

Mémoire présenté le :
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : Kevine TCHAKOUTE NJEUNKAM

Titre : Calibrage du risque de sinistralité future en assurance IARD et étude des corrélations

Confidentialité : NON OUI (Durée : 1 an 2 ans)

Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

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

*Entreprise : AVIVA Assurances
Nom : Antoine CHANH*

Signature : 

*Directeur de mémoire en entre-
prise :
Nom :*

Signature :

*Membres présents du jury de
l'ISFA*

*Invité :
Nom :*

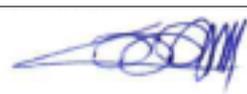
Signature :

*Autorisation de publication et
de mise en ligne sur un site de
diffusion de documents actua-
riels (après expiration de l'éventuel
délai de confidentialité)*

Signature du responsable entreprise



Signature du candidat



Remerciements

Ce travail est la résultante du soutien, de l'assistance et de la collaboration de plusieurs personnes auxquelles je tiens à exprimer ma profonde gratitude. Il s'agit de :

- Antoine CHANH et Brice CARLES, pour l'accueil chaleureux qu'ils m'ont accordé, leurs conseils et encouragements ;
- Hoan BUY, Ioan MARA et Sarah DAMI pour leur disponibilité et collaboration ;
- Tout le personnel de la fonction actuarielle d'AVIVA pour son accueil et sa convivialité, qui ont favorisé un cadre agréable et propice à la recherche ;
- Tout le corps enseignant de l'ISFA pour les connaissances qu'ils m'ont inculquées tout au long de ma formation ;
- Tous ceux qui de près ou de loin m'ont soutenue tout au long de ce projet de mémoire.

Résumé

Mots-clés : Solvabilité 2, SCR, modèle interne, risque de sinistralité future, copules, facteur de causalité, diversification, matrice QIS, EIOPA.

Dans le cadre de la réglementation solvabilité II, les sociétés d'assurance doivent disposer d'un niveau de capital au moins égal au capital de solvabilité requis appelé SCR afin de pouvoir résister à un choc bicentenaire. La société d'assurance peut faire le choix de calculer son SCR soit en utilisant la formule standard avec paramètres fixes calibrés sur le marché européen, soit avec un modèle interne avec des paramètres plus adaptés à son profil de risque.

Chez AVIVA Assurances, le risque de prime est subdivisé en trois sous-risques : le risque de portefeuille, le risque de sinistralité future et le risque de frais. Dans ce mémoire, nous traiterons uniquement de la composante sinistralité future du risque de prime qui représente plus de 70% du SCR du risque de prime.

Dans un processus d'élaboration d'un modèle interne pour AVIVA Assurances, nous sommes emmenés dans ce mémoire à calibrer le risque de sinistralité future qui mesure la déviation de la sinistralité future provenant de l'aléa sur la fréquence et la sévérité des sinistres. Le calibrage sera fait par branches de garanties homogènes appelées *Model Calculation Class* (MCC).

Etant donné que le calibrage se base sur des données historiques des sinistres de la société, celles-ci ont été mises en as-if en utilisant des indices marché adaptés aux différentes MCC. En effet, la mise en as-if permet de prendre en compte la valeur temps de l'argent et par conséquent, d'obtenir des valeurs comparables entre elles. Par la suite, ces données ont été mises à l'ultime par la méthode de Chain-Ladder afin d'estimer la valeur totale des sinistres. Les retraitements ainsi effectués sur les données historiques ont conduit à une hausse de la charge de sinistres de 14,52% permettant ainsi de ne pas sous-estimer la charge de sinistres de la société.

Au sein des MCC, il peut être possible d'observer deux types de sinistres : les sinistres dits attritionnels qui ont une fréquence de survenance élevée mais des coûts relativement faibles et des sinistres dits graves qui ont une fréquence de survenance faible mais des coûts relativement élevés. La théorie des valeurs extrêmes a été utilisée pour déterminer pour chaque MCC le seuil qui sépare la sinistralité attritionnelle de la sinistralité grave. L'approche utilisée est celle des excès au dessus d'un seuil et l'analyse a été basée sur les indicateurs tels que le *Mean excess plot*, *Gerstengabe plot*, l'*AMSE* de l'estimateur de Hill. Il en est ressorti que le seuil de grave n'est pas toujours le même pour les différentes branches de garanties et aussi qu'il n'est pas nécessaire de distinguer les deux sinistralités pour certaines branches de garanties.

Pour chaque MCC, un modèle agrégé de charge ultime en unité d'exposition (charge ultime/nombre de contrats) a été utilisé pour modéliser la sinistralité attritionnelle et un modèle fréquence*sévérité a été utilisé pour modéliser la sinistralité grave. Pour chaque modèle, a été calibré une loi de distribution, un coefficient de variation et le SCR associé. Les SCR ainsi obtenus doivent être agrégés dans un premier temps au sein de chaque branche de garanties pour les sinistralité attritionnelles et graves et dans un second entre les différentes branches de garanties.

La petitesse de la profondeur des données historiques a été un obstacle au calibrage des corrélations entre les branches de garanties. En effet, l'historique à disposition, constitué de 19 points, n'est pas suffisant pour refléter la structure de corrélation qui existe entre les différentes branches de garanties. Une alternative a été de faire recours au Bootstrap en appliquant une structure de ré-échantillonnage identique aux triangles des résidus de Chain-Ladder de chaque MCC. Les copules ont été utilisées pour calibrer les corrélations entre les différentes branches de garanties. Ces copules ont été calibrées sur les données issues du ré-échantillonnage. Cependant, vu le nombre élevé de paires de corrélation à calibrer, les copules ont été calibrées sur les paires de MCC pour lesquelles le tau de Kendall est significatif. De plus, parmi les paires pour lesquelles les copules ont été calibrées, les corrélations qui en découlent sont parfois faibles.

Une approche qualitative basée sur les facteurs de causalité a également étudiée pour donner une autre vision de la corrélation, en effet, elle interprète la dépendance entre deux branches de garanties (ou MCC) comme le fait que ces branches de garanties partagent les mêmes risques (facteurs de causalité) et réagissent de façon similaire quant à la survenance de ces facteurs de causalité.

Les résultats de l'approche par la méthode des copules et ceux de l'approche par facteurs de causalité ont été conciliés pour obtenir la matrice de corrélation finale. Il en est ressorti que, sur 105 paires calibrées, seules 33 sont égales à celles préconisée par la formule QIS ; ceci confirme que la matrice de corrélation préconisée par l'EIOPA n'est pas adaptée au portefeuille d'AVIVA Assurances. La matrice de corrélation ainsi calibrée a conduit à un bénéfice de diversification du SCR total au titre de la sinistralité future de 26 m€ soit de 11% par rapport à la matrice QIS.

Abstract

Key words : Solvency 2, SCR, internal model, futur claim risk, copula, causal driver, diversification, QIS matrix, EIOPA.

Under Solvency II regulations, insurance companies must have a level of capital at least equal to the Solvency Capital Requirement (SCR) in order to be able to meet its obligations over the next 12 months with a probability of at least 99.5% (a 1 in 200 event). The insurance company can choose to calculate its SCR either using the Standard Formula with fixed parameters calibrated on the European market, or using an internal model with parameters more suited to its risk profile.

At *AVIVA Assurances*, premium risk is subdivided into three sub-risks : portfolio risk, future claims risk and expense risk. In this paper, we will focus only the future claims component of the premium risk, which represents more than 70% of the SCR of the premium risk.

In order to develop an internal model for *AVIVA Insurance*, we calibrated the future claims risk which measures the deviation of future claims from the uncertainty on the frequency and severity of claims. The calibration is done by homogeneous classes called Model Calculation Class (MCC).

Since the calibration is based on the company's historical claims data, these have been put into as-if using market indexes which are relevant to the different MCCs. Indeed, the as-if allows to take into account the time value of money and therefore to obtain comparable values. Subsequently, in order to estimate the total value of claims, these data were putted into their ultimate values by using the Chain-Ladder method. The adjustments made to the historical data led to a 14.52% increase in the claims amount , thus ensuring that the company's claims amount was not underestimated.

Within MCC, it may be possible to observe two types of claims : attritional claims, which have a high frequency but relatively low costs, and Large claims, which have a low frequency but relatively high costs. Extreme value theory was used to determine for each MCC the threshold between attritional and large claims. The approach used is that of Peack over threshold and the analysis was based on indicators such as the Mean excess plot, Gerstengabe plot and the AMSE (Asymptotic Mean Square error) of the Hill estimator. we obtained that the threshold of large claims is not always the same for the classes and also that it is not necessary to separate the two type of claims for certain classes.

For each MCC, an aggregated model of ultimate claim in exposure unit (ultimate /number of contracts) was used to model attritional claims and a frequency*severity model was used to

model large claims. For each model, a distribution, a coefficient of variation and the associated SCR were calibrated. The SCRs obtained must be aggregated first within each class for attritional and large claims and secondly between the different classes.

The small length of historical data used has been an obstacle for the calibration of correlations between classes. In fact, the 19 observations available in the history is not sufficient to reflect the correlation structure between classes . An alternative was to use the Bootstrap by applying an identical re-sampling structure to the triangles of the Chain-Ladder residuals of each MCC. Copulas were used to calibrate the correlations between the different classes. This was performed on the data from the re-sampling. However, copulas were calibrated on the pairs of MCCs for which the Kendall's tau is significant. In addition, among the pairs for which the copula were calibrated, the resulting correlations are sometimes low.

A qualitative approach based on causal driver has also been performed in order to give another view of the correlation, in that it interprets the dependence between two classes as the fact that these classes share the same risks (causal driver) and react in a similar way to the occurrence of these causal driver.

The results of copulas approach and the causal driver approach were reconciled to obtain the final correlation matrix. we obtained that, out of 105 calibrated pairs, only 33 are equal to those recommended by the QIS matrix of the Standard formula ; this confirms that the correlation matrix recommended by the EIOPA is not adapted to the portfolio of AVIVA insurance. The correlation matrix thus calibrated led to a diversification benefit of the total SCR for future claims of €26m, i.e. 11% in comparison to the QIS matrix.

Sommaire

Remerciements	i
Résumé	ii
Abstract	iv
Introduction	2
0.1 Contexte et justification de l'étude	2
0.2 Problématique	3
0.3 Objectifs	3
0.4 Plan de travail	4
1 Cadre de l'assurance non-vie	5
1.1 Réglementation solvabilité 2	5
1.2 La formule standard	7
1.3 Le risque de prime dans le modèle d'AVIVA : définition du risque de sinistralité future	9
2 Retraitement des données	16
2.1 Description de la base	16
2.2 Mise en as-if	18
2.3 Projection à l'ultime	23
3 Étude du seuil de sinistre grave par branche de garanties	28
3.1 Théorie des valeurs extrêmes	28
3.2 Estimation de l'indice de queue	31
3.3 Méthodes de détermination du seuil	32
3.4 Application	34
4 Modélisation de la sinistralité	41
4.1 Tests d'adéquation à une loi	41
4.2 Modélisation	44
4.3 Calcul des SCR stand-alone	51
5 Théorie quantitative sur l'agrégation des risques	53
5.1 Histoire des corrélations	53
5.2 Théorie des copules	55

6	Détermination de la structure de dépendance par une approche quantitative	63
6.1	Ré-échantillonnage des données	63
6.2	Structure de dépendance	66
7	Une approche par facteur de causalité	84
7.1	Liste des facteurs de causalité	84
7.2	Quantification des facteurs de causalité	92
7.3	Calcul des corrélations	92
8	Matrice de corrélation idéale	98
8.1	Comparaison des différentes méthodes	99
8.2	Avantages et inconvénients des différentes méthodes	104
8.3	Choix de la matrice de corrélation idéale	106
	Limites de l'étude	109
	Conclusion	110
	Bibliographie	110
	Annexes	113
	Table des matières	131
	Liste des figures	133
	Liste des tableaux	135

Introduction

0.1 Contexte et justification de l'étude

Dans un contexte où l'avenir est incertain, l'assureur a pour vocation de protéger ses clients (assurés) moyennant le paiement d'une prime contre des événements qui pourront leur porter préjudice. L'assureur vend des produits dont il ne connaît pas à l'avance le prix de revient : la prime d'assurance est calculée sur la base des hypothèses faites sur la sinistralité future et sur le volume de primes futures anticipée. Cependant, si le processus de tarification utilisé sous-estime le risque, ou s'il conduit à attirer que des mauvais risques ou encore s'il y a des perturbations sur le marché, alors la sinistralité va être supérieure à celle anticipée. Ce cas de figure est problématique pour l'assureur, car il risquera de ne pas pouvoir honorer ses engagements vis à vis des assurés.

Pour garantir la solidité financière des assureurs et des ré-assureurs, le régulateur soumet les sociétés d'assurance et de réassurances à un ensemble de règles dites « prudentielles » de solvabilité. En effet, elles doivent disposer d'un niveau de fonds propres au moins égal au capital de solvabilité requis (SCR) afin de pouvoir résister à un choc bicentenaire ; c'est à dire une perte que subirait la société avec une probabilité de 0,5% sur un horizon d'un an.

Pour déterminer son SCR, l'entreprise peut soit utiliser la formule standard (avec des paramètres fixes, calibrés sur le marché européen), soit développer un modèle interne partiel (formule standard et calibrages propres pour un ou plusieurs risques spécifiques), soit développer son propre modèle interne (avec des paramètres adaptés à son profil de risque). AVIVA s'inscrit dans le dernier cas de figure.

Le groupe AVIVA utilise un modèle interne « partiel »¹ pour la détermination de son SCR depuis l'entrée en vigueur de solvabilité 2. Ce modèle interne couvre certaines de ses filiales vie et non-vie ; c'est pour cette raison qu'il est qualifié de « partiel ». En France, deux sociétés d'assurance de la filiale française du groupe AVIVA à savoir « AVIVA Vie » et « AVIVA Épargne Retraite » utilisent déjà ce modèle interne. Ceci n'est pas encore le cas d'« AVIVA Assurances », la principale entité française du groupe AVIVA avec des activités non-vie, qui calcule son SCR à l'aide de la formule standard. AVIVA Assurances souhaite utiliser un modèle interne complet sur l'ensemble de ses risques.

Dans un processus d'élaboration d'un modèle interne pour la branche non-vie de la filiale française « AVIVA Assurances », nous nous intéresserons dans ce mémoire au risque de sinistralité future qui capte la déviation de la sinistralité future. Ce risque mesure l'aléa sur la fréquence de survenance des sinistres et la sévérité de ceux-ci. Nous précisons que ce risque est

1. partiel en terme de filiales qui l'utilisent

le plus matériel de la société.

0.2 Problématique

La modélisation des pertes futures par une loi de probabilité n'est pertinente que si elle se restreint à un groupe homogène de contrats d'assurance. Chez AVIVA Assurances, la granularité utilisée est appelée MCC (Model Calculation Class) encore appelées « branches de garanties » ; ce sont des groupes intra homogènes de produits d'assurance.

Dans le cadre de la modélisation de la sinistralité future, deux types de sinistres se distinguent pour chaque MCC à savoir : des sinistres qui surviennent assez souvent mais avec des coûts relativement moins élevés (sinistres attritionnels) et des sinistres qui surviennent moins souvent mais avec des coûts relativement plus élevés (sinistres graves). Il n'est donc pas judicieux de proposer une seule loi de probabilité pour les pertes futures, mais plutôt d'en proposer 2 : une pour les sinistres attritionnels et une autre pour les sinistres graves. La difficulté repose sur la question suivante : « **A partir de quel seuil pouvons-nous qualifier un sinistre de grave ?** » Ce seuil devra dépendre de la branche de garanties considérée.

Pour une branche de garanties donnée, connaissant le seuil qui sépare la sinistralité attritionnelle de la sinistralité grave, il sera dès lors possible de modéliser les deux sinistralités et calculer les SCR associés. Le capital de solvabilité requis (SCR) au titre de la sinistralité future, s'obtient en agrégeant dans un premier temps les SCR des sinistralités attritionnelle et grave de chaque branche de garanties et dans un second temps par agrégation des SCR au titre de la sinistralité future obtenus sur les différentes branches de garanties.

Les branches qui subissent une perte élevée cette année pourront être compensées par les branches en pleine forme car la probabilité que toutes les branches subissent des pertes élevées la même année est très faible. Ainsi, le SCR de deux branches de garanties est inférieur à la somme des SCR individuels de ces deux branches du fait de la diversification. La difficulté ici est de déterminer de combien ce SCR sera inférieur par rapport la somme des SCR individuels. Il est alors nécessaire de répondre à la question suivante : « **Quelle est la structure de corrélation adéquate pour agréger les SCR du risque de sinistralité future des branches de garanties d'AVIVA Assurances ?** ». Cette structure devra être aussi simple que possible et aussi complexe que nécessaire.

0.3 Objectifs

L'objectif de ce mémoire est de « **calibrer le risque de sinistralité future** ». De manière spécifique, il s'agira de :

- Retraiter les données historiques : mise en as-if et mise à l'ultime. La mise en as-if permet de mettre les données historiques de sinistres sur un même référentiel de sorte à les rendre comparables. En effet, si un sinistre identique à un sinistre survenu en 2005 survenait aujourd'hui il n'aurait probablement pas le même coût. Quant à la mise à l'ultime, elle permet d'estimer le coût du sinistre après son écoulement total ;

- Déterminer le seuil qui sépare la sinistralité attritionnelle de la sinistralité grave pour chaque branche de garanties ;
- Modéliser les sinistralités attritionnelle et grave de chaque branche de garanties ;
- Définir une matrice de corrélation entre les SCR du risque de sinistralité future des différentes branches de garanties.

0.4 Plan de travail

Notre travail s'articule autour de 8 chapitres répartis comme suit : Le premier chapitre pose les bases de la réglementation solvabilité 2 en formule standard et en modèle interne ; le deuxième chapitre quant à lui, concerne les traitements effectués sur les données de sinistres à savoir les retraitements de l'inflation et la projection à l'ultime. Une étude pour distinguer la sinistralité attritionnelle de la sinistralité grave sera faite au chapitre 3. Nous allons ensuite modéliser les deux types de sinistralité au chapitre 4. Il ne restera plus qu'à agréger les deux types de sinistralité pour chaque branche de garanties ; puis les sinistralités des différentes branches de garanties. Pour cela deux approches seront utilisées : une approche quantitative (chapitres 5 et 6) et une approche qualitative (chapitre 7). Ces deux approches seront conciliées au chapitre 8 pour choisir la matrice de corrélation idéale.

Cadre de l'assurance non-vie

L'assurance est un système qui permet à une personne morale ou physique de se prémunir contre les conséquences financières et économiques liées à la survenance d'un risque aléatoire. Les sociétés d'assurance se distinguent des entreprises classiques par leur cycle inversé de production. En effet, l'assureur reçoit des primes pour couvrir des risques dont il ne connaît pas le coût (il ne connaît le coût qu'après la réalisation du risque).

Les directives communautaires européennes distinguent deux types d'assurances :

- Les Assurances « Non-Vie » : elles concernent les assurances de biens, les assurances de responsabilité civile et les assurances santé ;
- Les Assurances « Vie » : ce sont les assurances vie, décès, épargne, retraite...

Le périmètre de ce mémoire se limitera aux assurances de biens et de responsabilité civile.

1.1 Réglementation solvabilité 2

Depuis le 1^{er} janvier 2016, les sociétés d'assurance et de réassurance européennes sont soumises à la réglementation solvabilité II ; c'est un ensemble de règles dites « prudentielles »¹ de solvabilité qui portent sur les exigences en matière de solvabilité pour les assureurs et les ré-assureurs, de renforcement de la protection des assurés, visant à inciter les entreprises à améliorer leur connaissance et leur gestion des risques.

La réglementation Solvabilité II consiste en une approche de la solvabilité par les risques ; en effet, elle exige aux sociétés d'assurance de disposer d'un niveau de capital fonction des risques souscrits et de la stratégie d'investissement. Elle s'appuie sur trois piliers.

1.1.1 Le pilier 1 ou exigences quantitatives

Les engagements de l'assureur vis-à-vis des assurés sont représentés par les provisions techniques en valeur de marché ; elles sont estimées de sorte à représenter la valeur la plus juste des engagements de l'assureur. Cependant, en cas de survenance d'un choc exceptionnel, elles pourraient ne pas être suffisantes pour couvrir toutes les prestations ; c'est pour cette raison que l'assureur doit donc disposer d'un niveau de capital supplémentaire pour lui permettre de rester solvable en cas de chocs exceptionnels.

Le pilier 1 exige deux capitaux en fonction du niveau de prudence et/ou de contrôle : le SCR et le MCR.

1. <https://acpr.banque-france.fr>

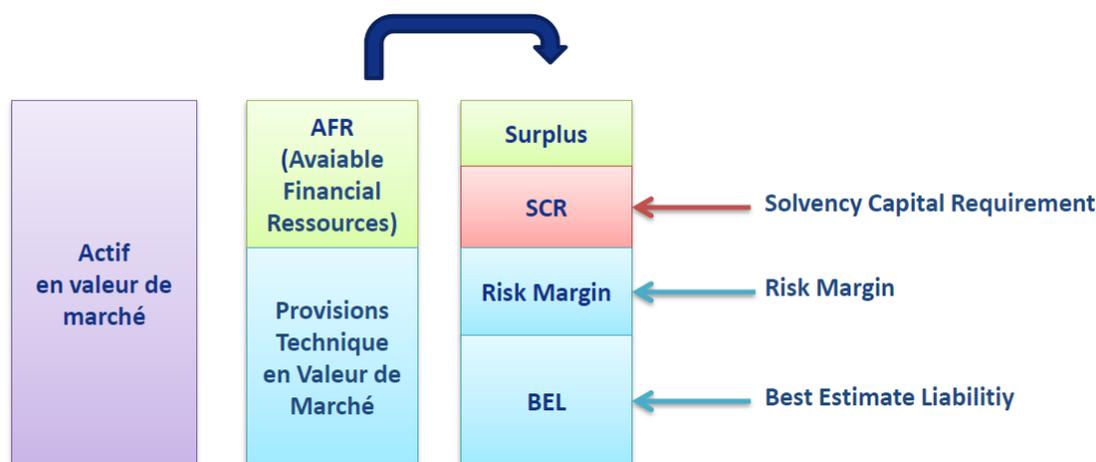


FIGURE 1.1 – Bilan vision Solvabilité 2

Le **SCR (Solvency Capital Requirement)** est le niveau de capital minimum que doit disposer une société d'assurance (en plus des engagements provisionnés) pour absorber des pertes économiques sur un horizon d'un an avec une probabilité de 99,5% ; pertes économiques par opposition aux pertes comptables c'est-à-dire en incluant les pertes non réalisées et les pertes de profits futurs inclus dans la valeur de la compagnie. En d'autres termes, le SCR est l'exigence de capital correspondant à une probabilité de faillite de 0.5% sur un horizon d'un an (scénario bicentenaire).

Le **MCR (Minimum Capital Requirement)** quant à lui représente le niveau de capital requis en dessous duquel L'ACPR (Autorité de Contrôle Prudentiel et de Résolution) pourrait prendre des décisions sévères vis-à-vis de l'assureur ; ceci pouvant aller jusqu'au retrait de l'agrément d'exercer en tant qu'assureur. Le plus souvent, le MCR est égal à un tiers du SCR.

L'**AFR (Available Financial Ressources)** ou capital disponible correspond aux fonds propres de l'entreprise en vision Solvabilité 2. Il est égal à l'actif éligible diminué des engagements valorisés en valeur de marché.

La **valeur de marché des provisions techniques** correspond au prix qu'exigerait une tierce partie pour honorer les engagements à la charge de l'assureur. Elle est constituée du BEL (Best Estimate Liabilities) et de la Risk Margin.

Le **BEL (Best Estimate Liabilities)** représente la meilleure vision économique des engagements de l'assureur. Il correspond à l'espérance mathématiques des cash flows représentant l'engagement de l'assureur tels que les prestations futures, les primes, les coûts associés à la gestion des contrats d'assurance.

La **Risk Margin** est le montant complémentaire qu'il faut ajouter au BEL de sorte à atteindre la valeur de marché des provisions techniques. Elle est généralement évaluée en terme de coût d'immobilisation du capital ; en effet, elle représente la rémunération du capital qu'un repreneur devrait immobiliser pour le portefeuille en run-off jusqu'à extinction totale des engagements.

1.2.2 Le risque de souscription non-vie

Le module « risque de souscription non-vie » reflète le risque qui découle des engagements d'assurance non-vie, compte tenu des périls couverts et des procédés appliqués dans l'exercice de cette activité. Il tient compte de l'incertitude qui pèse sur les résultats des entreprises d'assurance dans le cadre de leurs engagements d'assurance existants, ainsi que du nouveau portefeuille dont la souscription est attendue dans les douze mois à venir.

Le risque de souscription non-vie traite de la volatilité qui entoure les primes, les sinistres futurs, les frais de gestion de sinistres, les provisions, les rachats et la survenance d'événements catastrophiques. Il regroupe l'ensemble des risques pris par un assureur lors de la distribution des contrats d'assurance non-vie. Il comprend les modules de risques suivants :

a) Le risque de prime et réserve

Comme son nom l'indique, il comprend le risque de prime et le risque de réserve qui sont définis dans les paragraphes qui suivent.

Le risque de prime : Il résulte de l'inadéquation du tarif avec la sinistralité anticipée. Ce risque concerne la perte ou des changements défavorables résultant des fluctuations affectant la date de survenance, la fréquence et la gravité des événements assurés. En effet, si la sinistralité future est supérieure à celle anticipée dans le tarif, la compagnie doit faire face à un risque de solvabilité. Ce risque comprend également l'aléa sur le volume de primes anticipé (espéré) et sur les frais et commissions.

Le risque de réserve : C'est le risque lié à l'incertitude sur la capacité de l'entreprise à appréhender correctement la charge ultime de la sinistralité passée. Il provient d'une évolution défavorable des provisions constituées sur les années passées au titre des sinistres déjà survenus qu'ils aient été déclarés ou non. Il est dû à l'incertitude liée au montant des règlements et à la durée de développement des sinistres. Ce risque se mesure au 31/12/N par le boni/mali qui est défini comme la somme du montant de réserves au 31/12/N+1 destinées à couvrir les sinistres survenus avant le 31/12/N et du montant de prestations payées aux assurés l'année N+1 au titre des sinistres survenus avant le 31/12/N soustrait du montant de réserves au 31/12/N destinées à couvrir les sinistres survenus avant le 31/12/N.

b) Le risque de rachat

C'est le risque que les contrats pluriannuels soient résiliés en cours d'années. Ce risque est très courant en assurance vie et a fait l'objet de beaucoup de débats en assurance non-vie. L'EIOPA a confirmé l'existence de ce risque en assurance non vie lors du QIS 5.

c) Le Risque de catastrophe naturelle (CAT)

Il modélise les risques de survenance d'événements de type catastrophes, ne pouvant pas être calibrés sur l'historique de la société qui n'est pas assez long. En d'autres termes, c'est le risque de perte, ou de changement défavorable de la valeur des engagements d'assurance, résultant de l'incertitude importante liée aux événements extrêmes ou exceptionnels, qui pèse sur les hypothèses retenues en matière de prix et de provisionnement (risque de catastrophe en non-vie). Ce risque résulte d'événements naturels extrêmes ou à période de retour très élevée.

1.3 Le risque de prime dans le modèle d'AVIVA : définition du risque de sinistralité future

Le risque de prime tel que défini par le modèle standard est modélisé par AVIVA Assurances en deux modules distincts :

- **Le risque de portefeuille** : il modélise la déviation des primes réalisées par rapport à celles prévues/anticipées.
- **Le risque de sinistralité future ou risque de nouvelle souscription** : il modélise la volatilité liée à la déformation de la prime pure ; c'est l'aléa sur la fréquence de survenance des sinistres et sur la sévérité de ces sinistres. Pour ce risque, AVIVA modélise les sinistres correspondant aux primes qui seront émises pendant l'année qui suit la date d'arrêt.
- **Le risque de frais** : il modélise une déviation des frais par rapport à leur prévision moyennes.

Le risque de sinistralité future est le risque le plus matériel d'AVIVA Assurances ; il représente plus du 50% du SCR de souscription non-vie. C'est ce risque que nous nous proposons de modéliser dans cette étude.

1.3.1 Description des segments de risque

Le risque de sinistralité future est segmenté en 16 segments de risque appelés *Model Calculation Class (MCC)*. Une MCC est un groupe de produits d'assurance non-vie ; les MCC sont intra-homogènes et inter-hétérogènes. Chaque MCC peut être incluse dans une LoB (Line of Business) S2 tel qu'indiqué par le tableau suivant :

MCC	Libellé MCC	LoB S2
MCC01	Santé	Frais médicaux
MCC02	Accident	Protection du revenu
MCC03	Auto-Corp-Part (Automobile corporelle Particuliers)	Moteur (Responsabilité civile)
MCC04	Auto-Corp-Pro (Automobile corporelle Professionnels)	Moteur (Responsabilité civile)
MCC05	Auto-Mat-Part (Automobile Matérielle Particuliers)	Moteur (autres)
MCC06	Auto-Mat-Pro (Automobile Matérielle Professionnels)	Moteur (autres)
MCC07	Dommages-Part (Dommages Particuliers)	Incendie
MCC08	Divers	Divers
MCC09	Dommages-Pro (Dommages Professionnels)	Incendie
MCC10	Pr-Juridique (Protection Juridique)	Protection juridique
MCC11	RC-Part (Responsabilité Civile Particuliers)	Responsabilité civile
MCC12	RC-Pro (Responsabilité Civile Professionnels)	Responsabilité civile
MCC13	Grêle	Incendie
MCC14	Cons-DO (Construction Dommages)	Incendie
MCC15	Cons-RCD (Construction Responsabilité Civile)	Responsabilité civile
MCC16	Cons-Autre (Construction Autre)	Incendie

TABLE 1.1 – Segments de risques

Les garanties couvertes par les différents segments de risque sont les suivantes :

Santé : ce segment couvre les risques liés aux frais de soins de santé.

Accident : il couvre les sinistres causés par les accidents de la vie privée.

Auto-Corp-Part : ce segment concerne les dommages corporels causés par tout type de véhicule motorisé pour les particuliers et tel que la responsabilité civile automobile de l'assuré est engagée.

Auto-Corp-Pro : il couvre les dommages corporels causés par tout type de véhicule motorisé utilisé dans le cadre d'une activité professionnelle chez les professionnels et tel que la responsabilité civile automobile de l'assuré est engagée.

Auto-Mat-Part : il couvre les dommages matériels sur tout type de véhicule motorisé pour les particuliers assurés.

Auto-Mat-Pro : ce segment concerne les dommages matériels sur tout type de véhicule motorisé utilisé dans la cadre d'une activité professionnelle chez les professionnels assurés.

Dommages-Part : il concerne l'assurance habitation personnelle.

Dommages-Pro : couvre le risque dûs aux dommages matériels sur les locaux commerciaux et les stocks mais aussi le risque sur les pertes d'exploitation.

Pr-Juridique : il concerne les risques sur les frais juridiques.

RC_Part : il couvre les risques pour lesquels la responsabilité civile (hors automobile) des particuliers assurés est engagée.

RC_Pro : ce segment couvre les risques pour lesquels la responsabilité civile (hors automobile) des professionnels assurés est engagée par exemple la responsabilité civile des professions médicales, des professions réglementées, des commerçants,...

Grêle : il Couvre les dommages causés par la grêle, le gel ou les inondations sur les produits agricoles.

Cons-DO : c'est une assurance de dommages spécifique à la France couvrant les locataires d'un mauvais travail de construction (garantie 10 ans).

Cons-RCD : c'est une assurance de responsabilité civile spécifique à la France couvrant les entrepreneurs du bâtiment en cas de dommages causés par leurs travaux de construction (10 ans de garantie).

Cons-Autre : couvre les risques liés à la construction mais qui ne sont pas couverts par les segments Cons-DO et Cons-RCD.

Divers : ce segment couvre tous les autres risques qui ne seront compris dans les 15 autres MCCs.

Afin de pouvoir mieux les contrôler, une société d'assurance doit pouvoir mesurer ses risques. Plusieurs outils ont été développés pour mesurer les risques ; ces outils feront l'objet de la section suivante.

1.3.2 Les mesures de risque

Dans le but d'estimer le montant de capital requis pour absorber les pertes futures imprévues, les sociétés d'assurance font le plus souvent appel à des outils très utiles en assurance et en finance : les mesures de risques. La plupart des mesures de risque utilisées actuellement sont des quantités statistiques qui décrivent la distribution de pertes d'un portefeuille sur un horizon déterminé à l'avance.

i) Les axiomes de cohérence

Les propriétés souhaitables d'une mesure de risque ρ définie sur l'espace des variables aléatoires, et prenant ses valeurs dans \mathbb{R} sont :

- **Axiome de monotonie** : Pour tous risques X et Y , si $X \geq Y$ alors $\rho(X) \geq \rho(Y)$. Ceci se traduit par le fait que la mesure de risque doit pouvoir reproduire l'ordre des risques ; en d'autres termes, plus un risque est grand, plus il doit avoir une mesure de risque élevée. On peut également traduire cet axiome par le fait que les plus gros risques doivent avoir les mesures de risques les plus importantes.
- **Axiome d'homogénéité positive** : $\forall \lambda \in \mathbb{R}_+$, $\rho(\lambda X) = \lambda \rho(X)$; cet axiome signifie qu'une bonne mesure de risque doit pouvoir capter l'amplitude du risque ; si un risque est multiplié par 2, sa mesure doit également l'être.
- **Axiome de l'invariance par translation** : $\forall k \in \mathbb{R}$ $\rho(X + k) = \rho(X) + k$; il n'y a pas d'incertitude sur les constantes ; donc l'ajout (ou le retrait) d'une quantité déterministe k au risque augmentera (ou diminuera) sa mesure de risque de k également.
- **Axiome de sous-additivité** : Pour tous risques X et Y , $\rho(X + Y) \leq \rho(X) + \rho(Y)$; c'est le bénéfice de diversification. En effet, les risques agrégés doivent avoir une mesure de risque inférieure ou égale à la somme de leurs mesures de risques individuelles. Ceci trouve tout son sens dans ces propos d'Artzner « une fusion ne crée pas de risque supplémentaire ». On peut également comprendre cet axiome par le fait que la probabilité que deux risques surviennent simultanément est au plus égale à 1 ; plus cette probabilité est faible, plus le bénéfice de diversification est important.

On qualifiera de **mesure cohérente** une mesure qui vérifie ces 4 axiomes.

ii) Exemples de mesures de risque

Plusieurs mesures de risque ont été développées dans la littérature. Nous présenterons ici les mesures de risques les plus utilisées.

ii.1) La variance et l'écart-type

Ce sont des mesures de la dispersion des données par rapport à la tendance centrale. Cependant, elles ne sont pas bien adaptées au contexte de l'assurance car elles sont symétriques et ne permettent pas de faire la distinction entre les « écarts positifs » et les « écarts négatifs ». En effet, elles sont basées sur la distance quadratique.

La variance n'est pas sous additive et n'est pas homogène positive ; en effet, soit X un risque tel que $V(X) \neq 0$, on a :

$$\begin{aligned} V(X + X) &= V(2X) \\ &= Cov(2X, 2X) \\ &= 4V(X) \\ &> 2V(X) = V(X) + V(X) \end{aligned}$$

Donc si $V(X) \neq 0$ alors $V(X + X) > V(X) + V(X)$ et $V(2X) = 4V(X) > 2V(X)$.

L'écart-type est venu corriger la non-homogénéité positive et la non sous-additivité de la variance mais ne permet toujours pas de faire la distinction entre les « déviations positives » et les « déviations négatives ».

ii.2) Value-at-Risk

On appelle Value-at-Risk(VaR) de niveau α d'un risque X , le quantile de niveau α de la distribution de ce risque. On note $VaR(X, \alpha) = x_\alpha$ avec $P(X \leq x_\alpha) = \alpha$. On peut alors écrire $VaR(X, \alpha) = F_X^{-1}(\alpha)$

Elle permet de s'intéresser aux événements rares qui n'arrivent qu'avec une faible probabilité $1 - \alpha$. En assurance, on peut la définir comme le montant qui permettra de couvrir la perte engendrée par le risque avec une probabilité α . En d'autres termes, un assureur qui assure un risque X et qui dispose d'un niveau de capital égal à $VaR(X, \alpha)$ a une probabilité de ruine égale à $1 - \alpha$. Dans le cadre de solvabilité 2 par exemple, on calcule la $VaR(X, 99.5\%)$ c'est-à-dire le montant de perte qui sera dépassé avec une probabilité égale à 0,5%.

Propriété : La VaR est une fonction croissante de α .

Preuve : Soient $\alpha_1, \alpha_2 \in \mathbb{R} / \alpha_1 \leq \alpha_2$. Posons $VaR(X, \alpha_1) = x_{\alpha_1}$ et $VaR(X, \alpha_2) = x_{\alpha_2}$. On a :

$$\begin{aligned} P(X \leq x_{\alpha_1}) = \alpha_1 \leq \alpha_2 = P(X \leq x_{\alpha_2}) &\Rightarrow P(X \leq x_{\alpha_1}) \leq P(X \leq x_{\alpha_2}) \\ &\Rightarrow \{X \leq x_{\alpha_1}\} \subseteq \{X \leq x_{\alpha_2}\} \\ &\Rightarrow x_{\alpha_1} \leq x_{\alpha_2} \text{ ie } VaR(X, \alpha_1) \leq VaR(X, \alpha_2) \end{aligned}$$

Le problème principal de la VaR est qu'elle n'est pas toujours sous-additive . En effet, si on considère deux risques X et Y indépendants de lois de Pareto de paramètres respectifs (2, 1) et (2, 2), les représentations de $VaR(X + Y, \alpha)$ et $VaR(X, \alpha) + VaR(Y, \alpha)$ en fonction de α montrent que les deux courbes se touchent donc :

$$\exists \alpha \in [0, 1] / VaR(X + Y, \alpha) \geq VaR(X, \alpha) + VaR(Y, \alpha)$$

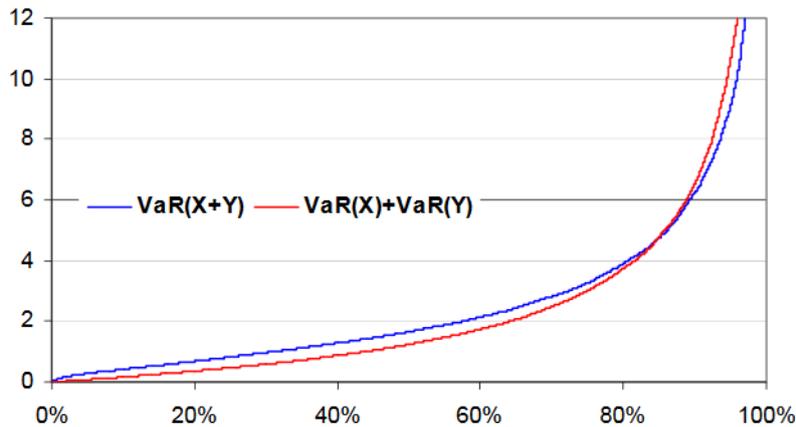


FIGURE 1.3 – La non sous-additivité de la VaR

Bien qu'elle ne soit pas sous-additive, la VaR est la mesure de risque répandue sous la réglementation solvabilité 2. Une alternative à la VaR serait la Tail Value-at-risk définie par $TVaR(X, \alpha) = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR(X, t) dt$; La $TVaR(X, \alpha)$ n'est alors rien d'autre que la moyenne des VaR de niveau supérieur à α ; elle majore la $VaR(X, \alpha)$, est également croissante en α et est **toujours sous-additive**. La TVaR est une mesure de risque cohérente; c'est la mesure de risque utilisée sous la « Swiss Solvency Test ».

1.3.3 Le SCR de sinistralité future

Dans le cadre de la modélisation du risque de sinistralité future et du calcul du SCR associé, une distinction est faite entre sinistralité attritionnelle et sinistralité grave. Pour chaque MCC, un SCR attritionnel et un SCR grave seront calculés.

Les sinistres attritionnels sont plus fréquents et relativement moins coûteux par rapport aux sinistres graves qui ont des coûts relativement plus élevés avec des fréquences de survenance plus faibles.

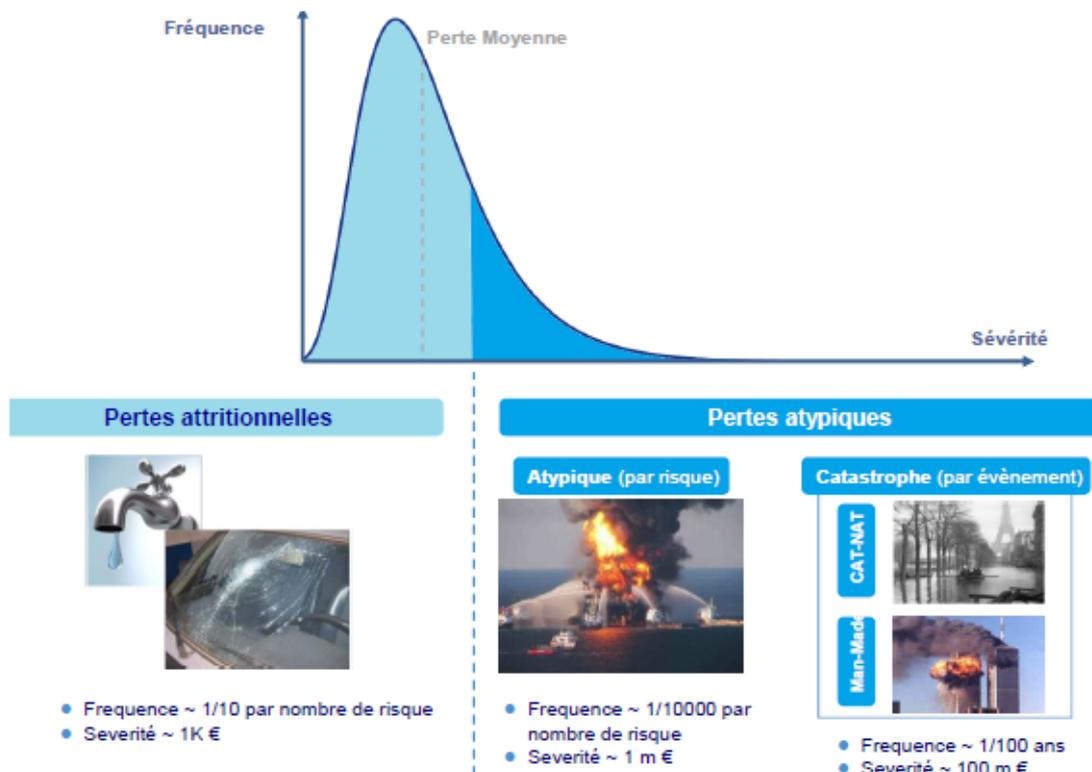


FIGURE 1.4 – Les pertes attritionnelles, les pertes graves et les pertes catastrophiques

Définition du modèle

La charge totale de sinistres S d'une MCC est donnée par :

$$S = \sum_{i=0}^N X_i$$

où N est la variable aléatoire qui représente le nombre de sinistres de la MCC et $(X_i)_{i \geq 1}$ la suite de variables aléatoires représentant les montants individuels des sinistres de la MCC considérée.

La charge totale S se décompose comme suit :

$$S = \sum_{i=0}^N X_i = S_{att} + S_{grave} = \sum_{i=0}^{N_{att}} X_i^{att} + \sum_{i=0}^{N_{grave}} X_i^{grave}$$

Où :

i) S_{att} est la charge de la sinistralité attritionnelle donnée par : $S_{att} = \sum_{i=0}^{N_{att}} X_i^{att}$; N_{att} est la variable aléatoire qui représente le nombre de sinistres attritionnels et X_i^{att} la charge du $i^{\text{ème}}$ sinistre attritionnel. Les sinistres attritionnels étant des sinistres de masses, ils seront modélisés de façon agrégée ; de plus, pour ces sinistres, le nombre de sinistres N_{att} est très lié à l'exposition acquise. Pour cette raison, nous faisons le choix de modéliser directement la « **prime pure attritionnelle** » que nous définissons comme étant le rapport entre la charge attritionnelle S_{att} et l'exposition acquise. On considérera un modèle simpliste dans lequel l'exposition est supposé fixe.

ii) S_{grave} est la charge de la sinistralité grave ; elle est égale à $S_{grave} = \sum_{i=0}^{N_{grave}} X_i^{grave}$ avec N_{grave} la variable aléatoire qui représente le nombre de sinistres graves et X_i^{grave} la charge individuelle du $i^{\text{ème}}$ sinistre grave. Les variables aléatoires $(X_i^{grave})_{i \geq 1}$ sont supposées i.id et indépendantes de N_{grave} . Les sinistres graves sont modélisés de façon individuelle par une approche **fréquence*sévérité**.

Le modèle retenu est alors :

$$S = \sum_{i=0}^N X_i = S_{att} + \sum_{i=0}^{N_{grave}} X_i^{grave}$$

Calcul du SCR

Pour une MCC donnée, le SCR attritionnel s'obtient par une simulation de 10000² observations à partir de la loi de la prime pure attritionnelle ; la $VaR(99,5\%)$ de la prime pure attritionnelle étant la 50^{ème} plus grande valeur parmi les 10000. Le SCR au titre de la sinistralité attritionnelle s'en déduit par la relation :

$$SCR^{att} = Expo * [VaR(99,5\% \text{ des } 10000 \text{ simulations}) - \text{moyenne des } 10000 \text{ simulations}]$$

Pour le SCR de la sinistralité grave, on simule 10000 montants de pertes graves, le SCR des pertes graves étant la 50^{ème} plus grande valeur parmi les 10000 et on a la relation :

$$SCR^{grave} = VaR(99,5\% \text{ des } 10000 \text{ simulations}) - \text{moyenne des } 10000 \text{ simulations}$$

Les SCR ainsi obtenus tant sur les pertes attritionnelles que sur les pertes graves dépendent fortement des 10000 simulations faites ; pour une question de robustesse des résultats, il serait judicieux de répéter le processus un nombre important de fois, par exemple 100 fois ; on obtiendra alors 100 SCR attritionnels (respectivement graves) pour chaque MCC. Le SCR attritionnel (respectivement grave) final de chaque MCC s'obtient en prenant la moyenne des 100 SCR. Le principal avantage de ce procédé est qu'il permet d'obtenir un intervalle de confiance du SCR.

2. Ce nombre de simulations a été choisi pour des contraintes de temps de calcul

Le SCR de chaque segment de risque s'obtient en agrégeant via des structures de dépendance les SCR grave et attritionnel. Les SCR obtenus sur chacun des segments de risques seront à leur tour agrégés via des structures de dépendance pour obtenir le SCR au titre de la sinistralité future.

La réglementation Solvabilité 2 recommande une agrégation linéaire des SCR. En effet, si on dispose de n segments de risques, le SCR total est donné par :

$$SCR = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \rho_{ij} * SCR_{segment_i} * SCR_{segment_j}}$$

Où ρ_{ij} est le coefficient de corrélation linéaire entre les segments de risque i et j .

Démarche

Au regard de la problématique de l'étude ainsi que des éléments présentés dans ce chapitre nous comprenons que pour calculer le SCR de sinistralité future, il est indispensable de :

- Vérifier la qualité des données à disposition les mettre en en as-if et à l'ultime ;
- Déterminer pour chaque segment de risque le seuil qui sépare la sinistralité attritionnelle de la sinistralité grave ;
- Modéliser la sinistralité attritionnelle et la sinistralité grave ;
- Déterminer la structure de dépendance premièrement entre les pertes attritionnelles et graves de chaque segment de risque et deuxièmement entre les différents segments de risque.

Retraitement des données

Tout travail scientifique n'a de poids que si les données utilisées sont de bonne qualité et ont du sens. Dans ce chapitre, nous allons dans un premier temps présenter la base de données qui sera utilisée dans ce mémoire et analyser quelques statistiques descriptives. Dans un second temps, il sera question de retraiter les données de l'inflation et de les projeter à l'ultime pour avoir une vision plus juste du risque.

2.1 Description de la base

La base de données utilisée dans cette étude est celle des sinistres d'AVIVA Assurances. Elle contient tous les sinistres (hors santé) survenus de 2000 à 2018 observés à différentes années d'exercice (du 31/12/2000 au 31/12/2018). Les variables qui figurent dans cette base sont renseignées dans le tableau suivant :

Variables	Libellés
REPORTING_YEAR	Fin d'année d'observation
LIB_CLASS	Libellé d'activité défini pour le modèle interne
CLAIM_IDENTIFIER	Code sinistre
ORIGIN_PERIOD	Fin d'année de survenance du sinistre
PT	Paiement de l'exercice (en €)
PT_A	Paiement de l'exercice antérieur (en €)
SAP	Sinistre à payer (en €)
RE	Recours encaissé sur l'exercice (en €)
RE_A	Recours encaissé sur l'exercice antérieur (en €)
RAE	Recours à encaisser (en €)

TABLE 2.1 – Description de la base

Afin de mieux comprendre l'information que renvoie chacune de ces variables, considérons l'exemple suivant : soit un sinistre survenu le 20/03/2015 ; au 31/12/2018, on a pour ce sinistre : `reporting_year=31/12/2018`, `origin_period=31/12/2015`. `PT` est le montant de règlement effectué en 2018 au titre de ce sinistre ; `PT_A` est le montant cumulé des règlements effectués de 2015 à 2017 au titre de ce sinistre.

