



AON France - Reinsurance solutions

Master ACTUARIAT, Université de Strasbourg

**Modélisation sur exposition de la sinistralité incendie
dans un contexte de réassurance non-proportionnelle
par risque**

12 avril 2024

Auteur :
Bastien Camboulive

Supervisé par :
Franck Lopez

Résumé

Par nature de la réassurance non-proportionnelle par risque s'appliquant généralement aux risques de pointe, les données historiques utilisées pour modéliser la sinistralité d'une cédante sont peu nombreuses. Les méthodes sur expérience et exposition de tarification des traités en excédent de sinistre en incendie présentées en première partie de ce mémoire ne suffisent parfois pas à obtenir des modélisations robustes. Notamment, la méthode utilisée en interne (se basant sur l'historique de sinistralité de la cédante) permet une caractérisation totale de la sinistralité brute au-delà d'un seuil et l'estimation de différents indicateurs statistiques. L'enjeu ici n'est pas uniquement de modéliser une prime pour une tranche de réassurance donnée, mais bien d'estimer la distribution de la sinistralité brute des graves individuels des cédantes.

L'objectif de ce mémoire est de développer des méthodologies de modélisation sur exposition de la sinistralité incendie qui permettront de s'affranchir d'une insuffisance de statistiques sinistres inhérente à la nature de la réassurance non-proportionnelle et de caractériser la distribution de la sinistralité brute au-delà d'un seuil des cédantes. Pour répondre à cette problématique, deux méthodologies sont proposées. La première s'apparente à une modélisation marché, réajustée par l'exposition de la cédante. La seconde s'inscrit dans la continuité de la première, en rajoutant une dimension explicative à la sinistralité par l'utilisation de la relation liant somme assurée et sinistralité en incendie. Aussi, les méthodes de tarifications sur exposition 'classiques' seront implémentées et confrontées aux méthodes que nous proposons. Les limites et apports de ces méthodes sont étudiés, notamment grâce à une utilisation en interne de la méthode pour le premier modèle.

Abstract

Due to the nature of non-proportional risk reinsurance, which generally applies to peak risks, the historical data used to model a cedant's losses is limited. The experience and exposure pricing methods for excess of loss fire treaties presented in the first part of this report are sometimes insufficient to produce robust models. In particular, the method used internally (based on the cedant's claims history) enables us to fully characterize the gross claims experience above a certain threshold, and to estimate various statistical indicators. The challenge here is not simply to model a premium for a given reinsurance bracket, but to estimate the distribution of the gross claims experience of individual cedants.

The aim of this dissertation is to develop methodologies for modelling fire claims on exposure, which will make it possible to overcome the inadequacy of claims statistics inherent in the nature of non-proportional reinsurance, and to characterize the distribution of cedants' gross claims experience above a threshold. Two methodologies are proposed to address this issue. The first is similar to a market model, adjusted by the cedant's exposure. The second uses the relationship between sum insured and pure fire loss experience to model severity more precisely, policy by policy. Classical exposure pricing methods will be implemented and compared with our proposed methods. The limits and contributions of these methods are examined, in particular through in-house use of the method for the first model.

Remerciements

Ce mémoire ayant été réalisé au sein du service Analytics d'Aon France, j'aimerais remercier Fabien Ramaharobanbro de m'avoir accueilli dans sa belle équipe.

Mes remerciements vont aussi à tous les membres de l'équipe Actuariat pour leur bonne humeur, les bons moments passés ensemble durant cette année d'alternance, et pour leurs conseils avisés sur la rédaction de ce mémoire. Plus particulièrement, je tiens à remercier mon tuteur professionnel, Franck Lopez, sans qui ce mémoire n'aurait pas vu le jour.

Aussi, je souhaite remercier tout le corps enseignant du DUAS pour leur engagement, et pour m'avoir permis d'acquérir toutes les connaissances nécessaires à la rédaction d'un mémoire d'actuariat.

Plus largement, je remercie toutes les personnes qui ont contribué de près ou de loin à la rédaction et à l'aboutissement de ce mémoire.

Enfin, je tiens à exprimer toute ma gratitude et tout mon amour envers celle qui m'apporte stabilité, amour et soutien inconditionnel depuis plus de 4 ans maintenant, Vanessa.

Table des matières

Introduction	12
I Contexte de l'étude	14
1 Généralités sur la réassurance	15
1.1 Le mécanisme de réassurance	15
1.2 Les différents types de réassurance	15
1.3 Les principales formes de réassurance	16
1.3.1 La réassurance proportionnelle	16
1.3.2 La réassurance non-proportionnelle	18
1.3.3 Les clauses	22
2 Méthodes de tarification d'un traité XS	24
2.1 La notion de tranches	24
2.2 Tarification à l'expérience	25
2.2.1 Mise en <i>as-if</i>	25
2.2.2 La méthode du <i>Burning Cost</i>	26
2.2.3 La méthode probabiliste	28
2.3 Tarification sur exposition en incendie	30
3 Le risque incendie : couverture en réassurance et modélisation utilisée en interne	37
3.1 Les traités Dommages Aux Biens	37
3.2 Le rôle d'AON France	39
3.2.1 Le courtage en réassurance	39
3.2.2 Le service Analytics	40
3.3 Méthode AON de modélisation de la sinistralité incendie	41
3.3.1 Les données	41
3.3.2 Calibration des lois de fréquence et sévérité	45

4 Problématique du mémoire	49
II Première approche de modélisation sur exposition : le modèle 0D	51
5 Méthodologie	53
5.1 Construction d'une base marché	53
5.2 Gestion des seuils de communication	53
5.3 Ajustement des lois de fréquence et de sévérité par segment	57
5.4 La question de l'inflation	58
6 Application à un cas pratique	60
6.1 Présentation du portefeuille de la cédante et de la structure étudiée	60
6.1.1 Le portefeuille	60
6.1.2 La structure de réassurance	62
6.2 Modélisation sur expérience	63
6.3 Modélisation sur exposition : utilisation du modèle 0D	65
6.4 Comparaison des résultats avec les méthodes sur expérience et sur exposition	72
6.5 Conclusion	75
7 Limites du modèle	76
III Deuxième approche de modélisation sur exposition : implémentation des méthodes classiques de tarification à l'aide des courbes d'exposition	78
8 Estimation de la prime pure d'un traité à l'aide des courbes d'exposition	80
8.1 Méthodologie et choix des paramètres pour l'application à la cédante A	80
8.2 Résultats obtenus lors de l'application au cas pratique de la cédante A	85
9 Approche stochastique	87
9.1 Méthodologie et choix des paramètres pour l'application à la cédante A	87
9.1.1 Estimation de la fréquence	87
9.1.2 Estimation de la sévérité	89
9.1.3 Paramètres utilisés	89
9.2 Résultats obtenus lors de l'application au cas pratique de la cédante A	90
10 Les limites des méthodes sur exposition classiques	92
IV Troisième approche de modélisation sur exposition : modèle 1D sur somme assurée	95
11 Méthodes de régressions pour la calibration	97
11.1 Les différents modèles	97

11.1.1 Régressions linéaires	98
11.1.2 Modèles linéaires généralisés - GLM	98
11.1.3 Modèles additifs généralisés - GAM	100
11.1.4 Modèles additifs généralisés pour la position, l'échelle et la forme - GAMLSS	100
11.2 Critères de sélection des modèles	101
12 Méthodologie	104
12.1 Calibration des modèles	104
12.1.1 Construction d'une base d'apprentissage	104
12.1.2 Calibration de la fréquence	105
12.1.3 Calibration de la sévérité	106
12.2 Utilisation des modèles	107
12.2.1 Premier cas	107
12.2.2 Deuxième cas	108
12.3 La question de l'inflation	108
13 Développement de l'outil R et validation des méthodes	110
13.1 Génération des données	110
13.1.1 Génération d'un portefeuille d'exposition	110
13.1.2 Génération des sinistres	113
13.2 Gestion des données	115
13.3 Calibration et validation des modèles	115
13.3.1 Calibration de la fréquence	115
13.3.2 Calibration de la sévérité	115
13.4 Simulations	118
14 Étude sur données réelles	120
14.1 Les données	120
14.2 Premier cas : sinistres transmis sans le détail de la répartition dommage/responsabilité	121
14.3 Deuxième cas : sommes assurées non disponibles	125
15 Limites du modèle	127
Conclusion	130
Bibliographie	134
Annexes	ii
A La méthode de Chain-Ladder	ii
A.1 Présentation	ii
A.2 Théorie	ii

A.3	Limites de la méthode	iii
B	Méthodes d'estimations paramétriques	v
B.1	Rappels sur les estimateurs	v
B.2	La méthode des moments	vi
B.3	Le maximum de vraisemblance	vii

Table des figures

1	Illustration de la répartition cédante/réassureur dans le cadre d'un traité QP	17
2	Illustration de la répartition cédante/réassureur dans le cadre d'un traité QP	18
3	Fonctionnement d'un traité XS	19
4	Structure de réassurance en tranches d'XS	20
5	Traité 5 XS 5	21
6	Récapitulatif des différentes formes de réassurance	22
7	Ajustement commun à toutes les méthodes par expérience	25
8	Mise en As-If	26
9	Fonctionnement d'une méthode sur exposition	31
10	Exemple d'une courbe d'exposition	32
11	Courbe d'exposition Swiss Re et Lloyd's	35
12	Couverture XS DaB - Incendie et Risques annexes	38
13	Rôle d'un courtier en réassurance	39
14	Processus de modélisation AON	41
15	Hétérogénéité de la base sans indexation du seuil de communication	42
16	Impact du seuil de modélisation sur les cadences de développements moyennes estimées	44
17	Étapes globales du processus de modélisation AON	49
18	Segmentation de la base	54
19	Seuils de modélisation	55
20	Affectation des sinistres en fonction de la charge indexée maximale et du seuil de communication	56
21	Gestion de la segmentation après projection à l'ultime	57
22	Capitaux sous risques des souscriptions Dommages aux biens	61

23	Répartition des biens et des sommes assurées en portefeuille	61
24	Profil de risques de la cédante A	62
25	Structure de réassurance étudiée	63
26	Fonctions de répartition de différentes lois ajustées sur les données de la cédante . .	64
27	Résultat de la calibration de la loi de fréquence avec les données de la cédante . .	64
28	Répartition des risques par portefeuille	65
29	Seuils de communication de la base marché ainsi contruite	66
30	Coefficients de Chain-Ladder	68
31	Sinistres supérieurs au seuil indexé par année de survenance	69
32	Sinistres supérieurs au seuil indexé par année de survenance	69
33	Mise à l'échelle des paramètres de fréquence marché	69
34	Fonction de répartition empirique de la charge des sinistres	70
35	Modèle 0D - Lois de sévérité choisies pour la modélisation	71
36	Récapitulatif des lois de fréquence et sévérité utilisées pour chaque modélisation . .	72
37	Sinistralité brute	73
38	Résultats de modélisation du traité 47 XS 3 en fonction de l'approche utilisée	74
39	Loss Ratios marché aux États-Unis par LoB	83
40	Méthode d'estimation de la prime pure d'un traité de réassurance par les courbes d'exposition appliquée à la cédante A	85
41	Résultats de modélisation du traité 47 XS 3 en fonction de l'approche utilisée	90
42	Estimation de la prime pure du traité en fonction des paramètres utilisés	93
43	Illustration graphique d'une régression linéaire	98
44	Illustration graphique d'un GLM	99
45	Illustration graphique d'un GAMLSS	101
46	Schéma des relations entre les différentes régressions paramétriques	101
47	AIC en fonction des degrés de liberté d'un modèle	102
48	Jointure de la base sinistres aux bases d'exposition ligne à ligne en fonction de l'année de survenance des sinistres et de la date d'extraction des portefeuilles	105
49	Fonction de répartition empirique des valeurs assurées	111
50	Fonction de répartition empirique des sommes assurées avec interpolation linéaire	112
51	Histogramme des sommes assurées simulées obtenues	113
52	Sinistres simulés	114
53	Espérance et Variance en fonction de la somme assurée	116
54	Distribution des résidus	117
55	Quantiles du modèle	117
56	Fonctionnement du modèle sur somme assurée	118
57	Histogrammes des sinistres générés par le modèle	118

58	Sinistres d'origine (en noir), sinistres simulés (en rouge)	119
59	Répartition des charge ultimes en fonction de la somme assurée	122
60	Répartition des charge ultimes en fonction de la somme assurée après différentes transformations	123
61	Répartition des taux de dommage en fonction de la somme assurée	124
62	Répartition des sinistres Dommages (en rouges) vs Responsabilité (en bleus)	125
63	Fonctionnement de la méthode sur exposition 'classique'	131
64	Triangle de développement	iii
65	Remplissage en cascade d'un triangle de développement	iii

Liste des tableaux

1	Exemple d'un quote-part	16
2	Exemple d'un excédent de plein	18
3	Exemple d'un excédent de plein	20
4	Données de l'exemple	23
5	Application de l'XS avec clauses	23
6	Données sinistres pour le calcul du <i>Burning Cost</i>	27
7	Lois utilisées pour modéliser la fréquence dans la méthode AON	29
8	Lois utilisées usuellement pour modéliser la sévérité des graves	29
9	Exemple d'un profil de risques à T bandes	31
10	Échantillon de fréquence illustratif	47
11	Descriptif des cédantes de la base marché constituée	66
12	Seuils de communication des cédantes de la base marché avant et après indexation	67
13	Tableau des segments obtenus	67
14	Paramètres c en fonction de la somme assurée en 2012	81
15	Paramètres c en fonction de la somme assurée indexée	82
16	Prime pure du traité en fonction du modèle utilisé	85
17	Bandé fictive d'un profil de risque pour application de la méthode	88
18	Lois à support $[0, +\infty]$	106
19	Lois à support $[0, 1]$	106
20	Profil de risque fictif utilisé pour la génération du portefeuille d'exposition	111
21	AIC des modèles testés	116
22	Données disponibles	121

23	Données obtenues après jointure sur les bases d'exposition et suppression des évènements	122
24	Récapitulatif des données disponibles	125
25	Statistiques sinistres	125

Introduction

Introduction

L'inversion du cycle de production est caractéristique de l'activité d'assurance. Les primes sont encaissées avant que les règlements n'interviennent. Dès lors, le rôle de l'actuaire est de probabiliser et quantifier au mieux ces risques, qui, par nature, sont incertains. La gestion des risques est le cœur du métier de l'actuaire.

Le contexte incertain et risqué de l'assurance a entraîné au fil des années la mise en place d'un certain nombre de normes réglementaires (Solvabilité II, IFRS 9/17 pour ne citer qu'elles...), qui ont pour principal objectif de prévenir le risque de ruine. En effet, un assureur doit toujours pouvoir honorer ses engagements. Pour rester solvable face aux risques en toutes circonstances et augmenter ses capacités de souscription, l'assureur peut avoir recours à la réassurance. En échange d'une prime que l'assureur versera au réassureur, certains risques définis contractuellement seront couverts par le réassureur : on parle de transfert de risque. Les assureurs cherchent à obtenir une structure de réassurance offrant un transfert de risque optimal, tout en maintenant des coûts maîtrisés. De leur côté, les réassureurs ont pour objectif de fixer un prix pour la réassurance qui reflète véritablement le risque pris, tout en restant compétitifs sur le marché.

Le courtier en réassurance assume un rôle d'analyste et d'exécutant pour la cédante, tout en jouant un rôle de représentation auprès des réassureurs. En particulier, le courtier agit en tant que conseiller stratégique, travaillant en étroite collaboration avec les cédantes pour optimiser leur structure de réassurance. Cette collaboration implique une analyse approfondie de leur exposition aux risques, de leur sinistralité brute, ainsi que de leurs besoins spécifiques en matière de protection et de gestion des risques. La responsabilité du courtier comprend également la modélisation de l'ensemble de l'activité des cédantes, y compris les encaissements, les frais et autres paramètres pertinents, afin de concevoir des programmes de réassurance sur mesure répondant pleinement à leurs exigences et aux spécificités de leur profil de risque. L'approche privilégiée chez AON met en avant la modélisation du brut comme pilier central de la démarche. Cette modélisation permet non seulement d'évaluer de manière exhaustive les différents impacts de la réassurance, mais aussi

d'identifier les opportunités d'optimisation et de gestion des risques pour les clients et est utilisée par les commerciaux comme support de négociation avec les réassureurs pendant la période de renouvellement.

Néanmoins, cette diminution du risque à un coût pour les assureurs. En réassurance non-proportionnelle, la tarification de ces protections peut varier d'une méthode à l'autre. En fonction de la structure modélisée et des données disponibles, la tarification peut être faite à l'aide de méthodes sur expérience ou à l'aide de méthodes sur exposition. Les méthodes sur expérience reposent sur les données historiques des cédantes, ce qui permet d'évaluer le risque en se basant sur des événements passés. Cependant, l'insuffisance ou la qualité limitée des données historiques peut rendre la tarification sur expérience non pertinente dans certains cas. Les méthodes sur exposition, quant à elles, sont généralement basées sur des données de sinistralité marché, offrant ainsi une perspective plus globale du risque. En réassurance non-proportionnelle s'appliquant aux risques de pointe, il n'est pas rare d'observer une insuffisance statistique qui peut rendre la tarification sur expérience non pertinente : les méthodes sur exposition peuvent alors être utilisées pour combler l'insuffisance des données de sinistralité historiques d'une cédante. Ainsi, elles offrent une alternative précieuse pour évaluer et tarifer le risque, contribuant à une prise de décision éclairée dans le secteur de la réassurance.

En incendie, des méthodes sur exposition se basant sur la somme assurée des polices en portefeuille existent, mais elles se focalisent sur l'obtention d'un prix de réassurance. La méthode AON de modélisation de la sinistralité incendie s'appuie sur une méthode sur expérience qui permet de modéliser la sinistralité brute au-delà d'un seuil (nous travaillons avec des données tronquées) de la cédante pour avoir une vision de la distribution de la sinistralité brute, du prix de la réassurance, de la rétention de la cédante, etc... Face à des données de sinistralité historique insuffisantes, l'utilisation d'une méthode sur expérience n'est pas judicieuse : une méthode se basant sur l'exposition de la cédante serait bien plus pertinente et représentative du risque réel de la cédante.

L'objectif de ce mémoire est de concevoir et mettre en œuvre des méthodes de modélisation de la sinistralité brute incendie (au-delà d'un seuil) basées sur l'exposition, offrant une représentation plus précise du risque encouru par les cédantes dont les données de sinistralité sont limitées. Ces méthodes visent idéalement à fournir une estimation détaillée des distributions de la sévérité brute au-delà du seuil de modélisation, ainsi que des caractéristiques clés telles que les cessions de réassurance et la rétention.

Première partie

Contexte de l'étude

1 Généralités sur la réassurance	15
1.1 Le mécanisme de réassurance	15
1.2 Les différents types de réassurance	15
1.3 Les principales formes de réassurance	16
2 Méthodes de tarification d'un traité XS	24
2.1 La notion de tranches	24
2.2 Tarification à l'expérience	25
2.3 Tarification sur exposition en incendie	30
3 Le risque incendie : couverture en réassurance et modélisation utilisée en interne	37
3.1 Les traités Dommages Aux Biens	37
3.2 Le rôle d'AON France	39
3.3 Méthode AON de modélisation de la sinistralité incendie	41
4 Problématique du mémoire	49

CHAPITRE 1

Généralités sur la réassurance

1.1 Le mécanisme de réassurance

La réassurance est souvent présentée comme « l'assurance des assureurs ». Plus formellement, il s'agit d'une opération par laquelle un assureur (la cédante) fait garantir par un réassureur tout ou une partie des risques qu'il a lui-même couverts, en échange d'une prime. Le réassureur joue ainsi un rôle très important dans l'écosystème assurantiel. En effet, il permet aux cédantes d'une part de se protéger contre les gros risques qu'elles ne pourraient pas forcément assurer seules, et d'autre part d'augmenter leurs capacités de souscription, et donc la taille de leurs portefeuilles, tout en respectant les exigences de solvabilité imposées par la norme Solvabilité II. De la même manière qu'un assureur, un réassureur va essayer de se construire un portefeuille diversifié en mutualisant les grands risques. Cependant, par définition de la réassurance non proportionnelle, les réassureurs seront uniquement concernés par les risques exceptionnels : il est alors plus délicat de récolter suffisamment de données de bonne qualité, permettant l'étude des phénomènes concernés et assurant ainsi la qualité du portefeuille. C'est pour pallier cette faiblesse que différents modèles et méthodes ont été développés. Ils seront présentés succinctement dans la suite de cette partie.

1.2 Les différents types de réassurance

Réglementairement parlant, il existe deux types de réassurance :

- La réassurance facultative
- La réassurance obligatoire

La réassurance obligatoire

On parle de traité de réassurance obligatoire lorsqu'une cédante souhaite couvrir tout son portefeuille pour une certaine catégorie de risque (Property, par exemple). Le terme de réassurance

obligatoire vient de l'obligation des deux parties au contrat de céder ou d'accepter tous les risques couverts dans le traité de réassurance conclu. Nous nous intéresserons uniquement à ce type de réassurance dans le cadre de cette étude.

La réassurance facultative

La réassurance facultative vient souvent en complément de la réassurance obligatoire, lorsque la cédante veut couvrir un risque spécifique, non pris en compte dans les traités obligatoires. On parle de réassurance facultative car le réassureur n'est pas tenu d'accepter de porter le risque concerné. De la même manière, la cédante choisit les risques qu'elle souhaite céder au réassureur.

1.3 Les principales formes de réassurance

Il existe deux grandes familles de contrats en réassurance : la réassurance proportionnelle, et la réassurance non-proportionnelle. Il est à la charge de la cédante et du réassureur de décider de la forme que la couverture de réassurance prendra, après négociations directs (ou par l'intermédiaire d'un courtier).

1.3.1 La réassurance proportionnelle

Un traité de réassurance est dit proportionnel lorsque la cédante cède au réassureur un pourcentage des primes qu'il touche pour un portefeuille donné, et le même pourcentage de la sinistralité associée à ce portefeuille. Ce pourcentage, ainsi que les modalités de fonctionnement du traité sont définis contractuellement. Il existe deux grandes catégories de traités de réassurance proportionnelle.

Les traités en quote-part

Il s'agit d'une forme très simple de réassurance. La cédante fixe un pourcentage, appelé taux de cession, et le réassureur prend en charge ce pourcentage de sinistralité pour chaque sinistre couvert par le traité. La prime versée au réassureur est obtenue en appliquant le taux de cession à l'assiette des primes encaissées par la cédante pour les garanties concernées.

Considérons un traité QP avec un taux de cession $\tau = 20\%$ appliqué aux données du tableau 1 :

N° de sinistre	Montant	Taux de cession	Part cédante	Part réassureur
1	10	0.2	8	2
2	20	0.2	16	4
3	30	0.2	24	6

TABLE 1 – Exemple d'un quote-part

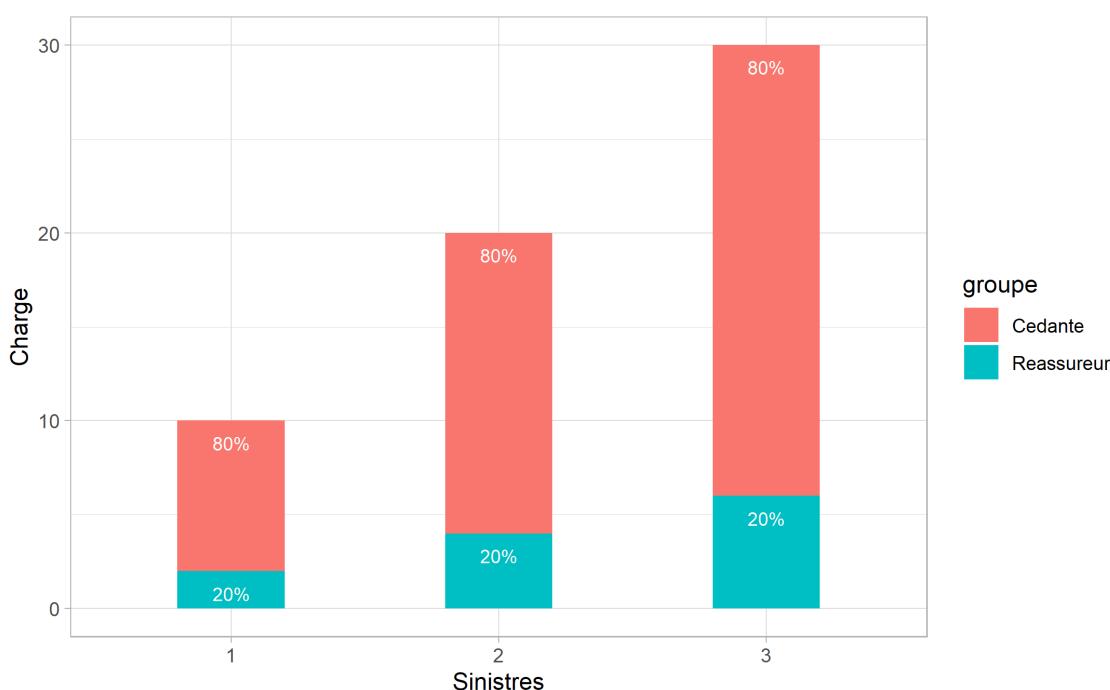


FIGURE 1 – Illustration de la répartition cédante/réassureur dans le cadre d'un traité QP

Graphiquement, la proportionnalité du traité est observable (figure 1).

Un tel transfert de risque a pour conséquence de transférer au réassureur un portefeuille de qualité similaire à celui souscrit par la cédante. Néanmoins, les traités en quote-part entraînent une forte cession du volume de primes. Ces traités sont souvent utilisés par les jeunes assureurs, ou les compagnies s'attaquant à un nouveau marché : ils n'ont pas encore une connaissance suffisante du marché et de sa sinistralité, et souhaitent ainsi se couvrir plus largement.

Les traités en excédent de plein

Un traité en excédent de plein fonctionne de manière similaire à un traité quote-part, à la seule différence que le taux de cession n'est pas fixe : il dépend de la somme assurée. La cédante fixe une limite, appelée plein, et la part au-dessus de cette limite sera cédée au réassureur. Pour chaque risque, le pourcentage de valeur assurée supérieur au plein est calculé : cela correspond au taux de cession du risque. Pour obtenir la prime, les taux de cessions sont appliqués aux primes encaissées pour chaque risque.

Le taux de cession se calcule de la manière suivante :

$$\tau_i = \max\left(0, \frac{C_i - P}{C_i}\right)$$

avec

- τ_i le taux de cession associé au risque i
- C_i le capital assuré associé au risque i
- P le plein

Considérons un traité en excédent de plein avec un plein $P = 100$ appliqué aux données du tableau 2 :

N° de sinistre	Somme assurée	Taux de cession	Montant	Part cédante	Part réassureur
1	130	0.23	100	77	23
2	220	0.55	160	73	87
3	70	0	20	20	0

TABLE 2 – Exemple d'un excédent de plein

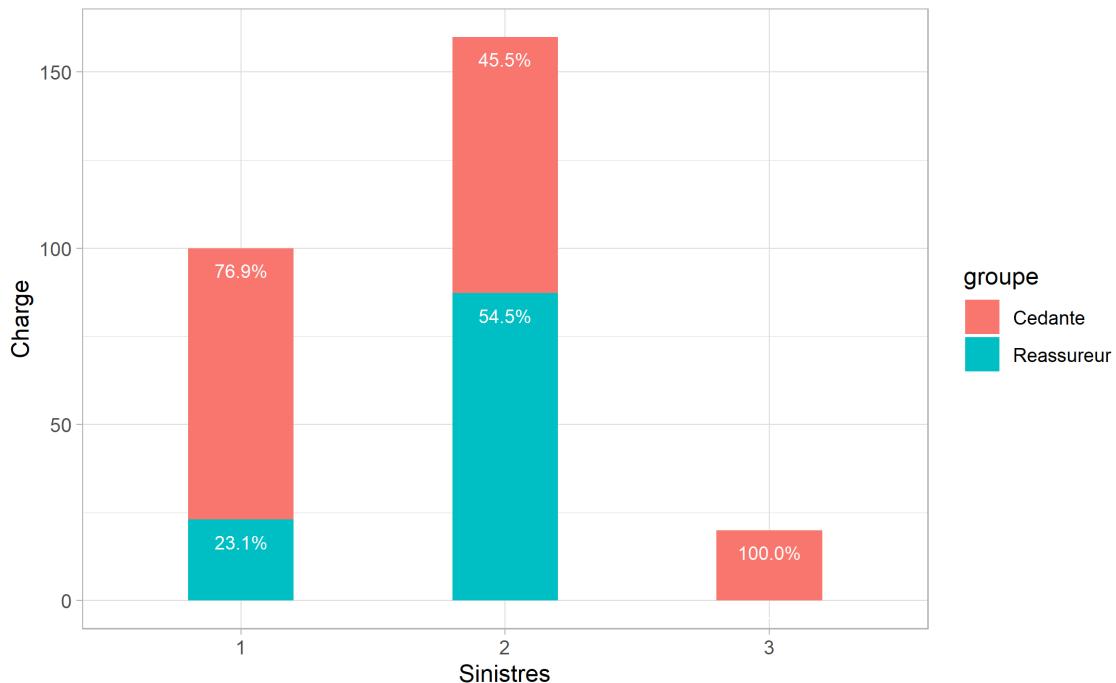


FIGURE 2 – Illustration de la répartition cédante/réassureur dans le cadre d'un traité QP

Le réassureur couvre ainsi uniquement les polices à risque (en termes de somme assurée) important, et plus les polices à faible somme assurée. Ainsi, les résultats sont plus volatils.

1.3.2 La réassurance non-proportionnelle

Les traités de réassurance non proportionnels sont appelés ainsi car la charge cédée au réassureur n'est pas proportionnelle à la charge brute des sinistres. La cédante va établir un seuil, appelé priorité, qui aura le rôle d'une franchise, à partir duquel le réassureur interviendra et prendra en charge les sinistres. Généralement, la cédante définit un deuxième seuil, appelé limite, au-delà duquel le réassureur ne prendra plus en charge la sinistralité : c'est la limite d'intervention du réassureur. Le réassureur aura donc à sa charge au maximum la portée du traité, définie comme la différence entre limite et priorité. Comme pour la réassurance proportionnelle, il existe différents types de traités non-proportionnels.

Les traités en excédent de sinistre

Les traités en excédent de sinistre (ou traités XS) sont les plus communément utilisés. Ils fonctionnent comme des contrats d'assurance avec franchise, la franchise étant appelée la priorité (D). A cette priorité vient généralement se rajouter un plafond, qui correspond au montant au-delà duquel le réassureur ne couvrira plus le risque. Ainsi, la part du sinistre comprise entre le plafond et la priorité est à la charge du réassureur. L'engagement maximum de celui-ci est donc la différence entre le plafond et la priorité : on l'appelle la portée (C). Ces traités sont généralement décrits de la manière suivante : C XS D.

Le montant à la charge du réassureur est alors :

$$S = \min(\max(X - D, 0), C)$$

avec X le montant du sinistre.

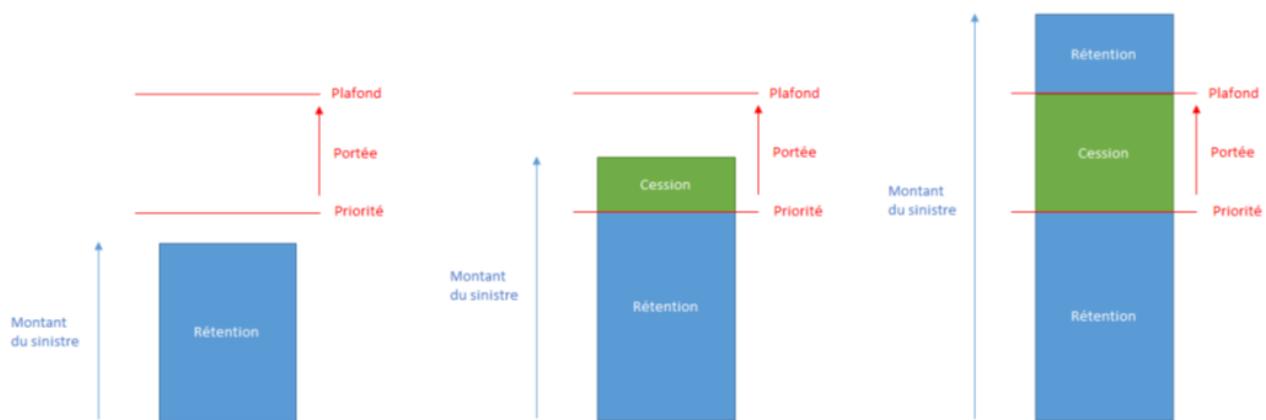


FIGURE 3 – Fonctionnement d'un traité XS

Le traité est appliqué individuellement à chaque sinistre déclaré par la cédante en fin d'exercice. Lorsque la cédante choisit de construire une structure de réassurance en utilisant des traités XS, elle définit plusieurs tranches – des traités successifs – et paye une prime pour chacune des tranches.

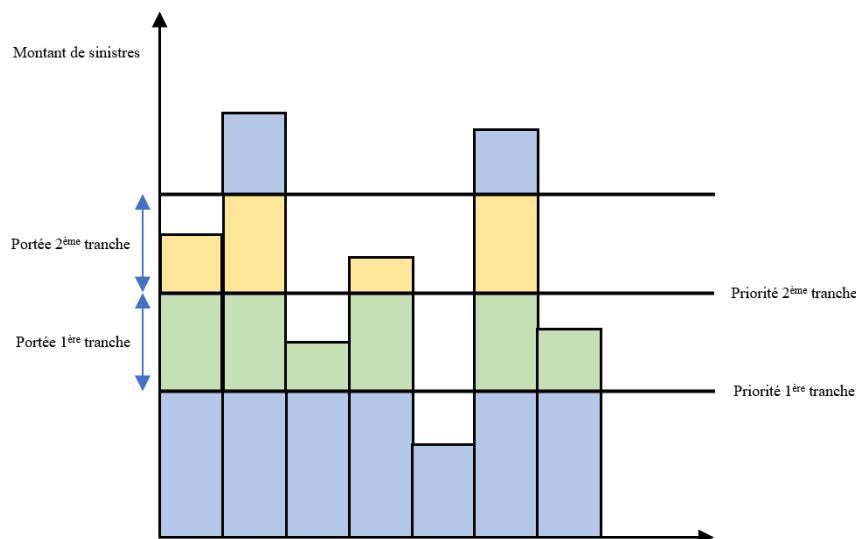


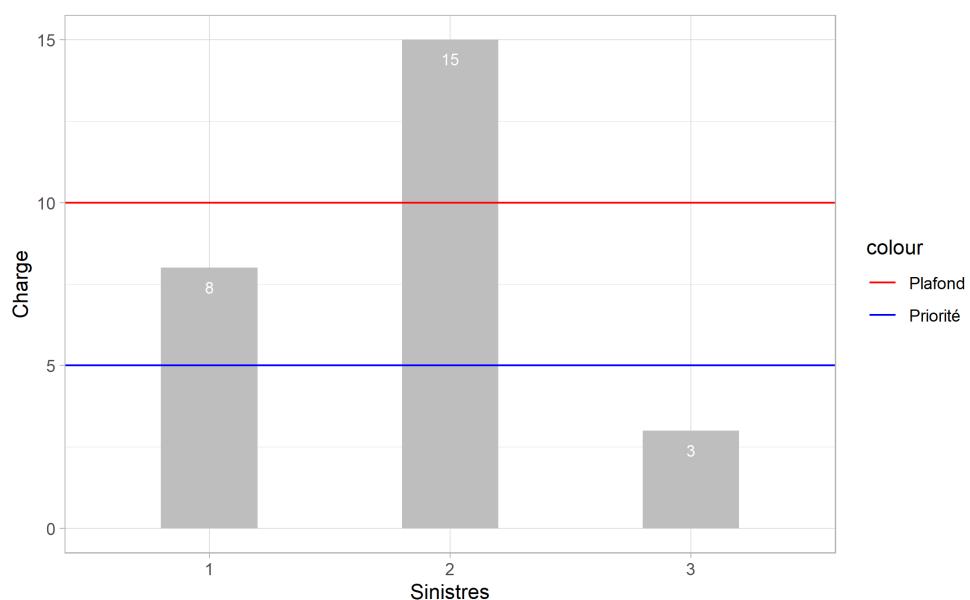
FIGURE 4 – Structure de réassurance en tranches d'XS

La prime est, contrairement à la réassurance proportionnelle, déterminée par le réassureur en fonction des statistiques de la cédante et du marché sur la branche considérée. Ces traités sont utilisés par les cédantes pour réduire leur exposition par risque, et acquérir une plus forte capacité de souscription.

Considérons un traité 5 XS 5 appliqué aux données suivantes :

N° de sinistre	Montant	Part cédante	Part réassureur
1	8	5	3
2	15	10	5
3	3	3	0

TABLE 3 – Exemple d'un excédent de plein

**FIGURE 5 – Traité 5 XS 5**

Pour le sinistre 1, on observe un dépassement de la priorité, mais pas du plafond. Après règlement de la franchise par la cédante, le réassureur prend en charge le reste. Pour le sinistre 2, le montant dépasse le plafond : le réassureur prend donc en charge la portée, tandis que la cédante réglera la priorité et le surplus au-dessus du plafond. Le sinistre 3 n'atteignant pas la priorité, il est intégralement à la charge de la cédante.

Notons que les traités XS peuvent être définis par risque – on applique alors le traité individuellement à chaque sinistre – , ou par évènement. Un évènement est un sinistre impactant plusieurs risques en même temps. Les évènements sont définis contractuellement dans les traités.

Les traités en excédent de perte

Les traités en excédent de perte (ou traités SL) fonctionnent de la même manière que les traités XS, mais ne sont plus appliqués sinistre par sinistre. Ils sont appliqués à la sinistralité agrégée du portefeuille de la cédante sur l'année. Ils peuvent être définis en pourcentage du Loss Ratio.

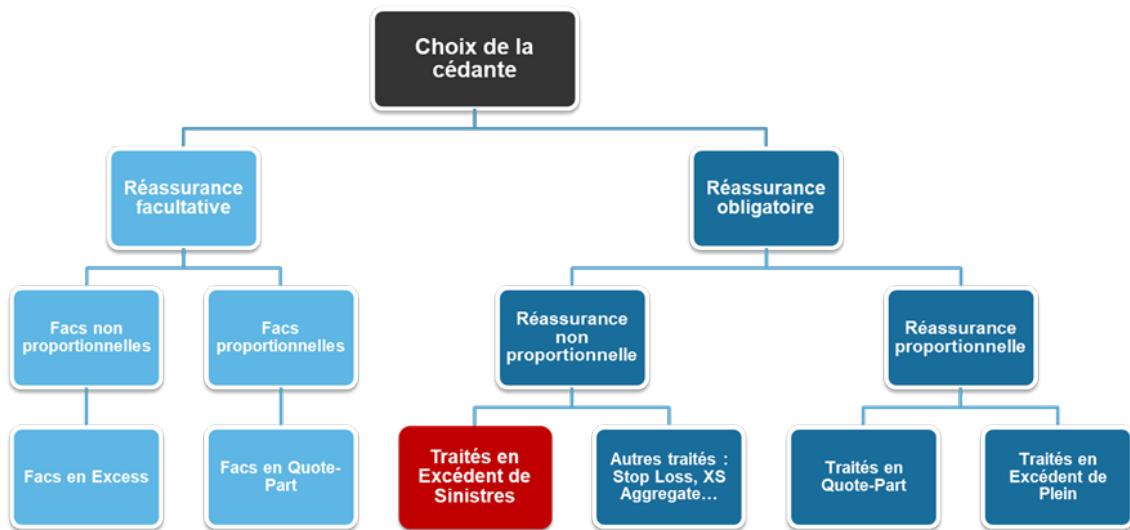


FIGURE 6 – Récapitulatif des différentes formes de réassurance

1.3.3 Les clauses

Ces traités sont les outils de base avec lesquels cédantes et réassureurs mettent en place des couvertures. Dans la pratique, ils ne sont jamais souscrits tels quels : ils présentent généralement des clauses. Il est important de les connaître et les comprendre, car elles peuvent modifier de manière non négligeable le résultat financier de l'opération de réassurance. Nous présenterons brièvement les clauses afférentes aux traités de réassurance non-proportionnelle.

