Mémoire présenté devant le jury de l' EURIA en vue de l'admission à l'Institut des Actuaires

le 25 Septembre 2019

Par : Maxime BEN-BRIK
Titre : Provisionnement des sinistres graves en responsabilité civile médicale

Confidentialité : Oui - (Durée : 2 ans)

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

Membres présents du jury de l'Institut des Actuaires :
Romain LAÏLY
Jérémie DEVUN
Anne BONTOUX
Khalil SAID

Signatures :

Entreprise :
Prim'Act

Signature :

Membre présent du jury de l' EURIA :
Franck VERMET

Signature :

Directeurs de mémoire en entreprise :
Frédéric PLANCHET
Gilles DEPOMMIER

Signatures :

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

Signature du responsable entreprise :

Signature du candidat :

EURIA
EURo Institut
d'Actuariat

6, avenue le Gorgeu
CS 93837
29238 Brest Cedex 3

T +33 (0)2 98 01 66 55
F +33 (0)2 98 01 66 57
euria@univ-brest.fr

Résumé

Pour une branche à développement long comme la responsabilité civile professionnelle des médecins, en moyenne, les 1% de sinistres les plus graves représentent jusqu'à 50% de la charge totale. Leur bonne évaluation *a priori* constitue alors un enjeu majeur pour l'assureur.

L'objet de ce mémoire est d'identifier et de provisionner ces sinistres dans le cadre de la responsabilité civile médicale (abrégé « RCM »). Les garanties de la RCM concernent en grande partie toute erreur de diagnostic ou de traitement qui est considérée comme une faute, c'est-à-dire, dans le cadre juridique français, dès lors qu'elle a été causée par un défaut d'information, un défaut de consentement, un manquement au devoir d'assistance ou par un manquement au devoir de secret professionnel de la part du praticien.

L'identification des sinistres graves s'articule autour de la théorie des valeurs extrêmes. Elle aide à obtenir la définition d'un seuil statistiquement stable à partir duquel un sinistre est considéré comme grave, et ce à partir de l'évaluation du coût unitaire de chaque sinistre, en fonction de ses caractéristiques. Ces informations sont complétées par une évaluation du nombre de sinistres graves non encore déclarés, pour le calcul des IBNR et, le cas échéant, d'une évaluation des sinistres non encore survenus.

Dû à leur faible nombre, la modélisation du provisionnement de sinistres graves, que ce soit pour les comptes sociaux ou pour les comptes prudentiels (Solvabilité II), implique une mise en œuvre d'ajustements actuariels pour limiter au maximum l'impact de la forte volatilité que présentent leurs charges. Sous ces conditions, les méthodes de provisionnement simples, telles que *Chain-Ladder*, peuvent réaliser des résultats satisfaisants.

Mots clefs : responsabilité civile médicale, sinistre grave, provisionnement, Solvabilité II, inflation, trajectoire, mise en *as-if*, théorie des valeurs extrêmes, *Peak-Over-Threshold*, *Chain-Ladder*, *London-Chain*, Bornhuetter-Ferguson, Mack, *Bootstrap*, Monte-Carlo

Summary

For a long-running branch, such as the medical professionnal for doctors, 1% of the most severe losses represents, on average, up to 50% of total charges. Their good *a priori* evaluation is a major stake for the insurer.

This thesis' purpose is to identify and provision these losses as part of the medical liability (shorten "ML"). The ML's waranties largely concern any misdiagnosis or treatment that is considered as a fault, which means, in the French legal framework, since it was caused by an information defection, a lack of consent, a duty of assistance's breach or a professional secrecy's violation from the practitioner.

Severe losses identification is based on extreme value theory. It helps getting a threshold that is statistically stable above which a loss is considered as severe, and this from each loss' unit cost evaluation, according to its characteristics. These information are completed by number of severe losses that are not yet reported for IBNR calculation and, where applicable, an assessment of losses not yet incurred.

Due to their small number, severe losses provision modeling, whether it is for corporate accounts or for prudential accounts (Solvency II), implies the implementation of actuarial adjustments in order to minimize the high volacity's impact of their charges. Under these conditions, simple provisioning methods, such as Chain-Ladder, can achieve satisfying results.

Keywords : medical professional liability insurance, severe medical mispractice, provisioning, Solvency II, inflation, trajectory, as-if implementation, extreme value theory, Peak-Over-Threshold, Chain-Ladder, London-Chain, Bornhuetter-Ferguson, Mack, Bootstrap, Monte-Carlo

Synthèse

En France, la responsabilité civile médicale (abrégée « RCM ») oblige réparation contre tout dommage subi dans le cadre de l'exercice d'une profession de santé. Une assurance prévue à cet effet doit couvrir tout préjudice réclamé par un tiers (généralement un patient) et où la responsabilité du praticien assuré est prouvée et décidée par la Justice comme engagée.

Peu d'établissements proposent un contrat d'assurance RCM, ce qui se traduit par un marché français très fermé avec des assurances spécialisées sur des classes de risque distinctes au sein du secteur de la santé : on relève entre autres des assurances dédiées aux médecins libéraux, par spécialité, ou encore des assurances spécialisées sur la responsabilité civile des établissements de santé.

Une analyse de marché montre que les indemnisations moyennes par sinistre sont très hétérogènes : les chirurgiens, gynécologues obstétriques et médecins généralistes représenteraient à eux trois plus de la moitié des indemnisations totales identifiées sur le marché.

La responsabilité civile médicale a subi plusieurs évolutions juridiques qui ont cherché à atténuer l'impact financier des sinistres graves. En effet, il est observé qu'en moyenne, les 1% de sinistres les plus graves représentent jusqu'à 50% de la charge totale. La réforme la plus significative de ces dernières décennies est la loi « About [1] », qui ne considère plus les sinistres RCM en date de survenance mais en date de réclamation.

Pour provisionner en accord avec cette réforme, les données sont alors divisées sous trois bases de rattachement :

- une base survenance, qui contient les sinistres clos avant la loi « About » ;
- une base réclamation, qui contient les sinistres ouverts ou reconduits après l'entrée en vigueur de cette loi ;
- et une base subséquente qui contient les sinistres post-réforme réclamés après fermeture d'un contrat.

Dans le cadre de ce mémoire, le calcul des provisions pour sinistres graves s'articulera sur un portefeuille de professionnels de santé libéraux. Les retraitements et études de cohérences effectués montrent une bonne qualité des données utilisées.

1. [Journal Officiel, 2002a]

L'identification des sinistres graves s'appuie sur l'estimation d'un « seuil de gravité ». Une première approche, naïve, serait d'appliquer une méthode de dichotomie où ce seuil serait déterminé tel que le montant total de charges sur les dossiers qui le dépassent converge vers les 50% de charges cumulées sur tout le portefeuille. Cette méthode est efficace en temps de calcul mais elle se montre très volatile avec l'évolution du portefeuille dans le temps.

Une seconde approche, mathématiquement plus fiable, s'inspire de la méthode « *Peak Over Threshold* » et est élaborée suivant la théorie des valeurs extrêmes. Son objectif est de déterminer un seuil tel que la queue de distribution, caractérisée par l'ensemble des dossiers qui ont une charge totale supérieure à ce même seuil, soit en adéquation avec une loi d'*extremum*.

La recherche d'une telle loi s'appuie sur des méthodes d'analyses graphiques, renforcées par des tests statistiques. Leur application sur le portefeuille étudié démontre que la queue de distribution des charges cumulées suit une loi de Pareto et que son paramètre de position estimé fournit un seuil de gravité statistiquement robuste.

Les sinistres RCM étant réputés comme longs, les montants ont été mis en *as-if* en s'appuyant sur l'indice des prix à la consommation français afin de limiter les effets liés à la dérive du seuil de gravité dans le temps.

Une fois ce seuil obtenu, les triangles de liquidation des sinistres graves RCM peuvent être construits. Avec une analyse d'adéquation des méthodes de provisionnement déterministes sur la sinistralité antérieure observée, on montre que :

- la méthode de *Chain-Ladder* semble fournir des provisions correctement estimées pour la base survenance ;
- la méthode de *London-Chain* semble fournir des provisions correctement estimées pour la base réclamation ;
- et la méthode de Bornhuetter-Ferguson, en fixant comme paramètre exogène les ultimes estimés par *Chain-Ladder*, fournit la meilleure estimation des provisions parmi toutes les méthodes testées sur la base subséquente.

Dans l'ensemble, ces contrôles effectués sur les méthodes de provisionnement déterministes montrent l'importance de la méthode d'identification des sinistres graves pour obtenir des calculs de provisions satisfaisant sans avoir à faire intervenir un modèle particulièrement complexe.

Quant aux méthodes de provisionnement stochastiques, elles fournissent des résultats équivalents mais les erreurs de prédiction estimées soulignent la forte volatilité des charges liées aux sinistres graves. L'aspect exceptionnel de ces sinistres doit justifier cette observation, ce qui amène alors à réfléchir sur le besoin de mutualiser les sinistres graves dans cette branche de la responsabilité civile professionnelle pour palier ce problème de volatilité.

Abstract

In France, medical civil liability (shorten "ML") is legally required to compensate any damage encountered while exercising healthcare activites. An insurance product dedicated for this purpose has to cover any loss claimed by a third party (usually a patient) and where the responsibility of the insured practitioner is proven and decided by the law as committed.

Few companies provide ML policies, which leads to a very restricted French market with specialized insurances products on separated risk classes within healthcare sector : there are insurers, among others, dedicated to private doctors, depending on the speciality, or civil liability insurers for health institutions.

A market analysis shows that the average compensation per loss is very heterogeneous : surgeons, obstetric gynecologists and general practitioners together represent more than half of the total identified compensations.

Medical civil liability has undergone several legal developments that aimed to lighten severe losses' financial impact. In fact, it is observed that, on average, 1% of the most severe losses represents up to 50% of total charges. The most significant reform in the past decades is the "About" law, which no longer considers ML losses from occurence date but from claim date.

According to this reform, datas are divided into three attachment subdatas for provisionning :

- An occurence database, which contains losses closed before "About [2]" law.
- A claim database, which contains losses opened or renewed after this law came into force.
- And a subsequent database which contains post-reform losses claimed after the policy's closure.

As part of this thesis, severe losses' provision calculation will be based on a private health professionnals' portfolio. The data processing followed by a data analysis underline a good quality database.

2. [Journal Officiel, 2002a]

Serious losses' identification is based on a "severity threshold" estimation. A naive first approach consists on applying a dichotomic search where this threshold would be determined such that the total charges on records exceeding it converges to 50% of accumulated charges over the entire portfolio. This method is efficient when it comes to computation and execution ; however, it represents a considerable volatility over the time with the portfolio's evolution.

A second approach that is mathematically more reliable is inspired by the "Peak Over Threshold" method and is developed according to extreme values theory. Its objective is to determine a threshold such as the tail of a distribution, characterized by the set of records that have a total charge greater than the same threshold, that is in line with an extreme value distribution.

Searching such a distribution is based on graphical analysis methods strengthened by statistical tests. Their application to the studied portfolio shows that the cumulated charges' tail of a distribution fits with a Pareto distribution and its estimated location parameter provides a statistically robust severity threshold.

As severe LM losses are renowned for being long, amounts have been implemented "as-if" using the French consumer price index in order to limit the threshold ?s severity time drifting.

Once we get this threshold, severe ML losses' run-off triangles are built. With an adequacy analysis over deterministic provisioning methods on previously observered losses, we show that :

- The Chain-Ladder method seems to provide correctly estimated provisions for the occurence data.
- The London-Chain method seems to provide correctly estimated provisions for the claim data.
- And the Bornhuetter-Ferguson method, by fixing Chain-Ladder's ultimates as an exogeneous parameter, seems to provide the best provision estimation among the other tested methods, for the subsequent data.

Overall, these controls on deterministic provisioning methods show that the identification of severe losses models method's importance for obtaining satisfying provisioning calculations without the need of a particular complex model.

Stochastic provisionning methods provide similar results regardless of how the estimated error of prediction underlines the high severe losses charges' volatility. These losses' exceptional aspect must justify this observation, which leads us to think about the need to mutualize severe losses in this professional liability branch in order to solve this volatility problem.

Remerciements

Je remercie du fond du cœur mes parents pour leur soutien inconditionnel apporté tout au long de la rédaction de ce mémoire.

Je remercie également Frédéric PLANCHET et Gilles DEPOMMIER, actuaires associés de Prim'Act, pour leur encadrement, leur expertise et leur disponibilité sans failles qui m'ont beaucoup aidé pour ce mémoire.

Je remercie de même Danielle KHOUGEA, actuaire consultante chez Prim'Act, qui, dans sa grande générosité, a pris de son temps pour relire dans les moindres recoins ce mémoire.

Je remercie aussi Patrick TANZILLI, responsable du département Finances - Réassurance de La Médicale, pour son accord vis-à-vis de l'utilisation des données, sans lequel ce mémoire n'aurait pas existé.

Enfin, je remercie tous les associés et collègues Prim'Act pour leur immense sympathie et leurs conseils aiguisés qui m'auront ainsi aidé à avancer ce mémoire en toutes circonstances, en particulier les stagiaires et consultants qui préparent aussi leur mémoire pour cette année et avec qui un esprit d'entraide s'est instauré : Olivier CALZA, Alexandre EXTRAT, Julie FABOUX, Pierre RUNDSTADLER et Auriol WABO FOKA.

Table des matières

Introduction **1**

1 Responsabilité civile médicale et sinistre grave **3**
1.1 Cadre juridique de la responsabilité civile médicale en France 4
1.1.1 Notion de responsabilité : frontière entre sanction et indemnisation 4
1.1.2 Principes de la responsabilité civile professionnelle 5
1.1.3 Application au secteur médical 6
1.2 Le marché français de la responsabilité civile médicale 7
1.2.1 Un faible volume d'informations 9
1.2.2 État des lieux de la sinistralité 10
1.2.3 Contributions des assureurs spécialistes de la responsabilité civile médicale 13
1.3 Notion de sinistre grave et enjeux 22
1.3.1 Spécificités de la responsabilité civile médicale 23
1.3.2 Provisionnement du sinistre RCM : une question de temps 24
1.3.3 Périmètre d'étude du mémoire 26

2 Modalités de définition d'un sinistre grave **31**
2.1 Construction de la base de données 32
2.1.1 Importation et mise en forme des bases initiales 32
2.1.2 Agrégations et jointures 32
2.1.3 Étude de cohérence et finalisation 37
2.2 Définition du choix du seuil des sinistres graves 47
2.2.1 Approche par dichotomie 47
2.2.2 Approche basée sur la théorie des valeurs extrêmes 51
2.2.3 Adaptation de la théorie des valeurs extrêmes à la problématique et analyse des résultats 57
2.3 Pistes d'amélioration sur la détermination du seuil 66
2.3.1 Réflexion sur la dérive du seuil dans le temps 66
2.3.2 Mise en *as-if* du portefeuille 68
2.3.3 Le *GW-plot*, une solution simple et fiable ? 71

3 Provisionnement des sinistres graves **75**
3.1 Principes de modélisation adaptés à la responsabilité civile médicale . . . 76
3.1.1 Conventions de provisionnement autour de la loi « About » 77
3.1.2 Méthodes d'estimation déterministes 78
3.1.3 Méthodes d'estimation stochastiques 83
3.2 Application sur les sinistres graves 87
3.2.1 Construction des triangles de sinistres graves 89
3.2.2 Utilisation de méthodes déterministes 90
3.2.3 Utilisation de méthodes stochastiques 97
3.3 Interprétation des résultats et pistes de réflexion 100
3.3.1 Validation de la pertinence des provisions : adaptation du *standardized mortality ratio* à la problématique 100
3.3.2 Application sur la base d'étude 101
3.3.3 Piste de réflexion : de l'importance de mutualiser 102

Conclusion **105**

A Présentation de l'ONIAM **107**

B Présentation des bases de données initiales **109**

C Présentation des bases de données finales **115**

D Méthodes d'estimation semi-paramétriques de queue de distribution **117**
D.1 Estimateur de Pickands 117
D.2 Estimateur de Hill 118
D.3 Estimateur de Dekkers-Einmahl-de Haan 118

E Méthodes d'analyse graphiques sur les valeurs extrêmes **119**
E.1 Diagramme quantile-quantile 119
E.2 Diagramme moyenne des excès 121
E.3 Diagramme des estimateurs 122
E.4 Diagramme de Gerstengarbe et Werner 124

F Tests d'adéquation statistique de lois **127**
F.1 Test du Khi-2 127
F.2 Test de Kolmogorov-Smirnov 128

Glossaire **129**

Table des figures **133**

Liste des tableaux **135**

Bibliographie **137**

Introduction

La responsabilité civile médicale, abrégée « RCM », est une branche de la responsabilité civile professionnelle portée sur le secteur de la santé. Elle possède des caractéristiques juridiques particulières et une sinistralité généralement longue à indemniser.

La RCM couvre en effet les professionnels de santé du risque d'être condamnées au civil à verser des compensations financières suite à une erreur médicale, ou toute autre forme de préjudice en lien avec l'activité exercée, et qui engage leur responsabilité. Ces sinistres traduisent en règle générale des séquelles sur l'humain.

Les cas les plus graves peuvent signifier une invalidité permanente, le besoin de l'assistance d'une tierce personne, ou une hospitalisation prolongée. De telles situations représentent un coût trop important pour être supportable intégralement par le professionnel de santé en tort et justifient en grande partie l'obligation de souscrire à une assurance.

Pourtant, en France, peu d'assureurs se sont engagés dans cette branche de la responsabilité civile professionnelle, en partie parce que sa sinistralité lourde s'avère très volatile et coûteuse : en moyenne, les 1% dossiers RCM les plus graves cumulent une charge totale qui peut aller jusqu'à 50% de l'ensemble du portefeuille.

Comment alors appréhender la sinistralité grave en responsabilité civile médicale ?

L'objectif de ce mémoire sera d'expliciter une démarche actuarielle qui sera portée sur l'identification et le provisionnement des sinistres graves sur un portefeuille d'assurance RCM. Elle introduira entre autres :

- la définition de responsabilité civile médicale, dans le sens du Droit français, et les évolutions juridiques de ces dernières décennies qui en découlent ;
- la théorie des valeurs extrêmes qui aidera à déterminer un « seuil de gravité » statistiquement stable et adapté au portefeuille étudié ;
- et des notions actuarielles de provisionnement de sinistres.

Cette démarche sera accompagné :

- d'une étude de marché sur la sinistralité RCM ;
- d'analyses détaillées des bases de données à disposition pour cette étude (qualité des données, retraitements, *etc.*) ;
- et des procédures de validation qui jaugeront la pertinence des résultats obtenus et des hypothèses émises (graphiques, statistiques ou empiriques).

Chapitre 1

Responsabilité civile médicale et sinistre grave

1.1 Cadre juridique de la responsabilité civile médicale en France

En France, il est obligatoire pour tout professionnel de santé de souscrire une assurance qui garantit leur responsabilité civile professionnelle [1]. Il convient de définir cette notion de la manière la plus exacte possible.

- Qu'est-ce que la responsabilité civile, au sens du droit français ?
- Sous quelles conditions la responsabilité professionnelle est-elle engagée ?
- Qu'est-ce qui distingue le domaine médical des autres activités professionnelles en termes de responsabilité ?

1.1.1 Notion de responsabilité : frontière entre sanction et indemnisation

La responsabilité se distingue sous deux objectifs selon les conséquences des actes qui sont mis en cause :

la responsabilité pénale : elle répond à des infractions qui sont propres à la loi ;

la responsabilité civile : elle répond à des dommages qui sont causés à autrui.

La responsabilité pénale sanctionne les comportements qui sont réprouvés par la loi. En ce sens, et par le principe de légalité des peines, elle ne peut être engagée que dans le cadre d'un texte légal qui décrit le ou les comportements en question comme une infraction. Les infractions se classent selon leur gravité :

la contravention : infraction la moins grave, elle est généralement assortie d'une amende et est jugée par le tribunal de police.

le délit : infraction intermédiaire, elle peut conduire à une peine d'emprisonnement, qui peut aller jusqu'à dix ans, ou à d'autres formes de sanction comme les travaux d'intérêts généraux ou le sursis avec mise à l'épreuve. Elle est jugée par le tribunal correctionnel.

le crime : infraction la plus grave, elle est passible de réclusion jusqu'à perpétuité et est jugée par la cour d'assises.

Par leur fonction répressive, il est interdit aux assurances de garantir les conséquences de toute forme d'infraction.

En France, toute personne physique ou morale doit réparer le dommage causé à autrui : c'est la responsabilité civile [2]. Elle se décline de deux façons :

la responsabilité civile contractuelle : réparation suite à un manquement à une obligation prévue dans un contrat conclu librement entre deux parties ;

la responsabilité civile délictuelle : réparation d'un acte préjudiciable commis en dehors de tout contrat par un tiers.

1. [Code des assurances, 2011]
2. [Code civil, 2016]

Trois critères engagent la responsabilité civile :

le préjudice : il peut être causé aux biens ou au patrimoine de la victime (préjudice matériel), relatif à l'intégrité physique, physiologique ou fonctionnelle de la victime (préjusice corporel) ou il peut découler d'une atteinte à l'affection, l'honneur ou la réputation de la victime (préjudice moral). Pour que le préjudice soit établi, il doit être certain, personnel, et direct. De surcroît, il est impératif que le préjudice demandé soit conforme à la loi, ou au moins à l'ordre public et aux bonnes mœurs, pour ne pas être rejeté par les tribunaux.

le fait : il correspond aux causes de la réalisation du préjudice. Le fait peut être un acte (dans ce cadre involontaire ou volontaire avec des conséquences involontaires) commis par l'auteur du préjudice (fait personnel), par le biais d'une chose dont il aurait eu la garde au moment du préjudice (fait des choses), ou par un tiers avec un pouvoir de direction, de surveillance et de contrôle au moment du préjudice (fait d'autrui).

le lien de causalité : pour engager la responsabilité civile, une fois le préjudice et le fait établis, il doit obligatoirement exister un lien clair entre ces deux éléments.

Les assureurs ont la possibilité de garantir la réparation de préjudices où une responsabilité civile est engagée.

Ainsi, la responsabilité pénale a une fonction répressive tandis que la responsabilité civile ne porte que sur le dédommagement d'une victime. En ce sens, les assurances type « responsabilité civile » prémunissent contre le risque de subir de lourdes réparations à verser qui pourraient aller jusqu'à la confiscation de tout bien ou patrimoine personnel si le condamné se trouvait insolvable. Indirectement, elles offrent une meilleure indemnisation des victimes et elles n'incitent pas à enfreindre la loi.

1.1.2 Principes de la responsabilité civile professionnelle

Comme son nom le sous-entend, la responsabilité civile professionnelle est engagée lorsque l'origine d'un préjudice remonte à l'exercice d'une profession. En d'autres termes, c'est une transposition de la responsabilité civile au monde du travail. Sous le droit français, il est stipulé que tout professionnel est systématiquement responsable de tout préjudice qui est causé par son activité.

De manière générale, la responsabilité civile d'un professionnel est engagée lorsqu'il a commis une faute clairement identifiée, que celle-ci a causé un préjudice sur un tiers, et que le lien de causalité entre la faute commise et ce préjudice a été prouvé. La faute doit avoir été comise de façon involontaire et sans intentions de nuire, ou la responsabilité civile basculerait en responsabilité pénale [3]. La faute involontaire se décline sous plusieurs formes :

l'imprudence : faute causée par une action irréfléchie et accomplie sans souci des conséquences préjudiciables qu'elle portait ;

3. [Code pénal, 2000]

la négligence : faute non intentionnelle qui résulte d'un manque de vigilance ;

le manquement à une obligation : faute généralement commise dans le cadre de la responsabilité civile contractuelle. Elle peut aussi correspondre au manquement à une obligation de prudence ou de sécurité.

En France, l'assurance « responsabilité civile professionnelle » n'est pas obligatoire pour tous les professionnels mais elle reste fortement recommandée. Les professions réglementées ont obligation de souscrire une assurance de ce type [4]. Les principaux secteurs d'activité qui sont concernés par cette obligation sont la santé, le droit, le bâtiment, et toute activité qui fait régulièrement usage de véhicules terrestres à moteur.

1.1.3 Application au secteur médical

La responsabilité civile professionnelle appliquée dans l'ensemble du secteur de la santé, abrégée « responsabilité civile médicale » (RCM), oblige réparation contre tout dommage subi par un ou plusieurs tiers dans le cadre de l'exercice de cette profession. Une assurance prévue à cet effet doit alors couvrir tout préjudice réclamé par un patient ou un tiers et qui serait prouvé comme étant en lien avec l'activité exercée par l'entité assurée [5]. Lorsque la responsabilité d'un hôpital est engagée, on parle également de « responsabilité civile administrative médicale ».

En ce sens, le modèle français d'assurance RCM doit se prémunir contre toute responsabilité engagée en cas de faute médicale, mais aussi contre les risques qui sont imbriqués par les pratiques exercées dans cette profession.

Depuis l'arrêté « Mercier » de 1936, la relation professionnelle entre un médecin et son patient se caractérise par un contrat médical. Ce contrat est :

civil : il n'est pas soumis au droit commercial, contrairement à un contrat commercial ;

oral : il n'est pas nécessaire d'établir un formalisme quelconque pour que le contrat médical soit établi ;

conclu *intuitu personnae* : le professionnel de santé doit traiter ses patients personnellement ;

synallagmatique : les deux parties du contrat médical ont chacun des obligations différentes. Le patient doit payer les honoraires et suivre les prescriptions, et le praticien doit soigner et informer.

Au-delà du contrat médical, la responsabilité civile d'un professionnel de santé peut être également engagée suite à un dommage non-contractuel ou à une invalidation du contrat médical qui a été établi.

4. [Bercy Infos, 2018]
5. [Code de la santé publique, 2011]

Le professionel de santé français a obligation de moyen, pas de résultat. Dès lors qu'il a répondu à une demande d'un patient, il doit [6] :

- respecter le secret professionnel, institué dans l'intérêt du patient ;
- assurer des soins consciencieux, dévoués et fondés sur les données acquises de la science ;
- procurer une information loyale, claire et appropriée sur l'état, les investigations et les soins proposés ;
- tenir compte de la personnalité du patient dans ses explications et veiller à leur compréhension ;
- agir dans les limites de ses compétences et faire appel, dans un quelconque doute ou si nécessaire, à un professionnel de santé spécialisé ;
- ne faire courir au patient aucun risque injustifié ;
- avoir le consentement du patient dans toutes les circonstances.

Le non-respect d'au moins une de ces conditions peut engager une réclamation du patient ou d'un tiers pour préjudice causé par le praticien en faute.

Par exemple, une telle assurance doit pouvoir garantir :

- un acte de prévention erroné (défaut d'information de la part du praticien) ;
- une erreur de diagnostic ;
- une faute commise dans le cadre des soins ;
- un accident médical ;
- une affection iatrogène (effet secondaire induit par un traitement) ;
- une infection nosocomiale (en milieu hospitalier).

À l'inverse, si une réclamation d'un patient ou d'un tiers sur des dommages médicaux est démontrée comme n'engageant pas la responsabilité du professionnel de santé ou de l'établissement de santé assuré, alors les frais de procédures qui sont liés à cette réclamation (intervention d'un avocat notamment) ne sont pas couverts.

À titre d'exemple, les cas suivants ne sont pas couverts par une assurance responsabilité civile médicale pure :

- infection nosocomiale qui serait causée par des facteurs extérieurs à l'établissement de santé concerné ;
- accusations à tort d'un patient contre un patricien ;
- défaut de fabrication d'un médicament prescrit.

D'autres garanties sont généralement ajoutées aux contrats RCM (en particulier la protections juridiques) pour couvrir ces cas-là.

1.2 Le marché français de la responsabilité civile médicale

Le marché de la responsabilité civile médicale a la particularité d'être à la fois étroit et onéreux par la complexité de sa sinistralité. En France, les assureurs sont confrontés depuis plusieurs décennies à deux sources de craintes :

6. [Code de la santé publique, 2012]

l'existence de « trous de garantie » : les montants d'indemnisation à garantir par les assureurs sont *in fine* fixés librement par la Justice lorsqu'un accord à l'amiable n'a pas été trouvé entre la victime et le responsable. Les écarts entre les montants d'indemnisation assurés et les montants d'indemnisation issus de décisions de Justice peuvent alors contraindre les assureurs à refuser de verser des indemnités, car sortent des clauses du contrat d'assurance. Ces situations sont rares mais augmentent significativement la difficulté de construire des dispositifs de garanties à la fois mathématiquement performants et accessibles à une plus large population possible.

l'augmentation régulière et inégale des primes : l'indemnisation des victimes repose sur le principe de la réparation intégrale. Les assurances étant segmentées par spécialité, les spécialités à risque se retrouvent alors confrontées à des hausses de prime importantes. De surcroît, ce principe génère également des expositions au risque différentes selon la spécialité du praticien couvert.

Pour atténuer ces risques, le cadre juridique de la responsabilité civile médicale a été modifié à plusieurs reprises :

loi « Kouchner » : (mars 2002 [7]) obligation des professionnels de santé libéraux et des établissements de santé de souscrire une assurance RCM et transfert partiel de l'indemnisation des sinistres à la solidarité nationale avec la création de l'Office National d'Indemnisation des Accidents Médicaux (ONIAM) ;

loi « About » : (décembre 2002 [8]) instauration d'un régime de contrats d'assurance fondé sur la réclamation et non plus sur le fait générateur, prolongement des garanties d'assurance sur une période de dix ans après cessation d'activité et absorption des risques non assurés *via* le Bureau Central de Tarification (BCT) avec la création du Groupement Temporaire d'Assurance Médicale (GTAM) ;

réponse à la crise assurantielle de 2006 : instauration d'un système de subventions par la CPAM destinées aux praticiens conventionnés et accrédités ;

mission « Johanet » : (juillet 2010 puis janvier 2011 [9]) sollicitation d'experts pour apporter de nouvelles propositions d'amélioration de la responsabilité civile médicale. Une importante vague de débats dans les institutions politiques a suivi cette mission et a abouti, entre autres, sur la loi « Fourcade » (modifiation du plafond de garantie des contrats d'assurance des professionnels de santé libéraux) ;

loi de finance 2012 : (décembre 2011 [10]) création d'un fonds de garantie qui intervient en cas de dépassement des garanties contractuelles ou en cas d'une intervention sur un sinistre au-delà d'une cessation des garanties, une cessation d'activité, ou un décès.

etc.

7. [Journal Officiel, 2002b]
8. [Journal Officiel, 2002a]
9. [Johanet, 2011]
10. [Journal Officiel, 2011]

Malgré ces évolutions régulières, les assureurs restent peu nombreux à proposer des produits qui garantissent la responsabilité civile médicale.

1.2.1 Un faible volume d'informations

L'aspect « cas par cas » imprégné au travers des décisions de Justice, et le faible volume de données qui est disponible par spécialité médicale sont sources d'incertitude conséquentes sur les sinistres à couvrir. Une tarification mathématiquement robuste sous ces conditions présente cette branche de la responsabilité civile comme plus difficile à gérer que les autres ; les dommages causés sur le corps humain, socle de l'activité des professionnels de santé, sont plus complexes à quantifier que les dommages matériels.

Peu de statistiques publiques sont disponibles pour un état des lieux récent et complet du marché français des assurances responsabilité civile médicale. Ce secteur très fermé regroupe aujourd'hui une quinzaine d'acteurs qui se découpent en trois catégories :

les compagnies spécialistes de la responsabilité civile médicale :
- Mutuelle d'Assurance du Corps de Santé Français (MACSF) - le Sou Médical ;
- Société Hospitalière d'Assurances Mutuelles (SHAM) ;
- la Médicale ;
- Groupe Pasteur Mutualité ;
- *etc.*

les compagnies généralistes françaises :
- AXA ;
- Mutuelles du Mans Assurances (MMA) ;

les compagnies étrangères :
- Medical Insurance Company (MIC), société spécialiste de droit irlandais crée en 2002 ;
- AmTrust Financial Services, société américaine fondée en 1998 ;
- Chubb Limited, *consortium* d'assurance et de réassurance mondial qui siège à Zurich en Suisse ;
- CNA-Hardy, société américaine fondé en 1897 ;
- The Travelers Companies, société américaine fondée en 1853 ;
- *etc.*

En France, le marché de la responsabilité civile médicale représentait 2,8% du marché de l'assurance IARD des professionnels sur l'année 2016, avec un total de primes encaissées de 550 M€ [11]. Selon les éléments rapportés dans le cadre de la mission « Johanet [12] », les primes émises en 2009 représentaient 0,75% des primes émises en assurance non-vie. Au niveau des principales spécialités assurées, les répartitions suivantes ont été relevées :
- 68% des primes versées en gynécologie obstétrique proviendraient des deux premiers opérateurs ;
- 78% en anesthésie réanimation ;

11. [Banga, 2017]
12. [Johanet, 2011]

- et 83% en chirurgie.

Ces trois spécialités, fortement concentrées sur les principaux acteurs de la responsabilité civile médicale, sont les plus risquées de ce marché. Les principaux acteurs de ce marché sont :

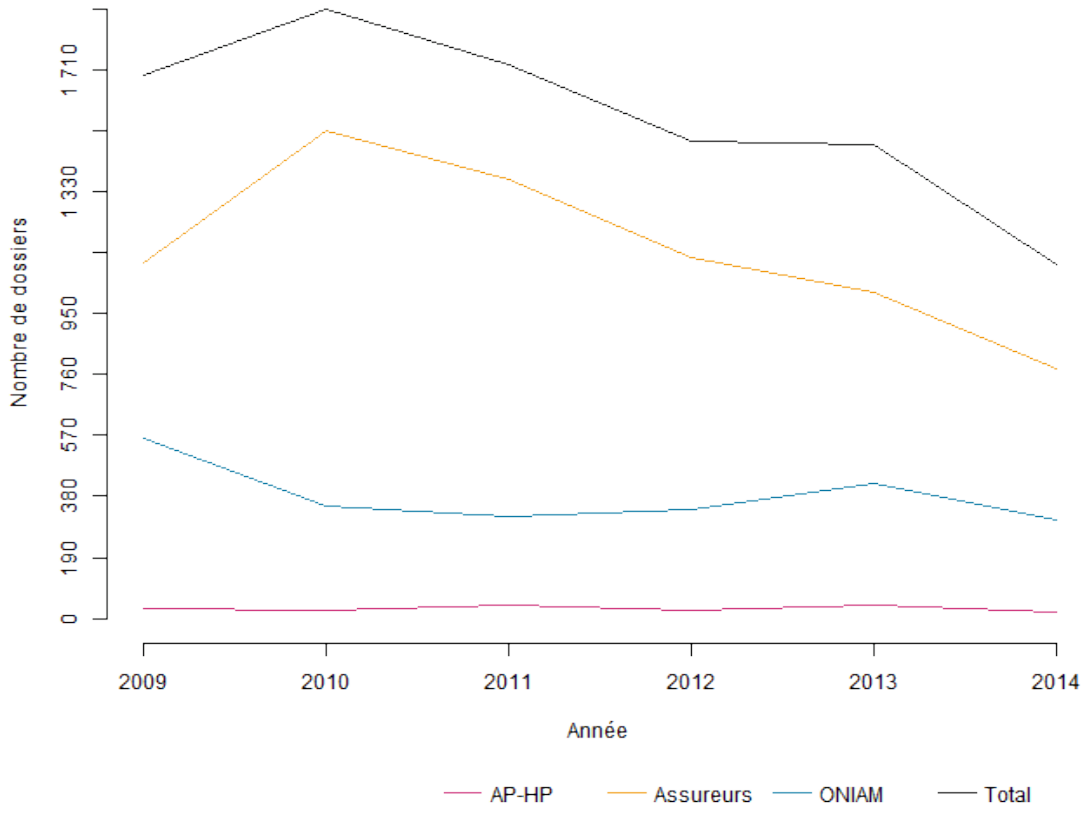
- MACSF-Le Sou Médical avec les professionnels de santé du secteur libéral ;
- et SHAM avec les établissements de santé.

Ces deux entités publient chacune un rapport annuel sur la sinistralité qu'elles observent sur le marché ; pour l'année 2016, le taux de sinistralité relevé sur les professionnels de santé s'élève à 0,99% et celui des médecins, toutes spécialités confondues, s'élève à 1,63%.

1.2.2 État des lieux de la sinistralité

Le rapport d'activité 2015 [13] de l'Observatoire des Risques Médicaux (ORM, placé auprès de l'ONIAM) est l'un des derniers rapports en date publié par cette institution qui relève et compare des données financières entre principaux acteurs de la responsabilité civile médicale. La période d'observation de ce dossier se situe sur 2009-2014.

FIGURE 1.1 – Nombre de dossiers relevés par l'ORM entre 2009 et 2014



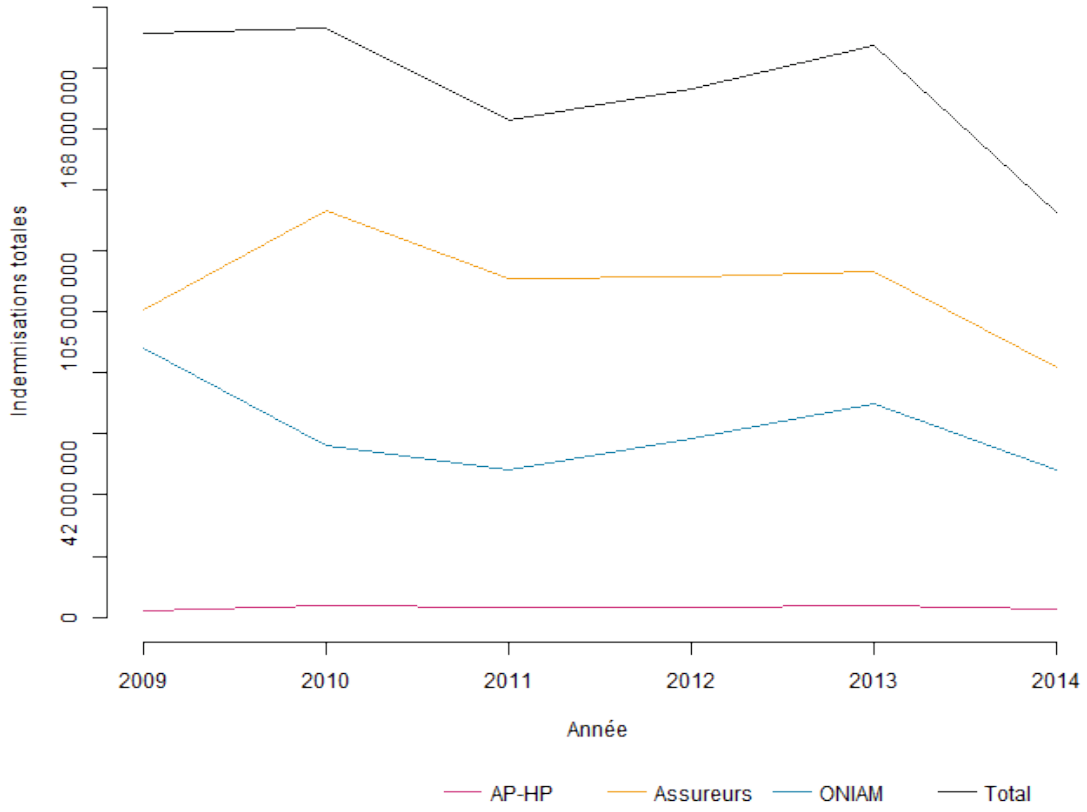
13. [Observatoire des Risques Médicaux, 2015]

Le nombre de dossiers indiqués sur les années 2013 et 2014 ne sont pas interprétables dans la mesure où les sinistres indemnisés tardivement ne sont pas encore connus en 2015. Les sinistres en responsabilité médicale peuvent durer jusqu'à plusieurs années, et les assureurs ont tendance à remonter tardivement de tels dossiers à l'ORM. L'organisme a relevé qu'en moyenne, pour une même année de survenance du sinistre :

- 61% des dossiers sont déposés la même année ;
- près de 75% dans les deux ans ;
- et malheureusement environ 11% de l'ensemble est réclamé plus de cinq ans après.

En se référant au dernier rapport d'activité de l'ONIAM [14], on peut déduire que la hausse dans le temps des dossiers pris en charge par l'ONIAM soulignerait l'amélioration en continu de leurs services rendus aux victimes et de leur collaboration avec les Commissions de Conciliation et d'Indemnisation (CCI) : meilleur suivi des avis émis par les CCI, réduction des délais de traitement des demandes d'indemnisation par les CCI, hausse de l'acceptation des offres amiables proposées par l'ONIAM, *etc.*

FIGURE 1.2 – Indemnisations totales relevées par l'ORM et versées entre 2009 et 2014



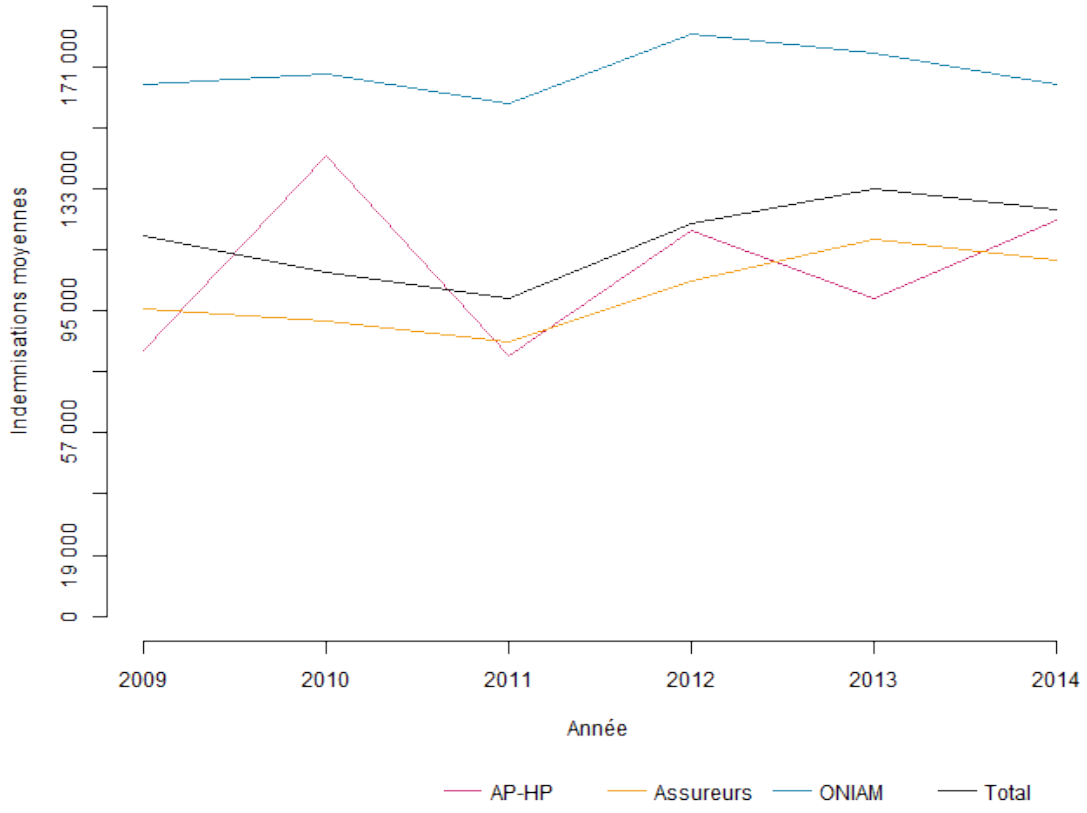
Pour les mêmes raisons que la figure précédente, les indemnisations totales indiquées

14. [Office National d'Indemnisation des Accidents Médicaux, 2018]

sur les années 2013 et 2014 ne sont pas interprétables car les sinistres tardifs de ces années sont inconnus en 2015, année d'élaboration de ces statistiques.

