | C1   | 0.00         | 0.02        | 0.03        | 0.75 |  |
| 0,25 - 0,50                              | 2.00     | 7.00        | 9.00        | 0.00     | C2   | 0.50         | 0.06        | 0.24        | 0.00 |  |
| 0,50 - 0,75                              | 3.00     | 5.00        | 6.00        | 4.00     | C3   | 1.08         | 0.07        | 1.09        | 0.75 |  |
| 0,75 - 1                                 | 0.00     | 1.00        | 3.00        | 14.00    | C4   | 0.06         | 2.08        | 0.00        | 0.00 |  |
| Effectifs théoriques de la copule Gumbel |          |             |             |          | C5   | 0.01         | 0.15        | 0.05        | 0.91 |  |
|  | 0 - 0,25 | 0,25 - 0,50 | 0,50 - 0,75 | 0,75 - 1 | C6   | 0.16         | 2.35        | 0.00        | 0.01 |  |
| 0 - 0,25                                 | 12.41    | 4.54        | 0.97        | 0.09     | C7   | 6.16         | 4.15        | 4.49        | 4.48 |  |
| 0,25 - 0,50                              | 4.54     | 8.41        | 4.50        | 0.55     | C8   | 0.24         | 0.03        | 0.05        | 0.05 |  |
| 0,50 - 0,75                              | 0.97     | 4.50        | 9.15        | 3.37     | C9   | 0.16         | 2.35        | 0.00        | 0.01 |  |
| 0,75 - 1                                 | 0.09     | 0.55        | 3.37        | 13.99    | C10  | 0.01         | 0.15        | 0.05        | 0.91 |  |
|  |          |             |             |          | <b>Chi-deux = somme des écarts au carré pondérés</b> |              |             |             |      |  |
|  |          |             |             |          | <b>8.38</b>  | <b>11.39</b> | <b>5.99</b> | <b>7.87</b> |      |  |
|  |          |             |             |          | <b>p-value à 8 DL</b>                                |              |             |             |      |  |
|  |          |             |             |          | <b>0.40</b>  | <b>0.18</b>  | <b>0.65</b> | <b>0.45</b> |      |  |
|  |          |             |             |          | <b>p-value à 9 DL</b>                                |              |             |             |      |  |
|  |          |             |             |          | <b>0.50</b>  | <b>0.25</b>  | <b>0.74</b> | <b>0.55</b> |      |  |

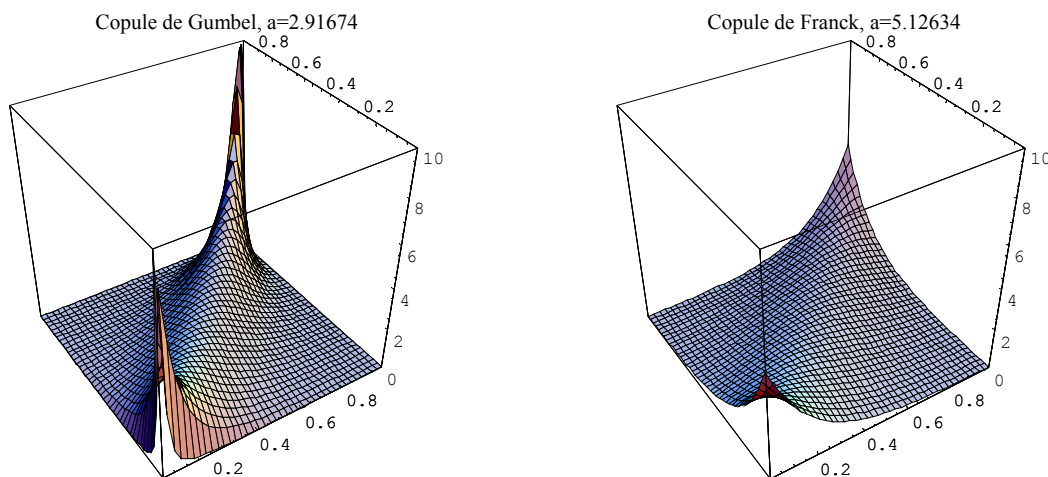
Ici, les uniformes ont été découpées en 4 intervalles conduisant à 16 cases, ensuite regroupées en 10 classes. Toutes les copules sont acceptées au seuil de 5%. La copule de Gumbel dont la p-valeur est maximale est sélectionnée comme copule décrivant le mieux la structure de dépendance de ce couple.