Dans le cadre de cette étude, la variable d'intérêt est la charge de sinistre nette de recours calculée comme suit : $Charge_Net_RC=PT+PT_A+SAP-(RE+RE_A)$. Dans la suite, quand

nous parlerons de charge, nous précisons qu'il s'agit de la charge de sinistre nette de recours.

La base contient au total 7,1 millions de sinistres survenus entre 2000 et 2018 ; ces sinistres correspondent à une charge totale de 10,3 milliards d'euro. Afin de se familiariser avec les données, il serait nécessaire de présenter les répartitions du nombre de sinistres et de la charge de sinistres par segments de risque (MCC).

		Nombre		Charge		Coût moyen en k€
		Valeur en milliers	proportion	Valeur en M€	proportion	
Auto_Mat_Part	MCC05	3 491	48,9%	2 634	26%	0,8
Dommages_Part	MCC07	1 102	15,4%	1 901	18%	1,7
Dommages_Pro	MCC09	301	4,2%	1 303	13%	4,3
Auto_Mat_Pro	MCC06	996	13,9%	1 165	11%	1,2
Auto_Corp_Part	MCC03	105	1,5%	1 000	10%	9,6
RC_Pro	MCC12	155	2,2%	484	5%	3,1
Cons_RCD	MCC15	100	1,4%	405	4%	4,0
Grele	MCC13	73	1,0%	392	4%	5,4
Auto_Corp_Pro	MCC04	24	0,3%	364	4%	15,3
Pr_Juridique	MCC10	497	7,0%	199	2%	0,4
RC_Part	MCC11	183	2,6%	171	2%	0,9
Accident	MCC02	58	0,8%	155	2%	2,7
Cons_DO	MCC14	51	0,7%	90	1%	1,8
Cons_Autre	MCC16	5	0,1%	36	0%	7,9
Divers	MCC08	1	0,0%	0	0%	0,6
Total		7 140	1	10 298	1	1,4

TABLE 2.2 – Répartition des données historiques de sinistres

Les cinq segments de risques les plus matériels sont par ordre de matérialité :

- **Automobile Matérielle Particuliers** : Ce segment représente à lui seul 49% du nombre de sinistres et 26% de la charge totale. La charge maximale vaut 7,1 M€ et la charge moyenne vaut 0,8 k€.
- **Dommages Particuliers** : Il représente 15% du nombre de sinistres et 18% de la charge totale. La charge maximale vaut 4,9 M€ et la charge moyenne vaut 1,7 k€.
- **Dommages Professionnels** : Ce segment ne représente que 4% du nombre de sinistres mais a un poids de 13% dans la charge totale. La charge maximale vaut 24,1 M€ et la charge moyenne vaut 4,3 k€.
- **Automobile Matérielle Professionnels** : Il représente 14% du nombre de sinistres et 11% de la charge totale. La charge maximale vaut 9 M€ et la charge moyenne vaut 1,2 k€.
- **Automobile Corporelle Particuliers** : Ce segment ne représente que 1% du nombre de sinistres mais a un poids de 10% dans la charge totale. La charge maximale vaut 20,1 M€ et la charge moyenne vaut 9,6 k€.

Les coûts moyens historiques par année de survenance sont renseignés dans le tableau suivant :

Survenance	Book value (M€)	Nombre en milliers	Coût moyen (k€)
2000	431	396	1,1
2001	452	381	1,2
2002	446	358	1,2
2003	463	360	1,3
2004	412	353	1,2
2005	468	353	1,3
2006	431	355	1,2
2007	452	346	1,3
2008	452	339	1,3
2009	548	380	1,4
2010	519	381	1,4
2011	549	375	1,5
2012	573	385	1,5
2013	624	399	1,6
2014	642	400	1,6
2015	656	391	1,7
2016	737	402	1,8
2017	710	407	1,7
2018	733	382	1,9
Total	10 298	7 140	1,4

TABLE 2.3 – Coûts moyens historiques par année de survenance

Ces coûts moyens ne peuvent pas être comparés entre eux car il s’agit des données historiques non retraitées de l’inflation. Il sera question dans la section suivante de retraiter ces données historiques de l’inflation afin d’obtenir des statistiques comparables.

2.2 Mise en as-if

Les données comptables de deux années ne peuvent pas être directement comparées ; elles doivent être mises sur un même référentiel. En effet, la valeur d’un euro il y a 5 ans n’est pas la même que celle d’un euro aujourd’hui. Une solution est la mise en as-if qui consiste à retraiter les données de l’inflation. Il s’agit en fait d’une translation de tous les sinistres à l’année de projection (2019 ici), afin d’évaluer la charge de chacun de ces sinistres comme s’ils étaient survenus durant l’année de projection.

2.2.1 Méthodologie

Pour la mise en as-if, nous avons le choix entre deux méthodologies :

- **La mise en as-if rétrospective** : qui corrige uniquement l’inflation passée et suppose que l’inflation future n’est pas différente de l’inflation passée
- **La mise en as-if prospective** : elle permet de considérer une inflation future différente de l’inflation passée.

Pour cette étude, le choix s’est porté sur la mise en as-if prospective car elle permet de considérer une inflation future différente de l’inflation passée, surtout que nous ne disposons pas d’arguments solides pour affirmer que l’inflation future serait similaire à l’inflation passée.

Le procédé est le suivant : soit un sinistre k survenu l’année i . Désignons par $X_{i,j}^k$ la charge non cumulée évaluée l’année $i + j$ (développement j) au titre de ce sinistre ; $X_{i,0}^k$ étant la somme du montant réglé l’année de survenance et de la provision constituée au titre de ce sinistre la

même année. Le vecteur des incréments de charges de ce sinistre est alors $(X_{i,0}^k, \dots, X_{i,j}^k, \dots, X_{i,n}^k)$.

La charge $X_{i,j}^k$ mise en as-if est donnée par :

$$X_{i,j}^{k,as-if} = X_{i,j}^k * \frac{I_{2019+j}}{I_{i+j}}$$

Dans ces notations, i varie de 2000 à 2018, j varie de 0 à 18 et I_l désigne la valeur de l'indice d'inflation l'année l . Les indices utilisés seront présentés dans la suite.

Les charges cumulées mises en as-if sont obtenues tout simplement en faisant le cumul des incréments de charges mis en as-if.

2.2.2 Indices utilisés

Tel qu'explicité ci-dessus, la mise en as-if des données historiques est basée sur l'évolution d'un indice. L'indice utilisé doit être approprié au segment de risque considéré.

2.2.2.1 Indice interne : évolution du coût moyen des sinistres

Une première alternative serait de capter l'inflation par l'évolution du coût moyen par année de survenance. Cette méthode consiste à calculer le coût moyen par année de survenance et à ajuster l'évolution du coût moyen par une tendance (linéaire, exponentielle, etc.). L'évolution de l'inflation est captée par cette tendance.

Pour la MCC la plus matérielle (Auto_Mat_Part) par exemple, nous avons représenté le coût moyen par année de survenance. Le coût moyen semble évoluer suivant une tendance exponentielle (au regard du R^2).

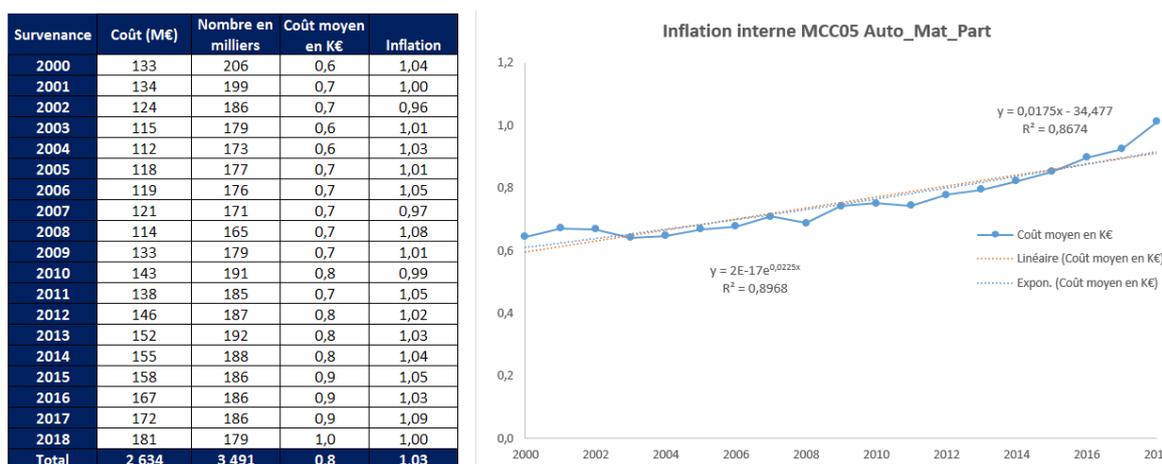


FIGURE 2.1 – Inflation interne du segment Automobile matérielle pour les particuliers

Il ressort des données que le taux d'inflation moyen pour ce segment de risque est de 3%. Ce taux correspond à un peu plus au double de celui observé sur le marché (Sur la période allant de 2000 à 2018, le tau d'inflation moyen calculé à partir des données de l'INSEE¹ est de 1,44%).

1. <https://www.insee.fr>

On constate alors que cette méthode comporte un biais en raison du fait que les sinistres ne se développent pas de la même manière ; mais aussi, en raison du fait que si un sinistre important (c'est-à-dire au coût élevé) survient une année, il aura tendance à augmenter le coût moyen de cette année et par conséquent l'indice d'inflation de cette année aura tendance à augmenter. De plus, le portefeuille peut évoluer au cours du temps ; par exemple le fait d'augmenter le montant franchise augmente, augmentera le coût moyen des sinistres.

2.2.2.2 Indices de marché

L'alternative précédente n'ayant pas été jugée appropriée, nous avons opté de faire recours aux indices de marché. Il s'agit en fait d'utiliser des indices sur lesquels sont indexés les contrats. Ces indices varient d'un segment à un autre. Nos choix ont été appuyés par des jugements d'experts.

a) MCC Accident

L'indice retenu est le **PASS** (Plafond Annuel de la Sécurité Sociale). Le PASS est l'indice généralement utilisé comme base de calcul des avantages sociaux ; il est mis à jour chaque année par les pouvoirs publics et prend effet au 1^{er} janvier. Le PASS 2019 a été fixé à 40 524€². Le PASS est pris en compte pour calculer :

- les indemnités journalières (IJ) pour maladie, accident du travail, maternité, paternité ;
- les pensions d'invalidité ;
- les pensions d'assurance vieillesse du régime général.

Les garanties accidents de la vie privée permettent aux assurés de se couvrir contre les dommages corporels liés aux accidents de la vie privée et entraînant une incapacité / une invalidité permanente importante. C'est pour cette raison que nous avons jugé qu'il serait pertinent d'indexer la garantie «Accident» sur le PASS.

b) Les MCC Automobiles corporelles et de responsabilité civile

L'indice choisi est celui des prix à la consommation (IPC). L'IPC est une mesure de la tendance générale des prix des biens et services que nous consommons. On dit qu'il y a inflation (respectivement déflation) si les prix ont une tendance à la hausse (respectivement à la baisse) et dans ce cas, l'IPC est supérieur (respectivement inférieur) à 100%.

D'après L'INSEE, l'IPC est « l'instrument de mesure de l'inflation. Il permet d'estimer, entre deux périodes données, la variation moyenne des prix des produits consommés par les ménages. L'indice des prix hors tabac sert à indexer de nombreux contrats privés, des pensions alimentaires, des rentes viagères et aussi à indexer le SMIC. »³

On peut donc comprendre à travers cette description que l'IPC est adapté aux MCC Automobiles corporelles et de responsabilité civile car les contrats sur ces garanties sont indexés sur l'IPC (hors tabac).

2. www.ameli.fr

3. <https://www.insee.fr>

c) Les MCC Automobiles matérielles

Nous avons retenu l'IPC évalué sur les services pour l'entretien et la réparation de véhicules tels que : montage de pièces, équilibrage des roues, visite de contrôle, services de dépannage, vidange, graissage, lavage ; mais aussi les coûts de la main d'œuvre et des fournitures. Puisque les services pour l'entretien et la réparation de véhicules font partie des garanties couvertes par les MCC Automobiles matérielles, il est donc logique de les indexer sur l'IPC évalué sur les services pour l'entretien et la réparation de véhicules.

d) Les MCC Dommages

L'indice FFB du coût de la construction a été choisi. En effet, le prix d'un contrat d'assurance habitation par exemple, dépend non seulement de la durée du contrat, mais aussi de l'indice ICC FFB⁴. C'est un indice trimestriel base 1 au 1^{er} janvier 1941, calculé par la FFB (Fédération Française du Bâtiment). Il « enregistre les variations de coût des différents éléments qui entrent dans la composition de l'ouvrage. Ce calcul ne prend pas en compte la valeur des terrains. »⁵ D'après la FFB, l'objet initial de cet indice est l'indexation des polices d'assurance c'est pour cette raison que l'avons trouvé pertinent pour les MCC dommages (car elles traitent aussi des dommages matériels aux locaux commerciaux et d'habitation).

e) MCC Divers et PR-Juridique

L'indice des Prix à la Consommation (IPC) a été retenu pour ces segments de risques. En effet, ces deux branches de garanties sont très génériques.

f) La MCC grêle

L'indice retenu est celui des prix des produits agricoles à la production (IPPAP). Il est calculé à partir des prix mensuels d'un échantillon de produits agricoles représentatifs. D'après l'INSEE, il « permet de mesurer l'évolution des prix perçus par les agriculteurs en contrepartie des produits qu'ils livrent. »⁶

Cet indice a été jugé pertinent pour la MCC grêle car les fortes tempêtes de grêle ont des conséquences néfastes sur l'agriculture et par conséquent entraînent une hausse des prix des produits agricoles (du fait de la rareté de ceux-ci).

g) Les MCC Construction

Le prix d'un contrat d'assurance construction est étroitement lié à la valeur du bâtiment assuré. C'est l'indice BT01 qui est utilisé pour actualiser ou réviser les prix des marchés de construction.

3.2.2.3 Projection des indices de marché

La méthodologie de mise en as-if que nous avons retenue prend en compte l'inflation future. pour chaque indice de marché, nous disposons des informations jusqu'en 2018. Afin de prendre

4. <https://reassurez-moi.fr>

5. <https://www.ffbatiment.fr>

6. <https://www.insee.fr>

en compte l'inflation future, nous allons projeter les indices. Plus concrètement, nous allons estimer les tendances dans les séries des indices et projeter ces indices dans le futur suivant les tendances estimées.

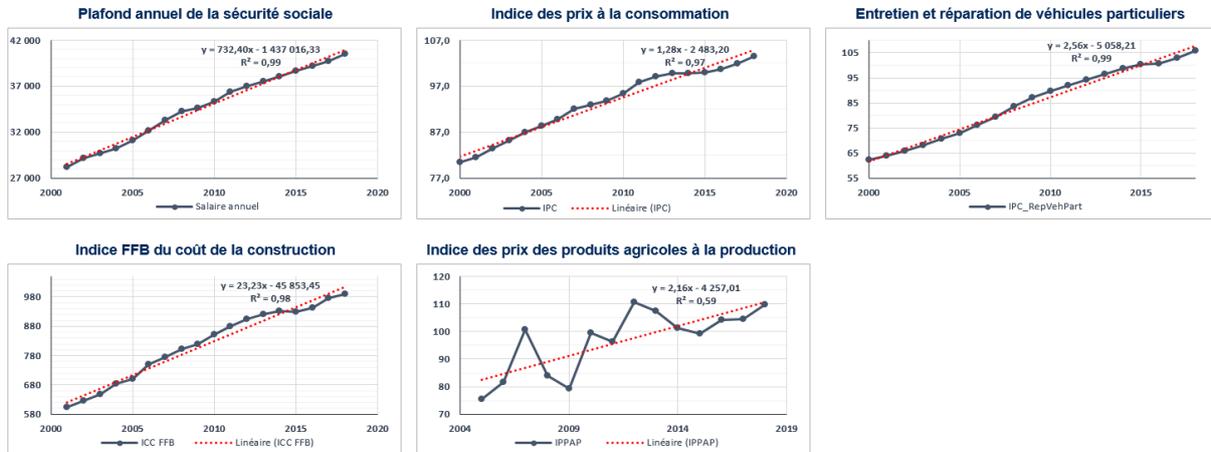


FIGURE 2.2 – Projection des indices de marché

Le R^2 est globalement très bon, à l'exception de l'indice des prix des produits agricoles à la production où les points sont volatiles d'une année à l'autre. Pour chacun de ces indices, nous avons utilisé la tendance linéaire obtenue pour les projeter dans le futur.

2.2.3 Résultats et impact de la mise en as-if

Nous avons ajusté de l'inflation les données de sinistres individuels pour chaque segment de risque par le procédé décrit plus haut. Le taux de variation obtenu est présenté ci-après :

	Book value (M€)	as-if value (M€)	Variation
2000	431	570	32%
2001	452	589	30%
2002	446	567	27%
2003	463	577	25%
2004	412	504	22%
2005	468	562	20%
2006	431	511	19%
2007	452	522	16%
2008	452	516	14%
2009	548	620	13%
2010	519	577	11%
2011	549	597	9%
2012	573	615	7%
2013	624	665	6%
2014	642	683	6%
2015	656	696	6%
2016	737	777	5%
2017	710	739	4%
2018	733	752	3%
Total	10 298	11 638	13,0%

	Book value (M€)	as-if value (M€)	Variation
Accident	155	193	24,5%
Auto_Corp_Part	1 000	1 143	14,3%
Auto_Corp_Pro	364	409	12,3%
Auto_Mat_Part	2 634	2 989	13,5%
Auto_Mat_Pro	1 165	1 305	12,0%
Dommages_Part	1 901	2 136	12,3%
Divers	0,35	0,40	12,6%
Dommages_Pro	1 303	1 467	12,6%
Pr_Juridique	199	222	12,0%
RC_Part	171	194	13,6%
RC_Pro	484	549	13,5%
Grele	392	436	11,1%
Cons_DO	90	100	11,0%
Cons_RCD	405	455	12,2%
Cons_Autre	36	41	15,2%
Total	10 298	11 638	13,0%

TABLE 2.4 – Impact de la mise en as-if

Les résultats par année de survenance sont cohérents car l'impact de l'inflation est décroissant avec la survenance. En effet, pour une vision au 31 décembre 2018, les si-

nistres survenus en 2000 sont plus développés par rapport à ceux survenus en 2018; et par conséquent les sinistres survenus en 2000 sont plus sujets à l'inflation que ceux survenus en 2018.

La mise en as-if a eu un impact plus grand sur les MCC Accident, Automobile matérielle et corporelle pour les particuliers et les MCC de responsabilité.

De façon globale, la mise en as-if a eu un impact à la hausse de 13% sur la charge de sinistre totale.

Le tableau suivant nous donne le détail de l'impact de la mise en as-if sur les cinq segments de risques les plus matériels. L'impact de l'inflation sur ces segments de risque n'est pas très différent de l'impact global.

	MCC5			MCC7			MCC9			MCC6			MCC3		
	Book value (M€)	as-if value (M€)	Variation	Book value (M€)	as-if value (M€)	Variation	Book value (M€)	as-if value (M€)	Variation	Book value (M€)	as-if value (M€)	Variation	Book value (M€)	as-if value (M€)	Variation
2000	133	175	32,0%	62	82	31,8%	47	62	32,2%	43	57	31,7%	52	68	31,8%
2001	134	174	30,1%	63	82	30,3%	50	65	30,7%	44	57	30,0%	66	85	28,8%
2002	124	158	27,3%	74	93	26,8%	63	80	27,1%	44	55	27,1%	56	71	26,5%
2003	115	143	24,6%	99	123	24,1%	62	78	24,9%	46	57	24,4%	50	61	23,7%
2004	112	137	22,0%	61	75	21,9%	53	65	22,4%	46	57	22,0%	47	57	21,4%
2005	118	142	20,1%	74	89	20,0%	55	66	20,1%	47	56	20,1%	59	70	19,8%
2006	119	141	18,4%	74	87	18,3%	48	57	18,4%	49	58	18,3%	45	54	18,2%
2007	121	140	15,3%	77	89	15,3%	47	55	15,3%	48	56	15,3%	45	52	15,6%
2008	114	130	14,2%	86	98	14,3%	53	61	14,2%	48	54	14,2%	36	41	14,1%
2009	133	150	13,1%	124	140	13,2%	75	85	13,3%	49	55	13,1%	62	70	13,4%
2010	143	159	11,2%	115	128	11,2%	61	67	11,3%	58	65	11,0%	38	43	10,9%
2011	138	149	8,5%	120	131	8,6%	53	58	8,5%	55	59	8,5%	52	56	8,4%
2012	146	156	7,0%	129	138	7,1%	67	71	7,0%	57	61	7,0%	55	59	7,3%
2013	152	162	6,3%	122	130	6,2%	96	101	6,2%	62	66	6,4%	42	44	6,3%
2014	155	164	6,2%	132	140	6,2%	87	92	6,2%	73	77	6,3%	53	57	6,3%
2015	158	168	6,1%	108	115	6,1%	96	102	6,0%	81	86	6,1%	77	82	6,4%
2016	167	176	5,4%	134	141	5,4%	89	94	5,4%	97	102	5,4%	70	73	5,3%
2017	172	179	4,2%	122	127	4,2%	91	95	4,2%	110	114	4,2%	50	52	4,1%
2018	181	185	2,6%	124	128	2,6%	111	113	2,6%	108	111	2,6%	45	46	2,6%
Total	2 634	2 989	13,5%	1 901	2 136	12,3%	1 303	1 467	12,6%	1 165	1 305	12,0%	1 000	1 143	14,3%

TABLE 2.5 – Impact de la mise en as-if sur les 5 MCC les plus matérielles

2.3 Projection à l'ultime

La mise à l'ultime est très importante car elle permet d'avoir une vision plus juste du risque. Elle n'est faite que sur les sinistres non clos (provision non nulle). En ce qui concerne les sinistres clos, la valeur à l'ultime est égale à la dernière valeur connue inflatée.

Pour la projection à l'ultime des sinistres non clos, nous avons utilisé la méthode de Chain-Ladder; c'est une méthode qui s'applique aux triangles de charges cumulées ou de règlements cumulés. L'objectif étant de dérouler le triangle de charges cumulées et d'en déduire le montant de provision à constituer.

L'étape incontournable de la méthode de Chain-Ladder est le calcul des facteurs de développement.

2.3.1 Calcul des facteurs de développement

Ce calcul sera effectué pour chaque segment de risque sur les charges cumulées de sinistres mises en as-if. Désignons par $C_{i,j}$ la charge cumulée mise en as-if évaluée l'année 2000 + $i + j$ (développement j) au titre des sinistres survenus l'année 2000 + i ; avec $i, j = 0, \dots, n - 1$; où $n + 1$ correspond au nombre d'années de survénance. Dans cette étude, on a 19 années de survénance; par conséquent $n = 18$. Les facteurs de développement individuels sont donnés par $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}, j = 0, \dots, n - 1$.

L'hypothèse du modèle : Le modèle de Chain-Ladder repose sur l'hypothèse selon laquelle le ratio $\frac{C_{i,j+1}}{C_{i,j}}$ ne dépend pas de l'année de survénance i . Cette hypothèse est équivalente à :

- **H1 :** $\forall j = 0, \dots, n - 1$, les facteurs de développement individuels $f_{i,j}$ ne dépendent pas de l'année d'origine i . On peut alors écrire $f_j = \frac{C_{i,j+1}}{C_{i,j}} \forall i$.
- **H2 :** $\forall j = 0, \dots, n - 1$, on a les égalités : $\frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = \dots = \frac{C_{17,j+1}}{C_{17,j}}$.

Sous ces hypothèses, un estimateur sans biais du facteur de développement est donné par :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}, j = 0, \dots, n - 1$$

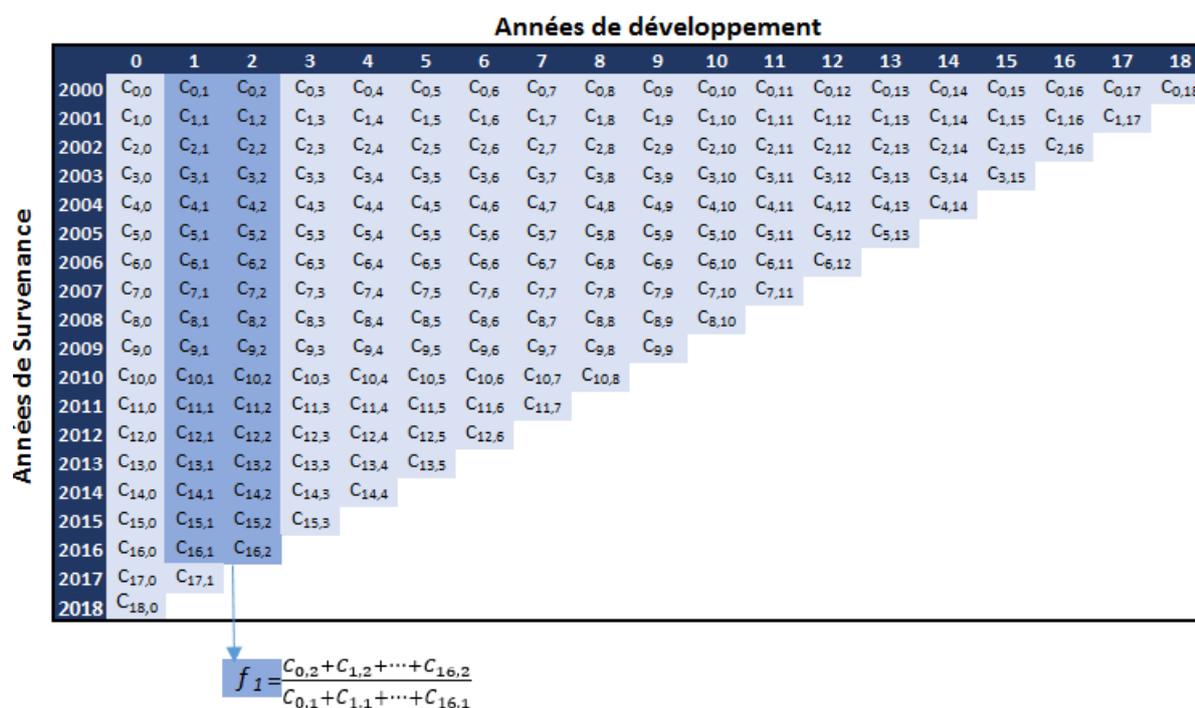


TABLE 2.6 – Triangle de développement

La méthode déterministe de Chain-Ladder permet alors d'en déduire les estimations des charges cumulées futures : $\forall j \geq n - i + 1, \hat{C}_{i,j} = C_{i,n-i} \prod_{k=n-i}^{j-1} f_k$

Validation de l'hypothèse du modèle :

L'hypothèse H1, suppose que $\forall j = 0, \dots, n - 1, C_{i,j+1} = f_j * C_{i,j}$; donc les couples

$(C_{i,j}, C_{i,j+1})_{i=0,\dots,n-j-1}$ doivent être alignés sur une droite d'ordonnée à l'origine 0. Le graphique suivant concerne le test fait sur la MCC05 (Auto_Mat_Part).

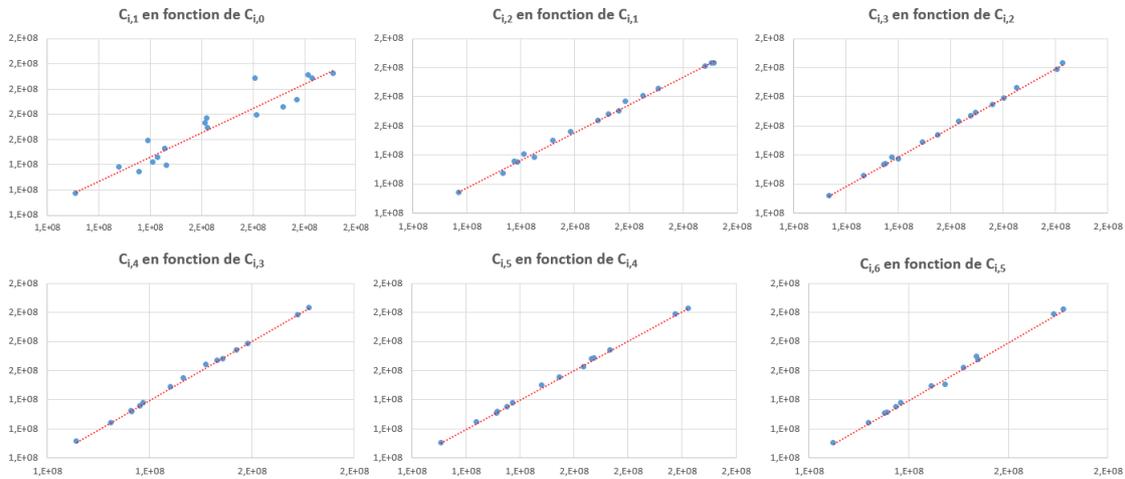


FIGURE 2.3 – Test de l’hypothèse H1 du modèle de Chain-Ladder

Pour l’année de développement 0, l’hypothèse H1 semble ne pas être totalement vérifiée ; mais à partir de l’année de développement 1, on voit que les points sont quasiment alignés sur une droite qui passe par l’origine. Les R^2 associés aux régressions linéaires des vecteurs formés des couples $(C_{i,j}, C_{i,j+1})_{i=0,\dots,n-j-1}$ sont supérieurs à 90%. Les statistiques de Fisher associées sont toutes significatives au seuil de 5%. Nous validons ainsi l’hypothèse H1 de Chain-Ladder pour la MCC Auto-Mat-Part.

Nous avons également vérifié cette hypothèse sur tous les segments de risque ; il en est ressorti que cette hypothèse est également vérifiée pour les autres segments de risques.

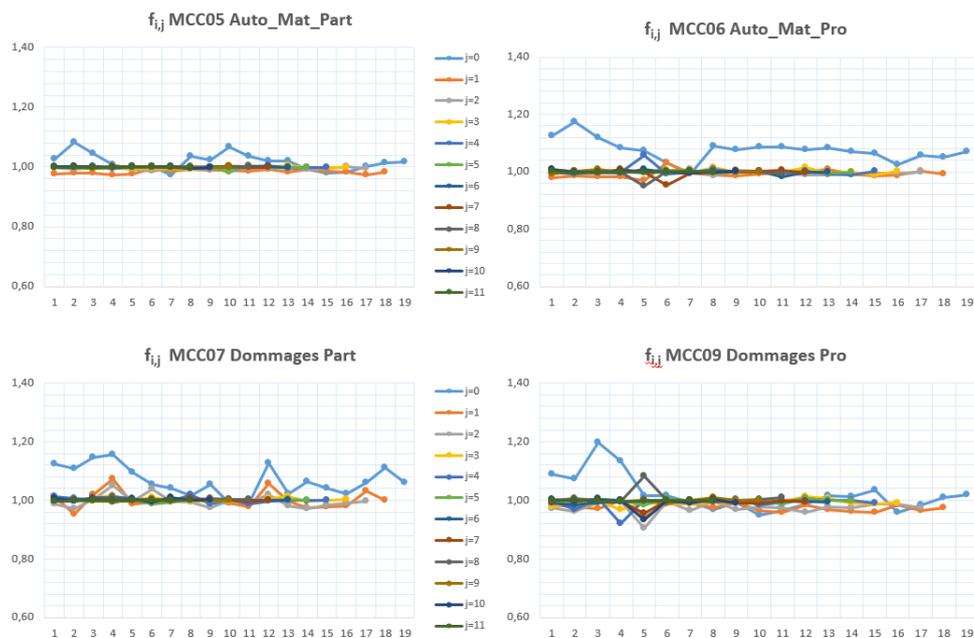


FIGURE 2.4 – Test de l’hypothèse H2 du modèle de Chain-Ladder

L'hypothèse H2 suppose que $\forall j = 0, \dots, n - 1$, les facteurs de développement ($f_{i,j}, i = 0, \dots, n - 1$) sont identiques. Les coefficients de développement individuels des quatre MCC les plus matérielles sont représentés sur le graphique ci-dessus. Pour chacune de ces MCC, les coefficients de développement sont sensiblement constants à partir du développement 1. Nous avons obtenu le même résultat sur les autres segments de risques. Cette hypothèse étant équivalente à l'hypothèse H_1 , nous pouvons affirmer au seuil de 5% que $\forall j = 0, \dots, n - 1, \forall i = 0, \dots, n - 1, f_{i,j} = f_j$ Nous validons ainsi l'hypothèse H2 du modèle de Chain-Ladder.

Les coefficients de développement de Chain-Ladder calculés sur les triangles de développement des différentes branches de garanties sont donnés dans le tableau ci-après.

fj	MCC5	MCC7	MCC9	MCC6	MCC3	MCC2	MCC4	MCC8	MCC10	MCC11	MCC12	MCC13	MCC14	MCC15	MCC16
0	1,016	1,061	1,019	1,071	1,424	1,208	1,502	1,060	0,888	0,880	1,277	1,051	1,016	1,216	1,603
1	0,984	1,001	0,974	0,992	1,049	1,023	1,025	0,822	1,055	1,001	1,112	0,996	1,020	1,116	1,307
2	0,995	0,996	0,976	0,997	0,979	1,009	0,981	0,855	1,071	1,016	1,089	0,996	1,038	1,105	1,100
3	0,999	1,002	0,992	1,001	1,003	0,999	1,016	0,876	1,026	1,017	1,032	0,996	1,031	1,064	1,023
4	1,000	1,001	0,990	1,001	1,000	1,006	1,016	1,015	1,009	1,005	1,016	0,999	1,004	1,042	1,072
5	0,998	0,998	0,996	1,000	1,007	1,004	0,992	1,087	1,004	0,985	1,007	0,999	1,031	1,026	1,004
6	1,000	1,001	0,994	0,999	0,994	1,003	1,029	0,915	1,001	0,995	0,990	1,000	1,024	1,019	1,008
7	1,000	1,000	0,993	0,996	0,995	0,998	1,006	0,995	0,999	0,983	0,993	1,000	1,060	1,009	1,005
8	1,000	1,003	1,008	0,994	1,000	1,001	1,007	1,020	0,999	1,001	0,992	1,000	1,003	1,006	1,043
9	1,000	1,000	0,999	1,000	1,005	0,999	1,003	1,009	0,998	0,993	1,002	1,000	0,998	0,999	1,044
10	1,000	1,002	0,991	1,001	1,013	1,000	0,970	0,980	0,998	0,996	1,014	1,000	1,020	0,981	0,995
11	1,000	1,000	0,999	1,000	1,021	1,004	0,998	1,000	0,999	1,003	0,975	1,000	1,007	0,993	0,996
12	1,000	1,002	0,999	1,000	1,012	1,000	1,002	1,000	1,000	0,998	0,989	1,000	0,999	0,992	0,951
13	1,000	1,000	1,002	1,000	1,007	1,000	1,000	0,970	1,000	1,001	1,003	1,000	1,012	1,000	1,066
14	1,000	1,002	1,000	1,000	1,003	1,001	1,007	1,055	0,999	1,002	1,002	1,000	1,000	0,997	1,140
15	1,000	1,000	1,000	1,001	1,000	1,002	1,028	1,000	0,998	1,002	0,999	1,000	0,998	1,005	1,001
16	1,000	1,000	0,999	1,000	1,001	1,000	1,000	1,000	0,999	1,002	0,996	1,000	1,020	0,995	1,002
17	1,000	1,000	1,000	1,000	1,001	1,000	1,007	1,000	1,000	1,000	0,988	1,000	0,999	1,000	1,000

TABLE 2.7 – Coefficients de développement

2.3.2 Résultats de la projection à l'ultime

L'hypothèse de Chain-Ladder étant validée, nous allons utiliser les coefficients de développement calculés précédemment, pour la projection à l'ultime.

Nous avons projeté à l'ultime les données de sinistres corrigées de l'inflation. Le procédé est le suivant : Soit un sinistre k survenu l'année $2000 + i$; si le sinistre est clos c'est-à-dire $SAP^k = 0$, on reconduit la dernière valeur connue du sinistre corrigée de l'inflation comme sa valeur ultime ; sinon on utilise les coefficients de développement ; Ceci peut se traduire par la formule suivante :

$$\hat{C}_{i,n}^{k,ultime} = \begin{cases} C_{i,n-i}^{k,inflate} & Si \quad SAP_{i,n-i}^k = 0 \\ C_{i,n-i}^{k,inflate} * \prod_{j=n-i}^{n-1} f_j & Si \quad SAP_{i,n-i}^k \neq 0 \end{cases}$$

Le résultat global est résumé dans le tableau ci-dessous. On peut y voir que la projection à l'ultime a un impact plus prononcé sur les années les plus récentes ; ceci est tout à fait cohérent car ces années sont moins développées.

La projection à l'ultime impacte à la hausse de montant total de sinistres corrigé de l'inflation

de 1,33%. L'impact conjoint de la correction de l'inflation et de la projection à l'ultime sur le montant total des données de sinistres historiques est de 14,52% à la hausse.

	Book value (M€)	as-if value (M€)	Ultimate Value (M€)	Variation Book -> as-if	Variation as-if -> ultimate	Variation Book -> ultimate & as-if
2000	431	570	570	32,23%	0,00%	32,23%
2001	452	589	588	30,26%	-0,02%	30,23%
2002	446	567	567	27,22%	-0,02%	27,20%
2003	463	577	578	24,57%	0,09%	24,68%
2004	412	504	505	22,27%	0,23%	22,55%
2005	468	562	564	20,15%	0,37%	20,60%
2006	431	511	513	18,51%	0,41%	18,99%
2007	452	522	524	15,51%	0,40%	15,97%
2008	452	516	518	14,35%	0,34%	14,73%
2009	548	620	623	13,24%	0,49%	13,80%
2010	519	577	580	11,21%	0,39%	11,65%
2011	549	597	600	8,62%	0,61%	9,28%
2012	573	615	617	7,21%	0,41%	7,65%
2013	624	665	667	6,43%	0,32%	6,77%
2014	642	683	687	6,33%	0,68%	7,06%
2015	656	696	705	6,14%	1,30%	7,51%
2016	737	777	790	5,43%	1,55%	7,07%
2017	710	739	758	4,16%	2,51%	6,78%
2018	733	752	839	2,56%	11,63%	14,49%
Total	10 298	11 638	11 793	13,01%	1,33%	14,52%

TABLE 2.8 – Résultat de la projection à l'ultime

Nous avons à ce niveau pour chaque segment de risque, des données corrigées de l'inflation et projetées à l'ultime. Cependant, nous ne pouvons pas encore faire une modélisation de ces données car nous ne savons pas encore lesquels de nos sinistres peuvent être qualifiés de sinistres attritionnels ou de sinistres graves. Cette étude fera l'objet du chapitre suivant.

Étude du seuil de sinistre grave par branche de garanties

La théorie des valeurs extrêmes est une branche de la statistique qui s'intéresse aux queues de distribution. Autrement dit, la théorie des valeurs extrêmes permet de détecter et de modéliser les comportements atypiques c'est-à-dire ceux qui sont très différents de la tendance générale (ventre de la distribution). L'objectif étant souvent de répondre aux problématiques du type :

- Déterminer le seuil à partir duquel on est dans les valeurs extrêmes ;
- Quelle est la probabilité p de dépasser une valeur donnée ;
- Calculer le seuil qui sera dépassé avec une probabilité p très petite ;
- Recherche d'une dépendance entre les valeurs exceptionnelles.

Dans le cadre du risque de sinistralité future, la modélisation distincte de la sinistralité attritionnelle et de la sinistralité grave fait appel à la théorie des valeurs extrêmes en ce sens où il est primordial de déterminer le seuil qui sépare les deux sinistralités.

3.1 Théorie des valeurs extrêmes

Il s'agira dans cette section d'explicitier la théorie autour de la détermination du seuil à partir duquel on peut qualifier un sinistre de « sinistre grave ».

Il existe deux approches en théorie des valeurs extrêmes : l'approche par « maximum » et l'approche par « excès au delà d'un seuil ».

3.1.1 Loi du maximum

Soient X_1, X_2, \dots, X_n , n variables aléatoires indépendantes et identiquement distribuées suivant une loi F . Si on note $S_n = X_1 + X_2 + \dots + X_n$ et si F admet un moment d'ordre 2, alors le théorème centrale limite garantit que :

$$\frac{S_n - nE(X)}{\sigma\sqrt{n}} \xrightarrow[n \rightarrow +\infty]{\text{loi}} N(0, 1)$$

Lorsque le moment d'ordre 2 n'existe pas, le théorème centrale limite généralisé garantit que pour deux suites a_n et b_n bien choisies, on peut connaître le comportement asymptotique de $\frac{S_n - a_n}{b_n}$.

Ce théorème peut être transposé à la loi du maximum $M_n = \text{Max}(X_1, X_2, \dots, X_n)$ centré et normalisé, sauf que la loi limite n'est plus nécessairement normale.

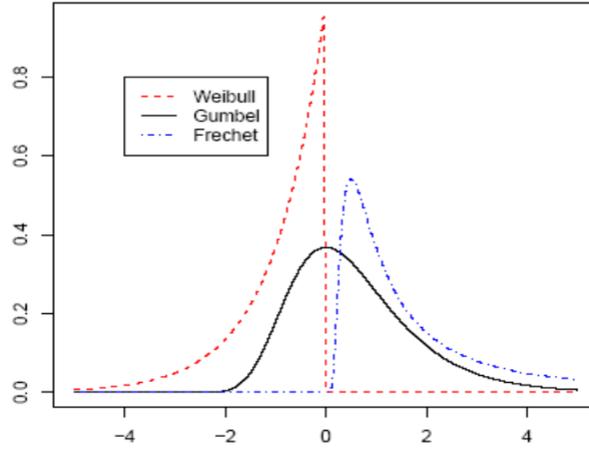
Théorème de Fischer et Tippett

S'il existe deux suites de normalisation $(a_n)_{n \in \mathbb{N}}$ (avec $a_n > 0 \forall n \in \mathbb{N}$) et $(b_n)_{n \in \mathbb{N}}$ et une distribution non dégénérée G telles que :

$$\forall x, \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F(a_n x + b_n)^n \xrightarrow{n \rightarrow +\infty} G(x)$$

alors G est du même type que l'une des trois distributions suivantes :

- Loi de Fréchet ($\alpha > 0$) : $\Phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0. \end{cases}$
- Loi de Weibull ($\alpha > 0$) : $\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0. \end{cases}$
- Loi de Gumbel : $\Lambda(x) = \exp(-\exp(-x)) \quad x \in \mathbb{R}$



Distribution généralisée des valeurs extrêmes

Von Mises (1954) et Jenkinson (1955) ont proposé la distribution $GEV(\mu, \sigma, \xi)$ qui regroupe les trois précédentes dont la fonction de répartition est définie par :

$$G(x) = \begin{cases} \exp(-[1 + \xi(\frac{x-\mu}{\sigma})]_+^{-1/\xi}) & \text{si } \xi \neq 0 \\ \exp(-\exp(-(\frac{x-\mu}{\sigma}))) & \text{si } \xi = 0. \end{cases}$$

ξ est le paramètre de forme, μ le paramètre de position et σ est le paramètre d'échelle.

pour $\xi \neq 0$, on notera $G_\xi(x) = \exp(-(1 + \xi x)_+^{-1/\xi})$.

Les distributions GEV ont la particularité d'être **max-stables** ; en effet si X_1, X_2, \dots, X_n sont des variables aléatoire IID de loi $GEV(\mu, \sigma, \xi)$, alors $\exists (a_n > 0)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites de normalisation telles que :

$$\frac{M_n - b_n}{a_n} \stackrel{d}{=} X_1$$

Il ne s'agit plus ici d'une convergence en loi comme dans le théorème de Fischer et Tippett, mais d'une égalité des distributions et ceci pour tout n .

Reformulation du théorème de Fischer et Tippett

S'il existe $(a_n > 0)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites de normalisation telles que :

$$\forall x, \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow +\infty} G(x)$$

pour une distribution non dégénérée G , alors G est du même type qu'une distribution $GEV(\mu, \sigma, \xi)$.

Domaine d'attraction

On dit qu'une distribution F appartient au domaine d'attraction d'une distribution G et on note $F \in D(G)$ s'il existe $(a_n > 0)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ deux suites de normalisation telles que :

$$\forall x, \mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{n \rightarrow +\infty} G(x)$$

Pour une distribution $GEV(\mu, \sigma, \xi)$, le paramètre ξ caractérise le comportement de la distribution dans les extrêmes, il est aussi appelé l'indice de queue. Si $\xi > 0$ alors on a une queue épaisse (domaine d'attraction de Fréchet ; exemple : Pareto, Student, Fréchet) ; si $\xi = 0$ alors on a une queue intermédiaire (domaine d'attraction de Gumbel ; exemple : Normale, Exponentielle, Gamma, Log-normale) ; si $\xi < 0$ alors on a une queue fine (domaine d'attraction de Weibull ; exemple : Uniforme, Bêta).

En assurance non-vie, les lois usuelles pour la modélisation des sinistres graves correspondent au cas $\xi \geq 0$. Une technique pour savoir si on est dans le cas $\xi = 0$ ou $\xi > 0$ consiste à tracer les graphes suivants :

- **Graphe pour $\xi > 0$**

La loi de Fréchet a pour fonction de répartition $\Phi_\alpha(x) = \exp(-x^{-\alpha})$ ($x > 0$) ; il serait équivalent d'écrire : $\ln(\ln(\frac{1}{\Phi_\alpha(x)})) \simeq \alpha \ln(x) + \beta$. Autrement dit, $\ln(\ln(\frac{1}{\Phi_\alpha(x)}))$ est linéaire en $\ln(x)$. Il suffira alors de représenter $\ln(\ln(\frac{1}{\Phi_\alpha(x)}))$ en fonction de $\ln(x)$.

- **Graphe pour $\xi = 0$**

La loi de Gumbel a pour fonction de répartition $\Lambda(x) = \exp(-\exp(-x))$; il serait équivalent d'écrire : $\ln(\ln(\frac{1}{\Lambda(x)})) \simeq \alpha x + \beta$. Autrement dit, $\ln(\ln(\frac{1}{\Lambda(x)}))$ est linéaire en x . Il suffira alors de représenter $\ln(\ln(\frac{1}{\Lambda(x)}))$ en fonction de x .

3.1.2 Loi des dépassements d'un seuil

Plutôt que de s'intéresser à la loi du maximum, on s'intéresse ici à la loi des observations qui dépassent un seuil. Cette approche a été proposée suite aux critiques de l'approche précédente. En effet l'approche par « loi du maximum » tient compte d'une seule observation (la plus grande) ; il y a alors une perte de l'information contenue dans les autres plus grandes valeurs.

Soient X_1, X_2, \dots, X_n , n variables aléatoires indépendantes et identiquement distribuées suivant une loi F . Soit un seuil $u > 0$; la distribution conditionnelle de survie au delà du seuil u est définie par :

$$\bar{F}_u(x) = \mathbb{P}(X - u > x / X > u) = \frac{\bar{F}(x + u)}{\bar{F}(u)}$$

Pour un seuil u approprié, les observations $(X_i - u)_+, i = 1, \dots, n$ sont caractérisées par une distribution de Pareto généralisée.

Distribution de Pareto Généralisée

Elle est notée $GPD(\beta, \xi)$ et est définie par : $G_{\xi, \beta}(x) = \begin{cases} 1 - [1 + \xi(x/\beta)]_+^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - e^{-x/\beta} & \text{si } \xi = 0. \end{cases}$ Elle

est définie sur $[0, +\infty[$ si $\xi > 0$ et sur $[0, -\beta/\xi[$ si $\xi < 0$.

Le seuil u doit être suffisamment grand pour qu'il y ait convergence de la loi des dépassements vers une GPD, mais il doit aussi être suffisamment petit pour qu'il y ait un nombre suffisant d'observations pour bien estimer les paramètres de la GPD.

Quelques propriétés la Distribution de Pareto Généralisée

- Si $X \sim GPD(\beta, \xi)$ et $\xi < 1$, alors $E(X) = \frac{\beta}{1-\xi}$
- Si $X \sim GPD(\beta, \xi)$ et $\xi < 1$, alors $X - u | X > u \sim GPD(\beta + \xi u, \xi)$ et $E(X - u | X > u) = \frac{\beta + \xi u}{1-\xi}$. Cette propriété signifie tout simplement que lorsqu'il y a convergence vers une GPD, la fonction de dépassement moyen $e(u) = E(X - u | X > u)$ est linéaire de pente $\frac{\xi}{1-\xi}$.
- **Seuil stabilité** : si pour un seuil u , $(X - u)_+ \sim GPD(\beta_u, \xi)$, alors $\forall v \geq u$, $(X - v)_+ \sim GPD(\beta_v, \xi)$ avec $\beta_v = \beta_u + \xi(v - u)$

Théorème de Pickands

Une distribution F appartient au domaine d'attraction d'une loi GEV de paramètre de forme ξ si et seulement s'il existe une fonction positive $\beta(\cdot)$ telle que :

$$\sup_{0 < x < X^F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| \xrightarrow{u \rightarrow x^F} 0$$

En d'autres termes, pour un seuil assez grand, la loi de Pareto généralisée est une bonne approximation de la loi des excès.