Quoi qu'il en soit, on note que sur toute la période 2009-2014 les assureurs contribuent à 63% au montant global d'indemnisation avec 74% des dossiers du marché. Les dossiers restants sont partagés entre l'ONIAM (36% des indemnisations) et l'Assistance Publique - Hôpitaux de Paris (AP-HP, 2% des indemnisations), seul établissement public de santé français qui a obtenu avec la loi « Kouchner [15] » la compétence pour indemniser à l'amiable et au contentieux les dommages consécutifs à l'activité médicale.

FIGURE 1.3 – Indemnisations moyennes déduites des données fournies par l'ORM, période 2009-2014



En combinant les nombres de dossiers et les indemnisations totales, on obtient les indemnisations moyennes par dossier. Plusieurs remarques :

forte volatilité de l'AP-HP : avec un écart-type deux fois plus élevé que les autres catégories, le faible nombre de dossiers traités annuellement par le centre hospitalier universitaire d'Île-de-France (environ 30 sinistres à indemniser par an) explique la forte variabilité des indemnisations moyennes par année ;

15. [Journal Officiel, 2002b]

hausse des indemnisations moyennes : ce constat est conforté par la hausse du nombre de réclamations relevée *a posteriori* au sein de MACSF-Le Sou Médical et SHAM. Ils enregistrent en effet une hausse de la sinistralité de, respectivement, 8,69% et 1,8% pour l'année 2016 [16] ;

indemnisations moyennes élevées pour l'ONIAM : cette observation laisse entendre que l'ONIAM se chargerait des cas les plus sévères que les assureurs ne seraient probablement pas capables de supporter seuls.

1.2.3 Contributions des assureurs spécialistes de la responsabilité civile médicale

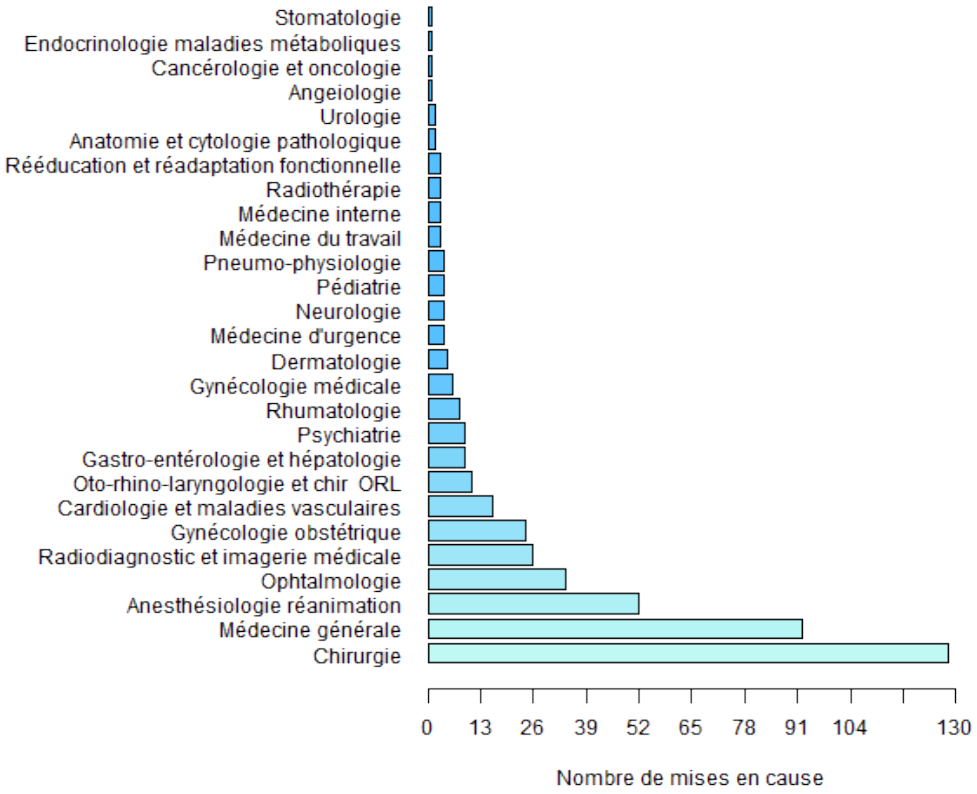
Comme mentionné précédemment, les assureurs spécialistes publient annuellement un bilan de la sinistralité à laquelle ils doivent faire face, indépendamment des tendances du marché.

Sinistralité des professions de santé libérales

Avec 475 633 sociétaires en 2017, dont 145 328 médecins, MACSF-Le Sou Médical relève une sinistralité de 1,02% toutes déclarations confondues, soit une hausse d'environ 3% de sa sinistralité depuis 2016. Ces sinistres représentent un montant total d'indemnités allouées aux victimes par les juridictions civiles de 42 286 422€ [17] .

16. [Banga, 2017]
17. [Mutuelle d'Assurance du Corps de Santé Français - Le Sou Médical, 2017]

FIGURE 1.4 – Nombre de mises en cause par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical



Les nombres présentés ci-dessus se réfèrent à des mises en cause de médecins uniquement et non à des mises en cause sur l'ensemble des professionnels de santé. MACSF-Le Sou Médical souligne à propos dans son rapport d'activité que le taux de médecins condamnés en 2017 était de 45% , en légère diminution mais stable par rapport à 2016 où le taux était de 47% .

La chirurgie figure comme la spécialité la plus exposée au risque. Parmi ces déclarations, les sous-branches de la chirurgie qui ont le plus de sinistralité sont :

- la chirurgie générale, qui comporte notamment les opérations sur l'appendicite et les hernies (38% des dossiers) ;
- la chirurgie orthopédique et traumatologique (32% des dossiers) ;
- et la chirurgie viscérale et digestive (11% des dossiers).

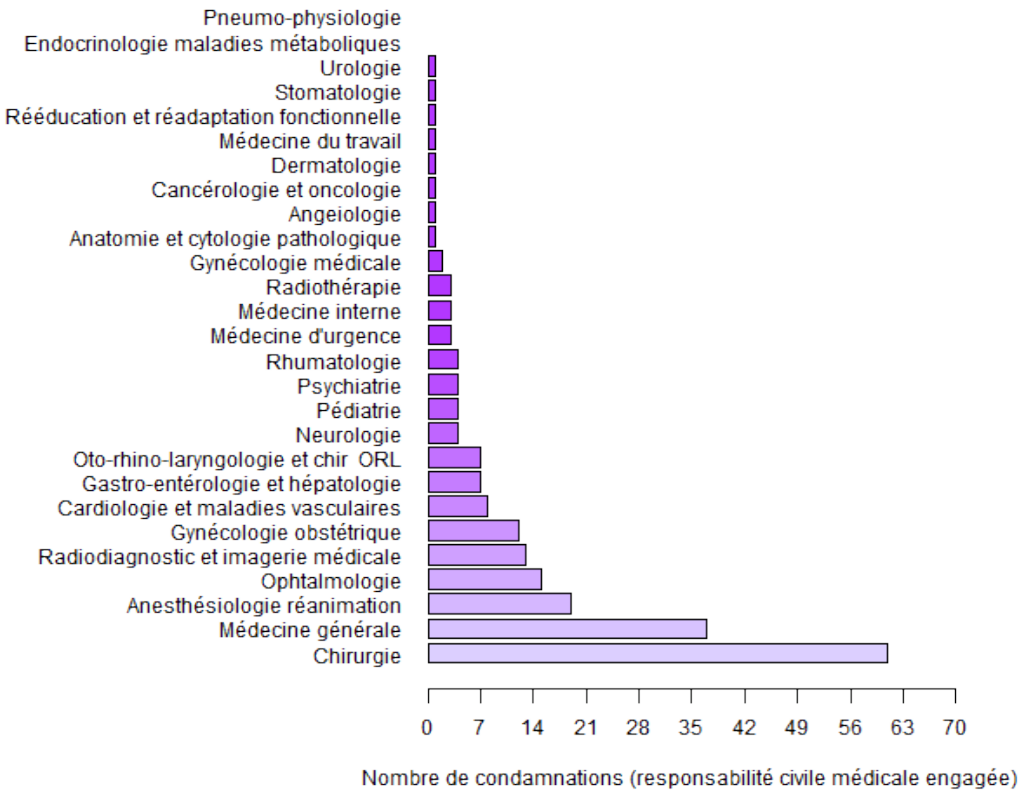
Néanmoins, les mises en cause n'aboutissent pas toutes à des condamnations, élément déclencheur de l'intervention de la responsabilité civile médicale et donc de l'engagement les assurances concernées. En accord avec les données mises à disposition par MACSF-Le Sou Médical, le coefficient de corrélation de Pearson entre le nombre de mises en cause

et le nombre de condamnations par spécialité est très proche de 1 :

$$\text{Corr}_{\text{MACSF}}(\text{mises en cause}, \text{condamnations}) \simeq 0,991 \tag{1.1}$$

Ce résultat montre que le taux de condamnation est, dans sa globalité, indépendant de la spécialité observée puisque le nombre de mises en cause et le nombre de condamnations ont une dépendance très forte d'une spécialité à une autre.

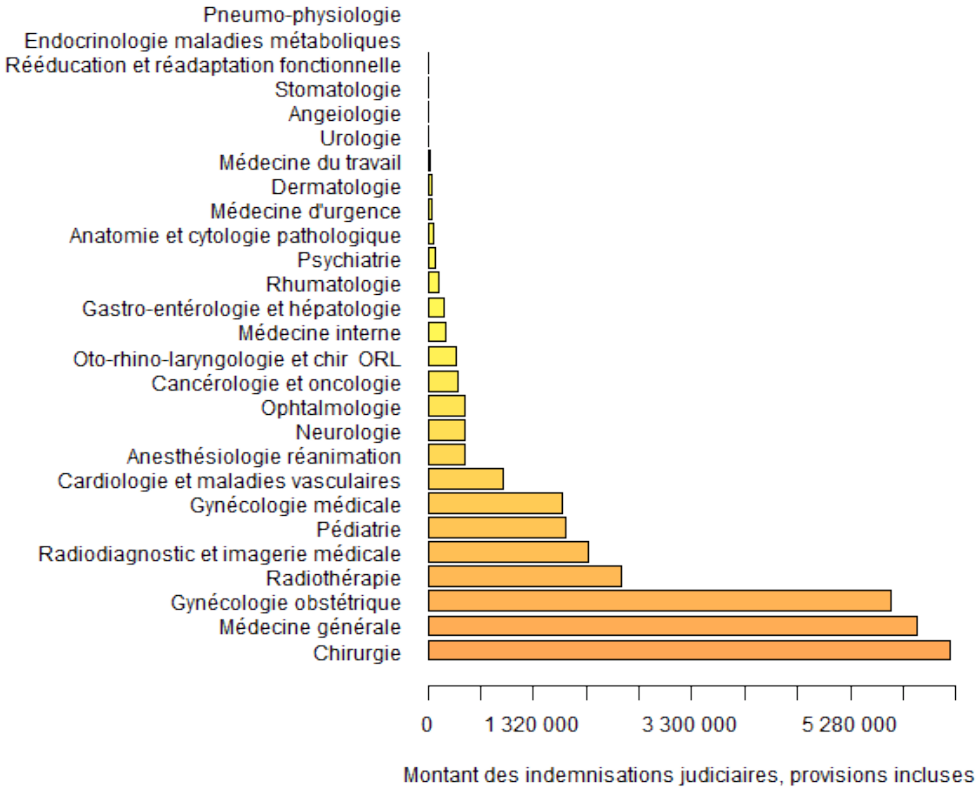
FIGURE 1.5 – Nombre de médecins condamnés par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical



Chirurgie et médecine générale sont les deux spécialités les plus exposées, tout comme les plus mises en cause. Parmi les sous-branches de la chirurgie, les tendances sur les condamnations restent également équivalentes à celles observées avec les mises en cause :

- la chirurgie générale représente 44,3% des condamnations ;
- la chirurgie orthopédique et traumatologique représente 31,1% des condamnations ;
- et la chirurgie viscérale et digestive 13,1% des condamnations.

FIGURE 1.6 – Indemnisations totales mises à charge aux médecins condamnés par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical



Les montants présentés ci-dessous comprennent les indemnités versées et les provisions pour restant à payer.

On note que chirurgie et médecine générale sont les domaines les plus exposés au risque financier, comme observé précédemment en nombre de dossiers (mises en cause ou condamnations). De même, parmi ces indemnisations, les sous-branches de la chirurgie qui sont le plus impactés sont :

- la chirurgie générale (49% des indemnisations) ;
- la chirurgie orthopédique et traumatologique (31% des indemnisations) ;
- et la chirurgie viscérale et digestive (9% des indemnisations).

Au sein de la chirurgie, les classements par nombre de dossiers ou par indemnisations versées au total sont relativement corrélées. Toutefois, cette corrélation ne se vérifie pas sur l'ensemble du reste des spécialités. À titre d'exemple, la gynécologie obstétrique est la sixième spécialité en termes de fréquence mais la troisième en termes de coûts.

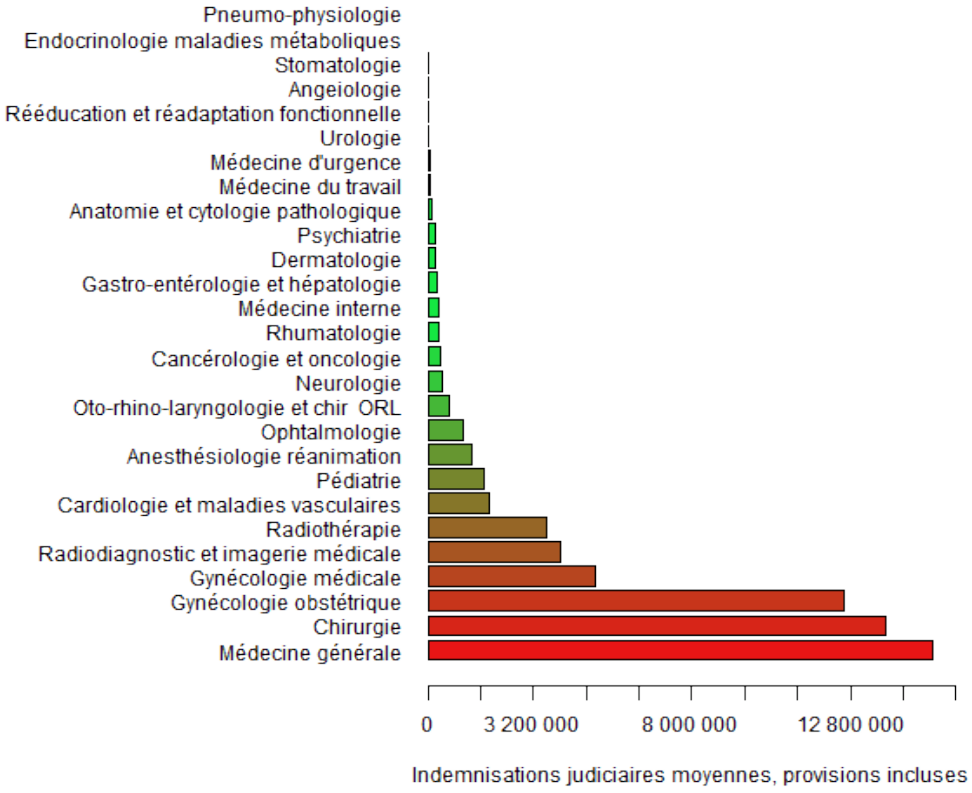
Le coefficient de corrélation de Pearson calculé entre le nombre de mises en cause et le montant total des indemnisations est plus faible que celui entre mises en cause et

condamnations, mais reste toutefois assez élevé :

$$\text{Corr}_{\text{MACSF}}(\text{condamnations}, \text{indemnisations totales}) \simeq 0,795 \tag{1.2}$$

Cette « baisse » dans les corrélations étudiées démontre que les coûts des sinistres sont davantage propres aux spécialités que les fréquences des sinistres.

FIGURE 1.7 – Indemnisations moyennes par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical



En se rapportant à des indemnisations moyennes, le classement dénote que chirurgie, médecine générale, et gynécologie obstétrique sont largement les spécialités qui détiennent les sinistres les plus coûteux.

Il faut cependant garder à l'esprit que les provisions pour restant à payer sont incluses dans ces chiffres, et MACSF-Le Sou Médical explique à juste titre que dans ces domaines les séquelles présentées par les victimes sont généralement graves et avec des répercussions importantes sur leur vie personnelle et professionnelle. De plus, en chirurgie, la responsabilité est bien plus souvent portée à toute l'équipe soignante qu'au chirurgien seul.

Finalement, l'indemnisation moyenne par médecin, toutes spécialités confondues, s'élève à 141 336€ pour l'année 2017. À titre de comparaison, MACSF-Le Sou Médical a relevé que l'indemnisation moyenne par professionnel de santé, médecin ou non, toute spécialité confondue est de 32 628€ sur la même année [18]. Cependant, ces indicateurs sont peu représentatifs de la fréquence et du poids du risque : les sinistres les plus graves peuvent entraîner des charges très lourdes qui traduisent des répercutions significatives et irréparables sur le mode de vie des victimes (perte d'autonomie partielle, totale, voire patient devenu handicapé).

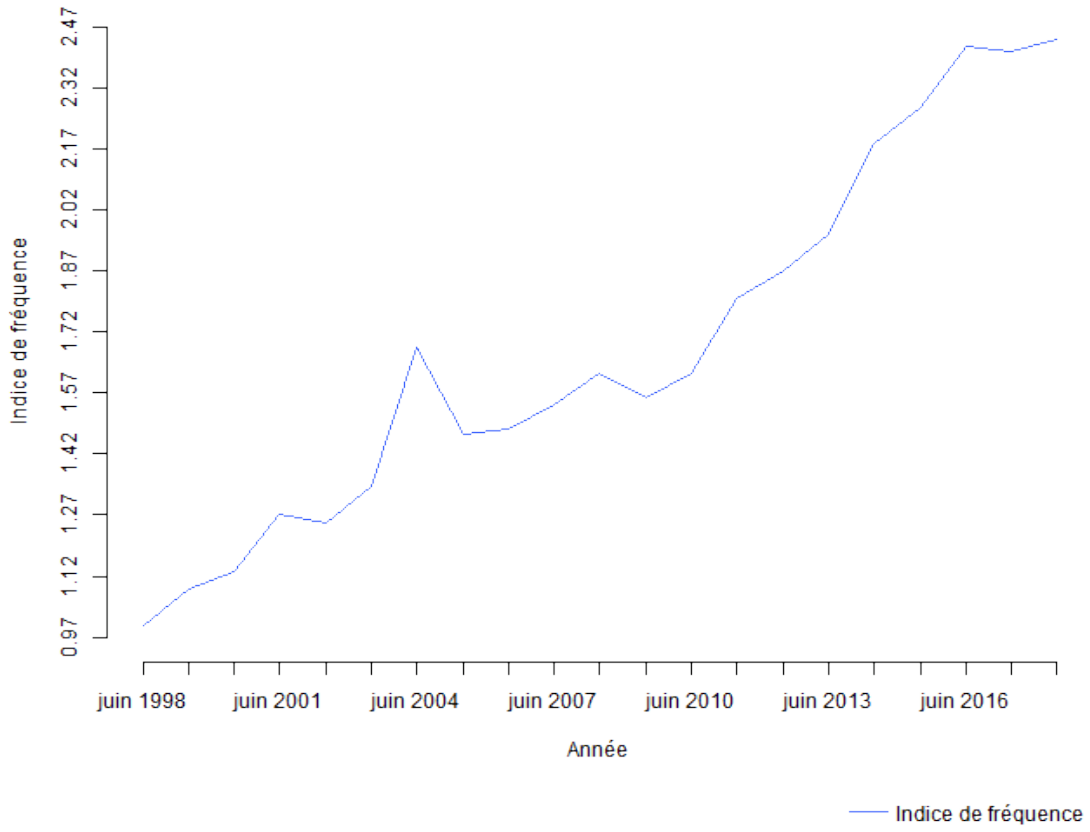
La distance importante constatée entre les coûts moyens des sinistres sur les spécialités à risque et les autres spécialités laisse penser qu'une segmentation des tarifs par spécialité condamne les praticiens les plus exposés au risque à payer des primes beaucoup plus élevées que les autres ; cela s'est effectivement observé dans le passé, sur la période 2004-2009, avec une augmentation des primes émises de 59,6% et une hausse des charges nettes de recours de 31% sur l'ensemble des disciplines les plus à risque (anesthésie-réanimation, gynécologie obstétrique et chirurgie) [19].

Sinistralité des établissements de santé

En 2018, SHAM a diffusé un communiqué de presse qui résume la sinistralité à laquelle cet assureur a dû faire face avec les établissements de santé qu'il assurait sur l'année 2017 [20]. Ce rapport révèle une évolution de la sinistralité en milieu hospitalier qui se stabilise depuis 2016 mais avec une augmentation sur tout l'historique qui est significative : en vingt ans, les sinistres RCM auraient plus que doublé au sein des établissements de santé.

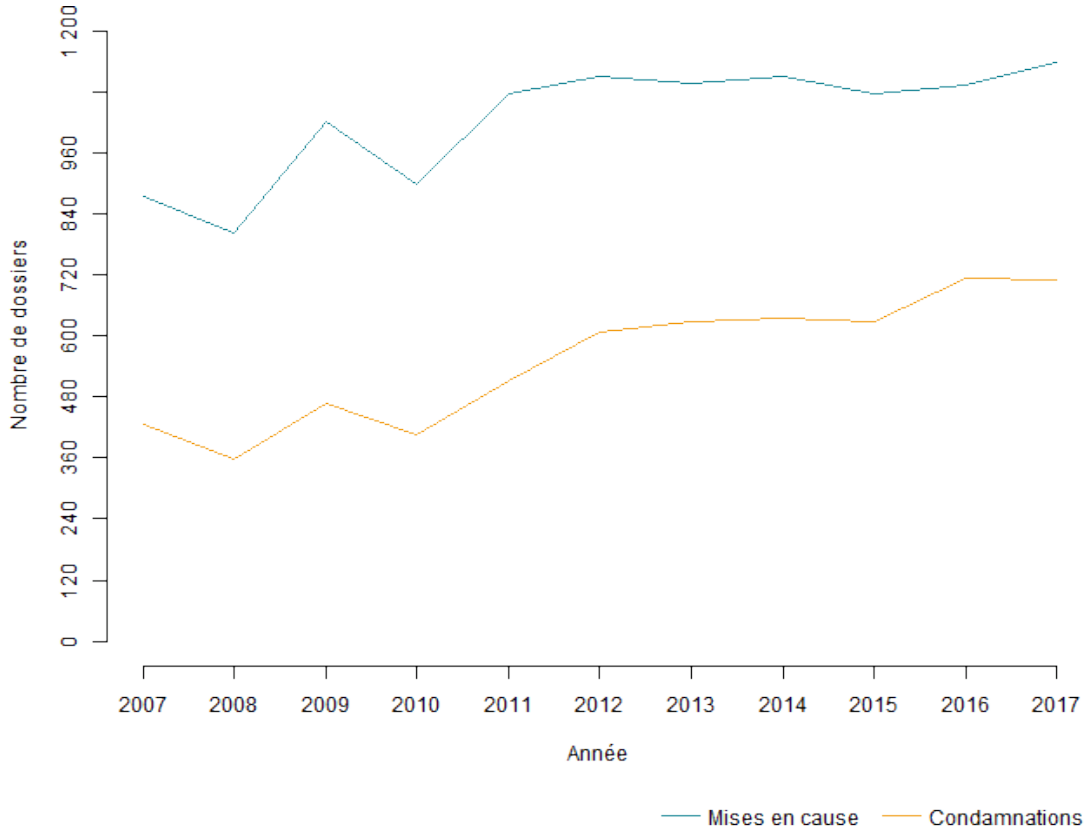
18. [Mutuelle d'Assurance du Corps de Santé Français - Le Sou Médical, 2017]
19. [Johanet, 2011]
20. [Société Hospitalière d'Assurances Mutuelles, 2018]

FIGURE 1.8 – Évolution de la sinistralité en établissement de santé sur la période 1998-2018 relativement à la fréquence renseignée au 30 juin 1998, portefeuille SHAM



Le pic observé en juin 2004 serait probablement lié aux conséquences de la loi « About », où les sinistrés de longue date ont eu la possibilité de réclamer eux-même indemnisation. Avec 11 000 sociétaires et 4 300 structures sociales et médico-sociales assurées par la SHAM en 2017, leur montant total d'indemnités allouées aux victimes par les juridictions civiles s'élève à 221,2M€, dont 219,2M€ dans les établissements de santé.

FIGURE 1.9 – Évolution du nombre de mises en cause et du nombre de condamnations en établissement de santé sur la période 2007-2017, portefeuille SHAM



Le coefficient de corrélation de Pearson entre nombre de mises en cause et nombre de condamnations est également proche de 1 :

$$\text{Corr}_{\text{SHAM}}(\text{mises en cause}, \text{condamnations}) \simeq 0,903 \tag{1.3}$$

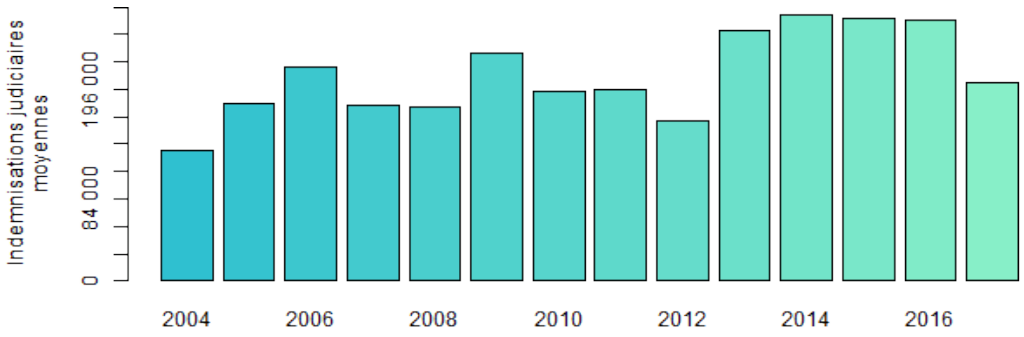
On déduit de ce résultat qu'en milieu hospitalier, la variation entre le nombre de mises en cause et le nombre de condamnations ne présente pas de déformation significative dans le temps, notamment à partir de 2012 et de 2011 respectivement.

La volatilité élevée qui est observée avant 2010, qui s'observe aussi sur le taux de sinistralité, pourrait s'expliquer en partie par le fait que le GTAM, devenu le Groupement Temporaire de la Réassurance Médicale (GTREM), n'est plus alimenté en affaires nouvelles depuis le 1er janvier 2009 [21]. Le marché de la responsabilité civile médicale ayant été considéré comme « assaini », l'absorption des sinistres les plus graves s'est estompée. Toutefois, les variations des nombres de sinistres peuvent aller jusqu'au-delà de 10% du total de dossiers ouverts au cours d'une année, ce qui montre que le retour des sinistres

21. [Association des Professionnels de la Réassurance en France, 2011]

graves dans les portefeuilles, qui sont bien plus rares en nombre, ne peut expliquer à lui seul la différence de variation entre avant 2010et après 2010.

FIGURE 1.10 – Évolution des indemnisations moyennes en établissement de santé sur la période 2007-2017, portefeuille SHAM



En comparaison, les montants moyens d'indemnisations versées par dossier présentés par SHAM soulignent une importante volatilité, positionnés après 2010 ou non. En notant que ces montants n'incluent pas de provisions pour restant à payer et qu'ils ne sont pas cumulés dans le temps, cette volatilité s'explique par le poids important des sinistres lourds, peu nombreux mais longs et coûteux. À titre d'exemple, SHAM a traité :

en 2017 : 19 dossiers dont le coût estimé de la condamnation à verser dans l'année qui suit est supérieur à 1M€, dont 7 en obstétrique ;

en 2016 : 27 dossiers, dont 14 en obstétrique.

En ramenant ces montants moyens au totaux, on observe un coefficient de corrélation plus proche de 1 :

$$\text{Corr}_{\text{SHAM}}(\text{condamnations}, \text{indemnisations totales}) \simeq 0,886 \tag{1.4}$$

Les coûts des sinistres seraient alors relativement corrélés avec les fréquences de sinistres dans le temps, mais leur volatilité causée par les sinistres graves ne garanti pas un provisionnement stable dans le temps.

Des tendances de marché similaires, mais des portefeuilles spécialisés

En conclusion de ces analyses du marché, les évolutions de sinistralité entre les principaux acteurs de la responsabilité civile médicale semblent avoir des tendances dans le temps et au travers des principaux segments parmi les assurés qui sont interprétables : la part des sinistres semble stable, la sinistralité en elle-même est en hausse mais ne présente pas de vagues imprévisibles. En revanche, le coût moyen des sinistres semble très variables aussi bien dans le temps que d'une spécialité à une autre.

Selon le dernier rapport de la mission « Johanet [22] », ces conclusions peuvent être représentatives du marché de la responsabilité civile médicale car celui-ci est caractérisé comme oligopole. En effet, sur l'exercice 2009, les deux premiers opérateurs cumulaient :

- 68% des primes en gynécologie obstétrique ;
- 78% en anesthésie-réanimation ;
- et 83% en chirurgie.

On peut alors supposer que les rapports d'activité de MACSF-Le Sou Médical (dont le groupe se revendique *leader* du marché français avec plus de 950 000 sociétaires en 2018) et de SHAM (qui se revendique premier assureur RCM en France, Espagne et Italie du Nord avec 829M€ de primes collectées en 2017) peuvent être suffisamment représentatifs des tendances du marché *a minima* à l'échelle nationale.

Cependant, la volatilité des coûts reste marquée sur tous les rapports. Elle se justifierait en grande partie par l'impact des sinistres graves.

1.3 Notion de sinistre grave et enjeux

Depuis la modernisation des autorités de contrôle, effectuée dans le cadre de la loi du 1er août 2003 de sécurité financière, le Code des Assurances décrit un sinistre en assurance de responsabilité comme suit [23] :

> « [...] *constitue un sinistre tout dommage ou ensemble de dommages causés à des tiers, engageant la responsabilité de l'assuré, résultant d'un fait dommageable et ayant donné lieu à une ou plusieurs réclamations.* »

En ce sens, d'un point de vue probabiliste, on qualifie un sinistre de grave en responsabilité civile tout sinistre engageant la responsabilité de l'assuré sur une forme de dommage qui possède une très faible probabilité d'occurrence mais qui est relativement conséquente à indemniser par les assureurs.

La particularité des sinistres graves en responsabilité civile médicale est qu'ils sont non seulement coûteux et surtout trop rares pour établir des calculs mathématiquement fiables sur du court terme (en particulier respect des hypothèses sur la loi des grands nombres et du théorème central limite), mais ils sont susceptibles d'être longs : les délais de traitement des demandes d'indemnisation peuvent prendre plusieurs mois [24], et les périodes d'indemnisation des cas graves peuvent s'étendre sur des années voire des décennies.

Les enjeux liés au traitement des sinistres graves reposent alors sur :

- leur identification ;
- leur mode de provisionnement ;
- et la gestion de leur dérive dans le temps.

22. [Johanet, 2011]
23. [Code des assurances, 2003a]
24. [Office National d'Indemnisation des Accidents Médicaux, 2018]

1.3.1 Spécificités de la responsabilité civile médicale

Les évolutions juridiques qu'ont connues les assurances en responsabilité civile médicale sur ces deux dernières décennies ont impacté significativement non seulement les modalités d'indemnisation des sinistres mais aussi leurs modalités de gestion.

Avant la loi « About » : les sinistres en responsabilité civile professionnelle étaient pris en compte en fonction de leur date de survenance. Un sinistre était pris en charge par l'assureur dès lors que le fait générateur du dommage intervenait au cours de la période garantie par le contrat. La prescription civile était de trente ans.

Après l'entrée en vigueur de la loi « About » : les sinistres sont désormais pris en compte selon leur date de réclamation [25], avec reprise du passé inconnu et l'introduction de la garantie dite subséquente. Il est rendu possible de garantir des réclamations qui sont postérieures à l'expiration d'un contrat pour des faits dommageables au moment où ce même contrat était actif.

Ainsi, depuis cette réforme, tout contrat de responsabilité civile professionnelle couvre les sinistres dont la première réclamation est formulée :

- soit pendant la période de validité du contrat d'assurance, quelle que soit la date du fait générateur (principe de la « base réclamation ») ;
- soit, en cas de changement d'assureur avec continuité d'activité, pendant une période de cinq ans à compter de l'expiration ou de la résiliation du contrat d'assurance. En cas de cessation définitive d'activité, ce délai est porté à dix ans.

Il va sans dire que les sinistres antérieurs au contrat, connus de l'assuré lors de la souscription de son contrat, ne sont pas pris en charge par l'assureur. Toutefois, sous le principe de la base réclamation, les sinistres qui sont inconnus et antérieurs à la souscription du contrat peuvent faire l'objet d'une indemnisation : on parle de « sinistre repris du passé ».

À l'opposé de l'échelle de temps, les sinistres qui sont réclamés après la cessation d'activité de l'assuré peuvent aussi faire l'objet d'une indemnisation grâce à la notion de garantie subséquente : on parle de « sinistres subséquent ».

En conséquence de ces évolutions juridiques et des particularités qu'elles procurent, la gestion des contrats RCM fait figure d'exception parmi les modes de fonctionnement des contrats d'assurance privée qui existent par ses mécanismes par répartition. Une segmentation de nature juridique se dessine aux sinistres RCM en fonction de l'état de leur contrat, qui est désormais conditionné par la date de réclamation et non plus la date de survenance du sinistre en question.

25. [Journal Officiel, 2002a]

Ainsi, les modèles de provisionnement propres à la loi « About » sont impactés sur le reclassement des sinistres dans les triangles de règlements à construire avec une division de la base des sinistres en deux segments :

une base de survenance, retenue pour :
- les contrats qui sont clôturés avant la réforme, quelle que soit leur date de réclamation ;
- ou les contrats qui sont renouvelés après la réforme mais avec une date de réclamation qui date d'avant la réforme ;

une base de réclamation, retenue pour les contrats souscrits ou renouvelés après la réforme et avec des dates de réclamation postérieures à la réforme.

De ces modifications majeures dans les modalités de gestion des sinistres, le cycle de vie d'un sinistre RCM se déroule en quatre grandes phases :

1. la survenance, lorsque le fait dommageable a eu lieu ;
2. la déclaration, lorsque le professionnel de santé assuré signale le sinistre à leur assureur ;
3. la réclamation, effectuée par la victime (si l'assuré ne s'est pas manifesté, la déclaration est considérée en même temps que la réclamation) ;
4. et la clôture, lorsque les indemnisations ont été intégralement versées.

Il est à noter qu'une réouverture d'un sinistre reste possible, lorsqu'une victime fait une rechute en lien avec le fait dommageable du sinistre qui a été clos par exemple.

1.3.2 Provisionnement du sinistre RCM : une question de temps

Les sinistres graves RCM sont avec un développement long, et le changement de gestion des bases imputées par les évolutions juridictionnelles de ce milieu (et tout particulièrement la loi « About »), les assureurs se retrouvent face à un historique tronqué à l'année 2003 et à des contraintes juridiques liées au mécanisme de « base réclamation » qui rendent le provisionnement à long terme incertain.

Au niveau social[26], une provision pour sinistres à payer (PSAP) doit être constituée à chaque exercice pour représenter au bilan une valeur estimative des dépenses futures totales auxquelles les assureurs seraient confrontés pour indemniser les sinistres garantis par leurs contrats.

Au niveau prudentiel[27] , la segmentation des provisions à constituer est plus fine : les assureurs qui couvrent de telles garanties se doivent de constituer des provisions qui couvrent à la fois, pour chaque année d'exercice :
- les sinistres ouverts aux indemnisations avant 2003 selon leur date de survenance ;
- les sinistres réclamés dans l'année, indépendamment de leur date de survenance ;
- et les engagements qui sont liés aux garanties subséquentes.

26. [Code des assurances, 2015]
27. [Code des assurances, 2003b]

La provision sur la garantie subséquent doit couvrir les sinistres probables réclamés après résiliation du contrat d'assurance, caractérisée sous trois principaux risques :
- réclamation après le départ à la retraite de l'assuré ;
- réclamation après changement d'activité de l'assuré ;
- réclamation après changement d'assurance par l'assuré.

Au-delà de ces risques, le changement de référentiel au sein des méthodes de provisionnement a généré deux situations qui provoquent un trou de garantie :
- lorsque le fait générateur du dommage est survenu et connu dans la période d'ouverture du contrat d'assurance, mais la première réclamation a lieu au-delà de la période de garantie subséquente de cinq ans ;
- et lorsque le fait générateur du dommage est survenu dans la période d'ouverture du contrat d'assurance, connu ou inconnu, mais la première réclamation a lieu dix ans après la cessation d'activité ou le décès de l'assuré.

Ces contraintes se traduisent finalement par trois types de provisions pour sinistres inconnus :
- les provisions pour les contrats considérés en date de survenance ;
- les provisions pour les contrats considérés en date de réclamation ;
- et les provisions pour la garantie subséquente.

En d'autres termes, le provisionnement repose sur l'état juridique du sinistre, conditionné par la date du fait générateur du dommage, la date de réclamation, la période d'ouverture du contrat d'assurance et la situation actuelle de l'assuré : ce sont les classes de risque.

De manière générale, les principales classes de risque rencontrées en responsabilité civile médicale sont :

fréquence et coût moyen faibles : professions de santé qui ont une faible intervention sur le corps humain comme les hospitaliers, les remplaçants et les paramédicaux ;

fréquence élevée et coût moyen faible : odontologie, médecine vétérinaire ;

fréquence élevée mais coût moyen très variable : médecine générale ;

fréquence et coût moyen relativement élevés : otorhinolaryngologie, stomatologie, radiologie ;

fréquence et coût moyen très élevés : chirurgie obstétrique, anesthésiologie.

Cette classification offre un indicateur de suivi de risque, mais les évènements graves peuvent perturber sa fiabilité. Des analyses portées sur les probabilités de résiliation, la part des sinistres post-résiliation, ou la probabilité de changement peuvent également aider au provisionnement des sinistres RCM. Enfin, une analyse de la dérive dans le temps

des provisions corrigerait convenablement la volatilité des sinistres longs, comme recommandé par le Code des Assurances [28], afin de converger vers une provision « *prudente, fiable et objective* ».

TABLE 1.1 – Synthèse des modalités de provisionnement en responsabilité civile médicale

Groupe de risque	**État du contrat**	**Date de réclamation**	**Base de rattachement**	**Nature juridique**	**Provisions à constituer**
1	Clos ou non renouvelé avant le 02/11/2003	À toute date	Survenance	Survenance	IBNER
2	Renouvelé ou souscrit après le 02/11/2003	Avant la réforme « About »	Réclamation		IBNER, IBNYR, IBNR
3		Après la réforme « About »		Réclamation	

Remarque : le 02/11/2003 correspond à la date d'entrée en vigueur de la loi « About [29] ».

Ainsi, sous ces conditions, le provisionnement des sinistres RCM sont impactés par les sinistres graves pour deux raisons :

- la réévaluation de la charge des sinistres graves connus ;
- et la charge des sinistres qui ne sont pas encore déclarés, du fait de la garantie subséquente.

1.3.3 Périmètre d'étude du mémoire

L'ensemble des études qui sont menées dans ce mémoire s'appuient sur les données d'inventaire d'un principal assureur des professions de santé libérales.

Les données sont vues au 31/03/2019 et comportent l'état des dossiers de sinistres au 31/03/2019 ainsi qu'aux 30 juin et 31 décembre des années 1986 à 2018. L'historique paraît conséquent, mais il faut garder à l'esprit le fait que les sinistres recensés avant 2004 ne reflètent pas la réalité des évolutions juridiques récentes [30].

Les dossiers en question sont relatifs au contrat d'assurance RCM proposé à des praticiens libéraux, toutes spécialisations confondues. Les montants regroupés dans ces

28. [Code des assurances, 2003c]
29. [Journal Officiel, 2002a]
30. [Journal Officiel, 2002a]

données portent sur les préjudices corporels dont les déclinaisons sont établies par un groupe de travail de la Cour de cassation en juillet 2015 [31].

Au niveau du service de gestion des sinistres, les postes de préjudice provisionnés avant la date de consolidation sont les postes de préjudice suivants :

dépenses de santé liées au préjudice : ensemble des frais hospitaliers, médicaux, paramédicaux et pharmaceutiques qui ont été engagés par les organismes sociaux ou restés à la charge de la victime. Les frais qui sont susceptibles d'être exposés par la victime directe sont aussi provisionnés (honoraires médecin-conseil, frais de déplacement, *etc.*) ;

tierce personne temporaire : aide humaine apportée par une tierce personne pour suppléer une perte d'autonomie de la victime ;

perte de gains professionnels actuels : perte de revenus subie par la victime ;

déficit fonctionnel temporaire, partiel ou total : invalidité subie par la victime, dégagée de toute incidence sur sa rémunération professionnelle. Ce poste de préjudice fait plus largement référence à la perte de qualité de vie et aux périodes d'hospitalisation ;

souffrances endurées : souffrances physiques et psychiques (troubles compris) endurées par la victime ;

préjudice esthétique temporaire : altération de l'apparence physique temporaire de la victime.

