



**Mémoire présenté devant le jury de l'ENSAE en vue de  
l'obtention du  
Diplôme d'Actuaire ENSAE  
et de l'admission à l'Institut des Actuaires  
le 31 mai 2013**

Par : Octavie MBIAKOUP, Hoang Vu NGUYEN

**Titre** : Approche multivariée de la théorie de la ruine : Applications en assurance et risque alimentaire.

Confidentialité : non.

**Membres présents du jury de l'Institut  
des Actuaires**

**Florence PICARD**

Signature :

**Alexandre YOU**

Signature :

**Marc PHAMTRONG**

Signature:

**Entreprise  
ENSAE**

**Directeurs de mémoire en entreprise  
Olivier WINTENBERGER**

Signature :

**Jessica TRESSOU**

Signature :

**Membres présents du jury de l'ENSAE**

Xavier MILHAUD

Emmanuel BRUNEL

Bibliothèque :

Secrétariat :

# Sommaire

<b>Remerciements</b> .....	4
<b>Introduction</b> .....	5
<b>Chapitre I Cadre général</b> .....	7
I- Théorie des valeurs extrêmes .....	7
1) Présentation de la théorie des valeurs extrêmes .....	8
a) Loi généralisée des extrêmes.....	8
b) Excès au-delà d'un certain seuil (Distribution conditionnelle des excès) .....	11
2) Exemple de comportements limites dans les queues.....	14
3) Estimation statistique .....	17
II- Théorie de la ruine.....	24
1- Modèle classique de la théorie de la ruine.....	24
2- Le Modèle KDEM (Kinetic Dietary Exposure Model) .....	26
3- Lien entre modèle de ruine et KDEM .....	27
III- Approches multivariées .....	29
1. Introduction aux copules .....	29
2. Copules archimédiennes.....	33
<b>Chapitre II Application au risque sanitaire</b> .....	35
I) Contexte.....	35
1) Choix des substances.....	35
2) Choix des aliments étudiés .....	36
3) Analyse des résultats .....	36
4) Limites de l'interprétation des résultats et de l'étude.....	37
II) Présentation des contaminants.....	38
1) Les effets du méthyle-mercure .....	38
2) Les effets du Cadmium.....	38
3) Les effets des polychlorobiphényles .....	40
4) Description des données .....	41
III) Estimation du risque.....	44
1) Cas d'un seul risque .....	44
a) Simulation des données .....	44
b) Calcul de la charge corporelle .....	45
c) Détermination de la dose tolérable de long terme (TBB).....	49
2) Cas de plusieurs risques .....	51

<b>Chapitre III Application au risque assurantiel.....</b>	<b>59</b>
I- Contexte.....	59
II- Solvabilité II.....	61
A- Pilier 1 : les exigences quantitatives en capital .....	61
a) Calcul des provisions économiques.....	61
b) Calcul du Minimum de Capital Requis et du Capital de Solvabilité Requis.....	62
c) Les différents types de risques .....	65
1) Le risque de souscription.....	65
2) Le risque opérationnel .....	65
3) Le risque de crédit .....	65
4) Le risque de marché.....	65
d) Exigence en capital lié au risque de marché : le SCR <sub>Mkt</sub> .....	66
B- Pilier 2 : gestion des risques et principes de gouvernance.....	68
C- Pilier 3 : discipline de marché .....	68
III- Estimation du risque.....	68
A- La formule standard.....	68
B- Le modèle interne.....	69
a) Description des données .....	72
b) SCR par la méthode paramétrique (gaussienne).....	77
c) VAR par la méthode Monte Carlo.....	78
d) VAR historique.....	79
e) SCR par la méthode de McNeil et Frey.....	80
C- Constitution du portefeuille optimal.....	83
<b>Conclusion.....</b>	<b>86</b>
Liste des Tableaux .....	88
Table des figures.....	89
Bibliographie.....	90

## Remerciements

Nous tenons à remercier nos encadrants, Jessica TRESSOU et Olivier WINTENBERGER, pour les précieux conseils qu'ils nous ont fournis tout au long de l'année, et surtout pour leur patience et leur écoute.

Un grand merci également aux personnes qui ont eu la gentillesse de répondre à nos questions concernant ce mémoire.

Nous pensons aux intervenants professionnels et académiques de l'ENSAE. Nous avons particulièrement apprécié l'aide de Victor BRUNEL, notre correspondant ENSAE. Enfin, nous souhaitons remercier l'ensemble de nos proches pour leur soutien et leur patience.

## Introduction

Une compagnie d'assurance a souvent plusieurs branches d'activité (assurance automobile, incendie, tempête, etc.) ou des filiales dans différents pays. Les filiales sont soumises à des normes de solvabilité distinctes avec une fongibilité du capital<sup>1</sup> limitée. Pour chaque branche, il est nécessaire de calculer un montant de capital à geler pour faire face au risque de ruine à un horizon un an. Dans ce contexte, il est pertinent d'étudier l'évolution conjointe des réserves des différentes branches d'activité de la compagnie, dans le but d'estimer le niveau de capital total requis pour une meilleure gestion actif/passif. Or, à cause de la complexité croissante des marchés financiers, il est de plus en plus difficile de faire la gestion actif/passif avec le portefeuille typique des compagnies d'assurance. La littérature récente a montré l'avantage des stratégies de Hedge Fund dans la gestion Actif/Passif mais la réglementation en vigueur rend difficile son intégration dans la répartition de l'actif des sociétés d'assurance. Elle exige en effet une marge de 49% pour les risques inhérents à des stratégies de Hedges Funds (HF), pourtant des études<sup>2</sup> ont montré qu'une exigence de 25% est suffisante pour une allocation de fonds de couverture diversifiée. Une analyse du risque assurantiel vise donc en premier lieu à estimer le SCR (Solvency Capital Requirement) qui est le montant de capital requis pour faire face à un risque de ruine à horizon un an avec une probabilité de 99.5% sur différentes stratégies de Hedge Fund. Il s'agit en fait de déterminer le montant de capital que la compagnie d'assurance doit détenir pour chaque stratégie de HF afin d'absorber les pertes imprévues importantes et donc d'assurer la solvabilité de la compagnie d'assurance. Ensuite, l'idée finale est d'élaborer une combinaison optimale de stratégies permettant de réduire ce montant de capital.

De manière similaire, les individus sont exposés via leur alimentation à une multitude de contaminants. Dans ce cadre, toute mesure visant à réduire l'exposition à un contaminant peut avoir un impact indésirable quant à l'exposition à un autre contaminant, les processus étant naturellement dépendants via le comportement alimentaire. En effet, plusieurs contaminants peuvent se retrouver à la fois dans le même aliment. Une analyse de risque alimentaire vise en premier lieu à déterminer si une substance donnée peut poser un problème de santé publique. Dans ce cas, il s'agit d'abord d'identifier les caractéristiques des individus

---

<sup>1</sup> capacité de transférer des réserves d'une branche ou filiale n'ayant pas de problème vers une branche en difficulté

<sup>2</sup> **Mathieu Vaissié**, «Solvency II : Regulation change and Hedge Fund Evolution »

les plus exposés, puis d'élaborer les moyens de réduction du risque les plus efficaces et de mettre en œuvre éventuellement certaines mesures de sécurité sanitaire. Les composants nocifs qu'on rencontre fréquemment dans l'alimentation sont les polychlorobiphényles (Pcb), le cadmium (Cd) et le méthyle-mercure (MeHg).

L'objet de notre étude est d'évaluer la probabilité pour que le rendement d'une stratégie de Hedge Fund soit inférieur à un seuil donné dès lors que celle d'une autre stratégie l'est aussi. L'idée est de trouver une combinaison de stratégies permettant de réduire le montant du capital requis à mobiliser. De manière similaire, il s'agit de déterminer la probabilité pour un contaminant donné de dépasser un niveau d'exposition jugé dangereux pour la santé dès lors que l'exposition à un autre contaminant est très élevée.

Nous allons commencer par exposer le cadre théorique général de notre étude (Chapitre I). Ensuite, nous appliquerons ce cadre aux risques sanitaire (Chapitre II) et assurantiel (Chapitre III).

## Chapitre I Cadre général

Dans ce chapitre, nous présentons le cadre général de notre étude. Il s'agit d'abord de modéliser les queues de distribution des risques présents en utilisant la théorie des valeurs extrêmes. Ensuite, on modélisera la probabilité de ruine qu'un seul risque peut provoquer sachant sa distribution de probabilité connue. Enfin, on exposera les structures d'agrégation pertinentes sachant les marginales connues, c'est-à-dire les distributions des risques individuels connues.

### I - Théorie des valeurs extrêmes

La modélisation des événements extrêmes (ouragan, tremblement de terre ou inondation, crues, crises financières, krachs, chocs pétroliers) est aujourd'hui un champ de recherche particulièrement actif, notamment par l'importance de leurs impacts économiques et sociaux. En particulier, depuis quelques années, on note un intérêt croissant pour l'application de la Théorie des Valeurs Extrêmes (TVE) pour la modélisation de tels événements.

Prévoir certains événements ou comportements à partir de l'étude des valeurs extrêmes d'une série, est donc un des principaux objectifs pour ceux qui tentent d'appliquer la TVE. Cette théorie est apparue entre 1920 et 1940, grâce à Fréchet, Fisher et Tippett, Gumbel et Gnedenko. Lorsque l'on modélise le maximum d'un ensemble de variables aléatoires, alors, sous certaines conditions que nous préciserons plus loin, la distribution ne peut appartenir qu'à l'une des trois lois suivantes: Weibull (à support borné), Gumbel (à support non borné et à queues fines) et Fréchet (à support non borné et à queues épaisses). Ces trois lois définissent une famille de distributions statistiques appelée « famille parétienne », dont les applications sont innombrables et très diverses ; nous nous limiterons dans ce rapport à des exemples liés aux risques alimentaire et assurantiel.

La théorie des valeurs extrêmes permet d'évaluer les événements rares et les pertes associées à leur apparition. En d'autres termes, lorsqu'une perte importante survient, cette théorie permet d'en évaluer l'ampleur. Les résultats théoriques sur le comportement stochastique des extrêmes d'échantillon qu'offre la théorie des valeurs extrêmes permettent de proposer un cadre mathématique rigoureux pour réaliser de telles extrapolations. De plus,

cette théorie acquiert une importance particulière du fait qu'elle s'intéresse directement à la queue de la loi. En fait, seules les données extrêmes sont utilisées pour estimer les paramètres des modèles de la TVE ce qui assure un meilleur ajustement du modèle à la queue de la loi et donc une meilleure estimation de la Value-at-Risk (VaR). La VaR est une notion utilisée généralement pour mesurer le risque de marché d'un portefeuille d'instruments financiers. Elle correspond au montant de pertes qui ne devrait être dépassé qu'avec une probabilité donnée sur un horizon temporel donné. L'utilisation de la VaR n'est désormais plus limitée aux instruments financiers : on peut en faire un outil de gestion des risques dans tous les domaines (estimation du risque couru par une entreprise lié à la grippe A, par exemple).

## 1) Présentation de la théorie des valeurs extrêmes

### a) Loi généralisée des extrêmes

La théorie des valeurs extrêmes a pour but d'étudier la loi du maximum d'une suite de variables aléatoires réelles même si, et spécialement si, la loi du phénomène n'est pas connue. Formellement, considérons  $(X_1, \dots, X_n)$  une suite de  $n$  variables aléatoires indépendantes et identiquement distribuées de fonction de répartition  $F$ .

Pour étudier le comportement extrême des événements extrêmes, on considère la variable aléatoire  $M_n = \max(X_1, \dots, X_n)$ , le maximum d'un échantillon de taille  $n$ .

Comme les variables aléatoires sont indépendantes et identiquement distribuées, alors la fonction de répartition de  $M_n$  est donnée par :

$$F_{M_n}(x) = P(M_n \leq x) = P(X_1 \leq x, \dots, X_n \leq x) = P(X_1 \leq x) \dots P(X_n \leq x) = (F(x))^n. \quad (1)$$

L'équation (1) présente un intérêt très limité. De plus, la loi d'une variable aléatoire parente  $X$  (On dit que  $X$  est une aléatoire parente de la suite indépendante et identiquement distribuée  $(X_1, \dots, X_n)$  si cette suite a la même loi que  $X$ ) est rarement connue avec précision et, même si la loi de cette variable parente est connue avec exactitude, la loi du



terme maximum n'est pas toujours facilement calculable. Pour ces raisons, il est intéressant de considérer les comportements asymptotiques du maximum convenablement normalisé.

**Définition :** On dit que deux variables aléatoires réelles  $X$  et  $Y$  sont de même type s'il existe des constantes réelles  $a > 0$  et  $b \in \mathbb{R}$  telles que  $Y$  et  $aX + b$  suivent une même loi.

De façon analogue au théorème central limite, peut-on trouver des constantes de normalisation  $a_n > 0$  et  $b_n \in \mathbb{R}$  et une loi non-dégénérée  $H$  telles que :

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(a_n x + b_n)^n \rightarrow H(x) \quad (2)$$

Fisher et Tippett (1928) trouvent une solution à ce problème au moyen d'un théorème qui porte leur nom et qui est l'un des fondements de la théorie des valeurs extrêmes.

**Théorème 1 :** S'il existe deux suites de constantes de normalisation avec  $(a_n) > 0$  et  $(b_n) \in \mathbb{R}$  et une loi non-dégénérée  $H$  telles que  $\frac{M_n - b_n}{a_n} \rightarrow H$  (en loi), alors  $H$  est de l'une de ces trois limites :

la distribution de Gumbel :  $\Lambda(x) = \exp(-\exp(-x))$ ,  $x \in \mathbb{R}$ .

La distribution de Fréchet :  $\Phi_\alpha(x) = \exp(-x^{-\alpha}) \mathbb{1}_{]0, +\infty[}(x) \begin{cases} 0 & \text{si } x < 0 \\ \exp(-x^{-\alpha}) & \text{si } x \geq 0, \alpha > 0 \end{cases}$

la distribution de Weibull :  $\Psi_\alpha(x) = \exp(-(-x)^{-\alpha}) \mathbb{1}_{]-\infty, 0[}(x) + 1 \begin{cases} \exp(-(-x)^{-\alpha}) & \text{si } x < 0, \alpha < 0 \\ 1 & \text{si } x \geq 0 \end{cases}$

Ce théorème donne un résultat très intéressant. En effet, quelle que soit la loi limite de la variable parente  $X$ , la loi limite des extrêmes a toujours la même forme. Bien que le comportement de ces lois soit complètement différent, elles peuvent être combinées en une seule paramétrisation contenant un unique paramètre qui contrôle l'épaisseur de la queue de loi appelé indice des valeurs extrêmes (ou indice de queue) :

$$H_\xi(x) = \begin{cases} \exp\left(-\left(1 + \xi x\right)^{-\frac{1}{\xi}}\right), & \xi \neq 0, \quad 1 + \xi x > 0 \\ \exp(-\exp(-x)), & \xi = 0 \end{cases}$$

où  $H$  est une fonction non-dégénérée. Cette loi est appelée loi de valeurs extrêmes généralisée (GEV). En introduisant les paramètres de localisation  $\mu$  et la dispersion  $\sigma$  dans la

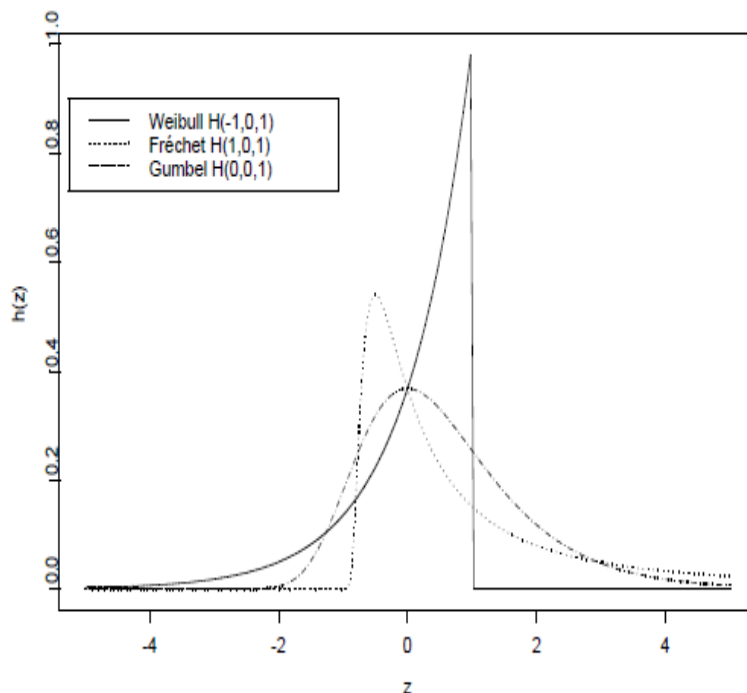
paramétrisation, on obtient la forme la plus générale de la distribution des valeurs extrêmes généralisées (GEV) :

$$H_{\xi,\mu,\sigma}(x) = \exp\left(-\left(1 + \xi \frac{x - \mu}{\sigma}\right)^{-\frac{1}{\xi}}\right), \quad \xi \neq 0, \quad 1 + \xi \frac{x - \mu}{\sigma} > 0$$

$\xi$  est un paramètre de forme (shape parameter).

Le théorème de Fisher-Tippett (1928) fournit, en quelque sorte, la contrepartie du Théorème Central Limite (TCL) dans le cas d'événements extrêmes. Cependant, contrairement au TCL, où la loi normale est la seule loi limite possible, dans le cas des extrêmes, trois types de loi limite sont possibles (Figure I.1) :

- la loi de Gumbel :  $\xi = 0$ .
- la loi de Fréchet :  $\xi > 0$  correspond au paramètre de Fréchet  $\alpha = 1/\xi$ .
- la loi de Weibull :  $\xi < 0$  correspond au paramètre de Fréchet  $\alpha = -1/\xi$ .



**Figure I.1 : Densités des lois de valeurs extrêmes**

Quand  $\xi = 0$  , la loi de X présente une décroissance de type exponentiel dans la queue de la distribution, on dit que l'on est alors dans le domaine d'attraction de Gumbel (c'est le cas par

exemple de la loi normale, exponentielle, gamma ou lognormale) ; le domaine de Fréchet, correspondant à une loi de  $X$  non bornée, est le domaine où  $\xi > 0$  (c'est le cas par exemple de la loi de Cauchy ou Pareto) et enfin quand  $\xi < 0$ , la loi de  $X$  est bornée et on dit que l'on est dans le domaine d'attraction de Weibull (c'est le cas par exemple de la loi uniforme ou béta).

La loi la plus courante rencontrée sur des données financières ou macroéconomiques est la loi de Fréchet. La loi de Weibull étant écartée car une loi de Weibull ne peut être rare or les variations en finance ou en économie sont généralement non bornées dont la mesure où on peut toujours rencontrer des variations extrêmes. La loi de Gumbel est extrêmement rare également. Néanmoins, cette dernière loi sert de référence et permet ainsi de mesurer, notamment graphiquement, l'écart entre les lois empiriques et la loi normale.

L'approche basée sur la GEV a été critiquée dans la mesure où l'utilisation d'un seul maxima conduit à une perte d'information contenue dans les autres grandes valeurs de l'échantillon. Pour pallier ce problème, la méthode POT (Peaks-over-Threshold) ou méthode des excès au-delà d'un seuil élevé a été introduite dans Pickands (1975).

#### **b) Excès au-delà d'un certain seuil (Distribution conditionnelle des excès)**

La méthode des excès au-delà d'un seuil (ou Peak Over Threshold, POT) repose sur le comportement des valeurs observées au-delà d'un seuil donné. En d'autres termes, elle consiste à observer non pas le maximum ou les plus grandes valeurs mais toutes les valeurs des réalisations qui excèdent un certain seuil élevé. L'idée de base de cette approche consiste à choisir un seuil suffisamment élevé et à étudier les excès au-delà de ce seuil.

On définit un seuil  $u \in \mathbb{R}$ ,  $N_u = \text{card}\{i: i = 1, \dots, n, X_i > u\}$  et  $Y_j = X_i - u > 0$  pour  $0 \leq j \leq N_u$  où  $N_u$  est le nombre de dépassements du seuil  $u$  par les  $X_{i \leq n}$  et les  $Y_{j \leq N_u}$  sont les excès correspondants (Figure I. 2).

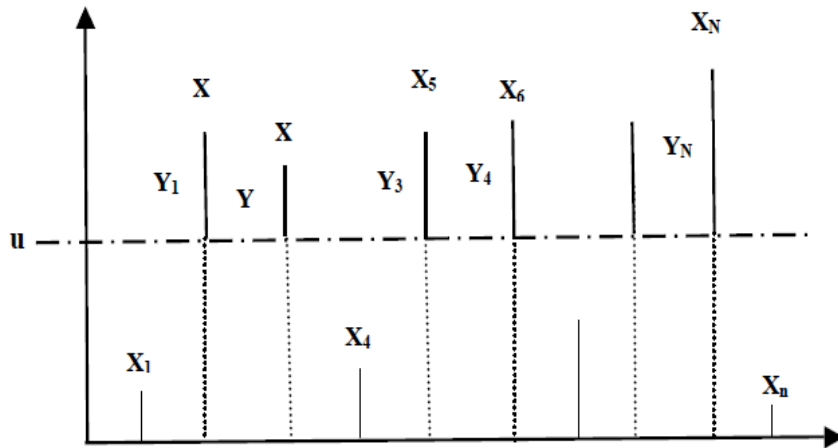


Figure I.2 : Méthode des excès

On cherche à partir de la loi  $F$  de  $X$  à définir une loi conditionnelle  $F_u$  par rapport au seuil  $u$  pour les variables aléatoires dépassant ce seuil. On définit alors la loi conditionnelle des excès  $F_u$  par :

$$F_u(y) = \mathbb{P}(X - u \leq y | X > u) = \frac{F(u + y) - F(u)}{1 - F(u)}, \quad y \geq 0$$

Le théorème de Pickands-Balkema-de Haan (1974, 1975) ci-après donne la forme de la loi limite pour les valeurs extrêmes : sous certaines conditions de convergence, la loi limite est une loi de Pareto généralisée que l'on notera GPD (Generalized Pareto Distribution).

**Théorème 2 :** Une fonction de répartition  $F$  appartient au domaine d'attraction maximale de  $H_\xi$  si et seulement si, il existe une fonction positive  $\beta(u)$  telle que :

$$\lim_{u \rightarrow x_F} \sup_{0 \leq y \leq x_F - u} |F_u(y) - G_{\xi, \beta(u)}(y)| = 0$$

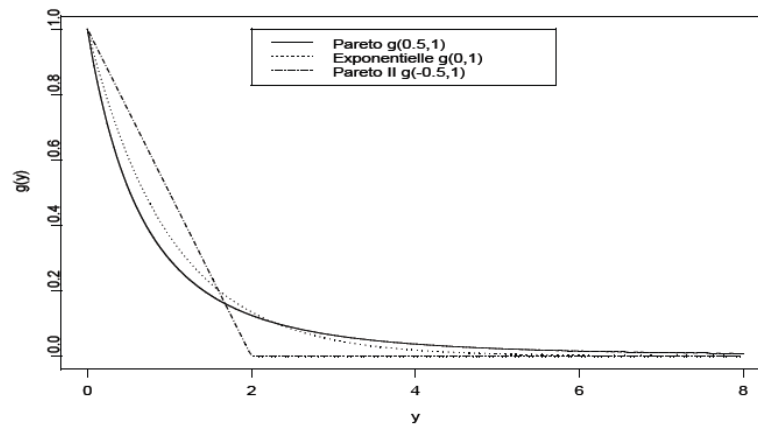
où  $F_u$  est la fonction de répartition conditionnelle des excès pour  $u$  élevé,  $x_F$  est le point terminal de  $F$ ,  $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$  et  $G_{\xi, \beta(u)}(y)$  est la GPD donnée par :

$$G_{\xi, \beta(u)}(y) = \begin{cases} 1 - (1 + \xi \frac{y}{\beta(u)})^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\beta(u)}\right), & \xi = 0 \end{cases}$$

où  $y \geq 0$ , pour  $\xi \geq 0$  et  $0 \leq y \leq -\frac{\beta(u)}{\xi}$  pour  $\xi < 0$ .

