

# **MODÉLISATION AVANCÉE DES RISQUES**

---

Pierre Clauss

DiFiQ  
Université Paris-Dauphine  
Ensaе

## OBJECTIF DE L'ENSEIGNEMENT

---

L'objectif de ce cours de 12h composé de 4 séances de 3h est ambitieux : assimiler les techniques de modélisation avancée des risques dans le but de pouvoir les appliquer à de nombreux domaines d'application. En effet, mesurer les risques est essentiel en finance pour la gestion d'actifs, la détermination des fonds propres des banques et assurances ou encore le pricing de produits complexes.

Ce cours est composé de 3 parties : la première porte sur les risques extrêmes avec un focus sur les fondements de la Value-at-Risk (VaR) et une application des méthodes statistiques de la Théorie des Valeurs Extrêmes (TVE) au calcul de la VaR.

La seconde partie traite des risques multiples et plus particulièrement de l'outil probabiliste *copule* ainsi que ses propriétés mathématiques et statistiques. Cet outil a une définition mathématique relativement simple mais son heuristique n'est pas si aisée à appréhender.

Enfin, la troisième partie a pour objectif de maîtriser les techniques probabilistes de modélisation du risque de crédit.

La *philosophie* de ce cours est d'aborder en profondeur les techniques statistiques et probabilistes appliquées à la mesure du risque de marché et du risque de crédit et de les appliquer à des enjeux financiers. C'est pour cela qu'une partie importante du cours est dédiée à la programmation informatique, à l'aide du logiciel R<sup>1</sup>. Le repo sur ma page [GitHub](#) dédié à ce cours permettra d'y retrouver ce polycopié, les corrigés des exercices, les données financières et les codes R permettant de faire les ateliers du cours.

---

1. [R Core Team](#) (2020). R : A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

# TABLE DES MATIÈRES

<b>RISQUES EXTRÊMES</b>	<b>6</b>
<b>INTRODUCTION</b>	<b>6</b>
<b>1 VALUE-AT-RISK : FONDEMENTS ET PREMIÈRES FORMALISATIONS</b>	<b>8</b>
1.1 Fondements de la VaR . . . . .	8
1.1.1 Mesure synthétique du risque extrême . . . . .	8
1.1.2 Définition mathématique . . . . .	9
1.1.3 Évaluation . . . . .	10
1.2 Premières formalisations . . . . .	11
1.2.1 VaR historique et VaR bootstrap . . . . .	11
1.2.2 VaR Gaussienne . . . . .	11
1.2.3 VaR Student . . . . .	12
<b>2 VALUE-AT-RISK ET THÉORIE DES VALEURS EXTRÊMES</b>	<b>15</b>
2.1 Théorie des valeurs extrêmes . . . . .	15
2.1.1 Théorèmes . . . . .	15
2.1.2 Méthodes statistiques . . . . .	17
2.2 Application à la Value-at-Risk . . . . .	20
<b>CONCLUSION</b>	<b>22</b>
<b>RISQUES MULTIPLES 1</b>	<b>24</b>
<b>INTRODUCTION</b>	<b>24</b>
<b>3 DÉFINITION ET PROPRIÉTÉS DES COPULES</b>	<b>25</b>
3.1 Définition . . . . .	25
3.2 Propriétés . . . . .	26
3.2.1 Théorème de Sklar . . . . .	26
3.2.2 Bornes de Fréchet . . . . .	26
3.2.3 Théorème d'invariance . . . . .	27
3.2.4 Dépendance de queue . . . . .	27
3.3 Exemples de copules . . . . .	28
3.3.1 Copule Gaussienne . . . . .	28
3.3.2 Copule Student . . . . .	28
3.3.3 Copules Archimédiennes . . . . .	29

<b>4</b>	<b>MISE EN ŒUVRE DES COPULES</b>	<b>32</b>
4.1	Méthodes de simulation . . . . .	32
4.1.1	Méthode des distributions . . . . .	32
4.1.2	Méthode des distributions conditionnelles . . . . .	33
4.2	Méthodes d'inférence statistique . . . . .	34
4.2.1	Méthode des moments . . . . .	34
4.2.2	Méthode du maximum de vraisemblance . . . . .	34
4.2.3	Outils graphiques d'ajustement . . . . .	35
	<b>CONCLUSION</b>	<b>38</b>
	<b>RISQUES MULTIPLES 2</b>	<b>40</b>
	<b>INTRODUCTION</b>	<b>40</b>
<b>5</b>	<b>MODÉLISATION DE LA PROBABILITÉ DE DÉFAUT</b>	<b>41</b>
5.1	Approche par les ratings . . . . .	41
5.1.1	Notation des agences . . . . .	41
5.1.2	Scoring des émetteurs . . . . .	43
5.2	Modèle de Merton . . . . .	44
5.2.1	Modélisation de la firme . . . . .	45
5.2.2	Probabilité de défaut . . . . .	45
5.2.3	Estimation . . . . .	46
5.3	Modèle à intensité constante . . . . .	46
5.3.1	Probabilité de défaut . . . . .	46
5.3.2	Évaluation d'une obligation risquée . . . . .	47
5.3.3	Évaluation d'un CDS . . . . .	47
<b>6</b>	<b>MODÉLISATION DE LA PERTE EN CAS DE DÉFAUT</b>	<b>49</b>
6.1	Mesure de la LGD . . . . .	49
6.2	Modélisation probabiliste et estimation . . . . .	50
<b>7</b>	<b>MODÉLISATION DE LA DÉPENDANCE DES DÉFAUTS</b>	<b>51</b>
7.1	Mesurer la dépendance des événements de défaut . . . . .	51
7.1.1	Dépendance à partir du modèle de Merton . . . . .	51
7.1.2	Dépendance à partir du modèle à intensité constante . . . . .	51
7.2	Copules et dépendance des défauts . . . . .	52
<b>8</b>	<b>CREDITVAR</b>	<b>53</b>
	<b>CONCLUSION</b>	<b>55</b>
	<b>BIBLIOGRAPHIE</b>	<b>55</b>

# **RISQUES EXTRÊMES**

**APPLICATION À LA MESURE DU RISQUE DE MARCHÉ**

# INTRODUCTION

La Théorie des Valeurs Extrêmes (TVE) a révolutionné la finance quantitative il y a 25 ans, en permettant d'affiner la précision des résultats de mesures de risque fournis aux banques. En effet, la TVE a permis de modéliser rigoureusement la queue de distribution des pertes d'un portefeuille d'investissement. Néanmoins, la grande crise financière de 2007-2011 révèle combien il reste de chemin pour généraliser ses résultats et définir des mesures des risques extrêmes robustes.

Historiquement, les besoins en modélisation des risques se sont fait sentir à la suite des différentes crises des institutions bancaires dans les années 70, 80 et 90 (Herstatt, Barings, Natwest, LTCM, entre autres). Alors, des réglementations ont été définies pour prévenir les pertes irréversibles des banques, dues à des chutes de valeur sur les marchés des actifs financiers, à des faillites de contreparties ou à des risques opérationnels, et ainsi définir un montant de réserve.

C'est l'objectif du Comité de Bâle qui, en 1988, définit un ratio international de solvabilité d'exigence d'un certain montant de fonds propres. Il permet de sécuriser les prises de risque des banques en renforçant l'adéquation entre les risques et leurs fonds propres. Le ratio Cooke qui correspond au rapport entre le montant des fonds propres et celui des encours pondérés de crédit doit être supérieur à 8%.

Le Comité de Bâle a développé en janvier 2001 une réglementation plus approfondie concernant les risques à évaluer sur un portefeuille d'actifs. Suivant la spécificité de ces actifs, trois risques fondamentaux et modélisables se distinguent :

- le risque de marché, correspondant à tout risque de perte dû aux fluctuations des marchés financiers (baisse des prix des actions, des taux, des devises, etc). Ce risque est relatif aux crises que peuvent subir les marchés financiers (crise de 1929, krach de 1987, bulle technologique des années 2000, crise des subprimes de 2007).
- le risque de crédit, relatif au risque de défaut de paiement d'une entreprise, d'une institution financière ou encore d'un Etat, qui a contracté des contrats de dettes et qui ne peut rembourser ses créanciers dans les temps. Nous pouvons citer pour les plus médiatisés Enron, les ménages subprimes, Lehman Brothers, la Grèce.
- le risque opérationnel, concernant tous les problèmes relatifs aux opérations de transaction comme les erreurs, les pannes de système, les incendies, les fraudes, etc.

Le ratio de solvabilité devient alors celui de MacDonough correspondant au rapport entre les fonds propres et les encours issus des activités de marché, de crédit et opérationnelles. Il doit être aussi supérieur à 8%.

Depuis, la gestion du risque pris par les banques est devenue fondamentale. La future réforme Bâle IV qui interviendra début 2022 à partir d'un texte écrit en décembre 2017 par le Comité de Bâle contraindra plus fortement les banques. Et les exigences sur les modèles internes seront accrues s'inscrivant dans la droite ligne du document SR 11-7 de la Fed, *Guidance on Model Risk Management*, publiée en 2011.

Et cette gestion s'est répandue au-delà des banques chez les assureurs (Solvabilité 2), les gérants de fonds (DICI ou Document d'Informations Clés pour l'Investisseur). Modéliser, mesurer, gérer les risques est devenu essentiel dans les activités financières modernes. L'objet de ce cours est d'étudier précisément l'apport de la TVE en ce qui concerne la mesure du risque de marché.

Il contient deux chapitres.

Le premier chapitre s'attachera à définir l'indicateur le plus important pour mesurer le risque de marché :

la Value-at-Risk. Nous étudierons aussi les premières formalisations de cet indicateur de risque : VaR historique, VaR bootstrap, VaR Gaussienne, VaR Student.

Le second chapitre développera les méthodologies statistiques issues de la TVE qui s'appliqueront alors au cas du calcul de Value-at-Risk et nous permettront de définir des mesures de risques extrêmes robustes.

Nous nous plaçons dans un cadre univarié de mesure du risque, c'est-à-dire faisant abstraction de la structure de dépendance des actifs financiers (cet aspect sera développé dans les parties 2 et 3 de ce cours).

# CHAPITRE 1

## VALUE-AT-RISK : FONDEMENTS ET PREMIÈRES FORMALISATIONS

### 1.1 Fondements de la VaR

#### 1.1.1 Mesure synthétique du risque extrême

Le risque de marché correspond à un risque de perte des portefeuilles des investisseurs dues aux variations des marchés financiers : marchés des instruments de base (actions, obligations, devises, matières premières) mais aussi marchés des produits dérivés (contrats à terme, options). Pour gérer le risque de marché, il faut donc mesurer de manière précise ce risque extrême.

Une première mesure du risque est la **sensibilité** du produit financier détenu relativement à son marché. Ainsi, puisque le risque de marché est le risque de perte consécutivement aux évolutions des marchés, si, naïvement, nous possédons une action européenne et que le marché des actions européen subit une baisse, il semble naturel de penser qu'il va en être de même pour notre action. Cette première mesure s'assimile à la sensibilité des portefeuilles ou titres détenus relativement à des facteurs de risque de marché. Ainsi, des modèles de risque se sont développés comme le Capital Asset Pricing Model (cf. [Sharpe \(1964\)](#)) pour le plus connu. Ils mesurent les sensibilités (coefficients de régression) des titres ou portefeuilles de titres financiers à des facteurs de risque tels le marché dans son ensemble par exemple.

Néanmoins, le *risk manager* a besoin d'une mesure plus synthétique du risque encouru par son portefeuille d'investissement. En effet, lorsque plusieurs instruments très différents composent le portefeuille, il est difficile d'agrèger les nombreuses sensibilités. Nous pouvons ainsi appréhender le risque à partir des profils de la dispersion des pertes et profits des actifs. Nous pouvons faire ressortir deux mesures de risque à partir de la distribution des rentabilités des actifs : la **volatilité** et la **Value-at-Risk** ou **VaR**.

Ces mesures ne captent pas la même chose :

- la volatilité mesure la dispersion moyenne des rentabilités autour de leur moyenne : en effet, son expression en fonction du vecteur des rentabilités  $R$  est  $\sigma = \sqrt{\mathbb{E}[(R - \mathbb{E}(R))^2]}$ . Elle prend donc en compte toutes les rentabilités, positives ou non, extrêmes ou non.
- la Value-at-Risk est le quantile de perte déterminant la plus grande perte que peut subir un portefeuille avec une probabilité d'occurrence faible sur un horizon donné : elle mesure ainsi le risque extrême.

La logique entre les deux mesures est donc différente. En effet, la volatilité peut être élevée et seulement capturer des risques **moyens**, certes importants, mais pas **extrêmes**. Tout l'enjeu d'une mesure du risque synthétique pertinente est d'estimer correctement la perte potentielle que peut subir un investissement. Or un actif peut avoir une volatilité de 20% mais ne pas connaître de pertes plus élevées en valeur absolue que 5% par exemple sur une journée. Au contraire, un actif avec une volatilité de 15% peut connaître des pertes,



certes rares, de plus de 10% sur une journée. Le second actif nous semble donc plus risqué, même si ses risques moyens s'avèrent être moins importants que ceux du premier.

En fait, déterminer le risque par le moment d'ordre 2 (la volatilité) présuppose que les moments suivants (skewness et kurtosis) ne nécessitent pas d'être incorporés dans une mesure de risque. Le paradigme sous-jacent est la normalité des rentabilités d'un actif financier (la loi Normale étant caractérisée par les deux premiers moments). Utiliser la Value-at-Risk permet de résoudre ce problème car un quantile de distribution n'est pas une mesure moyenne de la dispersion mais prend seulement en compte les pertes qui nous intéressent ici, c'est-à-dire les pertes extrêmes.

Nous venons d'introduire le fait que la volatilité n'est certainement pas la meilleure mesure de risque extrême. C'est pour cela que la Value-at-Risk est utilisée par la plupart de risk managers.

Nous allons maintenant la définir et la formaliser précisément.

### 1.1.2 Définition mathématique

La VaR représente une perte potentielle dont il faut définir la **probabilité**  $\alpha$  d'occurrence, ou de manière équivalente le niveau de confiance  $1 - \alpha$ , ainsi que l'**horizon**  $h$  de durée d'investissement. Par exemple, une probabilité de 1% pour un horizon  $h$  de un jour revient à accepter que la perte potentielle ne dépasse la VaR que 2 à 3 fois par an (1 an  $\approx$  260 jours ouvrés). L'horizon représente la durée sur laquelle nous estimons le risque, c'est-à-dire le temps nécessaire pour couvrir la position du portefeuille : 1 jour en trading, 1 mois en gestion d'actifs.

Soit  $V_t^P$  la valeur en  $t$  du portefeuille et  $V_{t+h}^P$  la valeur à l'horizon  $h$ . La rentabilité du portefeuille  $R_{t|t+h}^P$  investi à la date  $t$  et détenu sur un horizon de durée  $h$  a pour expression,  $\forall t \geq 1$  :

$$R_{t|t+h}^P = \frac{V_{t+h}^P}{V_t^P} - 1.$$

Supposons que  $R_{t|t+h}^P$  est issu d'une variable aléatoire notée  $R^P$ . La  $\text{VaR}(h, \alpha)$  est alors définie par :

$$\mathbb{P}[R^P \leq \text{VaR}(h, \alpha)] = \alpha$$

c'est-à-dire la probabilité que la rentabilité du portefeuille soit inférieure à la VaR sur un horizon  $h$  est égale à  $\alpha$ <sup>1</sup>.

**Définition 1.1 (VaR).** Notons  $F$  la fonction de répartition de la variable aléatoire  $R^P$  des rentabilités du portefeuille de périodicité  $h$ . La Value-at-Risk de probabilité d'occurrence  $\alpha$  et d'horizon d'investissement  $h$  est alors le quantile défini par :

$$\text{VaR}(h, \alpha) = F^{-1}(\alpha) \tag{1.1}$$

Il nous reste à caractériser la forme de  $F$ . La première solution consiste à estimer  $F$  par sa distribution empirique : nous avons alors la VaR historique. La seconde est d'utiliser un modèle paramétrique : les praticiens utilisent classiquement la VaR Gaussienne. Mais d'autres paramétrisations (Student, TVE) se révéleront plus robustes.

---

1. Les praticiens utilisent régulièrement le niveau de confiance  $1 - \alpha$  qui correspond au quantile de probabilité  $\alpha$  : ainsi, une VaR de niveau de confiance 99% par exemple correspond au quantile de probabilité  $\alpha = 1 - 99\% = 1\%$ .

### 1.1.3 Évaluation

Une étape importante dans la mesure du risque de marché est l'évaluation de la Value-at-Risk. Deux méthodes d'évaluation s'offrent à nous :

- une méthode ex-ante, utilisant des tests d'adéquation statistiques des modélisations adaptées aux rentabilités financières dans le cas de VaR paramétriques,
- une méthode ex-post, utilisant des backtests : pour les réaliser, il nous faut séparer l'échantillon des données entre un échantillon d'apprentissage et un échantillon de backtest<sup>2</sup>. Nous calculons la VaR sur l'échantillon d'apprentissage et nous déterminons le nombre d'*exceptions* (rentabilités inférieures à la VaR) sur l'échantillon de backtest. Pour que le VaR soit adéquate aux données de marché, il faut que le pourcentage d'exceptions soit proche du niveau d'occurrence  $\alpha$  de la VaR. La détermination des exceptions de la Value-at-Risk est une étape essentielle de sa construction. Elle va permettre au risk manager de communiquer sur la robustesse des indicateurs qu'il a construits.

Un conseil pour appliquer ces méthodes d'évaluation : faire coïncider périodicité des données à partir lesquelles on définit la VaR et horizon d'investissement  $h$ . Ceci n'est malheureusement pas toujours possible. En effet, ceci devient impossible lorsqu'il est demandé un calcul de VaR sur un horizon long (1 an par exemple). Alors, il est difficile dans ce cas de construire un échantillon important de données. Les données viennent à manquer.

Deux solutions sont alors possibles :

- la première est celle du *scaling* : pour obtenir la VaR à horizon  $h * \delta$ , il suffit de déterminer la VaR à horizon  $h$  (que l'on peut évaluer) que l'on multiplie par le facteur  $\sqrt{\delta}$ .
- la seconde, qui demande un temps de calcul plus important mais qui est plus robuste, est l'utilisation de simulations Monte Carlo : nous simulons alors plusieurs trajectoires du portefeuille d'investissement sur la période  $h * \delta$  et nous obtenons pour chaque trajectoire une rentabilité à l'horizon souhaité. Nous déterminons la VaR sur l'échantillon de ces rentabilités simulées.

Concernant l'évaluation ex-ante, un outil graphique intéressant peut être utilisé : le QQ-plot. Rappelons ici sa définition. Cet outil examine graphiquement l'hypothèse que les rentabilités suivent une loi particulière. Nous traçons le nuage de points associant les quantiles de la distribution empirique avec ceux de la distribution théorique.

**Définition 1.2** (QQ-plot). *Le QQ-plot trace pour un échantillon de taille  $n$  les couples :*

$$\left\{ \left( F^{-1} \left( \frac{n - k + 1}{n + 1} \right), R_{k:n}^P \right), k = 1, \dots, n \right\}$$

avec  $F$  la fonction de répartition de la loi théorique et  $R_{k:n}^P$  la statistique d'ordre  $k$  sur les  $n$  données.

**Propriété 1.1.** *Une adéquation parfaite entre la distribution de l'échantillon des rentabilités et la loi testée se caractérise par une représentation parfaitement linéaire du QQ-plot.*

---

2. Cette construction prend son sens lorsque la modélisation est homoscédastique ; dans le cas contraire, il est nécessaire de réaliser un backtest récursif : la VaR est dynamique et évolue avec le temps.