Les résultats ci-dessous correspondent à l'ensemble des dépendances analysées.

|  | copule | $\hat{a}_{CML}$ |
|--|--------|-----------------|
| MR PART. – Tempête / MR PROF. – Tempête      | Gumbel | 2,917           |
| MR PART. – Tempête / AUTO – Tempête          | Gumbel | 2,194           |
| AUTO – Cat Nat / MR PROF. – Cat Nat          | Franck | 5,126           |
| MR PART. – Cat Nat / MR PROF. – Cat Nat      | Gumbel | 1,801           |
| MR PROF. – Pertes Fin. / MR PROF. – Incendie | Franck | 5,235           |
| AUTO – Vol / MR PROF. – Vol                  | Franck | 3,163           |
| AUTO – RC Mat / AUTO – Tous Risques          | HRT    | 0,779           |

Ces résultats montrent que les copules de Gumbel et de Franck décrivent le mieux les structures de dépendance présentes au sein du portefeuille. La copule HRT n'est sélectionnée qu'une fois. En raison des propriétés des copules sélectionnées, ceci signifierait que les dépendances exhibées se caractérisent avant tout, soit par une absence de dépendance de queue, soit par une dépendance de queue supérieure.

Ceci est illustré dans les graphiques suivants qui représentent la densité théorique des copules bivariées : Gumbel pour « MR PART–Tempête / MR PROF–Tempête », et Franck pour « AUTO–Cat Nat / MR PROF–Cat Nat ».



## **Partie 4. MODELISATION DU BESOIN EN FONDS PROPRES**

### **4.1 Hypothèses du modèle**

Les travaux visent à mesurer l'impact sur le besoin en fonds propres des dépendances entre risques modélisés à l'aide de diverses copules. Plus précisément, l'objectif est d'estimer l'erreur commise lorsqu'on suppose l'indépendance entre risques. Le besoin en fonds propres est évalué à travers plusieurs mesures de risque, et on rappelle que seul le risque de fluctuation des sinistres bruts de réassurance est pris en compte.

Avant tout, il convient de définir un modèle qui régitte l'évolution des fonds propres de la compagnie. Le modèle présenté vise à évaluer simplement la situation financière de la compagnie afin d'obtenir des résultats facilement interprétables et apportant, dans la mesure du possible, une réponse claire à la problématique donnée. Un certain nombre d'hypothèses de travail ont donc été formulées.

L'aléa lié à la fluctuation des charges sinistres est analysé dans deux cas. Dans un cas, les charges sinistres sont supposées indépendantes entre garanties et sont agrégées les unes aux autres. Dans l'autre cas, les charges sinistres sont additionnées mais en intégrant entre certaines garanties des dépendances positives modélisées à l'aide de plusieurs copules.

Le modèle projette la situation financière de la compagnie avec un horizon annuel qui est l'année civile. Le modèle est discret en ce sens qu'on ne modélise pas la situation financière à un instant  $t$  quelconque de l'année. Une analyse dynamique de l'évolution des fonds propres sur plusieurs exercices successifs n'est pas étudiée mais pourrait faire l'objet de travaux ultérieurs.

En raison du risque modélisé, les fonds propres réellement détenus par la compagnie ne peuvent être utilisés dans ce modèle. En effet, ils sont destinés à couvrir tous les risques susceptibles d'affecter la situation financière de la compagnie. Une approche normative a donc été retenue afin de fixer le niveau des fonds propres en début de période. Ces derniers sont supposés égaux à la marge de solvabilité réglementaire européenne, et sont déterminés dans notre cas à partir du volume de cotisations.

Par ailleurs, on suppose qu'aucun mouvement ne les affecte en cours d'année. Les fonds propres de fin de période, avant allocation du résultat annuel, sont donc identiques à ceux de début de période. Ces fonds propres sont appelés  $FP$ .

Les contrats sont supposés émis au 1<sup>er</sup> janvier pour une durée d'exposition d'un an. Les cotisations commerciales acquises sont donc égales aux cotisations commerciales émises. De plus, on considère que les chargements pour l'acquisition, l'administration des contrats et la gestion des sinistres couvrent exactement les frais, et sont donc exclus des cotisations<sup>5</sup> pour simplifier les flux. Les cotisations acquises incluses dans la modélisation sont donc égales aux primes pures incluant un chargement de sécurité. Ces cotisations sont notées  $P$ .