L'Annual Aggregate Deductible - AAD

Cette clause fonctionne comme une franchise annuelle. Les premiers sinistres qui dépassent la priorité du traité seront à la charge de la cédante. Une fois l'entièreté de l'AAD consommée, le traité fonctionne normalement. Les cédantes utilisent cette clause pour diminuer le coût de la réassurance. La charge du réassureur devient alors :

$$S = \max(Charge Nette d'XS - AAD, 0)$$

L'Annual Aggregate Limit - AAL

C'est une limite annuelle d'engagement, qui permet au réassureur de limiter son engagement maximal au montant fixé par l'AAL. Cela a pour effet de diminuer le coût de la réassurance. La charge du réassureur devient alors :

$$S = \max(Charge Nette d'XS, AAL)$$

La reconstitution de garantie

Les réassureurs définissent des limites d'engagement par traité. Elles sont définies contractuellement, et une fois dépassées, le risque n'est plus couvert par le traité. Toutefois, le contrat peut

prévoir des reconstitutions de garantie. Ces reconstitutions, qui réinitialisent le traité, peuvent être gratuites ou payantes, limitées ou illimitées

Exemple d'application

On considère les données suivantes :

N° de sinistre	1	2	3	4	5
Montant	8	15	3	10	7

TABLE 4 – Données de l'exemple

Appliquons un traité 5 XS 5, avec AAD=4 et AAL=10.

N° de sinistre	Montant	Charge de l'XS avec AAD	Charge de l'XS avec AAL	Charge de l'XS avec AAD+AAL
1	8	3	0	3
2	15	5	4	5
3	3	0	0	0
4	10	5	5	2
5	7	2	2	0
Total	43	15	11	10

TABLE 5 – Application de l'XS avec clauses

Les clauses du traité ont bien pour effet de réduire la charge du réassureur. Le coût du traité en sera donc amoindri.

Les traités et clauses présentés sont les outils principaux avec lesquelles assureurs et réassureurs mettent en place différentes structures de réassurance. Ensuite, ces couvertures doivent être tarifées. Le chapitre suivant présente les méthodes utilisées pour côter des traités en excédent de sinistre.

CHAPITRE 2

Méthodes de tarification d'un traité XS

Comme expliqué en 1.3.2, les traités XS fonctionnent en Portée et Priorité. Les cédantes choisissant d'avoir recours à des traités XS pour couvrir un portefeuille de risques définissent alors une certaine structure, un empilement de traités dont la priorité sera de plus en plus haute. Le tarif associé à chaque tranche sera alors différent.

2.1 La notion de tranches

Pour calculer un tarif, il faut différencier trois types de tranche :

- Les tranches basses, dites travaillantes;
- Les tranches moyennes, dites semi-travaillantes;
- Les tranches hautes, dites non-travaillantes.

Une tranche est considérée comme travaillante si elle est fortement sinistrée. Dès lors, les données sinistres associées à ces tranches seront nombreuses. Une tranche est considérée comme semi-travaillante si elle est régulièrement sinistrée, mais rarement transpercée (une tranche est transpercée quand le plafond est atteint). Les données sinistres de ces tranches seront donc moins robustes. Enfin, une tranche est dite non-travaillante lorsqu'elle n'est généralement pas atteinte. Le manque de données pousse donc le réassureur à employer des méthodes de tarification sur exposition.

Plus on monte dans les tranches, plus les données se font rares, et moins les méthodes basées sur l'historique de sinistralité de la cédante sont robustes. Ainsi, en fonction de la tranche, les méthodes de cotation employées seront différentes :

- Les tranches travaillantes seront tarifées à partir de méthodes déterministes ou probabilistes basées sur l'expérience;

- Les tranches semi-travaillantes seront tarifées à partir de méthodes probabilistes ;
- Les tranches non travaillantes seront dans l'idéal tarifées à partir de modèles sur exposition.

2.2 Tarification à l'expérience

Les méthodes de tarification à l'expérience se basent uniquement sur l'historique de sinistralité de la cédante. L'idée derrière ces méthodes est de considérer que la sinistralité passée avec des ajustements peut être représentative de la sinistralité future. Par construction, il faut disposer d'un nombre suffisamment important de données pour appliquer la méthode. Avec ce type d'approche, les bases de données utilisées doivent être retraitées afin de représenter au mieux la sinistralité passé de la cédante. La figure 7 schématise le procédé commun à toutes les méthodes par expérience.

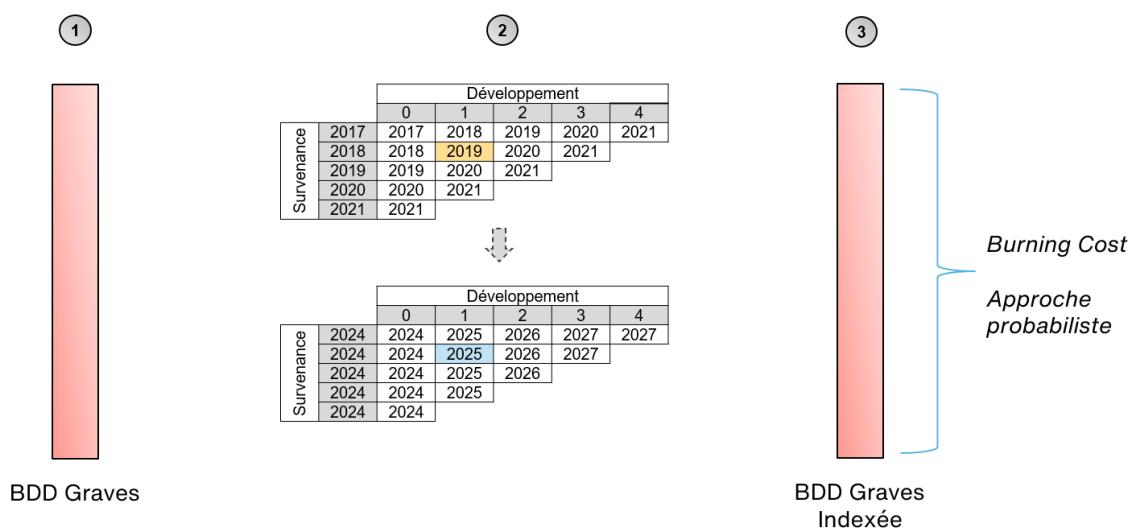


FIGURE 7 – Ajustement commun à toutes les méthodes par expérience

Le mécanisme d'indexation ou de mise en *as-if* est détaillé ci-après.

2.2.1 Mise en *as-if*

Avant application de n'importe quelle méthode sur expérience, les données doivent être mises en *as-if*. La mise en *as-if* répond à la question suivante : combien coûterait un sinistre survenu dans le passé s'il survenait durant l'année de référence (on prend généralement l'année de cotation) ? Par exemple, pour l'année de cotation n , un sinistre survenu en $k < n$ mis en *as if* est vu en année de survenance n . Ce travail préliminaire sur les données est primordiale pour l'homogénéisation de la base. En effet, un sinistre de 1M€ survenu en 2005 n'est pas comparable à un sinistre de 1M€ survenu en 2024. Or, en réassurance, le volume de donnée étant par nature faible, les sinistres utilisés peuvent remonter à des dizaines d'années. L'inflation a un impact important sur le coût des sinistres. Il faut donc absolument revaloriser tous les sinistres pour qu'ils soient vus sur une même base monétaire. La prise en compte de l'inflation peut se faire par l'intermédiaire d'un indice de référence qui représente l'évolution dans le temps du coût des sinistres. En incendie, l'indice

l'indice du coût de la construction fourni par la Fédération Française du Bâtiment (indice FFB) est généralement utilisé.

En pratique, la méthode d'indexation des coûts des sinistres est présentée ci-après. Soient

- n l'année de cotation;
- i l'année de survenance;
- j le décalage ($i + j = \text{arrêté}$);
- $X_{i,j}$ un flux de sinistre survenu en i , au décalage j ;
- $X_{i,j}^n$ un flux de sinistre survenu en i vu en n , au décalage j .

Alors

$$X_{i,j}^n = X_{i,j} \frac{FFB_{n+j}}{FFB_{i+j}}$$

Tous les flux de sinistre sont indexés de cette manière, afin d'être vus en année de cotation. Tous les arrêtés comptables sont translatés dans le temps, des hypothèses de projection de l'indice dans le futur sont donc utilisées. Sur la figure 8, on observe à gauche un triangle non indexé, et à droite le même triangle indexé avec $n = 2024$ l'année de cotation.

		Développement				
		0	1	2	3	4
Survenance	2017	2017	2018	2019	2020	2021
	2018	2018	2019	2020	2021	
	2019	2019	2020	2021		
	2020	2020	2021			
	2021	2021				

		Développement				
		0	1	2	3	4
Survenance	2024	2024	2025	2026	2027	2027
	2024	2024	2025	2026	2027	
	2024	2024	2025	2026		
	2024	2024	2025			
	2024	2024				

FIGURE 8 – Mise en As-If

Après indexation, tous les sinistres sont bien vus en année de survenance 2024. Différentes méthodes statistiques peuvent alors être employées sur la base ainsi retraitée afin de coter des traités, ou estimer la distribution de la sinistralité brute au-delà du seuil de modélisation fixé.

2.2.2 La méthode du *Burning Cost*

Le *Burning Cost* est une méthode déterministe sur expérience. L'idée est de déterminer le taux de primes à verser au réassureur pour couvrir une tranche donnée sur l'année à venir.

Pour coter un traité XS à l'aide de la méthode du *Burning Cost*, il faut :

- Indexer les montants de primes encaissées pour avoir un référentiel commun. Il faut prendre en compte l'inflation, et l'évolution de la politique de souscription de la cédante au cours du temps;

- Indexer les montants de sinistre. Dans le cadre de l'Incendie, l'indice du coût de construction FFB est utilisé;
- Déterminer le montant total à la charge du réassureur pour la tranche que l'on souhaite tarifer;
- Déterminer le *Burning Cost*.

Soient :

- X_i la charge indexée associée au sinistre i ;
- P_i le montant de prime encaissée associée au sinistre i ;

Pour chaque sinistre i , le montant à la charge du réassureur dans le cadre du traité XS S_i est calculé à l'aide de la formule suivante :

$$S_i = \min(\max(X_i - D, 0), C)$$

Le *Burning Cost* est alors défini comme le rapport entre la sinistralité totale à la charge du réassureur et l'assiette de primes encaissée par la cédante pour les risques associés :

$$\hat{BC} = \frac{\sum_{i=1}^n S_i}{\sum_{i=1}^n P_i}$$

Traitons un exemple. On souhaite coter un traité 5 XS 5 en utilisant le *Burning Cost*. Nous disposons des données suivantes :

N° de sinistre	Charge indexée	Primes encaissées	Charge réassureur
1	8	10	3
2	3	12	0
3	15	20	5
4	4	12	0
5	6	15	1
Total	36	69	9

TABLE 6 – Données sinistres pour le calcul du *Burning Cost*

En appliquant la formule, on trouve le *Burning Cost* suivant :

$$\widehat{BC} = \frac{9}{69} = 13\%$$

La cédante reversera alors 13% des primes encaissées associées aux risques couverts par le traité 5 XS 5.

On remarque facilement les limites de cette méthode. Si on avait voulu coter un traité 10 XS 15, on aurait trouvé un *Burning Cost* de 0%. Or, bien que les 15 n'aient jamais été dépassés par le passé, le risque n'est pas pour autant inexistant. De plus, le *Burning Cost* permet uniquement une tarification des traités. C'est une méthode déterministe qui ne permet pas de simuler la sinistralité grave individuelle des cédantes et d'extrapoler les données.

2.2.3 La méthode probabiliste

L'approche probabiliste repose sur la théorie des modèles collectifs, qui permet l'étude indépendante de la fréquence et de la sévérité des sinistres. Cela offre une certaine flexibilité au niveau de la modélisation et simulation des structures de réassurance. On pourra ajuster différents paramètres impactant l'un ou l'autre des composantes du modèle.

Pour rappel, les modèles fréquence/coût se reposant sur la théorie des modèles collectifs définissent la sinistralité comme :

$$Z = \sum_{i=1}^T X_i$$

où

- les X_i sont i.i.d de loi X
- X est la variable aléatoire représentant la sévérité d'un sinistre
- T est la variable aléatoire représentant le nombre de sinistres
- $T \perp\!\!\!\perp X_i$

En réassurance non-proportionnelle par risque, les sinistres modélisés sont supérieurs à un seuil. Ce seuil sépare la sinistralité attritionnelle de forte fréquence et de faible intensité et la sinistralité des graves de faible fréquence mais forte intensité (et donc couverts par des traités en excédent de sinistre). Dans notre cas, le modèle collectif devient :

$$Z^{seuil} = \sum_{i=1}^{T^{seuil}} X_i^{seuil}$$

Les données communiquées par la cédante seront utilisées, après retraitement, pour calibrer une loi de fréquence et une loi de sévérité au-delà du seuil de modélisation défini. Ensuite, un nombre important de simulations stochastiques permettra d'estimer la distribution de sinistralité brute au-delà du seuil de modélisation fixé. Nous travaillons avec des données sinistres communiquées au-delà d'un seuil.

Modélisation de la loi de fréquence

Les traités XS couvrent en général une unique année. Les données sinistres transmises par la cédante sont utilisées pour calibrer une loi de fréquence sur une période d'un an. La fréquence est

modélisée au-delà d'un seuil de modélisation qui est suffisamment bas pour contenir un nombre suffisant de sinistre mais suffisamment haut pour être proche de la première tranche qu'on souhaite coter et donc comportant des sinistres dont la sévérité est représentative de la sinistralité entrante dans la tranche.

Pour chaque année de survenance de la base, le nombre de sinistres supérieurs au seuil de modélisation compté. L'échantillon obtenu est alors utilisé pour calibrer la loi. On estime les paramètres associés à plusieurs lois de fréquence et on choisit la loi qui correspond au mieux à nos données.

Dans le tableau ci-dessous les lois usuelles fréquemment utilisées sont présentées.

Loi	Paramètres	Support	Espérance	Variance
Poisson - $\mathcal{P}(\lambda)$	$\lambda \in \mathbb{R}_+^*$	\mathbb{N}	λ	λ
Binomiale - $\mathcal{B}(n, p)$	$n \in \mathbb{N}, p \in [0, 1]$	$\{0, \dots, n\}$	np	$np(1-p)$
Binomiale négative - $\mathcal{BN}(n, p)$	$n \in \mathbb{N}^*, p \in]0, 1]$	\mathbb{N}	$\frac{n(1-p)}{p}$	$\frac{n(1-p)}{p^2}$

TABLE 7 – Lois utilisées pour modéliser la fréquence dans la méthode AON

Le ratio espérance/variance observée sur les données peut être utilisé pour choisir la distribution de fréquence.

Modélisation de la loi de sévérité

Pour modéliser la sévérité des sinistres, diverses lois continues usuelles sont ajustées par maximum de vraisemblance à l'échantillon de données retraitées. Tous les sinistres de l'échantillon supérieurs au seuil de modélisation sont utilisés pour calibrer la loi.

Le tableau 8 présente les différentes lois utilisées usuellement pour modéliser la sévérité des sinistres graves :

Loi	Paramètres	Support	Espérance	Variance
Log-normale - $\mathcal{LN}(\mu, \sigma)$	$\mu \in \mathbb{R}, \sigma > 0$	\mathbb{R}_+^*	$e^{\mu + \frac{\sigma^2}{2}}$	$(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$
Log-logistique - $\mathcal{LL}(\alpha, \beta)$	$\alpha > 0, \beta > 0$	\mathbb{R}_+	$\alpha \frac{\pi/\beta}{\sin(\pi/\beta)}$	$\alpha^2 \frac{2\pi/\beta}{\sin(2\pi/\beta)}$
Pareto - $\mathcal{P}(\lambda, k)$	$\lambda > 0, k > 0$	$[\lambda; \infty[$	$\frac{k\lambda}{k-1}$	$\frac{\lambda^2 k}{(k-1)^2(k-2)}$ pour $k > 2$
Weibull - $\mathcal{W}(\lambda, k)$	$\lambda > 0, k > 0$	\mathbb{R}_+	$\lambda \Gamma(1 + \frac{1}{k})$	$\lambda^2 \Gamma(1 + \frac{2}{k}) - \mu^2$

TABLE 8 – Lois utilisées usuellement pour modéliser la sévérité des graves

Simulations stochastiques

Les structures de réassurance en XS comprennent généralement de nombreuses clauses (Reconstitutions, AAD, AAL, ...) : il devient alors difficile d'obtenir une formule analytique pour l'espérance de la charge du traité. Les actuaires ont donc recours à des simulations stochastiques qui permettront d'estimer cette quantité pour une année donnée. La méthode consiste à réaliser un grand nombre

de scénario (n simulations) pour une même année de couverture. Pour chacun des scénarios, on réalise les étapes suivantes :

1. N sinistres sont générés aléatoirement à l'aide de la loi de fréquence calibrée en amont sur les données ;
2. Pour chaque sinistre ainsi généré, un coût est simulé en utilisant la loi de sévérité ;
3. La réassurance est pleinement appliquée à l'échantillon généré pour calculer la charge du réassureur ;
4. On réitère les trois premières étapes n fois.

Cette méthode est très utilisée car elle a l'avantage de générer de nombreux indicateurs statistiques et financiers à l'issue des simulations. Les récupérations moyenne, l'écart-type, les périodes de retour, les S/P bruts et nettes de réassurance, etc... sont calculés et permettent de déterminer la prime pure de la structure, mais aussi l'efficacité de celle-ci. Aussi, la modélisation du brut permet de mesurer, estimer la plupart des structures de réassurance qu'une cédante voudraient mettre en place sans être limité par un certain type de traité. Les reconstitutions, AAD, AAL, traités structurés, etc... peuvent ainsi facilement être pris en compte. Ces simulations permettent aux cédantes d'appréhender aux mieux leurs risques, et d'optimiser leur couverture de réassurance. Cependant, la qualité des résultats obtenus dépend intégralement de l'étape précédente de calibration des lois de fréquence et de sévérité qui est basée sur l'expérience. La faiblesse de la donnée de sinistralité impacte donc directement la qualité des résultats obtenus. Il faut dans ce cas recourir à d'autres méthodes en complément qui sont basées sur des données externes.

2.3 Tarification sur exposition en incendie

La tarification sur exposition se base, comme son nom l'indique, sur l'exposition de la cédante sur la période couverte par le traité. Elle consiste à traduire cette exposition en sinistralité probable à charge de la réassurance afin d'obtenir un tarif pour celle-ci. Le lien entre exposition et sinistralité à charge doit être construit à partir de données externes à la cédante, par exemple des données marché de la sinistralité. La figure 9 schématise l'utilisation d'une méthode sur exposition.

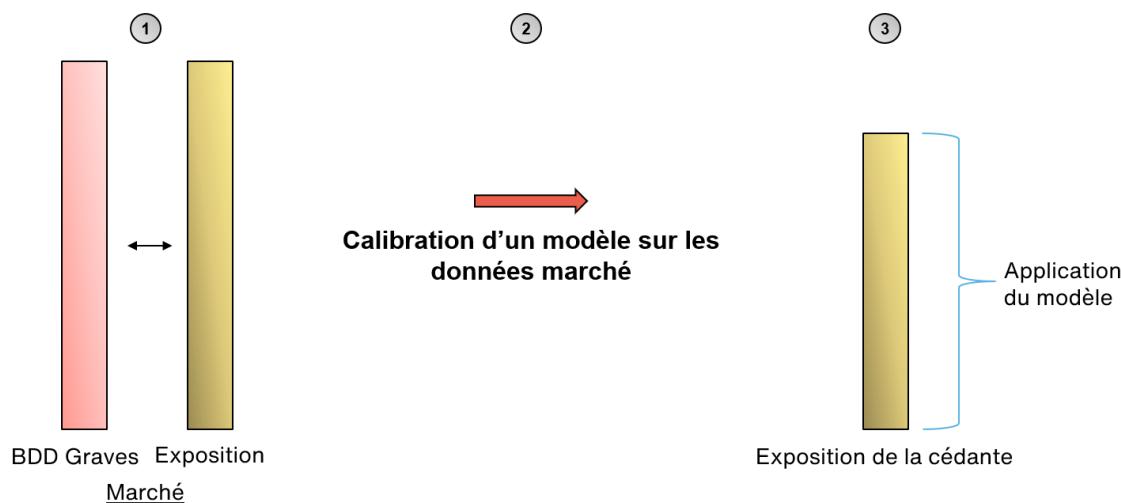


FIGURE 9 – Fonctionnement d'une méthode sur exposition

Les méthodes sur exposition sont généralement utilisées pour tarifer des structures pour un nouveau portefeuille ne disposant pas d'historique de sinistralité, ou dans le cas de traités en excédents de sinistres avec des priorités très hautes.

En incendie, la caractérisation de l'exposition peut se faire de manière synthétique à l'aide d'un profil de risque :

Profil de risque

Le profil de risque d'une cédante, transmis au réassureur dans le cadre des renouvellements, est la description du portefeuille de la cédante sous forme d'un tableau segmenté en bande de sommes assurées. Chaque ligne contient la prime, le nombre de risques et la somme assurée totale. Aussi, les bornes inférieure et supérieure de la bande sont communiquées. Ci-dessous l'illustration d'un profil de risque.

Borne Inférieure	Borne Supérieure	Prime totale	Nombre de risques	Somme assurée totale
LB_1	UB_1	P_1	N_1	TSI_1
$LB_2 = UB_1$	UB_2	P_2	N_2	TSI_2
...
$LB_T = UB_{T-1}$	UB_T	P_T	N_T	TSI_T

TABLE 9 – Exemple d'un profil de risques à T bandes

A partir du profil, nous pouvons déterminer pour chaque bande t la prime et les sommes assurées moyennes.

$$\text{Le taux de prime, } PM_t = \frac{P_t}{TSI_t}$$

$$\text{La somme assurée moyenne, } ASI_t = \frac{TSI_t}{N_t}$$

Si la somme assurée totale par bande ou le nombre de risques ne sont pas disponibles, on définira alors la somme assurée moyenne comme la moyenne entre la borne inférieure et la borne supérieure de la bande, ie :

$$ASI_t = \frac{LB_t + UB_t}{2}$$

Dans la suite, on émet l'hypothèse que chaque risques d'une bande ont même somme assurée (la moyenne).

Une application traditionnelle de la méthode de tarification sur exposition en incendie peut se faire via l'utilisation de courbes d'exposition, généralisées par Stefan BERNEGGER ([1], 1997) dans les années 90. Elles permettent de répartir les primes entre l'assureur et le réassureur, sans dépendre des fluctuations statistiques de la sinistralité.

Dans la suite, nous présenterons le concept général des courbes d'exposition, et nous nous intéresserons plus particulièrement aux courbes MBBEFD introduites par BERNEGGER dans [1].

Courbes d'exposition

Les courbes d'exposition sont des courbes estimant la part de prime nécessaire à la couverture d'un risque en fonction du degré de dommage couvert. En voici un exemple :

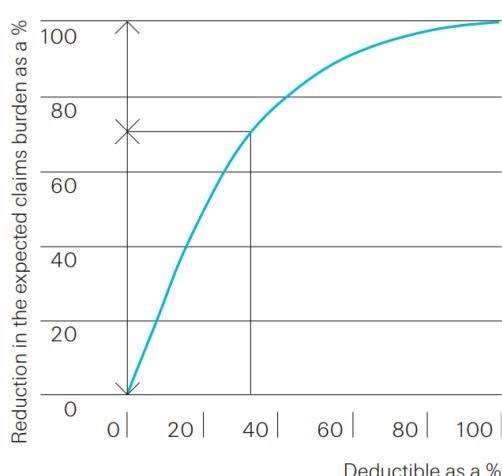


FIGURE 10 – Exemple d'une courbe d'exposition

Supposons que cette courbe d'exposition soit associée à une bande d'un profil de risques transmis par une cé-dante. En abscisse nous retrouvons une franchise normalisée par la somme assurée moyenne de la bande. En ordonnée nous avons la part de prime à conserver en fonction du niveau de franchise. Considérons une tranche Ill XS A; la priorité normalisée du traité est alors définie comme $a = \min(1, \frac{A}{ASI})$. La courbe peut se lire de la manière suivante : l'assureur a besoin de 70% des primes de la bande pour couvrir sa rétention (limitée à A par sinistre); la part de prime nécessaire au réassureur pour couvrir son engagement dans le cadre de l'ILL XS A est alors de 30% des primes de la bande.

Formalisme mathématique des courbes d'exposition

Différentes variables doivent être introduites. Dans cette partie, on considère :

$$\begin{cases} x = \frac{X}{SI} \in [0; 1] \text{ le taux de dommage d'un sinistre} \\ m = \min\left(\frac{M}{SI}, 1\right) \in [0; 1] \text{ la rétention de la cé-dante exprimée en pourcentage de la somme assurée} \end{cases}$$

avec

- X une variable aléatoire représentant le montant de sinistre,
- SI la somme assurée du bien correspondant
- M la priorité du traité considérée

La fonction d'espérance du taux de dommage d'un sinistre limité à la priorité normalisée m est introduite et définit comme :

$$L(m) = \mathbb{E} [\min(m, x)]$$

A partir de cette fonction on peut exprimer l'espérance de rétention de la cédante et la part à charge du réassureur :

$$\mathbb{E}[X_{rét}] = SI \times L(m)$$

$$\mathbb{E}[X_{réc}] = SI \times (L(1) - L(m))$$

La courbe d'exposition G est alors défini comme $G(m) = \frac{L(m)}{L(1)}$. Il s'agit de la part de prime nécessaire à la cédante pour couvrir les sinistres d'un montant limité à M , et on a :

$$G(m) = \frac{L(m)}{L(1)} = \frac{\int_0^m (1 - F(x)) dx}{\int_0^1 (1 - F(x)) dx} = \frac{\int_0^m (1 - F(x)) dx}{\mathbb{E}[x]}$$

où $F(\cdot)$ est la fonction de répartition du taux de dommages x .

La distribution des taux de dommages peut alors être exprimée en fonction de la dérivée de la courbe d'exposition. En effet, on a :

$$G'(m) = \frac{1 - F(m)}{\mathbb{E}[x]}$$

ce qui nous donne avec $F(0) = 0$ et $G'(0) = 1/\mathbb{E}[x]$

$$F(x) = \begin{cases} 1 & \text{si } x = 1 \\ 1 - \frac{G'(x)}{G'(0)} & \text{si } 0 \leq x < 1 \end{cases}$$

Le cas des courbes MBBED et leur lien avec les courbes Swiss RE

Dans son article de 1997, BERNEGGER [1] introduit les distributions MBBED (Maxwell-Boltzmann, Bose-Einstein, Fermi-Dirac) pour modéliser les taux de dommages. Il va montrer que cette famille de distribution est pertinente pour modéliser la distribution des taux de dommages dans l'intervalle $[0, 1]$. Malgré l'ancienneté de l'article, cette approche est toujours utilisée par certains acteurs du

marché.

Il introduit alors des courbes d'exposition analytiques à 2 paramètres b et g de la forme :

$$G_{b,g}(m) = \begin{cases} m & \text{si } g = 1 \vee b = 0 \\ \frac{\ln(1+(g-1)d)}{\ln(g)} & \text{si } b = 1 \wedge g > 1 \\ \frac{1-b^d}{1-b} & \text{si } bg = 1 \wedge g > 1 \\ \frac{\ln\left(\frac{(g-1)b+(1-gb)b^d}{1-b}\right)}{\ln(gb)} & \text{si } b > 0 \wedge b \neq 1 \wedge bg \neq 1 \wedge g > 1 \end{cases}$$

avec

- $b \in [0, \infty[$
- $g \in [1, \infty[$

La fonction de répartition du taux de dommage x correspondant est la suivante :

$$F(x) = \begin{cases} 1 & \text{si } x = 1 \\ 0 & \text{si } x < 1 \wedge (g = 1 \vee b = 0) \\ 1 - \frac{1}{1+(g-1)x} & \text{si } x < 1 \wedge b = 1 \wedge g > 1 \\ 1 - b^x & \text{si } x < 1 \wedge bg = 1 \wedge g > 1 \\ 1 - \frac{1-b}{(g-1)b^{1-x}+(1-gb)} & \text{si } x < 1 \wedge b > 0 \wedge b \neq 1 \wedge bg \neq 1 \wedge g > 1 \end{cases}$$

En théorie, il est alors possible à partir des distributions MBBEFD de caractériser des courbes d'exposition à l'aide de différentes méthodes. On peut par exemple estimer les paramètres b et g des courbes à l'aide de la méthode des moments ou du maximum de vraisemblance sur un échantillon de données.

Dans son article, Bernegger introduit également des courbes d'exposition MBBEFD à un paramètre c telles que :

$$G_c(m) = G_{b_c, g_c}(m)$$

avec

- $b_c = e^{3,1-0,15(1+c)c}$
- $g_c = e^{(0,78+0,12c)c}$

Ce paramétrage permet notamment de retrouver les célèbres courbes d'exposition Swiss Re, introduites par Peter Gasser. Respectivement, les paramètres $c = 1.5, 2, 3, 4$ approchent très bien les courbes Swiss Re 1, 2, 3, 4 et le paramètre $c = 5$ correspond à la courbe Lloyd's. Ces courbes sont présentées dans la figure 11 ainsi que les distributions des taux de dommages correspondantes.

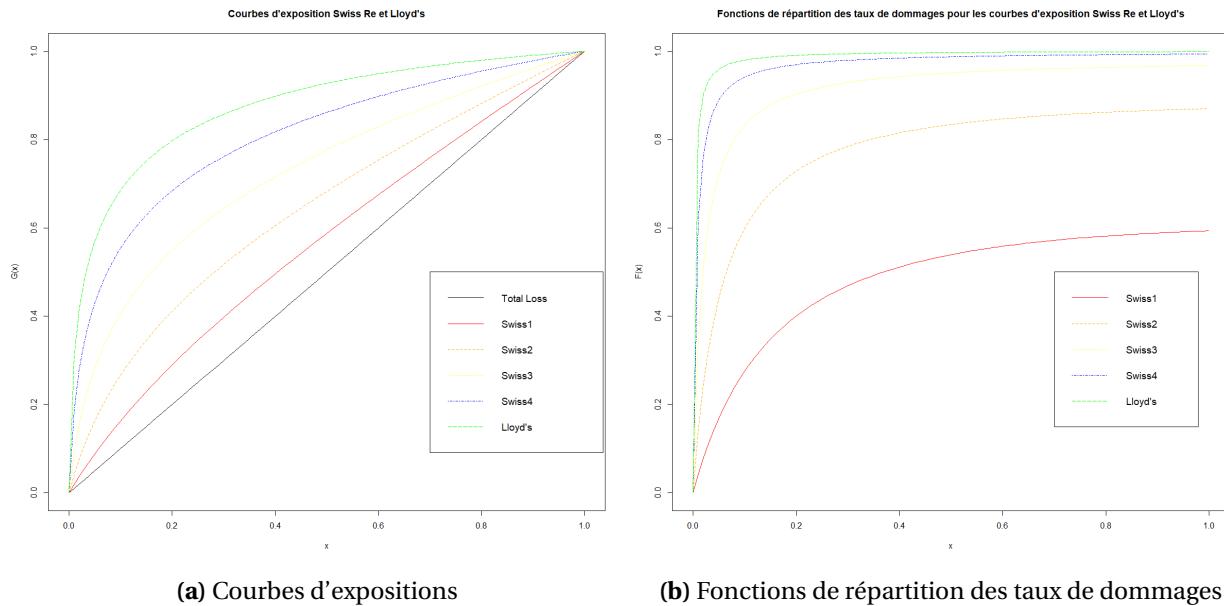


FIGURE 11 – Courbe d'exposition Swiss Re et Lloyd's

Plus les courbes d'exposition tendent vers la diagonale, plus on aura tendance à observer des taux de destruction proches de 1 (la diagonale étant le cas extrême où tous les sinistres auraient un taux de dommage égal à 1). Le choix de la courbe d'exposition utilisée pour les bandes du profil de risque devra représenter au mieux les risques sous-jacents. Si une courbe proche de la diagonale est choisie, cela signifie qu'on considère que les risques sous-jacents seront sujets à des taux de dommages assez élevés. A contrario, si la courbe choisie s'éloigne fortement de la diagonale (comme la courbe Lloyd's), alors la probabilité d'observer des taux de dommages élevés sera très faibles. Plusieurs critères sont pris en compte pour choisir le paramètre c de la courbe d'exposition à associer à une bande du profil de risque d'une cédante, les principales étant :

- la somme assurée moyenne de la bande,
- la typologie de risque (risque résidentiel, commercial ou industriel par exemple).

La question de l'inflation

En ce contexte de forte inflation, il est légitime de se poser la question de l'impact de l'inflation sur les différentes méthodes et modèles utilisées. Dans le cas des méthodes reposant sur l'utilisation de courbes d'exposition, les courbes sont théoriquement robustes face à l'inflation. En effet, les courbes d'exposition reposent sur une estimation de la distribution des taux de dommages en

fonction de la somme assurée. Comme nous réfléchissons en taux de dommages, et sous réserve d'émettre l'hypothèse que les sinistres évoluent de la même manière que les sommes assurées, les distributions de taux dommages estimées sur des données qui peuvent être anciennes restent d'actualité. C'est d'ailleurs le cas des courbes Swiss Re qui restent encore aujourd'hui utilisées par certains acteurs du marché. La relation modélisée par les courbes d'exposition étant relative (lien entre sévérité et somme assurée), les courbes ne sont pas forcément impactées par l'inflation.

Les courbes d'exposition permettent, dans leur utilisation la plus simple, de déterminer la part de prime d'assurance qui revient à la réassurance dans le cadre d'un traité en excédent de sinistre. Le profil de risques transmis par les cédantes est utilisé pour la tarification. La méthode sera présentée plus loin dans le mémoire et illustrée à l'aide d'un cas pratique. Aussi, ces courbes permettent à partir d'hypothèses fixées en amont de générer la sinistralité brute au-delà d'un seuil.

CHAPITRE 3

Le risque incendie : couverture en réassurance et modélisation utilisée en interne

3.1 Les traités Dommages Aux Biens

En réassurance non-proportionnelle, le risque Incendie est généralement couvert par des traités en excédent de sinistre dits Dommages aux Biens (DAB). En fait, ce sont les sinistres "Incendie et Risques annexes" qui sont couverts. Deux garanties sont donc à considérer. La garantie Incendie va concerter les dommages matériels causés par certains évènements. La garantie Risques Annexes va couvrir dans ce cas la responsabilité civile issues d'évènements Incendie.

La garantie incendie

La garantie incendie couvre notamment les dommages matériels causés par :

- Une explosion ou implosion,
- Une chute de foudre,
- Les effets du courant électrique,
- L'enfumage provenant d'un appareil défectueux ou d'un incendie au voisinage,
- Le choc par un appareil aérien ou spatial,
- Un incendie.

Les sinistres incendie sont susceptibles de causer la destruction totale du bien assuré : la réassurance est donc très utilisée pour couvrir ces risques. Cependant, c'est une garantie dommages, donc le montant du sinistre (limité à cette garantie) est borné par la valeur assurée du bien. Pour rappel, c'est cette observation qui permet d'utiliser la méthode de tarification se basant sur les courbes d'exposition.

La garantie risques annexes

La garantie risques annexes couvre la responsabilité des sinistres ayant pour origine l'une des causes couvertes par la garantie incendie. Notamment, les risques annexes couvrent :

- Le recours des voisins et des tiers,
- Les risques locatifs,
- etc...

Contrairement à la partie dommage des sinistres, le montant de la part RC des sinistres n'est pas limité par la valeur assurée mais peut être limité par un plafond contractuel.

Ces garanties vont donc de pair (les risques annexes couvrant les conséquences directes de l'incendie) mais n'exposent pas la cédante aux mêmes risques, l'une couvrant du dommage pur et l'autre la responsabilité des assurés. La figure 12 schématise ce qui vient d'être dit.

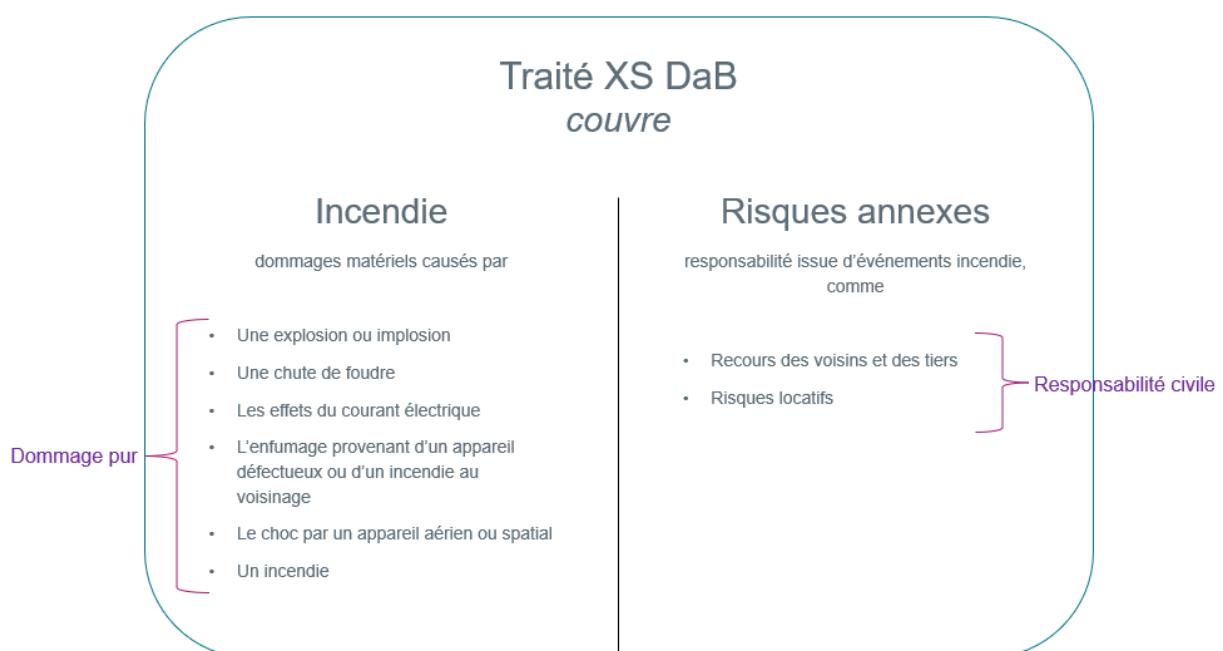


FIGURE 12 – Couverture XS DaB - Incendie et Risques annexes

Il est donc très important de garder en tête que les sinistres incendie communiqués par les cédante dans le cadre du renouvellement de leur traité DaB se composent à la fois de dommage pur de la police sinistré et de responsabilité. Ce point a des conséquences importantes en terme de modélisation du risque car la garantie Incendie pure devrait dépendre du bien et de la valeur assurées contrairement à la garantie Risque annexes.

3.2 Le rôle d'AON France

AON France, filiale française d'AON plc, est le leader français du courtage en réassurance et du conseil en gestion de capital des sociétés d'assurance. La société compte plus de 1000 collaborateurs, répartis principalement entre Paris, Marseille et Angoulême. Elle propose aussi une activité de courtage en assurance. Cette étude a été réalisé au sein du service réassurance, nous nous intéresserons donc uniquement à ce secteur d'activité.

3.2.1 Le courtage en réassurance

Lorsqu'une cédante souhaite mettre en place une structure de réassurance, deux options s'offrent à elle :

- Démarcher des réassureurs et mettre en place des structures de réassurance directement avec eux. Cette option a pour avantage de développer de véritables relations cédante/réassureur. Néanmoins, cela représente une forte charge de travail pour la cédante.
- Mandater un courtier en réassurance, comme AON, qui se chargera de mettre en place cette structure et servira de relais entre cédante et réassureur.