## 1.2 Premières formalisations

### 1.2.1 VaR historique et VaR bootstrap

Pour la VaR historique, nous estimons  $F$  de l'équation 1.1 non-paramétriquement par sa distribution empirique  $\hat{F}$ . Supposons que nous disposons d'un historique de  $T$  observations des rentabilités. La distribution empirique  $\hat{F}$  des rentabilités du portefeuille est alors caractérisée par le vecteur  $(\hat{R}_{1|1+h}^P, \dots, \hat{R}_{T|T+h}^P)'$  et  $\hat{F}^{-1}(\alpha)$  correspond au quantile de probabilité  $\alpha$  de  $(\hat{R}_{1|1+h}^P, \dots, \hat{R}_{T|T+h}^P)'$ .

L'avantage de cette méthode est de ne pas imposer d'hypothèse sur la loi de distribution des rentabilités. Mais le problème qui se pose est la longueur  $T$  de l'historique qui, si elle est trop faible, ne fournit pas un calcul précis de la VaR. Nous pouvons alors utiliser la technique du ré-échantillonnage ou *bootstrap* (cf. Efron (1995)).

Nous construisons pour cela une version aléatoire de même taille du vecteur  $(\hat{R}_{1|1+h}^P, \dots, \hat{R}_{T|T+h}^P)'$  dans laquelle les rentabilités du portefeuille peuvent apparaître pour certaines aucune fois, pour d'autres une fois, pour d'autres encore deux fois, etc. Ce nouveau vecteur est une version ré-échantillonnée ou *bootstrap* du vecteur d'origine. Nous pouvons alors déterminer la VaR historique de cet échantillon. Nous réalisons  $B$  répliques *bootstrap* et la VaR historique *bootstrap* correspond à la moyenne des  $B$  VaR historiques.

L'inconvénient de ces VaR non paramétriques est l'*over-fitting*, soit la dépendance trop forte à l'échantillon sur lequel elle a été déterminée, et donc la difficulté de cette VaR à être robuste dans le temps. C'est pour cela que les VaR paramétriques, certes moins ajustées aux données étudiées, peuvent permettre une meilleure prévision.

### 1.2.2 VaR Gaussienne

De manière classique, le choix du modèle paramétrique se porte sur la loi Normale. Supposons donc que les rentabilités du portefeuille  $R_{t|t+h}^P$  sont issues d'une variable aléatoire Gaussienne : ainsi,  $F$  est définie par  $\mathcal{N}(\mu, \sigma^2)$  avec  $\mu$  l'espérance des rentabilités et  $\sigma^2$  la variance des rentabilités.

Nous obtenons ainsi à partir de l'équation (1.1) :

$$\Phi\left(\frac{\text{VaR}(h, \alpha) - \mu}{\sigma}\right) = \alpha$$

avec  $\Phi$  la fonction de répartition de la loi Normale centrée réduite.

**Définition 1.3** (VaR Gaussienne). *La VaR Gaussienne est définie de la manière suivante :*

$$\text{VaR}(h, \alpha) = \mu + \sigma \Phi^{-1}(\alpha)$$

Cette méthode repose sur des hypothèses non satisfaisantes pour modéliser les rentabilités des actifs financiers puisque trois faits stylisés caractérisent leur distribution et ne sont pas captés par la modélisation Gaussienne :

1. l'asymétrie : l'occurrence de pertes extrêmes est plus forte que celle de profits extrêmes,
2. les queues épaisses,
3. le *clustering* : par exemple, les rentabilités élevées en valeur absolue sont suivies par des rentabilités élevées en valeur absolue.

Le deuxième fait stylisé peut s'observer sur la Figure 1.1 à l'aide des QQ-plot et densité de la queue des pertes de l'indice de marché français CAC 40 relativement à la loi Normale. En effet, on observe une queue plus épaisse dans l'échantillon des rentabilités du CAC 40 que dans un échantillon gaussien.

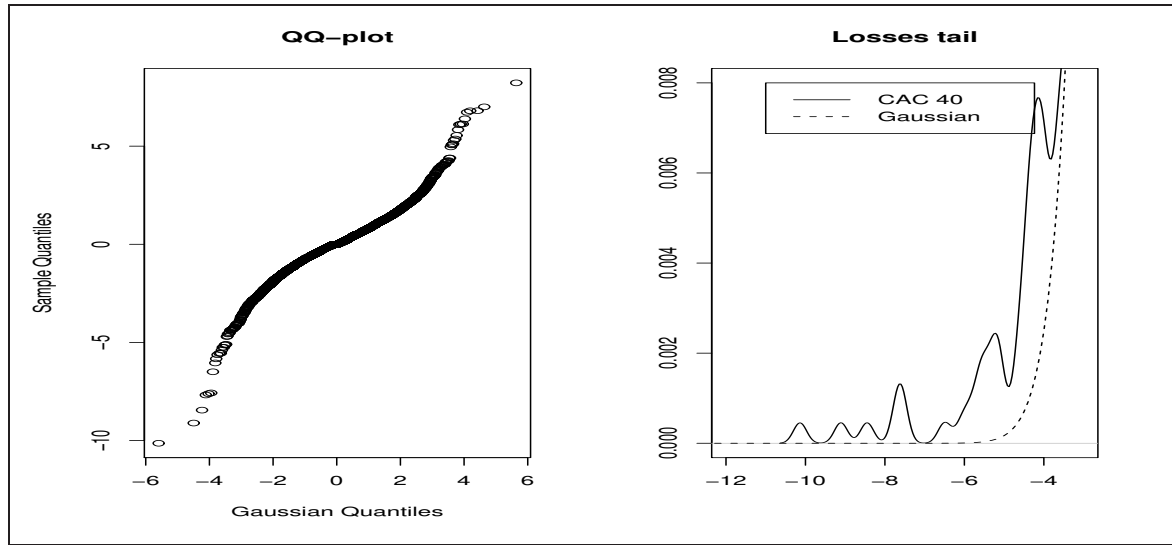


FIGURE 1.1 – QQ-plot et distribution de la queue des pertes du CAC 40 relativement à la loi Normale

### 1.2.3 VaR Student

Pour améliorer la précision de la VaR de marché, nous allons présenter une paramétrisation plus robuste de la loi des rentabilités d'un actif financier : la loi *Skew Student*. De nombreuses autres lois peuvent bien sûr aussi être utilisées (nous en verrons d'autres avec la TVE).

L'utilisation de la loi de Student permet d'appréhender de manière simple le kurtosis élevé des rentabilités financières, via le degré de liberté. Mais, pour une modélisation optimale, il faut maîtriser les autres paramètres de localisation, dispersion et asymétrie. C'est ce que nous allons faire grâce à la loi *skew Student* développée par [Azzalini et Capitanio \(2003\)](#).

Pour appréhender cette loi, nous allons auparavant déterminer la loi de Student centrée réduite, puis la loi de Student généralisée.

#### Loi de Student centrée réduite

La loi de Student qui est classiquement utilisée dans la littérature n'est pas centrée réduite.

**Définition 1.4** (Loi de Student). *Elle a pour densité :*

$$f_k(x) = A(k) \left[ 1 + \frac{x^2}{k} \right]^{-\frac{k+1}{2}}$$

avec  $A(k) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)}$ ,  $k$  le nombre de degrés de liberté (réel strictement positif) et la fonction

$$\text{Gamma} : \Gamma(a) = \int_0^{+\infty} e^{-t} t^{a-1} dt \text{ pour } a > 0^3.$$

**Propriété 1.2.** *L'espérance de la loi de Student est 0 (elle est non définie pour  $k \leq 1$ ).*

*La variance est égale à  $\frac{k}{k-2}$  (elle est non définie pour  $k \leq 2$ ).*

*Le skewness est 0 et il est défini pour  $k > 3$ .*

3. Les résultats essentiels de la fonction Gamma sont :  $\Gamma(x+1) = x\Gamma(x)$ ,  $\Gamma(n) = (n-1)!$  pour  $n$  entier, et  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ .

Le kurtosis est égal à  $\frac{3(k-2)}{k-4}$  et il n'est défini que pour  $k > 4$ .

Cette loi est donc déjà centrée mais non réduite. Ajoutons que l'épaisseur des queues de la distribution de Student augmente lorsque  $k$  diminue.

Notons la fonction de densité de la loi de Student classique  $f_k$ . Exhibons le paramètre de dispersion  $\sigma$  en calculant la densité  $f_{0,\sigma,k}$  de la variable  $Y = \sigma X$ . Nous utilisons un changement de variable<sup>4</sup> pour déterminer sa densité :

$$f_{0,\sigma,k}(y) = \frac{1}{\sigma} f_k\left(\frac{y}{\sigma}\right)$$

Calculons sa variance :

$$\begin{aligned} \mathbb{V}[Y] &= \mathbb{V}[\sigma X] \\ &= \sigma^2 \mathbb{V}[X] \\ &= \sigma^2 \frac{k}{k-2} \end{aligned}$$

Pour avoir  $\mathbb{V}[Y] = 1$ , il faut que  $\sigma = \sqrt{\frac{k-2}{k}}$ . La fonction de densité de la loi de Student centrée réduite est donc :

$$\begin{aligned} f_{0,1,k}(x) &= \sqrt{\frac{k-2}{k}} f_k\left(\sqrt{\frac{k-2}{k}} x\right) \\ &= \frac{1}{\sqrt{(k-2)\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left[1 + \frac{x^2}{k-2}\right]^{-\frac{k+1}{2}} \end{aligned}$$

### Loi de Student généralisée

Pour obtenir une loi de Student généralisée sur laquelle nous pouvons définir les paramètres de localisation  $\mu$  et de dispersion  $\sigma$ , nous utilisons à nouveau un changement de variable avec la fonction affine  $Y = \sigma X + \mu$  :

$$\begin{aligned} f_{\mu,\sigma,k}(y) &= \frac{1}{\sigma} f_{0,1,k}\left(\frac{y-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma} \sqrt{\frac{k-2}{k}} f_k\left(\sqrt{\frac{k-2}{k}} \left(\frac{y-\mu}{\sigma}\right)\right) \\ &= \frac{1}{\sigma} \frac{1}{\sqrt{(k-2)\pi}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)} \left[1 + \frac{1}{k-2} \left(\frac{y-\mu}{\sigma}\right)^2\right]^{-\frac{k+1}{2}} \end{aligned}$$

### Loi skew Student

Pour améliorer les performances d'ajustement de la distribution de Student, [Azzalini et Capitanio \(2003\)](#) intègrent un paramètre d'asymétrie réel  $\gamma$ . Lorsque  $\gamma = 0$ , nous retrouvons la loi symétrique ; lorsque  $\gamma > 0$ , le skewness de la loi est positif ; et il est négatif lorsque  $\gamma < 0$ <sup>5</sup>.

4. En notant  $f$  la densité de  $X$  et  $g$  celle de  $Y$ , si  $Y = \phi \circ X$  avec  $\phi$  un difféomorphisme, alors  $g(y) = f(\phi^{-1}(y)) |(\phi^{-1})'(y)|$ .

5. Ceci est valable lorsque le moment d'ordre 3 est défini, c'est-à-dire pour  $k > 3$ .

**Définition 1.5** (Loi skew Student). *La distribution de la loi skew Student s'écrit :*

$$f_{\mu,\sigma,\gamma,k}(x) = 2f_{\mu,\sigma,k}(x)F_{k+1} \left[ \gamma \frac{x-\mu}{\sigma} \sqrt{\frac{k+1}{\left(\frac{x-\mu}{\sigma}\right)^2 + k}} \right]$$

Une autre formulation de la loi *skew Student* va nous être utile lorsqu'il s'agira de simuler des nombres aléatoires engendrés par cette loi :

$$SkT = \frac{SN}{\sqrt{X/k}}$$

avec  $ST$  suivant une loi *skew Student*,  $SN$  suivant une loi *skew Normal* de paramètres  $\mu$ ,  $\sigma$  et  $\gamma$ , et  $X$  suivant une loi du  $\chi^2$  à  $k$  degrés de liberté.

La loi *skew Normal* a pour fonction de densité :

$$g_{\mu,\sigma,\gamma}(x) = 2\varphi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\gamma \frac{x-\mu}{\sigma}\right)$$

avec  $\varphi(\cdot)$  et  $\Phi(\cdot)$  respectivement les fonctions de densité et de répartition d'une loi Normale  $\mathcal{N}(0, 1)$ .

Nous obtenons ainsi un outil de modélisation qui s'avérera efficace et complet (nous pouvons maîtriser les 4 moments et donc l'asymétrie et le caractère leptokurtique des rentabilités d'un actif financier) pour la distribution des rentabilités d'un actif financier. L'estimation de cette loi se fait à l'aide de la méthode classique du Maximum de Vraisemblance résolu à l'aide de méthodes numériques (cf. [Azzalini et Capitanio \(2003\)](#)).

**Exercice 1.1.** *Calculer les différentes VaR historique, bootstrap, Gaussienne et Student avec  $h = 1$  jour pour  $\alpha = 1\%$  et  $\alpha = 0.1\%$  des rentabilités d'indices actions proposés et dont vous comparerez les résultats. Quelle est la VaR paramétrique la plus précise selon vous ?*

## CHAPITRE 2

# VALUE-AT-RISK ET THÉORIE DES VALEURS EXTRÊMES

### 2.1 Théorie des valeurs extrêmes

La Théorie des Valeurs Extrêmes s'intéresse non pas à la modélisation totale d'une distribution mais seulement aux queues de cette distribution à partir de lois spécifiques que nous allons préciser dans ce qui suit (cf. [Reiss et Thomas \(2001\)](#) et [Beirlant et al. \(2004\)](#) pour plus d'approfondissements).

Deux théorèmes sont essentiels à la compréhension de la Théorie des Valeurs Extrêmes : celui de [Fisher et Tippet \(1928\)](#) et celui de [Balkema et de Haan \(1974\)](#) et [Pickands \(1975\)](#).

#### 2.1.1 Théorèmes

##### Théorème de Fisher-Tippett

**Définition 2.1** (Loi GEV). *Nous supposons avoir une suite d'observations  $Z_1, \dots, Z_n$  issue d'une fonction de distribution inconnue  $F$  (aucune hypothèse n'est présupposée). Cet échantillon peut être séparé en  $k$  blocs disjoints de même longueur  $s$  (un bloc peut correspondre à un mois, un an, etc.) formant les données  $Z_1^{(i)}, \dots, Z_s^{(i)}$ ,  $i = 1, \dots, k$ . Nous nous intéressons aux maxima de ces  $k$  blocs  $Y_i^s = \max(Z_1^{(i)}, \dots, Z_s^{(i)})$ , qui forment l'échantillon de données supposées iid  $Y_1^s, \dots, Y_k^s$  de distribution  $F^s$ . La loi fondamentale à la modélisation des maxima est la Generalized Extreme Value (GEV) définie par la fonction de répartition suivante :*

$$H_\xi(x) = \begin{cases} \exp\left(-(1 + \xi x)^{-1/\xi}\right) & \text{si } \xi \neq 0 \\ \exp(-e^{-x}) & \text{si } \xi = 0 \end{cases}$$

où  $x$  est tel que  $1 + \xi x > 0$ .  $\xi$  est le paramètre de forme ou de queue. La GEV rassemble trois distributions particulières : si  $\xi > 0$ , c'est la distribution de Fréchet ; si  $\xi < 0$ , nous avons la distribution de Weibull ; enfin,  $\xi = 0$  donne la distribution de Gumbel. La famille peut être étendue en ajoutant les paramètres de localisation  $\mu$  et de dispersion  $\sigma > 0$  : la GEV  $H_{\xi, \mu, \sigma}(x)$  est en fait  $H_\xi((x - \mu)/\sigma)$ .

Le théorème essentiel à la modélisation des maxima est celui de [Fisher et Tippet \(1928\)](#) prouvé en 1928.

**Théorème 2.1** (Fisher-Tippett). *Supposons que nous pouvons trouver deux suites d'entiers réels  $a_i > 0$  et  $b_i$  telles que :*

$$a_i^{-1}(Y_i^s - b_i) \xrightarrow{\mathcal{L}} H \text{ pour } s \rightarrow \infty$$

ce qui est équivalent à

$$\mathbb{P}\{a_i^{-1}(Y_i^s - b_i) \leq y\} = F^s(a_i y + b_i) \rightarrow H(y) \text{ pour } s \rightarrow \infty \quad (2.1)$$

avec  $H$  une loi non dégénérée.  $a_i^{-1}(Y_i^s - b_i)$  sont les  $k$  maxima normalisés. Alors,  $F$  est dans le maximum domain of attraction (MDA) de  $H$ , ce qui s'écrit  $F \in \text{MDA}(H)$ .

Fisher-Tippett montre alors que  $F \in \text{MDA}(H)$  si, et seulement si,  $H$  est du type de  $H_\xi$ . La GEV est donc la seule distribution limite non dégénérée pour un échantillon de maxima normalisés<sup>1</sup>.

### Théorème de Balkema-de Haan-Pickands

**Définition 2.2** (Loi GPD). Nous supposons toujours avoir une suite d'observations  $Z_1, \dots, Z_n$  issue d'une fonction de distribution inconnue  $F$ , cette fois-ci présumées iid. Soit  $z_0$  la limite supérieure finie ou infinie du support de la loi  $F$ . Nous nous intéressons alors au nombre  $N_u$  d'extrêmes (strictement positifs) qui excèdent un seuil élevé  $u$ , et plus précisément aux excès au-dessus du seuil  $u$ ,  $X_i = Z_i - u$ , qui forment l'échantillon des excès  $X_1, \dots, X_{N_u}$ , supposés iid. Nous définissons la fonction de distribution des excès au dessus du seuil  $u$  par

$$F_u(x) = \mathbb{P}\{Z - u \leq x | Z > u\} = \mathbb{P}\{X \leq x | Z > u\} = \frac{F(x + u) - F(u)}{1 - F(u)} \quad (2.2)$$

pour  $0 \leq x < z_0 - u$ .  $F_u(x)$  est donc la probabilité qu'un extrême excède le seuil  $u$  par une quantité inférieure ou égale à  $x$ , sachant que le seuil est dépassé.

La loi essentielle à la modélisation des excès est la Generalized Pareto Distribution (GPD) définie par la fonction de répartition suivante :

$$G_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp(-x) & \text{si } \xi = 0 \end{cases}$$

où le support est  $x \geq 0$  lorsque  $\xi \geq 0$  et il devient  $0 \leq x \leq -1/\xi$  lorsque  $\xi < 0$ . La GPD regroupe trois distributions selon les valeurs du paramètre de forme. Lorsque  $\xi > 0$ , c'est la loi Pareto usuelle ; lorsque  $\xi < 0$ , nous avons la loi de Pareto de type II ; et  $\xi = 0$  donne la loi exponentielle. La famille peut être étendue en ajoutant le paramètre de localisation  $\mu$  : la GPD  $G_{\xi, \mu, \sigma}(x)$  est en fait  $G_\xi((x - \mu)/\sigma)$ . Si la loi des excès est  $G_{\xi, \sigma}(x)$ , alors celle des extrêmes est  $G_{\xi, \sigma}(x - u)$ .

Le théorème essentiel à la modélisation des excès est celui de [Balkema et de Haan \(1974\)](#) et [Pickands \(1975\)](#).

**Théorème 2.2** (Balkema-de Haan-Pickands). Pour une certaine classe de distributions<sup>2</sup>, la GPD est la distribution limite de la distribution des excès lorsque le seuil tend vers  $z_0$ . Formellement, nous pouvons trouver une fonction positive mesurable  $\sigma(u)$  telle que

$$\lim_{u \rightarrow z_0} \sup_{0 \leq x < z_0 - u} |F_u(x) - G_{\xi, \sigma(u)}(x)| = 0 \quad (2.3)$$

si, et seulement si,  $F \in \text{MDA}(H_\xi)$ .

1. Par exemple, la distribution de Student est dans le maximum domain of attraction de la distribution de Fréchet qui présente des queues épaisses. La distribution Gaussienne est quant à elle dans le MDA de la loi de Gumbel qui présente des queues de taille moyenne.