Liens entre la Distribution de Pareto Généralisée et la distribution GEV

Soient $\xi, x \in \mathbb{R}$, les propositions suivantes sont équivalentes :

- Il existe deux suites $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ telles que $:\mathbb{P}(\frac{M_n - a_n}{b_n} \leq x) \xrightarrow{n \rightarrow +\infty} G_\xi(x)$
- Il existe une fonction $a(\cdot)$ telle que :

$$\mathbb{P}\left(\frac{X - u}{a(u)} \geq x\right) \xrightarrow{u \rightarrow x^F} \begin{cases} [1 + \xi(x)]_+^{-1/\xi} & \text{si } \xi \neq 0 \\ e^{-x} & \text{si } \xi = 0. \end{cases}$$

- Il existe une fonction positive $\beta(\cdot)$ telle que :

$$\sup_{0 < x < X^F - u} |F_u(x) - G_{\xi, \beta(u)}(x)| \xrightarrow{u \rightarrow x^F} 0$$

Il en découle que l'indice de queue ξ est supposé être le même que l'on utilise une approche GEV ou GPD.

3.2 Estimation de l'indice de queue

L'indice de queue ξ contient l'information sur l'épaisseur de la queue de distribution. Il donne donc une indication sur l'importance des risques extrêmes pour une distribution. C'est pourquoi, il est bien de l'estimer. Afin de ne pas perdre l'information contenue dans les grandes valeurs, on préférera une approche GPD à une approche GEV dans cette étude.

Estimation par maximum de vraisemblance : Soient X_1, \dots, X_n , n variables aléatoires *i.i.d* de fonction de répartition F . Soit k le nombre d'excès obtenus avec un seuil u et Y_1, \dots, Y_k les excès associés. La log-vraisemblance des excès par rapport à une GPD(β, ξ) s'écrit :

$$\log L = -k \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^k \log \left(1 + \xi \frac{Y_i}{\beta}\right)$$

ξ et β s'en déduisent en résolvant numériquement le système d'équations obtenu suite à l'annulation des dérivées partielles de la log-vraisemblance.

Estimateurs de Hill et de Pickands : Soient X_1, \dots, X_n , n variables aléatoires *i.i.d* de fonction de répartition F . Soit $X_{i,n}$ la $i^{\text{ème}}$ statistique d'ordre de l'échantillon de taille n . On s'intéresse aux k plus grandes valeurs de notre échantillon ($X_{k,n} \leq \dots \leq X_{1,n}$). Les estimateurs de Hill et de Pickands sont donnés par :

$$\xi_{k,n}^{Hill} = \frac{1}{k} \sum_{j=1}^k \ln \left(\frac{X_{n-j+1,n}}{X_{n-k,n}} \right)$$

$$\xi_{k,n}^{Pickands} = \frac{1}{\ln(2)} \ln \left(\frac{X_{n-k,n} - X_{n-2k,n}}{X_{n-2k,n} - X_{n-4k,n}} \right)$$

Lorsque $n \rightarrow +\infty$, $k \rightarrow +\infty$ et $\frac{k}{n} \rightarrow 0$; ces estimateurs convergent vers des lois normales. En effet,

$$\sqrt{k} (\xi_{k,n}^{Hill} - \xi) \rightarrow N(0, \xi^2)$$

$$\sqrt{k} (\xi_{k,n}^{Pickands} - \xi) \rightarrow N\left(0, \xi^2 \frac{2^{2\xi+1} + 1}{(2(2^\xi - 1) \ln(2))^2}\right)$$

Remarque :

- L'estimateur de Hill n'est valable que pour le domaine d'attraction de Fréchet ($\xi > 0$), celui de Pickands par contre est valable pour tous les domaines d'attraction.
- L'estimateur de Hill a une variance asymptotique plus faible que celle de l'estimateur de Pickands; c'est l'estimateur d'indice de queue le plus utilisé en pratique.

Les estimateurs de Hill et Pickands sont fonction du nombre d'excès k à considérer. Ce nombre d'excès est étroitement lié au choix du seuil u qui, si très élevé, conduirait à une estimation peu robuste des paramètres de la GPD; par contre si u n'est pas suffisamment élevé, l'approximation de la loi des excès par une GPD risque d'être mauvaise. Il y a ainsi un arbitrage biais/robustesse.

3.3 Méthodes de détermination du seuil

La distribution GPD est très utilisée pour modéliser les dépassements de seuils. Le point crucial est la détermination du seuil u qui doit être suffisamment grand pour qu'il y ait convergence de la loi des dépassements vers une GPD, mais il doit aussi être suffisamment petit pour qu'il ait un nombre d'observations suffisant pour bien estimer les paramètres de la GPD. En pratique, plusieurs outils sont utilisés pour déterminer le seuil u .

3.3.1 Le graphique des dépassements moyens ou Mean Excess Plot

Il s'agit ici de faire varier le seuil, de calculer à chaque fois le dépassement moyen et de tracer le graphique des dépassements moyen en fonction des différents seuils. Le seuil approprié est celui à partir duquel le graphique est approximativement linéaire. L'indice de queue s'en déduit par la relation : $a = \frac{\xi}{1-\xi}$; où a est la pente de la droite. Il est important de préciser qu'en pratique, une approximation du dépassement moyen $e(u)$ sur un échantillon X_1, X_2, \dots, X_n est

$$\text{donnée par : } e_n(u) = \frac{\sum_{i=1}^n (X_i - u)_+}{\sum_{i=1}^n 1_{\{X_i > u\}}}.$$

Le graphe de $e(u)$ permet d'avoir une information sur la queue de distribution. En effet, si $e(u)$ est constante, alors on a une distribution de type exponentielle ; si $e(u)$ est une droite croissante, on a une distribution de type Pareto.

3.3.2 Le graphique de stabilité des paramètres d'échelle et de queue

La propriété de **seuil stabilité** présentée à la section 3.1.2 stipule que : si pour un seuil u , $(X - u)_+ \sim GPD(\beta_u, \xi)$, alors $\forall v \geq u$, $(X - v)_+ \sim GPD(\beta_v, \xi)$ avec $\beta_v = \beta_u + \xi(v - u)$.

En posant $\beta^* = \beta_v - \xi v$ on a $\beta^* = \beta_u + \xi(v - u) - \xi v = \beta_u - \xi u$ qui ne dépend pas de v .

Donc si $(X - u)_+ \sim GPD(\beta_u, \xi)$ pour un seuil u , alors $\forall v \geq u$ les paramètres (β^*, ξ) seront relativement constants. Les paramètres estimés $\hat{\beta}^*$ et $\hat{\xi}$ sont alors représentés sur un graphique pour différents seuils, le seuil à choisir est le seuil minimal à partir duquel les coefficients sont stables.

3.3.3 Le Hill plot

Soient X_1, \dots, X_n n variables aléatoires *i.i.d* de fonction de répartition F appartenant au domaine d'attraction de Fréchet. Cette méthode consiste à faire varier le nombre d'excès k et de calculer à chaque fois l'estimateur de Hill $\xi_{k,n}^{Hill}$ associé. On trace ensuite le graphique des couples $(k, \xi_{k,n}^{Hill})$. Le nombre d'excès approprié est celui à partir duquel le graphique devient horizontal.

3.3.4 AMSE (Asymptotic Mean Square Error)

Cette méthode consiste à retenir comme nombre d'excès approprié celui qui minimise l'erreur quadratique moyenne de l'estimateur de Hill. En guise de rappel, l'erreur quadratique moyenne d'un estimateur T_θ d'un paramètre θ est égale au carré du biais auquel on ajoute la variance ; ceci se traduit par :

$$E[T_\theta - \theta]^2 = [E(T_\theta) - \theta]^2 + Var(T_\theta)$$

On montre que : $Var(\xi_{k,n}^{Hill}) = \frac{(\xi^{Hill})^2}{k}$ et $E(\xi_{k,n}^{Hill}) = \xi + \frac{b_{k,n}}{1+\beta}$; où $b_{k,n}$ et β sont des paramètres.

Il en découle que $AMSE(\xi_{k,n}^{Hill}) = \frac{(\xi^{Hill})^2}{k} + \left(\frac{b_{k,n}}{1+\beta}\right)^2$ et $k_{optimal} = \underset{k}{\operatorname{argmin}} AMSE(\xi_{k,n}^{Hill})$.

3.3.5 Gertensgarbe plot

Cette méthode est issue des travaux de Gerstengarbe et Werner en 1989. Elle permet de déterminer le point de départ de la région extrême en donnant une estimation du seuil approprié.

L'idée est qu'à partir de la zone de sinistres extrêmes on aura sans doute une modification du comportement des écarts Δ_i (différence de coût entre deux sinistres). Plus précisément, soit X_1, \dots, X_n un échantillon de coûts de sinistres, on considère la série des différences $\Delta_i = X_{(i)} - X_{(i-1)}$, $i = 2, \dots, n$ de l'échantillon ordonné, $X_{(i)} \leq \dots \leq X_{(n)}$. L'idée est que le comportement des Δ_i pour les observations extrêmes est certainement différent du comportement des Δ_i pour les observations non extrêmes. Le procédé est le suivant : Pour chaque Δ_k on calcule le nombre n_k d'éléments précédents qui lui sont inférieurs. On calcule ensuite $U_i^* = \sum_{k=2}^i n_k$. La statistique de test est basée sur la version séquentielle du test de Mann-Kendall est donnée par :

$$U_i = \frac{U_i^* - E(U_i^*)}{\sqrt{Var(U_i^*)}} = \frac{U_i^* - \frac{i(i-1)}{4}}{\sqrt{\frac{i(i-1)(2i+5)}{72}}}, \quad i = 1, \dots, n-1$$

Le procédé ci-dessus est décrit sur la série décroissante des différences $\Delta_k, \dots, \Delta_2$. Le nombre d'excès approprié est le point d'intersection de ces deux séries.

La méthode Gerstengarbe peut être effectuée de manière récursive. Dans un premier temps, on l'applique à l'ensemble des données ; ceci permettra de déterminer le point départ du seuil atypique. Dans un second temps, on garde uniquement les sinistres supérieurs au seuil déterminé à l'étape 1 et on réitère l'opération.

Les éléments de la théorie des valeurs extrêmes étant bien explicités, il ne reste plus qu'à les implémenter.

3.4 Application

Il s'agira dans cette section d'appliquer les différentes méthodes présentées précédemment pour déterminer le seuil qui sépare les sinistres attritionnels des sinistres graves et ceci pour chaque segment de risque.

Travaillant sur 16 segments de risque, seul le processus de détermination de seuil sur le segment de risque « Automobile Corporelle particulier » sera explicité.

3.4.1 Statistiques descriptives

Afin de se familiariser davantage avec les données, il est nécessaire d'effectuer quelques statistiques descriptives sur les données de sinistres retraitées de l'inflation et projetées à l'ultime.

Min	q(25%)	Médiane	Moyenne	q(75%)	Max	std	CoV
0	1 336	2 851	11 567	5 940	27 282 687	221 377,1	19,13

TABLE 3.1 – Statistiques descriptives

Il ressort de ce tableau que les coûts de sinistres varient de 0 à 27 M€ et valent en moyenne 11 567€. Cependant, le coefficient de variation vaut 19,13 ; ce qui signifie qu'il y a une très grande dispersion des données autour de la moyenne. La médiane quant à elle, vaut 2 851€ ce qui signifie que 50% des sinistres ont un coût inférieur à 2 851€.

quantiles	q(85,85%)	q(97,5%)	q(98%)	q(98,5%)	q(99%)	q(99,5%)	q(99,7%)
Valeurs	11 567	79 551	99 267	132 040	196 778	369 424	571 688
Excès	9 785	1729	1383	1037	692	346	208

TABLE 3.2 – Quantiles

Le quantile d'ordre 85,85% correspond à la moyenne des sinistres et est supérieur au quadruple de la médiane. De plus, sur 100 sinistres, 98 ont un coût inférieur à 99 267€ et 99 ont un coût inférieur à 196 778€. On soupçonne une distribution à queue épaisse.

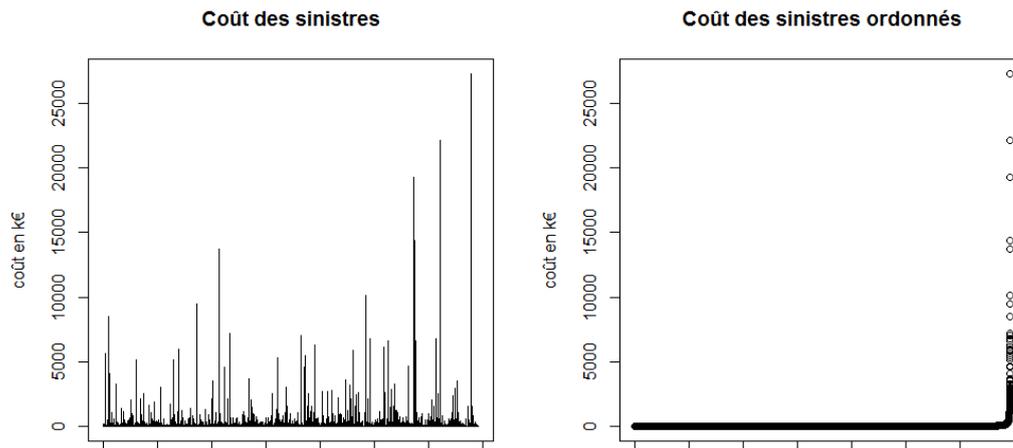


FIGURE 3.1 – Coûts individuels des sinistres en k€

Les deux graphiques ci-dessus donnent un premier aperçu de la série des coûts de sinistres. La structure de coude observée sur la série des coûts de sinistres ordonnés laisse présager que la queue de la distribution des données des coûts de sinistres est lourde. Il est fort probable que les données appartiennent au domaine d'attraction de Fréchet.

3.4.2 Détermination du domaine d'attraction

a) QQ plot exponentiel

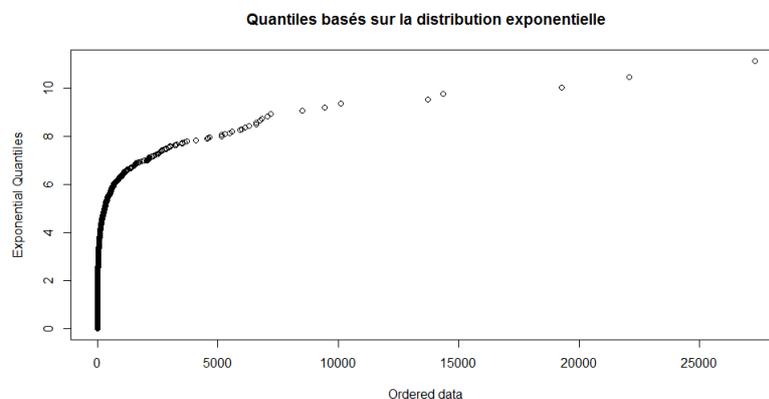


FIGURE 3.2 – QQ plot exponentiel (Coût en k€)

La structure concave du QQ-plot exponentiel des données s'interprète par le fait que la distribution des coûts de sinistres a une queue plus épaisse que celle de la loi exponentielle : il est très probable que la loi de distribution des données soit à queue épaisse du type Pareto, Fréchet, etc.

b) Test sur ξ

La courbe de gauche est strictement convexe, ce qui conduit au rejet de l'hypothèse selon laquelle $\xi = 0$.

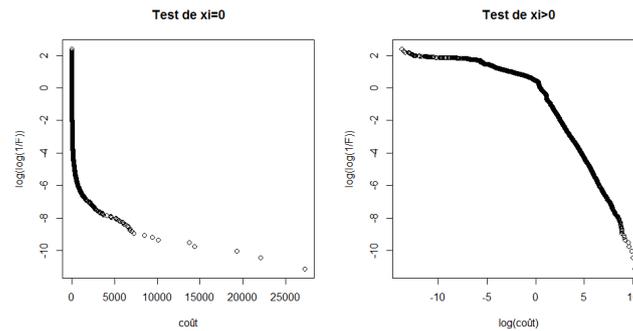


FIGURE 3.3 – Test sur ξ

Les résultats obtenus sur les données confirment le soupçon fait quant à l'épaisseur de la queue de distribution. En effet, on observe sur la figure ci-dessus une quasi-linéarité de $\ln(\ln(\frac{1}{F}))$ en $\ln(\text{coût})$; ce qui justifie le comportement à queue lourde de la distribution des données : la loi de distribution des données appartient au domaine d'attraction de Fréchet. Dans la suite, sera utilisé un estimateur qui n'est valide que dans un domaine d'attraction de Fréchet.

3.4.3 Détermination du seuil

Il s'agira dans cette sous-section de déterminer le seuil d'entrée dans la région extrême.

a) Mean excess plot

Le Mean excess plot est globalement croissant ; sa nette rapide croissance observée pour les premiers seuils permet de confirmer la conclusion d'une queue lourde.

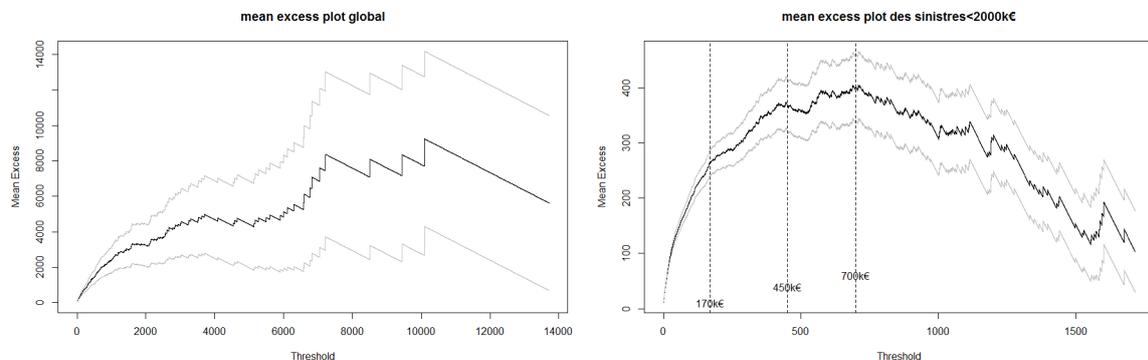


FIGURE 3.4 – Mean excess plot

L'espérance résiduelle est fortement influencée par les grandes valeurs. Si on exclut les sinistres dont le coût est supérieur à 2000 k€ (0,0091% de la base), on obtient le graphique de droite.

Plusieurs seuils possibles de linéarité du Mean excess plot à savoir : 170 k€, 450k€ et 700k€. Si un de ces seuils est retenu, alors tout seuil au dessus est également valable (propriété de seuil stabilité).

b) Hill Plot

Ce graphique donne pour chaque nombre d'excès k l'estimateur de Hill correspondant. Plus le nombre d'excès considéré est grand, plus l'estimateur de Hill se stabilise. L'estimateur de Hill semble stable à partir de 700 excès qui correspond à un seuil de 204 k€. Ainsi d'après cette méthode tout seuil inférieur ou égal à 204 k€ est adéquat.

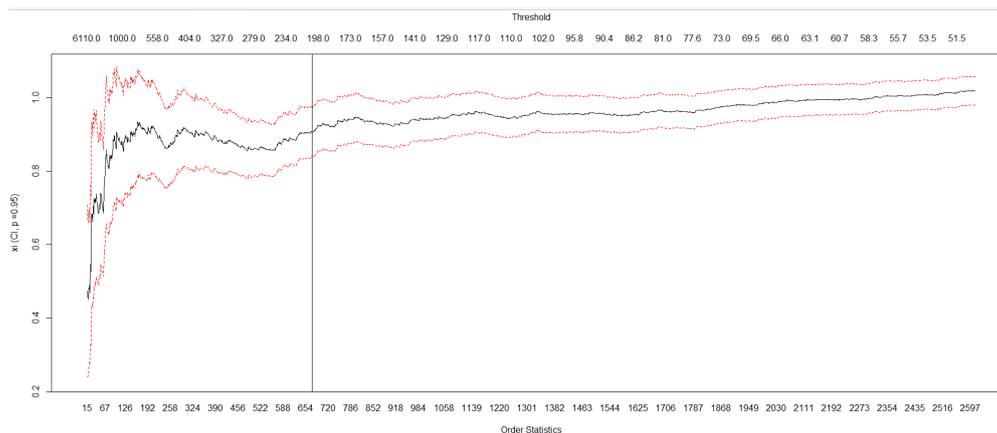


FIGURE 3.5 – Hill plot

En combinant ce résultat avec celui du Mean excess plot, nous pouvons dire que le seuil approprié serait compris entre 170k€ et 204k€.

c) AMSE

Le seuil qui minimise l'erreur quadratique de l'estimateur de Hill est 251 k€. Ce seuil correspond à un nombre d'excès égal à 565 excès.

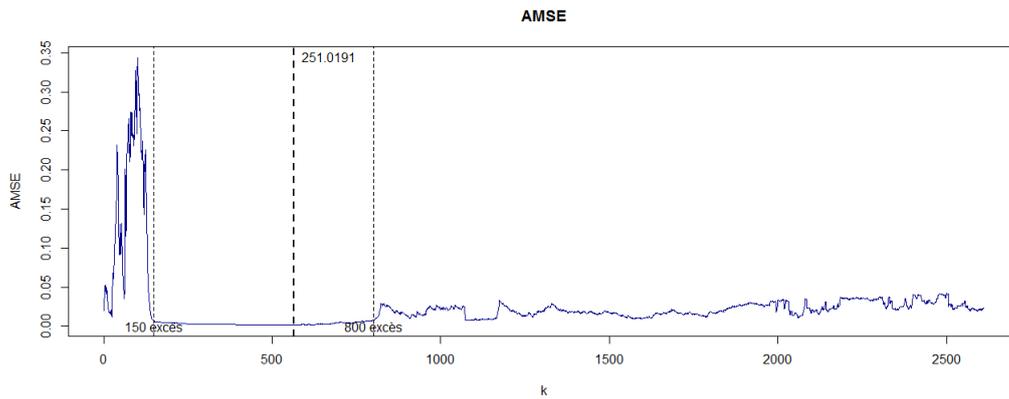


FIGURE 3.6 – AMSE

Cependant pour un nombre d'excès compris entre et 150 (seuil=780 k€) et 800 (seuil=165 k€) excès, l'AMSE est très proche de l'AMSE minimale. Ainsi, d'après ce critère, tout seuil compris entre 165 k€ et 780 k€ serait adéquat. Ce résultat est cohérent avec les précédents. En combinant ce résultat avec les précédents, il s'en suit que le seuil approprié serait compris entre 170k€ et 204k€.

d) Gerstengarbe plot

La méthode du Gerstengarbe plot a été appliquée de manière récursive. Dans un premier temps, le point de départ du seuil atypique a été déterminé sur l'ensemble de nos données.

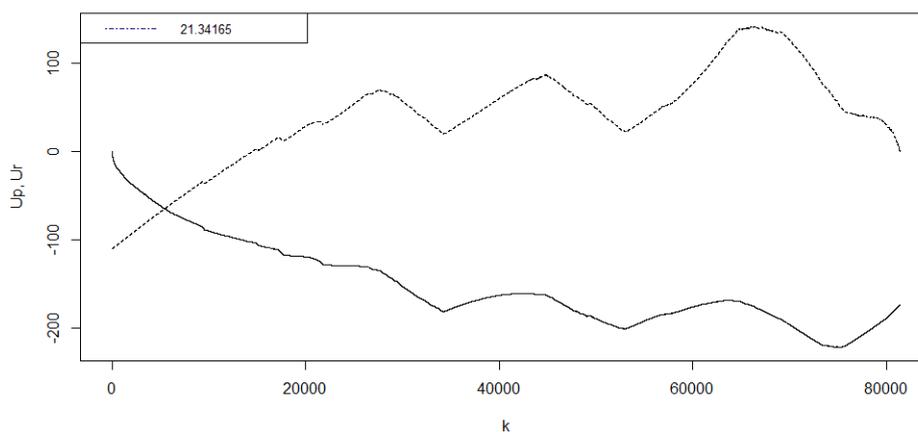


FIGURE 3.7 – Gerstengarbe plot

Uniquement sur les sinistres dont le coût est supérieur au seuil déterminé à l'étape 1 et l'opération a été réitéré et ainsi de suite. Les résultats obtenus sont les suivants ;

Étapes	Seuils en k€
1	21,34
2	54,53
3	195,73
4	708
5	2569,49

TABLE 3.3 – Seuils par la méthode de Gerstengarbe pour 5 itérations

Cette méthode fournit différents seuils; cependant, seul le seuil obtenu à la troisième itération (196 k€) est cohérent avec les résultats des autres méthodes car il est compris entre 170k€ et 204k€.

Seuil retenu

Les différentes méthodes étudiées permettent d'affirmer que le seuil $u = 200$ k€ serait adéquat. Toutefois, ce seuil doit être validé.

e) Validation du seuil retenu

Les paramètres de la GPD ont été estimés avec le seuil $u = 200$ k€. Les résultats qui en découlent sont les suivants :

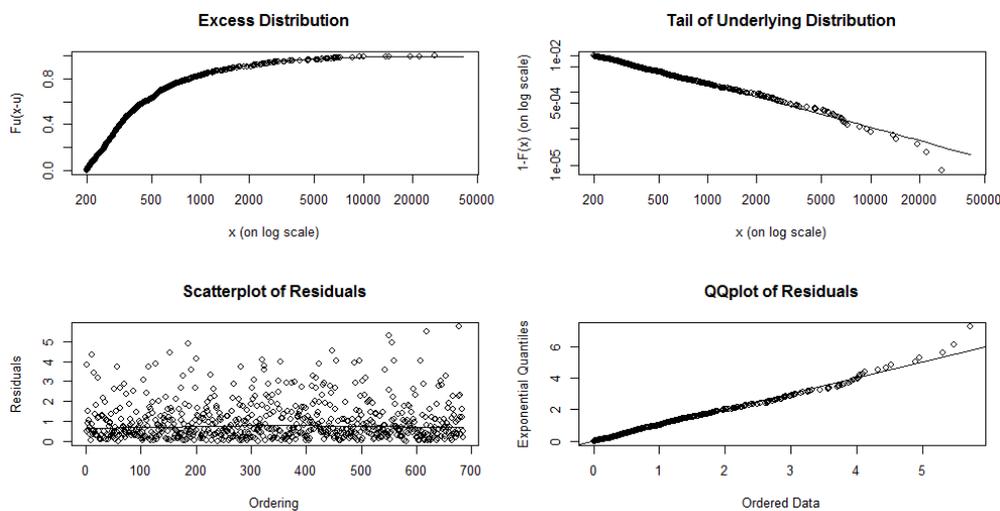


FIGURE 3.8 – Validation du seuil de sinistre grave

La loi GPD semble bien adaptée aux données; quant aux résidus, ils semblent purement aléatoires et aucune tendance particulière ne se décelé dans la série des résidus. Ceci confirme le fait que le seuil $u = 200$ k€ permet de bien distinguer les sinistres graves des sinistres attritionnels.

Le procédé a été implémenté sur tous les segments de risque pour déterminer le seuil qui sépare la sinistralité grave de la sinistralité attritionnelle. Les résultats suivants ont été obtenus :

MCC	Libellé MCC	Seuil retenu
MCC02	Accident	100 k€
MCC03	Auto-Corp-Part	200 k€
MCC04	Auto-Corp-Pro	200 k€
MCC05	Auto-Mat-Part	100 k€
MCC06	Auto-Mat-Pro	100 k€
MCC07	Dommmages-Part	200 k€
MCC08	Divers	
MCC09	Dommmages-Pro	200 k€
MCC10	Pr-Juridique	
MCC11	RC-Part	100 k€
MCC12	RC-Pro	150 k€
MCC13	Grêle	150 k€
MCC14	Cons-DO	
MCC15	Cons-RCD	
MCC16	Cons-Autre	

TABLE 3.4 – Seuils retenus pour chaque segment de risque

Modélisation de la sinistralité

Pour chaque segment de risque, il a été déterminé au chapitre précédent le seuil à partir duquel un sinistre est qualifié de sinistre grave. Il est dès lors possible de distinguer les sinistres attritionnels des sinistres graves. Les sinistres graves étant ceux dont le coût dépasse un seuil u_0 fonction du segment de risque.

Dans cette partie, il sera question de déterminer la loi qui modélise au mieux la volatilité de la sinistralité, d'en estimer les paramètres et d'en déduire le SCR. Les calculs seront faits à la granularité MCCxloss-type.

4.1 Tests d'adéquation à une loi

Les tests d'adéquation à une loi sont utilisés pour juger de la compatibilité d'un échantillon avec une distribution donnée. Ainsi, pour effectuer un test d'adéquation à une loi, il est indispensable de disposer d'un échantillon ainsi que d'une loi de référence entièrement définie. Le test d'adéquation sera ensuite effectué pour vérifier si l'échantillon correspond bien à cette loi. Si les paramètres de la loi à tester ne sont pas connus, il conviendrait de les estimer au préalable. Plusieurs méthodes peuvent être utilisées pour estimer les paramètres d'une loi ; les plus courantes sont la méthode des moments et la méthode du maximum de vraisemblance.

4.1.1 Estimation des paramètres de la loi

a) Méthode des moments

La méthode des moments est une méthode d'estimation très ancienne ; elle est utilisée pour faire de l'estimation ponctuelle.

Soit X_1, \dots, X_n une suite de variables aléatoires *i.i.d* de fonction de répartition F qui dépend d'un paramètre $\theta = (\theta_1, \dots, \theta_p)$. Les estimateurs des moments des paramètres $\theta_1, \dots, \theta_p$ sont obtenus en égalant au moins les p premiers moments théoriques et les p premiers moments empiriques. Ceci se traduit par :

$$E(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k$$

C'est une méthode d'estimation simple et intuitive, mais les estimateurs obtenus possèdent en général peu de belles propriétés.

b) Méthode du maximum de vraisemblance

Soit X_1, \dots, X_n une suite de variables aléatoires *i.i.d* de fonction de répartition F qui dépend d'un paramètre $\theta = (\theta_1, \dots, \theta_p)$. Soit x_1, \dots, x_n une réalisation de cette suite. La vraisemblance

de $F(\theta)$ sous l'échantillon x_1, \dots, x_n s'écrit :

$$L(\theta, x_1, \dots, x_n) = \prod_{i=1}^n f(x_i, \theta)$$

où f est la fonction de densité associée à F . On cherche la valeur de θ qui maximise la vraisemblance ou la log-vraisemblance ; pour cela, il faut résoudre le système d'équations obtenu en annulant des dérivés partielles de la log-vraisemblance par rapport aux composantes de θ c'est-à-dire :

$$\frac{\partial}{\partial \theta_j} \ln(L(\theta)) = 0, j = 1, \dots, p$$

L'estimateur du maximum de vraisemblance est défini de façon unique et peut être calculé explicitement ; lorsqu'il n'est pas possible d'avoir une expression explicite, on fait recours à des algorithmes de résolution numérique.

Sous certaines hypothèses, l'estimateur du maximum de vraisemblance est asymptotiquement sans biais et convergeant ; de plus sa variance est minimale : c'est donc un estimateur performant.

Une fois estimés les paramètres de la loi à tester, des tests d'adéquation peuvent être effectués afin de vérifier si cette loi est sous-jacente à l'échantillon. Plusieurs tests d'adéquation ont été développés dans la littérature. La plus part d'entre eux sont basés sur la fonction de répartition empirique.

Au préalable, il est nécessaire de rappeler les définitions des fonctions de répartition théorique et empirique. Soit X_1, \dots, X_n une suite de variables aléatoires *i.i.d* de fonction de répartition F . On rappelle que $\forall x, F(x) = \mathbb{P}(X \leq x)$. La fonction de répartition empirique associée est définie par $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$.

4.1.2 Test de Kolmogorov-Smirnov

Le test d'ajustement de Kolmogorov-Smirnov permet de tester l'hypothèse H_0 selon laquelle les données suivent une loi de probabilité théorique continue. Dans ce test, il est question de mesurer l'écart entre la fonction de répartition théorique et la fonction de répartition empirique. Soit X_1, \dots, X_n une suite de variables aléatoires *i.i.d* de fonction de répartition F continue et inconnue. On se donne une fonction de répartition F_0 , supposée continue. On veut résoudre le problème de test suivant :

$$\begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases}$$

La statistique de test est $D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F_0(x)|$.

Pour un seuil α , on rejette H_0 si $D_n > d_{n,1-\alpha}$; où $d_{n,1-\alpha}$ est le quantile d'ordre $(1 - \alpha)$ de la table de Kolmogorov-Smirnov ; il vérifie $\mathbb{P}_{H_0}(D_n \geq d_{n,1-\alpha}) \leq \alpha$.

4.1.3 Tests de Cramer-von Mises et d'Anderson-Darling

Le problème de test est le même que celui du test de Kolmogorov-Smirnov. Ces tests sont basés sur la même logique que le test de Kolmogorov-Smirnov à savoir mesurer l'écart entre la

fonction de répartition empirique et une fonction de répartition théorique connue. Leur particularité est qu'ils font intervenir l'écart quadratique $(F_n(x) - F_0(x))^2$. Leurs statistiques de test sont de la forme :

$$Q_n = n \int_{-\infty}^{+\infty} (F_n(x) - F_0(x))^2 \Psi(x) dF_0(x)$$

où Ψ est la fonction de pondération.

a) Tests de Cramer-von Mises

Pour ce test, la fonction de pondération est $\Psi(x) = 1$ et la statistique de test est approximée par :

$$W_n^2 = \sum_{i=1}^n \left(\frac{2i-1}{2n} - F_0(X_{(i)}) \right)^2 + \frac{1}{12n}$$

où $X_{(i)}$ est la $i^{\text{ème}}$ statistique d'ordre de l'échantillon. Pour un seuil α donné, l'hypothèse H_0 est rejetée si W_n^2 est supérieur à la valeur critique $w_{n,1-\alpha}$; où $w_{n,1-\alpha}$ est le quantile d'ordre $(1-\alpha)$ de la table de Cramer-von Mises.

b) Tests d'Anderson-Darling

La fonction de pondération est $\Psi(x) = \frac{1}{F_0(x)(1-F_0(x))}$ et la statistique de test est approximée par :

$$A_n^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln(F_0(X_{(i)})) + \ln(1 - F_0(X_{(n+1-i)}))]$$

Pour un seuil α , on rejette H_0 si $A_n^2 > a_{n,1-\alpha}$; où $a_{n,1-\alpha}$ est le quantile d'ordre $(1-\alpha)$ de la table d'Anderson-Darling.

La fonction de pondération utilisée dans le test d'Anderson-Darling donne plus de poids aux observations situées dans les queues de la distribution. Ce test est donc plus intéressant si on s'intéresse aux queues de distribution.

4.1.4 Graphique Quantiles-Quantiles (Q-Q plot)

Le graphique Quantiles-Quantiles (Q-Q plot) permet de tester graphiquement l'adéquation d'une distribution donnée à un modèle théorique. L'intuition du Q-Q plot est que les quantiles empiriques doivent approximativement être égaux aux quantiles de la loi sous-jacente aux données. Autrement-dit, si on représente sur un graphique les quantiles empiriques en abscisse et les quantiles théoriques correspondant en ordonnée, alors les points doivent être alignés sur la première bissectrice si la loi théorique est adéquate aux données.

Les tests présentés jusqu'à présent s'appliquent uniquement aux distributions continues. Dans le cas des distributions discrètes, le test le plus utilisé est le test du chi-deux d'adéquation à une loi. Cependant, ne disposant pas d'un historique assez long dans ce mémoire, le test du chi-deux ne serait pas robuste. Le critère du maximum de vraisemblance et le critère de minimisation des critères d'informations (AIC et BIC) seront utilisés pour faire le choix entre les distributions discrètes.

4.2 Modélisation

Disposant dès à présent d'un bon bagage théorique pour déterminer les lois qui modélisent au mieux la sinistralité, les différentes méthodes d'estimations et les différents tests d'adéquation à une loi présentés à la section précédente seront appliqués aux données d'étude.

En assurance non-vie, les principales quantités d'intérêt sont les coûts et fréquences des sinistres ; ainsi, les lois de probabilités les plus utilisées sont celles à support positif. Les trois lois positives les plus courantes pour la modélisation des coûts des sinistres sont les suivantes :

- **Loi log-normale** : elle a pour densité $f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\log(x)-\mu}{\sigma})^2}$; avec $x, \sigma^2 > 0, \mu \in \mathbb{R}$.
Si $X \sim LN(\mu, \sigma^2)$ alors $\ln(X) \sim N(\mu, \sigma^2)$
- **Loi gamma** : elle a pour densité $f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha-1}$; avec $x \geq 0, \alpha, \lambda > 0$ et Γ la fonction Gamma.
- **Loi weibull** : elle a pour densité $f(x) = \frac{\beta}{\eta^\beta} x^{\beta-1} e^{-\left(\frac{x}{\eta}\right)^\beta}$; avec $x > 0, \eta, \beta > 0$.

A ces lois, à été rajoutée la loi GPD (définie au chapitre précédent) pour la modélisation de la sinistralité grave. Cependant bien que la loi normale soit à support dans \mathbb{R} , elle a néanmoins été testée car elle a de « bonnes propriétés ».

Les lois de probabilités le plus souvent utilisées pour modéliser la fréquence des sinistres sont :

- **Loi binomiale négative** : elle est notée $BN(m, p)$ et a pour fonction de probabilité $\mathbb{P}(X = k) = \binom{m+k-1}{k} p^m (1-p)^k$; où $k \in \mathbb{N}$ et $p \in [0, 1]$
- **Loi de Poisson** : sa fonction de probabilité est $\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$; avec $\lambda > 0$ et $k \in \mathbb{N}$

L'ajustement aux données des lois présentées ci-dessus a été vérifiée en implémentant les tests d'adéquation présentés précédemment. Ce travail a été effectué sur chaque segment risque ; nous détaillerons dans la suite les résultats obtenus pour le segment « Automobile Corporelle pour les Particuliers »(MCC03).

4.2.1 Modélisation de la sinistralité attritionnelle

Estimation des paramètres

Les paramètres des lois ont été estimés par la méthode du maximum de vraisemblance. Les résultats obtenus sont présentés dans le tableau ci-après :

Lois ajustées	Normale	Log-normale	Gamma	Weibull
Paramètre 1	27,3	3,29	33,9	5,36
Paramètre 2	4,92	0,17	1,24	29,42
Log vraisemblance	-54,2	-52,7	-53,2	-55,7
AIC	112,4	109,5	110,3	115,4
BIC	114,2	111,3	112,1	117,1

TABLE 4.1 – Estimation des paramètres des lois théoriques

La vraisemblance que les données de prime pure de la sinistralité attritionnelle du segment « Auto_Corp_Part » soient issues d'une loi log-normale est plus grande ; les critères d'informations AIC et BIC associés sont les plus petits. Ceci signifie que sur la base de ces critères, la loi log-normale est celle qui ajuste le mieux les données de prime pure de la sinistralité attritionnelle de la MCC considérée.

Minimisation de la distance

Les distances entre la distribution empirique et les distributions théoriques estimées ont été calculées. Les métriques calculées sont : la distance de Kolmogorov-Smirnov, la distance de Cramer-Von Mises et la distance d'Anderson-Darling.

Loi ajustée	Normale	Log-normale	Gamma	Weibull
Statistique test de Kolmogorov S.	0,134	0,125	0,128	0,171
Statistique test de Cramer V.M.	0,087	0,055	0,064	0,128
Statistique test d'Anderson D.	0,67	0,45	0,51	0,88

TABLE 4.2 – Tests d'adéquation

Parmi les lois testées, la loi Log-normale est celle qui minimise les distances des tests de Kolmogorov, de Cramer-Von Mises et D'Anderson-Darling. De plus les p-valeurs des tests d'adéquation associés ne permettent pas de rejeter l'hypothèse d'adéquation à la loi Log-normale.

Choix de la loi

Les métriques considérées sont : la log-vraisemblance, la distance de Kolmogorov-Smirnov, la distance de Cramer-Von Mises et la distance d'Anderson-Darling.

Pour chaque métrique, un rang est attribué aux différentes lois théoriques. Par exemple, le rang 1 sera attribué à la loi qui a la plus petite distance au sens de la métrique considérée. Le score total de chaque loi est la somme des carrés de ses rangs obtenus sur chaque métrique. La loi retenue est celle qui a le plus petit score. En d'autres termes, on cherche à résoudre le problème de minimisation suivant :

$$\min_{i \in \{1, \dots, I\}} \sum_{j=1}^J \left(\text{rang}_j^{l_i} \right)^2$$

Où J est le nombre de métriques, I le nombre de lois théoriques testées, l_i la loi i et $\text{rang}_j^{l_i}$ le rang de la loi i au sens de la métrique j .

Loi ajustée	Normale	Log-normale	Gamma	Weibull
Log vraisemblance	3	1	2	4
Statistique test de Kolmogorov S.	4	1	2	3
Statistique test de Cramer V.M.	3	1	2	4
Statistique test d'Anderson D.	3	1	2	4

TABLE 4.3 – Tests d'adéquation : choix de la loi

Parmi les lois testées, la loi Log-normale est celle qui a le plus petit score.

Tests graphiques

Par la suite, nous avons comparé les distributions des différentes lois théoriques estimées à la distribution empirique.

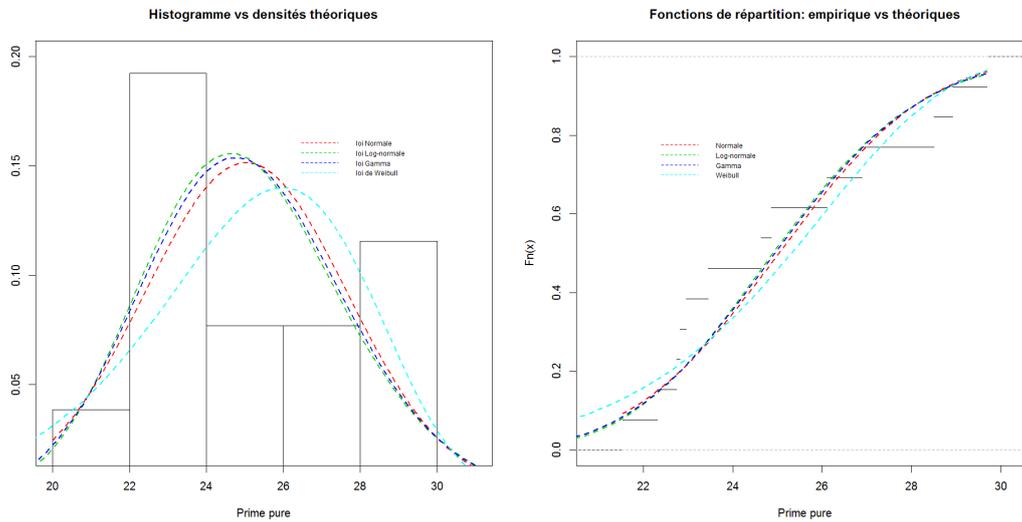


FIGURE 4.1 – Ajustement graphique

Les densités et les fonctions de répartition des lois Log-normale et Gamma sont celles qui semblent le mieux s'ajuster aux données. Toutefois, il est difficile à partir de ce graphique de déterminer laquelle de ces deux lois est la plus adéquate aux données.

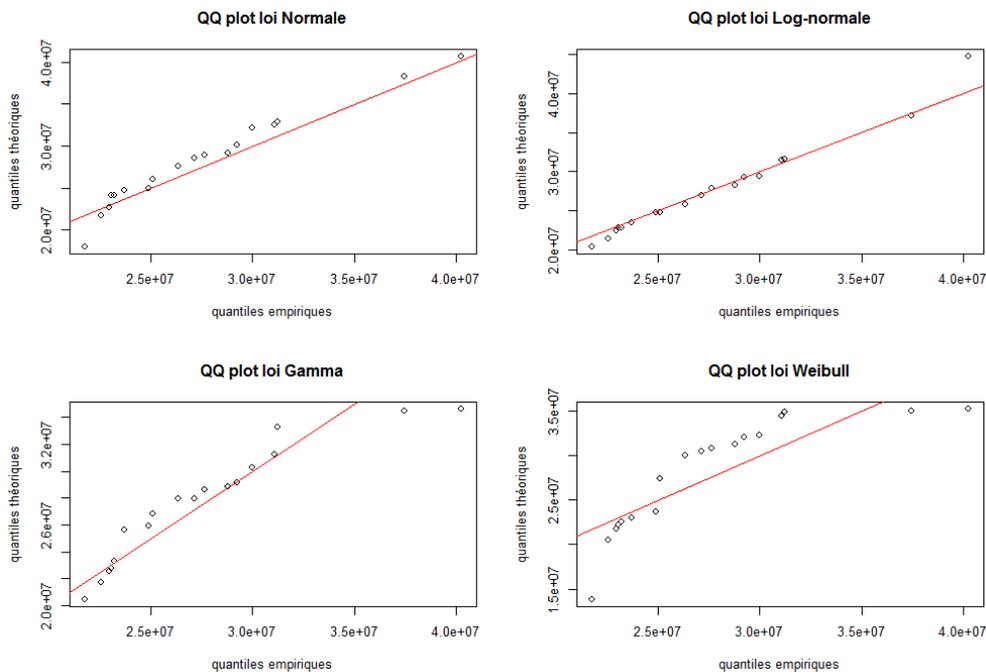


FIGURE 4.2 – Graphiques Quantiles-Quantiles

Le Q-Q plot confirme le fait que la loi Log-normale ajuste le mieux les données : c'est cette loi qui sera retenue pour modéliser la charge attritionnelle du segment de risque « Automobile Corporelle pour les particuliers ».

Par un raisonnement analogue, les distributions théoriques qui ajustent au mieux la sinistralité attritionnelle des différents segments de risque ont été déterminées. Les résultats obtenus sont résumés dans le tableau suivant :

MCC	Libellé MCC	Loi retenue	moyenne	écart-type	CoV
MCC02	Accident	Log-Normale	67,23	21	31,24%
MCC03	Auto-Corp-Part	Log-Normale	27,29	5,1	18,53%
MCC04	Auto-Corp-Pro	Gamma	112,18	33,4	29,8%
MCC05	Auto-Mat-Part	Log-Normale	159,4	16,8	10,54%
MCC06	Auto-Mat-Pro	Weibull	460,6	100,4	21,89%
MCC07	Dommages-Part	Weibull	144,02	21,6	15,72%
MCC08	Divers	Gamma	1,21	0,71	59,1%
MCC09	Dommages-Part	Log-Normale	380,3	56,5	14,9%
MCC10	Pr-juridique	Log-Normale	6,2	1,3	21,4%
MCC11	RC-Part	Gamma	7,4	0,8	11,2%
MCC12	RC-Pro	Gamma	108,8	29,3	26,9%
MCC13	Grêle	Log-Normale	1515,6	725,9	47,9%
MCC14	Cons-DO	Weibull	1332,4	549	41,2%
MCC15	Cons-RCD	Gamma	2577,3	570,3	22,1%
MCC16	Cons-Autre	Gamma	431	262,9	60,9%

TABLE 4.4 – Modélisation de la charge ultime attritionnelle en unité d'exposition

4.2.2 Modélisation de la sinistralité grave

L'approche utilisée pour modéliser la sinistralité grave est une approche « fréquence*sévérité ». La modélisation a été effectuée sur les données de coûts ultimes individuels et de fréquences ultimes des sinistres.

Nous tenons à préciser que les données historiques des nombres de sinistres graves ont été mis à l'ultime par la méthode de Chain-Ladder ; elles ont par la suite été mises en *as-if* 2019 c'est-à-dire corrigées de l'évolution du portefeuille de la société. Soit n_i le nombre ultime de sinistres graves survenus l'année 2000 + i ce nombre mis en *as-if* 2019 est donné par :

$$n_i^{as-if} = n_i * \frac{Exposition\ acquise\ 2000 + i}{Exposition\ acquise\ future\ 2019}$$

Plutôt que de modéliser la fréquence ultime de sinistre grave, on modélisera le nombre ultime de sinistre grave ; en effet, le modèle considéré est un modèle simpliste qui suppose une exposition fixe. La fréquence s'en déduira par la relation : $Frequence = \frac{Nombre}{Exposition}$

Estimation des paramètres

Les paramètres des différentes lois ont été estimés par la méthode du maximum de vraisemblance. Les résultats sont présentés dans le tableau ci-après :

Loi ajustée	Normale	Log-normale	Gamma	Weibull	GPD
Paramètre 1	0,9 M€	13,14	0,94	0,85	0,825
Paramètre 2	2,20 M€	0,88	9,82E-07	8,55E+05	2,1E+05
Log vraisemblance	- 7 531	- 6 783	- 6 939	- 6 922	- 6 618
AIC	15 066	13 570	13 883	13 848	12 298
BIC	15 075	13 579	13 891	13 856	12 308

TABLE 4.5 – Estimation des paramètres de la loi de la sévérité grave

La vraisemblance que les données de la sévérité grave pour la MCC03 soient issues d'une loi GPD est plus grande ; les critères d'informations AIC et BIC associés sont les plus petits. Ceci signifie que sur la base de ces critères, la loi de Pareto Généralisée semble être celle qui ajuste au mieux les données de la sévérité grave.

