

**Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuares**

Par : Julie GAUDIN		
Titre : Modélisation de la sévérité des sinistres en réassurance : de la théorie des valeurs extrêmes à l'approche par distorsion.		
Confidentialité : ☒ NON ☐ OUI (Durée : ☐ 1 an ☐ 2 ans)		
Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.		
<u>Membres présents du jury de la filière :</u> Maud THOMAS	Signature :	<u>Entreprise : CCR RE</u> Nom : Sharmini MANICKAM Signature : 
<u>Membres présents du jury de l'Institut des Actuares :</u>		<u>Directeur de mémoire en entreprise</u> Nom : Sharmini MANICKAM Signature : 
		<u>Invité :</u> Nom : Signature :
		Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité) <u>Signature du responsable entreprise : Alpha BAH</u> A. Bah
		<u>Signature du candidat :</u> 

Remerciements

Je tiens tout d'abord à exprimer ma profonde gratitude à ma tutrice chez CCR Re, Sharmini MANICKAM, pour m'avoir offert l'opportunité de m'initier au monde de la réassurance, ainsi que pour la transmission de ses savoirs et ses conseils avisés. Je remercie également Pierre MOURIER pour son accompagnement précieux et ses orientations tout au long de l'élaboration de ce mémoire.

Je souhaite également adresser mes remerciements à Alpha BAH et Lucas RICHARD pour leur relecture éclairée de ce mémoire, et, plus largement, aux membres de la direction actuariat et risques, pour leur bienveillance et leur disponibilité continues au cours de mon alternance.

Ce mémoire, constituant l'aboutissement concret de cinq années d'études, me donne aussi l'occasion d'exprimer ma reconnaissance aux enseignants et aux membres de l'administration universitaire pour leur engagement et leur accompagnement tout au long de ce parcours.

Enfin, je remercie mes parents pour leur soutien et leurs encouragements constants au cours de mes études.

Résumé

Dans un marché de la réassurance en pleine expansion, la nécessité d'une tarification précise et fiable est cruciale pour maintenir la compétitivité des prix. Ce mémoire s'inscrit dans cette dynamique en proposant une évolution de la tarification des traités en excédent de sinistres, avec un accent sur l'amélioration de la modélisation de la sévérité dans le cadre du modèle collectif.

Le modèle collectif a pour objectif d'estimer les coûts futurs en modélisant la fréquence et la sévérité des sinistres à l'aide de lois statistiques. Cependant, la modélisation traditionnelle de la sévérité, souvent focalisée sur la partie centrale de la distribution, tend à négliger les sinistres extrêmes, situés dans la queue de la distribution. Or, ces événements, bien que rares, entraînent des coûts significatifs et posent un défi majeur pour la tarification des tranches élevées.

Pour surmonter cette difficulté, la sévérité a été décomposée en deux parties : le corps et la queue de la distribution, modélisés de façon distincte. Cette approche, inspirée de la théorie des valeurs extrêmes, permet une meilleure estimation des sinistres rares, assurant ainsi des résultats plus fiables pour les tranches élevées.

Ce mémoire examine également un cas particulier de la modélisation de la queue de la distribution : celui où elle appartient au domaine d'attraction de Gumbel, caractérisé par des queues fines. Bien que ce scénario soit moins fréquent en réassurance, il présente des défis importants. En effet, même si les données de la cédante suggèrent un faible risque de sinistres extrêmes, l'historique de tels événements sur le marché où ses risques sont exposés met en évidence l'existence d'un risque réel. Ainsi, l'utilisation d'une loi issue du domaine de Gumbel, comme la loi exponentielle, conduit à une sous-estimation de ce risque. Pour corriger ce biais, des fonctions de distorsion ont été introduites, permettant de réajuster la modélisation exponentielle vers la loi de marché adaptée. Un facteur de crédibilité a également été développé pour calibrer cette distorsion, en tenant compte du niveau de confiance dans les données de la cédante comme reflet de sa sinistralité future.

En conclusion, cette étude propose une méthode affinée pour modéliser les sinistres extrêmes, offrant ainsi une tarification plus robuste et explicable pour les réassureurs.

Mots clés : *Excédent de sinistre, sévérité, théorie des valeurs extrêmes, Pareto Généralisée, tranches hautes, domaine d'attraction, fonctions de distorsion, loi de marché, facteur de crédibilité.*

Abstract

In a rapidly expanding reinsurance market, the need for accurate and reliable pricing is crucial to sustain price competitiveness. This thesis is part of this process by proposing an evolution in the pricing of excess of loss treaties, with a focus on improving the modelling of severity within the framework of the collective model.

The collective model aims to estimate future costs by modeling the frequency and severity of claims using statistical distributions. However, traditional severity modeling, often focused on the central part of the distribution, tends to overlook extreme claims, which occur in the distribution's tail. These events, while rare, result in significant costs and pose a major challenge for pricing the upper layers.

To overcome this difficulty, severity has been split into two parts : the body and tail, modelled separately. This approach, inspired by extreme value theory, allows better estimation of rare claims, thus ensuring more reliable results for high layers.

This paper also examines a particular case of modelling the tail of the distribution : when it belongs to the Gumbel attraction domain, characterised by thin tails. Although this scenario is less common in reinsurance, it presents significant challenges. Indeed, even if the data suggests a low risk of extreme losses, the historical occurrence of such events in the market where they are exposed demonstrates the presence of a real risk. Thus, the use of a distribution from the Gumbel domain, such as the exponential distribution, leads to an underestimation of this risk. To correct this bias, distortion functions have been introduced to readjust the exponential model to the appropriate market distribution. A credibility factor has also been developed to calibrate this distortion, taking into account the level of confidence in the cedant's data as a reflection of its future claims experience.

In conclusion, this study proposes a refined method for modelling extreme claims, offering more robust and explainable pricing for reinsurers.

Key words : *Excess of loss, severity, extreme value theory, Generalised Pareto, high layers, domain of attraction, distortion functions, market distribution, credibility factor.*

Synthèse

Face à la croissance soutenue du marché de la réassurance, la précision et la rigueur en matière de tarification sont toujours essentielles pour garantir la compétitivité et la pertinence des prix proposés. Dans ce contexte, il est impératif pour les réassureurs de disposer de méthodes de tarification justes et fiables, particulièrement dans un environnement où les risques sont de plus en plus complexes et les données parfois limitées.

CCR Re, dans son ambition de renforcer sa compétitivité et de s'aligner sur l'expansion du marché, s'efforce d'améliorer la précision de la tarification des traités en excédent de sinistres. Plus précisément, l'objectif est d'optimiser l'utilisation du modèle collectif afin d'obtenir des résultats fiables et robustes.

Le modèle collectif repose sur deux composantes essentielles : la fréquence et la sévérité des sinistres. La fréquence, souvent représentée par la loi de Poisson, modélise le nombre de sinistres attendus sur une période donnée, basé sur les données historiques de la cédante. La sévérité, quant à elle, modélise le coût des sinistres et est souvent représentée par la loi de Pareto, particulièrement efficace pour capturer les sinistres rares et coûteux. En combinant ces deux éléments, le modèle collectif permet d'estimer le coût total des sinistres futurs, offrant ainsi une base solide pour fixer une prime ajustée au risque.

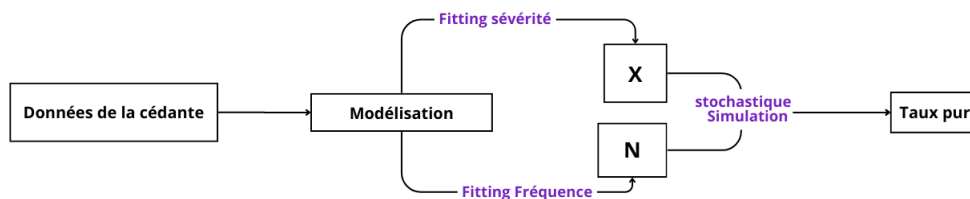


FIGURE 1 – Modèle collectif simplifié

Ce modèle présente une limite notable. En modélisant la totalité des sinistres de la cédante avec une loi unique, il devient difficile de capturer précisément les risques situés dans la queue de la distribution. En effet, lorsqu'une tranche de réassurance est travaillante, c'est-à-dire impactée par plusieurs sinistres sur la période d'observation, la sévérité des sinistres peut être correctement appréhendée par une distribution unique, couvrant l'ensemble des sinistres de la cédante. En revanche, pour les tranches non travaillantes, souvent situées dans les tranches hautes en raison de la faible fréquence des sinistres de grande ampleur, l'actuaire doit estimer la prime en se basant sur des valeurs situées dans la queue de la distribution de la sévérité. Or, le modèle actuel, qui utilise une

loi unique pour modéliser l'ensemble des sinistres, est principalement ajusté sur la partie basse de la distribution, où se concentre le plus grand nombre de sinistres. Par conséquent, la partie la plus extrême de la sinistralité, située dans la queue de la distribution, n'est pas bien modélisée, ce qui complique l'évaluation précise du risque dans les tranches hautes. Ainsi, l'objectif de ce mémoire est d'améliorer la modélisation de la sévérité des sinistres afin d'affiner la tarification des tranches élevées.

Pour atteindre cet objectif, dans la seconde partie de ce mémoire, la modélisation de la sévérité a été réexaminée et ajustée à l'aide de la théorie des valeurs extrêmes. La distribution de la sévérité des sinistres est divisée en deux parties distinctes : le corps et la queue de la distribution, séparés par un seuil des extrêmes, noté u . La détermination de ce seuil a été réalisée à l'aide des outils du mean excess plot et du threshold stability plot. Le corps de la distribution, généralement bien alimenté en données, est modélisé empiriquement. En revanche, la queue de la distribution est modélisée par une loi de Pareto généralisée (GPD), idéale pour les sinistres extrêmes au-dessus du seuil u . La jonction entre ces deux parties est assurée de manière linéaire pour garantir une transition fluide. Cette approche permet un ajustement plus fiable des sinistres extrêmes.

Le graphique ci-dessous illustre une fonction de répartition d'une loi de sévérité d'une cédante avec cette méthode. Cette méthode est nommée la sévérité composite.

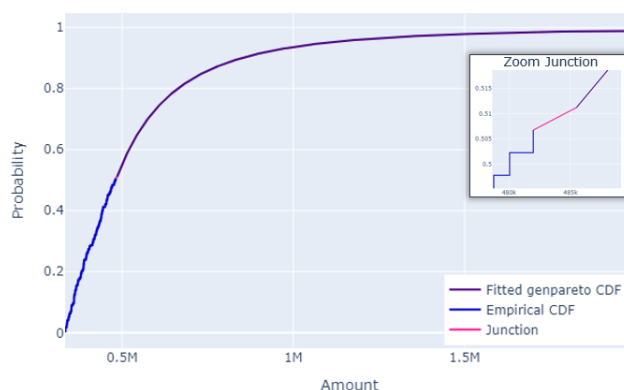


FIGURE 2 – Fonction de répartition de la sévérité composite

Cette approche permet ainsi d'obtenir des taux plus fiables et explicables dans les tranches hautes, offrant une plus grande confiance dans les taux purs calculés. Ce processus est résumée par le schéma suivant :

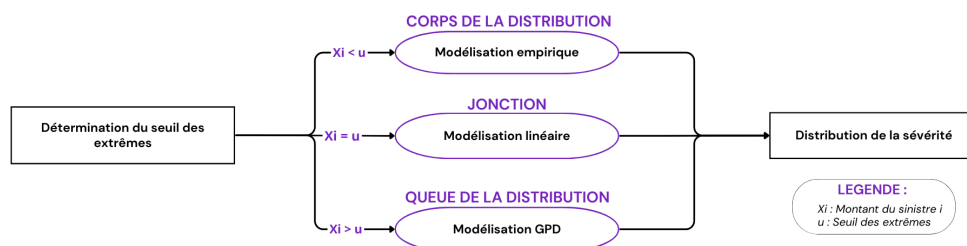


FIGURE 3 – Schéma récapitulatif de la modélisation de la sévérité

Dans la majorité des cas en réassurance, la queue de distribution de la sévérité des

sinistres de la cédante se situe dans le domaine de Fréchet, caractérisé par des lois à queues lourdes. Cependant, bien que ce domaine soit le plus couramment rencontré lors des tarifications, il arrive parfois que la queue de distribution appartienne au domaine de Gumbel, associé aux lois à queues fines. Cela implique une probabilité d'occurrence de sinistres extrêmes plus faible par rapport aux lois du domaine de Fréchet. Ce cas particulier est plus complexe en réassurance, car il suggérerait un besoin de réassurance très limité, voire inexistant. Cependant, l'historique des sinistres extrêmes dans le marché où les risques de la cédante sont exposés démontre clairement qu'un risque persiste, même si cela n'est pas pleinement reflété dans ses données historiques. Un sinistre extrême pourrait toujours se produire à l'avenir, malgré cette faible probabilité théorique.

Plusieurs exemples ont démontré que le domaine d'attraction de la queue de distribution de la cédante peut évoluer au fil du temps, passant du domaine de Gumbel à celui de Fréchet. Cette évolution justifie l'existence du risque et la pertinence du besoin de réassurance.

D'un point de vue théorique, lorsque la queue se situe dans le domaine de Gumbel, la loi de Pareto généralisée se réduit en une loi exponentielle, qui s'ajuste bien aux données historiques. Néanmoins, cette modélisation ne prend pas en compte adéquatement les risques associés aux tranches hautes, entraînant une sous-estimation des sinistres extrêmes futurs. Pour remédier à cette limite, des chargements arbitraires basés sur le jugement d'experts étaient ajoutés, mais cette méthode, bien que pratique, manquait de fondements mathématiques et n'offrait pas une confiance totale dans les taux calculés pour les tranches élevées..

C'est pourquoi, dans la troisième partie de ce mémoire, une approche basée sur les fonctions de distorsion a été introduite pour corriger de manière mathématique la sous-estimation des risques extrêmes associée à la loi exponentielle. Les fonctions de distorsion, en tant qu'outils statistiques, permettent de transformer les lois de probabilité. L'objectif étant ainsi d'appliquer une fonction de distorsion à la loi exponentielle ajustée. En modifiant la loi de cette manière, il deviendrait alors possible de mieux capturer les risques extrêmes qui sont sous-estimés par la loi exponentielle seule, tout en maintenant une base théorique solide.

La direction de la distorsion de la loi exponentielle est orientée vers la loi de marché. Une loi de marché est une distribution statistique élaborée pour représenter avec précision la sinistralité dans un marché spécifique, en fonction du pays et de la branche de réassurance concernés. Une telle loi bien établie constitue une base solide car elle intègre à la fois les tendances historiques et les particularités du marché. L'objectif est donc de distordre la loi exponentielle de manière à ce qu'elle converge vers la loi de marché correspondante. Ainsi, cette loi sert de point de référence pour déterminer la distorsion maximale que peut subir la loi exponentielle, garantissant que cette transformation reste pertinente et alignée sur les caractéristiques du marché.

Cependant, cette transformation ne peut être uniforme pour toutes les cédantes. Le

degré de distorsion doit être ajusté en fonction de la pertinence des données de chaque cédante sur l'avenir. Ainsi, un facteur de crédibilité est déterminé afin de calibrer la distorsion en fonction de la confiance accordée aux données de la cédante, les considérant comme un reflet plus ou moins fidèle de sa sinistralité future.

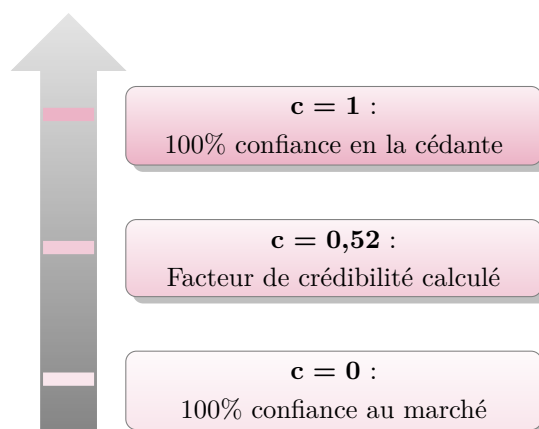


FIGURE 4 – Le facteur de crédibilité

Comme l'illustre le schéma ci-dessus, un facteur de crédibilité proche de 1 reflète une forte confiance dans les données de la cédante, caractérisées par l'observation de peu, voire aucun sinistre extrême. Cela suggère que, malgré l'absence ou la rareté de sinistralité extrême, les données de la cédante fournissent une estimation fiable des sinistres futurs. Ce scénario est pertinent si l'on considère que le marché dans lequel ces risques évoluent présente également un nombre limité de sinistres extrêmes, rendant ainsi les données de la cédante plus fidèles. Dans ce cas, la distorsion appliquée à la loi exponentielle sera minimale.

En revanche, un facteur de crédibilité proche de 0 indique une faible confiance dans les données, signifiant qu'elles ne sont pas perçues comme une bonne représentation de la sinistralité future. Ce scénario est envisageable lorsque le marché associé aux risques présente une forte sinistralité extrême, contrairement à ce que suggèrent les données de la cédante. Dans ce cas, une distorsion plus importante sera alors appliquée à la loi exponentielle. Les graphiques ci-dessous illustrent la fonction de répartition de la loi exponentielle distordue selon différents facteurs de crédibilité (en bleu) comparée à celle de la loi de marché (en violet). La fonction de distorsion utilisée est la fonction de hasard proportionnel (PH).

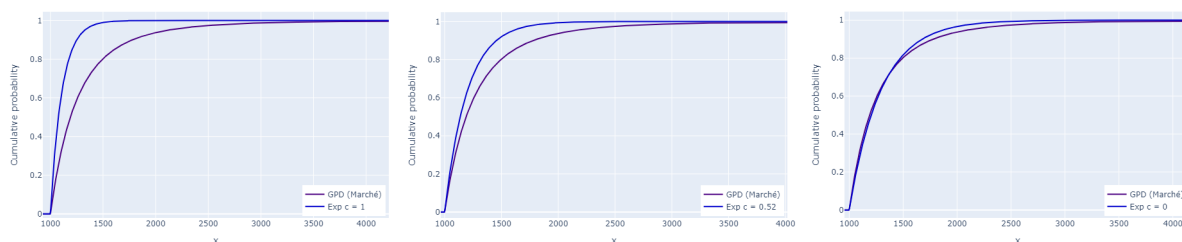


FIGURE 5 – Fonction de répartition distordue selon le facteur de crédibilité

Ainsi, plus le facteur de crédibilité c se rapproche de 0, plus la fonction de répartition

exponentielle distordue converge vers celle de la loi du marché.

Cette méthode a été testée via des simulations, afin de comparer les primes obtenues avec et sans distorsion. L'objectif était de comparer les estimations de la prime pure pour chaque tranche en fonction des lois suivantes :

- La loi exponentielle de la cédante (sans distorsion, $c = 1$).
- La loi exponentielle distordue ($c = 0,52$, le facteur de crédibilité calculé).
- La loi exponentielle distordue au maximum ($c = 0$).
- La loi du marché (GPD).

Le graphique suivant illustre les boxplots des primes pures possibles pour la tranche T1 : 1 000 XS 1 000, en fonction de la loi choisie. Ces boxplots offrent une vue d'ensemble de la moyenne et de la volatilité des primes pures pour chaque loi considérée. Ils permettent également d'observer l'évolution de la prime pure moyenne en fonction de la loi. Chaque point sur les boxplots représente une prime pure moyenne obtenue par simulation.

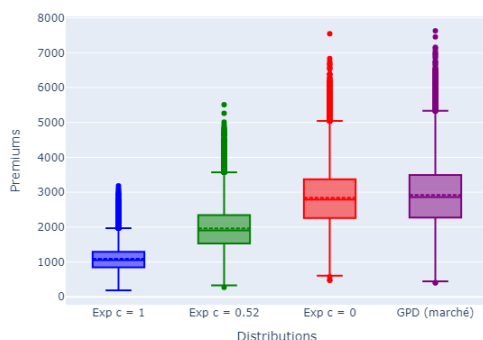


FIGURE 6 – Boxplot des primes de la T1 selon la loi

La distorsion de la loi exponentielle entraîne une augmentation de la prime pure moyenne, qui passe de 1 086 € pour la loi originale à 1 964 € après l'application du facteur de crédibilité calculé. Comme attendu, cette méthode permet d'obtenir des primes plus élevées, tout en maintenant la cohérence des résultats. La prime distordue se situe bien entre celle de la loi originale et celle du marché. Cette approche permet de prendre plus en considération les risques des tranches élevées, tout en restant fidèle aux données de la cédante.

En revanche, cette méthode présente certaines limites. Premièrement, bien que la distorsion de la loi exponentielle soit appliquée, l'écart entre les primes obtenues dans les tranches élevées s'accroît en raison des différences de domaine d'attraction dans les queues de distribution. Cette limite reste toutefois mineure, car elle n'apparaît que dans la queue de distribution de la loi exponentielle distordue, qui fait partie intégrante de la queue de distribution de la sévérité globale.

Secondement, la loi de marché utilisée ici repose sur une hypothèse simplificatrice et ne correspond donc pas à un marché réel ou spécifique. Cette simplification a été adoptée pour se concentrer principalement sur la mise en place de la méthode de distorsion. Toutefois, pour une mise en œuvre pratique de cette approche, il serait nécessaire de construire

une loi de marché reflétant fidèlement la sinistralité d'un marché particulier.

En conclusion, ce mémoire présente une méthode avancée pour la modélisation de la sévérité des sinistres. Les approches proposées offrent ainsi une base solide pour une tarification plus juste sur le plan théorique, augmentant la confiance dans les résultats obtenus. Le schéma ci-dessous récapitule la méthode développée.

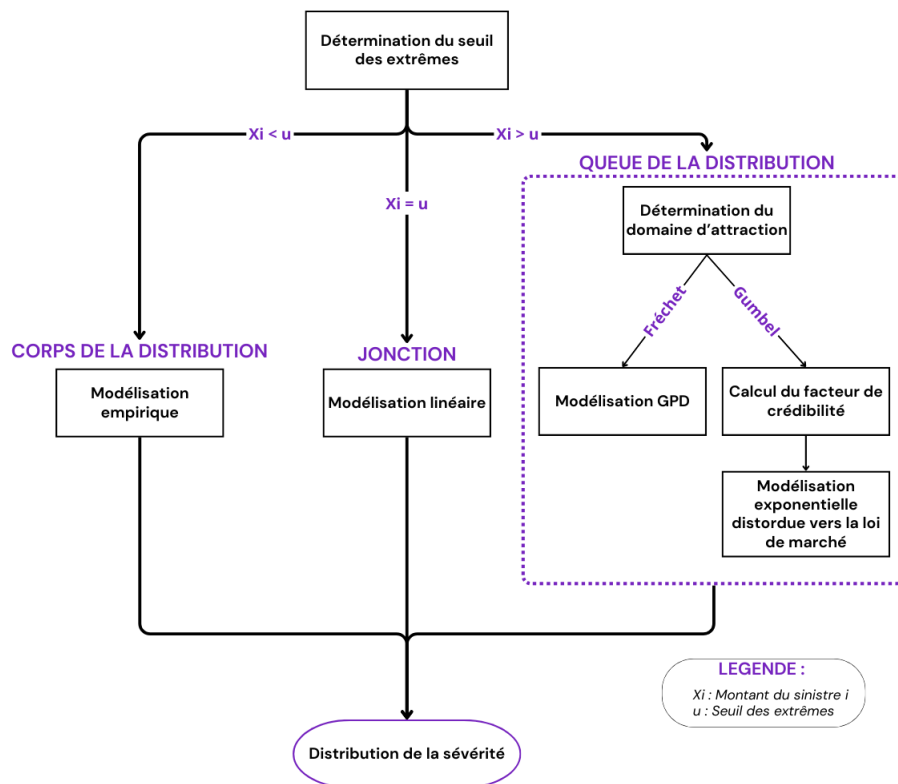


FIGURE 7 – Schéma récapitulatif de la méthode mise en place

Summary

In the face of sustained growth in the reinsurance market, accurate and rigorous pricing continues to be essential to ensure the competitiveness and relevance of the prices offered. In this context, it is imperative for reinsurers to have fair and reliable pricing methods, particularly in an environment where risks are increasingly complex and data is sometimes limited.

CCR Re, in its ambition to strengthen its competitiveness and align itself with market expansion, is working to improve the accuracy of pricing excess of loss treaties. More specifically, the aim is to optimise the use of the collective model in order to obtain reliable and robust results.

The collective model is based on two essential components : the frequency and severity of claims. Frequency, often represented by the Poisson distribution, models the number of claims expected over a given period, based on the cedant's historical data. Severity, on the other hand, models the cost of claims and is often represented by the Pareto distribution, which is particularly effective in capturing rare and costly claims. By combining these two elements, the collective model makes it possible to estimate the total cost of future claims, thus providing a solid basis for setting a risk-adjusted premium.

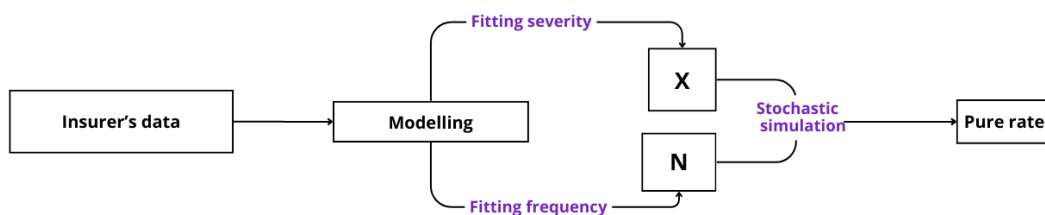


FIGURE 8 – Simplified collective model

This model has one notable limitation. By modelling all of the cedant's claims using a single distribution, it becomes difficult to accurately capture the risks located in the tail of the distribution. In fact, when a reinsurance layer is working, i.e. is impacted by several claims over the observation period, claim severity can be correctly captured by a single distribution covering all of the cedant's claims. On the other hand, for the non-working layers, which are often located in the high layers due to the low frequency of large claims, the actuary has to estimate the premium based on values located in the tail of the severity distribution. However, the current model, which uses a single distribution for all claims, is primarily calibrated on the lower part of the distribution, where most claims occur. As

a result, the most extreme part of the claims, found in the tail of the distribution, is not well modeled, making it difficult to accurately assess the risk in the upper layers. The aim of this thesis is to improve the claims severity distribution in order to refine the pricing of high layers.

To achieve this goal, in the second part of this thesis, the severity modelling was re-examined and adjusted using extreme value theory. The claims severity distribution is divided into two distinct parts : the body and tail, separated by an extremes threshold u . This threshold was determined using the mean excess plot and threshold stability plot tools. Body's distribution, which is generally well supplied with data, is modelled empirically. On the other hand, tail's distribution is modelled by a generalised Pareto distribution (GPD), which is ideal for extreme claims above the u threshold. The junction between these two parts is linear to ensure a smooth transition. This approach allows more reliable adjustment of extreme claims.

The graph below illustrates a cumulative function of a cedant's severity distribution using this method. This method is called composite severity.

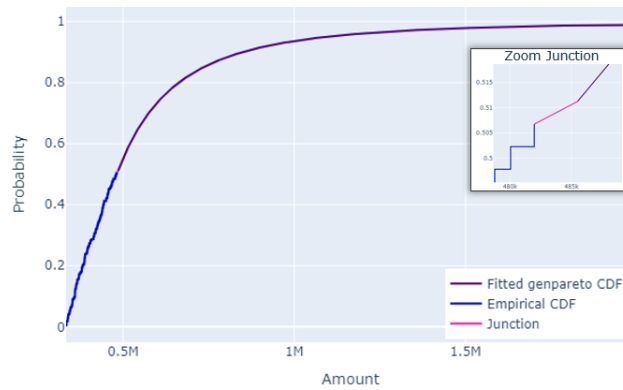


FIGURE 9 – Cumulative function for composite severity

This approach allows for more reliable and explainable rates in high layers, providing greater confidence in the calculated pure premiums. This process is summarized by the following diagram :

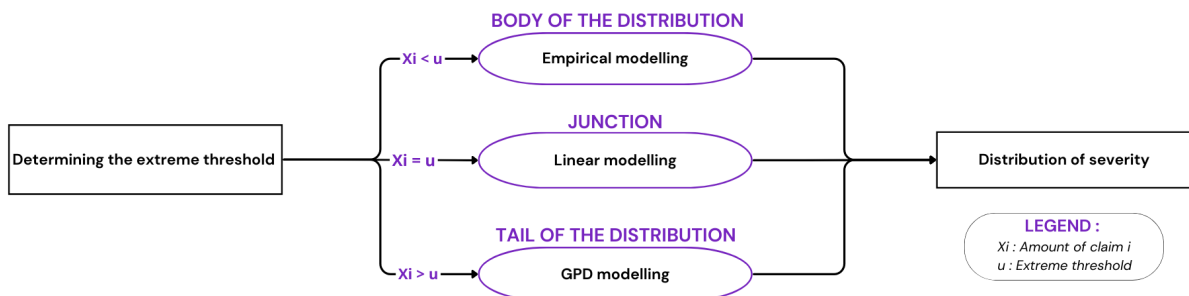


FIGURE 10 – Severity modelling diagram

In most reinsurance cases, the tail of cedant claim distribution lies in the Fréchet domain, characterised by heavy-tailed distributions. However, although this domain is the

most commonly encountered in pricing, it sometimes happens that the tail of the distribution belongs to the Gumbel domain, associated with thin-tailed distributions. This implies a lower probability of occurrence of extreme claims compared with Fréchet domain distributions. This particular case is more complex in reinsurance, as it would suggest a very limited or even non-existent need for reinsurance. However, the history of extreme claims in the market where the cedant's risks are exposed clearly shows that a risk still exists, even if this is not fully reflected in its historical data. An extreme loss could still occur in the future, despite this low theoretical probability.

Several examples have shown that the domain of attraction of the cedant's tail distribution can evolve over time, moving from the Gumbel domain to the Fréchet domain. This evolution justifies the existence of the risk and the relevance of the need for reinsurance.

From a theoretical point of view, when the tail is in the Gumbel domain, the Generalized Pareto Distribution reduces to an exponential distribution, which fits well with historical data. Nevertheless, this modeling does not adequately account for the risks associated with high layers, leading to an underestimation of future extreme claims. To address this limitation, arbitrary loadings based on expert judgment were added, but this method, while practical, lacked mathematical foundation and did not provide full confidence in the calculated rates for high layers.

This is the reason why, in the third part of this thesis, an approach based on distortion functions was introduced to correct mathematically the underestimation of extreme risks associated with the exponential distribution. Distortion functions, as statistical tools, can be used to transform probability distributions. The aim is to apply a distortion function to the adjusted exponential distribution. By modifying the distribution in this way, it would then be possible to better capture the extreme risks that are underestimated by the exponential distribution alone, while maintaining a solid theoretical foundation.

The distortion of exponential distribution is directed towards the market distribution. A market distribution is a statistical distribution developed to accurately represent claims behavior in a specific market, based on the country and the reinsurance branch involved. Such a well-established distribution serves as a solid basis because it incorporates both historical trends and market-specific features. The aim is therefore to distort the exponential distribution to converge with the corresponding market distribution. In this way, this distribution serves as a reference point for determining the maximum distortion that the exponential distribution can undergo, ensuring that this transformation remains relevant and aligned with the characteristics of the market.

However, this transformation cannot be uniform for all cedants. The degree of distortion must be adjusted according to the quality and relevance of each cedant's data. Therefore, a credibility factor is determined to calibrate the distortion based on the confidence placed in the cedant's data, considering them as a more or less accurate reflection of its future claims experience.

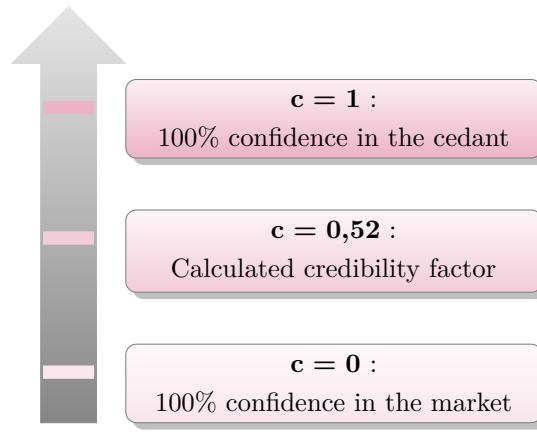


FIGURE 11 – Credibility factor

As illustrated in the diagram above, a credibility factor close to 1 reflects high confidence in cedant's data, characterised by few or no extreme claims. This suggests that, despite the lack or rarity of extreme claims, cedant's data provides a reliable estimate of future claims. This scenario is relevant if we consider that market in which these risks evolve also presents a limited number of extreme claims, making the cedant's data more reliable. In this case, the distortion applied to the exponential law will be minimal.