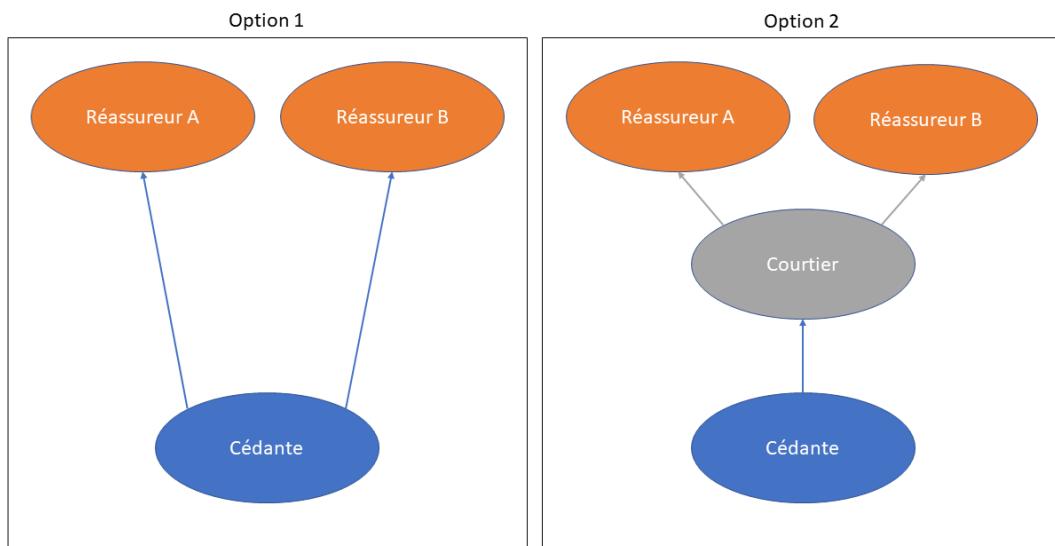


FIGURE 13 – Rôle d'un courtier en réassurance

Le rôle du courtier est donc de cerner les demandes et besoins du client quant à la couverture souhaitée, puis de le conseiller ,l'accompagner et le représenter tout au long de cette démarche : il a un véritable rôle d'exécution.

L'analyse technique revêt un caractère de plus en plus important dans les relations entre clients, courtiers et réassureurs. Cela amène les équipes actuariels et Cat modeler a occuper une place plus déterminante que dans le passé. Le courtier, à l'aide de ses équipes de techniciens, va alors

analyser le risque des clients afin de les conseiller et les accompagner au mieux dans la conception des structures de réassurance.

3.2.2 Le service Analytics

Le service Analytics d'AON Reinsurance Solutions répond à cette demande d'analyse des risques et des structures des cédantes. L'équipe est structurée en deux pôles : Actuariat et Cat Modeling. Elle accompagne les équipes commerciales pour l'ensemble des services délivrés, tels que :

- Analyse des données de portefeuille et de sinistralité
- Évaluation des expositions catastrophes (Tempête, Grêle, Inondations, Sécheresse, etc.) et modélisation sur la base de ces expositions ;
- Modélisation de la charge technique des traités de réassurance (Automobile, Incendie, Construction, Cyber, Crédit-Caution, Forces de la nature, CatNat, etc.) et mesure de l'efficacité des structures au regard de l'appétence au risque de la compagnie ;
- Études de marché Forces de la Nature et RC automobile ;
- Conseil relatif aux nouvelles règles de solvabilité, soutien technique pour les calculs de la formule standard, aide opérationnelle lors de la mise en place de la documentation de modèles internes ou de modèles d'évaluation interne des risques et de la solvabilité ("ORSA") dans le cadre de Solvabilité II ;
- Apport d'expertise technique et de services de modélisation dans le cadre de transactions ou de prestations de conseil pour les clients.

Ce mémoire, réalisé au sein du pôle actuariat, vise à améliorer les travaux de modélisation et de mesure de l'efficacité des structures de réassurance sur la branche Incendie. A l'heure actuelle, seule la méthode de modélisation sur expérience est implantée pour cette branche de risque.

La section suivante expose la méthodologie de modélisation employée actuellement sur l'incendie.

3.3 Méthode AON de modélisation de la sinistralité incendie

Dans cette section, la méthode de modélisation de la sinistralité utilisée par AON sera détaillée. Les principes présentés seront pour la plupart aussi employés lors de la calibration des modèles sur exposition. La méthode probabiliste vue précédemment est utilisée. C'est une méthode de modélisation sur expérience. La charge totale des sinistres est modélisée sans distinction de la part de dommage pure et de responsabilité civile. Les étapes de modélisation sont schématisées dans la figure 14 ci-dessous.

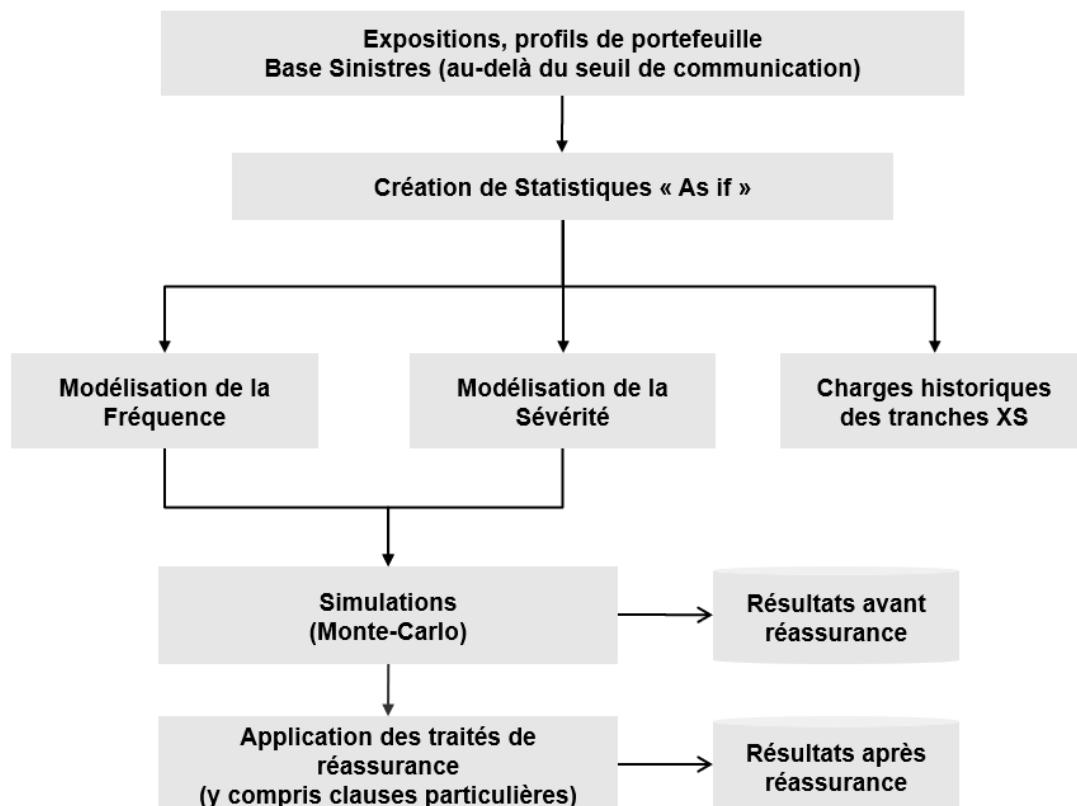


FIGURE 14 – Processus de modélisation AON

3.3.1 Les données

Dans un premier temps, la méthode étant basée sur l'historique de sinistralité de la cédante, les données de sinistralité historique doivent faire l'objet de différents ajustements pour tenir compte de l'évolution de l'exposition de la cédante en terme de coût et de fréquence. Le mécanisme d'indexation des sinistres a été détaillé en 2.2.1.

Le seuil de communication

Le seuil de communication est le seuil fixé par la cédante à partir duquel elle inclura ou non un sinistre dans la base transmise. Le seuil de communication des avis de sinistres est généralement fixé à 75% de la priorité du traité. Dans la pratique, la base de donnée de la sinistralité historique transmise lors du renouvellement des traités est communiquée à un seuil plus bas afin de s'assurer

que l'intégralité des sinistres atteignant la priorité une fois mise en as-if soit présent dans la base

Le seuil doit lui aussi être retraité avant modélisation. Étant le même pour toutes les années de survenance de la base, il doit être indexé afin qu'il soit vu sur une même base monétaire pour tous les sinistres.

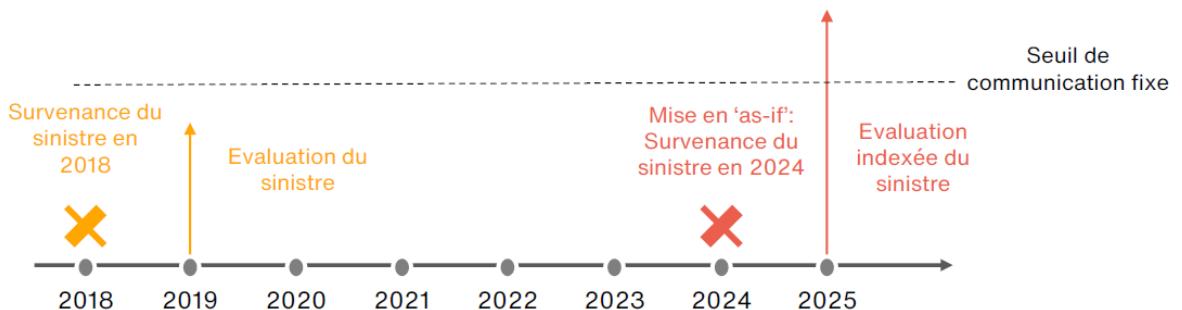


FIGURE 15 – Hétérogénéité de la base sans indexation du seuil de communication

Comme l'illustre la figure 15, si le seuil de communication n'est pas indexé, des problèmes d'hétérogénéité dans la base peuvent apparaître (une perte d'information).

Soient

- $n_{cotation}$ l'année de cotation;
- n_{min} l'année de survenance minimale de la base;
- SC le seuil de communication;
- SC^{ind} le seuil de communication indexé;

Alors

$$SC^{ind} = SC \frac{FFB_{n_{cotation}}}{FFB_{n_{min}}}$$

Le seuil de communication étant indexé avec comme année de référence la survenance minimale de la base, le problème présenté au paragraphe précédent est résolu.

Cette nouvelle valeur est comparée au seuil de modélisation SM choisi en amont, et le seuil de modélisation final SM' est défini comme :

$$SM' = \max(SC^{ind}, SM)$$

Les sinistres sont filtrés sur cette valeur pour déterminer s'ils feront partie ou non de la modélisation.

Projection à l'ultime

Les sinistres étant indexés et le seuil de modélisation défini, les sinistres sont projetés à l'ultime avant ajustement des lois de fréquence et sévérité. Il est très important d'estimer la charge ultime

des sinistres : bien que l'incendie soit une branche à développement court, des bonis et malis sont tout de même observés tout au long de la vie du sinistre.

Pour déterminer la charge ultime de chaque sinistre de la base, des coefficients de passage associés à chaque année de décalage sont calculés à l'aide de la méthode de Chain-Ladder présentée en annexe A. Les sinistres utilisés dans le triangle de développement sont ceux dont la charge indexée a déjà dépassé au moins une fois le seuil de modélisation défini précédemment, i.e :

$$\max_{k \in \{0, \dots, n-i\}} (X_{i,k}^n) \geq SM'$$

Soit $X_{i,ultime}^n$ la charge ultime d'un sinistre survenu en i indexé en n . On a alors :

$$\forall i + j > n, X_{i,ultime}^n = \hat{\lambda}_{n-i} \times \dots \times \hat{\lambda}_{j-1} \times X_{i,j}^n$$

avec

- i l'année de survenance
- n l'année de cotation
- j le dernier décalage connu

Les sinistres utilisés pour la modélisation de la sinistralité vérifient $X_{i,ultime}^n \geq SM'$

Il est important d'utiliser le seuil de modélisation pour filtrer les sinistres qui seront utilisés dans le calcul des facteurs de développement ; en effet, en fonction de la gravité des sinistres, les développements observés pourront être significativement différents. Dans le cas que nous étudions, c'est bien la cadence de développement des sinistres qui ont au moins une fois dépassé le seuil de modélisation que l'on souhaite estimer. Les sinistres projetés à l'ultime seront ensuite utilisés pour calibrer les lois de fréquence et de coût : la projection doit donc se faire avec beaucoup de prudence. L'impact du seuil utilisé sur les facteurs de développement obtenus est étudié dans le paragraphe ci-après.

Impact du seuil de modélisation dans l'estimation de la charge ultime des sinistres

Pour cette étude benchmark, les données de sinistralité incendie historique d'une cédante sont utilisées. Le volume de données est initialement assez important (1588 sinistres distincts, avec des survenances allant de 1984 à 2022). Le seuil de communication de la cédante est égal à 300k, et les montants de sinistre sont communiqués *from ground up* (FGU), c'est à dire que la charge brute du sinistre est connue.

La base a été transmise lors des travaux pour le renouvellement 2023, le dernier arrêté connu est donc le 2022 et l'année de cotation (pour l'indexation) a été fixé à 2023. L'étude est réduite à un horizon de 15 ans, à la fois pour utiliser des données suffisamment récentes pour être considérées comme représentatives de la sinistralité future, et pour assurer un volume de données suffisamment

important. L'indice FFB est utilisé pour indexer les sinistres et le seuil de communication. Après indexation, le seuil de modélisation minimum est défini à 450k.

Pour montrer l'impact du seuil de modélisation dans l'estimation du développement à l'ultime des sinistres avec la méthode de Chain-Ladder ainsi que la différence de cadence de développement moyenne des sinistres en fonction de leur gravité, des coefficients sont calculés à différents niveaux de seuil de modélisation à partir des données décrites ci-avant. La figure 16 présente les résultats obtenus (les facteurs de développement sont fixés à 1 à partir du onzième décalage).

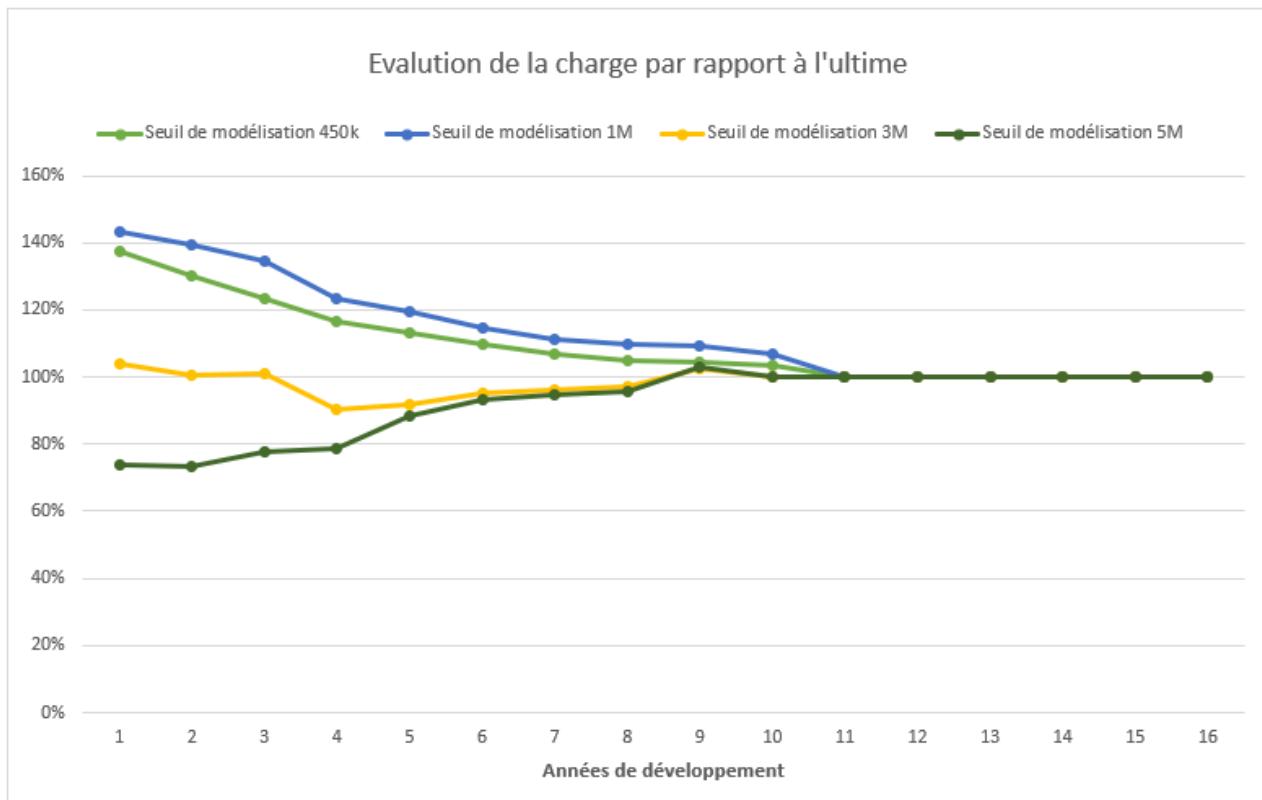


FIGURE 16 – Impact du seuil de modélisation sur les cadences de développements moyennes estimées

En observant le graphique, il apparaît assez clairement qu'en fonction du seuil de modélisation utilisé, le développement à l'ultime moyen n'est pas la même. Cette étude permet de montrer que les sinistres ne se développent pas de la même manière en fonction de leur gravité, donc le seuil de modélisation est à choisir avec beaucoup de soin et doit être cohérent avec la modélisation effectuée.

Par exemple, dans le cas de la cédante étudiée

- avec un seuil de modélisation fixé à 450k, on observe en moyenne des bonus jusqu'à la charge à l'ultime ;
- avec un seuil de modélisation fixé plus haut, à 5M, on observe cette fois plutôt une dégradation des sinistres jusqu'à l'ultime.

Les facteurs de développement calculés à partir du seuil 450k ne sont donc pas pertinents pour

projeter à l'ultime les sinistres dont la charge indexée maximale est supérieure au seuil de 5M. Des bonus seraient projetés alors qu'on voit bien que ces sinistres présentent une tendance à l'aggravation : il est primordial de filtrer les données utilisées pour calculer les facteurs de développement avec le seuil de modélisation choisi.

3.3.2 Calibration des lois de fréquence et sévérité

Une fois la base de donnée mise en as-if et projetée à l'ultime, un modèle fréquence-sévérité peut être calibré à partir de ces données.

3.3.2.1 Calibration de la loi fréquence

La fréquence est défini par deux composantes. L'ensemble des sinistres connus dépassant le seuil de modélisation à l'ultime en est la composante principale. La deuxième composante provient des déclarations tardives et des dépassement tardifs du seuil de communication

Fréquence des sinistres connus

Pour définir l'échantillon, le nombre de sinistres dont la charge ultime est supérieure au seuil de modélisation au dernier arrêté est observé pour chaque année de survenance, i.e :

$$\text{NUltimateSupSeuil}_i = \sum_{k=1}^N \mathbb{1}_{\{X_{i,ultime}^{(k)} \geq SM\}}$$

avec

- i l'année de survenance
- N le nombre de sinistre survenu en i
- $X_{i,ultime}^{(k)}$ la charge ultime du k-ème sinistre survenu en i

Ajustements pour tenir compte des effets de l'exposition

L'échantillon doit être traité afin de tenir compte de l'évolution de l'exposition du portefeuille. En effet, une fréquence de 2 sinistres supérieurs au seuil n'a pas la même signification lorsque le nombre de biens assurés est de 1M contre 2M. Pour supprimer les effets de l'exposition, le ratio de l'exposition de l'année de cotation n par rapport à l'exposition d'une année de survenance i est appliqué aux nombres de sinistres supérieurs au seuil par année.

$$\text{NCorrigé}_i = \text{NUltimateSupSeuil}_i \times \frac{E_n}{E_i}$$

Ensuite, pour déterminer le nombre final de sinistres supérieurs au seuil de modélisation, il faut modéliser les tardifs.

Les tardifs

Il existe deux types de tardifs : les déclarations tardives et les dépassements tardifs. Les dépassements tardifs sont plus fréquents.

Les déclarations tardives correspondent aux sinistres survenus en année n , mais déclarés au plus tôt en $n+1$. Ils sont à prendre en compte, et à estimer car si l'on veut coter un traité une année donnée, les déclarations tardives sont incluses dans la charge finale du traité. Les déclarations tardives peuvent être confondues avec des arrêtés manquants dans la base de données, il est donc important de s'assurer de l'origine de ces tardifs pour ne pas faire une mauvaise interprétation des données. Pour les déterminer, il suffit d'observer les sinistres dont la charge ultime est supérieure au seuil de modélisation, et dont l'année du premier arrêté présent dans la base est différente de l'année de survenance.

Les dépassements tardifs sont des sinistres dont la charge est inférieure au seuil de modélisation à la survenance, mais le dépasse au cours de sa vie. Pour les déterminer, on comptabilise un dépassement tardif pour une survenance et un décalage donnée lorsqu'un sinistre dépasse pour la première fois le seuil de modélisation.

Alors, le nombre final de tardifs par année de survenance et décalage est la somme des déclarations et dépassements tardifs aux années de survenance et décalages concernés. Ensuite, il faut retraitier les tardifs afin de prendre en compte les évolutions de l'exposition de la cédante et calculer des tardifs ajustés.

Ajustements pour tenir compte des effets de l'exposition sur les tardifs

Dans nos modélisations, on considère que le nombre de tardifs pouvant survenir pour une année de survenance dépend de sa maturité (son développement) et de son exposition. La calibration de ce modèle de tardif se fait en observant les données historiques. On calcule pour cela un taux par année de survenance d'apparition des tardifs (nombre de tardifs divisé par l'exposition) en fonction de la maturité :

$$\text{TauxTardifs}_{i,j} = \frac{\text{Tardifs}_{i,j}}{E_i}$$

Ensuite, pour chaque décalage, une moyenne de ces taux est calculée :

$$\text{TauxTardifsMoyen}_j = \sum_{k=i_{min}}^{i_{max}} \text{TauxTardifs}_{k,j} / (i_{max} - i_{min} + 1)$$

avec

- i l'année de survenance;
- j le décalage;
- E_i le nombre de biens couverts en année i .

Le taux obtenus sont des taux moyens sans effets d'exposition. Ils peuvent alors être utilisés pour estimer les tardifs par survenance et décalage.

Finalement, pour chaque décalage et année de survenance, le calcul du nombre de tardifs par survenance i et décalage j est donné par :

$$\text{NewTardifs}_{i,j} = \text{TauxTardifsMoyen}_j * E_i$$

et pour chaque année de survenance, le nombre final de tardifs est la somme des NewTardifs sur tous les décalages :

$$\text{NTardifsFinal}_i = \sum_{j=0}^{n-i} \text{NewTardifs}_{i,j}$$

Les tardifs sont ensuite sommés par année de survenance aux nombres de sinistres supérieurs au seuil de modélisation à l'ultime déterminés précédemment pour obtenir l'échantillon final :

$$\text{NFinal}_i = \text{NTardifsFinal}_i + \text{NCorrigé}_i$$

Calibration de la loi de fréquence

Un échantillon de la forme suivante est obtenu (les données sont illustratives) :

Survenance	Exposition	Tardifs	NUltimateSupSeuil	NCorrigé	NFinal
1	1M	0,6	5,2	6	6,6
...
n-1	3M	0,3	12	12,4	12,7

TABLE 10 – Échantillon de fréquence illustratif

Pour ajuster une loi de fréquence sur l'échantillon, des poids sont affectés à chaque année de survenance en fonction de l'exposition, et une moyenne et une variance pondérés par l'exposition sont calculées. Cela permet de donner plus de poids aux années fortement exposées.

Le poids associé à l'année de survenance i est :

$$\omega_i = \frac{E_i}{\sum_{k=\text{Surv}_{min}}^n E_k}$$

avec Surv_{min} l'année de survenance minimale définie.

Les estimateurs pondérés suivants sont obtenus :

$$\hat{\mu} = \sum_{i=\text{Surv}_{min}}^n \omega_i \times \text{NFinal}_i$$

$$\hat{\sigma}^2 = \frac{n - \text{Surv}_{min}}{n - \text{Surv}_{min} - 1} \times \sum_{i=\text{Surv}_{min}}^n \omega_i \times (\text{NFinal}_i - \hat{\mu})^2$$

Finalement, le ratio Variance/Espérance permet de choisir la loi modélisant au mieux l'échantillon. Les paramètres de celle-ci sont ensuite déterminés à l'aide de la méthode des moments.

$$\left\{ \begin{array}{l} \text{Binomiale si } \frac{\hat{\sigma}^2}{\hat{\mu}} < 0,9 \\ \text{Poisson si } 0,9 \leq \frac{\hat{\sigma}^2}{\hat{\mu}} \leq 1,1 \\ \text{Binomiale négative si } \frac{\hat{\sigma}^2}{\hat{\mu}} > 1,1 \end{array} \right.$$

3.3.2.2 Calibration de la loi de sévérité

Pour modéliser la sévérité, les lois usuelles pour la sévérité présentées au tableau 8 sont ajustés par maximum de vraisemblance sur un échantillon de données crée au préalable. La calibration de la loi de sévérité se fait sur les montants de sinistre qui dépassent à l'ultime le seuil de modélisation fixé. Pour choisir une loi de sévérité parmi celles testées, l'espérance et la variance théoriques des lois sont comparées à l'espérance et la variance de l'échantillon de sinistres utilisé. On vérifie surtout que l'espérance et la variance des récupérations des tranches de réassurance que l'on cherche à coter obtenues avec la loi calibrée sont suffisamment proches du Burning Cost.

Si les écarts observés sont trop importants, l'usage de lois de sévérité par morceaux peut être envisagé afin de réduire ces écarts. Concrètement, un ou plusieurs seuils de coupure sont définis, et autant de lois que de morceaux obtenus sont calibrées : une courbe de sévérité par morceaux est alors obtenue. La loi de fréquence reste quant à elle unique.

3.3.2.3 Conclusion

Le modèle fréquence-sévérité obtenu peut être utilisé pour réaliser des simulations stochastiques qui visent à générer un nombre important de scénarios de sinistralité brute qui soient cohérents avec les données observées. Cela permet par la suite d'appliquer des traités de réassurance et de déterminer la prime pure ainsi que des indicateurs techniques comme les périodes de retour.

CHAPITRE 4

Problématique du mémoire

La modélisation sur expérience permet d'obtenir des résultats satisfaisant lorsque le volume de données de sinistralité historique est suffisamment important pour garantir une bonne calibration des différentes lois. Dans le cas contraire, l'absence de donnée doit être comblée par le recours à des données externes via la construction d'un modèle sur exposition. A l'heure actuelle, nous ne disposons pas en interne de méthode permettant d'estimer la sinistralité brute incendie au-delà d'un seuil d'une cédante à partir de son exposition.

L'objectif de ce mémoire est donc d'implémenter des méthodes de modélisation sur exposition de la sinistralité brute incendie au-delà d'un seuil reflétant plus justement le risque des cédantes dont les données de sinistralité sont insuffisantes.

En s'appuyant sur des données marché - des données sinistres et d'exposition provenant de différentes cédantes - nous souhaitons développer un modèle qui permettrait de simuler la sinistralité de toutes les cédantes en utilisant leurs données d'exposition.

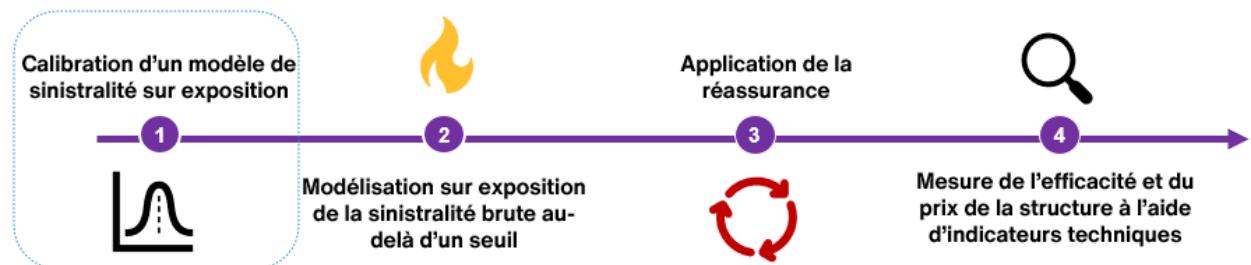


FIGURE 17 – Étapes globales du processus de modélisation AON

L'objectif est donc de développer l'étape '1' du schéma ci-dessus (figure 17) qui présente le processus global de modélisation employé en interne. L'implémentation des méthodes s'est voulue

graduelle. Le premier modèle qui sera mis en place, le modèle 0D, est un modèle fréquence/coût marché par segments dont les lois de fréquence sont rapportées à l'exposition de la cédante (le facteur d'exposition utilisé est le nombre de biens assurés). Ensuite, pour considérer le profil de risque des cédantes, les méthodes de tarification sur exposition 'classiques' reposant sur des courbes d'exposition marché ont été implémentées et étudiées. Notamment, les courbes MBBEFD introduites par Stefan Bernegger seront utilisées. Elles seront utilisées dans un premier temps pour calculer une prime pure, puis le modèle sera détourné afin de pouvoir employer une approche stochastique. Enfin, nous proposerons un modèle s'inscrivant dans la continuité du modèle 0D et reposant sur la relation liant somme assurée et sinistralité sur cette branche de risque. L'objectif serait d'obtenir un modèle de fréquence et un modèle de sévérité dont les paramètres des lois seraient des fonctions de la somme assurée.

Deuxième partie

Première approche de modélisation sur exposition : le modèle 0D

5 Méthodologie	53
5.1 Construction d'une base marché	53
5.2 Gestion des seuils de communication	53
5.3 Ajustement des lois de fréquence et de sévérité par segment	57
5.4 La question de l'inflation	58
6 Application à un cas pratique	60
6.1 Présentation du portefeuille de la cédante et de la structure étudiée	60
6.2 Modélisation sur expérience	63
6.3 Modélisation sur exposition : utilisation du modèle 0D	65
6.4 Comparaison des résultats avec les méthodes sur expérience et sur exposition	72
6.5 Conclusion	75
7 Limites du modèle	76

Intuition

On cherche à construire une modélisation basée sur l'exposition du client pour compléter la modélisation sur expérience qui, on l'a vu, peut s'avérer insuffisante. On dispose par ailleurs de la statistique sinistre d'un grand nombre de cédantes. On souhaiterait donc pouvoir exploiter la statistique sinistre des cédantes comparables à celle qu'on cherche à modéliser. Une première approche consisterait à réaliser une modélisation sur expérience sur une base de sinistralité marché et d'appliquer ensuite ce modèle sur la cédante à modéliser en ajustant la fréquence à l'exposition de la cédante.

Cette approche s'inspire de la méthode de modélisation classique de la sinistralité incendie décrite en 3.3. Seulement, la constitution d'une base de données marché vient avec des défis supplémentaires : la disparité des seuils de communication entre les cédantes et la consolidation d'une base de données présentant un profil de risque similaire à celui du client. La gestion des seuils différents impose une modélisation différenciée en fonction du seuil afin d'exploiter l'intégralité de l'information disponible. Une approche simple serait de tronquer de la base sinistre au-delà du seuil maximale. Cela aurait pour principal désavantage de ne pas pouvoir modéliser la sinistralité en dessous du seuil maximale. La consolidation de la base de données se fait principalement par appréciation qualitative en amont des modélisations.

CHAPITRE 5

Méthodologie

Ce chapitre explique en détail la méthodologie mise en place afin de proposer une modélisation marché applicable à toute cédante dont l'exposition (le nombre de biens assurés en portefeuille) est connue. Nous parlerons de modèle 0D car il s'agira d'un modèle calibré sur une exposition constante, sans ajout d'une dimension explicative permettant d'affiner la modélisation (que ce soit la fréquence ou la sévérité des sinistres). Notamment, la gestion des seuils de communications différents pour chaque cédante de la base est expliquée.

5.1 Construction d'une base marché

La première étape du modèle est de constituer une base de données sinistres avec des cédantes dont les profils de risque sont similaires. L'objectif est d'obtenir une base représentative du marché sur le segment de risque étudié. Il s'agit d'une étape se reposant sur le jugement de l'actuaire qui doit identifier des cédantes qui ont un profil au maximum similaire à celui de la cédante étudiée. Cela se traduit par une étude du portefeuille des cédantes, des différentes couvertures, etc... Cette étape induit un biais assez fort dans les modélisations et doit être réalisée avec minutie, en faisant un compromis entre quantité de données pour une calibration précise et qualité de la donnée (la plus proche possible du risque étudié). Une étape de mise en forme des données est nécessaire pour constituer une base marché unique. Ensuite, la modélisation s'effectue en suivant les mêmes étapes que la modélisation sur expérience détaillé dans le chapitre 3.3. La particularité d'une telle modélisation marché est la disparité des seuils de communication.

5.2 Gestion des seuils de communication

Dans cette partie, nous expliquerons la marche choisie pour gérer les disparités au niveau des seuils de communications induites par la présence de plusieurs cédantes dans la base.

POURQUOI SEGMENTER LA BASE ?

La segmentation des données et des modélisations permet de s'assurer de l'homogénéité des données utilisées. La base utilisée est une base marché ; par définition, les sinistres la constituant sont issus de cédantes différentes. Or, les cédantes ne fixent pas toutes leur seuil de communication au même niveau. Certaines auront tendance à communiquer les sinistres assez bas, tandis que d'autres les communiqueront à partir d'un seuil bien plus élevé. Le niveau d'information transmis par les cédante n'est pas le même. Pour travailler sur des bases homogènes, les données doivent alors être segmentées. La figure 18 ci-dessous schématise la segmentation effectuée.

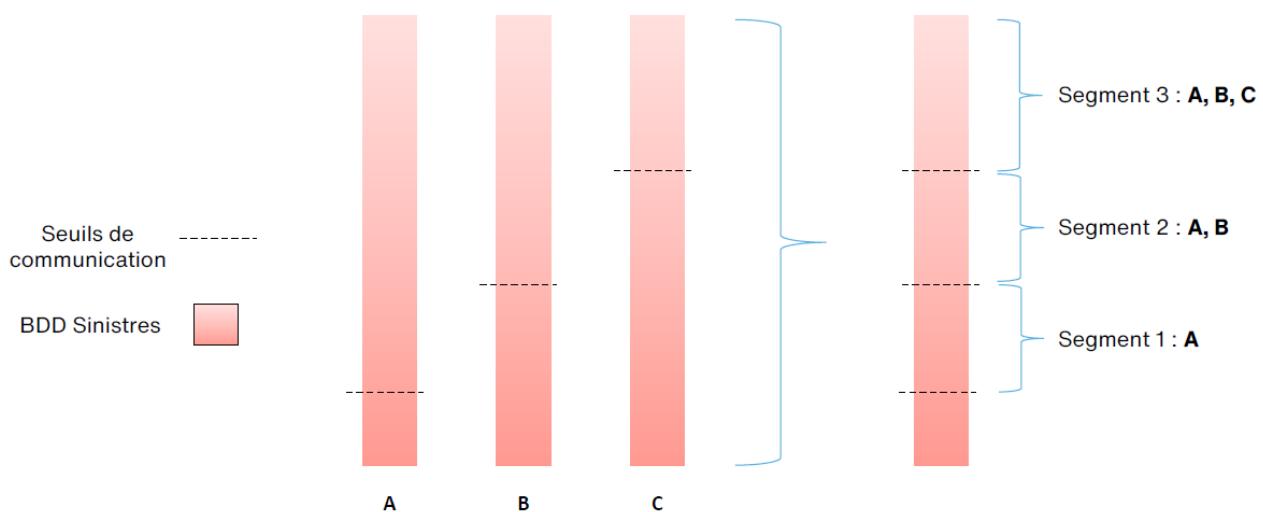


FIGURE 18 – Segmentation de la base

Les segments obtenus seront alors homogènes, et des modèles fréquence/coût pourront être calibrés sur chacun d'entre eux.

Mise en as-if

En premier lieu, les sinistres sont retraités pour être vus à la même année de survenance (mise en as-if). Ce retraitement est le même pour toutes les cédantes de la base, il est alors appliqué sinistre par sinistre, sans se soucier de la cédante d'origine.

Les seuils de communication des différentes cédantes de la base sont aussi indexés, en utilisant l'indice *FFB*.

Soient

- k le nombre de cédantes constituant la base de données,
- n^c l'année de cotation,
- n^{min} l'année de survenance minimale de la base de données,

- SC_i le seuil de communication de la cédante i ,
- SC_i^{ind} le seuil de communication indexé de la cédante i

Pour chaque cédante $i \in \{1, \dots, k\}$

$$SC_i^{ind} = SC_i \times \frac{FFB_{n^c}}{FFB_{n^{min}}}$$

A cette étape, toutes les données de la base sont indexées.

Segmentation de la base sinistres en fonction des seuils de communication

Les données étant indexées, une segmentation va maintenant être définie. La base sera alors segmentée en sous-bases indépendantes les unes des autres. Les traitements et modélisations seront ensuite réalisés indépendamment sur chaque segment. Les seuils de communication indexés des cédantes de la base sont utilisés pour définir les segments.

Prenons l'exemple de trois cédantes, A, B et C, dont les seuils de communications indexés sont respectivement 0,5M, 1M et 2M. Alors, la figure 19 montre comment les seuils seront utilisés pour définir nos segments :

Seuils de communication indexés	
A	500 000 €
B	1 000 000 €
C	2 000 000 €

→

Segment	Borne Inf	Borne Sup
1	500 000 €	1 000 000 €
2	1 000 000 €	2 000 000 €
3	2 000 000 €	Infini

FIGURE 19 – Seuils de modélisation

Après définition des bornes, il faut affecter chaque sinistre au segment lui correspondant. Pour cela, la charge indexée maximale par sinistre est comparée aux bornes inférieures et supérieures des segments définis précédemment, ainsi qu'au seuil de communication indexé de sa cédante d'origine.

Soient

- C_{max}^{ind} la charge indexée maximale du sinistre X
- LB_i la borne inférieure du segment i
- UB_i la borne supérieure du segment i
- Sc_X le seuil de communication indexé de la cédante d'origine de X

Pour chaque segment $i \in \{1, k\}$, on a :

- * si $C_{max}^{ind} \geq LB_i$ et $Sc_X \leq LB_i$, alors X est affecté au segment i .
- * si $C_{max}^{ind} \geq LB_i$ mais $Sc_X > LB_i$, alors X n'est pas affecté au segment i . Dans ce cas, on identifie le segment j tel que $Sc_X = LB_j$ et X est affecté au segment j .

EXEMPLE

Pour illustrer l'affectation des sinistres aux segments de modélisation, reprenons l'exemple précédemment utilisé avec les trois cédantes A, B et C. Pour rappel, on a

$$\begin{cases} Sc_A^{ind} = 0,5M \\ Sc_B^{ind} = 1M \\ Sc_C^{ind} = 2M \end{cases}$$

Les seuils ainsi définis entraînent la création de trois segments distincts (figure 19) :

Segment 1 - $LB_1 = 0,5M$, $UB_1 = 1M$

Segment 2 - $LB_2 = 1M$, $UB_2 = 2M$

Segment 3 - $LB_3 = 2M$, $UB_3 = \infty$

En fonction de leur charge indexée maximale, les sinistres de la cédante A seront répartis entre les 3 segments de modélisation. Les sinistres de la cédante B seront répartis entre les segments 2 et 3, et tous les sinistres de la cédante C seront affectés au segment 3. La figure 20 illustre le procédé.

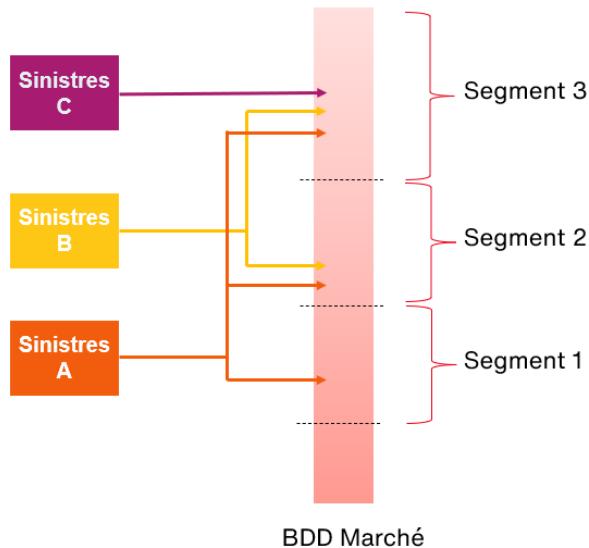


FIGURE 20 – Affectation des sinistres en fonction de la charge indexée maximale et du seuil de communication

Les sinistres sont ensuite projetés à l'ultime.