Les postes de préjudice provisionnés après consolidation sont les suivants :

dépenses de santé futures : ensemble des frais hospitaliers, médicaux, paramédicaux et pharmaceutiques qui sont médicalement prévisibles, répétitifs et rendus nécessaires en vue de l'état pathologique permanent et chronique de la victime (installation de prothèses incluses) ;

frais de véhicule : dépenses nécessaires pour procéder à l'adaptation d'un véhicule aux besoins de la victime (surcoût lié au renouvellement du véhicule inclus) ;

aménagement du logement : frais déboursés par la victime directe pour adapter son logement à son handicap ;

préjudice esthétique permanent : altération de l'apparence physique définitive de la victime ;

déficit fonctionnel permanent : incapacité, constatée médicalement, qui établit que le dommage subi a une incidence sur les fonctions du corps humain (IPP ou AIPP). Elle comporte les atteintes aux fonctions physiologiques, la douleur permanente, la perte de qualité de vie et les troubles ;

tierce personne définitive : aide humaine apportée par une tierce personne pour suppléer la perte d'autonomie de la victime ;

31. [Dintilhac, 2005]

perte de gains professionnels futurs : perte ou diminution des revenus causée par l'incapacité permanente à laquelle la victime est confrontée ;

incidence professionnelle : incidences périphériques du dommage qui impactent la sphère professionnelle de la victime. C'est une provision qui vient compléter le poste de préjudice « perte de gains professionnels futurs » ;

préjudice scolaire, universitaire ou de formation : équivalent de l'incidence professionnelle pour une victime adulte ;

préjudice d'agrément : impossibilité pour la victime de pratiquer régulièrement une activité spécifique sportive ou de loisirs ;

préjudice sexuel : préjudice lié au système génital, l'acte sexuel en lui-même, ou la difficulté voire l'impossibilité à procréer subi par la victime ;

préjudice d'établissement : perte d'espoir ou de possibilité de réaliser un projet de vie familial en raison de la gravité du handicap permanent dont est atteint la victime ;

préjudice permanent exceptionnel : regroupe les préjudices atypiques permanents, hors préjudices patrimoniaux ;

préjudice lié à des pathologies évolutives : concerne les pathologies incurables, en tenant compte de l'angoisse liée à la connaissance de la pathologie ;

autres préjudices : tient compte d'autres postes qui ne sont pas prévus par la nomenclature du rapport « Dintilhac [32] » et qui sont impactants en RCM car reconnus par la jurisprudence, comme le défaut d'information (préjudice d'impréparation, perte de change, *etc.*), le préjudice moral (complète le préjudice des souffrances endurées) ou le préjudice d'angoisse de mort imminente.

Les charges et provisions des sinistres étudiés prennent en compte l'ensemble de ces postes de préjudice, indépendamment de l'activité exercée par l'assuré. Des montants forfaitaires variant entre 1€ (cas des sinistres déclarés par l'assuré mais pas encore réclamés par la victime) et 100 000€ sont fixés à l'ouverture d'un dossier selon la spécialité considérée et la gravité probable que peut représenter le sinistre en question. À l'inverse, en termes de limites, les dommages corporels exceptionnellement graves qui dépassent 8 000 000€ de charges par sinistre et 15 000 000€ de charges par année d'assurance sont exclus des garanties du contrat d'assurance RCM étudié.

Selon le même service de gestion, les sinistres graves sont décelés en moyenne un an et demi après la date de consolidation. Lorsqu'ils font l'objet d'un traitement à l'amiable, les règlements associés sont déclenchés en moyenne trois ans après la date de réclamation du sinistre, dont un an environ pour la mise en accord sur les montants à payer. Enfin, il a été indiqué que les sinistres qui se concluent sur un accord à l'amiable sont plus nombreux que les sinistres qui font l'objet d'une procédure judiciaire, mais le suivi des règlements en fonction de leur forme, du nombre et de la durée des procédures est difficilement traçable.

32. [Dintilhac, 2005]

En somme, les sinistres et leurs indemnisations qui sont considérées regroupent les praticiens libéraux qui font face à des dommages corporels où leur responsabilité est mise en cause. L'indemnisation des assurés se fait tardivement, en partie du fait de la durée de traitement des dossiers, et cette contrainte exige la mise en place de modèles d'estimation pour maîtriser cet aspect incertain de la responsabilité civile médicale.

Chapitre 2

Modalités de définition d'un sinistre grave

2.1 Construction de la base de données

L'ensemble des traitements et calculs effectués dans le cadre de ce mémoire seront appliqués en langage R. Les données utilisées seront importées et traitées en trois temps :

- importation sous R des bases de données initiales ;
- agrégations, jointures, et retraitements des données ;
- analyses descriptives des bases de données finales obtenues.

2.1.1 Importation et mise en forme des bases initiales

Les bases de données reçues sont issues de plusieurs extractions de la branche « responsabilité civile professionnelle » :

- pour la date d'arrêté 31/03/2019, trois tables utilisées pour des analyses techniques sont extraites : table des recours, table des règlements, et table des provisions de règlements ;
- pour les années antérieures, une table scindée en deux fichiers est utilisée et recense tous les sinistres en cours avec un pas semestriel.

TABLE 2.1 – Volumétrie des bases initiales utilisées

Nom de la base	Base sinistres historiques, partie 1	Base sinistres historiques, partie 2	Base règlements	Base provisions	Base recours
Nombre d'observations	900 000	975 974	138 100	293 192	1 955
Nombre de variables	22		13		
Date(s) d'arrêté	Tous les 30/06 et 31/12 entre le 30/06/1986 et le 30/06/2018		31/03/2019		

Pour chaque base initiale, une ligne correspond à une date d'opération. Un sinistre développé sur plusieurs semestres de flux aura une ligne par semestre. Les descriptions des bases importées sont disponibles en annexe B.

2.1.2 Agrégations et jointures

Cette première phase vise à rendre les bases de données « R-compatibles » et exploitables pour la construction des triangles de développement.

Étape 1 : import des bases et sélection des variables

Après importation des fichiers, la première chose à faire consiste à sélectionner les seules variables qui sont communes à toutes les bases et qui seront utilisées tout au long de cette étude. Les variables qui nous intéressent dans ce cadre sont :

- la date d'arrêté, point de repère temporel qui sert à l'observation de l'évolution des dossiers dans le temps ;
- les dates de survenance, de réclamation, de « flux », et de clôture des sinistres pour l'élaboration des triangles de développement ;
- l'assignation des dossiers en base « sinistre » ou « réclamation », établie par le service de gestion ;
- et les montants de règlements, provisions, et recours.

TABLE 2.2 – Volumétrie des variables sélectionnées

Nom de la base	Base sinistres historiques, partie 1	Base sinistres historiques, partie 2	Base règlements	Base provisions	Base recours
Nombre de variables sélectionnées	8 (-14)		9 (-4)		

Étape 2 : mise en forme des données

Une fois cette sélection effectuée, le contenu des bases allégées doit être disposé sur un même format exploitable par R : une adaptation le plus en amont possible des formats nombres, chaînes de caractères et dates limite au maximum le risque de perte d'information ou d'incompatibilité dans les calculs qui seront effectués par la suite.

Étape 3 : regroupement des bases homogènes

Après la mise en forme des bases, les bases « sinistres historiques » sont concaténées pour former une unique base.

TABLE 2.3 – Volumétrie de bases concaténées

Nom de la base	Base sinistres historiques, partie 1	Base sinistres historiques, partie 2	Base concaténée « sinistres historiques »
Nombre d'observations	900 000	975 974	1 875 974

Il est à noter que les trois autres bases seront regroupées dans un second temps car elles sont disposées sur des périmètres différents qui nécessitent des jointures au cas par

cas.

Étape 4 : suppression des dates aberrantes

On suppose que si les dates sélectionnées (survenance, réclamation, flux et clôture) sont au-delà de la date d'arrêté considérée, elles n'ont alors pas de sens car ces dates ne sont pas censé être des dates prédictives : on « voit » les bases en date d'arrêté, pas au-delà. Les dates aberrantes sont remplacées par « NA », sens R de la valeur manquante.

TABLE 2.4 – Volumétrie des dates aberrantes retraitées

Nom de la base	Base sinistres historiques	Base règlements	Base provisions	Base recours
Dates de survenance aberrantes	0	0	0	0
Dates de réclamation aberrantes	1 197 163 (63,82%)	0	0	0
Dates de flux aberrantes	*Pas de dates de flux*	0	0	0
Dates de clôture aberrantes	*Pas de dates de clôture*	31 432 (22,76%)	49 346 (16,83%)	476 (24,35%)

Les dates aberrantes correspondent, pour la totalité d'entre elles sur les bases règlements, provisions, recours et pour la quasi-totalité sur la base sinistres historiques, à des dates par défaut lorsque l'information est manquante (31/12/9999 ou semestre 16 100). Sur cet aspect, les données à disposition sont d'une grande qualité.

Étape 5 : agrégation des opérations de gestion

Dans un semestre, plusieurs opérations comptables peuvent avoir été menées et chacune d'entre elles représente une ligne dans les bases de données. Les montants seront donc sommés et regroupés par numéro de sinistre, date d'arrêté et semestre de flux.

TABLE 2.5 – Volumétrie des agrégations sur opérations

Nom de la base	Base sinistres historiques	Base règlements	Base provisions	Base recours
Nombre d'observations avant agrégation des opérations	1 875 974	138 100	293 192	1 955
Nombre d'observations après agrégation des opérations	1 875 921	99 422	214 871	1 474
Évolution	-53 (0,003%)	-38 678 (28,01%)	-78 321 (26,71%)	-481 (24,60%)

La proportion d'opérations agrégées est relativement stable entre règlements, provisions et recours. Quant aux données sinistres historiques, la différence de proportions retraitées est justifiée par son extraction du système d'informations qui suit une procédure différente et moins détaillée que les autres bases (pas de dates d'opération renseignées, les montants sont déjà agrégés par arrêté).

Étape 6 : jointures, calculs semestriels et décomposition des bases par année d'arrêté

Les bases règlements, provisions et recours sont jointes par le couple :

$$\text{numéro de sinistre} \times \text{survenance} \times \text{rélamation} \times \text{date d'arrêté}$$

et sans élimination de lignes. Par exemple, un sinistre présent dans la base règlements mais pas dans la base recours reste dans la nouvelle base mais avec un montant de recours renseigné à « NA » (conséquence d'une jointure complète sur R).

Une fois cette jointure effectuée, les variables suivantes sont rajoutées dans un souci d'harmoniser les informations à disposition :

- la date d'arrêté « 31/03/2019 » est indiquée dans la base jointe, les bases qui la précèdent ayant été extraites à cette même date mais sans variable qui l'indique ;
- les numéros de semestre de survenance et de réclamation sont calculés dans les bases sinistres historiques pour être en phase avec les semestres indiqués dans la base jointe ;
- à l'aide des semestres de flux, les semestres de développement sont calculés dans la base jointe :

$$\text{semestre de développement}(i) = \text{semestre de flux}(i) - \text{semestre initial} \quad (2.1)$$

Le « semestre initial » peut correspondre au semestre de survenance ou au semestre de réclamation.

La nouvelle base et la base sinistres historiques sont ensuite distinctes par date d'arrêté pour visualiser l'évolution du nombre de dossiers dans le temps.

TABLE 2.6 – Volumétrie des bases jointes puis scindées par année d'arrêté

Nom de la base	Base sinistres historiques		Base jointe « finale »
Nombre d'observations après jointure complète et découpage par année d'arrêté	1986 : 4 819	1987 : 6 505	2019 : 215 377
	1988 : 8 274	1989 : 10 090	
	1990 : 12 235	1991 : 14 464	
	1992 : 17 031	1993 : 19 774	
	1994 : 22 927	1995 : 26 090	
	1996 : 29 315	1997 : 32 824	
	1998 : 36 407	1999 : 39 974	
	2000 : 43 559	2001 : 47 087	
	2002 : 50 571	2003 : 54 248	
	2004 : 58 697	2005 : 63 376	
	2006 : 68 163	2007 : 73 252	
	2008 : 78 800	2009 : 84 222	
	2010 : 89 796	2011 : 95 778	
	2012 : 102 068	2013 : 108 445	
	2014 : 115 292	2015 : 122 714	
	2016 : 129 940	2017 : 137 535	
	2018 : 71 649		
Nombre de variables	10 (+2)		12 (+3)

On remarque une irrégularité sur l'année 2018. Celle-ci s'explique par le fait que les arrêtés sont semestriels et que nous ne disposons pas des données arrêtées au 31/12/2018. Les autres bases, de 1987 à 2017, contiennent deux arrêts (au 30/06 et au 31/12). On note également que le volume de la base jointe est bien plus important que la tendance affichée par les bases sinistres historiques. Ceci s'explique par le fait que les sinistres renseignés dans la base jointe sont décomposés par semestre de flux depuis l'origine, tandis que les bases historiques fournissent une « image » de tous les sinistres en date d'arrêté, ouverts ou clos et indépendamment des semestres de développement antérieurs.

Une étude de cohérence couplée à quelques ajustements permettront de rendre l'ensemble des bases construites jusqu'ici comparables et *in fine* prêtes pour les travaux qui suivront.

2.1.3 Étude de cohérence et finalisation

Cette seconde phase cherche à identifier les incohérences dans les bases de données obtenues puis faire des retraitements et corrections lorsque les tests effectués montrent des irrégularités.

Construction d'une base de référence pour analyser les incohérences

À partir de la base finale, une base agrégée par numéro de sinistre est constituée. Pour chaque numéro de sinistre, elle renseigne :

- le nombre de semestres de flux recensés ;
- les premier et dernier semestre de flux ;
- les nombres de semestres de survenance, de réclamation, et de clôture uniques ;
- et les montants totaux de règlements, provisions et recours encaissés.

Avec cette table, les tests de cohérence suivants sont appliqués :

- unicité du semestre de survenance, de réclamation et de clôture (si existence) pour chaque numéro de sinistre ;
- montants totaux supérieurs ou égaux à 10^{-3} [1] ;
- continuité des semestres de flux dans la période d'indemnisation, où on doit s'attendre à avoir la relation suivante :

$$\begin{aligned}\text{nombre de semestres de flux avec un règlement} > 0 = \\ &\max(\text{semestres de flux avec un règlement} > 0) \\ &- \min(\text{semestres de flux avec un règlement} > 0) \\ &+ 1\end{aligned} \tag{2.2}$$

Ce calcul utilisé pour la vérification de la continuité des semestres est adapté aux règles de gestion propres à l'alimentation des bases de données tout au long de la vie d'un sinistre :

1. lorsqu'un sinistre est ouvert, c'est-à-dire déclaré, survenu ou réclamé en fonction de son positionnement par rapport à la loi « About », une ligne est ajoutée dans la base de données avec une première constitution de provisions propres au sinistre ;
2. et lorsqu'un sinistre est considéré par le service de gestion comme clos, une ligne est ajoutée pour défaire les provisions restantes.

En ce sens, si aucune opération comptable est effectuée, comme le versement des règlements, alors aucune ligne ne sera ajoutée dans la base de données. Étudier la continuité des semestres de flux entre leur minimum et leur maximum au sens large peut alors être source de faux positifs.

Quelques exemples de modes d'alimentation de la base en fonction du type de sinistre rencontré :

1. Le zéro exact n'a pas été pris en compte pour limiter les erreurs d'arrondi liées à la machine, générées suite aux agrégations précédemment appliquées.

lorsqu'un sinistre est déclaré mais n'aboutit pas à une condamnation : une première ligne est ajoutée dans la base au moment de la déclaration avec un montant de provisions forfaitaire. Aucun versement n'est effectué le temps de la procédure judiciaire. Dès lors que l'assureur est informé de la non-condamnation de son assuré, une seconde ligne est rajoutée pour restituer les provisions préalablement constituées ;

lorsqu'un sinistre est déclaré puis aboutit à une condamnation : une première ligne est ajoutée dans la base au moment de la déclaration avec un montant de provisions forfaitaire. Au moment de la condamnation, une ligne de plus est rajoutée chaque semestre avec :
- le montant des règlements versés ;
- les éventuels recours prélevés ;
- et les provisions recalculées en fonction des circonstances.

Il se peut que le sinistre soit clos par le service au-delà de la fin des indemnisations, ce qui génère une ligne supplémentaire au semestre de clôture avec la restitution des provisions indiquées ;

lorsqu'un sinistre est réclamé avec condamnation civile à l'appui : la procédure est identique à la précédente, à l'exception que les règlements sont versés dès le premier semestre.

Sous ces conditions, la période d'indemnisation est à distinguer de la période totale « de flux » que peut présenter un sinistre dans les bases de données.

TABLE 2.7 – Volumétrie des études de cohérence

Nom de la base	Base finale
Nombre de sinistres	74 986
Plusieurs semestres de survenance	0
Plusieurs semestres de réclamation	0
Plusieurs semestres de clôture	0
Somme des règlements strictement négative	2 (0,003%)
Somme des provisions strictement négative	59 (0,079%)
Somme des recours strictement négative	1 (0,001%)
Sinistres avec semestres de flux discontinus	10 854 (14,47%)

Les sinistres avec semestres discontinus peuvent correspondre à des dossiers qui ne requièrent que des indemnisations ponctuelles. Leur faible proportion, avec les autres résultats de l'étude de cohérence menée, montre une base de données de bonne qualité.

Retraitements à l'issue de l'étude de cohérences

Conditionnés par les résultats précédemment obtenus, les retraitements suivants sont alors effectués sur la base finale :

- les montants renseignés « NA », montants manquants suite aux jointures des bases de règlements, provisions et recours, sont remplacés par zéro ;
- les sinistres qui ont un montant cumulé total négatif (individuellement sur règlements, provisions, ou recours), sont corrigés sur le dernier semestre de flux corrigé. La quasi-totalité de ces erreurs sont causées par des erreurs d'arrondi.

TABLE 2.8 – Volumétrie des retraitements à l'issue de l'étude de cohérences

Nom de la base	Base finale		
Montant	Règlements	Provisions	Recours
Montants « NA » corrigés	115 955 (53,84%)	506 (0,23%)	213 903 (99,32%)
Montants négatifs corrigés	2 (0,001%)	59 (0,03%)	1 (0,0005%)

Les retraitements suivants préparent le terrain pour la construction des triangles de développement, et donc les calculs de provisionnement :

1. calcul des charges par sinistre :

$$\text{charges} = \text{règlements} + \text{provisions pour règlements} \tag{2.3}$$

2. calcul des coûts par sinistre :

$$\text{coûts} = \text{charges} - \text{recours} - \text{provisions pour recours} \tag{2.4}$$

3. calcul des années en fonction des semestres ;

$$\text{année} = \left\lceil \text{année}(01/07/1949) + \frac{\text{semestre}}{2} \right\rceil \tag{2.5}$$

4. déduction des années de développement :

$$\text{année de développement}(i) = \left\lceil \frac{\text{semestre de développement}(i)}{2} \right\rceil \tag{2.6}$$

5. cumul des montants dans le temps par numéro de sinistre et montants cumulés.

TABLE 2.9 – Volumétrie des bases après retraitements

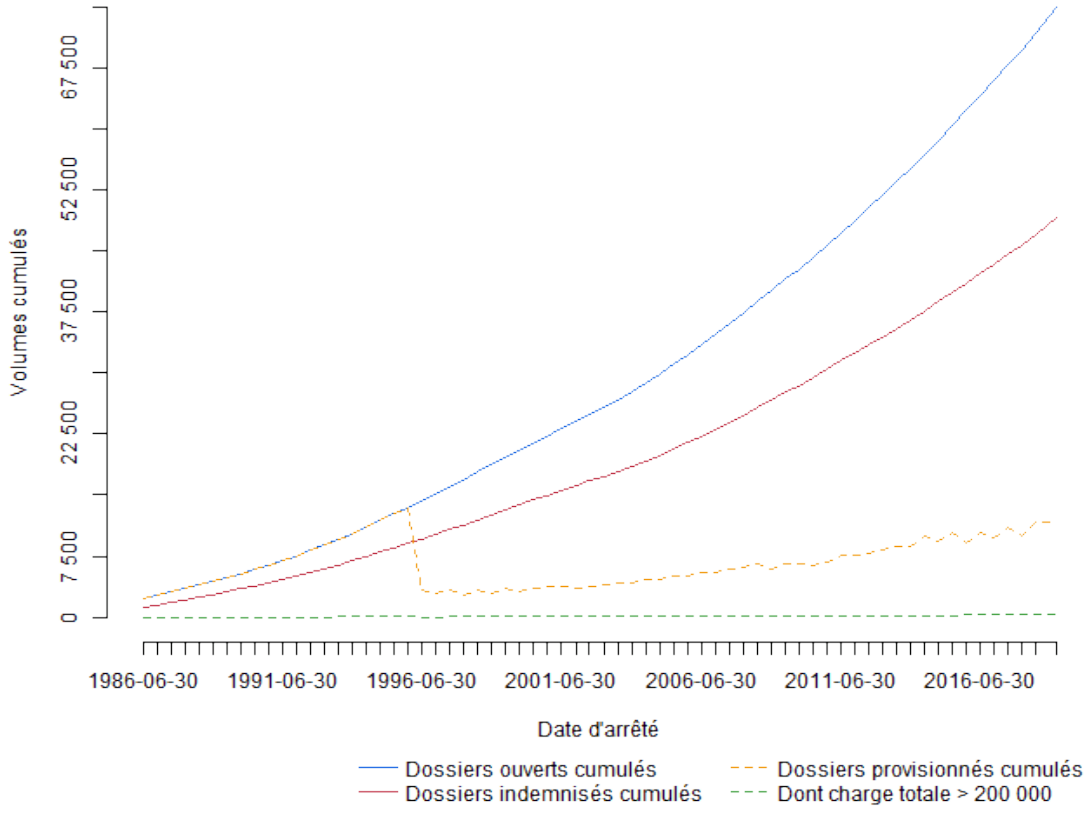
Nom de la base	Base sinistres historiques « retraitées »	Base finale « retraitée »
Nombre d'observations après retraitements	*cf.* tableau 2.6	
Nombre de variables après retraitement	20 (+10)	25 (+13)

Une synthèse descriptive des bases obtenues est disponible en annexe C.

Analyse descriptive des données retraitées

Les bases de sinistres « historiques » permettent d'observer les évolutions des grandeurs de la base dans le temps.

FIGURE 2.1 – Volumétrie du portefeuille



La mission « Johannet [2] » explique que l'ACP déterminait en 2009 les sinistres graves comme étant ceux qui ont une charge totale, nette de recours, qui dépasse 200 000€.

De ce graphique (figure 2.1), on observe :
- une progression stable des dossiers cumulés dans le temps, quelle que soit leur nature (indemnisé, « grave » au sens ACP ou non) ;
- et une « cassure » sur le cumul des dossiers provisionnés en 1996. Celle-ci s'explique par un probable changement de méthode de provisionnement des sinistres déclarés (tables réglementaires, règles de gestion, *etc.*).

Malgré ces volumes cumulés vraisemblablement constants dans le temps, la proportion de sinistres à indemniser sur l'ensemble des sinistres déclarés est constatée en hausse :

TABLE 2.10 – Proportion des dossiers en fonction des charges et règlements

Date d'arrêté	30/06/1986	...	30/06/2002	...	30/06/2018	31/03/2019
Part des dossiers indemnisés parmi tous les dossiers (règlement cumulé $>$ 0)	55,28%	...	67,29%	...	65,53%	65,54%
Part des dossiers indemnisés « grave » parmi les dossiers indemnisés (charge cumulée $>$ 200 000)	0,65%	...	0,45%	...	0,70%	0,68%

Le nombre de dossiers graves cumulés dans le temps peut être observé à la baisse car cette baisse-ci est une conséquence des corrections imputées sur les dossiers graves sur-provisionnés. À titre d'exemple, un cas grave qui requiert le versement de rentes sur du long terme aura d'importantes provisions constituées sur son dossier, ce qui fait gonfler la charge totale et donc dépasser le seuil. Si la victime présumée ne peut plus prétendre aux indemnisations initialement établies de façon prématurée (décès, *etc.*), les provisions restantes seront défaites, opération qui est susceptible de ramener la charge totale en-dessous du seuil l'arrêté suivant.

L'allure des courbes obtenues laisse supposer que la tendance la plus pertinente observée sur l'évolution du nombre de dossiers serait polynomiale de degré 2 de la forme suivante :

$$\begin{aligned} \hat{f} : \mathbb{N} &\to \mathbb{R} \\ x &\mapsto \alpha \cdot x^2 + \beta \cdot x + \gamma \end{aligned} \tag{2.7}$$

Où x représente le nombre de jours cumulés depuis le 01/01/1900 (cette date étant le jour 1), et α, β et γ sont des coefficients obtenus par la méthode des moindres carrés.

2. [Johanet, 2011]

TABLE 2.11 – Régressions polynomiales de degré 2 sur le volume des dossiers en fonction de leur type

Régression sur...	Nombre de dossiers ouverts	Nombre de dossiers indemnisés	Nombre de dossiers provisionnés	Nombre de dossiers graves
α	$\approx 0,000355$	$\approx 0,000220$	$\approx 0,000090$	$\approx 0,000003$
β	$\approx -20,731182$	$\approx -12,633158$	$\approx -6,045258$	$\approx -0,178818$
γ	$\approx 303\ 788$	$\approx 181\ 110$	$\approx 103\ 483$	$\approx 2\ 974$
Coefficient de détermination	$\approx 0,9994$	$\approx 0,9995$	$\approx 0,9799$	$\approx 0,9867$
Coefficient de corrélation	$\approx 0,9997$	$\approx 0,9997$	$\approx 0,9899$	$\approx 0,9933$

Remarque : la régression polynomiale du nombre de dossiers provisionnés a été mesurée sur les données à partir de 1996 pour ne pas inclure l'effet du changement de méthode de provisionnement. Les coefficients ont été calculés sur cette même période.

TABLE 2.12 – Projections polynomiales des volumes du portefeuille

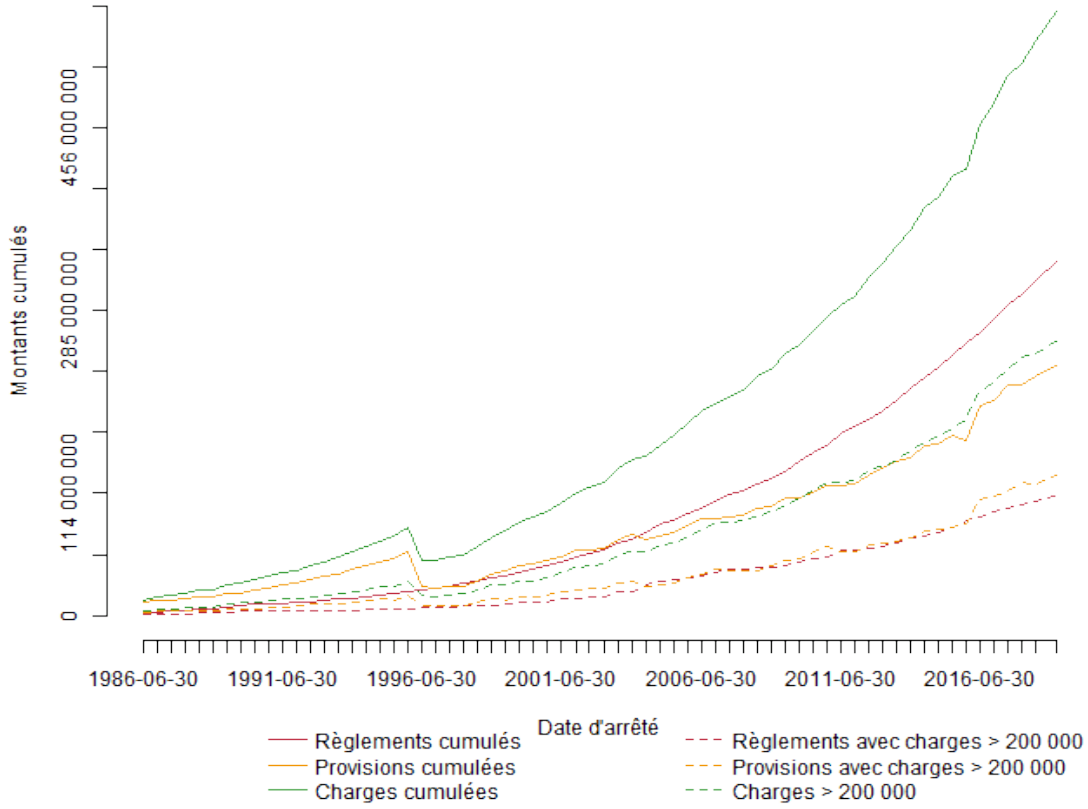
Horizon de projection	31/03/2019 (données actuelles)	31/12/2019 (fin de l'année)	31/12/2029 (+10 ans)	31/12/2039 (+20 ans)
Nombre de dossiers ouverts	74 986	76 429	118 994	171 008
Évolution	-	+1,92%	+58,69%	+128,05%
Nombre de dossiers indemnisés	49 143	50 148	77 404	110 523
Évolution	-	+2,05%	+57,51%	+124,90%
Nombre de dossiers provisionnés	11 762	12 373	20 477	30 993
Évolution	-	+5,19%	+74,09%	+163,50%
Nombre de dossiers graves	332	337	587	908
Évolution	-	+1,45%	+76,66%	+173,61%

Selon ces projections, les sinistres graves auraient tendance à évoluer plus rapidement que l'ensemble des sinistres à indemniser. En d'autres termes, la part de sinistres graves augmenterait, *i.e.* qui auraient une charge $> 200\ 000$€, augmenterait avec le temps. Cette idée pourrait être cohérente dans le sens où le seuil utilisé jusqu'ici ne varie pas dans le temps, donc ne prend pas en compte des paramètres influents sur la sinistralité à long terme comme l'inflation. Ce sujet sera abordé à la section 2.3.1.

Les régressions polynomiales obtenues donnent ainsi une idée de l'évolution future du portefeuille, mais il faut garder à l'esprit la locution latine « *Cum hoc ergo propter hoc* » (avec ceci, donc à cause de ceci) : ce n'est pas parce que les coefficients de détermination obtenus sont très satisfaisants que les modèles de régression établis seront pertinents dans la durée. Corrélation ne signifie pas causalité.

Il aurait été intéressant d'analyser la fréquence des sinistres, mais n'ayant pas connaissance du nombre d'adhérents total dans le temps, il n'est pas possible de poursuivre cette réflexion sur la projection de la sinistralité du portefeuille étudié.

FIGURE 2.2 – Montants cumulés du portefeuille



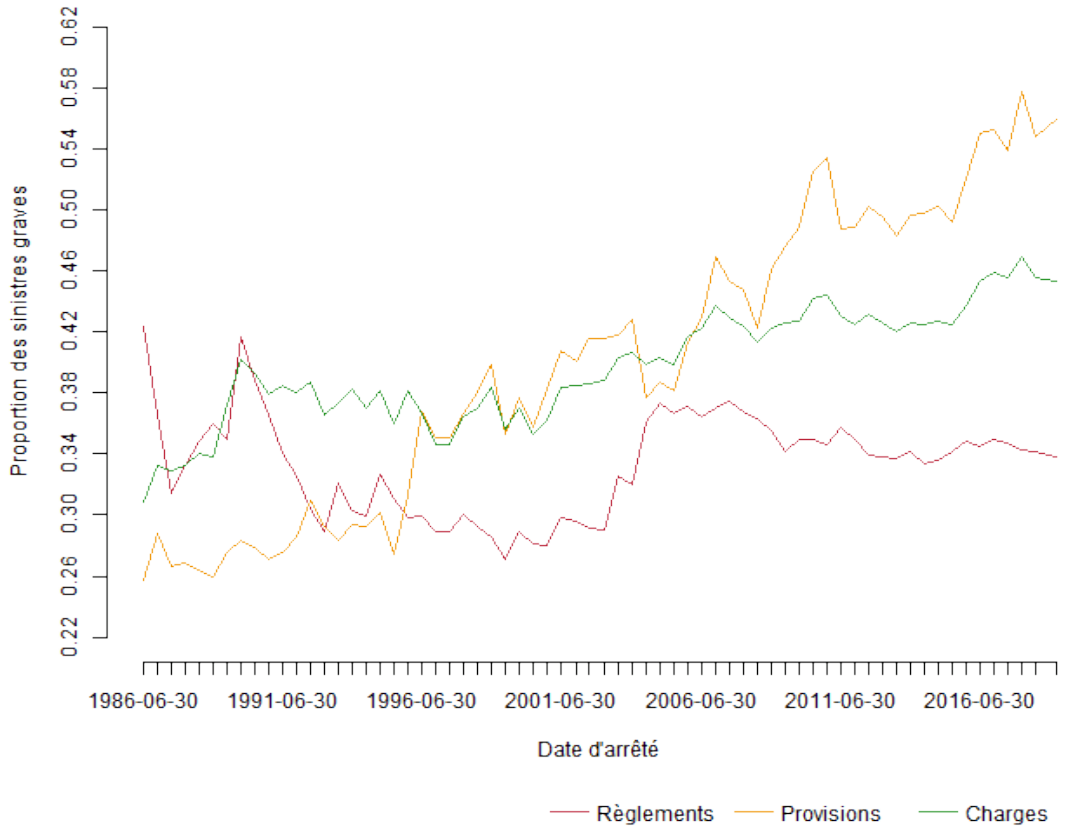
On retrouve la chute qui était observée sur les nombres de dossiers provisionnés sur les montants de provisions cumulés dans le temps, avec en amont un éloignement conséquent

des provisions cumulées par rapport aux règlements cumulés qui est observé. Ces analyses renforcent l'hypothèse d'un changement de gestion des provisions qui aurait été effectué cette année-là dans un but de limiter le risque de sur-provisionnement.

Il est à noter que cet impact ne concerne que les provisions et non les règlements, constatés ici aussi avec une progression quasi-linéaire dans le temps. Les charges sont quant à elles impactées parce qu'elles contiennent les provisions.

Concernant les sinistres graves, considérés comme ayant une charge totale qui dépasse 200 000€, on note qu'ils représentent en moyenne 40% des charges cumulées mais avec des évolutions par composante de la charge qui sont volatiles et hétérogènes.

FIGURE 2.3 – Proportion des « graves » dans le portefeuille en fonction du montant



La part des sinistres graves sur le total au niveau des charges a une progression à la hausse dans le temps, tandis que les règlements sont stables dans le temps et les provisions augmentent. La forte volatilité des montants de règlements sur les premières années s'explique par le très faible volume de sinistres graves présent dans le portefeuille. Présenter les tendances du portefeuille en « cumul » cherche à réduire cette volatilité avec le temps.

En appliquant les mêmes méthodes de régression que sur les volumes de dossiers, on obtient les coefficients suivants :

TABLE 2.13 – Régressions polynomiales de degré 2 sur les montants cumulés en fonction de leur type

Régression sur...	Montants cumulés totaux			Dont charge totale > 200 000		
	Règlements	Provisions	Charges	Règlements	Provisions	Charges
α	$\approx 2,941$	$\approx 1,705$	$\approx 4,645$	$\approx 1,024$	$\approx 1,164$	$\approx 2,301$
β	≈ -195k	≈ -112k	≈ -307k	≈ -68k	≈ -78k	≈ -154k
γ	$\approx 3\ 256$M	$\approx 1\ 859$M	$\approx 5\ 115$M	$\approx 1\ 134$M	$\approx 1\ 322$M	$\approx 2\ 591$M
Coefficient de détermination	$\approx 0,9976$	$\approx 0,9749$	$\approx 0,9928$	$\approx 0,9961$	$\approx 0,9766$	$\approx 0,9883$
Coefficient de corrélation	$\approx 0,9988$	$\approx 0,9874$	$\approx 0,9964$	$\approx 0,9981$	$\approx 0,9882$	$\approx 0,9941$

TABLE 2.14 – Projections polynomiales des montants cumulés du portefeuille

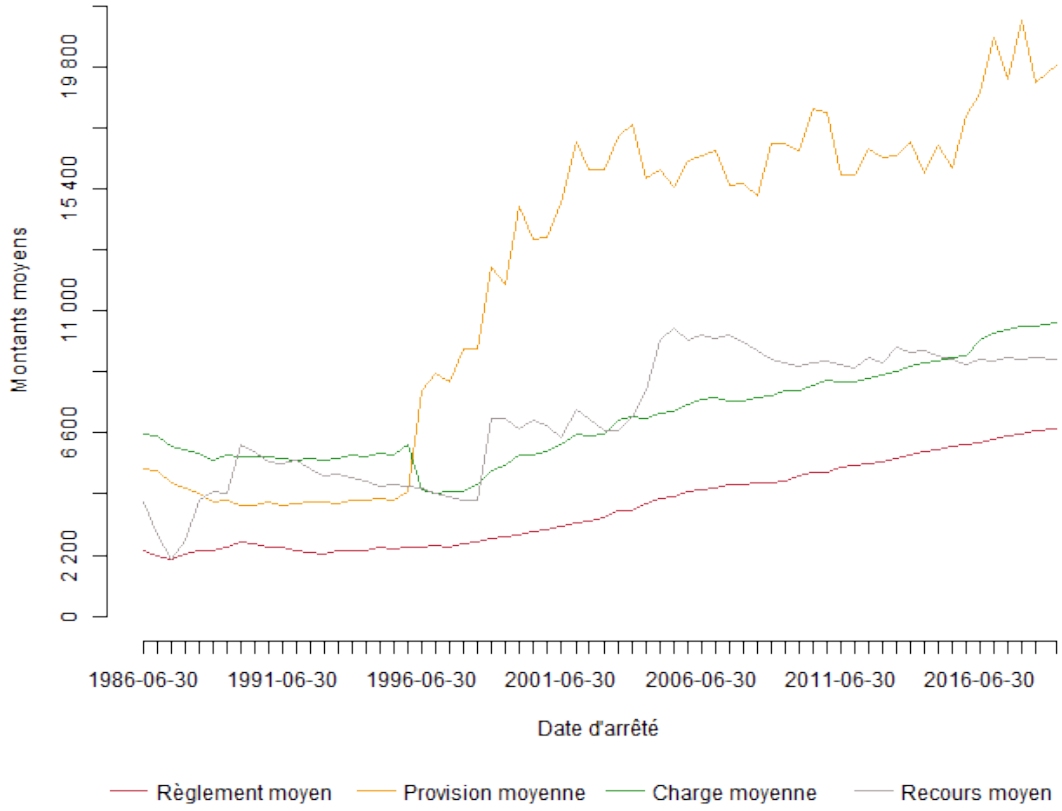
Horizon de projection		31/03/2019 (données actuelles)	31/12/2019 (fin de l'année)	31/12/2029 (+10 ans)	31/12/2039 (+20 ans)
Montants cumulés totaux	**Règlements**	331 161 125	337 360 240	604 250 371	949 524 479
	Évolution	-	+1,87%	+82,46%	+186,73%
	Provisions	233 342 258	230 369 845	390 296 340	595 651 572
	Évolution	-	-1,27%	+67,26%	+155,27%
	Charges	564 503 383	567 730 093	994 546 733	1 545 176 089
	Évolution	-	+0,57%	+76,18%	+173,72%
Dont charges totales > 200 000	**Règlements**	111 836 489	116 495 707	209 249 065	329 291 788
	Évolution	-	+4,17%	+87,10%	+194,44%
	Provisions	130 333 121	127 314 847	229 633 562	362 980 010
	Évolution	-	-2,32%	+76,19%	+178,50%
	Charges	255 879 719	256 694 691	461 296 537	727 237 263
	Évolution	-	+0,32%	+80,28%	+184,21%

Tout comme avec les volumes, en accord avec les projections obtenues par une régression polynomiale d'ordre 2, on remarque que les montants cumulés « graves » évolueraient

plus rapidement que les totaux. Pour les mêmes raisons que citées précédemment, même si les projections effectuées jusqu'ici ont une pertinence dans les grandes lignes, mettre en place un seuil qui dérive avec le temps rendrait ces résultats plus cohérents.

À titre d'information, en divisant les montants cumulés par les nombres de dossiers cumulés qui contiennent au moins lesdits montants, on obtient les montants moyens suivants :

FIGURE 2.4 – Montants moyens par dossier du portefeuille



La forte hausse du montant de provision moyen s'explique par l'impact de 1996. Les recours n'ont pas été évoqués jusqu'ici car leur faible quantité (moins de 2% aussi bien en volume qu'en montant total) n'est pas déterminante sur la compréhension des données, d'autant plus qu'ils n'impactent pas le calcul de la charge. Toutefois, on note la hausse par paliers de leur montant moyen qui doit probablement souligner des « vagues » de réévaluation des dossiers au niveau de la gestion.

Avec les montants moyens, le sinistre grave défini jusqu'ici se montre financièrement très impactant : la charge moyenne d'un sinistre grave est évaluée à 770 722€ au 31/03/2019, soit 73 fois plus que la charge moyenne globale à la même date (10 566€).

Finalement, ces analyses descriptives montrent une base de données retraitée qui paraît cohérente et stable en termes d'évolutions. De plus, en partant d'un seuil de définition d'un sinistre grave fixé par l'ACP, on note que les sinistres graves :

- sont très faibles en termes de nombre (moins de 1%) ;
- mais représentent une charge totale face à laquelle l'assureur ne peut pas passer outre (presque 50% du total).

De telles caractéristiques sont synonymes de forte volatilité avec d'important impacts financiers, cauchemars de l'assureur. Pour limiter au maximum de subir un tel risque, il faut alors adapter la définition d'un sinistre grave au portefeuille en question.

Ainsi, nous allons par la suite appliquer des méthodes actuarielles pour adapter ce seuil, et donc la problématique du sinistre grave.

2.2 Définition du choix du seuil des sinistres graves

Pour distinguer les sinistres exceptionnels des sinistres attritionnels, des observations effectuées sur les données et sur ce qui a comme habitude de se faire au sein des assureurs en RCM, nous allons chercher à définir un « seuil de gravité » qui scinderait la base de données en deux parties selon la médiane des charges cumulées. En d'autres termes, les sinistres attritionnels et exceptionnels couvriraient chacun la moitié des charges imputées au portefeuille, mais les fréquences seront de part et d'autre nettement différentes.

Une première approche sera effectuée par dichotomie sur le total des charges, puis une seconde approche utilisera des lois d'extremum pour estimer ce seuil de gravité sur les quantiles. La cohérence de ces deux approches sera vérifiée.

2.2.1 Approche par dichotomie

Pour avoir une idée globale d'où se situe le seuil de gravité cherché, nous appliquons dans un premier temps une série de seuils sur le portefeuille arrêté au 31/03/2019 et observons la répartition obtenue :

TABLE 2.15 – Volumétrie du portefeuille arrêté au 31/03/2019 en fonction du seuil de charges cumulées fixé

Seuil fixé	Nombre de dossiers	**Part des dossiers**	Charge cumulée	Charge moyenne	**Part des charges**
1	52 406	100%	564 502 313	10 772	100%
1 000	30 152	57,54%	558 162 861	18 512	98,88%
5 000	13 732	26,20%	515 861 509	37 566	91,38%
10 000	5 751	10,97%	451 465 916	78 502	79,98%
20 000	3 243	6,19%	415 918 626	128 251	73,68%
40 000	1 822	3,48%	375 673 418	206 187	66,55%
60 000	1 177	2,25%	343 157 431	291 553	60,79%
80 000	908	1,73%	324 550 318	357 434	57,49%
100 000	712	1,36%	306 839 137	430 954	54,36%
120 000	548	1,05%	289 143 035	527 633	51,22%
140 000	472	0,90%	279 367 274	591 880	49,49%
160 000	413	0,79%	270 529 405	655 035	47,92%
200 000	332	0,63%	255 879 719	770 722	45,33%
400 000	152	0,29%	204 643 919	1 346 342	36,25%
500 000	123	0,23%	191 871 355	1 559 930	33,99%

On constate ainsi que la médiane des charges cumulées situerait un seuil entre 120 000 et 140 000.

Pour affiner ce seuil, nous appliquerons une recherche par dichotomie adaptée à cette problématique. Puisque l'évolution de la répartition des charges cumulées est fonction décroissante du seuil et que le volume de données à disposition est suffisamment conséquent, on peut approximer la fonction « part des charges cumulées selon le seuil fixé » comme continue, ce qui donne un sens à l'applicabilité d'une méthode type dichotomie.