Conversely, a credibility factor close to 0 indicates low confidence in the data, meaning they are not seen as a good representation of future claims experience. This scenario is plausible when the market associated with this risks shows high extreme claims, contrary to what the cedant's data suggests. A more significant distortion will then be applied to the exponential distribution. The graphs below illustrate the cumulative distribution function of the distorted exponential distribution under different credibility factors (in blue) compared to that of the market distribution (in purple). The distortion function used is the proportional hazard (PH) function.

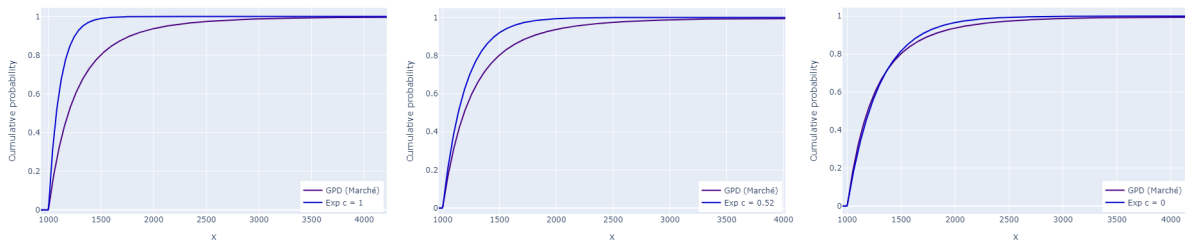


FIGURE 12 – Distorted cumulative function according to the credibility factor

Thus, the closer the credibility factor c is to 0, the more the distorted exponential distribution converges towards that of the market distribution.

This method was tested using simulations to compare the premiums obtained with and without distortion. The goal was to compare the estimates of the pure premium for each layer based on the following distributions :

- The cedant's exponential distribution (without distortion, $c=1$).
- The distorted exponential distribution ($c=0.52$, the calculated credibility factor).

- The maximally distorted exponential distribution ($c=0$).
- The market distribution (GPD).

The following graph illustrates the boxplots of the possible pure premiums for layer L1 : 1,000 XS 1,000, based on the chosen distribution. These boxplots provide an overview of the mean and volatility of pure premiums for each considered distribution. They also allow us to observe the evolution of the average pure premium depending on the distribution. Each point on the boxplots represents an average pure premium obtained through simulation.

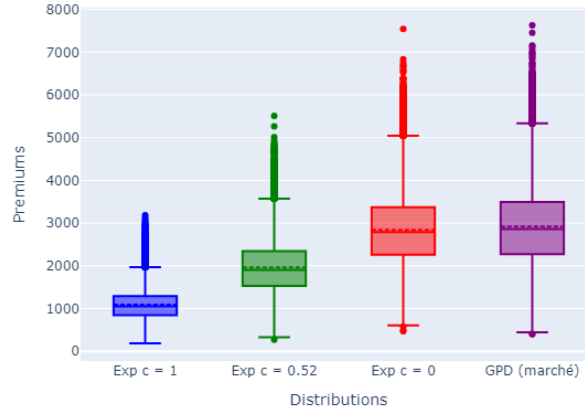


FIGURE 13 – Boxplot of T1 premiums according to the distribution

The distortion of the exponential distribution leads to an increase in the average pure premium, rising from €1,086 for the original distribution to €1,964 after applying the calculated credibility factor. As expected, this method results in higher premiums while maintaining consistency in the results. The distorted premium falls well between the original distribution premium and the market premium. This approach allows for better consideration of risks in the higher layers, while remaining faithful to the cedant's data.

However, this method has some limitations. First, although the distortion of the exponential distribution is applied, the gap between the premiums obtained in the higher layers increases due to differences in the domain of attraction in the distribution tails. This limitation remains minor, as it only appears in the tail of the distorted exponential distribution, which is part of the overall severity tail distribution.

Second, the market distribution used here is based on a simplifying assumption and does not correspond to a real or specific market. This simplification was adopted to focus primarily on implementing the distortion method. However, for a practical application of this approach, it would be necessary to construct a market distribution that accurately reflects the claims behavior of a particular market.

In conclusion, this thesis presents an advanced method for modelling the severity of claims. The proposed approaches provide a solid basis for more theoretically accurate pricing, increasing confidence in the results obtained. The diagram below summarises the method developed.

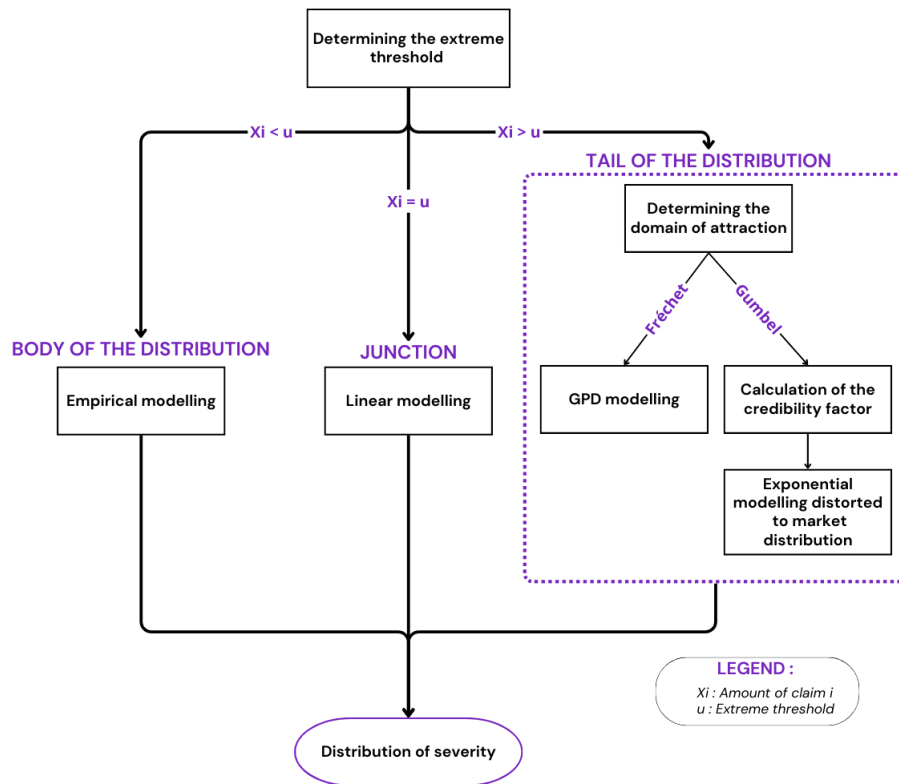


FIGURE 14 – Diagram of the method used

Sommaire

Remerciements	1
Résumé	2
Abstract	3
Synthèse	4
Summary	10
Table des figures	21
Introduction	22
 I Approches traditionnelles de tarification	 23
1 La réassurance	24
1.1 Présentation générale	24
1.2 Présentation de CCR Re	24
1.3 La réassurance proportionnelle	25
1.3.1 Quote Part	25
1.3.2 Excédent de plein	26
1.3.3 Les différentes clauses possibles	27
1.4 La réassurance non-proportionnelle	28
1.4.1 Excédent de sinistres	28
1.4.2 Excédent de perte annuelle	30
1.4.3 Les différentes clauses possibles	31
1.5 Domaine d'application	32
1.5.1 La réassurance non-vie	32
1.5.2 La distinction entre les branches	32
1.6 La prime de réassurance	33
1.6.1 Le taux pur	33
1.6.2 Le taux technique	34
1.6.3 Le taux commercial	34
 2 La tarification par expérience	 35
2.1 Le processus général de tarification	35
2.2 Mise en "as-if" des données	36
2.3 L'approche classique : Le burning cost	37
2.3.1 La méthode	37

2.3.2	Les limites	38
2.4	L'approche probabiliste : Le modèle collectif	38
2.4.1	Le modèle	38
2.4.2	Choix du seuil de modélisation	39
2.4.3	Modélisation de la fréquence	40
2.4.4	Modélisation de la sévérité	40
2.4.5	Les limites	42
3	Application du modèle collectif	43
3.1	Présentation du traité	43
3.2	Mise en "as-if"	44
3.2.1	Primes	44
3.2.2	Sinistres	44
3.2.3	Seuil de la statistique	45
3.3	Burning cost	45
3.3.1	Horizon de modélisation	46
3.3.2	Burning cost détaillé de la T1	46
3.3.3	Résultats burning cost	47
3.4	Modélisation	47
3.4.1	Choix du seuil de modélisation	47
3.4.2	Fréquence	48
3.4.3	Sévérité	48
3.5	Calcul de la prime pure	49
3.5.1	Simulation	49
3.5.2	Résultats	50
3.6	Limite et enjeu	51
II	L'approche par la théorie des valeurs extrêmes	52
4	La théorie des valeurs extrêmes	53
4.1	Convergence en loi du maximum d'un échantillon	53
4.1.1	Théorie classique	53
4.1.2	Etude de la loi du maximum	54
4.1.3	Loi des valeurs extrêmes généralisés	54
4.1.4	Les domaines d'attraction	55
4.2	La théorie des excédents au-delà d'un seuil	56
4.2.1	Etude de la loi des excès	56
4.2.2	Loi de pareto généralisé	56
4.3	Détermination du seuil des extrêmes	57
4.3.1	Le mean excess plot	58
4.3.2	Le threshold stability plot	59
4.4	Détermination du domaine d'attraction	60
4.4.1	QQ plot	60
4.4.2	Return level plot	62
4.4.3	Test du rapport de vraisemblance	63
4.5	Estimation paramétrique de la GPD	63
4.5.1	La méthode du maximum de vraisemblance	63
4.5.2	La méthode des moments	64
4.5.3	La méthode des moments pondérés	65

4.6	Estimation semi-paramétrique de l'indice des valeurs extrêmes	65
4.6.1	Estimateur de Hill	65
4.6.2	Estimateur de Pickands	66
5	Modélisation de la sévérité par la théorie des valeurs extrêmes	67
5.1	Détermination du seuil des extrêmes	67
5.1.1	Mean excess plot	67
5.1.2	Threshold stability plot	68
5.2	Domaine d'attraction	68
5.2.1	QQ plot	68
5.2.2	Test du rapport de vraisemblance	69
5.2.3	Return level plot	69
5.3	Construction de la courbe de sévérité	70
5.3.1	Le corps de la distribution	70
5.3.2	La queue de la distribution	70
5.3.3	Jonction	71
5.4	Calcul de la prime pure	71
5.4.1	Simulation	71
5.4.2	Résultats	75
5.5	Limites et enjeux	75
III	L'approche par distorsion	76
6	Modélisation de la queue de distribution pour le domaine de Gumbel	77
6.1	Problématique	77
6.1.1	Le domaine de Gumbel en réassurance	77
6.1.2	Exemple : du domaine de Gumbel au domaine de Fréchet	77
6.2	Cas pratique et théorique	79
6.2.1	Cas pratique : la tarification actuelle	79
6.2.2	Cas théorique : la tarification à partir d'une loi de marché	81
7	Les fonctions de distorsion	83
7.1	Présentation générale des fonctions de distorsion	83
7.1.1	Définition	83
7.1.2	Rôle dans la tarification	83
7.2	Exemples de fonction de disorsion	84
7.2.1	Fonction de distorsion de Wang	84
7.2.2	Fonction de distorsion de hasard proportionnel (PH)	84
7.3	Ecart minimal entre deux distributions	86
7.3.1	Divergence de Kullback-Leibler	86
7.3.2	Distance quadratique L^2	87
8	Application : l'approche de la distorsion pour la tarification	88
8.1	Approximation des données	88
8.2	La distorsion maximale	89
8.3	La distorsion idéale : le facteur de crédibilité	90
8.3.1	Interprétation du facteur	90
8.3.2	Construction du facteur	91
8.3.3	Distorsion selon le facteur de crédibilité	92

8.4	Résultats	93
8.5	Limites	95
8.5.1	L'écart des primes dans les tranches élevées	95
8.5.2	La loi de marché	96
Conclusion		97
Bibliographie		99
Annexes		101

Table des figures

1	Modèle collectif simplifié	4
2	Fonction de répartition de la sévérité composite	5
3	Schéma récapitulatif de la modélisation de la sévérité	5
4	Le facteur de crédibilité	7
5	Fonction de répartition distordue selon le facteur de crédibilité	7
6	Boxplot des primes de la T1 selon la loi	8
7	Schéma récapitulatif de la méthode mise en place	9
8	Simplified collective model	10
9	Cumulative function for composite severity	11
10	Severity modelling diagram	11
11	Credibility factor	13
12	Distorted cumulative function according to the credibility factor	13
13	Boxplot of T1 premiums according to the distribution	14
14	Diagram of the method used	15
1.1	Organigramme du département direction actuariat et risques	25
1.2	Répartition du portefeuille de la cédante pour un QP	26
1.3	Répartition du portefeuille de la cédante pour un EP	27
1.4	Répartition de la sinistralité pour un XS par risque	29
1.5	Répartition de la sinistralité pour un XS par événement	30
1.6	Répartition de la perte annuelle pour un SL	31
2.1	Le processus de tarification	35
2.2	Les différentes étapes de la méthode burning cost	38
2.3	Densité de la loi de Pareto selon α	41
2.4	Récapitulatif de l'approche probabiliste	42
3.1	Traité de réassurance	43
3.2	Evolution des primes	44
3.3	Evolution des sinistres	45
3.4	Burning cost T1	46
3.5	Récapitulatif BC	47
3.6	Fonction de répartition empirique et de la loi de Poisson	48
3.7	Tableau ajustement sévérité	49
3.8	Fonction de répartition empirique et de la loi de Pareto	49
3.9	Résultats de la tarification du traité	50
4.1	Densité de la loi GEV selon γ	55
4.2	Densité de la loi GPD selon γ et σ	57
4.3	Mean excess plot	59
4.4	Mean excess plot d'une GPD(0.7, 600 000) avec 500 observations (50 à droite).	59

4.5	Threshold stability plot (pour γ à gauche et pour σ à droite)	60
4.6	QQ Plot : Fréchet à gauche, Gumbel au centre et Weibull à droite	61
4.7	Return level plot	62
5.1	Mean excess plot	67
5.2	Threshold stability plot	68
5.3	QQ plot exponentiel	69
5.4	Return level plot	69
5.5	Tableau ajustement GPD	70
5.6	QQ plot GPD	71
5.7	Fonction de répartition composite	71
5.8	Tableau taux et primes pures	75
6.1	QQ plot (gauche) et return level plot (droite)	78
6.2	Return level plot	79
7.1	Fonction de distorsion de Wang sur l'exponentielle	84
7.2	PH transform sur l'exponentielle	85
7.3	Fonction de répartition distordue maximale (KL)	86
7.4	Fonction de répartition distordue maximale (L2)	87
8.1	Fonction de répartition cédante et marché	89
8.2	Fonction de répartition distordue maximale (KL)	89
8.3	Facteur de crédibilité	90
8.4	Probabilité de γ	91
8.5	Fonction de répartition de l'exponentielle selon le facteur de crédibilité	93
8.6	Structure fictive du traité	94
8.7	Boxplot des primes de la T1 selon la loi	94
8.8	Boxplot des primes de la T3 selon la loi	95
8.9	mean excess plot exemple	104
8.10	Stability plot exemple	104

Introduction

Le marché de la réassurance connaît une expansion continue. Selon les prévisions, le volume des primes émises devrait passer de 444 milliards USD en 2024 à 592 milliards USD d'ici 2029 [MI24]. Cette croissance est alimentée par l'augmentation de la demande en assurance à l'échelle mondiale et par la nécessité de mieux se protéger contre les catastrophes naturelles. Face à ces évolutions, les réassureurs, comme CCR Re, doivent s'adapter en optimisant la tarification des risques pour maintenir leur compétitivité et garantir une gestion rigoureuse des sinistres futurs.

L'un des principaux enjeux pour les réassureurs est de proposer une tarification aussi précise et rigoureuse que possible. Une tarification juste et fiable est essentielle pour établir des taux pertinents, reflétant fidèlement les risques couverts. Cet enjeu de précision est crucial pour tous les types de traités, car une tarification inadéquate peut non seulement exposer les réassureurs à des pertes significatives ou à une sous-couverture des risques, mais également entraîner une perte de compétitivité du réassureur sur le marché. Une évaluation précise permet ainsi d'appréhender le risque de manière optimale.

Cet enjeu devient particulièrement complexe dans le cas des traités en excédent de sinistres (XS), notamment lorsqu'il s'agit des tranches élevées. Ces dernières sont fréquemment exposées à des sinistres extrêmes, qui sont rares mais très coûteux. Le principal défi ici réside dans la modélisation de ces sinistres, car la faible fréquence de ces événements conduit à un manque de données historiques, rendant la précision et la fiabilité de la tarification de ces tranches difficile. En conséquence, les réassureurs doivent faire face à une grande incertitude lorsqu'ils évaluent les risques et fixent les taux pour ces tranches.

Traditionnellement, la tarification des traités XS par expérience est réalisée avec le modèle collectif. Ce modèle combine l'estimation de la fréquence des sinistres et de leur sévérité, entre la priorité et la portée du traité. Bien que cette méthode soit largement utilisée, elle présente certaines limites, notamment pour modéliser de manière adéquate les sinistres extrêmes qui ont un impact majeur sur les tranches élevées des traités XS.

L'objectif de ce mémoire est d'optimiser cette méthode traditionnelle en améliorant la modélisation des sinistres extrêmes afin d'obtenir des taux plus fiables et plus explicables. Cela permettra de mieux refléter la sinistralité future possible, et de proposer une tarification plus robuste pour ces tranches.

Pour répondre à cette problématique, le mémoire est structuré en trois sections. La première section présente en détail la méthode traditionnelle de tarification des XS et ses limites dans le contexte des tranches élevées. La deuxième section explore l'utilisation de la théorie des valeurs extrêmes pour améliorer la modélisation des sinistres rares. Elle vise ainsi à mieux relever les défis posés par ces tranches. Enfin, la troisième section propose une approche innovante, en introduisant le principe de distorsion. Celle-ci sera appliquée à un cas particulier de la modélisation des sinistres extrêmes, pour renforcer la fiabilité des taux dans un cadre particulier.

Première partie

Approches traditionnelles de tarification

Chapitre 1

La réassurance

1.1 Présentation générale

La réassurance, souvent décrite comme "l'assurance des assureurs", constitue un mécanisme financier essentiel. Elle offre à une compagnie d'assurance, appelée cédante, la possibilité de transférer une partie de ses risques à une autre entité spécialisée, le réassureur.

Concrètement, la réassurance se matérialise par un accord contractuel entre la cédante et la société de réassurance, où la cédante peut céder une fraction de ses risques assurés en échange du paiement d'une prime. En vertu de cet accord, le réassureur s'engage à rembourser l'entière ou une partie des sommes versées aux assurés en cas de sinistre, soulageant ainsi la pression financière sur la cédante.

Le contrat qui encadre les relations entre la cédante et le réassureur est généralement désigné sous le terme de traité de réassurance. Ce dernier définit les conditions, les limites et les obligations de chaque partie, offrant ainsi un cadre clair pour la protection des intérêts de la cédante et du réassureur.

Les principaux avantages de la réassurance

Fondamentalement, son rôle premier est avant tout de répartir les risques, réduisant ainsi l'exposition financière des assureurs. De plus, elle leur offre une flexibilité accrue pour faire face à des événements majeurs tels que les catastrophes naturelles, les risques extrêmes, ainsi que les défis émergents tels que les cyber-risques. Ainsi, la réassurance agit comme un soutien financier, permettant aux assureurs de couvrir des risques qui dépassent leur capacité de souscription.

De plus, la réassurance offre la possibilité de diversifier géographiquement les risques, tirant avantage de la portée internationale de ses opérations.

Pour finir, la réassurance joue un rôle crucial pour stabiliser le résultat annuel de l'assureur en atténuant la volatilité de sa sinistralité, assurant ainsi une vision financière plus stable à long terme.

1.2 Présentation de CCR Re

CCR Re (future Arundo Re), anciennement une filiale du groupe CCR, a été établie en 2016 à l'initiative de l'État français. Depuis le 1^{er} juillet 2023, le groupe CCR a cédé 75% de CCR Re au consortium formé par les groupes SMABTP et MACSF. Ces deux entités renforcent la position de CCR Re sur le marché de la réassurance et contribuent

à son expansion en apportant leur savoir-faire et leurs particularités dans le domaine de l'assurance.

CCR Re est un acteur mondial de la réassurance, présent dans plus de 90 pays et offrant une couverture complète dans les principales branches de l'assurance : Vie, Non-Vie et spécialités. En 2023, CCR Re a généré un chiffre d'affaires de 1 186 millions d'euros avec un résultat net de 56 millions d'euros, marquant une croissance de 20 % par rapport à l'année précédente.

Le service de tarification et aide à la souscription non-vie

Le service de tarification et aide à la souscription non-vie, auquel je suis rattachée, appartient à la direction actuariat et risques. Il est supervisé par le département cotation et cellule CAT. Son objectif principal est de fournir une expertise technique sur les traités de réassurance notamment en proposant un prix pour chaque contrat. En parallèle, des projets de recherche sont effectués afin d'améliorer les outils et les méthodes utilisées. C'est au sein de ce service que ce mémoire a été réalisé.

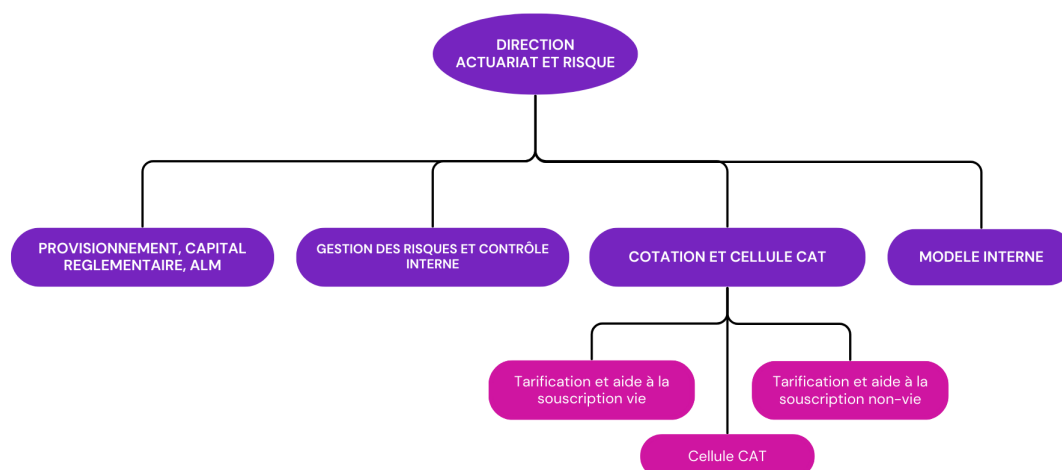


FIGURE 1.1 – Organigramme du département direction actuariat et risques

En réassurance, il convient de différencier deux types de traités : les traités dits de réassurance proportionnelle et ceux de réassurance non proportionnelle.

1.3 La réassurance proportionnelle

Comme son nom l'indique, un réassureur prend en charge une proportion d'un risque moyennant une proportion identique de la prime payée par l'assuré et paie, en cas de survenance du risque, le sinistre dans la même proportion.

Il existe deux formes de réassurance proportionnelle :

1.3.1 Quote Part

Désigné par l'acronyme "QP" ou "QS" (pour *quota share* en anglais), ce type de traité représente la forme la plus intuitive de la réassurance. Dans ce cadre, l'assureur cède une proportion constante, désignée par le taux de cession, sur tous les risques définis de son portefeuille. En échange de ce service, le réassureur reçoit la même part de primes.

Ainsi, l'assureur transfère la même proportion de risque, que ce soit pour des risques faibles ou élevés. Cela signifie que le profil du portefeuille conservé par l'assureur est similaire à celui accepté par le réassureur.

Exemple :

- QP taux de cession 70%.
- Portefeuille composée des profils de risque de A à F.

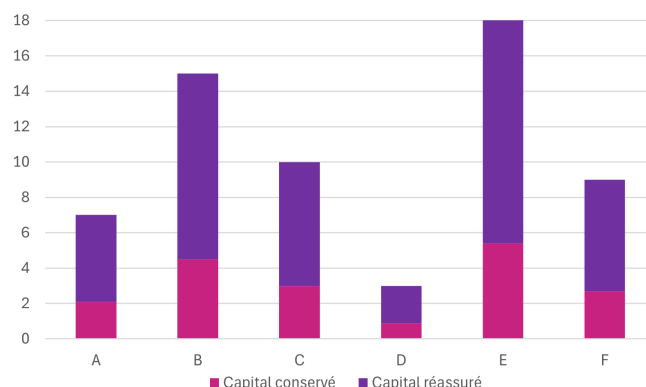


FIGURE 1.2 – Répartition du portefeuille de la cédante pour un QP

Cet exemple illustre que le QP permet de réduire les risques supportés par l'assureur sans pour autant qu'il y ait un phénomène de nivellement. Les sinistres, qu'ils soient de faible ou de grande sévérité, sont partagés proportionnellement entre les deux parties.

1.3.2 Excédent de plein

Désigné par l'acronyme "EP" ou "SP" (pour *surplus* en anglais), ce type de traité permet à l'assureur de conserver les risques à faible coûts dans son portefeuille. Dans ce cadre, le réassureur prend en charge uniquement la portion des risques dépassant un niveau de capital, appelé plein de rétention R . En échange il reçoit, sur chaque risque concerné, la proportion de prime correspondant à la part du dépassement acceptée dans le capital assuré.

Le mécanisme de l'excédent de plein fonctionne de manière similaire à celui du QP, à l'exception près que le taux de cession varie pour chaque risque. Un pourcentage d'excédent par rapport au capital assuré est déterminé pour chaque risque, définissant ainsi le taux de cession applicable à chaque cas. Ce taux entre en jeu uniquement si le capital assuré CA du risque dépasse le seuil de rétention R , et il est calculé dans les limites de la capacité de l'excédent de plein C . Mathématiquement, le taux de cession du i ème risque τ_i est défini comme suit :

$$\tau_i = \begin{cases} 0 & \text{si } CA_i < R \\ \frac{CA_i - R}{C} & \text{si } R \leq CA_i < C \\ \frac{C}{CA_i} & \text{sinon.} \end{cases}$$

Exemple : (en euros)

- Traité EP avec $R = 3$ millions.
- 3 pleins, $C = 9$ millions.
- Portefeuille identique à l'exemple du QP.

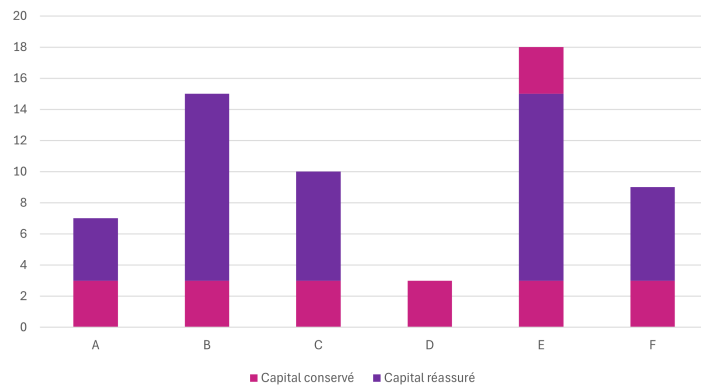


FIGURE 1.3 – Répartition du portefeuille de la cédante pour un EP

Ainsi, le portefeuille retenu a été nivelé et les pointes ont bien disparu. En effet, étant donné que $CA < R$ pour le risque D, ce dernier n'est pas inclus dans la couverture.

L'EP permet à l'assureur de garder un engagement limité tout en ne cédant pas tous ses risques. Il permet donc à l'assureur de déterminer par nature d'affaires la rétention qu'il estime souhaitable de conserver pour chaque compte.

1.3.3 Les différentes clauses possibles

Dans un traité proportionnel, de nombreuses clauses peuvent être appliquées afin de définir les modalités de partage des risques entre l'assureur et le réassureur. Bien qu'il existe de nombreuses clauses possibles, seules les principales sont présentées dans ce mémoire.

La commission

La commission représente une partie de la prime d'assurance reversée par le réassureur à la cédante. Elle participe aux frais d'acquisition et de gestion des affaires cédées. En d'autres termes, en payant une commission de réassurance, le réassureur contribue aux frais déployés par l'assureur pour l'acquisition de ses polices. Il est également important de noter que le véritable coût de la réassurance se négocie en grande partie à travers la commission.

Le taux de commission varie selon la branche, la nature du traité, la pratique du marché concerné et les résultats des exercices passés. En réassurance, la commission peut être de nature fixe ou à échelle, en fonction des accords entre les parties.

- Commission fixe : Dans ce cas, la commission est un montant fixe prédéterminé convenu entre les parties. Cette participation aux frais se fait sous la forme d'une commission égale à un pourcentage des primes versées.
- Commission à échelle : Contrairement à la commission fixe, la commission à échelle varie en fonction des sinistres. L'échelle de commission évolue en fonction inverse de la sinistralité enregistrée. La sinistralité est déterminée par le ratio $\frac{S}{P}$.

$$\frac{S}{P} = \frac{\text{Coût des sinistres}}{\text{Assiette de primes perçues}}$$

Exemple : La commission à échelle peut se représenter de la manière suivante.

Minimum	30%	Pour un S/P de	58%
A	43.5%	Pour un S/P de	46%
A	52.5%	Pour un S/P de	36%

TABLE 1.1 – Commission à échelle du traité

Le choix entre une commission fixe ou à échelle dépend souvent des préférences des compagnies impliquées et de la complexité des arrangements de réassurance.

La participation aux bénéfices

La participation aux bénéfices, abrégée par l'acronyme PB, est un accord établi entre le réassureur et l'assureur pour partager une partie des bénéfices réalisés par le réassureur. Selon les termes de cette clause, si le réassureur génère des bénéfices sur les contrats de réassurance, une part de ces bénéfices sera partagée avec la compagnie d'assurance. Ce partage peut être calculé en fonction d'un pourcentage prédéfini ou d'une formule spécifique convenue entre les deux parties. La formule utilisée est la suivante :

$$PB = \max(0, XX\% [P - (C + E + DC)])$$

XX % : taux de PB C : sinistres DC : report de pertes
P : primes cédées E : chargements

Cette clause encourage une souscription responsable et incite la cédante à améliorer la qualité de son portefeuille pour accroître ses profits.

La participation aux pertes

La participation aux pertes, abrégée par l'acronyme PP, permet au réassureur de partager les pertes subies avec la cédante. Lorsque les sinistres dépassent un certain seuil défini dans le contrat, la participation aux pertes entre en jeu. Ainsi l'assureur s'engage à contribuer financièrement pour couvrir une partie des pertes excédentaires.

1.4 La réassurance non-proportionnelle

La réassurance non-proportionnelle constitue le deuxième type de contrat de réassurance, offrant une solution aux limites de la réassurance proportionnelle. Contrairement à cette dernière, elle ne se base pas sur les capitaux assurés mais sur les sinistres survenus.

Ce modèle implique qu'un réassureur s'engage à verser à l'assureur un certain montant uniquement si un événement spécifique se produit et dépasse un certain seuil. En échange de cette couverture, le réassureur perçoit une prime visant à compenser le risque assumé.

Il convient de noter qu'il existe deux grandes catégories de couverture pour cette forme de réassurance : les traités en excédent de sinistres et les traités en excédent de pertes.

1.4.1 Excédent de sinistres

Désigné par l'acronyme "XS" ou "XL" (pour *excess of loss* en anglais), ce type de traité implique que le réassureur s'engage à indemniser la portion du montant du sinistre

dépassant un seuil prédéfini, appelé franchise F .

Cependant, le réassureur ne versera pas plus que la couverture fixée, connue sous le nom de portée P .

Ainsi, la notation de la couverture du traité prend la forme de "Portée XS Franchise" ou encore "P XS F".

La charge du réassureur C_{XS} peut être calculée à l'aide de la formule suivante (où X représente le montant total du sinistre) :

$$C_{XS} = \min(\max[X - F; 0]; P)$$

Trois formes d'XS sont distinguées :

1. **L'XS par risque** : Dans ce type de contrat, l'assureur s'engage à verser une indemnisation dès qu'une police est touchée par un sinistre excédant le montant de la franchise.

Exemple : (en euros)

- XS par risque : 100 M XS 50 M.
- Liste des sinistres : [20 ; 90 ; 50 ; 170 ; 100].