Ce théorème montre l'existence d'une relation étroite entre la GPD et la GEV (Generalized Extreme Value) ; Pickands (1975) a montré que pour toute loi  $F$ , l'approximation GPD définie ci-dessus n'est vérifiée que s'il existe des constantes de normalisation et une loi non-dégénérée telle que l'équation 2 est vérifiée. Dans ce cas, si  $H$  est écrite sous la forme d'une GEV, alors l'indice de queue  $\xi$  est le même que celui de la GPD.

De même, pour la GPD, le cas où  $\xi > 0$  correspond aux lois à queues épaisses, pour lequel  $1 - G$  se comporte comme une puissance  $x^{-1/\xi}$  pour  $x$  assez élevé. Si  $\xi = 0$ , on trouve  $1 - \exp\left(-\frac{y}{\beta(u)}\right)$  : c'est une loi exponentielle de paramètre  $\beta$  et enfin  $\xi < 0$ , c'est la loi de Pareto de type II à support borné (Figure I.3).



**Figure I.3 : Densités des lois de valeurs extrêmes**

La GPD a les propriétés suivantes :

$$\mathbb{E}(Y) = \frac{\beta}{1 - \xi}, \quad (\xi < 1)$$

$$V(Y) = \frac{\beta^2}{(1 - \xi)^2(1 - 2\xi)}, \quad (\xi < \frac{1}{2})$$

En pratique, le choix du seuil constitue une difficulté. En fait,  $u$  doit être assez grand pour que l'approximation GPD soit valide, mais pas trop élevé afin de garder un nombre suffisant de dépassements pour estimer les paramètres du modèle. Le seuil doit être choisi de façon à faire un arbitrage, traditionnel en statistiques, entre le biais et la variance.

Généralement,  $u$  est déterminé graphiquement en exploitant la linéarité de la fonction d'excès moyenne  $e(u)$  pour la GPD (Embrechts et al., 1997). La fonction d'excès moyenne est donnée par la relation :

$$e(u) = \mathbb{E}(X - u | X > u > 0) = \frac{\beta + \xi u}{1 - \xi}, \quad \xi < 0$$

Cette technique fournit une aide précieuse, cependant, il ne faut pas attendre d'elle la bonne valeur de  $u$ . En pratique, plusieurs valeurs de  $u$  doivent être testées. Ce problème du choix a suscité de nombreux travaux dans la littérature. Beirlant *et al.*(1976) suggèrent de choisir le seuil  $u$  qui minimise l'erreur quadratique moyenne asymptotique de l'estimateur de Hill de l'indice de queue, tout en supposant que  $F$  appartient au domaine d'attraction maximal de Fréchet.

## 2) Exemple de comportements limites dans les queues

Dans cette sous-section, nous proposons trois exemples illustrant comment se manifestent les lois limites de la GEV et la GPD en pratique, compte tenu de différentes hypothèses sur la loi  $F$  de la variable parente  $X$  de l'échantillon.

### i) Cas de la loi exponentielle

Pour la loi exponentielle de paramètre  $\lambda = 1$ , la fonction de répartition de cette loi est  $F(x) = 1 - e^{-x}$  pour  $x \geq 0$ . En posant  $b_n = \log(n)$  et  $a_n = 1$  alors,

$$F^n(a_n x + b_n) = (1 - e^{-x - \log(n)})^n = (1 - \frac{e^{-x}}{n})^n \rightarrow \exp(-e^{-x}) = \Lambda(x),$$

C'est-à-dire que le maximum convenablement normalisé de la loi exponentielle converge vers la loi de Gumbel. Cette loi appartient au domaine d'attraction maximal de Gumbel.

Aussi, les lois dans le domaine d'attraction de la loi de Gumbel sont parfois appelées lois de type exponentiel.

En ce qui concerne la méthode des excès au-delà d'un seuil  $u$ , en prenant  $\beta_u = 1$ , alors, pour tout  $y > 0$ ,

$$\begin{aligned}
F_u(y) &= \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y \geq 0 \\
&= \frac{e^{-u} - e^{-u-y}}{e^{-u}} \\
&= 1 - e^{-y}.
\end{aligned}$$

Aussi, la loi limite est la loi GPD de paramètre  $\xi = 0$  avec  $\beta_u = 1$ . Notons que dans ce cas, la loi GPD n'est pas simplement la loi limite, mais il s'agit de la loi exacte pour tout  $u$ .

## ii) Cas de la loi de Pareto

Pour la loi de la fonction de répartition  $F(x) = 1 - cx^{-\alpha}$ , où  $c > 0$  et  $\alpha > 0$ . En posant  $b_n = 0$  et  $a_n = (nc)^{1/\alpha}$  alors on a pour  $x > 0$  :

$$F^n(a_n x + b_n) = (1 - c(a_n x)^{-\alpha})^n = \left(1 - \frac{x^{-\alpha}}{n}\right)^n \rightarrow \exp(-x^{-\alpha}) = \Phi_\alpha(x)$$

qui est la loi de Fréchet. La loi de Pareto appartient au domaine d'attraction maximal de Fréchet. Aussi, les lois dans le domaine d'attraction de la loi de Fréchet sont parfois appelées lois de type Pareto.

En se basant sur la méthode des excès au-delà d'un seuil  $u$  et en considérant  $\beta_u = ub$  pour  $b > 0$ , on a alors,

$$\begin{aligned}
F_u(y) &= \frac{F(u + uby) - F(u)}{1 - F(u)}, \quad y \geq 0 \\
&= \frac{cu^{-\alpha} - c(u + uby)^{-\alpha}}{cu^{-\alpha}} \\
&= 1 - (1 + by)^{-\alpha}
\end{aligned}$$

Pour  $\xi = 1/\alpha$  et  $b = \xi$ , la limite est la loi GPD de paramètre  $\xi$ .

## iii) Cas de la loi normale

Soit  $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-y^2/2} dy$  la fonction de répartition de la loi  $N(0; 1)$ . Un résultat mathématique dit que  $1 - F(x) \sim \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}$  au voisinage de  $+\infty$ , donc :

$$\lim_{u \rightarrow +\infty} \frac{1 - F(u + \frac{z}{u})}{1 - F(u)} = \lim_{u \rightarrow +\infty} [(1 + \frac{z}{u^2})^{-1} \exp\{-\frac{1}{2}(u + \frac{z}{u})^2 + \frac{1}{2}u^2\}] = e^{-z}$$

Si on suppose dans un premier temps que  $\beta_u = \frac{1}{u}$ , alors,

$$1 - \frac{1 - F(u + \frac{z}{u})}{1 - F(u)} = \frac{F(u + \beta_u z) - F(u)}{1 - F(u)} \rightarrow 1 - e^{-z}, u \rightarrow +\infty$$

et par la suite la loi limite des excès au-delà d'un seuil est la loi exponentielle.

Dans un second temps si on considère  $b_n$  solution de  $F(b_n) = 1 - \frac{1}{n}$  et  $a_n = \frac{1}{b_n}$ , on obtient :

$$n\{1 - F(a_n x + b_n)\} = \frac{1 - F(a_n x + b_n)}{1 - F(b_n)} \rightarrow e^{-x}$$

Et donc :

$$\lim_{n \rightarrow +\infty} F^n(a_n x + b_n) = \lim_{n \rightarrow +\infty} (1 - \frac{e^{-x}}{n})^n \rightarrow \exp(-e^{-x}) = \Lambda(x)$$

qui est la loi de Gumbel.

Bien que la loi de Gumbel et la loi exponentielle soient respectivement les limites exactes des maxima et des excès au-delà d'un seuil, les meilleures approximations sont obtenues en utilisant la loi GEV et la loi GPD. Ces lois sont très générales et permettent d'inclure même le cas où  $\xi \neq 0$ . En fait, il est généralement préférable de considérer ces deux lois même si on pense que la loi de Gumbel ou la loi exponentielle sont les vraies lois limites.

Bien que la loi de Gumbel et la loi exponentielle soient respectivement les limites exactes des maxima et des excès au-delà d'un seuil, les meilleures approximations sont obtenues en utilisant la loi GEV et la loi GPD. Ces lois sont très générales et permettent d'inclure même le cas où  $\xi \neq 0$ . En fait, il est généralement préférable de considérer ces deux lois même si on pense que la loi de Gumbel ou la loi exponentielle sont les vraies lois limites.



### 3) Estimation statistique

En se référant à la littérature, on note diverses méthodes qui ont été proposées pour estimer les paramètres des lois GEV et GPD, on trouve la méthode basée sur le maximum de vraisemblance (Smith, 1987), la méthode des moments (Christopeit, 1994), la méthode des moments pondérés (Greenwood et al. 1979), ou encore les méthodes bayésiennes (Lye et al. 1993). Il existe également des approches non paramétriques pour l'estimation de l'indice de queue. Les plus utilisées en pratique sont l'estimateur de Pickands (1974), l'estimateur de Hill (1975) (pour le cas des lois de type Fréchet uniquement) et l'estimateur de Dekkers-Einmahl-De Hann (1989). La méthode la plus populaire et qui sous certaines conditions est la plus efficace est la méthode de maximum de vraisemblance.

Dans ce qui suit, nous allons présenter dans un premier temps cette dernière méthode d'estimation paramétrique pour la GPD. Par la suite, nous présenterons l'estimation de la Value-at-Risk selon une approche basée sur la méthode POT (Peak Over Thershold). Et enfin, nous présenterons une autre méthode non paramétrique, le modèle de McNeil et Frey (2000) qui s'applique sur les données financières. Pour une description plus complète, voir Embrechts et al, (1997).

#### i) Estimation des paramètres de la GPD par le maximum de vraisemblance

Considérons à nouveau la loi GPD dont l'expression analytique est donnée par l'équation (2). La densité bi-paramétrique de la GDP est donnée par :

$$g(y) = \frac{1}{\beta} \left(1 + \xi \frac{y}{\beta}\right)^{-\frac{1}{\xi}-1},$$

pour  $y \geq 0$  si  $\xi \geq 0$ , et  $0 \leq y \leq -\frac{\beta}{\xi}$  si  $\xi < 0$ .

L'estimation de la GPD, par la méthode de maximum de vraisemblance, porte sur l'indice de queue  $\xi$  ainsi que sur le paramètre d'échelle  $\beta$ . L'expression de la log-vraisemblance est donc :

$$l((\beta, \xi): Y) = -N_u \ln(\beta) - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \ln\left(1 + \frac{\xi}{\beta} y_i\right)$$

pour un échantillon  $(Y_1, \dots, Y_{N_u})$  des excès. A partir d'ici, on peut dériver par rapport à chaque paramètre, obtenir les équations à résoudre, et dégager l'estimateur de maximum de vraisemblance (EMV),  $\hat{\theta}_{ML}$  de  $\theta = (\beta, \xi)$ .

Pour  $\xi > -\frac{1}{2}$ , Smith (1985, 1987), Hosking et Wallis (1987) prouvent que les conditions de régularité de la fonction de vraisemblance sont remplies et que l'estimateur du maximum de vraisemblance conduit à un estimateur non biaisé, asymptotiquement normal.

## ii) Estimation de la VaR ou quantiles extrêmes

Rappelons que la loi des excès au-delà de  $u$  suffisamment élevé est :

$$F_u(y) = \mathbb{P}(X - u \leq y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{\bar{F}(u) - \bar{F}(u+y)}{\bar{F}(u)}, \quad y \geq 0$$

où  $\bar{F} = 1 - F$  est la fonction de survie. Cela peut être réécrit comme :

$$F_u(y)\bar{F}(u) = F(u+y) - F(u)\bar{F}(u) - \bar{F}(u+y).$$

Ce qui équivaut à :

$$\bar{F}(u+y) = \bar{F}(u) - F_u(y)\bar{F}(u) = \bar{F}(u)\bar{F}_u(y)$$

On aura donc :

$$\bar{F}(u+y) = \bar{F}(u)\bar{F}_u(y) \tag{3}$$

Rappelons que l'équation du théorème de Balkema et Haan (1974) et Pickands (1975) donne un résultat limite pour  $F_u(y)$ . Ceci nous amène à proposer l'approximation suivante, pour  $u$  assez grand :

$$F_u(y) \approx G_{\xi, \beta(u)}(y), \quad u \rightarrow +\infty, y \geq 0$$

Ou encore :

$$\bar{F}_u(y) \approx \bar{G}_{\xi, \beta(u)}(y), \quad u \rightarrow +\infty, y \geq 0$$

où  $G_{\xi, \beta(u)}(y)$  est une GPD.

L'équation (3) montre que l'estimation de la queue de  $F$  pourrait être obtenue en estimant  $\bar{F}(u)$  et  $\bar{F}_u(y)$  séparément. Un estimateur naturel de  $\bar{F}(u)$  est la loi empirique :

$$\hat{F}_n(u) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{X_j > u\}} = \frac{N_u}{n}$$

où  $N_u$  est le nombre d'excès au-delà du seuil  $u$ .

D'autre part, l'approximation précédente permet de proposer un estimateur pour  $\bar{F}_u(y)$ , soit :

$$\hat{F}_u(y) = \bar{G}_{\hat{\xi}, \hat{\beta}}(y) = \left(1 + \hat{\xi} \frac{y}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

pour des  $\hat{\xi} = \hat{\xi}(N_u)$  et  $\hat{\beta}(u) = \hat{\beta}(N_u)$  adéquats.

L'estimateur résulte de la queue  $\bar{F}(u+y) = \bar{F}(x)$  (pour  $y > 0$ ) et a donc la forme :

$$\hat{F}(u+y) = \hat{F}(u) \hat{F}_u(y) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{y}{\hat{\beta}}\right)^{-1/\hat{\xi}}$$

Dans le cas de Gumbel et Fréchet ( $\xi \geq 0$ ), la restriction au domaine pour l'équation précédente est  $y \geq 0$  du fait que l'on estime la partie supérieure de la queue de  $\bar{F}$ .

En inversant cette équation, on obtient l'estimateur du quantile supérieur  $x_p$

$$\hat{x}_p = u + \frac{\hat{\xi}}{\hat{\beta}} \left[ \left( \frac{n}{N_u} (1-p) \right)^{-\hat{\xi}} - 1 \right],$$

pour  $p > F(u)$ .

La VaR n'est rien d'autre que le quantile extrême calculé à partir de la loi asymptotique des extrema (Generalized Pareto Distribution), obtenu en modélisant les pertes (ou les profits) extrêmes par la méthode des excès. Ainsi, la  $VaR_p$ , correspondant au modèle GPD inconditionnel, pour un horizon donné  $t$  et avec une certaine probabilité  $p$ , ou de façon équivalente à un niveau de confiance (typiquement 95% ou 99%)  $c = 1 - p$  est :

$$\overline{VaR}_p = u + \frac{\hat{\xi}}{\hat{\beta}} \left[ \left( \frac{n}{N_u} (1-p) \right)^{-\hat{\xi}} - 1 \right],$$

où  $\hat{\xi}$  et  $\hat{\beta}$  représentent des estimateurs de la loi GPD s'ajustant au  $N_u$  excès de rendements au-delà d'un seuil  $u$ , délimitant le début de la queue gauche (ou droite) de la loi parente.

### iii) *Le modèle de McNeil et Frey (2000)*

Il s'agit du modèle GPD conditionnel qui permet de combler les inconvénients des modèles standards de la TVE (Théorie des Valeurs Extremes). En effet, les données financières étant caractérisées par une variance qui n'est pas constante au cours du temps. McNeil et Frey (2000) suggèrent la prise en compte de la volatilité dans l'application de la TVE. L'estimation se fait ici en deux étapes :

**Etape 1:** Estimation du modèle de type GARCH (Generalized Auto Regressive Conditional Heteroscedastic) pour les rendements.

Les chroniques des rendements des données du risque assurantiel sont caractérisées par le « clustering » ou regroupement de volatilité, à savoir les périodes de fortes volatilités alternent avec les périodes de faibles volatilités. Ce phénomène, que nous appelons aussi hétéroscédasticité conditionnelle, est particulièrement fréquent dans les données financières.

Or, pour ce type de chroniques, les modèles linéaires utilisés auparavant, tels que les modèles autorégressifs moyenne mobile (ARMA), développés par Box et Jenkins sont limités : ils ne permettent pas la prise en compte des phénomènes de variabilité de la volatilité en fonction du temps et des mécanismes d'asymétrie. En effet, dans la modélisation ARMA, l'espérance conditionnelle varie au cours du temps alors que la variance conditionnelle ne change pas.

Face aux lacunes représentant ARMA(p,q) surtout pour les problèmes financiers, Engle a proposé une nouvelle classe de modèles autorégressifs conditionnellement hétéroscédastiques (ARCH) apte à capter le comportement de la volatilité dans le temps. Elle a été par la suite généralisée par Bollerslev pour obtenir la classe des modèles GARCH. Depuis ces travaux précurseurs, de nombreuses spécifications de la volatilité ont été proposées.

Dans ce qui suit et sans entrer dans les détails de la méthodologie GARCH, nous précisons les principaux éléments utilisés par la suite.

Ainsi, supposons la dynamique de la chronique des rendements  $r_t$  modélisée par :

$$r_t = \mu_t + \sigma_t z_t,$$

où  $\mu_t$  est la moyenne conditionnelle à l'ensemble de l'information disponible, à savoir les données de rendements, jusqu'au temps  $t - 1$  et  $z_t$  désigne les innovations supposées iid, d'espérance  $E(z_t) = 0$  et de variance  $V(z_t) = 1$  et de fonction de répartition marginale  $F_z$ . Pour simplifier la présentation, nous supposons que la variance conditionnelle  $\sigma_t^2$  de la chronique ajustée  $\varepsilon_t = r_t - \mu_t$  est un processus GARCH(1,1) :

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \gamma \sigma_{t-1}^2,$$

Où  $\beta_0 > 0$ ,  $\beta_1 > 0$  et  $\gamma > 0$ . De plus, le processus GARCH est stationnaire au sens faible si et seulement si la somme des coefficients  $(\beta_1 + \gamma)$  est inférieure à 1.

Sous l'hypothèse de normalité de la loi conditionnelle des innovations, la fonction log-vraisemblance d'un échantillon de  $n$  observations indépendantes et identiquement distribuées est donnée par :

$$L(\theta) = -2 \ln(2\Pi) - \frac{1}{2} \sum_{i=1}^n \ln(\sigma_t) - \frac{1}{2} \sum_{i=1}^n \frac{(r_t - \mu_t)^2}{\sigma_t}.$$

La maximisation de  $L$  permet d'obtenir l'estimateur de maximum de vraisemblance (EMV)  $\hat{\theta}_{ML}$  des paramètres inconnus ainsi que d'extraire des résidus standardisés :

$$\frac{r_{t-k} - \mu_{t-k}}{\sigma_{t-k}},$$

pour  $k$  allant de 0 à  $n-1$ .

**Etape 2:** La GPD pour l'estimation des quantiles des résidus

Les résidus standardisés ainsi obtenus sont utiles pour vérifier si la spécification GARCH retenue est adéquate. En particulier, les innovations doivent être indépendantes et identiquement distribuées. En général, cette hypothèse est plus plausible pour les innovations que pour la chronique des rendements. Si c'est le cas, la GPD, fondée sur l'hypothèse que les données sont indépendantes et identiquement distribuées, permettra d'estimer les queues de la loi de ces innovations. Par la suite, nous déduisons le quantile extrême  $z_p = \overline{VaR}_p(Z)$  en nous appuyant sur la formule :

$$\overline{VaR}_p = u + \frac{\hat{\xi}}{\hat{\beta}} \left[ \left( \frac{n}{N_u} (1-p) \right)^{-\hat{\xi}} - 1 \right],$$

appliquée sur les innovations.

Ainsi, la  $\overline{VaR}_{p,t+1}$  est donnée par :

$$\overline{VaR}_{p,t+1} = \hat{\mu}_{t+1} + \overline{VaR}_p(Z) \hat{\sigma}_{t+1}.$$

Cette approche combinée GARCH-GPD a divers avantages: elle permet d'une part de capturer la nature hétéroscédastique des chroniques temporelles étudiées via une spécification GARCH. D'autre part, elle permet une meilleure description des quantiles extrêmes via la TVE. Depuis les travaux précurseurs de McNeil et Frey [2000], différentes études ont adopté cette approche conditionnelle. Bali et Neftçi [2003] ont appliqué l'approche GARCH-TVE sur le taux d'intérêt à court terme des États-Unis. Ils prouvent que l'approche considérée est plus précise en matière d'estimation de la VaR qu'un modèle GARCH à erreurs Student (GARCH-T).

## II- Théorie de la ruine

### 1- Modèle classique de la théorie de la ruine

La théorie du risque remonte à Lundberg (1903) qui a réalisé l'importance des processus de Poisson dans les modèles d'assurance IARD<sup>3</sup>. Cramér a posé les fondations d'une analyse probabiliste des problèmes d'assurance, et toutes ces contributions sont cristallisées dans le modèle de Cramér-Lundberg (1926, 1930). Ce modèle représente l'évolution des réserves d'une compagnie et repose sur les processus de Poisson composés. Explicitement, si on note

$(R(t))_{t \geq 0}$  le processus qui modélise les réserves, on a :

$$R(t) = u + ct - \sum_{i=1}^{N(t)} X_i .$$

où

- $u$  et  $c$  sont deux réels supposés positifs, représentant respectivement les réserves initiales de la compagnie et le taux de cotisation demandé aux assurés, qui est supposé constant,
- $(N(t))_{t \geq 0}$  est un processus de Poisson homogène d'intensité ou fréquence  $\lambda$  représentant le nombre de sinistres,
- $X_i$  est une variable aléatoire positive représentant le montant du  $i^{\text{ème}}$  sinistre. Les  $(X_i)_{i \geq 1}$  sont supposés indépendants, identiquement distribués (noté i.i.d par la suite) et indépendants de  $(N(t))_{t \geq 0}$ , le nombre de sinistre. Ils sont positifs, de même fonction de répartition  $F_X$  et de moyenne  $\mu$ .

Les durées entre deux sinistres, notées ,  $V_k = U_k - U_{k-1}$  sont des variables aléatoires iid distribuées selon la loi exponentielle de paramètre  $\lambda$ .

Le concept de la théorie de la ruine (TR) se comprend comme étant la survenance d'un scénario défavorable pouvant conduire à l'impossibilité pour la société d'assurance de faire face à certains de ses engagements, aussi bien envers ses assurés que ses actionnaires, voire à devoir cesser son activité pour cause d'insolvabilité.

---

<sup>3</sup> Incendie, Accidents et Risques Divers



Le but premier de la théorie de la ruine a donc logiquement été de modéliser l'évolution de la richesse de la société par un processus stochastique, d'évaluer la probabilité de la ruine, soit la probabilité que le scénario traduisant un échec se réalise, et bien entendu d'estimer le niveau de réserve initiale pour rendre cette probabilité de ruine suffisamment faible.

Le modèle le plus connu ou classique de la TR représente le fonctionnement d'une société d'assurance de la manière suivante : On suppose que la société d'assurance encaisse des cotisations de ses assurés -appelés primes- de manière continue, à raison de  $c$  unités de comptes par unité de temps (1000 euros par heure par exemple). Elle dispose d'une réserve initiale  $u$  pour absorber un éventuel excès de sinistralité et doit indemniser ses assurés pour les sinistres qui la concernent.