La charge sinistres annuelle est la seule variable aléatoire du modèle. Elle est calculée sur l'ensemble du portefeuille et est brute de réassurance<sup>6</sup>. Elle somme les charges sinistres annuelles des 23 garanties. Pour chaque garantie, la charge sinistres annuelle est égale à la somme de 12 charges mensuelles simulées. Dans notre cas de figure, on rappelle que les mois ont été considérés comme indépendants et suivant la même loi. La charge sinistres annuelle reflète un seul exercice de survenance et les sinistres survenus sont entièrement réglés en milieu d'année. La charge sinistres annuelle est appelée  $S$ .

Les actifs de début de période sont égaux à la somme des cotisations et du capital de début de période. Ils sont par ailleurs supposés stables au cours de l'année.

Le résultat financier exclut toute réalisation de plus ou moins-value, et n'est constitué que des produits financiers dégagés par les actifs de début de période. Les actifs du segment fonds propres rapportent un taux d'intérêt annuel de 2% alors que les actifs du segment assurance, en l'occurrence les cotisations, génèrent un rendement annuel de 3% duquel il faut soustraire un manque à gagner, lié au règlement des sinistres intervenant en milieu d'année. Il n'y a pas de modélisation de besoin en fonds de roulement, et les rendements sont supposés sans aléa. Le résultat financier est noté  $RF$ .

Le résultat annuel sera appelé  $X$  et le besoin en fonds propres évalué en fin d'année sera fonction de cette variable. Par simplification,  $X$  sera vu avant impôts et taxes. En raison des hypothèses précédemment formulées,  $X$  s'écrit alors

$$X = \text{résultat technique} + \text{résultat financier}$$
$$X = (P - S) + (2\% \times FP + 3\% \times P - 1.5\% \times S) = 0.02 \times FP + 1.03 \times P - 1.015 \times S$$

On note que  $X$  est très peu sensible au niveau de fonds propres de début de période.  $X$  est une variable aléatoire dont la loi est inconnue. Sa distribution est approchée de façon empirique à travers les simulations effectuées. On note  $F_X$  sa distribution empirique.

## 4.2. Mesures de risque

La sensibilité du besoin en fonds propres au phénomène de dépendances entre garanties est appréhendée à travers plusieurs mesures de risque faciles à mettre en œuvre pour un assureur dans un modèle interne de gestion des risques.

---

<sup>5</sup> De façon symétrique, le respect de l'hypothèse d'un jeu à somme nulle entre les chargements et les frais conduit à ne retenir aucun élément lié aux frais généraux dans la formation du résultat annuel.

<sup>6</sup> Ce qui revient à ne pas intégrer de résultat de réassurance.

### Probabilité de ruine :

La compagnie est dite en ruine lorsque les fonds propres de fin de période ne suffisent pas à absorber un résultat déficitaire, autrement dit la situation de ruine correspond à l'événement où  $X < -FP$ . On calculera sur la base des simulations la probabilité empirique de cet événement, soit  $F_X(-FP)$ .

### VaR, TVaR, XTVaR :

Afin de tester la sensibilité des résultats au seuil de sécurité retenu, quatre valeurs sont données à  $\alpha$  : 0.5%, 0.1%, 0.05% et 0.01%.

## 4.3. Simulations

On se propose d'exposer brièvement le principe de fonctionnement des algorithmes de simulation. Pour ce faire, on introduit les notations suivantes :

$X_i$  = charge sinistres de la garantie  $i$ ,

$U_i = F_{X_i}(x)$ , avec  $F_{X_i}$  fonction de répartition de  $X_i$  et donc  $U_i \rightarrow U[0;1]$ .

L'algorithme utilisé s'appuie sur la méthode de simulation par copule conditionnelle décrite dans Cadoux *et al* [2004]. Il est mis en œuvre afin que l'écart sur la charge agrégée,  $X_1 + X_2$ , entre la situation d'indépendance et celle de dépendance, soit uniquement dû à la prise en compte d'une dépendance entre les deux garanties<sup>7</sup>. Dans un premier temps, on calcule la charge agrégée en simulant indépendamment  $X_1$  et  $X_2$ , puis on choisit librement de déduire  $U_1$  ou  $U_2$ . Supposons qu'on déduise  $U_1$ , dans un second temps, on génère une variable aléatoire uniforme  $p$  représentant la fonction de répartition de la copule, que l'on inverse afin de déduire  $U_2$ , fonction de  $p$  et de  $U_1$ . Il ne reste plus alors qu'à inverser  $U_2$  pour obtenir  $X_2$  et avoir ainsi la charge agrégée de sinistres dans le cas dépendant.