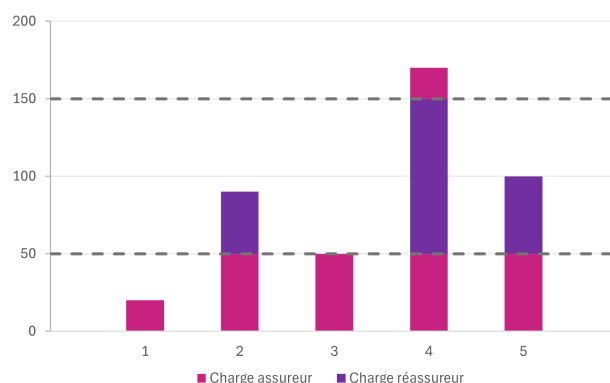


FIGURE 1.4 – Répartition de la sinistralité pour un XS par risque

La charge totale de l'assureur s'élève ainsi à 240 millions d'euros, les sinistres de moindre importance ne sont pas couverts par le réassureur. En effet, le contrat est déclenché pour les sinistres n°2, n°4 et n°5 car ils dépassent la franchise. Néanmoins, le sinistre n°4 excède également la limite de couverture du contrat. Par conséquent, l'assureur supporte une charge de 70 millions d'euros pour ce sinistre.

2. **L'XS par événement** : Ce type d'XS diffère du précédent en ce sens que l'événement n'est plus déterminé par une seule police sinistrée. Un événement est représenté par la survenance de plusieurs sinistres résultant d'une même cause. Cette cause peut être un incendie, une inondation, des émeutes, etc. Il est essentiel de déterminer clairement l'événement à travers des clauses précises afin de définir les sinistres pris en charge.

Exemple : (en euros)

- XS par événement : 100 M XS 50 M.
- Événement n°1 = 20 + 90 + 50 = 160 M.
- Événement n°2 = 170 + 100 = 270 M.

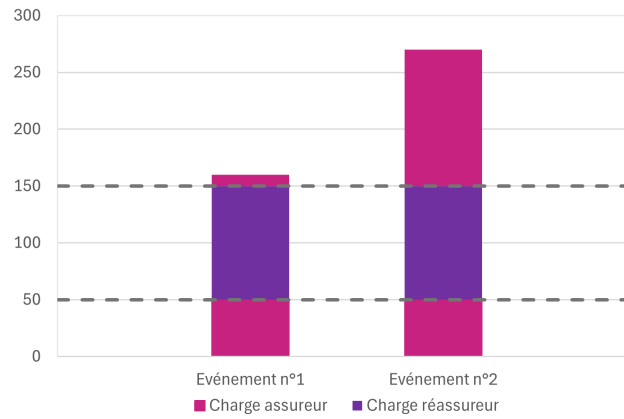


FIGURE 1.5 – Répartition de la sinistralité pour un XS par événement

Comme les sinistres sont cumulés dans des événements distincts, les montants de ces événements sont considérables. Cela entre donc en jeu dans la couverture de l'XS. Par conséquent, la charge supportée par l'assureur s'élève à 230 millions d'euros.

3. **L'XS par risque et par événement** : Cette dernière forme d'XS combine les deux types de sinistralité afin de fournir à l'assureur tous les avantages possibles. Elle lui offre la flexibilité de choisir entre le traitement par risque ou par événement selon ses besoins et préférences.

1.4.2 Excédent de perte annuelle

Désigné par l'acronyme "SL" (pour *stop loss* en anglais), ce type de contrat fonctionne de manière similaire à l'XS. Cependant, son objectif principal est de protéger l'assureur contre des résultats défavorables. Contrairement à l'XS, les bornes de ce traité ne concernent pas les montants des sinistres, mais la sinistralité globale de l'assureur. En général, ils sont exprimés avec le rapport sinistres sur primes (ou plus rarement sinistres sur capitaux).

La charge du réassureur C_{SL} peut être calculée à l'aide de la formule suivante :

$$C_{SL} = \min \left(\max \left(\frac{\sum_{i=1}^n X_i}{EPI} - F, 0 \right), P \right) \times EPI$$

Avec :

- X_i : le montant du $i^{\text{ème}}$ sinistre.
- n : le nombre annuel de sinistres.
- EPI : la somme des primes acquises par la cédante.

Exemple : (en euros)

- Traité SL : 50% XS 50%.
- Sinistralité annuelle sur 5 ans (en millions) : [20 ; 90 ; 50 ; 170 ; 100].
- EPI annuelle sur 5 ans (en millions) : [100 ; 140 ; 130 ; 80 ; 90].

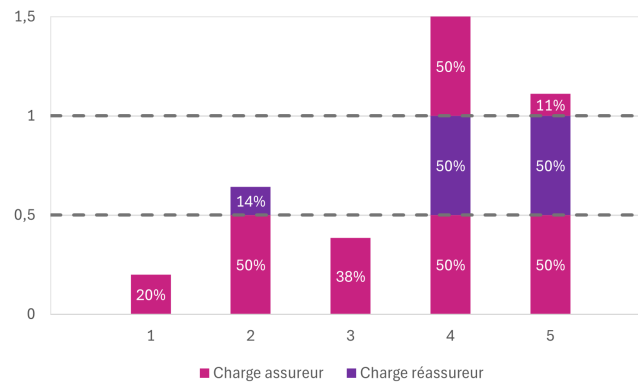


FIGURE 1.6 – Répartition de la perte annuelle pour un SL

Une fois les ratios sinistres/primes calculés, la répartition des pertes annuelles entre l'assureur et le réassureur s'effectue de la même manière que pour l'XS.

Pour l'XS et le SL, il est fréquent que les traités soient divisés en plusieurs tranches s'imbriquant les unes sur les autres. Ce processus de « layering » offre la possibilité de couvrir différents risques selon les tranches. Cela permet également une tarification plus précise et détaillée. En effet, certains sinistres peuvent déclencher uniquement certaines tranches, ce qui permet de mieux distinguer les sinistres de grande et de faible sévérité. Enfin, cette méthode permet d'optimiser l'allocation de capital.

1.4.3 Les différentes clauses possibles

Dans un traité non-proportionnel, de nombreuses clauses peuvent être appliquées afin de définir les modalités de partage des risques entre l'assureur et le réassureur. Bien qu'il existe de nombreuses clauses possibles, seules les principales sont présentées dans ce mémoire.

La franchise annuelle

La clause de franchise annuelle agrégée, connue sous l'acronyme "AAD" (pour *annual aggregate deductible*), restreint l'engagement du réassureur. En vertu de cette clause, l'assureur supporte les premiers sinistres dépassant la priorité avec pour limite l'AAD. Par conséquent, l'intervention du réassureur est retardée, ce qui contribue à réduire la prime de réassurance.

La limite annuelle

La limite annuelle agrégée, désignée par l'acronyme "AAL" (pour *annual aggregate limit*), est une clause permettant au réassureur de limiter son engagement. Elle représente la garantie annuelle maximale accordée par le réassureur, devant être au minimum égale à la portée du traité. Une fois que le réassureur a atteint cette garantie maximale, il cesse de prendre en charge tout sinistre supplémentaire. Cette clause contribue ainsi à réduire la prime de réassurance.

La clause de reconstitution

Cette clause vise à restreindre l'engagement du réassureur. Contrairement à l'AAL, l'engagement du réassureur est déterminé par un nombre de reconstitutions n variant de 0 à ∞ . Ce dernier définit combien de fois la portée P du traité est applicable. L'AAL est

calculée par la formule suivante ici :

$$AAL = (1 + n) * P$$

La prime de base de réassurance est conçue pour couvrir la portée initiale d'un XS. En cas de reconstitution gratuite, l'assureur peut prolonger la couverture sans paiement supplémentaire, ce qui permet au réassureur de limiter son exposition sans modifier la prime initiale. En revanche, lors d'une reconstitution payante, l'assureur verse une prime de reconstitution (P_r), proportionnelle à la portion de couverture supplémentaire absorbée par la tranche. Cette prime se calcule ainsi :

$$P_r = \frac{P_c}{P} \times T \times c$$

Avec :

- P_c : la partie de la portée consommée par le sinistre.
- T : la prime commerciale.
- c : le coefficient forfaitaire $\geq 0\%$.

Dans un traité non-proportionnel, il est envisageable que la clause de stabilisation ou la clause d'indexation soit également incluse. Ces clauses peuvent être appliquées lorsque le risque couvert appartient à une "branche longue" (notion expliquée ultérieurement, cf. 1.5.2). Toutefois, ce mémoire se focalise sur l'étude des risques en "branche courte". Par conséquent, ces clauses ne seront pas abordées.

1.5 Domaine d'application

1.5.1 La réassurance non-vie

De la même manière qu'en assurance « traditionnelle », la réassurance se divise en deux axes distincts :

- La réassurance vie.
- La réassurance non-vie comprenant :
 - Le domaine IARD (Incendie, Accidents et Risques Divers).
 - Le domaine de la santé, prévention et dépendance.

La réassurance non-vie représente une part significative, totalisant 68% des primes brutes cédées en réassurance en 2022 à l'échelle mondiale [AP23]. Ce secteur englobe plusieurs branches telles que l'automobile, l'incendie, la construction, etc.

C'est sur cet axe que se concentre ce mémoire.

1.5.2 La distinction entre les branches

Il est essentiel de distinguer deux types de branches en réassurance non-vie : la branche à développement long (branche longue) et celle à développement court (branche courte). Cette notion est nécessaire car la tarification sera différente selon la branche concernée.

Pour ce faire, il est nécessaire de connaître plusieurs dates clés dans le processus d'assurance :

- **La date de souscription** : qui marque le début de la couverture pour l'assuré.
- **La date de survenance du sinistre** : qui doit être postérieure à la date de souscription pour être prise en charge.
- **La date de déclaration du sinistre** : qui peut différer de la date de survenance du sinistre.

- **La date de règlement du sinistre :** Le moment où l'assureur et le réassureur effectuent le versement des montants qui leur sont respectivement dus pour régler le sinistre.

La différenciation entre une branche longue et une branche courte s'évalue selon les délais entre ces dates.

Branche courte

En branche courte, la déclaration des sinistres se fait rapidement et l'estimation du montant des dommages est fixée en quelques mois. De plus, le règlement est généralement rapide, sans imprévus.

Les assurances incendie et vol font partie de cette catégorie.

Branche longue

En branche longue, la déclaration d'un sinistre peut apparaître 10 ans après sa survenance. De plus, l'évaluation du montant final des dommages peut prendre plusieurs années. La responsabilité civile (abrégé par l'acronyme "RC") est l'une des branches appartenant à ce domaine.

Par exemple, dans le cas d'une assurance RC professionnelle, les conséquences d'une erreur commise par un obstétricien lors de la naissance d'un nouveau-né peuvent ne se manifester qu'à l'âge de la puberté de l'enfant. De plus, l'état de santé de l'enfant peut se détériorer avec le temps, entraînant une augmentation du montant total à la charge de l'assureur.

En outre, le règlement complet du sinistre peut prendre un temps considérable. Généralement, les paiements s'étalent sur le long terme et fluctuent au fil des années. Une distinction est établie entre les sinistres déjà indemnisés, désignés par "SP", et le solde restant à payer, appelé "SAP". Le montant total "incurred" du sinistre est ainsi la somme de SP et SAP. Ce montant n'est pas fixe et peut connaître des fluctuations significatives au fil du temps.

Dans les branches d'assurance à développement long, les clauses de stabilisation et d'indexation peuvent être appliquées sur long terme.

Cette branche d'assurance exige une méthodologie particulière de la part de l'actuaire pour évaluer le montant total du sinistre. Toutefois, cette méthode ne sera pas abordée dans ce mémoire, car l'analyse se concentre sur une branche d'assurance à développement court.

1.6 La prime de réassurance

En réassurance, le service de tarification a pour but de déterminer la prime de réassurance. Celle-ci revêt une importance capitale pour évaluer avec précision le risque du traité. Habituellement exprimée sous forme de taux, elle est ensuite appliquée à l'assiette de prime estimée **EPI** pour l'année de couverture afin d'obtenir la prime finale.

Trois types de taux sont utilisés pour la construction de cette prime :

1.6.1 Le taux pur

Premièrement, la prime pure représente l'estimation du coût moyen des sinistres assumés par le réassureur. Son évaluation relève de l'expertise de l'actuaire tarificateur, qui a recours à diverses méthodes statistiques et probabilistes adaptées au type de traité et

aux données disponibles.

Pour les traités proportionnels, le calcul de la prime pure est direct. Il suffit en effet, de déterminer le taux de cession pour prélever la part correspondante sur les primes d'assurance de la cédante.

En revanche, pour les traités non-proportionnels, la méthodologie varie selon la quantité et la nature des données. La détermination du taux pur n'est pas immédiate et requiert une ou plusieurs méthodes de tarification.

1.6.2 Le taux technique

Le taux technique est calculée à partir du taux pur et prend en considération divers chargements, coûts et frais supportés par le réassureur. Les éléments suivants sont pris en compte :

1. **Le chargement de sécurité** : Noté C_s , ce supplément est ajouté au taux pur pour compenser la volatilité des sinistres de ce traité. Plus les données seront limitées, plus la volatilité du résultat sera grande.
2. **Les frais de gestion** : Noté F_g , ces frais couvrent les diverses dépenses supportées par le réassureur pour la gestion de ce traité. Par exemple, la gestion des sinistres représente un coût supplémentaire pour lui.
3. **Les frais de courtage** : Noté F_c , ces frais permettent au courtier de percevoir une rémunération pour les contrats qu'il a réussi à souscrire pour la cédante.
4. **Le coût du capital** : Noté C_k , ce coût représente la rémunération exigée par les actionnaires du réassureur pour l'immobilisation des fonds nécessaires à la couverture des risques. Lors de la souscription d'un contrat, le réassureur doit allouer un capital suffisant pour couvrir la volatilité des résultats attendus. Ce capital, essentiel pour absorber les éventuelles pertes, doit être rémunéré de manière adéquate aux actionnaires.

Le calcul du taux technique est donné par la formule suivante :

$$\text{Taux technique} = \frac{\text{Taux pur} \times (1 + C_s) + C_k}{1 - F_c - F_g}$$

1.6.3 Le taux commercial

Le taux commercial correspond au taux de réassurance facturée à la cédante sur son EPI. Ce taux peut différer du taux technique en fonction de la politique de souscription en place. La prime commerciale est ensuite calculé en multipliant le taux commercial sur l'EPI estimé pour l'année de couverture.

Dans ce mémoire, l'étude se focalisera sur la détermination du taux pur des traités non-proportionnels. À présent, examinons quelques méthodes de tarification qui permettent d'aboutir à ce taux pur.

Chapitre 2

La tarification par expérience

La tarification des contrats de réassurance revêt une importance capitale pour appréhender et évaluer les risques couverts. Différentes méthodes sont utilisées pour évaluer ces contrats, notamment la tarification par expérience et la tarification par exposition. Ces deux approches distinctes peuvent également être combinées.

La tarification par exposition repose sur l'évaluation du risque en fonction des profils de risque de la cédante, afin de déterminer son exposition. Cette méthode n'est pas abordée dans ce mémoire.

La tarification par expérience représente l'approche la plus simple et intuitive en réassurance. Elle repose sur l'analyse de l'historique des sinistres et des primes de la cédante. Cette méthode suppose que l'expérience passée en matière de sinistralité reflètera au mieux le comportement du portefeuille de l'assureur.

Deux approches peuvent être utilisées :

1. L'approche classique, le burning cost ¹.
2. L'approche probabiliste, le modèle collectif.

2.1 Le processus général de tarification

Il est primordial de comprendre le processus de tarification avant d'examiner ces différentes approches.

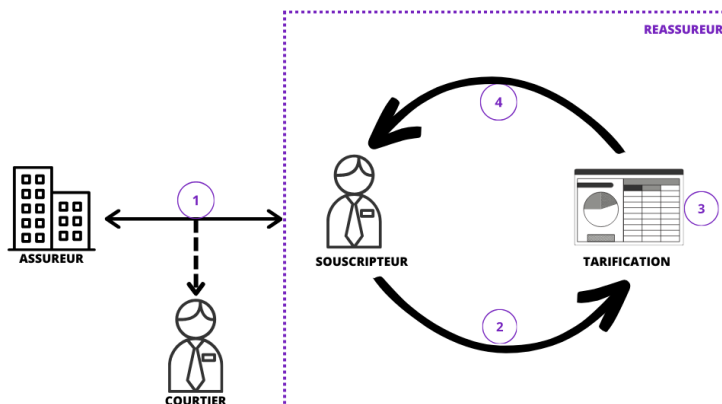


FIGURE 2.1 – Le processus de tarification

1. "Coût de brûlure" en français.

PROCESSUS DE TARIFICATION CHEZ CCR RE :

Etape 1 : Le souscripteur reçoit la demande de réassurance d'une cédante.

Il est en contact avec le courtier ou avec la cédante directement.

Ils déterminent les termes du contrat : Le slip et le wording.

Etape 2 : Le souscripteur envoie une demande de cotation. Il transmet les éléments suivants :

Le slip : détails du type de couverture et des spécificités, clauses du traité.

Le wording : contrat final signé entre les deux parties.

L'historique des primes et sinistres et dans certains cas sa mesure d'exposition.

Etape 3 : Cotation du traité par le tarificateur.

Le tarificateur étudie et analyse les données.

Il détermine le taux technique du traité.

Etape 4 : Résultat envoyé au souscripteur.

Il peut choisir d'accepter ou de refuser de signer ce traité.

Il décide des placements qu'il veut faire sur différents traités dans le marché.

Il ne peut pas dépasser sa capacité annuelle de souscription fixé par la direction.

2.2 Mise en "as-if" des données

Étant donné que la tarification par expérience repose sur l'analyse de l'historique des sinistres et des primes de la cédante, il est crucial que chaque élément soit représentatif du portefeuille à protéger. Toutefois, les données fournies par l'assureur couvrent généralement plusieurs années d'historique. Ainsi, un sinistre survenu il y a 6 ans ne peut être directement comparé à un sinistre survenu l'année dernière. Il est essentiel de prendre en compte certains facteurs qui influent sur ces montants. Cette étape est appelée la mise en "as-if" ². Elle se décompose en deux parties :

Revalorisation des données :

Il est essentiel de revaloriser les données pour prendre en compte toute inflation potentielle. Habituellement, on utilise l'indice des prix à la consommation (abrégi par l'acronyme "IPC") pour estimer l'évolution monétaire du pays concerné. D'autres indices comme l'indice de l'évolution des salaires, celui du coût de la construction ou encore des coûts d'hospitalisation peuvent également être utilisés, selon la branche couverte et la disponibilité des indices dans le pays concerné.

L'actualisation des sinistres s'obtient par la formule suivante :

$$S_k^n = S_k \times \frac{I_n}{I_k}$$

Avec :

- n : l'année de couverture.
- I_k : la valeur de l'indice pour l'année k .
- S_k : le montant d'un sinistre pour l'année k .
- S_k^n : le montant du même sinistre mis en "as-if" pour l'année n .

2. "Comme si" en français.

Redressement des données :

Cette étape est essentielle pour assurer que les données utilisées soient cohérentes et pertinentes pour l'année de couverture. Elle prend en compte deux aspects principaux :

D'une part, elle considère les éventuels changements dans la politique de souscription du traité au fil des années. En effet, les garanties et les limites peuvent évoluer avec le temps, et certains risques couverts peuvent même changer ou disparaître. Il est donc crucial de rendre le portefeuille homogène pour l'année de couverture en tenant compte de ces évolutions.

D'autre part, elle intègre les évolutions juridiques et les changements dans la jurisprudence, qui peuvent avoir un impact significatif sur le montant des sinistres. En tenant compte de ces facteurs, on s'assure que les données reflètent correctement le contexte juridique actuel.

Les données mise en "as-if" sont systématiquement prises en compte dans l'analyse d'une tarification par expérience.

2.3 L'approche classique : Le burning cost

2.3.1 La méthode

La méthode du burning cost, abrégée par l'acronyme "BC", est l'approche la plus simple à appréhender. Elle se définit comme le rapport entre les sinistres à la charge du réassureur et l'assiette de primes perçue par l'assureur chaque année.

Pour appliquer cette méthode, la charge du réassureur est calculée pour chaque année du traité en utilisant les données "as-if". Cette étape permet par la suite de calculer le BC annuel selon la formule suivante :

$$BC_i = \frac{\sum_{j=1}^{n_i} \min(\max[X_j - F, 0], P)}{EPI_i}$$

avec :

- n_i le nombre de sinistres de l'année i .
- EPI_i l'assiette de prime de l'année i .
- X_j le montant mis en as-if du $j^{\text{ème}}$ sinistre.

Cette formule est présentée dans le cas classique (avec reconstitutions illimitées et gratuites, sans AAD ni AAL).

Le taux pur peut ensuite être calculé en prenant la moyenne pondérée de tous les BC_i des années prises en compte dans l'étude. Cette moyenne permet de refléter de manière précise les variations des sinistres au fil des années. Enfin, la prime pure de réassurance est obtenue en multipliant ce taux pur par l'assiette de prime estimée pour l'année de couverture.



FIGURE 2.2 – Les différentes étapes de la méthode burning cost

Cette méthode est efficace lorsque les tranches de l'XS à tarifier sont travaillantes. La définition d'une tranche travaillante est subjective et dépend de chaque compagnie de réassurance, mais elle peut être résumée comme suit : une tranche est considérée comme travaillante lorsqu'elle régulièrement sinistrée.

2.3.2 Les limites

En raison de sa simplicité apparente, cette méthode présente certaines limites. Tout d'abord, elle exige une grande quantité de données dans l'historique de la cédante pour être efficace. En effet, le taux pur calculé ne sera pas robuste si les données sont limitées. Par conséquent, pour les nouveaux contrats, par exemple, l'absence d'historique rend l'utilisation de cette méthode impossible.

Enfin, cette méthode n'est pas fiable pour les tranches non travaillantes du traité. Dans de tels cas, la valeur du BC pour ces tranches est nulle, ce qui entraîne un taux pur nul également. Cette situation n'a pas de sens, car la prime pure de réassurance ne peut être gratuite, le risque est bien réel.

2.4 L'approche probabiliste : Le modèle collectif

2.4.1 Le modèle

La seconde méthode exploitant l'expérience de la cédante est le modèle collectif. Conçu pour pallier le défi de l'estimation des primes de réassurance dans les tranches non-travaillantes, ce modèle permet une analyse distincte de la fréquence et de la sévérité des sinistres de l'assureur.

Sa structure se décompose ainsi :

- N : la fréquence. Cette variable aléatoire discrète, prenant ses valeurs dans $\mathbb{N}$, décrit le nombre annuel de sinistres dans le portefeuille.
- $(X_i)_{i \geq 0}$: la sévérité. Cette suite de variables aléatoires continues représente les coûts individuels des sinistres par ordre de survenance.

La charge totale des sinistres S est obtenue par la formule suivante :

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

Hypothèses du modèle :

- Les variables X_i et N sont indépendantes.
- Les X_i sont indépendants et identiquement distribués.
- Les $X_0 = 0$.

Il en résulte une prime pure reliant la fréquence et la sévérité des sinistres de l'assureur :

$$\mathbb{E}[S] = \mathbb{E}[N] \times \mathbb{E}[X]$$

2.4.2 Choix du seuil de modélisation

Avant d'examiner les différentes distributions possibles pour modéliser la fréquence et la sévérité des sinistres, il est nécessaire de se pencher sur le choix du seuil de modélisation S_m . Ce seuil détermine le montant à partir duquel l'analyse et la modélisation sont effectuées.

Ce choix est essentiel pour filtrer les données pertinentes dans le calcul de la prime et influe sur la modélisation de la fréquence des sinistres.

En général, les données transmises au réassureur ne représentent pas la totalité des informations de la cédante, un tri préalable a déjà été effectué pour éviter de surcharger le réassureur d'informations non pertinentes. Ce seuil est connu sous le nom de seuil de déclaration de sinistres S_d . Il est essentiel que ce seuil, aussi bien avant qu'après la mise en "as-if", ne dépasse pas la franchise du traité, afin de s'assurer qu'aucun sinistre susceptible d'impacter le traité ne soit exclu. Si cette condition est respectée et que ce seuil convient au tarificateur, il sera utilisé pour la modélisation. Autrement, un seuil de modélisation plus élevé sera choisi.

Il n'existe pas de règles strictes pour choisir ce seuil, mais il doit respecter la contrainte suivante :

$$S_d \leq S_m \leq F$$

En effet, S_m ne peut pas être supérieur à la franchise F , sinon la modélisation ne couvrira pas correctement toute la tranche.

Le choix du seuil doit également équilibrer entre ne pas opter pour un seuil trop bas, afin de ne pas concentrer la modélisation sur des sinistres de faibles montants par rapport à la tranche, et ne pas choisir un seuil trop élevé, pour ne pas perdre des informations importantes.

2.4.3 Modélisation de la fréquence

Après avoir défini S_m , la modélisation de la fréquence des sinistres peut être abordée. Chez CCR Re, cette fréquence est représentée par une loi de Poisson de paramètre λ , bien que certaines compagnies de réassurance optent également pour la loi binomiale négative.

La loi de Poisson

La loi de Poisson est particulièrement bien adaptée en réassurance car elle permet de caractériser la survenance d'événements rares au sein d'un portefeuille composé de multiples polices.

Définition : Soit N une variable aléatoire discrète suivant une loi de Poisson de paramètre λ que l'on note $N \sim \text{Pois}(\lambda)$. Si le nombre moyen d'occurrences dans un intervalle de temps fixé est de λ , alors la probabilité qu'il survienne exactement k sinistres est :

$$\mathbb{P}[N = k] = \frac{\lambda^k e^{-\lambda}}{k!}$$

Son espérance vaut :

$$\mathbb{E}[N] = \lambda$$

Pour estimer le paramètre de cette loi, il suffit d'utiliser la méthode des moments ou le maximum de vraisemblance, ajustés selon les données d'expérience. Les primes et les sinistres de la cédante sont prétraitées en utilisant un coefficient d'échelle déterminé selon l'exposition du portefeuille (EPI de l'année, nombre de risques, etc.).

La modélisation de la fréquence n'est pas approfondie dans ce mémoire. Par conséquent, l'étude de la fréquence ne sera pas davantage développée.

2.4.4 Modélisation de la sévérité

Cette partie se concentre sur la modélisation de la sévérité des sinistres. Étant donné que le seuil de modélisation est établi au préalable, seuls les montants des sinistres dépassant ce seuil sont pertinents pour la tarification. L'objectif est d'ajuster au mieux une loi continue à l'échantillon de données mis en as-if $(x_i)_{i \geq 0}$.

Plusieurs lois usuelles peuvent être utilisées pour modéliser la sévérité en réassurance :

La loi de Pareto

La loi de Pareto est souvent utilisée en réassurance en raison de sa capacité à modéliser les grands sinistres et de sa facilité de manipulation.

Définition : Une variable aléatoire X suit une loi de Pareto de paramètres α et A , notée $X \sim \mathcal{P}(\alpha, A)$, si elle possède une fonction de densité de la forme suivante :

$$f(x) = \alpha \frac{A^\alpha}{x^{\alpha+1}} \mathbb{1}_{x \geq A}$$

La fonction de répartition pour $x \geq A$ est représentée par la formule suivante :

$$F(x) = 1 - \left(\frac{A}{x}\right)^\alpha$$

Son espérance pour $\alpha > 1$ est donnée par :

$$\mathbb{E}[X] = \frac{A\alpha}{\alpha - 1}$$

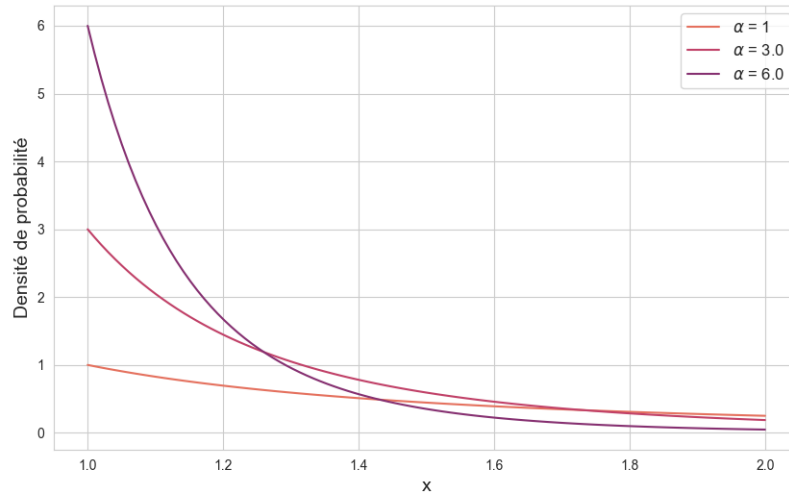


FIGURE 2.3 – Densité de la loi de Pareto selon α

Remarques :

- Lorsque $\alpha \leq 1$, l'espérance est infinie. Par conséquent, cette loi ne peut être utilisée dans ce cas pour la tarification, car cela implique que le risque est non-assurable.
- Plus la valeur de α se rapproche de 1, plus la distribution aura une queue épaisse. La figure 2.3 permet de visualiser que plus α est proche de 1 plus la probabilité qu'un grand sinistre se réalise sera élevée.

La loi Exponentielle

Définition : Une variable aléatoire $X \sim \text{Exp}(\lambda)$ avec $\lambda > 0$, si elle possède une fonction de densité de la forme suivante :

$$\forall x > 0, f(x) = \lambda \exp(-\lambda x) \mathbb{1}_{x \geq 0}$$

La fonction de répartition est représentée par la formule suivante, avec $x \geq 0$:

$$F(x) = 1 - \exp(-\lambda x)$$

Son espérance est donnée par :

$$\mathbb{E}[X] = \frac{1}{\lambda}$$

Pour estimer les paramètres de ces lois, il suffit d'utiliser la méthode des moments ou le maximum de vraisemblance, ajustés selon les données d'expérience.

Pour déterminer la loi qui s'ajuste le mieux aux données, divers tests et critères statistiques sont utilisés. Parmi les tests figure le test de Kolmogorov-Smirnov. En complément, les modèles sont évalués à l'aide de critères d'information tels que le critère d'information d'Akaike ("AIC") et le critère d'information bayésien ("BIC"). Ces procédures sont détaillées en annexe (cf. Annexe B [pages 104](#)).

Une fois que la loi de fréquence et la loi de sévérité sont établies, une simulation

stochastique de ces deux composantes sera réalisée afin d'estimer le taux pur désiré pour chaque tranche.

Toutes les étapes de l'approche du modèle fréquence-sévérité peuvent être représentées comme suit :

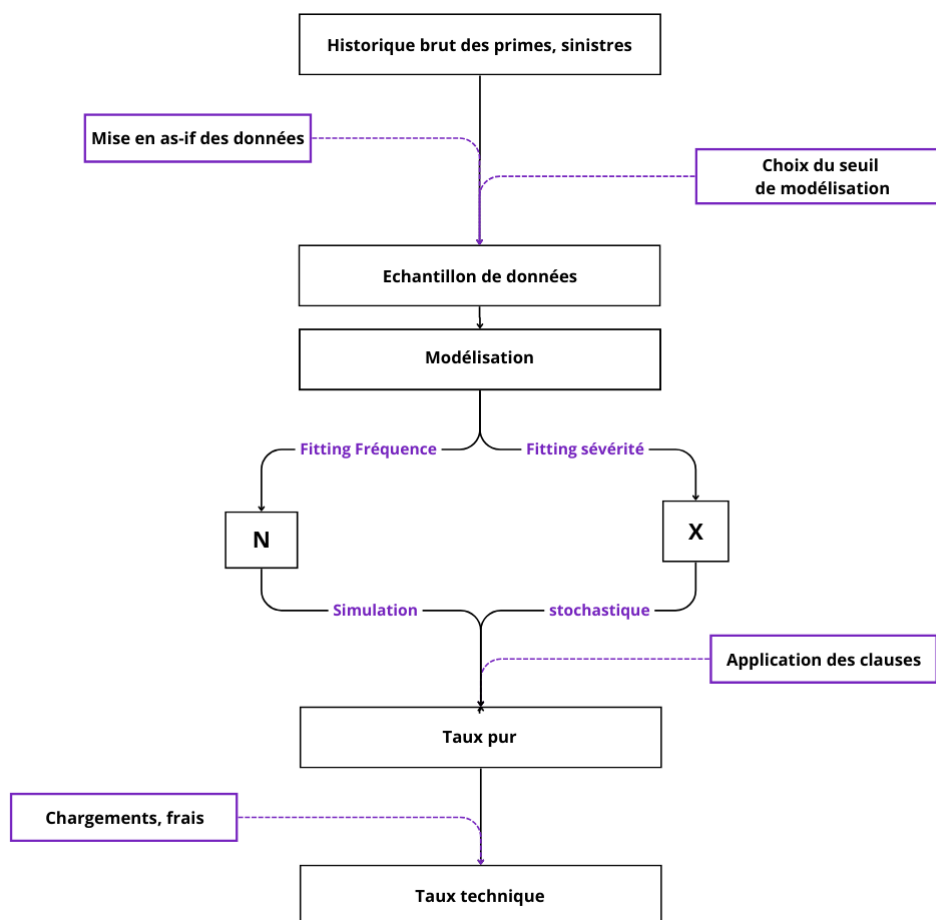


FIGURE 2.4 – Récapitulatif de l'approche probabiliste

2.4.5 Les limites

Cette méthode comporte également des limites. Tout d'abord, l'hypothèse d'indépendance entre la fréquence et la sévérité des sinistres n'est pas toujours vérifiée, ce qui peut fausser le modèle et conduire à des estimations biaisées. Ensuite, elle nécessite un nombre conséquent d'observations pour estimer précisément les paramètres des lois, car la précision statistique repose également sur la taille de l'échantillon. Lorsqu'il y a peu de données disponibles, les estimations deviennent moins fiables, ce qui accroît l'incertitude dans le choix des distributions ainsi que dans l'estimation des paramètres, rendant les prédictions finales plus fragiles.