Soit $(x_i)_{i\in\mathbb{N}}$ l'ensemble des charges cumulées renseignées dans le portefeuille. Nous allons étudier la convergence d'une suite de seuils $(s_i)_{i\in\mathbb{N}}$ qui respecte la relation de récurrence suivante :

$$\begin{aligned} s_{min} &= 120\ 000 \\ s_{max} &= 140\ 000 \\ s_0 &= s_{min} \\ s_1 &= \frac{s_{max} + s_{min}}{2} \end{aligned}$$

Et pour chaque $i \geq 1$:

$$\begin{aligned}
\text{si : } & |s_i - s_{i-1}| < \epsilon \\
\text{alors : } & s_{i+1} = s_i \\
\text{sinon, si : } & \frac{\sum\limits_{\substack{j \in \mathbb{N} \\ x_j > s_i}} x_j}{\sum\limits_{j \in \mathbb{N}} x_j} > \frac{1}{2} \\
\text{alors : } & \begin{cases} s_{i+1} = \frac{s_i + s_{max}}{2} \\ s_{min} = s_i \end{cases} \\
\text{sinon, si : } & \frac{\sum\limits_{\substack{j \in \mathbb{N} \\ x_j > s_i}} x_j}{\sum\limits_{j \in \mathbb{N}} x_j} < \frac{1}{2} \\
\text{alors : } & \begin{cases} s_{i+1} = \frac{s_i + s_{min}}{2} \\ s_{max} = s_i \end{cases} \\
\text{sinon : } & s_{i+1} = s_i
\end{aligned}$$

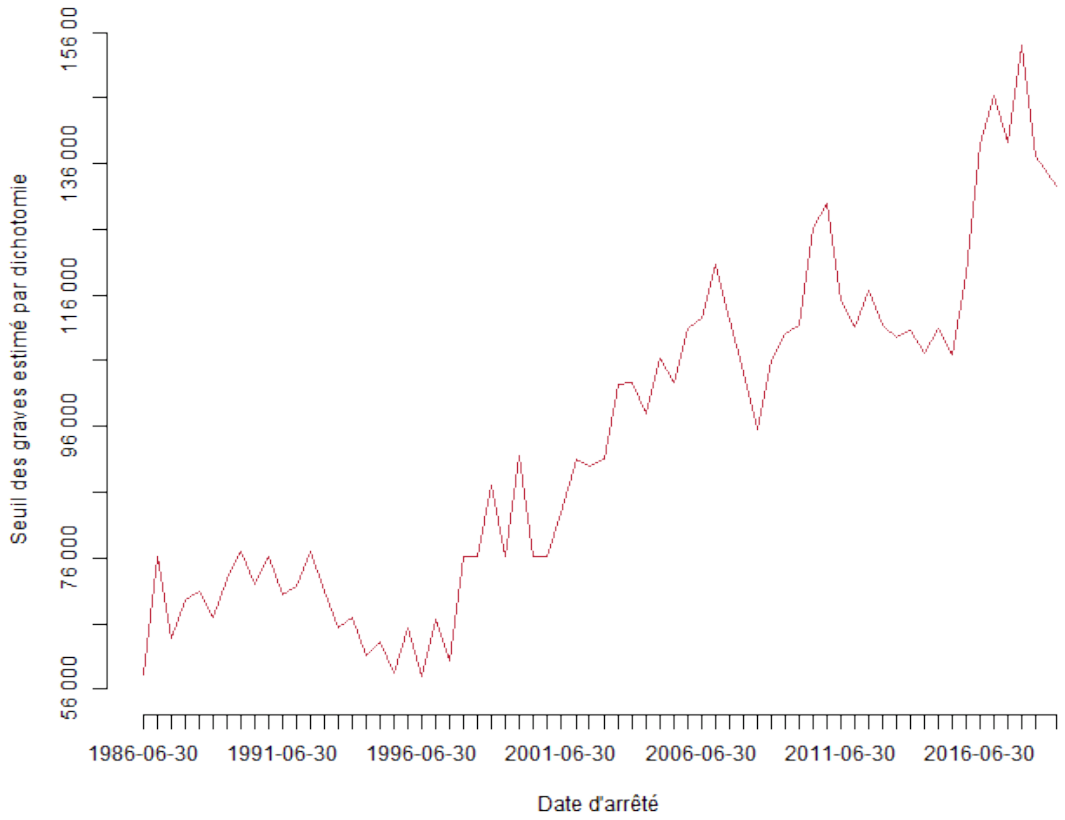
En considérant $\epsilon = 10^{-3}$, on détermine que cette suite converge vers :

$$s_i \underset{i \to +\infty}{\longrightarrow} 132\ 592,93 \text{ au bout de } 25 \text{ itérations} \tag{2.8}$$

En-dessous du seuil « de marché » qui était suggéré par l'ACP en 2009 [3] (200 000€), ce résultat paraît malgré tout pertinent, à première vue. En effet, en remontant l'historique des sinistres, on s'aperçoit que cette méthode a un caractère volatile comme le témoigne le graphique ci-après.

3. [Johanet, 2011]

FIGURE 2.5 – Évolution du seuil de gravité déterminé par dichotomie en fonction de la date d'arrêté



À titre d'information, ces résultats ont été établis en adaptant les bornes par :

$$x_{min} = 10\ 000$$
$$x_{max} = 1\ 000\ 000$$

Tous ont convergé en 30 itérations.

Bien que cette approche soit robuste en temps de calcul (ici 1,22 secondes pour tout exécuter), elle montre ainsi ses limites sur sa prédictibilité des sinistres graves : la dichotomie ne se base que sur les données à disposition sans chercher à considérer le comportement des queues de distribution. D'un autre point de vue, ce genre de méthode empirique nécessite un volume de données tellement important pour être efficace qu'il n'est jamais disponible dans la réalité [4].

Mais en ce sens, elle reste un bon indicateur pour la compréhension des données : les augmentations brutales du seuil laissent songer à l'ouverture de sinistres exception-

4. [Planchet et Thérond, 2007]

nels, et *a contrario* les baisses brutales peuvent être synonymes d'ouvertures de sinistres attritionnels plus élevées que d'habitude.

2.2.2 Approche basée sur la théorie des valeurs extrêmes

Le principe de cette approche est d'analyser le comportement de la queue de distribution des charges cumulées observées afin d'obtenir un seuil de gravité plus pertinent statistiquement et moins volatile dans le temps. Pour aboutir à cela, il sera question d'ajuster l'ensemble du portefeuille étudié à une loi paramétrique puis d'estimer son quantile au niveau de probabilité le plus représentatif de la queue de distribution.

Notations

On considère $(X_i)_{i\in\mathbb{N}^*}$ un échantillon de n variables aléatoires indépendantes et identiquement distribuées (i.i.d.). On notera F la fonction de répartition de la loi de X et $\overline{F} = 1 - F$ sa fonction de survie.

En ordonnant l'ensemble des $n \in \mathbb{N}^*$ variables aléatoires, on note $X_{k,n}$ la statistique d'ordre associée :

$$\begin{aligned} \min_{i\in[\![1,n]\!]}(X_i) &= X_{n,n} \\ \max_{i\in[\![1,n]\!]}(X_i) &= X_{1,n} \\ \forall \begin{cases} (k_1,k_2) \in [\![1,n]\!]^2 \\ k_1 > k_2 \end{cases} &: X_{k_1,n} \leq X_{k_2,n} \end{aligned} \tag{2.9}$$

On notera F_s la fonction de répartition du surplus des variables aléatoires X qui dépassent le seuil s :

$$\begin{aligned} \forall s \in \mathbb{R},\ F_s : \mathbb{R} &\to [0,1] \\ x &\mapsto \mathbb{P}[X - s \leq x | X > s] \end{aligned} \tag{2.10}$$

On notera aussi x_F la borne supérieure du support de X et de fonction de répartition paramétrique F :

$$\forall x \in \mathbb{R} : x_F = \sup\{x | F(x) < 1\} \tag{2.11}$$

Enfin, on notera N_s le nombre d'observations $(x_i)_{i\in\mathbb{N}^*}$ qui dépassent le seuil s :

$$\forall s \in \mathbb{R} : N_s = \sum_{i=1}^{n} \mathbf{1}_{x_i > s} \tag{2.12}$$

Éléments de théorie des valeurs extrêmes

La théorie des valeurs extrêmes, dans son sens le plus stricte, étudie la distribution asymptotique du maximum.

Sous les notations qui précèdent, pour considérer le comportement extrême de évènements, on considère le maximum d'un échantillon de taille n $X_{1,n}$. Comme les variables aléatoires sont i.i.d., la fonction de répartition associée à ce maximum s'écrirait :

$$F_{X_{1,n}} = \mathbb{P}[X_{1,n} \leq x] = (F(x))^n \tag{2.13}$$

Cependant, cette relation présente peu d'intérêt dans notre situation pour plusieurs raisons :

- la loi associée aux données ne sera jamais connue avec précision car requiert une quantité de données inatteignable ;
- et même si elle était connue, la loi du maximum n'est pas facilement calculable puisqu'elle converge $\mathbb{P}$-presque sûrement vers une loi dégénérée.

$$X_{1,n} \xrightarrow[n\to+\infty]{\text{p.s.}} x_F \tag{2.14}$$

$$\forall x \in \mathbb{R} : F_{X_{1,n}}(x) \xrightarrow[n\to+\infty]{\text{p.s.}} \begin{cases} 0 \text{ si } F(x) < 1 \\ 1 \text{ sinon} \end{cases} \tag{2.15}$$

Pour éviter ceci, il serait alors plus pertinent de considérer des comportements asymptotiques du maximum normalisé que le maximum en lui-même.

À ce titre, la théorie des valeurs extrêmes a vu ses débuts par ce théorème fondamental établi en 1928 par Ronald Aylmer Fisher et Leonard Henry Caleb Tippett [5] :

Définition 1 (Lois de même type) *Deux variables aléatoires X et Y sont de même type s'il existe un couple de deux constantes réelles $(a, b) \in \mathbb{R}^2$ tels que :*

$$Y \overset{\text{loi}}{=} a \cdot X + b$$

Si X et Y sont de même type, alors leur fonction de répartition, respectivement F_X et F_Y sont liées par la relation suivante :

$$\forall x \in \mathbb{R} : F_Y(x) = F_X(a \cdot x + b)$$

En d'autres termes, X et Y sont de même type s'ils ont la même loi à un facteur de localisation et d'échelle près.

Pour cette « normalisation », à l'image du théorème central limite, on introduit deux constantes de normalisation qui dépendent de l'échantillon analysé $a_n > 0$ et $b_n \in \mathbb{R}$ telles qu'il existe une loi non-dégénérée de fonction de répartition H qui respecte la relation suivante :

$$\forall x \in \mathbb{R}, \mathbb{P}\left[\frac{X_{1,n} - b_n}{a_n} \leq x\right] = (F(a_n \cdot x + b_n))^n \xrightarrow[n\to+\infty]{} H(x) \tag{2.16}$$

Fisher et Tippett ont alors trouvé une solution à ce problème et ont établi le théorème fondamental de la théorie des valeurs extrêmes suivant :

5. [Fisher et Tippett, 1928]

Théorème 1 (Fisher-Tippett-Gnedenko) *Pour un échantillon de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ i.i.d., s'il existe :*

- *deux suites de constantes de normalisation $(a_n)_{n\in\mathbb{N}} > 0$ et $(b_n)_{n\in\mathbb{N}} \in \mathbb{R}$;*
- *et une loi non-dégénérée de fonction de répartition H telle que :*

$$\forall x \in \mathbb{R}, \mathbb{P}\left[\frac{X_{1,n} - b_n}{a_n} \leq x\right] = (F(a_n \cdot x + b_n))^n \underset{n\to+\infty}{\longrightarrow} H(x)$$

alors cette loi de fonction de répartition H est du même type que l'une de ces trois lois d'extremum :

les lois d'extremum de type I : *(lois détaillées par Emil Julius Gumbel)*

$$\forall x \in \mathbb{R} : \Delta(x) = \exp(-\exp(-x)) \tag{2.17}$$

les lois d'extremum de type II : *(lois établies par Maurice Fréchet en 1927[6])*

$$\forall x \in \mathbb{R}, \forall \alpha > 0 : \Phi_\alpha(x) = \begin{cases} 0 & si\ x \leq 0 \\ \exp(-x^{-\alpha}) & sinon \end{cases} \tag{2.18}$$

les lois d'extremum de type III : *(lois identifiées par par Maurice Fréchet en 1927 puis détaillées par Ernst Hjalmar Waloddi Weibull en 1951[7])*

$$\forall x \in \mathbb{R}, \forall \alpha > 0 : \Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & si\ x \leq 0 \\ 1 & sinon \end{cases} \tag{2.19}$$

Les trois types de loi d'extremum résultent de travaux effectués par Boris Vladimirovitch Gnedenko en 1948[8], d'où la présence de son nom. Ce théorème démontre que toute loi limite d'un échantillon de variables aléatoires a une loi des extrêmes issue d'une même famille de lois d'extremum, et ce indépendamment de son comportement : on parlera de domaine d'attraction maximum (DAM).

Définition 2 (Domaine d'attraction maximum) *La fonction de répartition F d'une loi qui répond aux hypothèses du théorème de Fisher-Tippett-Gnedenko est dite comme appartenir au DAM de H.*

Il aura été démontré plus tard que ces trois groupes de distributions sont liés à la représentation de Jenkinson-Von Mises, dénommée ainsi « loi d'extremum généralisée » (GEV), par la combinaison de trois paramètres :

- un paramètre de position $\mu \in \mathbb{R}$;
- un paramètre d'échelle $\sigma > 0$;

6. [Fréchet, 1927]
7. [Weibull, 1951]
8. [Gnedenko, 1948]

- et un paramètre de forme $\xi \in \mathbb{R}$ qui s'avère nous renseigner sur l'épaisseur de la queue de distribution.

Définition 3 (Loi d'extremum généralisée) *Une loi qui appartient à une famile de distributions GEV(μ, σ, ξ) est non-dégénérée avec la fonction de répartition $H_{\mu,\sigma,\xi}$ suivante :*

$$\forall x \in \mathbb{R} : H_{\mu,\sigma,\xi}(x) = \exp\left(-\left[1+\xi\cdot\frac{x-\mu}{\sigma}\right]_+^{-\frac{1}{\xi}}\right) \tag{2.20}$$

Cette fonction de répartition n'est pas définie en $\xi = 0$ mais elle peut s'entendre en ce point comme la limite suivante :

$$\forall x \in \mathbb{R} : H_{\mu,\sigma,\xi}(x) \underset{\xi\to 0}{\longrightarrow} \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \tag{2.21}$$

La fonction de densité associée à cette loi est la suivante :

$$\begin{aligned} \forall x \in \mathbb{R} : h_{\mu,\sigma,\xi}(x) &= \frac{1}{\sigma}\cdot\left(1+\xi\cdot\frac{x-\mu}{\sigma}\right)^{-\left(1+\frac{1}{\xi}\right)}\cdot\exp\left(-\left(1+\xi\cdot\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) \\ h_{\mu,\sigma,\xi}(x) &\underset{\xi\to 0}{\longrightarrow} \frac{1}{\sigma}\cdot\exp\left(-\frac{x-\mu}{\sigma}\right)\cdot\exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right) \end{aligned} \tag{2.22}$$

Son lien avec les trois familles de lois d'extremum se base sur la valeur de ξ :

Table 2.16 – Caractéristiques de la GEV en fonction du paramètre ξ

Valeurs de ξ	$\xi = 0$	$\xi = \alpha^{-1} > 0$	$\xi = -\alpha^{-1} < 0$
Domaine d'attraction maximum	Gumbel (type I)	Fréchet (type II)	Weibull (type III)
Épaisseur de la queue de distribution	Fine ou moyenne	Épaisse	Fine ou distribution bornée à droite
Mode de décroissance de la queue	Exponentielle	Puissance	-
Support	Non borné		Borné
Lois usuelles correspondantes	Normale, gamma, exponentielle, log-normale	Cauchy, Pareto, chi-deux, Student	Uniforme, beta

Ainsi, pour adapter ces notions théoriques à la problématique d'identification des sinistres les plus graves du portefeuille étudié, une première approche envisageable consiste à ajuster l'ensemble des données à une loi d'extremum puis estimer son quantile au niveau de probabilité souhaité. De ce quantile extrait, un seuil « quantile » pourra être déduit et distinguera les sinistres « en queue de distribution » des autres.

Cette méthode d'estimation suit la chronologie suivante :

1. choix d'une ou de plusieurs lois d'extremum ;
2. estimation des paramètres selon différentes méthodes (*cf.* annexe D) ;
3. tests d'adéquation ;
4. et inversion de la fonction de répartition.

Utilisation de lois d'excès pour choisir le seuil

Une seconde approche, plus analytique, serait de modéliser l'ensemble des sinistres qui dépasse un seuil suffisamment élevé pour considérer ces sinistres comme graves. On cherche alors non plus à étudier le comportement de la queue de distribution mais un ensemble de valeurs observées au-delà d'un certain seuil modélisé.

Cette méthode, initiée par Pickands en 1975, considère pour un seuil $s \in \mathbb{R}$ la suite des dépassements du seuil s par les $(X_i)_{i\in\mathbb{N}}$ suivante :

$$\forall s \in \mathbb{R}, (Y_j)_{j\in[\![1,N_s]\!]} = (X_i - s)_{i\in\{i\in\mathbb{N}^*, X_i - s>0\}}$$

La fonction de répartition de Y vaut F_s, et il a été démontré que, sous certaines conditions de convergence, la loi limite de Y serait une Pareto généralisée (GPD).

Définition 4 (Loi de Pareto généralisée) *Une loi qui appartient à une famille de distributions GPD(μ, σ, ξ) est non-dégénérée avec la fonction de répartition $G_{\mu,\sigma,\xi}$ suivante :*

$$\begin{aligned} \forall x \in \mathbb{R} : G_{\mu,\sigma,\xi}(x) &= 1 - \left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}} \\ G_{\mu,\sigma,\xi}(x) &\underset{\xi\to 0}{\longrightarrow} 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) \end{aligned} \tag{2.23}$$

avec $\sigma > 0$ et $x \geq \mu$ si $\xi \geq 0$, et $0 \leq x \leq \mu - \frac{\sigma}{\xi}$ sinon.

La densité associée à cette loi est la suivante :

$$\begin{aligned} \forall x \in \mathbb{R} : g_{\mu,\sigma,\xi}(x) &= \frac{1}{\sigma} \cdot \left(1 + \xi \cdot \frac{x-\mu}{\sigma}\right)^{-\left(1+\frac{1}{\xi}\right)} \\ g_{\mu,\sigma,\xi}(x) &\underset{\xi\to 0}{\longrightarrow} \frac{1}{\sigma} \cdot \exp\left(-\frac{x-\mu}{\sigma}\right) \end{aligned} \tag{2.24}$$

Théorème 2 (Pickands-Balkema-de Haan) *Soient H_ξ la fonction de répartition d'une GEV de paramètre de forme $\xi \in \mathbb{R}$, $s \in \mathbb{R}$ un seuil fixé et G_ξ une fonction de répartition d'une GPD qui dépend du même ξ.*

Une fonction de répartition F appartient au domaine d'attraction maximum associé à G_ξ si, et seulement si, il existe une fonction positive $\beta : \mathbb{R} \to \mathbb{R}_+^*$ telle que :

$$\sup_{0 \leq y \leq x_F - s} \left\{ |F_s(y) - H_{\xi,\beta(s)}| \right\} \underset{s \to x_F}{\longrightarrow} 0 \tag{2.25}$$

Ce théorème démontre ainsi l'existence d'un lien entre le paramètre d'une loi d'un DAM et le comportement limite des excès au-delà d'un seuil : les queues de distribution peuvent, sous certaines conditions, s'approximer par une GPD. Cette conclusion est le fondement de la méthode de modélisation d'évènements rares qui dépassent un seuil *Peak Over the Threshold* (POT), initiée par Laurens de Haan et Holger Rootzén en 1993.

En considérant la loi « excès qui dépassent un seuil » on observe avec la formule de Bayes que :

$$\forall (x,s) \in \mathbb{R}^2 : F_s(x) = \mathbb{P}[X - s \leq x | X > s] = \begin{cases} \frac{F(x+s)-F(s)}{1-F(s)} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \tag{2.26}$$

On en déduit la relation suivante :

$$\forall (x,s) \in \mathbb{R}^2 : \overline{F}(x) = \overline{F}(s) \cdot \overline{F_s}(x-s) \tag{2.27}$$

Le théorème précédent nous démontre que $\forall s \in \mathbb{R}, x \mapsto F_s(x-s)$ est approximable par une GPD de fonction de répartition de la forme $G_{s,\hat{\beta}(s),\hat{\xi}}(x)$ pour un s suffisamment grand. De ces formules, on peut alors estimer F selon la relation suivante [9] :

$$\forall x > s, n \in \mathbb{N} : \overline{\hat{F}}(x) = \frac{N_s}{n} \cdot \left(1 + \hat{\xi} \cdot \frac{x-s}{\hat{\beta}(s)} \right)^{-\frac{1}{\hat{\xi}}} \tag{2.28}$$

Il suffirait alors d'estimer les paramètres de cette distribution théorique pour obtenir une approximation optimale de $\overline{F}$, et *in fine* une estimation précise de la fonction quantile F^{-1} à des ordres p élevés :

$$\hat{x}_p = s + \frac{\hat{\xi}}{\hat{\beta}(s)} \cdot \left(\left(\frac{n \cdot (1-p)}{N_s} \right)^{-\hat{\xi}} - 1 \right) \tag{2.29}$$

La difficulté de cette méthode réside dans le choix du seuil, socle du POT. Il doit être suffisamment grand pour que l'approximation par une GPD ait un sens mais sans être trop proche du quantile à estimer pour disposer d'un volume statistiquement satisfaisant. Pour cela, il existe une méthode similaire à l'estimation de Hill d'une queue de distribution qui consiste à tracer la fonction d'espérance empirique résiduelle de $(X_i)_{\substack{i \in [\![1,n]\!] \\ n \in \mathbb{N}^*}}$ en fonction

9. [Planchet et Thérond, 2007]

de s :

$$\begin{aligned} \forall n \in \mathbb{N}^*, e_n : \mathbb{R} &\to \mathbb{R}_+ \\ s &\mapsto \frac{1}{N_s} \cdot \sum_{i=1}^{n} [X_i - s]_+ \end{aligned} \tag{2.30}$$

et d'ensuite identifier le plus petit s tel que la fonction soit approximativement linéaire sur $[s, +\infty[$.

Cette démarche se justifie par la linéarité de l'espérance résiduelle d'une GDP de paramètre de forme $\xi < 1$:

$$\begin{aligned} \forall X \in GDP(\beta(s), \xi), e : \mathbb{R} &\to \mathbb{R}_+ \\ s &\mapsto \mathbb{E}[X - s] = \frac{\beta(s) + \xi \cdot s}{1 - \xi} \end{aligned} \tag{2.31}$$

Toutefois, cette méthode restant empirique, la valeur de s serait plus ou moins cohérente en fonction du nombre de valeurs testées pour la déterminer.

2.2.3 Adaptation de la théorie des valeurs extrêmes à la problématique et analyse des résultats

Dans ce qui suit, nous allons utiliser ces théories pour les appliquer aux sinistres graves RCM. L'approche adaptée à notre problématique suivra le protocole suivant :

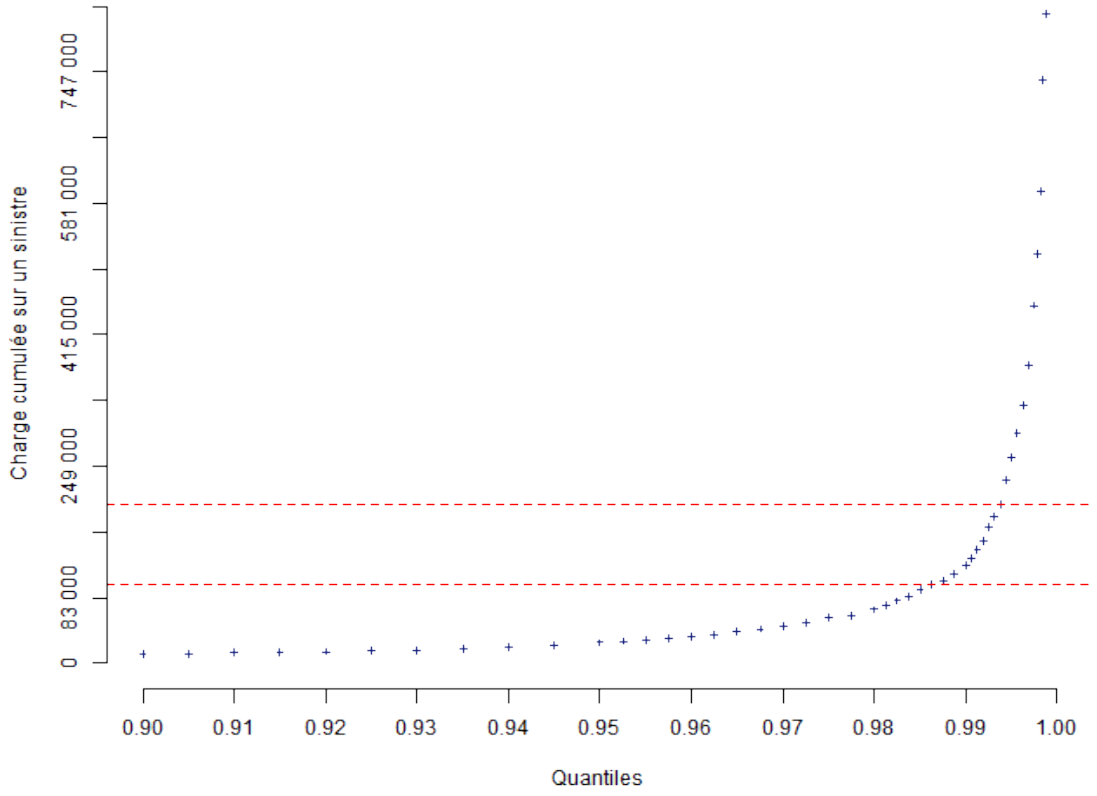
identification du type de loi d'extremum : cette recherche consiste à déterminer dans quelle famille de distributions d'extremum se rangent les observations étudiées. Elle s'effectuera graphiquement à l'aide d'un diagramme quantile-quantile (*QQ-plot*, *cf.* annexe E.1) par rapport à l'exponentielle ;

estimation des paramètres : une analyse graphique de la moyenne des excédents en fonction d'un seuil sur les données (*ME-plot*, *cf.* annexe E.2) identifiera un seuil minimum « stable » à appliquer aux lois simulées pour être en adéquation avec les données. En fonction du type de loi d'extremum qui aura été précédemment déterminé, des analyses graphiques similaires seront appliquées sur des estimateurs semi-paramétriques (*cf.* annexe D) ;

adéquation des données avec une loi d'extremum : une fois les paramètres estimés, il sera à nouveau question d'effectuer une analyse *QQ-plot* pour identifier quelle loi d'extremum connue est la plus en adéquation avec les données ;

validation des résultats : une fois la loi trouvée, des tests d'adéquation seront effectués entre le seuil estimé et ladite loi pour savoir si la modélisation de sinistres graves selon cette loi est statistiquement cohérente.

FIGURE 2.6 – Relevé de quantiles sur la charge cumulée par sinistre de la base d'étude, quantiles situés entre 99,0% et 99,9%



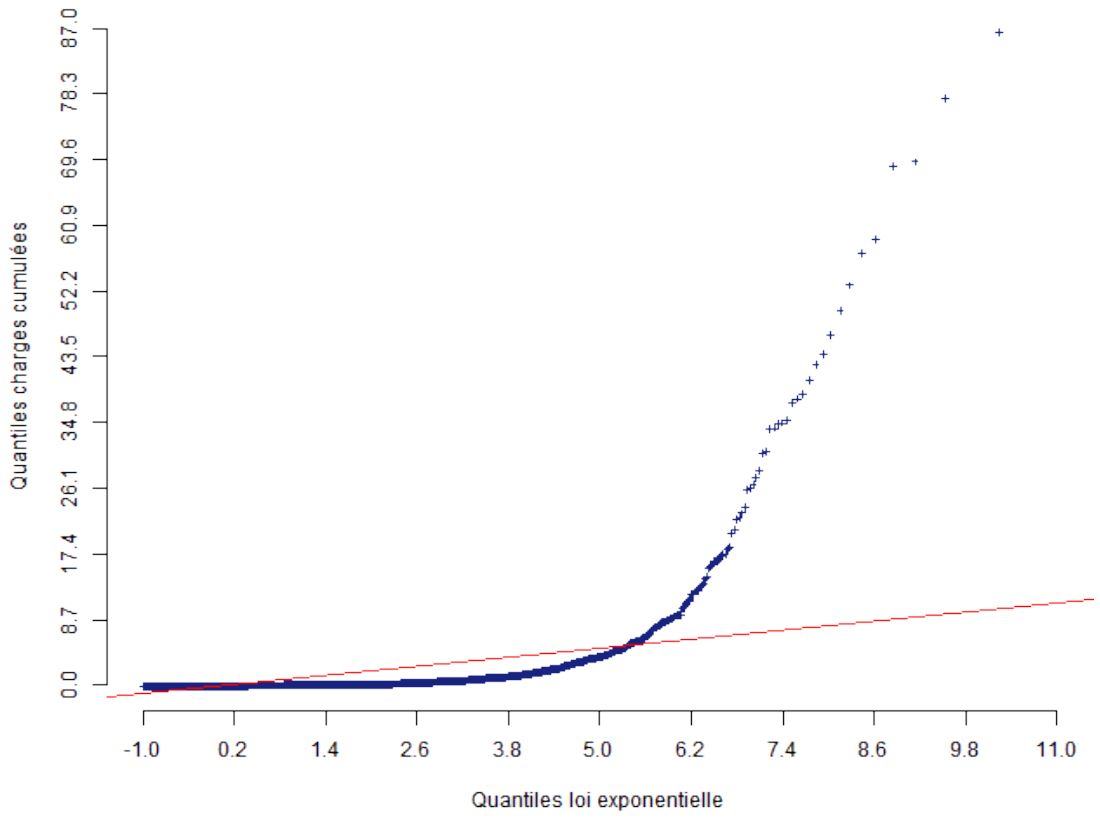
En amont de l'application de cette procédure, afin de se faire une première idée de la queue de distribution du portefeuille analysé, on construit l'évolution des quantiles des charges cumulées par sinistre. On observe graphiquement une relative stabilité du montant « quantile des charges cumulées » jusqu'aux alentours du quantile à 95% . Entre 95% et 99% , la charge cumulée croît de plus en plus rapidement jusqu'à augmenter quasi-verticalement au-delà des 99% . À ce stade, on pourrait alors imaginer que ce décrochage observé aux alentours de 99% symbolise le passage des sinistres attritionnels aux sinistres graves, ce qui situerait alors un seuil de gravité approximativement entre 100 000€ et 200 000€.

Pour une meilleure lisibilité des résultats, les *QQ-plot* seront centrés et réduits sur une échelle commune qui, pour un échantillon de variables aléatoires ou d'observations $(x_i)_{i\in\mathbb{N}^*}$, suit la relation suivante :

$$\forall i \in [\![0,n]\!], n \in \mathbb{N}^* : x_i^* = \frac{x_i - \overline{x}}{\sigma_x}$$

où $\overline{x}$ et σ_x représentent respectivement la moyenne et l'écart-type de l'échantillon.

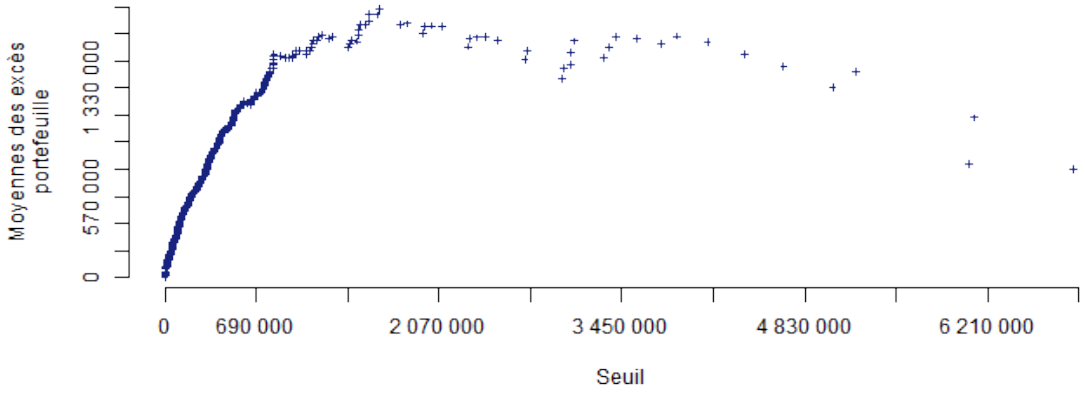
FIGURE 2.7 – *QQ-plot* entre charges cumulées observées et la loi exponentielle standard ($\lambda = 1$)



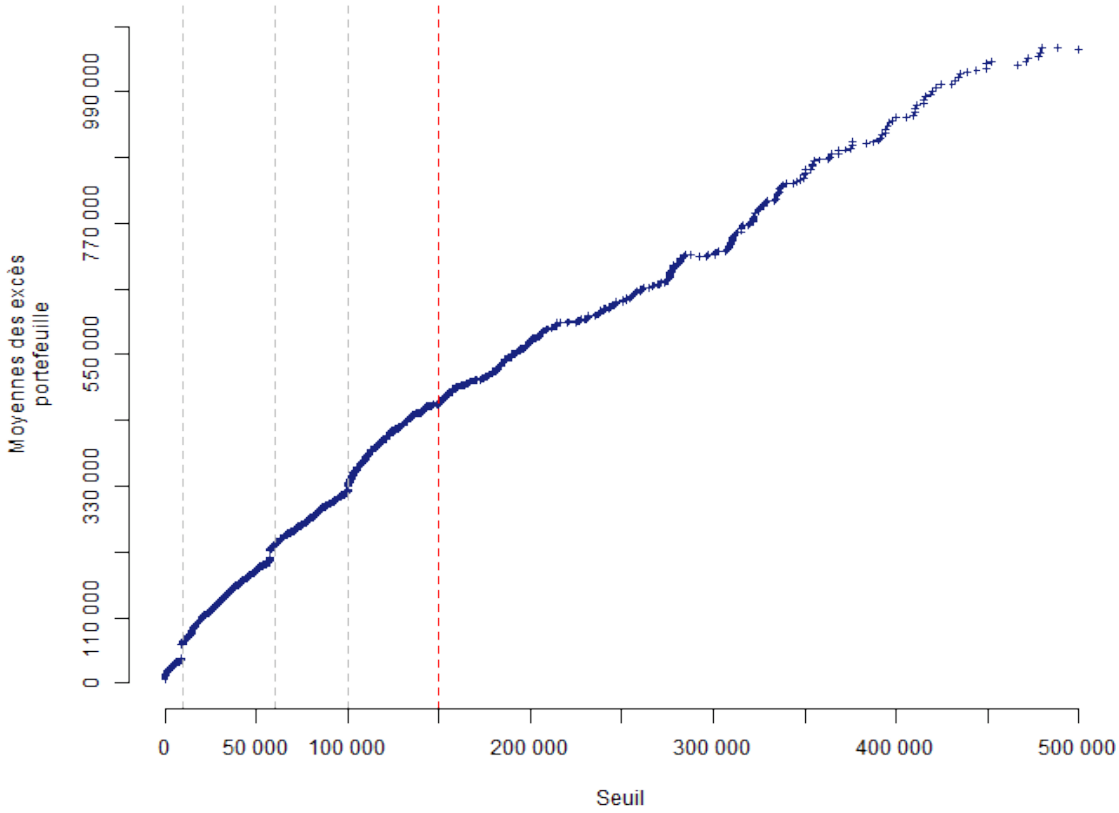
La courbe obtenue est convexe : la queue de distribution du portefeuille étudié est lourde. Le comportement des sinistres graves suivrait alors une loi d'extremum de type II (distributions de Fréchet).

FIGURE 2.8 – *ME-plot* sur charges cumulées observées

(a) Vue d'ensemble



(b) Analyse sur la plage de seuils $[0, 500\ 000]$



Les moyennes d'excès du portefeuille forment une fonction croissante du seuil : cette observation confirme à nouveau que la queue de distribution étudiée est lourde.

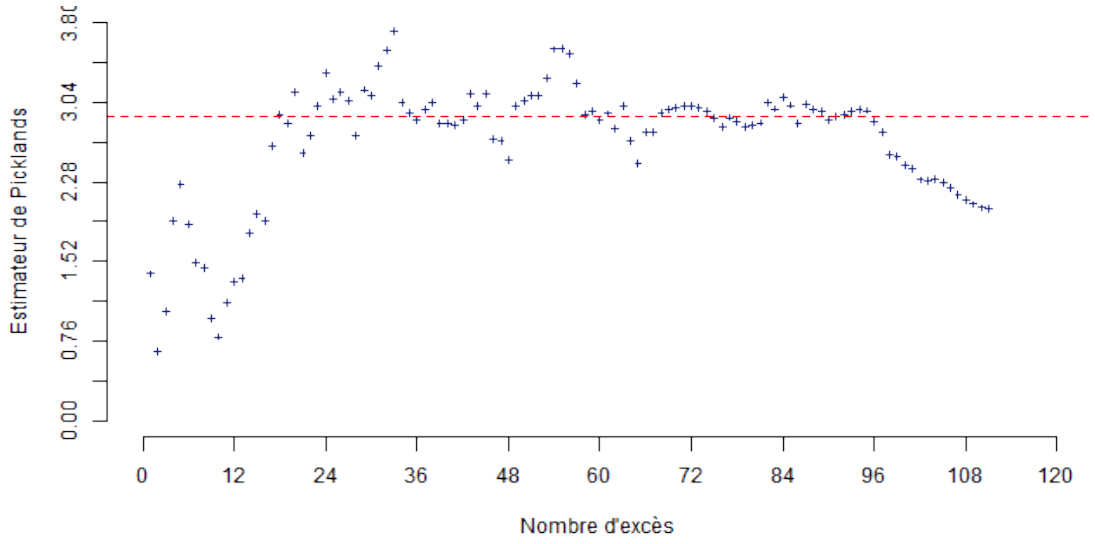
Sur l'ensemble des seuils inférieurs à 500 000€, l'évolution des excédents moyens

montre des cassures (indiquées par des droites verticales) jusqu'aux environs de 150 000€. Au-delà de ce seuil, la tendance est relativement linéaire, et le nombre de points relevés diminue suffisamment au-delà des 300 000€ pour ne plus représenter un impact significatif sur la tendance « linéaire » qui est précédemment observée.

On déduit alors de ce *ME-plot* que le seuil optimal de stabilité pour modéliser des excédents de seuil au niveau de nos données se situerait à **150 000€**.

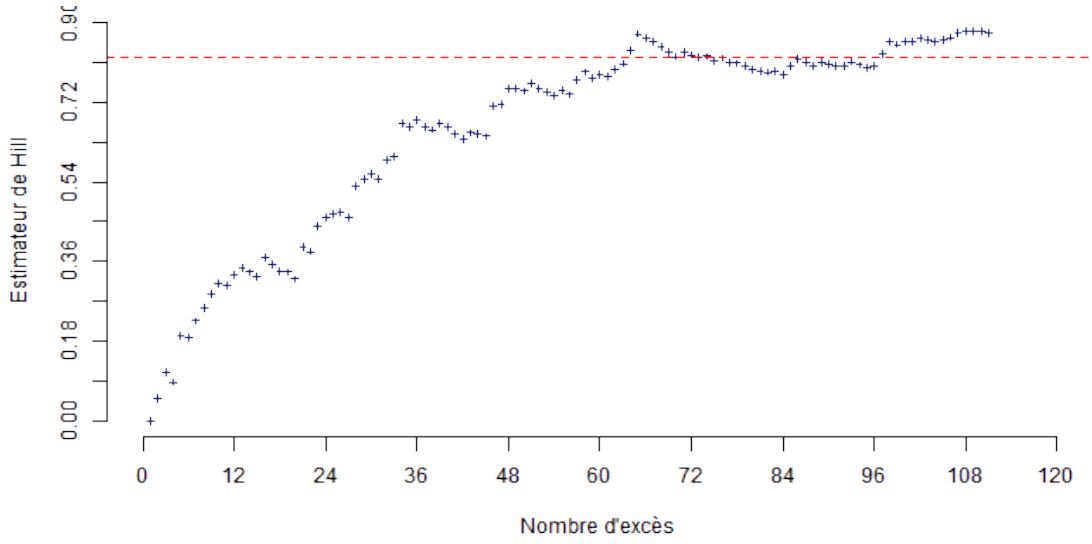
À présent, il est question de déterminer un paramètre de forme relativement stable à appliquer sur la base d'étude. Ayant déterminé que le DAM le plus en adéquation avec celle-ci est de Fréchet, *i.e.* nous situons $\xi > 0$, des diagrammes de Pickands ainsi que de Dekkers-Eimbahl-de Haan sont établis. Pour respecter la condition « $\frac{k}{n} \to 0$ », nous considèrerons $k = \frac{\text{Nombre de charges cumulées qui dépassent le seuil estimé}}{4}$.

FIGURE 2.9 – *Pickands-plot* sur charges cumulées observées



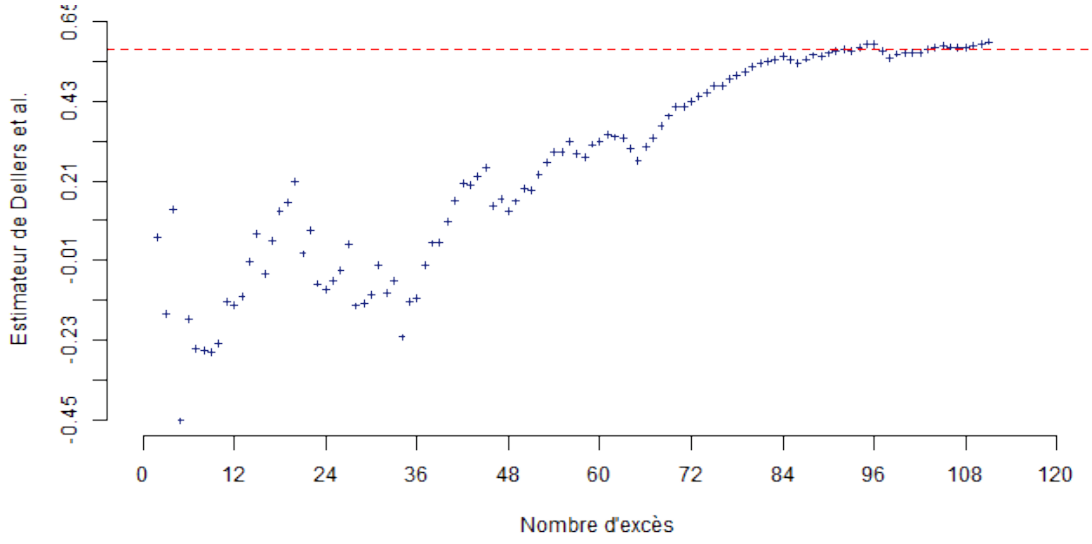
Le diagramme de Pickands présente une importante volatilité et situerait un paramètre de forme aux alentours de $\hat{\xi}^P \approx 2,75$.