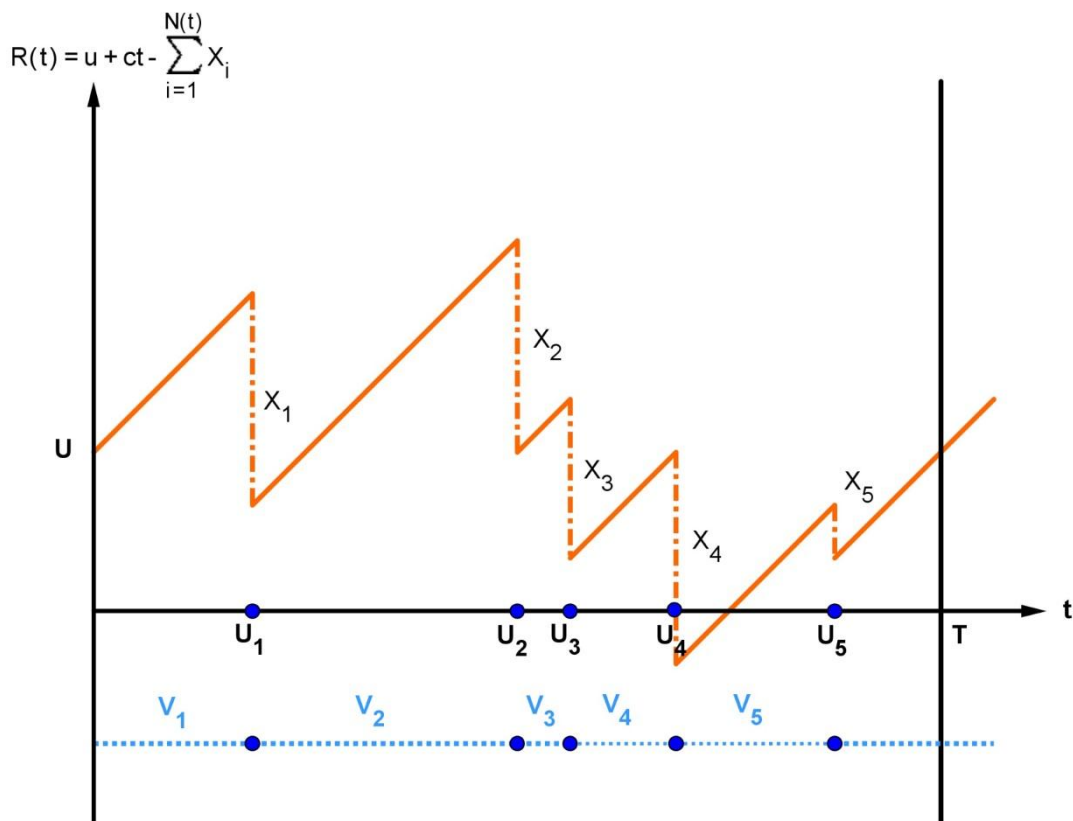


Figure I.4 : Un exemple de trajectoire du processus  $(R(t))_{t \geq 0}$

Le modèle de base, dit de Cramér-Lundberg, présente donc l'avantage d'être accessible à une large population: une compagnie d'assurance dispose d'un capital initial, qui augmente régulièrement grâce aux cotisations des assurés, mais diminue ponctuellement lors des remboursements de sinistres. La ruine intervient lorsque le montant des réserves tombe à 0.

De manière similaire, chaque individu dispose dans son organisme d'une charge corporelle initiale de contaminants, qui augmente régulièrement à cause de l'alimentation, mais diminue ponctuellement entre deux prises selon la théorie pharmacocinétique. Le risque sanitaire ici intervient lorsque la quantité de contaminant dans l'organisme dépasse un certain seuil.

## 2- Le Modèle KDEM (Kinetic Dietary Exposure Model)

Un modèle à temps continu, baptisé Kinetic Dietary Exposure Model (KDEM), permet de modéliser de manière parcimonieuse ces phénomènes d'accumulation et d'élimination. L'accumulation par voie alimentaire  $y$  est décrite par un processus ponctuel marqué  $(T_n, U_n)_{n \in \mathbb{N}}$ ,  $T_n$  désignant la date de la  $n$ -ième prise et  $U_n$  la quantité de contaminant ingéré correspondante. Les durées inter-prises,  $\Delta T_n = T_n - T_{n-1}$  avec  $n \geq 1$ , sont supposées indépendantes et identiquement distribuées, de loi  $G(dt) = g(t)dt$ . La suite  $\{T_n\}$  forme un processus de renouvellement standard. Par convention l'origine des temps correspond à la première prise considérée  $T_0 = 0$ . Les quantités de contaminants  $\{U_n\}_{n \in \mathbb{N}}$  sont supposées former une suite de variables aléatoires indépendantes et identiquement distribuées de loi  $F_U(dt) = f_U(t)dt$  indépendante de la série chronologique  $\{T_n\}$ . Les sauts du processus d'exposition sont ainsi les mêmes que ceux du processus à trajectoires (continu à droite et limite à gauche) défini par:

$$W_t = \sum_{n \leq N(t)} U_n, \quad \forall t \geq 0,$$

où  $N(t) = \sum_{n \in \mathbb{N}} I_{\{T_n \leq t\}}$  désigne le nombre de prise alimentaire avant  $t$ .

Un modèle pharmacocinétique linéaire à un compartiment est utilisé pour décrire le processus d'élimination. Précisément, entre deux prises consécutives, la quantité de contaminant présente dans l'organisme décroît avec le temps selon l'équation différentielle linéaire :

$$\frac{dx}{dt}(t) = -\theta x(t)$$

avec  $\theta > 0$ . Même si ce modèle est d'une grande simplicité puisqu'il ne dépend en effet que du paramètre d'accélération  $\theta$ , il est largement utilisé en pratique et rend compte des propriétés pharmacocinétiques d'un grand nombre de composés chimiques. Il est généralement décrit par la demi-vie biologique du contaminant, i.e. le paramètre  $\log(2) / \theta$  représentant le temps nécessaire pour que la quantité de contaminant au sein de l'organisme soit réduite de moitié en l'absence de prise supplémentaire.

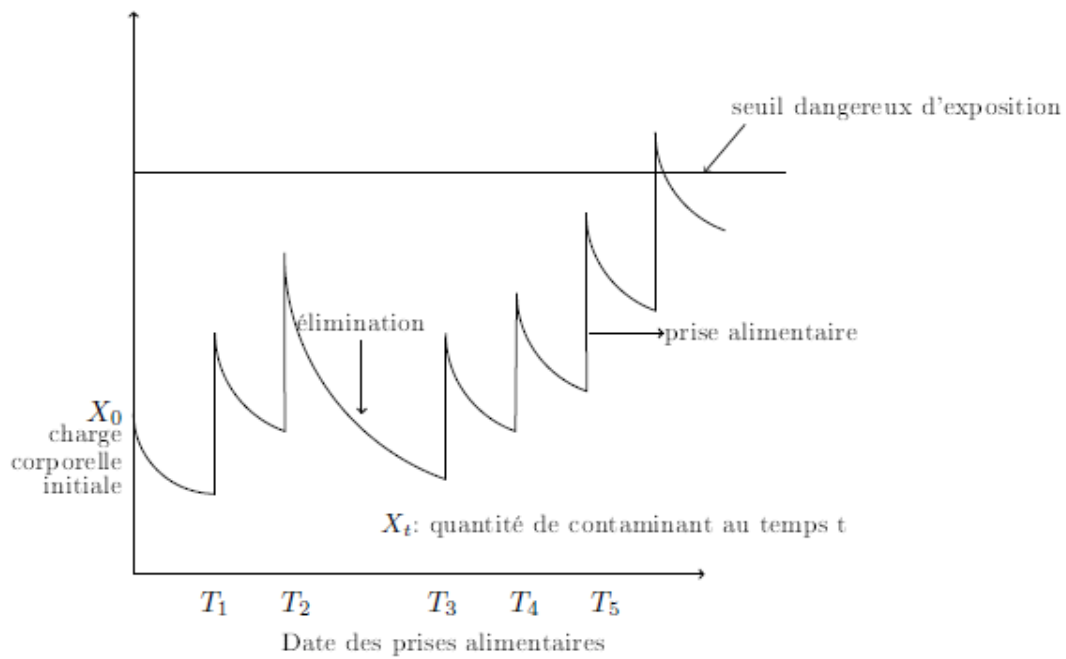
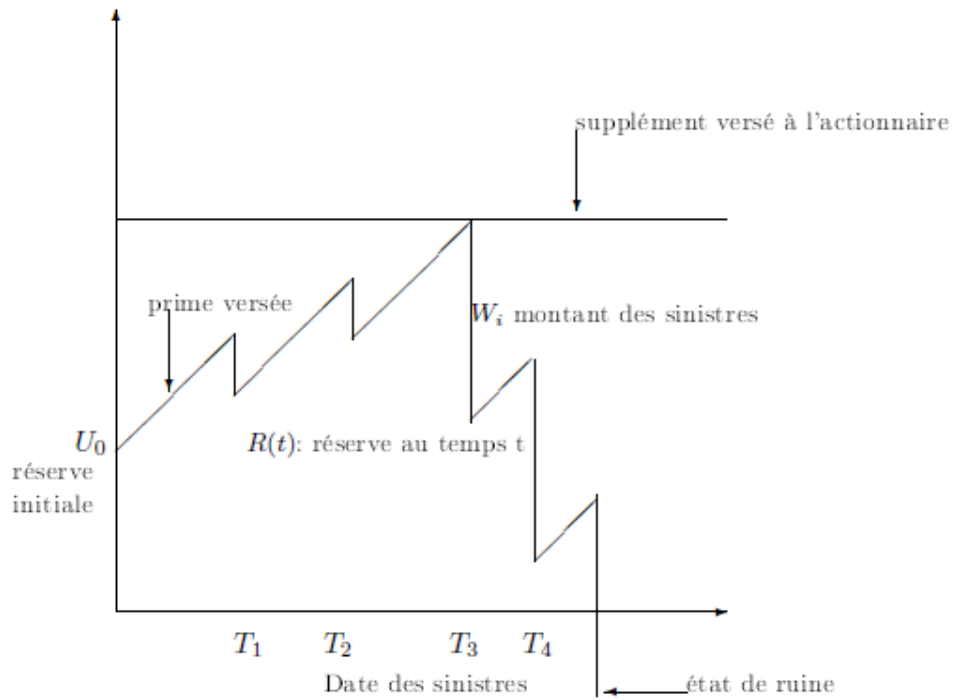
La dynamique du processus à temps continu « KDEM » est ainsi régie par trois paramètres : les distributions  $G$  et  $F_U$  caractérisant le comportement alimentaire au regard du phénomène de contamination et la demi-vie biologique du contaminant. Ce processus est dit « déterministe par morceaux » dans la mesure où il évolue de façon déterministe entre les prises successives.

### 3- Lien entre modèle de ruine et KDEM

Le modèle de ruine est un modèle dynamique à temps continu, qui décrit l'évolution de la réserve d'une compagnie d'assurance. La modélisation de la ruine permet d'évaluer la probabilité de la ruine, c'est-à-dire la probabilité que la réserve de la compagnie d'assurance soit nulle ou négative. Dans le cadre du risque alimentaire, le modèle KDEM permet d'évaluer la probabilité de dépasser le seuil, c'est à dire la probabilité que la quantité de contaminant atteigne ou dépasse le seuil dangereux pour la santé. Il en ressort, Figure I.5, que KDEM est le modèle "renversé" du modèle de ruine. La correspondance entre les différentes quantités est donnée dans le Tableau I.1.

<b>Risque Assurantiel</b>	<b>Risque alimentaire</b>
Montant des sinistres	Prise alimentaire : exposition
Reserve initiale	Charge corporelle initiale
Prime collectée	Elimination
Reserve en temps $i$	Quantité du contaminant en temps $i$

**Tableau I.1 : « Concordance » entre modèle de ruine et KDEM**



**Figure I. 5 : Modèle de Ruine (en haut) et modèle KDEM (en bas)**

En utilisant les mêmes notations dans la théorie de la ruine que dans le modèle KDEM, le montant cumulé des sinistres sera donné par :

$$W_t = \sum_{n \leq N(t)} U_n, \quad \forall t \geq 0,$$

où  $N(t) = \sum_{n \in \mathbb{N}} \mathbb{I}_{\{T_n \leq t\}}$  désigne le nombre de sinistres survenus jusqu'en  $t$  et  $U_n$  le montant du sinistre  $n$ . La richesse de la compagnie est donc :

$$R_t = u - \sum_{n \leq N(t)} Z_n, \quad \forall t \geq 0$$

avec  $u$  la richesse initiale,  $Z_n = U_n - c\Delta T_n$  où  $c\Delta T_n$  est le montant de prime collecté entre les instants  $T_n$  et  $T_{n-1}$ , ce qui correspond ici à la variable  $x$  que l'on peut modéliser par

$$\frac{dx}{dt}(t) = \theta x(t)$$

Notons qu'on n'a plus  $-\theta$  mais  $\theta$  car les primes augmentent entre deux sinistres alors que dans le risque alimentaire, la charge corporelle décroît entre deux prises alimentaires.

### III- Approches multivariées

Dans les risques alimentaire et assurantiel, on a affaire à plusieurs sources de risques interdépendants qui font qu'on a besoin d'une structure d'agrégation pour estimer le risque total. C'est à ce niveau que peuvent intervenir les copules. En effet, la recherche d'une optimalité du niveau des fonds propres par exemple 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. Dans cette sous-section, nous allons présenter les copules et les problématiques liées à leur choix et à leur estimation.

#### 1. Introduction aux copules

Commençons par rappeler la différence entre la dépendance et la corrélation : si  $X$  et  $Y$  sont indépendantes, alors  $X$  et  $Y$  sont 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, sur les risques assurantiel comme alimentaire, le coefficient de corrélation doit être utilisée avec précaution car il n'est pertinent qu'en présence de distributions elliptiques (distribution multivariée normale ou de Student) ou de dépendance linéaire.

***Définition d'une copule***

**Définition** : 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 de Sklar** : 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,  $\mathbb{R}^n$ ,

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

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

**Théorème** : 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 de Sklar. Si les marginales  $F_1, \dots, F_n$  sont continues, alors  $C$  est unique, autrement  $C$  est uniquement déterminée sur  $Im(F_1) \times \dots \times Im(F_n)$ ,  $Im(X)$  représentant l'ensemble des valeurs prises par  $X$

A partir du théorème ci-dessus, 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 :

$$C(u_1, \dots, u_n) = \mathbb{P}(U_1 \leq u_1, \dots, U_n \leq u_n).$$

Avec les copules, on ne travaille plus en nombre ou montant mais en rang. Autrement dit, dans la pratique, on transforme linéairement réalisations  $(x_1, \dots, x_n)$  en uniformes empiriques  $(u_1, \dots, u_n)$  où

$$u_i = \frac{\text{rang}(x_i)}{n + 1},$$

pour tout  $i$  de 1 à  $n$ .

**Définition de la densité d'une copule :** La densité  $c$  d'une copule  $C$ , si elle existe, est définie comme suit :

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

### Mesure de concordance

Rappelons la notion de concordance. Soient  $(x_1, y_1)$  et  $(x_2, y_2)$  deux réalisations d'un vecteur aléatoire continu  $(X, Y)$ , alors  $(x_1, y_1)$  et  $(x_2, y_2)$  sont dites :

- concordantes si  $(x_1 - x_2)(y_1 - y_2) > 0$
- discordantes si  $(x_1 - x_2)(y_1 - y_2) < 0$

**Théorème :** Soit  $\kappa$  une mesure de concordance pour les variables aléatoires  $X$  et  $Y$ .

- Si  $Y$  est une fonction croissante de  $X$ , alors  $\kappa(X, Y) = 1$ .
- Si  $Y$  est une fonction décroissante de  $X$ , alors  $\kappa(X, Y) = -1$
- Si  $\alpha$  et  $\beta$  sont des fonctions strictement croissantes,  $\kappa(\alpha(X), \beta(Y)) = \kappa(X, Y)$

Il est facile de construire des exemples où le coefficient de corrélation linéaire de Pearson n'est pas invariant par transformation strictement croissante. 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 :** Soit  $(X, Y)$  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 Kendall est :

$$\tau(X, Y) = \mathbb{P}[(X - X')(Y - Y') > 0] - \mathbb{P}[(X - X')(Y - Y') < 0]$$

Un estimateur du tau de Kendall, à partir d'un échantillon  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  peut s'écrire comme :

$$\hat{\tau}(X, Y) = \frac{2}{n(n-1)} \sum_{j=2}^n \sum_{i=1}^{j-1} \text{signe}[(x_j - x_i)(y_j - y_i)]$$

où  $\text{signe}(x) = 1$  si  $x \geq 0$ , et  $-1$  sinon.

**Définition :** Soit  $(X, Y)$  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 rho de Spearman est :

$$\rho_S(X, Y) = 3 \mathbb{P}[(X - X')(Y - Y') > 0] - \mathbb{P}[(X - X')(Y - Y') < 0]$$

Un estimateur du rho de Kendall, à partir d'un échantillon  $\{(x_1, y_1), \dots, (x_n, y_n)\}$  peut s'écrire comme :

$$\hat{\rho}_S(X, Y) = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

où  $R_i = \text{rang}(x_i)$ ,  $S_i = \text{rang}(y_i)$ ,  $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$  et  $\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i$ .

### Dépendance de queues

La dépendance de queue donne 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** : Soient  $X$  et  $Y$  deux variables aléatoires de fonctions de répartition respectives  $F_X$  et  $F_Y$ . On définit comme suit, lorsque la limite concernée existe, les coefficients de dépendances de queue :

- supérieur :  $\lambda_U(X, Y) = \lim_{\alpha \rightarrow 0^-} \mathbb{P}[(X > F_X^{-1}(\alpha))(Y > F_Y^{-1}(\alpha))]$
- inférieur :  $\lambda_L(X, Y) = \lim_{\alpha \rightarrow 0^-} \mathbb{P}[(X \leq F_X^{-1}(\alpha))(Y \leq F_Y^{-1}(\alpha))]$

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. Une interprétation similaire vaut pour les charges corporelles de deux contaminants.

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

**Définition** de la copule de survie : Soit la copule  $C(u_1, \dots, u_n) = \mathbb{P}(U_1 \leq u_1, \dots, U_n \leq u_n)$ , alors la copule de survie est :

$$\bar{C}(u_1, \dots, u_n) = C(1 - u_1, \dots, 1 - u_n) = \mathbb{P}(U_1 > u_1, \dots, U_n > u_n)$$

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  $\bar{C}$ . 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. Copules archimédiennes

L'intérêt des copules archimédiennes vient du fait qu'elles 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** des copules archimédiennes : Soit  $\varphi$  une fonction convexe, continue, strictement décroissante  $[0; 1]$  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 de générateur  $\varphi$ .

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

**Tableau I.2 : Caractéristiques de quelques copules**

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 aux risques assurantiel et alimentaire pour étudier l'impact de la survenance d'événements de forte intensité sur la dépendance entre branches d'assurance ou de risques alimentaires.

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.

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.

## Chapitre II Application au risque sanitaire

### I) Contexte

La connaissance de la contamination éventuelle des aliments et de leur composition en nutriments constitue un outil majeur de la sécurité sanitaire. Elle contribue en effet à documenter les expositions par voie alimentaire aux agents microbiologiques, chimiques et physiques ainsi que les apports nutritionnels. Ces expositions permettent d'évaluer les risques pour la population et ainsi d'éclairer les prises de décision en matière de gestion des risques (contrôle et réglementation), aux niveaux national, européen et international.

En France, la surveillance de la contamination des aliments par les substances est régulièrement assurée dans un cadre réglementaire au travers de plans de contrôle et de plans de surveillance, pilotés par les ministères compétents.

Cette connaissance peut être complétée et renforcée par des études de l'alimentation totale (EAT). Ces études s'appuient sur une méthode standardisée, recommandée par l'Organisation mondiale de la santé (OMS). Elles visent à rechercher différentes substances susceptibles d'être présentes dans les aliments « tels que consommés ».

En plus de la caractérisation des niveaux d'exposition des consommateurs aux contaminants présents dans leur alimentation, ces études permettent également d'évaluer, pour certains minéraux, la bonne adéquation des apports avec les besoins nutritionnels de la population. Elles permettent enfin d'identifier les aliments qui contribuent le plus aux apports et aux expositions.

L'étude de l'alimentation totale (EAT1) a été réalisée entre 2000 et 2004 par l'Institut national de recherche agronomique (INRA), en collaboration avec l'Agence française de sécurité sanitaire des aliments (AFSSA). Elle a permis de dresser un bilan de l'exposition de la population, des adultes et des enfants, aux contaminants inorganiques et minéraux, ainsi qu'aux mycotoxines.

#### 1) Choix des substances

La sélection des substances d'intérêt s'appuie sur les enjeux d'évaluation des risques pour différentes substances et aussi sur une revue de la littérature afin d'identifier les

substances émergentes pour lesquelles il y aurait un enjeu d'évaluation des risques. Certaines substances font aussi l'objet d'une réglementation fixant les teneurs maximales dans certaines denrées alimentaires ou dans les eaux des boissons.

Des EAT sont réalisées dans une trentaine de pays à travers le Monde (Australie, Cameroun, Canada, Chine, Espagne, Etats-Unis, Nouvelle-Zélande, République Tchèque, Royaume-Uni...).

## **2) Choix des aliments étudiés**

L'enquête s'appuie sur les données de consommations alimentaires de l'étude INCA2 (AFSSA, 2009a; Dubuisson et al., 2010; Lioret et al., 2010). Cette étude décrit les habitudes alimentaires des adultes et des enfants de plus de 3 ans en France, en qualité et en quantité.

Deux critères principaux ont été considérés :

- (i) les aliments les plus consommés et
- (ii) des aliments peu consommés mais susceptibles d'être fortement contaminés.

Les différents types d'aliments sélectionnés couvrent environ 90% de la consommation alimentaire des adultes et des enfants et sont choisis de manière à ne présenter que peu de variabilité interrégionale.

Pour chaque aliment, seule la partie comestible a été utilisée, puis les aliments ont été préparés « tels que consommés ». Par exemple, les fruits et légumes ont été lavés. Les légumes, la viande et les produits de la mer ont été cuits : braisés, poêlés, grillés, cuits au four, frits... Les aliments ont ensuite été mixés en plusieurs échantillons composites représentatifs des paniers de consommations et d'achats des consommateurs pour les huit inter-régions enquêtées, et analysés par des laboratoires accrédités. Ces analyses ont conduit à la production de plus de 230 000 résultats analytiques, consécutifs à la recherche des substances pertinentes dans les différents échantillons : chaque substance a ainsi été recherchée dans les aliments connus ou supposés les contenir d'après la littérature scientifique.

## **3) Analyse des résultats**

Les résultats des analyses des échantillons ont permis de calculer d'une part les apports nutritionnels et d'autre part l'exposition de chaque consommateur à des contaminants

chimiques. Ce calcul a été réalisé selon les recommandations internationales de l’OMS, en combinant les données de consommation de l’étude INCA2 avec les résultats analytiques.

Les expositions et apports ainsi calculés ont été comparés à des valeurs de référence afin de caractériser les risques pour différentes populations :

- pour les contaminants : dose journalière admissible (DJA) ou tolérable (DJT), dose hebdomadaire tolérable provisoire (DHTP), dose mensuelle tolérable provisoire (DMTP), dose sans effet ou *benchmark dose limit* (BMDL)... établies par des instances scientifiques françaises, européennes ou internationales : ces différentes notions sont couvertes par le terme plus générique de valeur toxicologique de référence (VTR);
- pour les apports nutritionnels : besoin nutritionnel moyen (BNM) dérivé de l’apport nutritionnel conseillé (ANC), pour évaluer le risque d’insuffisance d’apports, et la limite de sécurité pour évaluer le risque d’apports excessifs.

#### 4) Limites de l’interprétation des résultats et de l’étude

L’enquête permet d’évaluer les apports et expositions par la voie alimentaire et l’eau de boisson de la population générale. Les évaluations de risque ne prennent en compte que cette voie alimentaire, alors que, bien que cette voie constitue la principale voie d’exposition, des interrogations subsistent sur certaines substances pour lesquelles il convient d’étudier les expositions par les autres voies (air, contact).

L’enquête rend compte des apports et expositions de la population générale, c’est-à-dire des enfants de plus de 3 ans jusqu’aux adultes de 18 à 79 ans, mais pas de groupes de population particuliers comme les enfants de moins de 3 ans ou les femmes enceintes.

L’échantillonnage alimentaire de l’étude couvre environ 90% de la consommation alimentaire en France. Les 10% restant correspondent à des aliments peu consommés par la population générale et qui n’apparaissent pas susceptibles de contribuer de façon importante aux apports ou à l’exposition de la population générale aux substances ciblées, car les aliments à forte contribution ont été inclus dans le plan d’échantillonnage.

Par ailleurs, la durée de recueil des consommations (7 jours) induit une incertitude sur la caractérisation des très faibles et très forts apports ou expositions.

Cette base de données de la consommation alimentaire permet ainsi de disposer d'une vision actualisée des apports nutritionnels de la population vivant en France et d'évaluer leur conformité aux apports recommandés (identification des populations pour lesquelles les apports sont en adéquation par rapport aux besoins) et aux limites de sécurité lorsqu'elles existent.

Les individus sont donc exposés à des risques à travers leur alimentation. Les contaminants contenus dans les aliments sont ingérés tous les jours et s'y accumulent. Au-delà d'un seuil, les risques représentent un vrai danger pour l'individu qui risque de développer des maladies chroniques et non la fin de vie comme en assurance.