Chapitre 3

Application du modèle collectif

Ce chapitre est consacré à la tarification d'un traité en utilisant le modèle collectif présenté précédemment.

3.1 Présentation du traité

CCR Re intervient pour couvrir les risques associés à la branche dommages aux biens ("DAB") pour une cédante française. Le traité se présente sous la forme d'un contrat non proportionnel de type XS et est effectif pour une durée d'un an à partir du 1^{er} janvier 2024. Les risques couverts par ce traité incluent :

- les risques locatifs,
- la responsabilité civile du propriétaire,
- le recours des voisins et des tiers,
- les risques de conflagration,
- les risques technologiques.

La branche DAB Incendie couvre les coûts liés à la réparation ou au remplacement de biens endommagés à la suite d'un incendie. En réassurance, cela inclut, par exemple, la destruction d'un bâtiment assuré à la suite d'un sinistre incendie. Cependant, ce contrat ne couvre pas les sinistres causés par des catastrophes naturelles.

Ce programme de réassurance est structuré en quatre tranches d'XS, exprimées en millions d'euros, comme illustré ci-dessous :

T4	8 m XS 17 m Reconstitution : 1 - 100%
T3	7 m XS 10 m Reconstitution : 1 - 100%
T2	8 m XS 2 m Reconstitution : 1 - 100%
T1	1,2 m XS 0,8 m Reconstitutions : 1 - gratuite 1 - 100%

FIGURE 3.1 – Traité de réassurance

Il n'y a pas de conditions spécifiques supplémentaires à considérer dans le cadre de la tarification. Ce traité sera l'objet de l'analyse qui suit dans ce mémoire.

3.2 Mise en "as-if"

Avant de procéder à l'application du modèle collectif, il est primordial de mettre en "as-if" les données transmises par la cédante (cf. 2.2).

3.2.1 Primes

Pour la mise en "as-if" des primes, la cédante a fourni à la fois le montant total des primes par année et le nombre de contrats assurés correspondant. Ainsi, la prime moyenne d'un contrat peut être simplement calculée. La formule de la mise en "as-if" des primes est donc la suivante :

$$P_i^n = \frac{P_n}{C_n} * C_i$$

Avec :

- P_i^n : L'assiette de primes de l'année i en as-if pour l'année n.
- P_i : L'assiette de primes de l'année i.
- C_i : Le nombre de contrat assurés l'année i.

La prime "as-if" d'une année i correspond donc à la valeur de la prime moyenne d'un contrat cette année multiplié par le nombre de contrats signés l'année i.

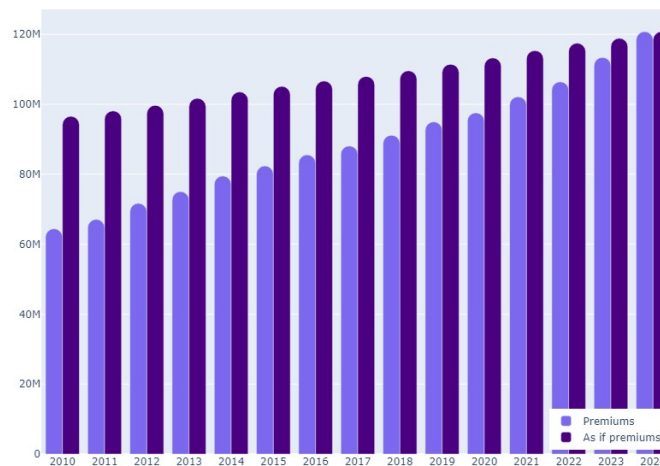


FIGURE 3.2 – Evolution des primes

Le graphique ci-dessus, qui présente l'évolution des primes mise en as-if de la cédante, montre une légère croissance de son portefeuille depuis 2010, passant d'environ 95 millions à près de 120 millions en 2023. Pour l'année 2024, l'assiette de primes est estimée à 120 millions.

3.2.2 Sinistres

Pour les sinistres, la mise en "as-if" est effectuée à l'aide de l'indice du bâtiment BT01, qui mesure l'évolution des prix des travaux de construction en France. Cet indice est particulièrement pertinent pour les sinistres DAB incendie, car les coûts de reconstruction ou de réparation sont directement influencés par les fluctuations des prix dans ce secteur.

L'histogramme ci-dessous représente le montant maximal du sinistre revalorisé pour chaque année de l'historique. Il indique également le montant médian revalorisé des sinistres par année.

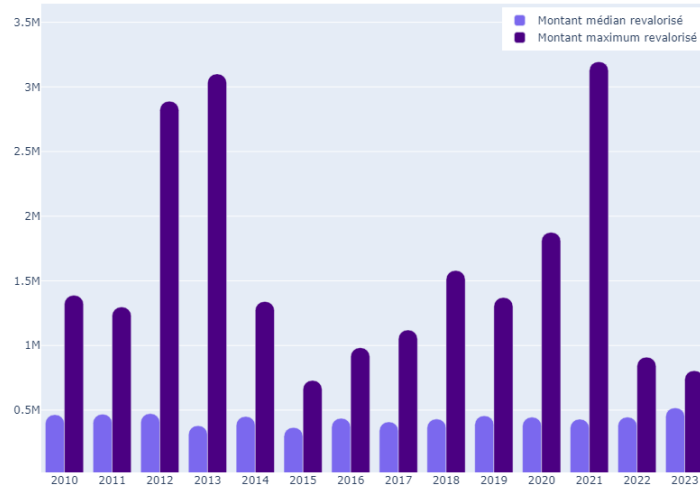


FIGURE 3.3 – Evolution des sinistres

L'histogramme retraçant l'évolution des sinistres permet d'analyser la sinistralité historique de la cédante et d'identifier les tranches de réassurance touchées. Globalement, le montant médian des sinistres par année se maintient autour de 500 000 € entre 2010 et 2023, un niveau inférieur au seuil de 800 000 €, qui marque le début du programme de réassurance. Cependant, certaines années se distinguent par des sinistres plus importants. En particulier, les années 2012, 2013 et 2021 ont connu des sinistres particulièrement élevés, proches de 3 millions d'euros.

La première tranche de réassurance (T1), avec une franchise de 800 000 €, a été déclenchée la plupart des années, tandis que la deuxième tranche (T2), dont la franchise est fixée à 2 millions d'euros, n'a été impactée que lors des sinistres des années 2012, 2013 et 2021. Les tranches supérieures, quant à elles, n'ont jamais été affectées par la sinistralité observée dans l'historique de la cédante.

3.2.3 Seuil de la statistique

Dans ce traité, le seuil de déclaration, noté S_d , fourni par la cédante est fixé à 250 000 €. Cela signifie que tous les sinistres historiques transmis par la cédante dépassent ce montant. Cependant, afin d'adapter ces sinistres à la situation actuelle, une mise en as-if est réalisée, ce qui modifie les montants des sinistres et, par conséquent, le seuil de déclaration. La mise en as-if, effectuée à l'aide de l'indice BT01, donne un nouveau seuil de 332 410,57 €.

Cette mise à jour permet de vérifier que tous les sinistres impactant le traité ont bien été transmis, car le seuil de déclaration mis en as-if reste inférieur à la franchise de la tranche T1. Ainsi, tous les sinistres supérieurs à la franchise de la T1 sont bien inclus. Ce calcul fournit également une estimation de la troncature à gauche des données, permettant d'évaluer la précision des informations transmises.

Dans le cadre de la tarification, le seuil de modélisation sera donc situé entre S_d et la franchise.

3.3 Burning cost

Une fois les données de la cédante mises en "as-if", un BC est calculé pour chaque tranche du traité. L'étude du BC (présentée dans le chapitre 2.3) est primordiale pour analyser la sinistralité de la cédante.

3.3.1 Horizon de modélisation

Lors de l'étude d'un BC, il est important de choisir un horizon de modélisation correspondant aux années sélectionnées pour l'analyse de la sinistralité de la cédante. Ce choix se fait selon plusieurs critères :

Premièrement, le choix de l'horizon de modélisation dépend de la nature de la branche, qu'elle soit à développement court ou long (comme expliqué dans le chapitre n°1 1.5.2). Pour une branche à développement court, il est possible de considérer uniquement les cinq dernières années d'historique. En revanche, pour une branche à développement long, il est préférable de prendre en compte au moins dix ans d'historique.

La branche concernée par le traité présenté est classée comme une branche à développement court, étant donné que les dommages aux biens relèvent de sinistres matériels. Le montant total des réparations est donc généralement déterminé rapidement. Il serait donc possible de ne prendre en compte que les cinq dernières années d'historique. Toutefois, chez CCR Re, une période plus longue est privilégiée si nécessaire. Par conséquent, l'horizon de modélisation choisi pour ce traité couvre une période de dix ans, de 2013 à 2022.

Secondement, les données fournies par la cédante ont souvent une date d'arrêté antérieure au 31 décembre de l'année en cours. C'est pourquoi la dernière année de statistiques (2023 dans ce cas) n'est généralement pas incluse. En effet, la sinistralité annuelle de la cédante peut encore évoluer, rendant les données de cette année partiellement fiables. Cependant, si l'année en question est marquée par une sinistralité significativement plus élevée par rapport aux autres, elle doit être prise en considération, malgré son incomplétude. Dans le cadre de ce traité, ce n'est pas le cas, donc l'année 2023 n'est pas incluse dans l'horizon de modélisation.

3.3.2 Burning cost détaillé de la T1

Le tableau ci-dessous présente un BC détaillé pour la T1. Les colonnes indiquent le nombre et le montant des sinistres à charge pour le réassureur chaque année. La colonne du BC par année est également incluse. Ces informations sont essentielles pour identifier les années les plus sinistrées et observer la tendance du portefeuille. La ligne moyenne du tableau calcule la moyenne pondérée par les primes selon les colonnes concernées. Cette moyenne est calculée sur les années sélectionnées dans l'horizon de modélisation.

Années	Primes revalorisées	Nombre sinistres à charge	Montant à charge annuel	BC
2010	96 552 008	3	870 357	0,90%
2011	98 108 420	4	795 223	0,81%
2012	99 689 920	4	2 971 513	2,98%
2013	101 696 932	1	1 200 000	1,18%
2014	103 539 662	4	1 340 924	1,30%
2015	105 130 916	0	-	0,00%
2016	106 649 380	3	329 354	0,31%
2017	107 960 345	8	1 347 979	1,25%
2018	109 566 510	4	843 783	0,77%
2019	111 436 537	3	828 101	0,74%
2020	113 256 268	2	1 203 329	1,06%
2021	115 355 277	4	2 242 770	1,94%
2022	117 470 209	1	109 219	0,09%
2023	118 858 765	1	4 670	0,00%
Moyenne	109 206 204	3,32	1 044 605	0,86%
			BC net de recs	0,74%

FIGURE 3.4 – Burning cost T1

L'analyse de ce tableau permet de mettre en évidence la consommation de la tranche T1 et de confirmer qu'il s'agit bien d'une tranche travaillante. Comme expliqué précédemment, une tranche travaillante est caractérisée par une fréquence élevée de sinistres. Cela se manifeste ici par une moyenne pondérée de 3,32 sinistres à charge du réassureur, ce qui est relativement élevé. De plus, la moyenne pondérée du montant annuel à charge du réassureur s'élève à 1 044 604,83 €, ce qui est proche de la portée maximale de la tranche. Ces éléments confirment que la tranche T1 est effectivement une tranche travaillante.

Le BC net de reconstitution pour cette tranche s'élève donc à 0,74 %. Étant donné que cette tranche est travaillante, cette valeur peut être utilisée comme taux pur.

3.3.3 Résultats burning cost

Le tableau ci-dessous présente le BC pour toutes les tranches du programme. Les valeurs affichées sont les moyennes pondérées calculées sur l'horizon de modélisation.

Tranches	Nombre moyen de sinistres par an	Montant à charge de la réassurance	BC net de recs
T1	3,32	1 044 604,83	0,742%
T2	0,22	253 776,96	0,209%
T3	-	-	0,000%
T4	-	-	0,000%

FIGURE 3.5 – Récapitulatif BC

Les chiffres ci-dessus permettent de conclure que seule la tranche T1 est une tranche travaillante. La tranche T2 n'est pas nulle mais n'est pas suffisamment sinistrée pour être considérée comme travaillante. Les tranches T3 et T4 ne sont jamais touchées par des sinistres, ce qui illustre clairement la notion de tranche non travaillante.

Les BC nets de reconstitution des tranches T3 et T4 ne peuvent pas être utilisés comme taux pur, car une tranche ne peut être gratuite : un risque existe toujours. Ainsi, l'application du modèle collectif est nécessaire.

3.4 Modélisation

3.4.1 Choix du seuil de modélisation

Avant de modéliser la sévérité et la fréquence des sinistres, certains paramètres doivent être pris en compte.

Premièrement, il est essentiel de déterminer l'horizon de modélisation, comme expliqué dans la partie précédente. Pour notre étude, cet horizon est fixé de 2013 à 2022.

Deuxièmement, il faut définir le seuil à partir duquel la modélisation sera effectuée. Ce seuil doit répondre à la condition suivante, présentée dans le chapitre précédent (2.4.2) :

$$S_d \leq S_m \leq F$$

Le choix de ce seuil doit prendre en considération le nombre de sinistres modélisés, car il est crucial d'avoir un nombre minimum de sinistres pour pouvoir appliquer des lois statistiques de manière fiable. En général, le seuil de modélisation est fixé entre la moitié de la franchise et la franchise complète.

Dans le cadre de cette tarification, un seuil de 600 000 € a été retenu pour une franchise fixée à 800 000 €, ce qui permet d'inclure 29 sinistres dans l'analyse.

3.4.2 Fréquence

Pour la modélisation de la fréquence des sinistres, les données peuvent être ajustées uniquement à une loi de Poisson chez CCR Re. Ainsi, la loi de fréquence obtenue est la suivante : $N \sim \text{Poi}(6, 54)$.

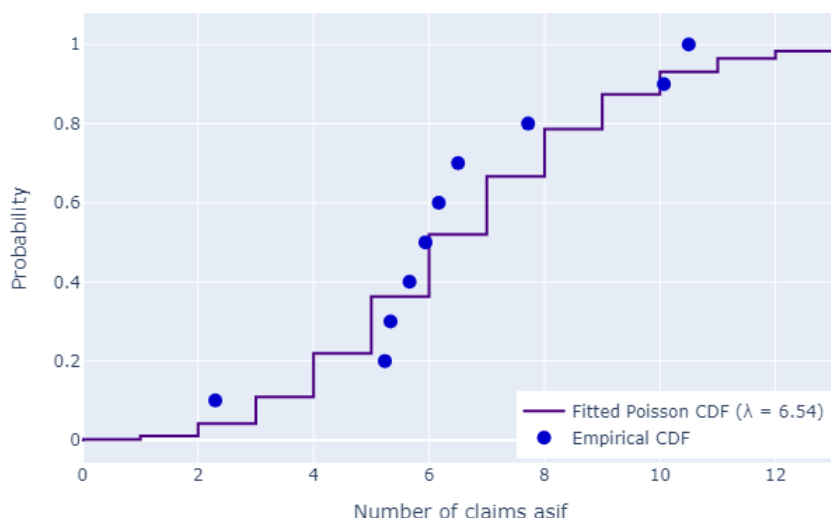


FIGURE 3.6 – Fonction de répartition empirique et de la loi de Poisson

Le graphique ci-dessus compare la fonction de répartition du nombre de sinistres annuels (ajustée en fonction de l'évolution de l'exposition de la cédante) à celle de la loi de Poisson ajustée. Le nombre de sinistres annuels suit de près la fonction de répartition théorique, suggérant que l'ajustement est satisfaisant pour poursuivre avec cette loi. Étant donné que ce mémoire se concentre sur la modélisation de la sévérité, les autres méthodes pour affiner la modélisation de la fréquence ne seront pas explorées ici. Néanmoins, la compagnie reconnaît les limites actuelles et travaille sur des projets visant à améliorer la modélisation de la fréquence.

3.4.3 Sévérité

Pour modéliser la sévérité des sinistres, les données peuvent être ajustées à une loi de Pareto ou à une loi exponentielle. L'ajustement se fait à l'aide de la méthode du maximum de vraisemblance ("MLE") et de la méthode des moments ("MM") afin d'estimer les paramètres de ces distributions pour plus de précision.

De plus, pour déterminer la loi la plus adéquate aux données, le test de Kolmogorov-Smirnov ("KS") est appliqué, suivi du calcul du critère AIC. Le test KS permet d'évaluer l'adéquation entre les distributions ajustées et les données empiriques. Si la p-valeur obtenue est supérieure à 0.05, l'hypothèse selon laquelle il existe une différence significative entre la distribution ajustée et les données empiriques peut être rejetée. Le critère AIC, quant à lui, est un indicateur à minimiser : plus son score est faible, plus le modèle est jugé pertinent pour les données observées.

Les résultats obtenus sont présentés dans le tableau suivant :

Lois	Méthodes	Paramètre	Test KS	AIC
Pareto	MLE	2.72	0.82	1615.48
	MM	2.77	0.79	1615.50
Exponentielle	MLE	2.96 e^{-6}	0.21	1622.19
	MM	2.96 e^{-6}	0.21	1622.19

FIGURE 3.7 – Tableau ajustement sévérité

Les deux lois présentent une p-valeur supérieure à 0.05, ce qui conduit à rejeter l’hypothèse d’une différence significative entre chaque distribution et les données empiriques. Ainsi, le test KS est validé pour les deux distributions. Toutefois, la loi de Pareto affiche un critère AIC plus faible. Par conséquent, elle est choisie et modélisée via la méthode du maximum de vraisemblance.

Par conséquent, la loi de sévérité retenue est la suivante : $X \sim \text{Pareto}(2.72)$

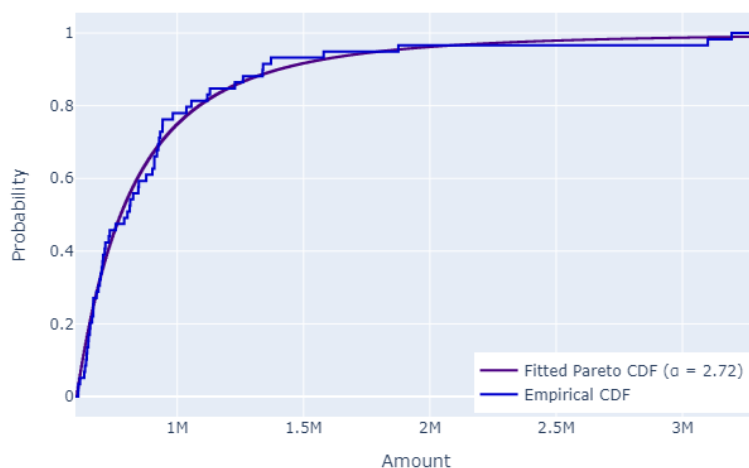


FIGURE 3.8 – Fonction de répartition empirique et de la loi de Pareto

L’ajustement de la loi de Pareto sur les points empiriques semble adéquat, comme le montre le graphique ci-dessus.

3.5 Calcul de la prime pure

Pour calculer les taux et les primes pures de chaque tranche, la formule présentée dans le chapitre n°2 (2.4.1) peut être directement appliquée. Toutefois, il est courant de réaliser des simulations afin d’obtenir une perspective plus complète sur la sinistralité future de la cédante pour ces tranches.

3.5.1 Simulation

Afin de réaliser les simulations, il est d’abord nécessaire de déterminer le nombre de simulations à réaliser. Ce choix dépend du niveau de confiance dans la modélisation des lois basées sur les données de la cédante. Plus la modélisation est incertaine, plus il est important d’observer la volatilité de la sinistralité de la cédante. La robustesse de la modélisation repose principalement sur l’étendue de l’historique disponible, en particulier sur la fréquence des sinistres. Pour ce traité, le nombre de simulations retenu est de 250 000.

Les simulations sont générées en utilisant les lois ajustées précédemment. La loi de Poisson ajustée permet d'obtenir la fréquence des sinistres, tandis que la loi de Pareto ajustée détermine le montant de chacun de ces sinistres. Ainsi, une matrice nommée YLT (*Years Loss Table* en anglais¹) est créée, représentant au mieux la sinistralité future de la cédante. Elle est définie comme suit :

$$\begin{pmatrix} \text{Années} & \text{Pertes} \\ \dots & \dots \\ \dots & \dots \\ \dots & \dots \\ \dots & \dots \end{pmatrix}$$

Après la création des YLT, il suffit de calculer la charge annuelle du réassureur pour chaque tranche sur toutes années simulées, puis de déterminer la charge moyenne du réassureur.

3.5.2 Résultats

Le tableau ci-dessous présente les taux purs ainsi que les primes pures obtenus pour chaque tranche avec cette méthode de tarification.

Tranche	Taux pur	Prime pure
T1	0,748%	903 146
T2	0,220%	265 963
T3	0,009%	11 599
T4	0,003%	3 554
Total	0,980%	1 184 262

FIGURE 3.9 – Résultats de la tarification du traité

La modélisation révèle un taux pur de 0,98% pour l'ensemble du traité, correspondant à une prime pure de 1 184 262 €. Ce taux est principalement influencé par les résultats des tranches T1 et T2, tandis que ceux des tranches T3 et T4 restent très faibles en raison de leur caractère extrême.

Les taux purs obtenus pour les tranches T1 et T2 sont considérés comme fiables, car ils s'appuient sur un historique de sinistres significatif ayant impacté ces tranches. La modélisation est donc complète et solide, étant fondée sur des données concrètes et représentatives de la sinistralité passée de la cédante.

De plus, les taux calculés sont quasiment identiques à ceux obtenus par le BC, ce qui est à la fois cohérent et rassurant. Cette correspondance valide la méthode utilisée, en démontrant que les taux obtenus sont en adéquation avec l'expérience historique. Ainsi, les résultats pour les tranches T1 et T2 sont jugés fiables et pertinents, reflétant fidèlement la sinistralité de la cédante.

En revanche, un problème se pose pour les tranches supérieures T3 et T4, qui n'ont jamais été impactées par des sinistres dans le passé. Les taux calculés pour ces tranches

1. "Tableau de pertes annuelles" en français.

manquent de fiabilité. En effet, la modélisation traditionnelle de la sévérité des sinistres, basée sur une seule loi statistique, se concentre essentiellement sur la partie centrale de la sinistralité de la cédante. Or, l'historique des sinistres de cette dernière ne dépasse pas 3 millions d'euros (le sinistre maximum, revalorisé, survenu en 2021). Étant donné que la tranche T3 ne se déclenche qu'à partir de sinistres supérieurs à 10 millions d'euros, et la T4 à partir de 17 millions d'euros, la modélisation traditionnelle ne prend pas en compte ces sinistres potentiels, nettement au-dessus de la sinistralité observée.

Par conséquent, les estimations des taux pour ces tranches reposent sur des sinistres bien en dessous des seuils de franchise, ce qui réduit considérablement la fiabilité des résultats obtenus.

L'objectif est donc d'améliorer la précision de la tarification en renforçant la modélisation de la queue de distribution de la sévérité des sinistres, afin d'obtenir des taux plus fiables et précis pour ces tranches élevées.

3.6 Limite et enjeu

L'application du modèle collectif tel qu'il est utilisé ici présente une limite centrale qui impacte directement la fiabilité des résultats obtenus.

La modélisation de la sévérité des sinistres à travers une unique loi statistique constitue une hypothèse simplificatrice qui, bien qu'efficace pour la sinistralité centrale de la cédante, limite la prise en compte des sinistres extrêmes. En effet, cette approche se concentre principalement sur la partie centrale de la distribution de la sinistralité, ce qui entraîne une modélisation imprécise des sinistres situés dans la queue de la distribution, où se trouvent les événements les plus rares et de grande ampleur. Dans ce contexte, les taux obtenus pour les tranches élevées sont donc moins fiables et peuvent ne pas refléter pleinement le risque potentiel, compromettant ainsi la justesse de l'évaluation.

L'enjeu principal de cette modélisation est de garantir des taux qui soient non seulement représentatifs du risque réel, mais également compétitifs. Pour le réassureur, il est essentiel de disposer d'estimations précises afin de mesurer adéquatement le risque encouru. En effet, des taux trop élevés pourraient rendre le traité moins attractif, tandis que des taux sous-évalués exposent le réassureur à des pertes significatives en cas de sinistres extrêmes. La capacité à modéliser précisément les risques rares devient donc essentielle pour assurer la viabilité et la compétitivité du réassureur.

Pour répondre à cette limite, l'introduction de la théorie des valeurs extrêmes dans la modélisation de la sévérité des sinistres s'avère pertinente. Cette approche permet d'aller au-delà de la simple modélisation de la sinistralité centrale en se concentrant également sur les risques présents dans la queue de la distribution. En améliorant la précision des estimations pour les sinistres extrêmes, elle vise à renforcer la fiabilité des taux dans les tranches élevées, permettant ainsi une meilleure quantification des risques rares et offrant une tarification de plus grande confiance.

Ainsi, ce mémoire a pour objectif d'améliorer la modélisation de la sévérité des sinistres afin d'obtenir des résultats plus précis et explicables, particulièrement pour les tranches élevées.

Deuxième partie

L'approche par la théorie des valeurs extrêmes

Chapitre 4

La théorie des valeurs extrêmes

Contrairement à l'approche "classique" qui se concentre sur la distribution centrale des sinistres, la théorie des valeurs extrêmes permet une analyse plus approfondie de la queue de distribution de ces sinistres. Ceci est crucial dans la tarification d'un traité de réassurance afin d'assurer une estimation fiable des tranches non travaillantes, comme expliqué dans le chapitre précédent.

La théorie des valeurs extrêmes se décline en deux approches distinctes :

- L'approche des maxima par bloc.
- L'approche du dépassement de seuil.

L'objectif ici est de présenter les principaux résultats de la théorie des valeurs extrêmes afin d'appliquer cette méthode à la modélisation de la sévérité des sinistres.

4.1 Convergence en loi du maximum d'un échantillon

4.1.1 Théorie classique

Considérons $X_1, \dots, X_n$ un échantillon de n variables aléatoires indépendantes et identiquement distribuées représentant les montants des sinistres de la cédante.

Dans le cadre de la théorie "classique", la loi faible des grands nombres conduit au résultat suivant :

$$\frac{S_n}{n} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \mathbb{E}[X]$$

Avec :

- X intégrable.
- $S_n = \sum_{i=1}^n X_i$.

De plus, si X est de carré intégrable, le théorème central limite donne le résultat suivant :

$$\frac{S_n - n\mathbb{E}[X]}{\sqrt{n\mathbb{V}[X]}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Ce résultat permet d'obtenir des informations sur le comportement de la partie centrale des sinistres de la cédante.

4.1.2 Etude de la loi du maximum

En réassurance, les sinistres de grande ampleur exercent une influence significative sur le total des sinistres. Par conséquent, l'actuaire doit se pencher sur la partie extrême de la distribution des montants de sinistres pour évaluer les tranches non travaillantes. Une approche intuitive consiste à examiner le comportement du montant maximal des sinistres, désigné par $X_{(n)} = \max(X_1, \dots, X_n)$.

La fonction de répartition F des X_i est définie comme suit :

$$\mathbb{P}[X_{(n)} \leq x] = F^n(x)$$

De plus, la loi limite de $X_{(n)}$ est la suivante :

$$F^n(x) \xrightarrow{n \rightarrow +\infty} \begin{cases} 0 & \text{si } x < x_F \\ 1 & \text{si } x \geq x_F \end{cases}$$

Avec $x_F = \sup\{x \in \mathbb{R}, F(x) < 1\}$ le point terminal de F .

Ainsi, le maximum converge en probabilité vers le point terminal de l'échantillon.

4.1.3 Loi des valeurs extrêmes généralisés

Afin d'obtenir des informations sur la convergence en loi, la définition d'un domaine d'attraction est introduite.

Définition : Domaine d'attraction Soit H une fonction de répartition non-dégénérée (c'est-à-dire qu'elle n'est pas associée à une variable constante presque sûrement). La fonction F appartient au domaine d'attraction de H , noté $F \in \mathcal{DA}(H)$ s'il existe deux suites $a_n > 0$ et b_n telles que en tout point de continuité x de H .

$$\lim_{x \rightarrow +\infty} F^n(a_n x + b_n) = H(x) \Leftrightarrow \frac{X_{(n)} - b_n}{a_n} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} Y$$

où la variable aléatoire Y admet pour fonction de répartition H .

Par conséquent, en prenant en compte $a_n > 0$ et b_n comme constantes de normalisation, la théorie des valeurs extrêmes se concentre sur la distribution limite H qui répond à la convergence en loi suivante :

$$\mathbb{P}\left(\frac{X_{(n)} - b_n}{a_n} \leq x\right) = F^n(a_n x + b_n) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} H(x)$$

Le théorème de Fischer-Tippet-Gnedenko permet d'obtenir la famille de distributions à laquelle appartient H .

Théorème de Fischer-Tippet-Gnedenko : Si $F \in \mathcal{DA}(G)$ où G est une fonction de répartition non-dégénérée, alors G est de même type que la fonction de répartition de la loi Generalized Extreme Value (GEV) notée H_γ définie pour tout x tel que $1 + \gamma x > 0$:

$$H_\gamma(x) = \begin{cases} \exp[-(1 + \gamma x)^{-\frac{1}{\gamma}}] & \text{si } \gamma \neq 0 \\ \exp(-e^{-x}) & \text{si } \gamma = 0 \end{cases}$$

La loi limite du maximum est ainsi déterminée par un unique paramètre de forme γ , appelé l'indice des valeurs extrêmes, qui modifie la forme de la queue de distribution.

Remarque : La loi GEV peut également être représentée comme suit $H_{\gamma,\mu,\sigma}(x) = H_{\gamma}(\frac{x-\mu}{\sigma})$ avec μ le paramètre de position et σ le paramètre de dispersion.

4.1.4 Les domaines d'attraction

Le paramètre de forme γ définit le domaine d'attraction de la loi :

- Pour $\gamma > 0$, le domaine d'attraction correspond à celui de Fréchet, caractérisé par des distributions à queue épaisse. Dans ce cas, sa fonction de survie $1 - F(x)$ décroît vers 0 à une vitesse polynomiale, ce qui signifie que la probabilité d'événements extrêmes reste significative.
- Pour $\gamma = 0$, il s'agit du domaine d'attraction de Gumbel, où les distributions présentent une queue fine. Ici, la fonction de survie converge vers 0 à une vitesse exponentielle, indiquant une décroissance rapide des probabilités d'événements extrêmes.
- Pour $\gamma < 0$, le domaine d'attraction est celui de Weibull, correspondant à des distributions à queue fine à droite, avec un point terminal fini. La fonction de survie $1 - F(x)$ tend vers 0 à une vitesse polynomiale lorsque x s'approche d'une valeur finie.