Projection à l'ultime

Des coefficients de Chain-Ladder sont calculés sur chaque segment comme décrit en 3.3.1, et les sinistres correspondants sont projetés à l'ultime. Pour chaque segment, le seuil de modélisation i.e le seuil filtrant les données utilisées pour le calcul des facteurs de développement est égal à la borne inférieure du segment. Après la projection à l'ultime, il est possible qu'un sinistre change de segment. Si tel est le cas, il y a deux possibilités : soit le sinistre passe en dessous de son seuil de communication et on l'exclut, soit il change simplement de segment.

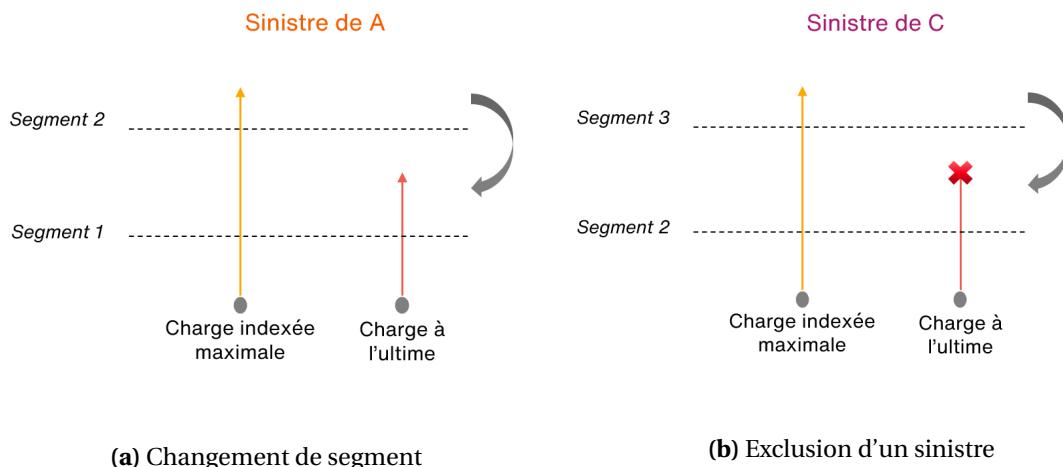


FIGURE 21 – Gestion de la segmentation après projection à l'ultime

La figure (a) présente un cas de changement de segment, et la figure (b) présente un cas d'exclusion de la modélisation d'un sinistre après projection à l'ultime.

Les retraitements de la base sont maintenant terminés. Les sous-bases obtenues vont être utilisées pour ajuster des lois modélisant la sinistralité au sein de chaque segment.

5.3 Ajustement des lois de fréquence et de sévérité par segment

La dernière étape de la méthode est l'ajustement de lois de fréquence et de sévérité sur chaque segment de modélisation. La méthode employée dans le cas classique et présentée au chapitre 3.3.2 est alors appliquée à chaque sous-base.

Ajustement des lois de fréquence

Pour chaque segment i , la loi de la variable aléatoire T_i du nombre de sinistres sur le segment est calibrée grâce aux données en reprenant les étapes détaillées en 3.3.2.1. Néanmoins, les lois obtenues sont des lois calibrées sur une exposition marché. Pour les utiliser sur une cédante individuelle, il est nécessaire de mettre à l'échelle la loi de fréquence à la taille de la cédante. Nous utilisons alors pour chaque segment i le facteur suivant :

$$\alpha_i = \frac{E}{E_i}$$

avec

- E l'exposition de la cédante,
- E_i l'exposition marché sur le segment i .

Ensuite, on fait l'hypothèse que la fréquence évolue linéairement avec l'exposition. La fréquence pour un nombre E de polices est le résultat de l'agrégation de E fréquences individuelles qui sont supposément indépendantes entre elles et identiquement distribuées. L'espérance du nombre de sinistres d'un échantillon E de polices est donc la somme des espérances pour chaque police individuelle. De même la variance du nombre de sinistres d'un échantillon E de polices est la somme des variances individuelles (car on suppose qu'il y a indépendance entre les polices). Autrement dit, l'espérance et la variance varient linéairement avec le nombre de polices. On a alors :

$$\begin{cases} \mathbb{E}[T'_i] = \alpha_i \mathbb{E}[T_i] \\ \mathbb{V}[T'_i] = \alpha_i \mathbb{V}[T_i] \end{cases}$$

Les lois de fréquences de chaque segment sont alors définies par ces nouveaux paramètres.

Ajustement des lois de sévérité

Des lois de sévérité sur chaque segment sont ajustées. Pour rappel, la loi est ajustée par maximum de vraisemblance sur les charges à l'ultime des sinistres supérieures au seuil de modélisation.

Finalement, les lois de fréquence et sévérité de chaque segment peuvent être resimulées dans ReMetrica afin de générer la sinistralité brute de la cédante. Une caractérisation complète des lois de la sinistralité brute permet, par l'intermédiaire d'une simulation stochastique, d'appliquer les traités de réassurance de manière détaillée et d'obtenir ainsi la distribution des cessions en réassurance. Cette distribution détaillée permet la tarification des traités de réassurance mais aussi plus généralement la modélisation des résultats techniques de la branche.

5.4 La question de l'inflation

La mise en place d'une méthodologie de modélisation requiert une réflexion approfondie sur sa robustesse face à l'inflation, en particulier dans le contexte actuel caractérisé par une inflation marquée. Le fondement de la méthode exposée ici repose sur une calibration des fréquences et des coûts sur des données de marché, ajustée à l'échelle de l'exposition de la cédante.

Il est essentiel de noter que les lois de fréquence et de coût ont été calibrées sur les données observées au cours de l'année de cotation N. Pour illustrer, dans le cas d'une cotation en 2024, tous les sinistres sont considérés comme survenant en 2024. Ces lois, bien que représentatives de l'année de cotation étudiée, ne tiennent pas compte de l'évolution potentielle des montants de sinistres liée à l'inflation.

Ainsi, chaque année, il devient impératif de calibrer de nouvelles lois en se basant sur des données ajustées comme si elles provenaient de l'année de cotation étudiée. Cette démarche permet de prendre en considération les variations dues à l'inflation et d'assurer la pertinence continue du modèle au fil des renouvellements. Ce modèle a donc vocation à être mis à jour tous les ans.

CHAPITRE 6

Application à un cas pratique

Ce modèle a déjà pu être utilisé dans le cadre des travaux de modélisation des cessions de traités de réassurance de l'année 2023. L'objectif des travaux était donc dans un premier temps de modéliser la sinistralité brute au-delà d'un seuil du client, puis d'appliquer le traité de réassurance étudié aux données ainsi simulées. Cette section permettra de présenter concrètement les étapes de modélisation et les résultats obtenus avec le modèle dans le cas d'une cédante, que nous appellerons cédante A. Les résultats obtenus avec une modélisation basée uniquement sur l'expérience sinistre de la cédante sont aussi présentés; nous pourrons alors comparer quantitativement les deux méthodes.

6.1 Présentation du portefeuille de la cédante et de la structure étudiée

Cette section permettra de contextualiser les travaux qui ont été réalisés. Dans un premier temps, le portefeuille couvert sera décrit; puis, le programme de réassurance étudié sera également décrit.

6.1.1 Le portefeuille

Le portefeuille étudié est un portefeuille Dommages aux biens des particuliers et des professionnels. La figure 23 présente globalement les capitaux sous risques de ce portefeuille, évalués en 2023 :

	Biens assurés	Sommes assurées (en Milliards d€)
MRH	3 091 054	484,8
MRPRO	109 081	30,6
TOTAL	3 200 135	515,4

FIGURE 22 – Capitaux sous risques des souscriptions Dommages aux biens

Le portefeuille est très largement résidentiel. Il est important de préciser que ces sommes assurées ne contiennent pas les frais et pertes, les frais de déblais et de démolition, les pertes indirectes ou encore les recours des voisins et des tiers (bien qu'ils soient couverts par le traité qui sera présenté dans la section suivante). La figure suivante est un graphique de la répartition des contrats MRH/MRPRO et des sommes assurées par type de contrats.

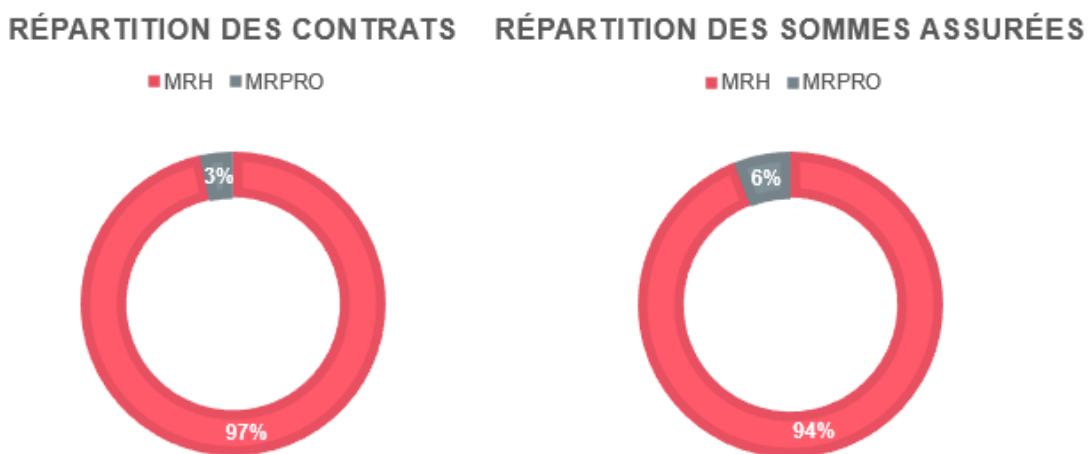


FIGURE 23 – Répartition des biens et des sommes assurées en portefeuille

Bien qu'ils ne représentent que 3% des contrats totaux, les risques professionnelles comptent pour 6% des sommes assurées globales du portefeuille : cela s'explique par le fait que les sommes assurées des risques professionnelles sont en général plus élevées que les sommes assurées des risques résidentiels.

Le client nous a également transmis son profil de risques pour les risques de somme assurée supérieure à 3M€. Il est présenté en figure 24.

LB	UB	Nombre de risques	Somme assurée totale	Prime totale
3 000 000	4 000 000	310	1 048 254 138	923 644
4 000 000	5 000 000	125	556 842 520	498 364
5 000 000	6 000 000	57	309 437 398	227 392
6 000 000	7 000 000	39	252 002 432	211 793
7 000 000	8 000 000	12	89 219 760	54 248
8 000 000	9 000 000	13	109 082 746	71 749
9 000 000	10 000 000	14	132 275 132	78 289
10 000 000	11 000 000	12	125 586 720	114 145
11 000 000	12 000 000	8	91 759 371	70 348
12 000 000	13 000 000	8	99 775 543	54 947
13 000 000	14 000 000	5	67 249 884	56 325
14 000 000	15 000 000	7	101 276 805	72 455
15 000 000	16 000 000	5	77 391 219	62 024
16 000 000	17 000 000	6	98 080 819	82 614
17 000 000	18 000 000	1	17 643 600	18 774
18 000 000	19 000 000	6	110 922 235	69 647
19 000 000	20 000 000	0	-	-
20 000 000	21 000 000	2	40 774 538	19 469
21 000 000	22 000 000	2	43 534 561	23 914
22 000 000	23 000 000	4	90 502 770	67 945
23 000 000	24 000 000	0	-	-
24 000 000	25 000 000	2	49 993 760	60 131
25 000 000	30 000 000	40	1 047 178 364	511 367
30 000 000	35 000 000	8	256 688 340	173 922
35 000 000	40 000 000	8	299 653 258	196 315
40 000 000	45 000 000	12	512 799 452	412 227
45 000 000	50 000 000	8	380 739 255	510 977
50 000 000	1 000 000 000	28	2 238 965 552	2 987 059

FIGURE 24 – Profil de risques de la cédante A

Au total, 742 risques sont concernés pour une exposition en terme de somme assurée d'environ 8,2 milliards d'euros et 7,6 millions d'euros de primes. Le profil a été transmis tel quel et inclut à la fois les risques résidentiels et professionnels : la distinction n'est pas communiquée.

6.1.2 La structure de réassurance

Le traité étudié dans ce cas d'étude est un traité *Man Made* Incendie. Autrement dit, le traité couvre les risques qui ne résultent pas d'évènements naturels. Trois tranches parallèles composent le traité :

- 47M XS 3M par risque couvrant l'incendie,
- 180M XS 5M par évènement (2 risques ou plus sinistrés) couvrant la conflagration,
- un Stop Loss 180M XS 30M couvrant le risque terrorisme.

Un AAL commun à 175M est également pris en compte. Il n'y a pas de reconstitutions pour les tranches contribuant à l'AAL. Le traité couvre l'ensemble des souscriptions Dommages aux biens de particuliers ou de professionnels. La figure 25 schématisse la structure de réassurance en question.

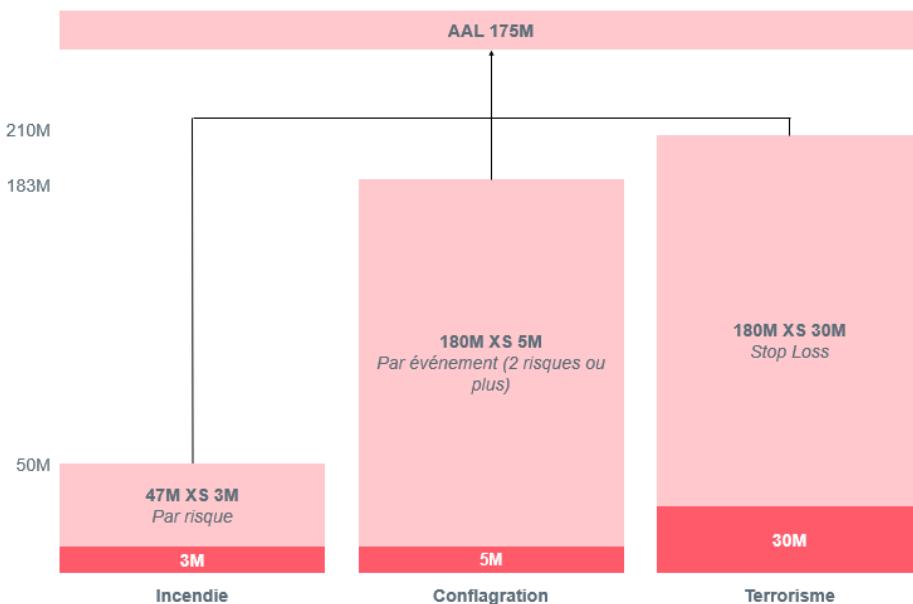


FIGURE 25 – Structure de réassurance étudiée

La prime commerciale payée pour le traité global était de 3,76M€.

La conflagration et le terrorisme ont été modélisés par l'équipe Cat Modelling. Nous modéliserons donc uniquement la tranche incendie par risque, l'impact de l'AAL sera alors difficile à estimer (il faudrait grouper toutes les modélisations). Néanmoins, cette étude permettra de mettre en avant l'utilisation de méthodes sur exposition, et l'enjeu est surtout ici centré sur la modélisation de la sinistralité grave incendie dans le cas d'une statistique sinistre historique insuffisante. A partir de nos simulations de sinistres, nous pouvons envisager l'étude de tout type de traité. L'objectif est d'estimer la contribution de la tranche incendie par risque au traité global.

Historiquement, la cédante n'a pas de sinistres ayant attaché la tranche. Ce cas d'étude sera donc très pertinent pour montrer les limites des méthodes sur expérience et l'intérêt d'utiliser des méthodes sur exposition pour modéliser la sinistralité dans certains cas.

6.2 Modélisation sur expérience

Les données transmises par le client contiennent 46 sinistres, dont le plus ancien est survenu en 1998. Le seuil de communication est de 1M€. Les sinistres transmis sont donc les sinistres impliquant une unique police et ayant déjà été évalués au moins une fois au dessus du seuil à 1M. Ils sont transmis bruts de tout. L'objectif ici est de modéliser la sinistralité incendie au-delà d'un seuil par risque individuel du client, et mesurer son impact sur la réassurance souscrite, ainsi que l'efficience de la réassurance.

Dans un premier temps, nous appliquons la méthode de modélisation classique sur expérience d'AON en reproduisant toutes les étapes détaillées dans le chapitre 3.3.

Après réduction de la fenêtre de survenance des sinistres [2009-2022], indexation et projection à l'ultime, seuls 10 sinistres dépassent le seuil de modélisation choisi (qui est égal dans ce cas au seuil de communication indexé), dont seulement 1 dépassant la barre des 3M et faisant travailler l'XS. L'échantillon obtenu est utilisé pour ajuster par maximum de vraisemblance les lois de sévérité usuelles présentées dans le tableau 8. La figure 26 présente les fonctions de répartition obtenues à la suite de la calibration.

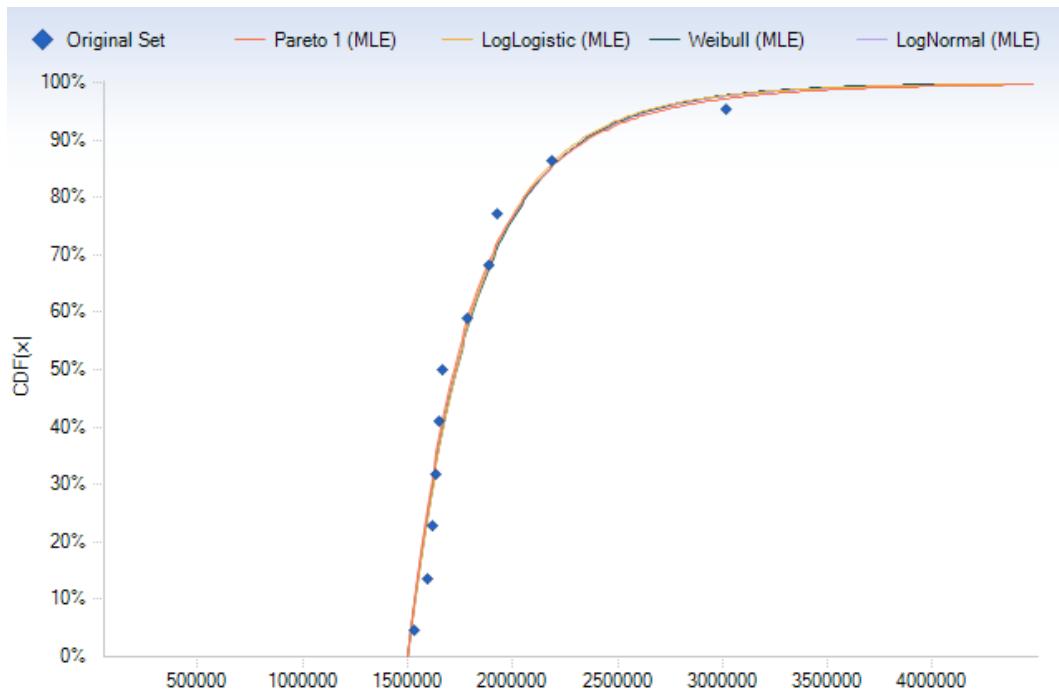


FIGURE 26 – Fonctions de répartition de différentes lois ajustées sur les données de la cédante

Concernant la fréquence, on obtient une moyenne pondérée par l'exposition de 1,14.

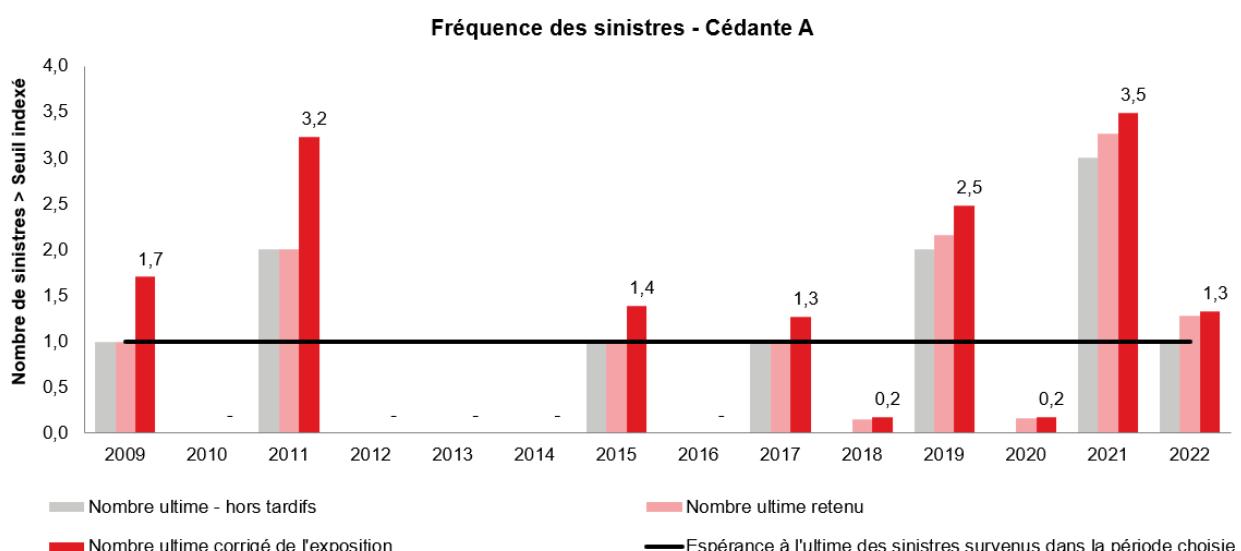


FIGURE 27 – Résultat de la calibration de la loi de fréquence avec les données de la cédante

La calibration de la sévérité sur la sinistralité historique de la cédante ne permet pas de générer des

sinistres de montants importants qui ne sont pas présents dans la statistique. Les récupérations calculées à partir de cette sévérité dans une tranche 47 XS 3 M€ seront artificiellement basses. De même, l'estimation des VaR extrêmes ne peut pas être réalisée à partir de cette modélisation. Le modèle reflète correctement les données historiques mais les données historiques ne sont pas suffisamment représentatives du risque que l'on cherche à modéliser. Dans ce contexte, nous avons décidé de modéliser le risque à l'aide du modèle 0D pour compléter les données de la cédante par des données marché et obtenir un modèle de sinistralité plus pertinent.

6.3 Modélisation sur exposition : utilisation du modèle 0D

Pour pallier la pauvreté de la statistique sinistre historique de la cédante, nous utilisons le modèle 0D.

1. Construction de la base marché

La première étape de l'utilisation du modèle est de construire une base marché, à priori homogène en terme de risque. Dans notre cas, la cédante est principalement exposé à du risque résidentiel, mais également à des risques professionnels (à hauteur de 3% des souscriptions globales Dommages Aux Biens). Nous avons réussi à identifier parmi nos clients 6 cédantes au profil de risque en partie similaire à celui de la cédante A. Pour les identifier, nous nous sommes appuyés sur la répartition des risques des portefeuilles clients. Nous nous sommes donc concentrés sur les cédantes couvrant un portefeuille majoritairement résidentiel, avec une très faible part de risque professionnel. Le but était après agrégation de retrouver la même répartition MRH/MRPRO de la base marché et de la cédante A, à savoir 97% de contrats résidentiels, et 3% de contrats professionnels. Nous avons réussi à obtenir une base très similaire à la base de la cédante A, comme le montre la figure 28 ci-dessous :

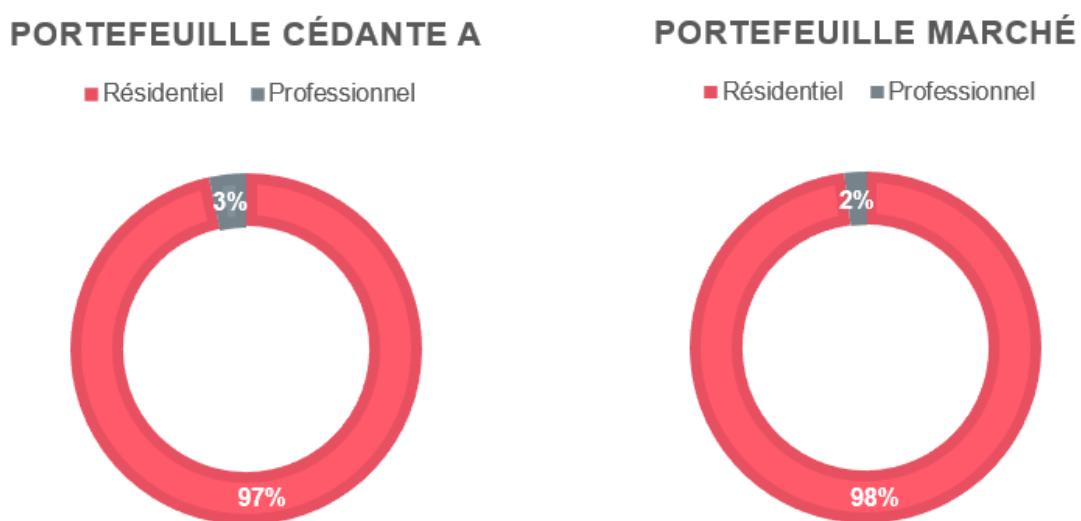


FIGURE 28 – Répartition des risques par portefeuille

L'exposition de la cédante A est de 3,2M de biens assurés, tandis que l'exposition de la base marché

est de 13,4M de biens assurés (y compris la cédante A). Nous obtenons alors une base composée de 7 cédantes aux profils de risques similaires, mais aux seuils de communications différents. On définit la survenance minimale de la base à 2012 : c'est un compromis entre nombre de sinistres et sinistres pas trop anciens.

Cédante	Nombre de sinistres	Seuil de communication (en K€)
1	673	75
2	714	150
3	285	250
4	912	300
5	214	350
6	86	500
A	23	1000

TABLE 11 – Descriptif des cédantes de la base marché constituée

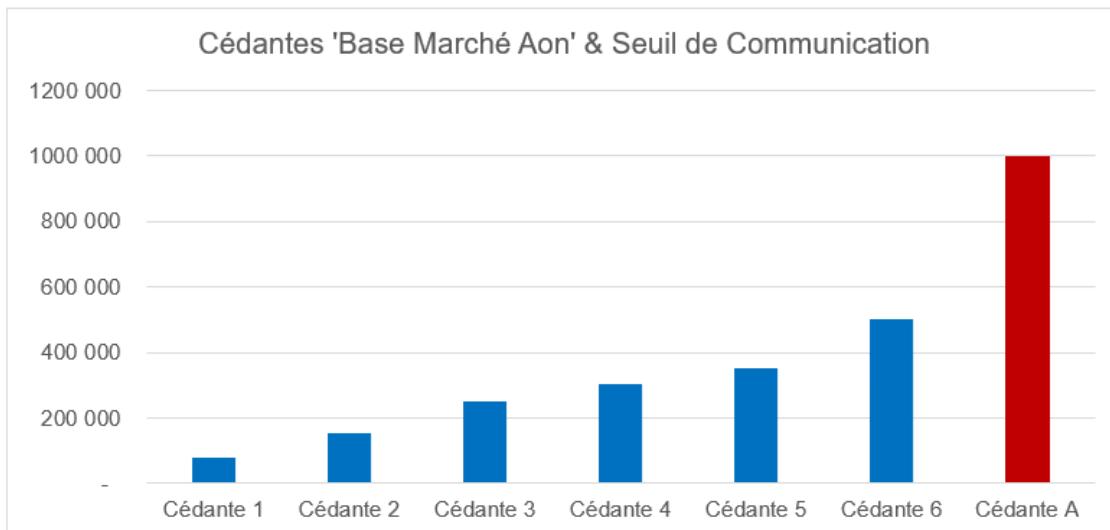


FIGURE 29 – Seuils de communication de la base marché ainsi contruite

Dans le cadre de cette étude, nous nous intéressons uniquement à la modélisation au dessus du plus haut seuil de communication de la base (correspondant au seuil de la cédante A).

2. Mise en As-if et projection à l'ultime de la base de donnée

La base "marché" étant maintenant construite, et la fenêtre de survenance clairement définie [2012-2021], nous pouvons passer à l'étape de gestion des données : l'indexation et la projection à l'ultime. L'outil R que nous avons crée automatise tous ces traitements, en utilisant des segmentations par cas pour respecter les disparités au sein des seuils de communication et ainsi construire des bases à priori homogènes.

Pour une cotation 2024, l'indice d'indexation des seuils est calculé comme :

$$i = \frac{FFB_{2024}}{FFB_{2012}} = 1,385$$

Alors, on a les résultats suivants :

Cédante	Seuil de communication (en K€)	Seuil de communication indexé (en K€)
1	75	104
2	150	208
3	250	346
4	300	416
5	350	485
6	500	693
A	1000	1385

TABLE 12 – Seuils de communication des cédantes de la base marché avant et après indexation

La segmentation de la base de données est alors définie par les seuils de communication indexés :

Segment	Borne Inf (en K€)	Borne Sup (en K€)
1	104	208
2	208	346
3	346	416
4	416	485
5	485	693
6	693	1385
7	1385	∞

TABLE 13 – Tableau des segments obtenus

La suite de l'étude se concentre sur le segment 7 qui est composé des sinistres issus de la cédante A et de toutes les cédantes de la base marché. Calculons maintenant les coefficients de Chain-Ladder pour ce segment. Nous pouvons comparer les coefficients calculés sur les données sinistres de la cédante A avec ceux calculés sur la base marché.

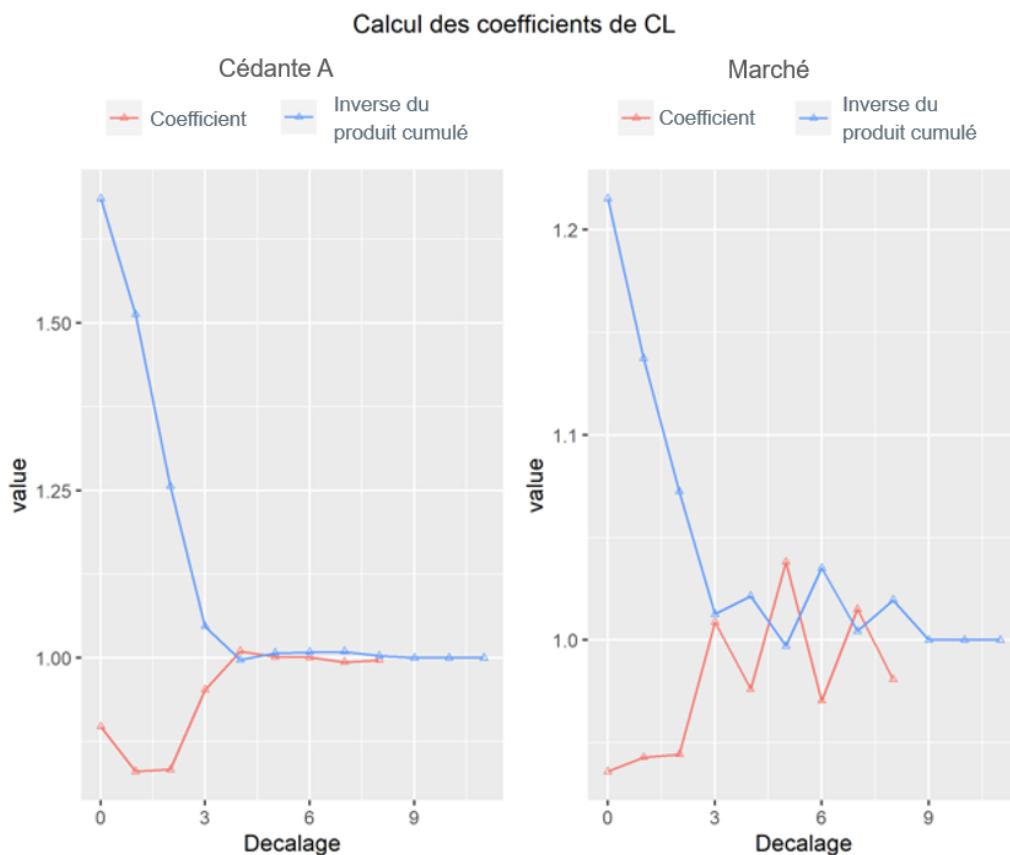


FIGURE 30 – Coefficients de Chain-Ladder

Les coefficients Chain-Ladder calculés à partir de la base marché reposent sur un volume de données plus important que ceux calculés à partir des données de la cédante A uniquement. On peut dès lors supposer qu'ils sont plus représentatifs de la cadence de développement moyenne des sinistres de cette branche au-delà du seuil de 1,385M. Après projection à l'ultime des sinistres, nous comptons 49 sinistres supérieurs à 1,385M, le seuil de modélisation du segment 7. Ces sinistres seront utilisés dans la calibration des lois de fréquence et coût.

3. Modélisation de la fréquence

Pour chaque année de survenance, nous estimons le nombre de sinistres supérieurs au seuil indexé. Des tardifs sont ensuite calculés pour chaque année de survenance en suivant la méthodologie développée en 3.3.2.1. Enfin, le nombre de sinistres par survenance est corrigé de l'exposition.

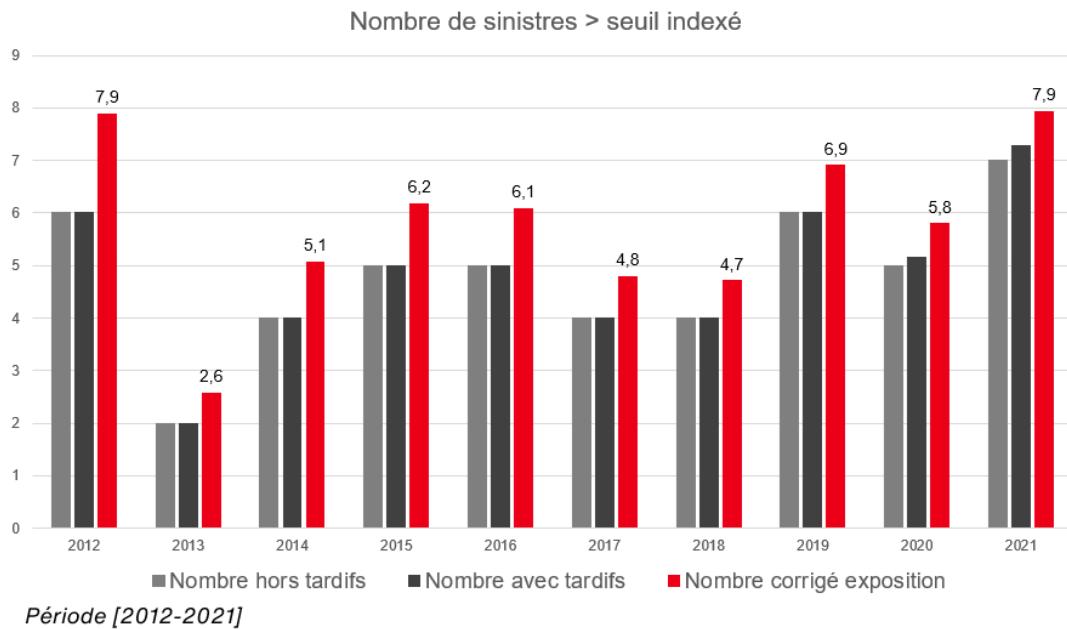


FIGURE 31 – Sinistres supérieurs au seuil indexé par année de survenance

Plus en détail, nous obtenons le tableau de résultats suivant :

Survenance	AnneeCotation	Exposition	Exposition Cotation	NewTardif	NUltimeSup Seuil	NRetenu	NCorrigé	Poids Exposition	Espérance	Variance
2012	2024	10 198 424	13 413 951	-	6,00	6,00	7,89	0,09		
2013	2024	10 431 592	13 413 951	-	2,00	2,00	2,57	0,09		
2014	2024	10 579 730	13 413 951	-	4,00	4,00	5,07	0,09		
2015	2024	10 854 296	13 413 951	-	5,00	5,00	6,18	0,10		
2016	2024	11 021 357	13 413 951	-	5,00	5,00	6,09	0,10		
2017	2024	11 187 657	13 413 951	-	4,00	4,00	4,80	0,10		
2018	2024	11 394 845	13 413 951	-	4,00	4,00	4,71	0,10		
2019	2024	11 653 886	13 413 951	-	6,00	6,00	6,91	0,10		
2020	2024	11 936 668	13 413 951	0,16	5,00	5,16	5,80	0,11		
2021	2024	12 291 289	13 413 951	0,28	7,00	7,28	7,94	0,11	5,82	2,56

FIGURE 32 – Sinistres supérieurs au seuil indexé par année de survenance

L'espérance et la variance obtenues correspondent à une fréquence marché, pour une exposition de 13,4M environ. Au vu du ratio $\frac{\sigma^2}{\mu}$, la loi Binomiale est choisie pour modéliser la fréquence.

$$\frac{\sigma^2}{\mu} = \frac{2,56}{5,82} < 0,9 \implies \text{Loi Binomiale}$$

Il faut ensuite recalculer les moments de la loi en les mettant à l'échelle de l'exposition de la cédante A. Plus formellement, on utilise le ratio de l'exposition de la cédante A par rapport à l'exposition marché (soit 24%) pour effectuer cette mise à l'échelle.

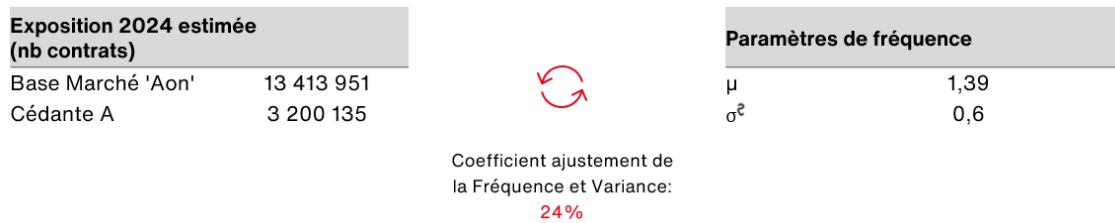


FIGURE 33 – Mise à l'échelle des paramètres de fréquence marché

Finalement, en posant T_7' la variable aléatoire du nombre de sinistres supérieurs au seuil de modélisation de 1,385M pour l'exposition de la cédante A, il vient que :

$$T_7' \sim \mathcal{B}(2; 0,695)$$

Il faut maintenant calibrer une loi de sévérité.