Les paramètres des lois théoriques testées sur le nombre de sinistres graves ont également été estimés par la méthode du maximum de vraisemblance. Les résultats sont présentés dans le tableau ci-après :

Loi ajustée	Loi de Poisson	Loi Binomiale Négative
Paramètres	59,167	(42,06 ; 59,165)
Log vraisemblance	-74,83	-69,90
AIC	151,65	143,81
BIC	152,55	145,59

TABLE 4.6 – Estimation des paramètres de la loi du nombre de sinistres graves

La vraisemblance que le nombre de sinistres graves de la MCC03 (Automobile corporelle pour les particuliers) soit issu d'une loi Binomiale Négative est plus grande ; les critères d'informations AIC et BIC associés bien que proches de ceux de la loi de Poisson sont également plus petits.

Minimisation de la distance

Les distances entre la distribution empirique de la sévérité grave et les distributions théoriques estimées ont été calculées. Les métriques calculées sont : la distance de Kolmogorov-Smirnov, la distance de Cramer-Von Mises et la distance d'Anderson-Darling. Les résultats obtenus sont renseignés dans le tableau suivant :

Loi ajustée	Normale	Log-normale	Gamma	Weibull	GPD
Stat test de Kolmogorov S.	0,36	0,15	0,23	0,25	0,50
Stat test de Cramer V.M.	20,62	3,68	8,99	7,48	30,5
Stat test d'Anderson D.	$+\infty$	21,26	46,273	41,1	136,53

TABLE 4.7 – Tests d'adéquation

La loi qui minimise les distances des tests de Kolmogorov-Smirnov, de Cramer-Von Mises et d'Anderson-Darling est la loi Log-normale. De plus, les p-valeurs des tests d'adéquation associés ne permettent pas de rejeter l'hypothèse d'adéquation à la loi Log-normale.

Choix de la loi

On utilise le principe décrit pour le choix de la loi de la sinistralité attritionnelle. Les métriques considérées sont : la log-vraisemblance, la distance de Kolmogorov-Smirnov, la distance de Cramer-Von Mises et la distance d'Anderson-Darling.

Loi ajustée	Normale	Log-normale	Gamma	Weibull	GPD
Log vraisemblance	5	2	4	3	1
Test de Kolmogorov S.	4	1	2	3	5
Test de Cramer V.M.	4	1	3	2	5
Test d'Anderson D.	4	1	3	2	5
Score	73	7	38	26	76

TABLE 4.8 – Tests d'adéquation : choix de la loi

Il ressort que la loi Log-normale est celle qui a le plus petit score. Au sens des 4 métriques utilisées, elle est celle qui ajuste le mieux les coûts des sinistres graves de la MCC03 (Automobile Corporelle pour les Particuliers).

Tests graphiques

Le graphique ci-dessous compare les distributions des lois théoriques estimées à la distribution empirique des données de coûts individuels de sinistres graves.

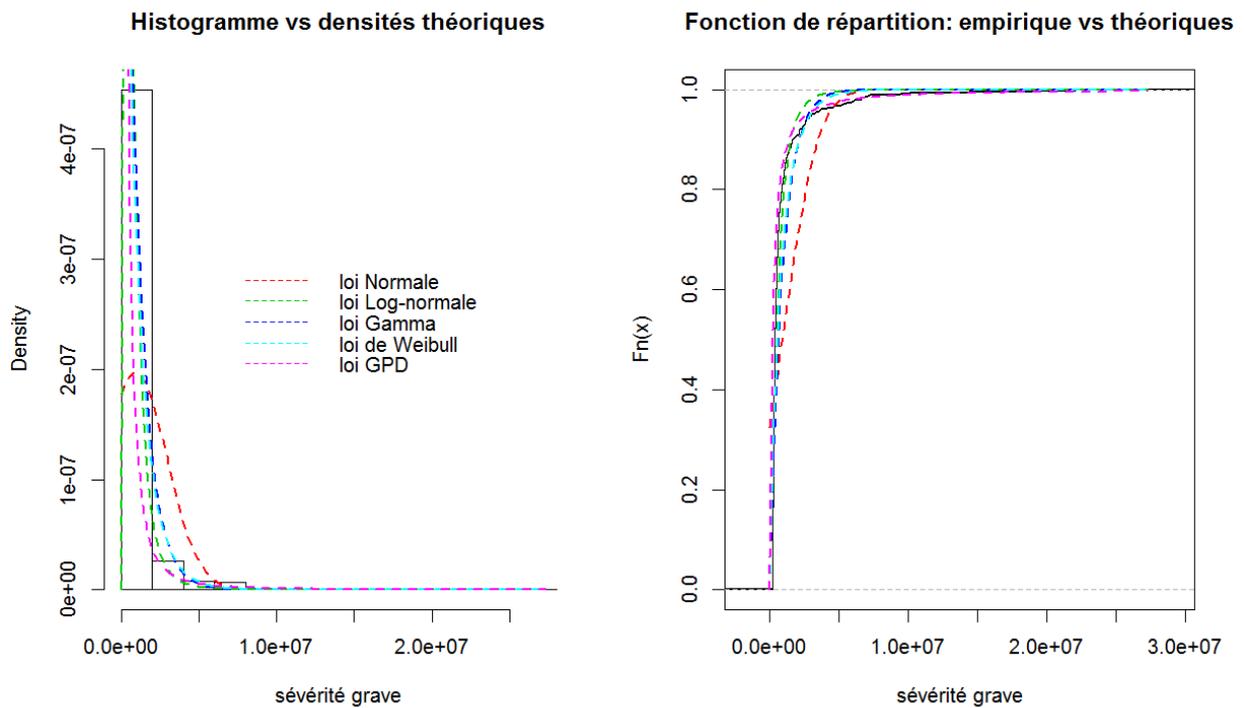


FIGURE 4.3 – Ajustement loi de la sévérité grave

Le graphique ci-dessus ne permet pas de rejeter l'hypothèse selon laquelle les données de la sévérité grave seraient issues d'une loi log-normale.

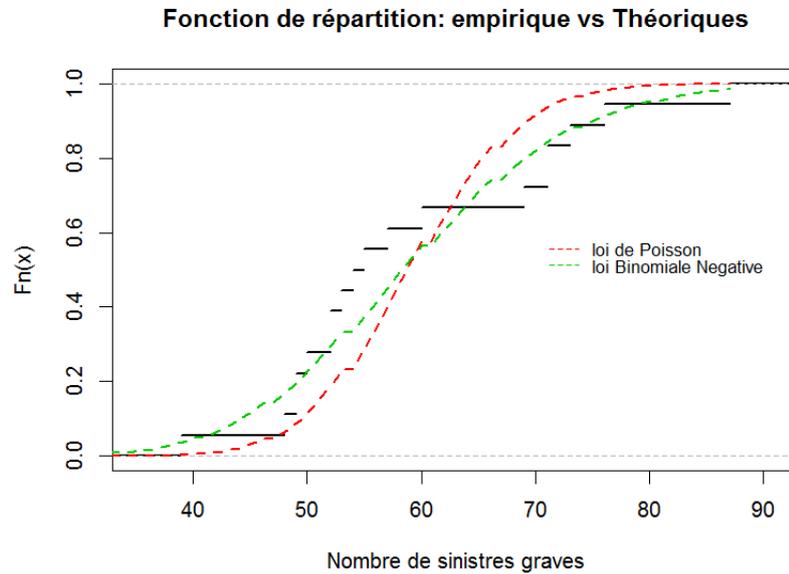


FIGURE 4.4 – Ajustement loi du nombre de sinistres graves

En ce qui concerne le nombre de sinistres graves, la loi Binomiale négative semble plus proche de la fonction de répartition empirique de la loi du nombre de sinistres graves.

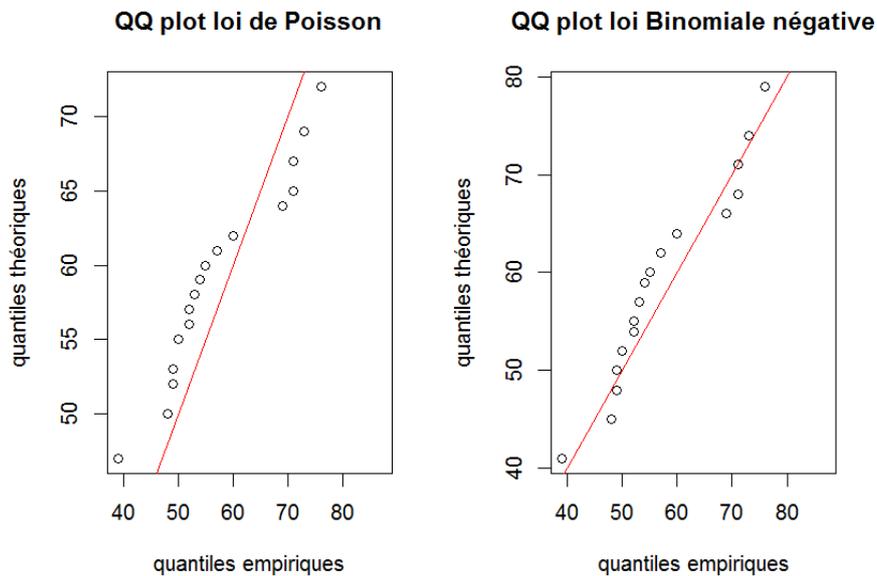


FIGURE 4.5 – Q-Q plot fréquence des sinistres graves

Les Q-Q plot de la distribution du nombre de sinistres graves confirment le fait que la loi Binomiale négative ajuste le mieux le nombre de sinistres graves : c'est cette loi que sera retenue

pour modéliser le nombre de sinistres graves du segment de risque « Automobile Corporelle pour les particuliers ».

Par un raisonnement analogue, nous avons déterminé les distributions théoriques qui ajustent le mieux la sinistralité grave des différents segments de risque. Les résultats obtenus sont résumés dans le tableau suivant :

MCC	Libellé MCC	Loi retenue	Moyenne	Ecart-type	CoV
MCC02	Accident	Gamma	0,255	0,18	70,62%
MCC03	Auto-Corp-Part	Log-normale	0,92	0,93	102%
MCC04	Auto-Corp-Pro	Gamma	0,89	0,89	100,4%
MCC05	Auto-Mat-Part	Gamma	0,33	0,27	81,2%
MCC06	Auto-Mat-Pro	Gamma	0,24	0,19	76,8%
MCC07	Dommages-Part	Gamma	0,40	0,20	49,7%
MCC09	Dommages-Pro	Gamma	0,68	0,58	84,9%
MCC11	RC-Part	Gamma	0,33	0,27	83,7%
MCC12	RC-Pro	Gamma	0,50	0,44	87,8%
MCC13	Grêle	Gamma	0,25	0,11	42,5%

TABLE 4.9 – Modélisation de la charge des sinistres graves (en millions d’euros)

MCC	Libellé MCC	Loi retenue	Moyenne	Ecart-type	CoV
MCC02	Accident	Binomiale négative	5	2,3	43,34%
MCC03	Auto-Corp-Part	Binomiale négative	59	11,93	20,17%
MCC04	Auto-Corp-Pro	Binomiale négative	25	7,85	31,82%
MCC05	Auto-Mat-Part	Poisson	44	6,59	15,14%
MCC06	Auto-Mat-Pro	Binomiale négative	23	7,73	34,00%
MCC07	Dommages-Part	Poisson	66	8,09	12,36%
MCC09	Dommages-Pro	Binomiale négative	69	13,14	19,14%
MCC11	RC-Part	Poisson	7	2,73	36,64%
MCC12	RC-Pro	Binomiale négative	32	7,88	24,7%
MCC13	Grêle	Binomiale négative	8	8,43	101,51%

TABLE 4.10 – Modélisation de la fréquence grave

4.3 Calcul des SCR stand-alone

Les lois qui ajustent le mieux les sinistralités attritionnelle et grave étant maintenant « connues », ainsi que les paramètres associés, il est dès lors possible de calculer les VAR(99,5%) associées et d’en déduire les SCR.

Soit un segment de risque donné; la sinistralité attritionnelle étant modélisée par une approche « prime pure », pour obtenir le SCR attritionnel de ce segment, on simule 10000 observations à partir de la loi de sa « prime pure », la $Var(99,5\%)$ de la « prime pure » étant la 50^{ème}

plus grande valeur parmi les 10000 ; on en déduit le SCR attritionnel par la relation :

$$SCR^{att} = Exposition * (VaR(99,5\% \text{ des } 10000 \text{ simulations}) - \text{moyenne des } 10000 \text{ simulations})$$

Pour ce qui concerne le SCR de la sinistralité grave, on simule 10000 montants de pertes graves, le SCR des pertes graves étant la 50^{ème} plus grande valeur parmi les 10000 et on a la relation :

$$SCR^{grave} = VaR(99,5\% \text{ des } 10000 \text{ simulations}) - \text{moyenne des } 10000 \text{ simulations}]$$

Nous avons répété le processus 100 fois ; nous avons obtenus 100 SCR attritionnels et 100 SCR graves pour cette MCC. Le SCR attritionnel (respectivement grave) final de cette MCC sera la moyenne des 100 SCR.

MCC	Libellé MCC	SCR attritionnel	SCR grave
MCC02	Accident	8,3	2,3
MCC03	Auto-Corp-Part	14,4	31,2
MCC04	Auto-Corp-Pro	7,1	25,9
MCC05	Auto-Mat-Part	46,3	7,9
MCC06	Auto-Mat-Pro	40,4	6,5
MCC07	Dommages-Part	36,6	9,9
MCC08	Divers	0,096	
MCC09	Dommages-Pro	22,3	29,3
MCC10	Pr-Juridique	9,5	
MCC11	RC-Part	2,1	3,7
MCC12	RC-Pro	13,2	14
MCC13	Grêle	40,3	9,1
MCC14	Cons-DO	6,7	
MCC15	Cons-RCD	17,6	
MCC16	Cons-Autre	5,1	

TABLE 4.11 – SCR par type de sinistralité et par Branche (en millions d’euros)

Pour obtenir le SCR de sinistralité future, nous ne pouvons pas tout simplement additionner les SCR obtenus car il faut prendre en compte la diversification du portefeuille. Il faut alors définir une structure de dépendance entre les différents segments de risques.

Théorie quantitative sur l'agrégation des risques

Les études faites au chapitre précédent ont permis de calculer le SCR de sinistralité future pour chaque type de perte de calculer les SCR des différentes MCC. Ces SCR ne peuvent pas être simplement additionnés pour obtenir le capital requis au titre de la sinistralité future car il n'y a pas forcément une dépendance parfaite entre les segments de risques. En effet, le capital requis pour deux segments de risques très faiblement corrélés est relativement faible par rapport à la situation où ces deux segments de risques seraient très corrélés. En effet, plus deux segments de risque sont faiblement corrélés, plus la probabilité qu'ils subissent simultanément un choc est faible. Il faut alors définir la structure de dépendance qui existe entre les différents segments de risques afin de prendre en compte la diversification qui existe entre eux.

La structure de dépendance a un impact important sur les résultats du module de sinistralité future. en 2018, le SCR du module de sinistralité future était 2,33 fois moins élevé en cas de non corrélation entre les risques qu'en cas de corrélation entre les risques ; ceci montre à quel point la structure de corrélation utilisée pour agréger les risques est d'une importance capitale pour AVIVA. L'agrégation étant hiérarchique, nous allons étudier la structure de dépendance dans un premier temps entre les sinistralité attritionnelle et grave de chaque MCC , puis dans un second temps entre les MCC.

5.1 Histoire des corrélations

5.1.1 Une approche hiérarchique

Le SCR de sinistralité future chez AVIVA Assurances est calculé de façon hiérarchique. A chaque niveau de la hiérarchie, est utilisée une matrice de corrélation tel que spécifié par le graphique suivant :

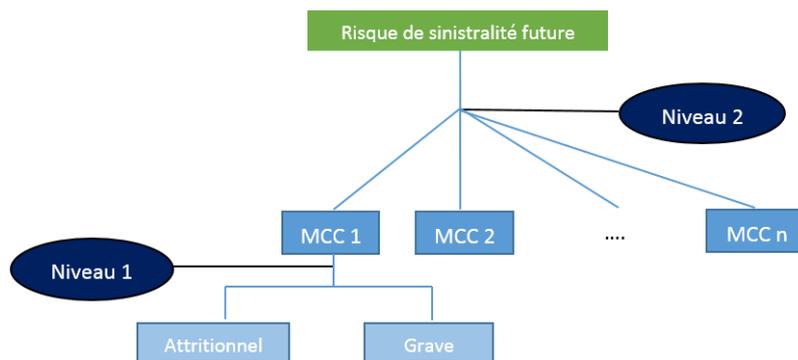


FIGURE 5.1 – Hiérarchie de l'agrégation de la sinistralité future

- Agrégation de niveau 1 : Il s'agit ici d'agréger les sinistralité attritionnelle et grave au sein de chaque segment de risque.
- Agrégation de niveau 2 : Ce niveau concerne la retranscription de la diversification entre les branches (MCC).

On peut se demander s'il serait mieux d'agréger pour chaque type de perte la sinistralité des segments de risque et ensuite d'agréger la sinistralité des types de pertes. Nous avons jugé cette façon de faire pas adéquate car le fait d'être attritionnel ou grave dépend du segment de risque.

5.1.2 Les outils courants de mesure de dépendance

Plusieurs outils ont été développés dans la littérature pour mesurer la dépendance entre variables. Avant de présenter les différentes méthodes de mesure de la dépendance, il est nécessaire de définir au préalable les propriétés que doit vérifier une mesure de dépendance.

Dans l'idéal, une mesure de dépendance δ doit vérifier pour toutes variables aléatoires X et Y les propriétés suivantes :

- **P.1 Symétrie** : $\delta(X, Y) = \delta(Y, X)$
- **P.2 Normalisation** : $-1 \leq \delta(X, Y) \leq 1$
- **P.3** : $\delta(X, Y) = 1 \iff X$ et Y sont comonotones et $\delta(X, Y) = -1 \iff X$ et Y sont antimonotones
- **P.4** : Pour toute fonction croissante f , $\delta(f(X), Y) = \delta(X, f(Y)) = \delta(X, Y)$ et pour toute fonction décroissante g , $\delta(g(X), Y) = \delta(X, g(Y)) = -\delta(X, Y)$
- **P.5** : $\delta(X, Y) = 0 \iff X$ et Y sont indépendantes.

Les outils courants de mesure de dépendance ne vérifient pas pour la plupart les propriétés P.4 et P.5.

a) Le coefficient de corrélation linéaire de Spearman

Soient X et Y deux variables aléatoires réelles. Le coefficient de corrélation linéaire de Spearman est défini par :

$$\rho(X, Y) = \frac{CoV(X, Y)}{\sigma_X \sigma_Y}$$

Où $CoV(X, Y)$ est la covariance entre X et Y et σ_X et σ_Y les écart-types des variables X et Y . Le coefficient de corrélation linéaire de Spearman est la forme de dépendance la plus simple et est facile à calculer. Cependant il n'est pas invariant par transformation strictement monotone (sauf par transformation affine) et ne mesure que la dépendance linéaire; ainsi, le fait qu'il soit nul ou très faible ne donne aucune information sur la structure de dépendance entre les variables considérées. Toutefois, pour des distributions gaussiennes, c'est une réelle mesure de la dépendance.

b) Le Tau de Kendall

Soit (X, Y) un couple de variables aléatoires réelles. Soient (X_1, Y_1) et (X_2, Y_2) deux vecteurs aléatoires indépendants de vecteur mère (X, Y) . Le tau de Kendall est défini par :

$$\tau(X, Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

C'est la différence entre la probabilité que X et Y évoluent dans le même sens et celle qu'ils n'évoluent pas dans le même sens.

Le tau de Kendall est indépendant des marginales et vérifie les propriétés P.1, P.2, P.3 et P.4 mais ne vérifie pas la propriété P.5; en effet si X et Y sont indépendants alors $\tau(X, Y) = 0$ mais la réciproque est fautive.

c) Le Rho de Spearman

Soit (X, Y) un couple de variables aléatoires réelles. Soient (X_1, Y_1) , (X_2, Y_2) et (X_3, Y_3) trois vecteurs aléatoires indépendants de vecteur mère (X, Y) . Le Rho de Spearman est défini par :

$$\rho_s(X, Y) = 3[\mathbb{P}((X_1 - X_2)(Y_1 - Y_3) > 0) - \mathbb{P}((X_1 - X_2)(Y_1 - Y_3) < 0)]$$

Il vérifie les mêmes propriétés que le tau de Kendall et en plus, $\rho_s(X, Y) = \rho(F_X(X), F_Y(Y))$; où ρ est le coefficient de corrélation linéaire de Pearson. Cependant, si X et Y sont indépendants alors $\rho_s(X, Y) = 0$ mais la réciproque est fautive.

Remarque :

On peut constater que l'inconvénient commun à ces coefficients de corrélation est qu'ils ne nous donnent pas une idée sur le caractère symétrique ou non de la dépendance. D'un autre côté, si ces coefficients sont très faibles ou nuls, cela ne veut pas dire que les variables sont indépendantes : ce ne sont pas de « vraies mesures de dépendance ».

5.2 Théorie des copules

Nous avons vu que les mesures de dépendance courantes n'étaient toujours pas adaptées pour décrire la dépendance entre les variables aléatoires. En effet, elles mesurent la dépendance par un scalaire et ne nous donnent aucune information sur celle-ci (dépendance dans les extrêmes par exemple ou des dépendances asymétriques). Un outil pour palier à ces limites : **les copules !**

La notion de « copule » a été introduite par Sklar en 1959. Les copules sont des fonctions de répartition multidimensionnelles des vecteurs aléatoires dont les marginales sont des lois

uniformes sur $[0, 1]$.

Ce concept cadre avec notre problématique car nous nous intéressons au comportement de la queue de distribution. En effet, nous cherchons à évaluer le risque de sinistralité future dans un contexte bicentenaire. Pour cela, nous essayerons de répondre aux questions du type : Si un segment de risque subit une grande perte est-ce que ce sera le cas pour un autre segment de risque donné? Lorsqu'un segment de risque génère une grande perte attritionnelle est-ce qu'il génère aussi une très grande perte grave? Les copules, qui permettent de rendre compte de ces différentes structures de dépendances, deviennent particulièrement intéressantes dans ce contexte.

Rappels et Notations

Dans ce chapitre nous désignons par X_1, \dots, X_n une suite de variables aléatoires *i.i.d* de fonction de répartition F . On rappelle que $\forall x, F(x) = \mathbb{P}(X \leq x)$ et que la variable $U = F(X)$ suit une loi uniforme sur $[0, 1]$.

La fonction de répartition empirique associée est définie par $F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq x}$.

On appelle **inverse généralisé** de F la fonction F^{-1} définie par :

$$F^{-1}(y) = \inf_x \{x / F(x) \geq y\}$$

5.2.1 Définitions et résultats de base

Une copule $C : [0, 1]^n \rightarrow [0, 1]$ en dimension n est une fonction de répartition d'un vecteur aléatoire (Y_1, \dots, Y_n) dont les lois marginales sont des lois uniformes sur $[0, 1]$. Elle possède les propriétés suivantes :

- $\forall u_1, \dots, u_n \in [0, 1], C(u_1, \dots, u_n) = 0$ si au moins un $u_i = 0$
- $\forall u \in [0, 1], C(1, \dots, 1, u, 1, \dots, 1) = u$
- $\forall u^1 = (u_1^1, \dots, u_n^1), u^2 = (u_1^2, \dots, u_n^2) \in [0, 1]^n$ tels que $\forall i \in \{1, \dots, n\} u_i^1 \leq u_i^2$, on a :

$$\sum_{k_1=1}^2 \dots \sum_{k_n=1}^2 (-1)^{k_1+\dots+k_n} C(u_1^{k_1}, \dots, u_n^{k_n}) \geq 0$$

En dimension $n = 2$, nous avons la caractérisation suivante :

- $\forall u \in [0, 1], C(0, u) = C(u, 0) = 0$
- $\forall u \in [0, 1], C(1, u) = C(u, 1) = u$
- Pour tous vecteurs $u = (u_1, u_2)$ et $v = (v_1, v_2)$ tels que $\forall i \in \{1, 2\} u_i \leq v_i$; on a :

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

Théorème de Sklar

Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de loi de distribution F dont les lois marginales F_1, \dots, F_n sont continues; alors il existe une copule unique C telle que :

$$\forall x = (x_1, \dots, x_n) \in \mathbb{R}^n \quad F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n))$$

Ce résultat est important en pratique, car il indique que l'analyse d'une problématique multi-variée peut être décomposée en deux étapes :

- l'identification des distributions marginales
- l'analyse de la structure de dépendance entre les composantes

Remarque : Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de loi de distribution F dont les lois marginales F_1, \dots, F_n sont continues et soit C la copule associée. On a :

- $\forall u_1, \dots, u_n \in [0, 1], C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))$
- Dans le cas de variables aléatoires indépendantes, la copule associée est la copule indépendante C^\perp définie par : $\forall u_1, \dots, u_n \in [0, 1] C^\perp(u_1, \dots, u_n) = \prod_{i=1}^n u_i$

Propriété :

Les copules sont invariantes par transformation strictement croissante des variables aléatoires. Autrement dit, soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de loi de distribution F dont les lois marginales F_1, \dots, F_n sont continues et C la copule associée. Pour tout vecteur $g = (g_1, \dots, g_n)$ de fonctions strictement croissantes, C est également la copule associée au vecteur aléatoire $g(X) = (g_1(X_1), \dots, g_n(X_n))$.

Définition : Soit C une copule.

- La **copule de survie** de C est la fonction \bar{C} définie par :

$$\forall u_1, \dots, u_n \in [0, 1], \bar{C}(u_1, \dots, u_n) = C(1 - u_1, \dots, 1 - u_n)$$

- La **densité** de C est la fonction c définie par :

$$\forall u_1, \dots, u_n \in [0, 1] c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n}$$

5.2.2 Principales copules

Il existe plusieurs familles de copules ; les plus célèbres étant la famille des copules elliptiques (par exemple la copule gaussienne et la copule de Student) et la famille des copules archimédiennes (par exemple les copules de Clayton, Franck, Gumbel).

Etant donné que notre étude porte sur l'étude de corrélation entre segment de risque, nous travaillerons dans la suite en dimension 2.

a) La copule gaussienne

Comme son nom l'indique, la copule gaussienne est sous-jacente à la loi normale multi-variée. Elle est définie par :

$$\begin{aligned} \forall u_1, u_2 \in [0, 1], C(u_1, u_2) &= \Phi_\Sigma(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \\ &= \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left(-\frac{x^2 - 2rxy + y^2}{2\sqrt{1-r^2}}\right) dx dy \end{aligned}$$

où Φ_Σ est la fonction de répartition de la loi normale bivariée centrée réduite de matrice de variance covariance $\Sigma = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$ et Φ^{-1} l'inverse de la distribution normale centrée réduite uni-variée.

b) La copule de Student

La copule de Student est sous-jacente à la loi de Student multi-variée. Soit $r \in]-1, 1[$ le coefficient de corrélation linéaire et ν le nombre de degrés de liberté. La copule de student de paramètres r et ν est définie par :

$$C(u_1, u_2) = \frac{1}{\pi\nu\sqrt{1-r^2}} \int_{-\infty}^{T_\nu^{-1}(u_1)} \int_{-\infty}^{T_\nu^{-1}(u_2)} \frac{\Gamma(\frac{\nu}{2} + 1)}{\Gamma(\frac{\nu}{2})} \left(1 + \frac{x^2 - 2rxy + y^2}{\nu(1-r^2)}\right)^{-(\frac{\nu}{2}+1)} dx dy$$

où T_ν est la fonction de répartition de la loi de Student uni-variée à ν degrés de libertés et Γ est la fonction Gamma.

c) Les Copules archimédiennes

Une copule C sera dite archimédienne stricte s'il existe une fonction $\phi : [0, 1] \rightarrow [0, +\infty[$ continue, strictement décroissante, convexe, vérifiant $\phi(0) = +\infty$, $\phi(1) = 0$ et telle que $C(u_1, \dots, u_n) = \phi^{-1}(\phi(u_1) + \dots + \phi(u_n))$. ϕ est le générateur de C . Les copules de Clayton, Franck et Gumbel sont des copules archimédiennes.

Copule	Générateur ϕ	Expression
Clayton	$\phi(\theta) = \frac{u^{-\theta}-1}{\theta}, \theta \geq 0$	$C(u_1, u_2) = (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$
Franck	$\phi(\theta) = -\ln\left(\frac{e^{-\theta u}-1}{e^{-\theta}-1}\right), \theta \neq 0$	$C(u_1, u_2) = -\frac{1}{\theta} \ln\left(1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right)$
Gumbel	$\phi(\theta) = (-\ln(u))^\theta, \theta \geq 1$	$C(u_1, u_2) = \exp(-([\ln(u_1)]^\theta + [\ln(u_2)]^\theta)^{\frac{1}{\theta}})$

TABLE 5.1 – Copules archimédiennes

d) La copule HRT

C'est la copule de survie de Clayton ; elle n'est pas archimédienne. Elle a été inventée par Gary Venter en 2001 pour modéliser la dépendance sur les évènements de forte intensité. Elle présente une structure de dépendance inversée par rapport à la copule de Clayton, soit une densité de probabilité très élevée aux abords du point (1,1).

En dimension 2, la fonction de répartition de la copule HRT pour $\theta > 0$ est définie par :

$$C(u_1, u_2) = u_1 + u_2 - 1 + (1 - u_1)^{-\theta} + (1 - u_2)^{-\theta} - 1)^{-\frac{1}{\theta}}$$

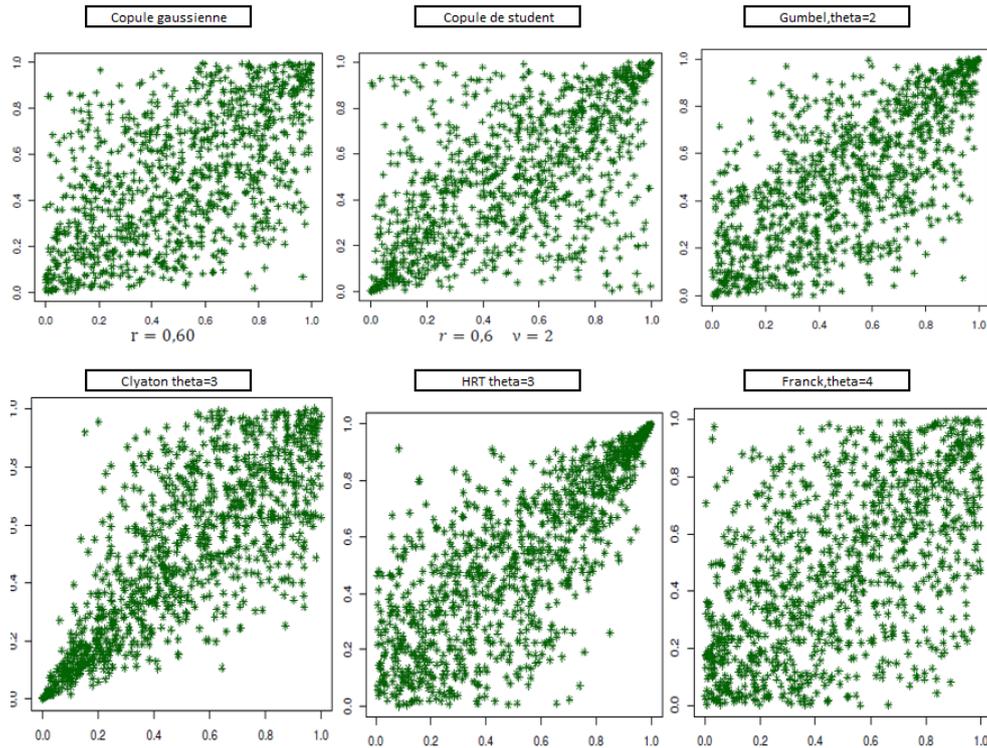


FIGURE 5.2 – Distribution des copules archimédiennes et de la copule de survie de Clayton

Ce graphique montre très clairement que :

- Les copules gaussiennes et de Student permettent de capter les dépendances symétriques.
- La copule gaussienne ne présente pas de dépendance de queue et n'est donc pas adaptée à des valeurs extrêmes.
- La copule de Student permet de capter les dépendances extrêmes positives et négatives.
- La copule de Franck permet de modéliser les dépendances aussi bien positives que négatives.
- La copule de Gumbel n'appréhende que des dépendances positives et a la caractéristique de pouvoir représenter une structure de dépendance plus accentuée sur la queue supérieure.
- La copule de Clayton permet de modéliser uniquement les dépendances positives et a la caractéristique de capter les dépendances à faible intensité dans la queue inférieure.
- La copule de survie de Clayton permet de modéliser la dépendance positive sur des événements extrêmes de forte intensité.

5.2.3 Estimation des paramètres d'une copule

Il existe différentes méthodes, paramétriques ou non paramétriques, qui permettent d'estimer les paramètres d'une copule à partir des données disponibles. Nous introduirons uniquement les méthodes utilisées dans la suite de ce mémoire.

a) Méthode des moments

Cette méthode consiste à évaluer l'expression paramétrique analytique des moments d'une mesure de dépendance considérée avec un estimateur non paramétrique de cette même mesure. Le tau de Kendall et le rho de Spearman d'un couple (X_1, X_2) sont reliés à la copule C de ce couple par les relations :

$$\begin{aligned}\tau(X, Y) &= 4 \int_0^1 \int_0^1 C(u, v) c(u, v) du dv - 1 \\ \rho(X, Y) &= 12 \int_0^1 \int_0^1 C(u, v) du dv - 3\end{aligned}$$

Les estimateurs des paramètres de la copule s'obtiennent en remplaçant τ et ρ_s par leurs valeurs empiriques.

- Dans le cas d'une copule gaussienne C d'un vecteur aléatoire (X, Y) de corrélation linéaire r , on a la relation $\tau = \frac{2}{\pi} \text{Arcsin}(r)$ et $\rho_s = \frac{6}{\pi} \text{Arcsin}(\frac{r}{2})$
- Pour une copule de Student C d'un vecteur aléatoire (X, Y) de corrélation linéaire r , on a la relation $\tau = \frac{2}{\pi} \text{Arcsin}(r)$
- Dans le cas d'une copule de Gumbel de paramètre θ , $\tau = \frac{\theta-1}{\theta}$
- Pour une copule de survie de Clayton ou de Clayton de paramètre θ , $\tau = \frac{\theta}{\theta+2}$
- Dans le cas d'une copule de Franck de paramètre $\theta = 1 - 4 \frac{1-D_1(\theta)}{\theta}$ où $D_k(x)$ est la fonction de Debye définie par $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t-1} dt$

b) Méthode IFM

Soit $X = (X_1, X_2)$ un vecteur aléatoire de copule C . On a l'égalité

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

On peut alors déduire la densité en dérivant cette expression. On obtient :

$$f(x_1, x_2) = c(F_1(x_1), F_2(x_2)) \prod_{i=1}^2 f_i(x_i)$$

Pour un échantillon $(x_1^k, \dots, x_n^k), k = 1, 2$, la log-vraisemblance s'écrit alors :

$$\log(f[(x_1^1, x_1^2), \dots, (x_n^1, x_n^2), \theta]) = \sum_{i=1}^n \log(c(F_1(x_i^1, \theta), F_2(x_i^2, \theta))) + \sum_{i=1}^n \sum_{k=1}^2 \log(f_k(x_i^k, \theta))$$

La méthode IFM consiste à décomposer le paramètre θ sous la forme $\theta = (\theta_1, \theta_2, \theta_C)$; où θ_i est le paramètres de la loi marginale F_i et θ_C est le paramètre de la copule C . Cette méthode se fait en deux étapes :

- Estimation des paramètres θ_1, θ_2 des lois marginales.
- Estimation du paramètre θ_C de la copule. On remplace dans l'expression de la log-vraisemblance les paramètres des lois marginales par leurs valeurs estimées et on détermine la valeur de θ_C qui maximise la vraisemblance.

5.2.4 Choix de la meilleure copule

Plusieurs critères peuvent être utilisés pour choisir la copule qui capte le mieux la structure de dépendance dans un vecteur aléatoire.

a) Le maximum de vraisemblance

Il s'agit tout simplement de retenir la copule qui a la plus grande vraisemblance.

b) Le critère AIC

D'après ce critère, la meilleure copule est celle pour laquelle le critère d'information d'Akaike est le plus petit avec $AIC = -2\ln(L) + 2k$ où L est la vraisemblance maximisée et k le nombre de paramètres de la copule.

c) La fonction de Kendall $K(u)$

Soit (U_1, U_2) un vecteur aléatoire de lois uniformes sur $[0,1]$ dont la copule associée est C . On définit la fonction de Kendall K par :

$$K(u) = \mathbb{P}(C(U_1, U_2) \leq u), u \in [0, 1]$$

Si la copule C est une copule archimédienne de générateur ϕ , alors on l'égalité :

$$K(u) = \mathbb{P}(C(U_1, U_2) \leq u) = u - \frac{\phi(u)}{\phi'(u)}$$

En pratique, la fonction K est utilisée comme méthode d'analyse graphique permettant d'identifier la copule archimédienne théorique qui s'ajuste le mieux à la copule empirique issue de l'échantillon de données.

d) Les « fonctions de concentrations dans les queues » $L(z)$ et $R(z)$

Soit (U, V) un vecteur aléatoire de lois uniformes sur $[0,1]$.

La fonction de concentration dans la queue supérieure est définie par :

$$\begin{aligned} R(z) &= \mathbb{P}(U > z/V > z) \\ &= \frac{\mathbb{P}(U > z, V > z)}{\mathbb{P}(V > z)} \\ &= \frac{\mathbb{P}(U > z, V > z)}{1 - z} \\ &= \frac{1 - \mathbb{P}(U \leq z, V > z) - \mathbb{P}(U > z, V \leq z) - \mathbb{P}(U \leq z, V \leq z)}{1 - z} \\ &= \frac{1 - \mathbb{P}(U \leq z) + \mathbb{P}(U \leq z, V \leq z) - \mathbb{P}(V \leq z) + \mathbb{P}(U \leq z, V \leq z) - \mathbb{P}(U \leq z, V \leq z)}{1 - z} \\ &= \frac{1 + C(z, z) - 2z}{1 - z} \end{aligned}$$

La fonction de concentration dans la queue inférieure est définie par :

$$\begin{aligned} L(z) &= \mathbb{P}(U < z/V < z) \\ &= \frac{\mathbb{P}(U < z, V < z)}{\mathbb{P}(V < z)} \\ &= \frac{C(z, z)}{z} \end{aligned}$$

En pratique, le processus de sélection de la meilleure copule consiste à comparer graphiquement les fonctions $R(z)$ et $L(z)$ de la copule empirique, aux fonctions $R(z)$ et $L(z)$ des copules théoriques à tester.

e) Minimisation de la distance

Ce critère consiste à retenir la copule qui minimise la distance à la copule empirique. Deux distances sont utilisées :

- La distance de Cramer-Von Mises basée sur la distance L^2 . La statistique associée est définie par : $S_n = n \int_{[0,1]^d} (C_n(u) - C_{\hat{\theta}}(u))^2 dC_n(u)$
- La distance de Kolmogorov-Smirnov basée sur la distance L^∞ . La statistique associée est définie par : $S_n = n \sup_{u \in [0,1]^d} |C_n(u) - C_{\hat{\theta}}(u)|$

5.2.5 Copules et simulation

Les copules peuvent être utilisées comme outil à la simulation de distributions conjointes possédant une structure de dépendance donnée.

Soit $X = (X_1, X_2)$ un vecteur aléatoire dont les fonctions de répartition marginales F_1 et F_2 sont connues ainsi que la copule associée C . On souhaite simuler les réalisations (x_1, x_2) de ce vecteur aléatoire. La démarche consiste à simuler dans un premier temps les trajectoires (u_1, u_2) d'un vecteur aléatoire de variables uniformes sur $[0,1]$ dont la fonction de répartition jointe est la copule C . Les réalisations (x_1, x_2) s'en déduisent par la relation $(x_1, x_2) = C(F_1^{-1}(u_1), F_2^{-1}(u_1))$.

Remarque : Pour simuler une trajectoire (u_1, u_2) d'un vecteur aléatoire (U_1, U_2) de variables uniformes sur $[0, 1]$ dont la fonction de répartition jointe est une copule C , on peut procéder comme suit :

- (1) Simuler une valeur u_1 suivant une loi uniforme sur $[0, 1]$
- (2) Simuler une valeur u_2 suivant une loi de fonction de répartition $C_2 = C(U_2 = u_2/U_1 = u_1)$ sur $[0, 1]$; avec $C_2 = \frac{\partial C}{\partial u_1}(u_1, u_2)$

Détermination de la structure de dépendance par une approche quantitative

Nous disposons maintenant d'éléments théoriques pour déterminer la structure de dépendance entre les branches de garanties. Cependant, l'historique des données n'étant pas assez profond, nous devons tout d'abord ré-échantillonner les données tout en conservant la structure de corrélation entre les différentes branches de garanties.

6.1 Ré-échantillonnage des données

Dans cette étude, nous disposons d'un historique de données allant de 2000 à 2018 ; ce qui fait un total de 19 années ; il n'est pas judicieux de paramétrer des copules avec ce nombre d'années. Pour illustrer cela, nous avons simulé 19 réalisations d'une copule gaussienne de paramètre 0.5 en dimension 2 ; nous avons ensuite estimé le paramètre de la copule normale à partir de ces 19 observations. Ce processus a été répété 100 fois. Nous avons obtenu une moyenne de 0,48 sur les 100 estimations ; ceci signifie qu'en moyenne le paramètre estimé est proche de sa vraie valeur. Toutefois, pour certaines réalisations de la copule, le paramètre estimé vaut 0.

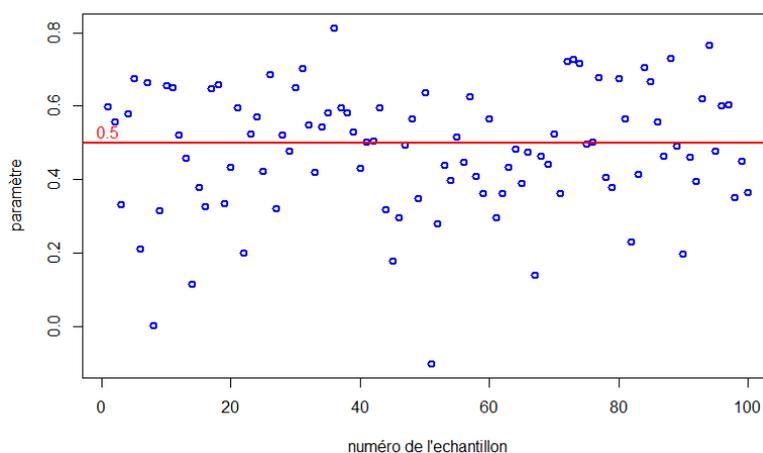


FIGURE 6.1 – Estimation du paramètre d'une copule normale bvariée de paramètre 0,5 à partir de 100 échantillons de 19 simulations issues de cette copule

Afin de palier à ce problème de l'insuffisance de la taille des données, une solution serait d'utiliser des données mensuelles, trimestrielles ou semestrielles. Cependant, nous ne disposons que

des données annuelles. Nous avons opté pour l'utilisation d'un outil très célèbre : le **Bootstrap**.

6.1.1 Théorie du Bootstrap

Le Bootstrap est une technique de ré-échantillonnage imaginée par Efron en 1979; le but premier étant de réduire le biais d'un estimateur et d'estimer sa variance. On peut aussi l'utiliser pour calculer les intervalles de confiance et faire des tests d'hypothèses.

Soit X_1, \dots, X_n une suite de variables aléatoires réelles *i.i.d* de fonction de répartition F qui dépend d'un paramètre θ inconnu. Soit $x = (x_1, \dots, x_n)$ une réalisation de (X_1, \dots, X_n) ; on cherche à calculer l'estimateur $\hat{\theta}(x)$ de θ . On définit la fonction de répartition empirique par

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

Un échantillon Bootstrap est un échantillon de la fonction de répartition empirique F_n équi-pondéré $\frac{1}{n}$; en d'autres termes, c'est un tirage aléatoire avec remise de n éléments dans (x_1, \dots, x_n) .

La méthode du Bootstrap consiste à tirer de manière indépendante B échantillons Bootstrap $(x_1^b, \dots, x_n^b)_{b=1, \dots, B}$, de calculer sur chaque échantillon Bootstrap l'estimation $\hat{\theta}^b(x^b)$ de θ . On en déduit l'estimation de θ par :

$$\hat{\theta}^B(x) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^b(x^b)$$

La variance correspondante est donnée par :

$$V(\hat{\theta}^B(x)) = \frac{1}{B-1} \sum_{b=1}^B [\hat{\theta}^B(x) - \hat{\theta}^b(x^b)]^2$$

6.1.2 Mise en oeuvre du Bootstrap

Afin de recalculer autant de fois que nous le souhaitons les charges ultimes, pour chaque branche de garanties, nous allons utiliser le Bootstrap. Cependant nous ne pouvons pas directement appliquer le Bootstrap sur les triangles de charges attritionnelles et de charges graves car la méthode du Bootstrap s'applique sur des données indépendantes et identiquement distribuées, ce qui n'est pas le cas des données de triangles de développement.

Nous avons opté pour une combinaison Bootstrap et Chain-Ladder. Nous expliquerons chaque étape de ce processus de Bootstrap associé à la méthode chain-Ladder sur les triangles de charges. Le procédé est le suivant :

a) Cas des attritionnels

On utilise la méthode de Chain-Ladder. Soit une branche de garanties donnée; on constitue le triangle de charges attritionnelles cumulées. Désignons par $C_{i,j}^{att}$ (respectivement $P_{i,j}^{att}$) la charge attritionnelle cumulée (respectivement incrémentale) évaluée l'année $2000 + i + j$ (développement j) au titre des sinistres survenus l'année $2000 + i$; avec $i, j = 0, \dots, n - 1$ où $n + 1$ correspond au nombre d'années de survénance.

Dans cette étude, on a 19 années de survénance; par conséquent $n = 18$. Les facteurs de développement de Chain-Ladder sont donnés par :

$$\hat{f}_j^{att} = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}^{att}}{\sum_{i=0}^{n-j-1} C_{i,j}^{att}}, j = 0, \dots, n-1$$

On fait ensuite une reconstitution rétrospective du triangle de charges attritionnelles cumulées à partir des valeurs de la diagonale $C_{i,n-i}^{att}$, $i = 0, \dots, n$ du triangle; en effet, on utilise les coefficients de développement de Chain-Ladder pour estimer les charges attritionnelles cumulées antérieures :

$$\forall j < n - i + 1, \hat{C}_{i,j}^{att} = \frac{C_{i,n-i}^{att}}{\prod_{k=n-i}^{j-1} \hat{f}_k^{att}} \quad i = 0, \dots, n$$

On en déduit le triangle de charges attritionnelles incrémentales antérieures $\hat{P}_{i,j}^{att}$, puis les résidus par la formule :

$$r_{i,j}^{att} = \frac{P_{i,j}^{att} - \hat{P}_{i,j}^{att}}{\sqrt{\hat{P}_{i,j}^{att}}} \quad i = 0, \dots, n \text{ et } j \leq n - i + 1$$

Afin de repérer d'éventuelles tendances qui iraient à l'encontre des hypothèses de Chain-Ladder, nous avons tracé les résidus de chaque segment par année de développement.

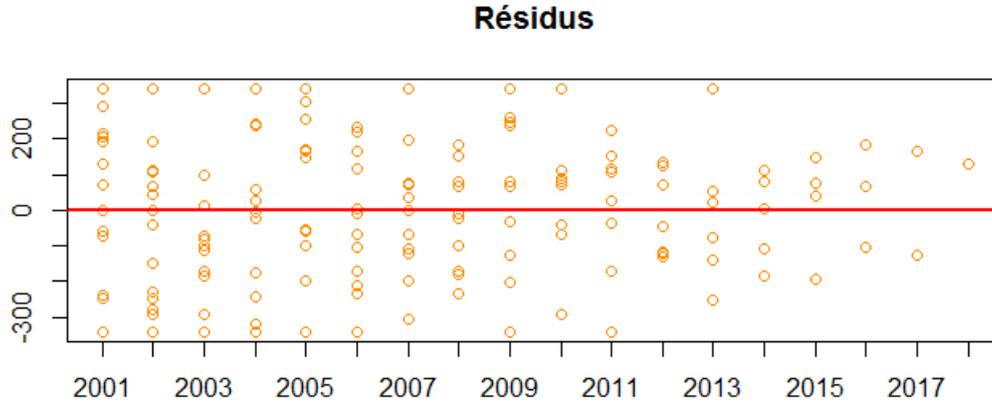


FIGURE 6.2 – Résidus du segment "Accident"

Une fois les résidus calculés, les étapes du ré-échantillonnage sont les suivantes :

- (1) : Tirage aléatoire avec remise des résidus pour chaque case (i, j) (survéance i développement j); on obtient alors un échantillon Bootstrap $(r_{i,j}^{att} \quad i = 0, \dots, n \quad j \leq n - i + 1)$ de résidus.
- (2) : Reconstruction du triangle supérieur de charges attritionnelles incrémentales

$$P'_{i,j}^{att} = \hat{P}_{i,j}^{att} + r_{i,j}^{att} * \sqrt{\hat{P}_{i,j}^{att}}$$

- (3) : Reconstruction du triangle supérieur de charges attritionnelles cumulées $(C'_{i,j}^{att})$
- (4) : Calcul de nouveaux facteurs de développement $(f'_{i,j}^{att})$
- (5) : Construction du triangle inférieur de charges attritionnelles cumulées par la méthode de Chain-Ladder
- (6) : Déduction des charges attritionnelles ultimes pour chaque année de survénance
- (7) : Répéter les étapes (1) à (6) $B - 1$ fois.

b) Cas des graves

Nous procédons de la même façon sur les triangles de charges graves.

c) Prise en compte de la structure de dépendance

Si on applique ce procédé de ré-échantillonnage sur une branche 1, ensuite sur une branche 2, il y a de fortes chances que l'on perde la structure de dépendance qui existe entre ces deux branches.