FIGURE 2.10 – *Hill-plot* sur charges cumulées observées



Le diagramme de Hill est moins volatile que le diagramme de Pickands précédemment présenté. L'estimateur « stable » se situerait ici aux alentours de $\hat{\xi}^H \approx 0,82$.

FIGURE 2.11 – *DEdH-plot* sur charges cumulées observées



Le diagramme de Dekkers-Einmahl-de Haan situerait un paramètre de forme aux alentours de $\hat{\xi}^{DEdH} \approx 0,57$, seuil plus proche que déterminé avec le *Hill-plot* qu'avec le *Pickands-plot*.

Toutefois, nous observons des résultats relativement différents d'une méthode à une

autre. Ces écarts de résultats observés peuvent s'expliquer par le choix de k. En effet, la condition $\frac{k}{n} \to 0$ impose un k suffisamment petit pour estimer un ξ stable et intrinsèquement avec le moins de biais possible. Néanmoins, cette faible quantité implique d'effectuer des analyses sur un faible volume, d'où la volatilité des résultats.

Bien que le diagramme de Hill soit réputé pour être adapté aux distributions type « de Fréchet », le paramètre de forme retenu sera $\xi = 0,57$ car le *DEdH-plot* présente graphiquement la meilleure convergence. Les deux autres seront retenus à titre d'information pour se faire une idée de l'évolution de l'adéquation des lois avec la base d'étude en fonction du paramétrage fixé.

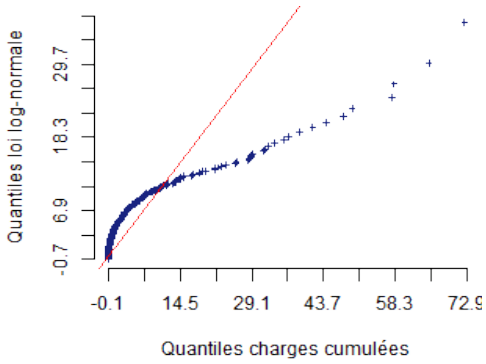
Ainsi, avec ces informations qui nous fournissent une estimation convenable du paramètre de position et du paramètre de forme associés à la base d'étude, nous pouvons à présent songer à tester l'adéquation statistique de lois issues de la famille dite de Fréchet avec des paramètres d'échelle et de forme standards pour identifier quelle loi se rapproche le plus du comportement de la queue de distribution du portefeuille. De la même façon que précédemment, les *QQ-plot* seront présentés sur une échelle commune normalisée.

Quant au paramètre d'échelle, il sera calculé en fonction des deux autres paramètres estimés et des grandeurs classiques associées aux lois usuelles (moyenne, variance, *etc.*). Sous ces trois paramètres, sept lois ont été testées :

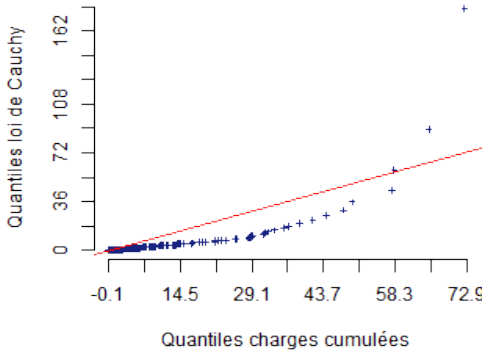
- loi Log-normale ;
- loi de Cauchy ;
- loi de Student (avec n degrés de liberté, n étant le nombre de sinistres) ;
- loi du Khi-2 (*idem*) ;
- loi de Fréchet ;
- loi de Pareto ;
- loi de Pareto généralisée.

FIGURE 2.12 – *QQ-plot* sur des lois d'extremum usuelles de type II

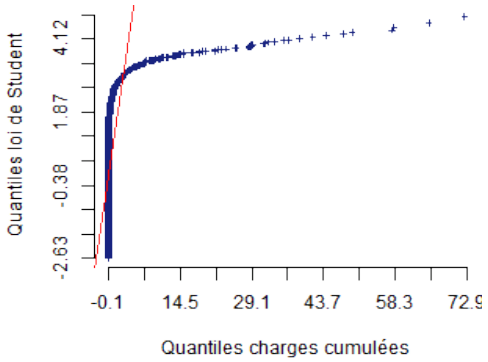
(a) Loi log-normale



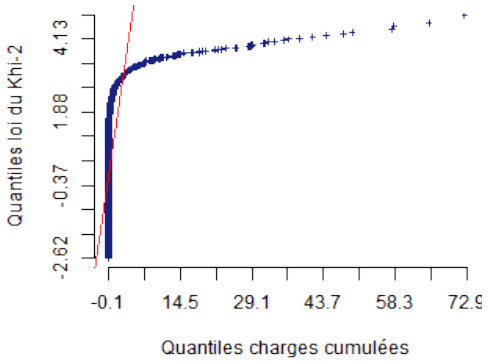
(b) Loi de Cauchy



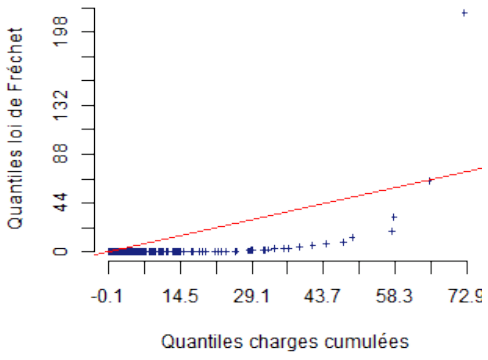
(c) Loi de Student



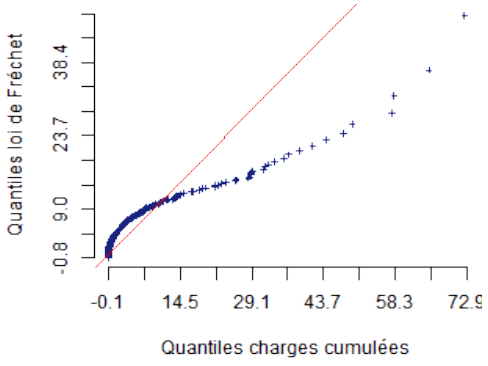
(d) Loi du Khi-2



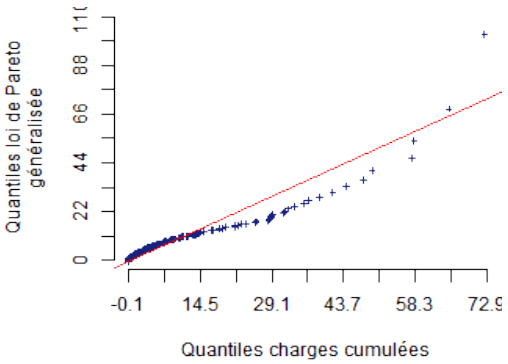
(e) Loi de Fréchet, $\xi = 0,57$



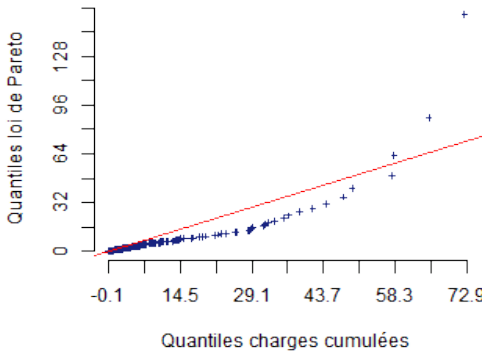
(f) Loi de Fréchet, $\xi = 2,75$



(g) Loi de Pareto généralisée, $\xi = 0,57$



(h) Loi de Pareto, $\xi = 0,82$



Les comparaisons de diagrammes quantile-quantile obtenues montrent que, sous ces conditions, la loi usuelle de type II qui s'accommode le mieux au comportement des sinistres graves du portefeuille est la **loi de Pareto généralisée**. Cette conclusion fait

écho au théorème de Pickands-Balkema-de Haan qui stipule que le comportement d'une queue de distribution proche d'un DAM peut s'approximer par une GPD.

Toutefois, dans un souci d'effectuer des analyses prospectives, comme notre GPD retenue a un paramètre de forme $\xi = 0,57 > \frac{1}{2}$, cela signifie qu'elle possède une variance infinie. Il serait alors préférable de basculer sur une **loi de Pareto**, cas particulier de loi de Pareto généralisée qui est beaucoup plus stable.

Afin de trancher sur cette question, et intrinsèquement confirmer l'adéquation du comportement du portefeuille de sinistralité sur les cas rares avec une loi d'extremum, nous allons effectuer des tests statistiques (*cf.* annexe F) qui valideront la pertinence de la démarche poursuivie jusqu'ici.

TABLE 2.17 – Valeurs p issues de tests statistiques d'adéquation des lois étudiées avec la sinistralité grave du portefeuille étudié

Loi testée pour adéquation	Pareto généralisée, $\xi = 0,57$	Pareto, $\xi = 0,57$	Pareto, $\xi = 0,82$	Fréchet, $\xi = 0,82$	Fréchet, $\xi = 2,75$
Valeur p (Kolmogorov-Smirnov)	$\approx 7,74\%$	$\approx 0,00\%$	$\approx 5,11\%$	$\approx 0,06\%$	$\approx 0,00\%$
Valeur p (Khi-2)	$\approx 23,96\%$	$\approx 23,96\%$	$\approx 23,96\%$	$\approx 0,00\%$	$\approx 0,00\%$

La contradiction des résultats observés en fonction du test statistique effectué est intéressante car met en garde contre l'utilisation abusive de tests pour chercher naïvement une *p-value* la plus satisfaisante qu'elle soit.

En effet, le test du Khi-2 cherche une adéquation entre un échantillon et une loi de probabilité discrète, alors que le test de Kolmogorov-Smirnov cherche lui une adéquation entre un échantillon et une loi de probabilité continue. La GPD étant une loi continue, nous nous fions alors aux résultats du test de Kolmogorov-Smirnov.

Quoi qu'il en soit, avec un seuil de première espèce fixé à 5%, nous observons ainsi qu'il n'y a pas de présomption contre l'hypothèse « la queue de distribution associée à la base d'étude suit une GPD ». De même, en utilisant l'estimateur de Hill, on constate qu'il n'y a également pas de présomption contre l'hypothèse « la queue de distribution associée à la base d'étude suit une loi de Pareto ». Par contre, l'hypothèse « la queue de distribution associée à la base d'étude suit une loi de Fréchet » présente une forte suspicion de rejet. Ces analyses confirment les conclusions issues de l'étude graphique d'adéquation des lois de type II sur la sinistralité grave observée.

En remontant toute la procédure appliquée dans cette section, le non-rejet du test de Kolmogorov-Smirnov signifie le non-rejet de l'hypothèse « 150 000€ est un seuil de gravité statistiquement fiable ».

2.3 Pistes d'amélioration sur la détermination du seuil

Dans un souci de rapprocher l'interprétation du mode de détermination d'un seuil de gravité de la réalité, nous pouvons adapter la problématique de définition d'un sinistre grave à l'évolution de la sinistralité dans le temps. En effet, les sinistres RCM les plus importants étant longs (jusqu'à plusieurs années, voire décennies), il pourrait être intéressant d'inclure dans nos méthodes d'identification des sinistres graves cet aspect « temps ». Cet aspect-ci peut s'analyser sous deux pistes de réflexion :

- la dérive des montants d'indemnisation dans le temps ;
- et la dérive des nombres de sinistres dans le temps.

Enfin, une méthode alternative sera présentée ci-contre pour déterminer un seuil sans appliquer la théorie des valeurs extrêmes.

2.3.1 Réflexion sur la dérive du seuil dans le temps

Le coût de la sinistralité évolue dans le temps, et son impact peut être conséquent dans les situations où l'assureur doit indemniser sur une durée relativement longue. Dans le milieu de la santé, les séquelles qui font intervenir la responsabilité civile sont susceptibles de l'être à vie et peuvent faire l'objet d'indemnisations étalées sur plusieurs années, voire décennies sous ces conditions pour les cas les plus jeunes.

En RCM, la dérive des sinistres dans le temps devient alors un enjeu majeur, tant pour les cas attritionnels que les cas graves, pour améliorer la qualité des données et *in fine* la qualité prédictive des évènements rares de cette branche.

FIGURE 2.13 – Évolution dans le temps des charges cumulées moyennes avec un seuil de gravité 150 000€

Remarque : pour une meilleure comparaison entre charges cumulées moyennes totales, attritionnelles et graves, les séries de données ont été mises à une même échelle.

À titre d'illustration, appliquer « naïvement » le seuil estimé jusqu'à présent sur les arrêtés antérieurs montre une importante variation des charges graves. De plus, étant donné que les charges dépendent des provisions appliquées sur les sinistres en vue de prédire la sinistralité future, les pics de variation des moyennes s'expliquent en partie par l'évolution des modalités de gestion du portefeuille et les entrées et sorties de sinistres exceptionnels (*cf.* analyses descriptives en section 2.1.3).

On note également que cette variation est relativement indépendante de l'échantillon de sinistres observés, ce qui relève en ce sens l'importance que prend le rôle du temps dans l'évolution de l'indemnisation moyenne des sinistres. Ainsi, l'objectif de cette adaptation serait de déterminer un indice d'inflation sur les indemnisations observées dans le temps tel qu'un sinistre survenu dans le passé puisse présenter les mêmes impacts financiers que s'il était survenu aujourd'hui : c'est la mise en *as-if* .

Pour concevoir un tel indice, l'idéal serait de pouvoir quantifier tous les facteurs externes susceptibles d'impacter le coût d'un sinistre. Par exemple, dans le cas de la responsabilité civile médicale, on peut supposer que le coût des sinistres dépend non seulement de l'inflation monétaire mais aussi d'inflations « non-monétaires » telles que :

- l'augmentation du prix du matériel médical, pas nécessairement corrélé avec l'inflation monétaire car dépend en majeure partie des progrès de la Médecine ;
- l'augmentation des salaires, qui impacterait non seulement les frais d'expertise médicale mais aussi le coût d'intervention d'une tierce personne dans la durée ou encore le « manque à gagner salarial » que subirait une victime en arrêt de travail ;
- ou l'augmentation de dépenses de santé comme les frais d'hospitalisation, probablement plus corrélée avec l'indice des prix à la consommation qu'avec l'inflation monétaire.

Mais avec les bases de données à disposition, nous n'avons pas accès à un niveau d'information suffisamment précis pour permettre d'établir de façon fiable une différenciation de tels cas en fonction du type de sinistre rencontré. Des hypothèses de simplification seront alors effectuées.

En termes de nombre de dossiers, étant donné que les provisions influencent le niveau de charge, il peut arriver des cas où un sinistre perd sa « gravité » suite à une correction du niveau de provisionnement pour diverses raisons qui sont propres à chaque dossier. Pour limiter ce risque, il peut être également intéressant de relever l'évolution des sinistres graves entrants afin de conserver en tout temps l'ensemble des sinistres qui ont un volume de données stable dans le temps : on dénommera cette règle de gestion « grave un jour, grave toujours ».

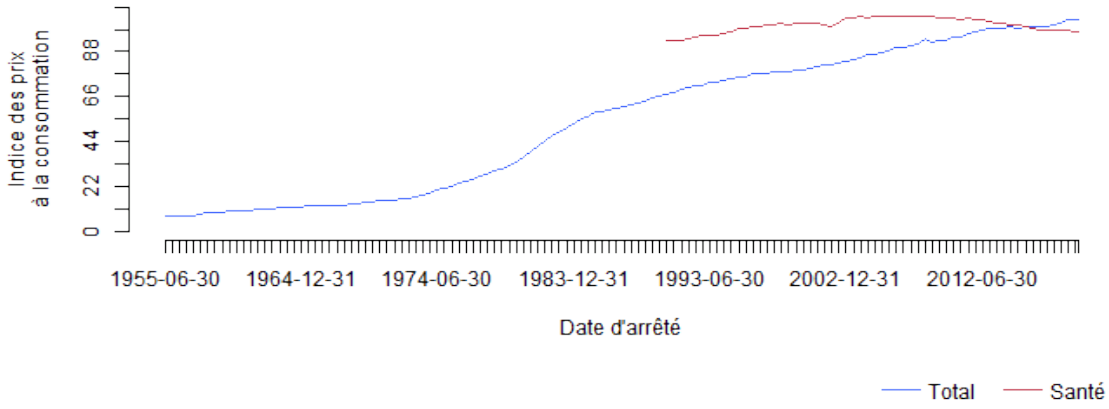
2.3.2 Mise en *as-if* du portefeuille

La mise en *as-if* des charges cumulées du portefeuille se déroulera en deux étapes :

1. construction d'un indice d'inflation adapté aux données ;
2. et application de cet indice.

Comme évoqué précédemment, le faible niveau d'informations du portefeuille à disposition ne permet pas une analyse relativement complète qui identifierait un indice d'inflation qui dépendrait idéalement des garanties activées par le contrat d'assurance RCM souscrit. Ne disposant que des données financières sans aucune information claire liée aux contrats, la problématique de la construction d'un indice d'inflation se résumera à la restitution de l'indice des prix à la consommation.

FIGURE 2.14 – Évolution des principaux indices de prix à la consommation (source : data.imf.org)



Il aurait été intéressant d'affiner cet indice sur les frais de santé, mais il s'avère que ces derniers ne sont pas disponibles sur un historique suffisamment long pour couvrir l'ensemble des dates de survenance de sinistres (*cf.* annexe B). Afin d'éviter le risque de créer des biais par changement de méthode de construction de l'indice d'inflation, nous nous contenterons alors de considérer l'indice « général » des prix à la consommation comme indice de référence.

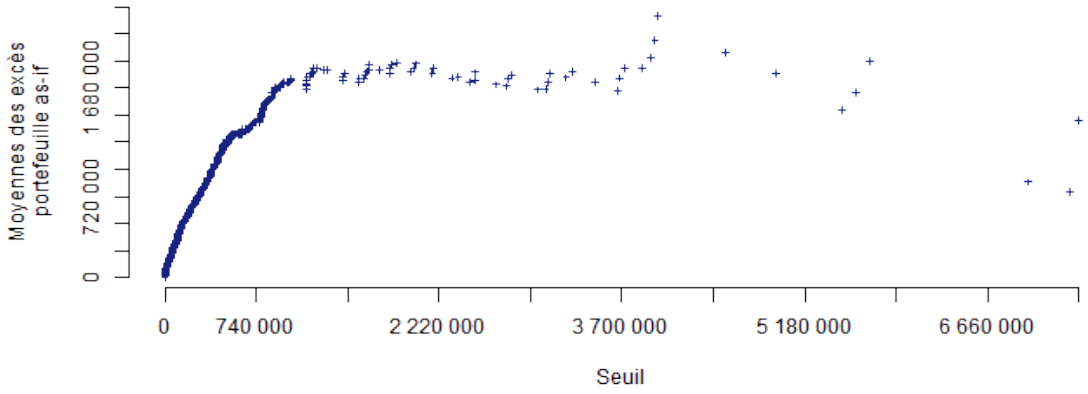
En posant $IPC(i)$ l'indice des prix à la consommation au semestre $i \in \mathbb{N}^*$, sachant que le semestre « 1 » débute au 01/07/1949, les montants enregistrés aux semestres de flux i $(x_k(i))_{\substack{(i,k)\in\mathbb{N}^*\times[\![1,n]\!] \\ n\in\mathbb{N}^*}}$ sont mis en *as-if* selon la relation suivante :

$$\forall (i,j,k) \in (\mathbb{N}^*)^3, i \geq j : x_{k,\text{asif}}(i) = x_k(j) \cdot \frac{IPC(i)}{IPC(j)} \tag{2.32}$$

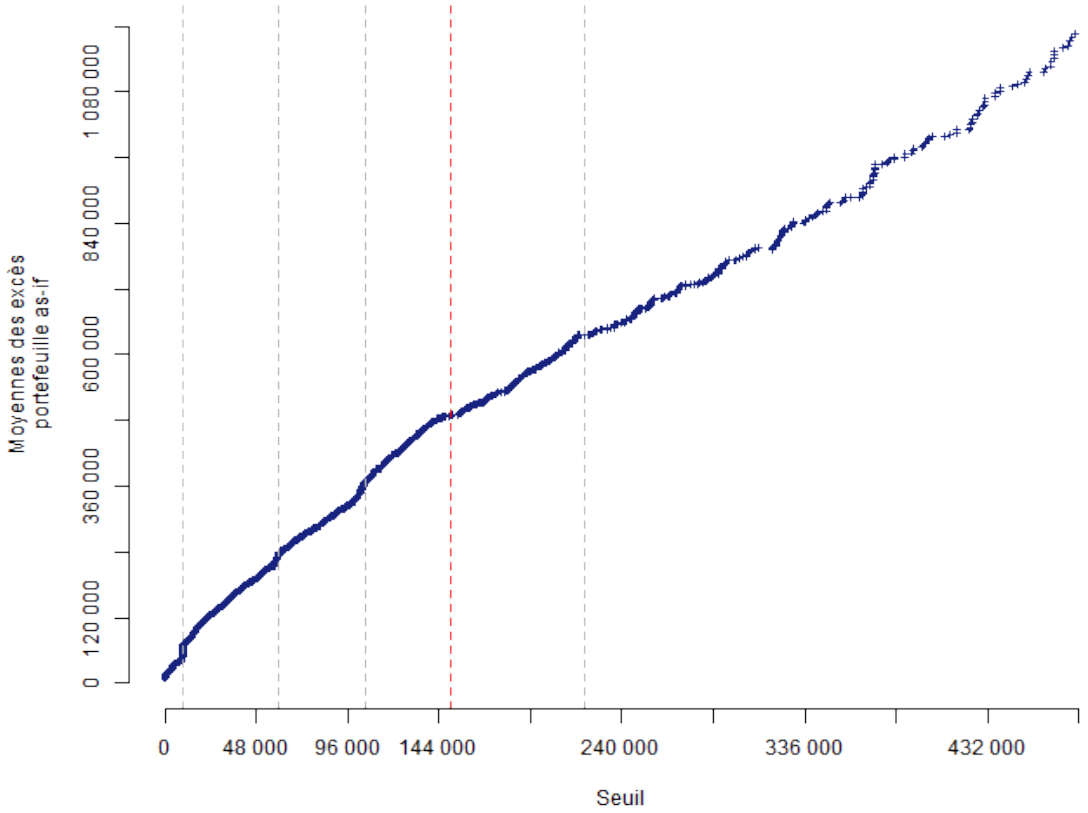
Une fois les montants réajustés, ils sont recumulés et on réapplique les méthodes d'estimation du seuil comme précédemment.

FIGURE 2.15 – *ME-plot* sur charges cumulées *as-if*

(a) Vue d'ensemble



(b) Analyse sur la plage de seuils $[0, 500\ 000]$



Sous ces conditions, on remarque dans l'ensemble que les points s'alignent légèrement « mieux » que sans mise en *as-if* du portefeuille. On peut imaginer en ce sens que ce procédé d'ajustement des montants influe positivement la modélisation des extrêmes.

TABLE 2.18 – Distinction selon la mise en *as-if* ou non des répartitions conditionnées par le seuil de gravité

Base de données	Historique	*as-if*
Part des dossiers avec charge cumulée $\geq$ 150 000	0,851%	0,759%
Part des charges cumulées $\geq$ 150 000	48,82%	50,08%

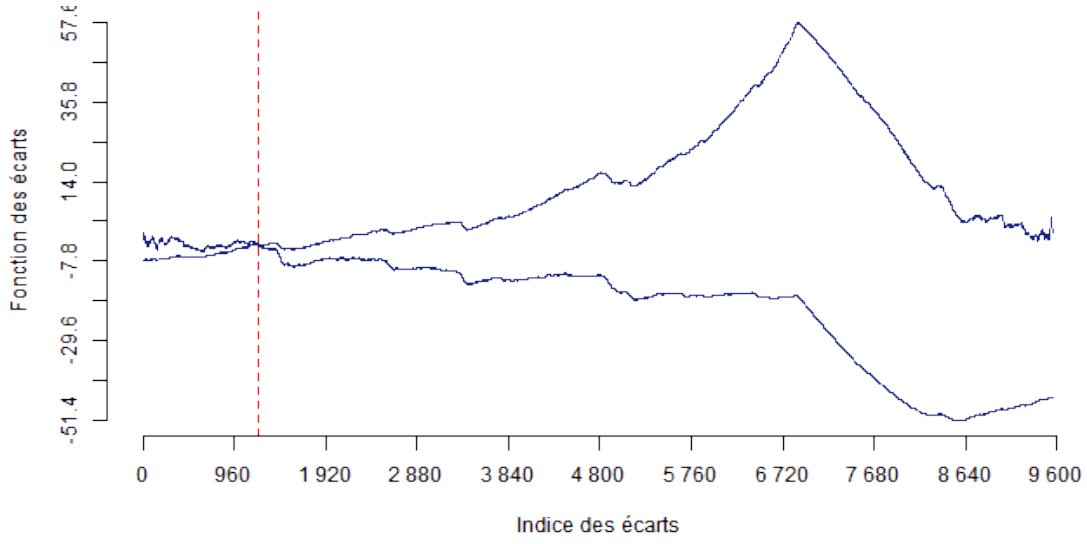
Toutefois, ces modifications de tendances maintiennent le seuil 150 000€ comme un seuil graphiquement fiable. L'impact de la mise en *as-if* devrait alors être, dans le cadre de cette étude, plus significative sur le provisionnement que sur la modélisation du seuil de gravité seule.

En effet, le seuil de gravité étant estimé au 31/03/2019 et plutôt que de re-estimer le seuil à chaque date d'arrêté antérieure, on pourra se contenter d'indexer directement ce seuil dans le temps et provisionner rétrospectivement grâce à cette indexation établie si nécessaire.

2.3.3 Le *GW-plot*, une solution simple et fiable ?

Comme présenté en annexe E.4, le diagramme de Gerstengarbe et Werner peut être envisageable comme solution alternative à l'analyse statistique d'un seuil de gravité.

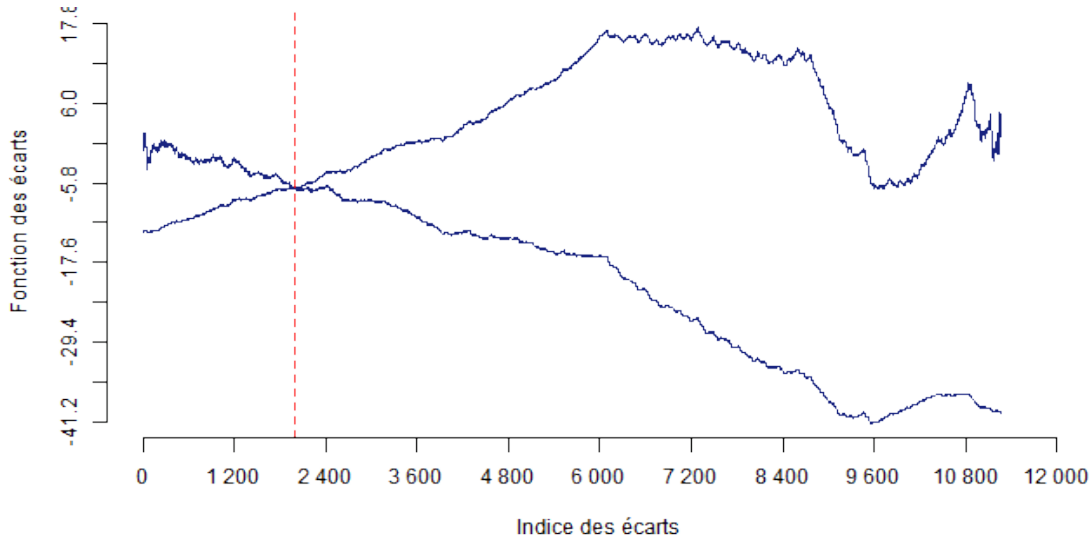
FIGURE 2.16 – *GW-plot* sur charges cumulées observées, réparties sur la tranche [50 000, 250 000]



Remarque : le diagramme de Gerstengarbe et Werner n'est pas effectué sur la totalité des données par souci de temps de calcul. L'algorithme associé à cette méthode graphique présente en effet une complexité vraisemblablement polynomiale.

De ce graphe, on voit par un simple croisement de deux courbes la frontière qui délimite un changement de comportement sur la sinistralité. En relevant les distances absolues, le croisement correspondant à l'endroit où celle-ci est minimale, on déduit de cette méthode que le seuil de gravité se situe au 1 213ème écart ordonné sur les données sélectionnées. Ce rang correspond à un seuil de 157 950€.

FIGURE 2.17 – *GW-plot* sur charges cumulées *as-if*, réparties sur la tranche [50 000, 250 000]



De même, ici le croisement se déroule au 2 004ème écart ordonné sur la plage de charges cumulées définies. Il correspond à un montant de 143 549,20€.

Avec ces deux résultats, on s'aperçoit ainsi que :
- non seulement les seuils estimés sont relativement proches, ce qui conforte la conjecture précédemment émise qui supposait un faible impact de la mise en *as-if* du portefeuille sur la détermination du seuil de gravité ;
- mais surtout que ces seuils sont également proches du seuil estimé par méthodes statistiques.

Dans de telles conditions, quelle méthode doit alors être privilégiée ?

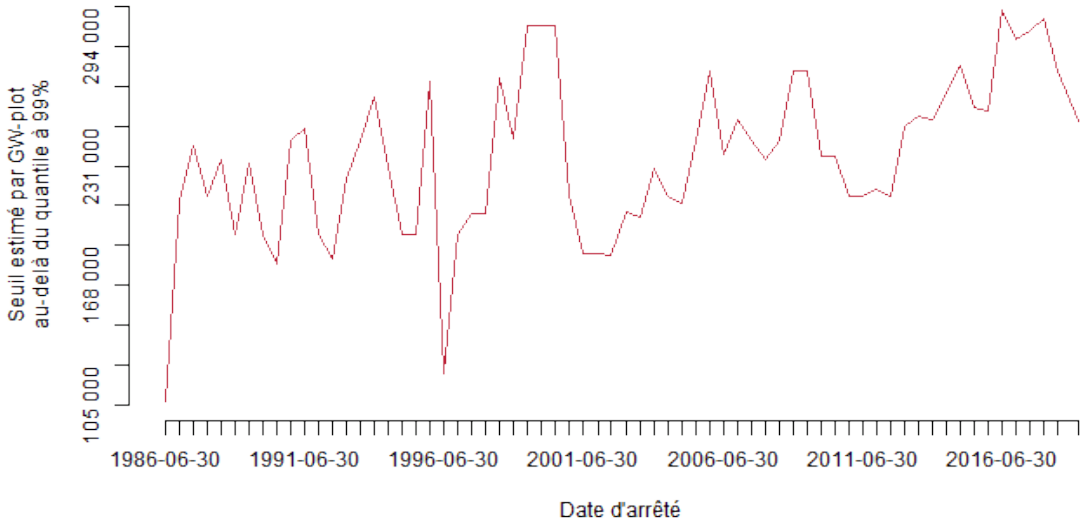
Le *GW-plot* est simple à comprendre et fait preuve d'une interprétabilité très accessible : en ordonnant les données, à partir du moment où on rentre dans la région des évènements rares, la volatilité augmente de manière significative. Des écarts entre les données ordonnées sont plus alors importants que le reste, que ce soit dans le sens

croissant ou décroissant. Graphiquement, cette analyse du comportement des queues de distribution se traduit par le croisement des deux courbes ascendantes et descendantes selon une fonction des écarts proche du test séquentiel de tendance de Mann-Kendall.

Toutefois, comme l'approche dichotomique présentée en section 2.2.1, cette méthode ne fait pas appel aux théories des valeurs extrêmes et reste uniquement dépendante des données observées. Elle peut alors :

- présenter une forte volatilité dans le temps ;
- et fournir des résultats biaisés en cas d'absence ou de censure de données inadéquate ou involontaire.

FIGURE 2.18 – Évolution des seuils estimés par la méthode de Gerstengarbe et Werner sur le dernier centile dans le temps



À titre d'exemple, si on ne considère que le dernier centile des bases de sinistres historiques, on obtient par cette méthode non seulement des seuils de gravité qui sont doubles comparés à ceux estimés jusqu'ici, mais également très volatiles dans le temps. C'est principalement pour cette raison que les résultats de cette méthode ne peuvent suffire à eux seuls. La bonne connaissance du portefeuille étudié est nécessaire pour une manipulation des données adéquate aux méthodes d'analyses employées. Sous ces conditions, cette méthode demeure un bon indicateur.

Finalement, pour l'application des méthodes de provisionnement, nous fixerons le seuil de gravité à **150 000€**, démontré statistiquement fiable.

Chapitre 3

Provisionnement des sinistres graves

3.1 Principes de modélisation adaptés à la responsabilité civile médicale

Les compagnies d'assurances ont une activité qui repose sur un cycle de production inversé : les primes sont encaissées pour compenser des sinistres inconnus en date de versement. Ceci explique pourquoi les bilans et comptes de résultats de telles structures sont constitués différemment de ceux des entreprises qui exercent une activité propre à d'autres secteurs. Les assureurs doivent constituer des provisions techniques en vue d'anticiper la sinistralité future à laquelle ils devront faire face.

Dans un tel contexte, des réglementations spécifiques aux assurances sont adoptées pour cadrer des conditions dans lesquelles les compagnies, mutuelles et autres entités juridiques qui exercent une activité d'assurance doivent se conformer pour minimiser au mieux le risque de faire défaut face à l'incertain.

TABLE 3.1 – Évolution du bilan sur activité d'assurance entre les directives Solvabilité I et Solvabilité II

(a) Bilan Solvabilité I (comptes sociaux)

<table>
<tr><th>Actif</th><th colspan="2">Passif</th></tr>
<tr><td rowspan="2">Plus-values latentes</td><td rowspan="2">Fonds propres</td><td>Excédent de fonds propres</td></tr>
<tr><td rowspan="2">Actifs en valeur nette comptable</td><td>Marge de solvabilité</td></tr>
<tr><td>Provisions techniques à charge des réassureurs</td><td colspan="2">Provisions techniques</td></tr>
</table>

(b) Bilan Solvabilité II (comptes prudentiels)

<table>
<tr><th>Actif</th><th colspan="2">Passif</th></tr>
<tr><td rowspan="3">Actifs en valeur de marché</td><td rowspan="3">Fonds prop.</td><td>Excédent de fonds propres</td></tr>
<tr><td>SCR</td></tr>
<tr><td>MCR</td></tr>
<tr><td rowspan="2">Provisions techniques à charge des réassureurs</td><td rowspan="2">Prov. tech.</td><td>Marge pour risque</td></tr>
<tr><td>Flux futurs Best Estimate</td></tr>
</table>

Nos applications de méthodes de provisionnement s'effectueront dans ce chapitre en comptes prudentiels. Selon l'article 75 de la directive Solvabilité II[1], ces provisions doivent être évaluées de façon cohérente au marché avec :

- une composante qui calcule la meilleure estimation des flux futurs (*Best Estimate*) ;
- et une composante qui calcule la marge pour risque (*Risk Margin*).

Toujours selon la même directive, l'article 77 stipule qu'une meilleure estimation correspond à « *la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinents* ». Pour

1. [Parlement Européen et Conseil de l'Union europénne, 2009]

cela, il est possible d'effectuer cette estimation à partir d'un triangle de liquidation $(X_{i,j})_{\substack{(i,j)\in[\![0,n]\!]\times[\![0,m]\!] \\ (n,m)\in(\mathbb{N}^*)^2}} \in \mathbb{R}^{(n+1)\times(m+1)}$, objet de ce chapitre.

TABLE 3.2 – Illustration d'un triangle de liquidation

		Périodes de développement				
		0	$\cdots$	j	$\cdots$	m
Périodes de sinistralité	0	$X_{1,1}$	$\cdots$	$X_{1,j}$	$\cdots$	$X_{1,m}$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\iddots$	
	i	$X_{i,1}$	$\cdots$	$X_{i,j}$		
	$\vdots$	$\vdots$	$\iddots$		*Liquidations à estimer*	
	n	$X_{n,1}$				

Un triangle de liquidation est une représentation de la liquidation des sinistres dans le temps. Pour chaque année de survenance, il sert au traitement comptable des PSAP et à l'application de méthodes actuarielles pour l'estimation des IBNR et ultimes de sinistres. Un triangle de liquidation peut être en représentation de montants décumulés $(X_{i,j})_{i,j}$ ou cumulés $\left(C_{i,j} = \sum_{k=0}^{j} X_{i,k}\right)_{i,j}$.

Les estimations dans un triangle de liquidation consistent à compléter la partie « inférieure » avec des cadences de développement, définies selon diverses méthodes actuarielles, afin d'obtenir un total de montants à l'ultime. Ces méthodes peuvent faire intervenir des calculs stochastiques ou non.

Un processus d'actualisation de ces sommes prédites nous fournit un montant de *Best Estimate*.

3.1.1 Conventions de provisionnement autour de la loi « About »

Dans le cadre de la responsabilité civile médicale, la construction de triangles de liquidation se doit de distinguer la nature juridique des sinistres, modifiée depuis la loi « About » (*cf.* tableau 1.1). Les triangles obtenus pour le calcul des provisions seront alors scindés en deux sous-triangles de provisionnement :

un triangle « base survenance » : il concerne l'ensemble des sinistres survenus avant la loi « About » et font l'objet de provisions IBNER et IBNYR ;

et un triangle « base réclamation » : il concerne les sinistres survenus après la loi « About » et font l'objets de provisions IBNER uniquement car les sinistres inconnus de l'assureur à la clôture d'un contrat referont surface dans les modèles de provisionnement au moment de la réclamation.

En fonction de leur nature juridique, des reclassements sont effectués afin de constituer les provisions pour couvrir la garantie subséquente. Une fois les triangles constitués et les ultimes obtenus, les montants seront alors répartis par type de provision puis restitués au bilan prudentiel.

Toutefois, une contrainte majeure persiste pour pouvoir distinguer les triangles correctement : nous ne disposons pas de la date de déclaration des sinistres, synonyme de date d'ouverture, ni les dates de validité des contrats (ouverture et fermeture éventuelles). Par contre, la base de données fournit un indicateur de classification des sinistres par le service de gestion. Nous nous contenterons alors de se fier à cette donnée car chercher à la redéterminer, au vu du peu de données à disposition, peut potentiellement être source d'erreurs.

Une fois la distinction effectuée, les triangles seront construits selon les conventions suivantes :

- provisionnant sur des sinistres graves, donc avec un faible volume de données, les périodes de flux des triangles de liquidation seront semestrielles afin de se procurer un volume statistique « artificiellement » convenable ;
- les flux de la base survenance sont rattachés au semestre de survenance pour les triangles de liquidation ;
- les flux des bases réclamation et subséquente sont rattachés au semestre de réclamation s'il est renseigné, survenance sinon (on fera en ce sens l'hypothèse que survenance et réclamation sont confondus) ;
- règlements, provisions pour règlements et recours seront pris en compte dans des triangles de liquidation regroupés pour définir les triangles de charges et de coûts (*cf.* respectivement les équations 2.3 et 2.4 pour leur formule de calcul). On notera à ce titre que les provisions pour recours ne sont pas prises en compte car disponibles uniquement dans les bases historiques et non dans la base à provisionner.

Après application de ces conventions, il ne restera plus qu'à estimer les charges et coûts ultimes et en déduire les provisions à constituer.

En normes sociales, les provisions doivent être ici segmentées entre base survenance et base réclamation, hors subséquentes futures. Par contre, en normes prudentielles, les provisions restent segmentées mais elles doivent inclure les subséquentes futures. Le calcul des provisions Solvabilité II s'effectuera alors sur la base d'un modèle de provisionnement comptable ajusté des provisions pour risques subséquents (soit le triangle *base subséquente*) et de réévaluations éventuelles au cas-par-cas des sinistres particuliers, en concertation avec le service de gestion.

3.1.2 Méthodes d'estimation déterministes

Pour évaluer les charges et coûts à l'ultime, il existe diverses méthodes relativement simples et populaires qui n'utilisent aucun calcul stochastique et s'appuient directement sur les données.

Pour la suite, on considèrera un triangle de liquidation de dimension finie $n \times m, (n,m) \in (\mathbb{N}^*)^2$ qui suit les notations suivantes :

- $C_{i,j}$ correspond à un montant cumulé en période d'exercice i et en période de développement j ;
- $\mathcal{D}_{n,m} = \{C_{i,j}; (n,m) \in (\mathbb{N}^*)^2, i+j \leq n, 0 \leq j \leq m\}$ l'ensemble des montants cumulés connus (autrement dit le « triangle supérieur »).

Méthode de *Chain-Ladder*

La méthode déterministe la plus répandue est la méthode de *Chain-Ladder*. Elle puise sa popularité dans sa grande simplicité et adaptabilité : elle fonctionne en effet avec n'importe quel type de triangle de liquidation (nombre de dossiers, charges, coûts, *etc.*) et dans tous les secteurs de l'assurance.

Définition 5 (Facteur de développement) *Soit $(C_{i,j}) \in \mathcal{D}_{n,m}$ un ensemble de montants cumulés. Les facteurs de développement suivent la relation suivante :*

$$\begin{aligned} f : [\![0,n]\!] \times [\![0,m-1]\!] &\to \mathbb{R} \\ (i,j) &\mapsto \frac{C_{i,j+1}}{C_{i,j}} \end{aligned} \tag{3.1}$$

Pour que la méthode de *Chain-Ladder* soit applicable, deux hypothèses doivent être vérifiées :

- les montants cumulés $C_{i,j}$, pour une période d'exercice i retenue, sont indépendants ;
- en considérant les moments d'ordre 1, il existe des facteurs de développement individuels $(f_j)_{j \in \mathbb{N}}$ qui vérifient :

$$\begin{aligned} \forall (i,j) \in [\![0,n]\!] \times [\![0,m]\!] : \mathbb{E}[C_{i,j} | C_{i,0}, \cdots, C_{i,j-1}] &= \mathbb{E}[C_{i,j} | C_{i,j-1}] \\ &= f_{j-1} \cdot C_{i,j-1} \end{aligned} \tag{3.2}$$

Ces deux hypothèses peuvent se résumer au fait que les facteurs de développement individuels doivent être indépendants de la période de sinistralité considérée (ici semestre de survenance ou de réclamation en fonction de la base de rattachement considérée). Cette hypothèse forte se traduit par la relation suivante :

$$\forall (n,j) \in \mathbb{N}^* \times [\![0,n-1]\!] : f(0,j) = f(1,j) = \cdots = f(n-j-1,j) \tag{3.3}$$

Sous ces hypothèses, nous obtenons alors :

$$\begin{aligned} \forall i \in [\![0,n]\!] : \mathbb{E}[C_{i,m} | \mathcal{D}_{n,m}] &= \mathbb{E}[C_{i,m} | C_{i,n-i}] \\ &= C_{i,n-i} \cdot \prod_{k \in [\![n-i,m-1]\!]} f_k \end{aligned} \tag{3.4}$$

Nous pouvons ainsi constituer les réserves *Best Estimate* pour une période d'exercice i, vues en période m et sachant $\mathcal{D}_{n,m}$ selon la relation suivante :

$$\forall i \in [\![0,n]\!] : \mathbb{E}[C_{i,m}|\mathcal{D}_{n,m}] - C_{i,n-i} = C_{i,n-i} \cdot \left(\prod_{k \in [\![n-i,m-1]\!]} f_k - 1 \right) \tag{3.5}$$

De cette démarche, on déduit ainsi les coefficients de *Chain-Ladder* qui permettent de calculer les montants à l'ultime.