2. Cette classe est large : elle regroupe toutes les distributions continues classiques.



En résumé, nous avons :

Si pour une distribution  $F$  inconnue<sup>3</sup>, l'échantillon des maxima normalisés converge en loi vers une distribution non dégénérée, alors il est équivalent de dire que  $F$  est dans le *maximum domain of attraction* de la GEV  $H_\xi$  (**Fisher-Tippett**). Dans ce cas, il s'en déduit que la distribution des excès au-dessus d'un seuil élevé converge vers une GPD  $G_{\xi,\sigma}$  lorsque le seuil tend vers la limite supérieure du support de  $F$  (**Balkema-de Haan-Pickands**).

A partir de ces deux résultats de probabilité, nous obtenons deux méthodes différentes de modélisation statistique : le premier théorème étudie le comportement asymptotique d'un échantillon de maxima (modélisation *Block Maxima*) ; alors que le second étudie celui d'un échantillon d'excès au-dessus d'un seuil élevé (modélisation *Peaks Over Threshold*).

### 2.1.2 Méthodes statistiques

Deux méthodes principales de modélisation des événements rares sont possibles : la méthode *Block Maxima* (BM) qui modélise la distribution des extrêmes par la distribution *Generalized Extreme Value* (GEV) ; et la méthode *Peaks Over Threshold* (POT) qui modélise la distribution des excès au-dessus d'un seuil<sup>4</sup> (en anglais *threshold*) élevé par la *Generalized Pareto Distribution* (GPD).

#### Modélisation paramétrique de la distribution des maxima par blocs

Cette modélisation des queues de distribution s'appuie sur le théorème de Fisher-Tippett et nous supposons que l'échantillon de maxima suit exactement une loi GEV.

##### Sélection de la taille des blocs

Il n'y a pas d'outils statistiques d'aide à la sélection de la taille des blocs. Il faut cependant que la condition asymptotique (2.1) soit vérifiée et donc que  $s$  soit suffisamment grand. Mais, il faut aussi que nous ayons un nombre suffisant de maxima pour que l'estimation des paramètres de la GEV soit assez précise : c'est le compromis biais-variance. Il est donc courant de prendre  $s = 21$  (pour un mois) ou  $s = 252$  (pour un an), valeurs qui représentent un bon compromis suivant la taille de l'échantillon d'origine.

##### Estimation du modèle BM

A partir de l'échantillon des maxima construit précédemment, nous pouvons estimer les paramètres de la GEV. Deux méthodes d'estimation sont envisageables : l'Estimation par Maximum de Vraisemblance (EMV) et celle par les Moments Pondérés (EMP). Ici, nous ne verrons que l'EMV.

Soit l'échantillon de maxima supposé iid  $Y = (Y_1^s, \dots, Y_k^s)$  et  $h_{\xi,\mu,\sigma}$  la densité de la loi GEV  $H_{\xi,\mu,\sigma}$ . Cette dernière s'écrit pour  $\xi \neq 0$  :

$$h_{\xi,\mu,\sigma}(y) = \frac{1}{\sigma} \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]^{-\frac{1+\xi}{\xi}} \exp \left\{ - \left[ 1 + \xi \left( \frac{y - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}$$

Et la vraisemblance de l'échantillon  $Y^s$  est égale à :

$$\mathcal{L}(\xi, \mu, \sigma; Y^s) = \prod_{i=1}^n h_{\xi,\mu,\sigma}(Y_i^s) \text{ (car l'échantillon est iid)}$$

Il est fait appel à des procédures numériques pour la maximisation de la vraisemblance. Alors le calcul des estimateurs ne pose pas de sérieux problèmes. En revanche, rien ne nous assure de leurs régularités

3. Si  $F$  était connue, nous n'aurions pas besoin de nous intéresser au comportement asymptotique des maxima puisque  $F^n$  serait très simplement identifiable.

4. Dans la littérature, il se peut que la méthode POT comprenne aussi la modélisation de la fréquence des excès : ceci est une extension du modèle POT classique, qui décrit le comportement des excès ainsi que leur fréquence par un *point process* de Poisson à deux dimensions. Nous ne l'étudierons pas ici.

(estimateurs asymptotiquement efficaces et normaux) surtout lorsque l'échantillon est de petite taille. [Smith \(1989\)](#) montre qu'il suffit que  $\xi \geq -0.5$ <sup>5</sup> pour que les conditions de régularité de l'EMV soient remplies. Dans le cas où  $\xi = 0$ , la log-vraisemblance est égale à :

$$l(0, \mu, \sigma; Y) = -n \ln \sigma - \sum_{i=1}^n \exp\left(-\frac{Y_i - \mu}{\sigma}\right) - \sum_{i=1}^n \frac{Y_i - \mu}{\sigma}$$

En dérivant cette fonction relativement aux deux paramètres, nous obtenons le système d'équations à résoudre suivant :

$$\begin{aligned} n - \sum_{i=1}^n \exp\left(-\frac{Y_i - \mu}{\sigma}\right) &= 0 \\ n + \sum_{i=1}^n \frac{Y_i - \mu}{\sigma} \left[ \exp\left(-\frac{Y_i - \mu}{\sigma}\right) - 1 \right] &= 0 \end{aligned}$$

Précisons cependant qu'il n'existe pas de solution explicite à ces équations de maximisation (utilisation de méthodes numériques).

### Modélisation paramétrique de la distribution des excès

Cette modélisation des queues de distribution s'appuie sur le théorème de Balkema-de Haan-Pickands et nous supposons que l'échantillon des excès au-dessus du seuil  $u$  suit exactement une loi GPD.

#### Sélection du seuil

Avant de pouvoir estimer le modèle, il faut construire un échantillon adéquat. Il nous faut trouver un seuil  $u$  de sélection des données extrêmes suffisamment élevé pour que l'approximation asymptotique (2.3) soit applicable (si nous choisissons un seuil trop bas, les estimations seront biaisées) et au-dessus duquel nous conservons assez de données pour des estimations précises (si le seuil est trop élevé, les écarts-types des estimateurs seront très importants) : c'est le compromis biais-variance.

Un des outils de choix du seuil est le graphe de la *sample mean excess function*  $e_n(u)$  (ME-plot).

**Définition 2.3** (ME-plot). *Le ME-plot est défini de la manière suivante :*

$$\{(u, e_n(u)), Z_{n:n} < u < Z_{1:n}\}$$

où  $Z_{1:n}$  et  $Z_{n:n}$  sont respectivement les maximum et minimum de l'échantillon. Et  $e_n(u)$  est définie par

$$e_n(u) = \frac{\sum_{i=1}^n (Z_i - u)^+}{\sum_{i=1}^n 1_{\{Z_i > u\}}} = \frac{1}{N_u} \sum_{i=1}^n (Z_i - u)^+$$

c'est-à-dire la somme des excès au-dessus du seuil  $u$  divisé par le nombre  $N_u$  de données qui excèdent  $u$ . La *sample mean excess function*  $e_n(u)$  est l'estimateur empirique de la *mean excess function* :

$$e(u) = \mathbb{E}[Z - u | Z > u]$$

Comment interprète-t-on ce graphe ? Il faut savoir tout d'abord que la *mean excess function* de la GPD est :

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \tag{2.4}$$

---

5. Ce qui est le cas généralement des distributions des rentabilités des actifs financiers.

où  $\sigma + \xi u > 0$ . Ainsi, si le ME-plot a un comportement linéaire au-dessus d'une certaine valeur de  $u$ , cela signifie que les excès au-dessus de ce seuil suivent une GPD.

Dans l'exemple de la Figure 2.1 (pertes en valeur absolue d'un indice boursier), nous pouvons définir le seuil à 1 : en effet, cette valeur correspond à un judicieux arbitrage entre le souhait d'avoir un seuil élevé et celui d'obtenir un échantillon d'excès de taille importante.

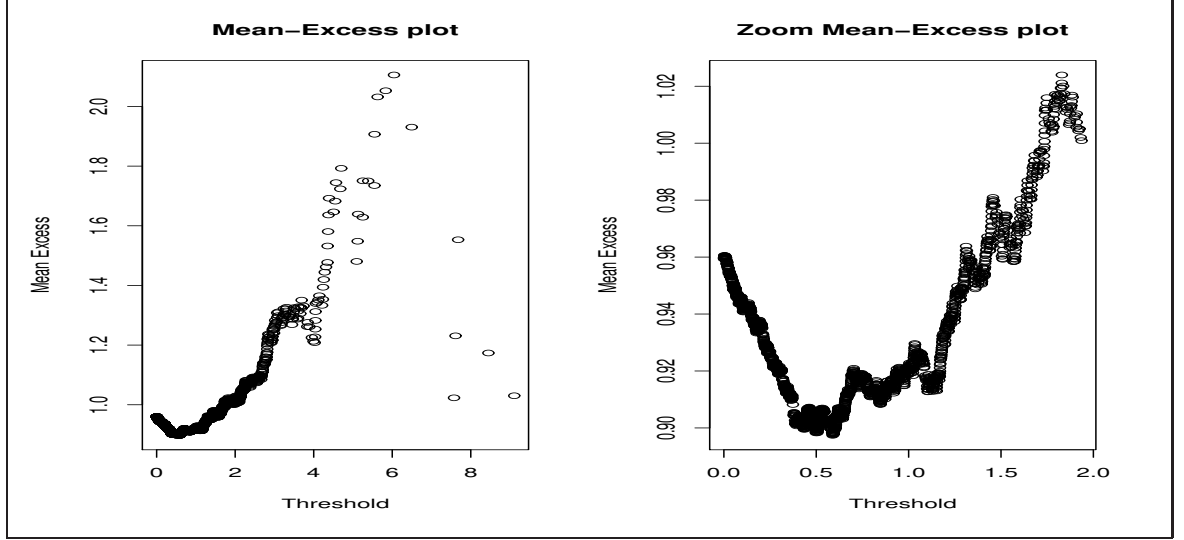


FIGURE 2.1 – Mean-Excess plots pour la queue des pertes

#### Estimation du modèle POT

Deux méthodes d'estimation sont ici encore envisageables : l'Estimation par Maximum de Vraisemblance (EMV) et celle par les Moments Pondérés (EMP). Comme pour la méthode BM, nous nous intéresserons ici qu'à l'EMV.

Supposons que notre échantillon des excès  $X = (X_1, \dots, X_{N_u})$  est iid avec comme fonction de distribution la GPD  $G$ . La fonction de densité  $g$  de  $G$  est alors pour  $\xi \neq 0$  :

$$g(x) = \frac{1}{\sigma} \left(1 + \xi \frac{x}{\sigma}\right)^{-\frac{1}{\xi}-1}$$

La log-vraisemblance est donc égale à :

$$l(\xi, \sigma; X) = -N_u \ln \sigma - \left(\frac{1}{\xi} + 1\right) \sum_{i=1}^{N_u} \ln \left(1 + \frac{\xi}{\sigma} X_i\right)$$

En dérivant cette fonction en  $\xi$  et  $\sigma$ , nous obtenons les équations de maximisation à partir desquelles nous calculons les estimateurs du Maximum de Vraisemblance  $(\hat{\xi}_{N_u}, \hat{\sigma}_{N_u})$  (à l'aide de méthodes numériques).

Et pour  $\xi = 0$ , nous avons :

$$g(x) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right)$$

$$l(0, \sigma; X) = -N_u \ln \sigma - \frac{1}{\sigma} \sum_{i=1}^{N_u} X_i$$

Nous obtenons alors  $\hat{\sigma}_{N_u} = \sum_{i=1}^{N_u} X_i / N_u$  qui n'est autre que la moyenne empirique des excès (pour  $\xi = 0$ , la GPD est la loi exponentielle).

Hosking et Wallis (1987) ont montré que lorsque  $0 \leq \xi \leq 0.4$  et pour des échantillons de taille petite, l'EMP obtient des estimateurs plus précis que l'EMV (avec des écarts-types plus faibles). Néanmoins, cette différence s'atténue avec l'augmentation de la taille de l'échantillon. En outre, Rootzen et Tajvidi (1996) révèlent que pour  $\xi \geq 0.5$ , l'EMP calcule des estimateurs fortement biaisés contrairement aux estimateurs de l'EMV qui sont efficaces. Enfin, pour  $\xi \geq -0.5$ , les conditions de régularité de l'EMV sont remplies et les estimateurs du Maximum de Vraisemblance  $(\hat{\xi}_{N_u}, \hat{\sigma}_{N_u})$  calculés sur l'échantillon des  $N_u$  excès sont asymptotiquement normaux (Hosking et Wallis (1987)).

## 2.2 Application à la Value-at-Risk

Nous avons vu que deux méthodes statistiques de modélisation des queues s'offrent à nous : la méthode BM et la méthode POT. En Finance, néanmoins, une méthode va être préférée à cause du phénomène de *clustering* ou grappes de volatilité des rentabilités des actifs financiers.

En effet, nous allons privilégier la modélisation des queues de distribution par la méthode *Peaks Over Threshold* avec l'Estimation par Maximum de Vraisemblance qui possède plusieurs avantages sur les autres méthodes. Tout d'abord, elle est assez flexible et réaliste par rapport à la méthode *Block Maxima*, qui ne prend pas en compte toutes les valeurs susceptibles d'être extrêmes. En effet, la méthode BM extrait le maximum de chaque période définie au préalable (mois, année, etc.). Elle peut donc perdre certaines valeurs extrêmes qui pourraient survenir autour du maximum de la période (phénomène de cycles financiers ou *clustering*) alors qu'à la période suivante, le maximum pourrait être relativement faible. Au contraire, la méthode POT évite ce problème puisqu'elle extrait les maxima au-dessus d'un seuil fixé à l'avance. Cette méthode prend donc en compte les phénomènes de *clusters* ou grappes de volatilités caractéristiques des rentabilités des actifs financiers. Elle est donc particulièrement adaptée en Finance, lorsque dans d'autres domaines, où les cycles sont absents, la méthode BM sera préférée.

Enfin, l'EMV a une robustesse et consistance certaines pour les séries qui nous concernent (pour lesquelles  $\xi \geq -0.5$ ).

La justesse des résultats de cette modélisation s'observent sur la Figure 2.2.

La modélisation des rentabilités d'un portefeuille d'actifs financiers à l'aide d'une GPD et de la méthode POT va permettre d'ajuster un modèle robuste aux trois faits stylisés des actifs financiers. Ceci s'observe clairement sur la Figure 2.2.

Il faut ajouter une précision lorsque l'on calcule une VaR à partir de la TVE. Puisque la modélisation est ajustée à partir de l'échantillon des pertes et que nous voulons obtenir un quantile sur l'échantillon total, il nous faut réaliser une translation de la probabilité  $\alpha$  pour calculer la VaR de notre échantillon.

**Propriété 2.1.** La probabilité d'occurrence  $\alpha$  appliquée à la VaR calculée à l'aide d'une loi issue de la TVE est égale à :

$$\alpha_{TVE} = 1 - \kappa * \alpha$$

avec  $\kappa$  le ratio entre le nombre de données de l'échantillon total et le nombre de données de l'échantillon des extrêmes (approché par la taille du bloc  $s$  pour la méthode BM).

**Exercice 2.1.** Calculer les VaR GEV et GPD avec  $h = 1$  jour pour  $\alpha = 1\%$  et  $\alpha = 0.1\%$  des rentabilités d'indices actions proposés et dont vous comparerez les résultats. Quelle est la VaR la plus précise selon vous ?

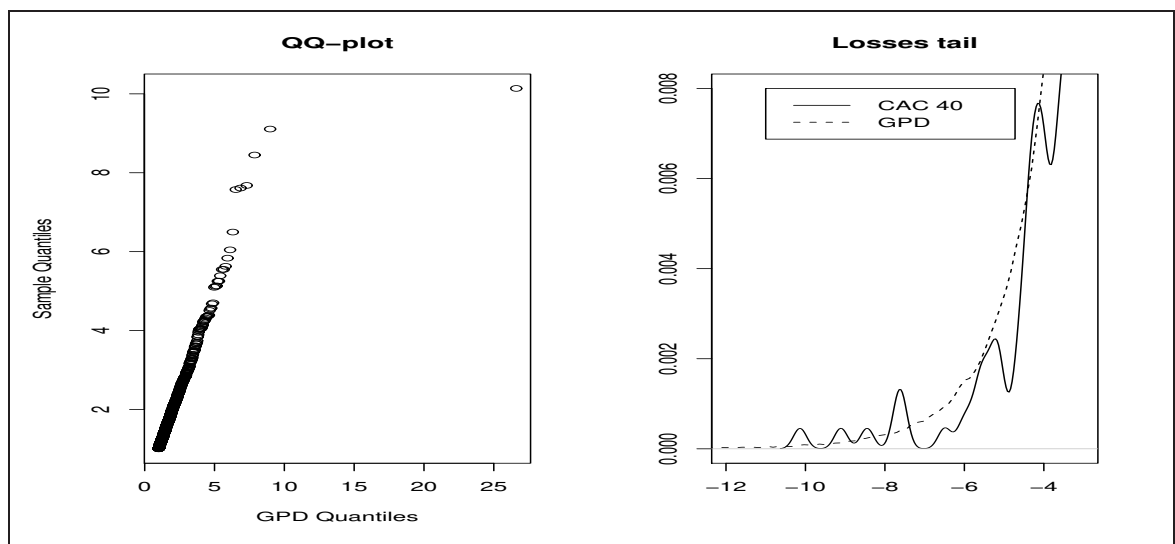


FIGURE 2.2 – QQ-plot et distribution de la queue des pertes du CAC 40 relativement à la loi GPD

# CONCLUSION

Cette partie a permis d'appliquer différentes méthodes statistiques (*bootstrap*, *skew Student*, TVE) à la modélisation et à la mesure du risque de marché. Nous avons alors calculé des Value-at-Risk robustes et ajustées à la réalité des marchés financiers.

La modélisation TVE peut être aussi combinée avec une modélisation dynamique des rentabilités financières dans le but d'améliorer les performances de la VaR paramétrique. Ceci est fait dans l'article de [Frey et McNeil \(2000\)](#) via l'utilisation d'une modélisation ARMA-GARCH avec des résidus suivant une loi de la TVE. Modéliser les rentabilités de manière dynamique va nous permettre d'affiner la modélisation de la VaR. Nous nous rapprocherons encore un peu plus de la réalité des marchés financiers, notamment de la notion de *persistance* ou *clustering* des marchés (une forte volatilité est suivie d'une autre forte volatilité).

Il faut néanmoins préciser que la VaR a ses inconvénients et il peut être judicieux de déterminer d'autres mesures de risque telle la Conditional Value-at-Risk (CVaR), appelée aussi Expected Shortfall. Cette mesure est en effet une mesure cohérente du risque, comme définie par [Artzner et al. \(1999\)](#), contrairement à la VaR, car elle respecte le critère de sous-additivité. Le principe de diversification est donc satisfait, ce qui n'est pas le cas de la VaR : en effet, la VaR globale d'un portefeuille peut être supérieure à la somme des VaR des sous-portefeuilles qui le composent. La CVaR se définit comme l'espérance des pertes au-delà de la VaR et dans le cadre de la TVE, et plus précisément de la modélisation par une GPD, elle s'écrit :

$$\text{CVaR}(h, \alpha) = \mathbb{E}[R^P | R^P \leq \text{VaR}(h, \alpha)] = \text{VaR}(h, \alpha) \left( \frac{1}{1 - \xi} + \frac{\sigma - \xi u}{(1 - \xi) \text{VaR}(h, \alpha)} \right)$$

Enfin, ces calculs de VaR présupposent la linéarité des rentabilités du portefeuille : un actif linéaire a une relation linéaire avec son facteur de risque sous-jacent. Les options par exemple ont au contraire une relation *non linéaire* avec leur facteur de risque synthétisé par le sous-jacent de l'option. Il est alors difficile de synthétiser un portefeuille de plusieurs actifs dont certains sont non-linéaires par une seule loi synthétique. L'alternative est l'utilisation de simulations Monte Carlo distinguant les modélisations des différents actifs linéaires et non-linéaires.