Fréchet	Gumbel	Weibull
Pareto	Normale	
Log-gamma	Exponentielle	Uniforme
Student	Log-normale	Beta
Cauchy	Gamma	
	Weibull	

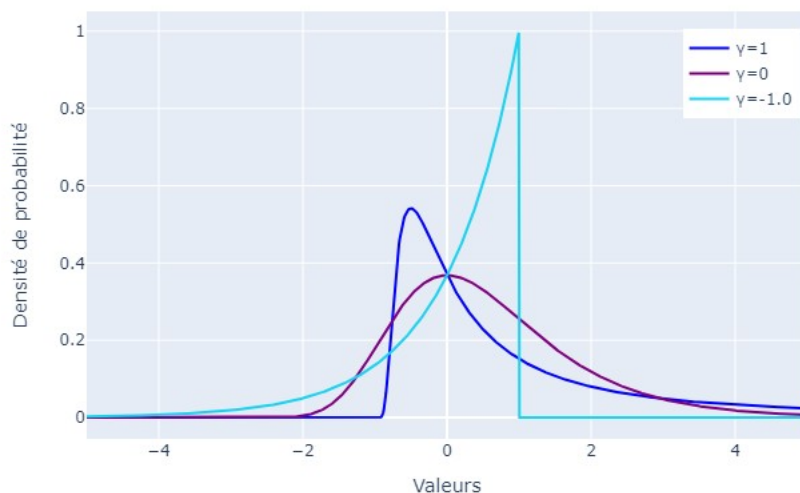


FIGURE 4.1 – Densité de la loi GEV selon γ

Remarque : En réassurance, les distributions appartenant au domaine de Fréchet suscitent un plus grand intérêt pour l'actuaire. Comme mentionné précédemment, celui-ci regroupe les distributions présentant une queue de distribution plus épaisse. Par conséquent, comme le montre la figure ci-dessus, la probabilité d'occurrence d'un sinistre extrême sera plus élevée dans ce scénario.

Cette étude sur la loi du maximum est appelée l'approche des maxima par bloc. Néanmoins, cette méthode ne sera pas approfondie davantage car elle n'est pas la plus adaptée au contexte de réassurance. En effet, les actuaires ont besoin de comprendre plus globalement les sinistres extrêmes de la cédante plutôt que le maximum possible. Par conséquent, la suite se concentrera sur l'approche du dépassement de seuil.

4.2 La théorie des excédents au-delà d'un seuil

L'approche du dépassement de seuil vise à examiner l'intégralité de la queue de distribution des sinistres plutôt que de se concentrer uniquement sur les sinistres maximum. L'accent est donc mis sur tous les sinistres qui dépassent un montant spécifique, noté u .

4.2.1 Etude de la loi des excès

Pour analyser l'évolution asymptotique de la distribution des excès des sinistres au-dessus du seuil u , la loi suivante est définie.

Définition : Loi des excès

Soit X une variable aléatoire de fonction de répartition F . Pour tout seuil $u < x_F$, la loi des excès de X au dessus du seuil u est :

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

4.2.2 Loi de pareto généralisé

En appliquant la fonction de répartition $H_{\gamma,\mu,\sigma}$ de la GEV à la loi des excès, la distribution obtenue est une Pareto généralisée (GPD).

Loi de Pareto généralisée : X suit une distribution Pareto généralisée (GPD) de paramètres $\gamma \in \mathbb{R}$ et $\sigma > 0$, noté $X \sim \mathcal{GPD}(\sigma, \gamma)$, si elle admet une fonction de répartition de la forme :

$$G_{\gamma,\sigma}(x) = \begin{cases} 1 - (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}} & \text{si } \gamma \neq 0 \\ 1 - e^{-\frac{x}{\sigma}} & \text{si } \gamma = 0 \end{cases}$$

Avec $1 + \frac{\gamma x}{\sigma} > 0$.

Sa fonction de densité est représentée par la formule suivante :

$$g_{\gamma,\sigma}(x) = \begin{cases} \frac{1}{\sigma} (1 + \frac{\gamma x}{\sigma})^{-(\frac{\gamma+1}{\gamma})} & \text{si } \gamma \neq 0 \\ \frac{1}{\sigma} e^{-\frac{x}{\sigma}} & \text{si } \gamma = 0 \end{cases}$$

Son espérance et sa variance sont les suivantes :

$$\begin{aligned} \forall \gamma < 1, \mathbb{E}[X] &= \frac{\sigma}{1 - \gamma} \\ \forall \gamma < \frac{1}{2}, \mathbb{V}[X] &= \frac{\sigma^2}{(1 - \gamma)^2(1 - 2\gamma)} \end{aligned}$$

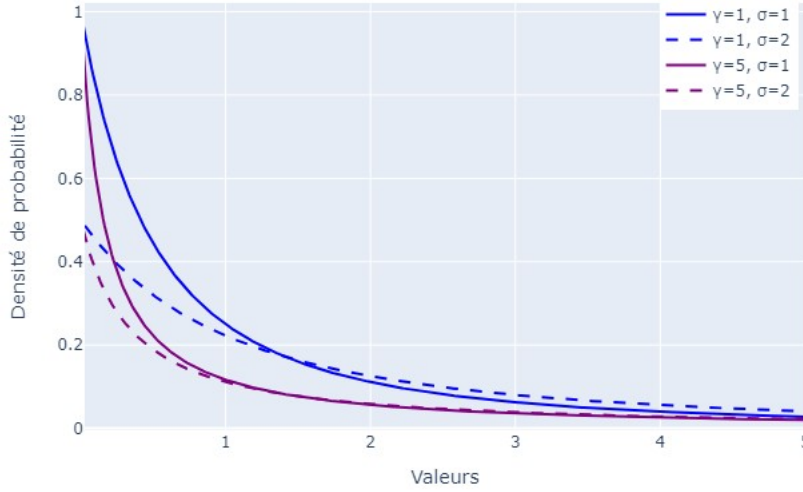


FIGURE 4.2 – Densité de la loi GPD selon γ et σ

Théorème de Pickands, Balkema et De Hann : *La fonction de répartition F de point terminal x_F appartient au domaine d'attraction de H_γ si et seulement si il existe une fonction positive σ telle que :*

$$\lim_{x \rightarrow x_F} \sup_{0 < x < x_F - u} |F_u(x) - G_{\gamma, \sigma(u)}(x)| = 0$$

Ce théorème suggère qu'à partir d'un certain seuil u , suffisamment élevé, la loi de Pareto Généralisée constitue une approximation fiable de la distribution des excès. Ainsi, la fonction de répartition $F_u(x - u)$ peut être approchée par $G_{\gamma, \sigma}(u)$.

Remarque :

- Comme le présente la figure ci-dessus 4.2, le choix des paramètres γ et σ modifie la forme de la queue de distribution.
- Lorsque $\gamma = 0$, alors la loi est exponentielle de paramètre $\frac{1}{\sigma}$.
- La loi GPD peut être écrite avec un paramètre de position μ également. $G_{\gamma, \sigma, \mu}(x) = G_{\gamma, \sigma}(x - \mu)$.
- L'indice des valeurs extrêmes γ est le même pour la GEV que pour la GPD.

4.3 Détermination du seuil des extrêmes

Pour mettre en place l'approche du dépassement de seuil, il est crucial de définir le seuil des extrêmes, noté u , à partir duquel l'analyse sera réalisée. Dans ce cas, u représente le montant de sinistre au-delà duquel les sinistres sont considérés comme étant extrêmes.

Le choix de ce seuil est d'une importance capitale et ne doit pas être pris à la légère. Un seuil u trop bas risque de biaiser l'étude en incluant des données qui ne sont pas véritablement extrêmes. En revanche, un u trop élevé réduit la quantité de données disponibles pour la modélisation, entraînant une estimation moins précise et une augmentation de la variance de l'étude. Il est important de souligner qu'en réassurance, l'accès à un grand nombre de données est souvent limité, ce qui rend chaque donnée essentielle pour la précision de l'analyse.

L'objectif est donc de trouver un équilibre optimal entre le biais et la variance.

4.3.1 Le mean excess plot

Le mean excess plot¹ est un graphique utilisé pour déterminer le seuil des valeurs extrêmes.

Il repose sur la fonction des excès moyens définie pour $X > 0$ par :

$$e(u) = \mathbb{E}[X - u | X > u] = \frac{\int_u^{+\infty} (1 - F(x)) dx}{1 - F(u)}$$

Avec F la fonction de répartition de X .

Cette fonction permet de quantifier l'écart moyen entre les sinistres et le seuil u , à condition que ces sinistres soient strictement supérieurs à ce seuil. Une estimation empirique de la fonction des excès moyen est définie par :

$$\hat{e}(u) = \frac{\sum X_i \mathbb{1}_{X_i > u}}{\sum \mathbb{1}_{X_i > u}} - u$$

Cet estimateur calcule donc la somme des excès des sinistres au-delà du seuil u , divisée par le nombre de sinistres dépassant ce seuil.

Le mean excess plot consiste donc à tracer **(u , $\hat{e}(u)$)** où u représente chaque montant de sinistre de la cédante. L'objectif est de déterminer le montant du sinistre à partir duquel les données adoptent un comportement similaire à celui d'une distribution de Pareto généralisée.

Proposition : Soit X une variable aléatoire suivant une $GPD(\gamma, \sigma)$. La fonction des excès moyens de X s'écrit pour tout $\gamma < 1$:

$$e(u) = \frac{\gamma}{1 - \gamma} \times u + \frac{\sigma}{1 - \gamma}$$

Par conséquent, si $X \sim GPD(\gamma, \sigma)$ alors la fonction des excès moyens est affine à partir de u .

Ainsi si la distribution des excès moyens des sinistres de la cédante présente une tendance linéaire au-delà d'un sinistre spécifique u , cela suggère que les données dépassant ce seuil suivent une loi de Pareto généralisée. Ainsi, il est nécessaire de déterminer le seuil à partir duquel la fonction devient approximativement linéaire pour identifier le seuil des valeurs extrêmes.

Exemple : Considérons un échantillon de données simulées représentant des montants de sinistre. L'objectif est de retrouver le seuil des valeurs extrêmes déjà fixé. L'échantillon est construit comme suit :

- $n = 150$.
- $u = 500\,011$, seuil des extrêmes.
- $X_i \sim \text{Lognorm}(1.40, 2)$, $\forall X_i < u$.
- $X_i \sim GPD(2, 1)$, $\forall X_i \geq u$.

Dans cet échantillon de sinistres, la queue de distribution suit une loi GPD. Par conséquent, une tendance linéaire devrait être observée sur le mean excess plot lorsque l'échantillon suit cette loi.

1. "Le graphique de la moyenne des excès" en français.

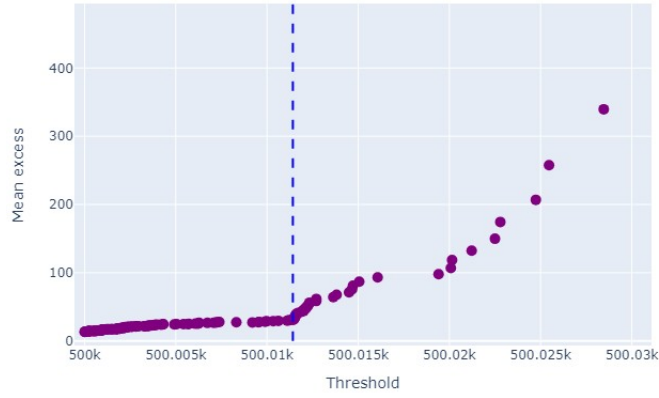


FIGURE 4.3 – Mean excess plot

Le graphique ci-dessus illustre clairement qu'à partir du seuil u , le reste de la distribution présente une tendance linéaire. Ainsi on retrouve bien le seuil u sur la ligne verticale.

Remarque : Identifier la linéarité peut s'avérer plus complexe lorsqu'on dispose d'un volume de données limité, comme le montrent les graphiques suivants. Il est donc crucial de prendre cet aspect en compte lorsqu'on cherche à déterminer le seuil.

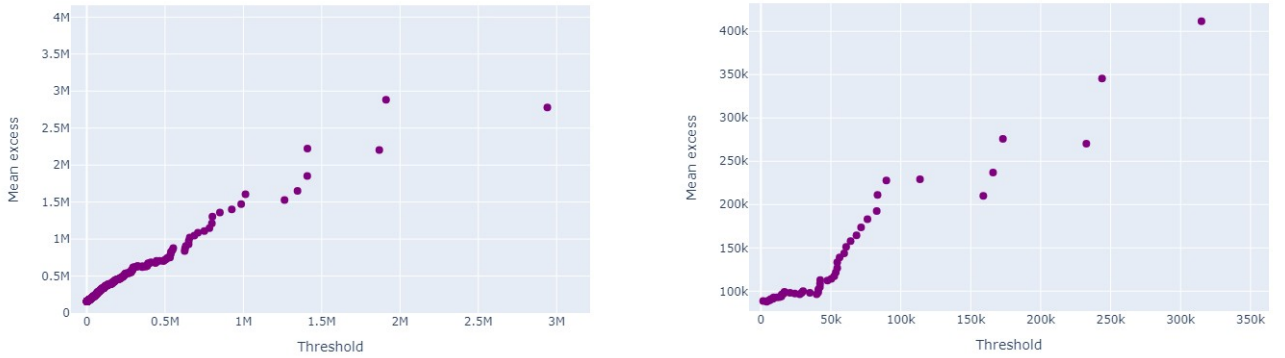


FIGURE 4.4 – Mean excess plot d'une GPD(0.7, 600 000) avec 500 observations (50 à droite).

4.3.2 Le threshold stability plot

Le threshold stability plot² est un graphique permettant également de déterminer le seuil des extrêmes u .

Il affiche l'estimation du paramètre γ de la loi GPD pour chaque seuil possible. Un seuil u_0 adéquat pour modéliser les dépassements est trouvé lorsque l'estimation du paramètre γ se stabilise. En d'autres termes, γ reste approximativement inchangé pour tous les seuils plus élevés que u_0 .

Ce graphique permet également d'analyser l'estimation du paramètre d'échelle σ . Le seuil u_0 qui permet une bonne modélisation des dépassements présente une relation connue avec tous les seuils u plus élevés que u_0 .

L'objectif de ce graphique est donc de trouver un seuil u_0 qui assure une modélisation fiable des dépassements, en recherchant la stabilité des paramètres.

2. "Le graphique du seuil stabilisé" en français.

Exemple :

Reprenons l'échantillon simulé dans l'exemple précédent, qui possède un seuil des extrêmes $u_0 = 500011$. Le threshold stability plot est affiché ci-dessous, avec l'estimation du paramètre de forme γ à gauche et celle du paramètre d'échelle σ à droite.

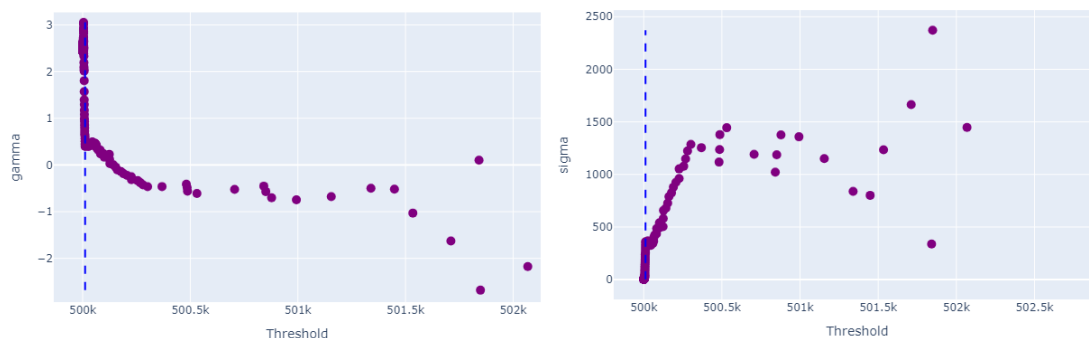


FIGURE 4.5 – Threshold stability plot (pour γ à gauche et pour σ à droite)

L'estimation du paramètre γ se stabilise bien à partir de u_0 , ce qui confirme que ce seuil est adéquat pour modéliser les dépassements. En ce qui concerne l'estimation du paramètre σ , une relation linéaire est observable à partir du point u_0 .

Ainsi, le seuil des extrêmes (représenté par la ligne bleue pointillée) peut être retrouvé à l'aide de ce graphique.

Remarque :

Lors de l'interprétation graphique, les derniers points ne doivent pas être pris avec une grande considération en raison de leur forte instabilité, car les estimations sont réalisées avec de moins en moins de données.

De plus, l'estimation du paramètre d'échelle σ est plus difficile à interpréter pour des seuils élevés en raison de la grande variance de l'estimation. Par conséquent, dans la suite, seules les estimations du paramètre γ seront étudiées dans le threshold stability plot.

4.4 Détermination du domaine d'attraction

Avant de procéder à l'estimation numérique des paramètres de la GPD qui ajuste le mieux les données, il est pertinent d'examiner le domaine d'attraction auquel cet échantillon appartient. Cette analyse est essentielle pour choisir la méthode appropriée d'estimation des paramètres, car certaines méthodes présentent des contraintes sur le paramètre γ .

4.4.1 QQ plot

Le QQ plot³ est une méthode graphique permettant de comparer deux distributions en affichant leurs quantiles respectifs. Si les points sont bien alignés, cela signifie que les deux distributions sont similaires. Cette méthode est souvent utilisée pour vérifier si les points empiriques peuvent être bien ajustés avec une distribution théorique choisie. Dans ce contexte, les quantiles empiriques sont placés en abscisse et les quantiles théoriques en ordonnée.

Dans ce cas, le QQ plot est utilisé pour déterminer le domaine d'attraction de l'échantillon de données. Pour ce faire, il suffit de comparer les quantiles empiriques aux quantiles

3. "Le graphique quantiles-quantiles" en français.

théoriques d'une loi exponentielle. En effet, la loi exponentielle ayant un $\gamma = 0$ (domaine de Gumbel) se situe entre les deux autres domaines d'attraction. Si les données empiriques sont dans le domaine d'attraction de Fréchet ($\gamma > 0$), le QQ plot sera convexe. Inversement, si les données empiriques sont dans le domaine d'attraction de Weibull ($\gamma < 0$), le QQ plot sera concave. L'analyse de la tendance de la courbe du QQ plot permet de connaître le domaine d'attraction des points empiriques.

Pour tracer le QQ plot, la formule suivante est utilisée :

$$\left(X_i, -\frac{1}{\lambda} \ln \left(1 - \frac{i}{n+1} \right) \right)$$

Avec :

- X_i : les quantiles empiriques triées dans l'ordre croissant.
- λ : le paramètre de l'exponentielle.
- n : le nombre total de données.

Exemple :

Considérons trois échantillons simulés de 50 données provenant de distributions GPD, représentant les trois domaines d'attraction :

- **Echantillon n°1** : $X \sim GPD(1000, 250, 1)$ (QQ plot de gauche).
- **Echantillon n°2** : $X \sim GPD(1000, 250, 0)$ (QQ plot du milieu).
- **Echantillon n°3** : $X \sim GPD(1000, 250, -1)$ (QQ plot de droite).

Les trois QQ plots ci-dessous représentent chacun de ces échantillons comparés à une loi exponentielle de paramètre $\lambda = 1$.

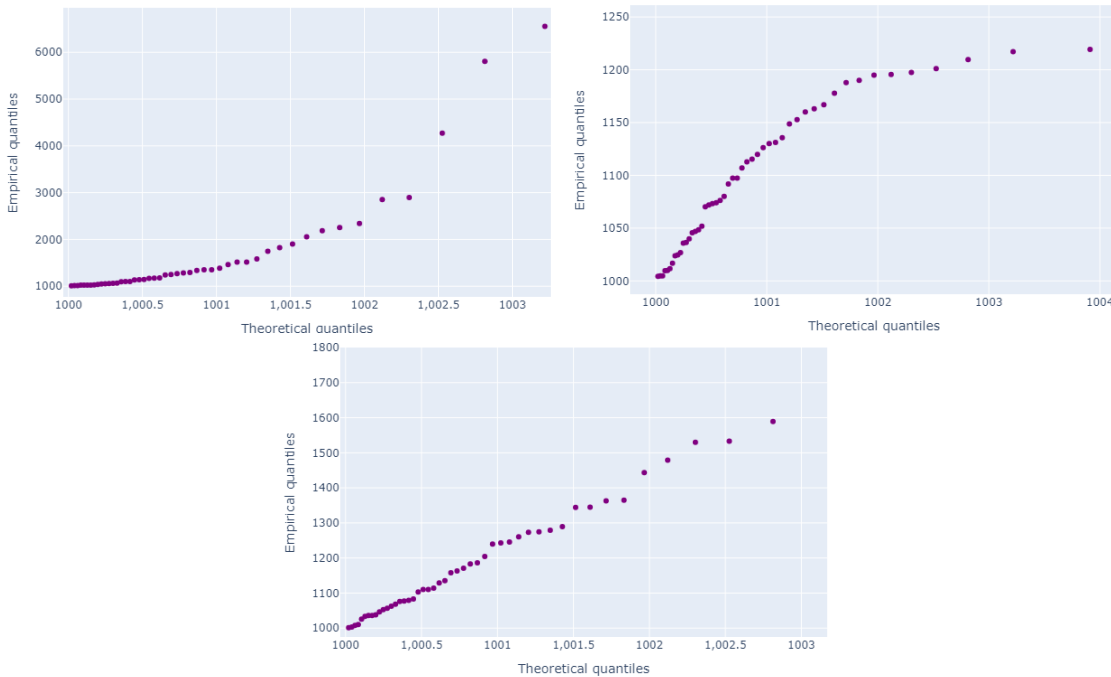


FIGURE 4.6 – QQ Plot : Fréchet à gauche, Gumbel au centre et Weibull à droite

Le QQ plot de gauche est convexe car l'échantillon n°1 est dans le domaine d'attraction de Fréchet. Le QQ plot du milieu est linéaire car l'échantillon n°2 est dans le domaine de Gumbel. Le QQ plot de droite est concave car l'échantillon n°3 est dans le domaine d'attraction de Weibull.

Ces observations confirment que le QQ plot est un outil efficace pour identifier le domaine d'attraction des données empiriques.

4.4.2 Return level plot

Le return level plot⁴ est un outil graphique permettant également de déterminer le domaine d'attraction d'un échantillon de données.

Ce graphique affiche les niveaux de retour z_p en fonction de la période de retour associée, exprimée en $\frac{1}{p}$. Les niveaux de retour sont obtenus en utilisant l'équation suivante :

$$G_{\gamma,\sigma,\mu}(z_p) = 1 - p$$

où $G_{\gamma,\sigma,\mu}(x)$ est la fonction de répartition de la loi GPD. Ainsi, les niveaux de retour sont de la forme suivante :

$$z_p = \begin{cases} \mu + \frac{\sigma}{\gamma}((1-p)^{-\gamma} - 1) & \text{si } \gamma \neq 0 \\ -\mu - \sigma \log(1-p) & \text{si } \gamma = 0 \end{cases}$$

Le niveau de retour z_p correspond à la valeur pour laquelle la probabilité de dépassement est p .

Le return level plot peut s'afficher avec les axes suivants : $(-\log(y_p), z_p)$ avec $y_p = 1-p$.

Exemple :

Le return level plot ci-dessous permet d'analyser le comportement de ce dernier selon le domaine d'attraction d'un échantillon. Les échantillons simulés de 50 données sont les suivants :

- **Echantillon n°1** : $X \sim GPD(1000, 250, 0.5)$ - domaine de Fréchet.
- **Echantillon n°2** : $X \sim GPD(1000, 250, 0)$ - domaine de Gumbel.
- **Echantillon n°3** : $X \sim GPD(1000, 250, -1)$ - domaine de Weibull.

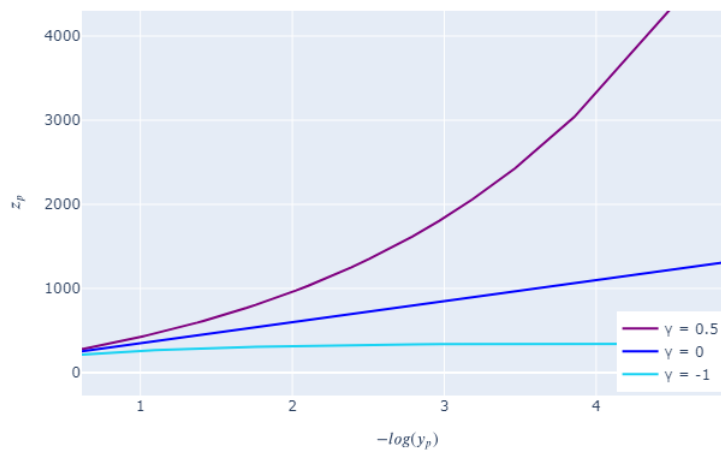


FIGURE 4.7 – Return level plot

Ainsi, l'interprétation des courbes est la suivante :

- Si la courbe est convexe, l'échantillon est dans le domaine de Fréchet.
- Si la courbe est linéaire, l'échantillon est dans le domaine de Gumbel.
- Si la courbe est concave, l'échantillon est dans le domaine de Weibull.

4. "Le graphique du niveau de retour" en français.

Remarque :

Les niveaux de retour empiriques peuvent également être affichés sur le graphique afin d'analyser l'ajustement et le domaine d'attraction de l'échantillon.

4.4.3 Test du rapport de vraisemblance

Le test du rapport de vraisemblance est un test statistique permettant de comparer deux modèles entre eux. Lors de la détermination du domaine d'attraction, ce test permet de déterminer si l'hypothèse selon laquelle le domaine d'attraction de l'échantillon est de Gumbel peut être rejetée. Ce test est formulé de la manière suivante :

- **Hypothèse H_0 :** $\gamma = 0$ (domaine de Gumbel).
- **Hypothèse H_1 :** $\gamma \neq 0$.

Ici, $\hat{\gamma}$ est l'estimateur du maximum de vraisemblance pour γ et $\hat{\gamma}_0$ est l'estimateur du maximum de vraisemblance sous H_0 .

La statistique de test λ est calculée selon l'expression :

$$\Lambda = -2 \log \frac{L(\hat{\gamma}_0)}{L(\hat{\gamma})}$$

où L est la fonction de vraisemblance.

Sous l'hypothèse H_0 , λ suit une loi du χ^2 à 1 degré de liberté (correspondant au nombre de contraintes imposées par H_0).

La p-valeur est calculée comme suit :

$$\text{p-valeur} = P(\chi_1^2 \geq \lambda) = 1 - F_{\chi_1^2}(\lambda)$$

L'hypothèse H_0 est rejetée à un niveau de confiance de 95% si la statistique de test λ est supérieure au quantile 0,05 de la loi $\chi^2(1)$. Autrement dit, si la p-valeur est inférieure à 0,05, on peut rejeter l'hypothèse H_0 et ainsi l'échantillon de données n'appartient pas au domaine de Gumbel.

Cette méthode fournit des preuves supplémentaires aux méthodes graphiques pour déterminer le domaine d'attraction de l'échantillon.

4.5 Estimation paramétrique de la GPD

Une fois le seuil des valeurs extrêmes et le domaine d'attraction déterminés, il est nécessaire d'estimer les paramètres de la GPD pour l'échantillon des sinistres dépassant le montant u . Cet échantillon est constitué de la manière suivante :

$$\forall i \in [1, N_u], Y_i = X_{(n-N_u+i)} - X_{(n-N_u)}$$

N_u : le nombre de sinistres dépassant le montant u

4.5.1 La méthode du maximum de vraisemblance

Les paramètres γ et σ de la GPD peuvent être estimés en utilisant la méthode du maximum de vraisemblance. Pour ce faire, il s'agit de déterminer les valeurs de ces paramètres

qui maximisent la fonction de log-vraisemblance suivante :

$$l(\gamma, \sigma | y_1, \dots, y_{N_u}) = \begin{cases} -N_u \ln(\sigma) - \left(\frac{1}{\gamma} + 1\right) \sum_{i=1}^{N_u} \ln(1 + \gamma \frac{y_i}{\sigma}) & \text{si } \gamma \neq 0 \\ -N_u \ln(\sigma) - \frac{1}{\sigma} \sum_{i=1}^{N_u} y_i & \text{si } \gamma = 0 \end{cases}$$

Les estimateurs $\hat{\gamma}_{N_u}^{ML}$ et $\hat{\sigma}_{N_u}^{ML}$ sont obtenus en dérivant la log-vraisemblance par rapport à chacun de ces paramètres, puis en résolvant :

$$\begin{cases} \frac{\partial l(\gamma, \sigma | y_i)}{\partial \gamma} = 0 \\ \frac{\partial l(\gamma, \sigma | y_i)}{\partial \sigma} = 0 \end{cases}$$

Dans le cas où $\gamma = 0$, l'estimateur $\hat{\sigma}^{ML}$ est donné par :

$$\hat{\sigma}^{ML} = \frac{\sum_{i=1}^{N_u} y_i}{N_u}$$

En revanche, dans le cas contraire, la solution du système n'est pas aussi directe. Cependant, des méthodes numériques itératives sont employées pour résoudre le système.

De plus, Smith (1984) a montré la normalité asymptotique des estimateurs $\hat{\gamma}$ et $\hat{\sigma}$ dans le cas où $\gamma > -\frac{1}{2}$.

4.5.2 La méthode des moments

L'estimation des paramètres peut également être réalisée avec la méthode des moments. Cette approche implique d'égaliser les deux premiers moments théoriques de la GPD avec les deux premiers moments empiriques. Il suffit ensuite de résoudre le système à deux équations. Il convient de noter que cette méthode est applicable uniquement lorsque $\gamma < \frac{1}{2}$, condition nécessaire pour garantir l'existence des deux premiers moments empiriques.

Les deux premiers moments d'une loi GPD sont les suivants :

$$\begin{aligned} m_1 &= \mathbb{E}[X] = \frac{\sigma}{1 - \gamma} \\ m_2 &= \mathbb{E}[X^2] = \frac{2\sigma^2}{(1 - \gamma)(1 - 2\gamma)} \end{aligned}$$

Ainsi, il est simplement nécessaire d'égaliser ces moments théoriques avec les deux premiers moments empiriques des montants des sinistres.

$$\begin{aligned} \hat{m}_1 &= \frac{1}{N_u} \sum_{i=1}^{N_u} y_i = \bar{Y} \\ \hat{m}_2 &= \frac{1}{N_u} \sum_{i=1}^{N_u} y_i^2 = \hat{s}^2 + \bar{Y} \end{aligned}$$

Avec $\hat{s}$ l'écart-type des sinistres supérieur à u .

Ainsi, les estimateurs de la loi GPD sont les suivants :

$$\hat{\gamma}^{MM} = \frac{1}{2} \left[1 - \frac{\bar{Y}^2}{\hat{s}^2} \right]$$

$$\hat{\sigma}^{MM} = \bar{Y} \left[\frac{1 + \frac{\bar{Y}^2}{\hat{s}^2}}{2} \right]$$

4.5.3 La méthode des moments pondérés

La méthode des moments pondérés est légèrement moins contraignante que la méthode précédente car elle s'applique pour $\gamma < 1$ (nécessite uniquement un moment d'ordre un). Pour commencer, définissons ce qu'est un moment pondéré :

Définition : Moment pondéré

Soit Z une variable aléatoire avec F comme fonction de répartition. Si Z est intégrable, le moment pondéré de Z d'ordre $r \in \mathbb{N}$ et $s \in \mathbb{N}$ est :

$$WM_Z(r, s) = \mathbb{E}[ZF^r(Z)(1 - F(Z))^s]$$

Pour la suite, il est considéré que $r = 0$. Ainsi pour $Z \sim GPD(\gamma, \sigma)$, pour $\gamma < 1$:

$$WM_Z(0, s) = \mathbb{E}[Z(1 - F(Z))^s] = \frac{\sigma}{(s+1)(s+1-\gamma)} = v_s$$

Ainsi en résolvant le système suivant :

$$\begin{cases} \frac{\sigma}{1-\gamma} = v_0 \\ \frac{\sigma}{4-2\gamma} = v_1 \end{cases}$$

Les estimateurs suivants sont obtenus :

$$\hat{\gamma} = \frac{4\hat{v}_1 - \hat{v}_0}{2\hat{v}_1 - \hat{v}_0} \text{ et } \hat{\sigma} = \frac{2\hat{v}_0\hat{v}_1}{\hat{v}_0 - 2\hat{v}_1}$$

Avec $\hat{v}_s$ représentant les moments pondérés empiriques des données :

$$\hat{v}_s = \frac{1}{N_u} \sum_{i=1}^{N_u} Z_i \frac{(N_u - i + s)! N_u!}{(N_u - i)! (N_u + s)!}$$

4.6 Estimation semi-paramétrique de l'indice des valeurs extrêmes

4.6.1 Estimateur de Hill

L'estimateur de Hill, introduit en 1975, est utilisé pour déterminer l'indice des valeurs extrêmes γ . Néanmoins, cette méthode est applicable uniquement aux distributions qui appartiennent au domaine d'attraction de Fréchet, c'est-à-dire lorsque $\gamma > 0$.

Il est calculé de la manière suivante :

$$\hat{\gamma}^H(N_u) = \frac{1}{N_u} \sum_{i=1}^{N_u} \ln(X_{(n-i+1)}) - \ln(X_{(n-N_u)})$$

Théorème : Loi limite pour $\hat{\gamma}^H$

Si $F \in DA(\gamma)$, $\gamma > 0$ et (N_u) suite intermédiaire, c'est-à-dire $N_u \rightarrow +\infty$ et $\frac{N_u}{n} \rightarrow 0$.

$$\sqrt{N_u} \left(\hat{\gamma}^H(N_u) - \mathbb{E} \left[\hat{\gamma}^H(N_u) \right] \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \gamma^2)$$

Remarque :

Cet estimateur est construit à partir de l'estimateur du maximum de vraisemblance d'une loi de Pareto de paramètre α . La relation $\alpha = \frac{1}{\gamma}$ lie ces deux estimateurs.