Afin de ne pas perdre la structure de corrélation entre les branches de garanties, nous avons procédé comme suit : si le résidu attritionnel d'indices (i, j) est tiré sur la première branche alors :

- Ce sera le résidu grave d'indices (i, j) qui sera retenu sur la première branche ;
- Ce sera le résidu attritionnel et le résidu grave d'indices (i, j) qui sera retenu sur toutes les branches

En calculant les charges attritionnelles ultimes et les charges graves ultimes avec exactement la même structure de ré-échantillonnage appliqués aux résidus de Chain-Ladder des différentes branches de garanties, nous pouvons raisonnablement penser que les corrélations, si elles existent, seront conservées.

6.2 Structure de dépendance

Nous avons maintenant un historique « suffisant » de données pour estimer les paramètres des copules.

6.2.1 Structure de dépendance entre la sinistralité attritionnelle et la sinistralité grave par branche

L'étude du seuil qui sépare la sinistralité attritionnelle de la sinistralité grave, a permis de voir que pour certaines branches de garanties, il n'est pas nécessaire de faire la distinction entre sinistralité attritionnelle et sinistralité grave. Pour de telles branches, le SCR de la branche est égal au SCR attritionnel. Pour ce qui concerne, les branches de garanties pour lesquelles nous avons une sinistralité attritionnelle et une sinistralité grave, nous allons agréger les SCR attritionnels et graves.

a) Visualisation graphique de la structure de dépendance

Un graphique est en général plus parlant qu'un chiffre. Avant d'estimer les paramètres des copules, nous allons tout d'abord visualiser la structure de dépendance ; pour cela, l'outil le plus utilisé est la **copule empirique**.

Soit $X = (X_1, X_2)$ un vecteur aléatoire dont les fonctions de répartitions marginales sont F_1 et F_2 . Soit $(x_1^1, x_1^2), \dots, (x_n^1, x_n^2)$ n observations du vecteur aléatoire X . On définit la **copule empirique** par :

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{x_i^1 \leq u, x_i^2 \leq v\}}$$

Cette copule correspond tout simplement au couple des rangs normalisés par le nombre d'observations. La copule empirique nous permet d'avoir une idée sur la structure de dépendance entre les variables étudiées mais ne permet pas d'estimer les paramètres de la copule.

Nous avons construit le graphe de la copule empirique entre la sinistralité attritionnelle et la sinistralité grave de la MCC « Accident ».

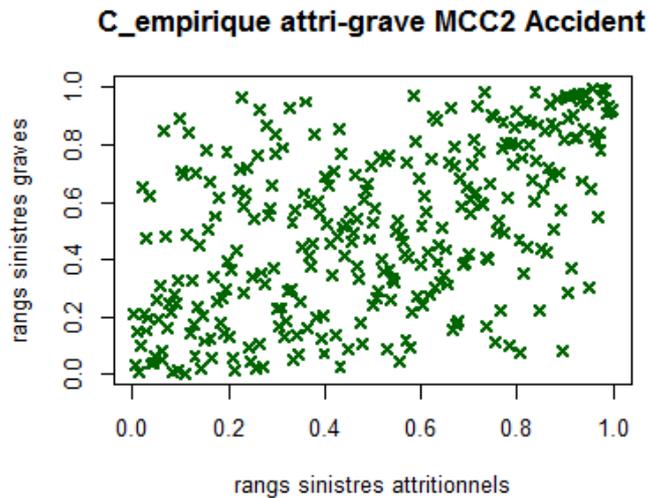


FIGURE 6.3 – Copule empirique sinistralité attritionnelle-sinistralité grave pour le segment Accident

Le graphe ci-dessus laisse présager une dépendance forte dans la queue supérieure.

b) Estimation

Les paramètres des lois marginales des sinistralités attritionnelles et graves estimés au chapitre 4 ont été introduits dans la log-vraisemblance de la densité de la copule. Nous avons ensuite déterminé la valeur du paramètre de la copule qui maximise la log-vraisemblance.

Pour la branche de garanties « Accident » par exemple, le tableau suivant donne les paramètres estimés des différentes copules théoriques testées entre la sinistralité attritionnelle et la sinistralité grave de la MCC considérée.

Copule	Paramètre (s)	Log vraisemblance	Intervalle confiance à 95%
Gaussienne	0,55	52,8	[0,47 ; 0,63]
Student	0,55 ;df=18	53,14	[0,46 ; NA]
Gumbel	1,55	53,4	[1,4 ; 1,7]
Franck	3,7	46,7	[2,9 ; 4,53]
Clayton	0,7	33,8	[0,52 ; 0,91]
HRT	0,91	48,7	[0,69 ; 1,13]

TABLE 6.1 – Copule sinistralité attritionnelle-sinistralité grave pour le segment Accident

Les intervalles de confiance à 95% des paramètres estimés ne contiennent pas 0 ; ce qui signifie que les paramètres sont significatifs au seuil de 5%.

c) Choix de la copule

L'estimation des paramètres des différentes copules ainsi des vraisemblances associées n'est pas suffisant pour déterminer la copule qui ajuste le mieux la structure de dépendance entre les sinistralités attritionnelle et grave de la MCC Accident.

Nous allons utiliser plusieurs critères :

- Les critères de maximisation de la vraisemblance et de minimisation du critère AIC.
- Minimisation de la distance de Cramer-Von-Mises.
- Minimisation de la distance de Kolmogorov.
- Comparaison des fonctions empiriques et théoriques de concentration dans les queues supérieures et inférieures.

Nous faisons le choix dans cette étude d'accorder un poids plus élevé à la fonction de concentration R ; en effet, ne perdons pas de vue que nous recherchons la $VAR(99, 5\%)$ et par conséquent, il est plus pertinent de se concentrer sur la copule pour laquelle il y a une bonne adéquation dans les queues supérieures.

Pour le segment « Accident », le tableau 6.1 indique que la copule de Gumbel a la plus grande vraisemblance.

Des tests d'adéquation basés sur les distances de Cramer-Von-Mises et de Kolmogorov ont également été effectués afin d'évaluer l'adéquation des différentes copules théoriques testées à la copule empirique. Les résultats obtenus sont renseignés dans le tableau suivant :

Copules	Cramer Von-Mises		Kolmogorov	
	Distance	P-value	Distance	P-value
Gaussienne	0.03	0,04	0,99	0,057
Student	0,04	0,01	0,94	0,07
Gumbel	0,02	0,24	0,68	0,53
Frank	0,05	0,01	1,08	0,05
Clayton	0,12	0,005	1,41	0,007
HRT	0,04	0,01		

TABLE 6.2 – Choix de la copule : test d'adéquation

Le test de Cramer Von-Mise d'adéquation rejette l'hypothèse que la structure de dépendance soit une copule gaussienne ou de Student ou de Frank ou de Clayton, ou HRT. Seule l'hypothèse selon laquelle la structure de dépendance serait une copule de Gumbel n'est pas rejetée. De plus la distance de Cramer-Von-Mises entre la copule empirique et la copule de Gumbel est la plus petite (relativement aux autres copules théoriques testées).

Il ressort des résultats du test de Kolmogorov d'adéquation que la copule Gumbel est celle qui a la plus petite distance au sens de Kolmogorov. Au sens de ce critère, la copule Gumbel est meilleure que les autres copules testées.

Plutôt que de regarder la structure de dépendance dans sa globalité, nous allons par la suite étudier les dépendances dans les queues inférieures et supérieures.

Les fonctions de concentrations dans les queues inférieure et supérieure de chaque copule ont été estimées; ces fonctions ont été comparées à leur équivalent empirique.

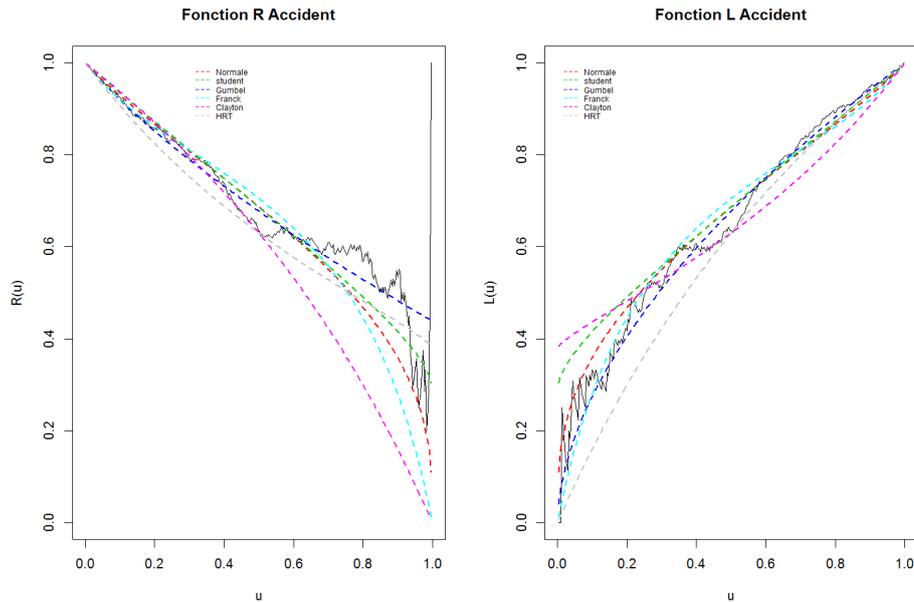


FIGURE 6.4 – Fonctions de concentration dans les queues

Le graphique ci-dessus montre une bonne adéquation entre les fonctions empiriques de concentration dans les queues inférieure et supérieure et les fonctions de concentration dans les queues inférieure et supérieure de la copule de Gumbel.

Pour confirmer ce constat graphique, nous avons calculé les écarts quadratiques moyens entre les fonctions empiriques de concentration dans les queues inférieure et supérieure et les fonctions de concentration dans les queues inférieure et supérieure des différentes copules testées. Le résultat obtenu est le suivant :

Copule	Fonction R	Fonction L
Gaussienne	0,0069	0,0014
Student	0,0038	0,0042
Gumbel	0,0029	0,00094
Frank	0,012	0,0017
Clayton	0,029	0,0071
HRT	0,0045	0,0069

TABLE 6.3 – Distance entre les Fonctions empirique et théoriques de concentration dans les queues

Il en ressort que la copule de Gumbel est celle qui a les plus petits écarts quadratiques moyens : La copule de Gumbel est celle qui capte le mieux la structure de dépendance tant dans la queue inférieure que dans la queue supérieure entre les sinistralités attritionnelle et grave de la MCC « Accident ».

Au regard des différents critères utilisés pour le choix de la copule, il en ressort que la copule de Gumbel est celle qui capte le mieux la structure de dépendance non seulement au global, mais aussi dans les queues inférieure et supérieure entre les sinistralités attritionnelle et grave de la MCC « Accident ».

Le processus utilisé pour déterminer la copule qui capte le mieux la structure de dépendance entre les sinistralités attritionnelle et grave de la MCC « Accident » a été reproduit pour déterminer les copules qui ajustent le mieux la structure de dépendance entre les sinistralités attritionnelle et grave des autres MCC.

Un graphique étant beaucoup plus parlant/interprétable, nous avons jugé pertinent de tracer le graphique des copules empiriques :

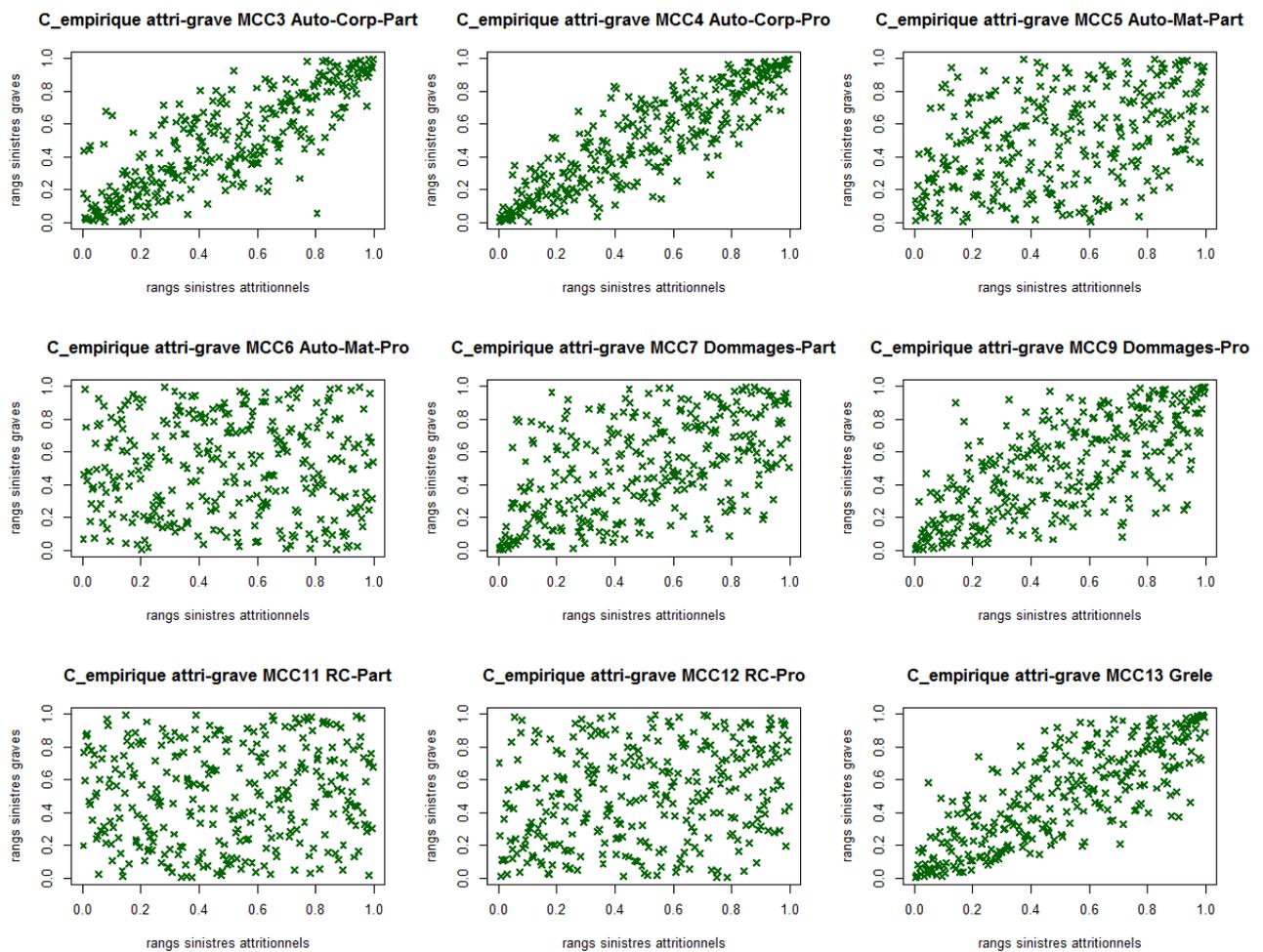


FIGURE 6.5 – Copules empiriques entre sinistralités attritionnelle et grave

Les résultats obtenus sur l'ensemble des MCC sont les suivants :

MCC	Copule retenue	Paramètres
Accident	Gumbel	1.55
Auto-Corp-Part	Student	0,83 ; df=12
Auto-Corp-Pro	Student	0,86 ; df=5
Auto-Mat-Part	Gaussienne	0,43
Auto-Mat-Pro	Clayton	0,2
Dommmages-Part	Clayton	0,86
Dommmages-Pro	Student	0,76 ; df=33
RC-Part	Clayton	0,14
RC-Pro	Clayton	0,44
Grêle	Student	0.84 ; df=5

TABLE 6.4 – Agrégation attritionnel et grave par branche

Taille de l'échantillon

La taille de l'historique des données à disposition étant très petite, les données ont été simulées par la technique du Bootstrap présentée à la section 6.1. Pour une question de temps de calcul, pour la calibration des copules, nous avons dû toutefois nous restreindre à un échantillon de 300 données. En effet, au delà, les calculs étaient trop lourds avec le logiciel utilisé (le logiciel R).

d) Calcul du SCR de chaque MCC

Pour chaque MCC, nous savons maintenant la copule qui modélise le mieux la structure de dépendance entre la sinistralité attritionnelle et la sinistralité grave. Pour calculer le SCR d'une MCC donnée, on procède comme suit :

- (1) Simulation de 10000 charges attritionnelles et simulation de 10000 charges graves
- (2) Simulation de 10000 couples d'observations par la copule retenue
- (3) Détermination des vecteurs des rangs des 10000 couples d'observations simulés par la copule
- (4) Rangement des 10000 simulations de charges attritionnelles et de charges graves suivant les vecteurs de rangs obtenus à l'étape précédente
- (5) Addition ligne à ligne des vecteurs de charges attritionnelles charges graves rangés (à l'étape précédente) ; on obtient alors un vecteur X de 10000 lignes
- (6) Rangement croissant du vecteur obtenu à l'étape (5). La $VAR(99,5\%)$ est la 9950^{ème} plus grande valeurs sur les 10000. Le SCR s'obtient en retranchant la moyenne à la $VAR(99,5\%)$.

Les étapes (1) à (6) peuvent être répétées plusieurs fois ; dans ce cas le SCR est égal à la moyenne des SCR obtenu à chaque répétition.

MCC	SCR attri	SCR grave	SCR agrégé
Accident	8,3	2,3	10,1
Auto-Corp-Part	14,4	38,3	44
Auto-Corp-Pro	7,1	25,9	32,1
Auto-Mat-Part	46,3	7,9	50,3
Auto-Mat-Pro	40,4	6,5	41,9
Dommages-Part	36,6	9,9	40,1
Divers	0,096		0,096
Dommages-Pro	22,3	29,3	48,1
Pr-Juridique	9,5		9,5
RC-Part	2,1	3,78	4,7
RC-Pro	13,2	14	21,5
Grêle	40,3	9,1	48,2
Cons-DO	6,7		6,7
Cons-RCD	17,6		17,6
Cons-Autre	5,1		5,1

TABLE 6.5 – SCR des MCC

e) Estimation de la corrélation entre la sinistralité attritionnelle et la sinistralité grave pour chaque branche

A partir des SCR attritionnels, graves et des SCR des MCC, il est possible de déduire une corrélation implicite entre les SCR attritionnels, graves. En effet, l'agrégation des SCR dans la solvabilité 2, est une agrégation linéaire. En effet pour deux risques X et Y de corrélation linéaire ρ , on a la relation :

$$SCR_{X,Y} = \sqrt{SCR_X^2 + SCR_Y^2 + 2\rho * SCR_X * SCR_Y}$$

Dans cette équation à seule inconnue est la corrélation ρ . On peut écrire :

$$\begin{aligned} SCR_{X,Y} &= \sqrt{SCR_X^2 + SCR_Y^2 + 2\rho * SCR_X * SCR_Y} \\ \Leftrightarrow SCR_{X,Y}^2 &= SCR_X^2 + SCR_Y^2 + 2\rho * SCR_X * SCR_Y \\ \Leftrightarrow SCR_{X,Y}^2 - SCR_X^2 - SCR_Y^2 &= 2\rho * SCR_X * SCR_Y \\ \Leftrightarrow \rho &= \frac{SCR_{X,Y}^2 - SCR_X^2 - SCR_Y^2}{2 * SCR_X * SCR_Y} \end{aligned}$$

Cette égalité peut être utilisée pour déterminer la corrélation linéaire implicite entre la sinistralité attritionnelle et la sinistralité sur chaque segment de risque. Les résultats obtenus sont regroupés dans le tableau suivant :

MCC	corrélation
Accident	73%
Auto-Corp-Part	84%
Auto-Corp-Pro	85%
Auto-Mat-Part	42%
Auto-Mat-Pro	23%
Dommages-Part	24%
Dommages-Pro	73%
RC-Part	25%
RC-Pro	25%
Grêle	83%

TABLE 6.6 – Corrélacion entre sinistralité attritionnelle et sinistralité grave

Les SCR de chaque MCC ayant été calculés, il faut également les agréger pour obtenir le SCR au titre de la sinistralité future. Pour cela, nous allons définir une structure de dépendance entre les différents segments de risque (MCC).

6.2.2 Structure de dépendance entre MCC

Dans cette étude, le portefeuille est segmenté en 15 branches de garanties (MCC) et par conséquent nous devons calibrer $\frac{15 \cdot (15+1)}{2} = 105$ copules bivariées entre les MCC. Ce nombre étant très élevé, il est nécessaire de faire une analyse préliminaire permettant de sélectionner les dépendances les plus significatives. Nous avons fait de le choix d'utiliser le tau de Kendall qui permet de mesurer la concordance entre les variables. Seront donc calibrer les structures de dépendance des couples qui ont un tau de Kendall significatif.

Le Tau de Kendall est calculé sur des variables aléatoires, il est donc lui-même une variable aléatoire. Son espérance est nulle et sa variance est égale à :

a) Pré-sélection par le tau de Kendall

Le tau de Kendall $\tau(X, Y)$ de deux variables aléatoires X et Y est également une variable aléatoire. Il est égal à : $\tau(X, Y) = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$. On cherche à résoudre le problème de test suivant :

$$\begin{cases} \tau(X, Y) = 0 (H_0) \\ \tau(X, Y) \neq 0 (H_1) \end{cases}$$

Sur un échantillon $(x_1, y_1), \dots, (x_n, y_n)$ de taille n du couple (X, Y) , on estime le tau de Kendall $\tau(X, Y)$ par :

$$\tau_n = \frac{2}{n(n-1)} \sum_{j=2}^n \sum_{i=1}^{j-1} \text{sign}(x_j - x_i)(y_j - y_i)$$

Où la fonction $\text{sign}(z)$ est égale à 1 si z est positif et à -1 si z est strictement négatif. Son espérance est nulle et On estime sa variance par :

$$V(\tau_n) = \frac{2(2n+5)}{9n(n-1)}$$

Le Tau de Kendall tend assez rapidement vers une loi normale. Dans la littérature, on estime que huit à dix observations suffisent pour qu'il y ait convergence. On peut alors écrire :

$$\tau_n \longrightarrow N\left(0, \sqrt{\frac{2(2n+5)}{9n(n-1)}}\right), \text{ avec } n \rightarrow \infty$$

La statistique de test sous l'hypothèse (H0) est :

$$T = \frac{\tau_n}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}} \longrightarrow N(0, 1), \text{ avec } n \rightarrow \infty$$

On rejettera l'hypothèse H_0 si $|T| > u_{1-\frac{\alpha}{2}}$

Nous avons travaillé au seuil de 5% et nous avons comparé la valeur absolue de statistique de test au quantile d'ordre $1 - \frac{5\%}{2} = 97,5\%$ de la loi normale centrée réduite. Nous avons obtenu que 54 couples sur 105 ont des taux de Kendall significatifs. Par ces 54 couples, 23 ont des tau de Kendall négatifs et 31 ont des tau de Kendall positifs. Ces couples sont donnés dans les tableaux suivants :

	Tau de Kendal	Statistique de test
Accident et Auto-Mat-Part	-0,17	-4,35
Auto-Corp-Part et Auto-Mat-Part	-0,14	-3,65
Auto-Corp-Part et Dommages-Pro	-0,12	-3,00
Auto-Corp-Part et Pr-Juridique	-0,09	-2,20
Auto-Corp-Part et Cons-DO	-0,24	-6,27
Auto-Corp-Part et Cons-RCD	-0,19	-4,81
Auto-Corp-Part et Cons-autre	-0,17	-4,44
Auto-Corp-Pro et Pr-Juridique	-0,20	-5,16
Auto-Corp-Pro et Cons-RCD	-0,08	-2,18
Auto-Mat-Part et Pr-Juridique	-0,09	-2,28
Auto-Mat-Pro et Divers	-0,12	-3,03
Auto-Mat-Pro et Pr-Juridique	-0,25	-6,54
Auto-Mat-Pro et Grele	-0,10	-2,49
Auto-Mat-Pro et Cons-DO	-0,16	-4,06
Auto-Mat-Pro et Cons-RCD	-0,13	-3,23
Dommages-Part et Pr-Juridique	-0,24	-6,29
Dommages-Part et Cons-DO	-0,08	-2,18
Divers et Dommages-Pro	-0,09	-2,20
Divers et RC-Part	-0,08	-2,04
Pr-Juridique et RC-Part	-0,11	-2,82
RC-Part et Cons-DO	-0,39	-10,17
RC-Part et Cons-RCD	-0,12	-3,16
RC-Pro et Cons-DO	-0,24	-6,08

TABLE 6.7 – Significativité du Tau de Kendall (A)

	Tau de Kendal	Statistique de test
Accident et Auto-Corp-Part	0,10	2,63
Accident et Auto-Corp-Pro	0,09	2,39
Accident et Dommages-Part	0,11	2,90
Accident et RC-Pro	0,22	5,77
Accident et Grele	0,21	5,55
Accident et Cons-RCD	0,19	4,95
Accident et Cons-autre	0,08	2,18
Auto-Corp-Part et Auto-Mat-Pro	0,12	3,23
Auto-Corp-Part et Dommages-Part	0,14	3,54
Auto-Corp-Part et RC-Part	0,32	8,30
Auto-Corp-Pro et Auto-Mat-Pro	0,19	4,89
Auto-Corp-Pro et Dommages-Part	0,18	4,59
Auto-Corp-Pro et Dommages-Pro	0,10	2,49
Auto-Corp-Pro et Grele	0,13	3,37
Auto-Mat-Part et Auto-Mat-Pro	0,20	5,10
Auto-Mat-Part et Dommages-Pro	0,23	6,02
Auto-Mat-Pro et Dommages-Part	0,27	6,86
Auto-Mat-Pro et Dommages-Pro	0,24	6,20
Auto-Mat-Pro et RC-Part	0,36	9,19
Dommages-Part et Dommages-Pro	0,24	6,24
Pr-Juridique et Grele	0,12	3,05
Pr-Juridique et Cons-DO	0,25	6,42
Pr-Juridique et Cons-RCD	0,32	8,16
RC-Part et RC-Pro	0,17	4,30
RC-Pro et Cons-RCD	0,08	2,04
RC-Pro et Cons-autre	0,08	1,98
Grele et Cons-DO	0,18	4,61
Grele et Cons-RCD	0,12	3,00
Grele et Cons-autre	0,09	2,22
Cons-DO et Cons-RCD	0,14	3,64
Cons-DO et Cons-autre	0,10	2,57

TABLE 6.8 – Significativité du Tau de Kendall (B)

b) Exclusion des couples avec des dépendances négatives

Des dépendances négatives entre segments de risque signifieraient que plus la probabilité qu'un segment subisse une perte élevée augmente, plus la probabilité que l'autre segment subisse une perte élevée diminue ; ce qui nous paraît illusoire. De plus, l'intérêt de la compagnie portant avant tout sur les dépendances susceptibles de générer un besoin additionnel en capital, nous avons décidé d'omettre les dépendances négatives.

Le procédé de détermination de la copule qui ajuste le mieux la structure de dépendance entre les MCC est le même que celui utilisé pour déterminer la structure de dépendance entre

la sinistralité attritionnelle et la sinistralité grave des différentes MCC.

Nous détaillerons le processus sur les paires suivantes :

- Auto-Corp-Part et RC-Part
- Dommages-Part et Dommages-Pro
- Pr-juridique et Cons-RCD

Le détail des résultats obtenus sur les autres paires de MCC seront présentés en annexes.

c) Visualisation graphique des structures de dépendance : copules empiriques

La copule empirique qui correspond au couple des rangs normalisés par le nombre d'observations permet d'avoir une idée sur la structure de dépendance entre les variables étudiées. Bien qu'elle ne permet pas d'estimer les paramètres de la copule, elle permet d'observer le caractère plus ou moins simultané des réalisations issues de l'échantillon. Il sera également utile d'analyser si ce caractère simultané est également présent dans les queues.

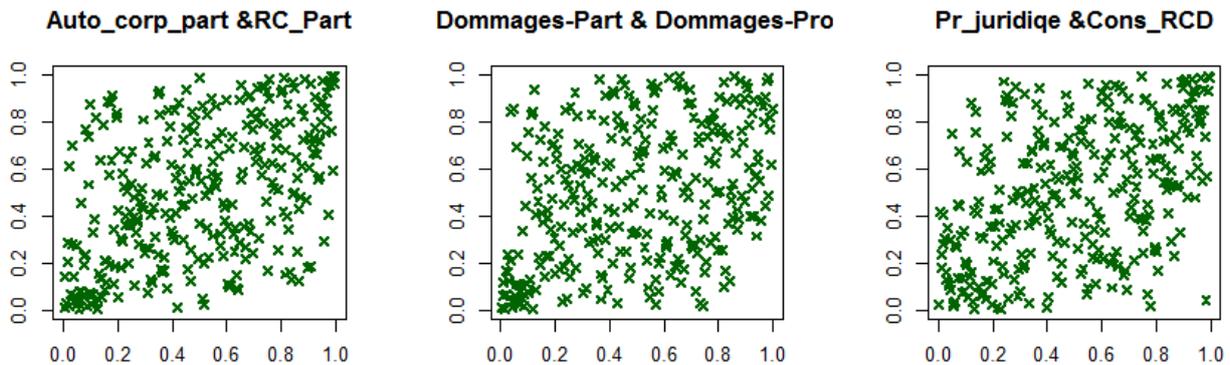


FIGURE 6.6 – Copules empiriques entre les sinistralités des MCC

Il ressort des graphiques ci-dessus que :

- Les Mcc Auto-Corp-Part et RC-Part sont fortement corrélés positivement dans la queue inférieure et dans la queue supérieure.
- La dépendance entre les MCC Auto-Mat-Pro et Dommages-Pro est plus accentuée dans la queue inférieure relativement à la queue supérieure où la dépendance n'est pas prononcée.
- Les MCC Pr-juridique et Cons-RCD présentent une dépendance positive de faible intensité dans la queue inférieure et dans la queue supérieure.

d) Estimation des paramètres des différentes copules

	Copule	Paramètre (s)	Log L	IC à 95%
Auto-Corp-Part et RC-Part	Gaussienne	0,52	45,24	[0,45 ; 0,59]
	Student	0,51 ;df=12	45,67	[0,41 ; 0,62]
	Gumbel	1,45	40,4	[1,3 ; 1,6]
	Franck	3,2	36,6	[2,5 ; 3,96]
	Clayton	0,8	39,2	[0,58 ; 1]
	HRT	0,71	33,5	[0,51 ; 0,9]
Dommages-Part et Dommages-Pro	Gaussienne	0,41	26,55	[0,32 ; 0,5]
	Student	0,41 ;df=4675	26,55	[0,31 ; 0,52]
	Gumbel	1,28	18,2	[1,2 ; 1,4]
	Franck	2,29	20,1	[1,58 ; 3]
	Clayton	0,64	28,2	[0,38 ; 0,89]
	HRT	0,64	11,01	[0,38 ; 0,89]
Pr-juridique et Cons-RCD	Gaussienne	0,49	38,5	[0,41 ; 0,56]
	Student	0,49 ;df=29	38,8	[0,39 ; 0,58]
	Gumbel	1,42	36,6	[1,3 ; 1,5]
	Franck	3,1	35,3	[2,4 ; 3,8]
	Clayton	0,6	26,6	[0,43 ; 0,80]
	HRT	0,7	32,7	[0,5 ; 0,89]

TABLE 6.9 – Estimation des paramètres des Copules entre les sinistralités des MCC

Les intervalles de confiance à 95% des paramètres estimés ne contiennent pas 0 ; ce qui signifie que les paramètres sont significatifs au seuil de 5%.

e) Choix de la copule

Nous avons utilisé les critères suivants :

- Les critères de maximisation de la vraisemblance et de minimisation du critère AIC.
- Minimisation de la distance de Cramer-Von-Mises.
- Minimisation de la distance de Kolmogorov.
- Comparaison des fonctions empiriques et théoriques de concentration dans les queues supérieures et inférieures.

Nous accorderons également un poids plus élevé à la fonction de concentration R ; en effet, ne perdons pas de vue que nous recherchons la $VAR(99, 5\%)$ et par conséquent, il est plus pertinent de se concentrer sur la copule pour laquelle il y a une bonne adéquation dans la queue supérieure.

Le tableau 6.9 indique qu'il est plus vraisemblable que :

- La structure de dépendance entre les MCC Auto-Corp-Part et RC-Part soit une structure de Student ; Toutefois, la vraisemblance de la Copule Normale est proche de celle de Student ;

- La structure de dépendance entre les MCC Dommages-Part et Dommages-Pro soit une structure de Clayton ;
- La structure de dépendance entre les MCC Pr-Juridique et Cons-RCD soit une structure de Student ; Toutefois, la vraisemblance de la Copule Normale est proche de celle de Student. De plus, le fait que le degré de liberté de la copule de Student soit élevé, nous laisse penser que la copule Normale serait meilleure.

Afin de pouvoir évaluer l'adéquation des différentes copules théoriques testées à la copule empirique, des tests d'adéquation basés sur les distances de Cramer-Von-Mises et de Kolmogorov ont été effectués. Les résultats obtenus sont renseignés dans le tableau suivant :

	Copules	Cramer Von-Mises		Kolmogorov	
		Distance	P-value	Distance	P-value
Auto-Corp-Part et RC-Part	Gaussienne	0,13	0,13	0,77	0,33
	Student	0,11	0,30	0,75	0,41
	Gumbel	0,33	0,003	1,39	0,007
	Frank	0,22	0,06	1,12	0,04
	Clayton	0,13	0,31	0,78	0,3
	HRT	0,4	0,005		
Dommages-Part et Dommages-Pro	Gaussienne	0,17	0,08	0,97	0,083
	Student	0,17	0,08	0,98	0,08
	Gumbel	0,45	0	1,61	0
	Frank	0,29	0,04	1,33	0,02
	Clayton	0,096	0,52	0,7	0,62
	HRT	0,4	0,005		
Pr-juridique et Cons-RCD	Gaussienne	0,04	0,87	0,48	0,95
	Student	0,04	0,87	0,52	0,89
	Gumbel	0,2	0,06	1,1	0,04
	Frank	0,08	0,45	0,73	0,44
	Clayton	0,23	0,045	0,998	0,14
	HRT	0,4	0,05		

TABLE 6.10 – Choix de la copule entre MCC : test d'adéquation

Il ressort des résultats des tests de Cramer-Von-Mises et Kolmogorov d'adéquation que :

- Pour les MCC Auto-Corp-Part et RC-Part, la copule de Student est celle qui a la plus petite distance au sens de ces deux tests. Au sens de ces critères, la copule de Student est meilleure que les autres copules testées.
- Pour les MCC Dommages-Part et Dommages-Pro, la copule qui minimise les distances de Cramer-Von-Mises et de Kolmogorov est la copule de Clayton.
- Pour les MC Pr-juridique et Cons-RCD, la copule qui minimise les distances de Cramer-Von-Mises et de Kolmogorov est la copule de Gaussienne.

Plutôt que de regarder la structure de dépendance dans sa globalité, nous allons par la suite étudier les dépendances dans les queues inférieures et supérieures.

Les fonctions de concentrations dans les queues inférieure et supérieure de chaque copule ont été estimées ; ces fonctions ont été comparées à leurs équivalents empiriques.

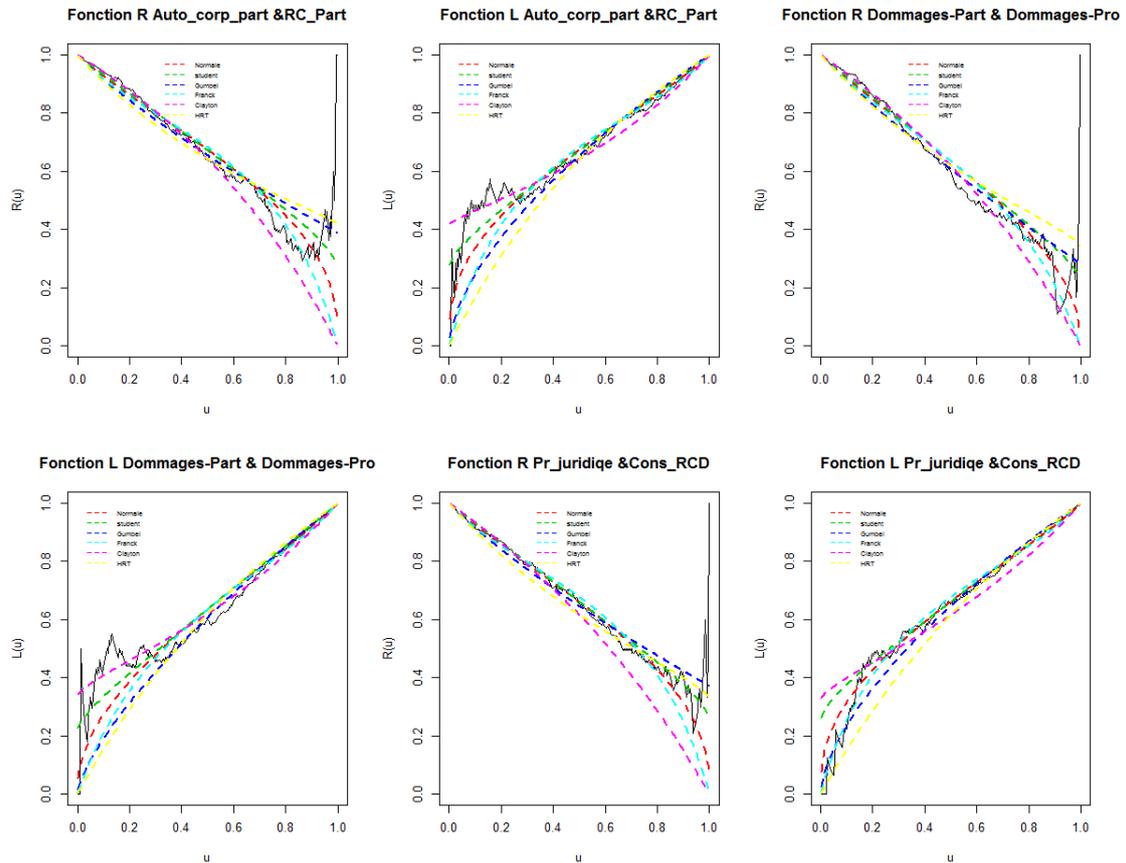


FIGURE 6.7 – Copule entre MCC : Fonctions de concentration dans les queues

A partir du graphique ci-dessus, il n'est toujours pas aisé de déterminer les copules dont l'adéquation entre les fonctions empiriques et théoriques de concentration dans les queues inférieure et supérieure est meilleure. Pour cela, nous avons calculé les écarts quadratiques moyens entre les fonctions empiriques et théoriques de concentration dans les queues inférieure et supérieure des différentes copules testées. Les copules qui ont les plus petits écarts quadratiques moyens sont renseignés dans le tableau suivant :

MCC	Fonction R	Fonction L
Auto-Corp-Part et RC-Part	Student	Student
Dommages-Part et Dommages-Pro	Clayton	Clayton
Pr-Juridique et Cons-RCD	Gaussienne	Gaussienne

TABLE 6.11 – Copule entre MCC : Distance entre les de fonctions concentration dans les queues empirique et théoriques

Au regard des différents critères utilisés pour le choix de la copule, nous retenons la copule de Student pour modéliser la structure de dépendance entre les MCC Auto-Corp-Part et RC-Part ; la copule de Clayton pour modéliser la structure de dépendance entre les MCC Dommages-Part et Dommages-Pro, et la copule Normale pour modéliser la structure de dépendance entre les MCC Pr-Juridique et Cons-RCD.

f) Copules retenues sur les couples sélectionnés

Le procédé d'estimation des paramètres des copules et du choix de la « meilleure » copule a été répliqué sur tous les couples dont le tau de Kendall est positif et significatif. Nous avons obtenu les résultats suivants :

	Copule retenue	Paramètre (s)
Accident et Auto-Corp-Part	Student	0,18, df=18
Accident et Auto-Corp-Pro	Franck	0,83
Accident et Dommages-Part	Clayton	0,27
Accident et RC-Pro	Normale	0,37
Accident et Grele	Gumbel	1,27
Accident et Cons-RCD	Normale	0,30
Accident et Cons-autre	Franck	0,78
Auto-Corp-Part et Auto-Mat-Pro	Normale	0,22
Auto-Corp-Part et Dommages-Part	Franck	1,26
Auto-Corp-Pro et Auto-Mat-Pro	Gumbel	1,23
Auto-Corp-Pro et Dommages-Part	Franck	1,61
Auto-Corp-Pro et Dommages-Pro	Normale	0,25
Auto-Corp-Pro et Grele	Clayton	0,31
Auto-Mat-Part et Auto-Mat-Pro	Normale	0,30
Auto-Mat-Part et Dommages-Pro	Franck	2,2
Auto-Mat-Pro et Dommages-Part	Normale	0,41
Auto-Mat-Pro et Dommages-Pro	Student	0,37 ; df=6
Auto-Mat-Pro et RC-Part	Normale	0,53
Pr-Juridique et Grele	Normale	0,19
Pr-Juridique et Cons-DO	Student	0,38 ; df=24
Pr-Juridique et Cons-RCD	Normale	0,49
RC-Part et RC-Pro	Normale	0,25
RC-Pro et Cons-RCD	Gumbel	1,01
RC-Pro et Cons-autre	Franck	0,69
Grele et Cons-DO	Clayton	0,43
Grele et Cons-RCD	Franck	1,05
Grele et Cons-autre	Student	0,15 ; df=20
Cons-DO et Cons-autre	Clayton	0,24

TABLE 6.12 – Copules entre MCC

g) Calcul du SCR agrégé de deux branches (MCC)

Pour chaque couple (A,B) de segments de risque, nous savons maintenant la copule qui modélise le mieux la structure de dépendance entre les segments de risque A et B. Pour calculer le SCR d'un couple de segments de risque (A,B), on procède comme suit :

- (1) Simulation de 10000 charges attritionnelles et simulation de 10000 charges graves du segment de risque A
- (2) Simulation de 10000 couples d'observations par la copule qui modélise la structure de dépendance entre les sinistralités attritionnelles et graves du segment A
- (3) Détermination des vecteurs des rangs des 10000 couples d'observations simulés par la copule qui modélise la structure de dépendance entre les sinistralités attritionnelles et graves du segment A
- (4) Rangement des 10000 simulations de charges attritionnelles et de charges graves suivant les vecteurs de rangs obtenus à l'étape précédente
- (5) Addition ligne à ligne des vecteurs de charges attritionnelles charges graves rangés (à l'étape précédente) ; obtient alors un vecteur X_A de 10000 lignes
- (6) On déroule les étapes (1) à (5) sur le segment B ; obtient alors un vecteur X_B de 10000 lignes
- (7) Simulation de 10000 couples d'observations par la copule qui modélise la structure de dépendance entre les segments A et B
- (8) Détermination des vecteurs des rangs des 10000 couples d'observations simulés par la copule qui modélise la structure de dépendance entre les segments A et B ; on obtient un vecteur des rangs $R = (R_A, R_B)$
- (9) Rangement de X_A suivant le vecteur R_A et de X_B suivant le vecteur R_B
- (10) Addition ligne à ligne des vecteurs X_A et X_B rangés à l'étape (9). On obtient alors un vecteur X .
- (11) Rangement croissant du vecteur X obtenu à l'étape (10). La $VAR(99,5\%)$ est la 9950^{ème} plus grande valeurs sur les 10000. Le SCR s'obtient en retranchant la moyenne à la $VAR(99,5\%)$.

Les étapes (1) à (11) peuvent être répétées plusieurs fois ; dans ce cas le SCR est égal à la moyenne des SCR obtenu à chaque répétition.

h) Estimation de la corrélation implicite entre les SCR des paires de MCC

L'agrégation des SCR dans la norme solvabilité 2 étant linéaire, nous utiliserons le même procédé que celui utilisé pour estimer la corrélation entre la sinistralité attritionnelle et la sinistralité grave des branches de garanties. Les résultats obtenus sont regroupés dans le tableau suivant :

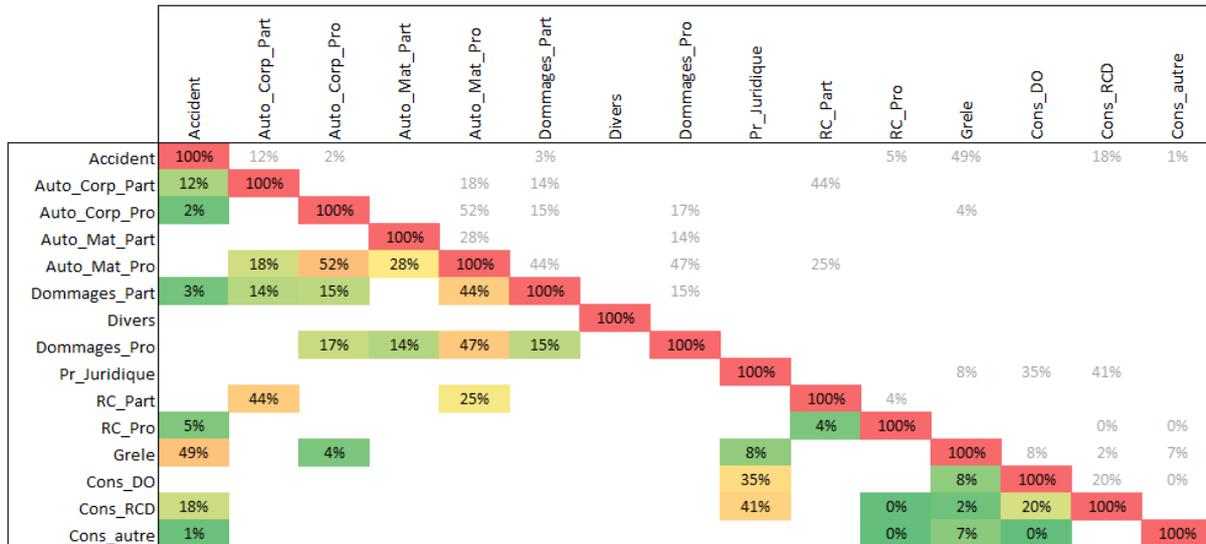


FIGURE 6.8 – Corrélations (méthode des copules)

Pour rappel, l'estimation de la dépendance par la théorie des copules a été faite sur les données obtenues en appliquant un ré-échantillonnage identique sur l'ensemble des branches. La structure de dépendance entre les branches est a priori prise en compte par cette structure identique de ré-échantillonnage.

Il ressort que :

- Les branches de garanties construction (Cons-DO et Cons-RCD) sont relativement corrélées à la branche de garanties Protection-juridique, ce qui est cohérent en ce sens où les sinistres construction coûtent très chers et sont le plus souvent sources de conflits juridiques.
- Les branches de garanties Auto-Mat-Pro et Dommages-Pro sont corrélées à hauteur de 47% ce qui est cohérent car les produits de ces deux branches de garanties sont généralement vendus ensemble ; ainsi, une perte sur l'une, constitue 50% de chance de perte sur l'autre.
- les branches de garanties Auto-Corp-Part et RC-Part sont corrélés à hauteur de 44%, ce qui nous paraît logique d'autant plus que les garanties sont similaires (responsabilité civile).

Malgré ces cohérences, on peut observer que les corrélations obtenues ne dépassent pas 53%. Le couple de branches de garanties le plus corrélé est Auto-Corp-Pro et Auto-Mat-Pro. De plus, certaines corrélations nous paraissent très faibles. En effet, on se serait attendu à avoir des corrélations relativement élevées entre les branches de garanties similaires telles que Dommages-Part et Dommages-Pro, Auto-Mat-Part et Auto-Mat-Pro, RC-Pro et Cons-RCD.

Ces observations nous poussent à penser que pour certaines paires de branches de garanties, cette méthode ne retranscrit pas assez bien les structures de corrélation entre celles-ci. De plus, il faut noter que cette méthode fournit les résultats des corrélations uniquement pour les couples de branches de garanties dont le tau de Kendall était significatif. Pour les autres couples, la corrélation n'a pas pu être estimée par cette méthode. Seules 31 sur 105 paires de

corrélations ont pu être obtenues par cette méthode.