Il est enfin possible d'ajouter la modélisation de la structure de dépendance au sein du calcul de VaR via les simulations Monte Carlo des distributions des rentabilités des actifs composant le portefeuille d'investissement. Nous étudierons les modèles de structures de dépendance dans la partie suivante.

Pour terminer cette partie, il est intéressant de remarquer le caractère pro-cyclique du calcul de la VaR : en effet, lors de crises, la demande en capital va augmenter et inversement lors de marchés haussiers et moins volatils. Ceci est un problème que les nouvelles réglementations résolvent via des ajustements contra-cycliques plus vertueux.

# **RISQUES MULTIPLES 1**

**THÉORIE DES COPULES**

# INTRODUCTION

La copule est un outil relativement innovant de modélisation de la structure de dépendance de plusieurs variables aléatoires. La connaissance de cet outil probabiliste est essentielle à l'appréhension de nombreux domaines d'application de la finance quantitative : mesure du risque de crédit, évaluation de produits de crédits structurés, réplique de la performance des hedge funds, mesure du risque de marché, gestion de portefeuille à l'aide de simulations Monte Carlo, pricing d'options à plusieurs sous-jacents, etc. Ainsi, à chaque fois qu'il est nécessaire de modéliser une structure de dépendance, nous pouvons faire appel aux copules.

Une mesure de dépendance régulièrement utilisée est la corrélation linéaire de Bravais-Pearson. Cet indicateur est performant lorsque la relation de dépendance est linéaire et l'univers considéré gaussien, cadre d'analyse malheureusement rare en Finance. Pour remédier à cela, nous avons recours à d'autres indicateurs de dépendance se fondant sur les discordances et concordances observées sur un échantillon. Alors, nous utilisons des coefficients de corrélation non linéaires et non paramétriques, comme le tau de Kendall et le rho de Spearman. Ce sont de bons indicateurs globaux de la dépendance entre variables aléatoires. En outre, ils sont compris entre  $-1$  et  $1$  comme le coefficient de corrélation linéaire : une valeur de  $1$  signifie une concordance parfaite.

Mesurer la dépendance à l'aide d'indicateurs statistiques est une chose, la modéliser par une *fonction de dépendance* en est une autre. La copule répond à ce second objectif. Les indicateurs de dépendance (corrélation linéaire, tau de Kendall et rho de Spearman) pourront être définis dans le cadre de copules paramétriques à partir des paramètres de la copule.

Un intérêt pratique essentiel à l'utilisation de cet outil consiste en la possibilité de modéliser la dépendance d'événements extrêmes. Pour comprendre cela, prenons l'exemple du secteur bancaire et du cas de la faillite d'une banque. Ne pas répercuter cet événement au risque de faillite des autres banques paraîtrait peu réaliste dans une modélisation de la dépendance des défaillances. En effet, étant données les interdépendances existant entre ces sociétés, une faillite telle que celle de Lehman Brothers le 15 septembre 2008 a toutes les chances d'ébranler le secteur bancaire dans son entier.

Pour rendre compte de ce phénomène de contagion, nous avons à notre disposition bien entendu l'indicateur de dépendance *corrélation linéaire* et la fonction de dépendance qui y est associée, ie la *copule Gaussienne* (nous étudierons précisément sa définition et ses propriétés par la suite). Or cet outil suppose une dépendance des événements rares (le défaut d'une banque dans notre exemple) inexistante. Ainsi, modéliser la dépendance des phénomènes de faillite par la copule Gaussienne s'avèrerait peu probant. Cet outil a néanmoins été privilégié par de nombreux *quants* évaluant le prix d'un produit structuré comportant des emprunts subprimes ...

Pour remédier à cela, d'autres copules vont pouvoir être mobilisées. En effet, la grande diversité des copules existantes permettra de modéliser précisément la dépendance entre des événements extrêmes comme les événements de défaut.

Nous étudierons en premier lieu la définition et les propriétés fondamentales des copules en explicitant quelques exemples de copules paramétriques, pour ensuite, analyser leur mise en œuvre pratique via des méthodes de simulation et d'inférence statistique.



## CHAPITRE 3

# DÉFINITION ET PROPRIÉTÉS DES COPULES

La dépendance entre les variables aléatoires est parfaitement décrite par leur distribution jointe. Nous pouvons cependant distinguer les comportements des distributions marginales de la structure de dépendance : la copule est l'outil permettant d'extraire la structure de dépendance d'une distribution jointe et ainsi de séparer dépendance et comportement marginal.

En isolant cette structure de dépendance, il est alors possible de déterminer une loi multivariée originale, composée de lois marginales différentes. Prenons un exemple simple. La loi multivariée gaussienne est en fait la jonction d'une copule Gaussienne et de lois marginales gaussiennes. Si nous souhaitons modéliser plusieurs variables aléatoires à l'aide d'une structure de dépendance gaussienne mais avec des lois marginales différentes pour chacune des variables aléatoires, l'extraction de la copule Gaussienne à partir de la loi multivariée de même nom nous le permet. Ce sera l'objet du théorème 3.1 de Sklar.

Cette copule Gaussienne est donc extraite à partir d'une loi multivariée. Mais il est possible de définir des copules *ex-nihilo*. Il suffit que la fonction copule satisfasse la définition 3.1 suivante.

### 3.1 Définition

Nous nous limiterons dans ce cours à l'étude des copules bivariées, dont les résultats sont aisément généralisables dans la plupart des cas de dimension plus grande.

**Définition 3.1 (Copule).** La copule bivariée  $C$  fonction de  $[0, 1]^2 \rightarrow [0, 1]$  est définie par les caractéristiques suivantes :

- (i)  $C(u, 0) = C(0, u) = 0 \forall u \in [0, 1]$  : on dit que  $C$  est grounded,
- (ii)  $C(u, 1) = C(1, u) = u \forall u \in [0, 1]$  : les marges des distributions marginales sont des marges uniformes,
- (iii)  $C$  est 2-increasing :  $C(v_1, v_2) - C(v_1, u_2) - C(u_1, v_2) + C(u_1, u_2) \geq 0 \forall (u_1, u_2) \in [0, 1]^2, (v_1, v_2) \in [0, 1]^2$  tel que  $0 \leq u_1 \leq v_1 \leq 1$  et  $0 \leq u_2 \leq v_2 \leq 1$ .

Soient  $U_1$  et  $U_2$  deux variables aléatoires uniformes sur  $[0, 1]$ <sup>1</sup>, alors on a  $C(u_1, u_2) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2) \forall (u_1, u_2) \in [0, 1]^2$ . Cette définition assure donc que la copule est une distribution de probabilité avec des marges uniformes.

---

1. Rappelons qu'une loi uniforme  $U$  sur  $[0, 1]$  a pour fonction de répartition :

$$\mathbb{P}(U \leq u) = \begin{cases} 0 & \text{si } u \leq 0 \\ u & \text{si } 0 \leq u \leq 1 \\ 1 & \text{si } u \geq 1 \end{cases}$$

**Exercice 3.1.** Démontrer que la fonction  $C^\perp(u_1, u_2) = u_1 u_2$  avec  $(u_1, u_2) \in [0, 1]^2$  est bien une copule. Elle se dénomme la copule produit.

## 3.2 Propriétés

### 3.2.1 Théorème de Sklar

Une copule est donc déterminée soit *ex-nihilo* à partir de la définition 3.1, soit à l'aide d'une loi bivariée pré-existante.

Dans ce cas, il est fait appel au théorème de Sklar du nom de celui qui a introduit le concept *copule* en 1959. Ce théorème précise le lien défini par la copule  $C$ , déterminée à partir de la distribution jointe  $F$ , entre les fonctions de répartition marginales univariées  $F_1$  et  $F_2$  et la distribution bivariée  $F$ .

**Théorème 3.1** (Sklar). Soit  $F$  une distribution bivariée de marges  $F_1$  et  $F_2$ . La copule  $C$  associée à  $F$  s'écrit :

$$\begin{aligned} C(u_1, u_2) &= C(F_1(x_1), F_2(x_2)) \\ &= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \\ &= F(x_1, x_2) \end{aligned}$$

$C$  est unique lorsque les marges  $F_1$  et  $F_2$  sont continues.

La densité  $f$  d'une loi bivariée peut s'écrire aussi en fonction de la densité  $c$  de la copule associée et des densités des marginales  $f_1$  et  $f_2$  :

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \times f_1(x_1) \times f_2(x_2)$$

Dans la suite, nous illustrerons les exemples de copules à l'aide de représentations de leur densité en 3 dimensions. Nous y ajouterons les représentations de la densité, en 3 dimensions ou en 2 dimensions à partir des courbes de niveaux, d'une loi bivariée composée de marges prédéfinies et de la copule que l'on souhaite observer.

Précisons que  $c(u_1, u_2) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2)$ . Dans le cas de la copule produit  $C^\perp(u_1, u_2) = u_1 u_2$  par exemple, nous avons  $c^\perp(u_1, u_2) = 1$ . Ainsi, cette copule caractérise l'indépendance entre deux variables aléatoires puisque la densité bivariée d'un vecteur aléatoire  $(X_1, X_2)$  dont la structure de dépendance est déterminée par la copule produit s'écrit :

$$f(x_1, x_2) = f_1(x_1) \times f_2(x_2)$$

En outre, la caractéristique *2-increasing* d'une copule se résume à démontrer la positivité de sa densité, lorsque celle-ci existe.

**Exercice 3.2.** Soit la distribution logistique bivariée de Gumbel définie sur  $\mathbb{R}^2$  par  $F(x_1, x_2) = (1 + e^{-x_1} + e^{-x_2})^{-1}$ . Extraire la fonction copule à l'aide du théorème de Sklar.

### 3.2.2 Bornes de Fréchet

Définissons tout d'abord deux copules importantes :

1. la copule minimum qui a pour expression  $C^-(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$  avec  $(u_1, u_2) \in [0, 1]^2$ ,
2. la copule maximum qui s'écrit  $C^+(u_1, u_2) = \min(u_1, u_2)$  avec  $(u_1, u_2) \in [0, 1]^2$ .

Les copules minimum  $C^-$  et maximum  $C^+$  sont essentielles puisqu'elles définissent les copules extrémales de toute copule, que l'on appelle les bornes de Fréchet.

**Propriété 3.1.** Pour toute copule  $C$ , nous avons :

$$C^-(u_1, u_2) \leq C(u_1, u_2) \leq C^+(u_1, u_2)$$

### 3.2.3 Théorème d'invariance

L'un des théorèmes essentiels à la théorie des copules est celui de l'invariance par transformations strictement croissantes.

**Théorème 3.2** (Invariance). *Soient deux variables aléatoires continues  $X_1$  et  $X_2$  de marges  $F_1$  et  $F_2$  et de copule associée à la distribution  $F$  du vecteur aléatoire  $X = (X_1, X_2)$  que nous notons  $C_{\langle X_1, X_2 \rangle}$ . Si  $h_1$  et  $h_2$  sont deux fonctions strictement croissantes sur  $\text{Im}(X_1)$  et  $\text{Im}(X_2)$  respectivement, alors :*

$$C_{\langle h_1(X_1), h_2(X_2) \rangle} = C_{\langle X_1, X_2 \rangle}$$

La copule est invariante par transformations strictement croissantes des variables aléatoires.

*Démonstration.* Nous notons respectivement  $F$  et  $G$  les distributions des vecteurs aléatoires  $(X_1, X_2)$  et  $(h_1(X_1), h_2(X_2))$ .

Les marges de  $G$ , notées  $G_1$  et  $G_2$ , peuvent s'écrire en fonction des marges  $F_1$  et  $F_2$  de  $F$ , avec  $i \in \{1, 2\}$  :

$$G_i(x_i) = \mathbb{P}[h_i(X_i) \leq x_i] = \mathbb{P}[X_i \leq h_i^{-1}(x_i)] = F_i(h_i^{-1}(x_i))$$

car  $h_i$  est une fonction strictement croissante. Ainsi :

$$G_i^{-1}(u_i) = [F_i(h_i^{-1}(u_i))]^{-1} = h_i(F_i^{-1}(u_i)) \Leftrightarrow F_i^{-1}(u_i) = h_i^{-1}(G_i^{-1}(u_i)) \quad (3.1)$$

car  $(u \circ v)^{-1} = v^{-1} \circ u^{-1}$ .

Nous en déduisons, en utilisant le théorème de Sklar, que :

$$\begin{aligned} C_{\langle h_1(X_1), h_2(X_2) \rangle}(u_1, u_2) &= G(G_1^{-1}(u_1), G_2^{-1}(u_2)) \\ &= \mathbb{P}[h_1(X_1) \leq G_1^{-1}(u_1), h_2(X_2) \leq G_2^{-1}(u_2)] \\ &= \mathbb{P}[X_1 \leq h_1^{-1}(G_1^{-1}(u_1)), X_2 \leq h_2^{-1}(G_2^{-1}(u_2))] \\ &= \mathbb{P}[X_1 \leq F_1^{-1}(u_1), X_2 \leq F_2^{-1}(u_2)] \quad \text{d'après (3.1)} \\ &= F(F_1^{-1}(u_1), F_2^{-1}(u_2)) \\ &= C_{\langle X_1, X_2 \rangle}(u_1, u_2) \end{aligned}$$

□

### 3.2.4 Dépendance de queue

Définissons enfin un concept essentiel : la dépendance de queue pour une copule. Cette dernière mesure la probabilité de réalisations extrêmes simultanées.

L'indicateur de dépendance de queue d'une copule se déduit des probabilités conditionnelles suivantes :

$$\begin{aligned} \mathbb{P}[U_1 \leq u_1 | U_2 \leq u_2] &= \frac{\mathbb{P}[U_1 \leq u_1, U_2 \leq u_2]}{\mathbb{P}[U_2 \leq u_2]} = \frac{C(u_1, u_2)}{u_2} \\ \mathbb{P}[U_1 > u_1 | U_2 > u_2] &= \frac{\mathbb{P}[U_1 > u_1, U_2 > u_2]}{\mathbb{P}[U_2 > u_2]} \\ &= \frac{\mathbb{P}[U_1 \leq 1, U_2 > u_2] - \mathbb{P}[U_1 \leq u_1, U_2 > u_2]}{\mathbb{P}[U_2 \leq 1] - \mathbb{P}[U_2 \leq u_2]} \\ &= \frac{\mathbb{P}[U_1 \leq 1, U_2 \leq 1] - \mathbb{P}[U_1 \leq 1, U_2 \leq u_2] - \mathbb{P}[U_1 \leq u_1, U_2 > u_2]}{1 - u_2} \\ &= \frac{1 - u_2 - \mathbb{P}[U_1 \leq u_1, U_2 \leq 1] + \mathbb{P}[U_1 \leq u_1, U_2 \leq u_2]}{1 - u_2} \\ &= \frac{1 - u_1 - u_2 + C(u_1, u_2)}{1 - u_2} \end{aligned}$$

Nous pouvons alors définir les dépendances de queue à gauche et à droite.

**Définition 3.2** (Dépendance de queue à gauche). *Une copule  $C$  a une dépendance de queue à gauche (lower tail dependence) si :*

$$\lambda_L = \lim_{u \rightarrow 0^+} \frac{C(u, u)}{u}$$

existe et  $\lambda_L \in (0, 1]$ . Si  $\lambda_L = 0$  alors elle n'a pas de dépendance de queue à gauche.

**Définition 3.3** (Dépendance de queue à droite). *Une copule  $C$  a une dépendance de queue à droite (upper tail dependence) si :*

$$\lambda_U = \lim_{u \rightarrow 1^-} \frac{1 - 2u + C(u, u)}{1 - u}$$

existe et  $\lambda_U \in (0, 1]$ . Si  $\lambda_U = 0$  alors elle n'a pas de dépendance de queue à droite.

**Exercice 3.3.** Déterminer si la copule logistique de Gumbel a des dépendances de queue à gauche et à droite.

### 3.3 Exemples de copules

#### 3.3.1 Copule Gaussienne

**Définition 3.4** (Copule Gaussienne). *La copule Gaussienne bivariée est définie de la façon suivante :*

$$C(u_1, u_2 ; \rho) = \Phi_\rho (\Phi^{-1}(u_1), \Phi^{-1}(u_2))$$

$\rho$  est le coefficient de corrélation et  $\Phi_\rho$  la distribution normale bivariée standard de matrice de corrélation fonction de  $\rho$ .

Les Figures 3.1 et 3.2 permettent de visualiser la densité de la copule Gaussienne pour différentes valeurs du coefficient de corrélation linéaire, ainsi que la densité d'une loi gaussienne bivariée, qui est, rappelons-le, composée de marges gaussiennes et d'une copule Gaussienne.

Nous avons l'expression suivante du tau de Kendall en fonction du coefficient  $\rho$  de la copule Gaussienne :

$$\tau = \frac{2}{\pi} \arcsin \rho$$

Cette copule n'a pas de dépendance de queue (sauf pour une corrélation parfaite de 1) et ne permet donc pas de corréler des valeurs extrêmes.

Modéliser la structure de dépendance d'un échantillon par une copule Gaussienne est cohérent avec la mesure de cette dépendance par le coefficient de corrélation linéaire. Mais lorsque l'on souhaite modéliser une dépendance non linéaire ou entre événements extrêmes, nous allons faire appel à d'autres copules que nous explicitons dans la suite.

#### 3.3.2 Copule Student

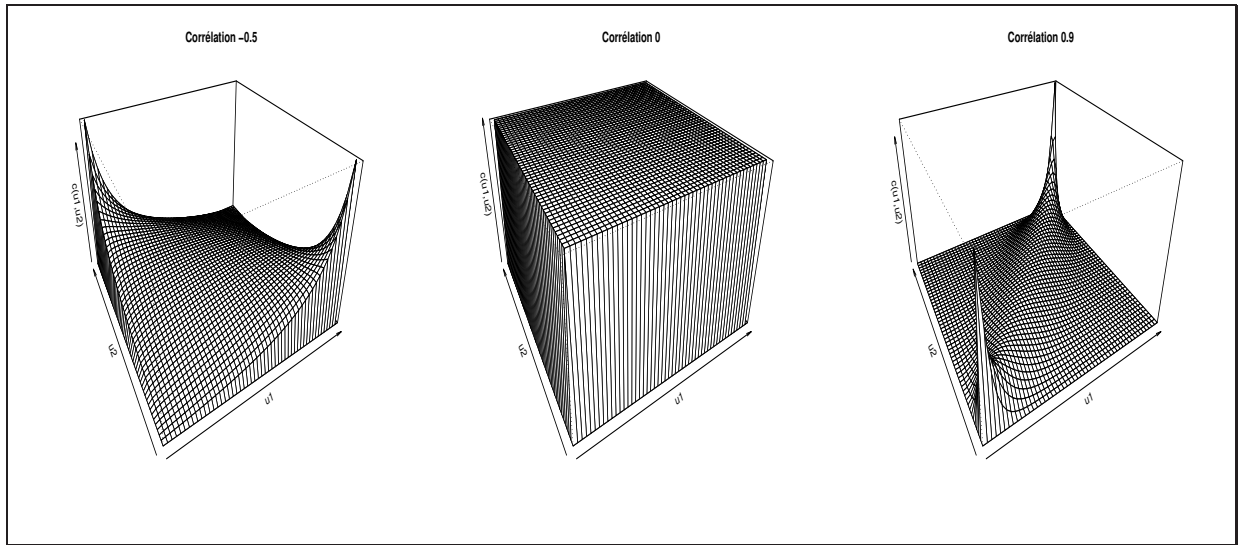
La copule Student est extraite de la même manière que la copule Gaussienne mais cette fois-ci à partir de la distribution de Student bivariée.