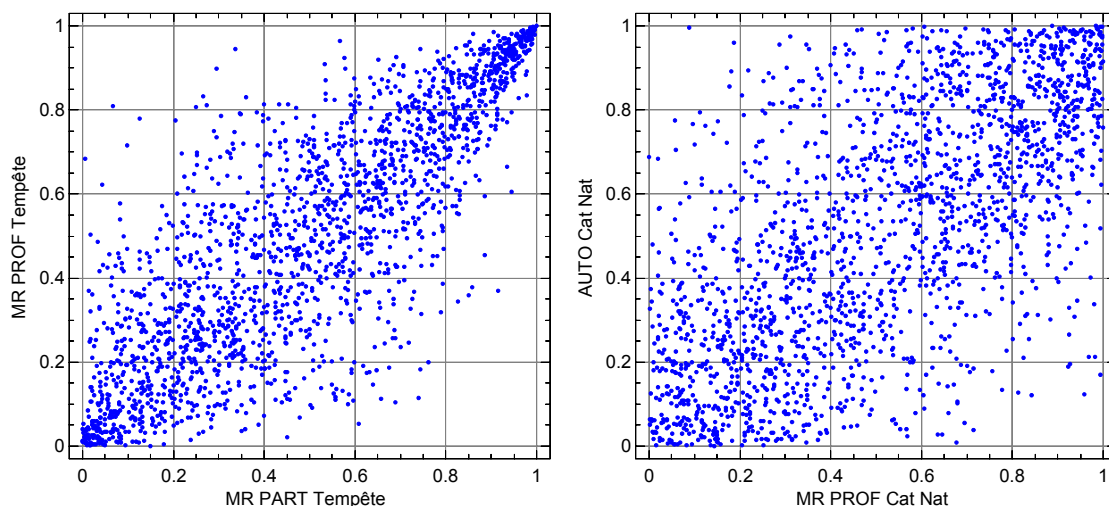
Il est important de noter que l'inversion de la copule n'admet pas de forme close dans le cas d'une copule de Gumbel. Il faut alors procéder par résolution numérique. Dans le cadre de ces travaux, la méthode de Newton a été utilisée. Elle donne de bons résultats mais nécessite de bien optimiser l'algorithme développé afin de limiter les temps de calcul, notamment lorsque des valeurs extrêmes sont générées.

Les deux graphiques suivants représentent 2000 observations issues des simulations de la copule de Gumbel retenue pour le couple « MR PART–Tempête / MR PROF–Tempête » et de la copule de Franck retenue pour le couple « AUTO–Cat Nat / MR PROF–Cat Nat ». Ils permettent d'appréhender la forme des dépendances générées, et d'apprécier leur adéquation avec le profil des copules paramétriques présentées précédemment pour ces deux couples.

---

<sup>7</sup> Si on simule deux fois  $U_1$ , une fois dans le cas dépendant, et une autre fois dans le cas indépendant, il est indispensable de procéder à un nombre de simulations très élevé pour considérer que les deux variables générées fournissent la même représentation d'un seul phénomène. Cela augmente considérablement les temps de calcul.





#### 4.4. Impact des dépendances sur le besoin en fonds propres

Cette section présente les résultats du modèle d'évaluation du besoin en fonds propres calculé à l'aide des mesures de risque sélectionnées. Afin d'avoir des résultats robustes, un nombre important de simulations de Monte-Carlo ont été menées : 12 millions de mois ont été simulés, soit 1 million d'années. On rappelle que l'on cherche avant tout à tester la sensibilité des mesures de risque à la prise en compte de dépendances modélisées à l'aide de copules.

##### 4.4.1 Impact des dépendances modélisées

Le périmètre d'analyse couvre l'ensemble des 23 garanties des trois branches. Les résultats sont regroupés dans le tableau ci-dessous, pour les valeurs  $\alpha$  des différents seuils de sécurité testés, et avec les écarts en pourcentage entre les cas d'indépendance et de dépendance des garanties.