4. Modélisation de la sévérité

Pour calibrer la sévérité, les sinistres projetés à l'ultime supérieurs au seuil de modélisation sont utilisés. Voici la fonction de répartition empirique des données dont nous disposons pour ajuster une loi :

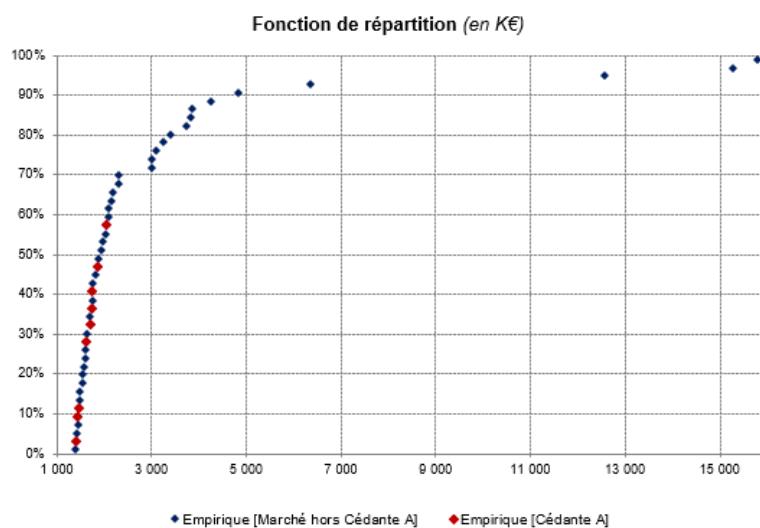


FIGURE 34 – Fonction de répartition empirique de la charge des sinistres

La méthode du maximum de vraisemblance est appliquée à ces données pour calibrer une loi de sévérité. L'utilisation d'un seuil de coupure à 3M permet une meilleure adéquation de la sinistralité (notamment pour le morceau au-delà de 3M, qui est celui auquel nous nous intéressons principalement). Voici les lois choisies :

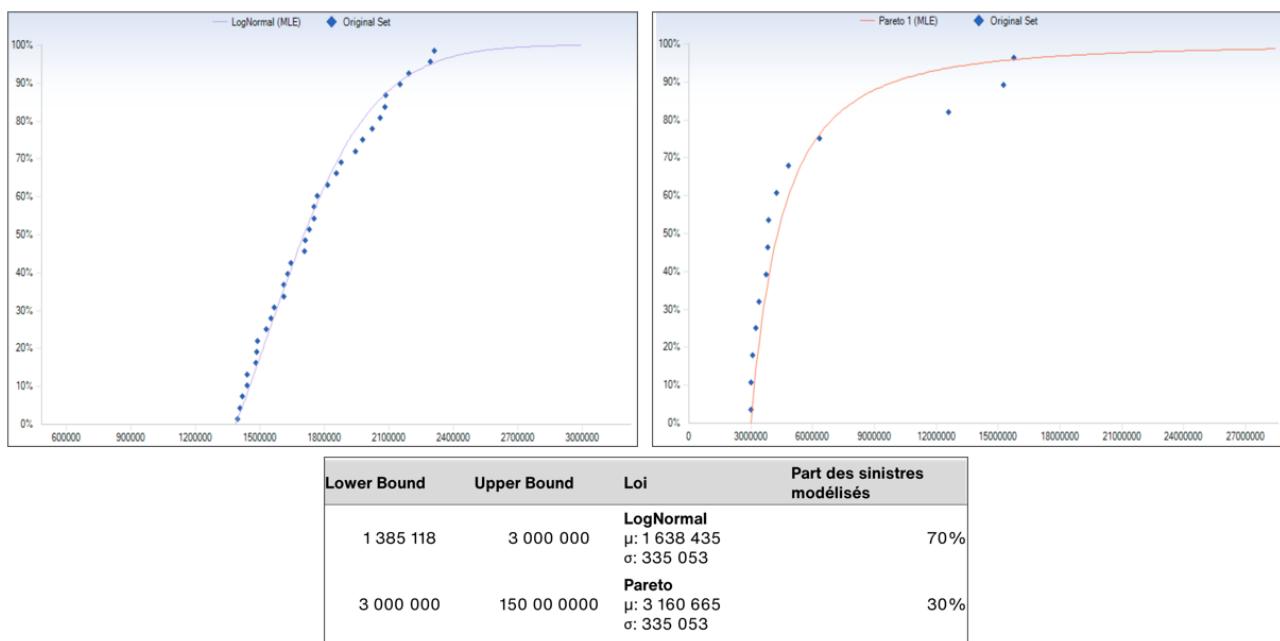


FIGURE 35 – Modèle 0D - Lois de sévérité choisies pour la modélisation

Le modèle obtenu est un modèle composite : nous combinons deux lois (Log-Normale et Pareto) avec une jonction à 3M. Ces modèles sont étudiés en détail dans GOFFARD, [9]. Il s'agit de la combinaison de deux modèles, l'un pour le 'ventre' de la distribution et l'autre pour la 'queue' de distribution, afin d'observer une meilleure adéquation du modèle aux données d'apprentissage. Ils sont définis comme suit :

$$f(x) = \begin{cases} p \frac{f_1(x)}{F_1(\gamma)}, & \text{si } x \leq \gamma \\ (1-p) \frac{f_2(x)}{1-F_2(\gamma)} & \text{si } x > \gamma \end{cases}$$

avec

- f_1, f_2 les fonctions de densité du 'ventre' et de la 'queue'.
- F_1, F_2 les fonctions de répartition du 'ventre' et de la 'queue'.
- p le paramètre de mélange. Dans notre cas, $p = 0,7$.
- γ le paramètre de seuil. Dans notre cas, $\gamma = 3M$

Il est possible de supposer continuité, continuité et dérivabilité ou discontinuité au point de jonction . Pour avoir continuité, il faut fixer un paramètre du modèle (le μ de la Log-Normale par exemple) ce qui n'est pas notre cas. Nous travaillerons donc avec un modèle sans vérifier la continuité de la densité au point de jonction; d'après l'auteur, l'hypothèse de continuité n'est pas nécessaire pour procéder aux simulations qui vont suivre, et il est tout à fait possible de travailler avec une densité potentiellement discontinue au point de jonction.

Les lois de fréquence et sévérité ayant été calibrées sur les données marché et recalculées à partir des données d'exposition de la cédante A, les simulations stochastiques peuvent être réalisées.

Dans le paragraphe qui suit, la sinistralité ainsi que l'application de la réassurance sous la forme d'un traité 47M XS 3M vont être simulés à l'aide d'une part des paramètres déterminés grâce à la modélisation sur expérience, et d'autre part avec les paramètres issus de la modélisation sur exposition présentée dans cette partie.

Pour l'étude d'un traité en excédent de sinistre ne comportant aucune clause, l'étape des simulations n'est pas nécessaire pour obtenir des estimations de la charge du traité et des périodes de retour; il s'agit en effet d'une simple multiplication d'espérances. Cependant, il est difficile d'obtenir des formules analytiques pour ces quantités si le traité étudié comprend diverses clauses. De plus, notre objectif est bien d'avoir une modélisation de la sinistralité brute grave incendie pour avoir matière à travailler sur les structures de réassurance de nos clients. L'étude de traités avec clauses (voir de solutions structurées) doit donc toujours rester envisageable : nous préférerons alors toujours employer une approche par simulations qui s'inscrit dans cette philosophie.

6.4 Comparaison des résultats avec les méthodes sur expérience et sur exposition

Mis à part les paramètres de sinistralité, les structures sont parfaitement identiques. Nous nous attendons à des résultats plus prudents pour le modèle 0D (car la sinistralité modélisée sera plus forte). La tableau ci-dessous rappelle les différentes lois (et leurs paramètres) utilisées pour chaque modèle.

<i>Approche</i>	<i>Loi de fréquence</i>	<i>Seuil de modélisation</i>	<i>Limite</i>	<i>Part des sinistres modélisés</i>	<i>Loi de sévérité</i>
Classique	Binomiale négative $\mu: 1,23$ $\sigma: 1,31$	1 529 405	∞	100%	LogNormal $\mu: 1 280 939$ $\sigma: 383 921$
Modèle 0	Binomiale $\mu: 1,39$ $\sigma: 0,61$	1 385 118	3 000 000	70%	LogNormal $\mu: 1 638 435$ $\sigma: 335 053$
		3 000 000	∞	30%	Pareto $\mu: 2 989 764$ $\sigma: 5 340 555$

FIGURE 36 – Récapitulatif des lois de fréquence et sévérité utilisées pour chaque modélisation

Pour rappel, nous réalisons l'étude d'un traité en excédent de sinistre 47M XS 3M.

Indicateurs statistiques étudiés

Dans un premier temps, nous comparerons les résultats de sinistralité brute. Pour cela, nous étudierons les distributions du sinistre maximal par scénario et de la charge brute globale par scénario. Dans un second temps, nous comparerons les résultats du traité en étudiant les récupérations à la

tranche de celui-ci.

Les indicateurs statistiques suivants seront utilisés :

- la moyenne,
- l'écart-type,
- les périodes de retour : niveaux de sinistralité associés à des périodes de retour données (1 fois tous les 2 ans, une fois tous les 25 ans, etc...),
- la TVaR à 100 ans : espérance de la perte étant donné qu'un événement en dehors d'un niveau de probabilité donné (événement centenaire ici) s'est produit.

Comparaison des sinistralités brutes

La figure ci-après présente les résultats de modélisation en terme de sinistralité brute en fonction de l'approche employée.

	Sinistre Maximal	Charge annuelle		Sinistre Maximal	Charge annuelle
Moyenne	1 249 027	2 254 027	Moyenne	3 292 884	4 200 964
Ecart-type	964 444	2 428 860	Ecart-type	4 688 569	5 102 284
Période de retour			Période de retour		
2 ans	1 678 087	1 728 161	2 ans	1 996 210	3 324 453
5 ans	2 034 652	3 839 630	5 ans	4 218 218	5 682 335
10 ans	2 247 936	5 523 272	10 ans	6 163 823	7 905 813
25 ans	2 505 000	7 406 348	25 ans	9 983 732	11 978 537
50 ans	2 688 656	8 857 597	50 ans	14 228 041	16 431 534
100 ans	2 870 146	10 179 400	100 ans	20 046 740	22 279 292
200 ans	3 055 835	11 392 675	200 ans	28 248 421	30 465 393
500 ans	3 306 870	13 082 816	500 ans	44 853 874	46 855 732
10 000 ans	4 129 282	18 112 944	10 000 ans	122 978 164	123 211 128
TVaR 100 ans	3 137 221	11 968 553	TVaR 100 ans	35 935 150	38 193 634

(a) Cédante A

(b) Cédante A - vue marché

FIGURE 37 – Sinistralité brute

En terme de sinistralité brute, la modélisation sur expérience génère en moyenne une charge modérée (2,25M), et des sinistres qui n'entrent que très rarement dans le traité (1 fois tous les 200 ans en moyenne). De plus, les périodes de retour estimées sont extrêmement faibles au vu de la couverture de réassurance souscrite et du profil de risque de la cédante. Cela s'explique par le peu de graves présents dans l'historique de la cédante. La modélisation n'est donc pas pertinente car elle n'est pas à l'image du risque porté par le client. Elle reflète l'historique mais pas le risque effectif du portefeuille du client que nous avons étudié avant d'entamer les modélisations. De plus, bien que le profil de risque du client ne prenne pas en compte les garanties RC, un certain nombre de risques assurés présentent des expositions bien supérieures au seuil de 3M€ ce qui milite en faveur

de l'utilisation de modèles sur exposition.

Les résultats issus du modèle sur exposition 0D permettent de combler l'absence de sinistralité historique majeure pour la cédante A en utilisant la statistique marché consolidée en amont afin de correspondre au maximum au risque du client. La sinistralité modélisée par ce biais est logiquement plus importante que celle modélisée à partir de la statistique historique de la cédante A. La période de retour à 200 ans du sinistre maximal par scénario est maintenant estimée à 28M. La période de retour à 5 ans est quand à elle estimée à 4,2M. Les lois ayant été calibrées sur une base marché reflétant le profil de risque de la cédante A, puis mises à l'échelle de celle-ci, nous pouvons supposer que cette sinistralité est représentative du risque porté par la cédante A, risque que sa statistique historique ne reflète pas.

A partir de la sinistralité générée par nos simulations, les cessions de réassurance vont pouvoir être estimées. Le traité décrit en 6.1.2 est simplement appliqué aux sinistres générés, et les résultats obtenus à chaque simulation permettent de calculer différents indicateurs statistiques. Notons qu'à partir des sinistres bruts générés par notre modèle nous pourrions estimer l'efficacité et les cessions de réassurance de structures diverses et variées : c'est ce qui fait la force de cette modélisation du brut et c'est bien là qu'est l'enjeu du mémoire.

Comparaison des cessions de réassurance

La figure suivante présente les résultats de modélisation du traité en fonction de l'approche employée.

Récupérations	Modélisation sur expérience	Modélisation sur exposition '0D'
Moyenne	1 650	1 210 785
Ecart Type	30 052	3 749 010
Période de retour		
2 ans	0	0
5 ans	0	1 259 487
10 ans	0	3 306 025
25 ans	0	7 295 305
50 ans	0	11 690 107
100 ans	0	17 640 311
200 ans	55 835	25 741 071
500 ans	306 870	42 290 994
1 000 ans	499 087	47 000 000
TVaR 100 ans	164 968	29 452 290

FIGURE 38 – Résultats de modélisation du traité 47 XS 3 en fonction de l'approche utilisée

La modélisation sur base expérience ne génère pratiquement pas de récupérations sur le traité de

réassurance. La calibration s'étant effectuée sur des données historiques inférieures à la priorité du traité, les résultats de cessions obtenus ne sont pas du tout représentatifs des récupérations attendues par le marché pour cette tranche. La période de retour à 200 ans des récupérations est estimée à 55 835, ce qui est négligeable pour une tranche dont la portée est de 47M en XS de 3M. La prime pure, quant à elle, est estimée à 1650. La prime commerciale pour le traité global a pourtant été payée 3,76M€. Le modèle sur expérience ne reflète donc pas le risque effectif de la cédante ainsi que le risque perçu par le marché.

La prime pure estimée par le modèle utilisant les lois calibrées à l'aide du modèle 0D est de 1,2M. Les résultats sont déjà plus cohérents, et confirment que l'utilisation du modèle était pertinente dans ce cas là. La période de retour des récupérations à 200 ans est estimée à 26M. Le point de sortie de la tranche a quant à lui une période de retour supérieure à 500 ans, ce qui pourrait soulever des questions. Il est alors bon de rappeler que la base de données marché ayant servi à la calibration du modèle a été consolidée de sorte à ce que la répartition MRH/MRPRO des biens assurés soit la même que le portefeuille de la cédante A (97% de MRH et 3% de MRPRO). Le profil de risque de la cédante A n'a pas été utilisé donc les effets individuels des souscriptions de la cédante ne sont pas captés à la perfection par le modèle, ce qui peut expliquer cette période de retour très élevée. Malgré tout, en estimant une prime commerciale grâce à un chargement sur écart-type de 40%, on obtient :

$$PC = PP + Chargement * \sigma_{rec} = 1,21M + 3,75M \times 40\% = 2,71M$$

Pour rappel, la prime commerciale du traité global (Incendie par risque / Conflagration / Terrorisme) est de 3,76M. On observe donc un différentiel d'environ 1M entre ces deux valeurs, mais notre estimation du prix commercial prend uniquement en compte l'incendie par risque. Les résultats obtenus à l'aide de notre modélisation semblent donc assez cohérents avec la vision des risques de la cédante A perçus par les réassureurs.

6.5 Conclusion

Ce cas d'étude permet d'illustrer les limites de la modélisation sur expérience face à l'absence de sinistres historiques suffisamment importants au regard de la tranche de réassurance que l'on cherche à modéliser. Plus précisément, l'absence de sinistre historique au-delà de 3 M€ ne permettait pas d'obtenir une modélisation satisfaisante de la sinistralité au-delà de ce seuil et donc ne permettait pas une modélisation satisfaisante d'une tranche de réassurance commençant à 3 M€.

Nous avons donc pu appliquer le modèle 0D à un cas réel, en répondant à un besoin de l'entreprise, et observer les avantages que cette méthodologie avait à offrir. La méthode permet de proposer une vision plus pertinente du risque dans le cas où celui-ci est clairement sous-estimé à cause d'une sinistralité historique insuffisante.

CHAPITRE 7

Limites du modèle

La mise en place de cette méthodologie nous a permis de répondre en partie à la problématique. Face à des données insuffisantes, nous pouvons maintenant utiliser ces fréquence/coût marché pour offrir une perspective nouvelle de la sinistralité à nos clients. En tant que courtier en réassurance, nous pouvons capitaliser sur notre vaste ensemble de données pour élaborer une vision plus précise du risque. En consolidant une base de données marché homogène, nous pouvons désormais modéliser la sinistralité brute incendie de chaque cédante au-delà de divers seuils (segments de modélisation), palliant ainsi le manque de données historiques.

L'utilisation du modèle repose sur une hypothèse assez forte : on suppose l'homogénéité des risques de la base marché avec ceux de la cédante. La qualité de la base obtenue et des sorties du modèle seront très largement influencés par la connaissance des données sous-jacentes.

Néanmoins, l'utilisation de l'exposition du client reste assez faible avec ce modèle. Lorsqu'il est appliqué à une cédante, seul le nombre de biens de son portefeuille permet de recalculer les paramètres des différentes lois de fréquence du modèle. Il s'agit d'une modélisation marché qui ne permet pas de prendre en compte les individualités de la cédante étudiée, bien que la base d'apprentissage marché se veuille au maximum homogène dans le risque. En incendie, les profils de risque permettent de visualiser ces individualités : ils sont représentatifs du portefeuille d'une cédante, et permettent d'appréhender au mieux ses risques.

De plus, l'adaptation du modèle à l'exposition de la cédante se fait uniquement au niveau de la fréquence des sinistres. Pourtant, en incendie, en fonction des caractéristiques des polices assurées (somme assurée, type de bâti, risque, ...) des études telles que [1] ou [5] précisent que la distribution de la sévérité des sinistres est variable en fonction de ces critères. Alors, dans la suite du mémoire, nous développerons et implémenterons des modèles prenant en compte ces caractéristiques pour

affiner nos modélisations et proposer une modélisation de la sinistralité brute au-delà d'un seuil la plus cohérente possible avec l'exposition des cédantes.

Dans la partie suivante, nous implémenterons les méthodes classiques de tarification sur exposition en incendie utilisant les courbes d'exposition présentées en 2.3 : les méthodes déterministes et stochastiques d'estimation de la prime pure d'un traité à l'aide des courbes d'exposition et du profil de risque de la cédante seront alors présentées. Ces méthodes seront employées au cas de la cédante A et les résultats pourront être comparés aux résultats obtenus avec la méthode sur expérience et avec le modèle 0D. Nous verrons toutefois à l'issue de cette implémentation que la logique d'utilisation sur laquelle repose cette méthode ne répond pas directement à la problématique de ce mémoire ; ces méthodes sont avant-tout utilisées dans un objectif de tarification de réassurance dans un contexte où la sinistralité attendue du portefeuille est connue (c'est une hypothèse du modèle) tandis que nous souhaitons dans ce mémoire caractériser à partir de l'exposition du client sa sinistralité brute au-delà d'un seuil puis calculer à partir du brut modélisé différents indicateurs techniques et financiers (VaR, Récupérations de réassurance, Volatilité, Volatilité transférée, etc...) pour évaluer et conseiller la cédante dans sa gestion des risques.

Troisième partie

Deuxième approche de modélisation sur exposition : implémentation des méthodes classiques de tarification à l'aide des courbes d'exposition

8 Estimation de la prime pure d'un traité à l'aide des courbes d'exposition	80
8.1 Méthodologie et choix des paramètres pour l'application à la cédante A	80
8.2 Résultats obtenus lors de l'application au cas pratique de la cédante A	85
9 Approche stochastique	87
9.1 Méthodologie et choix des paramètres pour l'application à la cédante A	87
9.2 Résultats obtenus lors de l'application au cas pratique de la cédante A	90
10 Les limites des méthodes sur exposition classiques	92

Intuition

Les courbes d'exposition, présentées dans la partie 1 de ce mémoire, sont très utilisées dans la tarification de traités en excédent de sinistres en incendie. L'approche la plus classique permet de déterminer la prime pure d'un traité de réassurance. Pour aller plus loin, HREVUŠ et MAREK présentent dans leur article *Exposure modelling in property reinsurance* ([10]) une utilisation des courbes permettant d'adopter une approche stochastique et ainsi d'obtenir des mesures de volatilité et de générer des sinistres individuels.

Ces méthodes vont être présentées dans les chapitres qui suivent, implémentées en interne et appliquées au cas étudié précédemment (le cas de la cédante A). Ensuite, les résultats seront comparés aux précédentes méthodes déjà employées, et nous discuterons des avantages et limites de ces méthodes et de leur employabilité en interne pour répondre à notre problématique.

CHAPITRE 8

Estimation de la prime pure d'un traité à l'aide des courbes d'exposition

L'objectif de ce chapitre est d'implémenter et présenter la méthodologie d'utilisation des courbes d'exposition pour estimer la prime pure d'un traité en excédent de sinistre par risque. Chaque étape de la méthode sera appliquée au cas de la cédante A : nous pourrons alors comparer le résultat obtenu à ceux obtenus avec la méthode sur expérience et avec le modèle 0D. L'idée de la méthode est de déterminer - à partir d'une hypothèse de sinistralité brute pour chaque bande du profil de risques et d'hypothèses sur la sévérité des sinistres - la part de prime pure revenant à la réassurance.

8.1 Méthodologie et choix des paramètres pour l'application à la cédante A

1. Estimation de la somme assurée moyenne par bande du profil de risque

Pour chaque bande h du profil de risques communiqué par la cédante, nous devons calculer la somme assurée moyenne ASI_h à l'aide de la somme assurée totale TSI_h et du nombre de risques N_h de la bande. On a pour un profil de risques à T bandes :

$$ASI_h = \frac{TSI_h}{N_h} \text{ pour tout } h = 1, \dots, T$$

Ce calcul est appliqué au profil de risques de la cédante A (introduit en 6.1.1).

2. Calculs de la limite l_h et la priorité m_h du traité en pourcentage de la somme assurée moyenne pour chaque bande h

Ensuite, il est nécessaire d'exprimer les bornes du traité étudié en pourcentage de la somme assurée moyenne pour chaque bande. Avec P la portée du traité et M sa priorité, on a :

$$\begin{cases} l_h = \min\left(\frac{P+M}{ASI_h}; 1\right) \\ m_h = \min\left(\frac{M}{ASI_h}; 1\right) \end{cases}$$

Dans le cas étudié, $P = 47M$ et $M = 3M$.

3. Choix de la courbe d'exposition $G_h(.)$ à utiliser pour chaque bande h

Ce point a été abordé lors de la présentation des courbes d'exposition en 2.3. Nous utiliserons les courbes d'exposition MBBEFD introduites par BERNEGGER dans [1]. Pour chaque bande de risque h nous devons fixer le paramètre c qui définit la courbe d'exposition de la bande. Pour rappel, la courbe d'exposition utilisée doit être représentative des risques de la bande, et plus précisément de la distribution des taux de dommages de ces risques. Le choix de c se fait en général en fonction de la somme assurée moyenne de la bande et de la typologie du risque considéré lorsqu'il est connu. Par exemple, pour des risques résidentiels de somme assurée assez faible, la courbe d'exposition SwissRe 1, i.e $c = 1.5$, peut être utilisée. A l'inverse, pour de très gros risques industrielles à valeur assurée très élevée, la courbe Lloyd's pourra être utilisée. De tels complexe sont généralement très bien équipés et protégés; les sinistres sont alors localisés et contenus, les taux de destruction observés en cas de sinistre sont donc en moyenne assez faibles.

Nous devons donc fixer les courbes d'exposition des différentes bandes du profil de risque de la cédante A que nous étudions tout au long de ce mémoire. C'est une étape cruciale qui influencera fortement les résultats en sortie. Pour fixer les courbes, nous nous appuierons sur des recommandations tirées de la littérature scientifique. Notamment, nous utiliserons les recommandations de DESMEDT et al. [5] présentées dans le tableau 14 ci-dessous.

Somme assurée	c
$\leq 100,000$	1,75
250,000	2,25
700,000	2,75
1,300,000	3,2
10,000,000	3,9
$\geq 50,000,000$	4,625

TABLE 14 – Paramètres c en fonction de la somme assurée en 2012

Ces recommandations ont été données en 2012. Afin de prendre en compte l'inflation dans le choix des courbes pour une cotation 2024, il est nécessaire d'indexer les sommes assurées de ce tableau. L'indice FFB sera utilisé pour l'indexation. Pour rappel, $i = \frac{FFB_{2024}}{FFB_{2012}} = 1,385$. Les paramètres obtenus sont présentés dans le tableau 15 ci-dessous.

Somme assurée	<i>c</i>
$\leq 138,500$	1,75
346,250	2,25
969,500	2,75
1,800,500	3,2
13,850,000	3,9
$\geq 69,250,000$	4,625

TABLE 15 – Paramètres *c* en fonction de la somme assurée indexée

Pour les sommes assurées entre celles du tableau, une interpolation linéaire est utilisée.

4. Calcul de $G_h(m_h)$ et $G_h(l_h)$

Les courbes d'exposition étant défini pour chaque bande, il s'agit maintenant d'évaluer la valeur de la courbe d'exposition en m_h et l_h . Le résultat obtenu correspond à la part de prime de la bande suffisant à couvrir les sinistres limités à m_h ou l_h (en taux de dommage).

5. Estimation de la sinistralité brute totale pour chaque bande h

Pour chaque bande h , les primes encaissées totales sont connus. Alors, la sinistralité attendue dans cette bande X_h peut se calculer de la façon suivante :

$$\mathbb{E}[Z_h] = P_h \times LR_h$$

avec

- P_h les primes encaissées pour couvrir les risques de la bande h ,
- LR_h le *Loss Ratio* estimé pour la bande h .

Le *Loss Ratio* peut être directement communiqué par le client; sinon, un *Loss Ratio* marché est utilisé. Dans notre cas, la cédante A ne nous ayant pas communiqué de *LR*, nous utiliserons un ratio marché qui sera le même pour toutes les bandes du profil de risque. Une étude réalisée par Swiss Re [14] présente les *loss ratios* observés sur le marché de l'assurance dommages aux États-Unis par ligne d'activité sur la première moitié de 2023. Nous utiliserons alors ces chiffres pour fixer notre hypothèse.

Line of business	Direct premium written, year to date, USD millions	YOY growth %	Direct loss ratio* YTD %
Fire & Allied Cmbnd	43,422	17.0	64.6
Other Liab (Occurrence)	36,834	3.6	63.7
Workers' Comp	29,696	3.1	48.0
Comm'l Multi Prl	30,495	10.6	64.2
Comm'l Auto Liab	23,663	-1.3	72.4
Other Liab (Claims)	19,215	-6.4	48.1
Inland Marine	18,167	7.3	42.2
Comm'l Auto Phys	7,111	5.1	68.4
Med Prof Liab	6,158	-1	57.6
Fidelity & Surety	5,360	7.3	22.5
Finl & Mrtg Grty	2,935	-1.4	2.8
Ocean Marine	2,980	5.4	48.3
Product Liability	2,498	-0.6	44.2
Earthquake	2,606	23.0	1.6
Warranty	1,822	-8.7	64.1
Credit	1,455	9.9	40.3
Aircraft	1,247	7.5	46.0
Boiler & Machinery	1,400	13.3	36.1
Commercial Total	235,086	4.7	56.3
Pvt Pass Auto Liab	45,175	8.6	75.6
Pvt Pass Auto Phys Damage	65,537	14.2	79.2
Homeowner, Farmowner	72,778	11.1	82.7
Personal Total	223,221	10.9	78.9

Source: S&P Global Capital IQ Pro, NAIC statutory filings, Swiss Re Institute

*Pure direct loss ratio (excluding LAE) = Direct losses incurred / Direct premiums earned

FIGURE 39 – Loss Ratios marché aux États-Unis par LoB

Le traité étudié couvrant l'incendie, nous utiliserons le *LR Fire & Allied Cmbnd*, soit 64,6%. Il est important ici de garder à l'esprit qu'il s'agit d'une valeur marché qui n'est pas forcément représentative de la cédante que nous étudions. Aussi, l'étude a été réalisée sur le marché d'assurance américain qui est différent du marché français. Nous verrons plus loin qu'à courbe d'exposition identique, l'estimation du *Loss Ratio* de la cédante utilisée dans les modélisations a un impact significatif sur les résultats obtenus. C'est une hypothèse très forte du modèle.

De plus, utiliser une hypothèse de *Loss Ratio* pour les bandes du profil de risque suppose que la sinistralité brute de la bande est connue. Ensuite, les différentes étapes décrites dans cette section permettent en fonction de la courbe d'exposition qui a été choisie d'estimer la part de cette sinistralité en moyenne à charge du traité. Or, tout l'enjeu de ce mémoire est justement d'estimer cette sinistralité brute 'grave' (donc au-delà d'un seuil) des bandes grâce à un modèle pour ensuite potentiellement estimer des primes pures de réassurance ou expérimenter différentes structures pour proposer la plus optimale. Nous visons à couvrir les cas où les cédantes ne disposent pas d'une statistique historique suffisante pour estimer leur sinistralité dans les tranches supérieures. De notre côté, nous ne disposons pas d'étude sur la sinistralité marché en incendie dans les tranches non travaillantes nous permettant d'utiliser une hypothèse de *Loss Ratio* AON qui soit assez robuste.

6. Estimation de la part à la charge de la réassurance pour chaque bande h

A partir des résultats précédents, les récupérations moyennes pour chaque bande ($Z_h^{réc}$) peuvent être calculées. On a :

$$\mathbb{E}[Z_h^{réc}] = (G_h(l_h) - G_h(m_h)) \times \mathbb{E}[Z_h]$$

7. Estimation de la prime pure du traité à partir des résultats précédents

Finalement, pour estimer la prime pure globale du traité, il suffit de sommer les récupérations estimées pour chaque bande $h = 1, \dots, T$ du profil de risque, i.e :

$$\mathbb{E}[Z^{réc}] = \sum_{h=1}^T \mathbb{E}[Z_h^{réc}]$$

Nous avons appliqué chacune de ces étapes au profil de risques de la cédante A, et nous présentons les résultats obtenus dans la section suivante.

8.2 Résultats obtenus lors de l'application au cas pratique de la cédante A

La figure 40 présente le détail de l'application des étapes et des résultats par bande du profil de risque.

LB	UB	Nombre de risques	Somme assurée totale	Prime totale	ASI	Sinistralité brute de la bande	mh	lh	c	G(mh)	G(lh)	Prime de réassurance
												1
3 000 000	4 000 000	310	1 048 254 138	923 644	3 381 465	596 674	0,89	1,00	3,29	0,96	1,00	22 378
4 000 000	5 000 000	125	556 842 520	498 364	4 454 740	321 943	0,67	1,00	3,35	0,89	1,00	36 761
5 000 000	6 000 000	57	309 437 398	227 392	5 428 726	146 895	0,55	1,00	3,41	0,84	1,00	23 776
6 000 000	7 000 000	39	252 002 432	211 793	6 461 601	136 818	0,46	1,00	3,47	0,80	1,00	27 259
7 000 000	8 000 000	12	89 219 760	54 248	7 434 980	35 044	0,40	1,00	3,53	0,77	1,00	7 932
8 000 000	9 000 000	13	109 082 746	71 749	8 390 980	46 350	0,36	1,00	3,58	0,75	1,00	11 474
9 000 000	10 000 000	14	132 275 132	78 289	9 448 224	50 575	0,32	1,00	3,64	0,73	1,00	13 474
10 000 000	11 000 000	12	125 586 720	114 145	10 465 560	73 737	0,29	1,00	3,70	0,72	1,00	20 726
11 000 000	12 000 000	8	91 759 371	70 348	11 469 921	45 445	0,26	1,00	3,76	0,71	1,00	13 314
12 000 000	13 000 000	8	99 775 543	54 947	12 471 943	35 496	0,24	1,00	3,82	0,70	1,00	10 746
13 000 000	14 000 000	5	67 249 884	56 325	13 449 977	36 386	0,22	1,00	3,88	0,69	1,00	11 302
14 000 000	15 000 000	7	101 276 805	72 455	14 468 115	46 806	0,21	1,00	3,91	0,68	1,00	15 021
15 000 000	16 000 000	5	77 391 219	62 024	15 478 244	40 068	0,19	1,00	3,92	0,67	1,00	13 315
16 000 000	17 000 000	6	98 080 819	82 614	16 346 803	53 369	0,18	1,00	3,93	0,66	1,00	18 217
17 000 000	18 000 000	1	17 643 600	18 774	17 643 600	12 128	0,17	1,00	3,95	0,65	1,00	4 290
18 000 000	19 000 000	6	110 922 235	69 647	18 487 039	44 992	0,16	1,00	3,96	0,64	1,00	16 249
20 000 000	21 000 000	2	40 774 538	19 469	20 387 269	12 577	0,15	1,00	3,99	0,62	1,00	4 733
21 000 000	22 000 000	2	43 534 561	23 914	21 767 280	15 448	0,14	1,00	4,00	0,61	1,00	5 966
22 000 000	23 000 000	4	90 502 770	67 945	22 625 692	43 892	0,13	1,00	4,01	0,61	1,00	17 200
24 000 000	25 000 000	2	49 993 760	60 131	24 996 880	38 845	0,12	1,00	4,05	0,59	1,00	15 776
25 000 000	30 000 000	40	1 047 178 364	511 367	26 179 459	330 343	0,11	1,00	4,06	0,59	1,00	136 270
30 000 000	35 000 000	8	256 688 340	173 922	32 086 042	112 353	0,09	1,00	4,14	0,56	1,00	49 306
35 000 000	40 000 000	8	299 653 258	196 315	37 456 657	126 819	0,08	1,00	4,21	0,54	1,00	57 923
40 000 000	45 000 000	12	512 799 452	412 227	42 733 288	266 299	0,07	1,00	4,28	0,53	1,00	125 256
45 000 000	50 000 000	8	380 739 255	510 977	47 592 407	330 091	0,06	1,00	4,34	0,52	1,00	158 537
50 000 000	1 000 000 000	28	2 238 965 552	2 987 059	79 963 055	1 929 640	0,04	0,63	4,63	0,47	0,94	902 058

FIGURE 40 – Méthode d'estimation de la prime pure d'un traité de réassurance par les courbes d'exposition appliquée à la cédante A

A partir du profil de risque de la cédante A et des hypothèses qui ont été faites sur le *Loss Ratio* et les courbes d'exposition, nous avons pu déterminer pour chaque bande la part de prime totale à céder au réassureur pour un 47M XS 3M par risque (colonne "Prime de réassurance" du tableau en figure 40). En sommant les primes de réassurance par bande on obtient l'estimation de la prime pure du traité :

$$\text{Prime pure} = 1,74 \text{M€}$$

Le tableau 16 permet de comparer les résultats obtenues à l'aide des différentes méthodes :

Modélisation sur expérience	Modélisation '0D'	Tarification sur exposition 'déterministe'
1,650	1,21M	1,74M

TABLE 16 – Prime pure du traité en fonction du modèle utilisé

La prime obtenue est encore une fois très supérieure à celle estimée à partir des modélisations sur expérience (qui sont quasiment nulles). Cela permet une fois de plus de montrer que les résultats des modèles sur expérience ne sont pas représentatifs du risque réel porté par la cédante A. La méthode de tarification sur exposition employée ici se reposant sur le profil de risque de la cédante, les individualités des polices de son portefeuille sont bien prises en compte.

Cependant, cette prime ne prend pas en compte les différentes clauses pouvant être incluses dans un traité en excédent de sinistre. Notamment, dans notre cas, le calcul tel qu'il est présenté correspond à la prime pure d'un 47M XS 3M avec reconstitutions illimitées. Dans son mémoire, BUTUN [2] explique comment prendre en compte les reconstitutions dans l'estimation par la méthode de tarification sur exposition de la prime pure d'un traité en excédent de sinistre. De plus, cette méthode est déterministe et ne permet pas d'obtenir des mesures de volatilité. Ces mesures sont pourtant primordiales d'une part pour mesurer l'efficacité des structures de réassurance et aider à la prise de décision, et d'autre part pour estimer une prime commerciale (prime pure avec un chargement sur la volatilité) du traité servant de référence aux commerciaux chargés de placer les traités sur le marché.

Dans le chapitre suivant, en nous appuyant sur [10] de HREVUŠ et MAREK, nous expliquerons comment utiliser les courbes d'exposition et profils de risque pour passer en vision stochastique et générer la sinistralité brute individuelle au-delà d'un seuil pour obtenir des mesures de volatilité et prendre en compte les différentes clauses pouvant être présentes dans les traités de réassurance. Nous l'appliquerons ensuite à la cédante A pour finalement comparer les résultats obtenus à ceux des modèles sur expérience et 0D.

CHAPITRE 9

Approche stochastique

Dans ce chapitre, la méthode décrite par HREVUŠ et MAREK dans *Exposure modelling in property reinsurance* [10] qui permet d'avoir une approche par simulation avec les courbes d'expositions et donc d'estimer la volatilité (ce qui n'est pas le cas avec la méthode précédemment décrite) est présentée. Notamment, la méthode sera appliquée au cas d'étude du mémoire, et les résultats comparés aux précédentes approches.

9.1 Méthodologie et choix des paramètres pour l'application à la cédante A

Pour procéder à ces simulations, une fois que les courbes d'exposition à utiliser pour chaque bande du profil de risque sont fixées, il est nécessaire de modéliser d'une part une fréquence N , et d'autre part les montants de sinistres X pour chaque bande h . L'idée est d'employer l'approche probabiliste décrite en première partie de ce mémoire et d'estimer les paramètres de fréquence et de coût à partir du profil de risque de la cédante et des courbes d'expositions correspondantes.

9.1.1 Estimation de la fréquence

Il existe plusieurs méthodes pour estimer la fréquence de sinistres au sein d'une bande h , deux d'entre elles vont être présentées :

Première méthode d'estimation de la fréquence

La première méthode s'appuie sur le modèle collectif et sur les résultats des courbes d'exposition présentés au début de ce mémoire. En effet, d'après le modèle collectif, on a :

$$\mathbb{E}[Z_h] = \mathbb{E}[N_h] \times \mathbb{E}[X_h]$$

avec N_h la fréquence de sinistres des risques de la bande h , et X_h la variable aléatoire du montant de sinistres.

Or, l'utilisation des méthodes de tarification à l'exposition reposent sur la connaissance à priori de l'espérance de sinistralité brute des bandes et de la distribution des taux de dommages des polices sinistrés (hypothèse de courbe d'exposition). De l'égalité précédente on obtient alors :

$$\mathbb{E}[N_h] = \frac{\mathbb{E}[Z_h]}{\mathbb{E}[X_h]} = \frac{\mathbb{E}[Z_h]}{ASI_h \mathbb{E}[x_h]} = \frac{\mathbb{E}[Z_h]}{ASI_h} G'(0)$$

car $\mathbb{E}[X_h] = ASI_h \mathbb{E}[x_h]$ et $\frac{1}{\mathbb{E}[x_h]} = G'(0)$

On peut également exprimer l'espérance du nombre de sinistres au-delà d'un seuil U comme :

$$\mathbb{E}\left[N_h | x > \frac{U}{ASI_h}\right] = \frac{\mathbb{E}[Z_h]}{ASI_h} G'\left(\frac{U}{ASI_h}\right) = \frac{\mathbb{E}[Z_h]}{ASI_h} G'\left(\frac{U}{ASI_h}\right)$$

Cette espérance peut alors être calculée pour chacune des bandes du profil de risques et on considérera que c'est le paramètre λ_h d'une loi de Poisson qui permettra de simuler des survenances de sinistre.

Deuxième méthode d'estimation de la fréquence

Une deuxième méthode d'estimation du nombre de sinistres au-delà d'un seuil U au sein d'une bande consiste à calculer la prime pure d'une 1 XS U pour la bande concernée à l'aide de la méthode d'estimation de prime pure via les courbes d'exposition présentée au chapitre précédent. En tarifant une tranche d'une portée de 1, la prime pure d'une 1 XS U donne la fréquence des sinistres supérieurs à U dans la bande.

Par exemple, si l'espérance du nombre de sinistres supérieurs à U dans une bande est de 2, alors la prime pure de la 1 XS U sera estimée à 2.

Exemple d'application de ces deux méthodes

Nous appliquons ces deux méthodes à la bande fictive suivante avec un seuil U fixé à 0 (on estime le nombre total de sinistres dans la bande) :

Somme assurée moyenne	Prime totale	Espérance de perte	Paramètre c de la courbe d'exposition
500,000	500,000	400,000	2

TABLE 17 – Bande fictive d'un profil de risque pour application de la méthode

Avec la première méthode, on a :

$$\mathbb{E}[N_h] = \frac{\mathbb{E}[Z_h]}{ASI_h} G'_2(0) = \frac{400,000}{500,000} G'_2(0)$$

$$\text{avec } G'_2(0) = \frac{1}{\mathbb{E}[x_h]} = \frac{1}{\int_0^1 (1 - F(x)) dx} = \frac{1}{22,61\%} = 4,423$$

$$\implies \mathbb{E}[N_h] = 3,538$$

Ensuite avec la deuxième méthode, on a :

$$\mathbb{E}[N_h] = \left(G_c\left(\frac{L}{ASI_h}\right) - G_c\left(\frac{M}{ASI_h}\right) \right) \times E[Z_h] = \left(G_2\left(\frac{1}{500,000}\right) - G_2\left(\frac{0}{500,000}\right) \right) \times 400,000 = 3,538$$

Les résultats sont bien identiques avec les deux méthodes : le nombre de sinistres par année de survenance dans cette bande est estimé à environ 3,54. Nous utiliserons par la suite uniquement la deuxième méthode avec un seuil U fixé à 3M (la priorité du traité étudié).

Remarque : en sommant les fréquences de sinistres estimées de toutes les bandes, on obtient la fréquence estimée au point d'attachement pour l'ensemble du profil de risque.