**Définition 3.5** (Copule Student). *La copule Student bivariée est définie de la façon suivante :*

$$C(u_1, u_2 ; \rho, k) = T_{\rho, k} (T_k^{-1}(u_1), T_k^{-1}(u_2))$$

avec  $\rho$  le coefficient de corrélation et  $T_{\rho, k}^{-1}$  la distribution de Student bivariée standard, de matrice de corrélation fonction de  $\rho$ , et de degré de liberté  $k$ .

Dans le cas où le coefficient de corrélation est différent de -1, la copule Student présente des dépendances de queues à droite et à gauche.

FIGURE 3.1 – Densités de trois copules Gaussiennes bivariées pour différentes valeurs de  $\rho$ 

### 3.3.3 Copules Archimédiennes

**Définition 3.6** (Copule Archimédienne). *Les copules Archimédiennes sont définies de la manière suivante :*

$$C(u_1, u_2) = \begin{cases} \varphi^{-1}(\varphi(u_1) + \varphi(u_2)) & \text{si } \varphi(u_1) + \varphi(u_2) \leq \varphi(0) \\ 0 & \text{sinon} \end{cases}$$

avec  $\varphi$  vérifiant  $\varphi(1) = 0$ ,  $\varphi'(u) < 0$  et  $\varphi''(u) > 0$  pour tout  $0 \leq u \leq 1$ .  $\varphi$  est appelée la fonction génératrice de la copule.

Le tau de Kendall  $\tau$  est égal pour les copules Archimédiennes à :

$$\tau = 1 + 4 \int_0^1 \frac{\varphi(u)}{\varphi'(u)} du$$

Quelques exemples de copules Archimédiennes bivariées sont, en notant  $\tilde{u} = -\ln u$  :

Nom	Générateur	Copule bivariée
Clayton ( $\theta > 0$ )	$u^{-\theta} - 1$	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-1/\theta}$
Gumbel ( $\theta \geq 1$ )	$(-\ln u)^\theta$	$\exp(-(\tilde{u}_1^\theta + \tilde{u}_2^\theta)^{1/\theta})$
Frank ( $\theta \neq 0$ )	$-\ln \frac{e^{-\theta u} - 1}{e^{-\theta} - 1}$	$-\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$

TABLE 3.1 – Exemples de copules Archimédiennes bivariées

Pour la copule Gumbel, nous avons l'expression suivante du  $\tau$  de Kendall en fonction de  $\theta$  :

$$\tau = 1 - \frac{1}{\theta}$$

**Exercice 3.4.** La copule logistique de Gumbel est un cas particulier d'une copule Archimédienne. Laquelle ? Même question pour la copule produit.

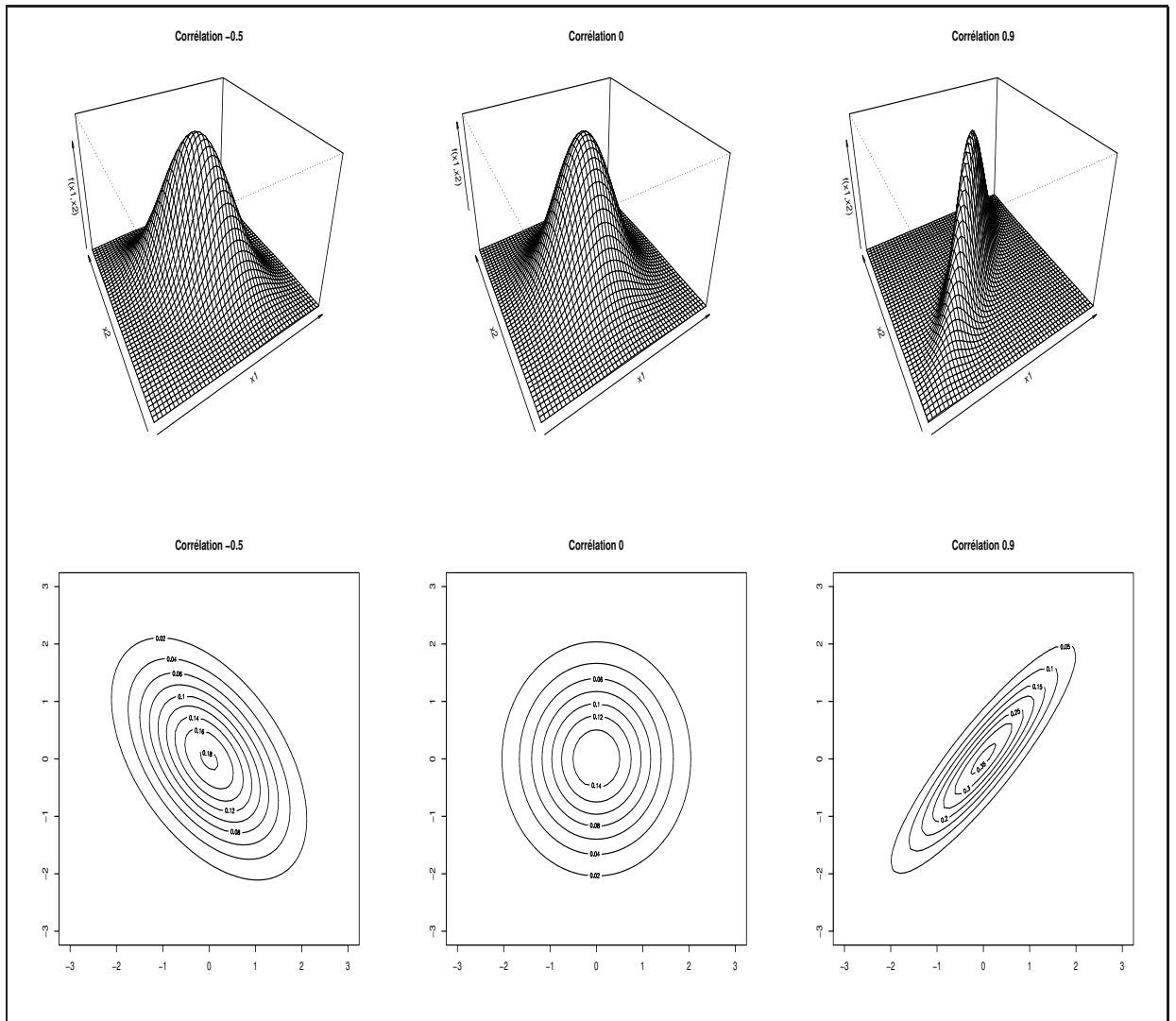


FIGURE 3.2 – Densités de trois lois bivariées construites à partir de marges gaussiennes centrées réduites et d'une copule Gaussienne pour différentes valeurs de  $\rho$ , avec leurs courbes de niveaux correspondantes

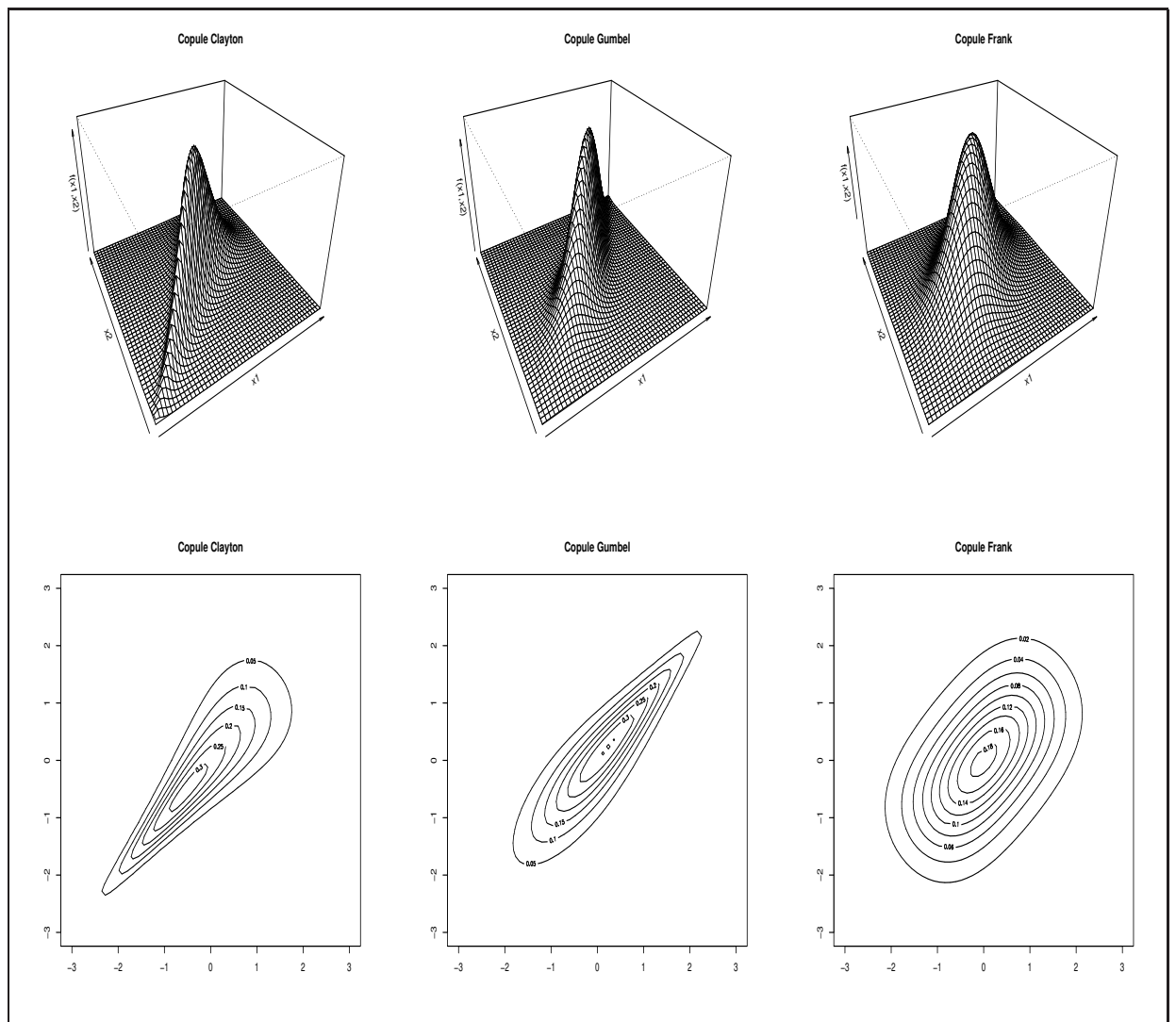


FIGURE 3.3 – Densités de trois lois bivariées construites à partir de marges gaussiennes centrées réduites et d'une copule Archimédienne de paramètre  $\theta = 3$ , avec leurs courbes de niveaux correspondantes

## CHAPITRE 4

# MISE EN ŒUVRE DES COPULES

Nous avons défini les copules, énoncé leurs propriétés fondamentales et décrit les copules les plus utilisées dans la pratique. Nous allons maintenant présenter deux méthodes essentielles aux statisticiens qui utilisent les copules : les méthodes de simulation et les méthodes d'inférence statistique.

### 4.1 Méthodes de simulation

Souhaiter simuler une copule bivariable  $C$  signifie simuler les arguments  $u_1$  et  $u_2$  de cette fonction tirés d'un vecteur aléatoire de loi uniforme  $U = (U_1, U_2)$  de distribution  $C$ . Ceci permet alors de déterminer par la suite la simulation d'un vecteur aléatoire  $X = (X_1, X_2)$  dont la structure de dépendance est définie par  $C$  et dont nous définissons des marges particulières  $F_1$  et  $F_2$ . Pour cela nous utilisons la transformation  $X = (F_1^{-1}(U_1), F_2^{-1}(U_2))$ .

Ceci est fondé sur la représentation copule de la distribution  $F$  donnée par le théorème 3.1 de Sklar :

$$C(u_1, u_2) = F(F_1^{-1}(u_1), F_2^{-1}(u_2))$$

Toute la difficulté est alors de simuler des nombres aléatoires issus de  $C$ . La simulation de la copule produit est relativement simple : on simule des réalisations de loi uniforme sur  $[0, 1]$  de manière indépendante. Mais pour la majorité des copules, cela s'avère plus complexe. Nous proposons alors 2 méthodes différentes selon la forme de la copule.

#### 4.1.1 Méthode des distributions

Dans le cas où la distribution  $F$  (générée par la copule  $C$ ) est plus facile à simuler que la copule  $C$ , nous utilisons la méthode des distributions. Nous avons alors :

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

Simuler les réalisations du vecteur aléatoire  $U = (U_1, U_2)$  revient à :

1. simuler les réalisations du vecteur aléatoire  $X = (X_1, X_2)$  de distribution  $F$ , puis à
2. appliquer la transformation  $U = (F_1(X_1), F_2(X_2))$ .

Ainsi, si nous voulons simuler une copule Gaussienne, nous utilisons cette méthode. En effet, nous pouvons simuler un vecteur gaussien bivariable  $X$  issu d'une loi Normale  $\mathcal{N}(0, \Psi)$ , avec  $\Psi$  la matrice de corrélation, en sachant que :

$$X = C_h Y$$

avec  $C_h$  la décomposition triangulaire inférieure de Cholesky qui vérifie  $C_h C_h' = \Psi$  et  $Y = (y_1, y_2)'$  un vecteur gaussien issu d'une loi  $\mathcal{N}(0, I)$  avec  $I$  la matrice identité.



Dans le cas bivarié, cette méthode de simulation revient à définir  $x_1 = y_1$  et déterminer  $x_2$  à l'aide de l'expression suivante, fonction du coefficient de corrélation linéaire  $\rho$  entre  $x_1$  et  $x_2$  :

$$x_2 = \rho y_1 + y_2 \sqrt{1 - \rho^2}$$

Il reste à appliquer aux réalisations  $(x_1, x_2)$  de ce vecteur gaussien  $X$  la transformation  $(F_1(x_1), F_2(x_2))$  avec  $F_1$  et  $F_2$  des marges gaussiennes standard, pour obtenir les réalisations  $(u_1, u_2)$  de la copule Gaussienne.

**Exercice 4.1.** *Etudier le fichier Excel dédié à la Copule Gaussienne sur ma page GitHub pour comprendre les différentes étapes de simulation d'une copule Gaussienne à partir de la méthode des distributions.*

Nous pouvons appliquer aussi la méthode des distributions à la simulation de la copule Student. Nous exploitons pour cela la relation entre un vecteur aléatoire  $Z$  de loi de Student  $T_{\Psi, k}$  et un vecteur aléatoire  $X$  issu d'une loi Normale  $\mathcal{N}(0, \Psi)$  :

$$Z = \frac{X}{\sqrt{\frac{\chi_k^2}{k}}}$$

#### 4.1.2 Méthode des distributions conditionnelles

Lorsque la copule n'est pas issue d'une distribution bivariée, il est impossible d'utiliser la méthode de simulation précédente. Une alternative est de recourir à la méthode des distributions conditionnelles qui simule directement la copule  $C$  et dont l'algorithme est le suivant :

1. Simuler 2 variables aléatoires uniformes  $v_1$  et  $v_2$ .
2. Imposer  $u_1 = v_1$ .
3. Déterminer  $u_2$  à l'aide de la distribution conditionnelle de la copule  $C$  sachant  $u_1$ . Cette dernière se définit de la manière suivante :

$$\begin{aligned} C_{2|1}(u_1, u_2) &= \mathbb{P}[U_2 \leq u_2 | U_1 = u_1] \\ &= \lim_{\Delta u \rightarrow 0^+} \frac{\mathbb{P}[U_2 \leq u_2, u_1 \leq U_1 \leq u_1 + \Delta u]}{\mathbb{P}[u_1 \leq U_1 \leq u_1 + \Delta u]} \\ &= \lim_{\Delta u \rightarrow 0^+} \frac{C(u_1 + \Delta u, u_2) - C(u_1, u_2)}{\Delta u} \\ &= \partial_1 C(u_1, u_2) \end{aligned}$$

Pour déterminer  $u_2$ , il nous reste à inverser la distribution conditionnelle et à y *plugger*  $v_2$ . Ainsi, nous avons la forme générale de l'inverse de la distribution conditionnelle suivante :

$$C^{-1}(u; u_1) = \{u_2 : C_{2|1}(u_1, u_2) = u\}$$

Au final, nous obtenons :  $u_2 = C^{-1}(v_2; u_1)$ .

Voici un exemple de cette méthode de simulation appliquée à la copule Frank.

La distribution de cette copule est :

$$C(u_1, u_2; \theta) = -\frac{1}{\theta} \ln \left( 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$$

avec  $(u_1, u_2) \in [0, 1]^2$  et  $\theta \neq 0$ .

Nous déterminons la distribution conditionnelle suivante :

$$\begin{aligned} C_{2|1}(u_1, u_2; \theta) &= \frac{\partial}{\partial u_1} C(u_1, u_2) = -\frac{1}{\theta} \left( \frac{-\theta e^{-\theta u_1} (e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right) \left( 1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)^{-1} \\ &= \frac{e^{-\theta u_1} (e^{-\theta u_2} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)} \end{aligned}$$

Comme :

$$\begin{aligned} C_{2|1}(u_1, u_2; \theta) = u &\Leftrightarrow \frac{e^{-\theta u_1} (e^{-\theta u_2} - 1)}{(e^{-\theta} - 1) + (e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)} = u \\ &\Leftrightarrow (e^{-\theta u_2} - 1) (e^{-\theta u_1} - u(e^{-\theta u_1} - 1)) = u(e^{-\theta} - 1) \\ &\Leftrightarrow e^{-\theta u_2} = 1 + \frac{u(e^{-\theta} - 1)}{u + (1 - u)e^{-\theta u_1}} \\ &\Leftrightarrow u_2 = -\frac{1}{\theta} \ln \left( 1 + \frac{u(e^{-\theta} - 1)}{u + (1 - u)e^{-\theta u_1}} \right) \end{aligned}$$

Nous obtenons alors :

$$u_2 = C^{-1}(v_2; u_1) = -\frac{1}{\theta} \ln \left( 1 + \frac{v_2(e^{-\theta} - 1)}{v_2 + (1 - v_2)e^{-\theta u_1}} \right)$$

## 4.2 Méthodes d'inférence statistique

Nous supposons avoir un échantillon de couples de données dont nous tentons d'estimer la fonction de dépendance. Nous présentons deux méthodes d'estimation classiques : la méthode des moments et la méthode du maximum de vraisemblance. Puis, nous déterminerons des outils graphiques d'ajustement de la copule estimée à l'échantillon des données bivariées.

### 4.2.1 Méthode des moments

Dans le cas des copules, la méthode des moments est fondée sur les relations existant entre le  $\tau$  de Kendall et les paramètres des copules.

Prenons l'exemple de l'estimation par la méthode des moments du paramètre  $\theta$  de la copule Gumbel. Rappelons la relation que nous avons entre  $\theta$  et le  $\tau$  de Kendall :

$$\tau = 1 - \frac{1}{\theta}$$

Nous en déduisons que l'estimation  $\hat{\theta}$  est égale à :

$$\hat{\theta} = \frac{1}{1 - \hat{\tau}}$$

### 4.2.2 Méthode du maximum de vraisemblance

Nous pouvons aussi utiliser la méthode du maximum de vraisemblance.

Prenons l'exemple de la copule Student. La densité de la copule Student est calculée, à partir de la matrice de corrélation  $\Psi$  et du degré de liberté  $k$ , sur les statistiques de rang de l'échantillon  $u_1$  et  $u_2$  :

$$c(u_1, u_2; \Psi, k) = |\Psi|^{-\frac{1}{2}} \frac{\Gamma(\frac{k+2}{2}) \Gamma(\frac{k}{2})}{[\Gamma(\frac{k+1}{2})]^2} \left( 1 + \frac{1}{k} y' \Psi^{-1} y \right)^{-\frac{k+2}{2}} \prod_{i=1}^2 \left( 1 + \frac{y_i^2}{k} \right)^{\frac{k+1}{2}}$$

avec  $y = (y_1, y_2)' = (T_k^{-1}(u_1), T_k^{-1}(u_2))'$  un vecteur de réalisations d'une variable aléatoire de Student de degrés de liberté  $k$ . À partir de cette densité, nous pouvons déterminer la log-vraisemblance de l'échantillon des couples  $(u_1, u_2)$  comme somme des log-densités associées à chacun des couples. Nous la maximisons en  $k$ , à l'aide de méthodes numériques classiques et en estimant au préalable la matrice de corrélation de manière empirique : ceci permettra une meilleure convergence de l'optimisation.