| Probabilité de ruine<br>avec fonds propres de début de période = 150.7 M€ |  |  |  | Indép. | Dépend. | % écart<br>dép. / ind. |
|---|--|--|--|--------|---------|------------------------|
|   |  |  |  | 0.005% | 0.012%  | 140%                   |

| en M€    |        |         |                        |        |         |                        |        |         |                        |
|----------|--------|---------|------------------------|--------|---------|------------------------|--------|---------|------------------------|
| $\alpha$ | VaR    |         |                        | TVaR   |         |                        | XTVaR  |         |                        |
|          | Indép. | Dépend. | % écart<br>dép. / ind. | Indép. | Dépend. | % écart<br>dép. / ind. | Indép. | Dépend. | % écart<br>dép. / ind. |
| 0,5%     | 22.1   | 29.5    | 33.8%                  | 38.2   | 52.1    | 36.3%                  | 16.2   | 22.6    | 39.7%                  |
| 0,1%     | 43.8   | 60.2    | 37.5%                  | 73.9   | 103.1   | 39.4%                  | 30.1   | 42.9    | 42.2%                  |
| 0,05%    | 58.8   | 81.6    | 38.9%                  | 97.9   | 137.0   | 39.9%                  | 39.2   | 55.4    | 41.5%                  |
| 0,01%    | 117.2  | 165.7   | 41.3%                  | 183.1  | 256.2   | 39.9%                  | 65.9   | 90.5    | 37.4%                  |

On note que la modélisation des dépendances à l'aide de copules augmente sensiblement le besoin en fonds propres, et ce quels que soient les seuils et les mesures de risque. En cas de dépendance entre garanties, le maintien d'un seuil de sécurité identique à celui du cas d'indépendance implique une augmentation de la VaR ou de la TVaR allant de +33% à +41%. La probabilité de ruine est considérablement augmentée dans le cas dépendant par rapport au cas indépendant, étant près de deux fois et demie supérieure.

#### 4.4.2 Contribution aux ruines de chaque garantie

Sur un périmètre limité aux 12 garanties issues des 7 couples dépendants, on cherche à cerner le comportement de chacune des garanties lors des résultats les plus défavorables. Pour cela, on étudie les 609 ruines issues du cas d'indépendance, et on se limite aux 609 plus fortes ruines du cas dépendant afin de rendre la comparaison pertinente. Etant donné que les cas de ruines correspondent aux résultats adverses les plus extrêmes, on se propose de regarder quelles charges sinistres ont le plus de valeurs extrêmes dans ces cas afin de détecter celles ayant la plus forte contribution à la ruine.

Le tableau suivant regroupe plusieurs indicateurs pour chaque garantie afin de comprendre leurs contributions respectives aux pertes les plus extrêmes :

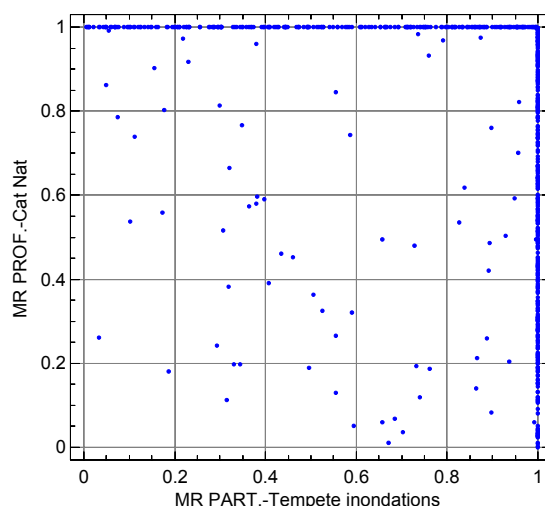
- La moyenne de chaque garantie ; et le coefficient de variation qui indique les garanties les plus volatiles.
- « *N Val. Extr.* » le nombre d'occurrences de valeurs extrêmes de la charge sinistres d'une garantie parmi les 609 ruines. La charge sinistres d'une garantie est dite extrême si son rang relatif – rang rapporté au nombre d'observations – est supérieur à 99,94%, c'est à dire si elle est parmi les 609 plus fortes valeurs de cette garantie.