## **II) Présentation des contaminants**

### **1) Les effets du méthyle-mercure**

Le méthyle-mercure (MeHg) alimentaire est un contaminant environnemental qui peut causer des lésions du système nerveux central du fœtus et de l'enfant de moins de trois ans s'il est ingéré à des doses importantes par la mère durant la grossesse et pendant l'allaitement. Le méthyle-mercure est principalement présent dans les graisses et la chair des produits de la mer et dans le poisson. Ces graisses représentent un danger pour les personnes qui en consomment. Au delà d'un certain seuil, il provoque des convulsions, paralysies spastiques, cécité, surdité et retard mental.

Selon l'OMS<sup>4</sup>, les personnes ingèrent 2,4 µg de méthyle-mercure quotidiennement et la dose à ne pas dépasser serait de 100 µg par semaine.

### **2) Les effets du Cadmium**

Le cadmium (Cd) est un métal blanc argent, qui se trouve principalement dans la croûte terrestre. Il se substitue au calcium dans le cristal osseux et en modifie les propriétés mécaniques. Le calcium est libéré dans le liquide extracellulaire.

Le cadmium entraîne une importante fuite calcique dans les selles, même avec des expositions très faibles. Il est toujours présent dans la nature en combinaison avec le zinc. Il

---

<sup>4</sup> Organisation Mondiale de la santé

est présent aussi dans l'industrie comme sous-produit inévitable de l'extraction du zinc, du plomb et du cuivre.

On le trouve dans les pesticides et les engrais, il peut donc pénétrer dans l'environnement par le sol. L'absorption de cadmium se fait essentiellement par la nourriture. Les aliments qui sont riches en cadmium peuvent augmenter de façon importante la concentration en cadmium du corps humain.

On peut citer des aliments comme le foie, les champignons, les moules, les mollusques, les crustacés, la poudre de cacao, les légumes et les algues séchées. La cigarette est aussi une voie de contamination assez dangereuse, les fumeurs faisant partie du sous-groupe qui a un niveau significativement plus élevé que la normale. En effet, les fumées de tabac transportent le cadmium dans les poumons, ensuite le sang le transporte à travers tout le reste du corps où il peut augmenter les effets du cadmium déjà présent dans l'organisme du fait d'une alimentation riche en cadmium. Le cadmium est d'abord transporté jusqu'au foie par le sang. Puis il se lie aux protéines pour former des complexes qui sont transportés jusqu'aux reins. Le cadmium s'accumule dans les reins où il endommage le mécanisme de filtration. Cela entraîne l'excrétion des protéines essentielles et de sucre hors de l'organisme par les urines et d'autres dommages au niveau des reins. Les autres problèmes que le cadmium peut provoquer sont :

- Diarrhée, douleurs d'estomac et vomissements importants
- Fracture des os
- Echec de reproduction pouvant aller jusqu'à l'infertilité
- Problème au système nerveux central
- Problème au niveau du système immunitaire
- Désordre psychologique
- Probablement altération de l'ADN ou développement du cancer

En France, la quantité moyenne annuelle de cadmium ingérée par personne et par an est évaluée à 12 mg. L'absorption par voie digestive de plus de 0,9 g ou par voie respiratoire (poussières) peut entraîner des troubles graves.

### 3) Les effets des polychlorobiphényles

Les polychlorobiphényles (Pcb), sont des composés aromatiques chlorés. Les PCB sont des mélanges industriels fabriqués et utilisés en agriculture et dans l'industrie à partir des années 30 pour leurs propriétés isolantes (transformateurs électriques) et leur stabilité chimique et physique (encres, peinture). Leur présence dans l'environnement provient de ces utilisations, qui, au cours des années 1970, ont été restreintes aux systèmes clos (transformateurs, condensateurs). La production et l'utilisation des PCB sont interdites en France depuis 1987.

En raison de leur grande stabilité chimique et physique et leur faible biodégradabilité, la présence des Pcb demeure remarquablement stable dans certains "réservoirs" comme les sédiments marins ou des rivières. Ils s'accumulent dans les tissus graisseux des animaux et ce, tout au long de la chaîne alimentaire jusqu'à l'homme. On les retrouve particulièrement dans les aliments riches en graisses tels que poissons, crustacés, lait et produits laitiers, œufs. L'alimentation constitue donc la principale voie de contamination de la population générale (plus de 90 % de l'exposition totale).

Dans l'organisme humain, ces molécules s'accumulent préférentiellement dans le foie et le tissu adipeux. Leur élimination se fait par les selles et le lait maternel. Le temps mis par l'organisme pour éliminer 50% des dioxines accumulées (demi-vie) est en moyenne de 7 ans.

Les Pcb ont des effets dermatologiques sur l'homme (chloracné, affection cutanée observée après une exposition aiguë à forte dose) et les augmentations transitoires des enzymes hépatiques ; des observations sont par ailleurs de plus en plus en faveur d'un risque cardiovasculaire. Les autres problèmes que le cadmium peut provoquer sont :

- Le risque tératogène (malformations du nouveau-né). Il est suspecté en cas de forte exposition mais non démontré. On ne retrouve pas de malformation spécifiquement induite par les Pcb.
- De faibles excès de risque pour tous les cancers confondus ont été trouvés dans les études épidémiologiques menées auprès de populations professionnellement exposées aux dioxines.



L'exposition moyenne chez les adultes est estimée à 53,7 pg /kg p.c./mois<sup>5</sup>, soit 1,8 pg /kg p.c./jour et chez les enfants, à 82,7 pg /kg p.c./mois, soit 2,8 pg /kg p.c./j. Ce niveau plus important pour les enfants de 3-14 ans est à mettre en rapport avec le niveau de consommation alimentaire plus important proportionnellement au poids corporel. Cependant, 28 % de la population ont une exposition supérieure à la DMTP, rapportée à la vie entière. D'une manière générale, les poissons et autres produits de la mer sont plus contaminés que les produits animaux terrestres. Ici, nous nous baserons sur la deuxième EAT et choisirons donc une dose hebdomadaire tolérable de 17,5 pg.

#### 4) Description des données

Le laboratoire INRA Met@risk nous a fourni les expositions des adultes pour chaque jour de la semaine d'enquête pour les trois contaminants étudiés. Elles ont été calculées en multipliant les consommations des individus de l'enquête, par les contaminations mesurées dans les aliments dans la deuxième étude de l'alimentation totale (EAT2) réalisée par l'Anses pour les trois contaminants décrits ci-dessus. Nous considérons par la suite que l'individu consomme une fois par semaine ces contaminants et la consommation hebdomadaire est donc obtenue en additionnant les quantités journalières sur une semaine. Le tableau II.1 suivant nous donne les statistiques descriptives des valeurs hebdomadaires des consommations. Les données sont en microgramme, sauf pour les Pcb, qui sont exprimées en picogramme.

<b>Contaminants</b>	<b>Pcb (en pg)</b>	<b>Cd (en µg)</b>	<b>MeHg (en µg)</b>
Dose hebdomadaire Tolérable	17,5	2,5	1,6
Minimum	0,16	0,10	0,00
Mediane	2,51	0,98	0,02
Moyenne	2,91	1,04	0,10
1 <sup>er</sup> quantile	1,77	0,75	0,00
Maximum	21,19	3,55	4,47
3 <sup>ième</sup> quartile	3,58	1,28	0,09
Kurtosis	15,93	5,85	75,47
Skewness	2,77	1,14	7,52
<b>Expositions au-delà de la DHT</b>	<b>0,12%</b>	<b>0,60%</b>	<b>0,80%</b>

Source : Données INRA Mét@risk

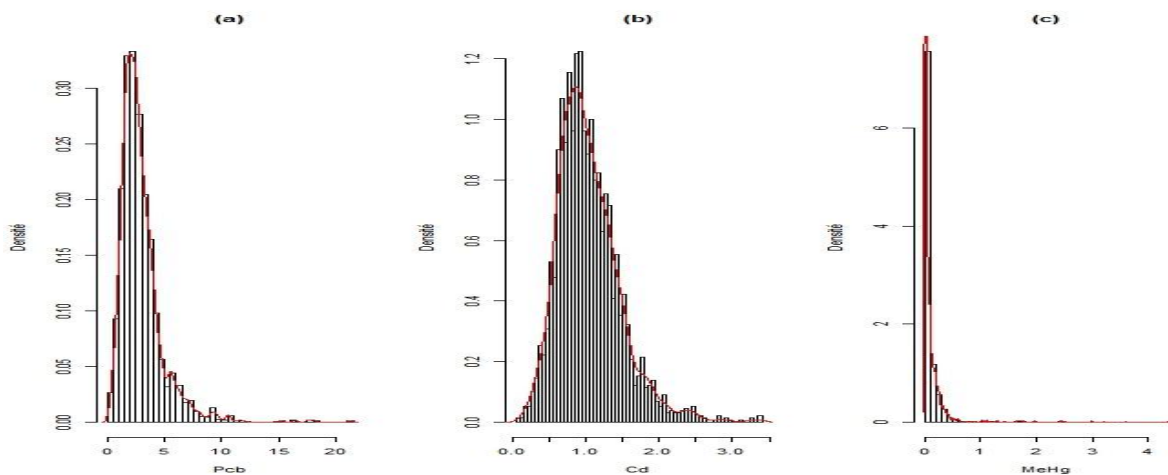
**Tableau II.1 : Statistiques descriptives hebdomadaires des contaminants**

<sup>5</sup> Picogramme par kilogramme de poids corporel par mois.  $1 \text{ pg} = 10^{-6} \mu\text{g}$  (microgramme)

La moyenne de consommation hebdomadaire des différents contaminants est plus faible que la dose hebdomadaire tolérable (DHT)<sup>6</sup>. Par ailleurs, la quantité maximale qu'un individu a déjà consommée par semaine selon notre base de données est supérieure à la dose hebdomadaire tolérable. Le coefficient d'aplatissement (kurtosis) de la série d'exposition à chaque contaminant est très élevé, de loin supérieur à 3 (kurtosis de la loi normale). Ceci signifie que la série des expositions est leptokurtique, caractéristique des queues de distributions plus épaisses que celles de la loi normale.

D'autre part, le coefficient d'asymétrie est positif et différent de 0 (skewness de la loi normale). Ceci signifie que la série des expositions est dégagée à gauche de la moyenne, et donc une queue de distribution étalée vers la droite.

Les histogrammes permettent de mieux visualiser ces informations :



**Figure II 1 : Histogramme des Pcb (a), du Cd (b) et du MeHg (c)**

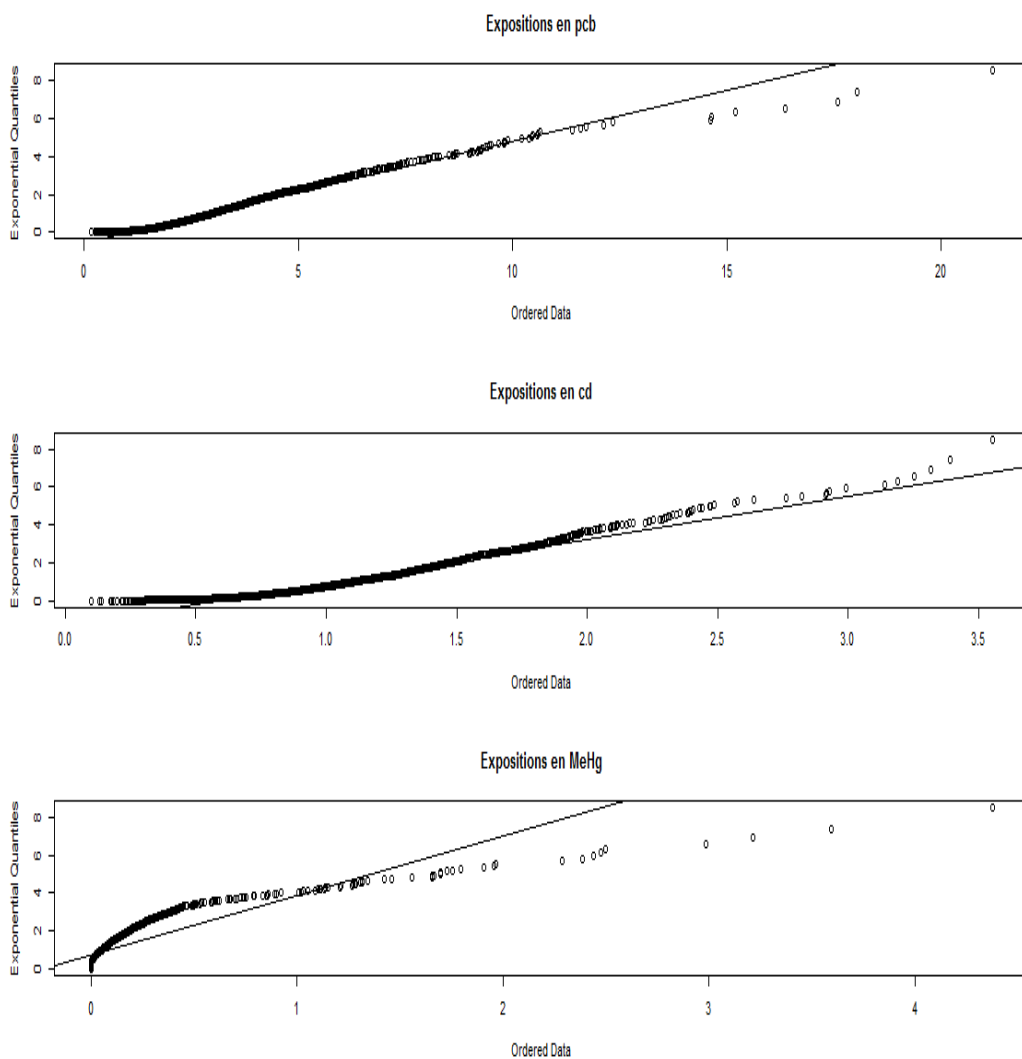
Enfin, afin de savoir si la trajectoire des expositions suit une loi particulière, nous traçons les QQplot. Le QQplot est un outil qui permet de voir si la distribution d'une variable dans un échantillon provient d'une distribution théorique spécifique. C'est un graphique qui oppose les quantiles de la distribution empirique aux quantiles de la distribution théorique envisagée. Si l'échantillon provient bien de cette distribution théorique, alors le QQplot sera linéaire.

Dans la théorie des valeurs extrêmes, on utilise souvent le QQplot pour tester si les données suivent la distribution exponentielle. L'intérêt de ce graphique est de ne nous

<sup>6</sup> La DHT est la quantité d'un contaminant qu'un individu peut ingérer tous les jours durant sa vie sans effet nocif sur sa santé.

permettre d'obtenir la forme de la queue de distribution. Trois cas de figures sont possibles :

- Les données suivent la loi exponentielle : la distribution présente une queue très légère, les points du graphique présentent une forme linéaire.
- Les données suivent une distribution à queue épaisse « fat-tailed distribution » : le graphique QQplot est concave.
- Les données suivent une distribution à queue légère « short-tailed distribution » : le graphique QQplot a une forme convexe.



**Figure II.2 : QQplot des valeurs des expositions**

L'analyse du QQplot (Figure II.2) révèle que les queues de distribution des expositions aux Pcb et au Cd semblent pouvoir être modélisées par une loi exponentielle, tandis que le

MeHg a une forme concave, signe d'une queue plus épaisse. Nous constatons également que nous avons des valeurs extrêmes dans la série des valeurs des expositions.

### III) Estimation du risque

#### 1) Cas d'un seul risque

Nous voulons étudier le risque sanitaire provenant de l'accumulation d'un seul élément dans l'organisme. Il s'agit en fait d'évaluer le risque dû à l'accumulation des Pcb, du Cd ou du MeHg.

Les données décrites ci-dessus sont relatives à la consommation de 2488<sup>7</sup> individus au cours d'une semaine. L'idée étant de calculer la probabilité pour que la quantité du contaminant  $j$  ( $j$  allant de 1 à 3) dans l'organisme au cours d'une année soit supérieure à un certain seuil, on doit d'abord simuler les séries d'expositions pour avoir une période d'observation plus longue. En fait, les trajectoires des processus d'expositions ne sont en général pas observables en pratique et l'estimation des quantités ne peuvent être faite que par simulation. On va donc considérer ces 2488 expositions hebdomadaires comme des réalisations possibles de l'exposition hebdomadaire d'un individu moyen de la population adulte.

##### a) Simulation des données

Dans un premier temps, nous avons supposé que les expositions hebdomadaires suivaient une loi exponentielle. On note  $E_j \sim \exp(\lambda_j)$ , pour  $j$  allant de 1 à 3,  $j$  étant le contaminant. Partant de cette hypothèse, simuler la trajectoire des expositions revient à estimer le paramètre de la loi exponentielle pour un individu moyen.

Si  $E_{ij}$  suit une loi exponentielle de paramètre  $\lambda_j$ , alors  $\widehat{\lambda}_j$  peut être estimé par la méthode des moments.  $(E_{ij})$  est un  $n$ -échantillon de la loi de  $E_j$  ( $n=2488$ ). Comme  $\mathbb{E}(E_j) = \frac{1}{\lambda_j}$  alors un estimateur de  $\lambda_j$  est donné par  $\widehat{\lambda}_j = \frac{1}{\frac{1}{n} \sum_{i=1}^n E_{ij}}$  où  $n$  est le nombre d'observations. A

---

<sup>7</sup> On a seulement 2488 individus car les données brutes ont été préalablement nettoyées. En particulier, les femmes enceintes ou allaitantes ne sont pas prises en compte en raison des besoins nutritionnels trop spécifiques. En outre, sont éliminés de la base de données les individus présentant des valeurs manquantes sur les variables d'intérêt afin de ne pas biaiser les résultats

partir de ce paramètre, on peut simuler les valeurs des expositions hebdomadaires d'un individu moyen à un contaminant choisi durant une période donnée (une vie de 50 ans par exemple) et calculer la charge corporelle y afférant comme nous le montrons dans le paragraphe suivant.

Dans un second temps, nous avons fait appel à la méthode de tirage selon la distribution empirique. Cette méthode a été motivée par la corrélation des consommations : en effet, un individu qui mange beaucoup de poisson aura du MeHg et du Cd alors que celui qui n'en mange pas n'aura que du Cd via les aliments autre que le poisson. Aussi, l'enquête a été réalisée en mesurant dans chaque assiette de nourriture la teneur en chacun des trois contaminants. Cela montre donc qu'il y a une dépendance entre les contaminants, ou plus précisément, qu'un aliment peut contenir les trois types de contaminants à la fois. Or simuler les valeurs des expositions en supposant que chaque exposition suit une loi exponentielle de paramètre l'inverse de sa moyenne, revient à ignorer la dépendance entre les contaminants. Si nous voulons calculer la probabilité pour que la charge corporelle en un contaminant soit supérieure à un seuil  $u$  sachant que la charge corporelle en un autre contaminant est supérieure à un seuil  $v$  alors, il est nécessaire de garder cette structure de dépendance entre les données d'expositions.

Trouver une distribution paramétrique trivariée est difficile car les données présentent des zéros. La méthode de tirage selon la méthode empirique consiste alors à tirer de façon aléatoire avec remise dans une base de données. Cette méthode a l'avantage de tirer chaque individu avec toutes les données le concernant, ce qui permet de garder la structure de dépendance. Ainsi un individu quelconque est stocké dans la nouvelle base de données avec les quantités de Pcb, de Cd et de MeHg qu'il a ingérées au cours de la même semaine. Cette méthode est efficace tant dans l'étude d'un seul risque que dans celle des risques liés.

Dans la suite de notre étude, nous utilisons les deux techniques de simulations des trajectoires des expositions pour le calcul des trajectoires de charge corporelle quand l'application se fera sur un seul risque, mais nous nous limitons uniquement aux trajectoires d'expositions obtenues par la deuxième méthode pour calculer les charges corporelles dans le cas des risques liés.

### *b) Calcul de la charge corporelle*

Nous avons simulé les données pour un individu moyen sur une durée de 50 ans, soit

2500 valeurs d'expositions.

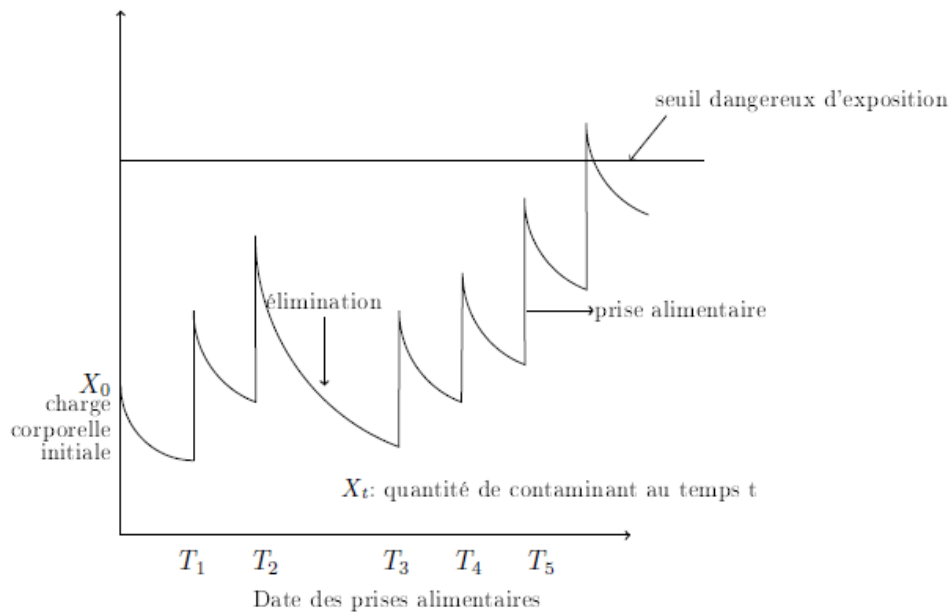
Les données relatives aux charges corporelles sont obtenues à partir des expositions. Notons  $E_{ij}$  l'exposition au contaminant  $j$  au cours de la  $i^{\text{ème}}$  semaine. Alors la charge corporelle, qui est la quantité totale du contaminant  $j$  dans l'organisme au cours de cette  $i^{\text{ème}}$  semaine est donnée par :

$$X_{i,j} = X_{i-1,j} \exp(-\theta_j) + E_{i,j}$$

Et l'état initial  $X_{0,j}$ .

Cette relation suppose qu'entre la  $(i - 1)^{\text{ème}}$  et la  $i^{\text{ème}}$  semaine, la charge corporelle de  $j$  est passée de  $X_{i-1,j}$  à  $X_{i-1,j} * \exp(-\theta_j)$ ,  $\exp(-\theta_j)$  étant le coefficient d'élimination (processus d'élimination) qui est donc inférieur à 1.  $\theta_j$  est donné par la relation  $\theta_j = \ln 2 / DV_j$  où  $DV_j$  est appelée la demi-vie du contaminant. La demi-vie d'un contaminant est le temps nécessaire pour que la quantité de ce contaminant dans l'organisme soit éliminée de moitié en l'absence d'une prise supplémentaire d'un produit contenant ce contaminant. Dans le cadre de notre étude, nous travaillons avec les demi-vies données dans le Tableau II.2. Ces données ne sont pas idéales (elles changent en particulier selon l'organe cible considéré), mais elles nous permettent de conduire ces premières analyses.

La charge corporelle en Pcb par exemple au cours de la semaine  $i$  dépend de la charge corporelle en Pcb de la semaine précédente qu'on multiplie par un coefficient d'élimination et à laquelle on rajoute la nouvelle exposition, c'est-à-dire la quantité des Pcb qu'on a ingérée au cours de cette semaine. Pour simplifier, on considère une ingestion par semaine plutôt que de suivre les apports exacts qui, pour les Pcb et le Cd présents dans de nombreux aliments, sont quotidiens. L'évolution de la charge corporelle peut être visualisée sur le schéma de la figure II.3 :



**Figure II.3 : Evolution de la charge corporelle**

Sur la Figure II.3, à l'instant initial, l'individu considéré a une quantité de contaminant dans l'organisme, que nous notons génériquement  $X_0$ . Entre  $T_0$  et  $T_1$ , l'individu élimine la différence entre  $X_0$  et  $X_0 * \exp(-\theta)$  et ingère en  $T_1$  la quantité  $E_1$ . On remarque donc un saut dans la trajectoire de la charge corporelle. La vitesse d'élimination dépend de la durée de la demi-vie. Cette dernière est assez longue pour les Pcb et le Cd et plus courte pour le MeHg. Ce sont les demi-vies très longues de ces deux premiers contaminants qui favorisent l'accumulation du contaminant dans l'organisme, car ces contaminants se retrouvent dans la plupart des aliments consommés et ne seront que peu éliminés vu la fréquence des prises.