### 4.2.3 Outils graphiques d'ajustement

Pour l'évaluation de l'ajustement d'une copule à la structure de dépendance d'un échantillon, nous pouvons utiliser les outils graphiques suivants :

1. la comparaison des densités empirique (estimée sur l'échantillon non paramétriquement par la méthode du noyau) et théorique en 3 dimensions ou sous forme de courbes de niveaux,
2. le dépendogramme,
3. et le *Kendall plot* ou *K-plot* : cf. [Genest et Boies \(2003\)](#).

Ces graphes permettent notamment d'observer si la dépendance de queue est présente dans les données et si elle est bien modélisée par la copule choisie. Nous allons préciser les deux derniers outils graphiques.

#### Dépendogramme

Le dépendogramme représente la structure de dépendance sous la forme du nuage de points des marges uniformes  $(u_1, u_2)$  extraites de l'échantillon ou encore de simulations d'une copule théorique. Précisons que les couples  $(u_1, u_2)$  issues de l'échantillon forment ce que nous appelons la copule empirique. Cette dernière est définie par les statistiques de rang issues de l'échantillon. Nous comparerons le dépendogramme de cette copule empirique à celui d'autres copules théoriques estimée sur l'échantillon.

Le dépendogramme permet d'observer la dépendance des réalisations issues de l'échantillon. Dans les queues plus précisément, il sera utile d'analyser si la dépendance est forte et donc s'il est nécessaire d'estimer sur notre échantillon une copule avec une dépendance de queue.

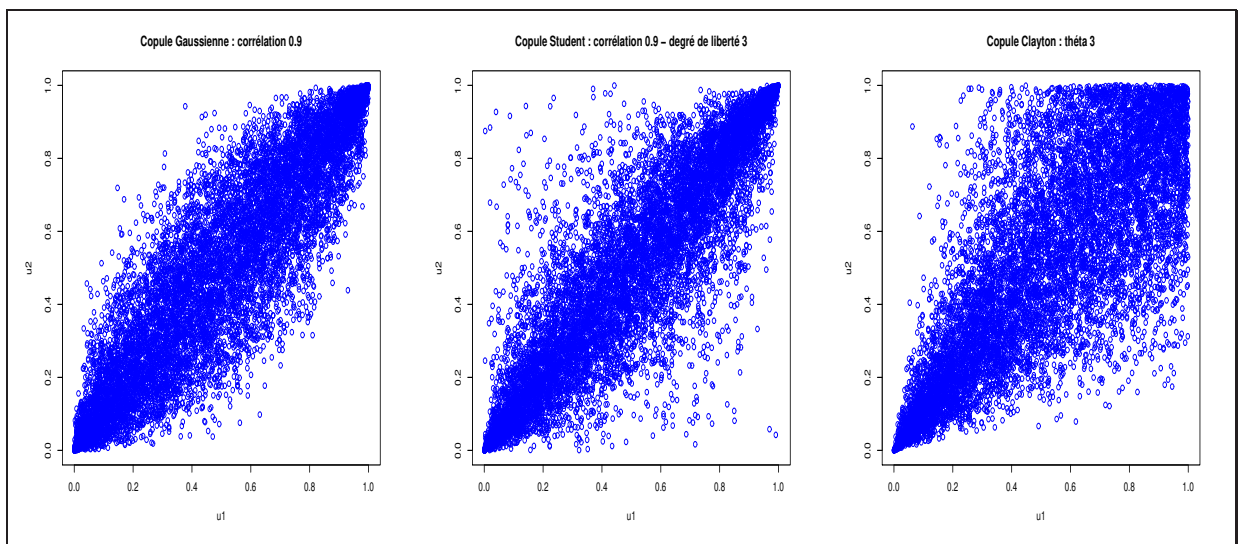


FIGURE 4.1 – Dépendogrammes pour des données simulées à partir de 3 copules différentes

### Kendall plot

Le Kendall plot va permettre une comparaison plus directe entre la copule empirique et la copule théorique. L'algorithme de construction d'un Kendall plot est le suivant :

1. Calculer pour chaque paire des observations des rangs de l'échantillon  $(u_{1i}, u_{2i})$  la fonction  $H_i$  suivante :

$$H_i = \frac{1}{n-1} \text{card}\{j \neq i : u_{1j} \leq u_{1i}, u_{2j} \leq u_{2i}\}$$

pour  $1 \leq i \leq n$  avec  $n$  la longueur de l'échantillon. Il faut ensuite les ordonner pour obtenir  $H_{(1)} \leq \dots \leq H_{(n)}$ . Nous obtenons ainsi la partie empirique du K-plot issue de l'échantillon.

2. Il faut la comparer à celle correspondant à la copule théorique que nous estimons sur l'échantillon. Pour cela, il faut déterminer  $H_{i:n}^{\text{th}} = E\left(H_{(i)}^{\text{th}}\right)$  avec :

$$H_i^{\text{th}} = \frac{1}{n-1} \text{card}\{j \neq i : u_{1j}^{\text{th}} \leq u_{1i}^{\text{th}}, u_{2j}^{\text{th}} \leq u_{2i}^{\text{th}}\}$$

pour chaque couple  $(u_{1i}^{\text{th}}, u_{2i}^{\text{th}})$  issu de la copule théorique. On détermine ces couples à l'aide de simulations Monte Carlo de la copule théorique. Ainsi, on simule  $n$  couples réalisations de la copule théorique  $m$  fois, on détermine  $H_i^{\text{th}}$  pour chaque  $m$ -ième simulation, que l'on ordonne ensuite, et enfin on calcule l'espérance des  $m$   $H_{(i)}^{\text{th}}$  pour chaque  $i$ . Et l'on obtient  $H_{i:n}^{\text{th}}$ .

3. Il nous reste à grapher les paires  $(H_{(i)}, H_{i:n}^{\text{th}})$  pour obtenir le K-plot.

Plus le Kendall plot se rapproche d'une droite, plus l'ajustement entre la structure de dépendance de l'échantillon et la copule estimée sur ce même échantillon est bon.

Il existe aussi des tests statistiques d'ajustement de type Cramér-von Mises : cf. [Genest et al. \(2009\)](#).

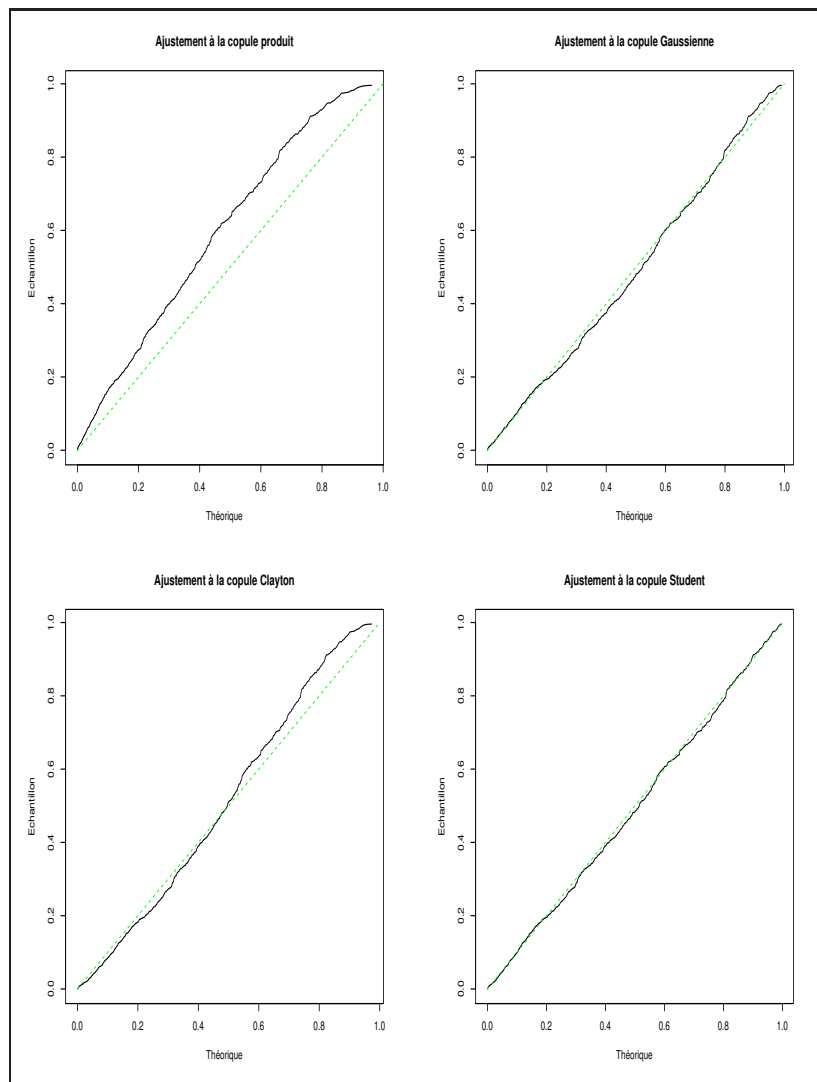


FIGURE 4.2 – Kendall plots comparant un échantillon de données simulées à partir d'une copule Student (corrélation 0.5 - degré de liberté 3) à plusieurs copules estimées sur ce même échantillon

# CONCLUSION

Nous espérons que cette partie a permis d'appréhender précisément les tenants et les aboutissants de la modélisation de la structure de dépendance d'un échantillon par les copules.

Malgré sa relative simplicité mathématique, l'outil copule n'est pas toujours utilisé dans toutes ses potentialités. En Finance, par exemple, de nombreux travaux sont encore à réaliser pour appliquer cet objet rigoureusement et de manière adéquate dans des domaines aussi variés que le pricing de produits structurés ou encore la modélisation des risques multiples.

# **RISQUES MULTIPLES 2**

**APPLICATION À LA MESURE DU RISQUE DE CRÉDIT**

# INTRODUCTION

La gestion du risque pris par les banques est devenue fondamentale. Et un élément essentiel est la mesure du risque de contagion ou risque multiple. Ce risque survient lorsqu'un portefeuille est composé de plusieurs actifs qui ont une structure de dépendance forte. Les différents risques réglementés par le Comité de Bâle pour les banques sont concernés par la présence de structure de dépendance. Nous allons dans cette partie nous intéresser plus particulièrement au risque de crédit.

Court rappel : alors que le risque de marché est défini comme le risque de perte lié aux variations des marchés, le risque de crédit est lié aux variations de la qualité de la signature d'un émetteur. Ainsi, un prêteur X est soumis via l'achat d'une obligation ou l'émission d'un prêt émis par Y au risque que ce dernier ne rembourse pas la capital investi à la maturité de l'obligation ou du prêt. Si Y ne peut pas faire face à ses engagements, cela signifie que Y a fait *défaut*. Les exemples historiques les plus marquants de défaut sont ceux d'Enron, Parmalat, Lehman Brothers, et plus récemment les ménages *subprimes* américains. Tout l'enjeu de cette partie est de modéliser la défaillance d'un émetteur de dette plus ou moins risquée dans le but de mesurer le risque de perte provoqué par cette potentielle défaillance.

Le risque de défaut se mesure à travers différents événements que peut subir un émetteur : l'évolution possible de sa notation ou *rating* (*upgrade* ou *downgrade*) émis par les agences de notation telles Moody's et Standard & Poor's par exemple, la modification de la cotation de l'entreprise, la variation du spread de crédit, le défaut de paiement, la faillite. Ces événements peuvent influencer plus ou moins sur la qualité de la signature de l'émetteur, augmentant alors, dans le cas d'une dégradation de la signature, le risque de défaut de l'entreprise.

Outre la probabilité de défaillance, deux autres paramètres interviennent dans la mesure du risque de crédit : la perte en cas de défaut (en anglais *Loss Given Default* ou LGD) qui quantifie ce que perd le créancier lorsque l'entreprise fait défaut, et la dépendance entre les défauts des émetteurs, qui peut être modélisée à l'aide d'une copule. Ces 3 paramètres essentiels à la modélisation du risque de crédit vont nous permettre de mesurer le risque multiple résultant de la détention d'un portefeuille de créances.

L'objet de cette partie est donc d'étudier rigoureusement les différentes méthodologies pour appréhender la probabilité de défaut ainsi que les paramètres de LGD et de dépendance des défauts. Cela nous permettra de déterminer une mesure du risque multiple d'un portefeuille de crédit à l'aide de la CreditVaR.



## CHAPITRE 5

# MODÉLISATION DE LA PROBABILITÉ DE DÉFAUT

Pour modéliser le risque de crédit, un élément essentiel est la probabilité de défaut. Trois possibilités s'offrent à nous :

- l'approche par les ratings et par les évolutions en terme de notations des émetteurs : à chaque *rating* est associée une probabilité de défaut d'autant plus élevée que le *rating* est mauvais,
- l'approche *structurelle* modélisant le défaut en prenant en compte le bilan comptable de l'entreprise et le processus financier conduisant au défaut,
- et l'approche *forme réduite* des modèles à intensité, modélisant le défaut sans référence au processus financier sous-jacent. Le défaut est imprévisible et sa loi est gouvernée par un processus stochastique.

### 5.1 Approche par les ratings

Cette méthodologie s'appuie sur le constat qu'un événement de défaut est consécutif à une dégradation progressive des ratings de l'émetteur. En outre, elle fait partie des fondements des réglementations du Comité de Bâle. Deux approches sont alors possibles pour déterminer une PD : la première élabore des notations à l'aide d'opinions d'experts et la seconde utilise les techniques quantitatives de *scoring*.

#### 5.1.1 Notation des agences

Les ratings des agences de notation sont déterminés sur la base d'audits comptables et financiers indépendants, objectifs, crédibles et transparents. Ces ratings sont en général présentés sous la forme d'une ou plusieurs lettres symbolisant la qualité de crédit de l'émetteur. L'ensemble de ces notations forme l'échelle de notation. Les ratings, qui s'appliquent aussi bien à l'émetteur qu'à l'instrument de dette émis, varient selon l'horizon auquel ils s'appliquent. Nous distinguons des ratings court terme (d'horizon égal ou inférieur à 1 an) et des ratings de long terme. Les agences ont des échelles différentes mais un point commun essentiel est la distinction entre les deux catégories *investment grade* et *sub-investment grade* ou *speculative grade* (cf. tableau 5.1). La première catégorie rassemble des contreparties ayant une relative stabilité dans leur modèle de développement et un niveau de risque modéré. La seconde catégorie inclut des émetteurs aux caractéristiques financières incertaines dont la probabilité de faire défaut est plus élevée.

Les ratings sont issus de la compétitivité de la contrepartie, de la qualité de son management, de la politique suivie, de la situation de son marché, ainsi que de ratios financiers, dépendant aussi du secteur de la société. Ils font l'objet d'un suivi régulier afin de refléter au mieux l'évolution du client.

Un point important est le fait que le rating ne reflète pas la valorisation d'une dette mais le jugement sur la qualité de la signature d'une dette et la capacité de remboursement de l'émetteur.

	Moody's	Stanbard & Poor's
<i>investment grade</i>	Aaa	AAA
	Aa	AA
	A	A
	Baa	BBB
<i>speculative grade</i>	Ba	BB
	B	B
	Caa	CCC

TABLE 5.1 – Ratings de Moody's et Stanbard &amp; Poor's

Les probabilités de défaut correspondent à la moyenne des fréquences de défaut par rapport à l'ensemble des contreparties de même rating et peuvent être calculées en fonction de la zone géographique, du secteur, de la période, de l'horizon.

Les agences de notation fournissent aussi des matrices de migration indiquant pour un horizon donné la probabilité de migration qu'un émetteur ou instrument appartenant à une classe de rating migre dans une autre classe. Ces matrices de migration peuvent être utilisées (i) pour déterminer des PD à horizon long ainsi que (ii) pour estimer les PD de très bons ratings.

En effet, si ces matrices sont stables dans le temps, nous pouvons faire l'hypothèse qu'elles sont markoviennes, c'est-à-dire que la seule connaissance de l'information en  $t$  permet de déterminer les valeurs de la matrice en  $t + 1$ . Cette hypothèse est néanmoins très discutée (cf. de Servigny et al. (2006) pour une revue des débats).

Nous supposons que l'on dispose de 7 niveaux de rating distincts hors défaut soit 8 états pour les créances. On note l'espace état  $N = \{AAA, AA, A, BBB, BB, B, CCC, D\}$  dénombrable et fini avec  $D$  la classe des titres en défaut. On note  $P = P(0, 1) = (p_{ij})_{i,j \in N \times N}$  la matrice  $8 \times 8$  de transition à 1 an et nous remarquons que  $P$  a la forme suivante, avec en lignes les ratings en  $t$  et en colonnes les ratings en  $t + 1$  :

$$P = P(0, 1) = \begin{pmatrix} R & L \\ 0 & 1 \end{pmatrix}$$

avec  $R$  une matrice de taille  $7 \times 7$  de transition d'une classe de crédit à l'autre,  $L$  un vecteur  $7 \times 1$  de probabilités de défaut,  $0$  le vecteur  $1 \times 7$  de probabilités nulles de sortir de l'état de défaut, et la valeur 1 traduisant l'état absorbant de défaut.  $P$  étant telle que  $0 \leq p_{ij} < \infty \forall i, j$  et  $\sum_j p_{ij} = 1 \forall i$ , la matrice de transition est qualifiée de *stochastique*.

Les probabilités de transition sont modélisées à l'aide d'une chaîne de Markov homogène à états finis.

**Définition 5.1** (Chaîne de Markov). *Un processus  $X = (X_n)_n$  est une chaîne de Markov homogène à états finis discrets  $N$  de matrice de transition  $P$  si on a :*

$$\mathbb{P}[X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i] = \mathbb{P}[X_{n+1} = j | X_n = i] = p_{ij}$$

**Propriété 5.1.** *L'équation de Chapman-Kolmogorov implique que :*

$$P(0, m + n) = P(0, n)P(0, m)$$

ou encore :

$$P(0, n) = P(0, 1)^n = P^n = (p_{ij}^{(n)})_{i,j}$$

avec  $P(0, n)$  la matrice de transition déterminée aujourd'hui pour un horizon de  $n$  années, et  $p_{ij}^{(n)}$  l'élément  $(i, j)$  de la matrice  $P^n$  (probabilité de passer de  $i$  à  $j$  en  $n$  années). Nous avons aussi :

$$\mathbb{P}[X_{n+m} = j | X_m = i] = p_{ij}^{(n)}$$

Les probabilités de défaut à n'importe quel horizon (et donc aussi pour un horizon long) peuvent être déterminées grâce à la modélisation de la matrice de migration par chaîne de Markov homogène.

Pour les probabilités de défaut de très bons ratings, nous pouvons régulièrement observer que ces derniers ont des PD nulles, surtout pour les *Low Default Portfolios* (LDP). Néanmoins, comme le précisent [Lando et Skødeberg \(2002\)](#), si une contrepartie AAA ne fait jamais défaut mais qu'est observée sur une contrepartie X une transition entre AAA et AA et que sur une autre contrepartie Y notée AA est observée une transition vers le défaut, alors la modélisation par chaîne de Markov permet de capter une PD non nulle entre la note AAA et le défaut.

### 5.1.2 Scoring des émetteurs

En général, les ratings des agences sont faits sur des États et entreprises de grande taille. Concernant les petites et moyennes entreprises, les probabilités de défaut peuvent être estimées à l'aide de méthodes de scoring.