| Charges sinistres annuelles en M€      | AUTO-RC mat.                    | AUTO-Tous risq. | AUTO-Vol | AUTO-Tempete | AUTO-Cat Nat | MR PROF.-Pertes Fin. | MR PROF.-Incendie | MR PROF.-Vol | MR PROF.-Tempete | MR PROF.-Cat Nat | MR PART.-Tempete | MR PART.-Cat Nat |     |
|--|---------------------------------|-----------------|----------|--------------|--------------|----------------------|-------------------|--------------|------------------|------------------|------------------|------------------|-----|
|  | <b>Ensemble des simulations</b> | Moyenne         | 172.0    | 210.5        | 69.7         | 0.6                  | 1.0               | 2.7          | 17.8             | 9.9              | 1.5              | 4.0              | 6.0 |
|  | Coef de var.                    | 4%              | 3%       | 4%           | 139%         | 157%                 | 22%               | 18%          | 4%               | 123%             | 107%             | 89%              | 29% |
| <b>Cas indép. 609 ruines</b>           | N val. Extr.                    | 2               | 0        | 0            | 5            | 26                   | 0                 | 1            | 1                | 39               | 222              | 301              | 0   |
| <b>Cas dép. 609 plus fortes ruines</b> | N val. Extr.                    | 4               | 0        | 0            | 316          | 16                   | 0                 | 2            | 0                | 390              | 131              | 421              | 28  |

#### Cas d'indépendance

Les garanties Tempête des trois branches et deux des garanties Cat Nat présentent des valeurs extrêmes dans les cas de ruine. Ce sont aussi celles qui ont les plus forts coefficients de variation. Les garanties « MR PART-Tempête » et « MR PROF-Cat Nat » présentent le plus de valeurs extrêmes en cas de ruine, respectivement 301 et 222.

Le graphique qui suit permet de visualiser les rangs relatifs de ces 2 garanties pour les 609 cas de ruine. On note qu'il n'existe aucune accumulation de points dans le coin supérieur droit, ce qui confirme que les valeurs extrêmes des deux garanties surviennent de façon indépendante. En conclusion, dans le cas d'indépendance, les ruines s'expliquent principalement par la survenance non concomitante de valeurs extrêmes sur les garanties les plus volatiles.

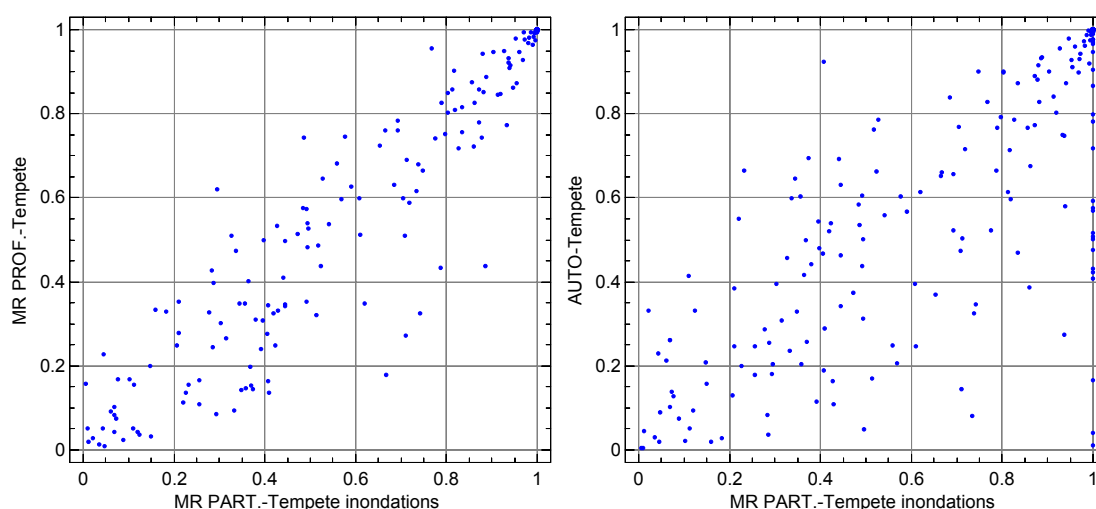


### Cas de dépendance

Au regard des résultats, les trois garanties Tempête contribuent encore plus fortement aux ruines que lorsqu'elles sont indépendantes. Les trois garanties Cat Nat ont aussi un effet important sur les ruines mais limité en raison de la part prépondérante des garanties Tempête.