On peut donc se rendre compte que ce n'est pas toujours facile de faire parler les données surtout quand l'historique utilisé est peu profond. Et si on essayait une approche qualitative pour calculer les corrélations ?

Une approche par facteur de causalité

Ce n'est pas toujours chose aisée de déterminer les corrélations à partir des données, d'autant plus que ces données pourraient conduire à des résultats pas satisfaisant ou inexplicables. Plutôt que de résoudre notre problématique par une approche quantitative tel que développé au chapitre précédent, nous allons aborder dans ce chapitre l'aspect qualitatif de la détermination des corrélations. Pour cela, nous avons faire appel à un panel d'experts qui a une connaissance assez large du portefeuille de la compagnie d'assurance. Le panel d'experts choisi est composé des experts de chaque branche de garanties.

L'exercice consiste dans un premier temps à définir avec les experts une liste d'évènements « inattendus » dont la survenance pourrait causer une déviation de la sinistralité future ; de tels évènements sont appelés « facteurs de causalité ». Dans un second temps, il a été question d'évaluer l'impact qu'aurait la survenance de ces évènements sur les différentes branches de garanties. Les impacts définis doivent être par la suite traduire en corrélations. L'idée fondamentale de cette approche est que plus deux branches de garanties sont impactées par les mêmes facteurs de causalité, plus elles sont corrélées.

L'approche par facteur de causalité ainsi décrite repose donc essentiellement sur des jugements d'experts. Pour chaque facteur de causalité, les experts de chaque branche de garanties devront répondre à la question suivante : « La survenance de ce facteur de causalité a-t-elle un impact sur le risque de sinistralité future porté par cette branche ? ». Si la réponse à cette question est oui, alors les experts devront attribuer le niveau d'impact (faible, moyen ou fort) qu'aurait la survenance du facteur de causalité sur la branche de garanties considérée.

A l'issu des différents ateliers effectués avec les experts, nous disposons de la liste des facteurs de causalité dont la survenance pourrait engendrer une déviation de la sinistralité des branches de garanties de la compagnie ainsi que les niveaux/ampleurs de cette déviation pour chaque branche de garantie. Un score de 0 sera attribué si le facteur de causalité n'a pas d'impact sur la branche de garanties, 1 si l'impact est faible, 2 si l'impact est moyen/modéré et 3 si l'impact est fort. Différentes méthodes seront décrites pour pouvoir traduire les impacts ainsi définis en corrélations.

7.1 Liste des facteurs de causalité

Les facteurs de causalité retenus doivent être exhaustifs et indépendants. En effet, si les facteurs de causalité ne sont pas exhaustifs alors les corrélations obtenues peuvent être sous-

estimées. Si par contre, les facteurs de causalité ne sont pas indépendants, alors les corrélations obtenues seront sur-estimées car il y aura un double comptage.

Les experts sont repartis de la liste des facteurs de causalité définis lors du calibrage de la norme solvabilité 2 et ont complété la liste avec d'autres facteurs de causalité qu'ils ont jugé pertinents pour la société. Une liste de 20 facteurs de causalité qui pourraient affecter la sinistralité future des branches de garanties a été définie.

7.1.1 La Super inflation

On parle de super-inflation lorsque le taux d'inflation de l'indice de marché sur lequel est indexé le contrat d'assurance est supérieur à celui de l'indice des prix à la consommation. Il s'agit donc du risque d'observer un taux d'inflation plus fort que celui de l'indice des prix à la consommation.

Un taux d'inflation plus fort que celui qui était prévu peut entraîner une déviation de la sinistralité future en ce sens où les montants de sinistres seront de plus en plus élevés soit à cause du fait que les valeurs des objets assurés seront plus élevées, soit à cause du fait que les factures, les montants de réparations (dans le cas de l'automobile par exemple) seront de plus en plus élevés.

Le graphique ci-dessous compare les variations annuelles des indices BT01, PASS et IPC.

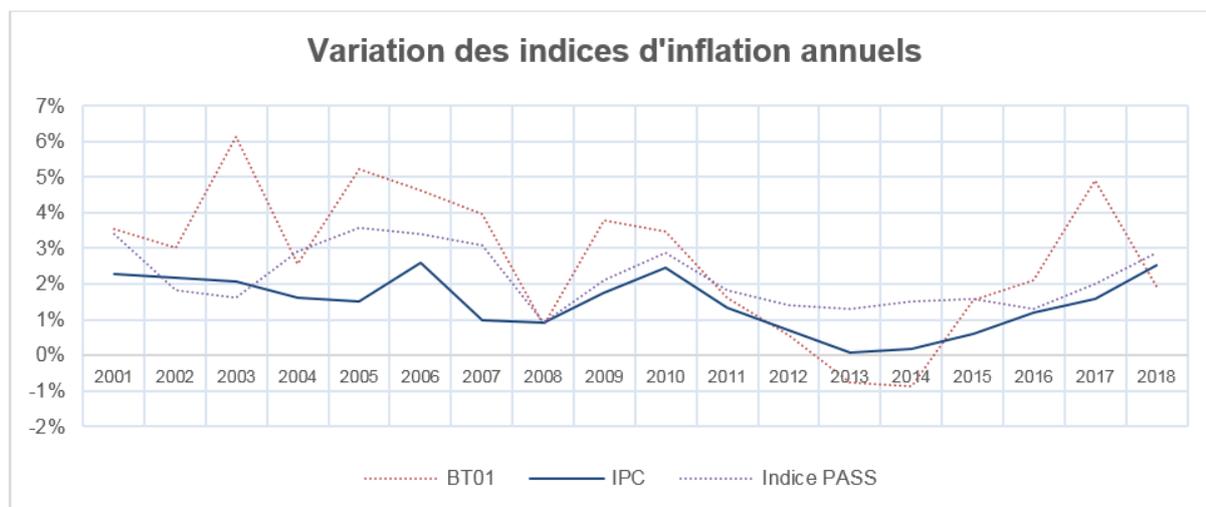


FIGURE 7.1 – Variation des indices d'inflation annuels

On peut observer que les branches dont les garanties sont indexées sur l'indice BT01 ou sur le PASS peuvent être impactées par la super-inflation. D'après le panel d'experts, les branches de garanties touchées par ce facteur de causalité sont :

- Auto-Corp-Pro et Auto-Corp-Part : Ces deux branches de garanties sont indexées sur l'IPC ; cependant, ce sont des branches longues, de plus, ces branches de garanties comportent un volet médical.
- Accident : cette garantie est indexée sur le PASS donc est impactée par la super-inflation.

- Cons-RCD et Cons-DO : Tel que l'indique le graphique ci-dessus, les garanties construction décennales qui sont indexées sur l'indice BT01 sont concernées par la super-inflation.

7.1.2 Augmentation du PIB

Il s'agit d'évaluer la déviation de la sinistralité due à l'augmentation de la richesse du pays.

Si les assurés sont plus riches, il peut y avoir un risque de changement de comportement de leur part. En effet, plus les assurés seront riches, plus ils auront tendance à s'offrir des loisirs luxueux, à augmenter probablement leur fréquence de voyage. ces éléments peuvent avoir un impact sur l'exposition au risque et donc sur la fréquence de survenance des sinistres.

D'autres part, Plus les assurés seront riches, plus ils auront tendance à acquérir des objets de valeurs plus élevées et auront également tendance à souscrire à de meilleures couvertures d'assurance.

Les branches de garanties touchées par ce facteur de causalité sont : Auto-Corp-Part, Auto-Corp-Pro, Auto-Mat-Part, Auto-Mat-Pro, Dommages-Part et Dommages-Pro, Cons-DO, Cons-RCD et Cons-autre.

7.1.3 Changement dans la pratique légale

Des changements réglementaires peuvent contraindre l'assureur à indemniser des sinistres qui n'étaient pas connaissables jusqu'alors. Nous pouvons citer quelques changements dans la pratique légale tels que :

- Les **nouvelles responsabilités médicales** qui ont vu le jour suite à l'arrêt perruche.
Arrêt Perruche : c'est l'histoire de Nicolas Perruche, né gravement handicapé à cause de la rubéole. En effet, sa mère a contracté une rubéole qui n'a pas été diagnostiquée au cours de sa grossesse, de ce fait, elle n'a pas pu recourir à une interruption volontaire de grossesse. En 1992, le tribunal de grande instance a reconnu « responsables de l'état de santé de l'enfant » le laboratoire et le médecin qui n'ont pas détecté la maladie de la mère et les « condamne in solidum avec leurs assureurs » à indemniser les parents et l'enfant. Dans cette affaire, le préjudice causé aux parents est incontestable ; cependant, le préjudice causé à l'enfant est contestable : un être humain n'est pas titulaire du droit de « naître ou de ne pas naître, de vivre ou de ne pas vivre ». Suite à une longue série de recours et de procès, en 2000, l'arrêt Perruche affirme le préjudice de l'enfant handicapé, contrairement aux jurisprudences précédentes qui indemnisaient les parents mais pas l'enfant handicapé.
- La **faute inexcusable** : la « faute inexcusable » de l'employeur est appliquée dans le cadre d'un accident du travail ou d'une maladie professionnelle ; elle vise à une meilleure indemnisation des victimes ou de leurs ayant droit. L'assurance de la faute inexcusable de l'employeur, longtemps interdite en France, a été autorisée par la loi de 1987. En 2002, la jurisprudence de la cours de cassation (arrêts du 28 février et du 11 avril) rend la recherche de la faute inexcusable encore plus facile en faisant peser sur l'employeur une « obligation de sécurité de résultat ».
- La **loi Hamon** votée en 2015, stipule que les assureurs ont l'obligation de rappeler à leurs clients leur libre choix du réparateur en cas de sinistres. Les assureurs n'ont donc plus le droit d'imposer aux assurés les réparateurs qui leurs sont agréés. Ce changement a un impact sur les montants de sinistres.

- Le **100% santé** ou le **reste à charge 0** qui donnent à tous les Français un accès à des soins de qualité, pris en charge à 100% par l'Assurance maladie obligatoire et complémentaire dans le domaine de l'optique, de l'audiologie et du dentaire. Il est valable hors dépassements d'honoraires et hors prestations non prises en charge.

Toutes les branches de garanties peuvent donc être impactées par des changements dans la pratique légale.

7.1.4 Erreur contractuelle

Ce facteur de causalité est relatif à une insuffisance des détails dans les contrats écrits conduisant à une couverture implicite des risques. D'après le panel d'experts interrogé, ce facteur de causalité n'a pas d'impact sur nos garanties ; en effet, ils considèrent que les contrats sont bien faits et que les choses qui ne sont pas couvertes explicitement par les contrats et qui peuvent l'être implicitement peuvent être dues à des avancées technologiques ou médicales et seront prises en compte dans les impacts de ces facteurs de causalité.

7.1.5 Évènements naturels

Les évènements naturels dûs aux conditions météorologiques défavorables tels que les inondations, les tremblements de terre, les tempêtes, la grêle ou encore le gel peuvent entraîner des demandes d'indemnisation plus nombreuses ou plus importantes. Ce facteur de causalité ne s'applique qu'aux branches de garanties relatives au dommage (Dommages-Part, Dommages-Pro) et à la grêle.

Ce facteur de causalité ne concerne pas les évènements naturels déclarés « catastrophe naturelle » ; de tels évènements sont gérés dans le module « Catastrophe naturelle ».

7.1.6 Vol et fraude interne

Il s'agit de la fraude et du vol interne. Par exemple, un vol interne des données ou des employés qui auraient de comportements frauduleux tel que l'indemnisation d'un sinistre qui ne devrait pas en principe être indemnisé. D'après le panel d'experts, ce risque bien que difficile à matérialiser, n'a pas été observé au sein de la compagnie ces dernières années. Ils ont affirmé qu'aucune branche de de garantie n'est impacté par ce facteur de causalité surtout qu'il est similaire aux facteurs de causalité « erreurs humaines » et « échecs de contrôle ».

7.1.7 Actions de l'homme

Ce facteur de causalité concerne les évènements provoqués par l'homme tels que les explosions, les actes de terrorisme, etc. Il s'agit de déterminer les branches de garanties sur lesquelles ont observera un grand volume ou coûts de sinistres à cause d'un acte de terrorisme par exemple. Ce facteur de causalité aura un impact uniquement sur les garanties dommages (Dommages-Part, Dommages-Pro) ; l'impact est jugé faible parce que les sinistres causés par les actes de terrorisme sont généralement couverts par la réassurance.

7.1.8 Erreurs humaines

Les erreurs humaines par exemple lors du traitement des informations sur les clients peuvent entraîner une déviation de la sinistralité future. Ce facteur de causalité a un impact sur toutes les branches de garanties. Les experts estiment que nous sommes toujours exposés aux erreurs.

7.1.9 Échec des contrôles

Des échecs dans le contrôle des données peuvent induire à une déviation de la sinistralité. C'est le cas du risque cyber-attaque par exemple qui est une menace de plus en plus importante pour les assureurs. Ce facteur de causalité est jugé redondant avec le facteur de causalité « erreurs humaines ».

7.1.10 Qualité des données et techniques

La qualité des données utilisées est d'une importance capital. En effet, si les données utilisées par l'assureur sont de mauvaises qualité, la tarification risque ne pas être en adéquation avec la sinistralité, les prévisions faites par l'assureur risque d'être erronées. Toutes les branches de garanties sont impactées par ce facteur de causalité.

7.1.11 Réseau de distribution

Un changement du réseau de distribution peut avoir une influence sur la survenance des sinistres surtout s'il s'agit du réseau de distribution de plusieurs produits, par exemple, les polices d'assurance achetées en gros (c'est le cas des assurances professionnelles).

Les garanties concernées sont :

- Auto-Corp-Part, Auto-Corp-Pro, Auto-Mat-Part, Auto-Mat-Pro : les garanties Auto-Corp-Pro et Auto-Mat-Pro sont principalement distribuées via le réseau des agents. Les garanties Auto-Corp-Part et Auto-Mat-Part sont distribuées par les agents et par les courtiers.
- Accident, RC-Part, RC-Pro ; elles sont distribuées par les agents et par les courtiers.
- Dommages-Part, Dommages-Pro, grêle : les produits de ces branches de garanties sont principalement distribuées via le réseau des agents
- Cons-DO, Cons-RCD , Cons-Autre : pour les garanties construction, on a beaucoup de courtage (50% du chiffre d'affaires)

7.1.12 Risque de longévité

Il s'agit ici de la déviation de la sinistralité due au fait que les assurés vivent plus longtemps. La survenance de ce facteur de causalité aura un impact uniquement sur les branches pour lesquelles il y a une possibilité de sortie en rente.

Avec l'évolution de plus en plus fulgurante de la médecine et de la communication sur la protection de la santé, l'espérance de vie n'a pas cessé de croître depuis 1960. En 2018, l'espérance de vie à la naissance est de 85,5ans pour les hommes et de 79,6 ans pour les femmes. D'après l'INSSE l'espérance de vie à la naissance va continuer de croître et atteindra 90 ans en 2070. Ce facteur de causalité a donc tout son sens pour une société d'assurance.

Les branches de garanties impactées par la survenance de facteur de causalité sont les suivantes : Auto-Corp-Part, Auto-Corp-Pro, RC-Part et RC-Pro.

7.1.13 Tendances sociales et culturelles

Des changements des tendances sociales et culturelles peuvent être une source de déviation de la sinistralité. C'est le cas des trottinettes électriques qui sont très tendances en ce moment et peuvent conduire à des sinistres de types responsabilité civile, contentieux juridiques, etc. Ces changements peuvent également conduire à nouveau besoin d'assurance et donc à des nouveaux sinistres.

Les garanties concernées sont :

- Auto-Corp-Part, Auto-Corp-Pro, Auto-Mat-Part et Auto-Mat-Pro : A cause des changements sociaux, de nouveaux types de sinistres peuvent voir le jour (exemple les sinistres causés par les trottinettes électriques). Il faut noter qu'en France, les trottinettes électriques sont considérées comme des véhicules terrestres à moteur au sens du code des assurances et par conséquent, elles sont donc soumises à l'obligation d'assurance responsabilité civile automobile.
- Dommages-Part, Dommages-Pro, RC-Part et RC-Pro : Les changements sociaux et des tendances peuvent avoir un impact sur la Responsabilité civile. Par exemple, un employé d'un salon de bien-être (les salons de bien-être sont très tendances de nos jours) peut par inadvertance appliquer un produit dangereux à un client ; dans ce cas, la RC-Pro du salon de bien-être sera engagée.
- Pr-Juridique : les changements sociaux et des tendances pourraient être source de nouveaux conflits juridiques.

7.1.14 Avancées médicales

L'amélioration des traitements médicaux, qui se traduit par une plus large gamme de produits que l'assuré peut utiliser peut entraîner une déviation de la sinistralité. En effet, les progrès médicaux entraînent souvent une augmentation des coûts des demandes de remboursements, car les nouvelles technologies médicales ont tendance à être onéreuses. C'est le cas des chirurgies réfractives et réparatrices par exemple, qui ont émergées avec les avancées médicales et qui, selon les conditions du contrat, peuvent coûter très chers à l'organisme d'assurance.

Les garanties impactées par ce facteur de causalité sont celles qui ont une composante corporelle c'est-à-dire : Auto-Corp-Part, Auto-Corp-Pro, RC-Part, RC-Pro.

7.1.15 Avancées technologiques

Les avancées technologiques (hors avancées médicales) peuvent augmenter la valeur des objets assurés et par conséquent, augmenter les montants des sinistres (par exemple les montants de réparation). Les branches de garanties concernées par ce facteur de causalité sont :

- Auto-Mat-Part, Auto-Mat-Pro : les avancées technologiques auront tendance à améliorer la sécurité des voitures ce qui va diminuer la fréquence des sinistres ; cependant, les coûts des sinistres seront élevés à cause de la valeur de l'objet assuré.
- Grêle : Les techniques d'agriculture peuvent être améliorées grâce aux avancées technologiques. L'arrivée des tracteurs par exemple, permettent aujourd'hui de cultiver de plus grandes surfaces et par conséquent, la perte si une tempête de grêle survient sera élevée.

7.1.16 Pratiques en matière d'emploi

Un manque de personnel qualifié peut conduire à une dérive de la sinistralité en ce sens où, il existe des branches qui nécessitent un niveau d'expertise élevé. Ce manque peut être dû à des difficultés de recrutement ou à un recrutement du personnel peu qualifié.

Les branches de garanties concernées par ce facteur de causalité sont :

- Auto-Corp-Part, Auto-Corp-Pro : un niveau d'expertise élevé est requis pour le gestionnaire des sinistres corporels qui peut intervenir pour démontrer la non responsabilité de l'assuré dans un accident automobile.
- Dommages-Part, Dommages-Pro : en cas d'incendie par exemple, il faudra des experts pour identifier la cause et la source de l'incendie.
- RC-Part, RC-Pro : Un niveau d'expertise est requis pour prouver la responsabilité ou non dans un sinistre.
- Cons-DO, Cons-RCD , Cons-autre : les garanties construction nécessitent un niveau d'expertise élevé. Par exemple en cas d'un sinistre lié au dommage ouvrage, l'assureur indemnisera l'assuré, mais fera appel aux experts pour rechercher la responsabilité des artisans qui ont intervenu sur le chantier.
- Grêle : les sinistres causés par la grêle nécessitent des compétences spécifiques
- Pr-Juridique : La loi réglemente le degré d'étude du personnel de gestion des sinistres relatifs la protection juridique.

7.1.17 Changement dans la gestion des réclamations des sinistres

Un changement dans le processus de gestion des réclamations relatif au niveau d'automatisation par exemple, peut causer une dérive de la sinistralité.

Toutes les branches de garanties à l'exception de la santé sont impactées par ce facteur de causalité. En effet, 95% des sinistres santé sont gérés de façon électronique et 5% sur papiers ; l'impact sur la santé est jugé nul.

7.1.18 Changement des actes de management

Les actes de management sont le plus souvent guidés par la compétition du marché. De la manière dont le panel d'experts a compris ce facteur de causalité, nous l'avons trouvé inclus dans le facteur de causalité « impact de la compétition du marché ».

7.1.19 Pandémie

Il s'agit ici d'évaluer l'impact qu'aurait la survenance d'une pandémie sur la sinistralité future.

Ce facteur de causalité est d'actualité avec le Covid-19 qui est une pandémie internationale et donc le bilan est de 1 793 224 cas confirmés et 110 052 décès à travers le monde¹. En France, on recense plus de 14 000 décès au 12/04/2020².

1. <https://www.sortiraparis.com>, consulté le 12/04/2020

2. op.cit.

Les mesures de confinement prises par le gouvernement français auront tendance à baisser la fréquence des sinistres de certaines branches de garanties ; c'est le cas par exemple des branches des garanties responsabilité civile (automobile et générale). En effet, moins les assurés seront en contact avec des gens, moins ils leur causeront des préjudices. Pour de telles branches, ce facteur de causalité ne constitue pas une source additionnelle de risque pour la société d'assurance.

Les contrats d'assurance complémentaire santé et les contrats de prévoyance par contre, peuvent intervenir pour indemniser les assurés (à hauteur des garanties souscrites). De même l'annulation de voyage ou le rapatriement sanitaire peuvent être couverts selon les contrats souscrits.

Pour ce qui concerne les entreprises, la pandémie du Covid-19 pourrait leur causer des préjudices tels que : la perte d'exploitation, la rupture de la chaîne d'approvisionnement, l'annulation d'événements et des défauts de livraison. En cas de pandémie ou d'épidémie, ces préjudices ne sont généralement pas couverts par les contrats d'assurances qui couvrent les entreprises. En effet, les conséquences économiques d'une pandémie de par son étendue, sont de fait inassurables.

Chez AVIVA, l'impact de la survenance d'une pandémie sur la sinistralité future est modélisée dans un module spécifique appelé *ENIDs* (*Events Not in Data*). Ce module modélise les événements susceptibles d'avoir un impact sur la sinistralité et sur les réserves, mais qui ne sont pas présents dans les données. Ainsi, d'une part, le Covid-19 aura un impact sur les garanties santé qui ne sont pas traitées dans ce mémoire, d'autre part, le Covid-19 sera traité dans le module *ENIDs*.

Nous tenons à préciser que l'impact économique du Covid-19 nous fait comprendre que les impacts qu'auraient la survenance d'événements extrêmes du type pandémie étaient sous-estimés. Par exemple, en réassurance, tant qu'il n'y a pas de catastrophe, les prix de réassurance sont relativement faibles et ceux-ci explosent après la survenance d'un gros événement ; pourtant le risque n'a pas intrinsèquement évolué.

7.1.20 Impact de la compétition du marché

Ce facteur de causalité concerne la déviation de la sinistralité future due à un changement de la stratégie du business à cause de la compétition du marché. Sur les produits d'appel comme les assurances automobiles et habitation, les évolutions du marché sont très rapides ; au risque de ne pas perdre ses clients, la société d'assurance doit suivre l'évolution non seulement en termes d'évolution informatique mais elle aussi, elle doit proposer des tarifs concurrentiels. Ce facteur de causalité a donc un impact direct sur le chiffre d'affaire ; l'impact est moins visible sur la sinistralité sur un horizon d'un an.

Toutefois, le panel d'experts a retenu un impact faible sur les garanties construction décennales car ce sont des engagements longs.

7.2 Quantification des facteurs de causalité

Après avoir établi la liste des facteurs de causalité ainsi que des branches de garanties qu'ils impactent, on doit quantifier les impacts de la survenance des ces facteurs de causalité.

L'impact d'un facteur de causalité sur une branche de garanties peut être évalué de plusieurs façons parmi lesquelles :

- Option 1 : utilisation d'une échelle binaire, il s'agit de faire une évaluation binaire (0 ou 1) de l'impact de la survenance du facteur de causalité sur la branche de garanties.
- Option 2 : les facteurs de causalité sont pris en compte via une échelle de notation à quatre niveaux de 0 à 3 (0 si pas d'impact, 1 si l'impact est faible, 2 si l'impact est moyen et 3 si l'impact est fort).
- Option 3 : elle découle de l'option 2 et traduit l'impact de la survenance du facteur de causalité en termes de proportion de la déviation de la sinistralité. Les niveaux suivants ont été retenus 0%, 5%, 10% et 15%. Le seuil de matérialité utilisé chez AVIVA étant de 5%, un facteur de causalité aura un impact sur une branche de garanties que s'il conduit à une déviation de la sinistralité d'au moins 5%.

Option 1	0	1	1	1
Option 2	0	1	2	3
Option 3	0%	5%	10%	15%

TABLE 7.1 – Quantification des facteurs de causalité

L'option 1 a l'avantage d'être simple et facile à implémenter. Cependant, elle ne permet pas de discriminer les facteurs de causalité en fonction de leurs niveaux d'impact sur les branches de garanties. L'option 1 sera appelée dans la suite par « **échelle binaire** ».

Les options 2 et 3 par contre, ont l'avantage de pouvoir discriminer les facteurs de causalité en fonction de leurs niveaux d'impact sur les branches de garanties. Cependant, elles rajoutent de la subjectivité dans l'évaluation de l'impact des facteurs de causalité. Dans la suite, l'option 2 sera appelée « **échelle ordinale** » et l'option 3 sera appelée « **échelle ordinale proportionnelle** »

Les impacts retenus pour l'échelle ordinale sont renseignés en annexe. Au risque d'être répétitif, nous n'avons pas renseignés les impacts obtenus avec les autres échelles car ceux ci peuvent être déduits des impacts de l'échelle ordinale.

Les trois échelles présentées ci-dessus seront utilisées pour calculer les corrélations entre les branches de garanties.

7.3 Calcul des corrélations

Il s'agit ici de répondre à la question suivante : « **comment traduire les informations récoltées en corrélation ?** »

Le calcul des corrélation entre les branches de garanties est juste une transcription des impacts des facteurs de causalité sur les branches de garanties. Cette transcription varie en fonction de l'échelle utilisée.

On peut considérer les branches de garanties comme des variables dont les observations sont les impacts des différents facteurs de causalité. Il s'agit alors d'un calcul de corrélations entre variables binaires ou ordinales.

Puisque les valeurs des variables sont des « nombres », on pourrait penser à utiliser les outils classiques de calcul des corrélations tels que la corrélation linéaire de Pearson, la corrélation des rangs de Spearman et le Tau de Kendall.

Il faut noter que le coefficient de corrélation linéaire de Pearson permet de mesurer les associations linéaires ; c'est une réelle mesure de corrélation si les variables sont Gaussiennes. Ce critère de normalité ne saurait être vérifiée dans notre cas car les variables prennent les valeurs dans $\{0,1\}$ ou $\{0,1,2,3\}$ ou $\{0\%, 5\%, 10\%, 15\%\}$.

Pour ce qui concerne le coefficient de corrélation des rangs de Spearman, il ne dépend pas des marginales car il est calculé sur les vecteurs des rangs. Cependant les valeurs prises par les variables sont $\{0,1\}$ ou $\{0,1,2,3\}$ ou $\{0\%, 5\%, 10\%, 15\%\}$, ce qui conduira à un nombre d'ex-aequo très important, rendant inappropriée cette métrique. En effet, le coefficient de corrélation des rangs de Spearman a du mal à caractériser correctement l'association entre les variables en présence d'un nombre d'ex-aequo très important.

Le Tau de Kendall classique ne peut pas également être appliqué à cause du fait que les variables ne sont pas continues, mais aussi à cause du grand nombre d'ex-aequo.

Nous présenterons différentes méthodes du calcul des corrélations qui pourraient s'appliquer dans le cas des variables dichotomiques et ordinales.

7.3.1 La méthode cardinale

C'est une mesure intuitive de la corrélation qui consiste à calculer le rapport entre le nombre de cas favorables et le nombre de cas possibles.

D'après cette méthode, la corrélation entre deux branches de garanties est égale au rapport du nombre de facteurs de causalité communs aux deux branches, divisé par le nombre total de facteurs de causalité. Plus deux branches de garanties auront des facteurs de causalité communs, plus elles seront corrélées.

Ceci peut être traduit comme suit : Soient deux branches de garanties B_i et B_j ; désignons par X_i et X_j l'ensemble des facteurs de causalité qui impactent les branches de garanties B_i et B_j respectivement. On a :

$$cor(B_i, B_j) = \begin{cases} \frac{card(X_i \cap X_j)}{card(X)} & Si \ B_i \neq B_j \\ 1 & Si \ B_i = B_j \end{cases}$$

Où X est l'ensemble des facteurs de causalité.

La méthode cardinale ne peut être utilisée que dans une évaluation binaire c'est-à-dire avec l'« échelle binaire ». Cette méthode a l'avantage d'être simpliste et facile à implémenter. Sa plus grosse faiblesse est de ne pas pouvoir différencier les paires (1,0), (0,1) et (0,0).

La matrice de corrélation obtenue avec la méthode cardinale et en utilisant une échelle binaire est la suivante :

	Accident	Auto_Corp_Part	Auto_Corp_Pro	Auto_Mat_Part	Auto_Mat_Pro	Dommmages_Part	Divers	Dommmages_Pro	PR_Juridique	RC_Part	RC_Pro	Grêle	Cons_DO	Cons_RCD	Cons_autre
Accident	100%	30%	30%	20%	25%	25%	0%	25%	20%	25%	25%	25%	30%	30%	25%
Auto_Corp_Part	30%	100%	55%	30%	35%	40%	0%	40%	30%	45%	45%	30%	40%	40%	35%
Auto_Corp_Pro	30%	55%	100%	30%	35%	40%	0%	40%	30%	45%	45%	30%	40%	40%	35%
Auto_Mat_Part	20%	30%	30%	100%	35%	30%	0%	30%	20%	25%	25%	25%	25%	25%	25%
Auto_Mat_Pro	25%	35%	35%	35%	100%	35%	0%	35%	25%	30%	30%	30%	30%	30%	30%
Dommmages_Part	25%	40%	40%	30%	35%	100%	0%	50%	30%	35%	35%	35%	35%	35%	35%
Divers	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Dommmages_Pro	25%	40%	40%	30%	35%	50%	0%	100%	30%	35%	35%	35%	35%	35%	35%
PR_Juridique	20%	30%	30%	20%	25%	30%	0%	30%	100%	30%	30%	25%	25%	25%	25%
RC_Part	25%	45%	45%	25%	30%	35%	0%	35%	30%	100%	45%	30%	30%	30%	30%
RC_Pro	25%	45%	45%	25%	30%	35%	0%	35%	30%	45%	100%	30%	30%	30%	30%
Grêle	25%	30%	30%	25%	30%	35%	0%	35%	25%	30%	30%	100%	30%	30%	30%
Cons_DO	30%	40%	40%	25%	30%	35%	0%	35%	25%	30%	30%	30%	100%	45%	40%
Cons_RCD	30%	40%	40%	25%	30%	35%	0%	35%	25%	30%	30%	30%	45%	100%	40%
Cons_autre	25%	35%	35%	25%	30%	35%	0%	35%	25%	30%	30%	30%	40%	40%	100%

FIGURE 7.2 – Méthode cardinale : échelle binaire

Toutes les corrélations obtenues sont positives. En effet, la méthode cardinale ne peut pas aboutir à des anti-corrélations.

On peut remarquer que la méthode cardinale aboutit à des corrélations pas très élevées (83% des corrélations sont inférieures ou égales à 40%). En effet, les corrélations obtenues par cette méthode peuvent être sous-estimées car tel que nous l'avons précisé, cette méthode ne distingue pas les paires (0,0), (0,1) et (1,0) ; aussi, elle suppose équivalentes toutes les paires (1,1) car elle ne prend pas en compte l'ampleur de l'impact des facteurs de causalité sur les branches de garanties.

Une analyse détaillée des résultats obtenus sera effectuée au chapitre suivant.

7.3.2 Le tau-b de Kendall

C'est une mesure de concordance pour les variables non continues ordinales. Il sera calculé entre les impacts des facteurs de causalité des différentes branches. Cette méthode peut être appliquée avec les trois options d'échelles décrites ci-dessus (l'échelle binaire, l'échelle ordinale et l'échelle ordinale proportionnelle).

Soient deux branches B_i et B_j , désignons par X_i (respectivement X_j) le vecteur d'impacts des facteurs de causalité sur la branche i (respectivement j). On a $X_i \in \{0, 1\}^{20}$ ou $X_i \in \{0, 1, 2, 3\}^{20}$ ou $X_i \in \{0, 5\%, 10\%, 15\%\}^{20}$ selon l'échelle utilisée. D'après cette méthode, la corrélation entre B_i et B_j est égal au tau-b de Kendall entre X_i, X_j . On peut alors écrire :

$$cor(B_i, B_j) = \tau_b(X_i, X_j) = 2 * \frac{P - Q}{\sqrt{D_{X_i} * D_{X_j}}}$$

Où P (respectivement Q) est le nombre de paires concordantes (respectivement discordantes) entre X_i et X_j ; $D_{X_i} = n * (n - 1) - d_{X_i}$ et $D_{X_j} = n * (n - 1) - d_{X_j}$. Enfin,

$d_{X_i} = \sum_{k=0}^h d_k^{x_i} (d_k^{x_i} - 1)$, avec h le nombre de modalités sans double comptage des ex-aequo et $d_k^{x_i}$ le nombre d'observations de la $i^{\text{ème}}$ modalité. De même, $d_{X_j} = \sum_{k=0}^l d_k^{x_j} (d_k^{x_j} - 1)$.

Nous tenons à préciser que cette méthode aboutira au même résultat que nous utilisons l'échelle ordinale ou l'échelle ordinale proportionnelle car $\{0,5\%,10\%,15\%\}=5\%*\{0,1,2,3\}$ et le tau-b de Kendall est invariant par transformation affine.

Les matrices de corrélation obtenues pour les différentes échelles sont les suivantes :

	Accident	Auto_Corp_Part	Auto_Corp_Pro	Auto_Mat_Part	Auto_Mat_Pro	Dommages_Part	Divers	Dommages_Pro	PR_Juridique	RC_Part	RC_Pro	Grêle	Cons_DO	Cons_RCD	Cons_autre
Accident	100%	29%	29%	23%	36%	15%	0%	15%	35%	25%	25%	36%	51%	51%	36%
Auto_Corp_Part	29%	100%	100%	15%	25%	23%	0%	23%	29%	66%	66%	4%	36%	36%	25%
Auto_Corp_Pro	29%	100%	100%	15%	25%	23%	0%	23%	29%	66%	66%	4%	36%	36%	25%
Auto_Mat_Part	23%	15%	15%	100%	82%	25%	0%	25%	23%	12%	12%	23%	12%	12%	23%
Auto_Mat_Pro	36%	25%	25%	82%	100%	36%	0%	36%	36%	23%	23%	35%	23%	23%	35%
Dommages_Part	15%	23%	23%	25%	36%	100%	0%	100%	39%	23%	23%	36%	23%	23%	36%
Divers	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Dommages_Pro	15%	23%	23%	25%	36%	100%	0%	100%	39%	23%	23%	36%	23%	23%	36%
PR_Juridique	35%	29%	29%	23%	36%	39%	0%	39%	100%	51%	51%	36%	25%	25%	36%
RC_Part	25%	66%	66%	12%	23%	23%	0%	23%	51%	100%	100%	23%	11%	11%	23%
RC_Pro	25%	66%	66%	12%	23%	23%	0%	23%	51%	100%	100%	23%	11%	11%	23%
Grêle	36%	4%	4%	23%	35%	36%	0%	36%	36%	23%	23%	100%	23%	23%	35%
Cons_DO	51%	36%	36%	12%	23%	23%	0%	23%	25%	11%	11%	23%	100%	100%	82%
Cons_RCD	51%	36%	36%	12%	23%	23%	0%	23%	25%	11%	11%	23%	100%	100%	82%
Cons_autre	36%	25%	25%	23%	35%	36%	0%	36%	36%	23%	23%	35%	82%	82%	100%

FIGURE 7.3 – Tau-b de Kendall : échelle binaire

	Accident	Auto_Corp_Part	Auto_Corp_Pro	Auto_Mat_Part	Auto_Mat_Pro	Dommages_Part	Divers	Dommages_Pro	PR_Juridique	RC_Part	RC_Pro	Grêle	Cons_DO	Cons_RCD	Cons_autre
Accident	100%	28%	25%	22%	33%	10%	0%	10%	33%	23%	23%	30%	43%	43%	31%
Auto_Corp_Part	28%	100%	90%	10%	23%	22%	0%	22%	33%	62%	62%	11%	28%	24%	20%
Auto_Corp_Pro	25%	90%	100%	15%	23%	13%	0%	13%	33%	68%	68%	6%	22%	24%	19%
Auto_Mat_Part	22%	10%	15%	100%	82%	21%	0%	20%	22%	12%	12%	22%	8%	8%	21%
Auto_Mat_Pro	33%	23%	23%	82%	100%	30%	0%	30%	33%	21%	21%	29%	16%	20%	34%
Dommages_Part	10%	22%	13%	21%	30%	100%	0%	99%	30%	14%	14%	35%	19%	13%	27%
Divers	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Dommages_Pro	10%	22%	13%	20%	30%	99%	0%	100%	30%	14%	14%	34%	19%	13%	27%
PR_Juridique	33%	33%	33%	22%	33%	30%	0%	30%	100%	54%	54%	38%	21%	20%	33%
RC_Part	23%	62%	68%	12%	21%	14%	0%	14%	54%	100%	100%	23%	6%	7%	19%
RC_Pro	23%	62%	68%	12%	21%	14%	0%	14%	54%	100%	100%	23%	6%	7%	19%
Grêle	30%	11%	6%	22%	29%	35%	0%	34%	38%	23%	23%	100%	23%	18%	31%
Cons_DO	43%	28%	22%	8%	16%	19%	0%	19%	21%	6%	6%	23%	100%	93%	77%
Cons_RCD	43%	24%	24%	8%	20%	13%	0%	13%	20%	7%	7%	18%	93%	100%	83%
Cons_autre	31%	20%	19%	21%	34%	27%	0%	27%	33%	19%	19%	31%	77%	83%	100%

FIGURE 7.4 – Tau-b de Kendall : échelle ordinale et échelle ordinale proportionnelle

Les niveaux de corrélations obtenus avec l'échelle binaire sont assez proches de ceux obtenus avec l'échelle ordinale.

Les cohérences observées entre les deux matrices de corrélation sont les suivantes :

- Les branches de garanties qui couvrent la responsabilité civile (Auto-Corp-Part, Auto-Corp-Pro, RC-Part et RC-Pro) sont fortement corrélées entre elles (de 62% à 68%)
- les branches de garanties qui couvrent la responsabilité civile (hors automobile) sont fortement corrélées à la garantie protection juridique.
- Les branches de garanties sur la Construction sont très corrélées entre elles ; il en est de même des branches de garanties qui couvrent les mêmes garanties mais qui diffèrent selon la nature de l'assuré (exemple : Auto-Corp-Part et Auto-Corp-Pro).

Une analyse détaillée des résultats obtenus sera effectuée au chapitre suivant.

7.3.3 La méthode additive

Cette méthode a la spécificité de pouvoir prendre en compte le fait que les facteurs de causalité utilisés n'expliquent pas nécessairement toute la déviation de sinistralité des différentes branches de garanties et calcule les corrélations uniquement sur la proportion de la déviation de la sinistralité qui est expliquée par les facteurs de causalité. Elle s'applique uniquement à l'échelle ordinale proportionnelle :

$$cor(B_i, B_j) = \mathbb{1}_{B_i \neq B_j} * \sum_{k=1}^{20} \min(X_i^k, X_j^k) + \mathbb{1}_{B_i=B_j}$$

Où X_i^k est l'impact du facteur de causalité k sur la branche de garanties i

Avec cette méthode, si les facteurs de causalité retenus expliquent $x\%$ de la déviation de la sinistralité d'une branche B , alors la corrélation de la branche B avec les autres branches de garanties sera au plus égale à $x\%$.

La matrice de corrélation obtenue pour cette approche est la suivante :

	Accident	Auto_Corp_Part	Auto_Corp_Pro	Auto_Mat_Part	Auto_Mat_Pro	Dommmages_Part	Divers	Dommmages_Pro	PR_Juridique	RC_Part	RC_Pro	Grêle	Cons_DO	Cons_RCD	Cons_autre
Accident	100%	30%	30%	20%	25%	25%	0%	25%	20%	25%	25%	25%	30%	30%	25%
Auto_Corp_Part	30%	100%	60%	30%	35%	45%	0%	45%	35%	50%	50%	40%	45%	40%	35%
Auto_Corp_Pro	30%	60%	100%	30%	35%	40%	0%	40%	35%	50%	50%	35%	40%	40%	35%
Auto_Mat_Part	20%	30%	30%	100%	35%	30%	0%	30%	20%	25%	25%	25%	25%	25%	25%
Auto_Mat_Pro	25%	35%	35%	35%	100%	35%	0%	35%	25%	30%	30%	30%	30%	30%	30%
Dommmages_Part	25%	45%	40%	30%	35%	100%	0%	60%	30%	35%	35%	45%	40%	35%	35%
Divers	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Dommmages_Pro	25%	45%	40%	30%	35%	60%	0%	100%	30%	35%	35%	50%	40%	35%	35%
PR_Juridique	20%	35%	35%	20%	25%	30%	0%	30%	100%	35%	35%	30%	25%	25%	25%
RC_Part	25%	50%	50%	25%	30%	35%	0%	35%	35%	100%	50%	35%	30%	30%	30%
RC_Pro	25%	50%	50%	25%	30%	35%	0%	35%	35%	50%	100%	35%	30%	30%	30%
Grêle	25%	40%	35%	25%	30%	45%	0%	50%	30%	35%	35%	100%	40%	35%	35%
Cons_DO	30%	45%	40%	25%	30%	40%	0%	40%	25%	30%	30%	40%	100%	60%	50%
Cons_RCD	30%	40%	40%	25%	30%	35%	0%	35%	25%	30%	30%	35%	60%	100%	50%
Cons_autre	25%	35%	35%	25%	30%	35%	0%	35%	25%	30%	30%	35%	50%	50%	100%

FIGURE 7.5 – Méthode additive

La tendance des corrélations obtenues est assez similaire à celle obtenue en utilisant la méthode cardinale. Cependant, la méthode additive parvient à prendre en compte l'ampleur de l'impact des facteurs de causalité ; c'est ce qui justifie que parfois ces corrélations sont plus élevées que celles de la méthode cardinale.

Une analyse détaillée des résultats obtenus sera effectuée au chapitre suivant.

Matrice de corrélation idéale

Diverses méthodes ont été utilisées pour calculer les corrélations entre les branches de garanties à savoir : la méthode des copules, la méthode méthode cardinale, le τ_b de Kendall, et la méthode additive. Nous allons comparer les corrélations obtenues avec chacune des différentes méthodes et nous allons également sur la base des résultats, proposer une matrice de corrélation que nous jugerons plus adaptée, en précisant le critère de choix utilisé.

Tout d'abord, le premier critère de référence est la matrice de corrélation QIS préconisée par l'EIPOA.

	1	2	3	4	5	6	7	8	9	10	11	12
1: Moteur (Responsabilité civile)	100%	50%	50%	25%	50%	25%	50%	25%	50%	25%	25%	25%
2: Moteur (autres)	50%	100%	25%	25%	25%	25%	50%	50%	50%	25%	25%	25%
3: Marine, aviation transport	50%	25%	100%	25%	25%	25%	25%	50%	50%	25%	50%	25%
4: Incendie	25%	25%	25%	100%	25%	25%	25%	50%	50%	25%	50%	50%
5: Responsabilité civile	50%	25%	25%	25%	100%	50%	50%	25%	50%	50%	25%	25%
6: Crédit	25%	25%	25%	25%	50%	100%	50%	25%	50%	50%	25%	25%
7: Protection juridique	50%	50%	25%	25%	50%	50%	100%	25%	50%	50%	25%	25%
8: Assistance	25%	50%	50%	50%	25%	25%	25%	100%	50%	25%	25%	50%
9: Divers	50%	50%	50%	50%	50%	50%	50%	50%	100%	25%	50%	25%
10: Réassurance non proportionnelle (corporelle)	25%	25%	25%	25%	50%	50%	50%	25%	25%	100%	25%	25%
11: Réassurance non proportionnelle (Mat)	25%	25%	50%	50%	25%	25%	25%	25%	50%	25%	100%	25%
12: Réassurance non proportionnelle (Dommage aux biens)	25%	25%	25%	50%	25%	25%	25%	50%	25%	25%	25%	100%

FIGURE 8.1 – Matrice de corrélation QIS préconisée par l'EIPOA

Il faut noter que plusieurs branches de garanties de notre segmentation sont regroupées en une seule branche de garanties dans la formule standard. C'est le cas des branches de garanties construction qui ne sont pas traitées spécifiquement dans la formule standard et qui pourtant, ont un profil de risque bien spécifique. De façon plus générale, on a les branches de garanties :

- Auto-Corp-Part et Auto-Corp-Pro qui appartient à la branche 1 Moteur (Responsabilité civile)
- Auto-Mat-Part et Auto-Mat-Pro qui appartient à la branche 2 Moteur (autres)
- Dommages-Part, Dommages-Pro, Grêle, Cons-DO et Cons-autre qui appartient à la branche 4 incendie
- RC-Part, RC-Pro et Cons-RCD qui appartient à la branche 5 Responsabilité civile

Si l'on s'en tient strictement à la formule standard, les branches de notre segmentation qui ne sont pas différenciées par la formule standard seraient corrélées à 100%. Nous avons estimée une matrice de corrélation que nous avons appelée « matrice QIS bis » dans laquelle, nous avons fait

le choix de mettre une corrélation de 75% entre toutes les branches de notre segmentation qui appartiennent à une même branche de la segmentation QIS. La matrice QIS bis sera utilisée pour faire des études de sensibilité.

8.1 Comparaison des différentes méthodes

8.1.1 Comparaison des SCR

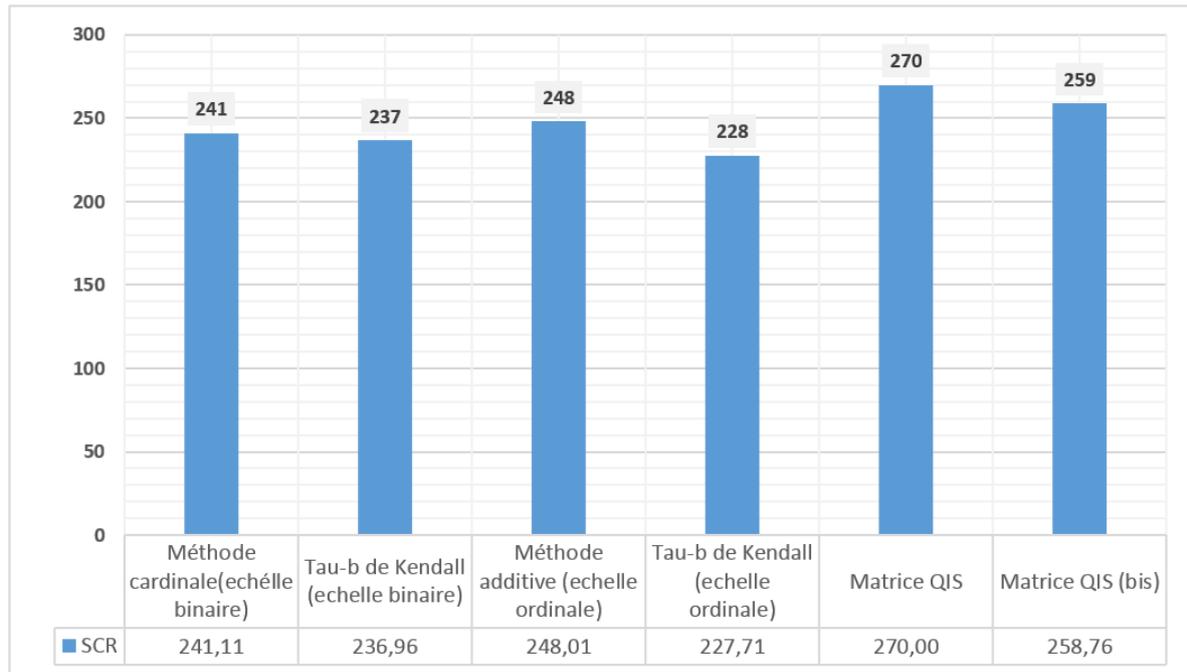


FIGURE 8.2 – Comparaison des SCR

Le SCR au titre de la sinistralité future obtenu avec les corrélations QIS est plus élevé que celui obtenu avec les corrélations calibrées.