Les variables explicatives, issues d'informations comptables, sont des ratios économiques et financiers statistiquement significatifs pour différencier les entreprises susceptibles d'être en difficulté de celles qui le sont moins (performances, solvabilité, autonomie financière, etc.). Un *mapping* spécifique permet de relier le *scoring* continu et l'échelle de notation discrète.

Plusieurs techniques de *scoring* sont utilisées pour estimer la probabilité de défaut : l'analyse discriminante linéaire de Fisher pour la plus ancienne, la régression logistique pour la plus utilisée encore aujourd'hui, mais aussi des techniques de machine learning *black box* : nous renvoyons pour une revue de ces méthodes à [Lessmann et al. \(2015\)](#). Nous allons nous concentrer sur la régression logistique et nous donnons dans la suite un rappel de sa méthodologie (cf. [Davidson et Mackinnon \(1993\)](#) et [Thomas \(2000\)](#)).

Nous supposons que le défaut est modélisé par une variable dépendante  $y_i$  binaire, égale à 0 (pas de défaut), ou égale à 1 (défaut).

Considérons en premier lieu le modèle suivant :

$$y_i = x_i\beta + u_i$$

avec  $x_i$  un vecteur  $(1 \times K)$  de variables explicatives pour l'individu  $i$  associées au vecteur  $(K \times 1)$  de paramètres  $\beta$  et  $u_i$  le résidu.

L'hypothèse sous-jacente est que  $E(y_i|x_i) = x_i\beta$  et donc que la probabilité de défaut conditionnelle est une fonction linéaire des  $x_i$ .

Ce modèle à probabilité linéaire présente de nombreux inconvénients. La technique des Moindres Carrés Ordinaires (MCO) classique utilisée dans les modèles de régression à variable dépendante continue ne peut être appliquée.

Tout d'abord, rien ne contraint le modèle à ce que la probabilité estimée soit comprise entre 0 et 1.

Ensuite, les résidus de la régression ne peuvent prendre que 2 valeurs :  $1 - x_i\beta$  avec la probabilité  $x_i\beta$  et  $-x_i\beta$  avec la probabilité complémentaire  $1 - x_i\beta$ . Le cadre d'analyse des MCO considérant les résidus continus se prête donc mal à la modélisation de  $y_i$ .

Enfin, le calcul de la variance des résidus donne :

$$\begin{aligned} V(u_i) &= (u_i^2|y_i = 1) * P(y_i = 1) + (u_i^2|y_i = 0) * P(y_i = 0) \\ &= (1 - x_i\beta)^2 * (x_i\beta) + (-x_i\beta)^2 * (1 - x_i\beta) \\ &= x_i\beta(1 - x_i\beta) \end{aligned}$$

Le modèle est donc hétéroscédastique par construction, la variance des résidus dépendant des  $x_i$  et n'étant ainsi pas constante.

Il est donc nécessaire d'utiliser une technique alternative et pour cela nous allons introduire une variable latente  $y_i^*$  (inobservable) telle que :

$$\begin{cases} y_i = 1 & \text{si } y_i^* > c \\ y_i = 0 & \text{si } y_i^* \leq c \end{cases}$$

où  $y_i^* = x_i\beta + u_i$  et  $c$  correspond à la valeur seuil que nous normalisons à 0.

Nous considérons alors la réalisation de la variable dépendante binaire comme provenant d'une règle de décision s'écrivant :

$$\begin{cases} \mathbb{P}(y_i = 1) = \mathbb{P}(u_i > -x_i\beta) = 1 - F(-x_i\beta) = F(x_i\beta) \\ \mathbb{P}(y_i = 0) = \mathbb{P}(u_i \leq -x_i\beta) = F(-x_i\beta) = 1 - F(x_i\beta) \end{cases}$$

avec  $F$  une fonction de répartition symétrique autour de 0.

Il faut bien remarquer que dans la modélisation de la variable binaire, nous ne modélisons plus la variable en elle-même, mais les probabilités de réalisation des événements  $\{y_i = 1\}$  et  $\{y_i = 0\}$ .

Il nous faut définir  $F$  la fonction de répartition des résidus. Classiquement, nous avons le choix entre la loi logistique (modèle Logit), la loi normale (modèle Probit) et la loi de Gompertz (modèle Gompit). Nous ne verrons ici que le cas de la loi logistique, plus facile à manipuler, qui a pour fonction de répartition  $\Lambda$  :

$$\Lambda(x_i\beta) = \frac{\exp(x_i\beta)}{1 + \exp(x_i\beta)}$$

Nous estimons le modèle par maximum de vraisemblance. La vraisemblance s'écrit :

$$\mathcal{L}(y, x; \beta) = \prod_{i=1}^n \left[ \frac{1}{1 + \exp(x_i\beta)} \right]^{1-y_i} \left[ \frac{\exp(x_i\beta)}{1 + \exp(x_i\beta)} \right]^{y_i}$$

avec un échantillon de  $n$  individus. Et la log-vraisemblance a pour expression :

$$\log \mathcal{L} = - \sum_{i=1}^n [\log[1 + \exp(x_i\beta)] - y_i x_i \beta]$$

**Propriété 5.2.** Nous obtenons alors les estimations des probabilités de défaut, en notant  $p_i$  la probabilité que  $\{y_i = 1\}$  (apparition d'un défaut) :

$$\hat{p}_i = \frac{\exp(x_i\hat{\beta})}{1 + \exp(x_i\hat{\beta})} \Leftrightarrow \ln \frac{\hat{p}_i}{1 - \hat{p}_i} = x_i\hat{\beta}$$

La probabilité de l'événement  $\{y_i = 0\}$  est la probabilité complémentaire.

Précisons que les valeurs des paramètres estimés  $\hat{\beta}$  n'ont pas d'intérêt en soi, seul leur signe est une information utilisable : en effet, il indique si la variable associée influence la probabilité de défaut à la hausse ou à la baisse.

## 5.2 Modèle de Merton

De manière complémentaire à l'approche par les ratings, le recours aux modèles structurels (dont Merton est à l'origine) de valorisation de la firme a permis, à l'aide de nombreuses approximations empiriques, de dégager une estimation de la probabilité de défaut des entreprises cotées. L'approche structurelle a été initiée par Merton (1974) en 1974. Elle est fondée sur la modélisation de l'évolution du bilan, économique et non comptable, de l'entreprise. Le défaut intervient alors si l'émetteur de la dette est dans l'impossibilité d'honorer ses engagements. Le risque de défaut est modélisé en utilisant les principes de *pricing* des options dans le cadre de Black et Scholes (1973). Cette modélisation est largement utilisée par les praticiens : par exemple Moody's KMV (cf. Crosbie et Bohn (2003)) ou encore JP Morgan avec le modèle *Credit Grades* (cf. Finger (2002)).

### 5.2.1 Modélisation de la firme

L'hypothèse fondamentale est la modélisation de la dynamique de la valeur de la firme  $V$  par un mouvement Brownien géométrique sous la probabilité historique :

$$dV_t = V_t(\mu dt + \sigma dW_t) \quad (5.1)$$

avec  $\mu$  et  $\sigma$  les paramètres de tendance et de diffusion constants, et  $W_t$  un mouvement Brownien standard.

En outre, la valeur de la firme  $V_t$  en  $t$  (l'actif du bilan de la firme) est définie par la somme de la valeur actionnariale de la firme  $E_t$  et de la valeur de la dette  $D_t$  (somme représentant le passif du bilan) : ainsi, l'actif étant égal au passif,  $V_t = E_t + D_t$ . La valeur actionnariale se présente sous la forme d'actions et la dette sous la forme d'un zéro-coupon de maturité  $T$  et de nominal  $D$ . A maturité, il y a défaut si la valeur de la firme  $V_T$  est inférieure à la dette  $D$  que la firme doit payer aux créanciers. Ainsi, les détenteurs d'obligations reçoivent  $\min(V_T, D)$  et le reliquat est versé aux actionnaires :  $\max(V_T - D, 0)$ . Nous reconnaissons pour ce dernier versement le payoff d'un call européen. Le revenu des actionnaires s'assimile donc à celui d'une option d'achat d'échéance  $T$  et de prix d'exercice  $D$  (le remboursement du zéro-coupon à maturité). Étudions plus en détail le payoff de la dette  $\min(V_T, D)$ . Il est égal aussi à :

$$\min(V_T, D) = D - \max(D - V_T, 0)$$

La dette risquée se confond donc avec une dette sans risque (de même maturité  $T$  et de même nominal  $D$ ) à laquelle on retranche un put (option de vente) appelée *put-to-default* écrit sur la valeur de l'entreprise.

**Exercice 5.1.** Sous les hypothèses classiques de la théorie des options dans le cadre de Black et Scholes, écrire  $D_t$  sous la probabilité risque-neutre  $\mathbb{Q}$ , en notant  $P_t$  (respectivement  $C_t$ ) le prix du put (call) sur la valeur de la firme  $V_t$  de prix d'exercice  $D$  et de maturité  $T$ .

Indication : utiliser la formule de parité call-put  $C_t - P_t = V_t - Le^{-r(T-t)}$  et  $D_T = D - \max(D - V_T, 0)$ .

### 5.2.2 Probabilité de défaut

Si l'on passe à la probabilité risque-neutre  $\mathbb{Q}$ , l'équation différentielle stochastique (EDS) (5.1) devient :

$$dV_t = V_t(r dt + \sigma d\bar{W}_t) \quad (5.2)$$

avec  $r$  le taux sans risque constant et  $\bar{W}_t$  un mouvement Brownien standard sous  $\mathbb{Q}$ .

Et la solution de l'EDS est :

$$V_t = V_0 \exp \left[ \left( r - \frac{\sigma^2}{2} \right) t + \sigma \bar{W}_t \right]$$

où  $\bar{W}_t \sim \mathcal{N}(0, \sqrt{t})$ .

**Exercice 5.2.** La probabilité de défaut, notée  $PD$ , est définie de la manière suivante sous la probabilité risque-neutre  $PD = \mathbb{Q}[V_T \leq D]$ . Calculer cette probabilité de défaut  $PD$  en  $t$ .

On peut à partir de ce modèle déterminer le spread *implicite* de crédit, c'est-à-dire induit par le modèle, à partir de la formule suivante :

$$\text{spread}_t = -\frac{1}{T-t} \ln \left( \Phi(d_2) + \frac{\Phi(-d_1)}{L} \right)$$

avec  $\Phi(\cdot)$  la fonction de répartition de la loi Normale centrée et réduite,  $d_1 = \frac{1}{\sigma\sqrt{T-t}} \left[ \left( r + \frac{\sigma^2}{2} \right) (T-t) + \ln \frac{V_t}{D} \right]$

et  $d_2 = d_1 - \sigma\sqrt{T-t}$  et avec  $L = \frac{De^{-rT}}{V_0}$  le levier d'endettement (cf. Hull (2009) pour plus de précisions).

### 5.2.3 Estimation

Tout d'abord, le paramètre  $D$  peut être estimé à partir de l'évaluation de la dette de l'entreprise. Il se pose néanmoins des problèmes pour l'estimation des paramètres  $V_t$  et  $\sigma$  qui ne sont pas directement observables. Ils peuvent être estimés à partir des données historiques de la valeur de l'action  $E_t$ .

La valeur de l'action (call sur la valeur de la firme) est donnée par :

$$E_t = f(t, V_t) = V_t \Phi(d_1) - D e^{-r(T-t)} \Phi(d_2) \quad (5.3)$$

Supposons que le processus  $E_t$  suit une EDS de la forme  $dE_t = E_t[\bar{\mu}_E dt + \sigma_E d\bar{W}_t]$  sous  $\mathbb{Q}$ .

**Exercice 5.3.** Déterminer la volatilité  $\sigma_E$  du processus  $E_t$  en fonction de la volatilité du processus de la valeur de l'entreprise  $V_t$ .

Indication : utiliser le lemme d'Itô.

Nous avons présenté les différents enjeux sous-jacents à la modélisation structurelle de Merton (cf. aussi [Haworth \(2004\)](#)). Néanmoins, ce modèle présente des limites dont la principale est le fait que le défaut de l'émetteur ne peut intervenir qu'à la maturité de la dette. Les modèles de premier instant de passage résolvent ce problème. Nous ne développerons pas cette modélisation dans le cadre de ce cours (cf. la description du modèle de *Credit Grades* [Finger \(2002\)](#) par exemple comme application possible).

## 5.3 Modèle à intensité constante

L'approche *forme réduite* des modèles à intensité dissocie de manière explicite défaut et valeur de la firme, au contraire des modèles structurels. La date de défaut ressort d'un processus exogène et ne dépend pas des actifs de la firme. Cette modélisation permet de rendre compte des ruptures : le défaut survient lorsqu'un saut d'un processus de Poisson apparaît (cf. [Kurtz et Pignard \(2004\)](#) pour plus de précisions sur la modélisation des modèles à forme réduite). Nous étudions dans ce cours le cas d'un modèle à intensité constante.

### 5.3.1 Probabilité de défaut

**Définition 5.2** (Intensité de défaut). Soit  $\tau$  le temps de défaut. Nous définissons la probabilité de défaut instantanée comme l'intensité de défaut  $\lambda$  (ou taux de hasard) sur un intervalle  $(t, t + h]$  sachant qu'il n'y a pas eu de défaut auparavant :

$$\lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}[\tau \in (t, t + h] \mid \tau > t] = \lambda \quad (5.4)$$

Cette formulation revient à dire que  $\tau$  est le premier instant de saut d'un processus de Poisson d'intensité  $\lambda$ .

**Définition 5.3** (Processus de Poisson). Un processus de Poisson standard  $N = (N_t)_t$  est un processus de comptage<sup>1</sup> défini par le fait que :

1. les durées entre 2 dates  $t_{i-1}$  et  $t_i$  sont iid et distribuées selon une loi exponentielle de paramètre  $\lambda$ ,
2. de façon équivalente les variations disjointes de  $N$  sont indépendantes et distribuées pour  $N_t - N_s$  selon une loi de Poisson de paramètre  $\lambda(t - s)$  pour  $s < t$  :

$$\mathbb{P}[N_t - N_s = k] = \frac{1}{k!} [\lambda(t - s)]^k e^{-\lambda(t-s)}$$

1. Un processus de comptage est déterminé par le fait que  $N_t$  vaut 1 si l'événement *compté* apparaît, 0 sinon. Ici le processus compte l'événement de défaut qui est un état absorbant : lorsqu'il apparaît, le comptage est donc stoppé.

La probabilité d'un saut sur un instant  $\Delta t$  est défini par :

$$\frac{1}{\Delta t} \mathbb{P}[N_{t+\Delta t} - N_t = 1] = \lambda e^{-\lambda \Delta t} \rightarrow \lambda \quad \text{pour } \Delta t \rightarrow 0$$

Nous retrouvons donc bien la formulation de l'intensité de défaut (5.4) avec le processus de défaut  $N$  défini par  $N_t = \mathbb{1}_{\tau \leq t}$ . En effet, le défaut est le premier saut du processus de Poisson.

**Exercice 5.4.** Déterminer l'expression de la probabilité de défaut en  $T$  en fonction de l'intensité de défaut.

### 5.3.2 Évaluation d'une obligation risquée

Une obligation zéro-coupon risquée de nominal 1 et de maturité  $T$  ayant un recouvrement nul, c'est-à-dire dont le détenteur ne reçoit rien en cas de défaut, a une valeur  $D_t$  en  $t$ <sup>2</sup> sous la probabilité risque-neutre  $\mathbb{Q}$ <sup>3</sup> :

$$D_t = \mathbb{1}_{t < \tau} \mathbb{E}_{\mathbb{Q}}[e^{-r(T-t)} \mathbb{1}_{T < \tau} | \mathcal{F}_t]$$

**Exercice 5.5.** Donner la valorisation d'une obligation risquée en 0, en notant  $\bar{\lambda}$  l'intensité de défaut sous la probabilité risque-neutre.

Il est important de remarquer que cette intensité joue le rôle d'une rémunération supplémentaire : elle est assimilable à une prime de risque. Dans le cas d'une obligation zéro-coupon sans risque de défaut, nous avons  $\bar{\lambda} = 0$ . L'incertitude s'accroissant, la rentabilité exigée est plus importante. Nous pouvons estimer ainsi les intensités à partir de la structure par terme des spreads de crédit (équivalent à la différence entre le taux de l'obligation risquée et le taux sans risque).

Nous avons supposé le taux de recouvrement nul. Nous allons développer les formules de *pricing* des zéro-coupon avec un taux de recouvrement constant mais non nul. Nous nous intéresserons par la suite à l'estimation du taux de recouvrement de manière plus précise.

L'hypothèse de recouvrement la plus courante s'appelle *fractional recovery of par value* et consiste en un recouvrement à la maturité  $T$ <sup>4</sup> d'une fraction  $\delta$  du nominal du titre. Nous avons la valeur  $D_t^\delta$  de l'obligation suivante :

$$\begin{aligned} D_t^\delta &= \mathbb{1}_{t < \tau} \mathbb{E}_{\mathbb{Q}} \left[ e^{-r(T-t)} \mathbb{1}_{T < \tau} + \delta e^{-r(T-t)} \mathbb{1}_{\tau \leq T} | \mathcal{F}_t \right] \\ &= D_t + \mathbb{1}_{t < \tau} \delta e^{-r(T-t)} \mathbb{Q}[\tau \leq T | \mathcal{F}_t] \end{aligned}$$

**Exercice 5.6.** Donner la valorisation d'une obligation risquée en 0 avec un taux de recouvrement  $\delta$ .

### 5.3.3 Évaluation d'un CDS

Le modèle à intensité constante permet de valoriser le plus simple des produits dérivés de crédit et qui doit être considéré comme l'élément de base de produits plus exotiques : le Credit Default Swap ou CDS (cf. Braouezec et Brun (2007) pour plus de précisions sur les produits dérivés de crédit).

C'est un contrat financier conclu de gré à gré, pour une durée et un montant (notionnel) déterminés, et référencé sur un emprunteur bien défini. Dans ce contrat, l'acheteur de protection paie périodiquement lors de  $N$  dates une prime  $s$  (appelée aussi spread ou marge) tandis que le vendeur de protection s'engage à payer, en cas de défaut de l'emprunteur de référence, un montant compensant la perte résultant du défaut à hauteur du montant notionnel. Le CDS permet ainsi le transfert de risque de crédit. La protection est valable jusqu'à la maturité du swap  $T$ .

Un raisonnement simple d'absence d'opportunité d'arbitrage permet d'approximer la marge d'un CDS par le spread d'une obligation à taux variable de même maturité, ayant les mêmes dates de tombée de coupons et émise par l'emprunteur de référence du CDS.

2. On utilise l'indicatrice  $\mathbb{1}$  pour modéliser le fait que lorsque  $\tau < t$ , il y a défaut et l'obligation a une valeur nulle.

3. Les prix actualisés sont alors des martingales.

4. Nous n'étudions pas la formule d'évaluation pour un recouvrement à l'instant de défaut mais à la maturité.

Mais il est possible d'évaluer aussi cette marge plus précisément à l'aide des hypothèses du modèle à intensité constante. L'objectif est alors de calculer la marge  $s^*$  du CDS en 0 qui s'obtient en égalisant la jambe fixe, correspondant aux flux payés par l'acheteur de protection tant qu'il n'y a pas de défaut, et la jambe variable, correspondant au flux payé par le vendeur de protection en cas de défaut.