4.6.2 Estimateur de Pickands

L'estimateur de Pickands, introduit en 1975, est un autre outil pour déterminer l'indice des valeurs extrêmes γ . De plus, cet estimateur a l'avantage d'être valide dans les trois domaines d'attraction. Il est calculé de la manière suivante :

$$\hat{\gamma}^P(N_u) = \frac{1}{\ln(2)} \ln \left(\frac{X_{(n-N_u)} - X_{(n-2N_u)}}{X_{(n-2N_u)} - X_{(n-4N_u)}} \right)$$

Théorème : Normalité asymptotique

Si $F \in DA(\gamma)$ avec $\gamma > \frac{-1}{2}$,

$$\sqrt{N_u} \left(\frac{\hat{\gamma}^P(N_u) - \gamma}{\sigma(\gamma)} \right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Avec : $\sigma(\gamma) = \frac{\gamma \sqrt{2^{2\gamma+1} + 1}}{2(2^\gamma - 1) \ln(2)}$

Cet estimateur est calculé à partir des distances entre deux statistiques d'ordre sans recourir au maximum de l'échantillon. Cela entraîne une perte d'information concernant la queue de distribution. Par conséquent, l'estimateur est plus volatile que l'estimateur de Hill.

Pour conclure ce chapitre, une fois le seuil des valeurs extrêmes u pour les sinistres de la cédante défini, la distribution de la sévérité des sinistres sera segmentée en deux parties. Le corps de la distribution sera modélisé empiriquement, tandis que la queue de distribution sera modélisée par une GPD. L'estimation de cette loi sur les données sera effectuée exclusivement à l'aide de la méthode du maximum de vraisemblance. Le chapitre suivant détaillera l'application de cette méthode à une tarification type.

Chapitre 5

Modélisation de la sévérité par la théorie des valeurs extrêmes

Les notions théoriques des valeurs extrêmes introduites, il est possible de reprendre la tarification du traité XS DAB abordée dans le chapitre 3. Cette section se concentre exclusivement sur la modélisation de la sévérité des sinistres de cette cédante. La méthode pour la fréquence reste inchangée.

5.1 Détermination du seuil des extrêmes

La première étape pour modéliser la sévérité consiste à déterminer le seuil des valeurs extrêmes. Pour cela, les deux méthodes présentées dans le chapitre 4 (4.3) sont appliquées.

5.1.1 Mean excess plot

Le mean excess plot des sinistres de la cédante est affiché ci-dessous.

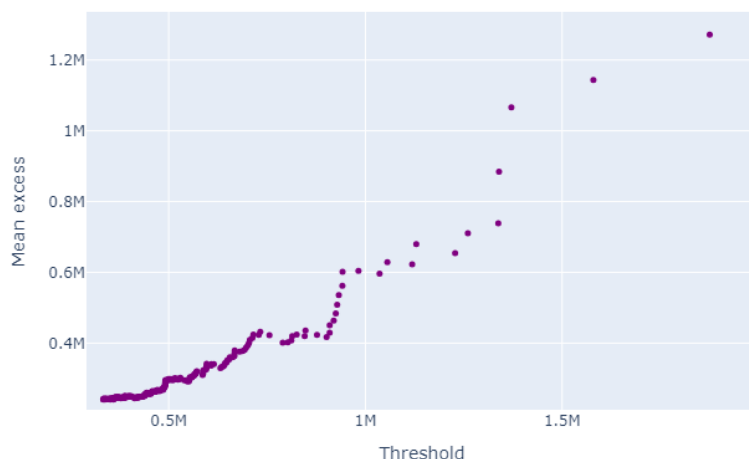


FIGURE 5.1 – Mean excess plot

Le mean excess plot expose plusieurs informations importantes.

Premièrement, la partie gauche du graphique montre une linéarité relativement stable. Cependant, à partir de 485 488 sur l'axe des abscisses, un changement plus marqué dans la pente est observé. Ce montant indique une transition par rapport à la partie inférieure des seuils du graphique, où la tendance reste peu pentue. Il est également important de noter que les derniers points du graphique présentent des valeurs instables, considérées comme des valeurs aberrantes.

Ainsi, avec ces observations, un seuil des extrêmes fixé à $u_0 = 485488$ semble être un choix approprié.

5.1.2 Threshold stability plot

La seconde méthode pour déterminer le seuil des extrêmes est le threshold stability plot. Cette méthode va permettre de confirmer ou non le choix du seuil des extrêmes fait avec le mean excess plot.

Le threshold stability plot de la cédante pour l'estimation du paramètre γ est affiché ci-dessous.

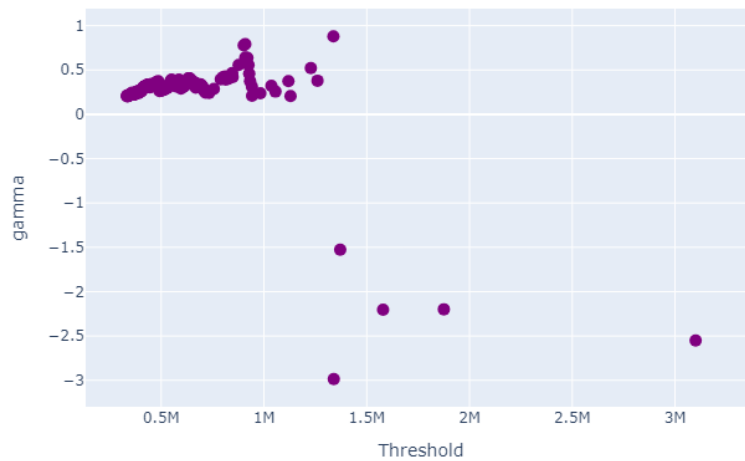


FIGURE 5.2 – Threshold stability plot

Ce graphique permet d'arriver aux observations suivantes :

Premièrement, la partie gauche du graphique présente une légère augmentation linéaire de la courbe. À partir de 485 488.5 sur l'axe des abscisses, l'estimation du paramètre γ semble se stabiliser autour d'environ 0,40. De plus, les points situés plus à droite ne sont pas pris en compte en raison de l'instabilité de l'estimation du paramètre γ pour les montants élevés.

Ainsi, le seuil choisi à $u_0 = 485488$ par le mean excess plot est confirmé ici avec le threshold stability plot.

5.2 Domaine d'attraction

Une fois le seuil des extrêmes déterminé, il est maintenant possible d'étudier le domaine d'attraction de la queue de la distribution. Pour cela, les trois méthodes présentées dans le chapitre n°4 (4.4) sont appliquées.

5.2.1 QQ plot

Le but du QQ plot exponentiel ici est d'observer la tendance de la courbe permettant de connaître le domaine d'attraction de la queue de distributions des sinistres de la cédante. Le QQ-plot est affiché ci-dessous.

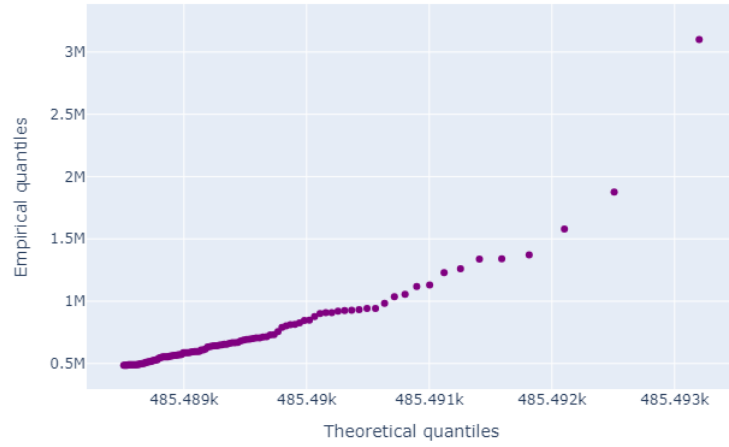


FIGURE 5.3 – QQ plot exponentiel

La tendance de la courbe ici n'est pas très explicite. Il en ressort qu'elle peut être légèrement convexe ou plutôt linéaire. Ainsi, cette méthode ne permet pas de donner le domaine d'attraction avec certitude ici mais au moins d'en éliminer un : le domaine de Weibull (courbe concave).

5.2.2 Test du rapport de vraisemblance

Le test du rapport de vraisemblance permet de déterminer si l'échantillon présent appartient au domaine de Gumbel ou non. Ce test est réalisé de la même manière qu'il a été présenté dans le chapitre n°4 (4.4.3).

La p-valeur obtenue est de 1.086×10^{-11} . Ainsi, l'hypothèse H_0 est rejetée, ce qui signifie que la queue de distribution des sinistres de la cédante n'appartient pas au domaine de Gumbel. Ce test permet donc de lever les doutes soulevés par le QQ plot précédent.

5.2.3 Return level plot

Le return level plot est plus explicite pour déterminer le domaine d'attraction des sinistres de la cédante ici. Le graphique est présenté ci-dessous.

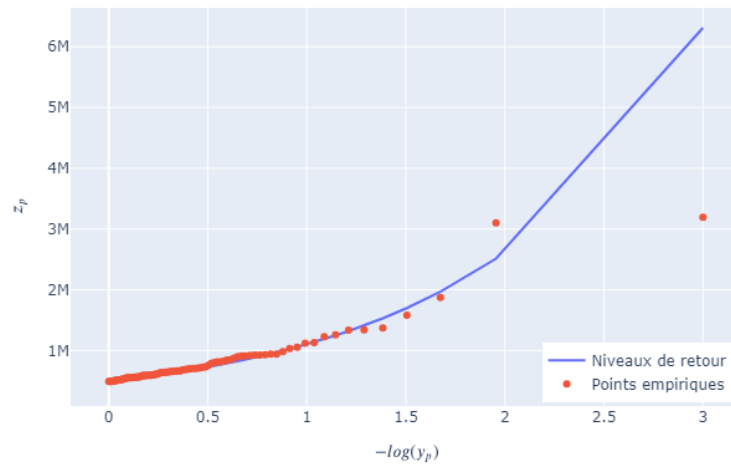


FIGURE 5.4 – Return level plot

La courbe des niveaux de retour est convexe, ce qui permet d'affirmer que la queue de la distribution appartient au domaine de Fréchet. De plus, les sinistres empiriques de

la cédante s'ajustent bien à cette courbe, ce qui valide que l'estimation des niveaux de retour est représentative des données.

Ainsi, il est possible de conclure que la queue de la distribution des sinistres de la cédante appartient au domaine de Fréchet.

5.3 Construction de la courbe de sévérité

Après avoir effectué les étapes précédentes, il est possible de construire la courbe représentant la sévérité des sinistres de la cédante. Cette courbe est décomposée en trois parties :

- le corps de la distribution,
- la queue de la distribution,
- la jonction reliant les deux parties.

5.3.1 Le corps de la distribution

Pour modéliser le corps de la distribution, la fonction de répartition empirique est utilisée.

Cette approche est particulièrement adaptée pour le corps de la distribution, qui regroupe un grand volume de données représentant les sinistres se produisant avant la franchise ou au sein des tranches travaillantes. Étant donné que ces tranches sont souvent évaluées sur la base du BC, il n'est pas essentiel d'appliquer des modèles très complexes pour cette section de la distribution. Par conséquent, une modélisation empirique du corps de la distribution est tout à fait appropriée et justifiable dans ce contexte.

5.3.2 La queue de la distribution

La modélisation de la queue de la distribution commence à partir du seuil des extrêmes. À partir de u_0 , une loi GPD est ajustée à l'échantillon des données restantes. Pour cet ajustement, la méthode du maximum de vraisemblance est utilisée.

Pour les sinistres au-dessus du seuil des extrêmes, l'ajustement suivant est obtenu :

Loi	Méthode	Paramètres			Test KS	AIC
		μ	σ	γ		
GPD	MLE	485 488.5	165 759.16	0.389	Vrai	2 953.56

FIGURE 5.5 – Tableau ajustement GPD

Ainsi, $\forall X_i \geq u_0, X \sim GPD(485488.5, 165759.16, 0.389)$.

Le tableau précise que le test de Kolmogorov-Smirnov est validé, confirmant que l'ajustement avec la GPD est adéquat et représente bien les données.

De plus, un QQ plot est affiché ci-dessous pour vérifier la qualité de l'ajustement :

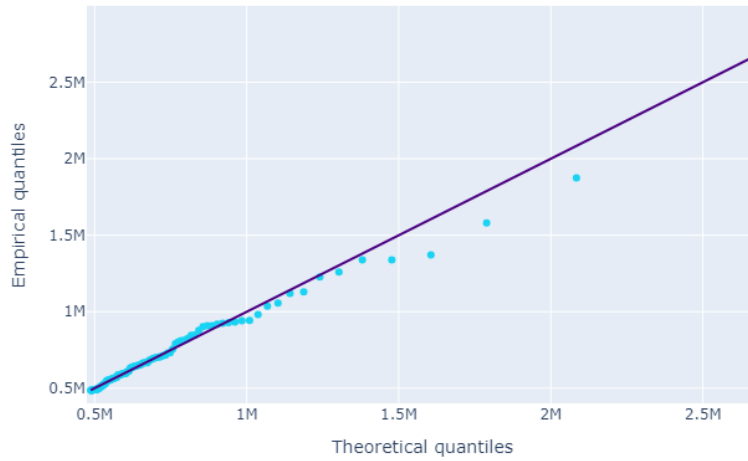


FIGURE 5.6 – QQ plot GPD

Les quantiles sont bien alignés sur la ligne théorique. Par conséquent, l'ajustement de la GPD représente fidèlement la queue de la distribution des sinistres de la cédante.

5.3.3 Jonction

La dernière étape pour construire la courbe de sévérité des sinistres est la jonction reliant le corps et la queue de la distribution. Cette jonction est cruciale pour garantir la continuité de la distribution. Elle est réalisée à l'aide d'une fonction linéaire reliant les deux parties de la distribution.

Ainsi, la construction de la courbe de sévérité donne la fonction de répartition suivante.

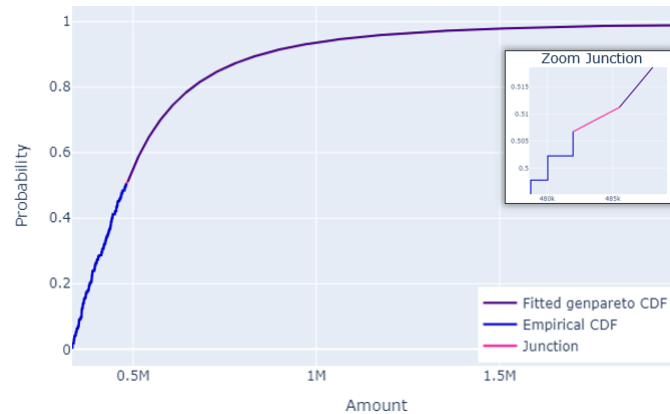


FIGURE 5.7 – Fonction de répartition composite

5.4 Calcul de la prime pure

Une fois la fréquence et la sévérité modélisées, il est possible de calculer les taux purs pour chaque tranche.

5.4.1 Simulation

Comme pour la modélisation classique (3.5.1), des simulations sont effectuées sur les données afin d'obtenir une estimation plus large et plus précise de la sinistralité future de la cédante.

La méthode de simulation des YLT a été ajustée pour intégrer les modifications apportées à la modélisation de la sévérité. Comme dans l’approche classique (3.5.1), la fréquence des sinistres est simulée à l’aide de la loi de Poisson ajustée, générant ainsi un vecteur représentant le nombre de sinistres par année. En revanche, la simulation des montants des sinistres a été modifiée pour prendre en compte le type du sinistre (corps, jonction, queue).

Identification du type de sinistre

Pour chaque sinistre simulé au cours de ces années, un tirage supplémentaire est effectué afin de déterminer son type, c’est-à-dire si ce sinistre appartient au corps, à la jonction ou à la queue de la distribution. Le but étant ainsi de garantir que la répartition simulée reste fidèle à la sinistralité observée de la cédante.

Les proportions des types de sinistres sont calculées à partir de l’échantillon empirique de la cédante à l’aide des notations suivantes :

- N_{total} : Nombre total de sinistres observés.
- X_i : Montant du i -ème sinistre.
- u_{jonction} : Montant du plus grand sinistre avant le seuil des extrêmes.
- u : Seuil des extrêmes.

Les proportions sont déterminées de la manière suivante :

- **Le corps** : Proportion des sinistres inférieurs à u_{jonction} .

$$P_{\text{corps}} = \frac{1}{N_{\text{total}}} \sum_{i=1}^{N_{\text{total}}} \mathbb{1}_{\{X_i < u_{\text{jonction}}\}} * 100$$

- **La jonction** : Proportion des sinistres égaux à u_{jonction} .

$$P_{\text{jonction}} = \frac{1}{N_{\text{total}}} \sum_{i=1}^{N_{\text{total}}} \mathbb{1}_{\{X_i = u_{\text{jonction}}\}} * 100$$

- **La queue** : Proportion de sinistres supérieurs à u .

$$P_{\text{queue}} = \frac{1}{N_{\text{total}}} \sum_{i=1}^{N_{\text{total}}} \mathbb{1}_{\{X_i \geq u\}} * 100$$

Ces proportions sont ensuite utilisées pour effectuer un tirage aléatoire permettant de déterminer le type de chaque sinistre simulé.

Simulation des montants des sinistres

Une fois le type de chaque sinistre déterminé, les montants sont simulés en fonction de la classe correspondante :

- **Pour les sinistres appartenant au corps de la distribution** : Le montant est tiré directement de l’échantillon empirique de la cédante, en cohérence avec la modélisation empirique du corps.
- **Pour les sinistres de la jonction** : Un tirage est réalisé à partir d’une loi uniforme sur l’intervalle correspondant (entre u_{jonction} et u).

- **Pour les sinistres situés dans la queue de distribution :** Le montant est simulé en utilisant la loi GPD ajustée aux données de la queue de distribution de la cédante.

Cette méthodologie permet d'obtenir un montant simulé précis pour chaque sinistre, aboutissant à la création d'une matrice YLT regroupant les sinistres simulés par année.

Une fois les simulations terminées, la matrice YLT est complétée. Le reste de la méthode utilisée pour calculer les taux purs par tranche reste inchangé, garantissant ainsi la continuité du processus de tarification.

Exemple :

Considérons à présent un exemple illustrant la simulation de la sinistralité future pour la cédante sur une période de 3 ans, avec un seuil u fixé à 200 000 €. Détaillons ci-dessous le déroulement de cette méthode :

Étape 1 : Simulation de la fréquence des sinistres.

Un vecteur représentant le nombre de sinistres par année est généré à partir de la loi de Poisson ajustée à la fréquence des sinistres de la cédante. Le vecteur obtenu est le suivant :

$$\text{Vecteur initial} = \begin{pmatrix} \text{Années} & \text{Nombre de sinistres} \\ 1 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 3 & 1 \\ 3 & 1 \end{pmatrix}$$

Ainsi, il a été simulé que 2 sinistres surviennent la première année, 1 sinistre la deuxième année et 3 sinistres la dernière année.

Étape 2 : Attribution des types de sinistres.

Une fois le nombre de sinistres obtenu, il faut également connaître le type de chaque sinistre (corps, jonction, queue). Pour cela, la proportion de chaque type est calculée à partir de l'échantillon empirique de la cédante :

$$P_{\text{corps}} = 60\% \quad P_{\text{jonction}} = 10\% \quad P_{\text{queue}} = 30\%$$

Ainsi, chaque sinistre simulé peut être assigné de manière aléatoire à un type en fonction de ces proportions :

$$\text{Vecteur avec types} = \begin{pmatrix} \text{Années} & \text{Nombre de sinistres} & \text{Type} \\ 1 & 1 & \text{Corps} \\ 1 & 1 & \text{Queue} \\ 2 & 1 & \text{Corps} \\ 3 & 1 & \text{Jonction} \\ 3 & 1 & \text{Queue} \\ 3 & 1 & \text{Corps} \end{pmatrix}$$

Étape 3 : Simulation des montants selon le type.

Ensuite, pour chaque sinistre, un montant est simulé selon son type. Les simulations sont les suivantes :

- **Corps** : Tirage dans l'échantillon empirique de la cédante.
- **Jonction** : Tirage à partir d'une loi uniforme entre $u_{jonction}$ et u .
- **Queue** : Tirage suivant la loi GPD ajustée à la queue de distribution de la cédante.

Les montants simulés sont ensuite intégrés dans la matrice YLT, exprimés ici en milliers d'euros, comme illustré ci-dessous :

$$\text{Matrice YLT} = \begin{pmatrix} \text{Années} & \text{Pertes} & \text{Type} \\ 1 & 160 & \text{Corps} \\ 1 & 220 & \text{Queue} \\ 2 & 170 & \text{Corps} \\ 3 & 190 & \text{Jonction} \\ 3 & 310 & \text{Queue} \\ 3 & 150 & \text{Corps} \end{pmatrix}$$

Étape 4 : Obtention de la matrice YLT finale.

Pour obtenir la matrice finale utilisée pour les calculs YLT, la colonne "Type" est supprimée, ne laissant que les années et les montants simulés.

$$\text{Matrice YLT finale} = \begin{pmatrix} \text{Années} & \text{Pertes} \\ 1 & 160 \\ 1 & 220 \\ 2 & 170 \\ 3 & 190 \\ 3 & 310 \\ 3 & 150 \end{pmatrix}$$

Ainsi, cette simulation permet d'obtenir une matrice YLT ajustée en tenant compte de la modélisation du corps, de la jonction, et de la queue de distribution.

Remarque :

Dans la tarification actuelle du traité XS DAB, les proportions observées sont les suivantes :

$$P_{corps} = 47.42\% \quad P_{jonction} = 0.04\% \quad P_{queue} = 52.54\%$$

La proportion élevée de sinistres appartenant à la queue de distribution peut sembler inhabituelle, mais elle est courante en réassurance. En effet, la cédante ne transmet pas l'ensemble de sa sinistralité, mais plutôt les sinistres déjà considérés comme "extrêmes", ce qui signifie qu'un tri préalable a été effectué. Cela explique pourquoi le seuil des extrêmes peut être positionné relativement bas dans la distribution empirique, entraînant ainsi une proportion plus importante de sinistres dans la queue par rapport à un portefeuille classique, où tous les sinistres sont pris en compte.

5.4.2 Résultats

Le tableau ci-dessous présente les taux purs ainsi que les primes pures proposées pour chaque tranche avec cette méthode de tarification.

Tranche	Taux pur	Prime pure
T1	0,735%	888 368
T2	0,224%	269 781
T3	0,011%	14 716
T4	0,004%	5 311
Total	0,974%	1 178 176

FIGURE 5.8 – Tableau taux et primes pures

La modélisation donne un taux pur de 0,974 % pour l'ensemble du traité, correspondant à une prime pure de 1 178 176 €. Ce taux pur est presque identique à celui obtenu par la modélisation classique (3.5.2). Néanmoins, les taux purs obtenus selon la tranche ont été modifiés par rapport à la modélisation classique. En effet, le taux pur de la T1 avec le modèle composite a légèrement diminué par rapport au modèle classique, tandis que les taux purs des T2, T3, et T4 ont augmenté.

L'application de la théorie des valeurs extrêmes dans la modélisation de la sévérité des sinistres a permis d'obtenir des résultats plus justifiables. Cette approche offre une meilleure précision dans la queue de la distribution, où se situent les sinistres extrêmes. En conséquence, la tarification des tranches élevées devient plus fiable et reflète plus fidèlement le risque associé à ces sinistres.

Ainsi, une modélisation plus complète de la sévérité des sinistres permet d'obtenir des résultats plus justifiables, renforçant ainsi la confiance envers les estimations réalisées.

5.5 Limites et enjeux

La méthode présentée ici comporte néanmoins certaines limites.

Premièrement, l'un des problèmes fondamentaux en réassurance est le nombre souvent limité de données disponibles pour modéliser des lois statistiques. La loi GPD, qui possède deux paramètres à estimer, est particulièrement sensible à cette limitation. Lorsque les données sont peu nombreuses, les estimations des paramètres de la GPD peuvent être moins robustes et sujettes à des erreurs plus importantes. Bien que la GPD soit théoriquement plus appropriée pour modéliser la queue de distribution, l'erreur de modélisation ne se mesure pas aussi bien que l'erreur d'estimation.

Chez CCR Re, la méthode utilisant la théorie des valeurs extrêmes a été implémentée avec une loi de Pareto pour modéliser la queue de distribution en cas de données limitées. Néanmoins, ce mémoire continuera à se focaliser sur la GPD afin de rester le plus rigoureux possible théoriquement.

Secondement, lorsque le domaine d'attraction de la queue de distribution est de Gumbel, la loi GPD se réduit à une loi exponentielle. Cette situation pose des problèmes spécifiques de tarification, qui seront abordés dans les chapitres suivants.

Troisième partie

L'approche par distorsion

Chapitre 6

Modélisation de la queue de distribution pour le domaine de Gumbel

Dans ce chapitre, le problème que pose le domaine de Gumbel lors de la tarification en réassurance est examiné. Comme pour toutes les parties suivantes, cette section se concentre uniquement sur la queue de distribution de la sévérité des sinistres de la cédante.

6.1 Problématique

6.1.1 Le domaine de Gumbel en réassurance

Le domaine d'attraction de Gumbel ($\gamma = 0$) inclut les distributions à queue fine, ce qui signifie que la probabilité d'occurrence d'un sinistre extrême est très faible par rapport au domaine de Fréchet. En théorie, lorsque la queue de distribution de la sévérité des sinistres de la cédante appartient au domaine d'attraction de Gumbel, cela implique une faible probabilité de survenance de sinistres extrêmes.

Cependant, l'analyse de l'historique des sinistres extrêmes dans le marché où les risques de la cédante sont exposés révèle clairement l'existence d'un risque, même si cela n'est pas totalement reflété dans ses données historiques. Un sinistre extrême peut toujours survenir à l'avenir, malgré cette probabilité théorique faible.

Ainsi, bien que cette queue de distribution soit dans le domaine d'attraction de Gumbel, il est crucial de tarifier le risque existant.

6.1.2 Exemple : du domaine de Gumbel au domaine de Fréchet

Pour démontrer l'existence d'un risque, il est possible de montrer qu'avec le temps, le domaine d'attraction peut évoluer, passant de Gumbel à Fréchet.

Exemple n°1 :

Prenons le traité examiné dans les chapitres n°3 et 5. Dans le chapitre n°5, il a été démontré que la queue de la distribution de la sévérité des sinistres appartenait au domaine de Fréchet. Cette analyse couvrirait un horizon de modélisation de 2013 à 2022.

Considérons maintenant un horizon de modélisation de 2014 à 2018, en supposant que ces données ont été transmises par la cédante pour souscrire ce traité en 2020. Le seuil des extrêmes devient alors $u_0 = 499387$ (cf. Annexe D [pages 104](#)).

Les graphiques ci-dessous montrent que le domaine d'attraction de la queue de distribution appartient alors au domaine de Gumbel. En effet, le QQ plot et le return level plot sont bien linéaires.

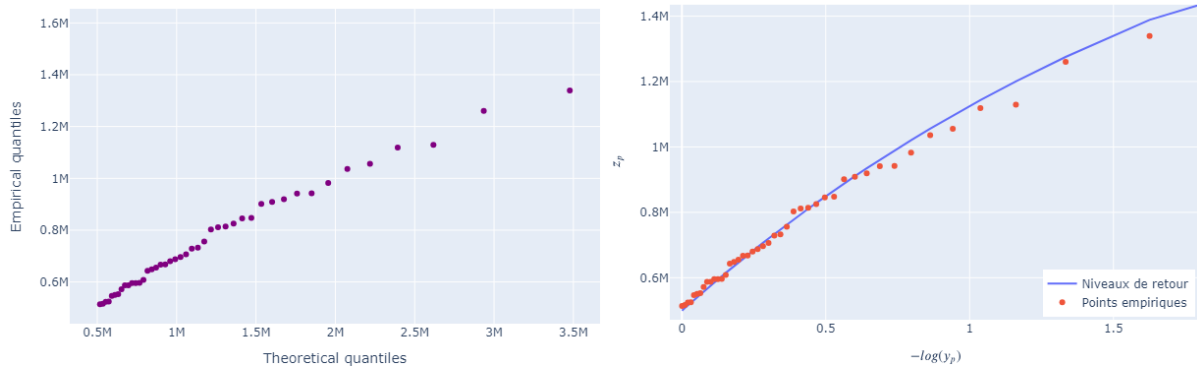


FIGURE 6.1 – QQ plot (gauche) et return level plot (droite)

De plus, le test du rapport de vraisemblance présenté dans le chapitre n°4 (4.4.3) donne une p-valeur d'environ 0,075. Bien que cette valeur soit proche du seuil conventionnel de 0,05, l'hypothèse nulle H_0 , qui affirme que le domaine d'attraction est de Gumbel, n'est pas rejetée. Par conséquent, les graphiques, combinés aux résultats du test, soutiennent l'idée que le domaine d'attraction de la queue de distribution est bien celui de Gumbel.

Ainsi, cet exemple montre que bien que la queue de distribution de la sévérité des sinistres de la cédante appartienne au domaine de Gumbel sur une période donnée, cela n'exclut pas que lorsqu'un échantillon plus large sera disponible dans les années à venir, le domaine d'attraction passe de Gumbel à Fréchet. Ce changement possible justifie donc le recours à la réassurance par la cédante, soulignant l'existence d'un risque.

Exemple n°2 :

Un exemple par simulation peut également illustrer ce phénomène.

Supposons un échantillon de 50 données simulées selon une loi GPD (1000, 250, 0.25) représentant des montants de sinistres. Cet échantillon, nommé "échantillon complet", se situe bien dans le domaine d'attraction de Fréchet avec un $\gamma > 0$.

Construisons ensuite un second échantillon, appelé "échantillon tronqué", qui correspond aux dix premières valeurs simulées du l'échantillon complet. Par conséquent, cet échantillon tronqué est également issu de la simulation de la GPD appartenant au domaine d'attraction de Fréchet.

Cependant, après avoir réalisé 200 simulations, une vingtaine de cas montrent que l'échantillon tronqué ne suit plus le domaine d'attraction de Fréchet, mais plutôt celui de Gumbel.

Le graphique des niveaux de retour présenté ci-dessous illustre l'un de ces cas, avec l'échantillon complet en bleu et l'échantillon tronqué en rouge.

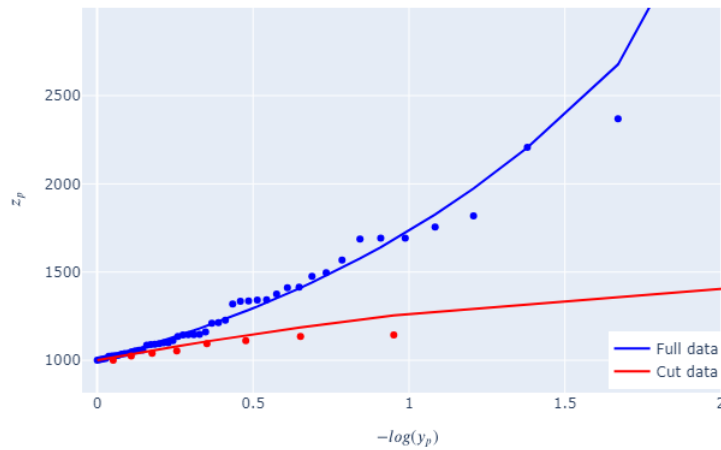


FIGURE 6.2 – Return level plot

La courbe du return level plot pour l'échantillon tronqué est linéaire, tandis que celle de l'échantillon complet présente une concavité.

De plus, le test du rapport de vraisemblance donne une p-valeur d'environ 0,87 ce qui ne permet pas de rejeter l'hypothèse H_0 . Ainsi, l'échantillon tronqué appartient au domaine d'attraction de Gumbel.

Ce deuxième exemple démontre donc aussi que le domaine d'attraction peut évoluer de Gumbel à Fréchet au fil du temps avec l'augmentation du nombre de données.

6.2 Cas pratique et théorique

6.2.1 Cas pratique : la tarification actuelle

Actuellement chez CCR Re, lorsque l'échantillon se situe dans le domaine d'attraction de Gumbel, plusieurs méthodes de tarification sont utilisées.