Le tableau ci-après présente la répartition des corrélations. Aucune corrélation négative n'a été obtenue. Les corrélations nulles sont celles qui concernent la MCC8 (Divers). La majorité des corrélations obtenues avec le τ_b de Kendall est comprise entre 0% et 25%; celle des méthodes cardinale et additive est comprise entre 25% et 50%.

	r=0]0 ;25%]]25% ;50%]]50% ;75%]]75% ;100%]
Méthode cardinale (échelle binaire)	14	21	69	1	0
Tau-b de Kendall (échelle binaire)	14	50	26	8	7
Méthode additive (échelle ordinale)	14	20	68	3	0
Tau-b de Kendall (échelle ordinale)	14	54	24	6	7
Matrice QIS	0	46	44	0	15
Matrice QIS (bis)	0	46	44	15	0

TABLE 8.1 – Comparaison des statistique des corrélations

8.1.2 Méthode cardinale (échelle binaire) vs Méthode additive (échelle ordinaire proportionnelle)

Les corrélations obtenues par la méthode cardinale sont très proches de celles obtenues par la méthode additive. En effet, sur 105 paires de corrélations, nous avons la comparaison suivante entre les deux méthodes : 76 sont égales, 21 sont distantes de 5%, 6 sont distantes de 10% et 2 sont distantes de 15%. Les écarts les plus élevés sont relatifs aux corrélations entre la branche de garanties grêle et les branches de garanties Auto-Corp-Part, Dommages-Part, Dommages-Pro et également entre les paires Cons-DO & Cons-RCD et Dommages-Part & Dommages-Pro. De façon générale, la matrice de corrélation obtenue par la méthode additive est plus prudente que celle de la méthode cardinale. Ceci nous semble cohérent dans la mesure où :

- La méthode cardinale ne prend pas en compte l'ampleur des facteurs de causalité sur les branches de garanties ; ce qui n'est pas le cas de la méthode additive.
- La méthode cardinale est plus sensible au nombre de facteurs de causalité retenus ; en effet, si aux 20 facteurs de causalité retenus, on ajoute d'autres facteurs de causalité qui n'ont aucun impact sur deux branches de garanties données, alors la corrélation entre ces deux branches de garanties calculée avec la méthode cardinale va baisser alors que celle de la méthode additive restera inchangée.

8.1.3 Tau-b de Kendall - Échelle binaire vs Échelle ordinaire

Les corrélations calculées avec le τ_b de Kendall varient en fonction de l'échelle utilisée. L'écart moyen entre les deux matrices de corrélation (échelle binaire vs échelle ordinaire) est de 3% ; les écarts varient de -8% à 14%, ce qui donne une étendue de 22%.

De manière générale, Le SCR agrégé obtenu en utilisant la matrice de corrélation calculée avec le τ_b de Kendall est de 4% plus élevé en utilisant une échelle binaire plutôt qu'en utilisant une échelle ordinaire. Il est beaucoup plus difficile de comprendre les écarts entre les deux matrices ; en effet le passage d'une échelle à l'autre peut avoir un impact important sur le nombre de paires concordantes et le nombre de paires discordantes.

Les matrices de corrélation obtenues, nous montrent que la plus grosse limite du Tau-b de Kendall est de ne pas discriminer les paires de garanties construction entre elles et les paires de branches qui couvrent la même garantie mais qui diffèrent selon le statut de l'assuré (Particuliers vs Professionnels).

8.1.4 Échelle binaire - Méthode cardinale vs Tau-b de Kendall

Le SCR agrégé avec les coefficients de corrélation du τ_b de Kendall obtenus avec l'échelle binaire est moins élevé que le SCR agrégé avec les corrélations obtenues par la méthode cardinale avec la même échelle. En effet, les corrélations obtenues avec le τ_b de Kendall sont pour la plus part inférieures à celles obtenues avec la méthode cardinale. Ceci s'explique par le fait que la méthode cardinale ne considère pas les paires (0,1), (1,0) et (0,0) alors que le τ_b de Kendall considère toutes les paires et s'intéresse à la concordance entre celles-ci. On peut observer à partir de corrélations obtenues que le τ_b de Kendall ne distingue pas bien les branches de garanties qui couvrent les mêmes garanties mais qui diffèrent par le statut

de l'assuré (particuliers vs professionnels); pour ces branches de garanties, les corrélations estimées par le τ_b de Kendall sont supérieures à 82% alors que la méthode cardinale fournit des corrélations inférieures ou égales à 55%.

Il faut noter que, les couples de branches de garanties pour lesquels la méthode cardinale est beaucoup plus prudente que la méthode du τ_b de Kendall sont ceux formés des branches de garanties responsabilité civile automobile et de la branche de garanties grêle (l'écart vaut 26%). Par contre, les couples de branches de garanties pour lesquels les corrélations obtenues avec la méthode du τ_b de Kendall sont très élevées par rapport à celles obtenues avec la méthode cardinale sont ceux formés des branches de garanties qui couvrent les mêmes garanties mais qui diffèrent par le statut de l'assuré (particuliers vs professionnels) et des branches de garanties construction.

De façon générale, l'écart moyen entre les deux matrices de corrélation est de 1%; il y'a de fortes raisons de penser que les écarts positifs compensent les écarts négatifs.

8.1.5 Échelle ordinale - Méthode additive vs Tau-b de Kendall

Tel que nous l'avons mentionné, la méthode additive a la spécificité de pouvoir prendre en compte le fait que les facteurs de causalité utilisés n'expliquent pas nécessairement toute la déviation de la sinistralité des différentes branches de garanties et calcule les corrélations en prenant en compte la part résiduelle de la proportion de la déviation de la sinistralité qui n'est pas expliquée par les facteurs de causalité.

On peut observer à partir des corrélations obtenues que, le τ_b de Kendall ne distingue pas bien les branches de garanties qui couvrent les mêmes garanties mais qui se distinguent par le statut de l'assuré (particuliers vs professionnels); pour ces branches de garanties, les corrélations estimées par le τ_b de Kendall sont supérieures à 82%. Pourtant, lorsqu'on traduit les impacts des facteurs de causalité en pourcentage de la déviation de la sinistralité future, on se rend compte que ces branches de garanties ne sont pas si corrélées. En effet, avec la méthode additive, la corrélation entre deux branches de garanties ne peut pas excéder les déviations totales qu'aurait la survenance des facteurs de causalité retenus sur chacune des deux branches.

L'écart moyen entre la matrice de corrélation obtenue avec la méthode additive et celle obtenue avec le τ_b de Kendall avec l'échelle ordinale est de 4%, ce qui signifie qu'en moyenne les corrélations obtenues avec le τ_b de Kendall sont inférieures à celles obtenues avec la méthode additive. Cependant, l'écart-type des écarts vaut 0,16 et on obtient un coefficient de variation égal à 3,9. Le coefficient de variation des écarts est très élevé; ce qui signifie qu'on a des « grands » écarts positifs et négatifs (car la moyenne est proche de 0) entre les corrélations obtenues par ces deux méthodes. Ceci est confirmé par l'étendue des écarts qui est de 79%; en effet les écarts sont compris entre -50% et 29%. Les couples de branches de garanties pour lesquelles on observe des écarts les plus élevés (-50% et -47%) concernent les paires de branches qui couvrent les mêmes garanties mais qui diffèrent par le statut de l'assuré.

8.1.6 Analyse détaillée des différentes corrélations

Dans cette partie, nous ferons une analyse au cas par cas des corrélations obtenues avec les différentes méthodes étudiées.

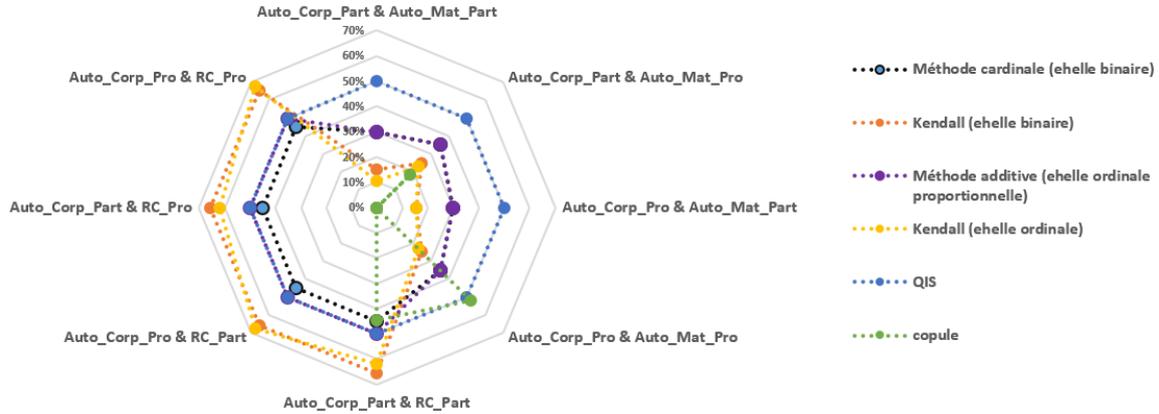


FIGURE 8.3 – Comparaison des corrélations (1)

Les corrélations estimées avec le τ_b de Kendall, pour les couples Auto-Corp-Part & RC-Part, Auto-Corp-Pro & RC-Part, Auto-Corp-Pro & RC-Pro et Auto-Corp-Part & RC-Pro sont très élevées et sont même supérieures aux corrélations QIS correspondantes. De plus, ces corrélations sont très proches; en effet, le τ_b de Kendall n'arrive pas à discriminer les paires de branches de garanties qui couvrent la même garantie mais qui diffèrent selon le statut de l'assuré (Particuliers vs Professionnels). Pour ces paires, les corrélations obtenues avec le τ_b de Kendall correspondent environ au 3/2 de celles obtenues avec les méthodes cardinale et additive, qui par contre sont très proches de celles de la matrice QIS.

Les corrélations estimées avec la méthode des copules, pour les couples Auto-Corp-Part & RC-Part et Auto-Corp-Pro & Auto-Mat-Pro sont cohérentes avec celles de la matrice QIS; ceci n'est pas le cas de la corrélation de la paire Auto-Corp-Part & Auto-Mat-Pro.

Pour toutes les paires de branches de garanties représentées sur le graphique ci-dessus, parmi les méthodes utilisées pour l'approche par facteurs de causalité, les corrélations obtenues avec la méthode additive sont celles qui se rapprochent le plus proches des corrélations de la matrice QIS.

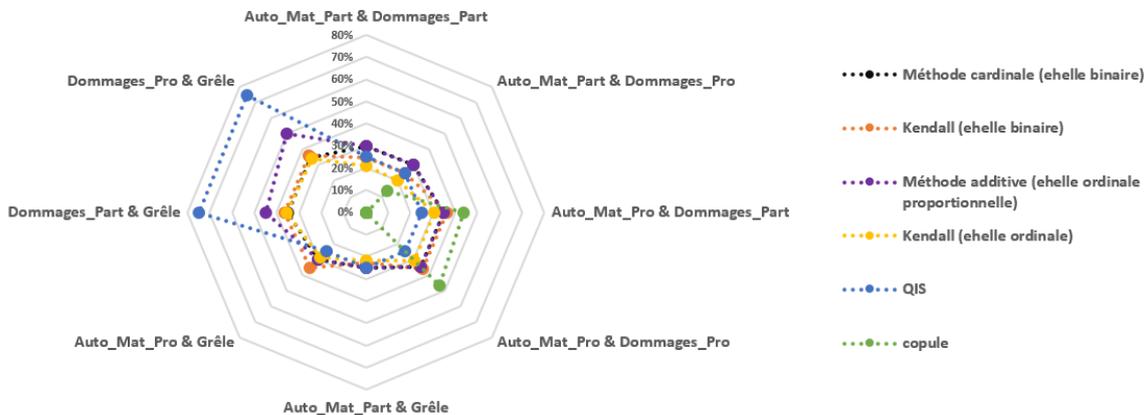


FIGURE 8.4 – Comparaison des corrélations (2)

Pour les paires représentées sur le graphique ci-dessus, les corrélations obtenues avec la méthode des copules ne sont pas proches de celles de la matrice QIS.

Pour ce qui concerne les autres méthodes, les corrélations obtenues pour les paires Auto-Mat-Part & Dommages-Part, Auto-Mat-Part & Dommages-Pro, Auto-Mat-Pro & Dommages-Part, Auto-Mat-Pro & Dommages-Pro, Auto-Mat-Part & Grêle et Auto-Mat-Pro & Grêle sont assez proches de celles de la matrice QIS. Ceci montre que pour ces paires, la corrélation de 25% que donne la matrice QIS peut-être cohérente bien qu'elle soit moins prudente que la méthode additive.

Les garanties Dommages-Part, Dommages-Pro et Grêle ne sont pas différenciées par la formule standard et appartiennent toutes à la classe « incendie » ; ceci justifie le fait que les corrélations QIS associées soient très élevées. Pour les paires Dommages-Part & Grêle et Dommages-Pro & Grêle, la méthode additive est plus prudente que les autres méthodes et est celle qui se rapproche le plus des corrélations QIS.

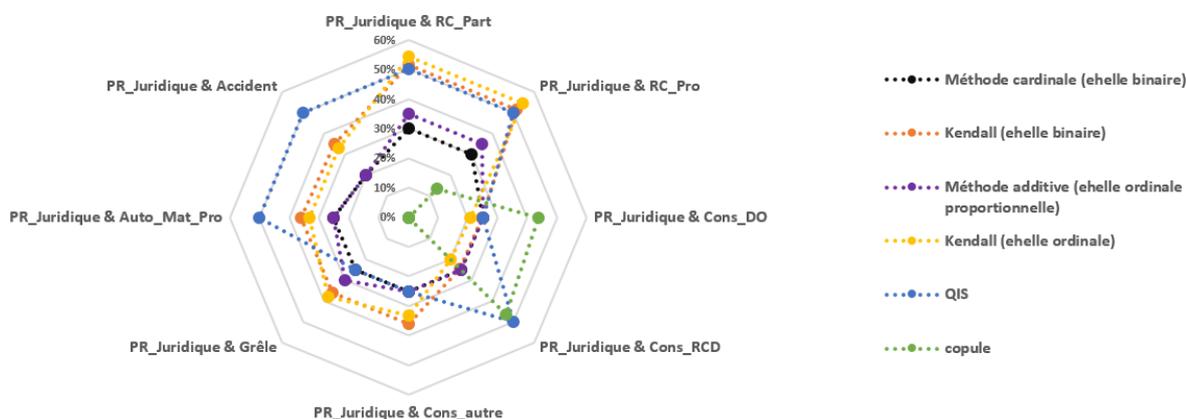


FIGURE 8.5 – Comparaison des corrélations (3)

Au regard du graphique ci-dessus, il ressort que les corrélations calculées avec les différentes méthodes sont très disparates. La corrélation de la paire Pr-Juridique & Cons-DO estimée avec la méthode des copules est très élevée par rapport à celle préconisée par la matrice QIS et également à celles estimées par les autres méthodes. Pour cette paire, les corrélations obtenues par la méthode des facteurs de causalité sont très proches de celles de la matrice QIS.

Pour ce qui concerne la paire Pr-Juridique & Cons-RCD par contre, la corrélation estimée par la méthode des copules est proche de celle de la matrice QIS, mais est très élevée par rapport à celles estimées par les autres méthodes.

Pour toutes les méthodes, les corrélations des paires Pr-Juridique & Auto-Mat-Pro et Pr-Juridique & Accident sont inférieures à celles de la matrice QIS. Ceci signifie que la corrélation de 50% que préconise l'EIOPA pour ces paires de branches de garanties est sur-estimée pour notre portefeuille.

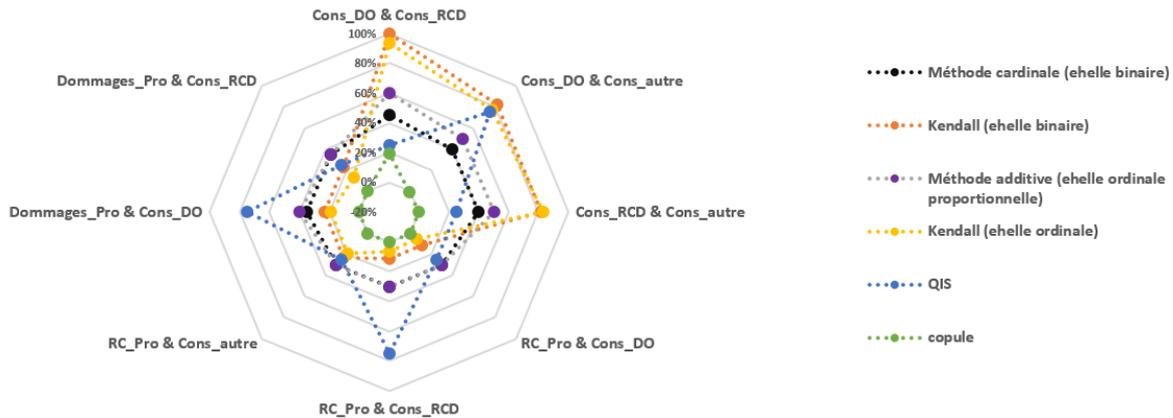


FIGURE 8.6 – Comparaison des corrélations (4)

Les branches de garanties RC-Pro et Cons-RCD ne sont pas différenciées par la formule standard et appartiennent toutes à la classe « Responsabilité civile » de la formule standard ; ceci justifie le fait que la corrélation QIS associée soit très élevée. C’est également le cas des garanties Dommages-Pro, Cons-DO et Cons-autre qui dans la formule standard, appartiennent à la classe « Incendie ».

Que ce soit avec une échelle binaire ou avec une échelle ordinale, le τ_b de Kendall ne parvient pas à discriminer les garanties construction ; en effet, les corrélations obtenues sont plus de 3 fois plus élevées que celles préconisées par la matrice QIS et plus de 1,5 fois plus élevées que celles obtenues avec les méthodes cardinale et additive.

Pour ce qui concerne les paires Dommages-Pro & Cons-RCD, RC-Pro & Cons-DO et RC-Pro & Cons-autre, la méthode additive aboutit à des corrélations plus élevées que celles obtenues avec les autres méthodes.

8.2 Avantages et inconvénients des différentes méthodes

Il est nécessaire d’analyser les avantages et les inconvénients des différentes méthodes utilisées pour calibrer les corrélations afin de faire une balance entre ces différentes méthodes et ainsi de pouvoir faire le choix de la corrélation la plus appropriée à chaque couple de branches de garanties.

8.2.1 Matrice QIS préconisée par l’EIOPA

La matrice de corrélation QIS a été estimée sur la base des données du marché des assureurs allemands entre les années 1998 et 2002 ; il s’agit en fait des corrélations entre les ratios de pertes des différentes branches.

Avantages :

- (1) Le calcul du SCR se fait avec des corrélations fixes et par une formule fermée explicitée en page 16.
- (2) Le fait d’utiliser la matrice QIS, pourrait faciliter la comparaison entre différents profils de risques.

Inconvénients :

- (1) Les données utilisées pour calibrer la matrice QIS peuvent ne pas refléter le profil de risque de la compagnie d'assurance étant donné que ces données ne couvrent que le marché des assureurs allemands. De plus, rien ne garantit que, si on reproduit l'étude sur des données récentes, on retrouvera les corrélations de la matrice QIS.
- (2) La matrice de corrélation QIS n'est pas tout à fait adaptée au portefeuille d'AVIVA Assurances. En effet, plusieurs branches de garanties de notre segmentation appartiennent à une seule branche de la segmentation QIS ; c'est le cas des garanties construction responsabilité civile décennale, responsabilité civile pour les particuliers et responsabilité civile pour les professionnels qui appartiennent toutes à la classe « Responsabilité civile » de la formule standard. Pourtant, la garantie construction responsabilité civile décennale a un profil de risque bien particulier. De plus la formule standard ne fait pas de distinction entre particuliers et professionnels.

8.2.2 La méthode des copules**Avantage :**

La méthode des copules permet de modéliser des structures de dépendances autres que linéaires (symétriques). En effet, elle permet de prendre en compte les dépendances de queue si celles-ci existent. Par exemple, les dépendances de forte intensité à droite c'est-à-dire dans les quantiles élevés peuvent être modélisées avec la copule de Gumbel.

Inconvénients :

- (1) Le calibrage des copules est intéressante quand l'historique de données à disposition est bien profond. Cependant, disposant de 19 années pour notre étude, nous avons dû faire recours à la technique du Bootstrap appliquée sur les résidus de Pearson du modèle de Chain-Ladder. Cependant, le choc bicentenaire n'est pas forcément pris en compte lors du ré-échantillonnage. En effet, les résidus de Chain-Ladder sont calculés sur un historique de 19 ans sur lequel la déviation de la sinistralité future n'a pas forcément atteint le choc bicentenaire.
- (2) Le calibrage des corrélations par la méthode des copules est très coûteux en temps. Dans cette étude par exemple, nous avons 105 couples de branches de garanties et donc 105 corrélations à calibrer. Nous avons dû restreindre l'étude des copules aux couples de branches de garanties pour lesquels le tau de Kendall est significatif.

8.2.3 Approche par facteurs de causalité**Avantages :**

- (1) Cette approche peut être utilisée pour calibrer la corrélation entre tout type de risques, également lorsque les données à disposition sont peu fiables ou ont une profondeur assez faible pour refléter la structure de corrélation qui existe entre les risques.
- (2) Elle repose moins sur des outils statistiques.
- (3) Elle est facile à implémenter.
- (4) Cette approche donne une autre vision de la corrélation. En effet, elle interprète la dépendance entre deux branches de garanties comme le fait que ces branches de garanties partagent les mêmes risques (facteurs de causalité) et réagissent de façon similaire quant à la survenance de ces facteurs de causalité.

Inconvénients :

- (1) Les résultats peuvent être erronés si le panel d'experts interrogé a une mauvaise compréhension des facteurs de causalité.
- (2) Cette approche est essentiellement basée sur des jugements d'experts ; par conséquent, elle peut être sujette à beaucoup de subjectivité.

8.3 Choix de la matrice de corrélation idéale

Nous avons montré à la section précédente, que la matrice QIS n'est pas nécessairement adaptée au portefeuille d'AVIVA Assurances et que les deux approches (quantitative et qualitative) utilisées pour calibrer les corrélations avaient chacune des avantages et inconvénients. Ces deux approches seront alors évaluées non seulement sur la base de la cohérence de leurs résultats mais aussi sur la base de la subjectivité de ceux-ci.

Les corrélations obtenues avec la méthode des copules sont pour la plus part très faibles, même sur certaines branches de garanties pour lesquelles on se serait attendu à avoir des corrélations relativement élevées. C'est le cas par exemple des paires Auto-Corp-Part & Dommages-Part (14%), Dommages-Part & Dommages-Pro (15%) et RC-Pro & Cons-RCD (0%). Ces résultats laissent croire que pour certaines branches de garanties, le re-échantillonnage utilisé dans la méthode des copules ne retranscrit pas toujours assez bien les structures de corrélation entre les branches de garanties. De plus, cette méthode fournit les résultats des corrélations seulement pour les couples de branches de garanties dont le Tau de Kendall était significatif. Pour les autres couples, la corrélation n'a pas pu être estimée avec cette méthode. Seules 31 sur 105 paires de corrélations ont pu être estimées par cette méthode.

Pour ce qui concerne l'approche par facteur de causalité, contrairement aux méthodes cardinale et additive, le τ_b de Kendall qui a l'avantage de pouvoir s'appliquer aussi bien à l'échelle binaire qu'à l'échelle ordinale et donc de prendre en compte l'ampleur de l'influence des facteurs de causalité sur les branches, est très instable. En effet, nous avons réalisé les sensibilités suivantes sur la paire Accident & Auto-Corp-Part :

Méthode	Cardinale	additive	Tau-b (échelle 01)	Tau-b (échelle 0123)
Scénario central	30%	30%	29%	30%
Ajout d'une paire (0,0)	29%	30%	31	30%
Ajout d'une paire (0,1)	29%	30%	22%	23%
Ajout d'une paire (1,0)	29%	30%	18%	18%
Ajout d'une paire (1,1)	33%	35%	31%	27%
Ajout d'une paire (2,0)	29%	30%	18%	17%
Ajout d'une paire (0,2)	29%	30%	22%	17%
Ajout d'une paire (2,2)	33%	40%	31%	32%
CD 20 (1,0)	30%	30%	15%	15%
CD 20 (2,0)	30%	30%	15%	14%

TABLE 8.2 – Sensibilité du Tau-b de Kendall

On peut constater que le Tau-b est vraiment très instable; effet, si aux 20 facteurs de causalité retenus, on ajoute un facteur causalité qui a un impact sur la branche Accident mais qui n'a pas d'impact sur la branche Auto-Corp-Part, alors le τ_b de Kendall de cette paire passe de 29% à 18%. Si l'impact du facteur de causalité « impact de la compétition du marché » devient (1,0) sur la paire Accident & Auto-Corp-Part, alors le τ_b de Kendall baisse quasiment de moitié.

Il ressort alors que le τ_b de Kendall est la méthode la plus subjective, la plus sensible au nombre de facteurs de causalité et par conséquent la plus instable. Pour ces raisons, nous n'allons pas considérer les résultats obtenus avec cette métrique.

La méthode cardinale quant à elle, dépend non seulement du nombre de facteurs de causalité qui ont un impact sur les branches de garanties, mais aussi du nombre de facteurs de causalité retenus pour l'étude. La méthode additive par contre est moins sensible au nombre de facteurs de causalité retenus, mais dépend beaucoup plus de l'ampleur attribué aux impacts des facteurs de causalité.

Pour le choix de la matrice de corrélation, nous avons adopté la démarche suivante :

(1) : Pour chaque paire de branches de garanties, si l'une des corrélations obtenues avec les méthodes additive et cardinale est proche (écart au plus égal à 5%) de la corrélation préconisée par la matrice QIS, alors, on gardera la corrélation QIS. Sinon, on calcule la moyenne des corrélations des deux méthodes pour la paire considérée.

(2) : Si la corrélation moyenne calculée à l'étape 1 est proche (écart au plus égal à 5%) de la corrélation préconisée par la matrice QIS, alors, on gardera la corrélation QIS.

(3) : Si la corrélation moyenne calculée à l'étape 1 n'est pas proche (écart supérieur à 5%) de la corrélation préconisée par la matrice QIS, alors si la corrélation obtenue par la méthode des copules est proche (écart supérieur à 5%) de celle préconisée par la matrice de corrélation QIS, alors, on gardera la corrélation QIS.

(4) : Sinon, on gardera la corrélation obtenue avec la méthode additive car elle est plus prudente que la méthode cardinale.

La matrice des corrélations obtenue pour le risque de sinistralité future est la suivante :

	Accident	Auto_Corp_Part	Auto_Corp_Pro	Auto_Mat_Part	Auto_Mat_Pro	Dommages_Part	Divers	Dommages_Pro	PR_Juridique	RC_Part	RC_Pro	Grêle	Cons_DO	Cons_RCD	Cons_autre
Accident	100%	30%	30%	20%	25%	25%	0%	25%	20%	25%	25%	50%	30%	30%	25%
Auto_Corp_Part	30%	100%	60%	30%	35%	45%	0%	45%	35%	50%	50%	25%	45%	40%	35%
Auto_Corp_Pro	30%	60%	100%	30%	50%	40%	0%	40%	35%	50%	50%	25%	40%	40%	35%
Auto_Mat_Part	20%	30%	30%	100%	35%	25%	0%	25%	20%	25%	25%	25%	25%	25%	25%
Auto_Mat_Pro	25%	35%	50%	35%	100%	35%	0%	35%	25%	25%	25%	25%	25%	25%	25%
Dommages_Part	25%	45%	40%	25%	35%	100%	0%	60%	25%	35%	35%	45%	40%	35%	35%
Divers	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
Dommages_Pro	25%	45%	40%	25%	35%	60%	0%	100%	25%	35%	35%	50%	40%	35%	35%
PR_Juridique	20%	35%	35%	20%	25%	25%	0%	25%	100%	35%	35%	25%	25%	25%	25%
RC_Part	25%	50%	50%	25%	25%	35%	0%	35%	35%	100%	50%	25%	25%	30%	25%
RC_Pro	25%	50%	50%	25%	25%	35%	0%	35%	35%	50%	100%	25%	25%	30%	25%
Grêle	50%	25%	25%	25%	25%	45%	0%	50%	25%	25%	25%	100%	40%	25%	35%
Cons_DO	30%	45%	40%	25%	25%	40%	0%	40%	25%	25%	25%	40%	100%	60%	50%
Cons_RCD	30%	40%	40%	25%	25%	35%	0%	35%	25%	30%	30%	25%	60%	100%	50%
Cons_autre	25%	35%	35%	25%	25%	35%	0%	35%	25%	25%	25%	35%	50%	50%	100%

FIGURE 8.7 – Matrice de corrélation retenue

Les paires en bleu clair sont celles pour lesquelles les corrélations obtenues avec au moins l'une des trois méthodes (copule, additive, cardinale) correspondent exactement à celles préconisées par la matrice QIS. Les paires en vert clair sont celles pour lesquelles les corrélations obtenues avec au moins l'une des trois méthodes sont proches (écart au plus égal à 5%) de celles préconisées par la matrice QIS. Les corrélations en rose ou en bordeaux sont celles pour lesquelles les corrélations retenues ne sont pas proches de celles préconisées par la matrice QIS.

De manière générale, il ressort que :

- Si les branches de garanties couvrent les mêmes garanties (par exemple la responsabilité civile générale) mais se distinguent par le statut de l'assuré (particulier/professionnel), les corrélations sont relativement élevées.
- Si les branches de garanties couvrent des garanties similaires mais différent selon le périmètre, les corrélations sont un peu plus faibles. C'est le cas des branches de garanties responsabilité civile automobile et responsabilité civile générale par exemple, également celui des garanties Dommages-Pro et Grêle.
- Lorsque deux branches de garanties ne couvrent pas des garanties similaires et différent le périmètre, les corrélations peuvent être relativement plus faibles.

Avec cette matrice, nous avons obtenu un SCR total de 244 M€ au titre de la sinistralité future, contre un SCR total de 270 m€ obtenu avec la matrice de corrélation QIS (259 m€ avec la matrice de corrélation QIS bis). Cette mise à jour de la matrice de corrélation a donc conduit à une bénéfice de diversification de 26 m€ (15 m€ si matrice QIS bis) soit de l'ordre de 11% (6% si matrice QIS bis).

Limites de l'étude

Aucune étude ne saurait être parfaite, nous avons identifié des limites relatives à notre étude :

(1) : Nous n'avons pas pris en compte les LOUB (le risques sur les contrats qui se renouvellent par tacite reconduction), la sinistralité au titre des primes émises non acquises et la déviation persistante de la sinistralité (la corrélation entre les années d'accident 1 et 2)

(2) : La Provision pour sinistre non encore manifesté n'a pas été prise en compte dans l'estimation de la charge ultime des sinistres des garanties construction décennales. Ce qui pourrait conduire à une sous-estimation de la charge ultime de ces branches de garanties et par conséquent du SCR

(3) : Le SCR a été calculé sur une vision ultime du risque.

Conclusion

Il était question pour nous dans ce travail de calibrer le risque de sinistralité future dans le cadre de l'élaboration d'un modèle interne pour AVIVA Assurances, ainsi que la structure de corrélation entre le risque de sinistralité future des branches de garanties.

Dans un premier temps, il était question de mettre en as-if les données historiques de sinistres afin de les rendre comparables entre elles et de prendre en compte la valeur temps de l'argent. Ces données ont par la suite été mises à l'ultime pour avoir la valeur la plus juste du risque. Ces retraitements ont abouti à une hausse de la charge totale de sinistre de 14,52%. Si ces retraitements n'avaient pas été effectués, il y aurait eu une sous-estimation de la charge de sinistre et donc une sous-estimation du risque.

Une étude du seuil qui sépare les sinistres attritionnels des sinistres graves a été effectuée sur chaque branche de garanties. Il en est ressorti qu'il n'est pas pertinent d'utiliser un seuil identique pour toutes les branches de garanties. Pour certaines branches de garanties, l'étude a abouti à la conclusion selon laquelle il n'est pas pertinent de distinguer la sinistralité attritionnelle de la sinistralité grave.

Nous avons par la suite modélisé la sinistralité. Pour la sinistralité attritionnelle, un modèle agrégé de charge ultime en unité d'exposition (Charge ultime/exposition) a été utilisé; pour ce qui concerne la sinistralité grave, une approche fréquence*sévérité a été retenue. Cette modélisation nous a permis de calibrer sur chaque modèle une loi de distribution, un coefficient de variation et un SCR; ceci à la granularité segment de risque x type de sinistralité.

Ne pouvant tout simplement additionner les SCR obtenus, nous avons combiné une approche quantitative de détermination des corrélations basée sur les copules et une approche qualitative basée sur des facteurs de causalité. Nous avons obtenu que la matrice de corrélation QIS préconisée par l'EIOPA n'était pas toujours adaptée au profil de risque de la société. La matrice de corrélation obtenue a conduit à un bénéfice de diversification de 26m€, soit de 11% par rapport à l'utilisation de la matrice QIS.

Bibliographie

- [1] Anaïs Belabed. Modélisation de la sinistralité atypique en RC automobile avec prise en compte des spécificités d'un versement en rente. Mémoire soutenu en 2017.
- [2] Anne Suchel. Etude des corrélations entre branches d'assurance IARD. Mémoire soutenu en 2015.
- [3] Anthony Derien. Solvabilité 2 : une réelle avancée? Hal archives ouvertes, septembre 2012.
- [4] Arthur Charpentier et Christophe Dutang. L'actuariat avec R, octobre 2012.
- [5] ACPR, Banque de France, NOTICE « Solvabilité II ». Calcul du SCR en formule standard, 17/12/2015.
- [6] CEIOPS-DOC-70/10. CEIOPS' Advice for level 2 Implementign Measures on Solvency II : SCR STANDARD FORMULA Article 111(d) Correlations. 2010.
- [7] David Cadoux et Jean-Marc Loizeau. Copules et dépendances : application pratique à la détermination du besoin en fonds propres d'un assureur non-vie.
- [8] Emilie Deprey et Alexandre Godzinski, Problématique de seuil dans la modélisation de la sinistralité en Réassurance Non Vie. Mémoire soutenu en 2007
- [9] Gwladys Marylou Noutong. Les Risk Drivers : modélisation de la dépendance entre branches d'activités d'une compagnie d'assurance Non-Vie. Mémoire soutenu en 2013.
- [10] Lluís Bermudeza, Antoni Ferri et Montserrat Guillen. A correlation sensitivity analysis for non-life underwriting risk module SCR. 2011.
- [11] Loiseau Audrey : Modélisation du capital requis pour le risque de souscription non-vie et étude de la dépendance. Mémoire soutenu en 2014.
- [12] RÈGLEMENT DÉLÉGUÉ (UE) 2015/35 DE LA COMMISSION. octobre 2014.
- [13] QIS5 technical Specifications.5 juillet 2010.
- [14] Romain LAILY Construction d'un modèle interne Partiel en Assurance non-vie. Mémoire soutenu en 2014.
- [15] Vincent Rouxel, Détermination du capital réglementaire de souscription non-vie sous Solvabilité II : pertinence d'une approche modèle interne. Mémoire soutenu en 2013.
- [16] Yu PEI, Modélisation de la sinistralité atypique en RC automobile avec prise en compte des spécificités d'un versement en rente. Mémoire soutenu en 2009.

Annexes

A1 : Etude de seuil

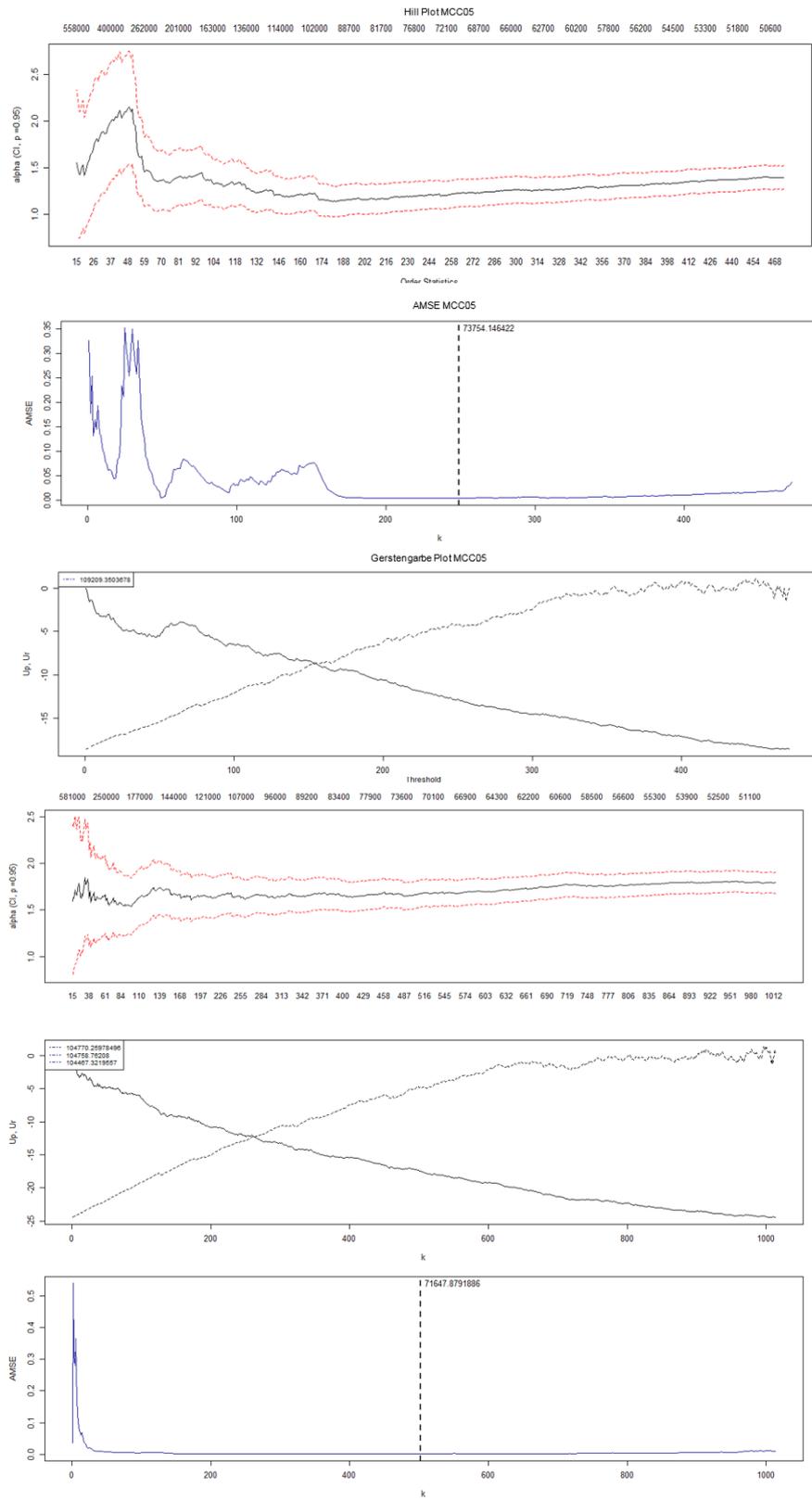


FIGURE 8.8 – MCC5 et MCC6

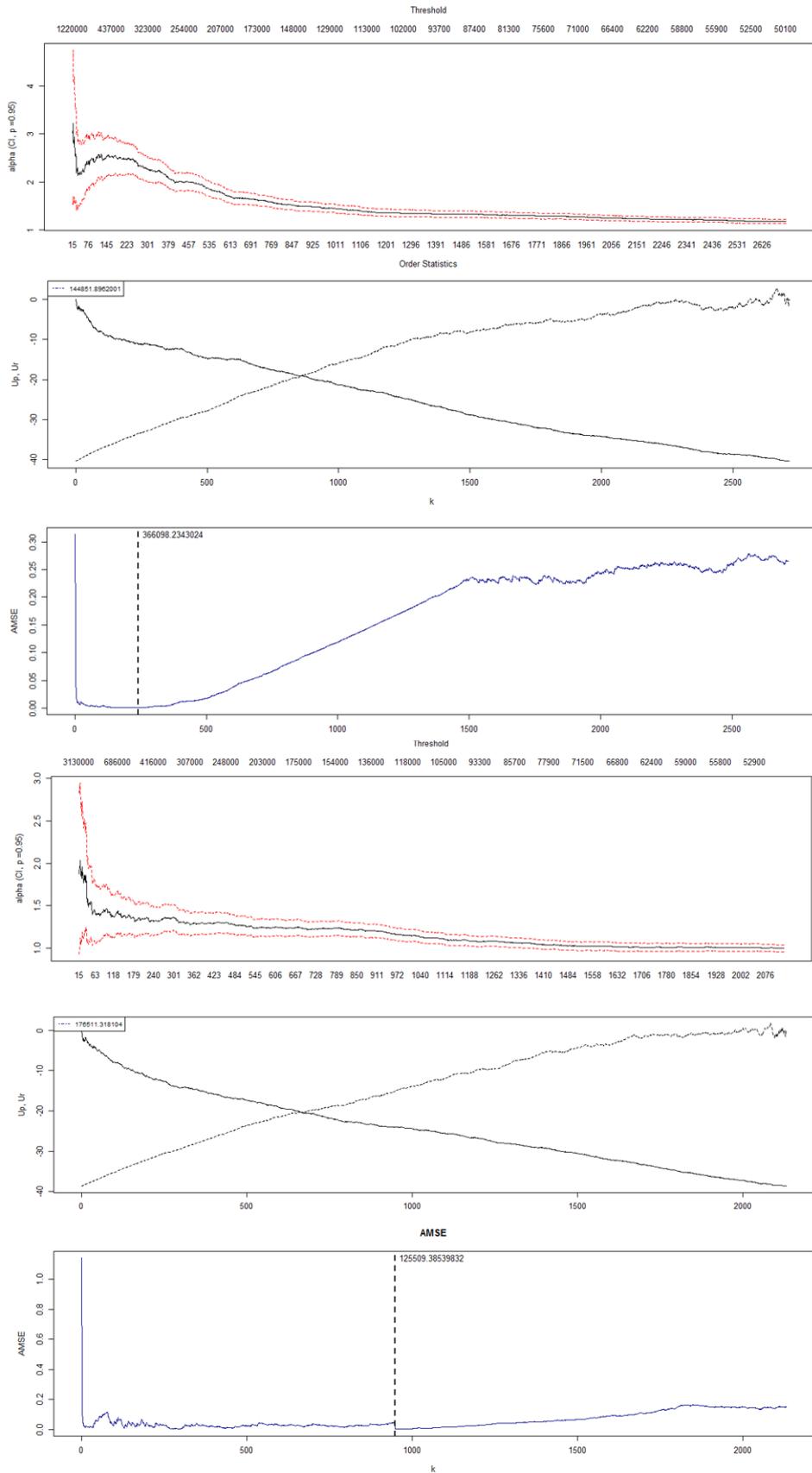
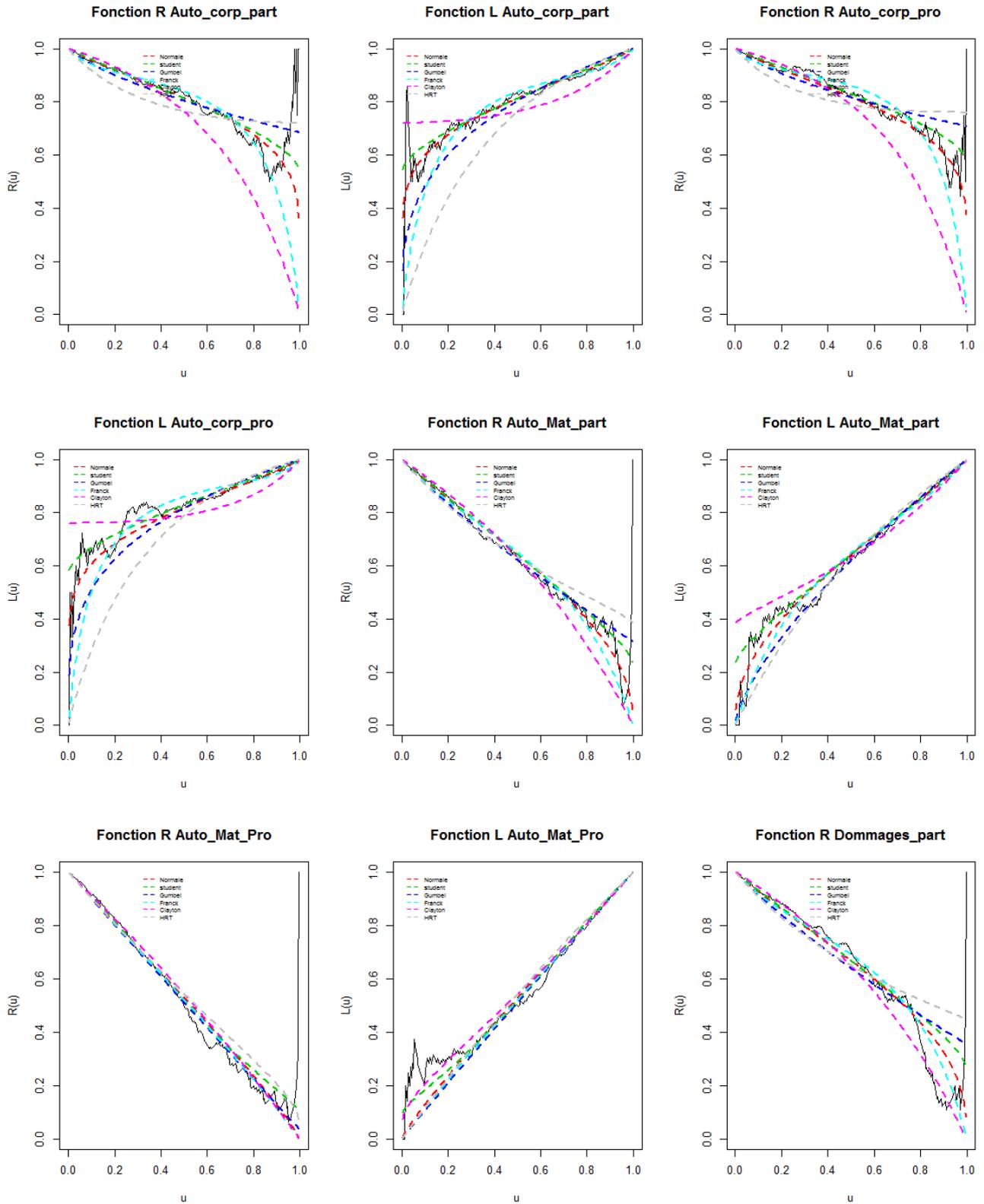
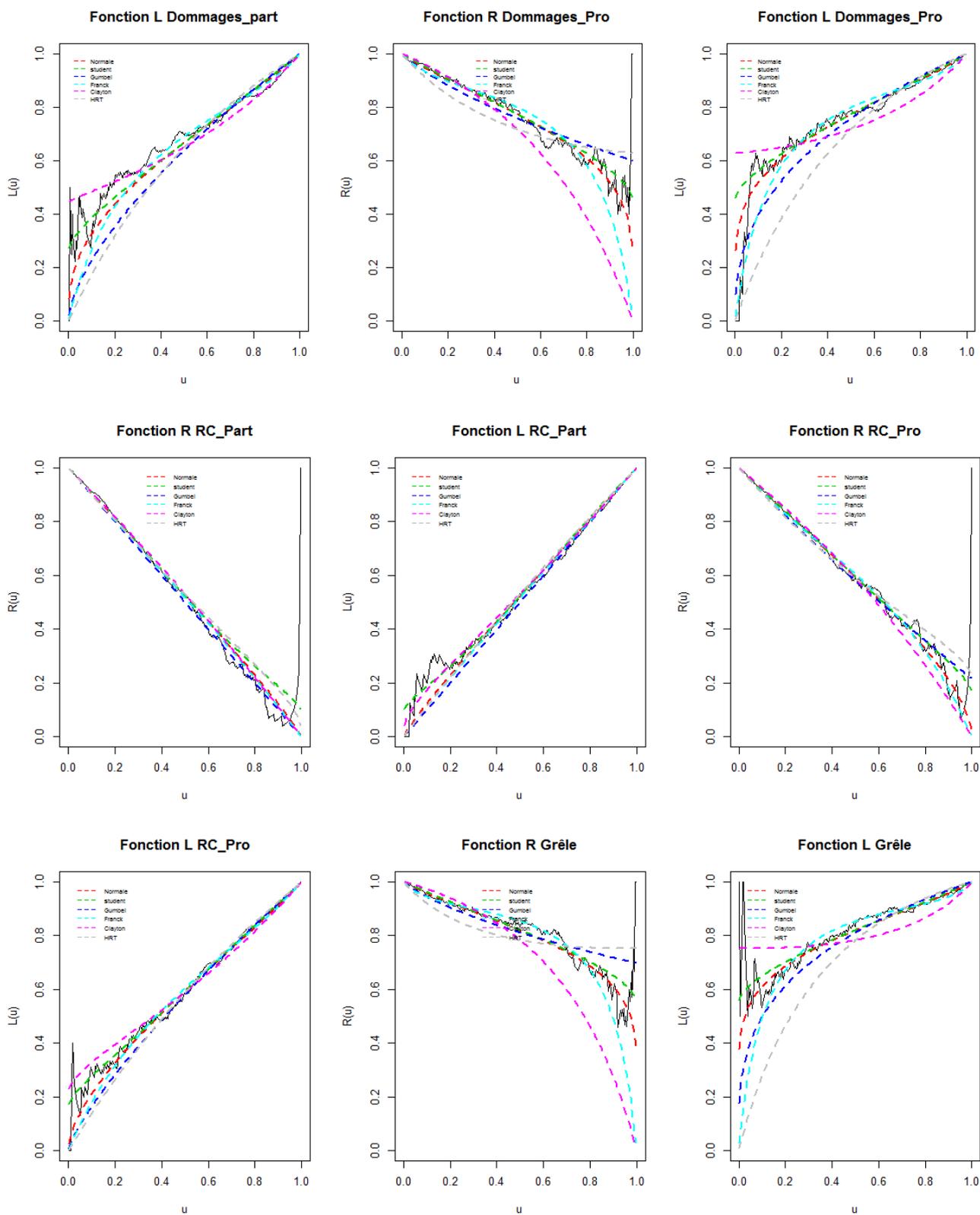