$X_0$  dépend de l'hypothèse qu'on se donne. En réalité, cet état initial dépend à la fois de la manière dont la mère transmet sa propre charge corporelle au fœtus pendant la grossesse (et éventuellement pendant l'allaitement) et aussi des contaminants. En pratique, on retient deux hypothèses. Soit on considère que la charge corporelle à l'instant initial est nulle c'est-à-dire que l'organisme ne contient aucun de ces contaminants à ce moment-là, soit on utilise la moyenne de long terme du processus (valeur autour de laquelle la trajectoire se stabilise).

Dans le second cas donc, la moyenne de long terme est obtenue en supposant  $(X_{i,j}) = \mathbb{E}(X_{i-1,j}) = X_{0,j}$ .

On a:

$$\begin{aligned} \mathbb{E}(X_{i,j}) &= \mathbb{E}[X_{i-1,j} \exp(-\theta_j) + E_{i,j}], \\ \Rightarrow \mathbb{E}(X_{i,j}) &= \mathbb{E}[X_{i-1,j} \exp(-\theta_j)] + \mathbb{E}(E_{i,j}), \\ \Rightarrow \mathbb{E}(X_{i,j}) &= \frac{\mathbb{E}(E_{i,j})}{[1 - \exp(-\theta_j)]} = X_{0,j} \end{aligned}$$

Ceci équivaut à considérer que la chaîne commence à l'état stationnaire. Nous avons donc, à partir des valeurs d'expositions simulées et de la valeur initiale de la charge corporelle, simulé des trajectoires de charge corporelle.

En Figure II.4, le QQplot des charges corporelles montrent des queues épaisses pour la trajectoire de Pcb et de Cd. Par contre, le MeHg semble suivre la loi exponentielle.

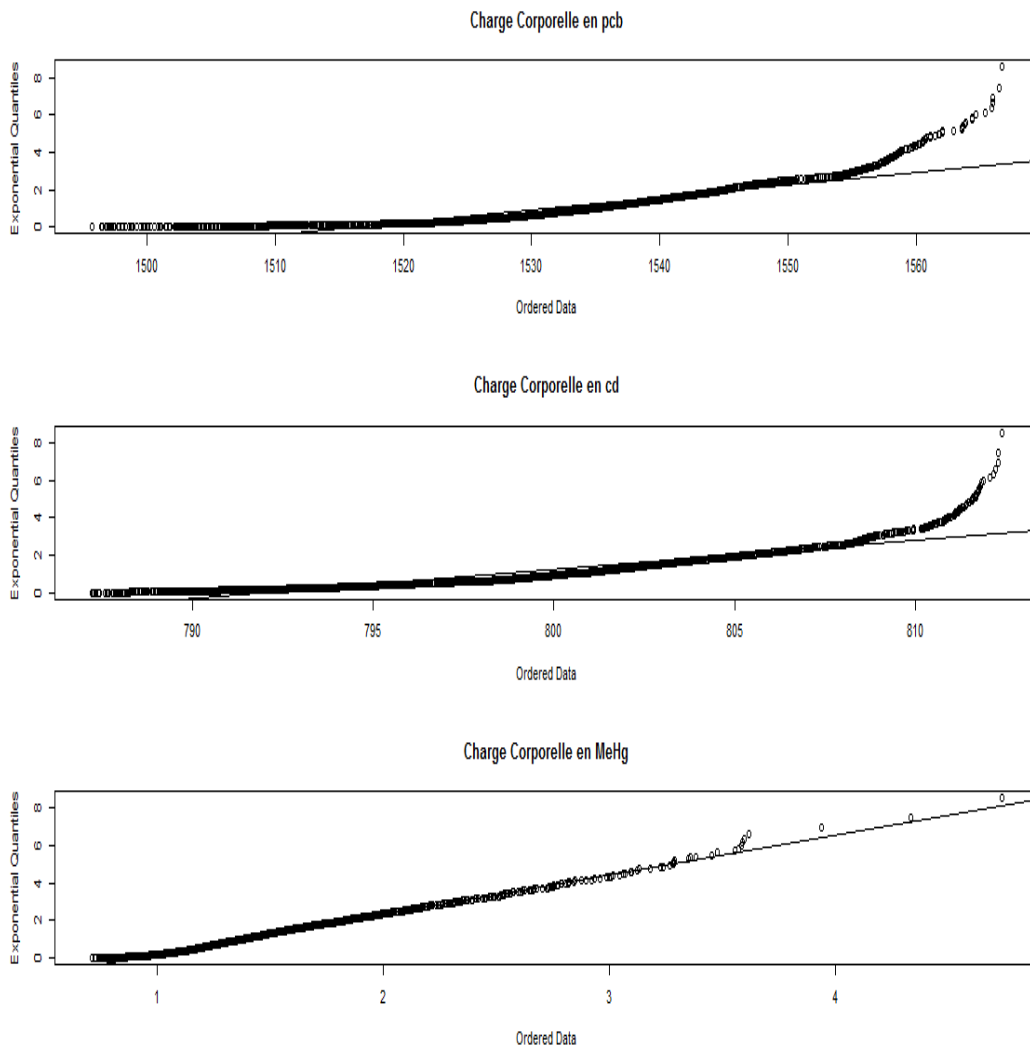


Figure II

Figure II.4 : QQplot des charges corporelles sur données simulées



<b>Contaminants</b>	<b>Pcb</b>	<b>Cd</b>	<b>MeHg</b>
Demi-vies (en semaines)	364	520	6
Dose hebdomadaire Tolérable en µg	17,5 pg <sup>8</sup>	2,5	1,6
Moyenne hebdomadaire en µg	2,93	1,04	0,10
Dose Tolérable de long terme en µg	8491pg	1876	14

**Tableau II 1: Demi-vies, DHT et Seuils des différents contaminants**

### c) Détermination de la dose tolérable de long terme (TBB<sup>9</sup>)

Le seuil ici est déterminé à partir de ce que l'on appelle la valeur toxicologique de référence (VTR). Cette valeur, fixée le plus souvent à partir d'effets toxiques observés chez l'animal de laboratoire, est une quantité maximale d'un contaminant que les experts estiment pouvoir être consommée sans que l'on puisse craindre d'effets néfastes pour la santé humaine durant la vie entière.

Les valeurs toxicologiques de référence incluent des marges de sécurité qui tiennent compte de l'extrapolation d'effets observés chez l'animal à d'éventuels effets chez l'homme et des différences entre les individus d'une même espèce (hommes et animaux).

Les toxicologues s'accordent sur le fait que pour des toxiques à effets chroniques, une exposition supérieure aux valeurs toxicologiques de référence chroniques pendant de courtes périodes de la vie d'un individu n'induit pas nécessairement un risque significatif pour la santé des personnes (Elles sont exprimées en masse du contaminant par kilogramme de poids corporel et par jour (ou par semaine ou par mois)).

L'Organisation Mondiale de la Santé a fixé les doses tolérables hebdomadaires des trois contaminants que nous étudions. Une dose hebdomadaire tolérable provisoire a été définie

<sup>8</sup> picogramme

<sup>9</sup> Tolerable Body Burden

par le JEFCA<sup>10</sup> pour le MeHg. Les valeurs que nous retenons finalement dans notre étude sont contenues dans le tableau II.2.

A partir de ces recommandations, nous calculons le seuil tolérable de long terme, obtenu comme celui correspondant à des prises hebdomadaires exactement égales à la dose hebdomadaire tolérable (DHT). Le processus se stabiliserait donc, en suivant un raisonnement similaire à celui du calcul de la charge corporelle initiale, autour du seuil tolérable de long terme (Tolerable Body Burden, TBB).

$$TBB_j = \frac{DHT_j}{1 - \exp(-\theta_j)}$$

Les charges corporelles retenues sont obtenues en utilisant les expositions issues du tirage selon la distribution empirique.

On regarde donc la probabilité pour que la charge corporelle en un contaminant soit supérieure au seuil du même contaminant au moins une fois sur une durée donnée et pendant au moins un an sur une durée donnée. Il s'agit en fait de calculer les probabilités

$$P_T(X_{ij} > TBB_j) \quad \text{et} \quad P_T \left( \sum_{i=1}^{52} X_{ij} > u_j \right),$$

où  $T$  désigne la durée sur laquelle on se place.

Les probabilités sont estimées sur les données simulées. Nous réalisons en fait 1000 simulations de vie, sachant qu'une vie correspond à une durée de 50 ans, soit 2600 semaines.

Ces probabilités sont très faibles, ce qui signifie que la quantité de contaminant dans l'organisme ne dépasse jamais le seuil, du moins dans la population enquêtée.

Pour quantifier ces probabilités presque nulles, nous avons fait appel à la méthode de multilevel splitting. Cette méthode consiste à multiplier les trajectoires qui approchent l'évènement (ici seuil) voulu et de laisser mourir les autres.

On suppose que l'on désire déterminer la probabilité d'atteindre un niveau  $u$ . On suppose que l'on connaisse les  $m$  niveaux intermédiaires  $u_1 < u_2 < \dots < u_m < u_{m+1}$ . On suppose que le temps d'arrêt vaut  $T_{max}$  et que le nombre de trajectoire à simuler vaut  $N$ .

<sup>10</sup> « Evaluation of certain food additives and contaminants for methylmercury », Sixty first report of Joint FAO/WHO Expert Committee on Food Additives, Technical Report Series, WHO, Geneva, Switzerland, 2003.

1. On simule  $N$  trajectoires partant de  $x_0 < u_1$ , de longueur  $T_{max}$ , indexées par  $k \in \{1, 2, \dots, N\}$ , qu'on note  $X^{[k]} = \{X_t^{[k]}, t \in [0, T]\}$

2. Pour  $j \in \{1, \dots, m\}$  :

(a) On note  $I_{1,j}$  le sous ensemble des indices des trajectoires ayant atteint le niveau  $u_j$  avant le temps  $T_{max}$ , c'est-à-dire au temps  $\tau_{u_j}^k = \{t \geq 0; X_t^{[k]} \geq u_j\}$  et on note  $I_{0,j}$  le sous ensemble des autres indices.

(b) Pour chaque trajectoire d'indice  $k'$  dans  $I_{0,j}$  on tire aléatoirement  $k$  dans  $I_{1,j}$  et on redéfinit la trajectoire  $X_t^{[k']}$  comme la trajectoire confondue avec  $X_t^{[k]}$  jusqu'au temps  $\tau_{u_j}^k$  puis on la prolonge ensuite jusqu'au temps  $T_{max}$  en simulant depuis l'état  $X_{\tau_{u_j}^k}^{[k]}$ .

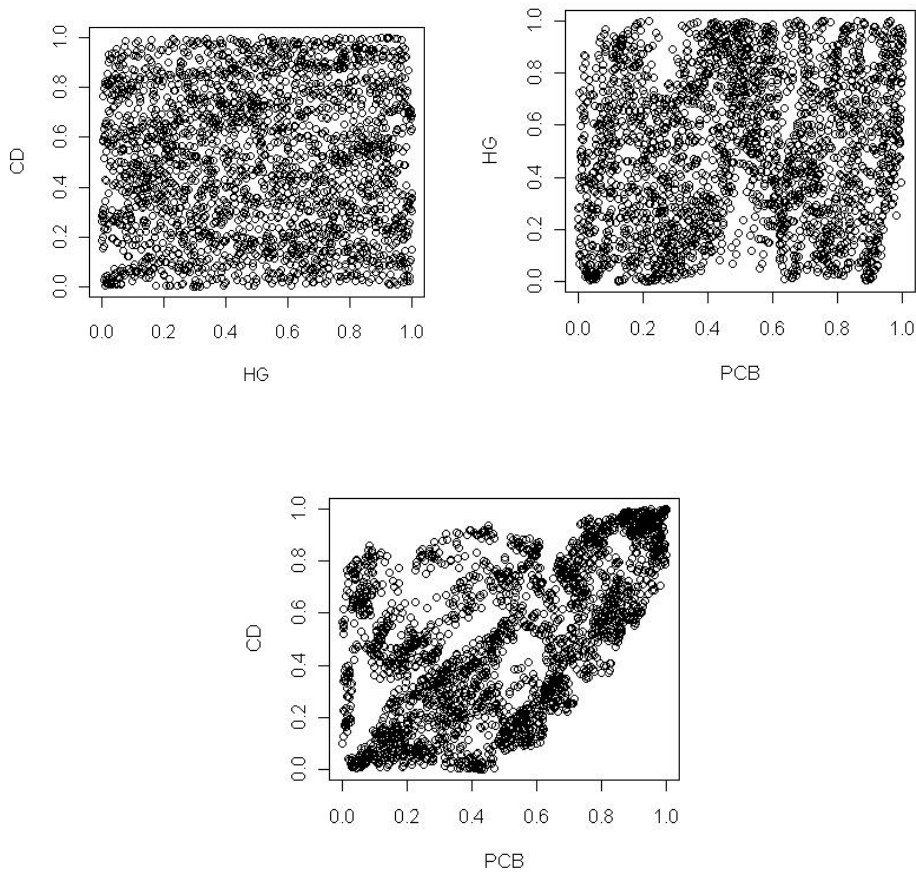
(c) On calcule ensuite  $P_j = \frac{|I_{1,j}|}{N}$  et on passe au niveau suivant  $u_{j+1}$ .

3. On calcul alors  $P = P_1 * P_2 * P_3 * \dots * P_m * P_{m+1}$  avec  $P_j = \frac{|I_{1,m+1}|}{N}$  correspondant à la proportion de particules ayant atteint le niveau  $u$  parmi celles qui avaient atteint le niveau précédent  $u_m$ .

Avec cette méthode, nous avons fixé  $N=10$ . Partant de 0, nous avons atteint le TBB au bout de 35 simulations, ce qui nous donne une probabilité presque nulle de l'ordre de  $\left(\frac{1}{10}\right)^{35}$ .

## 2) Cas de plusieurs risques

L'étude du dependogramme montre qu'il existe une dépendance entre les trajectoires de charges corporelles du Cd et celles des Pbc, mais, on ne note aucune dépendance avec le MeHg. On ne voit en effet aucune particularité sur les deux premières images de la figure II.5. Par contre sur la troisième, on note une concentration dans la cadran supérieur droit, ainsi que dans le cadran inférieur gauche.



**Figure II.5 : Dependogramme des charges corporelles**

Nous regardons la probabilité que la quantité du contaminant 1 dans le corps dépasse un seuil  $u_1$  au cours d'une semaine ou d'une année, sachant que le contaminant 2 a dépassé un seuil  $u_2$  au cours de la même semaine ou de la même année. Concrètement, nous calculons les probabilités :

$$P(X_{ij} > TBB_j \mid X_{ij'} > TBB_{j'})$$

et

$$P\left(\sum_{i=1}^{52} X_{ij} > u_j \mid \sum_{i=1}^{52} X_{ij+1} > u_{j+1}\right)$$

Notons  $E_i = \begin{pmatrix} E_{i1} \\ E_{i2} \\ E_{i3} \end{pmatrix}$  et  $X_i = \begin{pmatrix} X_{i1} \\ X_{i2} \\ X_{i3} \end{pmatrix}$  les vecteur des trajectoires des expositions aux

contaminants et de la charge corporelle correspondant au cours de la  $i^{\text{ème}}$  semaine. Alors nous pouvons écrire :

$$(X_i)' = (X_{i-1})' \begin{pmatrix} e^{-\theta_1} & 0 & 0 \\ 0 & e^{-\theta_2} & 0 \\ 0 & 0 & e^{-\theta_3} \end{pmatrix} + (E_i)'$$

$$\Rightarrow \begin{cases} X_{i1} = X_{i-1,1} \exp(-\theta_1) + E_{i1} \\ X_{i2} = X_{i-1,2} \exp(-\theta_2) + E_{i2} \\ X_{i3} = X_{i-1,3} \exp(-\theta_3) + E_{i3} \end{cases}$$

La matrice diagonale s'explique par l'hypothèse d'indépendance entre l'élimination des différents contaminants. La charge corporelle de chaque contaminant au cours d'une semaine dépend donc de la charge corporelle en ce contaminant la semaine précédente et de la nouvelle exposition.

Les probabilités estimées sont presque nulles, pour tous les couples de contaminants. La probabilité pour un couple de contaminants est la probabilité que chacun des deux contaminants soit au-dessus du seuil en même moment. A partir de cette probabilité, la probabilité conditionnelle est obtenue en divisant la probabilité pour un couple de contaminant d'être au-dessus du seuil par les probabilités individuelles, selon la condition :

$$P(X_{ij} > TBB_j | X_{ij'} > TBB_{j'}) = \frac{P(X_{ij} > TBB_j, X_{ij'} > TBB_{j'})}{P(X_{ij'} > TBB_{j'})}$$

L'utilisation de la méthode de multilevel splitting nous donne aussi des probabilités très faibles, voire nulles. Ce résultat n'est pas surprenant car aucun contaminant n'atteint le seuil.

Nous avons, dans ce qui précède, considéré une matrice diagonale  $A = \begin{pmatrix} e^{-\theta_1} & 0 & 0 \\ 0 & e^{-\theta_2} & 0 \\ 0 & 0 & e^{-\theta_3} \end{pmatrix}$

négligeant ainsi l'impact de l'élimination d'un contaminant sur l'autre. Partant du fait que le Pcb est stocké et donc éliminé par les graisses et le foie et le rein, le Cd est éliminé par le foie et les reins et le MeHg est éliminé par le sang et les graisses, et qu'il n'y a de ce fait aucun organe en commun entre le Cd et le MeHg, cette hypothèse d'indépendance entre les contaminants est remise en cause et nous reconsidérons la matrice A comme suit :

$$A = \begin{pmatrix} e^{-\theta_1} & a & b \\ c & e^{-\theta_2} & 0 \\ d & 0 & e^{-\theta_3} \end{pmatrix}$$

Dans cette nouvelle matrice, les 0 signifient qu'il n'y a aucun organe en commun entre le MeHg et le Cd. Le but du jeu est de déterminer les coefficients de la matrice et il n'y a à priori aucune raison que la matrice soit symétrique. En fait, nous postulons que lorsque deux contaminants sont dans le même organe, si l'organe commence à éliminer le premier contaminant, l'impact sur le second n'est pas forcément le même que si l'organe avait commencé par le second contaminant. Nous pouvons donc écrire :

$$X'_i = X'_{i-1} \begin{pmatrix} e^{-\theta_1} & a & b \\ c & e^{-\theta_2} & 0 \\ d & 0 & e^{-\theta_3} \end{pmatrix} + E_i$$

$$\Rightarrow \begin{cases} X_{i1} = X_{i-1,1} * \exp(-\theta_1) + c * X_{i-1,2} + d * X_{i-1,3} + E_{i1} \\ X_{i2} = a * X_{i-1,1} + X_{i-1,2} * \exp(-\theta_2) + 0 * X_{i-1,3} + E_{i2} \\ X_{i3} = b * X_{i-1,1} + 0 * X_{i-1,2} + X_{i-1,3} * \exp(-\theta_3) + E_{i3} \end{cases}$$

Avec cette nouvelle matrice, la charge corporelle d'un contaminant au cours de la semaine  $i$  dépend de la charge corporelle de ce contaminant et des autres au cours de la semaine précédente. Cette relation suppose en fait que le rythme d'élimination d'un contaminant est ralenti quand l'organe d'élimination contient d'autres contaminants, ce qui rend la charge corporelle en ce contaminant plus élevée que dans le cas de la matrice diagonale.

Les phénomènes de dépendance entre les coefficients d'élimination des contaminants sont assez mal connus des toxicologues, qui ne peuvent donc pas de donner des estimateurs.

Pour déterminer la matrice  $A$ , nous faisons comme hypothèse que pour que les charges corporelles puissent dépasser un seuil donné, elles doivent d'abord l'atteindre. Nous fixons ici que la charge corporelle de long terme de chaque contaminant est égale au seuil de ce contaminant, c'est-à-dire  $TBB_j = E(X_{i,j}) = E(X_{i-1,j})$ . Avec cette valeur initiale, nous pouvons obtenir les coefficients  $a$  et  $b$  de la matrice. Pour les deux coefficients restants, on a une équation à deux inconnues. De manière aléatoire, nous avons affecté des valeurs à  $c$  et  $d$ .

Avec cette nouvelle hypothèse, nous avons dans un premier temps gardé les niveaux de seuils  $TBB_j$  obtenus avec la matrice diagonale. Avec le seuil  $TBB$  hebdomadaire des trois contaminants comme point de départ dans notre simulation, nous calculons la probabilité de dépasser les seuils et d'y rester pendant un an pour les trois contaminants. Ces probabilités sont données dans le tableau suivant, ainsi que les probabilités pour chaque couple de contaminants :

Contaminants	Pcb	Cd	MeHg
Pcb	0,999	0,995	0,224
Cd	0,995	0,997	0,214
MeHg	0,224	0,214	0,263

**Tableau II.2: Probabilités avec seuil(j)= $TBB_j$  et A diagonale**

Nous avons aussi calculé la probabilité de dépasser le seuil des trois contaminants en même temps, qui est de 0.203.

Les probabilités ci-dessus sont calculées pour un individu qui a déjà atteint le seuil pour ces contaminants et nous avons regardé la probabilité de dépasser ce seuil. Il convient donc, dans un second temps, de déterminer les probabilités pour atteindre ce seuil et là, nous obtenons des probabilités relativement faibles. Ainsi, la probabilité d'atteindre le seuil du Pcb est de 0.05%, tandis que la probabilité d'atteindre le seuil du cadmium est de 0.5% et celle du MeHg est de 60%.

La probabilité très élevée pour un individu d'être contaminé par le MeHg nous amène à penser que les coefficients de la matrice ne sont pas bien calibrés et nécessitent sans doute des études expérimentales plus poussées.

Finalement, une autre méthode plus réaliste est proposée pour le calcul de la  $TBB$  en gardant la matrice d'élimination A diagonale comme au début de l'étude. A partir de la série des charges corporelles hebdomadaires, nous utilisons le Hill pour déterminer le seuil extrême pour chacun des contaminants. Le graphique Hill-plot, permet d'avoir des estimations du paramètre  $\alpha$  en fonction de l'ordre statistique le plus élevé (nombre des excès), nous choisissons ainsi l'indice le plus stable pour chaque série de charges corporelles.



Figure II.6 : Hill plot des Pcb, Cd et MeHg



La représentation Figure II.6 nous permet d'estimer le quantile extrême (que nous notons  $q$ ) en se servant de l'indice de queue de la distribution retrouvé par la même méthode.

Le Hill-plot est donc un outil à double utilité :

- L'estimation de l'indice de la queue de la distribution,
- L'estimation du seuil.

De Haan (1994) a montré que l'estimateur de Hill de l'indice de queue converge en probabilité vers sa vraie valeur et cela pour tout  $\alpha$  réel.

L'estimateur de Hill est plus fiable dans le domaine d'attraction de Fréchet pour lequel il fournit un estimateur de l'indice de queue plus efficace que d'autres méthodes.

Avec cette méthode, nous avons donc estimé les nouveaux seuils (TBB) et recalculé les probabilités. Le tableau ci-dessous nous donne pour chaque contaminant, le nombre d'excès au-delà du seuil, l'indice de queue et la valeur du seuil correspondant :

Contaminants	Pcb	Cd	MeHg
Nombre d'excès	42	24	34
Paramètre de queue	0,02 pg	2,5 $\mu$ g	0,20 $\mu$ g
Q	1550pg	788.9 $\mu$ g	2,43 $\mu$ g
Quantile 95%	1563pg	787,7 $\mu$ g	2,41 $\mu$ g

**Tableau II 4 : Excédances, seuils et quantiles des différents contaminants**

Les seuils obtenus par l'estimateur de Hill sont proches du quantile à 95%.

La matrice est diagonale mais on a baissé la TBB de moitié. Ici on calcule le quantile moyen de l'ordre 95% de la charge corporelle sur 1000 simulations de vie.

On suppose que le point de départ est la valeur moyenne des charges corporelles. Les probabilités obtenues pour un dépassement de seuil sur une année pour les Pcb est de 52.1%. Pour le Cd, cette probabilité est de 52.6% et elle est presque nulle pour le MeHg.