Définition 6 (Estimateurs de *Chain-Ladder*) *Soit $(C_{i,j}) \in \mathcal{D}_{n,m}$ un ensemble de montants cumulés. Les facteurs de développement de la méthode de Chain-Ladder sont définies, pour chaque période de développement, par la fonction suivante :*

$$\begin{aligned} \hat{f}^{CL} : [\![0, n-1]\!] &\to \mathbb{R} \\ j &\mapsto \frac{\sum_{i=0}^{n-j+1} C_{i,j+1}}{\sum_{i=0}^{n-j+1} C_{i,j}} \end{aligned} \tag{3.6}$$

Ces facteurs de développement nous fournissent directement :

- les montants ultimes par exercice i :

$$\hat{C}_{i,m}^{CL} = C_{i,m-i} \cdot \prod_{j=m-i}^{m-1} \hat{f}^{CL}(j) \tag{3.7}$$

- les provisions par exercice i :

$$\hat{R}_i^{CL} = \hat{C}_{i,m}^{CL} - C_{i,m-i} \tag{3.8}$$

- et les provisions totales :

$$\hat{R}^{CL} = \sum_{i=1}^{n} \hat{R}_i^{CL} \tag{3.9}$$

Méthode de *London-Chain*

Méthode introduite par S. Benjamin et L. M. Eagles dans une publication sur le calcul des réserves de l'assureur Lloyd's of London en 1986 [2], elle a l'avantage d'être moins contraignante que la méthode de *Chain-Ladder* avec une hypothèse, moins forte, de l'existence d'une fonction affine qui traduit la dynamique des montants cumulés par exercice :

$$\forall (n,j,i) \in \mathbb{N}^* \times [\![0,n-1]\!] \times [\![0,n-j-1]\!], C_{i,j+1} = \lambda_j \cdot C_{i,j} + \beta_j \tag{3.10}$$

où $(\lambda_j, \beta_j) \in \mathbb{R}^2$ sont des paramètres à estimer.

2. [Benjamin et Eagles, 1986]

Une fois cette hypothèse validée, les paramètres de la fonction affine sont déterminables selon la méthode des moindres carrés en minimisant la relation suivante :

$$\begin{aligned} &\forall n \in \mathbb{N}^*, \Delta : [\![0, n-1]\!] \to \mathbb{R} \\ &\qquad j \mapsto \sum_{i=0}^{n-j-1} (C_{i,j+1} - \beta_j - \lambda_j \cdot C_{i,j})^2 \end{aligned} \tag{3.11}$$

Définition 7 (Estimateurs de la méthode de *London-Chain*) *Soit* $(C_{i,j}) \in \mathcal{D}_{n,m}$ *un ensemble de montants cumulés. En posant les moyennes des montants cumulés sur les périodes de développement* j *et* $j+1$ *:*

$$\forall (n,j) \in \mathbb{N}^* \times [\![0, n-1]\!] : \begin{cases} \bar{C}_j &= \frac{1}{n-j} \cdot \sum_{i=0}^{n-j-1} C_{i,j} \\ \bar{C}_{j+1} &= \frac{1}{n-j} \cdot \sum_{i=0}^{n-j-1} C_{i,j+1} \end{cases} \tag{3.12}$$

les estimateurs des paramètres affines de la méthode de London-Chain λ_j *et* β_j *associés à la période de développement* j *sont définis par les fonctions suivantes :*

$$\begin{aligned} \hat{\lambda}^{LC} &: [\![0, n-1]\!] \to \mathbb{R} \\ &\qquad j \mapsto \frac{\frac{1}{n-j} \cdot \sum_{i=0}^{n-j-1} C_{i,j} \cdot C_{i,j+1} - \bar{C}_j \cdot \bar{C}_{j+1}}{\frac{1}{n-j} \cdot \sum_{i=0}^{n-j-1} C_{i,j}^2 - \bar{C}_j^2} \\ &\qquad n-1 \mapsto \frac{C_{0,n}}{C_{0,n-1}} \end{aligned} \tag{3.13}$$

$$\begin{aligned} \hat{\beta}^{LC} &: [\![0, n-1]\!] \to \mathbb{R} \\ &\qquad j \mapsto \bar{C}_{j+1} - \hat{\lambda}^{LC}(j) \cdot \bar{C}_j \end{aligned} \tag{3.14}$$

Remarque : en posant $\text{col}_j = (C_{i,j})_{i \in [\![0,j]\!]}$ la colonne du triangle de liquidation associée à la période de développement j, on observe que $\hat{\lambda}^{LC}(j) = \frac{\text{Cov}(\text{col}_j, \text{col}_{j+1})}{\text{Var}(\text{col}_j)}$.

Les charges à l'ultime selon cette méthode sont alors déterminées, pour un exercice donné i, en utilisant la relation de récurrence suivante :

$$\hat{C}_{i,m}^{LC} = \hat{\lambda}^{LC}(n) \cdot \hat{C}_{i,n-1}^{LC} + \hat{\beta}^{LC}(n) \tag{3.15}$$

Méthode de Bornhuetter-Ferguson

Développée par Ronald L. Bornhuetter et Ronald E. Ferguson en 1972[3], cette méthode introduit un paramètre exogène qui contrebalance la forte dépendance de la méthode de *Chain-Ladder* avec la dernière position connue des triangles de liquidation. En

3. [Bornhuetter et Ferguson, 1972]

effet, si cette dernière s'avère être nulle pour une quelconque raison, alors sa charge à l'ultime sera nulle. La méthode déterministe de Bornhuetter-Ferguson trouve alors tout son intérêt à être appliquée avec des triangles qui présentent des données instables.

Toutefois, pour que cette méthode soit applicable, les hypothèses suivantes doivent être vérifiées :

- pour chaque période d'exercice i, les montants cumulés $C_{i,j}$ doivent être indépendants de la période de développement j considérée ;
- il existe des paramètres $(\mu_i)_{i\in[\![0,n]\!]} \in (\mathbb{R}_+^*)^{n+1}$ et des cadences de montants cumulés $(\beta_i)_{i\in[\![0,m]\!]} \in (\mathbb{R}_+^*)^{m+1}$ avec $\beta_m = 1$ qui vérifient la relation suivante :

$$\begin{aligned} \forall (i,j,k) \in [\![0,n]\!] \times [\![0,m-1]\!] \times [\![0,m-j]\!] : \mathbb{E}[C_{i,0}] &= \beta_0 \cdot \mu_i \\ \mathbb{E}[C_{i,j+k} | C_{i,0}, \cdots, C_{i,j}] &= C_{i,j} + (\beta_{j+k} - \beta_j) \cdot \mu_i \end{aligned} \tag{3.16}$$

Sous ces hypothèses, on obtient alors les résultats suivants :

$$\forall (i,j) \in [\![0,n]\!] \times [\![0,m]\!] : \begin{cases} \mathbb{E}[C_{i,j}] &= \beta_j \cdot \mu_i \\ \mathbb{E}[C_{i,m}] &= \mu_i \end{cases} \tag{3.17}$$

Avec ces deux relations, la méthode de Bornhuetter-Ferguson déduit que les montants cumulés à l'ultime s'évalueraient ainsi :

$$\mathbb{E}[C_{i,m} | \mathcal{D}_{n,m}] = C_{i,n-i} + (1 - \beta_{n-i}) \cdot \mu_i \tag{3.18}$$

En faisant recours aux facteurs de développement issus de la méthode de *Chain-Ladder*, nous pouvons déduire les estimateurs de Bornhuetter-Ferguson.

Définition 8 (Estimateurs de la méthode de Bornhuetter-Ferguson) *Soit $(C_{i,j}) \in \mathcal{D}_{n,m}$ un ensemble de montants cumulés. Les estimateurs a priori μ_i^{BF} et a posteriori β_i^{BF} suivent les fonctions suivantes :*

$$\begin{aligned} \hat{\beta}^{BF} : [\![0,n-1]\!] &\to \mathbb{R} \\ j &\mapsto \prod_{k=j}^{n-1} \frac{1}{\hat{f}^{CL}(k)} \end{aligned} \tag{3.19}$$

$$\begin{aligned} \hat{\mu}^{BF} : [\![0,n]\!] &\to \mathbb{R} \\ i &\mapsto \hat{C}_{i,m} \end{aligned} \tag{3.20}$$

avec $\hat{C}_{i,m}$ un estimateur exogène du montant cumulé à l'ultime qui est fixé par avis d'expert.

En règle générale, cet estimateur exogène peut être déterminé à partir des *Loss-Ratio*. Néanmoins, dans le cadre de cet étude, n'ayant pas accès aux primes versées (bases de données exclusivement concentrées sur la sinistralité), il ne sera possible de fixer un tel estimateur selon cette méthode.

Ainsi, avec ces estimateurs, pour un exercice donné i, le montant à l'ultime se détermine également de façon récursive selon la relation suivante :

$$\hat{C}_{i,m}^{BF} = C_{i,n-i} + (1 - \hat{\beta}^{BF}(n-i)) \cdot \hat{\mu}^{BF}(i) \tag{3.21}$$

Contrairement aux méthodes de *Chain-Ladder* et de *London-Chain*, la méthode de Bornhuetter-Ferguson met à disposition un ajustement par avis d'expert qui ne rend pas cette approche déterministe complètement dépendante des données utilisées. Cependant, cet avantage signifie en contrepartie que l'issue de cette méthode est conditionnée par cet avis d'expert. En d'autres termes, nous courrons un risque de biaiser les estimations par un avis trop subjectif ou incohérent.

3.1.3 Méthodes d'estimation stochastiques

Les méthodes déterministes sont simple d'application, mais cette dimension non-prédictive nous empêche de quantifier un degré d'incertitude sur les réserves projetées. Ainsi, des méthodes d'estimation stochastiques ont progressivement vu le jour en considérant les montants connus comme des variables aléatoires. Dès lors, les théories basées sur les processus stochastiques sont applicables sur les triangles de liquidation en permettant notamment :

- d'émettre et de valider des hypothèses comportementales sur les triangles ;
- d'estimer et d'obtenir des paramètres à l'aide de démarches stochastiques, intervalles de confiance à l'appui ;
- ou encore de simuler la sinistralité future en connaissance des propriétés stochastiques qui émanent de cette simulation.

De l'erreur d'estimation

Pour mesurer l'incertitude liée aux méthodes de provisionnement stochastiques employées, nous utilisons l'erreur quadratique moyenne de prédiction (MSEP).

Définition 9 (Erreur quadratique moyenne de prédiction) *Soit $\mathcal{D} = (x_i)_{i\in\mathbb{N}}$ un ensemble d'observations issus d'une variable aléatoire X. La MSEP d'un prédicteur $\hat{X}$ de X se définit par la relation suivante :*

$$MSEP_{X|\mathcal{D}}(\hat{X}) = \mathbb{E}\left[(\hat{X} - X)^2 | \mathcal{D}\right] \tag{3.22}$$

Remarque : en fonction des propriétés propres au prédicteur $\hat{X}$ ou à la variable aléatoire associée X, diverses simplifications ou interprétations de la MSEP peuvent être déduites.

- Si $\hat{X}$ est $\mathcal{D}$-mesurable de X et un estimateur $\mathcal{D}$-mesurable de $\mathbb{E}[X|\mathcal{D}]$, alors la MSEP peut se décomposer ainsi :

$$MSEP_{X|\mathcal{D}}(\hat{X}) = \mathrm{Var}[X|\mathcal{D}] + \mathbb{E}\left[(\hat{X} - \mathbb{E}[X|\mathcal{D}])^2\right] \tag{3.23}$$

- Si X est indépendant de l'ensemble $\mathcal{D}$, alors on peut simplifier variance et espérance conditionnelles :

$$X \text{ indépendant de } \mathcal{D} \Rightarrow \begin{cases} \mathbb{E}[X|\mathcal{D}] & = \mathbb{E}[X] \\ \text{Var}[X|\mathcal{D}] & = \text{Var}[X] \end{cases} \tag{3.24}$$

- Si, de plus, $\hat{X}$ est un estimateur sans biais, l'erreur d'estimation se quantifie par la variance de $\hat{X}$:

$$\left.\begin{array}{l} X \text{ indépendant de } \mathcal{D} \\ \hat{X} \text{ estimateur sans biais de } \mathbb{E}[X] \end{array}\right\} \Rightarrow MSEP_{X|\mathcal{D}}(\hat{X}) = \text{Var}[X] + \text{Var}[\hat{X}] \tag{3.25}$$

- Sinon, dans le cas où X dépend de l'ensemble des observations $\mathcal{D}$, la MSEP peut s'écrire :

$$\begin{aligned} & X \text{ dépendant de } \mathcal{D} \\ \Rightarrow MSEP_{X|\mathcal{D}}(\hat{X}) =& \text{Var}[X] + \mathbb{E}[(\hat{X} - \mathbb{E}[X])^2] \\ & - 2 \cdot \mathbb{E}[(\hat{X} - \mathbb{E}[X]) \cdot (\mathbb{E}[X|\mathbb{D}] - \mathbb{E}[X])] \end{aligned} \tag{3.26}$$

- Si, de plus, $hatX$ est un estimateur de $\mathbb{E}[X]$ sans biais, l'erreur d'estimation se simplifie comme suit :

$$\begin{aligned} & \left.\begin{array}{l} X \text{ dépendant de } \mathcal{D} \\ \hat{X} \text{ estimateur sans biais de } \mathbb{E}[X] \end{array}\right\} \\ & \Rightarrow MSEP_{X|\mathcal{D}}(\hat{X}) = \text{Var}[X] + \text{Var}[\hat{X}] \\ & \qquad - 2 \cdot \text{Cov}(\hat{X}, \mathbb{E}[X|\mathcal{D}]) \end{aligned} \tag{3.27}$$

Finalement, de toutes ces relations, on dénote l'importance d'identifier correctement les estimateurs stochastiques afin d'en dégager convenablement des propriétés intéressantes qui guident à un juste calcul de la MSEP, socle de validation de la pertinence des méthodes de provisionnement stochastiques employées.

Pour la suite, on considèrera les mêmes notations que celles employées dans la section 3.1.2.

Méthode de Mack

Initiée par Thomas Mack en 1993 [4], cette méthode parvient à introduire une notion d'incertitude dans la méthode déterministe de *Chain-Ladder* tout en conservant les mêmes résultats. Il présuppose que les montants cumulés suivent un processus stochastique, avec une espérance et une variance notées respectivement $\mu_{i,j}$ et $\sigma^2_{i,j}$.

4. [Mack, 1993]

Sous cette méthode, trois hypothèses doivent être vérifiées :
- les paiements cumulés sont indépendants de l'année d'exercice observée, *i.e.* $C_{i,j}$ et $C_{k,j}$ sont indépendants si $i \neq k$;
- il existe des paramètres $(f_j)_{j\in[\![0,m-1]\!]} \in (\mathbb{R}_+^*)^m$ tels que :

$$\forall (i,j) \in [\![0,n]\!] \times [\![1,m]\!] : \mathbb{E}[C_{i,j}|C_{i,j-1}] = \mu_{i,j} = f_{j-1} \cdot C_{i,j-1} \tag{3.28}$$

- il existe des paramètres de variance $(\sigma_j^2)_{j\in[\![0,m-1]\!]} \in (\mathbb{R}_+^*)^m$ tels que :

$$\forall (i,j) \in [\![0,n]\!] \times [\![1,m]\!] : \mathrm{Var}[C_{i,j}|C_{i,j-1}] = \sigma_{i,j}^2 = \sigma_{j-1}^2 \cdot C_{i,j-1} \tag{3.29}$$

Les estimateurs des paramètres de la méthode de Mack se retrouvent par des procédés similaires à ceux appliqués dans les méthodes déterministes précédemment présentées.

Définition 10 (Estimateurs sans biais de Mack) *Soit $(C_{i,j}) \in \mathcal{D}_{n,m}$ un ensemble de montants cumulés. Les estimateurs sans biais de la méthode de Mack suivent, pour chaque période de développement, les fonctions suivantes :*

$$\begin{aligned} \hat{f}^M : [\![0,n-1]\!] &\to \mathbb{R} \\ j &\mapsto \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}} \end{aligned} \tag{3.30}$$

$$\begin{aligned} (\hat{\sigma}^M)^2 : [\![0,n-1]\!] &\to \mathbb{R} \\ j &\mapsto \frac{1}{n-j} \cdot \sum_{i=0}^{n-j-1} C_{i,j+1} \cdot \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}^M(j) \right)^2 \end{aligned} \tag{3.31}$$

De ces résultats, on remarquera la relation suivante, valable pour $m > 3$:

$$(\hat{\sigma}^M)^2(m-1) = \min\left(\frac{(\hat{\sigma}^M)^4(m-2)}{(\hat{\sigma}^M)^2(m-3)}, (\hat{\sigma}^M)^2(m-3), (\hat{\sigma}^M)^2(m-2) \right) \tag{3.32}$$

Les montants cumulés déduits $\hat{C}_{i,j}^M$ de ces estimateurs sont calculés de la même façon que la méthode de *Chain-Ladder*. Sous les hypothèses employées et les estimateurs posés, la MSEP de tels montants s'écrit ainsi :

$$\begin{aligned} MSEP_{\mathcal{D}_{n,m}}(\hat{C}_{i,n}^M) &= \mathbb{E}\left[(\hat{C}_{i,n} - C_{i,n})^2 | \mathcal{D}_{n,m} \right] \\ &= \mathrm{Var}[C_{i,n}|\mathcal{D}_{n,m}] + \left(\mathbb{E}\left[C_{i,n}|\mathcal{D}_{n,m} \right] - \hat{C}_{i,n}^M \right)^2 \end{aligned} \tag{3.33}$$

De manière équivalente à la méthode de *Chain-Ladder*, on pose :
- le montant des provisions estimées par exercice i selon la méthode de Mack :

$$\hat{R}_i^M = \hat{C}_{i,m}^M - C_{i,m-i} \tag{3.34}$$

- et le montant des provisions estimées totales :

$$\hat{R}^M = \sum_{i=1}^{n} \hat{R}_i^M \tag{3.35}$$

La provision à étudier $\hat{R}_i^M - R_i^M = \hat{C}_{i,m}^M - C_{i,m-i}$ nous fait déduire l'égalité entre la MSEP provisions et la MSEP montants estimés :

$$MSEP_{\mathcal{D}_{n,m}}(\hat{R}_i^M) = MSEP_{\mathcal{D}_{n,m}}(\hat{C}_{i,n}^M) \tag{3.36}$$

Ainsi, en posant $\hat{R}^M = \sum_{i=2}^{n} \hat{R}_i^M$ les provisions totales estimées par la méthode de Mack, l'erreur d'estimation sur ces provisions sont estimables par les relations suivantes :

$$\forall i \in [\![1,n]\!] : \widehat{MSEP}_{\mathcal{D}_{n,m}}(\hat{R}_i^M) = (\hat{C}_{i,n}^M)^2 \cdot \sum_{j=n-i-1}^{n-1} \frac{(\hat{\sigma}^M)^2(j)}{\left(\hat{f}^M(j)\right)^2} \cdot \left(\frac{1}{\hat{C}_{i,j}^M} + \frac{1}{\sum_{k=0}^{n-j} C_{k,j}} \right) \tag{3.37}$$

$$\widehat{MSEP}_{\mathcal{D}_{n,m}}(\hat{R}^M) = \sum_{i=1}^{n} \left(\widehat{MSEP}_{\mathcal{D}_{n,m}}(\hat{R}_i^M) + \hat{C}_{i,n}^M \cdot \sum_{k=i}^{n} \hat{C}_{k,n}^M \cdot \sum_{j=n-i}^{n-1} \frac{2 \cdot (\hat{\sigma}^M)^2(j)}{\left(\hat{f}^M(j)\right)^2 \cdot \sum_{k=0}^{n-j} C_{k,j}} \right) \tag{3.38}$$

Pour une meilleure interprétation du résultat, on pourra s'appuyer sur l'erreur de prédiction standard (SEP) relative, *i.e.* la racine de l'erreur quadratique moyenne de prédiction rapportée aux ultimes estimée :

$$\widehat{SEP}_{\mathcal{D}_{n,m}}^{rel}(\hat{R}^M) = \frac{\sqrt{\widehat{MSEP}_{\mathcal{D}_{n,m}}(\hat{R}^M)}}{\hat{R}^M} \tag{3.39}$$

Méthode *Bootstrap*

Cette méthode, initiée par Bradley Efron en 1979 [5], et de son appellation tirée de l'expression anglaise « *to pull oneself by one's bootstrap* » (se débrouiller sans l'aide d'autrui), tire sa renommée de sa capacité à générer des échantillons de données aléatoirement à partir d'un échantillon initial et avec un faible nombre d'hypothèses à respecter.

En effet, pour que cette méthode soit applicable, la seule condition à respecter est que les variables aléatoires utilisées pour cette méthode soient i.i.d.. Néanmoins, il s'avère que cette condition est rarement respectée dans la réalité, ce qui amène à ne considérer que les résidus des données qui ont beaucoup plus tendance à être i.i.d..

5. [Efron, 1979]

La méthode *Bootstrap* part d'une estimation déterministe choisie et applique des méthodes de ré-échantillonnage pour estimer les provisions futures. Pour un triangle de sinistres incrémentaux $(C_{i,j}) \in \mathcal{D}_{n,m}$ qui suivent une loi donnée, la méthode *Bootstrap* suit les étapes suivantes :

1. construction des estimateurs selon la méthode déterministe choisie ;
2. recalcul du triangle supérieur à partir de la dernière anti-diagonale connue $(C_{i,j})_{i+j=\min(n,m)+1}$ et des estimateurs ;
3. déduction des résidus $(r_{i,j})$ selon une méthode de calcul choisie. Les résidus de Pearson sont les plus répandus et seront appliqués dans le cadre de ce mémoire.

$$\forall (i,j) \in [\![1,n]\!] \times [\![i,m]\!] : r_{i,j}^P = \frac{C_{i,j} - \hat{C}_{i,j}}{\sqrt{\mathrm{Var}(\hat{C}_{i,j})}} \tag{3.40}$$

4. boucle selon un nombre de simulations fixé N :
   (a) ré-échantillonnage du triangle recalculé par replacement des résidus ;
   (b) calcul des facteurs de développement selon la méthode de *Chain-Ladder* ;
   (c) déduction du « triangle inférieur » ;
   (d) estimation de la provision et stockage du résultat ;
5. détermination à partir des N provisions obtenues de la distribution de la PSAP.

Cette démarche nous permet ainsi d'obtenir N estimations de la provision totale *Bootstrap* $(\hat{R}_k^B)_{k\in[\![1,N]\!]}$, et par une approximation de Monte-Carlo une espérance des provisions futures $\hat{R}^B = \frac{1}{N}\sum_{k=1}^{N} \hat{R}_k^B$. En fonction du modèle déterministe et des ré-échantillonnages, on dénote deux sources d'erreur de prédiction :

l'erreur d'estimation : s'introduit lors du ré-échantillonnage par remplacement des résidus ;

l'erreur de processus : s'introduit lors du tirage aléatoire des projections, obtenues après inversion des résidus ré-échantillonnés.

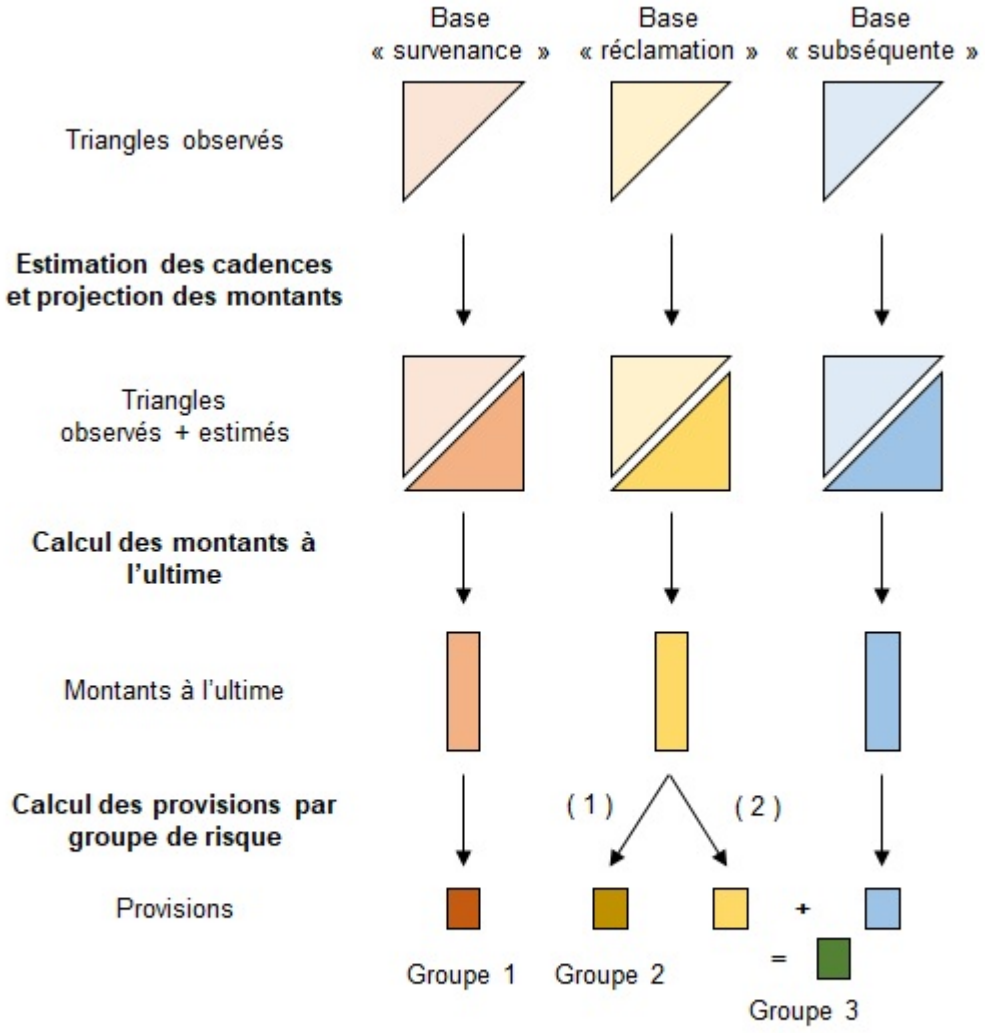
L'erreur de processus correspond à l'écart-type des provisions estimées.

3.2 Application sur les sinistres graves

Dans le contexte des sinistres graves en responsabilité civile médicale, le processus de provisionnement reste, dans les grandes lignes, identique aux autres branches de l'assurance IARD et responsabilité civile. Pour chaque modèle considéré, il faudra vérifier ses hypothèses, calculer les estimateurs et les appliquer pour obtenir une estimation des montants à l'ultime.

La spécificité des sinistres RCM réside dans les évolutions juridiques de cette branche qui nécessitent de distinguer plusieurs bases de provisionnement parmi l'ensemble des sinistres recensés, où des traitements différents sont effectués pour aboutir à différentes provisions prudentielles. En se basant sur les groupes de risque établis dans le tableau 1.1, les triangles de liquidation correspondants suivront la démarche suivante :

FIGURE 3.1 – Synthèse de la procédure de calcul des provisions en responsabilité civile médicale employée



(1) : sélection des sinistres avec date de réclamation, *i.e.* la ligne du triangle considéré, antérieure à la loi « About ».
(2) : sélection des sinistres avec date de réclamation postérieure à la loi « About ».

Nous pouvons noter que la division par groupe de risque pourrait s'effectuer dès la

construction des triangles supérieurs, mais on risquerait de perdre la notion de garantie subséquente qui présente un comportement *a priori* différent des autres garanties du contrat RCM. Cette disposition nous permettrait ainsi d'apporter de la prudence aux provisions constituées par une analyse de l'évolution des charges et coûts des sinistres dans le temps plus fine qu'en se basant uniquement sur les groupes de risque.

3.2.1 Construction des triangles de sinistres graves

Comme évoqué précédemment, les triangles de sinistralité seront décomposés en suivant un indicateur établi par le service de gestion. Des ajustements sont toutefois apportés sur cette segmentation pour adapter les triangles de liquidation à notre problématique :

- sélection des semestres de survenance / réclamation postérieurs au 01/01/1997 pour ne pas prendre en compte l'impact du changement de méthode de provisionnement de 1996 (*cf.* observations issues des figures 2.1 puis 2.2) ;
- sélection des dossiers qui ont, en date d'arrêté, une charge cumulée qui excède le seuil de gravité, fixé à 150 000€, que les montants soient historiques ou mis en *as-if*.

Une fois ces filtres appliqués, les triangles de sinistres graves sont construits en cumulés et en accord avec les conventions proposées en section 3.1.1. Par croisements entre la base de rattachement et le type de montant à provisionner, on recense un total de douze triangles de liquidation :

TABLE 3.3 – Montants totaux sur les triangles de liquidation construits

Base de rattachement	Survenance	Réclamation	Subséquente
Total triangle des charges	23 598 433	161 900 491	8 897 127
Total triangle des coûts	22 812 983	159 857 068	8 875 420
Total triangle des charges *as-if*	29 986 129	171 553 320	9 557 118
Total triangle des coûts *as-if*	29 076 249	169 302 267	9 534 055

On observe que les charges et les coûts présentent une différence relativement faible. En d'autres termes, les montants de recours paraissent peu significatifs sur l'ensemble du portefeuille de sinistralité grave. Nous nous contenterons pour la suite d'analyser que les résultats sur les charges.

3.2.2 Utilisation de méthodes déterministes

Vérification de l'hypothèse fondamentale

L'application de la quasi-totalité des méthodes de provisionnement citées précédemment requièrent une hypothèse commune : l'indépendance des périodes de sinistralité. Dans le cadre de ce mémoire, cette hypothèse est équivalente à s'assurer que les triangles considérés possèdent des charges cumulées qui sont indépendantes du semestre de survenance ou de réclamation considéré.

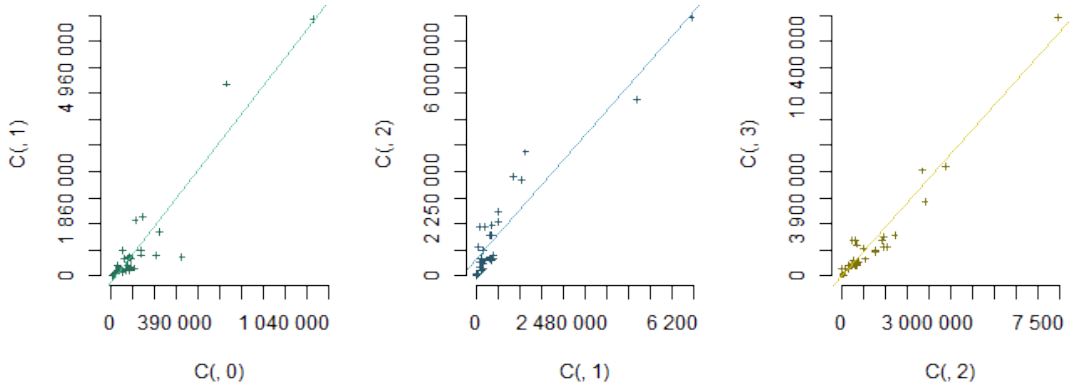
Cette vérification peut s'approximer par deux tests graphiques :

sur les couples de charges cumulées : les montants cumulés $(C_{i,j}, C_{i,j+1})_{i\in[\![0,n-j-1]\!]}$ doivent être sensiblement alignés avec une droite qui passe par l'origine ;

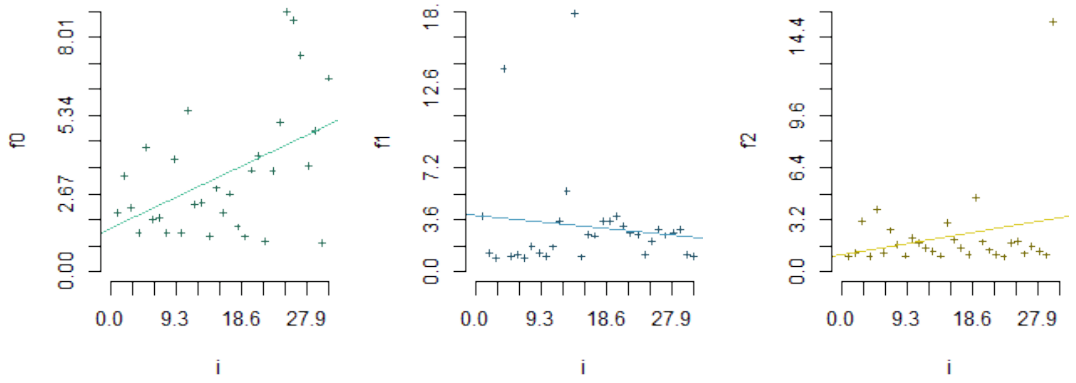
sur les facteurs de développement individuels : les $(f_j)_{j\in[\![0,m]\!]}$ doivent être sensiblement constants.

FIGURE 3.2 – Tests graphiques sur l'hypothèse de *Chain-Ladder*, triangle de liquidation des charges historiques, base réclamation

(a) Alignement des charges cumulées



(b) Alignement des facteurs de développement individuels



Remarque : les droites tracées dans les graphiques ci-contre correspondent aux régressions linéaires des points représentés.

À titre d'exemple, en considérant les charges cumulées de la base réclamation, nous observons sur les trois premiers semestres de développement que l'hypothèse n'est pas parfaitement vérifiée : les droites de régression sur les charges cumulées ne passent par exactement par l'origine et celles des facteurs de développement individuelles ne sont pas tout à fait plates. Néanmoins, plus nous avançons dans les semestres de développement, plus l'alignement des points observés est satisfaisant. Par principe, nous validons finalement cette hypothèse et poursuivons nos analyses.

Méthode de *Chain-Ladder*

TABLE 3.4 – Échantillon d'estimateurs *Chain-Ladder* obtenus

Semestre de développement	0	1	2	3	...	10	...	20	...
$\hat{f}^{CL}$ base survenance historique	2,864	3,266	1,451	1,446	...	1,048	...	0,998	...
$\hat{f}^{CL}$ base survenance *as-if*	3,310	2,989	1,426	1,430	...	1,043	...	0,998	...
$\hat{f}^{CL}$ base réclamation historique	4,051	1,832	1,544	1,416	...	1,049	...	1,055	...
$\hat{f}^{CL}$ base réclamation *as-if*	3,963	1,834	1,521	1,404	...	1,046	...	1,050	...
$\hat{f}^{CL}$ base subséquente historique	1,351	1,000	1,891	3,085	...	1,297	...	1,000	...
$\hat{f}^{CL}$ base subséquente *as-if*	1,351	1,000	1,809	3,097	...	1,282	...	1,000	...

D'un point de vue tendanciel, les estimateurs *as-if* des estimateurs « historiques » ne présentent pas de grande différence. La dérive du seuil dans le temps ne semblerait pas impacter significativement les évolutions des sinistres.

On constate l'existence de facteurs de développement *Chain-Ladder* qui sont < 1 dans les bases de survenance. Ceux-ci se justifient par le fait que les sinistres concernés précèdent la loi « About » et sont pour la plupart en fin de vie. Les dossiers clos avec un sur-provisionnement sont alors corrigés dans le négatif le semestre suivant et font alors baisser les charges cumulées, et de tels cas sont plus fréquemment rencontrés sur les sinistres rares comme ceux considérés dans ces triangles de liquidation.

On dénote aussi la variabilité des facteurs de développement de la base subséquente. Ceci s'explique simplement par le très faible volume de sinistres graves qui ont été activés par une garantie subséquente. Par exemple, seuls 7 cas de cette catégorie sont recensés sur les sinistres en charges cumulées historiques.

Enfin, sur la base réclamation où le volume de données est satisfaisant, on remarque une convergence rapide : les coefficients *Chain-Ladder* passent sous les 1,06 après 9 semestres de développement (charges *as-if* ou non). Cette convergence se corrèle avec l'alignement graphique des charges cumulées, soit implicitement la validation de l'hypothèse d'indépendance des charges par période de sinistralité.

Ces coefficients fournissent les provisions suivantes :

TABLE 3.5 – Provisions constituées selon la méthode *Chain-Ladder* par groupe de risque

Groupe de risque	1	2	3
Provisions *Chain-Ladder*	- 14 533	0	122 829 122
Provisions *Chain-Ladder as-if*	-14 802,34	0	92 796 830

Sur un nombre de dossiers importants, comme peut témoigner le groupe de risque 3 qui relève les sinistres « *post*-About » de la base réclamation plus les sinistres de la base subséquente, on observe une sensibilité sur les provisions de l'ordre de 25% .

On remarque la présence de montants de provisions négatifs sur le premier groupe de risque, *i.e.* la base « de survenance » qui contient l'ensemble des sinistres clos et non renouvelés avant la loi « About ». Cette observation rejoint le raisonnement expliqué vis-à-vis de la présence de facteurs de développement *Chain-Ladder* < 1 et prévoit, de surcroît, que d'autres provisions vont être probablement défaites dans le futur.

En effet, en regardant de plus près les dossiers en base survenance, on constate qu'il existe au 31/03/201918 sinistres graves survenus après 1996 (sur 38) qui cumulent encore 3 063 043€ de provisions sur règlements (soit ici environ 13% des charges totales en représentation). En admettant que la règle de gestion de rattachement des sinistres prise en compte fasse foi, la prédiction de provisions à la baisse semble alors justifiée.

On observe également qu'il n'y a pas de provisions sur le deuxième groupe de risque. Toujours en admettant que les rattachements des sinistres sont corrects, il est en effet observé dans les triangles de liquidation qu'en base de réclamation, les sinistres réclamés avant la loi « About » correspondent uniquement à la première ligne du triangle supérieur. En d'autres termes, les charges portées sur ces anciens sinistres ont toutes été liquidées.

Pour ces raisons-ci, nous détaillerons nos résultats par base de rattachement, car présente au vu des circonstances un meilleur intérêt en termes d'interprétation.

TABLE 3.6 – Provisions constituées selon la méthode *Chain-Ladder* par base de rattachement

Base de rattachement	Survenance	Réclamation	Subséquente
Provisions *Chain-Ladder*	- 14 533	122 263 835	565 286,50
Provisions *Chain-Ladder as-if*	-14 802,34	92 245 799	551 031,40

Méthode de *London-Chain*

Cette méthode est utilisable à la condition que les états cumulés $C_{i,j}$ et $C_{i,j+1}$ possèdent une relation affine indépendamment de la période de développement j choisie. Cette hypothèse est vérifiable graphiquement en montrant que ces quantités s'alignent sensiblement le long d'une droite affine. En reprenant les graphiques illustrés en figure 3.2 , nous observons que ces approximations sont vérifiables et acceptions alors cette hypothèse.

TABLE 3.7 – Échantillon d'estimateurs *London-Chain* obtenus

Semestre de développement	0	1	2	3	...	10	...	20	...
$\hat{\lambda}^{LC}$ base survenance historique	2,051	3,033	1,361	1,243	...	1,027	...	1,000	...
$\hat{\beta}^{LC}$ base survenance historique	6k	5k	7k	23k	...	10k	...	-2k	...
$\hat{\lambda}^{LC}$ base réclamation historique	5,204	1,163	1,631	1,153	...	1,074	...	1,081	...
$\hat{\beta}^{LC}$ base réclamation historique	-190k	461k	-99k	473k	...	-80k	...	-88k	...
$\hat{\lambda}^{LC}$ base subséquente historique	1,184	1,000	0,946	2,383	...	1,328	...	1,000	...
$\hat{\beta}^{LC}$ base subséquente historique	5k	0,00	43k	63k	...	-12k	...	0,00	...

On observe que les $\hat{\beta}^{LC}$ sont fortement variables d'un semestre de développement à un autre. Ceci illustre la capacité de ce modèle à tenir compte de tendances linéaires propres à chaque période de développement indépendamment des autres périodes sans remettre en cause la fiabilité du modèle.

À ce titre, on remarque également la convergence plus rapide des coefficients multiplicatifs $\hat{\lambda}^{LC}$ que les facteurs de développement *Chain-Ladder*. En se ressituant avec l'exemple précédent, on note sur la base de réclamation un passage des coefficients multiplicatifs sous les 1,06 au bout du cinquième semestre de développement, avec une décroissance plus prononcée sur les premières périodes qu'avec la méthode de *Chain-Ladder*.

Sur la base subséquente, on distingue plusieurs périodes de développement où les estimateurs se résument à une bissectrice (*i.e.* $\left(\hat{\beta}^{LC}, \hat{\lambda}^{LC}\right) = (1, 0)$). Ces cas-ci signifient que les charges cumulées évoluent, à ces semestres-ci, de manière parfaitement linéaire. Ceci traduit le faible volume de sinistres graves subséquents disponibles pour élaborer ce triangle de liquidation (ici 7 dossiers selon le seuil de gravité fixé et sans mise en *as-if* des charges cumulées). Il existe alors des semestres où il n'y a pas de nouveaux sinistres de cette catégorie voire de mouvements comptables liés à de telles garanties, d'où cette adéquation illusoirement parfaite des coefficients *London-Chain*.

TABLE 3.8 – Provisions constituées selon la méthode *London-Chain* par base de rattachement

Base de rattachement	Survenance	Réclamation	Subséquente
Provisions *London-Chain*	5 504 977	131 719 202	1 420 711
Provisions *Chain-Ladder as-if*	6 632 854	154 910 737	1 384 707

On observe que les nouvelles provisions estimées sont globalement plus élevées qu'avec la méthode *Chain-Ladder*, mais elles présentent des tendances par base de rattachement qui sont contradictoires.

En effet, la base survenance, dont les dossiers sont clos depuis au plus tard le 02/11/2003, prévoit ici une hausse des charges dans le futur alors que les facteurs de développement individuels indiquent une baisse des charges dans le temps, causé par la dé-constitution d'ultimes provisions sur les rares sinistres graves restants.

Les résultats *London-Chain* s'expliquent en fait par la structure du triangle de liquidation étudié, éloignée des formes habituellement rencontrées pour l'élaboration d'un tel modèle de provisionnement. La base de survenance ne considérant que les sinistres clos

avant la loi « About », le triangle de liquidation associé n'est plus alimenté depuis et présente un certain nombre de lignes de cumuls de charges nulles. La méthode *London-Chain* fournit alors des résultats biaisés car cherchent une droite de régression sur des données qui sont majoritairement composées de « 0 ». Cette méthode s'avèrerait donc inadaptée pour la base de rattachement de survenance.