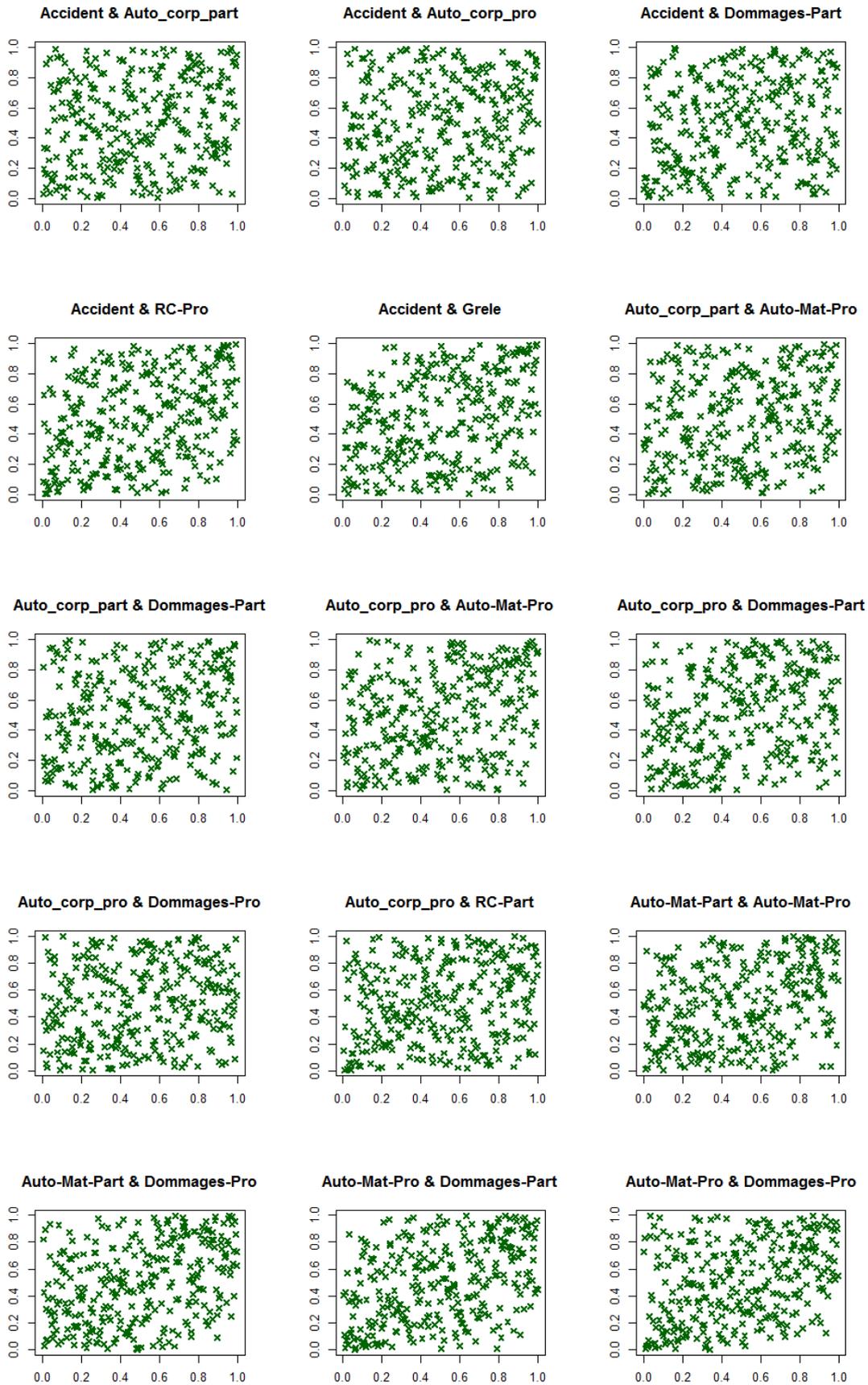
FIGURE 8.9 – MCC7 et MCC9

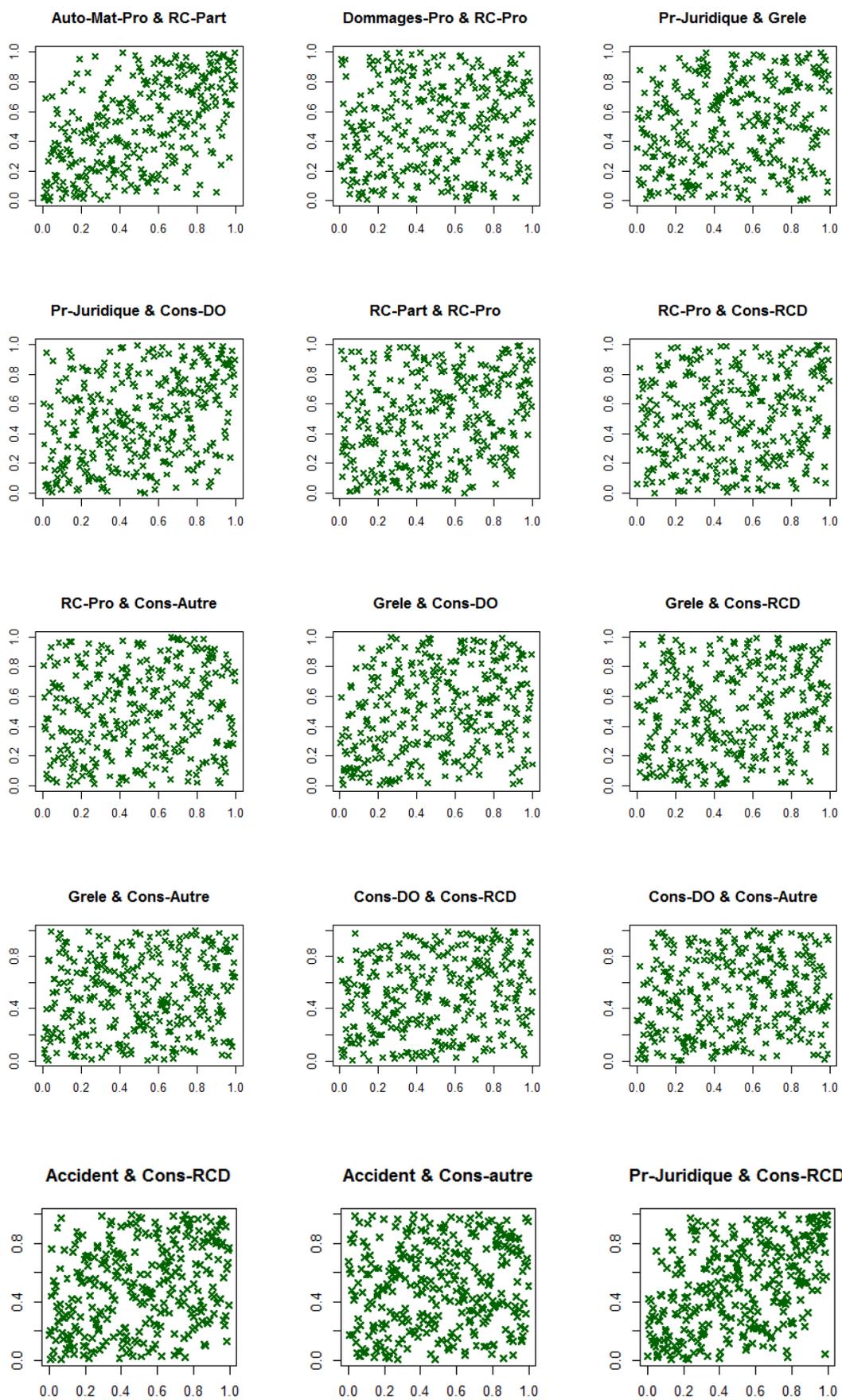
A2 : Dépendance entre sinistralités attritionnelle et grave : fonctions de concentration dans les queues



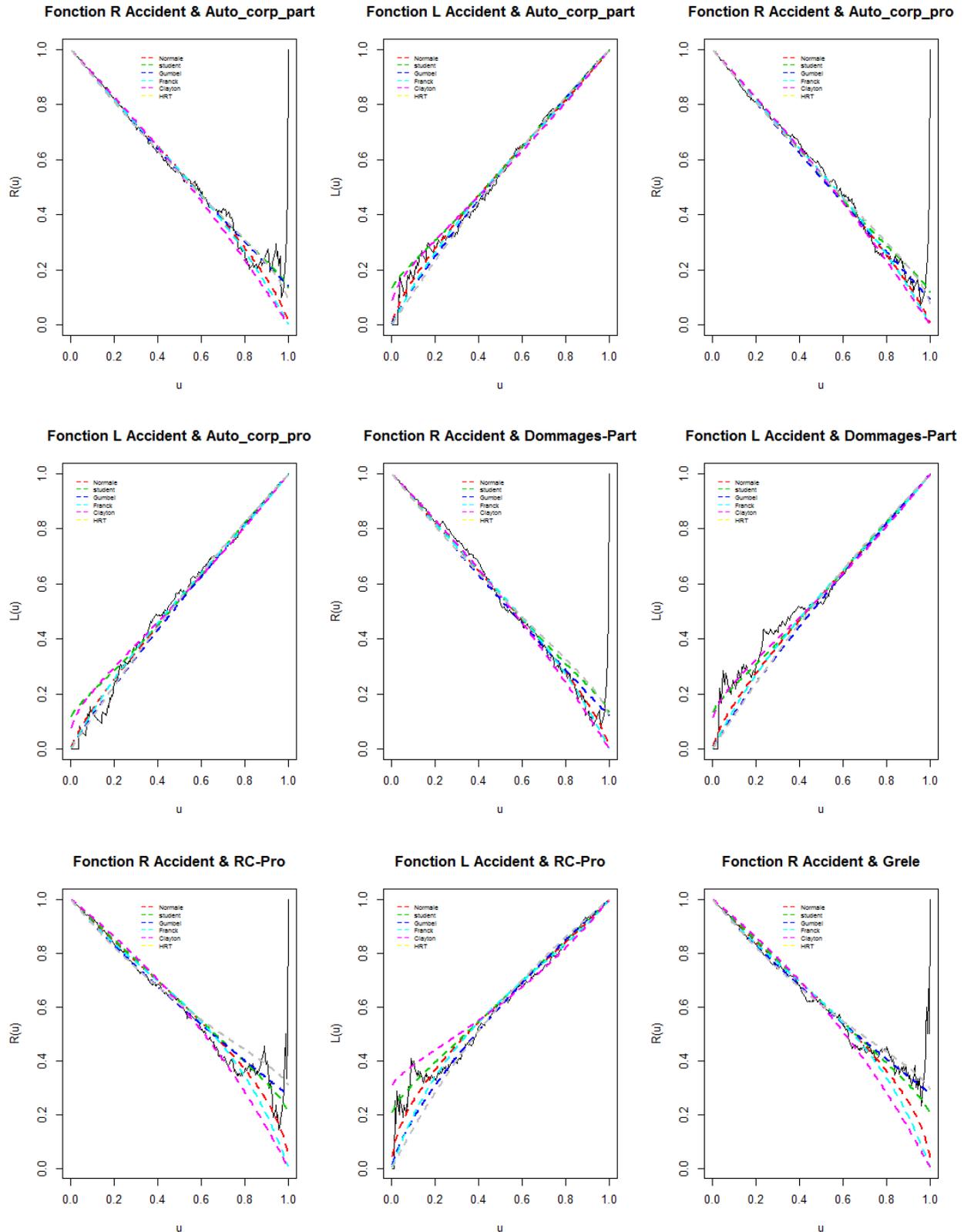


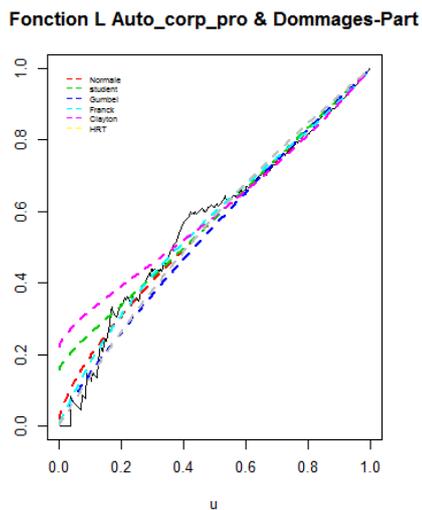
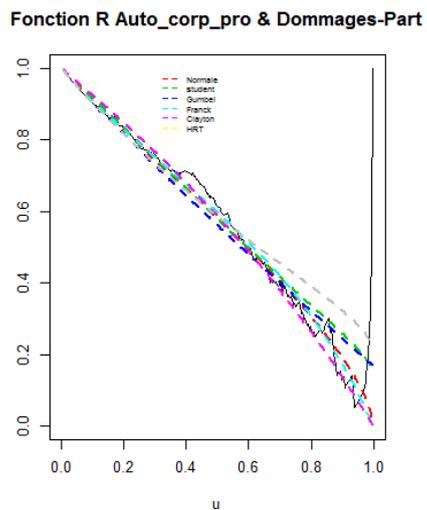
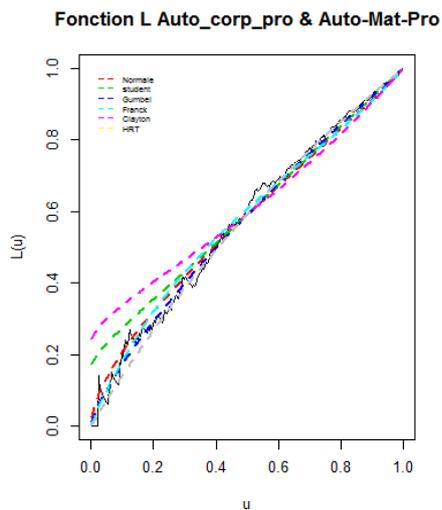
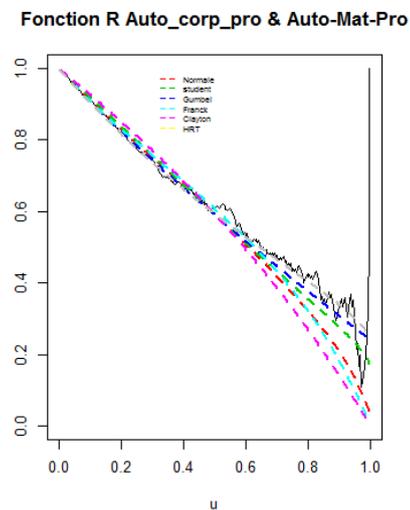
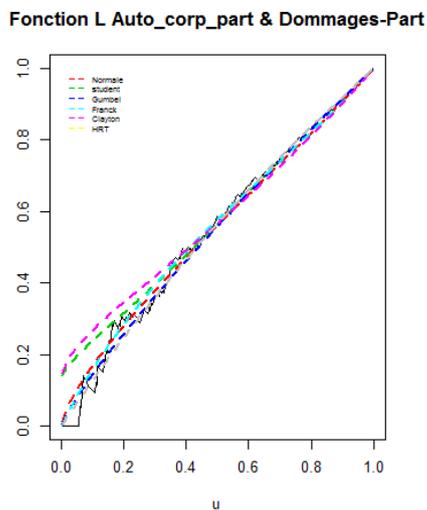
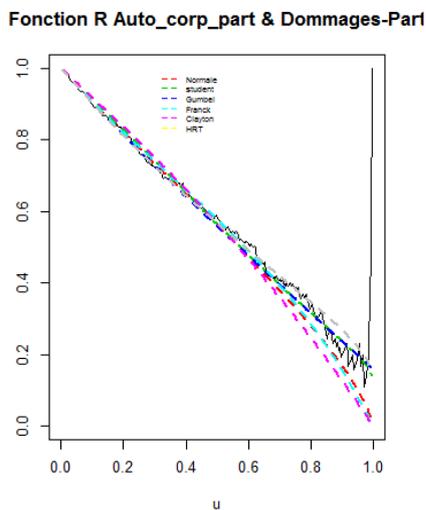
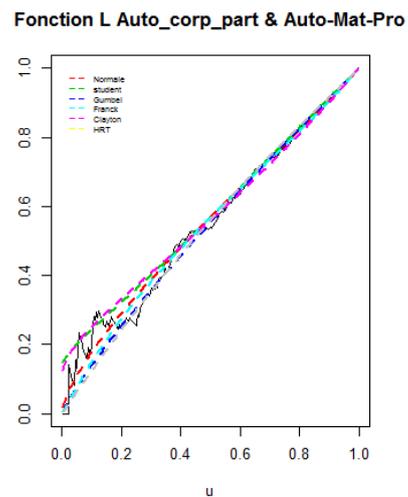
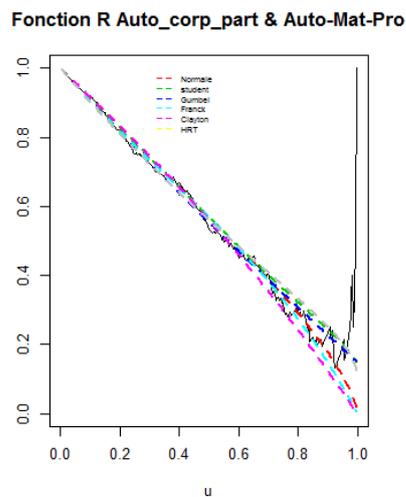
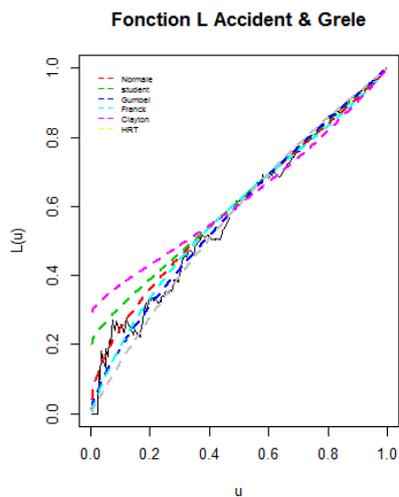
A3 : Dépendance entre MCC : Copules empiriques

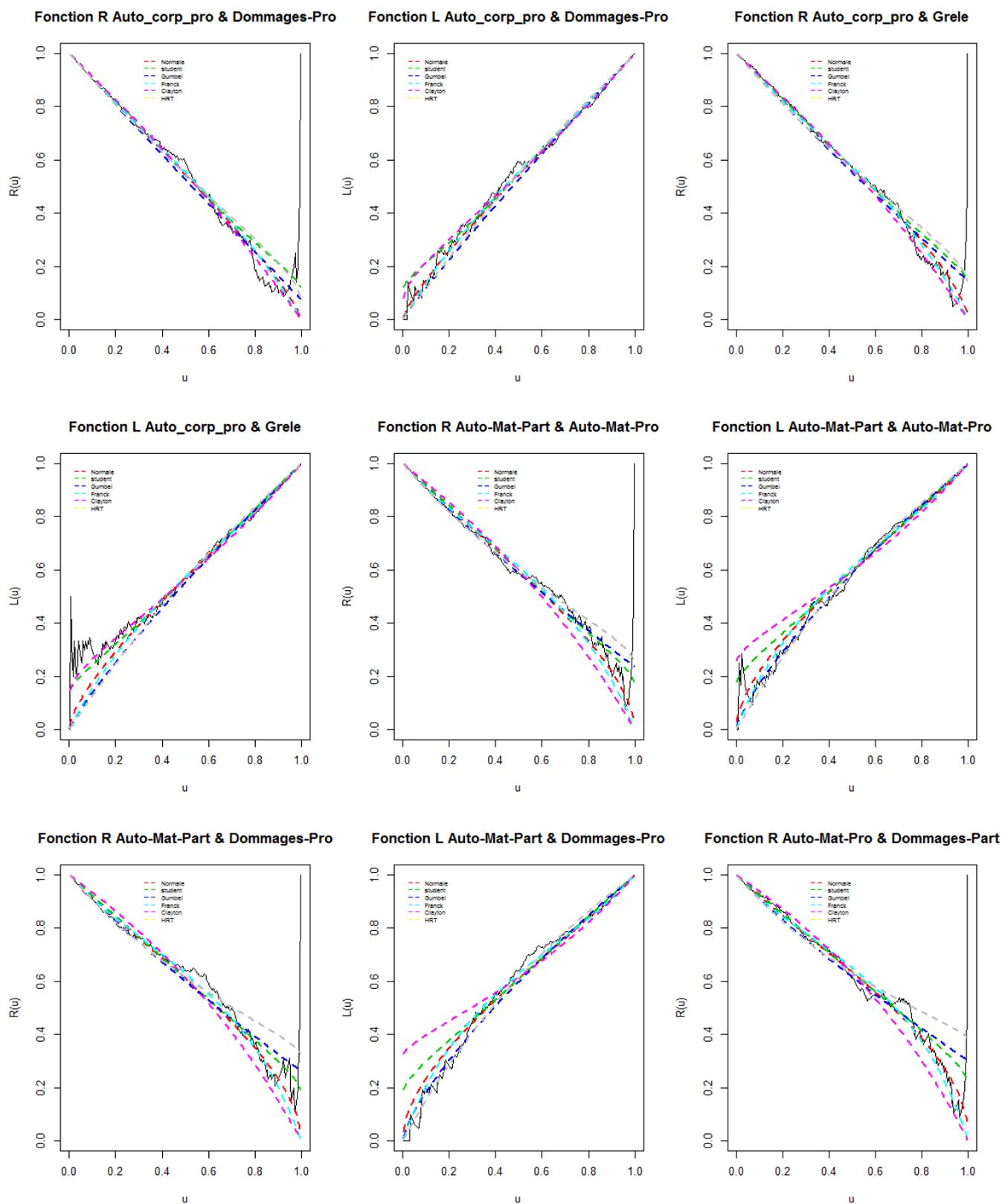


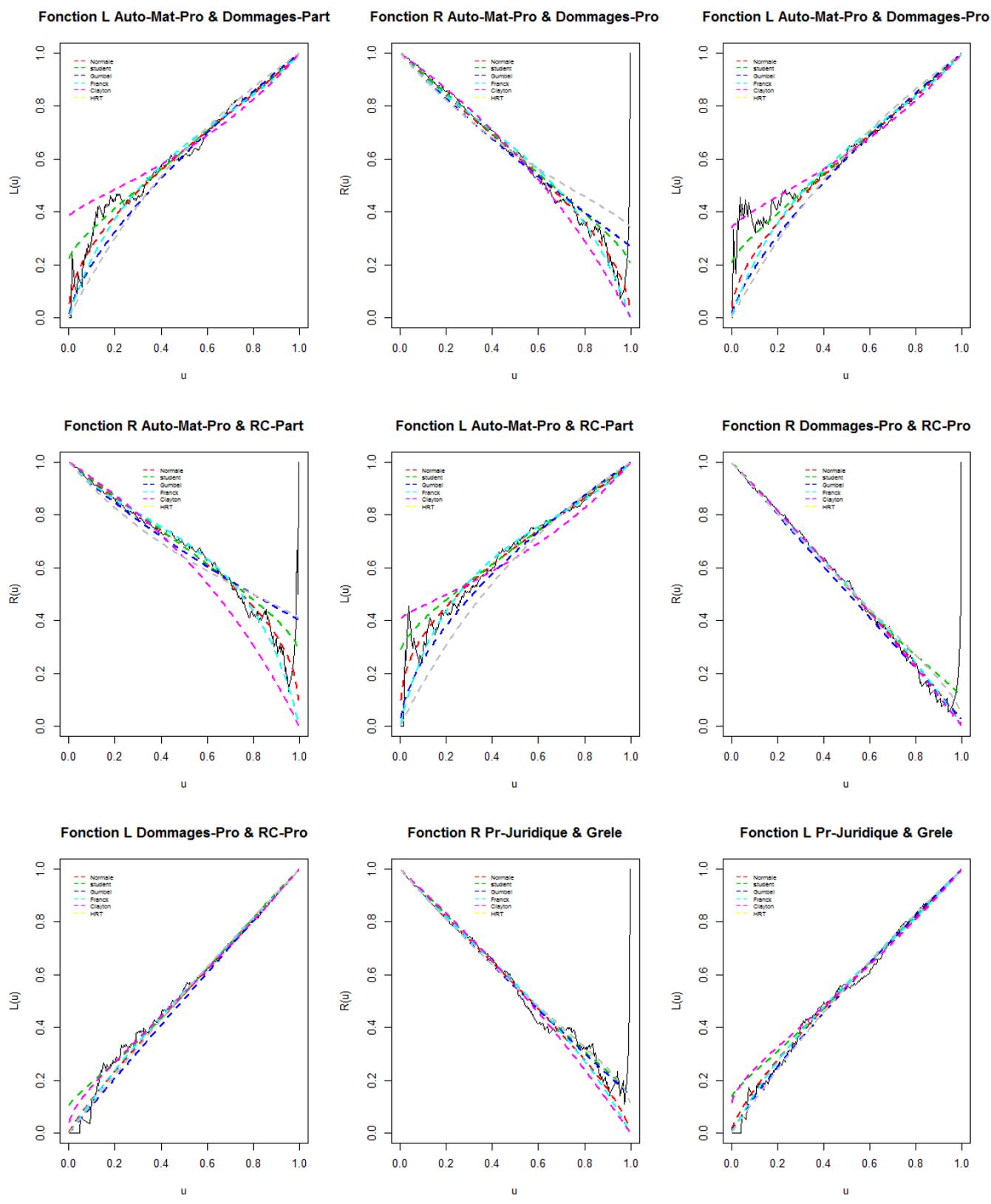


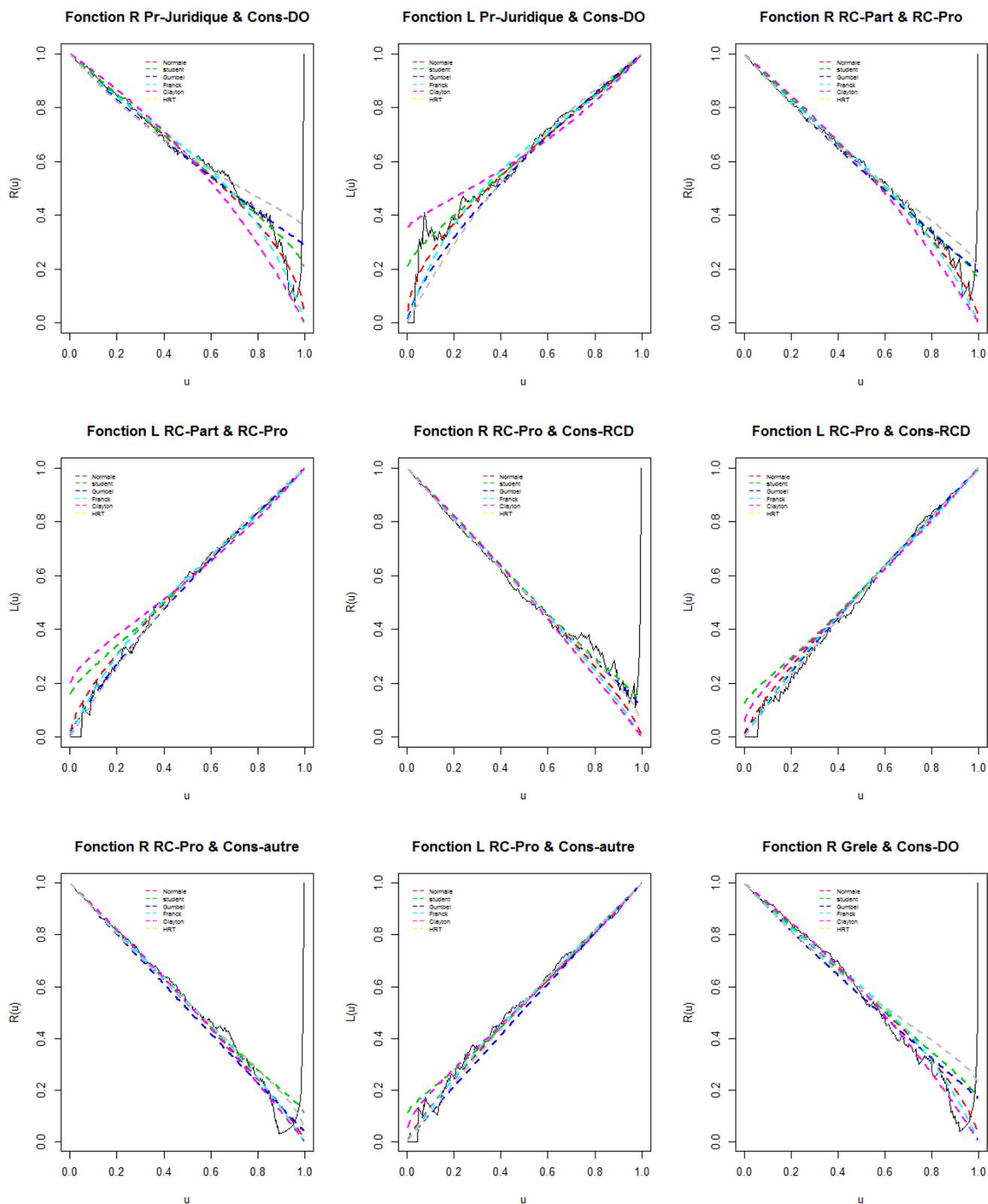
A4 : Dépendance entre MCC : fonctions de concentration dans les queues

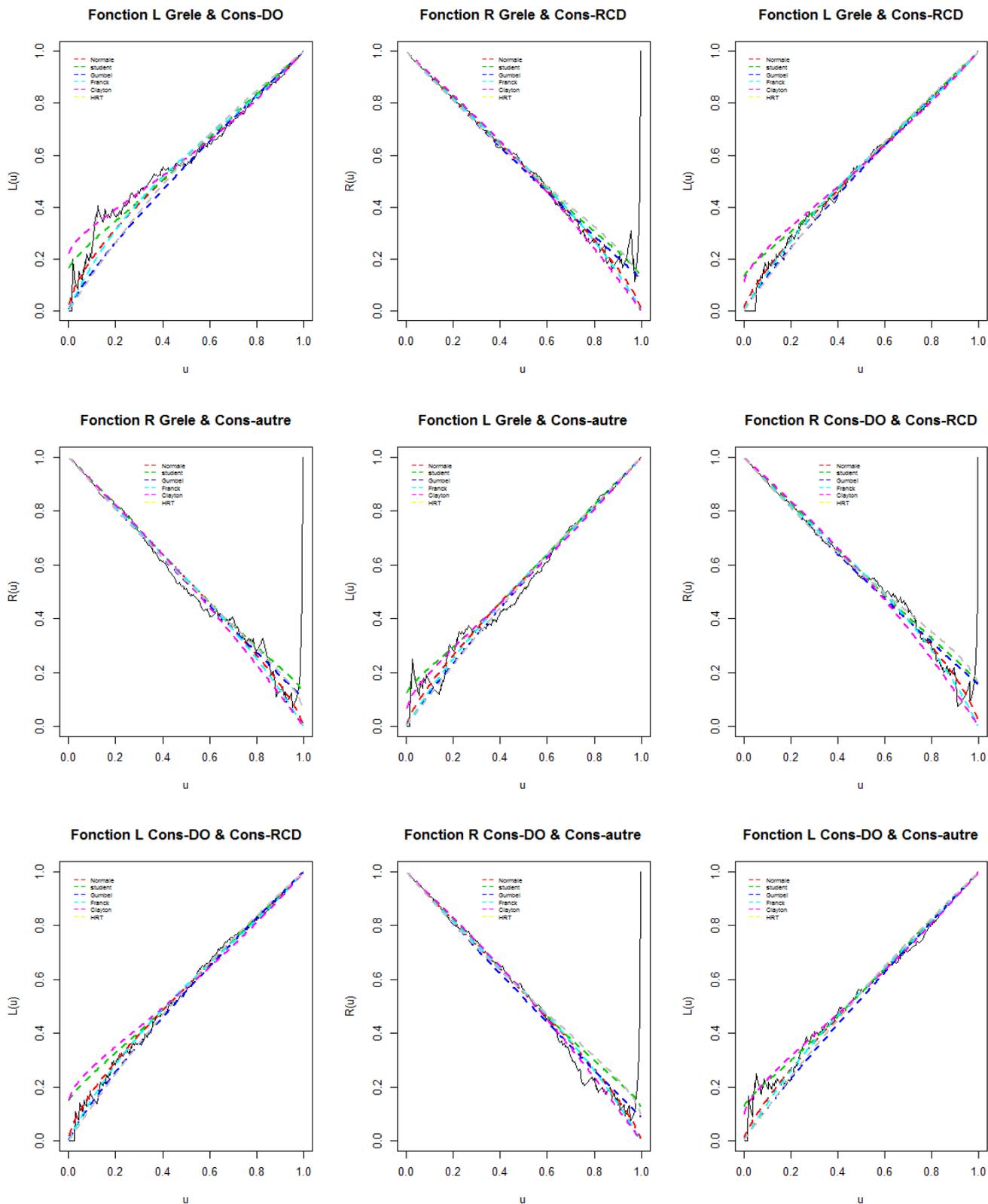












A5 : Approche par facteur de causalité

Facteurs de causalité	Accident	Auto_Corp_Part	Auto_Corp_Pro	Auto_Mat_Part	Auto_Mat_Pro	Dommmages_Part	Divers	Dommmages_Pro	Pr_Juridique	RC_Part	RC_Pro	Grêle	Cons_RCD	Cons_DO	Cons_Autre
Super inflation	5%	5%	5%	0%	0%	0%	0%	0%	0%	0%	0%	0%	10%	10%	0%
Augmentation de la richesse	0%	5%	5%	5%	5%	5%	0%	5%	0%	0%	0%	0%	5%	5%	5%
Changement dans la pratique légale	5%	10%	10%	5%	5%	5%	0%	5%	10%	10%	10%	15%	5%	5%	5%
Erreur contractuelle	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Evènements naturelles	0%	0%	0%	0%	0%	10%	0%	15%	0%	0%	0%	15%	0%	0%	0%
Vol et fraude interne	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Actions de l'homme	0%	0%	0%	0%	0%	5%	0%	5%	0%	0%	0%	0%	0%	0%	0%
Erreurs humaines	5%	5%	5%	5%	5%	5%	0%	5%	5%	5%	5%	5%	5%	5%	5%
Echec des contrôles	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Qualité des données	5%	10%	5%	0%	5%	10%	0%	10%	5%	5%	5%	10%	10%	5%	5%
Réseau de distribution	5%	5%	5%	5%	5%	5%	0%	5%	0%	5%	5%	5%	5%	5%	5%
Risque de longévité	0%	5%	5%	0%	0%	0%	0%	0%	0%	5%	5%	0%	0%	0%	0%
Tendances sociales et culturelles	0%	5%	5%	5%	5%	5%	0%	5%	5%	5%	5%	0%	0%	0%	0%
Avancées médicales	0%	5%	5%	0%	0%	0%	0%	0%	0%	5%	5%	0%	0%	0%	0%
Avancées technologiques	0%	0%	0%	5%	5%	0%	0%	0%	0%	0%	0%	5%	0%	0%	0%
Pratique en matière d'emploi	0%	5%	5%	0%	0%	5%	0%	5%	5%	5%	5%	10%	15%	15%	10%
Changement dans la gestion des réclamations des sinistres	5%	5%	5%	5%	5%	5%	0%	5%	5%	5%	5%	5%	5%	5%	5%
Impact des actes de management	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Pandémie	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Impact de la compétition du marché	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	5%	5%	5%

Table des matières

Remerciements	i
Résumé	ii
Abstract	iv
Introduction	2
0.1 Contexte et justification de l'étude	2
0.2 Problématique	3
0.3 Objectifs	3
0.4 Plan de travail	4
1 Cadre de l'assurance non-vie	5
1.1 Réglementation solvabilité 2	5
1.1.1 Le pilier 1 ou exigences quantitatives	5
1.1.2 Le pilier 2 ou exigences qualitatives	7
1.1.3 Le pilier 3 ou discipline de marché	7
1.2 La formule standard	7
1.2.1 Présentation générale	7
1.2.2 Le risque de souscription non-vie	8
a) Le risque de prime et réserve	8
b) Le risque de rachat	8
c) Le Risque de catastrophe naturelle (CAT)	8
1.3 Le risque de prime dans le modèle d'AVIVA : définition du risque de sinistralité future	9
1.3.1 Description des segments de risque	9
1.3.2 Les mesures de risque	10
i) Les axiomes de cohérence	11
ii) Exemples de mesures de risque	11
ii.1) La variance et l'écart-type	11
ii.2) Value-at-Risk	12
1.3.3 Le SCR de sinistralité future	13
Définition du modèle	13
Calcul du SCR	14
Démarche	15

2	Retraitement des données	16
2.1	Description de la base	16
2.2	Mise en as-if	18
2.2.1	Méthodologie	18
2.2.2	Indices utilisés	19
2.2.2.1	Indice interne : évolution du coût moyen des sinistres	19
2.2.2.2	Indices de marché	20
a)	MCC Accident	20
b)	Les MCC Automobiles corporelles et de responsabilité civile	20
c)	Les MCC Automobiles matérielles	21
d)	Les MCC Dommages	21
e)	MCC Divers et PR-Juridique	21
f)	La MCC grêle	21
g)	Les MCC Construction	21
3.2.2.3	Projection des indices de marché	21
2.2.3	Résultats et impact de la mise en as-if	22
2.3	Projection à l'ultime	23
2.3.1	Calcul des facteurs de développement	24
2.3.2	Résultats de la projection à l'ultime	26
3	Étude du seuil de sinistre grave par branche de garanties	28
3.1	Théorie des valeurs extrêmes	28
3.1.1	Loi du maximum	28
3.1.2	Loi des dépassements d'un seuil	30
3.2	Estimation de l'indice de queue	31
3.3	Méthodes de détermination du seuil	32
3.3.1	Le graphique des dépassements moyens ou Mean Excess Plot	33
3.3.2	Le graphique de stabilité des paramètres d'échelle et de queue	33
3.3.3	Le Hill plot	33
3.3.4	AMSE (Asymptotic Mean Square Error)	33
3.3.5	Gertensgarbe plot	33
3.4	Application	34
3.4.1	Statistiques descriptives	34
3.4.2	Détermination du domaine d'attraction	35
a)	QQ plot exponentiel	35
b)	Test sur ξ	36
3.4.3	Détermination du seuil	36
a)	Mean excess plot	36
b)	Hill Plot	37
c)	AMSE	37
d)	Gerstengarbe plot	38
e)	Validation du seuil retenu	39

4	Modélisation de la sinistralité	41
4.1	Tests d'adéquation à une loi	41
4.1.1	Estimation des paramètres de la loi	41
	a) Méthode des moments	41
	b) Méthode du maximum de vraisemblance	41
4.1.2	Test de Kolmogorov-Smirnov	42
4.1.3	Tests de Cramer-von Mises et d'Anderson-Darling	42
	a) Tests de Cramer-von Mises	43
	b) Tests d'Anderson-Darling	43
4.1.4	Graphique Quantiles-Quantiles (Q-Q plot)	43
4.2	Modélisation	44
4.2.1	Modélisation de la sinistralité attritionnelle	44
	Estimation des paramètres	44
	Minimisation de la distance	45
	Choix de la loi	45
	Tests graphiques	46
4.2.2	Modélisation de la sinistralité grave	47
	Estimation des paramètres	47
	Minimisation de la distance	48
	Choix de la loi	49
	Tests graphiques	49
4.3	Calcul des SCR stand-alone	51
5	Théorie quantitative sur l'agrégation des risques	53
5.1	Histoire des corrélations	53
5.1.1	Une approche hiérarchique	53
5.1.2	Les outils courants de mesure de dépendance	54
	a) Le coefficient de corrélation linéaire de Spearman	54
	b) Le Tau de Kendall	55
	c) Le Rho de Spearman	55
5.2	Théorie des copules	55
5.2.1	Définitions et résultats de base	56
5.2.2	Principales copules	57
	a) La copule gaussienne	57
	b) La copule de Student	58
	c) Les Copules archimédiennes	58
	d) La copule HRT	58
5.2.3	Estimation des paramètres d'une copule	59
	a) Méthode des moments	60
	b) Méthode IFM	60
5.2.4	Choix de la meilleure copule	61
	a) Le maximum de vraisemblance	61
	b) Le critère AIC	61
	c) La fonction de Kendall $K(u)$	61
	d) Les « fonctions de concentrations dans les queues » $L(z)$ et $R(z)$	61

	e) Minimisation de la distance	62
5.2.5	Copules et simulation	62
6	Détermination de la structure de dépendance par une approche quantitative	63
6.1	Ré-échantillonnage des données	63
6.1.1	Théorie du Bootstrap	64
6.1.2	Mise en oeuvre du Bootstrap	64
	a) Cas des attritionnels	64
	b) Cas des graves	66
	c) Prise en compte de la structure de dépendance	66
6.2	Structure de dépendance	66
6.2.1	Structure de dépendance entre la sinistralité attritionnelle et la sinistralité grave par branche	66
	a) Visualisation graphique de la structure de dépendance	66
	b) Estimation	67
	c) Choix de la copule	68
	d) Calcul du SCR de chaque MCC	71
	e) Estimation de la corrélation entre la sinistralité attritionnelle et la sinistralité grave pour chaque branche	72
6.2.2	Structure de dépendance entre MCC	73
	a) Pré-sélection par le tau de Kendall	73
	b) Exclusion des couples avec des dépendances négatives	75
	c) Visualisation graphique des structures de dépendance : copules empiriques	76
	d) Estimation des paramètres des différentes copules	77
	e) Choix de la copule	77
	f) Copules retenues sur les couples sélectionnés	80
	g) Calcul du SCR agrégé de deux branches (MCC)	81
	h) Estimation de la corrélation implicite entre les SCR des paires de MCC	81
7	Une approche par facteur de causalité	84
7.1	Liste des facteurs de causalité	84
7.1.1	La Super inflation	85
7.1.2	Augmentation du PIB	86
7.1.3	Changement dans la pratique légale	86
7.1.4	Erreur contractuelle	87
7.1.5	Évènements naturels	87
7.1.6	Vol et fraude interne	87
7.1.7	Actions de l'homme	87
7.1.8	Erreurs humaines	88
7.1.9	Échec des contrôles	88
7.1.10	Qualité des données et techniques	88
7.1.11	Réseau de distribution	88
7.1.12	Risque de longévité	88
7.1.13	Tendances sociales et culturelles	89
7.1.14	Avancées médicales	89

7.1.15	Avancées technologiques	89
7.1.16	Pratiques en matière d'emploi	90
7.1.17	Changement dans la gestion des réclamations des sinistres	90
7.1.18	Changement des actes de management	90
7.1.19	Pandémie	90
7.1.20	Impact de la compétition du marché	91
7.2	Quantification des facteurs de causalité	92
7.3	Calcul des corrélations	92
7.3.1	La méthode cardinale	93
7.3.2	Le tau-b de Kendall	94
7.3.3	La méthode additive	96
8	Matrice de corrélation idéale	98
8.1	Comparaison des différentes méthodes	99
8.1.1	Comparaison des SCR	99
8.1.2	Méthode cardinale (échelle binaire) vs Méthode additive (échelle ordinaire proportionnelle)	100
8.1.3	Tau-b de Kendall - Échelle binaire vs Échelle ordinaire	100
8.1.4	Échelle binaire - Méthode cardinale vs Tau-b de Kendall	100
8.1.5	Échelle ordinaire - Méthode additive vs Tau-b de Kendall	101
8.1.6	Analyse détaillée des différentes corrélations	102
8.2	Avantages et inconvénients des différentes méthodes	104
8.2.1	Matrice QIS préconisée par l'EIOPA	104
8.2.2	La méthode des copules	105
8.2.3	Approche par facteurs de causalité	105
8.3	Choix de la matrice de corrélation idéale	106
	Limites de l'étude	109
	Conclusion	110
	Bibliographie	110
	Annexes	113
	Table des matières	131
	Liste des figures	133
	Liste des tableaux	135

Table des figures

1.1	Bilan vision Solvabilité 2	6
1.2	Approche modulaire du calcul du SCR	7
1.3	La non sous-additivité de la VaR	12
1.4	Les pertes attritionnelles, les pertes graves et les pertes catastrophiques	13
2.1	Inflation interne du segment Automobile matérielle pour les particuliers	19
2.2	Projection des indices de marché	22
2.3	Test de l'hypothèse H1 du modèle de Chain-Ladder	25
2.4	Test de l'hypothèse H2 du modèle de Chain-Ladder	25
3.1	Coûts individuels des sinistres en k€	35
3.2	QQ plot exponentiel (Coût en k€)	35
3.3	Test sur ξ	36
3.4	Mean excess plot	36
3.5	Hill plot	37
3.6	AMSE	38
3.7	Gerstengarbe plot	38
3.8	Validation du seuil de sinistre grave	39
4.1	Ajustement graphique	46
4.2	Graphiques Quantiles-Quantiles	46
4.3	Ajustement loi de la sévérité grave	49
4.4	Ajustement loi du nombre de sinistres graves	50
4.5	Q-Q plot fréquence des sinistres graves	50
5.1	Hiérarchie de l'agrégation de la sinistralité future	54
5.2	Distribution des copules archimédiennes et de la copule de survie de Clayton	59
6.1	Estimation du paramètre d'une copule normale bivariée de paramètre 0,5 à partir de 100 échantillons de 19 simulations issues de cette copule	63
6.2	Résidus du segment "Accident"	65
6.3	Copule empirique sinistralité attritionnelle-sinistralité grave pour le segment Accident	67
6.4	Fonctions de concentration dans les queues	69
6.5	Copules empiriques entre sinistralités attritionnelle et grave	70
6.6	Copules empiriques entre les sinistralités des MCC	76
6.7	Copule entre MCC : Fonctions de concentration dans les queues	79
6.8	Corrélations (méthode des copules)	82

7.1	Variation des indices d'inflation annuels	85
7.2	Méthode cardinale : échelle binaire	94
7.3	Tau-b de Kendall : échelle binaire	95
7.4	Tau-b de Kendall : échelle ordinale et échelle ordinale proportionnelle	95
7.5	Méthode additive	96
8.1	Matrice de corrélation QIS préconisée par l'EIPOA	98
8.2	Comparaison des SCR	99
8.3	Comparaison des corrélations (1)	102
8.4	Comparaison des corrélations (2)	102
8.5	Comparaison des corrélations (3)	103
8.6	Comparaison des corrélations (4)	104
8.7	Matrice de corrélation retenue	108
8.8	MCC5 et MCC6	114
8.9	MCC7 et MCC9	115

Liste des tableaux

1.1	Segments de risques	9
2.1	Description de la base	16
2.2	Répartition des données historiques de sinistres	17
2.3	Coûts moyens historiques par année de survenance	18
2.4	Impact de la mise en as-if	22
2.5	Impact de la mise en as-if sur les 5 MCC les plus matérielles	23
2.6	Triangle de développement	24
2.7	Coefficients de développement	26
2.8	Résultat de la projection à l'ultime	27
3.1	Statistiques descriptives	34
3.2	Quantiles	35
3.3	Seuils par la méthode de Gerstengarbe pour 5 itérations	39
3.4	Seuils retenus pour chaque segment de risque	40
4.1	Estimation des paramètres des lois théoriques	44
4.2	Tests d'adéquation	45
4.3	Tests d'adéquation : choix de la loi	45
4.4	Modélisation de la charge ultime attritionnelle en unité d'exposition	47
4.5	Estimation des paramètres de la loi de la sévérité grave	48
4.6	Estimation des paramètres de la loi du nombre de sinistres graves	48
4.7	Tests d'adéquation	48
4.8	Tests d'adéquation : choix de la loi	49
4.9	Modélisation de la charge des sinistres graves (en millions d'euros)	51
4.10	Modélisation de la fréquence grave	51
4.11	SCR par type de sinistralité et par Branche (en millions d'euros)	52
5.1	Copules archimédiennes	58
6.1	Copule sinistralité attritionnelle-sinistralité grave pour le segment Accident	67
6.2	Choix de la copule : test d'adéquation	68
6.3	Distance entre les Fonctions empirique et théoriques de concentration dans les queues	69
6.4	Agrégation attritionnel et grave par branche	71
6.5	SCR des MCC	72
6.6	Corrélation entre sinistralité attritionnelle et sinistralité grave	73
6.7	Significativité du Tau de Kendall (A)	74

6.8	Significativité du Tau de Kendall (B)	75
6.9	Estimation des paramètres des Copules entre les sinistralités des MCC	77
6.10	Choix de la copule entre MCC : test d'adéquation	78
6.11	Copule entre MCC : Distance entre les de fonctions concentration dans les queues empirique et théoriques	79
6.12	Copules entre MCC	80
7.1	Quantification des facteurs de causalité	92
8.1	Comparaison des statistique des corrélations	99
8.2	Sensibilité du Tau-b de Kendall	106