Modélisation avec une loi exponentielle ou GPD

Lorsque les données se trouvent dans le domaine d'attraction de Gumbel, caractérisé par des queues fines, une approche consiste à modéliser l'extrémité de la distribution à l'aide d'une loi exponentielle ou d'une loi GPD, à condition que suffisamment de données soient disponibles. Dans ce cadre, la GPD se comporte en une loi exponentielle. Cette méthode présente les avantages et les limites suivantes :

Avantages :

- **Bon ajustement** : Cette méthode s'ajuste bien aux sinistres observés, offrant ainsi une représentation précise de la sévérité des sinistres de la cédante.
- **Précision** : La loi exponentielle s'adapte bien aux queues fines, permettant une estimation précise de la distribution des sinistres jusqu'à un certain point.

Limites :

- **Sous-estimation des risques extrêmes** : Cette méthode ne prend pas pleinement en compte les risques dans les tranches supérieures de la distribution. La loi exponentielle décroît rapidement, ce qui peut conduire à une sous-estimation de la probabilité et du montant des sinistres les plus graves.
- **Ajustement arbitraire** : Pour pallier cette sous-estimation, un ajustement arbitraire est appliqué aux tranches élevées. Bien que cet ajustement soit basé sur

des considérations pratiques ou un jugement d'expert, il est déterminé sans théorie mathématique. Cette approche peut introduire de l'incertitude et diminuer la confiance dans les résultats.

- **Fiabilité** : La dépendance à un ajustement arbitraire rend cette méthode moins fiable car elle ne repose pas sur une base méthodologique solide.

Utilisation d'une distribution de Pareto

Une alternative consiste à utiliser une loi de Pareto pour modéliser la queue de la distribution. La loi de Pareto, appartenant au domaine de Fréchet, est caractérisée par des queues lourdes, ce qui signifie qu'elle attribue une probabilité plus élevée aux sinistres extrêmes. Cette méthode présente également des avantages et des limites.

Avantage :

- **Taux plus élevés dans les tranches hautes** : En utilisant une loi de Pareto, les taux de sinistres dans les tranches hautes sont plus élevés, ce qui permet de mieux prendre en compte le risque existant. La nature de la loi de Pareto signifie qu'elle anticipe une probabilité plus élevée de sinistres extrêmes par rapport à la loi exponentielle. Cependant, bien que cette caractéristique puisse sembler bénéfique pour une évaluation précise du risque, elle peut également nuire à la compétitivité du réassureur sur le marché, rendant ses offres moins attractives.

Limites :

- **Non-conformité au domaine d'attraction** : La loi de Pareto n'appartient pas au même domaine d'attraction que l'échantillon de données initial, qui est celui de Gumbel. Cela signifie que l'ajustement de la distribution des sinistres avec une loi de Pareto peut introduire des biais de modélisation significatifs.
- **Mauvaise modélisation de la sévérité** : Étant donné que la loi de Pareto ne correspond pas à la distribution observée des sinistres, elle risque de ne pas modéliser correctement la sévérité des sinistres, ce qui peut entraîner des estimations biaisées.
- **Incertitude sur la couverture des risques** : Bien que la loi de Pareto permette d'obtenir des taux plus élevés, il n'y a pas de justification solide pour affirmer que ces taux reflètent le risque. En d'autres termes, les taux calculés peuvent ne pas couvrir de manière précise et fiable le risque des sinistres extrêmes.

Pour conclure, les deux méthodes présentent des limites importantes. La première, utilisant une loi exponentielle, modélise bien la sévérité des sinistres dans le cas présent. Toutefois, elle sous-estime les risques dans les tranches élevées et nécessite un ajustement arbitraire. La deuxième méthode, utilisant une loi de Pareto, capture mieux les risques élevés mais s'ajuste mal aux données initiales, ce qui peut entraîner des erreurs de modélisation. Les taux obtenus ne garantissent pas une couverture adéquate des risques.

La suite de ce mémoire vise à améliorer la première méthode pour pallier ses limites. La méthode n°1, basée sur la loi exponentielle, a été retenue en raison de sa solidité théorique. Elle permet de maintenir une base rigoureuse tout en s'ajustant au mieux aux données spécifiques de la cédante. L'objectif est ainsi d'affiner cette approche pour améliorer la modélisation des risques extrêmes tout en conservant une base théorique solide.

6.2.2 Cas théorique : la tarification à partir d'une loi de marché

Pour améliorer la méthode n°1, il est nécessaire de développer une approche plus avancée. L'objectif est de continuer à modéliser la queue de distribution avec une loi exponentielle lorsque celle-ci se trouve dans le domaine de Gumbel. Cette approche est priorisée car elle permet de maintenir une modélisation cohérente avec les données spécifiques de la cédante, offrant ainsi une représentation fidèle des risques. En utilisant la loi exponentielle, une base théorique solide est conservée tout en optimisant l'ajustement aux données observées, ce qui est essentiel pour garantir une tarification précise et rigoureuse.

Introduction de la distorsion

Pour répondre à la problématique de la sous-estimation des risques extrêmes avec la loi exponentielle, une approche fondée sur les fonctions de distorsion est mise en place. Les fonctions de distorsion sont des outils statistiques appliqués à des lois de probabilité afin de les transformer. Le choix du paramètre de distorsion et de la fonction de distorsion en elle-même détermine l'ampleur et la nature de cette transformation.

L'objectif est ainsi d'appliquer une fonction de distorsion à la loi exponentielle afin de la transformer et d'accroître la probabilité d'occurrence des sinistres extrêmes. En modifiant la loi de cette manière, il deviendrait alors possible de mieux capturer les risques rares mais significatifs qui sont sous-estimés par la loi exponentielle seule.

La théorie sur les fonctions de distorsion et leur application seront abordés dans le chapitre suivant.

Cible : la loi de marché

La distorsion de la loi exponentielle est essentielle, mais il est tout aussi important de déterminer la direction vers laquelle cette distorsion doit être effectuée. C'est ici qu'intervient la loi de marché, qui joue le rôle de référence essentielle pour guider cette transformation.

Une loi de marché est une distribution statistique élaborée pour représenter avec précision la sinistralité dans un marché spécifique, , que ce soit pour un pays, une branche de réassurance ou un type de sinistre particulier. Par exemple, dans le cadre du traité décrit dans le chapitre 3, la loi de marché pourrait correspondre à la sinistralité des sinistres incendie en France. Cette loi est construite en se basant sur des données historiques et actuelles du marché, afin de refléter au mieux les caractéristiques spécifiques du risque dans le contexte géographique et sectoriel donné. Une loi de marché bien établie constitue une base solide car elle intègre à la fois les tendances historiques et les particularités du marché.

L'objectif est donc de distordre la loi exponentielle de manière à ce qu'elle converge vers la loi de marché correspondante. Cette distorsion vise à aligner la modélisation des risques extrêmes avec une représentation réaliste et pertinente de la sinistralité telle qu'observée sur le marché. Cela permet ainsi d'assurer une évaluation plus précise et fiable des risques.

Ainsi, la loi de marché servira de point de référence pour déterminer la distorsion maximale que peut subir la loi exponentielle, assurant que cette transformation reste pertinente et alignée sur les caractéristiques du marché.

Remarque :

La tarification basée sur le marché est souvent employée lorsque les données fournies par la cédante sont insuffisantes ou de qualité inadéquate pour permettre une tarification

précise basée sur l'expérience. Si la sinistralité d'une cédante se trouve dans le domaine de Gumbel, cela indique probablement un manque d'informations suffisantes pour une tarification à l'expérience. En effet, la demande de réassurance témoigne de l'existence d'un risque. Ainsi, se rapprocher de la loi de marché peut être crucial lorsque les données disponibles sont insuffisantes.

Degré : facteur de crédibilité

Après avoir déterminé la direction de la distorsion en se basant sur la loi de marché, il sera essentiel de définir le degré de cette distorsion. Pour ce faire, un facteur de crédibilité sera construit afin de déterminer le pourcentage optimal de distorsion entre la loi exponentielle ajustée et la loi de marché.

Ce facteur de crédibilité permettra de mesurer la confiance accordée aux données de la cédante en tant que bonne représentation de sa sinistralité future. Autrement dit, il évaluera la fiabilité des données observées de la cédante pour prédire ses sinistres futurs en comparaison avec les standards du marché.

La méthode permettant de construire le facteur de crédibilité sera présentée dans les chapitres suivants.

Chapitre 7

Les fonctions de distorsion

Dans ce chapitre, les fonctions de distorsion sont introduites théoriquement.

7.1 Présentation générale des fonctions de distorsion

7.1.1 Définition

Définition : Une fonction de distorsion est une fonction non décroissante $g : [0,1] \rightarrow [0,1]$ telle que $g(0)=0$ et $g(1)=1$. Ainsi, F_g est la fonction distordue de F si $F_g(x) = g(F(x))$.

En termes simples, une fonction de distorsion permet de déformer la distribution d'une loi de probabilité dans le but de mieux représenter le risque perçu par les assureurs et réassureurs. Une application typique est la mesure du risque de Wang, qui modifie la distribution pour capturer l'aversion au risque de manière plus réaliste.

7.1.2 Rôle dans la tarification

Dans le contexte de la tarification, les fonctions de distorsion peuvent jouer un rôle important en modifiant la distribution des sinistres. Cette modification permet d'ajuster la prime pure pour mieux refléter les risques extrêmes.

Comme présenté dans le chapitre n°2, la prime de réassurance est obtenue en calculant le produit de l'espérance de la loi de sévérité, avec celle de la loi de la fréquence des sinistres (2.4.1). Dans notre cas, les fonctions de distorsion sont appliquées à la queue de distribution de la loi de sévérité des sinistres.

Cette application entraîne une modification de l'espérance de la sévérité, aboutissant à l'inégalité suivante lorsque la fonction de distorsion g est concave :

$$\mathbb{E}[X] = \int_0^\infty S(x)dx \leq \int_0^\infty g(S(x))dx = \mathbb{E}[X^{distortion}]$$

Avec :

- X : la loi de sévérité.
- $S(x)$: la fonction de survie de X .
- $g(x)$: la fonction de distorsion.

Ainsi, la distorsion entraîne une augmentation de l'espérance de la sévérité des sinistres, accentuant ainsi l'importance accordée aux risques extrêmes.

7.2 Exemples de fonction de distorsion

Il existe plusieurs types de fonctions de distorsion couramment utilisées. Dans ce mémoire, deux types de fonctions seront présentés.

7.2.1 Fonction de distorsion de Wang

La fonction de distorsion de Wang est définie par :

$$g(x) = \Phi(\Phi^{-1}(x) + \lambda)$$

où Φ est la fonction de répartition de la loi normale centrée réduite et λ est un paramètre de distorsion.

Le graphique ci-dessous représente la fonction de répartition d'une loi exponentielle initiale ainsi que la même fonction distordue par la fonction de distorsion de Wang avec un $\lambda = 0.6$.

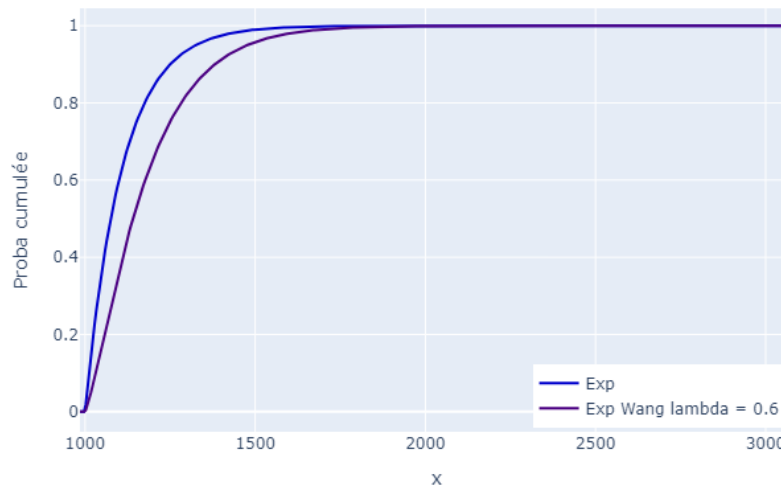


FIGURE 7.1 – Fonction de distorsion de Wang sur l'exponentielle

Le graphique illustre un déplacement de la fonction distordue vers la droite. Cela signifie que la probabilité d'occurrence de sinistres extrêmes augmente.

Remarque : Dans le cas de la fonction de distorsion de Wang, plus le paramètre λ augmente, plus la distorsion de la fonction est prononcée. Cela permet d'ajuster de manière flexible l'importance donnée aux sinistres extrêmes en fonction du niveau de risque perçu.

7.2.2 Fonction de distorsion de hasard proportionnel (PH)

La fonction de distorsion de hasard proportionnel, notée PH (*proportional hazard* en anglais), est définie par :

$$g(x) = x^r$$

où $0 < r < 1$ est le paramètre de distorsion. Cette fonction de distorsion modifie la probabilité en élevant la probabilité initiale à la puissance r , ce qui influe sur la pondération des sinistres extrêmes.

Le graphique ci-dessous représente la fonction de répartition d'une loi exponentielle basique ainsi que la même fonction distordue par la fonction de distorsion de PH avec un $r = 0.6$.

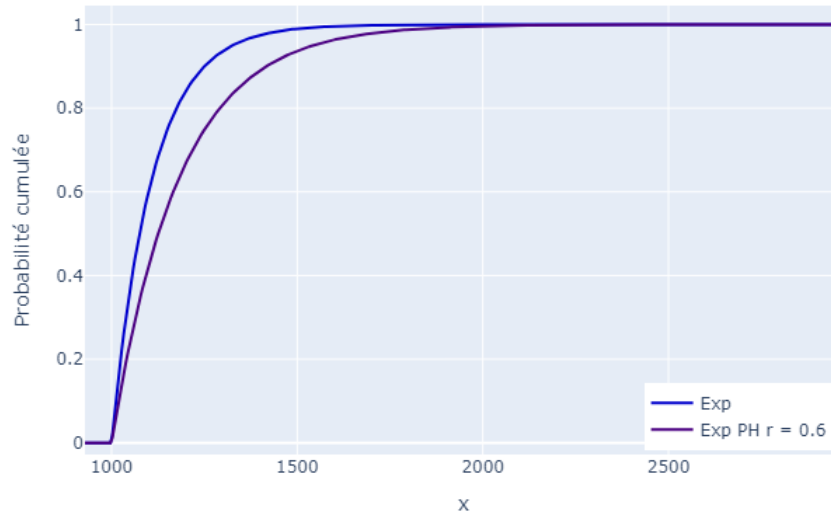


FIGURE 7.2 – PH transform sur l'exponentielle

De manière similaire à la fonction de distorsion de Wang, ce graphique illustre un déplacement de la fonction distordue vers la droite.

Remarque : Dans le cas de la fonction de distorsion de PH, plus le paramètre r se rapproche de 0, plus la distorsion de la fonction est prononcée. Tout comme la Wang, cela permet d'ajuster de manière flexible l'importance donnée aux sinistres extrêmes en fonction du niveau de risque perçu.

Exemple : Cas de la loi exponentielle

Lorsque la fonction à distordre est une loi exponentielle, la fonction de distorsion PH donne la fonction de survie distordue suivante :

$$S_g(x) = g(S(x)) = S(x)^r = e^{-\lambda x r}$$

Où $S(x)$ est la fonction de survie de la loi exponentielle de paramètre λ .

De même, la fonction de répartition distordue est la suivante :

$$F_g(x) = 1 - S_g(x) = 1 - e^{-\lambda r x}$$

Où $F(x)$ est la fonction de répartition de la loi exponentielle de paramètre λ .

La densité distordue est alors :

$$f_g(x) = (F_g(x))' = \lambda r e^{-\lambda r x}$$

Ainsi, dans le cas de la loi exponentielle, la fonction distordue suit également une loi exponentielle de paramètre λr . Cette propriété est particulièrement utile car elle permet de conserver les caractéristiques de la loi originale tout en ajustant la distribution pour mieux refléter les risques extrêmes.

Dans la suite de ce mémoire, l'étude sera focalisée sur la fonction de distorsion PH, en

raison de sa simplicité et de sa capacité à conserver la forme exponentielle de la distribution initiale.

7.3 Ecart minimal entre deux distributions

Comme évoqué précédemment, les fonctions de distorsion permettent de modifier la fonction originale selon les besoins. Toutefois, il est nécessaire non seulement de définir la direction de cette distorsion, mais aussi de déterminer la distorsion maximale que la fonction exponentielle peut atteindre, comme expliqué dans le chapitre précédent (6.2.2).

La tâche suivante consiste ainsi à identifier le paramètre de distorsion maximale (noté r_m), qui représente la plus grande déformation que la loi exponentielle peut subir tout en restant cohérente avec la loi de marché. Autrement dit, il s'agit de trouver le paramètre de distorsion qui minimise l'écart entre la distribution modélisée et la loi de marché, garantissant une convergence maximale. Pour y parvenir, deux méthodes sont mises en œuvre.

7.3.1 Divergence de Kullback-Leibler

La première méthode consiste à calculer la divergence de Kullback-Leibler. Cette mesure de probabilité quantifie l'écart entre deux distributions de probabilités. La formule de la divergence de Kullback-Leibler est la suivante :

$$D_{KL}(P\|Q) = \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx$$

Où P et Q sont deux distributions continues avec $p(x)$ et $q(x)$ les densités de probabilité respectives.

L'objectif est donc de minimiser cette divergence afin que la distorsion appliquée à la loi initiale permette une convergence maximale vers la loi de marché.

Exemple :

Considérons une loi exponentielle avec un paramètre $\lambda = 0.009$, ainsi qu'une loi de marché modélisée par une loi GPD avec les paramètres (1000, 250, 0.25). La fonction exponentielle est modifiée à l'aide de la fonction de distorsion PH. Le graphique ci-dessous présente la fonction distordue maximale obtenue en appliquant cette méthode.

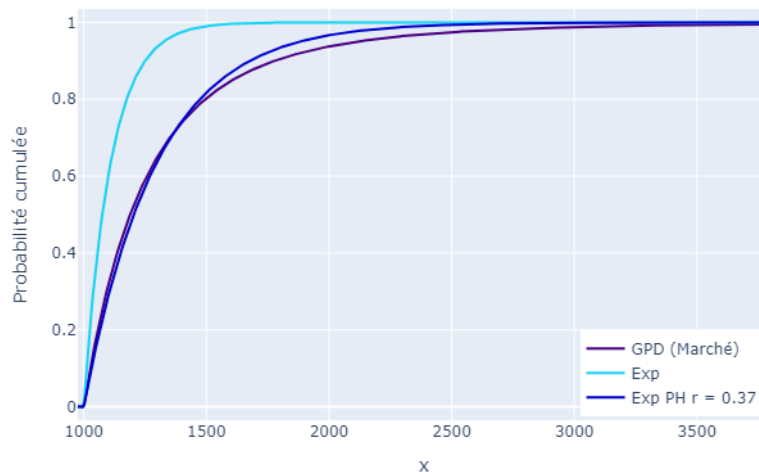


FIGURE 7.3 – Fonction de répartition distordue maximale (KL)

Dans cet exemple, le paramètre de distorsion maximale pour la fonction PH est $r_m = 0,37$. La fonction de répartition distordue au maximum est ainsi très proche de celle de la loi de marché.

7.3.2 Distance quadratique L^2

La deuxième méthode consiste à calculer la distance quadratique. Cette mesure de probabilité calcule la distance entre deux distributions de probabilités. La formule de la distance quadratique est la suivante :

$$D_{L^2}(p, q) = \int_{-\infty}^{\infty} (p(x) - q(x))^2 dx$$

Où $p(x)$ et $q(x)$ sont les densités de probabilité des deux distributions continues P et Q respectives.

Exemple :

Reprenons le même exemple que le précédent. Le graphique ci-dessous illustre la fonction distordue maximale obtenue en utilisant la méthode de la distance quadratique L^2 .

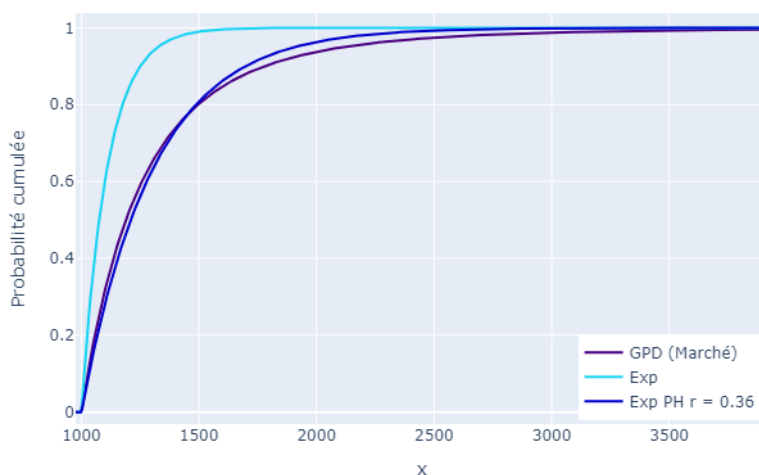


FIGURE 7.4 – Fonction de répartition distordue maximale (L^2)

Dans cet exemple, le paramètre de distorsion maximale pour la fonction PH est $r_m = 0.36$. La fonction de répartition distordue obtenue avec cette méthode est également très proche de celle de la loi de marché, garantissant ainsi une distorsion maximale en fonction de la distance quadratique.

Remarque :

Les deux méthodes aboutissent à des résultats similaires, comme le montrent les exemples précédents. Néanmoins, pour la suite de ce mémoire, seule la méthode de Kullback-Leibler sera privilégiée. Cette approche offre une mesure plus fine et plus détaillée des écarts entre les distributions. En effet, la divergence de Kullback-Leibler est particulièrement sensible aux variations des probabilités d'événements rares, un aspect crucial pour la modélisation des sinistres extrêmes. De plus, son caractère asymétrique permet de mieux capturer les écarts dans les queues de distribution, assurant ainsi un ajustement plus rigoureux de la loi exponentielle à celle du marché.

Chapitre 8

Application : l'approche de la distorsion pour la tarification

Après avoir exposé la théorie sur les fonctions de distorsion, il est désormais possible d'appliquer l'approche décrite dans la section [6.2.2](#).

8.1 Approximation des données

Pour illustrer l'approche mise en place dans ce mémoire, l'analyse sera effectuée sur le cas suivant :

- **La loi de marché :**

Déterminer une loi de marché représentative est un processus complexe et minutieux. Étant donné que la construction d'une telle loi n'est pas l'objet de ce mémoire, une loi statistique préexistante est utilisée à la place. La loi de marché choisie est une loi GPD avec les paramètres suivants : (1000, 250, 0.25). Cette loi est employée dans l'exemple n°2 de la section [6.1.2](#) pour illustrer les caractéristiques du domaine d'attraction, qui varient en fonction du nombre de données dans l'échantillon.

- **Données de la queue de distribution de la cédante :**

Les sinistres observés dans la queue de distribution de la cédante doivent appartenir au domaine d'attraction de Gumbel pour que l'application soit valide. Pour ce faire, l'échantillon tronqué de l'exemple n°2 de la section [6.1.2](#) est utilisé comme représentation des sinistres de la queue de distribution de la cédante. Cet échantillon correspond aux dix premiers tirages obtenus lors de la simulation des 50 sinistres selon la loi GPD (1000, 250, 0.25) (loi de marché retenue). Les sinistres retenus sont ensuite ajustés à l'aide d'une loi exponentielle avec un paramètre $\lambda = 0.009$, calibré sur les données de l'échantillon pour garantir une correspondance théorique précise avec le domaine d'attraction de Gumbel.

La figure suivante montre la fonction de répartition de la loi exponentielle ajustée aux sinistres de la cédante comparée à la loi de marché :

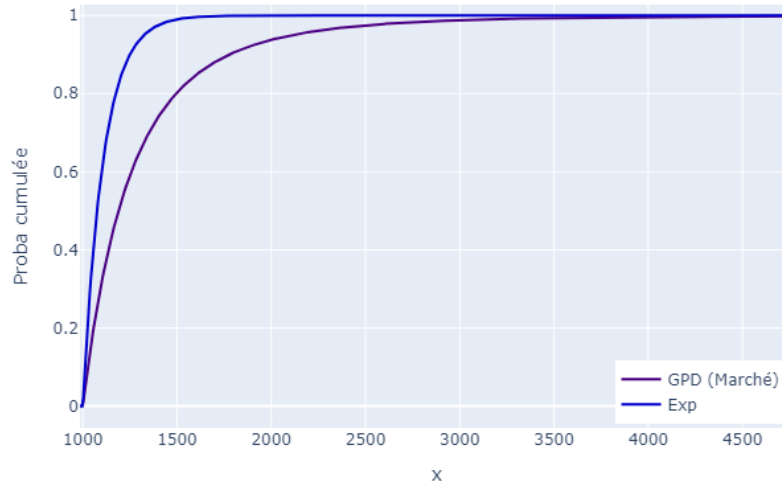


FIGURE 8.1 – Fonction de répartition cédante et marché

Il apparaît que la fonction de répartition de la loi exponentielle indique une probabilité d'occurrence des sinistres extrêmes bien inférieure à celle de la loi de marché. Par conséquent, il est nécessaire de distordre plus ou moins la loi exponentielle avec la fonction de distorsion PH, selon le facteur de crédibilité calculé, vers la loi de marché.

8.2 La distorsion maximale

Après la présentation des données, l'étape suivante consiste à déterminer le paramètre de distorsion maximale (r_m), qui correspond à la déformation maximale que la loi de la cédante peut subir tout en restant compatible avec la loi de marché.

Pour ce faire, la divergence de Kullback-Leibler est calculée afin d'identifier le plus faible écart entre ces deux distributions. Le graphique ci-dessous illustre la fonction distordue maximale obtenue par cette méthode.

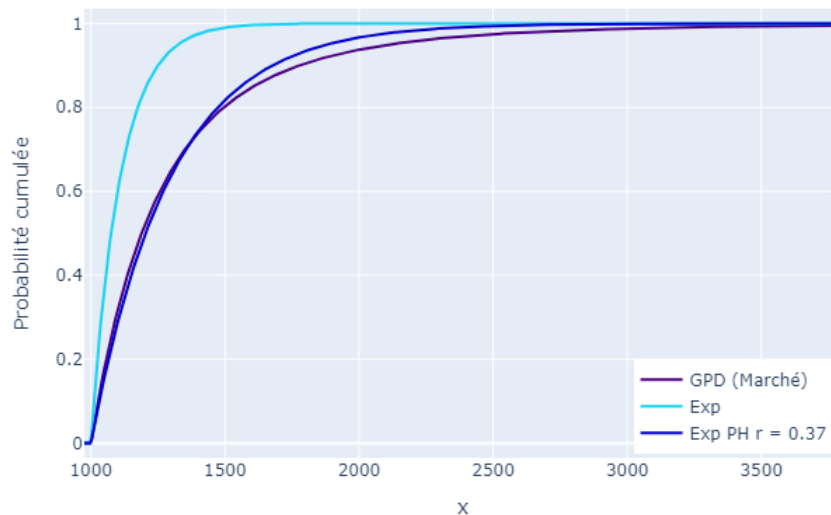


FIGURE 8.2 – Fonction de répartition distordue maximale (KL)

Le paramètre de distorsion maximale pour la fonction exponentielle de la cédante est $r_m = 0,37$ dans ce cas. Ainsi, la fonction de répartition distordue est la plus proche possible de la fonction de répartition de la loi de marché.

8.3 La distorsion idéale : le facteur de crédibilité

La notion du paramètre de distorsion maximale, calculé à l'aide du critère KL, a été introduite dans la section précédente, $r_m = 0,37$. Il est ensuite nécessaire de déterminer le paramètre de distorsion idéal (noté r_i) en fonction des informations fournies par la cédante. Ce paramètre doit respecter l'inégalité suivante :

$$r_m \leq r_i \leq 1$$

L'objectif est d'évaluer la fiabilité des données de la cédante comme un reflet fidèle de la réalité. Une confiance moindre dans les données entraînera une distorsion accrue de la loi exponentielle (loi cédante) vers la loi GPD (loi de marché). Un facteur de crédibilité sera ainsi construit pour quantifier cette confiance.

8.3.1 Interprétation du facteur

Le facteur de crédibilité de la cédante évalue la confiance accordée aux données fournies, en indiquant dans quelle mesure celles-ci sont jugées représentatives de sa sinistralité future. Un facteur de crédibilité proche de 1 reflète une forte confiance dans les données de la cédante, caractérisées par l'observation de peu, voire aucun sinistre extrême. Cela suggère que, malgré l'absence ou la rareté de sinistralité extrême, les données de la cédante fournissent une estimation fiable des sinistres futurs. Ce scénario est pertinent si l'on considère que le marché dans lequel ces risques évoluent présente également un nombre limité de sinistres extrêmes, rendant ainsi les données de la cédante plus fidèles. Dans ce cas, la distorsion appliquée à la loi exponentielle sera minimale, les données étant considérées suffisamment fiables pour ne pas nécessiter d'ajustement majeur.

En revanche, un facteur de crédibilité proche de 0 traduit une faible confiance dans les données, signifiant qu'elles ne sont pas perçues comme une représentation fiable de la sinistralité future. Ce scénario est envisageable lorsque le marché associé aux risques présente une forte sinistralité extrême, contrairement à ce que suggèrent les données de la cédante. Une distorsion plus importante sera alors appliquée, ajustant la loi exponentielle pour la rapprocher de la loi de marché, assurant ainsi une modélisation plus adéquate des risques extrêmes.

Le schéma ci-dessous illustre ce principe.

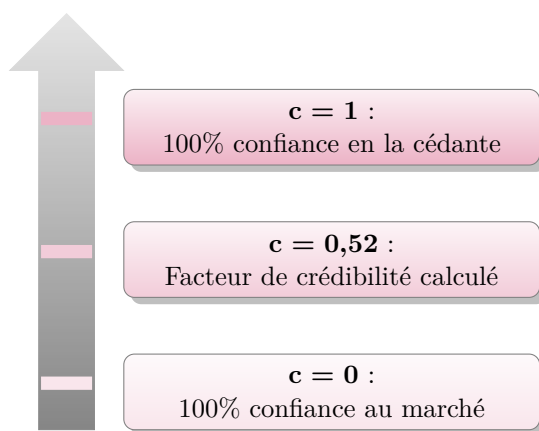


FIGURE 8.3 – Facteur de crédibilité

8.3.2 Construction du facteur

Test réalisé

Pour déterminer le facteur de crédibilité (noté c), l'objectif est d'estimer la probabilité que les données de la queue de distribution de la cédante appartiennent au domaine d'attraction de Gumbel dans le marché considéré. Si cette probabilité est faible, cela indiquerait une faible confiance dans les données de la cédante comme étant un bon reflet de sa sinistralité future. Inversement, une probabilité élevée traduirait une plus grande confiance dans ces données.

Cette probabilité sera obtenue à l'aide d'une courbe de densité représentant les valeurs possibles du paramètre γ dans ce marché. Le paramètre γ caractérise le domaine d'attraction d'une distribution statistique. Par conséquent, cette courbe permet d'évaluer la probabilité d'observer différents domaines d'attraction dans le marché, et notamment celui de Gumbel.

Données d'application

La construction de cette courbe nécessite la simulation de 100 échantillons, chacun suivant la loi de marché. Chaque échantillon est composé de 10 observations correspondant à des sinistres de la queue de distribution d'une cédante. Ces échantillons représentent ainsi la sinistralité de la queue de distribution de 100 cédantes différentes, toutes cohérentes avec la loi de marché.

Pour chaque échantillon, le paramètre γ est estimé à l'aide de la méthode du maximum de vraisemblance, permettant de déterminer le domaine d'attraction de la queue de distribution de chaque cédante. Une fois les 100 estimations de γ obtenues, la courbe de densité des gammas est construite, illustrant la répartition de ces paramètres dans le marché.

Résultat

La figure ci-dessous présente l'histogramme normalisé des probabilités des valeurs de gamma estimées, accompagné de la courbe de densité normalisée. Ce graphique permet d'évaluer la probabilité que les données de la cédante appartiennent au domaine d'attraction de Gumbel. Plus précisément, il illustre la distribution des valeurs de gamma obtenues dans le cadre de notre application.