9.1.2 Estimation de la sévérité

Le lien entre courbe d'exposition et distribution des taux de dommages a été explicité au début de ce mémoire. Dès lors que nous associons une courbe d'exposition MBBEFD aux risques d'une bande h , nous postulons également de la distribution des taux de dommages des risques de cette bande. Les distributions des taux de dommages de chaque bande seront alors utilisées pour simuler individuellement les montants de sinistres.

On simulera un taux de dommage x_i à l'aide d'une distribution MBBEFD : le montant du sinistre correspondant sera donc $X_i = x_i \times ASI$ (soit le taux de dommage simulé multiplié par la somme assurée moyenne de la bande du profil à laquelle appartient le risque). Ensuite, le traité est appliqué sinistre par sinistre. Cette étape est répétée un nombre très important de fois.

Les résultats obtenus permettront d'estimer les récupérations moyenne (la prime brute de la tranche) mais également d'avoir des mesures de volatilité comme l'écart-type et les périodes de retour.

9.1.3 Paramètres utilisés

Les paramètres utilisés pour les simulations seront les mêmes que dans le chapitre précédent, à savoir pour chaque bande h du profil de risque :

$$\begin{cases} LR = 64,6\% \\ c_h \text{ déterminé par la méthode décrite au chapitre précédent (voir table 15)} \end{cases}$$

Nous modélisons une tranche 47M€ XS 3M€, donc nous voulons connaître la fréquence des sinistres au-delà de 3M par bande du profil de risque. Nous tarifons par conséquent une tranche 1€ XS 3M€ pour chaque bande. Ensuite, pour la sévérité nous utiliserons des distributions de taux de

dommages tronquées : en effet, les sinistres simulés sont supérieurs au seuil de 3M donc les taux de dommages simulés une fois multipliés par la somme assurée moyenne de la bande concernée doivent être supérieurs au seuil. Pour chaque bande h du profil de risque, les bornes inférieures et supérieures de la distribution tronquées seront :

$$\begin{cases} \text{LimInf}_h = m_h \\ \text{LimSup}_h = 1 \end{cases}$$

Le nombre de simulations N est quant à lui fixé à 200,000.

9.2 Résultats obtenus lors de l'application au cas pratique de la cédante A

Après simulations, les résultats obtenus sont présentés en figure 41 et comparés aux sorties des modèles précédents.

Récupérations	Modélisation sur expérience	Modélisation sur exposition '0D'	Modélisation sur exposition 'classique'
Moyenne	1 650	1 210 785	1 729 420
Ecart Type	30 052	3 749 010	5 973 444
Période de retour			
2 ans	0	0	0
5 ans	0	1 259 487	787 422
10 ans	0	3 306 025	3 797 610
25 ans	0	7 295 305	12 373 589
50 ans	0	11 690 107	23 179 459
100 ans	0	17 640 311	38 077 282
200 ans	55 835	25 741 071	47 000 000
500 ans	306 870	42 290 994	47 000 000
1 000 ans	499 087	47 000 000	47 000 000
TVaR 100 ans	164 968	29 452 290	45 199 720

FIGURE 41 – Résultats de modélisation du traité 47 XS 3 en fonction de l'approche utilisée

Une tendance se dégage de ces résultats. On remarque assez aisément que les récupérations moyennes estimées par les deux modèles sur exposition sont très largement supérieures à celles estimées grâce à une modélisation sur expérience : le risque semble donc bel et bien sous-estimé par la méthode sur expérience et l'utilisation de méthodes sur exposition était dans ce cas, comme attendu, pertinente. Elles nous permettent d'avoir une vision du risque effectif de la cédante, risque qui est également perçu par les réassureurs.

Comparons maintenant les résultats de nos deux modèles sur exposition. Pour rappel, le modèle 0D a été calibré sur une base marché se voulant la plus proche possible du risque de la cédante A et consolidée à l'aide de données de sinistralité internes. De l'autre côté, la modélisation grâce aux courbes d'exposition repose d'une part sur une hypothèse de Loss Ratio tirée d'une étude Swiss RE du marché de l'assurance américain, et d'autre part sur des courbes d'exposition marché calibrées

sur d'anciennes données du marché de l'assurance et des recommandations pour le paramètres c à utiliser. Ce modèle a l'avantage d'utiliser le profil de risque de la cédante et donc de prendre en compte ses individualités.

Les récupérations moyennes du traité sont estimées à 1,73M par le modèle sur exposition classique, avec un écart-type de presque 6M. La sinistralité modélisé (et donc les récupérations à la tranche) est bien plus sévère que celle obtenue à l'aide du modèle 0D. Cela se traduit par une augmentation de +43% dans l'estimation de la prime pure par rapport au modèle 0D. De même, la volatilité modélisée est bien plus importante : +59%. En comparaison à la prime commerciale payée pour le traité (3,76M) le risque semble surestimé. En effet, en estimant la prime commerciale grâce à un chargement sur écart-type toujours de 40%, nous avons :

$$PC = PP + Chargement * \sigma_{rec} = 1,72M + 5,97M \times 40\% = 4,12M$$

La prime commerciale que nous estimons pour la tranche est de 4,12M, soit une prime supérieure à la prime payée pour le traité global qui couvre également la conflagration et le risque terrorisme. La vision du risque que nous préconisons est donc la vision de la sinistralité vue au travers du modèle 0D. Bien que ne prenant pas en compte les individualités de la cédante A observables à l'aide du profil de risque, nous avons fait en sorte de limiter l'incertitude du modèle en utilisant une base marché la plus proche du risque de la cédante A et avec des données internes. A contrario, les paramètres et courbes qui ont été utilisés dans notre modélisation sur exposition 'classique' reposent uniquement sur des recommandations et études externes. L'incertitude est ici très forte car l'utilisation du modèle ne repose pas sur des hypothèses construites en internes et maîtrisées ; les résultats obtenus ne semblent pas cohérents. La maîtrise de l'incertitude de ce modèle n'est pas bonne, et une étude spécifique visant à déterminer des paramètres de modèle pertinents seraient nécessaire pour exploiter la méthode. Les limites de la méthode reposant sur les courbes d'exposition marché sont présentées dans le chapitre suivant.

CHAPITRE 10

Les limites des méthodes sur exposition classiques

Ces méthodes permettent de modéliser la sinistralité à partir des souscriptions de la cédante sur l'année de cotation considérée et donc de combler une insuffisance de données sinistres historiques. Elles présentent toutefois des limites qui sont d'ailleurs abordées dans les articles *Exposure modelling in property reinsurance* (HREVUŠ et MAREK, [10]) et *Exposure and experience rating for property per risk excess of loss reinsurance revisited* (DESMEDT et al., [5])

- Pour utiliser ces méthodes, nous utilisons une estimation du *Loss Ratio* qui est dans l'idéal transmise par la cédante, ou une valeur marché sinon. Les résultats de la modélisation vont grandement dépendre de ce paramètre qui n'est qu'une estimation. De plus, il n'est pas évident que le *loss ratio* soit le même dans toutes les bandes du profil de risque.
- Aussi, le choix des courbes d'exposition à utiliser est très subjectif. Le choix doit reposer sur une très bonne connaissance des risques sous-jacents, ce qui n'est pas toujours le cas.
- Les courbes d'exposition marché utilisées lorsque nous ne disposons pas de courbes calibrées sur des données internes reposent sur des données externes assez anciennes. Bien que ces courbes soient théoriquement en partie résistantes à l'inflation (dans la mesure où on considère que les sinistres et les sommes assurées évoluent de la même manière avec le temps), les produits d'assurances ont eux évolués.
- Également, les primes communiquées dans le profil de risque peuvent inclure des couvertures autres que l'incendie par risque. Néanmoins, ce niveau d'information n'est généralement pas transmis, ce qui contribue à rajouter de l'incertitude aux modélisations.
- Enfin, les profils de risques peuvent être communiqués au global, sans distinction de faite sur la nature des risques. Pourtant, la distribution des taux de dommages est largement dépendante de la LoB étudiée.

De plus, la principale limite de cette méthode dans le cadre de ce mémoire est qu'elle est avant tout axée sur la tarification : l'objectif est d'estimer une prime de réassurance. Or, ce n'est pas l'enjeu

principal de ce mémoire ; la possibilité d'obtenir des primes de réassurance à partir de nos modèles est une conséquence de la modélisation du brut. Notre objectif premier est de modéliser les sinistres individuels incendie au-delà d'un seuil en nous basant sur l'exposition de la cédante. L'estimation de primes de réassurance découle ensuite de l'application des traités aux sinistres générés par nos modèles. La logique derrière les méthodes de tarification sur exposition employées dans cette partie diffère : elles supposent que la sinistralité brute de la cédante est connue (hypothèse de *Loss Ratio*), et en associant des distributions de taux de dommages aux bandes du profil de risque, on détermine la part de cette sinistralité brute à charge du traité en moyenne. Cette méthode est donc intéressante lorsque la connaissance des risques sous-jacents et de la cédante est très bonne.

Afin d'illustrer la sensibilité des résultats obtenus en fonction des paramètres de modèle choisis (*Loss Ratio* du portefeuille et courbe d'exposition), nous avons calculé la prime pure du traité étudié avec différentes combinaisons de paramètres. La figure 42 présente les résultats obtenus.

	Loss Ratio		
	60%	70%	80%
SwissRe 1 (c=1,5)	2,60 M	3,04 M	3,47 M
SwissRe 2 (c=2)	2,48 M	2,90 M	3,31 M
SwissRe 3 (c=3)	2,14 M	2,49 M	2,85 M
SwissRe 4 (c=4)	1,71 M	2,00 M	2,28 M
Lloyd's curve (c=5)	1,30 M	1,52 M	1,73 M

FIGURE 42 – Estimation de la prime pure du traité en fonction des paramètres utilisés

En définitive, les méthodes de modélisation basées sur les courbes d'exposition dépendent largement du choix des paramètres utilisés (comme le montre la figure 42 ci-dessus), ces derniers englobant deux aspects fondamentaux.

D'un côté, la sélection du paramètre c de la courbe d'exposition revient à fixer la distribution des taux de dommages des risques au sein de la bande. Bien que des recommandations existent pour ce choix en fonction de la somme assurée étudiée, ces conseils proviennent de données externes qui peuvent différer de celles que nous analysons. Ainsi, le choix de ce paramètre conserve une certaine dimension arbitraire. Dans cette optique, nous proposons le développement d'un modèle permettant de définir la distribution des taux de dommages (ou directement de la charge ultime) des sinistres sans être contraint par ce paramètre. Notre approche consisterait à calibrer un modèle à l'aide d'une régression GAMLSS, où les paramètres de la loi seraient des fonctions de la somme assurée. Cette approche se traduirait essentiellement par une redéfinition des distributions de sévérité (en explorant d'autres lois que celles utilisées précédemment) sans découpage par tranche de coût, mais plutôt en adoptant une perspective continue et sur des données plus récentes que celles ayant servi à la calibration des courbes d'exposition marché.

D'un autre côté, le deuxième paramètre utilisé dans les méthodes de courbe d'exposition est le "Loss Ratio". En d'autres termes, l'une des hypothèses sous-jacentes des modèles est de fixer a priori la sinistralité brute des différentes bandes de risque étudiées. Par la suite, les courbes d'exposition nous permettent de convertir cette sinistralité brute en sinistralité à la charge de la réassurance, en tenant compte des traités, des sommes assurées analysées, et des distributions de taux de dommages correspondantes. Toutefois, l'objectif central de ce mémoire consiste précisément à estimer la sinistralité brute (au-delà d'un seuil) pour les cédantes qui ne disposent d'aucune donnée historique et qui, par conséquent, ne connaissent pas leur Loss Ratio.

Notre approche consisterait alors à calibrer un modèle de fréquence cette fois dont les paramètres seraient toujours fonction des sommes assurées. Nous disposerions alors d'un modèle de fréquence et d'un modèle de sévérité dont les paramètres des lois seraient des fonctions de la somme assurée. Le modèle s'inscrirait dans la continuité du modèle 0D proposé en partie 2 de ce mémoire et permettrait bien à partir de l'exposition de la cédante d'estimer son risque effectif et de modéliser sa sinistralité brute incendie au-delà d'un seuil. A partir de nos modélisation nous pourrions alors appliquer assez simplement différentes structures de réassurance afin d'en mesurer l'efficacité, d'en estimer le coût, etc...

Quatrième partie

Troisième approche de modélisation sur exposition : modèle 1D sur somme assurée

11 Méthodes de régressions pour la calibration	97
11.1 Les différents modèles	97
11.2 Critères de sélection des modèles	101
12 Méthodologie	104
12.1 Calibration des modèles	104
12.2 Utilisation des modèles	107
12.3 La question de l'inflation	108
13 Développement de l'outil R et validation des méthodes	110
13.1 Génération des données	110
13.2 Gestion des données	115
13.3 Calibration et validation des modèles	115
13.4 Simulations	118
14 Étude sur données réelles	120
14.1 Les données	120
14.2 Premier cas : sinistres transmis sans le détail de la répartition dommage/responsabilité	121
14.3 Deuxième cas : sommes assurées non disponibles	125
15 Limites du modèle	127

Intuition

Ce modèle a pour vocation de s'inscrire dans la continuité du modèle 0D en intégrant une dimension explicative à la sinistralité par l'utilisation des sommes assurées. L'objectif est d'utiliser des données de sinistralité et d'exposition marché dont nous disposons en interne pour calibrer des modèles de régression permettant de simuler la sinistralité incendie individuelle au-delà d'un seuil à partir des profils de risque ou des polices ligne à ligne. Les méthodes de tarification sur exposition 'classiques' présentées dans la partie précédente ont été une source inspiration dans la mise en place de ce modèle : l'objectif était d'utiliser l'information des sommes assurées qui peut être reliée à la sinistralité en incendie tout en s'affranchissant des hypothèses très fortes nécessaires à l'utilisation des méthodes classiques et pour lesquels nous ne disposons pas d'études internes pour fixer des paramètres pertinents et minimisant l'incertitude.

CHAPITRE 11

Méthodes de régressions pour la calibration

Les méthodes de régressions paramétriques permettent d'estimer une variable, dite variable à expliquer, à partir d'une ou plusieurs autres variables explicatives. On parle de régression paramétrique car on fait l'hypothèse a priori d'une relation explicite reliant la variable à expliquer aux variables explicatives via des paramètres à déterminer. Les modèles GAMLSS (Generalized Additive Models for Location, Scale and Shape) développées par [Rigby et Stasinopoulos, 2010] seront utilisés. L'objectif sera, pour chaque risque d'un portefeuille d'exposition, de lui associer une probabilité d'occurrence de sinistre et une loi de sévérité pour les sinistres. Les paramètres de ces lois seront fonctions de nos variables explicatives.

Afin d'illustrer les différentes méthodes de régression paramétrique qu'on utilise par la suite, nous partirons du cas le plus simple, la régression linéaire simple, pour parcourir ensuite des modèles de plus en plus complexe jusqu'à arriver au modèle GAMLSS.

11.1 Les différents modèles

Dans cette partie nous présenterons de manière synthétique les différents modèles de régression paramétriques rencontrés tout au long de l'étude. La variable Y désignera la variable à expliquer, représentée dans notre modèle par la charge ultime d'un sinistre ou son taux de dommage. Le taux de dommage d'un sinistre est défini comme suit :

$$DR = \frac{X}{SA}$$

avec

- X la charge d'un sinistre,
- SA la somme assurée de la police sinistrée.

Les variables X_1, \dots, X_n désigneront les variables explicatives utilisées. Dans notre cas, il s'agit à minima de la somme assurée mais d'autres variables explicatives pourraient être ajoutées a posteriori.

11.1.1 Régressions linéaires

Une régression linéaire formule l'hypothèse que la relation liant la variable réponse Y aux variables explicatives X_1, \dots, X_n est linéaire.

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p} + \epsilon$$

où

- $\epsilon \sim \mathcal{N}(0, \sigma)$;
- Y_i pour $i = 1, \dots, n$ les variables aléatoires à expliquer;
- $(x_{i,1}, \dots, x_{i,p})$ les valeurs des variables explicatives observées à partir d'un jeu de données de n lignes.

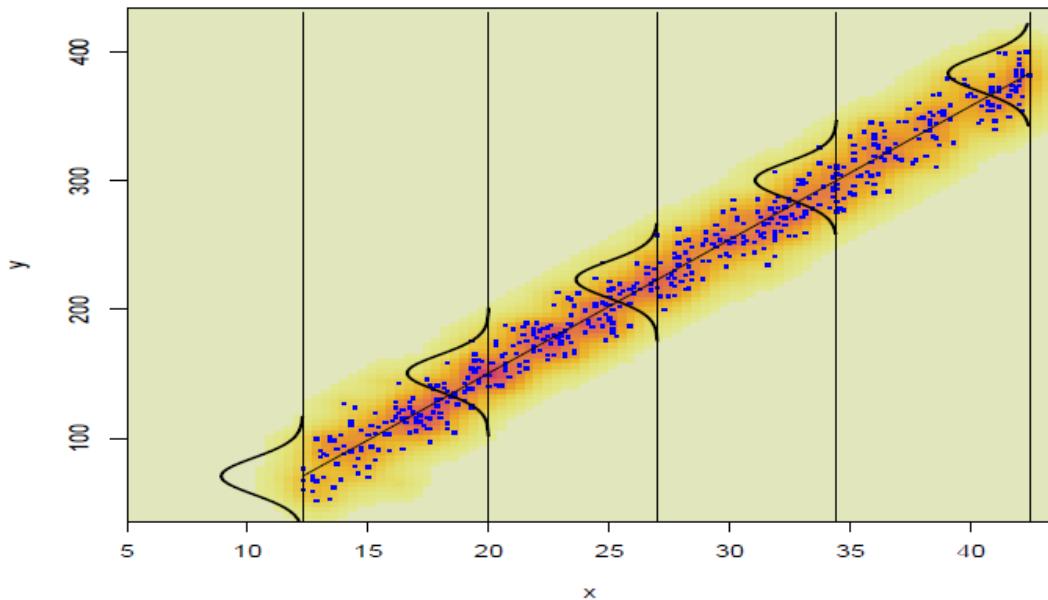


FIGURE 43 – Illustration graphique d'une régression linéaire

Ce modèle, très populaire dans le monde statistiques, est toutefois très limité. La relation entre Y et les X_1, \dots, X_n doit être strictement linéaire, et la variable prédictive Y suit supposément une loi normale de variance constante.

11.1.2 Modèles linéaires généralisés - GLM

Les modèles linéaires généralisés permettent de contourner en partie les limitations du modèle linéaire évoquées à l'instant. Il s'agit d'une généralisation plus souple des modèles linéaires.

- L'hypothèse de distribution normale de Y est remplacée par une hypothèse de famille de distribution exponentielle, notée EDF.
- La relation qui lie l'espérance de Y (μ) aux variables explicatives peut être non linéaire, via l'utilisation d'une fonction de lien $g(.)$.

Plus formellement,

$$Y_i \sim EDF(\mu_i, \phi)$$

$$g(\mu_i) = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_p x_{i,p}$$

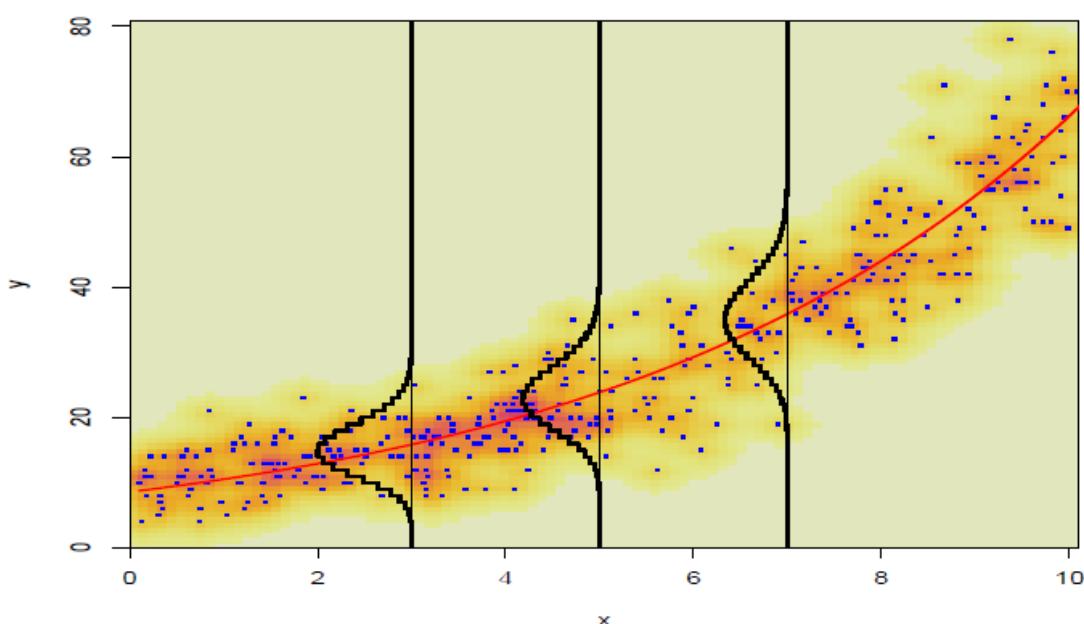


FIGURE 44 – Illustration graphique d'un GLM

La distribution de Y se limite aux distributions de la famille exponentielles (binomiale, poisson, normale, exponentielle, gamma, inverse-gaussienne, weibull etc...).

La prise en compte de la non linéarité se fait en déplaçant la contrainte de linéarité de μ sur $g(\mu)$. C'est pour dépasser cette contrainte que les modèles additifs généralisés (GAM) sont introduits.

Un cas particulier des GLM est la régression logistique. Elle est utilisée pour expliquer à partir de variables explicatives (x_1, \dots, x_k) une variable binaire. En d'autres termes, la variable à expliquer y est supposée suivre une loi de Bernoulli dont la probabilité p est dépendante des variables explicatives en entrée du modèle. Ce modèle de régression sera utilisé pour exprimer la probabilité d'occurrence d'un sinistre en fonction de la somme assurée des polices. L'idée est de ne plus segmenter les modèles en estimant la probabilité d'occurrence de sinistre par bande de somme assurée, mais de généraliser à l'aide d'une régression la relation qui existe entre occurrence de sinistre au-delà d'un seuil et somme assurée .

11.1.3 Modèles additifs généralisés - GAM

Les GAM, inspirés de la théorie des modèles additifs, viennent compléter les GLM par l'ajout de fonctions de lissage qui permettent la non-linéarité entre le prédicteur et les variables explicatives, i.e :

$$Y_i \sim EDF(\mu_i, \phi)$$

$$g(\mu_i) = \beta_0 + f_1(x_{i,1}) + \dots + f_p(x_{i,p})$$

où les f_1, \dots, f_p sont des fonctions de lissage

Les fonctions de lissage classiquement utilisées sont les splines cubiques. Pour un échantillon de points $\{X_i, Y_i\}$ donné, la construction de splines cubiques consistent à ajuster un polynôme de degré au plus 3 entre chaque points.

Les GAM permettent une plus grande liberté dans les modèles de régression.

11.1.4 Modèles additifs généralisés pour la position, l'échelle et la forme - GAMLSS

Introduits par [Rigby et Stasinopoulos, 2010], les GAMLSS permettent de s'affranchir des limitations dans le choix des distributions imposées par les GLM (famille exponentielle). De plus, tous les paramètres de la loi choisie peuvent être exprimés comme fonction des variables explicatives, avec possibilité d'utiliser des fonctions de lissage. Il s'agit donc d'un modèle semi-paramétrique.

Plus formellement,

$$Y_i \sim D(\theta^i) \text{ où } \theta^i = (\mu_i, \sigma_i, \nu_i, \tau_i)$$

$$g_k(\theta_k) = \beta_{k,0} + f_{k,1}(X_1) + \dots + f_{k,n}(X_n)$$

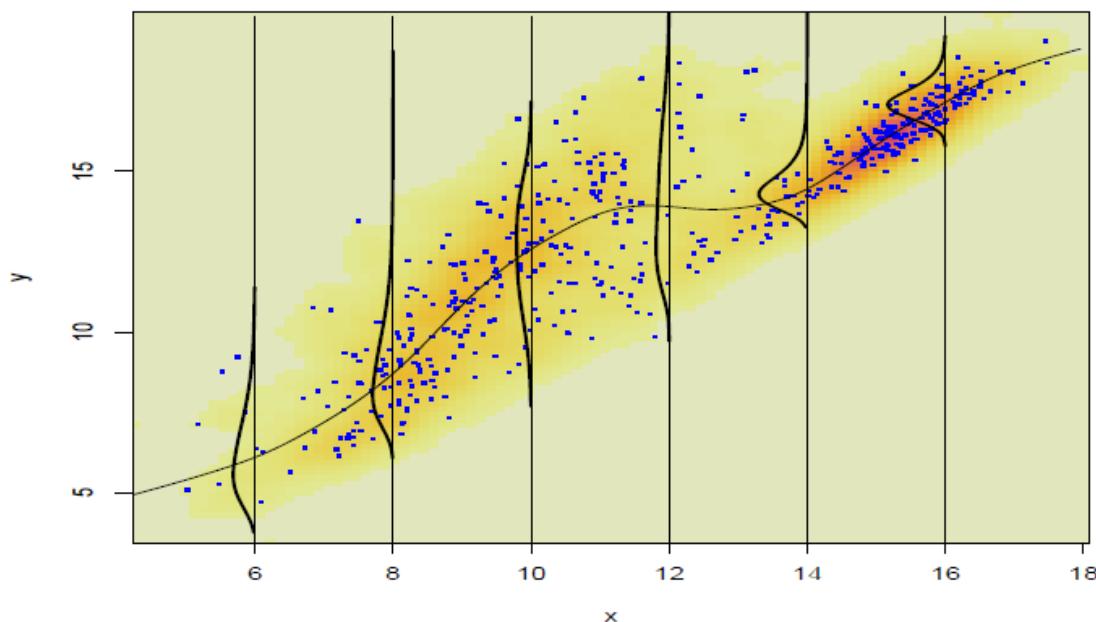


FIGURE 45 – Illustration graphique d'un GAMLSS

Tous les modèles présentés en amont peuvent être retrouvés dans la formulation d'un GAMLSS. En effet, on a l'arborescence suivante :

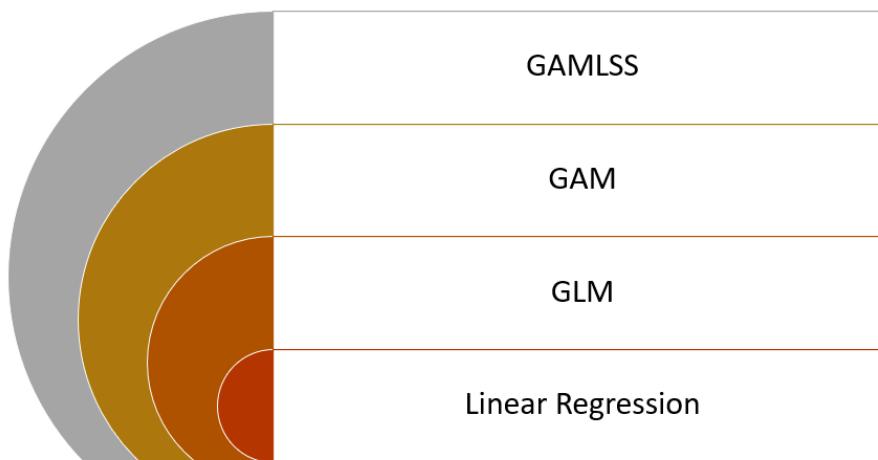


FIGURE 46 – Schéma des relations entre les différentes régressions paramétriques

11.2 Critères de sélection des modèles

Un large choix de modèles s'offre à nous. Il va donc falloir définir des critères de sélection, afin de sélectionner le modèle qui se généralise le mieux.

Critère d'information d'Akaike - AIC

Dans un premier temps, nous utiliserons un indicateur très couramment utilisé en statistiques : le critère d'information d'Akaike (AIC). L'idée derrière ce critère est de trouver un compromis entre maximisation de la vraisemblance du modèle, et nombre de paramètres à estimer. Il se définit

comme suit :

$$AIC = 2k - 2\ln(L)$$

avec

- k le nombre de paramètres à estimer
- L la vraisemblance du modèle

Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un paramètre. Le critère d'information d'Akaike permet de pénaliser les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible. La figure 47 présente à titre illustratif l'impact de l'ajout de variables explicatives dans le modèle sur l'AIC.

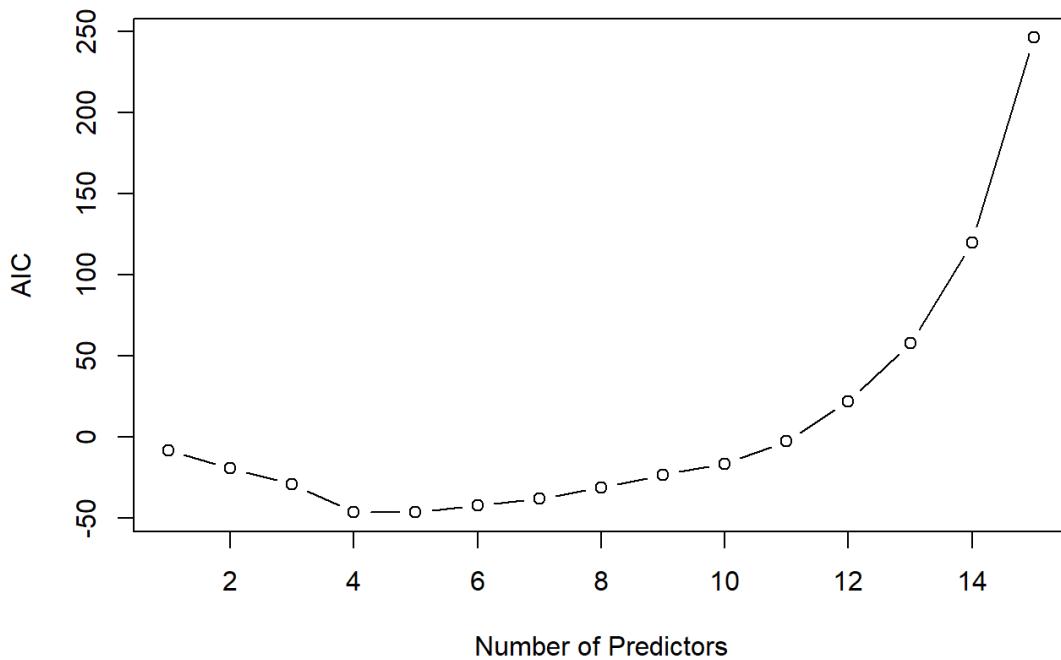


FIGURE 47 – AIC en fonction des degrés de liberté d'un modèle

On remarque bien sur la figure qu'à partir d'un certain nombre de prédicteurs (4), l'ajout de variables explicatives n'améliore plus l'AIC du modèle, au contraire.

Root Mean Squared Error - RMSE

Dans un second temps, nous nous intéresserons au RMSE du modèle, qui s'interprète comme l'erreur de prédiction moyenne du modèle. On mesure la qualité de prédiction du modèle en s'intéressant aux résidus induits par cette prédiction. Cet indicateur permet d'avoir une mesure synthétique de la qualité d'un modèle.

Formellement, il est donné par la formule suivante :

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

Nous lui accorderons toutefois moins d'importance que le critère AIC, car le RMSE n'atteste pas de la capacité de généralisation des modèles : il aura tendance à favoriser l'overfitting.

Critère R²

Enfin, nous utiliserons le R^2 de nos modèles. Il est défini par :

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Cette statistique, comprise entre 0 et 1 peut s'interpréter comme le pourcentage de variance expliquée par notre modèle par rapport à la variance totale du modèle, soit :

$$R^2 = \frac{\text{Variance expliquée par le modèle}}{\text{Variance totale}}$$

Néanmoins, par construction, le R^2 aura tendance à s'améliorer avec l'ajout de variables explicatives et de degrés de liberté. Pour autant, le modèle n'en sera pas forcément meilleur. Encore une fois, cet indicateur favorisera l'overfitting. Il est donc à considérer conjointement à tous les autres indicateurs définis précédemment.

CHAPITRE 12

Méthodologie

Dans ce chapitre, la méthodologie qui a été pensée pour utiliser l'information des sommes assurées est présentée. Dans un premier temps, la méthode de calibration des modèles est expliquée, et dans un second temps l'utilisation des modèles sur les données d'exposition d'une cédante.

12.1 Calibration des modèles

Tout d'abord, les modèles de fréquence et de coût doivent être calibrés.

12.1.1 Construction d'une base d'apprentissage

Avant toute calibration d'un modèle, il est primordiale de constituer une base de données propre, homogène et représentative de la modélisation prévue. C'est une étape chronophage mais déterminante dans la qualité de la calibration du modèle.

Dans le cas de cette approche, la base souhaitée est la jointure entre la base sinistre des cédantes utilisées et leurs portefeuilles d'exposition. Les portefeuilles d'exposition présentent ligne à ligne les polices de la cédante avec notamment : les caractéristiques du contrats, les caractéristiques du risque assuré (branche, garantie, produit, ...), les sommes assurées (en distinguant le bâti, le contenu et la perte d'exploitation pour les risques commerciaux ou industriels). Les bases d'exposition sont des photos à l'instant t du portefeuille d'une cédante. La prise en compte de la date d'extraction des bases sera donc à prendre en compte dans la jointure. L'année d'extraction de la base doit correspondre à la survenance des sinistres. La figure 48 ci-dessous illustre la jointure à effectuer.

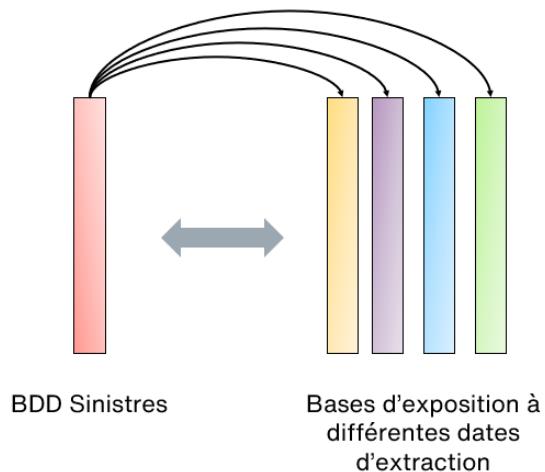


FIGURE 48 – Jointure de la base sinistres aux bases d'exposition ligne à ligne en fonction de l'année de survenance des sinistres et de la date d'extraction des portefeuilles

Les sinistres utilisés, issus de différentes cédantes, devront présenter des profils similaires.

12.1.2 Calibration de la fréquence

Une fois la base d'apprentissage consolidée, la calibration des modèles peut commencer. Dans un premier temps, la fréquence des sinistres est estimée. En théorie, on devrait utiliser un modèle de régression pour la fréquence en utilisant les sommes assurées comme variable explicative. En pratique, pour l'instant, on considérera une fréquence homogène pour l'ensemble du portefeuille d'exposition.

Comme dans l'approche classique, et le modèle 0D, on calcule dans un premier temps le nombre moyen de sinistres par année, corrigé de l'exposition et des tardifs.

$$\text{pour rappel, } \hat{\mu} = \sum_{i=SurvMin}^n \omega_i \times Nfinal_i$$

Cette fois, on ne veut pas simuler le nombre de sinistres, mais la probabilité d'occurrence d'un sinistre grave pour chaque assuré. La fréquence doit être définie individuellement pour chaque assuré afin d'utiliser les caractéristiques individuelles de la police sinistrée pour déterminer la sévérité du sinistre. On détermine alors la probabilité de sinistre \hat{p} :

$$\hat{p} = \frac{\hat{\mu}}{\text{Exposition}}$$

Cela revient à estimer le paramètre d'une loi de Bernoulli $\mathcal{B}(\hat{p})$. La fréquence d'apparition d'un sinistre par assurée étant bien définie, il est par la suite nécessaire de calibrer une sévérité dépendante des caractéristiques des polices sinistrées.

12.1.3 Calibration de la sévérité

La modélisation de la sévérité se base sur les modèles semi-paramétriques GAMLSS présentés au chapitre précédent, permettant d'ajuster sur les données une loi dont les paramètres sont fonctions de variables explicatives. Comme expliqué précédemment, la première variable utilisée dans le modèle est la somme assurée. Dans le cas où les cédantes transmettent également leur portefeuille ligne à ligne, il est tout à fait possible d'envisager utiliser d'autres variables comme le type de bâtiment, ou le statut juridique de l'assuré. Une étude des données avant calibration permettrait de sélectionner les variables potentiellement intéressantes.

En fonction des données utilisées, la Charge ultime ou les Taux de dommages des sinistres seront modélisés. Les GAMLSS permettent d'utiliser une grande variété de lois aux supports différents.

Si la charge ultime est modélisée, le choix d'une loi pourra être fait parmi les lois à support $[0, +\infty]$ suivantes :

Loi	Nombre de paramètres
Box-Cox Cole and Green	3
Box-Cox Power Exponential	4
Box-Cox-t	4
Exponential	1
Gamma	2
Generalized Beta type 2	4
Generalized Gamma	3
Generalized Inverse Gaussian	3
Inverse Gaussian	2
Log Normal	2
Log Normal family	2 (+1 fixé)
Reverse Generalized Extreme	3
Weibull	2

TABLE 18 – Lois à support $[0, +\infty]$

Si le taux de dommage est modélisé, les lois à support $[0, 1]$ suivantes pourront être choisies :

Loi	Nombre de paramètres
Beta	2
Beta original	2
Generalized beta type 1	4

TABLE 19 – Lois à support $[0, 1]$

Parmi tous les modèles testés, celui présentant l'AIC minimal sera choisi : c'est notre critère de

sélection du modèle. Les autres indicateurs présentés au chapitre précédent seront utilisés pour estimer à titre indicatif la qualité du modèle.

12.2 Utilisation des modèles

Une fois les 3 étapes précédentes réalisées, nous pouvons maintenant passer à l'utilisation des modèles : il ne faut pas perdre de vue que toutes ces étapes ont pour objectif de simuler la sinistralité d'une cédante à partir, dans le meilleur des cas, de son portefeuille d'exposition, ou de son profil de risque sinon. Les simulations ne se feront pas exactement de la même manière en fonction du cas.

12.2.1 Premier cas

Dans ce premier cas, considérons une cédante nous fournissant son portefeuille d'exposition afin que nous réalisions une étude de ses risques incendie.

1. Calcul des sommes assurées

La première étape consiste à évaluer les sommes assurées des polices d'assurances présentes dans la base. En effet, notre modèle de sévérité se base sur cette variable pour déterminer la loi de distribution des taux de dommages de chaque assuré. Pour ce faire, la méthodologie interne se reposant notamment sur le nombre de pièce et la surface des biens, et sur l'indice du Callon, est employée. Ce travail est réalisé par les équipes de *Cat Modelling*. Ainsi, le modèle de régression sur somme assurée calibré précédemment est utilisable sur cette base.

2. Génération des sinistres

La deuxième étape consiste, pour chaque ligne du portefeuille, à déterminer l'occurrence ou non d'un sinistre. Pour cela, on utilise la probabilité d'occurrence \hat{p} déterminée précédemment. Pour chaque police on simule une Bernoulli $\mathcal{B}(\hat{p})$. Dès lors, les taux de dommages des sinistres générés doivent eux aussi être simulés.

3. Génération de la gravité des sinistres

Il faut déterminer la sévérité des sinistres ayant été générés. Pour cela, le modèle de gravité de sinistres calibré précédemment est utilisé. Pour chaque assuré sinistré, on fait un tirage de la loi choisie, et dont les paramètres dépendront de la somme assurée de l'assuré (ou tout autre variable explicative pertinente et retenue lors de l'étape de la calibration). En fonction du modèle, la charge ultime ou le taux de dommages est simulé.

Ces étapes sont répétées un grand nombre de fois et les sinistres générés sont retenus dans une base de données appelée *Event Losses Table* (ELT). Cette base est ensuite directement utilisable dans ReMetrica pour mesurer la sinistralité, calculer les périodes de retour, l'écart-type, l'efficience de la réassurance, etc...

12.2.2 Deuxième cas

Pour ce deuxième cas, la cédante transmet uniquement son profil de risque (en distinguant les typologies de risque : résidentiel, industriel, etc...). Il a fallu développer une méthode pour appliquer ce modèle police par police à des profils de risque. L'idée est de simuler un portefeuille ligne à ligne ayant la même distribution que la cédante à partir de son profil risque.