<b>Contaminants</b>	<b>Pcb</b>	<b>Cd</b>	<b>MeHg</b>
Pcb	52,1	23,5	0
Cd	23,5	52,6	0
MeHg	0	0	0

**Tableau II 5: Probabilités en % avec seuil(j)= $q_{0,95}$  et A diagonale**

Nous pouvons conclure en disant que les probabilités de dépasser les seuils tolérables des contaminants de notre étude restent assez faibles, voire nulles. Il convient toutefois de préciser que nous avons calculé ces probabilités en considérant la charge corporelle résultant de l'accumulation des contaminants par voie alimentaire. Or d'autres voies de contaminations existent comme la peau, la poussière et la cigarette.

## Chapitre III Application au risque assurantiel

### I- Contexte

En avril 2009, deux ans après les premiers travaux de la commission européenne, le Parlement européen, dans la lignée du monde bancaire et des accords de Bâle II, a voté une réforme réglementaire du monde de l'assurance. Son objectif est d'obtenir une meilleure adéquation entre les fonds propres détenus par les compagnies d'assurances et les risques liés à leur activité. Cette directive, connue sous l'appellation Solvabilité II, entrera en vigueur d'ici 2016, et les travaux préparatoires constituent aujourd'hui un enjeu majeur pour les compagnies d'assurances européennes. Cette nouvelle réglementation impose le calcul d'indices de risques tels que la value-at-risk (VaR). La VaR à un an au seuil de 99.5% fournit un indice de ce qu'une compagnie d'assurance doit devoir geler pour couvrir le risque à horizon 1 an. Encore appelé SCR (Solvability Capital Requirement), il représente le capital cible nécessaire pour absorber le choc provoqué par un risque majeur (taux de mortalité, catastrophe naturelle, ...). L'activité d'assurance requiert donc 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 côté, les actionnaires espèrent maximiser le rendement de leur investissement en limitant le capital injecté dans la société.

L'assureur utilise l'épargne collectée auprès de sa clientèle pour constituer un portefeuille d'actifs destiné à lui permettre de verser des flux associés à son portefeuille d'épargne. Dans le cadre de Solvabilité II, il doit également associer à ce portefeuille un montant de fonds propres lui permettant d'assurer la pérennité de son activité à horizon d'un

an dans 99,5% des cas, ce qui représente une probabilité de ruine de 0.5%. Or, les activités traditionnelles des compagnies d'assurance ne permettent plus, à cause de la complexité croissante des marchés financiers, de faire une meilleure allocation actif/passif. C'est ainsi que les stratégies de Hedge Fund ont vu le jour, mais la réglementation rend difficile leur intégration dans la répartition de l'actif des sociétés d'assurance. En effet, la réglementation impose un montant élevé de capital à mobiliser (49%), alors que Mathieu Vaissié (2011), montre, en utilisant la méthode basée sur les avoirs (holdings-based analysis) et la méthode basée sur les rendements (returns-based analysis) que ce montant ne reflète pas les risques inhérents à ces stratégies car des informations existent désormais sur ces stratégies, ce qui permet de bien analyser les risques sous-jacents et donc de ne plus les classer dans la catégorie « other equities ».

C'est dans cette optique que nous voulons estimer les SCR sur les différentes stratégies de Hedge Funds du groupe Lyxor. Il est important de souligner à ce niveau que les sociétés de gestion telles que Lyxor ne sont pas soumises réglementairement au calcul de SCR. Seuls les assureurs (pour l'instant) doivent montrer qu'ils prennent bien en compte les problèmes de solvabilité. Le coût de chaque investissement dans un actif risqué se mesure pour l'assureur en SCR. Et il doit montrer qu'il est capable de faire ce calcul de manière efficace. Quand une société de gestion fait appel à un investisseur institutionnel (assurance, caisse de retraite, fond de pension, ...), la société de gestion vend moins un produit qu'un service. Elle doit convaincre que le produit proposé répond aux attentes de l'investisseur, et qu'en particulier:

- a) il complète efficacement les investissements à l'actif du client, c'est-à-dire le portefeuille de produit financier déjà détenu,
- b) il permet de mieux couvrir les engagements au passif du client.

Parmi les «services» annexes que propose la société de gestion, il y a, notamment pour les produits non standards, la méthode de calcul du SCR. En résumé, quand une société de gestion a des produits complexes à vendre, elle doit accompagner cette vente de tout un ensemble de services dont le calcul du SCR fait partie. Dans le cadre réglementaire, deux options sont possibles :

- la formule standard proposée par le régulateur de manière à convenir à l'ensemble du marché assurantiel
- le modèle interne développé par les compagnies d'assurance souhaitant calculer un capital de solvabilité requis le plus proche possible de leur profil de risque.

C'est cette seconde approche que nous allons utiliser pour estimer le niveau de SCR demandé par le régulateur.

Nous allons commencer par présenter les trois piliers de la réglementation Solvabilité II. Nous mettrons en avant le pilier I qui contient un des éléments clés de cette partie, puisque celui-ci introduit les principaux éléments quantitatifs pour évaluer la solvabilité d'une compagnie d'assurance.

## II- Solvabilité II

### A- Pilier 1 : les exigences quantitatives en capital

Le pilier 1 concerne les exigences quantitatives (fonds propres et provisions techniques). A cet effet, il instaure de nouvelles règles d'évaluation des éléments du bilan prudentiel des compagnies d'assurance :

	<b>Solvabilité I</b>	<b>Solvabilité II</b>
Evaluation actifs et passifs couvrables	Valeur comptable	Valeur de marché ( <i>fair value</i> )
Evaluation passifs non couvrables	Valeur comptable	<i>Best Estimate</i> et marge de risque
Exigence de marge	Pourcentage des primes, des sinistres ou des provisions	Quantile à 0,5 % du résultat de la société

**Tableau III. 1 : Méthode de valorisation du bilan prudentiel et de l'exigence de marge**

On voit dans le Tableau III.1 que le nouveau mode de comptabilisation marque une rupture par rapport à la réglementation antérieure, notamment en imposant une valorisation en Fair Value (valeur de marché) pour les actifs et en Best Estimate pour les engagements. Le pilier 1 fixe également les normes quantitatives de calcul des provisions techniques et des fonds propres. Il précise aussi comment l'assureur doit se prémunir contre les pertes : les pertes attendues (Expected Loss) doivent être intégrées dans les provisions mathématiques, tandis que les pertes exceptionnelles (unexpected loss) doivent être provisionnées par des réserves de capital appelées fonds propres devant lui permettre de continuer son activité.

#### a) Calcul des provisions économiques

On distingue deux types de risques: les risques couvrables et les risques non couvrables. Les provisions économiques pour risques couvrables correspondent au prix de la

couverture financière permettant de construire une stratégie d'investissement autofinancée dans laquelle les instruments financiers répliquent parfaitement les flux futurs dans tous les états du monde.

Dans le cas de provisions techniques non couvrables, le calcul se fait en deux étapes :

- La Meilleure Estimation (ou *Best Estimate*), qui correspond à la moyenne pondérée par leur probabilité des flux de trésorerie actualisés en utilisant une courbe de taux pertinente. Il s'agit donc de la valeur actuelle des flux futurs de trésorerie.
- La Marge de Risque qui est une marge de prudence dans l'estimation des provisions économiques. Elle est égale au montant que l'on devrait ajouter au *Best Estimate* afin qu'un autre assureur accepte de racheter les engagements d'assurance. Elle est calculée comme un coût du capital.

#### *b) Calcul du Minimum de Capital Requis et du Capital de Solvabilité Requis*

Solvabilité II introduit également deux mesures de garantie de la solvabilité :

- Le Minimum de Capital Requis (MCR) correspond au niveau minimum de fonds propres dont doit disposer une compagnie d'assurance. En dessous de ce niveau, l'intervention de l'autorité de contrôle - en France, l'Autorité de Contrôle Prudentiel - est automatique (retrait de l'agrément, plan de redressement, etc.). Il est calculé comme un pourcentage des provisions et des primes, et doit être compris entre 25 % et 45 % du Capital de Solvabilité Requis (d'après le QIS 5<sup>11</sup>).
- Le Capital de Solvabilité Requis (SCR) est défini comme le montant cible de fonds propres économiques dont doit disposer la compagnie d'assurance pour que la probabilité de ruine soit inférieure à 0,5 % à horizon un an. Autrement dit, il s'agit des fonds propres nécessaires à la compagnie pour qu'elle soit solvable dans un an, dans 99,5 % des cas.

Le calcul du SCR dépend des types de risques auxquels sont exposées les compagnies d'assurance: risque de souscription vie et non-vie, risque de souscription santé, risque de marché, risque de crédit et risque opérationnel. Il est basé sur une formule standard fournie par les autorités de régulation, ou sur un modèle interne que les compagnies peuvent mettre en place et qui devra être validé par les autorités de contrôle.

---

<sup>11</sup> Quantity Impact Study 5

Dans tous les cas, le niveau de confiance retenu de 99,5 % pour le SCR implique la nécessité d'estimer correctement un quantile d'ordre élevé, quantile que nous déterminerons dans ce mémoire en utilisant les outils que fournit la théorie des valeurs extrêmes et au besoin par la méthode de Monte Carlo.

Selon les spécifications du QIS 5, la formule standard définit le SCR de la manière suivante :

$$SCR = BSCR + Adj + Op,$$

BSCR (*Basic Solvency Capital Requirement*) est le SCR de base, Adj sont les ajustements liés à l'effet d'absorption des risques des futures participations aux bénéficiaires et des impôts différés et Op est l'exigence de capital au titre du risque opérationnel, déterminée de manière forfaitaire en fonction des provisions mathématiques et des primes.

Le BSCR est déterminé en utilisant une méthode modulaire dite *Bottom-Up* dans laquelle le calcul se divise en modules de risque qui se décomposent eux-mêmes en sous-modules (comme le montre la formule ci-dessus) : dans un premier temps, on calcule à l'intérieur de chaque module, le capital requis au titre des différents risques élémentaires puis ces capitaux sont agrégés à l'aide de matrices de corrélations et on obtient un SCR relatif à chaque module. Enfin, on agrège le tout avec une autre matrice de corrélation afin d'obtenir le BSCR.

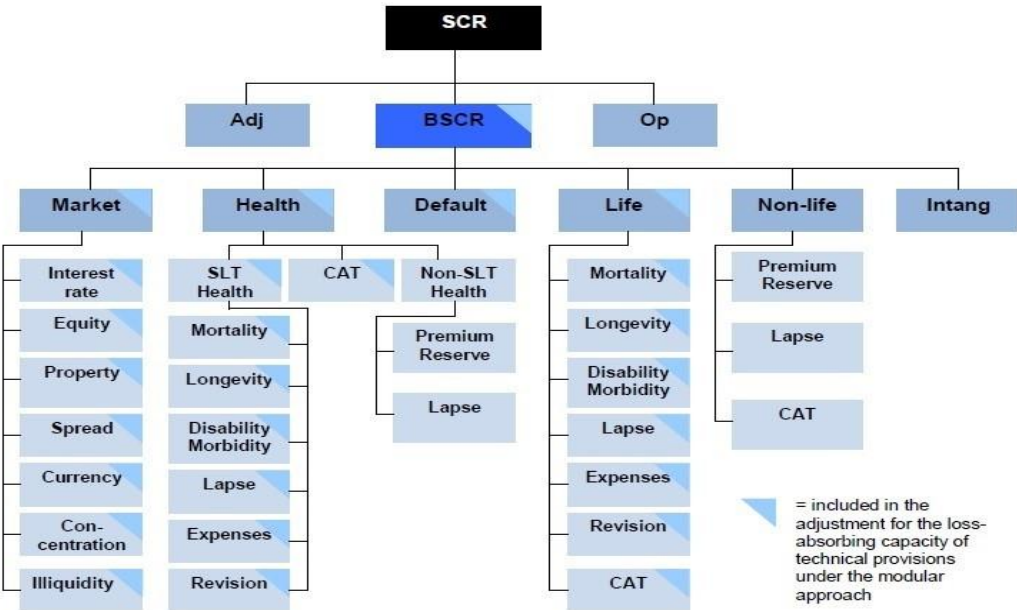


Tableau III.2 : Les modules de SCR d'après le QIS 5 - Formule Standard

Notons:

- *Op*: exigence de capital au titre du risque opérationnel, déterminée de manière forfaitaire en fonction de différents éléments tels que le BSCR, les provisions techniques brutes de réassurance pour les engagements vie et non-vie et le montant des primes acquises.
- *SCR<sub>Mkt</sub>* : exigence de capital au titre du risque de marché.
- *SCR<sub>Health</sub>* : exigence de capital au titre du risque souscription santé.
- *SCR<sub>Def</sub>* : exigence de capital au titre du risque de crédit.
- *SCR<sub>Life</sub>* : exigence de capital au titre du risque souscription vie.
- *SCR<sub>NL</sub>* : exigence de capital au titre du risque souscription non-vie.
- *SCR<sub>Intang</sub>* : exigence de capital au titre du risque des actifs incorporels.

Alors la corrélation entre les différents modules du SCR est prise en compte *via* la matrice de corrélation suivante fournie par le QIS 5 :

i \ j	Market	Default	Life	Health	Non-life
Market	1				
Default	0.25	1			
Life	0.25	0.25	1		
Health	0.25	0.25	0.25	1	
Non-life	0.25	0.5	0	0	1

**Tableau III. 3 : Matrice de corrélation des différents modules pour le calcul du BSCR - QIS 5**

$$BSCR = \sqrt{\sum_{ij} corr_{ij} * SCR_i * SCR_j} + SCR_{Intang}$$

Il faut aussi prendre en compte des facteurs permettant de limiter le risque, comme la réassurance (mais qui en contrepartie ajoute un risque de défaut), ou les instruments financiers qui doivent respecter un certain nombre de principes (répertoriés par la Commission Européenne) pour pouvoir être pris en compte dans le calcul du SCR.



### *c) Les différents types de risques*

#### **1) *Le risque de souscription***

Il s'agit de l'ensemble des risques pris par l'assureur lors de la distribution de contrats d'assurance. En assurance non-vie, on distingue le risque de prime ou risque de sous-tarification, le risque de réserve lié à la nature aléatoire de l'évaluation des sinistres et à l'erreur d'estimation, et le risque catastrophe qui résulte des événements extrêmes ou irréguliers. En assurance vie, ce sont les risques de longévité, de mortalité, de rachat...

#### **2) *Le risque opérationnel***

Ce risque est défini par l'EIOPA<sup>12</sup> comme le « risque de pertes provenant de processus internes inadéquats ou défaillants, de personnes et systèmes ou d'événements externes ». Il comprend donc les fraudes (interne ou externe), le dysfonctionnement de l'activité et des systèmes, les dommages aux actifs corporels, l'exécution et la gestion des processus et les pratiques en matière d'emploi et sécurité sur le lieu de travail.

#### **3) *Le risque de crédit***

Il s'agit du risque de pertes liées au défaut ou à la dégradation de la qualité financière d'une contrepartie. Dans le cadre d'une compagnie d'assurance, il peut s'agir d'un réassureur ou d'un courtier.

#### **4) *Le risque de marché***

Il s'agit du risque de pertes lors de mouvements adverses du marché. Dans les risques « quantifiables » identifiés par l'EIOPA, on distingue les risques liés à la baisse du marché des actions ou du marché immobilier, à la hausse ou baisse des taux, à une augmentation des spreads, aux taux de change ou aux concentrations. Le choix d'évaluer la solvabilité d'une entreprise dans Solvabilité II à partir de la valeur économique de ses fonds propres rend la solvabilité apparente de l'entreprise encore plus exposée au risque de marché. Par la suite, nous nous intéresserons dans cette partie assurantielle au module correspondant au risque de

---

<sup>12</sup> Autorité européenne des assurances et des pensions professionnelles

marché – surtout sur le risque des actions, des stratégies de Hedge Funds dont nous allons décrire les différents sous-modules dans le paragraphe suivant.

*d) Exigence en capital lié au risque de marché : le SCRMkt*

Le *SCRMkt* s’obtient en prenant en compte les chargements ci-dessous (toujours d’après le QIS 5) :

- *Mktint,up* : charge en capital relative au choc de hausse des taux.
- *Mktint,down* : charge en capital relative au choc de baisse des taux.
- *Mkteq* : charge en capital relative au choc sur les actions.
- *Mktprop* : charge en capital relative au choc sur les actifs immobiliers.
- *Mktup* : charge en capital relative au choc de spread.
- *Mktvonc* : charge en capital relative au choc de concentration.
- *Mktfx* : charge en capital relative au choc sur les taux de change.
- *Mktil* : charge en capital relative au choc sur la prime d’illiquidité.

Il se calcule en agrégeant les sous-risques avec la formule :

$$\max \left( \sqrt{\sum_{r,c} CorrMktUp_{r,c} * Mkt_r^{up} * Mkt_c^{up}}, \sqrt{\sum_{r,c} CorrMktDown_{r,c} * Mkt_r^{down} * Mkt_c^{down}} \right)$$

Où les corrélations *CorrMktUp* (corrélation pour les scénarios de hausse des niveaux de taux) et *CorrMktDown* (corrélation pour les scénarios de baisse des niveaux de taux) sont définies par :

CorrMktUp	Interest	Equity	Property	Spread	Currency	Concentration	Illiquidity premium
Interest	1						
Equity	0	1					
Property	0	0.75	1				
Spread	0	0.75	0.5	1			
Currency	0.25	0.25	0.25	0.25	1		
Concentration	0	0	0	0	0	1	
Illiquidity premium	0	0	0	-0.5	0	0	1

**Tableau III.4 : Matrice de corrélation pour le module risque de marché en cas de hausse des taux**

CorrMktDown	Interest	Equity	Property	Spread	Currency	Concentration	Illiquidity premium
Interest	1						
Equity	0.5	1					
Property	0.5	0.75	1				
Spread	0.5	0.75	0.5	1			
Currency	0.25	0.25	0.25	0.25	1		
Concentration	0	0	0	0	0	1	
Illiquidity premium	0	0	0	-0.5	0	0	1

**Tableau III.5 : Matrice de corrélation pour le module risque de marché en cas de baisse des taux**

Pour chaque facteur de risque, la charge en capital est calculée en mesurant l'impact d'un choc associé au facteur de risque sur la valeur nette de l'actif (NAV : *Net Asset Value*) qui correspond à la différence entre la valeur de l'actif et du passif.

Nous travaillons dans ce mémoire sur les données des rendements de plusieurs stratégies de Hedge Funds. Sans perte de la généralité, nous pouvons considérer chaque stratégie de Hedging comme une action sur le marché.

Le risque action est le risque induit par les variations du niveau des actions, ainsi que par celles de leur volatilité. On distingue deux types d'actions :

- les actions correspondant à l'indice dit « global », qui sont les actions cotées dans les pays de l'EEE et de l'OCDE.
- les actions « autres » qui regroupent les actions cotées dans les pays émergents, les actions non cotées et les investissements alternatifs.

Dans le QIS 5, le risque de concentration est lié à l'accumulation d'expositions à une même contrepartie. Il est constitué d'une part de la volatilité supplémentaire incluse dans le portefeuille d'actifs concentrés, et d'autre part de l'accumulation des risques de défaut (les obligations souveraines ne sont pas comprises dans ce module). Les chocs appliqués dépendent du rating du titre et du niveau de concentration. Les actifs considérés pour ce risque sont les produits de taux et les actions ainsi que les produits dérivés de ces actifs et les actifs immobiliers. Nous nous intéressons ici au problème de l'optimisation d'un portefeuille de plusieurs stratégies pour réduire ce risque de concentrations.

En dehors des deux risques cités ci-dessus, nous pouvons citer les principaux risques qui sont directement liés dans le calcul du SCR de marché comme : le risque de taux, le risque immobilier, le risque de change, le risque de liquidité et le risque de spread.

## **B- Pilier 2 : gestion des risques et principes de gouvernance**

Le pilier 2 complète les exigences quantitatives du pilier 1 par des exigences qualitatives. Il traite à la fois du contrôle par les autorités et du contrôle interne des compagnies et vise à s'assurer que :

- Les compagnies sont correctement gérées (gestion saine, prudente et efficace de l'activité), qu'elles sont capable d'évaluer, calculer et maîtriser leurs risques avec une quantité suffisante de capital à tout moment.
- Elles possèdent le niveau de capital requis pour faire face à leurs risques.

Le régulateur aura le pouvoir de contrôler la qualité des données et des procédures d'estimation et les systèmes mis en place pour mesurer et maîtriser les risques au cas où ils se matérialiseraient. Dans le cadre de la mise en œuvre du processus de contrôle, les autorités de supervision pourront imposer une exigence supplémentaire de capital, si le profil de risque de l'entreprise s'écarte significativement des hypothèses utilisées dans le calcul du SCR (formule standard ou modèle interne) ou si le système de gouvernance s'écarte significativement des standards établis.

## **C- Pilier 3 : discipline de marché**

Le troisième pilier fixe les règles de transparence financière et de la communication publique des compagnies d'assurances. Le but de ce pilier est de renforcer la discipline de marché et d'harmoniser les informations diffusées dans les Etats membres à différents niveaux (assurés, marché ou institutionnels, autorités de contrôle).

### **III- Estimation du risque**

#### **A- La formule standard**

La formule standard préconise de capitaliser en fonds propre 49% de la somme à investir dans n'importe quel hedge Fund (order equities). La raison est simple : les Hedge Fund sont classés dans la catégorie « autres actions », à cause du fait qu'il y a peu

d'informations sur ces stratégies ce qui ne permet pas (selon les régulateurs) de bien analyser les risques sous-jacents. Elle permet une approche analytique basée sur les données historiques de l'entreprise afin de déterminer de manière cohérente les risques les plus consommateurs en capital pour la société. Le capital unitaire nécessité par chaque risque est déterminé comme écart entre deux NAV (Net Asset Value) centrale et choquée. La valorisation de ces NAV nécessite généralement de disposer d'un modèle de gestion actif-passif (par exemple le modèle ALM : « Asset-Liability management ») simple.

## B- Le modèle interne

Le modèle interne est une formule de calcul réaliste du SCR que l'assureur doit donner au régulateur. Les structures préfèrent développer leur propre modèle interne quoique très coûteux, car cela leur permet de mieux prendre en compte leur profil de risque et de ce fait réduire le montant de SCR qu'elles auraient provisionné en utilisant la formule standard.

Soit  $P_t$  le montant du portefeuille d'une stratégie de hedge Fund à la date  $t$ . A  $t = 0$ , la probabilité pour que le rendement annualisé soit inférieur à 49% est de 0.5% et s'écrit  $P(P_{0\text{à}1\text{an}} - P_0 \leq 0.49P_0) = 0.5\%$ .

Avec les données hebdomadaires, on fait les approximations suivantes :

- Le rendement annuel

$$\frac{P_{0\text{à}1\text{an}} - P_0}{P_0} \sim \log \frac{P_{0\text{à}1\text{an}}}{P_0} = \sum_{t=0}^{51} \log \left( \frac{P_{t+1}}{P_t} \right) = \sum_{t=0}^{51} r_t$$

où  $(r_t)$  sont les log-rendements hebdomadaires.

- Les log rendements sont stationnaires

$$(r_{t_j+k})_{1 \leq j \leq k} \stackrel{\mathcal{L}}{=} (r_{t_j})_{1 \leq j \leq k} \quad \forall t_1 \leq t_2 \leq \dots \leq t_k$$

L'assureur utilise les rendements annualisés selon la relation :

$$RA = 52 * \bar{r}_T = 52 * \frac{1}{T} \sum_{t=1}^T r_t$$

La théorie moderne du portefeuille classique (Markovitz, 1952) fournit une approche pour calculer le SCR. L'idée de départ est de quantifier le risque de chaque stratégie grâce à l'écart-type des rendements

$$s_d = \sqrt{\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r}_T)^2}.$$

En supposant les log rendements non auto corrélés, la covariance  $cov(r_t, r_0) = 0 \forall t \geq 0$ , on peut avoir une approximation de l'écart-type annuel

$$\sqrt{\text{Var}(\sum_{t=1}^T r_t)} = \sqrt{52 * \text{Var}(r_1)} = \sqrt{52} * s_d$$

Cette approche repose sur des variances pour contrôler le risque (queue de distribution). Cela n'est possible que dans le cas gaussien. Supposons que  $(r_t)_{1 \leq t \leq T}$  constitue un vecteur gaussien. On peut écrire

$$\frac{\sum_{t=1}^{52} r_t - \mathbb{E}(\sum_{t=1}^{52} r_t)}{\sqrt{\text{Var}(\sum_{t=1}^{52} r_t)}} \sim \mathcal{N}(0,1)$$

Sous l'hypothèse de normalité, on calcule le SCR par la formule

$$\begin{aligned} \mathbb{P}\left[\frac{P_{T+52} - P_T}{P_T} \leq x\right] &= 99,5\% \quad \Leftrightarrow \quad \mathbb{P}(\sum_{t=1}^{52} r_t \leq x) = 99,5\%. \\ \Rightarrow \Phi\left(\frac{x - \mathbb{E}(\sum_{t=1}^{52} r_t)}{\sqrt{52} * s_d}\right) &= 99,5\%. \end{aligned} \quad (3.1)$$

La critique de cette méthode est que l'hypothèse de normalité n'est pas réaliste et sous-estime les queues de distribution.