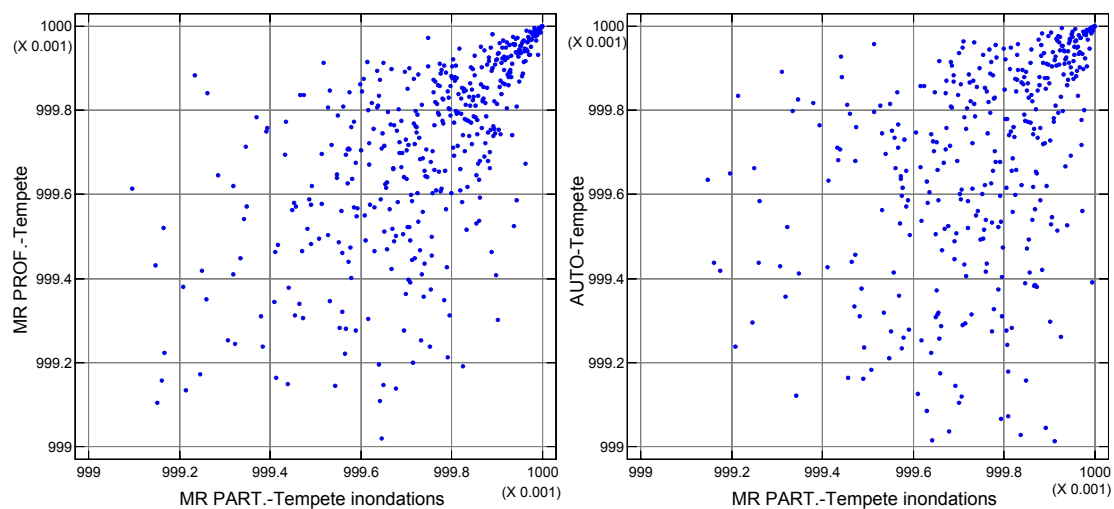
On voit ici tout l'effet de la modélisation de dépendances à l'aide de copules de Gumbel. En effet, plus d'une observation sur deux est désormais extrême pour les garanties « MR PROF–Tempête » et « AUTO–Tempête ». Cette forte contribution à la ruine s'explique par la liaison qui les fait dépendre de la garantie « MR PART–Tempête », modélisée par la copule de Gumbel qui accorde une probabilité plus forte à la survenance simultanée d'événements extrêmes. Un comptage effectué sur les cas de ruine indique qu'il existe 270 observations avec des valeurs extrêmes simultanées sur les trois garanties Tempête, auxquels s'ajoutent 103 autres événements extrêmes communs au couple « MR PART–Tempête » et « MR PROF–Tempête », ce qui représente au total plus de 60% des 609 plus fortes ruines.

Afin d'illustrer les résultats sur les garanties Tempête, on représente ci-dessous leurs rangs relatifs observés lors des 609 plus fortes ruines.



On observe facilement la dépendance positive existant entre les trois branches sur cette garantie, et surtout la forte concentration de valeurs supérieures extrêmes.

Afin de mieux visualiser cette accumulation, les graphiques ci-dessous reprennent les mêmes associations et effectuent un zoom sur les valeurs extrêmes situées dans l'intervalle bivarié  $[0,999;1]^2$ .



On note la très forte accumulation de sinistres extrêmes entre les branches MR PART et MR PROF. Ceci est aussi vrai entre les branches MR PART et AUTO, mais à un degré moindre en raison de la plus faible valeur du paramètre de la copule de Gumbel ajustée sur ce couple.

## **CONCLUSION**

Ce cas pratique montre que la prise en compte à l'aide de copules de dépendances positives entre charge sinistres augmente sensiblement le besoin en fonds propres d'un assureur non vie. Ce résultat est lié à l'utilisation de certaines copules qui permettent de modéliser la concomitance de situations fortement adverses, phénomène observé sur les données empiriques analysées mais à un degré moindre.

Les dépendances négatives, également présentes au sein du portefeuille étudié, pourraient aussi être modélisées et nuancer les conclusions de cette étude. Parmi les copules utilisées au cours de ces travaux, seule la copule de Franck appréhende ce type de dépendance. L'analyse devrait être élargie à d'autres familles de copules telles que les copules elliptiques entre autres, et notamment la copule t de Student qui permet en outre d'intégrer des dépendances de queue.

Bien évidemment, le modèle présenté ici n'est qu'une première étape et ne peut prétendre en l'état apporter une réponse complète à l'analyse de la solvabilité d'un assureur non vie. Il devrait être prolongé en mesurant l'impact de la réassurance et en analysant l'ensemble des risques supportés par la compagnie.

La sélection des copules représentant au mieux les structures de dépendance constitue un axe d'investigation supplémentaire. Le test d'adéquation du chi-deux est utile car il donne au critère de décision un fondement statistique et peut être utilisé dans un cadre multivarié. Cependant, il nécessite un découpage en classes qui lui fait perdre de la puissance. La recherche de tests multidimensionnels plus puissants appliqués aux copules pourrait offrir de nouvelles perspectives d'application.