Méthode de simulation d'un portefeuille ligne à ligne à partir du profil de risques

a. Fonction de répartition empirique de la distribution des sommes assurées du portefeuille

A partir du profil de risque de la cédante, la première étape consiste à approximer une fonction de répartition des sommes assurées du portefeuille étudié. Pour chaque bande du profil, une hypothèse de répartition linéaire des sommes assurées au sein de la bande est émise. En fonction du nombre de risques par bande, des points sont approximés par interpolation linéaire. En appliquant cette méthode à chaque bande du profil, on obtient autant de points que de polices dans le portefeuille de la cédante.

b. Inversion de la fonction de répartition

Afin d'appliquer la méthode de la transformée inverse, il faut inverser la fonction de répartition précédemment calculée. Cette inversion est réalisée à l'aide de la fonction *inverse* du package *GofKernel*.

c. Méthode de la transformée inverse

La méthode de la transformée inverse permet de simuler une variable aléatoire X à partir de sa fonction de répartition F . Pour modéliser un portefeuille d'exposition à partir du profil de risque de la cédante, on considère que SA , la somme assurée d'une police du portefeuille, est une variable aléatoire de fonction de répartition F déterminée précédemment, continue et strictement croissante. Alors si $U \sim \mathcal{U}(0, 1)$, $F^{-1}(U)$ a la même loi que SA .

Finalement, si n est la taille du portefeuille, n variables aléatoires de loi uniforme sur $[0, 1]$ (U_1, \dots, U_n) sont simulées et les $(F^{-1}(U_1), \dots, F^{-1}(U_n))$ obtenues représentent les sommes assurées des polices du portefeuille d'exposition.

Ensuite, les étapes 2 (Génération des sinistres) et 3 (Génération de la gravité des sinistres) du premier cas sont appliquées au portefeuille ainsi simulé.

12.3 La question de l'inflation

Dans le cadre de ce modèle, deux calibrations distinctes sont effectuées. D'un côté nous calibrons un modèle qui va associer à une somme assurée donnée la probabilité d'occurrence d'un sinistre. De l'autre côté, nous estimons à l'aide d'une régression GAMLSS la distribution des taux de dommages toujours en fonction de la somme assurée de la police sinistrée.

Dans les deux cas, la question de l'influence de l'inflation sur les modèles doit se poser. Les paramètres des lois de fréquence et sévérité calibrées dépendent de la somme assurée. De plus, les coefficients de régression ne sont pas forcément connus (notamment dans le cas de sévérité), les régressions GAMLSS permettant l'utilisation de fonctions de lissage non-paramétriques pour améliorer les résultats de modèle. Alors, une simple adaptation manuelle des lois utilisées en prenant en compte l'inflation au travers l'indexation des sommes assurées ne fonctionneraient pas. Toutefois, les modèles sauraient être utilisées malgré l'inflation. L'idée serait d'indexer (ou désindexer en l'occurrence) les sommes assurées pour les avoir en vision année de calibration du modèle. A partir de ces données désinflatées, les modèles pourraient être utilisés pour générer des taux de dommages; la conversion des taux de dommages en charge de sinistre se ferait ensuite avec la somme assurée originelle. Sinon, les courbes pourraient également être mises à jour chaque année avec des données de sinistralité actualisées afin de prendre en compte l'inflation.

CHAPITRE 13

Développement de l'outil R et validation des méthodes

Pour appliquer la méthodologie expliquée juste au-dessus, des outils ont été créés sous R. Pour les construire, des données fictives ont été utilisées : l'utilisation de données fictives a permis de contrôler les sorties et vérifier que les outils étaient fonctionnels et efficaces. De plus, cela a permis de préparer le terrain avant utilisation sur données réelles. Ainsi, cette section a été mise en place dans un objectif de déploiement d'un outil au sein de l'entreprise, et de validation des méthodes utilisées.

13.1 Génération des données

Dans un premier temps, nous allons générer des données à partir desquels les outils seront développés. Nous appliquerons alors la méthodologie développée et contrôleront les résultats en sortie. Les données seront générées aléatoirement, mais à partir d'une relation entre la variable à expliquer, les taux de dommages, et la variable explicative, la somme assurée. Ainsi, les données seront contrôlées et fiabilisées, et permettront de tester à la fois la robustesse des outils et méthodes utilisées et les outils développés sous R en vue d'une utilisation sur données réelles. En contrôlant les données d'entrées des modèles, nous pourront juger de la qualité des données en sortie.

13.1.1 Génération d'un portefeuille d'exposition

Tout d'abord, un portefeuille d'exposition est généré à partir du profil de risque présenté au tableau 20 :

Borne Inférieure	Borne Supérieure	Nombre de risques
5,000	200,000	20,000
200,000	500,000	12,000
500,000	1,000,000	6,000
1,000,000	2,000,000	2,000

TABLE 20 – Profil de risque fictif utilisé pour la génération du portefeuille d'exposition

En prenant uniquement en compte les bornes des bandes et le nombre de risques de chaque bande, la fonction de répartition empirique de ce portefeuille est présentée dans la figure 49 :

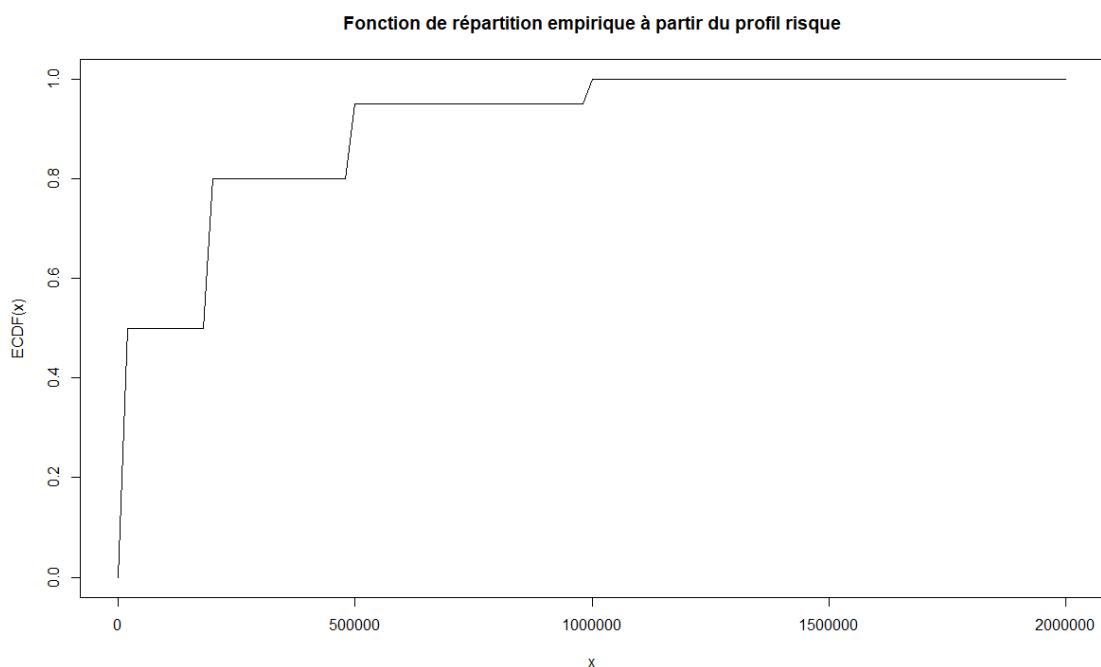


FIGURE 49 – Fonction de répartition empirique des valeurs assurées

Bien sûr, cela n'a pas vraiment de sens, et il est primordial d'estimer la répartition des sommes assurées de chaque tranche. L'hypothèse d'une répartition linéaire des sommes assurées au sein d'une même bande est émise, et nous prenons en compte le nombre de risque pour chaque bande. Nous aurions aussi pu, comme dans la méthode de tarification sur exposition, faire l'hypothèse que toutes les polices ont même somme assurée au sein d'une même bande (la somme assurée moyenne de la bande). Ici, par interpolation linéaire, la fonction de répartition suivante est obtenue :

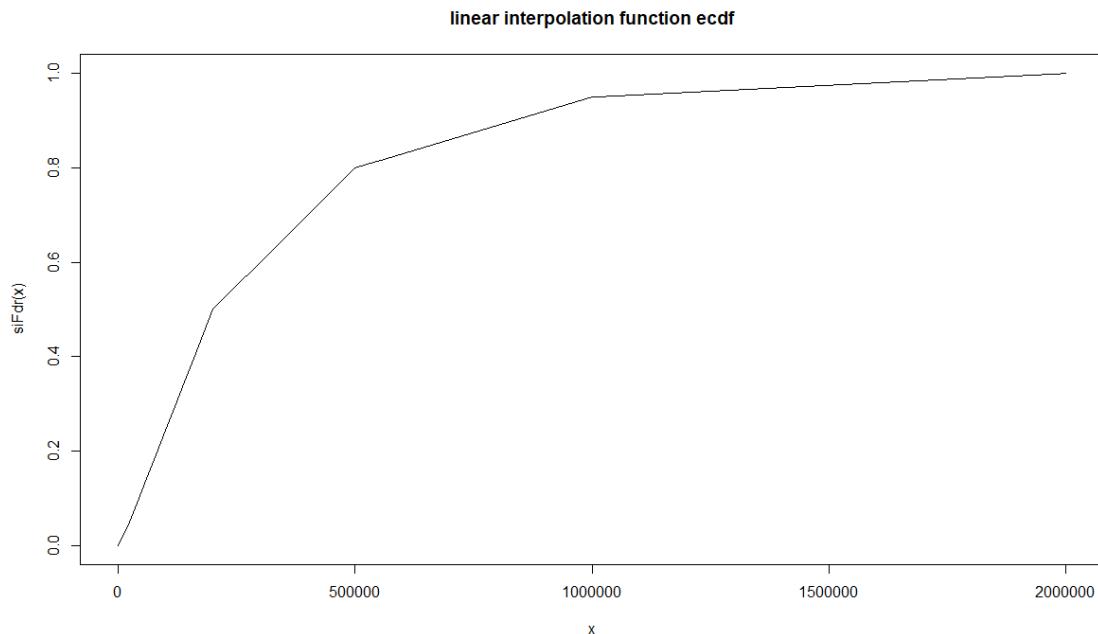


FIGURE 50 – Fonction de répartition empirique des sommes assurées avec interpolation linéaire

Ensuite, l'inverse de cette fonction de répartition des sommes assurées du portefeuille, obtenue à partir du profil de risque, est déterminée. Pour cela, c'est la fonction *inverse* qui est utilisée. La simulation d'un portefeuille à la répartition des sommes assurées similaire à celui étudié est maintenant possible.

Le portefeuille comporte 40,000 risques, donc on génère 40,000 nombres aléatoires entre 0 et 1 à l'aide d'une loi uniforme.

$$U = \begin{pmatrix} u_1 \\ \vdots \\ u_{40,000} \end{pmatrix}, U \sim \mathcal{U}([0, 1]^{40,000})$$

Puis, par la méthode de la transformée inverse, $F^{-1}(U)$ donne un portefeuille d'exposition telle que $F^{-1}(U_1, \dots, U_{40,000})$ ont même loi que SA .

Le portefeuille obtenu est le suivant :

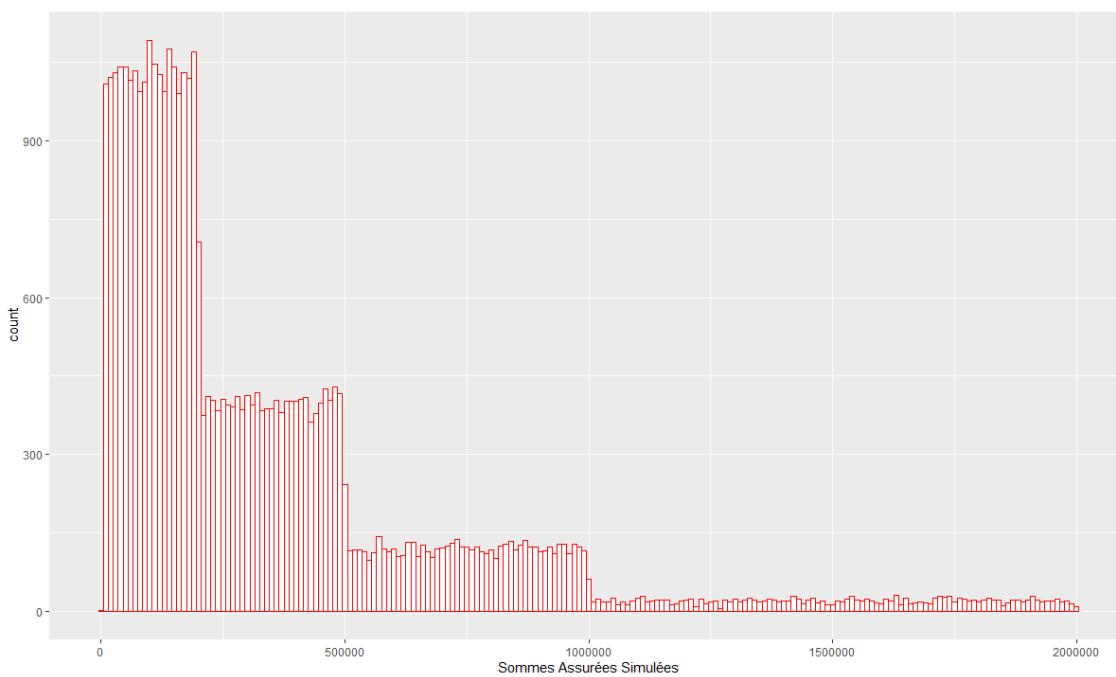


FIGURE 51 – Histogramme des sommes assurées simulées obtenues

La méthode mise en place a bien permis de générer un portefeuille d'exposition à partir d'un profil de risque (fictif dans ce cas). L'histogramme présenté représente bien la physionomie du profil de risque utilisé. Le résultat semble cohérent, la méthode de simulation d'un portefeuille ligne à ligne à partir d'un profil risque peut donc être validée (sous réserve d'une grosse hypothèse de répartition linéaire des sommes assurées au sein de chaque bande). Aussi, cette partie a permis de mettre en place les outils.

13.1.2 Génération des sinistres

A partir de ce portefeuille simulé, nous allons générer des sinistres qui serviront de base d'apprentissage à nos modèles. 10 années de sinistralité indépendantes seront simulées.

- Pour la fréquence :

Une simulation très simple est choisie. Une probabilité d'occurrence p d'un sinistre est fixée. Cette probabilité est la même pour toutes les polices du portefeuille. Pour chaque année de sinistralité que l'on souhaite simuler, on génère un vecteur aléatoire de dimension (1xN) dont chaque composante est une réalisation d'une Bernoulli $\mathcal{B}(p)$.

Ici, $p = 0.01$ et $N = 40,000$

- Pour la sévérité :

Pour ces données fictives, la décision a été prise de simuler les taux de dommages des sinistres. Pour cela, le choix s'est porté sur une loi Bêta. L'objectif est de contrôler la génération

des sinistres, et la relation qui lie la sévérité aux sommes assurées. Les paramètres de la loi sont définis comme des fonctions de la somme assurée.

Soient

$$\mu = 0,7 - 0,3 \times NormSA$$

$$\sigma^2 = 0,0005 + 0,005 \times NormSA$$

où $NormSA$ est la somme assurée normalisée, $NormSA = \frac{SA - SA_{min}}{SA_{max} - SA_{min}}$

A partir de cette moyenne et cette variance fixées, il faut déterminer les paramètres α et β qui satisfassent ces équations.

On pose $x = NormSA$ et on résout :

$$\begin{cases} 0,7 - 0,3x = \frac{\alpha}{\alpha+\beta} \\ 0,0005 + 0,005x = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{cases} \iff \begin{cases} \alpha = \frac{540x^3 - 1950x^2 + 353x + 2933}{100x+10} \\ \beta = 3 \frac{-180x^3 + 50x^2 + 649x + 419}{100x+10} \end{cases}$$

Avec ces paramètres, le taux de dommage des sinistres sera en moyenne décroissant par rapport à la somme assurée, tandis que la variance sera croissante par rapport à la somme assurée. 10 années de sinistralité sont simulées, et les résultats sont stockés dans une base qui sera utilisée pour entraîner le modèle.

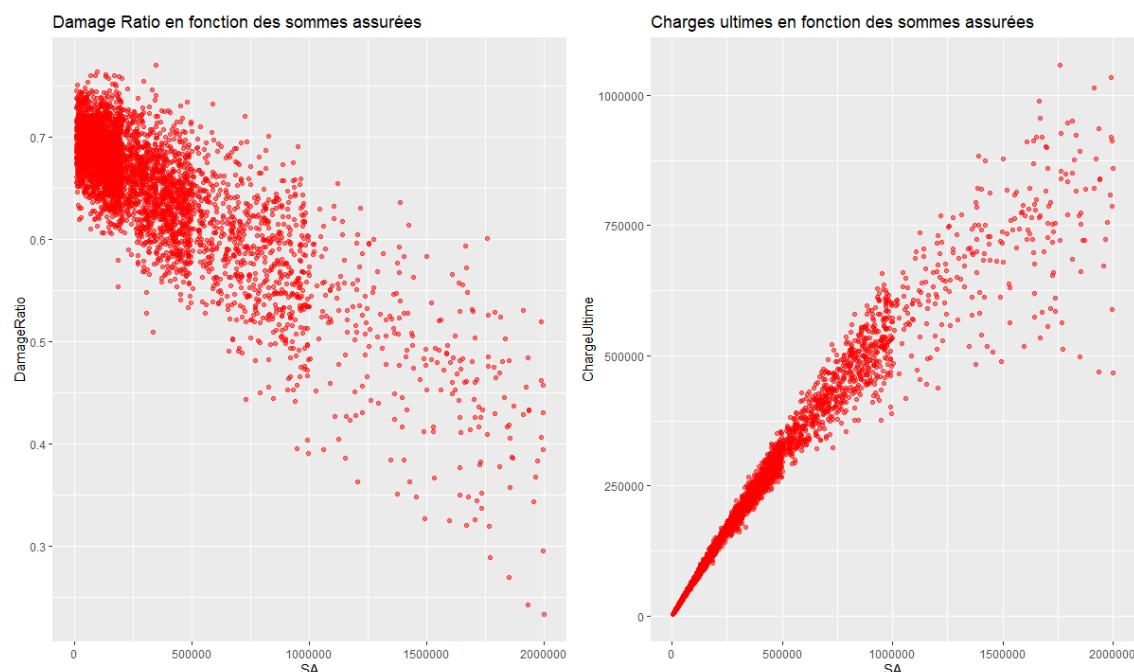


FIGURE 52 – Sinistres simulés

La figure 52 permet d'observer la décroissance des taux de dommages et l'augmentation de la variance en fonction de la somme assurée.

13.2 Gestion des données

Dans cette section, les données utilisées sont fictives, donc il n'y a pas de retraitements à effectuer.

Dans le cadre de données réelles, l'étape de création de la base d'entraînement n'est pas automatisable. En effet, les données utilisées sont issues de différentes cédantes, et de différentes années. Les retraitements doivent être manuels et dépendent de chaque cédante. Les formats des bases, noms de colonnes, années disponibles ne sont pas les mêmes entre chaque client et d'année en année.

Les traitements préliminaires à effectuer sur les données sont ensuite les mêmes que pour le modèle 0D. Avant toute modélisation, les sinistres doivent être indexés et projetés à l'ultime. Les outils sont donc déjà mis en place.

13.3 Calibration et validation des modèles

Dans cette section, les outils de calibration et validation des modèles seront construits à l'aide de la base de données sinistres simulées.

13.3.1 Calibration de la fréquence

L'estimation d'un paramètre de fréquence ne nécessite pas une refonte des outils déjà mis en place. En effet, nous pouvons simplement déterminer le nombre moyen de sinistres (corrigé de l'exposition) par année de survenance grâce aux méthodes déjà implémentées par le modèle 0D. Puis, il suffit de calculer le ratio de ce $\hat{\mu}$ par l'exposition de l'année que l'on souhaite simuler pour passer d'un nombre moyen de sinistres à une probabilité d'occurrence \hat{p} .

Dans ce cas fictif, on obtient : $\hat{p} = 0.009995$. Or, les sinistres ont été générés avec une proba $p = 0.01$. L'approximation est excellente, la méthode peut donc être validée.

13.3.2 Calibration de la sévérité

Un modèle de sévérité doit maintenant être calibré. Dans un premier temps, il faut calibrer tous les modèles envisageables, puis choisir le plus pertinent d'après les critères statistiques envisagés. Ensuite, il faut valider ce modèle (vérifier qu'il soit théoriquement viable).

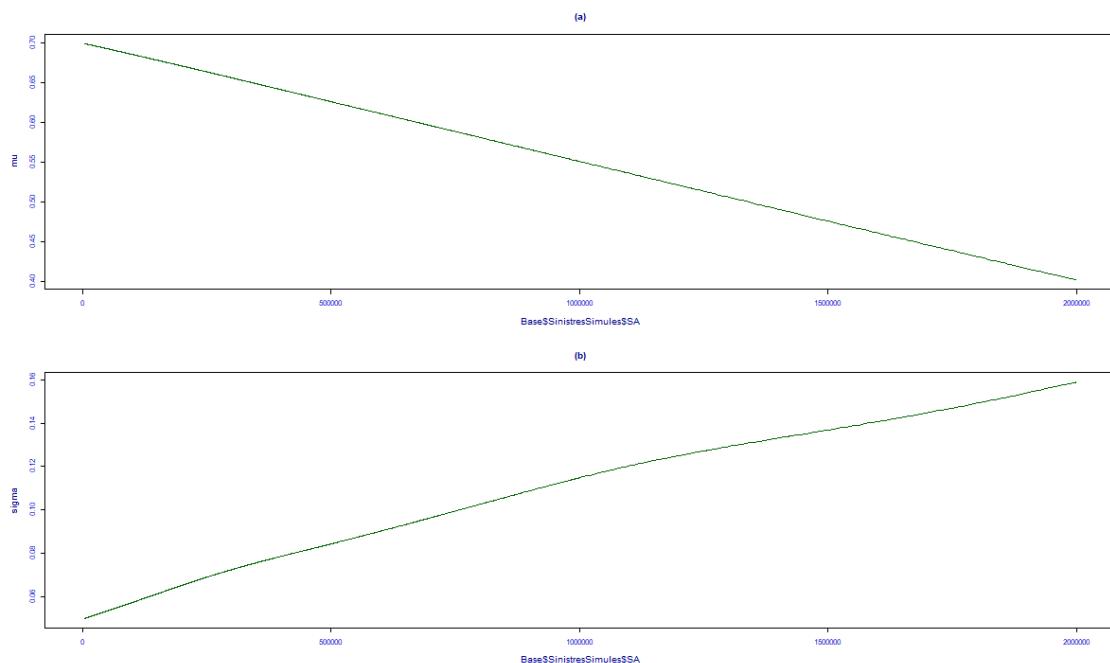
Pour la calibration, un script assez flexible permettant à l'utilisateur d'ajuster tous les modèles choisis est développé. Dans le cas de cet exemple sur données fictives, les lois à support $[0, 1]$ sont testées. La variable explicative utilisée est la Somme Assurée, et des fonctions de lissage sont utilisées, ce qui en font des modèles semi-paramétriques.

Le tableau ci-dessous présente les résultats obtenus en terme d'AIC :

Loi	AIC
Beta	-15698.30
Beta original	-15689.71
Generalized beta type 1	-12626.61

TABLE 21 – AIC des modèles testés

Le modèle Beta présente l'AIC minimal, c'est donc lui qui est choisi. Le résultat obtenu est bien le résultat que l'on attendait. En effet, les sinistres ont été générés avec une Beta dont les paramètres sont fonction de la Somme Assurée. Cet effet est facilement observable lorsqu'on trace les fonctions de l'espérance et de la variance estimées par le modèle en fonction de la somme assurée :

**FIGURE 53** – Espérance et Variance en fonction de la somme assurée

Vérifions maintenant que le modèle soit valide. La théorie dit que les résidus d'un modèle de régression suivent une loi normale centrée réduite. Vérifions graphiquement cette hypothèse :

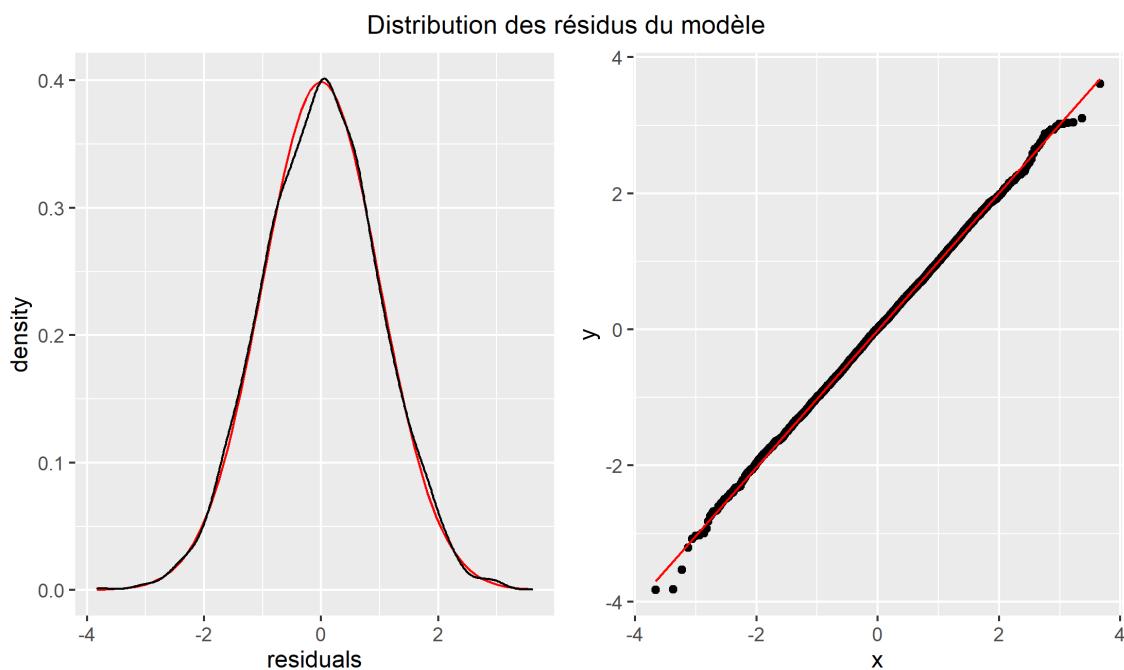


FIGURE 54 – Distribution des résidus

Sur le graphique de gauche, on observe en rouge la densité d'une normale centrée réduite, et en noir la densité des résidus du modèle. Les résidus suivent bien le tracé de la gaussienne. Le graphique de droite est un qq-plot. Les résidus ne s'éloignent de la gaussienne qu'au niveau des quantiles extrêmes, et cela reste léger. D'après ces graphiques, les résidus sont bien des variables aléatoires $\epsilon \sim \mathcal{N}(0, 1)$, et le modèle est valide.

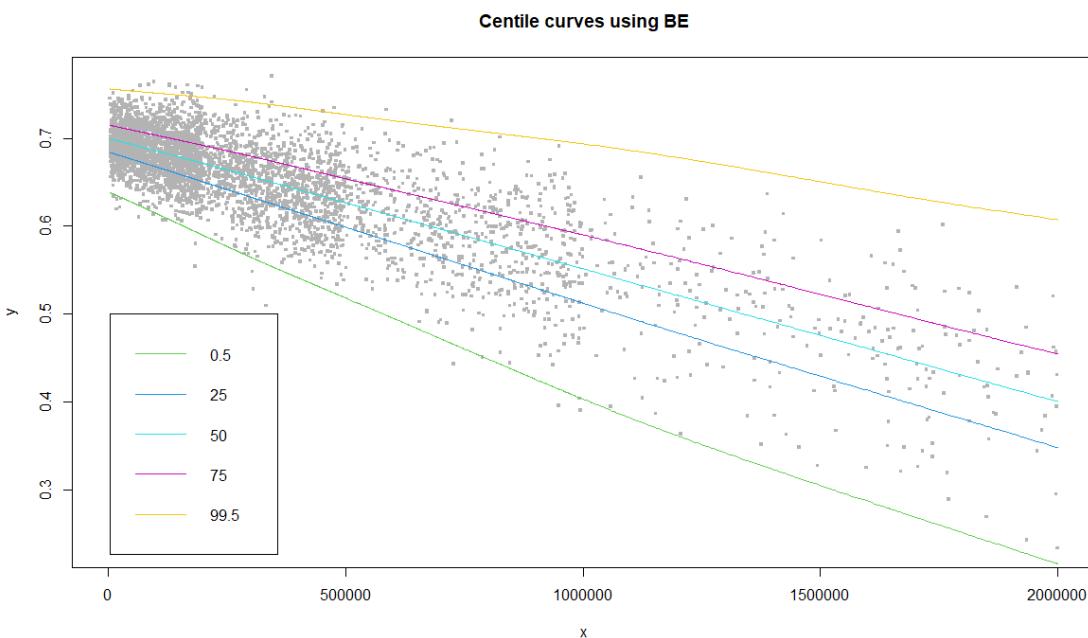


FIGURE 55 – Quantiles du modèle

La figure ci-dessus permet de mettre en avant la captation des effets de la somme assurée sur l'espérance et la variance. Au vu des résultats et des différents graphiques présentés, les modèles de

régression utilisés semblent être efficaces et pertinents.

13.4 Simulations

A partir des modèles de fréquence et sévérité calibrés, nous souhaitons maintenant générer une table d'évènement qui sera exploitable dans ReMetrica. Il s'agit là tout simplement de générer des sinistres sur le portefeuille de la cédante un certain nombre de fois.

La figure 56 présente schématiquement l'utilisation des modèles.

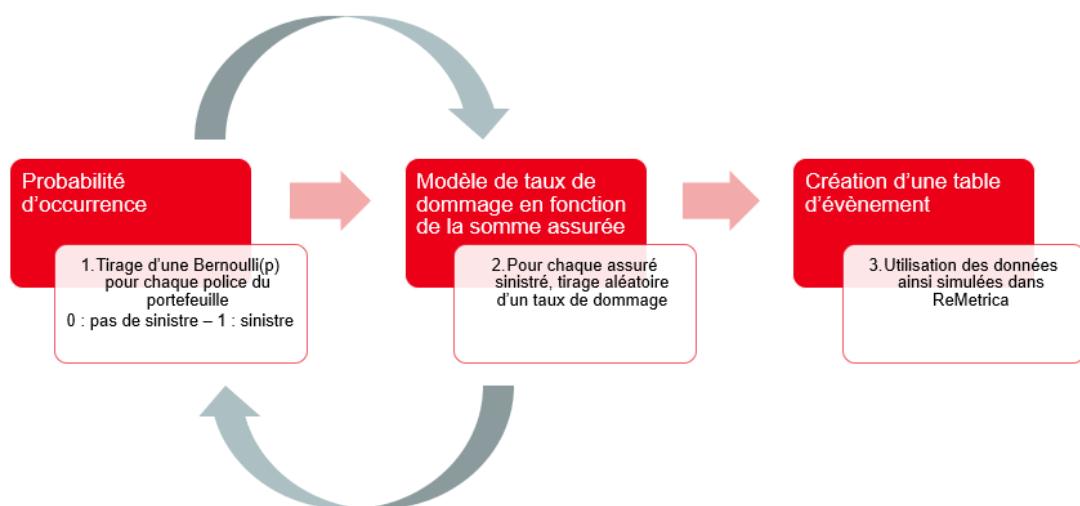


FIGURE 56 – Fonctionnement du modèle sur somme assurée

Soit $N = 100k$ le nombre de simulations pour ce cas fictif. Pour avoir un moyen de comparaison, nous utiliserons le modèle sur un portefeuille d'exposition généré à partir du même profil risque que le portefeuille utilisé pour la génération des sinistres et la calibration des modèles.

Vérifions que les simulations tendent bien en moyenne vers les données d'origine :

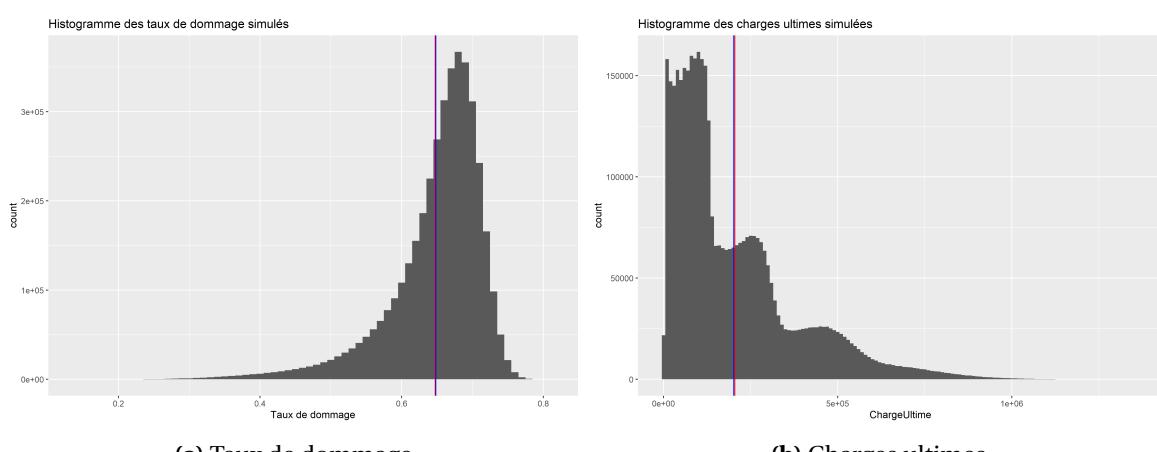


FIGURE 57 – Histogrammes des sinistres générés par le modèle

Sur chaque graphique,

- En bleu la moyenne observée sur la base d'apprentissage (base sinistres fictifs) ;
- En rouge la moyenne des sinistres générés par le modèle.

La moyenne des taux de dommage des sinistres générés converge bien vers celle des taux de dommage des sinistres de la base d'apprentissage.

Nous pouvons aussi visualiser les sinistres générés en comparaison aux sinistres d'origine.

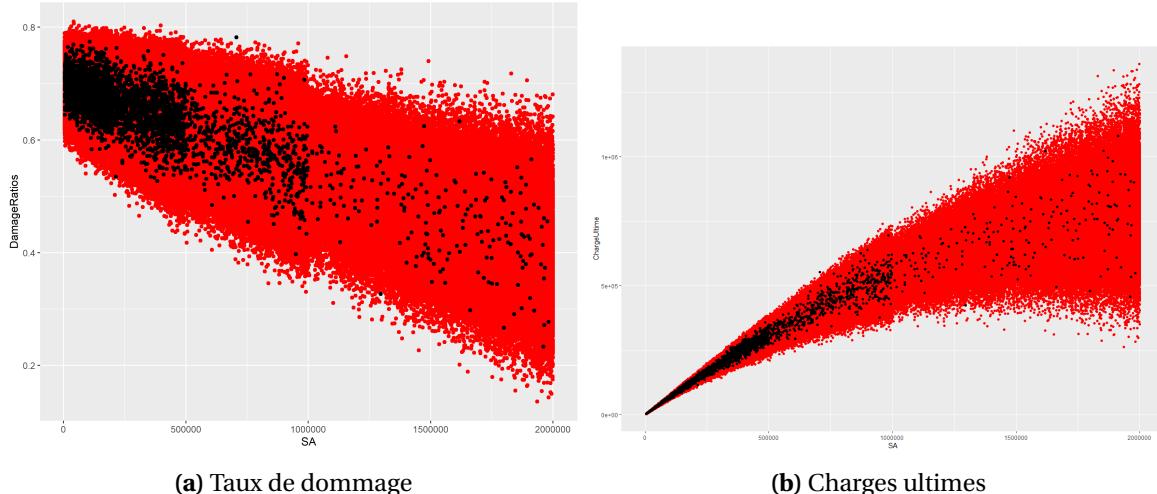


FIGURE 58 – Sinistres d'origine (en noir), sinistres simulés (en rouge)

Les simulations permettent d'estimer la distribution de la charge brute des sinistres et, après application individuelle de la réassurance, la distribution de la charge nette. L'estimation de ces distributions permet alors de calculer de nombreux indicateurs techniques (autant sur le brut que sur le net) comme les périodes de retour, la variance, la probabilité de toucher un XS...

L'utilisation de données fictives a permis de tester l'outil déployé au sein de l'équipe, mais sans données réelles, il est impossible d'estimer la capacité de généralisation du modèle. Nous allons voir dans le chapitre suivant que l'accès à la donnée est très compliqué, et ne nous permettra pas d'utiliser pour l'instant cette méthodologie.

CHAPITRE 14

Étude sur données réelles

Les outils et méthodes de modélisation étant maintenant validés et déployés, la suite logique de la démarche est de calibrer un modèle en utilisant des données marché. Historiquement, AON occupe une position de leader sur le marché du courtage en réassurance non-vie en France. Cette position nous permet d'avoir accès à un volume de données très important. Pourtant, nous allons le voir, la donnée restera LE facteur limitant de ce modèle. Les portefeuilles étudiés pour l'instant sont des portefeuilles uniquement résidentielles.

14.1 Les données

Dans un premier temps, un état des lieux des clients d'AON a été établi, avec pour objectif d'identifier les cédantes qui pourraient potentiellement être utilisées afin de calibrer le modèle. On parle ici de cédantes qui ont un portefeuille résidentiel.

Une liste de critères a été établie afin de faciliter l'identification des clients intéressants du point de vue de la modélisation :

- Disponibilité des portefeuilles d'expositions de nombreuses années; en effet, les sinistres présents dans les bases peuvent être particulièrement anciens. Le portefeuille d'exposition de l'année de survenance de ces sinistres doit donc être disponible pour faire la jointure entre les bases.
- Sommes assurées calculées au préalable; comme expliquée précédemment, les sommes assurées sont calculées en interne par les équipes de *Cat Modelers* dans le cadre de travaux clients. Les sommes assurées sont au coeur de cette méthode de modélisation, l'accès à cette variable est indispensable pour calibrer un modèle.
- Identification de la partie dommage et de la partie responsabilité des sinistres dans les base de données; rappelons que ce modèle a pour ambition d'améliorer la qualité de la modélisation

uniquement sur la partie dommage, grâce au lien entre taux de dommages et somme assurée.

Aucune cédante répondant aux 3 critères à la fois n'a pu être identifiée jusqu'à maintenant.

Toutefois, 7 cédantes répondant à au moins deux de ces critères ont été contactées afin d'obtenir la clé de jointure entre base sinistres et base d'exposition. Nous avons obtenu trois réponses positives qui nous ont permis de réfléchir et travailler sur données véritables.

Deux cas sont à distinguer :

- Le premier cas concerne deux cédantes : nous disposons de portefeuilles d'expositions avec le calcul des sommes assurées ligne à ligne effectué, mais les sinistres transmis par la cédante englobent l'incendie et les risques annexes sans distinction dommage/RC. A priori, notre méthodologie ne sera pas pertinente. Nous étudierons ces bases dans la partie 14.2.
- Le second cas semble le plus intéressant : la cédante concernée indique dans les bases sinistres transmises chaque année s'il s'agit d'incendie pure ou de responsabilité. Aussi, le nombre de sinistres communiqués est extrêmement important : c'est une cédante très implantée (une exposition de l'ordre de 4M de contrats) qui a fixé un seuil de communication à 300k, ce qui est assez bas au vu de sa taille et du niveau auquel elle se réassure. De plus, ils nous transmettent tous les 3 ans leur portefeuille ligne à ligne. Nous étudierons ce cas plus en détail dans la partie 14.3.

14.2 Premier cas : sinistres transmis sans le détail de la répartition dommage/responsabilité

Étudions les données récoltées en vue de la calibration du modèle de la sévérité des sinistres en fonction de la somme assurée.

Le tableau 22 récapitule les données utilisées pour constituer la base.