Néanmoins, sur des triangles à volume de données non-anormales dans le temps comme celle associée à la base de réclamation, on observe un rapprochement voire une légère inversion des tendances entre données historiques et mises en *as-if*. Ce constat doit s'expliquer par l'impact de l'indice d'inflation exploité sur les semestres récents.

Méthode de Bornhuetter-Ferguson

Comme évoqué précédemment, n'ayant pas accès à l'information sur les primes versées du contrat RCM analysé, l'estimateur exogène ne peut être établi selon les *loss-ratio* rencontrés. À défaut, pour au moins obtenir un aperçu du comportement de cette procédure sur les bases d'étudé, nous considèrerons cet estimateur égal à la charge à l'ultime estimée par la méthode de *Chain-Ladder* :

$$\forall i \in [\![0, n]\!], \hat{\mu}^{BF}(i) = \hat{C}^{CL}_{i,m} \tag{3.41}$$

Sous ces conditions, les estimateurs obtenus sont les suivants :

TABLE 3.9 – Échantillon d'estimateurs Bornhuetter-Ferguson obtenus

Semestre de développement	0	1	2	3	...	10	...	20	...
$\hat{\beta}^{BF}$ base survenance historique	0,014	0,041	0,133	0,194	...	0,678	...	1,103	...
$\hat{\mu}^{BF}$ base survenance historique	2 158k	1 403k	4 092k	372k	...	241k	...	0	...
$\hat{\beta}^{BF}$ base réclamation historique	0,019	0,075	0,137	0,212	...	0,546	...	0,893	...
$\hat{\mu}^{BF}$ base réclamation historique	3 049k	714k	3 363k	7 700k	...	2 977k	...	4687k	...
$\hat{\beta}^{BF}$ base subséquente historique	0,081	0,110	0,110	0,208	...	0,744	...	1,000	...
$\hat{\mu}^{BF}$ base subséquente historique	0	0	7 160k	0	...	0	...	0	...

On dénote l'existence d'estimateurs exogènes nuls au niveau de la base subséquente. Cela s'explique à nouveau par le faible volume de données de cette base qui génère des semestres sans sinistralité.

TABLE 3.10 – Provisions constituées selon la méthode Bornhuetter-Ferguson par base de rattachement

Base de rattachement	Survenance	Réclamation	Subséquente
Provisions Bornhuetter-Ferguson	16 533 554	53 607 817	6 438 116
Provisions Bornhuetter-Ferguson *as-if*	20 626 583	38 105 788	6 876 057

L'estimateur exogène qui a été fixé ne semble pas adapté à la base de survenance pour les mêmes raisons qu'explicitées précédemment. De plus, on constate que le provisionnement effectué en base réclamation est bien inférieur aux précédentes méthodes appliquées. En ce sens, par cette simple comparaison des résultats obtenus, on peut imager que le paramètre exogène fixé ici sous-provisionne le portefeuille de sinistralité grave étudié.

3.2.3 Utilisation de méthodes stochastiques

Méthode de *Mack*

Le modèle de Mack fournit les mêmes facteurs de développement que la méthode de *Chain-Ladder* avec l'insertion d'une variance qui fournit une hypothèse sur la loi suivie par la distribution.

TABLE 3.11 – Échantillon d'estimateurs Mack obtenus

Semestre de développement	0	1	2	3	...	10	...	20	...
$\hat{f}^M$ base survenance *as-if*	3,310	2,989	1,426	1,430	...	1,043	...	0,998	...
$(\hat{\sigma^M})^2$ base survenance *as-if*	1 133k	16 007k	297 258k	1 853k	...	12k	...	85,77	...
$\hat{f}^M$ base réclamation *as-if*	3,963	1,834	1,521	1,404	...	1,046	...	1,050	...
$(\hat{\sigma^M})^2$ base réclamation *as-if*	3 905k	11 677k	3 337k	1 305k	...	20k	...	69k	...
$\hat{f}^M$ base subséquente *as-if*	1,351	1,000	1,809	3,097	...	1,282	...	1,000	...
$(\hat{\sigma^M})^2$ base subséquente *as-if*	13k	0	107 981k	10 601k	..	4k	...	0	...

Sur l'ensemble des semestres de développement, on observe une convergence de la variance estimée vers 0. En revenant sur l'équation 3.31, on en déduit que cette convergence de la variance s'explique en particulier par la convergence des estimateurs de *Chain-Ladder* vers 1.

TABLE 3.12 – Provisions constituées selon la méthode de Mack par base de rattachement

Base de rattachement	Survenance	Réclamation	Subséquente
Provisions Mack	-14 533	122 263 835	565 286,50
$\widehat{MSEP}$	222 480M	142 810 510 632M	83 597M
$\widehat{SEP}$	471 678	377 902 779	289 131
$\widehat{SEP}$ rapportée sur les provisions	-3 245,57%	309,09%	51,15%
Provisions Mack *as-if*	-14 802,34	92 245 799	551 031,40
$\widehat{MSEP}$	237 905M	35 679 398 164M	85 293M
$\widehat{SEP}$	487 755	188 889 910	292 050
$\widehat{SEP}$ rapportée sur les provisions	-3 295,12%	204,77%	53,00%

Les provisions Mack sont effectivement identiques aux provisions *Chain-Ladder*. Néanmoins, on repère ici que les erreurs de prédiction estimées sont relativement élevées et

laissent penser que la structure des bases de rattachement biaise à nouveau la qualité prédictive des modèles utilisés. De plus, il faut garder à l'esprit que seuls les sinistres graves sont observés, ce qui peut aussi expliquer partiellement une volatilité plus importante que d'ordinaire sur le développement des charges ultimes dans le temps.

Toutefois, dans la mesure où le volume de données est satisfaisant, les estimations en *as-if* présentent une meilleure qualité prédictive que les estimations basées sur les données d'origine.

Méthode *Bootstrap*

Pour l'application de cette méthode, les conventions suivantes seront appliquéees :
- la méthode déterministe choisie est la méthode de *Chain-Ladder* ;
- l'indépendance des résidus est admise ;
- et le nombre de simulations est fixé à 10 000.

TABLE 3.13 – Provisions constituées selon la méthode de Mack par base de rattachement

Base de rattachement	Survenance	Réclamation	Subséquente
Provisions *Bootstrap* moyennes	-15 022,66	122 258 571,93	563 862,70
Variance empirique des provisions simulées	63M	6 758M	111M
Écart-type des provisions simulées rapportées aux provisions moyennes	-52,93%	0,067%	1,872%
Provisions *Bootstrap as-if* moyennes	-15 070,43	92 238 247,15	549 835,03
Variance empirique des provisions simulées	40M	5 029M	90M
Écart-type des provisions simulées rapportées aux provisions moyennes	-41,90%	0,077%	1,722%

La méthode *Bootstrap* employée ici faisant usage des hypothèses de *Chain-Ladder*, on observe que les provisions moyennes sont proches des résultats issus de la méthode de *Chain-Ladder* appliquée seule.

Concernant la qualité prédictive des résultats, on reste dans un même ordre de grandeur qu'avec la méthode de Mack, ce qui renforce les déductions émises précédemment quant à l'impact de la structure des triangles constitués dans le cadre de la responsabilité civile médicale. Cependant, en revenant sur les écarts absolus, on peut imaginer à ce stade que la qualité prédictive du modèle *Bootstrap* est plus pertinente sur la base d'étude considérée que le modèle de Mack.

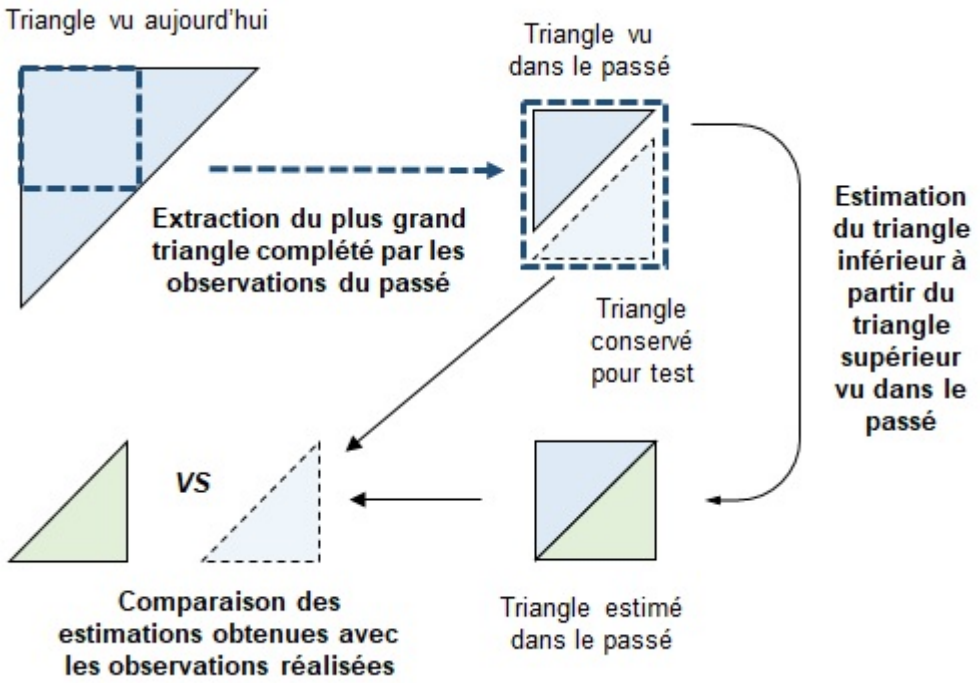
3.3 Interprétation des résultats et pistes de réflexion

3.3.1 Validation de la pertinence des provisions : adaptation du *standardized mortality ratio* à la problématique

Les modèles appliqués fournissent pour la majorité des provisions équivalentes car s'appuient plus ou moins directement sur une même méthode (à savoir la méthode de *Chain-Ladder*). Comment alors déceler parmi ces modèles celui qui fournirait le résultat le plus fiable pour les semestres à venir ?

Une idée serait de s'inspirer des méthodes de positionnement qui sont employées lors de la construction de tables d'expérience, dont le SMR en fait partie. En effet, on pourrait ré-appliquer les méthodes de provisionnement pré-établies dans un passé suffisamment loin pour que tous les estimés résultants puissent être comparables avec des observés.

FIGURE 3.3 – Synthèse de la procédure de validation des méthodes de provisionnement suggérée



Lorsque les écarts entre observés et estimés sont positifs, on peut conclure que le modèle aurait tendance à sur-provisionner, ce qui est prudent dans la limite d'un sur-provisionnement abusif. Par contre, s'il s'avère que les écarts entre observés et estimés sont négatifs, cela signifie que le modèle serait inadapté au portefeuille parce qu'il aurait tendance à sous-provisionner les sinistres à indemniser.

3.3.2 Application sur la base d'étude

TABLE 3.14 – Synthèse des écarts entre observés et estimés selon la méthode de provisionnement utilisée

Base de rattachement	Survenance	Réclamation	Subséquente
Écarts relatifs ***Chain-Ladder***	2,578%	7,698%	-26,560%
Écarts relatifs ***Chain-Ladder as-if***	1,219%	1,366%	-25,578%
Écarts relatifs ***London-Chain***	95,539%	8,082%	-23,475%
Écarts relatifs ***London-Chain as-if***	89,228%	9,944%	-21,686%
Écarts relatifs **Bornhuetter-Ferguson**	62,757%	112,509%	24,319%
Écarts relatifs **Bornhuetter-Ferguson *as-if***	63,186%	94,970%	26,375%
Écarts relatifs **Mack**	2,578%	7,698%	-26,560%
Écarts relatifs **Mack *as-if***	1,219%	1,366%	-25,578%
Écarts relatifs ***Bootstrap***	2,578%	7,696%	-26,577%
Écarts relatifs ***Bootstrap as-if***	1,219%	1,363%	-25,591%

De ces résultats, nous déduisons les points suivants :
- les écarts sont très proches entre les méthodes *Chain-Ladder*, Mack et *Bootstrap* car des facteurs de développement communs ont été utilisés ;
- les modèles sont plus en adéquation avec les triangles mis en *as-if* que les triangles originels ;
- seule la méthode de Bornhuetter-Ferguson fournit des écarts positifs sur la base subséquente, base très instable par son faible volume de données ;
- les forts écarts positifs constatés sur la base survenance soulignent l'inadéquation de certaines méthodes de provisionnement avec des triangles qui présentent des périodes de sinistralité sans indemnisations (ici parce que cette base concerne des sinistres clos avant le 02/11/2003).

Les tests stochastiques fournissent des résultats qui sont, certes, en adéquation avec des méthodes déterministes, mais ils permettent surtout d'identifier une volatilité de la sinistralité future. Leur intérêt est alors dans la quantification de la pertinence des

provisions constituées à l'instant t, les comparer avec les données du passé n'apportent pas un véritable intérêt.

Par contre, cette analyse type SMR permet, en s'appuyant sur l'expérience du portefeuille, de déceler parmi les méthodes déterministes celles qui seraient les plus adaptées au provisionnement futur. Toutefois, les résultats d'une telle analyse ne doivent pas systématiquement faire foi car le futur ne s'explique pas que par le passé : « *Cum hoc ergo propter hoc* ».

Finalement, par base de rattachement :

base survenance : on choisit la méthode de *Chain-Ladder* par défaut car les deux autres méthodes déterministes ne sont pas adaptés à cette base ;

base réclamation : *Chain-Ladder* fournit des résultats satisfaisants, mais les écarts sont trop proches du zéro pour être fiables sur du long terme ou face à une hausse inattendue future de la sinistralité. On opte donc pour la méthode de *London-Chain*, qui se montre plus prudente sans être dans les excès comme la méthode de Bornhuetter-Ferguson ;

base subséquente : on choisit la méthode de Bornhuetter-Ferguson par défaut car c'est la seule qui ne sous-estime pas la sinistralité.

Les analyses stochastiques fournissent des résultats satisfaisants, portent un intérêt statistique (comme le *Bootstrap* qui peut fournir une distribution des provisions). Cependant, pour être fiables en vérifiant notamment la loi des grands nombres et le théorème central limite, elles doivent être exécutées un très grand nombre de fois, chose coûteuse en temps.

Ainsi, les méthodes stochastiques peuvent être conservées dans un intérêt d'analyse du comportement des données et pouvoir apporter une vision prospective plus pertinente qu'en se contentant uniquement de résultats déterministes. Quant aux méthodes déterministes, dans le simple souci de calculer des provisions, elles paraissent dans ce contexte suffisantes.

3.3.3 Piste de réflexion : de l'importance de mutualiser

La mutualisation des risques est l'essence même de l'assurance : ses origines se trouvent dans l'expression de la solidarité d'un ensemble d'individus soumis à des risques identiques. [6]

L'émergence du numérique combinée à la gestion d'une très grande quantité de données de plus en plus accessible tente les assureurs vers l'individualisation des contrats. Toutefois, la segmentation dans le secteur assurantiel est une procédure à appliquer avec beaucoup de précaution.

6. [Dubuisson, 2008]

En effet, d'un point de vue purement mathématique, la segmentation en elle-même divise les volumes de donnée à disposition, ce qui implique qu'une segmentation trop poussée peut rendre les principes mathématiques fondamentaux de l'assurance inapplicables sur les modèles conçus, à savoir la loi des grands nombres et le théorème central limite.

Cette notion d'impact du volume s'est observée tout au court de ce mémoire : la volatilité du nombre de dossiers classés graves, tout comme leurs charges en décumulé, est bien plus importante dans le temps que celui des autres dossiers.

À titre d'exemple, au 31/03/2019, on relève un écart-type des charges décumulées de 294 278€ porté par 446 graves, alors que l'écart-type au global sur cette même base est de 28 124€ sur 74 986 sinistres. Cette importante variabilité a aussi été démontrée dans les calculs d'erreur d'estimation précédemment effectués sur les modèles de provisionnement stochastiques, et les faibles volumes ont leur part de responsabilité.

Sur l'aspect économique, une segmentation trop fine est bien souvent synonyme de modèle complexes, ce qui augmente significativement le risque opérationnel qui s'impute sur le coût des sinistres. Même si, d'un point de vue pratique, la segmentation a tendance a récompenser les « bons » risques en leur diminuant leurs primes, les classes de risques élevées se retrouvent condamnées à payer plus faute de mutualisation.

En effet, selon un rapport publié par l'Inspection Générale des Affaires Sociales en 2007[7], la perspective des primes était à la hausse sur l'ensemble des contrats d'assurance responsabilité civile médicale et cite le cas des médecins gynécologues-obstétriciens : en moyenne, cette spécialité versait 15 000€ de primes en 2007 contre 14 000€ en 2006. De plus, il a été remarqué que MACSF, acteur majeur de la RCM, souhaitait se retirer du marché des gynécologues-obstétriciens si leurs primes n'étaient pas augmentées de 15 000€.

Cela illustre à juste titre les dangers de la segmentation sur un aspect éthique : en perdant la notion de solidarité des contrats d'assurance, les classes fortement exposées au risque se retrouvent face à des primes exorbitantes, voire finissent rejetées par les assureurs. Mais les populations à risque ne le sont pas toujours par choix...

C'est en partie pour ces raisons-ci qu'il avait déjà été proposé dans le passé de faire évoluer la gestion des sinistres graves dans le sens de la mutualisation. M. Johanet[8] proposait en 2011 une liste de mesures dont la « [mutualisation de] *l'ensemble des professionnels de santé libéraux et conventionnels pour les sinistres supérieurs à 3M* ».

7. [Inspection générale des affaires sociales, 2007]
8. [Johanet, 2011]

Ainsi, il aurait été intéressant de poursuivre les analyses de la sinistralité grave en responsabilité civile médicale sur plusieurs portefeuilles issus de différentes structures et disciplines du milieu de la santé (pour rappel, le portefeuille étudié dans le cadre de ce mémoire regroupe un corps de médecins libéraux). *A contrario*, il aurait été intéressant d'obtenir un plus large panel d'informations issu du portefeuille étudié pour analyser les impacts sur le provisionnement d'une segmentation sur les sinistres graves interprétable mais trop prononcée.

Conclusion

En conclusion, cette analyse de la sinistralité grave sur le marché restreint qu'est la responsabilité civile médicale nous aura fait découvrir l'ensemble des démarches que doit poursuivre un actuaire pour mener à bien ses calculs de provisions. Il doit :

- faire preuve d'adaptabilité dans un environnement juridique et financier en constante évolution :
- avoir une connaissance profonde des portefeuilles d'assurance gérés, ce qui implique une communication avec l'ensemble des services de l'entreprise qui peuvent apporter une aide à la compréhension des données (service de gestion, *etc.*) ;
- et avoir un bagage scientifique et informatique suffisant pour être en capacité de concevoir et d'appliquer des modèles qui sont adaptés aux données étudiées.

Ces conditions fondamentales mettent en somme en avant la capacité requise d'un actuaire à garder un esprit critique sur toute les démarches pour fournir un travail mathématiquement fiable et adapté aux contraintes du réel.

Dans le cadre de ce mémoire, il a été question d'analyser la sinistralité grave en s'adaptant principalement aux contraintes de temps (les sinistres RCM graves étant généralement longs). Il a été question :

- d'adapter la problématique selon des critères juridiques comme la loi « About [9] », qui fait basculer une vision des sinistres en date de survenance à une vision des sinistres en date de réclamation ;
- d'adapter la problématique avec les données à disposition, sachant qu'elles ne représentaient que des informations élémentaires sur la sinistralité (dates et montants) par souci du respect du Règlement Général sur la Protection des Données (RGPD) ;
- et d'appliquer un savoir-faire mathématique pour établir de manière fiable et justifiée un modèle d'identification et de provisionnement adapté à une problématique spécifique.

Le sinistre grave, autrement dit exceptionnel, étant rare mais coûteux, ce mémoire aura souligné l'importance d'investir des approches stochastiques combinées avec une analyse en profondeur du portefeuille pour prédire au mieux la forte volatilité qu'engendrent de tels sinistres en responsabilité civile médicale.

Néanmoins, de grands enjeux financiers n'impliquent pas de grands modèles complexes. Un modèle simple mais justifié et en concordance avec les données étudiées peut suffire pour fournir des résultats très satisfaisants. La qualité des données est une condition *sine qua non* à de bons résultats issus d'un modèle, quelle que soit sa complexité.

On a en effet observé que la méthode de *Chain-Ladder* peut être relativement en adéquation avec les tendances d'un portefeuille de sinistres graves, sous réserve que ce

9. [Journal Officiel, 2002a]

dernier ait été correctement retraité et adapté à la problématique. La mise en *as-if* des indemnisations ainsi que le choix d'un seuil de gravité statistiquement stable sont des travaux sur les données qui auront fortement contribué à une bonne qualité prédictive des provisions calculées.

Annexe A

Présentation de l'ONIAM

Figure A.1 – Logo de l'ONIAM

L'Office National d'Indemnisation des Accidents Médicaux (ONIAM) est un établissement public créé par la loi « Kouchner [1] ». Son premier objectif est d'indemniser, de la façon la plus équitable possible, les victimes d'accident médical.

L'ONIAM intervient dans trois principaux cas de figure :
- lorsque l'assureur n'agit pas ou refuse explicitement de faire une offre d'indemnisation ;
- lorsque le professionnel de santé responsable n'est pas assuré ;
- ou lorsque la couverture d'assurance en question est épuisée, *i.e.* les indemnisations à verser sont supérieures au plafond de garantie ou les garanties subséquentes ont expiré.

Son dispositif d'indemnisation est accessible par tous et ne nécessite pas l'intervention d'un avocat. Il est établi en collaboration avec des Commissions de Conciliation et d'Indemnisation (CCI), institutions indépendantes et présidées par un magistrat. Les CCI favorisent la résolution des conflits entre usagers et professionnels de santé, notamment par la nomination d'un médiateur. Une fois l'avis émis par une CCI, lorsqu'il est favorable, la victime est indemnisée :
- soit par une ou plusieurs assurance d'un ou plusieurs professionnels de santé lorsque leur responsabilité civile est engagée ;

1. [Journal Officiel, 2002b]

- soit par l'ONIAM au titre de la solidarité nationale, décision qui est généralement prise quand il n'y a pas eu faute et que l'accident médical est anormal ;
- soit à la fois par le ou les assureurs et par l'ONIAM en cas de partage de l'indemnisation.

Côté ONIAM, ils apportent aux CCI un soutien technique et administratif, et les frais d'expertise nécessaires à l'instruction des demandes par les CCI sont à leur charge. Tous les dossiers reçus sont traité spécifiquement, au cas par cas, et dans un souci d'indemniser les dommages subis de manière rapide et gratuite, et ce sur tout le territoire.

Le champ d'activité d'ONIAM s'est progressivement diversifiée avec le temps ; aujourd'hui, cette structure traite :

- les accidents médicaux ;
- les activités de recherche biomédicale ;
- les affections iatrogènes, effets secondaires qui sont liés à un traitement médical ;
- et les infections nosocomiales, infections qui sont contractées dans un établissement de santé.

Annexe B

Présentation des bases de données initiales

TABLE B.1 – Synthèse des variables retenues de la base initiale « sinistres historiques, partie 1 »

Nom de la variable	**Pourcentage renseigné**	**Valeurs**	**Commentaires**
Numéro de sinistre	100%	Numéros d'identification	N'est pas la clé primaire de la base de données. Il s'agit du couple Numéro de sinistre × Date d'arrêté
Date d'arrêté	100%	Du 30/06/1986 au 30/06/2018	Sélectionner une date d'arrêté permet de visualiser tous les sinistres qui sont ouverts aux indemnisations à cette même date
Date de survenance	100%	Du 01/03/1956 au 21/06/2018	Date utilisée pour sélectionner les données en « base sinistre »
Date de réclamation du sinistre	52,42%	Du 04/02/1974 au 06/07/2018	Date utilisée pour sélectionner les données en « base réclamation » Les dates manquantes sont renseignées au 31/12/9999
Montant des règlements	100%	De 0 à 3 382 725,66	Montants nuls : 34,92%
Montant des recours	100%	De 0 à 690 121,58	Montants nuls : 98,50%
Montant des provisions pour règlements	100%	De 0 à 5 200 000	Montants nuls : 81,90%
Montant des provisions pour recours	100%	De -207,50 à 1 467 037	Montants nuls : 99,81% Montants négatifs : 0,001%

TABLE B.2 – Synthèse des variables retenues de la base initiale « sinistres historiques, partie 2 »

Nom de la variable	**Pourcentage renseigné**	**Valeurs**	**Commentaires**
Numéro de sinistre	100%	Numéros d'identification	N'est pas la clé primaire de la base de données. Il s'agit du couple Numéro de sinistre × Date d'arrêté
Date d'arrêté	100%	Du 30/06/1986 au 30/06/2018	Sélectionner une date d'arrêté permet de visualiser tous les sinistres qui sont ouverts aux indemnisations à cette même date
Date de survenance	100%	Du 01/01/1980 au 31/12/2017	Date utilisée pour sélectionner les données en « base sinistre »
Date de réclamation du sinistre	78,52%	Du 01/04/1981 au 22/01/2019	Date utilisée pour sélectionner les données en « base réclamation » Les dates manquantes sont renseignées au 31/12/9999
Montant des règlements	100%	De -102,94 à 4 039 334,89	Montants nuls : 32,42% Montants négatifs : 0,010%
Montant des recours	100%	De -198,36 à 516 843,91	Montants nuls : 98,85% Montants négatifs : 0,005%
Montant des provisions pour règlements	100%	De 0 à 4 600 000	Montants nuls : 73,77%
Montant des provisions pour recours	100%	De -25 231,61 à 3 815 000	Montants nuls : 99,91% Montants négatifs : 0,002%

Remarque : les montants négatifs correspondent soit à des erreurs de saisie (erreurs d'arrondi créées par le système qui génère les bases) soit à des corrections manuelles effectuées sur l'exercice suivant.

TABLE B.3 – Synthèse des variables retenues de la base initiale « base règlements », arrêtée au 31/03/2019

Nom de la variable	**Pourcentage renseigné**	**Valeurs**	**Commentaires**
Numéro de sinistre	100%	Numéros d'identification	N'est pas la clé primaire de la base de données. Il s'agit du couple Numéro de sinistre × Date d'opération
Type de sinistre	100%	0 : 14,03% 1 : 85,97%	Indique si le service de gestion a mis ce sinistre en « base réclamation » ou en « base sinistre »
Semestre de flux	100%	De 59 à 139	Sert à calculer les semestres de développement
Semestre de survenance	100%	De 13 à 139	Indice ligne du triangle de développement « base sinistre »
Semestre de réclamation	100%	De 16 à 139	Indice ligne du triangle de développement « base réclamation »
Semestre de clôture	77,24%	De 73 à 139	Les sinistres ouverts ont un semestre de clôture renseigné à 16 100
Semestre de développement du sinistre	100%	De 0 à 91	Indice colonne du triangle de développement « base sinistre »
Semestre de développement de la réclamation	100%	De 0 à 64	Indice colonne du triangle de développement « base réclamation »
Montant du règlement	100%	De -58 800,52 à 2 791 134,91	Montants nuls : 0,43% Montants négatifs : 0,439%g

TABLE B.4 – Synthèse des variables retenues de la base initiale « base provisions », arrêtée au 31/03/2019

Nom de la variable	**Pourcentage renseigné**	**Valeurs**	**Commentaires**
Numéro de sinistre	100%	Numéros d'identification	N'est pas la clé primaire de la base de données. Il s'agit du couple Numéro de sinistre × Date d'opération
Type de sinistre	100%	0 : 15,59% 1 : 84,41%	Indique si le service de gestion a mis ce sinistre en « base réclamation » ou en « base sinistre »
Semestre de flux	100%	De 16 à 139	Sert à calculer les semestres de développement
Semestre de survenance	100%	De 13 à 139	Indice ligne du triangle de développement « base sinistre »
Semestre de réclamation	100%	De 16 à 140	Indice ligne du triangle de développement « base réclamation »
Semestre de clôture	83,17%	De 73 à 139	Les sinistres ouverts ont un semestre de clôture renseigné à 16 100
Semestre de développement du sinistre	100%	De 0 à 98	Indice colonne du triangle de développement « base sinistre »
Semestre de développement de la réclamation	100%	De 0 à 77	Indice colonne du triangle de développement « base réclamation »
Montant de la provision pour règlement	100%	De -2 935 854 à 4 031 646	Montants nuls : 12,80% Montants négatifs : 54,55%g

TABLE B.5 – Synthèse des variables retenues de la base initiale « base recours », arrêtée au 31/03/2019

Nom de la variable	**Pourcentage renseigné**	**Valeurs**	**Commentaires**
Numéro de sinistre	100%	Numéros d'identification	N'est pas la clé primaire de la base de données. Il s'agit du couple Numéro de sinistre × Date d'opération
Type de sinistre	100%	0 : 25,78% 1 : 74,22%	Indique si le service de gestion a mis ce sinistre en « base réclamation » ou en « base sinistre »
Semestre de flux	100%	De 73 à 139	Sert à calculer les semestres de développement
Semestre de survenance	100%	De 29 à 136	Indice ligne du triangle de développement « base sinistre »
Semestre de réclamation	100%	De 49 à 136	Indice ligne du triangle de développement « base réclamation »
Semestre de clôture	75,65%	De 73 à 139	Les sinistres ouverts ont un semestre de clôture renseigné à 16 100
Semestre de développement du sinistre	100%	De 0 à 91	Indice colonne du triangle de développement « base sinistre »
Semestre de développement de la réclamation	100%	De 0 à 64	Indice colonne du triangle de développement « base réclamation »
Montant du recours	100%	De -40 693,13 à 638 056,18	Montants nuls : 0,10% Montants négatifs : 1,637%g

Remarque : le semestre de référence, celui qui correspond au semestre « 1 », débute au 01/07/1949.

Annexe C

Présentation des bases de données finales

Remarques préalables :

- la base finale, utilisée pour les calculs de provisionnement et qui comporte l'ensemble des sinistres depuis l'origine est composée de 215 377 observations, 25 variables et de 74 986 dossiers ;
- le semestre de référence, celui qui correspond au semestre « 1 », débute au 01/07/1949 ;
- les semestres de réclamation manquants correspondent à des sinistres qui ont été récemment déclarés mais pas encore réclamés. La règle de gestion est d'alors de constituer une provision forfaitaire en fonction de la gravité du sinistre signalé et supposer que la réclamation s'effectuera le semestre suivant.

TABLE C.1 – Synthèse des principales variables de la base finale obtenue, arrêtée au 31/03/2019

Nom de la variable	**Pourcentage renseigné**	**Valeurs**	**Commentaires**
Numéro de sinistre	100%	Numéros d'identification	N'est pas la clé primaire de la base de données. Il s'agit du couple Numéro de sinistre × Semestre de flux
Type de sinistre	100%	0 : 15,75% 1 : 84,25%	Indice ligne du triangle de développement « base sinistre »
Semestre de survenance	100%	De 13 à 139	Indice ligne du triangle de développement « base sinistre »
Semestre de réclamation	99,997%	De 16 à 139	Indice ligne du triangle de développement « base réclamation »
Semestre de développement de survenance	100%	De 0 à 98	Indice colonne du triangle de développement « base sinistre »
Semestre de développement de réclamation	99,997%	De 0 à 77	Indice colonne du triangle de développement « base réclamation »
Montant cumulé des règlements	100%	De -102,94 à 4 107 755,41	Les montants cumulés négatifs sont rectifiés sur le dernier semestre de flux à disposition
Montant cumulé des provisions	100%	De -0,56 à 5 200 000	
Montant cumulé des recours	100%	De -198,36 à 690 121,58	
Montant cumulé des charges	100%	De -101,42 à 7 614 308,78	Le cumul de charges négatif est corrigé sur le dernier semestre de flux à disposition
Montant cumulé des coûts	100%	De -137 252,12 à 7 614 308,78	Le cumul de coûts fortement négatif est causé par un dossier qui détient un recours plus élevé que la charge totale

Annexe D

Méthodes d'estimation semi-paramétriques de queue de distribution

L'analyse d'une queue de distribution se concentre sur le paramètre de forme ξ, issue de la loi du DAM dont la fonction de répartition étudiée F en fait partie. Les estimateurs de quantiles extrêmes font appel à ce paramètre, ce qui montre l'importance de l'estimer à une valeur la plus fiable possible.

Pour la suite, on considèrera $(X_i)_{\substack{i\in[\![1,n]\!]\\ n\in\mathbb{N}^*}}$ un n-échantillon d'une loi qui appartient au DAM de paramètre de forme $\xi \in \mathbb{R}$. Les notations employées ci-contre font référence aux notations posées dans la section 2.2.2.

D.1 Estimateur de Pickands

Définition 11 (Estimateur de Pickands) *Soit $k \in \mathbb{N}^*$ tel que $\frac{k}{n} \to 0$. L'estimateur de Pickands se définit comme suit :*

$$\hat{\xi}^P : k \mapsto \frac{1}{\ln(2)} \cdot \ln\left(\frac{X_{k,n} - X_{2\cdot k,n}}{X_{2\cdot k,n} - X_{4\cdot k,n}}\right) \tag{D.1}$$

L'estimateur de Pickands est un estimateur convergent de ξ. En 1989, Arnold L.M. Dekkers et Laurens de Haan démontrent que sous certaines conditions imposées à k et F, cet estimateur est asymptotiquement gaussien :

$$\sqrt{k} \cdot (\hat{\xi}^P(k) - \xi) \underset{\frac{k}{n}\to 0}{\longrightarrow} \mathcal{N}\left(0, \frac{\xi^2 \cdot (2^{1+\xi}+1)}{(2\cdot(2^\xi - 1)\cdot\ln(2))^2}\right) \tag{D.2}$$

D.2 Estimateur de Hill

Définition 12 (Estimateur de Hill) *Soit $k \in \mathbb{N}^*$ tel que $\frac{k}{n} \to 0$. L'estimateur de Hill se définit comme suit :*

$$\hat{\xi}^H : k \mapsto \frac{1}{k} \cdot \sum_{j=1}^{k} \ln\left(\frac{X_{j,n}}{X_{k,n}}\right) \tag{D.3}$$

L'estimateur de Hill est également un estimateur convergent de ξ. Il est considéré comme plus performant que l'estimateur de Pickands, au vu de ses variances asymptotiques, mais il n'est utilisable que pour les distributions de Fréchet (*i.e.* $\xi > 0$). En 1998, de Haan et Peng ont démontré que cet estimateur, sous certaines conditions imposées à k et F, cet estimateur est lui aussi asymptotiquement gaussien :

$$\sqrt{k} \cdot (\hat{\xi}^H(k) - \xi) \underset{\frac{k}{n}\to 0}{\longrightarrow} \mathcal{N}\left(0, \xi^2\right) \tag{D.4}$$

D.3 Estimateur de Dekkers-Einmahl-de Haan

Définition 13 (Estimateur de Dekkers-Einmahl-de Haan) *Soit $k \in \mathbb{N}^*$ tel que $\frac{k}{n} \to 0$. L'estimateur de Dekkers-Einmahl-de Haan se définit comme suit :*

$$\begin{aligned} \hat{\xi}^{DEdH} &: k \mapsto 1 + \hat{\xi}^{(1)}(k) - \frac{1}{2} \cdot \left(1 - \frac{(\hat{\xi}^{(1)}(k))^2}{\hat{\xi}^{(2)}(k)}\right)^{-1} \\ \forall i \in \{1,2\}, \hat{\xi}^{(i)} &: k \mapsto \frac{1}{k} \cdot \sum_{j=1}^{k} \left(\ln\left(\frac{X_{j,n}}{X_{k+1,n}}\right)\right)^i \end{aligned} \tag{D.5}$$

Cet estimateur, appelé aussi « estimateur des moments », est une généralisation de l'estimateur de Hill à l'ensemble des lois d'extremum (*i.e.* $\xi \in \mathbb{R}$). Les $\hat{\xi}^{(i)}$ sont des estimateurs convergents de ξ. En 1989, Dekkers, Einmahl et de Haan ont démontré que cet estimateur est asymptotiquement gaussien :

$$\sqrt{k} \cdot (\hat{\xi}^{DEdH}(k) - \xi) \underset{\frac{k}{n}\to 0}{\longrightarrow} \mathcal{N}\left(0, 1 + \xi^2\right) \tag{D.6}$$

Cet estimateur est également appelé « estimateur des moments », avec les $\xi^{(i)}$ qui s'interprètent comme des moments empiriques.

Annexe E

Méthodes d'analyse graphiques sur les valeurs extrêmes

Dans le cadre d'une recherche d'adéquation de loi d'extremum, les analyses graphiques offrent des identifications rapides et facilement interprétables de divers tests statistiques. Les notations employées ci-après font référence aux notations posées dans la section 2.2.2.

E.1 Diagramme quantile-quantile

Le *QQ-plot* est un diagramme qui permet de tester l'adéquation entre deux échantillons en fonction des quantiles qu'ils renvoient. Il sert en particulier à analyser l'adéquation entre les quantiles d'un échantillon d'observations et les quantiles d'une distribution théorique choisie.

En considérant un échantillon d'observations de taille $n \in \mathbb{N}^*$ $(x_i)_{i\in[\![1,n]\!]}$ et une loi d'inverse de fonction de répartition F^{-1}, la représentation quantile-quantile trace les points suivants :

$$\begin{aligned} [\![1,n]\!] &\to \mathbb{R}^2 \\ i &\mapsto \left(x_{i,n}, F^{-1}\left(\frac{n-i}{n}\right)\right) \end{aligned}$$

La droite séparatrice représentée correspond à la « droite de Henry », établie par Pierre Jean Paul Henri en 1894. Pour un échantillon de variables aléatoires X, cette droite est déterminée selon l'équation suivante :

$$\begin{aligned} \mathbb{R} &\to \mathbb{R} \\ x &\mapsto \frac{x-\overline{X}}{\sigma_X} \end{aligned} \tag{E.1}$$

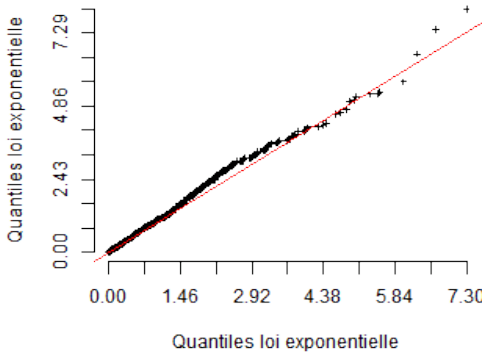
avec $\overline{X}$ et σ_X qui correspondent respectivement à la moyenne et l'écart-type de X. Si les données comparées sont normalisées, alors la droite de Henry du *QQ-plot* correspond à la bissectrice qui passe par l'origine.

En termes d'interprétation, le lien avec les valeurs extrêmes s'identifierait avec un *QQ-plot* exponentiel d'intensité $\lambda > 0$:

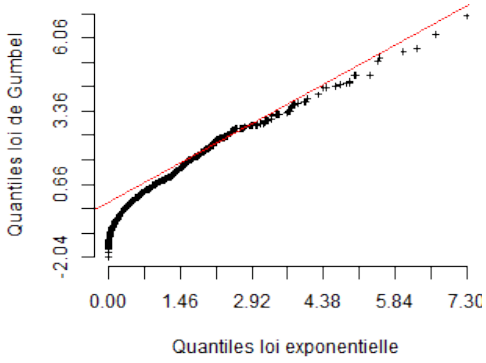
- si les points sont alignés le long de la droite de Henry (ici, elle correspondrait à une droite de pente $\frac{1}{\lambda}$), alors les données analysées suivent une loi exponentielle ;
- si l'allure générale est convexe, alors la queue de distribution des données analysées est plus lourde que la loi exponentielle ;
- et si l'allure générale est concave, alors la queue de distribution des données analysées est plus légère que la loi exponentielle.

FIGURE E.1 – Exemples de *QQ-plot* sur des lois usuelles

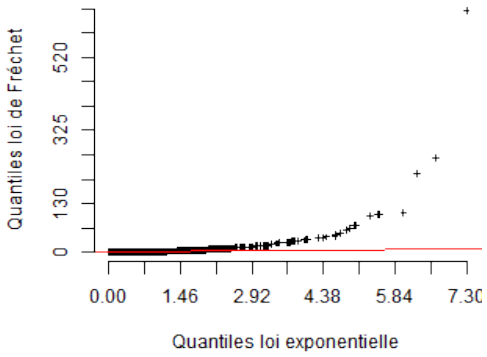
(a) Loi exponentielle (référence)



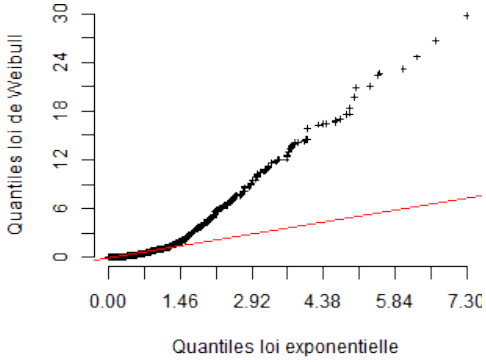
(b) Loi de Gumbel (type I)



(c) Loi de Fréchet (type II)



(d) Loi de Weibull (type III)



E.2 Diagramme moyenne des excès

Pour un échantillon de $n \in \mathbb{N}^*$ variables aléatoires d'espérance finie, on définit la moyenne des excès par la fonction suivante :

$$\begin{aligned} e : \mathbb{R} &\to \mathbb{R}_+ \\ s &\mapsto \frac{\mathbb{E}[X - s | X > s]}{s} \end{aligned} \tag{E.2}$$

Pour analyser la moyenne des excès par rapport à une loi donnée, on prélève un échantillon de n variables aléatoires qui suivent cette loi et on utilise l'estimateur suivant :

$$\begin{aligned} \hat{e} : \mathbb{R} &\to \mathbb{R}_+ \\ s &\mapsto \frac{1}{N_s} \cdot \sum_{i=1}^{n} [X_i - s]_+ \end{aligned} \tag{E.3}$$

Sur un ensemble de n données, la représentation moyenne des excès est tracée par les points suivants :

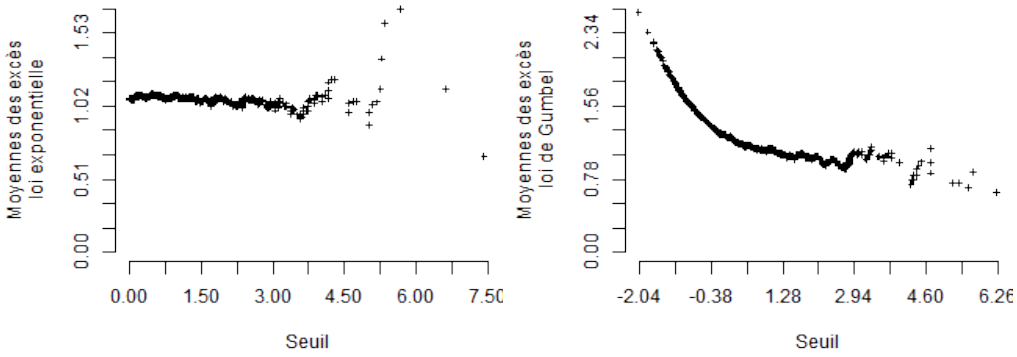
$$\begin{aligned} \mathbb{R} &\to \mathbb{R} \\ s &\mapsto (s, e(s)) \end{aligned}$$

En termes d'interprétation, si on observe un seuil s_0 où l'ensemble des points $s > s_0$ représentent une fonction affine du seuil, alors cela veut dire que s_0 est un seuil optimal dans le sens où la loi choisie est stable au-delà de ce seuil.