C'est pour cette raison qu'on a recours à l'approche de la théorie des valeurs extrêmes fondée sur le paradigme : « let the tails speak by themselves<sup>13</sup> ». Cette approche est efficace quand les données sont indépendantes, hypothèse pas toujours réaliste, à cause du phénomène d'apparition des valeurs extrêmes successives dans le temps du fait de la dépendance et du phénomène dit de clustering (regroupements).

Sous Solvabilité II, le SCR est la VaR de taux de couverture à 99,5% à 1 an. Le temps de retour moyen (temps d'attente moyen entre deux dépassements de seuil) est

<sup>13</sup> Laissons les queues de distributions s'exprimer

$$T(SCR) = \frac{1}{1-99,5\%} = 200 \text{ ans .}$$

En fait,

$$T(SCR) = \mathbb{E}[T_1(SCR)] = \frac{1}{p} = \frac{1}{1 - F(SCR)}$$

où  $p$  est la probabilité faible de dépasser le SCR(le seuil).

Une méthode empirique consiste à calculer  $\widehat{VaR} = F_n^{-1}(\alpha) = X_{([n\alpha])}$ .

$$F_n(SCR) = \frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq SCR\}}$$

où  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  et  $[n\alpha]$  est la partie entière de  $(n\alpha)$

La méthode empirique est inefficace car le temps de retour moyen est très long par rapport aux périodes d'observations. Il est donc nécessaire d'extrapoler, c'est-à-dire de déterminer un niveau pas encore atteint par nos observations. Pour cela, on fait appel à des théorèmes limites asymptotiques.

Soit  $X$  une variable aléatoire de fonction caractéristique :

$$\rho_X(t) = \mathbb{E}(e^{iXt}) \in \mathbb{C}.$$

On a l'équivalence suivante :

$$(X_n) \text{ converge en loi vers } X \iff X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X \iff \rho_{X_n}(t) \rightarrow \rho_X(t), \forall t \in \mathbb{R}.$$

D'après le théorème central limite :  $S_n = \sum_{i=1}^n X_i$  avec  $X_i$  (iid), il existe  $(a_n)$ ,  $(b_n) > 0$  tels que

$$b_n^{-1}(S_n - a_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0,v) \iff \mathbb{E}(X^2) < \infty, \text{ avec } v > 0$$

Les moments d'ordre 2 étant finis, on peut choisir

$$\begin{cases} v = \text{Var}(X) \\ a_n = n\mathbb{E}(X) \\ b_n = \sqrt{n} \end{cases}$$

On dit alors que le domaine d'attraction de la loi normale est l'ensemble des lois de carrés intégrables. L'utilisation de l'asymptotique  $n \rightarrow \infty$  permet de justifier l'utilisation de la loi normale dans la théorie moderne du portefeuille.

En pratique, les événements observés sont tellement dispersés que  $Var(X) = +\infty$ . La gaussienne doit donc être remplacée par son équivalent, une loi  $\alpha$ -stable.

Nous allons présenter les résultats de calcul de SCR en utilisant quatre méthodes: la méthode paramétrique (ou gaussienne), la méthode historique, la méthode de McNeil et Frey, et la méthode de Monte Carlo. Nous commençons d'abord par présenter les données.

#### *a) Description des données*

La base de données concerne des stratégies de Hedge Fund de la société Lyxor. Ce sont les rendements hebdomadaires de 18 stratégies de Hedge Fund allant du 16 avril 2002 au 31 décembre 2012; soit 560 points. Mais à cause des données manquantes, nous allons travailler avec 471 données. Le tableau III.6 donne les rendements annuels de ces différentes stratégies :

<b>Stratégies</b>	<b>Rendements</b>
COMPOSITE	3,00
Statistical, Arbitrage	0,73
Convertible, Bonds, Arbitrage	0,29
CTA, Long, Term	4,71
L, S, Credit, Arbitrage	1,32
Event, Driven... Special, Situations	5,16
Event, Driven... Merger, Arbitrage	6,19
L, S, Equity... Variable, Bias	1,80
Emerging, Markets	-3,04
Global, Macro	0,33
L, S, Equity... Market, Neutral	2,21
CTA, Short, Term	5,11
L, S, Equity... Short, Bias	-8,61
L, S, Equity... Long, Bias	4,61
Fixed, Income, Arbitrage	2,62
Credit, Strategies	1,73
Event, Driven... Distressed	0,72
Top, 10	5,19

**Tableau III.6 : Rendements des différentes stratégies de Hedge Fund**

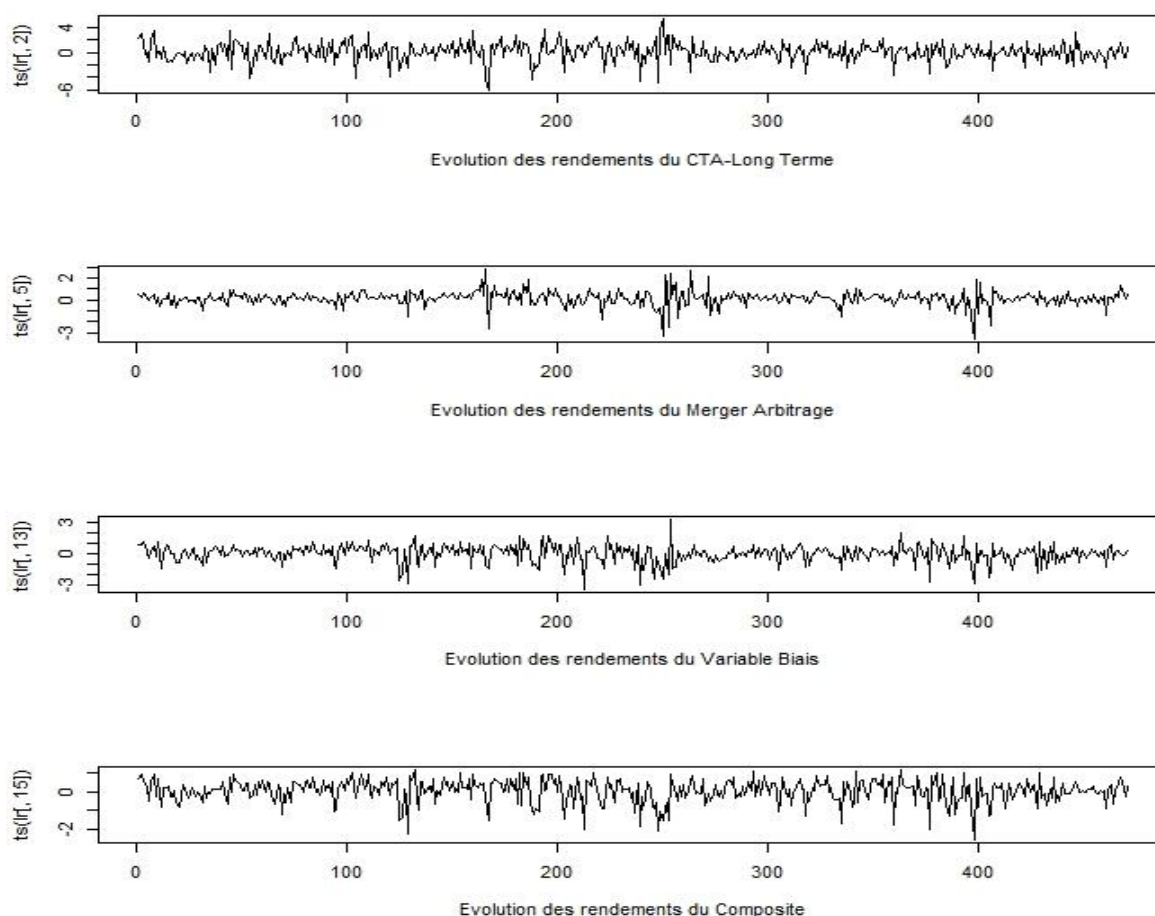


Nous allons nous limiter à quelques stratégies. Les stratégies que nous allons choisir sont celles qui mobilisent peu de capital individuellement, et nous voulons in fine trouver les combinaisons de stratégies permettant de réduire davantage le niveau du SCR tout en gardant un niveau de rentabilité convenable. Il convient de préciser qu'il n'est pas exclu que les stratégies non choisies à cause de leur SCR élevé puissent nécessiter peu de capital en faisant des combinaisons adéquates. Nous retenons pour notre étude les stratégies suivantes : CTA.Long.Term (CTALT), LS.Equity...Variable Bias(VB), Merger arbitrage ou Risk Arbitrage (MA) et Composite(COM). Le choix de ces stratégies a été motivé par le cours de formation par la recherche sur le clustering des volatilités dispensé par M. WINTENBERGER à l'ENSAE au cours de l'année scolaire 2012-2013<sup>14</sup>. Il montre en utilisant la même base de données que ces stratégies sont les plus rentables du groupe Lyxor et qu'elles nécessitent le moins de fonds propres. Nous précisons tout de même que le stratégie Top Ten, malgré son niveau de rentabilité élevé et sa faible consommation en capital n'a pas été retenue car non seulement elle est construite à partir des 10 plus gros fonds de la plateforme, mais aussi, elle n'est plus calculée depuis janvier 2012.

---

<sup>14</sup> <http://wintenberger.fr/ens>.

L'analyse graphique des indices des stratégies retenues montre des clusters de volatilité sur la Figure III-1.



**Figure III.1 : Evolutions hebdomadaire des stratégies**

D'une part, les fortes volatilités sont suivies par des fortes volatilités, et les faibles volatilités sont suivies par des faibles volatilités. D'autre part, les statistiques descriptives dénotent un profil de risque asymétrique du côté des pertes (skewness négatives) et à queue de distributions élevées (kurtosis plus élevés que celui de la loi normale qui est de 3) de la distribution des rendements hebdomadaires).

Stratégies	CTA,Long,Term	Merger,Arbitrage	Variable,Bias	COMPOSITE
Moyenne	0,091	0,119	0,035	0,058
Skewness	-0,401	-0,863	-0,843	-0,860
Kurtosis	4,332	9,727	5,366	4,684

**Tableau III. 7 : Statistiques descriptives des rendements hebdomadaires**

La réalisation du test de Jarque Bera de normalité conduit au rejet de l'hypothèse de normalité des rendements de hedge Fund. Enfin, l'application du test ARCH (Engel, 1982) conclut à la présence d'un effet de type ARCH. Ces caractéristiques représentatives des séries financières suggèrent la modélisation de la volatilité des hedge Fund par les modèles de type GARCH.

Le modèle GARCH, développé par Bollerslev(1986) est une généralisation du modèle ARCH développée par Engle (1982) pour modéliser le taux d'inflation au Royaume Uni.

Il s'écrit de la manière suivante, pour un GARCH (p,q) :

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2.$$

Les conditions du modèle sont que  $\alpha_i \geq 0, \beta_j \geq 0 \forall i$  et  $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$  (Condition de stationnarité de second ordre, c'est-à-dire variance non conditionnelle constante et finie).

Nous allons nous restreindre à l'étude du GARCH(1,1), le plus répandu en finance, et qui se révèle souvent très performant. Il s'écrit :

$$r_t = \sigma_t u_t. \quad \sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

Pour estimer les paramètres du GARCH(1,1), il faut déterminer la loi des innovations. En choisissant comme  $u_t$  la loi normale centrée réduite, nous avons  $F_{t-1}$  l'information en  $t - 1$ .

$$r_t / F_{t-1} \sim \mathcal{N}(0, \sigma_t^2)$$

Les paramètres du modèles GARCH peuvent alors être estimés en maximisant la vraisemblance :

$$L(R, \omega, \alpha, \beta) = \prod_{i=1}^n \rho(r_t / u_t, \sigma_t^2) = \prod_{t=1}^n \frac{1}{\sqrt{2\pi}\sigma_t} \exp \left[ -\frac{1}{2} \left( \frac{r_t - u_t}{\sigma_t} \right)^2 \right]$$

avec  $\rho(\cdot)$  la densité de la loi normale,  $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 = \omega + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2$  et  $R = (r_1, r_2, \dots, r_n)$

En Finance, la véritable loi des innovations  $u_t$  est rarement conditionnellement normale. Ainsi, nous utilisons une loi qui ne correspond pas à la loi utilisée pour calculer la vraisemblance, à savoir le pseudo maximum de vraisemblance (PMV).

Pour filtrer la volatilité à partir du GARCH(1,1), nous initialisons la première valeur de la variance en utilisant la variance de long terme  $\omega_0 = \frac{\omega}{1-\alpha-\beta}$ .

Les modèles GARCH sont très utilisés en finance pour modéliser le prix de différents produits financiers. Mais, pour un portefeuille, par exemple, ce modèle peut présenter des limites car la corrélation entre les titres n'est pas prise en considération. Du coup, les modèles GARCH multivariés satisfont cette condition, c'est-à-dire, ils incluent dans l'expression de la variance conditionnelle ces corrélations. Dans le cadre de cette étude, nous allons nous limiter au GARCH multivarié de type conditional correlation constant (CCC).

Ce modèle a été proposé par Bollerslev en 1990. Nous en donnons un bref aperçu.

Soit

$$r_t = H_t^{\frac{1}{2}} u_t$$

Avec  $r_t$  le vecteur des log-rendements de taille  $N \times 1$  et  $H_t = D_t R_t D_t$  de taille  $N \times N$ .  $u_t$  de taille  $N \times 1$  est la série des innovations qui suit une loi de moyenne 0 et de variance 1. Les  $u_t$  sont iid.

$$\begin{cases} H_t = & D_t R_t D_t \\ D_t = \text{diag} \left( h_{11t}^{\frac{1}{2}}, \dots, h_{NNt}^{\frac{1}{2}} \right) \\ R_t = \rho_{ijt}, \text{ avec } \rho_{iit} = 1 \end{cases}$$

où  $R_t$  est une matrice de dimension  $(N \times N)$  de corrélations conditionnelles et  $h_{iit}$  est donné par un modèle GARCH. On peut écrire  $h_{ijt}$  comme :

$$h_{ijt} = \rho_{ijt} \sqrt{h_{iit} h_{jtt}} \quad , \forall i \neq j \quad (3.2)$$

La condition de positivité de  $H_t$  implique la positivité de  $R_t$  et de chaque  $h_{iit}$ .

Dans ce cas, on

$$R_t = R = (\rho_{ijt}), \rho_{ii} = 1$$

C'est-à-dire la corrélation constante. L'équation (3.2) devient

$$h_{ijt} = \rho_{ij} \sqrt{h_{iit} h_{jtt}} \quad , \forall i \neq j$$

et la dynamique de la covariance est déterminée seulement par deux variances conditionnelles, ce qui implique un nombre de  $\frac{N(N-1)}{2}$  paramètres en  $R$ . Ce modèle est donc appelé conditional correlation constant GARCH (CCCGARCH).

Nous allons donc calculer le SCR à partir des données issues des simulations de rendements par le GARCH(1,1) et CCCGARCH(1,1) par quatre approches. La première approche considère que le rendement moyen annuel ( $\sum_{t=1}^{52} r_t$ ) suit une loi normale, la deuxième approche est le calcul du quantile 0.5% par la méthode Monte Carlo, la troisième approche s'appuie uniquement sur les données historiques, et la dernière méthode est celle de McNeil et Frey (2000).

Nous créons une série de 10400 rendements hebdomadaires pour chacune des deux méthodes de simulation, ce qui équivaut à 200 ans de rendements annuels ( $\sum_{j=1}^{52} r_j$ )<sub>t</sub> (Temps de retour moyen=200).

*b) SCR par la méthode paramétrique (gaussienne)*

D'après l'équation (3.1), le SCR calculé par la méthode gaussienne vérifie l'équation :

$$SCR_{gaussien} = \sqrt{52} * s_d * Z_{(0.995)} + \mathbb{E}(\sum_{t=1}^{52} r_t)$$

où  $Z_{(0.995)}$  est le quantile à 99.5% de la loi normale.

Les résultats de calcul de SCR en simulant le rendement de chaque stratégie par un GARCH(1,1) et un CCCGARCH(1,1) en considérant les innovations gaussiennes, sont donnés dans le Tableau III.8 :

Stratégie	SCR_Gaussien		Quantile 0,5%	
	GARCH(1,1)	CCCGARCH(1,1)	GARCH(1,1)	CCCGARCH(1,1)
CTA Long Term	26,08	21,96	19,08	20,7
Merger Arbitrage	13,7	6,74	14,01	7,88
LS Equity Variable Biais	7,65	17,9	5,92	18,64
Composite	7,64	10,1	8,3	10,61

**Tableau III. 8 : SCR et Quantile 0.5% sur rendements simulés**

Il ressort des résultats que, en simulant par le CCCGARCH, le SCR gaussien est très proche du quantile extrême. Ce qui n'est pas le cas quand les données sont simulées par le GARCH univarié.

Sur les données issues du GARCH multivarié, nous changeons la loi des innovations. Nous supposons d'abord que les résidus suivent la loi de Student, ensuite nous utilisons la méthode basée sur le bootstrap qui consiste, pour chaque stratégie  $i$ , à chaque temps  $t$ , à tirer l'innovation  $u_{it}$  de manière aléatoire avec remise dans la population des résidus standardisés obtenus en estimant les GARCH(1,1) sur chaque stratégie. On obtient les résultats donnés dans les Tableaux III.9 et III.10 ci-après :

Stratégie	Innovations suivent la loi de student	
	SCR_Gaussien	Quantile 0,5%
CTA Long Term	21,24	20,75
Merger Arbitrage	6,7	6,83
LS Equity Variable Biais	16,82	19,34
Composite	10,07	10,92

**Tableau III.9 : SCR et Quantile 0.5% avec innovations student**

Stratégie	Tirage avec remise des innovations	
	SCR_Gaussien	Quantile 0,5%
CTA Long Term	21,46	24,55
Merger Arbitrage	9,41	8,9
LS Equity Variable Biais	19,65	19,01
Composite	11,07	11,21

**Tableau III. 10: SCR et Quantile 0.5% avec innovations boostannées**

Il ressort des résultats que la loi des innovations n'a pas d'impact significatif sur le SCR. Du coup, les simulations se feront par la suite avec les innovations gaussiennes.

### *c) VaR par la méthode Monte Carlo*

Nous effectuons à présent 1000 simulations de 10400 rendements hebdomadaires  $r_t$  chacune. Nous calculons le SCR pour chaque simulation. C'est en fait le quantile de l'ordre 0.5% (car on s'intéresse aux pertes) pour la série  $(\sum_{t=1}^{52} r_t)$ . La VaR Monte Carlo finale est la moyenne des VaR calculées pour chaque simulation élémentaire.

Stratégie	Quantile 0.5%	
	GARCH(1,1)	CCCGARCH(1,1)
CTA Long Term	26,42	20,7
Merger Arbitrage	14,1	7,88
LS Equity Variable Biais	10,14	18,64
Composite	6,50	10,61

**Tableau III.11 : Quantile 0.5% sur rendements simulés**

Le SCR obtenu par la méthode de Monte Carlo sur les données multivariées est semblable au SCR paramétrique.

*d) VaR historique*

Nous calculons ensuite la somme des log-rendements pendant une année sur les données observées par blocs distincts de 52 données, c'est-à-dire de 1 à 52 puis de 53 à 104 etc... d'une part et par fenêtre glissante, c'est-à-dire de 1 à 52 puis de 2 à 53 etc...

Dans le premier cas, nous n'avons que 9 blocs distincts de rendements annuels (471 données !), ce qui n'est pas suffisant pour appliquer la théorie des valeurs extrêmes. Ici, nous retenons le minimum des neuf rendements pour chaque stratégie.

Dans le deuxième cas, les séries n'étant évidemment pas indépendantes, nous souhaitons voir si elles sont stationnaires. Un test de Phillips-Perron accepte la non-stationnarité. Le traitement des séries non-stationnaires par la théorie des valeurs extrêmes est un point largement discuté aujourd'hui et on ne dispose pas de méthodes simples comme dans le cas des séries stationnaires. Le résultat est consigné dans le Tableau III.12 suivant :

Stratégie	Minimum de la série	Quantile 0.5%
CTA Long Term	13.26	16.89
Merger Arbitrage	9.70	11.56
LS Equity Variable Biais	14.70	20.97
Composite	12.33	15.01

**Tableau III.12 : Quantile 0.5% sur fenêtre glissante d'un an**

La VaR historique sur fenêtres glissantes donne des mesures de VaR un peu plus prudentes. En dehors de la stratégie CTA Long Term, la VaR ici est supérieure à tous les autres quantiles calculés.

e) SCR par la méthode de McNeil et Frey

Il s'agit du modèle GPD conditionnel qui permet de combler les inconvénients des modèles standards de la TVE (Théorie des Valeurs Extrêmes). Dans la première étape, on simule la série CCCGARCH des rendements. Ensuite, on travaille sur des résidus standardisés de la série simulée qui sont calculés par la formule :

$$Z_t = \frac{r_t - \mu_t}{\sigma_t}$$

D'après le chapitre I ci-dessus, la  $VaR_p$ , correspondant au modèle GPD, pour un horizon donné  $t$  et avec une probabilité  $p$ , est obtenue par la relation :

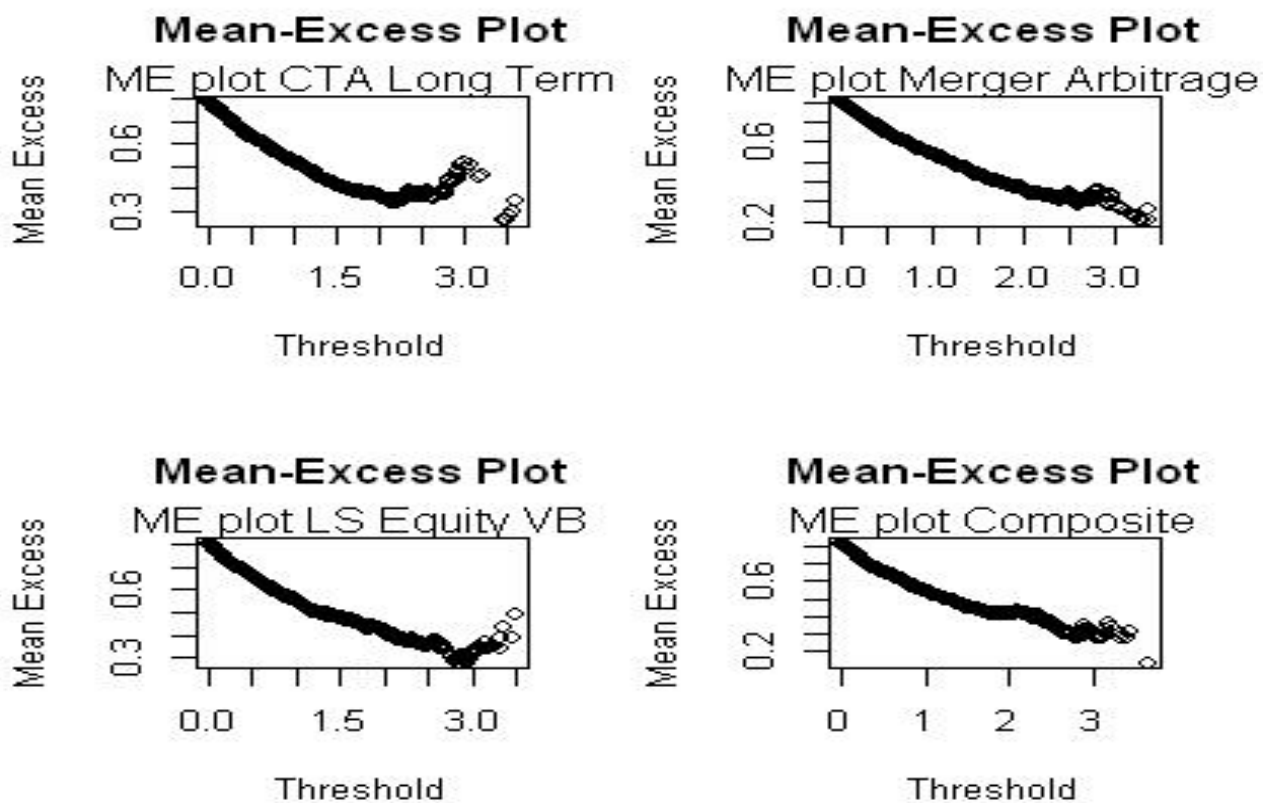
$$VaR_p = u + \frac{\beta}{\varepsilon} \left[ \left( \frac{n}{N_u} (1 - p) \right)^{-\varepsilon} - 1 \right] \quad (3.3)$$

Dans cette équation,  $\beta$  et  $\varepsilon$  représentent les estimateurs des paramètres de la loi GPD s'ajustant au  $N_u$  excès de rendements au-delà d'un seuil  $u$ .