Cédante	Sinistres	Bases d'exposition disponibles
1	212	2014,2016,2017,2018,2019,2020,2021
2	119	2015,2017,2018,2020,2021

TABLE 22 – Données disponibles

Nous disposons de 340 sinistres, parmi lesquels nombre d'entre eux ont une survenance ultérieure à 2014 (année de la première base d'exposition disponible). De plus, certains sont des évènements (i.e une seule cause génère plusieurs sinistres non indépendants) et ne rentrent donc pas dans le cadre de cette étude et de notre méthodologie. Après jointure des sinistres aux différentes bases d'exposition correspondantes et suppression des évènements, les données restantes sont résumées dans le tableau 23 :

Cédante	Sinistres
1	81
2	102

TABLE 23 – Données obtenues après jointure sur les bases d'exposition et suppression des évènements

La base est maintenant consolidée. Les sinistres et les sommes assurées sont indexées en utilisant l'indice du coût de la reconstruction. Ensuite, les sinistres sont projetés à l'ultime. Le premier graphique que l'on souhaite tracer est logiquement celui des charges de sinistre projetées à l'ultime en fonction des sommes assurées. Il permettra de mettre en évidence l'existence ou la non-existence d'une relation entre ces deux variables, qui pourrait être exploitée dans nos modèles. Comme les sinistres de cette base regroupent à la fois l'incendie pur et la responsabilité, nous ne nous attendons pas à des résultats probants. La figure 59 présente le graphique des charges ultimes en fonction de la somme assurée.

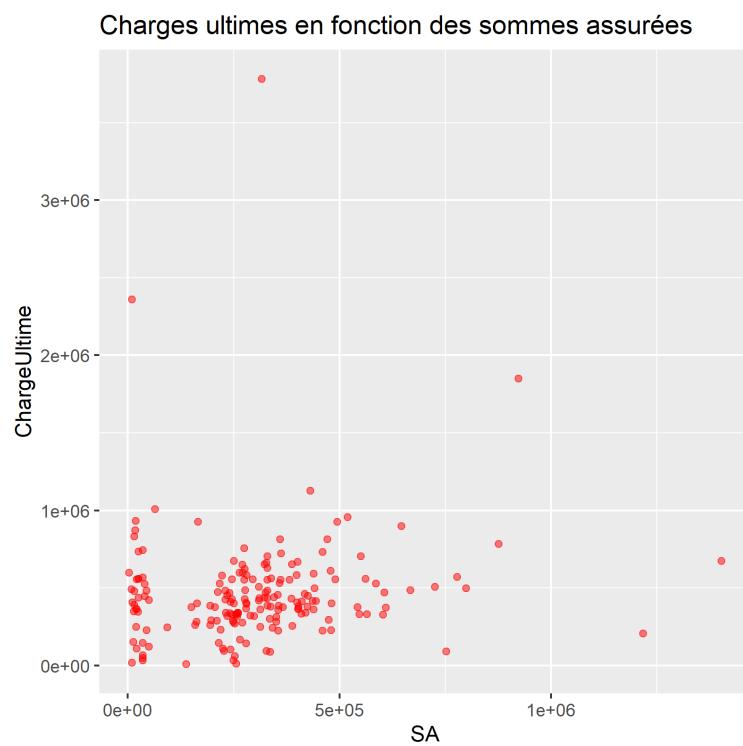


FIGURE 59 – Répartition des charge ultimes en fonction de la somme assurée

Comme attendu, aucune relation ne se dégage clairement de ce graphique. Pour aller plus loin, différentes transformations sont appliquées à nos variables, et les résultats sont présentés dans la

figure 60.

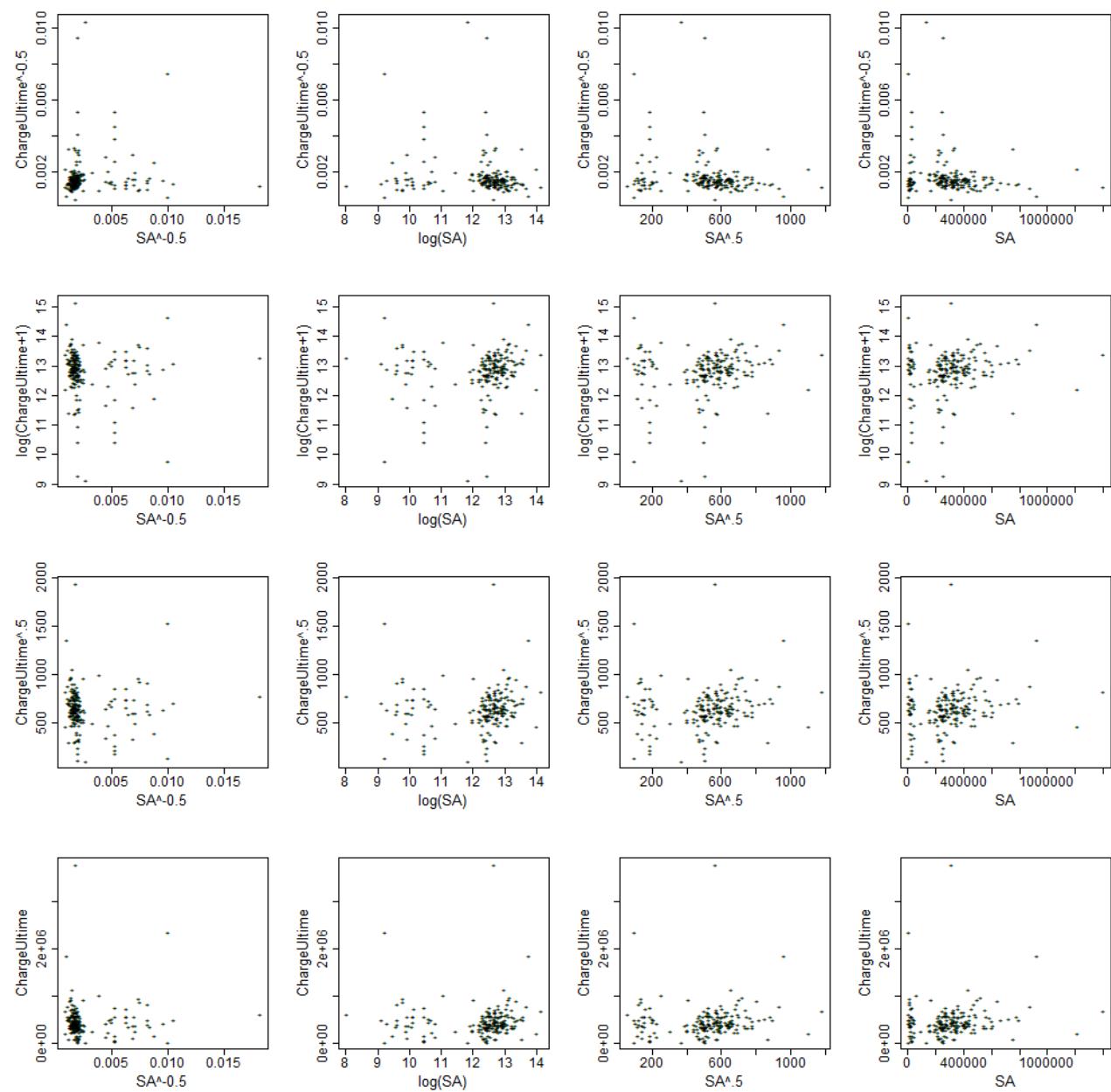


FIGURE 60 – Répartition des charge ultimes en fonction de la somme assurée après différentes transformations

Aucune relation entre les charges à l'ultime des sinistres et les sommes assurées ne semblent ressortir de ces différents graphiques. De plus, le graphique des taux de dommages en fonction de la somme assurée (figure 61) montre des taux de dommages systématiquement supérieurs à 1, et pouvant même s'élever à plus de 200.

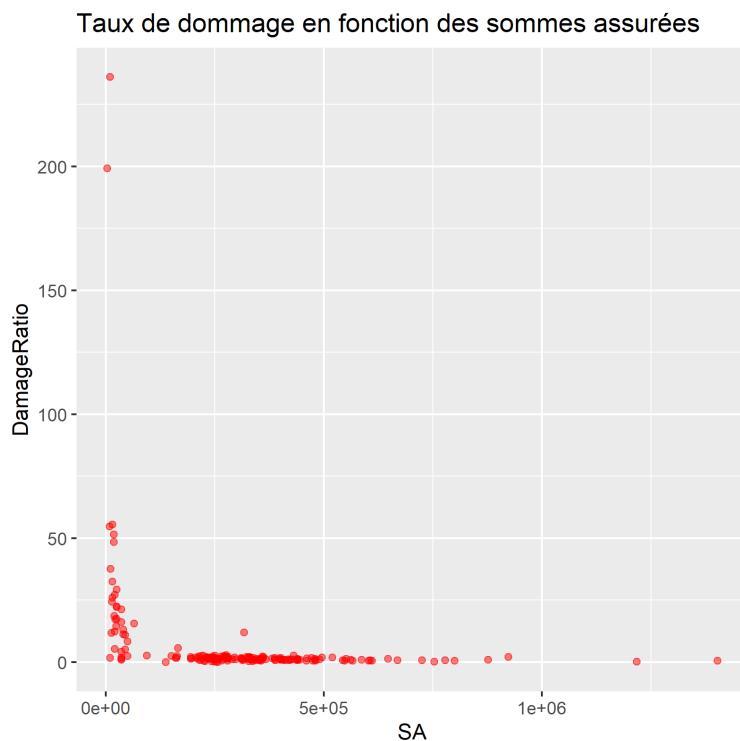


FIGURE 61 – Répartition des taux de dommage en fonction de la somme assurée

Nous travaillons pourtant bien ici avec des sommes assurées et non des 'Probable Maximum Loss' (PML). Le PML d'une police est l'estimation de la perte maximum probable; lorsque ce facteur d'exposition est utilisé, il est possible d'observer des taux de dommages supérieurs à 1. Cela ne devrait toutefois pas être le cas lorsque les sommes assurées sont utilisées, du moins pas dans les proportions observée ici (la majorité des taux de dommage sont supérieurs à 1).

L'une des hypothèses sur laquelle repose la méthode n'est pas vérifiée. Nous en déduisons que la composante RC de ces sinistres - composante non prise en compte dans le calcul de somme assurée - est très importante. Le modèle 0D semble donc être plus indiqué. Sans connaître la part de dommage pur des sinistres, un modèle de régression des taux de dommages sur sommes assurées ne semble pas pertinent.

L'étude du premier cas a montré que la connaissance du distinguo dommage pur/ responsabilité dans les sinistres est indispensable à l'utilisation de la méthode sur des risques résidentiels. Cette répartition ne peut être connu que grâce à une participation active des cédantes à l'étude : l'information n'est pas disponible en interne.

De plus, au vu des taux de dommages obtenus, les sinistres résidentiels semblent avoir une composante RC très importante ce qui milite une fois de plus pour l'utilisation du modèle 0D pour les risques résidentiels.

Alors, les cédantes qui transmettent leurs sinistres incendie sans l'information dommage/responsabilité sont pour l'instant laissées de côté.

14.3 Deuxième cas : sommes assurées non disponibles

Dans ce second cas, la répartition dommage/responsabilité des sinistres est connues et les bases d'exposition de certaines années sont disponibles. Le tableau 24 fait le point sur les données.

Cédante	Nombre de sinistres	Bases d'exposition disponibles
1	1444	2014,2017,2020

TABLE 24 – Récapitulatif des données disponibles

Parmi ces 1444 sinistres, 1146 sont flagués en incendie pure, et 298 en responsabilité. Des statistiques sont données dans le tableau 25 .

Type de sinistre	Effectif	1 ^{er} quartile	Médiane	Moyenne	3 ^{ème} quartile	Maximum
Dommage	1146	239k	328k	320k	398k	1,6M
Responsabilité	298	273k	340k	566k	489k	14,2M

TABLE 25 – Statistiques sinistres

Les sinistres dommages semblent moins volatils que les sinistres responsabilité, et sont plus concentrés autour de leur médiane. Visualisons graphiquement ces informations.

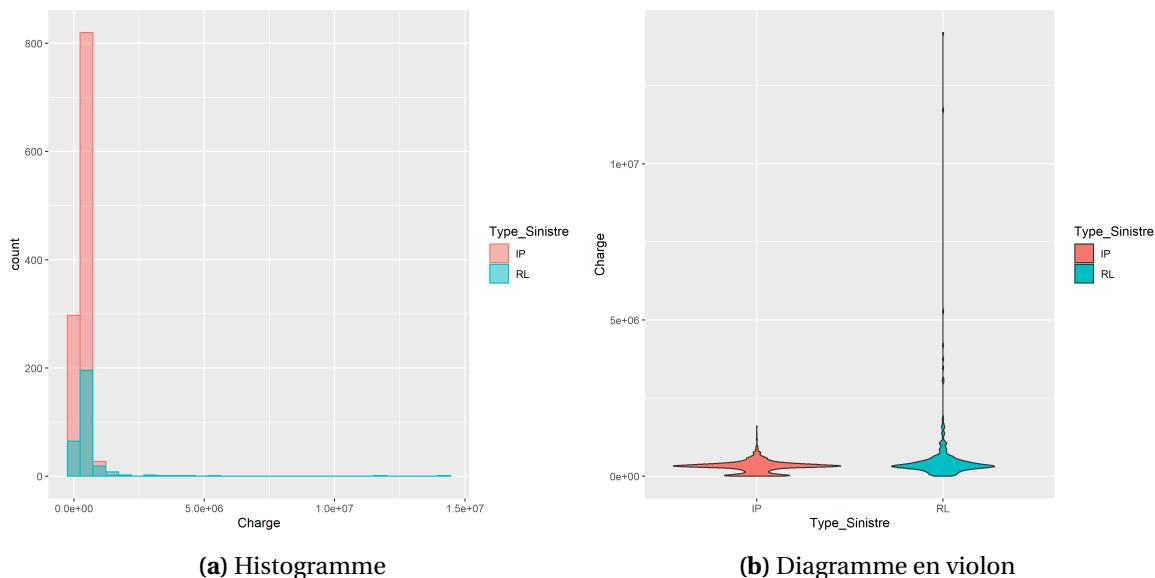


FIGURE 62 – Répartition des sinistres Dommages (en rouges) vs Responsabilité (en bleus)

Les sinistres en responsabilité n'étant pas bornés par leur somme assurée, il est logique d'observer une volatilité importante et des montants très élevés. Les sinistres dommages, quant à eux, sont bien concentrés autour de leur moyenne.

Les données de cette cédante semblent prometteuses. Le calcul des sommes assurées des polices ligne à ligne dans les bases d'exposition dont nous disposons devrait permettre de bâtir une base de données exploitable. De plus, il s'agit d'un calcul propre à AON qui se fait en interne, ce qui permet

de ne pas dépendre des clients pour aller plus loin dans l'étude. Cette étape nécessite toutefois l'intervention du pôle *Cat Modelling* et représente un temps de travail assez conséquent : elle s'inscrira alors dans un projet mené conjointement par l'équipe Actuariat et l'équipe *Cat Modelling* au sein d'Analytics.

CHAPITRE 15

Limites du modèle

Ce modèle a été pensé dans le but d'utiliser les données d'exposition transmises par les cédantes pour modéliser la composante dommage de la sinistralité incendie. Plus précisément, les profils de risque transmis chaque renouvellement par les cédantes nous ont conduits à mettre au point cette méthodologie. L'objectif initial était d'utiliser l'information des ces profils, à savoir les nombres de risques (1ère donnée d'exposition, autour de laquelle est construit le modèle 0D) et les sommes assurées (2ème donnée d'exposition, utilisée par le modèle de régression). Il permet de proposer aux clients une nouvelle approche, en lui appliquant une sinistralité marché qui va s'adapter à son profil. Néanmoins, il n'est pas exempt de biais.

Biais et limites identifiés

- Les données sont biaisées. En effet, les sommes assurées, variable explicative autour de laquelle fonctionne le modèle, sont une estimation. La valeur assurée réelle ne peut être connue qu'en cas de sinistre avec destruction totale.
De manière plus générale, une partie des sinistres n'est pas clos et leurs montants définitifs sont estimés à l'aide d'une méthode de projection à l'ultime Chain Ladder. Cela ajoute une incertitude dans la calibration du modèle.
- Pour modéliser la fréquence des sinistres, une unique loi est utilisée pour toutes les polices pour l'instant. L'objectif par la suite est de calibrer une régression logistique sur la somme assurée.
- Pour l'instant, seule la somme assurée des polices est utilisée pour expliquer en partie la sinistralité. D'autres variables pourraient être utilisées à l'avenir pour affiner le modèle.

- Les bases sinistres comptables présentent l'ensemble des arrêtés comptables pour chaque année de survenance. A contrario, les bases d'exposition sont constituées uniquement des données courantes au moment de l'extraction. Il est donc nécessaire de disposer d'un nombre conséquent de bases d'expositions à des dates d'extraction différentes. De plus, les numéros de polices des bases d'exposition ne correspondent en général pas aux numéros des sinistres : l'information doit nous être transmise par la cédante directement.
- Les sinistres utilisés pour la calibration doivent être bruts de tout (franchise, limite contractuelle, facts, ...). Ces paramètres doivent être pris en compte après modélisation du brut et varient en fonction des cédantes.
- Pour l'instant, l'étude pour la calibration a uniquement été menée sur le risque résidentiel. Pourtant, ces risques ne contribuent généralement que très peu aux cessions de réassurance de traités par risque. De plus, nous l'avons vu, les sinistres résidentiels ont une composante RC importante qui n'est elle pas reliée à la somme assurée de la police sinistrée. Le modèle n'est alors pas pertinent pour cette typologie de risque.

Enfin, pour le moment, le point bloquant l'utilisation du modèle 1D est l'accès aux données (voir chapitre 14). Nous avons vu qu'il existait des pistes pour avancer, mais tant que ce problème ne sera pas résolu, il sera impossible de calibrer et donc d'utiliser cette méthode de modélisation. Nous espérons pouvoir travailler à ce sujet dans le futur. Pour l'instant, le modèle 0D présenté en partie 1 reste le modèle que nous préconisons pour faire face à une insuffisance de statistique sinistre historique.

Conclusion

Conclusion

Résumé des travaux et réponse à la problématique

Ce mémoire s'est concentré sur la résolution d'une problématique fondamentale pour notre entreprise : le traitement du manque de données dans la modélisation des sinistres graves incendies. La nature de la réassurance non-proportionnelle, s'appliquant aux sinistres individuels graves, impose une modélisation individuelle des sinistres. En tant que courtier en réassurance de premier plan dans le secteur de la non-vie, l'équipe Analytics d'AON bénéficie d'un accès privilégié à des données marché. Nous avons donc entrepris de mettre en place des méthodologies permettant la modélisation de la sinistralité brute au-delà d'un seuil des cédantes, à partir d'un facteur d'exposition, afin de proposer à nos clients une vision du risque pertinente.

Dans un premier temps, le modèle 0D a été mis en place. Ce modèle repose sur la méthode probabiliste appliquée à une base de sinistralité marché en y ajoutant une gestion intelligente des seuils de communication des différentes cédantes de la base marché. Il permet l'obtention de fréquence/coût marché qui sont ensuite mis à l'échelle du portefeuille de la cédante à l'aide d'un facteur d'exposition : le nombre de biens assurés en portefeuille. Ce nombre est utilisé pour recalculer les paramètres des lois de fréquence. La méthode a été intégrée en interne, et a déjà eu l'occasion d'être utilisée à plusieurs reprises. Les résultats sur données clientes et le comparatif aux résultats obtenus avec la méthode sur expérience ont été présentés dans le chapitre 6 et ont permis de mettre en avant les apports du modèle pour faire face à une insuffisance de la statistique sinistre.

Par la suite, nous nous sommes penchés sur les méthodes classiques de tarification sur exposition en property par risque. Nous avons notamment implémenté la méthode présentée dans [10] permettant d'employer une approche par simulations et de générer individuellement les sinistres au-delà d'un seuil. Toutefois, après étude de la méthode, il s'est avéré qu'elle ne répondait pas à la problématique de ce mémoire. En effet, la sinistralité brute au-delà d'un seuil est définie à priori au travers des hypothèses de *Loss Ratio* et de courbes d'exposition. Ensuite, les courbes

d'exposition permettent de répartir cette sinistralité brute entre assureur et réassureur en fonction des traités étudiés pour estimer une prime de réassurance. Or, c'est justement la sinistralité brute au-delà d'un seuil que nous souhaitant obtenir en sortie de modèle. L'estimation d'une prime pure de réassurance découle ensuite de l'application de la structure au brut modélisé. Utiliser une hypothèse de *Loss Ratio* suppose que la sinistralité brute bande est connue, ce qui n'est pas le cas. Nous souhaitons que la sinistralité brute au-delà d'un seuil soit estimée par notre modèle à partir d'un facteur d'exposition, pas qu'elle soit fixée en amont. La figure 63 illustre ce propos.

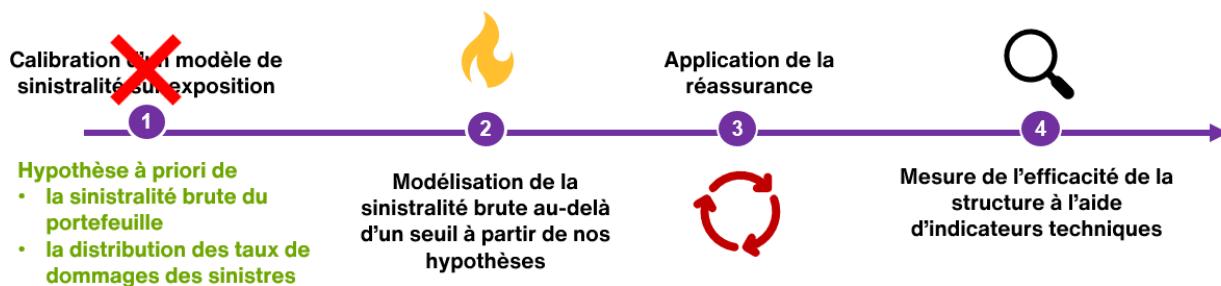


FIGURE 63 – Fonctionnement de la méthode sur exposition 'classique'

Actuellement, nous ne possédons pas l'expertise requise pour déterminer des paramètres pertinents en vue de l'application de cette méthode. Par conséquent, nous ne sommes pas en mesure de proposer des modélisations solides fondées sur cette approche, comme nous avons pu le voir lors de l'application au cas pratique de la cédante A.

Nous avons alors développé le modèle 1D, s'inscrivant dans la continuité du modèle 0D, pour la composante dommage pur des sinistres. Ce modèle intègre une dimension explicative supplémentaire par le biais des sommes assurées. Malgré des résultats encourageants avec des données simulées, l'application du modèle aux données réelles a mis en lumière des difficultés à obtenir des données exploitables pour sa calibration. En dépit des défis rencontrés dans l'obtention de données, nous avons pu mener des analyses significatives sur des portefeuilles résidentiels. Nous avons observé que le modèle 1D n'était pas adapté pour modéliser la sinistralité incendie grave des risques résidentiels. En effet, dans ce segment, les sinistres graves individuels semblent principalement résulter des garanties de responsabilité civile, et les montants de sinistres ne sont majoritairement pas expliqués par la somme assurée des biens. Dans ce contexte, le modèle 0D demeure plus approprié pour ce type de risque. Cette constatation souligne l'importance de prendre en compte les spécificités de chaque segment de risque lors du développement de modèles de prévision et de tarification. Elle met également en évidence la nécessité de continuer à explorer et à développer des approches innovantes pour modéliser la sinistralité incendie et mieux appréhender les dynamiques complexes de ce domaine dans le secteur de l'assurance non-vie. Ces résultats fournissent ainsi des orientations précieuses pour de futures recherches et développements dans ce domaine.

Discussion et développements futurs

En résumé, nous avons implantés en interne trois méthodologies distinctes de modélisation sur exposition, chacune présentant des forces et des faiblesses : le modèle 0D, le modèle sur exposition 'classique' et le modèle 1D sur somme assurée. A l'heure actuelle, la vision du risque que nous préconisons est celle du modèle 0D.

Toutefois, nous avons posé des bases qui pour chaque modèle pourraient être développées davantage.

Pour continuer le développement du modèle 0D, il serait très intéressant de consolider des bases marché par segment de risque (résidentiel, professionnel, industriel, ...) en identifiant dans nos bases sinistres les typologies correspondantes. Nous pourrions alors calibrer des courbes marchés par typologie de risque et améliorer la précision de nos modélisations en les segmentant par LoB. Cette étape repose principalement sur un travail d'identification qui ne peut être mené qu'avec l'aide de nos clients.

Ensuite, comme répété à plusieurs reprise dans le mémoire, la méthode de tarification sur exposition 'classique' repose sur des hypothèses de 'Loss Ratio' et courbes d'exposition à définir en amont. Nous pourrions alors mener une étude visant à définir des paramètres de modèle pertinents en fonction de critères (encore à définir) pour maîtriser au maximum la principale source d'incertitude de cette méthode et pouvoir l'utiliser.

Pour le dernier modèle, le modèle 1D sur sommes assurées, l'étude menée sur les risques résidentiels a montré d'une part qu'ils contribuent généralement assez peu aux cessions et d'autre part qu'ils sont fortement exposés à des garanties responsabilité qui ne sont pas expliqués par la somme assurée. Il serait alors intéressant de mener cette même étude sur un portefeuille industriel ou commercial, mais dont l'estimation des sommes assurées peut-être différente que pour le particulier (avec par exemple la nécessité d'estimer la garantie Perte d'Exploitation).

Bibliographie

- [1] Stefan BERNEGGER. « The Swiss Re Exposure Curves and the MBBEFD Distribution Class ». In : *ASTIN Bulletin* 27.1 (1997), p. 99-111.
- [2] Burhan BUTUN. « Property Excess of Loss Reinsurance Pricing in the Lloyd's market ». Mémoire d'actuariat, ISFA, 2016.
- [3] Adrien COSTEMALE. « Conseil pour la révision de l'optimalité de programmes de réassurance non-vie ». Mémoire d'actuariat, ISFA, 2015.
- [4] Michel DENUIT et Arthur CHARPENTIER. *Mathematiques de l'Assurance Non-Vie. Tome II: Tarification et Provisionnement.* 2005.
- [5] S DESMEDT et al. « Experience and exposure rating for property per risk excess of loss reinsurance revisited ». In : *ASTIN Bulletin: The Journal of the IAA* 42.1 (2012), p. 233-270.
- [6] Christophe DUTANG, Markus GESMANN et Giorgio SPEDICATO. *Exposure rating, destruction rate models and the MBBEFD package.* 2019.
- [7] Christine FINAS. « Les tarifs des réassureurs sont-ils crédibles ? » Mémoire d'actuariat, ISFA, 2015.
- [8] Victor GILLIER. « Calibration avancée du modèle de tarification à l'exposition dans le cadre de la réassurance incendie par risque ». Mémoire d'actuariat, ISFA, 2019.
- [9] Pierre-Olivier GOFFARD. « Echantillonneurs de Monte Carlo séquentiels pour la calibration et la sélection d'un modèle composite pour les pertes en assurance non-vie ». Mémoire d'actuariat, ISFA, 2021.
- [10] Jan HREVUŠ et Luboš MAREK. « Exposure Modelling in Property Reinsurance ». In : *Prague Economic Papers* 2019.2 (2019), p. 129-154.
- [11] Franck LOPEZ. « Provisionnement individuel stochastique appliqué à la réassurance non proportionnelle ». Mémoire d'actuariat, CEA, 2021.
- [12] David MALINGE. « Optimisation d'un portefeuille de réassurance non-vie : L'exemple du Property / Casualty ». Mémoire d'actuariat, ISFA, 2013.

- [13] Antoine PIERRE. « Modélisation de la sévérité des traités en excédent de sinistre, approche par la théorie des valeurs extrêmes ». Mémoire d'actuariat, DUAS, 2021.
- [14] Swiss RE. *US Property Casualty outlook: premiums in race to catch up with claims costs*. 2023.
URL : <https://www.swissre.com/institute/research/sigma-research/Insurance-Monitoring/us-property-casualty-outlook-september-2023.html>.
- [15] RA RIGBY et DM STASINOPoulos. *A flexible regression approach using GAMLSS in R* (2010).
2010.

Annexes

A La méthode de Chain-Ladder	ii
A.1 Présentation	ii
A.2 Théorie	ii
A.3 Limites de la méthode	iii
B Méthodes d'estimations paramétriques	v
B.1 Rappels sur les estimateurs	v
B.2 La méthode des moments	vi
B.3 Le maximum de vraisemblance	vii

ANNEXE A

La méthode de Chain-Ladder

A.1 Présentation

Nous présenterons ici l'approche classique de la méthode de Chain-Ladder. Nous serons amenés à l'utiliser pour projeter individuellement des sinistres à l'ultime lors de nos modélisations.

La méthode de Chain-Ladder est une méthode déterministe très largement utilisée dans l'actuariat. Elle est principalement utilisée en provisionnement non-vie. L'idée assez intuitive de cette méthode est de supposer que les sinistres survenus lors d'années de survenance différentes suivent la même cadence de règlement. Autrement dit, on suppose que le futur est reflet du passé.

A.2 Théorie

La méthode se base sur des triangles de développements. Définissons $C_{i,j}$ les règlements de sinistres cumulés :

$$C_{i,j} = \sum_{k=0}^j X_{i,k}$$

avec

- i l'année de survenance des sinistres
- $X_{i,k}$ les règlements en $i+k$ des sinistres survenus en i

	<i>i</i>	<i>j</i>	0	1	2	3	4	5
			C _{1,0}	C _{1,1}	C _{1,2}	C _{1,3}	C _{1,4}	C _{1,5}
	1							
	2		C _{2,0}	C _{2,1}	C _{2,2}	C _{2,3}	C _{2,4}	
	3		C _{3,0}	C _{3,1}	C _{3,2}	C _{3,3}		
	4		C _{4,0}	C _{4,1}	C _{4,2}			
	5		C _{5,0}	C _{5,1}				
	6		C _{6,0}					

FIGURE 64 – Triangle de développement

Les facteurs de développements, sous l'hypothèse d'indépendance entre les années de survenance, sont estimés par :

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

et par hypothèse, on a

$$\forall j \in \{0, \dots, n\}, \hat{C}_{i,j+1} = \hat{\lambda}_j \times C_{i,j}$$

La projection à l'ultime de la charge sinistre agrégée par survenance est réalisée ensuite en cascade en appliquant les facteurs de développements à partir de la diagonale du triangle.

	0	1	2	3	4	5
1	C _{1,0}	C _{1,1}	C _{1,2}	C _{1,3}	C _{1,4}	C _{1,5}
2	C _{2,0}	C _{2,1}	C _{2,2}	C _{2,3}	C _{2,4}	C _{2,5,estimé}
3	C _{3,0}	C _{3,1}	C _{3,2}	C _{3,3}	C _{3,4,estimé}	C _{3,5,estimé}
4	C _{4,0}	C _{4,1}	C _{4,2}	C _{4,3,estimé}	C _{4,4,estimé}	C _{4,5,estimé}
5	C _{5,0}	C _{5,1}	C _{5,2,estimé}	C _{5,3,estimé}	C _{5,4,estimé}	C _{5,5,estimé}
6	C _{6,0}	C _{6,1,estimé}				

	0	1	2	3	4	5
1	C _{1,0}	C _{1,1}	C _{1,2}	C _{1,3}	C _{1,4}	C _{1,5}
2	C _{2,0}	C _{2,1}	C _{2,2}	C _{2,3}	C _{2,4}	C _{2,5,estimé}
3	C _{3,0}	C _{3,1}	C _{3,2}	C _{3,3}	C _{3,4,estimé}	C _{3,5,estimé}
4	C _{4,0}	C _{4,1}	C _{4,2}	C _{4,3,estimé}	C _{4,4,estimé}	C _{4,5,estimé}
5	C _{5,0}	C _{5,1}	C _{5,2,estimé}	C _{5,3,estimé}	C _{5,4,estimé}	C _{5,5,estimé}
6	C _{6,0}	C _{6,1,estimé}	C _{6,2,estimé}	C _{6,3,estimé}	C _{6,4,estimé}	C _{6,5,estimé}

	0	1	2	3	4	5
1	C _{1,0}	C _{1,1}	C _{1,2}	C _{1,3}	C _{1,4}	C _{1,5}
2	C _{2,0}	C _{2,1}	C _{2,2}	C _{2,3}	C _{2,4}	C _{2,5,estimé}
3	C _{3,0}	C _{3,1}	C _{3,2}	C _{3,3}	C _{3,4,estimé}	C _{3,5,estimé}
4	C _{4,0}	C _{4,1}	C _{4,2}	C _{4,3,estimé}	C _{4,4,estimé}	C _{4,5,estimé}
5	C _{5,0}	C _{5,1}	C _{5,2,estimé}	C _{5,3,estimé}	C _{5,4,estimé}	C _{5,5,estimé}
6	C _{6,0}	C _{6,1,estimé}	C _{6,2,estimé}	C _{6,3,estimé}	C _{6,4,estimé}	C _{6,5,estimé}

FIGURE 65 – Remplissage en cascade d'un triangle de développement

A.3 Limites de la méthode

Cette méthode, très répandue de par sa simplicité de mise en place et son côté très intuitif, présente toutefois des limites évidentes :

- L'hypothèse d'indépendance des années de survenance est très limitante. Pour la vérifier, le passé doit être régulier. Le profil de l'assureur doit donc être le même d'année en année, ce qui n'est pas forcément vérifié;

- La base de données utilisée doit être suffisamment fourni. Un manque de donnée entraînerait des coefficients très volatiles, donc des résultats peu robustes.
- Chain-Ladder est une méthode déterministe qui ne permet donc pas d'estimer l'incertitude.

Des variantes de Chain-Ladder ont été développées pour pallier ces faiblesses. Néanmoins, dans le cadre de notre étude, l'utilisation de cette méthode sera suffisamment efficace. Le cœur de ce mémoire ne réside pas dans l'estimation de charges ultimes, mais dans la modélisation de la sinistralité.

ANNEXE B

Méthodes d'estimations paramétriques

Dans le monde statistique, on modélise souvent les comportements étudiés par des variables aléatoires qui suivent une loi usuelle dont on déterminera les paramètres en utilisant l'échantillon de données observé.

Cette partie portera sur les principales méthodes utilisées pour estimer les paramètres des lois envisagées lors d'une modélisation.

B.1 Rappels sur les estimateurs

Face à un échantillon, on peut reconnaître une famille de loi paramétrique, mais dont le paramètre nous reste inconnu. Les estimateurs nous permettront alors de donner une estimation de ce paramètre.

Estimateur ponctuel

Soit (X_1, \dots, X_n) un échantillon de n variables aléatoires i.i.d de loi L dépendant d'un paramètre θ . On appelle estimateur ponctuel de θ toute fonction $f(X_1, \dots, X_n)$. On appelle estimation toute valeur $f(x_1, \dots, x_n)$, avec (x_1, \dots, x_n) une réalisation de (X_1, \dots, X_n) .

Un estimateur ponctuel $\hat{\theta}_n$ est donc une variable aléatoire, et on appelle ses réalisations estimations.

Convergence/consistance d'un estimateur

On dit qu'un estimateur $\hat{\theta}_n$ est convergent ou consistant s'il converge en probabilité vers θ , i.e :

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = 0$$

On dit qu'un estimateur $\hat{\theta}_n$ est fortement convergent ou consistant s'il converge presque sûrement vers θ , i.e :

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta\right) = 1$$

Biais d'un estimateur

On appelle biais d'un estimateur la quantité

$$\mathbb{E}[\hat{\theta}_n] - \theta$$

Un estimateur est dit sans biais si et seulement si

$$\mathbb{E}[\hat{\theta}_n] = \theta \Leftrightarrow \text{Biais}(\hat{\theta}_n) = 0$$

Erreur Quadratique Moyenne

L'erreur quadratique moyenne d'un estimateur $\hat{\theta}$ de θ est une mesure de précision de cet estimateur, définie par :

$$EQM(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] \quad (\text{B.1})$$

$$= \text{Biais}(\hat{\theta})^2 + \text{Var}(\hat{\theta}) \quad (\text{B.2})$$

Dans le cas d'un estimateur sans biais, $EQM = \text{Variance}$.

Nous utiliserons ces différents indicateurs pour comparer les estimateurs issus de méthodes différentes. Nous présenterons ici deux différentes méthodes pour obtenir des estimateurs.

B.2 La méthode des moments

La méthode des moments est une méthode classique très largement utilisée dans l'estimation de lois discrètes. Elle repose sur la loi des grands nombres, qui affirme que si l'on considère un échantillon (X_1, \dots, X_n) de n variables aléatoires indépendantes et identiquement distribuées (i.i.d), alors la moyenne empirique de l'échantillon converge en probabilité (loi faible des grands nombres) / presque sûrement (loi forte des grands nombres) vers l'espérance.

Loi faible des grands nombres

Soit (X_n) une suite de v.a i.i.d admettant une espérance μ . Alors la moyenne empiriques $\overline{X_n} = \frac{1}{n} \sum_{i=1}^n X_i$ convergent en probabilité vers μ :

Pour tout $\epsilon > 0$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\overline{X_n} - \mu| > \epsilon) = 0$$

Loi forte des grands nombres

Si en plus les (X_n) sont intégrables, i.e $\mathbb{E}[|X_1|] < \infty$, alors la moyenne empirique converge presque sûrement vers μ :

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} \overline{X_n} = \mu\right) = 1$$

La méthode des moments consiste alors à exprimer les p premiers moments (en général, $p = 2$) de la loi choisie en fonction des paramètres $\theta = (\theta_1, \dots, \theta_p)$ de la loi. L'expression analytique des moments théoriques des lois paramétriques usuelles est connue. On obtient alors un système de la forme suivante :

$$\begin{cases} \mu_1 = h_1(\theta_1, \dots, \theta_p) \\ \dots \\ \mu_p = h_p(\theta_1, \dots, \theta_p) \end{cases}$$

En approximant les moments (μ_1, \dots, μ_p) par les estimateurs empiriques convergents $(\bar{X^1}, \dots, \bar{X^p})$, on peut résoudre le système et obtenir une estimation de $\theta = (\theta_1, \dots, \theta_p)$.

B.3 Le maximum de vraisemblance

La méthode des moments, bien que relativement intuitive et simple à mettre en place, ne propose pas toujours les estimateurs aux propriétés les plus intéressantes. Nous allons dans cette partie étudier les estimateurs du maximum de vraisemblance, qui présentent bien souvent des résultats plus intéressants. En effet, la méthode du maximum de vraisemblance donne des estimateurs sans biais, convergent et de variance minimale. L'idée derrière cette méthode est la suivante : on choisit comme estimation de θ la valeur la plus vraisemblable, i.e la valeur de θ pour laquelle la probabilité d'observer les valeurs de notre échantillon est la plus forte.

La loi d'une variable aléatoire est caractérisée par sa distribution de probabilités. De la même manière, la loi d'un échantillon de n réalisations (x_1, \dots, x_n) de variables aléatoires i.i.d (X_1, \dots, X_n) est caractérisée par $\prod_{i=1, \dots, n} f(x_i, \theta)$;

$f(x, \theta)$ est définie par :

$$f(x, \theta) = \begin{cases} f_\theta(x) & \text{si } X \text{ est une v.a continue} \\ \mathbb{P}_\theta(X = x) & \text{si } X \text{ est une v.a discrète} \end{cases}$$

$\prod_{i=1, \dots, n} f(x_i, \theta)$ est une fonction de (x_1, \dots, x_n) et θ que l'on note $L(x_1, \dots, x_n; \theta)$.

Estimateur du maximum de vraisemblance - EMV

L'estimateur du maximum de vraisemblance de θ est la valeur de t qui maximise la fonction de vraisemblance :

$$t \rightarrow L(x_1, \dots, x_n; t)$$

ou la log-vraisemblance :

$$t \rightarrow l(x_1, \dots, x_n; t) = \log(L(x_1, \dots, x_n; t))$$

En pratique, la solution de l'équation suivante est la valeur recherchée :

$$\frac{\partial L(x_1, \dots, x_n; \theta)}{\partial \theta} = 0 \quad \text{ou} \quad \frac{\partial l(x_1, \dots, x_n; \theta)}{\partial \theta} = 0$$

S'il n'existe pas de solution analytique, alors on n'utilise pas l'EMV.

Enfin, il faut vérifier que l'on ait bien affaire à un maximum. Pour cela, on détermine la dérivée seconde de la vraisemblance et on vérifie :

$$\frac{\partial^2 L(x_1, \dots, x_n; \theta)}{\partial \theta^2} < 0$$