En considérant que la prime est payée jusqu'au défaut, la valeur de la jambe fixe  $JF(s)$  est définie comme l'espérance actualisée du payoff sous la probabilité risque-neutre :

$$\begin{aligned}
 JF(s) &= \mathbb{E}_{\mathbb{Q}} \left[ \sum_{i=0}^{N-1} s(t_{i+1} - t_i) e^{-rt_{i+1}} \mathbb{1}_{\tau > t_{i+1}} \right] \\
 &= s \sum_{i=0}^{N-1} (t_{i+1} - t_i) e^{-rt_{i+1}} \mathbb{Q}[\tau > t_{i+1}] \\
 &= s \sum_{i=0}^{N-1} (t_{i+1} - t_i) e^{-(r+\bar{\lambda})t_{i+1}} \quad \text{car } \tau \text{ suit une loi exponentielle de fonction de survie } S(u) = e^{-\bar{\lambda}u} \\
 &\approx s \int_0^T e^{-(r+\bar{\lambda})u} du \quad \text{par approximation d'une intégrale par les sommes de Riemann} \\
 &= s \frac{1 - e^{-(r+\bar{\lambda})T}}{r + \bar{\lambda}}
 \end{aligned}$$

En supposant que le flux variable est payé à l'instant  $\tau$  d'occurrence du défaut, la valeur de la jambe variable  $JV$  s'écrit sous la probabilité risque-neutre<sup>5</sup> :

$$\begin{aligned}
 JV &= \mathbb{E}_{\mathbb{Q}} [(1 - \delta) e^{-r\tau} \mathbb{1}_{\tau \leq T}] \\
 &= (1 - \delta) \int_0^T e^{-ru} \bar{\lambda} e^{-\bar{\lambda}u} du \quad \text{car } \tau \text{ suit une loi exponentielle de densité } f(u) = \bar{\lambda} e^{-\bar{\lambda}u} \\
 &= (1 - \delta) \bar{\lambda} \int_0^T e^{-(r+\bar{\lambda})u} du \\
 &= (1 - \delta) \bar{\lambda} \frac{1 - e^{-(r+\bar{\lambda})T}}{r + \bar{\lambda}}
 \end{aligned}$$

Le *fair spread*  $s^*$  est tel que  $JF(s^*) = JV$ . Nous obtenons alors l'égalité suivante, que nous appelons le *triangle du crédit* :

$$s^* = (1 - \delta) \bar{\lambda}$$

Les CDS (et d'autant plus s'ils sont seniors) étant assez liquides, nous pouvons extraire l'intensité de défaut implicite au *pricing* des CDS pour estimer une probabilité de défaut implicite relativement juste pour l'émetteur sous-jacent et pour une maturité donnée. A partir du triangle du crédit, les PD implicites ont donc pour expression :

$$PD = 1 - \exp\left(-\frac{s}{1 - \delta} T\right)$$

Les prolongements du modèle à intensité définissent non plus l'intensité constante (qui a pour conséquence la génération de structures par terme *plates* des primes de CDS) mais des intensités avec des formes plus complexes (intensité constante par morceaux, polynomiale, stochastique, etc.). Les structures par terme générées seront alors plus proches de la réalité des marchés.

---

5. Pour une variable aléatoire réelle continue  $X$  et une application  $h$  de  $\mathbb{R}$  dans  $\mathbb{R}$ , rappelons le théorème de transfert qui s'écrit :  

$$\mathbb{E}[h(X)] = \int_{-\infty}^{+\infty} h(x) f_X(x) dx.$$



## CHAPITRE 6

# MODÉLISATION DE LA PERTE EN CAS DE DÉFAUT

La perte en cas de défaut (Loss Given Default ou LGD en anglais) est influencée par de nombreux facteurs qui peuvent être propres à la structure de l'entreprise ou découlant d'un état de l'activité économique ou encore d'une juridiction plus ou moins favorable aux créanciers.

### 6.1 Mesure de la LGD

Il existe principalement deux mesures retenues pour estimer la LGD :

- déduite du prix du titre de dette juste après que celui-ci a connu une situation de défaut : on parle de *LGD market-implied*,
- déduite de la valeur actualisée des actifs venant en compensation du défaut du titre de dette : on parle de *LGD workout* (valeur liée à la revente ou à la liquidation du fonds de commerce et des actifs).

La première de ces mesures présente l'inconvénient de ne pas pouvoir systématiquement être évaluée lorsqu'il n'existe pas de marché permettant de fournir un tel prix. Pour autant, lorsque cette mesure peut être appliquée, elle présente l'avantage de fournir une mesure de la LGD juste après l'événement de défaut. Mais cette valeur est plus élevée car provenant de *bonds* dont le lien entre prêteur et emprunteur est moins qu'avec un *loan*.

La seconde mesure présente l'inconvénient d'être parfois complexe à déterminer, en particulier lorsque les actifs venant en garantie sont illiquides. Le délai moyen séparant le défaut de la récupération peut être long et ceci pose donc la question de la détermination d'un taux d'actualisation approprié. Mais cette mesure est celle qui est privilégiée par le Comité de Bâle.

Il existe quatre grands types de risk drivers explicatifs du processus de recouvrement :

1. lié à la transaction
  - (a) la séniorité de l'instrument, qui attribue des priorités dans le remboursement à travers des clauses à l'émission (un prêt *senior* sera remboursé avant un prêt *junior* appelée aussi *subordonné*)
  - (b) le type de produit
  - (c) le type de sûreté (ou collatéral) qui est le nom donné aux actifs venant en garantie. En période de récession, la valeur de la sûreté aura tendance à décroître et la probabilité de défaut à augmenter.
  - (d) la zone géographique du collatéral
2. lié à l'émetteur
  - (a) sa taille

- (b) son secteur industriel. Le remboursement dépend largement de la valeur des actifs venant en garantie du défaut. Or la valeur de ces actifs dépend du secteur industriel auquel l'émetteur est attaché : un secteur où les actifs immatériels sont importants aura un taux de recouvrement plus faible que ceux dans lesquels ils seront moins présents.
- 3. lié à l'institution prêteuse
  - (a) sa capacité de négociation. Les droits de contrôle des créanciers sur l'entreprise sont prédéfinis dans le contrat. En outre, plus la part de l'investissement est importante, plus la capacité de négociation est forte pour le créancier.
  - (b) l'organisation interne et des services dédiés au recouvrement
- 4. lié à l'environnement externe
  - (a) le système de juridiction. Lorsqu'il y a défaut, le niveau de recouvrement est déterminé par un juge. Les différences de juridiction entre les pays est essentielle pour déterminer le taux de recouvrement, ainsi que les périodes de temps nécessaires au recouvrement (18 mois à 10 ans en France, 18 mois à 3 ans aux États-Unis).
  - (b) le cycle d'activité économique : une dépendance importante est révélée empiriquement entre le cycle économique, les taux d'intérêt et le taux de recouvrement.

Après avoir étudié la définition, ainsi que les risk drivers des LGD, nous allons pouvoir la modéliser.

## 6.2 Modélisation probabiliste et estimation

Nous avons précédemment introduit une LGD constante dans les modèles à intensité et l'évaluation d'obligation risquée. Ici, nous modélisons la LGD en la supposant aléatoire.

La forme de la distribution de la LGD nous pousse à supposer que les LGD sont distribuées selon une loi Bêta. C'est une distribution à support  $[0, 1]$  qui dépend de 2 paramètres  $a$  et  $b$ . L'expression de sa fonction de densité est :

$$f(x) = \frac{1}{\beta(a, b)} x^{a-1} (1-x)^{b-1} \mathbb{1}_{[0,1]}(x)$$

avec  $\beta(a, b) = \int_0^1 x^{a-1} (1-x)^{b-1} dx$ . La moyenne  $\mu$  et la variance  $\sigma$  sont égales à :

$$\mu = \frac{a}{a+b}$$

$$\sigma = \frac{ab}{(a+b)^2(a+b-1)}$$

Cette distribution est très flexible et permet de modéliser plusieurs formes de distribution. Ainsi, par exemple :

- si  $a = b = 1$ , nous obtenons la distribution uniforme,
- si  $a = b$ , nous avons une distribution symétrique par rapport à  $x = 1/2$ ; si les paramètres sont supérieurs à 1, la courbe est en cloche, alors qu'elle a la forme d'un U s'ils sont inférieurs à 1,
- si  $a > b$ , la courbe est asymétrique à droite (skewness négatif).

A partir de la LGD moyenne  $\hat{\mu}$  et la volatilité du taux  $\hat{\sigma}$ , la méthode des moments nous permet d'estimer les paramètres  $a$  et  $b$  de la manière suivante :

$$\hat{a} = \frac{\hat{\mu}^2(1-\hat{\mu})}{\hat{\sigma}^2} - \hat{\mu}$$

$$\hat{b} = \frac{\hat{\mu}^2(1-\hat{\mu})^2}{\hat{\mu}\hat{\sigma}^2} - (1-\hat{\mu})$$

La modélisation de la LGD ne doit pas être négligée. En effet, son impact sur la modélisation du risque de crédit est souvent plus important que celui de la probabilité de défaut.

## CHAPITRE 7

# MODÉLISATION DE LA DÉPENDANCE DES DÉFAUTS

Enfin, le dernier paramètre à considérer dans la mesure du risque de crédit est la dépendance entre les différents actifs d'un portefeuille de créances. Ce que l'on doit modéliser est plus précisément les dépendances entre les défaillances potentielles. Ceci permettra d'évaluer le risque de contagion ou risque multiple contenu dans un portefeuille de crédit. Pour cela, il nous faut tout d'abord déterminer des observations sur la structure de dépendances des défauts.

### 7.1 Mesurer la dépendance des événements de défaut

Les dépendances de défaillance ont la particularité de ne pas être en général observables directement. Néanmoins, les modèles structurels et à intensité vont nous permettre d'extraire une structure de dépendance approximée.

#### 7.1.1 Dépendance à partir du modèle de Merton

Rappelons qu'à partir du modèle de Merton (cf. section 5.2), la valeur de la firme peut être décomposée comme la somme de la part action et de la part dette :  $\text{Firme} = \text{Action} + \text{Dette}$ . Pour déterminer une structure de dépendance des défauts, nous allons faire l'hypothèse de l'équivalence entre le niveau de dépendance obtenu entre les rentabilités des actions et celui obtenu entre les rentabilités des firmes. La structure de dépendance entre les actions est facile à déterminer pour les contreparties cotées. Cette méthode d'extraction de la structure de dépendance des défaillances aura notre préférence malgré l'approximation réalisée.

#### 7.1.2 Dépendance à partir du modèle à intensité constante

Les dépendances de défaut peuvent être extraites aussi à partir du modèle à intensité étudiée en section 5.3. En effet, nous avons vu que l'intensité de défaut peut être approchée par le spread de crédit. Nous pouvons alors approcher la dépendance des défauts par celle des spreads. A préciser : la notion de spread recouvre à la fois la probabilité de défaut et le taux de recouvrement, difficiles à distinguer, et le risque de liquidité influence aussi les spreads en plus du risque de défaut.

## 7.2 Copules et dépendance des défauts

Maintenant que les données sur la structure de dépendance sont construites, il reste à modéliser cette dépendance. Un point essentiel est la nature des défauts. En effet, les événements de défaut sont des événements rares : il est donc nécessaire de pouvoir capter la dépendance extrême entre ces événements. Ceci peut être fait seulement par certaines copules ayant un indicateur de dépendance de queue à gauche non nul (cf. partie précédente).

Ces copules vont donc nous permettre de modéliser de manière robuste la structure de dépendance des défauts.

## CHAPITRE 8

# CREDITVAR

Nous avons étudié dans les chapitres précédents les éléments fondamentaux à la modélisation du risque de crédit. Les probabilité de défaut, LGD et dépendance des défauts vont permettre de mesurer le risque de crédit d'un portefeuille de créances. Pour cela, nous devons maintenant définir une mesure synthétique du risque pour capturer le risque de perte : ce sera la Value-at-Risk des pertes de risque de crédit, appelée CreditVaR.

Il nous faut tout d'abord caractériser la fonction de perte  $L$  d'un portefeuille de crédit pour un horizon donné. Supposons que le portefeuille porte sur  $I$  contreparties qui n'ont qu'une seule créance chacune. Le temps de défaut de la contrepartie  $i$  est noté  $\tau_i$ , le taux de recouvrement  $\delta_i$  et le notionnel  $x_i$ . Nous n'intégrons pas, par souci de simplification, le facteur d'actualisation des flux. Nous avons la fonction de pertes du portefeuille de créances égale à l'horizon  $h$  à :

$$L_h = \sum_{i=1}^I L_h^i = \sum_{i=1}^I x_i (1 - \delta_i) \mathbb{1}_{\tau_i \leq h}$$

avec  $L_h^i$  la perte associée à la contrepartie  $i$  à l'horizon  $h$ .

Nous utilisons les modélisations probabilistes étudiées précédemment des probabilités de défaut, des LGD et des dépendances des  $I$  pertes pour simuler les pertes du portefeuille.

La CreditVaR est alors définie comme le quantile de probabilité  $\alpha$  (égale par exemple à 99.9%) de la fonction de perte  $L_h$  :

$$\text{CreditVaR}(\alpha) = \inf\{L : \mathbb{P}[L_h \leq L] \geq \alpha\}$$

Nous l'obtenons à partir de simulations Monte Carlo de la distribution de pertes du portefeuille de crédit. Nous simulons alors  $N$  pertes  $L_h$ , comme somme des pertes sur chaque créance, et nous déterminons le quantile de probabilité  $\alpha$  sur ces  $N$  pertes.

**Exercice 8.1.** Vous disposez de 2 créances sur deux grandes entreprises de même notionnel 1000 EUR et de même maturité 1 an :

1. la première est un prêt d'escompte senior pour une grande entreprise dont le CDS de maturité 1 an a une prime de 300bp au moment du prêt (vous pouvez extraire la probabilité de défaut implicite à partir du triangle du crédit en supposant que le CDS a été évalué avec un taux de recouvrement égal à 40%); les LGD pour les créances ayant les mêmes caractéristiques que ce prêt et pour le secteur et le pays dans lequel évolue cette entreprise ont une moyenne de 32% et un écart-type de 41%.
2. la seconde est un prêt junior (ou subordonné) sans sûreté pour une grande entreprise dont le CDS de maturité 1 an a une prime de 50bp au moment du prêt; les LGD pour les créances ayant les mêmes caractéristiques que ce prêt et pour le secteur et le pays dans lequel évolue cette entreprise ont une moyenne de 68% et un écart-type de 33%.

3. la dépendance entre les défauts est à modéliser à partir des données de rentabilités actions simulées et se trouvant dans l'onglet CreditVaR du fichier Excel fourni. Il n'y a pas de dépendance entre les 2 LGD.

Il vous est demandé de déterminer la CreditVaR à 1 an à 99.9% par simulations Monte Carlo, en modélisant le plus précisément possible la structure de dépendance entre ces 2 créances (tester les copules produit, Gaussienne et Student), à partir des actions des 2 entreprises.

Précisons que cet exercice est à but strictement pédagogique. Dans la réalité, un portefeuille de crédit est exposé à de nombreuses créances permettant une meilleure diversification. Dans le cas étudié ici, le portefeuille est composé de seulement 2 créances : le risque de perte sera en conséquence très important. A titre de comparaison, avec 1000 fois plus de créances de nominal 1000 fois moins élevé, on peut diminuer la CreditVaR de 50%.

Ainsi, il y aura convergence entre la CreditVaR Gaussienne calculée dans ce cas et la CreditVaR déterminée via l'Asymptotic Single Risk Factor (ASRF) pour un niveau de confiance de 99.9% :

$$\text{CreditVaR} = \sum_{i=1}^I EAD_i * LGD_i * \Phi \left( \frac{\Phi^{-1}(PD_i) - \sqrt{\rho} \Phi^{-1}(1 - 99.9\%)}{\sqrt{1 - \rho}} \right)$$

avec  $\rho$  la corrélation supposée constante entre les actifs et  $\Phi$  la fonction de répartition Gaussienne standard. Pour les détails de la démonstration de l'ASRF, nous renvoyons à [Clauss et Pansard \(2021\)](#).

# CONCLUSION

Cette dernière partie a développé une introduction à la mesure du risque de crédit en étudiant précisément les probabilité de défaut, LGD et structure de dépendance des défauts.

L'utilisation des copules pour modéliser le risque multiple des défaillances des émetteurs de créances prend alors dans ce cadre tout son sens. En effet, la dépendance des défauts, qui sont des événements rares, ne peut être captée qu'à l'aide de certaines copules adéquates.

# BIBLIOGRAPHIE

- Artzner, P., Delbaen, F., Eber, J.M., Heath, D., 1999, Coherent Measures of Risk, *Mathematical Finance*, 9(3), 203-228.
- Azzalini, A., Capitanio, A., 2003, Distributions Generated by Perturbation of Symmetry with Emphasis on a Multivariate Skew t Distribution, *J.Roy.Statist.Soc* 65, 367-389.
- Balkema, A., de Haan, L., 1974, Residual Life Time at Great Age, *Annals of Probability* 2, 792-804.
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J. 2004, *Statistics of Extremes - Theory and Applications*, Wiley, England.
- Black, F., Scholes, M., 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81(3), 637-654.
- Balkema, A., de Haan, L., 1974, Residual Life Time at Great Age, *Annals of Probability* 2, 792-804.
- Braouezec, Y., Brun, J., 2007, *Dérivés de Crédit Vanille et Exotiques*, Revue Banque Édition.
- Clauss, P., Pansard, F., 2021, *Economie Bancaire - analyse financière, régulation et gestion des risques*, Ellipses.
- Crosbie, P., Bohn, J., 2003, *Modelling Default Risk*, KMV Corporation, San Francisco.
- Davidson, R., MacKinnon, J.G., 1993, *Estimation and Inference in Econometrics*, Oxford University Press.
- Efron, B., 1995, *Le Bootstrap et ses Applications*, CISIA.
- Finger, C., 2002, *Credit Grades : Technical Document*, RiskMetrics Group.
- Fisher, R., Tippett, L., 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.
- Frey, R., McNeil, J., 2000, Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series : an Extreme Value Approach, *Journal of Empirical Finance* 7, 271-300.
- Genest, Ch., Boies, J.C., 2003, Detecting Dependence with Kendall Plots, *American Statistical Association* 57, 275-284.
- Genest, Ch., Rémillard, B., Beaudoin, D., 2009, Goodness-of-fit Tests for Copulas : A Review and a Power Study, *Insurance : Mathematics and Economics* 44(2), 199-213.
- Haworth, H., 2004, *Structural Models of Default*, University of Oxford.
- Hosking, J., Wallis, J., 1987, Parameter and Quantile Estimation for the Generalized Pareto Distribution, *Technometrics* 29, 339-349.
- Hull, J.C., 2009, *Options, Futures, and other Derivatives*, Pearson, 7e édition.



- Kurtz, D., Pignard, T., 2004, *Modélisation du Risque de Crédit*, Groupe de Recherche Opérationnelle, Crédit Lyonnais.
- Lando, D., Skødeberg, T.M., 2002, Analyzing Rating Transitions and Rating Drift with Continuous Observations, *Journal of Banking & Finance* 26(2-3), 423-444.
- Lessmann, S., Baesens, B., Seow, H.V., Thomas, L.C., 2015, Benchmarking State-of-the-art Classification Algorithms for Credit Scoring : An Update of Research, *European Journal of Operational Research* 247(1), 124-136.
- Merton, R.C., 1974, On the Pricing of Corporate Debt : the Risk Structure of Interest Rates, *Journal of Finance* 29, 449-470.
- Pickands, J., 1975, Statistical Inference Using Extreme Order Statistics, *The Annals of Statistics* 3, 119-131.
- Reiss, R.D, Thomas, M., 2001, *Statistical Analysis of Extreme Values*, 2nd ed., Birkhäuser, Basel.
- Rootzen, H., Tajvidi, N., 1996, Extreme Value Statistics and Wind Storm Losses : a Case Study, *Scandinavian Actuarial Journal* 70-94.
- de Servigny, A., Metayer, B., Zelenko, I., 2006, *Le risque de Crédit*, Dunod.
- 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.
- Thomas, A., 2000, *Économétrie des Variables Qualitatives*, Dunod.