Le choix de seuil  $u$  est capital pour l'approche GPD. En effet, un bon estimateur des quantiles issus d'un modèle GPD dépend énormément du choix du seuil. Le seuil optimal est le résultat d'un arbitrage entre le biais et la variance. Une des méthodes pour déterminer le seuil  $u$  est l'application du graphique ME plot (Mean Excess plot). La figure III-2 donne des ME plots pour les quatre stratégies considérées :

Note : On s'intéresse aux pertes, du coup on ne considère ici que la valeur absolue des rendements négatifs.

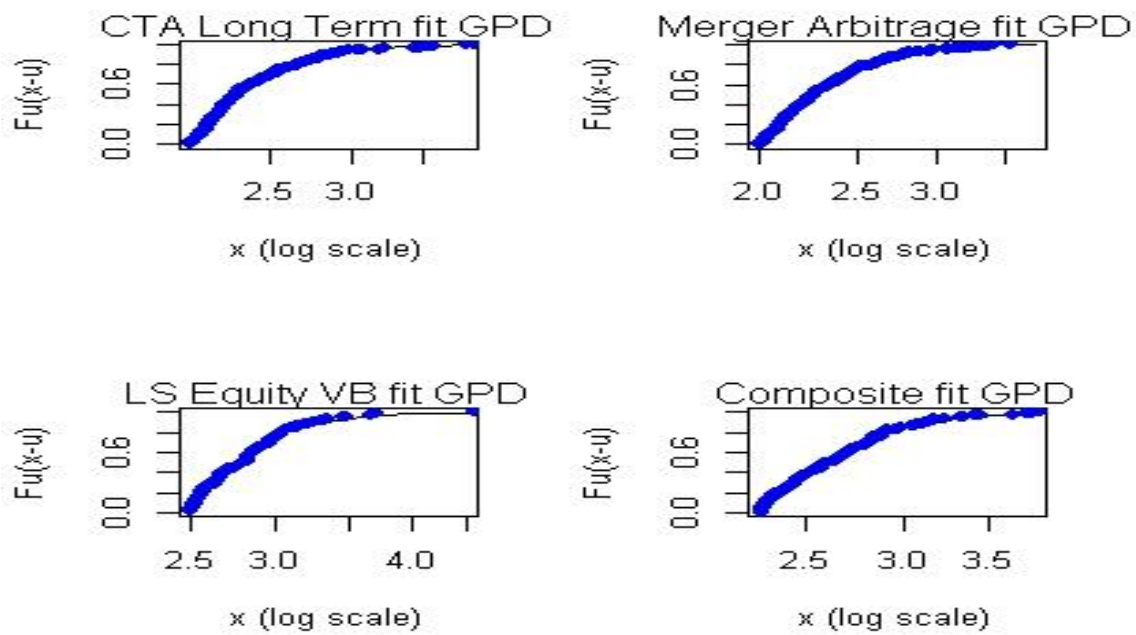




**Figure III.2 : ME plot des stratégies**

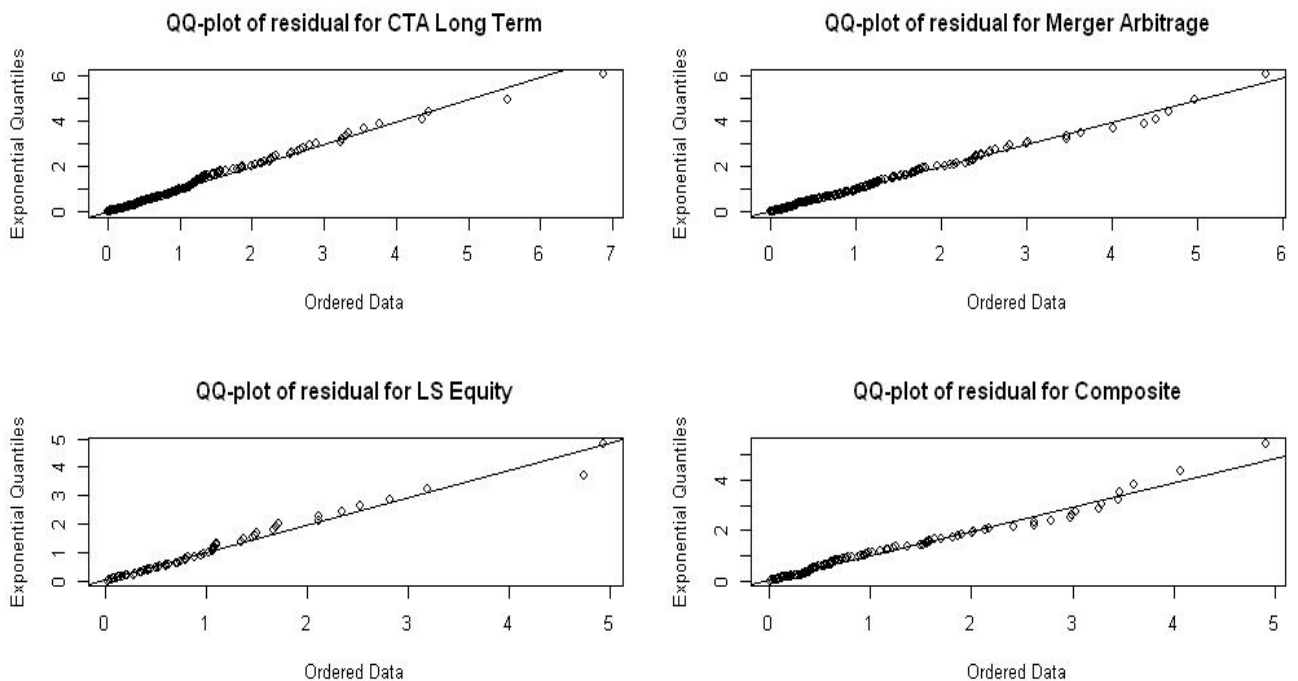
Le seuil  $u$  s'obtient graphiquement en zoomant sur le graphique et en regardant le point à partir duquel on a une tendance linéaire. Les seuils  $u_1$ ,  $u_2$ ,  $u_3$ ,  $u_4$  des quatre stratégies sont respectivement 2.1, 2, 2.5 et 2.3.

Les graphiques suivant nous permettent de vérifier si le choix du seuil  $u$  est efficace. La figure III.3 est l'estimation GPD de la distribution des excès avec respectivement les quatre seuils choisis ci-dessus. La droite noire de chaque graphique représente la fonction de répartition de la loi GPD estimée et les points bleus représentent la fonction de répartition empirique des excès.



**Figure III.3 : FitGPD versus Empirique Excess.**

Nous traçons le QQ-plot de ces résidus avec la distribution exponentielle comme référence. Une adéquation parfaite entre la distribution de l'échantillon des résidus et la loi exponentielle se caractérise par une représentation parfaitement linéaire du QQ-plot.



**Figure III.4: QQ plot of Residuals**

Les figures III.3 et III.4 ci-dessus montrent que notre choix du seuil  $u$  est acceptable et les données estimées sont donc bien distribuées selon la GPD. On remarque en effet sur la figure III.3 un ajustement adéquat de la queue de distribution. La GPD permettra alors d'estimer la  $VaR_p$  selon la formule (3.3).

Avec la  $VaR_p$  des innovations obtenue ci-dessus, on applique la formule de McNeil et Frey pour calculer le SCR:

$$SCR_{McNeil} = 52 * \frac{1}{T} \sum_{t=1}^T r_t + VaR_p * \sqrt{52} * \sqrt{\frac{1}{T} \sum_{t=1}^T (r_t - \bar{r}_T)^2}$$

On obtient les résultats donnés dans le tableau Tableau III.13

Stratégie	SCR_McNeil et Frey	SCR Gaussien
CTA Long Term	33,17	21,96
Merger Arbitrage	21,25	6,74
LS Equity Variable Biases	23,81	17,90
Composite	18,60	10,10

**Tableau III.13 : Comparaison SCR McNeil et Frey et SCR Gaussien**

Les résultats obtenus nous permettent de remarquer que nous avons sous-estimé la perte annuelle dans le cas Gaussien. En appliquant la méthode de McNeil et Frey, la  $VaR_{99,5\%}$  obtenue est bien plus importante que dans le cas gaussien. Il est plus prudent que le cas gaussien et donc préférable au sens de Solvabilité II. Ainsi, pour des investisseurs averses aux risques, c'est plus prudent de choisir la méthode McNeil pour le calcul de la provision.

### C- Constitution du portefeuille optimal

Nous voulons dans cette section résoudre le problème d'optimisation d'un portefeuille constitué des stratégies de Hedge fund que nous avons sélectionnées, à savoir CTALT, VB, MA et COM. Il s'agit en fait de trouver la valeur de  $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)$  qui satisfait la relation:

$$\text{Min}_{\alpha} SCR_{MC}(\alpha_1 \sum_{j=1}^{52} r_j^1 + \alpha_2 \sum_{j=1}^{52} r_j^2 + \alpha_3 \sum_{j=1}^{52} r_j^3 + \alpha_4 \sum_{j=1}^{52} r_j^4)$$

sous contrainte :

$$\alpha_1 E \left( \sum_{j=1}^{52} r_j^1 \right) + \alpha_2 E \left( \sum_{j=1}^{52} r_j^2 \right) + \alpha_3 E \left( \sum_{j=1}^{52} r_j^3 \right) + \alpha_4 E \left( \sum_{j=1}^{52} r_j^4 \right) \geq \text{Rendement défini}$$

Nous calculons le rendement annuel moyen des stratégies par la méthode de Monte Carlo. Nous effectuons 1000 simulations de 10400 observations chacune. Chaque observation représente le rendement hebdomadaire. A chaque simulation, nous calculons le rendement moyen et le rendement moyen final retenu est la moyenne des rendements moyens de chaque simulation primaire. Le tableau ci-dessous nous donne les rendements moyens (annuels) de chaque stratégie et aussi un rendement visé:

Stratégie	Rendements moyens
<b>Rendement visé</b>	<b>4,75</b>
CTA Long Term	4,72
Merger Arbitrage	6,19
LS Equity Variable Biais	1,81
Composite	3,01

**Tableau III.14 : Rendements moyens**

Nous cherchons donc à trouver une combinaison convexe optimale du portefeuille qui minimise le SCR, calculé par les deux méthodes à savoir la méthode gaussienne et celle de McNeil Frey. Cette optimisation du portefeuille se fait sous contrainte d'assurer un rendement annuel moyen supérieur ou égal à 4.75%. Le rendement de 4.75% est un minimum et pourrait changer en fonction de la constitution en proportion du portefeuille. Il est choisi dans ce contexte non seulement en fonction de l'étude des rendements moyens des quatre stratégies mais aussi parce que c'est le rendement moyen d'un Hedge Fund en réalité. Les proportions sont données dans les Tableaux III.15 et III.16, ainsi que le SCR nécessaire.

Stratégie	Portfolio Proportion (in %)				
	CTA Long Term	0,0007	100	0	0
Merger Arbitrage	99,97	0	100	0	0
LS Equity Variable Biais	0,027	0	0	100	0
Composite	0,0023	0	0	0	100
SCR- Gaussien method	6,23	21,96	6,74	17,9	10,1
Rendements du portfolio (in %)	6,18	4,72	6,19	1,81	3,01

**Tableau III.15 : Proportion du portfolio 1 avec SCR Gaussien**

Stratégie	Portfolio Proportion (in %)				
	CTA Long Term	0,735	100	0	0
Merger Arbitrage	78	0	100	0	0
LS Equity Variable Biais	9,5	0	0	100	0
Composite	11.765	0	0	0	100
SCR- McNeil Frey method	12,9	33,17	21,25	23,81	18,6
Rendements du portfolio (in %)	5.38	4,72	6,19	1,81	3,01

**Tableau III.16 : Proportion du portfolio 2 avec SCR McNeil et Frey**

Finally, with the proportions obtained for each method of calculation of the SCR, we perform a back-test to count the number of times that the annual return of each portfolio exceeds the SCR threshold, i.e. the annual loss exceeds the SCR to a negative value.

	Portfolio 1 (SCR Gaussien)	Portfolio 2 (SCR McNeil Frey)
SCR	6,23	12,9
Nombre de dépassement	5	1

**Tableau III.17 : Nombre de violation du SCR**

As seen previously, the SCR calculated by the Gaussian method underestimates the loss. The Gaussian method gives a smaller SCR which entails a risk for the investor. With the method of McNeil and Frey, the total SCR, although superior to that obtained by the Gaussian method, has been well reduced compared to the SCR of strategies taken individually. Statistically, it is preferable to have a portfolio constituted for the essential of the strategy « Merger Arbitrage » which requires a small SCR for a better return. But in reality, it is preferable to privilege diversification, and therefore prudence, under Solvability II. And as we see with the back-test, the number of times that the loss exceeds the SCR is lower in the case of McNeil and Frey than in the Gaussian case. This confirms the prudent strategy with the method of McNeil and Frey.

## Conclusion

Notre étude a porté sur la mise en œuvre de la théorie de la ruine dans deux cadres : le risque alimentaire et le risque assurantiel.

Concernant le risque alimentaire, l'absorption des quantités importantes des contaminants nocifs pour la santé contenus dans les aliments peut dans un horizon donné causer des problèmes de santé graves avec une certaine probabilité; nous avons modélisé la probabilité de tomber malade d'un individu. Nous avons d'abord estimé les paramètres de queues de distribution pour chacun des trois contaminants. Puis, nous avons calculé la probabilité pour que la quantité d'un contaminant dans l'organisme dépasse un seuil  $u_1$  au cours d'une semaine ou d'une année, sachant que la quantité d'un deuxième contaminant a dépassé un seuil  $u_2$  au cours de la même période. Les quantités de contaminants ont été déterminées en faisant des hypothèses sur les trajectoires des expositions. Les trajectoires de charge corporelles s'obtiennent en supposant que les processus d'éliminations sont indépendants. Nous avons montré que cette hypothèse n'est pas toujours vérifiée. Nous avons donc mis une dépendance entre ces processus. La détermination de la matrice non diagonale pour le calcul de la charge corporelle constituerait donc un bon avancement de la recherche sur ces problèmes de contaminants.

Dans un deuxième temps, nous avons travaillé sur des données financières (plus précisément sur quatre stratégies de Hedge Fund du groupe Lyxor). Nous avons d'abord modélisé indépendamment les 4 séries de manière dynamique en incluant dans chacune des paramètres de volatilité de type GARCH et nous avons appliqué nos méthodes sur les résidus indépendants identiquement distribués, issus de ce modèle, en les supposant gaussiens. Ceci nous a permis d'obtenir des projections dans l'avenir des stratégies étudiées, ce qui rend possible le calcul de la VaR (quantile 99.5% ou 0.5% selon qu'on regarde le gain ou la perte) ou d'un quantile extrême, par la méthode GEV.

Nous avons tenu compte ensuite de la dépendance des quatre séries de données en utilisant le modèle Garch multivarié (CCC Garch) pour la simulation et en supposant, comme dans le cas univarié, que le vecteur de résidus est un vecteur gaussien. Les SCR calculés par la méthode gaussienne et la méthode de quantile sont très proches pour les données simulées en multivarié, ce qui nous conforte bien dans l'hypothèse de dépendance. Nous avons

présenté ensuite l'approche conditionnelle de la théorie de valeurs extrêmes en comparant le SCR calculé par l'approche normale et l'approche GPD (McNeil et Frey). La comparaison du modèle GPD conditionnel à un ensemble de modèle laisse penser à la supériorité du modèle GPD parce qu'il produit une VaR qui reflète l'aspect dynamique de la volatilité et permet de capturer la lourdeur des queues (i.e. des pertes dans ce cas), ce que la VaR normale ne fait pas et aussi parce qu'il permet d'être assez prudent, ce qui est vivement recommandé dans Solvabilité II. Nous nous sommes ensuite intéressés à la constitution du portefeuille optimal, qui consiste à constituer un portefeuille contenant une proportion de chacune des stratégies dans le but de réduire le SCR sous contrainte d'une meilleure rentabilité. La méthode gaussienne fournit un SCR faible avec un rendement plutôt élevé, mais un portefeuille tourné vers une seule stratégie ou presque. La méthode de McNeil et Frey fournit certes un rendement moins intéressant que dans le cadre gaussien avec un SCR plus important, mais dans ce cas, on a une vraie diversification du portefeuille, ce qui semble plus recommandable pour un investisseur averse au risque.

Nous n'avons considéré dans cette application que quatre stratégies de Hedge Fund, et il serait sans doute intéressant de regarder ce qui se passe avec toutes les stratégies ensemble.

## Liste des Tableaux

Tableau I.1 : « Concordance » entre modèle de ruine et KDEM .....	27
Tableau I.2 : Caractéristiques de quelques copules.....	34
Tableau II.1 : Statistiques descriptives hebdomadaires des contaminants .....	41
Tableau II. 2 : Demi-vies, DHT et Seuils des différents contaminants .....	49
Tableau II. 3 : Probabilités avec seuils(j)=TBBj et A digonale.....	55
Tableau II. 4 : Probabilités en % avec seuil(j)= $q_{0,95}$ et A diagonale.....	57
Tableau II. 5 : Excédances, seuils et quantiles des différents contaminants.....	58
Tableau III. 1 : Méthode de valorisation du bilan prudentiel et de l'exigence de marge .....	61
Tableau III.2 : Les modules de SCR d'après le QIS 5 - Formule Standard .....	63
Tableau III. 3 : Matrice de corrélation des différents modules pour le calcul du BSCR - QIS 5.....	64
Tableau III.4 : Matrice de corrélation pour le module risque de marché en cas de hausse des taux .....	66
Tableau III.5 : Matrice de corrélation pour le module risque de marché en cas de baisse des taux.....	67
Tableau III.6 : Rendements des différentes stratégies de Hedge Fund.....	72
Tableau III. 7 : Statistiques descriptives des rendements hebdomadaires.....	74
Tableau III. 8 : SCR et Quantile 0.5% sur rendements simulés .....	77
Tableau III.9 : SCR et Quantile 0.5% avec innovations student .....	78
Tableau III. 10: SCR et Quantile 0.5% avec innovations bostrappées.....	78
Tableau III.11 : Quantile 0.5% sur rendements simulés.....	79
Tableau III.12 : Quantile 0.5% sur fenêtre glissante d'un an.....	79
Tableau III.13 : Comparaison SCR McNeil et Frey et SCR Gaussien.....	83
Tableau III.14 : Rendements moyens.....	84
Tableau III.15 : Proportion du portfolio 1 avec SCR Gaussien.....	84
Tableau III.16 : Proportion du portfolio 2 avec SCR McNeil et Frey.....	85
Tableau III.17 : Nombre de violation du SCR .....	85



## Table des figures

Figure I. 1 : Densités des lois de valeurs extrêmes.....	10
Figure I. 2 : Méthode des excès.....	12
Figure I. 3 : Densités des lois de valeurs extrêmes.....	13
Figure I. 4 : Un exemple de trajectoire du processus $(R(t))_{t \geq 0}$ .....	25
Figure I. 5 : Modèle de Ruine (en haut) et modèle KDEM en bas).....	28
Figure II.1 : Histogramme des Pcb(a), du Cd(b) et du MeHg(c).....	43
Figure II.2 : QQplot des valeurs des expositions.....	47
Figure II.3 : Evolution de la charge corporelle.....	48
Figure II.4 : QQplot des charges corporelles sur données simulées.....	48
Figure II.5 : Dependogramme des charges corporelles .....	51
Figure II.6 : Hill plot des Pcb, Cd, MeHg .....	55
Figure III.1 : Evolutions hebdomadaire des stratégies .....	74
Figure III.2 : ME Plot des stratégies .....	81
Figure III.3 : FitGPD versus Empirique Excess .....	82
Figure III.4 : QQ -plot of Residuals.....	82

## Bibliographie

- BALKEMA, A. &. (1974). Residual Life Time at Great Age,. *Annals of probability*, 792-804.
- BEIRLANT, J. G. (2004). *Statistics of Extremes -Theory and Applications*. England: Wiley.
- BERTAIL P., C. S. (2010). Statistical analysis of a dynamic model for food contaminant exposure with applications to dietary methylmercury contamination. *Journal of Biological Dynamics*, 212-234.
- BIARD, R. (2010). *Dépendance et événements extrêmes en théorie de la ruine : étude univariée et multivariée, problèmes d'allocation optimale*. Lyon: Thèse de l'Université Claude Bernard de Lyon I.
- BOLLERSLEV, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307-327.
- BUKLMANN, H. (1970). *Mathematical risk theory*. Berlin: Springer.
- CLEMENCON, C. &. (2009). Exposition aux risques alimentaires et processus stochastiques : le cas des contaminants chimiques. *Journal de la SFDS.*, 3-29.
- CRAMER, H. (1954). *The elements of probability theory and some of its applications*. New-York: Almqvist och Wiskell, Uppsala and Wiley.
- CRUZ, M. C. (1999). Modelling and measuring operational risk. *Journal of risk*, 63-72.
- EMBRECHTS P., K. C. (1997). *Modeling Extremal Events for Insurance and Finance*. Berlin: Springer.
- EMBRECHTS P., R. S. (1998). *Extreme value theory as a risk management tool*. Department of Mathematics ETH, Swiss Federal Technical University.
- DAROLLES, S et M., VAISSIE (2013), *Regulation: Threat or Opportunity for the Fund of Hedge Fund Industry? In Reconsidering Funds of Hedge Funds, The Financial Crisis and Best Practices in UCITS, Tail Risk, Performance, and Due Diligence, Academic Press*
- ENGEL, R. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50, 987-1007.
- FISHER, R. &. (1928). Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proceedings of the Cambridge Philosophical Society*, 24, , 180-190.
- FOUGERES, P. &. (2012). Estimation of conditional law given an extreme component. *Extremes*. 1-34.
- FREY, R. &. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series : an Extreme Value Approach. *Journal of Empirical Finance*, 7, 271-300.
- GARRIDO, M. (2002). *Modélisation des événements rares et estimations des quantiles extrêmes, Méthode de sélection de modèles pour les queues de distribution*. Grenoble: Thèse de doctorat, Université Grenoble 1.

- GOURIEROUX, C. M. (1984 ). Pseudo Maximum Likelihood Methods : Theory. *Econometrica*, 52, 680-700.
- HOSKING, J. &. (1987). Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29, 339-349.
- LUNDBERG, F. (1903). *Approximerad framställning av sannolikhetsfunktionen*. Uppsala: Akad. Afhandling. Almqvistoch Wiksell.
- MIKOSCH, T. &. (2012). *The cluster index of regularly varying sequences with applications to limit theory for functions of multivariate Markov chains*. Preprint.
- PICKANDS, J. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, 3, 119-131.
- ROOTZEN, H. &. (1996). Extreme Value Statistics and Wind Storm Losses : a Case Study. *Scandinavian Actuarial Journal*,, 70-94.
- SHARPE, W. (1964). Capital Asset Prices : a Theory of Market Equilibrium under Conditions of Risk . *Journal of Finance*, 19, 425-442.
- SMITH, R. (1989 ). Extreme Value Analysis of Environmental Time Series : an Application to Trend Detection in Ground-Level Zone. *Statistical Science*, 4, 367-393.
- VAISSIE, M. (2012), *Solvency II: Regulation Change and Hedge Fund Evolution*, *Journal of Alternative Investments*, 15, 86-97
- WHITE, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica*, 53 , 1-16.
- Wikipedia. (2013, 05 28). Récupéré sur <https://fr.wikipedia.org/wiki/Polychlorobiph%C3%A9nyle>.
- SMITH R. L.(2003), *Statistics of extremes, with application in environment, insurance and finance*,
- WINTERBERGER, O. (2013, 02). <http://wintenberger.fr/ens>.