## **BIBLIOGRAPHIE**

Artzner P., Delbaen F., Eber J-M. [1999], "Coherent measures of risk", *Mathematical Finance* 9.

Belguise O. [2001], "Tempêtes: Etude des dépendances entre les branches Auto et Incendie avec la théorie des copulas", *mémoire IA-USA*.

Besson J.L., Partrat C. [2005], "Assurance non-vie - Modélisation, simulation", *Economica*.

Bezat A., Nikeghbali A. [2000], "La théorie des extrêmes et la gestion des risques de marché", *groupe de travail ENSAE*.

Bouye E., Durrleman V., Nikeghbali A., Riboulet G., Roncalli T. [2000], "Copulas for finance : a reading guide and some applications", *GRO, Crédit Lyonnais, Paris*.

Cadoux D., Loizeau J.M. [2003], "Mesures de risque. Besoin en fonds propres", *Séminaire "Solvabilité, rentabilité en assurance non vie", Centre de Perfectionnement des Actuaire, Paris 2003*

Cadoux D., Loizeau J.M. [2004], "Dépendances au sein d'un portefeuille d'assurance non vie : cas pratique d'application de la théorie des copules à la détermination du besoin en fonds propres", *mémoire IA-CEA*

Charpentier A., Partrat C., Robert C. [2001], "Corrélation, tarification, pricing et solvabilité", *Journées de la FFA Biarritz 2001*.

Deheuvels P. [1979], "La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance", *Académie Royale de Belgique - Bulletin de la Classe des Sciences, 5<sup>ème</sup> série*.

De La Foata C., Odjo H. [2001], "Analyse d'un système de sécurité cohérent et optimal pour une compagnie d'assurance IARD", *ASTIN 2001, Washington*.

Durrleman V., Nikeghbali A., Roncalli T. [2000], "Which copula is the right one?", *GRO, Crédit Lyonnais, Paris*.

Embrechts P., McNeil A., Strauman D. [1999], "Correlation and dependence in risk management : properties and pitfalls", *Departement Mathematik, ETHZ*.

Embrechts P., Lindskog F., McNeil A. [2001], "Modelling dependence with copulas and applications to risk management", *Departement Mathematik, ETHZ*.

Faivre F. [2002], "Copules, besoin en fonds propres et allocation de capital pour une compagnie d'assurance non vie multibranches", *mémoire IA-ISFA*.

Fermanian J.D., Scaillet O. [2003], "Nonparametric estimation of copulas for time series", *Working paper*.

- Fermanian J.D. [2003], "Goodness-of-fit tests for copulas", *Working paper, CREST 2003-34, Paris*.
- Frees E., Valdez E. [1998], "Understanding relationships using copulas", *North American Actuarial Journal, Volume 2, N°1*
- Genest C., Rivest L. [1993], "Statistical inference procedures for bivariate archimedean copulas", *Journal of American Statistical Association, Volume 88, N°423*.
- Hurlimann W. [2001a], "Fitting bivariate cumulative returns with copulas", *Working paper, Winterthur Life and Pensions*.
- Hurlimann W. [2001b], "Analytical evaluation of economic risk capital and diversification using linear Spearman copulas", *Working paper, Winterthur Life and Pensions*.
- Joe H., Xu J.J. [1996], "The estimation method of inference functions for margins for multivariate models", *Department of Statistics, University of British Columbia, Technical Report*.
- Joe H. [1997], "Multivariate models and dependence concepts", *Chapman and Hall, London*.
- Klugman S., Parsa R. [1999], "Fitting bivariate loss distributions with copulas", *Insurance : Mathematics & Economics 24*
- Nelsen R. [1999], "An introduction to copulas", *Springer Lecture notes in statistics*.
- Peronnet F., Sellam B. [2001], "Allocation de fonds propres en assurance vie", *mémoire IA-ENSAE*.
- Robert C. [2001], "Gestion de risques multiples", *Journées Statistiques et Assurances Dommages, SFdS*.
- Saporta G. [1990], "Probabilités, analyse des données et statistiques", *Technip*.
- Sklar A. [1959], "Fonctions de répartition à n dimensions et leurs marges", *Publications de l'Institut de Statistique de Paris, 8, 229-231*.
- Venter G. [2001], "Tails of copulas", *ASTIN 2001*.
- Venter G. [2003], "Fit to a t - estimation, application and limitations of the t-copula Topic 1 : Risk evaluation", *ASTIN 2003*.