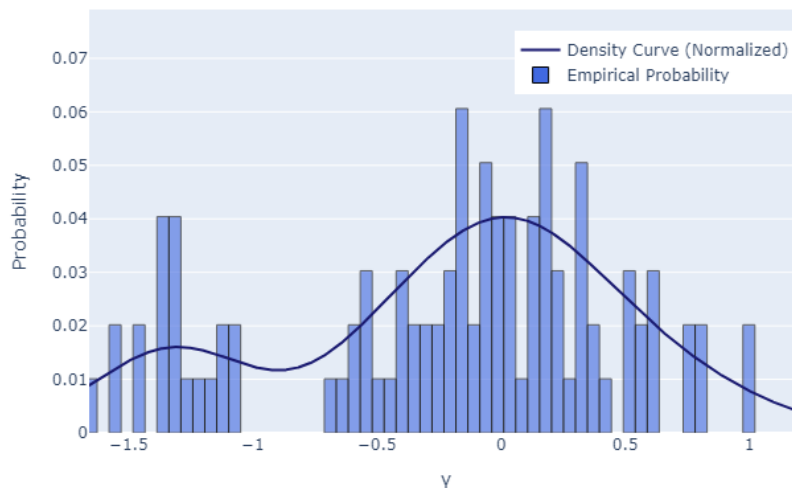


FIGURE 8.4 – Probabilité de γ

Le graphique montre que le paramètre γ , basé sur la loi de marché, peut présenter des fluctuations significatives. Bien que la loi de marché soit associée au domaine de Fréchet, certains échantillons de cédantes peuvent néanmoins se situer dans le domaine de Gumbel. La densité de probabilité pour ces valeurs n'est pas négligeable. Par exemple, pour un γ égal à -0,5, la probabilité estimée est de 0,02.

Hypothèse posée

Le facteur de crédibilité de la cédante se calcule ainsi en déterminant la probabilité que le paramètre γ soit inférieur ou égal au $\hat{\gamma}_{\text{cédante}}$ estimé de l'échantillon spécifique de la cédante. Mathématiquement, cette relation s'exprime par :

$$c = \mathbb{P}[\gamma \leq \hat{\gamma}_{\text{cédante}}]$$

Où c représente le facteur de crédibilité, et $\hat{\gamma}_{\text{cédante}}$ est la valeur estimée du paramètre γ pour l'échantillon de la cédante.

Dans le cadre de cette étude, le facteur de crédibilité de la cédante est le suivant :

$$c = \mathbb{P}[\gamma \leq -0,042] = 0,52$$

Ce résultat traduit une confiance modérée dans les données de la cédante par rapport aux caractéristiques du marché considéré dans cette application.

Remarque :

Pour rappel, la modélisation à l'aide d'une loi exponentielle distordue avec un facteur de crédibilité est effectuée uniquement lorsque l'échantillon appartient ou est très proche du domaine de Gumbel. Cette information est confirmée par le test du rapport de vraisemblance (présenté dans le chapitre 4.4.3). Par conséquent, si le domaine d'attraction de la queue de distribution de la cédante ne se trouve pas dans ce cas de figure, il n'est pas nécessaire de calculer un facteur de crédibilité ni de modéliser avec une loi exponentielle distordue. La modélisation par la GPD est alors suffisante.(cf. chapitre 5)

8.3.3 Distorsion selon le facteur de crédibilité

Une fois le facteur de crédibilité de la cédante déterminé, il est nécessaire de le relier au paramètre de distorsion approprié. Un facteur de crédibilité de $c = 1$, indiquant une confiance totale dans les données de la cédante, correspond à un paramètre de distorsion idéal $r_i = 1$. Dans ce cas, la fonction de distorsion g se réduit à la fonction identité, ce qui signifie qu'aucune distorsion n'est appliquée au modèle. Par conséquent, la distribution des sinistres est directement représentée par les données de la cédante, sans modification.

En revanche, lorsque $c = 0$, signifiant une confiance totale dans le marché, le paramètre de distorsion ne sera pas $r_i = 0$, mais plutôt $r_i = r_m$. En effet, si $r_i = 0$, la loi exponentielle distordue ne s'approche plus du tout de la loi de marché. Ainsi, $r_i = r_m$ représente la distorsion maximale possible vers la loi de marché. Le paramètre de distorsion idéal r_i est donc calculé selon la formule suivante :

$$r_i = r_m + c * (1 - r_m)$$

Les trois graphiques ci-dessous illustrent la fonction de répartition de la loi exponentielle distordue en fonction du paramètre de distorsion r_i , calculé à partir du facteur de crédibilité c . Ces graphiques comparent la loi exponentielle distordue à la fonction de répartition de la loi de marché.

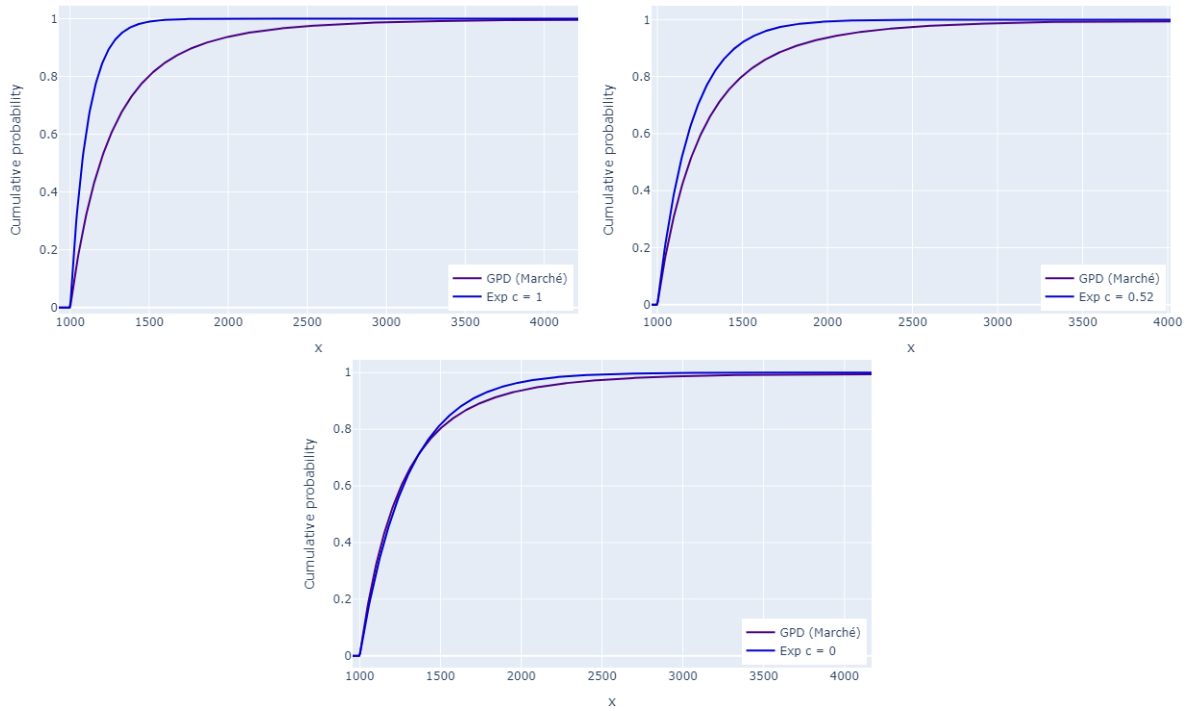


FIGURE 8.5 – Fonction de répartition de l'exponentielle selon le facteur de crédibilité

- Le graphique en haut à gauche présente la situation où le facteur de crédibilité est $c = 1$, ce qui indique une totale confiance en la cédante. Dans ce cas, la loi exponentielle de la cédante n'est pas distordue, correspondant à $r = 1$.
- Le graphique en haut à droite représente le facteur de crédibilité calculé dans notre application, $c = 0,52$. Ici, la fonction de répartition de la cédante est ainsi distordue avec un paramètre $r_i = 0.70$, rapprochant ainsi la fonction de répartition de la cédante de celle du marché.
- Le graphique en bas au milieu correspond à un facteur de crédibilité de $c = 0$, ce qui indique une grande confiance dans la loi de marché et une très faible confiance dans les données de la cédante. Dans ce cas, le paramètre de distorsion r est égal au paramètre de distorsion maximale déterminé par le critère KL ($r_m = 0,37$). La fonction de répartition de la cédante distordue est donc très proche de celle de la loi de marché.

Remarque :

Même avec un facteur de crédibilité de $c = 0$, signifiant une totale confiance dans la loi de marché, la loi de la cédante est distordue au maximum pour inclure légèrement les données de la cédante. Autrement, la tarification se baserait uniquement sur la loi de marché, rendant chaque cotation identique sans tenir compte des spécificités de la cédante. L'objectif de la distorsion est donc d'intégrer, même de manière minimale, les données de la cédante dans l'évaluation.

8.4 Résultats

Grâce au facteur de crédibilité, le paramètre de distorsion idéal r_i est déterminé, permettant ainsi de moduler la distorsion de la loi exponentielle de la cédante de manière non arbitraire.

Pour illustrer l'impact de cette distorsion sur une cotation, une structure de traité fictive a été créée. La structure est la suivante :

T3	2 000 XS 3 000
T2	1 000 XS 2 000
T1	1 000 XS 1 000

FIGURE 8.6 – Structure fictive du traité

L'objectif est de comparer les estimations de la prime pure pour chaque tranche en fonction de la loi choisie. Cette comparaison est effectuée en utilisant les quatre lois et facteurs de crédibilité suivants :

- la loi exponentielle de la cédante ($c = 1$),
- la loi exponentielle distordue ($c = 0.52$),
- la loi exponentielle distordue au maximum ($c = 0$),
- la loi de marché (GPD).

Le graphique ci-dessous présente les boxplots des primes pures possibles pour la tranche T1 en fonction de la loi sélectionnée.

Précisions :

Chaque point sur les boxplots représente une prime pure obtenue par simulation. Ces boxplots permettent de visualiser la moyenne et la volatilité des primes pures pour chacune des lois considérées. Ils permettent également d'observer la prime pure moyenne pour chaque loi.

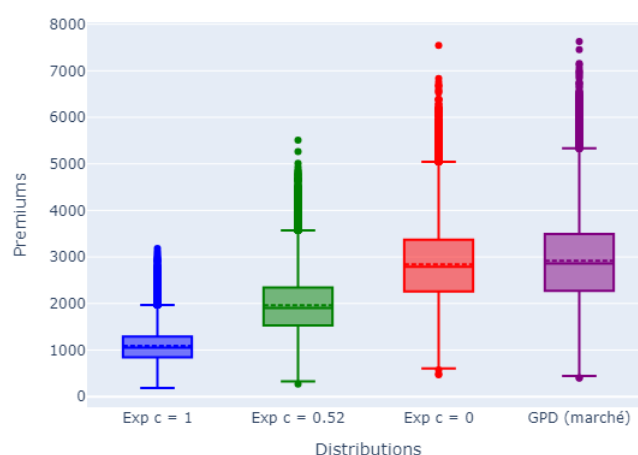


FIGURE 8.7 – Boxplot des primes de la T1 selon la loi

Le graphique des boxplots ci-dessus révèle les observations suivantes :

- **Effet de la distorsion sur la prime pure moyenne :** Plus la loi est distordue, plus la prime pure moyenne augmente. Par exemple, la prime pure moyenne pour la loi originale de la cédante (en bleu) est d'environ 1 086 €, tandis que pour la loi distordue avec le facteur de crédibilité calculé (en vert), elle est d'environ 1 964 €. Ce résultat est rassurant, car il confirme que la distorsion produit une prime plus

élevée, comme attendu.

- **Consistance des primes pures moyennes :** La prime pure moyenne sous la loi distordue se situe bien entre celle de la loi originale de la cédante (en bleu) et celle de la loi du marché (en violet). Ce constat est encourageant, car il assure la cohérence des résultats.
- **Variation de l'étendue des primes pures :** L'étendue des primes pures augmente avec la distorsion de la loi. Par exemple, la loi exponentielle distordue au maximum (en rouge) montre une étendue des primes pures très proche de celle de la loi du marché (en violet).
- **Proximité des primes pures maximales avec la loi du marché :** La prime pure moyenne pour la loi distordue au maximum ($c=0$, en rouge) est de 2 843 €, ce qui est proche de celle de la loi du marché (en violet), qui est d'environ 2 859 €. Cela indique que le critère KL est efficace pour déterminer le paramètre de distorsion maximale, rapprochant ainsi la loi exponentielle de la loi du marché.

En conclusion, la distorsion de la loi exponentielle permet d'éviter la sous-estimation du risque dans les tranches élevées en augmentant les primes pures, tout en maintenant une modélisation fidèle aux données de la cédante.

8.5 Limites

L'approche mise en place présente néanmoins certaines limites.

8.5.1 L'écart des primes dans les tranches élevées

Premièrement, l'écart entre les primes possibles dans les tranches les plus élevées tend à s'intensifier. En effet, la différence entre les domaines d'attraction devient de plus en plus marquée lorsque les queues de distribution sont prises en compte, ce qui est particulièrement le cas dans les tranches élevées. Par conséquent, malgré l'application de la distorsion à la loi exponentielle, la prime obtenue s'éloigne de plus en plus de celle issue de la loi de marché dans ces tranches élevées.

Pour illustrer cette limite, le graphique des boxplots de la tranche T3 est présenté ci-dessous.

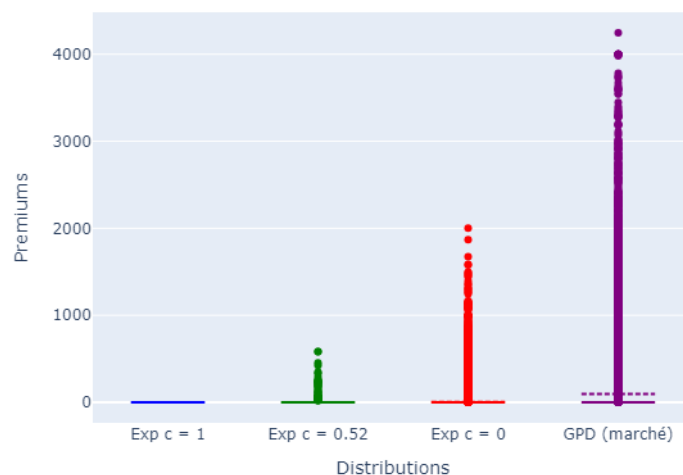


FIGURE 8.8 – Boxplot des primes de la T3 selon la loi

Dans cette tranche, les primes pures moyennes restent relativement faibles pour toutes

les lois concernées. Toutefois, l'écart dans l'étendue des primes pures possibles s'accroît en fonction de la loi choisie. Bien que la loi exponentielle distordue au maximum (en rouge) se rapproche de la loi de marché (en violet), la différence dans le domaine d'attraction entre ces deux lois entraîne un écart notable. En effet, la prime pure maximale obtenue avec la loi exponentielle distordue au maximum est d'environ 2 000 €, alors qu'elle atteint près de 4 000 € avec la loi de marché. Cela montre que la loi exponentielle distordue présente une étendue plus limitée dans les cas extrêmes.

Cette limite peut toutefois être considérée comme mineure, car elle n'apparaît que lorsqu'on examine la queue de distribution des deux lois ayant des domaines d'attraction différents. Cependant, l'approche sur la distorsion se concentre uniquement sur la queue de distribution de la sévérité. Par conséquent, ce cas se manifeste uniquement dans la queue de distribution de la loi exponentielle distordue, qui fait elle-même partie de la queue de distribution de la sévérité totale, représentant ainsi une portion plus infime.

8.5.2 La loi de marché

Secondement, l'approche proposée repose sur la distorsion de la loi exponentielle vers une loi de marché. Toutefois, la loi de marché utilisée dans cette application est une simulation basée sur une hypothèse simplificatrice. Autrement dit, elle ne correspond pas à un marché réel ou spécifique. Par conséquent, pour que cette approche soit effectivement mise en œuvre de manière pertinente, il serait nécessaire de construire une loi de marché reflétant avec précision la sinistralité d'un marché particulier.

La construction d'une telle loi de marché requiert un travail approfondi d'analyse des données historiques, actuelles et des spécificités du secteur concerné. Cette étape est essentielle, car une loi de marché mal représentée entraînerait des modélisations et des distorsions inexactes. Il est impératif de garantir que les hypothèses sous-jacentes à la loi de marché soient justes, car des erreurs dans cette représentation pourraient mener à des conclusions trompeuses. Dans ce contexte, il est pertinent de se demander si le marché reflète effectivement la réalité des sinistres. Toutefois, la construction d'une telle loi n'étant pas l'objet principal de ce mémoire, une hypothèse simplificatrice a été adoptée à des fins illustratives, permettant ainsi de se concentrer sur la méthode de distorsion elle-même.

Conclusion

L'enjeu de cette étude a été de solidifier la méthodologie appliquée lors de la tarification des traités XS utilisant le modèle collectif. L'objectif principal était d'affiner la modélisation de la sévérité des sinistres, afin de produire des résultats plus fiables et explicables, en particulier pour les tranches élevées.

L'étude a mis en lumière, grâce à la théorie des valeurs extrêmes, une décomposition efficace de la modélisation de la sévérité des sinistres en deux parties distinctes : le corps et la queue de distribution. Ces deux sections sont séparées par un seuil extrême, noté u . Le corps de la distribution a été modélisé empiriquement à partir des données de la cédante, tandis que la queue de distribution a été représentée par une loi GPD. Cette approche a permis d'améliorer la rigueur de la modélisation des sinistres extrêmes, garantissant ainsi des résultats plus justifiables et fiables pour les tranches élevées.

La modélisation de la queue de distribution a également mis en évidence une problématique spécifique en réassurance : l'appartenance au domaine d'attraction de Gumbel, qui concerne les distributions à queue fine. Lorsque la queue de distribution des sinistres de la cédante appartient à ce domaine, la faible probabilité d'occurrence d'un sinistre extrême peut poser un défi en matière de tarification. Malgré cette rareté théorique des événements extrêmes, le besoin de réassurance souligne l'importance de prendre en compte le risque potentiel de sinistres majeurs dans les tranches hautes.

Dans le domaine d'attraction de Gumbel, bien que la loi exponentielle s'ajuste correctement aux données, elle sous-estime la probabilité d'occurrence des sinistres extrêmes. Ainsi, la solution proposée repose sur l'introduction de fonctions de distorsion, qui ajustent la loi exponentielle de la cédante vers une loi de marché adaptée au traité. L'utilisation du critère de divergence de KL et la construction d'un facteur de crédibilité ont permis d'optimiser ce processus en déterminant le paramètre de distorsion idéal pour la cédante concernée. Ce cadre assure que le choix du paramètre de distorsion est rigoureux et non arbitraire, renforçant ainsi la robustesse du modèle.

En conclusion, ce mémoire a permis de faire évoluer la méthode de modélisation de la sévérité des sinistres lors de l'utilisation du modèle collectif. La méthode globale, qui consiste à séparer le corps et la queue de distribution grâce au seuil des valeurs extrêmes, se divise ensuite en deux approches en fonction du domaine d'attraction de la queue de distribution de la cédante. Les approches proposées offrent ainsi une base solide pour une tarification plus juste sur le plan théorique, augmentant la confiance envers les résultats obtenus.

Néanmoins, certaines limites ont été identifiées dans l'approche présentée, et plusieurs pistes de recherche pourraient permettre d'y remédier.

Tout d'abord, l'exploration de nouvelles fonctions de distorsion pourrait apporter des solutions aux limitations observées (cf. [8.5.1](#)), notamment en ce qui concerne l'écart dans l'étendue des sinistres selon le domaine d'attraction. Bien que ce cas soit rare, se produi-

sant uniquement lorsque les données se trouvent dans la queue de distribution de la loi distordue, appartenant elle-même à la queue de distribution de la sévérité, il demeure un problème moins fréquent mais intéressant à étudier.

Ensuite, la construction de la loi de marché, n'étant pas le sujet principal du mémoire, n'a pas été abordée en détail dans cette étude. En revanche, pour que l'approche de la distorsion soit pleinement efficace, il est nécessaire de développer une loi de marché qui reflète fidèlement la sinistralité du marché en question. Ainsi, une étude parallèle et complémentaire à ce mémoire pourrait consister en la construction de lois de marché pour toutes les branches courtes de réassurance et pour toutes les zones géographiques concernées. Cela permettrait de développer des représentations plus précises de la sinistralité pour chaque contexte spécifique, améliorant ainsi la pertinence et la fiabilité des résultats pour l'approche de distorsion.

Par ailleurs, ce mémoire ouvre également de nouvelles perspectives pour l'amélioration des pratiques de tarification en réassurance. Une première piste serait de modéliser le corps de la distribution de la sévérité par une approche paramétrique plutôt qu'empirique. Ceci permettrait d'étudier l'impact de ce changement sur les résultats. Par exemple, le mélange d'Erlang pourrait être exploré pour obtenir une modélisation plus précise.

Enfin, une étude similaire sur la modélisation de la fréquence des sinistres pourrait être menée. Cela permettrait d'approfondir la compréhension des interactions entre fréquence et sévérité, et ainsi de raffiner davantage les méthodes de tarification.

Ces pistes offrent des opportunités intéressantes pour enrichir les travaux actuels et faire évoluer les méthodes de tarification en réassurance.

L'étude menée dans ce mémoire a abouti au développement d'un outil de tarification des traités XS par expérience, conçu avec l'interface de programmation Python. Cet outil intègre l'approche du modèle composite détaillée dans ce travail et sera utilisé lors des prochaines périodes de renouvellement.

Bibliographie

- [AP23] APREF (*Association des professionnels de la réassurance en France*), site consulté le 2 septembre 2024.
- [BL20] BIBER, L. (2020). *Modélisation de la distribution de la sévérité en réassurance*. Mémoire d'actuaire, ISFA.
- [BE01] BILHAUD, E. (2001). *Bases techniques de l'assurance*. Groupe ENAss – AEA.
- [BE012] BILHAUD, E. (2001). *La tarification des traités de réassurance non proportionnels*. Groupe ENAss – AEA.
- [BE013] BILHAUD, E. (2001). *Techniques de la réassurance*. Groupe ENAss – AEA.
- [BA10] BORCHANI, A. (2010). *Statistiques des valeurs extrêmes dans le cas de lois discrètes*. HAL.
- [CE14] COHIGNAC, T., ECOTO, G. (2014). *La tarification de réassurance CAT-NAT*. CCR.
- [DA10] DELACROIX, A. (2021). *Réassurance Non-Vie*. ISFA.
- [DG07] DEPREY, E., GOLDZINSKI, A. (2007). *Problématique de seuil dans la modélisation de la sinistralité en Réassurance Non-Vie*. Mémoire d'actuariat, ENSAE.
- [ED18] ESCOBAR, D., PFLUG, G. (2018). *The distortion principle for insurance pricing : properties, identification and robustness*. Springer.
- [GL20] GARDES, L. (2020). *Théorie des valeurs extrêmes*. Université de Strasbourg.
- [GY11] GUILLOU, A., YOU, A. (2011). *Introduction à la théorie des valeurs extrêmes : Application en actuariat*. Université de Strasbourg et Société générale insurance.
- [MI24] MORDOR INTELLIGENCE (*Analyse de la taille et de la part du marché de la réassurance – Tendances et prévisions de croissance 2024-2029*), site consulté le 2 septembre 2024.
- [PJ75] PICKANDS, J. (1975). *Statistical inference using extreme order statistics*. Annals of statistics, 3.
- [PA21] PIERRE, A. (2021). *Modélisation de la sévérité des traités en excédent de sinistres, approche par la théorie des valeurs extrêmes*. Mémoire d'actuaire, DUAS.
- [PM12] POULIN, M. (2012). *Analyse des solutions actuarielles en tarification des traités de réassurance non-proportionnels Non-Vie*. Mémoire d'actuaire, centre d'études actuarielles.

- [SR84] SMITH, R. L. (1984). *Threshold Methods for Sample Extremes*. Springer.
- [TP04] THEROND, P. (2004). *Assurance non-vie, le modèle collectif*. ISFA.
- [TP042] THEROND, P. (2004). *Mesures et comparaison de risques*. ISFA.
- [WS95] WANG, S. (1995). *Insurance pricing and increases limits ratemaking by proportional hazards transforms*. *Mathematics and Economics* 17, 43-54.
- [WS97] WANG, S. (1997). *Implementation of proportional hazards transforms in rate-making*. PhD.

Annexes

A - Le test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov ("KS") est un test statistique permettant de déterminer si un échantillon de données empiriques suit une loi statistique modélisée. En d'autres termes, ce test compare la fonction de répartition empirique de la distribution observée d'un échantillon avec la fonction de répartition théorique de la loi statistique supposée. La procédure du test est la suivante :

Données : Soit $(x_1, x_2, \dots, x_n)$ un ensemble de n observations empiriques de la variable aléatoire X , et F sa fonction de répartition empirique. F_0 représente la fonction de répartition de la loi statistique théorique que l'on souhaite tester.

Hypothèse : L'hypothèse nulle (H_0) stipule que la fonction de répartition empirique de X est égale à F_0 , c'est-à-dire que l'échantillon suit bien la loi théorique.

Calcul : L'échantillon est d'abord ordonné tel que $x_1 \leq x_2 \leq \dots \leq x_n$. La statistique du test est calculée comme suit :

$$D_n = \sup_x |F_n(x) - F_0(x)|$$

où $F_n(x)$ est la fonction de répartition empirique et $F_0(x)$ la fonction de répartition théorique. D_n est le maximum des écarts absolus entre les fonctions de répartition empirique et théorique.

Décision : La comparaison de la statistique D_n avec la valeur critique d_n issue des tables de la distribution du test de KS se fait comme suit :

- Si $D_n < d_n$, l'hypothèse nulle H_0 est acceptée, ce qui indique que l'échantillon est compatible avec la loi théorique F_0 .
- Si $D_n \geq d_n$, l'hypothèse nulle est rejetée, suggérant que l'échantillon n'est pas compatible avec la loi théorique F_0 .

En conclusion, si l'hypothèse H_0 est acceptée, la loi statistique testée peut être considérée comme une bonne représentation de la distribution empirique des données.

B - Les critères statistiques

1 - Le critère d'information d'Akaike

Le critère d'information d'Akaike ("AIC") est une mesure utilisée pour évaluer la qualité d'un modèle statistique. Il est calculé selon la formule suivante :

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

où :

- k représente le nombre de paramètres à estimer.
- L est le maximum de vraisemblance du modèle.

L'AIC mesure la qualité de l'ajustement d'un modèle tout en intégrant la complexité de celui-ci. En effet, il pénalise les modèles comportant un grand nombre de paramètres pour éviter le sur-ajustement. Lorsque plusieurs modèles sont comparés, celui qui minimise l'AIC est considéré comme le meilleur choix.

2 - Le critère d'information bayésien

Le critère d'information bayésien ("BIC"), qui s'inspire de l'AIC, est une autre mesure d'évaluation. Sa formule est la suivante :

$$BIC = -2 \ln(L) + \ln(n)k$$

où :

- n est le nombre d'observations dans l'échantillon.
- k représente le nombre de paramètres.
- L est le maximum de vraisemblance du modèle.

Le BIC, comme l'AIC, quantifie la qualité de l'ajustement tout en tenant compte de la complexité du modèle, en pénalisant le nombre de paramètres à estimer. Cependant, le BIC introduit une pénalisation supplémentaire liée au nombre d'observations, ce qui le rend particulièrement utile lorsque le volume de données est limité. Ainsi, moins il y a de données dans l'échantillon, moins l'ajustement des paramètres est fiable, ce que le BIC prend en compte. Le modèle avec le plus faible BIC est considéré comme le meilleur.

Dans la modélisation statistique de la sévérité des sinistres de la cédante, ces deux critères sont calculés pour identifier le modèle statistique le plus approprié.

C - Linéarité du mean excess plot

Soit X une variable aléatoire avec $X \sim GPD(\sigma, \gamma)$. Soit $\gamma > 1$, la fonction des excès moyen s'écrit de la manière suivante :

$$\begin{aligned} e(u) &= \mathbb{E}[X - u | X > u] \\ &= \frac{\int_u^{+\infty} (1 - F(x)) dx}{1 - F(u)} \\ &= \frac{1}{\mathbb{P}(X < u)} \times \int_u^{+\infty} \mathbb{P}(X > x) dx \end{aligned}$$

Avec :

$$\mathbb{P}(X > u) = \left(1 + \frac{\gamma u}{\sigma}\right)^{-\frac{1}{\gamma}}$$

et :

$$\int_u^{+\infty} \mathbb{P}(X > x) dx = \int_u^{+\infty} \left(1 + \frac{\gamma x}{\sigma}\right)^{-\frac{1}{\gamma}} dx$$

Le changement de variable $y = 1 + \frac{\gamma x}{\sigma}$ donne : $\int_u^{+\infty} \mathbb{P}(X > x) dx = \frac{\sigma}{1-\gamma} \left[1 + \frac{\gamma u}{\sigma}\right]^{1-\frac{1}{\gamma}}$

Ainsi :

$$\begin{aligned} e(u) &= \frac{\sigma}{1-\gamma} \left[1 + \frac{\gamma u}{\sigma}\right]^{1-\frac{1}{\gamma}} \times \frac{1}{\left[1 + \frac{\gamma u}{\sigma}\right]^{\frac{-1}{\gamma}}} \\ &= \frac{\sigma}{1-\gamma} \left[1 + \frac{\gamma u}{\sigma}\right] \\ &= \frac{\gamma}{1-\gamma} \times u + \frac{\sigma}{1-\gamma} \end{aligned}$$

Ainsi, si $X \sim GPD(\gamma, \sigma)$ alors la fonction des excès moyen est affine en u .

D - Choix du seuil des extrêmes de l'exemple 6.1.2

Dans l'exemple 6.1.2, le seuil des extrêmes a été fixé à $u_0 = 499387,6$. Ce choix est justifié par l'analyse du mean excess plot et du Threshold Stability Plot.

Le graphique du mean excess plot est présenté ci-dessous :

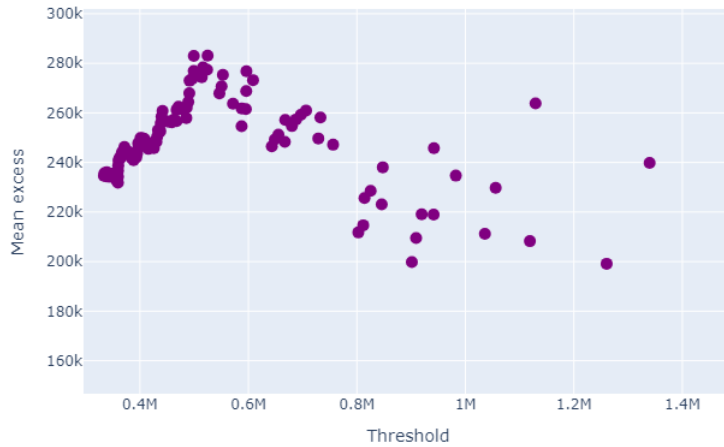


FIGURE 8.9 – mean excess plot exemple

Le graphique révèle plusieurs éléments clés :

- **Partie gauche du graphique** : La linéarité descendante relativement stable indique une tendance cohérente dans les données jusqu'à un certain point.
- **À partir de $u_0 = 499387,6$** : Un changement significatif est détecté, où la tendance linéaire passe d'une croissance à une décroissance. Ce point marque une transition claire par rapport aux seuils inférieurs, où la tendance était nettement décroissante.
- **Points extrêmes** : Les derniers points du graphique montrent des valeurs instables, souvent considérées comme des valeurs aberrantes.

Ces observations soutiennent que le seuil des extrêmes fixé à $u_0 = 499387,6$ est approprié.

Le threshold stability plot est affiché ci-dessous :

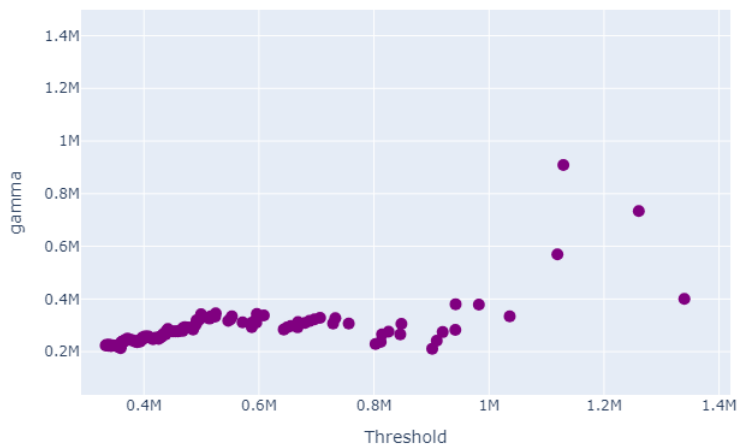


FIGURE 8.10 – Stability plot exemple

Le Threshold Stability Plot confirme cette sélection :

- **Stabilité de l'estimation du paramètre γ** : Dès le début sur l'axe des abscisses, l'estimation de γ reste stable, renforçant la confiance dans le choix du seuil.
- **Montants élevés** : Les points situés plus à droite montrent une instabilité dans l'estimation du paramètre γ , ce qui justifie l'exclusion de ces montants.

Ainsi, le seuil $u_0 = 499387,6$ déterminé par le mean excess plot est confirmé par le threshold stability plot, validant son choix pour l'analyse des extrêmes dans cet exemple.