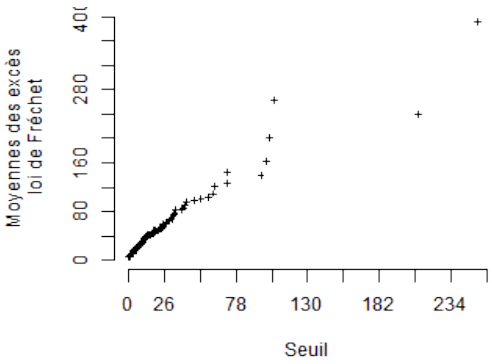
La queue de distribution peut être également interprétable à l'aide de ce diagramme :
- si la représentation est une fonction croissante, alors la queue de distribution des données analysées est lourde ;
- si la représentation est une fonction décroissante, alors la queue de distribution des données analysées est légère ;
- et si la représentation est constante, alors la queue de distribution des données est similaire à une queue de distribution exponentielle.

FIGURE E.2 – Exemples de *ME-plot* sur des lois usuelles

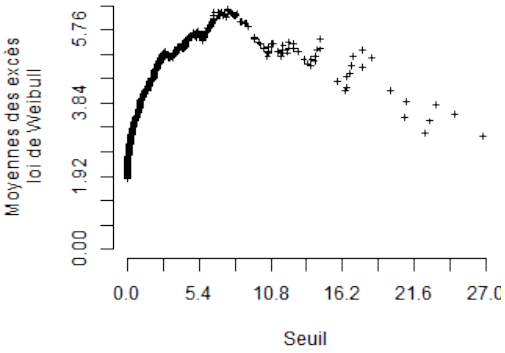
(a) Loi exponentielle (référence) (b) Loi de Gumbel (type I)



(c) Loi de Fréchet (type II)



(d) Loi de Weibull (type III)



E.3 Diagramme des estimateurs

Les diagrammes d'analyse des estimateurs reprennent les méthodes semi-paramétriques détaillées en annexe D et détaillent l'évolution des paramètres de forme estimés $\hat{\xi} \in \mathbb{R}$ en fonction du nombre de dépassements de seuil observés $k \in \mathbb{N}^*$ sur un échantillon de taille $n \in \mathbb{N}^*$, sachant que $\frac{k}{n} \to 0$.

Ainsi, le *Pickands-plot* trace les points suivants :

$$\begin{aligned} \mathbb{N}^* &\to \mathbb{R} \\ k &\mapsto (k, \hat{\xi}^P(k)) \end{aligned}$$

le *Hill-plot* trace les points suivants :

$$\begin{aligned} \mathbb{N}^* &\to \mathbb{R} \\ k &\mapsto (k, \hat{\xi}^H(k)) \end{aligned}$$

et le *Dekkers-Einmahl-de Haan-plot* trace les points suivants :

$$\begin{aligned} \mathbb{N}^* &\to \mathbb{R} \\ k &\mapsto (k, \hat{\xi}^{DEdH}(k)) \end{aligned}$$

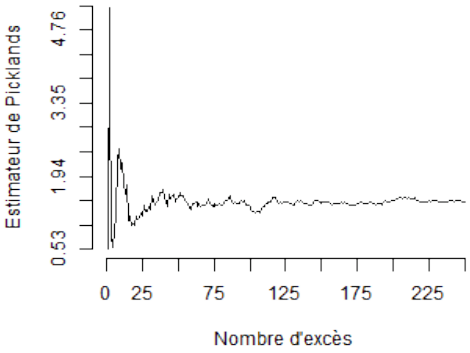
En termes d'interprétation, on a une logique analogue au *ME-plot* : si on observe un seuil k_0 à partir duquel pour tout $k > k_0$ la représentation converge vers une fonction constante, alors ce seuil k_0 fournit une estimation du paramètre de forme $\hat{\xi}$ stable.

Remarques :

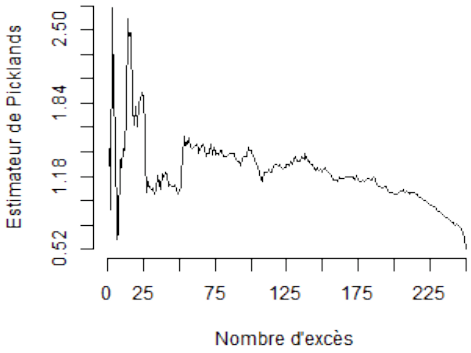
- Bien que les *Pickands-plot* et les *Hill-plot* sont généralement similaires, l'estimateur de Pickands est plus volatile que l'estimateur de Hill.
- Il faut garder à l'esprit les conditions d'applicabilité des estimateurs semi-paramétriques. Par exemple, faire un *Hill-plot* sur une distribution à queue légère n'a pas de sens.

FIGURE E.3 – Exemples de *Pickands-plot* sur des lois usuelles

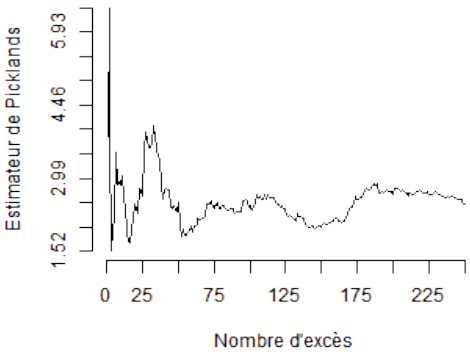
(a) Loi exponentielle (référence)



(b) Loi de Gumbel (type I)



(c) Loi de Fréchet (type II)



(d) Loi de Weibull (type III)

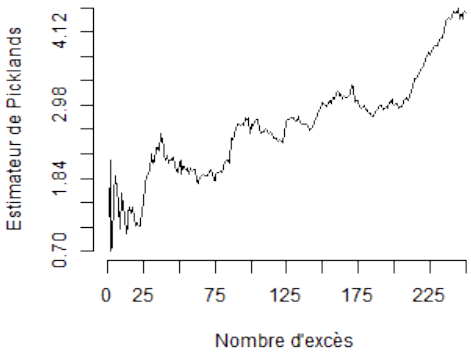
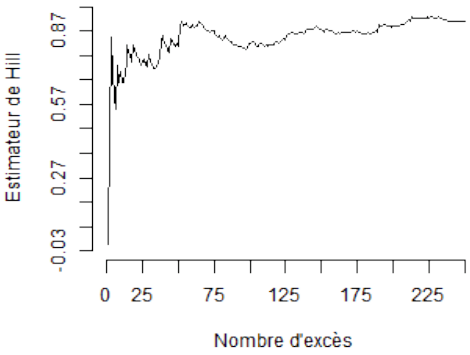


FIGURE E.4 – Exemples de *Hill-plot* sur des lois usuelles

(a) Loi de Fréchet (type II)



(b) Loi de Weibull (type III)

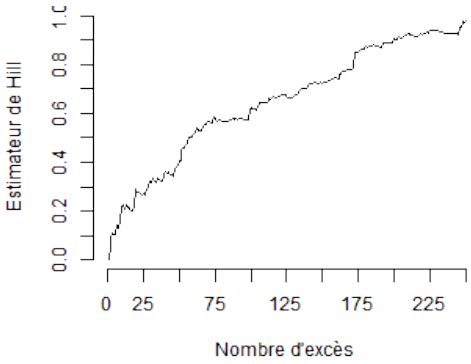
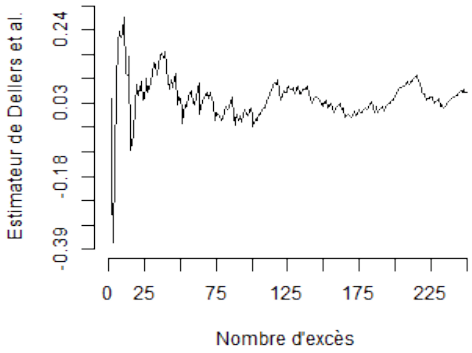
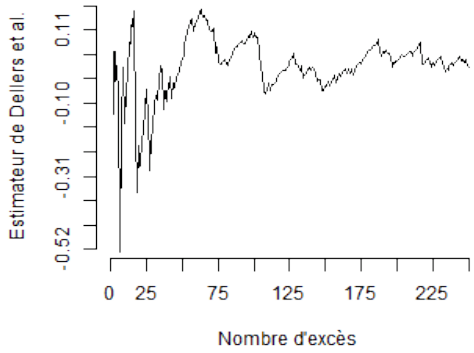


FIGURE E.5 – Exemples de *Dekkers-Einmahl-de Haan-plot* sur des lois usuelles

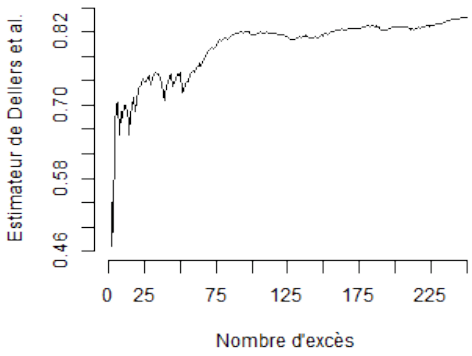
(a) Loi exponentielle (référence)



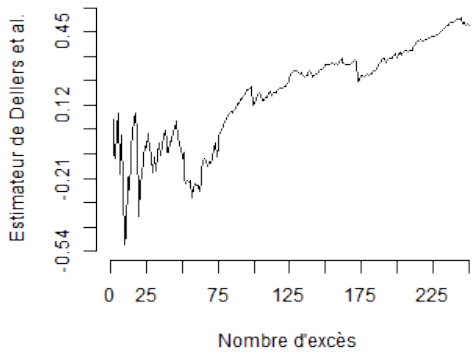
(b) Loi de Gumbel (type I)



(c) Loi de Fréchet (type II)



(d) Loi de Weibull (type III)



E.4 Diagramme de Gerstengarbe et Werner

En 1989, F.-W. Gerstengarbe et P. C. Werner ont établi une procédure qui identifie graphiquement un seuil d'identification des valeurs extrêmes sans utiliser la théorie des valeurs extrêmes. Cette méthode s'appuie sur le test séquentiel de tendance de Mann-Kendall en analysant le comportement des écarts entre différentes observations d'un échantillon donné.

Pour une série de $n \in \mathbb{N}^*$ variables aléatoires $(X_i)_{i\in[\![1,n]\!]}$, on pose la suite des écarts selon la relation suivante :

$$\forall i \in [\![2, n]\!] : \Delta_i = X_{i-1,n} - X_{i,n} \tag{E.4}$$

Avec cette suite d'écarts, on utilise la série des dépassements des écarts suivante :

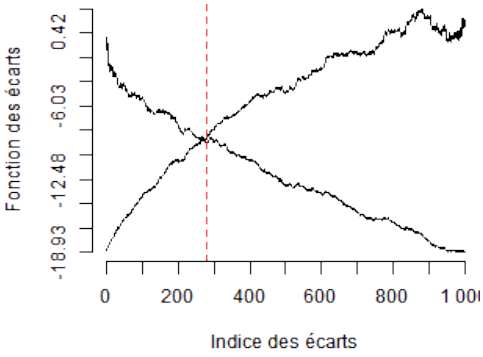
$$\forall i \in [\![2, n]\!] : U_i^* = \sum_{j=2}^{i} \left(\sum_{k=2}^{j} \mathbf{1}_{\Delta_k < \Delta_j} \right) \tag{E.5}$$

Le *GW-plot* trace alors la courbe des écarts, dans les sens des données croissant et décroissant, entre cette quantité et son espérance selon la relation suivante :

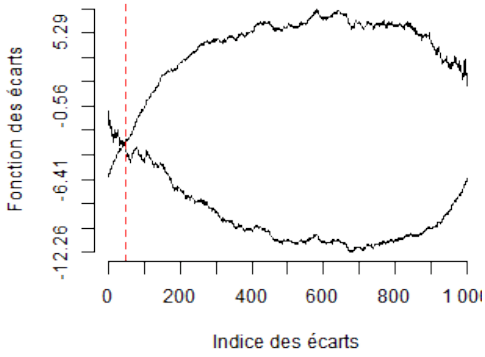
$$\begin{aligned} &[\![2, n]\!] \to \mathbb{R} \\ &U_i : i \mapsto \frac{U_i^* - \mathbb{E}[U_i^*]}{\sqrt{\mathrm{Var}[U_i^*]}} = \frac{U_i^* - \frac{1 \cdot (i-1)}{4}}{\sqrt{\frac{i \cdot (i-1) \cdot (2 \cdot i + 5)}{72}}} \end{aligned} \tag{E.6}$$

FIGURE E.6 – Exemples de *Gerstengarbe-Werner-plot* sur des lois usuelles

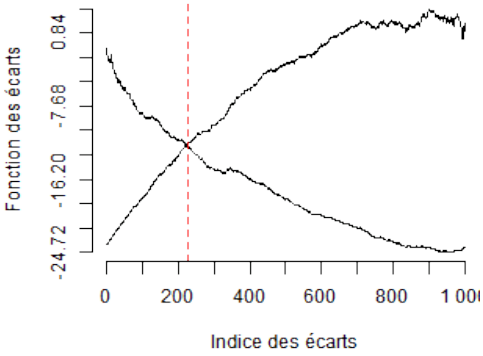
(a) Loi exponentielle (référence)



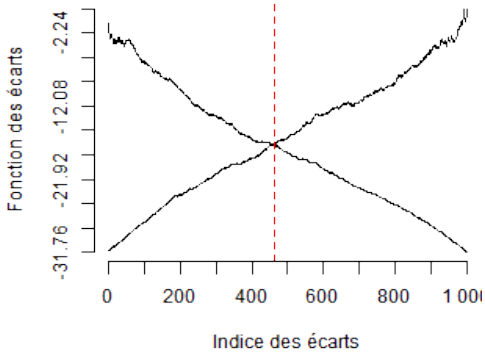
(b) Loi de Gumbel (type I)



(c) Loi de Fréchet (type II)



(d) Loi de Weibull (type III)



Annexe F

Tests d'adéquation statistique de lois

F.1 Test du Khi-2

Le test du Khi-2 est un test statistique qui offre une significativité sur l'adéquation entre un échantillon de données et une lois de probabilités discrète ou l'indépendance entre deux variables. Dans le cadre de ce mémoire, il sera question d'utiliser le test d'adéquation.

Définition 14 (Loi du Khi-2) *Soit $(X_i)_{i\in mathbbN^*}$ un ensemble de variables aléatoires indépendantes qui suivent une loi normale centrée réduite $\mathcal{N}(0,1)$. La loi du Khi-2 à n degrés de liberté correspond à la variable aléatoire suivante :*

$$\forall n \in \mathbb{N}^*, \chi^2_n = \sum_{i=1}^{n} X_i^2 \tag{F.1}$$

Pour deux échantillons de variables aléatoires $X = (X_i)_{\substack{i\in[\![1,n]\!]\\ n\in\mathbb{N}^*}}$ et $Y = (Y_i)_{\substack{i\in[\![1,n]\!]\\ n\in\mathbb{N}^*}}$, on teste l'hypothèse nulle H_0 « X suit la loi Y » contre l'hypothèse alternative « X suit une autre loi que Y ».

Soient :
- $\hat{p}_{X_i})_{i\in\mathbb{N}^*}$ les probabilités empiriques associées à l'échantillon observé X ;
- $(p_{Y_i})_{i\in\mathbb{N}^*}$ les probabilités associées à la loi de Y.

On applique la loi statistique suivante :

$$\forall k \in \mathbb{N}^*, D_k = \sum_{i=1}^{k} \frac{(p_{X_i} - p_{Y_i})^2}{p_{X_i}} \tag{F.2}$$

Sous H_0, cette loi suit asymptotiquement une loi du Khi-2 à $k-1$ degrés de liberté. Pour un niveau de test α, on prend alors comme décision :
- **rejet de** H_0 si $D_k > \chi^2_{k-1,1-\alpha}$;
- **le non-rejet de** H_0 si $D_k \leq \chi^2_{k-1,1-\alpha}$.

F.2 Test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov est un test statistique qui offre une significativité sur l'adéquation entre un échantillon de données et une lois de probabilités continue.

De même que pour le test du Khi-2, on considère deux échantillons de variables aléatoires $X = (X_i)_{\substack{i\in[\![1,n]\!] \\ n\in\mathbb{N}^*}}$ et $Y = (Y_i)_{\substack{i\in[\![1,n]\!] \\ n\in\mathbb{N}^*}}$ et on teste l'hypothèse nulle H_0 « X et Y ont la même distribution » contre l'hypothèse alternative H_1 « X et Y ont des distributions qui ont des fonctions de répartition différentes ».

Soient :
- $\hat{F}_X$ la fonction de répartition empirique associée à l'échantillon observé X ;
- F_Y la fonction de répartition associée à la loi de Y.

On applique la loi statistique suivante :

$$\forall n \in \mathbb{N}^*, D_n = \sup_{x\in\mathbb{R}} |\hat{F}_X(x) - F_Y(x)| \tag{F.3}$$

Sous H_0, cette loi suit asymptotiquement la relation suivante :

$$\sqrt{n} \cdot D_n \underset{n\to\infty}{\to} \sup_{x\in\mathbb{R}} |B(F_Y(x))| \tag{F.4}$$

Où B correspond à un pont brownien. La décision repose alors sur la convergence de cette quantité :
- **rejet de** H_0 si $\sqrt{n} \cdot D_n$ diverge ;
- **le non-rejet de** H_0 si $\sqrt{n} \cdot D_n$ converge.

Glossaire

ACP : abréviation de « Autorité de Contrôle Prudentiel », prédécesseur de l'ACPR jusqu'en 2010.

AIPP : abréviation de « Atteinte Permanente à l'Intégrité Physique et Psychique », taux qui sert à déterminer la réduction des capacités physiques ou intellectuelles d'une victime suite à un accident.

AP-HP : abréviation de « Assistance Publique - Hôpitaux de Paris », centre hospitalier universitaire composé de 39 hôpitaux répartis dans l'Île-de-France.

CCI : abréviation de « Comission de Conciliation et d'Indemnisation ».

Coefficient de corrélation de Pearson : coefficient de corrélation linéaire entre deux variables aléatoires réelles X et Y, d'espérances de variances finies (notées respectivement σ_X et σ_Y), qui suit la relation suivante :

$$\begin{aligned} \mathrm{Corr}(X,Y) &= \frac{\mathrm{Cov}(X,Y)}{\sigma_X \times \sigma_Y} \\ &= \frac{\mathbb{E}[(X-\mathbb{E}[X]) \times (Y-\mathbb{E}[Y])]}{\sigma_X \times \sigma_Y} \\ &= \frac{\mathbb{E}[X \times Y] - \mathbb{E}[X] \times \mathbb{E}[Y]}{\sigma_X \times \sigma_Y} \end{aligned} \tag{F.5}$$

Coefficient de détermination : pour une fonction $f : \mathbb{R} \to \mathbb{R}$ et une fonction de régression associée $\hat{f} : \mathbb{R} \to \mathbb{R}$, coefficient qui mesure la qualité prédictive de $\hat{f}$ sur un nombre de points fini $(x_i)_{i \in \mathbb{N}^*}$ par une méthode linéaire qui suit la relation suivante :

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(f(x_i) - \hat{f}(x_i) \right)^2}{\sum_{i=1}^{n} \left(f(x_i) - \overline{(f(x_i))}_{i \in [\![1,n]\!]} \right)^2} \tag{F.6}$$

Plus ce coefficient est proche de 1, plus il signifie que la relation linéaire entre f et $\hat{f}$ est forte. Ce coefficient peut s'interpréter par « $R^2\%$ des observations peuvent s'expliquer linéairement par le modèle régression étudié ».

DAM : abréviation de « Domaine d'Attraction Maximum », domaine où les lois respectent les hypothèses du théorème de Fisher-Tippett-Gnedenko.

GEV : abréviation de « *Generalized Extreme Value* », anglais de « loi d'extremum généralisée ».

GDP : abréviation de « *Generalized Pareto distribution* », anglais de « loi de Pareto généralisée ».

IBNER : abréviation de « *Incurred But Not Enough Reserved* », provisions constituées pour couvrir les provisions déjà constituées sur les sinistres survenus et déclarés à la date de clôture du bilan financier mais qui pourraient potentiellement s'avérer insuffisantes (aggravation probable du coût sinistre).

IBNYR : abréviation de « *Incurred But Not Yet Reported* », provisions qui estiment le coût ultime des sinistres survenus mais pas encore déclarés à la date de clôture (sinistres tardifs).

IBNR : abréviation de « *Incurred But Not Reported* », provision qui correspond à la relation suivante :

$$\text{IBNR} = \text{IBNER} + \text{IBNYR} \tag{F.7}$$

i.i.d. : abréviation de « [variables] indépendantes et identiquement distribuées ». Un ensemble de variables i.i.d. est un ensemble de variables aléatoires qui sont indépendantes et qui suivent toutes une même loi de probabilités.

Pour deux variables aléatoires réelles discrètes X et Y, elles sont i.i.d. si les relations suivantes sont respectées :

$$\forall (x,y) \in \mathbb{R}, \begin{cases} \mathbb{P}[X = x, Y = y] & = \mathbb{P}[X = x] \cdot \mathbb{P}[Y = y] \\ \mathbb{P}[X = x] & = \mathbb{P}[Y = x] \end{cases} \tag{F.8}$$

Dans le cas continu, deux variables aléatoires X et Y de fonctions de densité respectives f_X et f_Y sont i.i.d. si :

$$\forall (x,y) \in \mathbb{R}, \begin{cases} f_{X,Y}(x,y) & = f_X(x) \cdot f_Y(y) \\ f_X(x) & = f_Y(x) \end{cases} \tag{F.9}$$

où $f_{X,Y}$ représente la fonction de densité du couple (X, Y).

IPP : abréviation de « Invalidité Permanent Partielle ».

MCR : abréviation de « *Minimum Capital Requirement* », niveau minimum de fonds propres à constituer en bilan prudentiel. Un niveau de fonds propres en-dessous de ce niveau conduira à une intervention de l'autorité de contrôle.

MSEP : abréviation de « *Mean Square Error of Prediction* », erreur quadratique moyenne des prédictions.

ONIAM : abréviation de « Office National d'Indemnisation des Accidents Médicaux ».

POT : abréviation de « *Peak Over the Threshold* », méthode d'observations de réalisations d'une variable aléatoire au-delà d'un seuil.

PSAP : abréviation de « Provisions pour Sinistres À Payer », provision destinée à couvrir les coûts liés aux sinistres survenus et déclarés à la date de clôture des états financiers.

RCM : abréviation de « Responsabilité Civile Médicale », branche de la responsabilité civile professionnelle appliquée au secteur d'activité lié à la santé.

Résidus de Pearson : pour un ensemble de variables aléatoires observées X et estimées $\hat{X}$, représente la différence entre observations et estimations.

$$r^P = \frac{X - \hat{X}}{\sqrt{\mathrm{Var}[\hat{X}]}} \tag{F.10}$$

SCR : abréviation de « *Solvency Capital Requirement* », capital cible nécessaire à constituer en bilan prudentiel pour absorber l'ensemble des risques majeurs probables édictés par la directive Solvabilité II.

SEP : abréviation de « *Standard Error of Prediction* », erreur standard des prédictions.

SMR : abréviation de « *Standardized Mortality Ratio* », coefficient entre mortalité observée et mortalité estimée.

Théorème de Bayes : relation classique en théorie des probabilités qui stipule que pour deux évènements A et B, la relation suivante est respectée :

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A] \cdot \mathbb{P}[A]}{\mathbb{P}[B]} \tag{F.11}$$

Valeur p : en anglais « *p-value* », quantificateur de la significativité statistique d'un résultat issu d'un test basé sur une hypothèse nulle H_0 confrontée à une hypothèse alternative H_1. Sa valeur s'interprète en fonction de la décision qui veut être prise (accepter ou rejeter H_0 au seuil de significativité). Pour cela, les deux seuils suivants sont considérés :

- $\alpha \in [0, 1]$ le taux d'erreur de première espèce, *i.e.* la probabilité de rejeter à tort H_0 (communément fixé à 5%) ;
- $\beta \in [0, 1]$ le taux d'erreur de seconde espèce, *i.e.* la probabilité d'accepter à tort H_1 (défini à l'avance en fonction des données à disposition).

TABLE F.1 – Table de décision en fonction de la valeur p

Hypothèse à tester	H_0 vraie	H_1 vraie
Condition pour rejeter H_0	$p < \alpha$	$p < 1 - \beta$
Condition pour rejeter H_1	$p < 1 - \alpha$	$p < \beta$

Table des figures

1.1 Nombre de dossiers relevés par l'ORM entre 2009 et 2014 10
1.2 Indemnisations totales relevées par l'ORM et versées entre 2009 et 2014 . 11
1.3 Indemnisations moyennes déduites des données fournies par l'ORM, période 2009-2014 12
1.4 Nombre de mises en cause par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical 14
1.5 Nombre de médecins condamnés par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical 15
1.6 Indemnisations totales mises à charge aux médecins condamnés par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical 16
1.7 Indemnisations moyennes par spécialité sur l'année 2017, portefeuille MACSF-Le Sou Médical 17
1.8 Évolution de la sinistralité en établissement de santé sur la période 1998-2018 relativement à la fréquence renseignée au 30 juin 1998, portefeuille SHAM 19
1.9 Évolution du nombre de mises en cause et du nombre de condamnations en établissement de santé sur la période 2007-2017, portefeuille SHAM . . 20
1.10 Évolution des indemnisations moyennes en établissement de santé sur la période 2007-2017, portefeuille SHAM 21

2.1 Volumétrie du portefeuille 40
2.2 Montants cumulés du portefeuille 43
2.3 Proportion des « graves » dans le portefeuille en fonction du montant . . . 44
2.4 Montants moyens par dossier du portefeuille 46
2.5 Évolution du seuil de gravité déterminé par dichotomie en fonction de la date d'arrêté 50
2.6 Relevé de quantiles sur la charge cumulée par sinistre de la base d'étude, quantiles situés entre 99,0% et 99,9% 58
2.7 *QQ-plot* entre charges cumulées observées et la loi exponentielle standard ($\lambda = 1$) 59
2.8 *ME-plot* sur charges cumulées observées 60
2.9 *Pickands-plot* sur charges cumulées observées 61
2.10 *Hill-plot* sur charges cumulées observées 62
2.11 *DEdH-plot* sur charges cumulées observées 62
2.12 *QQ-plot* sur des lois d'extremum usuelles de type II 63

2.13 Évolution dans le temps des charges cumulées moyennes avec un seuil de gravité 150 000€ . . . 67
2.14 Évolution des principaux indices de prix à la consommation (source : data.imf.org) . . . 69
2.15 *ME-plot* sur charges cumulées *as-if* . . . 70
2.16 *GW-plot* sur charges cumulées observées, réparties sur la tranche [50 000, 250 000] . . . 71
2.17 *GW-plot* sur charges cumulées *as-if*, réparties sur la tranche [50 000, 250 000] 72
2.18 Évolution des seuils estimés par la méthode de Gerstengarbe et Werner sur le dernier centile dans le temps . . . 73

3.1 Synthèse de la procédure de calcul des provisions en responsabilité civile médicale employée . . . 88
3.2 Tests graphiques sur l'hypothèse de *Chain-Ladder*, triangle de liquidation des charges historiques, base réclamation . . . 90
3.3 Synthèse de la procédure de validation des méthodes de provisionnement suggérée . . . 100

A.1 Logo de l'ONIAM . . . 107

E.1 Exemples de *QQ-plot* sur des lois usuelles . . . 120
E.2 Exemples de *ME-plot* sur des lois usuelles . . . 121
E.3 Exemples de *Pickands-plot* sur des lois usuelles . . . 123
E.4 Exemples de *Hill-plot* sur des lois usuelles . . . 123
E.5 Exemples de *Dekkers-Einmahl-de Haan-plot* sur des lois usuelles . . . 124
E.6 Exemples de *Gerstengarbe-Werner-plot* sur des lois usuelles . . . 125

Liste des tableaux

1.1 Synthèse des modalités de provisionnement en responsabilité civile médicale 26

2.1 Volumétrie des bases initiales utilisées 32
2.2 Volumétrie des variables sélectionnées 33
2.3 Volumétrie de bases concaténées 33
2.4 Volumétrie des dates aberrantes retraitées 34
2.5 Volumétrie des agrégations sur opérations 35
2.6 Volumétrie des bases jointes puis scindées par année d'arrêté 36
2.7 Volumétrie des études de cohérence 38
2.8 Volumétrie des retraitements à l'issue de l'étude de cohérences 39
2.9 Volumétrie des bases après retraitements 40
2.10 Proportion des dossiers en fonction des charges et règlements 41
2.11 Régressions polynomiales de degré 2 sur le volume des dossiers en fonction de leur type 42
2.12 Projections polynomiales des volumes du portefeuille 42
2.13 Régressions polynomiales de degré 2 sur les montants cumulés en fonction de leur type 45
2.14 Projections polynomiales des montants cumulés du portefeuille 45
2.15 Volumétrie du portefeuille arrêté au 31/03/2019 en fonction du seuil de charges cumulées fixé 48
2.16 Caractéristiques de la GEV en fonction du paramètre ξ 54
2.17 Valeurs p issues de tests statistiques d'adéquation des lois étudiées avec la sinistralité grave du portefeuille étudié 65
2.18 Distinction selon la mise en *as-if* ou non des répartitions conditionnées par le seuil de gravité 71

3.1 Évolution du bilan sur activité d'assurance entre les directives Solvabilité I et Solvabilité II 76
3.2 Illustration d'un triangle de liquidation 77
3.3 Montants totaux sur les triangles de liquidation construits 89
3.4 Échantillon d'estimateurs *Chain-Ladder* obtenus 92
3.5 Provisions constituées selon la méthode *Chain-Ladder* par groupe de risque 93
3.6 Provisions constituées selon la méthode *Chain-Ladder* par base de rattachement 94

3.7 Échantillon d'estimateurs *London-Chain* obtenus 94
3.8 Provisions constituées selon la méthode *London-Chain* par base de rattachement 95
3.9 Échantillon d'estimateurs Bornhuetter-Ferguson obtenus 96
3.10 Provisions constituées selon la méthode Bornhuetter-Ferguson par base de rattachement 97
3.11 Échantillon d'estimateurs Mack obtenus 98
3.12 Provisions constituées selon la méthode de Mack par base de rattachement 98
3.13 Provisions constituées selon la méthode de Mack par base de rattachement 99
3.14 Synthèse des écarts entre observés et estimés selon la méthode de provisionnement utilisée 101

B.1 Synthèse des variables retenues de la base initiale « sinistres historiques, partie 1 » 110
B.2 Synthèse des variables retenues de la base initiale « sinistres historiques, partie 2 » 111
B.3 Synthèse des variables retenues de la base initiale « base règlements », arrêtée au 31/03/2019 112
B.4 Synthèse des variables retenues de la base initiale « base provisions », arrêtée au 31/03/2019 113
B.5 Synthèse des variables retenues de la base initiale « base recours », arrêtée au 31/03/2019 114

C.1 Synthèse des principales variables de la base finale obtenue, arrêtée au 31/03/2019 116

F.1 Table de décision en fonction de la valeur p 132

Bibliographie

[Association des Professionnels de la Réassurance en France, 2011] Association des Professionnels de la Réassurance en France (2011). Évolution de la Responsabilité Civile Médicale en France. https://www.apref.org/sites/default/files/espacedocumentaire/3_note_apref-rc_medicale_novembre_2011.pdf (Consulté le 05/08/2019).

[Banga, 2017] Banga, B. (2017). RC Médicale : un marché à la santé fragile. *La Tribune de l'Assurance*. https://www.tribune-assurance.fr/article/7719-rc-medicale-un-marche-a-la-sante-fragile (Consulté le 03/08/2019).

[Benjamin et Eagles, 1986] Benjamin, S. et Eagles, L. (1986). Reserves in Lloyd's and the London market. *Journal of the Institute of Actuaries*, 113(2):197–256.

[Bercy Infos, 2018] Bercy Infos (2018). Entreprises, quelles sont les assurances obligatoires ? https://www.economie.gouv.fr/entreprises/assurances-obligatoires (Consulté le 28/07/2019).

[Bornhuetter et Ferguson, 1972] Bornhuetter, R. et Ferguson, R. (1972). The Actuary and IBNR. *Proceedings of the Casualty Actuarial Society*, 59(112):181–192.

[Code civil, 2016] Code civil (2016). Chapitre Ier : La responsabilité extracontractuelle en général . Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idSectionTA=LEGISCTA000032021488&cidTexte=LEGITEXT000006070721&dateTexte=20190728 (Consulté le 28/07/2019).

[Code de la santé publique, 2011] Code de la santé publique (2011). Chapitre II : Risques sanitaires résultant du fonctionnement du système de santé. Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000020628252&idSectionTA=LEGISCTA000006185260&cidTexte=LEGITEXT000006072665&dateTexte=20190531 (Consulté le 31/05/2019).

[Code de la santé publique, 2012] Code de la santé publique (2012). Chapitre VII : Déontologie. Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000025843568&idSectionTA=LEGISCTA000006196408&cidTexte=LEGITEXT000006072665&dateTexte=20190601 (Consulté le 01/06/2019).

[Code des assurances, 2003a] Code des assurances (2003a). Chapitre IV : Les assurances de responsabilité. Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idSectionTA=LEGISCTA000006157248&cidTexte=LEGITEXT000006073984&dateTexte=20190806 (Consulté le 06/08/2019).

[Code des assurances, 2003b] Code des assurances (2003b). Section III : Provisions techniques des autres opérations d'assurance - Article R331-6. Journal Officiel. https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006073984&idArticle=LEGIARTI000006816253&dateTexte=20190807 (Consulté le 07/08/2019).

[Code des assurances, 2003c] Code des assurances (2003c). Sous-section 1 : Dispositions générales sur la valorisation des provisions techniques prudentielles. Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idSectionTA=LEGISCTA000030433272&cidTexte=LEGITEXT000006073984&dateTexte=20190807 (Consulté le 07/08/2019).

[Code des assurances, 2011] Code des assurances (2011). Titre V : L'assurance de responsabilité civile médicale. Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000025859349&idSectionTA=LEGISCTA000006157528&cidTexte=LEGITEXT000006073984&dateTexte=20190531 (Consulté le 31/05/2019).

[Code des assurances, 2015] Code des assurances (2015). Sous-section 3 : Provisions techniques des autres opérations d'assurance . Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idSectionTA=LEGISCTA000030584411&cidTexte=LEGITEXT000006073984&dateTexte=20190820 (Consulté le 20/08/2019).

[Code pénal, 2000] Code pénal (2000). Titre II : De la responsabilité pénale . Journal Officiel. https://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000006417201&idSectionTA=LEGISCTA000006149817&cidTexte=LEGITEXT000006070719&dateTexte=20190728 (Consulté le 28/07/2019).

[Dintilhac, 2005] Dintilhac, J. (2005). Rapport du groupe de travail chargé d'élaborer une nomenclature des préjudices corporels. https://solidarites-sante.gouv.fr/IMG/pdf/Rapport_groupe_de_travail_nomenclature_des_prejudices_corporels_de_Jean-Pierre_Dintilhac.pdf (Consulté le 10/08/2019).

[Dubuisson, 2008] Dubuisson, B. (2008). SolidaritÃ©, segmentation et discrimination en assurances, nouveau dÃ©bat, nouvelles questions. Colloque Cour de Cassation. https://www.courdecassation.fr/venements_23/colloques_4/2008_2484/intervention_bernard_dubuisson_11742.html (Consulté le 29/08/2019).

[Efron, 1979] Efron, B. (1979). Bootstrap methods : another look at the jack-knife. *The Annals of Statistics*, 7(1):1–26.

[Fisher et Tippett, 1928] Fisher, R. et Tippett, L. (1928). Limiting forms of the frequency distribution of the largest and smallest member of asample. *Mathematical Proceedings of the Cambridge Philosophical Society*, 24:180–190.

[Fréchet, 1927] Fréchet, M. (1927). Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématique, Cracovie*, 6:93–116.

[Gnedenko, 1948] Gnedenko, B. (1948). On a local limit theorem of the theory of probability. *Uspekhi Matematicheskikh Nauk*, 3(25):184–194.

[Inspection générale des affaires sociales, 2007] Inspection générale des affaires sociales (2007). L'assurance en responsabilité civile médicale. https://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/074000243.pdf (Consulté le 29/08/2019).

[Jaziri, 2011] Jaziri, S. (2011). Méthodes de provisionnement non-vie et risque de réserve à un an. Mémoire d'actuaire, ISFA. http://www.ressources-actuarielles.net/C12574E200674F5B/0/EB475A9EFCC6B761C12578C10069E42B.

[Johanet, 2011] Johanet, G. (2011). Rapport sur la Responsabilité Civile Médicale. https://solidarites-sante.gouv.fr/IMG/pdf/Rapport_Responsabilite_Civile_Medicale.pdf (Consulté le 31/07/2019).

[Journal Officiel, 2002a] Journal Officiel (2002a). Loi n2002-1577 du 30 décembre 2002 relative à la responsabilité civile médicale. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000234122&dateTexte=20190731 (Consulté le 31/07/2019).

[Journal Officiel, 2002b] Journal Officiel (2002b). Loi n2002-303 du 4 mars 2002 relative aux droits des malades et à la qualité du système de santé. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000227015&dateTexte=20190731 (Consulté le 31/07/2019).

[Journal Officiel, 2011] Journal Officiel (2011). Loi n2011-1977 du 28 décembre 2011 de finances pour 2012 - Article 146. https://www.legifrance.gouv.fr/eli/loi/2011/12/28/2011-1977/jo/article_146 (Consulté le 28/07/2019).

[Le Tesson *et al.*, 2014] Le Tesson, A., Lenain, A., Samba, S. et Ung, J. (2014). Estimation de l'erreur de prédiction dans le cas de l'utilisation d'une combinaison de méthodes pour le calcul de provisions en assurance IARD. Bureau d'études, EURIA. https://euria.univ-brest.fr/digitalAssets/27/27606_BE--Estimationdelerreurdeprediction-dans.pdf.

[Mack, 1993] Mack, T. (1993). Distribution-free calculation of the standard error of Chain-Ladder reserve estimates. *Astin Bulletin*, 23(2):213–225.

[Maniable, 2013] Maniable, E. (2013). Provisionnement Responsabilité Civile Médicale. Mémoire d'actuaire, ISFA. http://www.ressources-actuarielles.net/C12574E200674F5B/0/CE5241ADA1B02C78C1257BA5001AE1B9.

[Mann, 2015] Mann, L. (2015). Provisionnement et réassurance en RC Médicale. Mémoire d'actuaire, ISFA. http://www.ressources-actuarielles.net/C12574E200674F5B/0/0DEEBC5657DDB3D3C1257D69005820E8.

[Mutuelle d'Assurance du Corps de Santé Français - Le Sou Médical, 2017] Mutuelle d'Assurance du Corps de Santé Français - Le Sou Médical (2017). Le Risque Médical - Rapport annuel 2017. https://www.macsf-exerciceprofessionnel.fr/content/download/17922/203573/version/1/file/rapport%20annuel%202017%20VF.pdf (Consulté le 04/08/2019).

[Observatoire des Risques Médicaux, 2015] Observatoire des Risques Médicaux (2015). Rapport d'activité 2015 - Années 2009 à 2014. https://www.oniam.fr/medias/uploads/ORM/Rapport_ORM_2015__2009-2014.pdf (Consulté le 01/08/2019).

[Office National d'Indemnisation des Accidents Médicaux, 2018] Office National d'Indemnisation des Accidents Médicaux (2018). Rapport d'activité 2018. https://www.oniam.fr/medias/uploads/RAPPORT%20ACTIVITE%202015/ONIAM%20-%20Rapport%20activite%202018%20.pdf (Consulté le 03/08/2019).

[Oulidi, 2008] Oulidi, A. (2008). Apport de la statistique semi-paramétrique et non paramétrique à la résolution de certains problèmes en assurance non-vie. Mémoire d'actuaire, CEA. http://www.ressources-actuarielles.net/C12574E200674F5B/0/E638F0114EEEC3AEC125786500655D04.

[Parlement Européen et Conseil de l'Union europénne, 2009] Parlement Européen et Conseil de l'Union européenne (2009). Directive 2009/138/CE du Parlement européen et du Conseil du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II). Journal officiel de l'Union européenne. https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=OJ:L:2009:335:FULL&from=FR (Consulté le 25/08/2019).

[Planchet, 2017] Planchet, F. (2017). Modèles financiers en assurance et analyses dynamiques, Utilisation de la théorie des valeurs extrêmes dans le cadre de Solvabilité 2. http://www.ressources-actuarielles.net/EXT/ISFA/fp-isfa.nsf/0/72EE1310B7EBC2A2C1256FD2002E9C76/$FILE/Seance3-01.pdf (Consulté le 17/08/2019).

[Planchet et Thérond, 2007] Planchet, F. et Thérond, P. (2007). *Mesure et gestion des risques d'assurance, Analyse critique des futurs référentiels prudentiel et d'information financière*. Economica.

[Raggad, 2009] Raggad, B. (2009). Fondements de la théorie des valeurs extrêmes, ses principales applications et son apport à la gestion des risques du marché pétrolier. *Mathematics and Social Sciences*, 186(2):29–63. https://journals.openedition.org/msh/11069?file=1 (Consulté le 17/08/2019).

[Rouge-Maillart, 2018] Rouge-Maillart, Docteur C. (2018). La responsabilité médicale. https://sfml-asso.fr/images/docs/responsabilite.pdf (Consulté le 28/07/2019).

[Savarre et Payre, 2012] Savarre, S. et Payre, B. (2012). Charge ultime nette de réassurance en RC corporelle : 2 modèles stochastiques pour les flottes automobiles. Mémoire d'actuaire, CEA. http://www.ressources-actuarielles.net/C12574E200674F5B/0/B950959ACA4C0EA3C1257A3C0033818F.

[Société Hospitalière d'Assurances Mutuelles, 2018] Société Hospitalière d'Assurances Mutuelles (2018). Santé, social et médico-social - Chiffres clés de la Responsabilité Civile en établissements - Bilan 2017. https://www.sham.fr/sites/default/files/press-release-files/181002_SHAM_CP_CHIFFRES-CLES_RC_2018_FR_VDEF_0.pdf (Consulté le 04/08/2019).

[Weibull, 1951] Weibull, E. (1951). A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 18(3):293–297. http://www.barringer1.com/wa_files/Weibull-ASME-Paper-1951.pdf (Consulté le 26/08/2019).