

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaire
le 06/03/2025

Par : Léa Tenaille

Titre : Génération de scénarios de développement
de sinistres IBNeR et IBNyR individuels
en réassurance

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

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

Membres présents du jury de la filière

Entreprise : AXA GIE 

Nom : Nicolas Baradel

Signature :



*Membres présents du jury de l'Institut
des Actuaire*

Directeur du mémoire en entreprise :

Nom : Baptiste Voisin

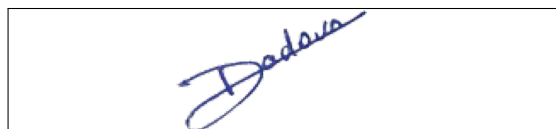
Signature :



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

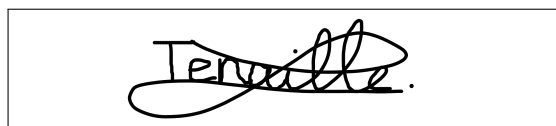
Secrétariat :

Signature du responsable entreprise



Bibliothèque :

Signature du candidat



Résumé

L'inversion du cycle de production est une particularité des compagnies d'assurance et de réassurance. Au moment où les contrats sont signés, les compagnies ne connaissent pas le montant de la prestation qu'elles devront verser dans le futur aux assurés. Les compagnies doivent donc constituer des provisions suffisantes afin de faire face à leurs engagements pris envers leurs clients.

La plupart des méthodes de provisionnement actuelles, dites classiques, exploitent des données agrégées sur des groupes de risques homogènes. Elles permettent d'identifier des tendances de développement des sinistres. Ce principe d'agrégation se base sur l'effet de mutualisation entre des montants de sinistres élevés et d'autres faibles, nécessitant ainsi l'utilisation d'un nombre important de sinistres.

En réassurance, ces approches sont moins robustes à cause du faible nombre de sinistres et de la variabilité très importante de leur montant. Pour nuancer les estimations de ces méthodes classiques, il est pertinent d'accéder aux informations individuelles des sinistres afin de générer des scénarios de sinistralité s'appuyant sur ces informations. Ces scénarios permettent de mettre en perspective les estimations des méthodes classiques et d'expliquer les résultats de ces estimations.

Ce souhait d'interprétabilité motive les travaux présentés dans ce mémoire et réalisés dans le cadre de la réassurance. Ils portent sur la génération de divers scénarios de développement futur des sinistres à travers une approche de provisionnement individuel. Ces scénarios permettent de déterminer la manière dont les provisions calculées en amont par des méthodes de provisionnement classiques pourraient être consommées par ces sinistres alors développés. Cette étude s'articule autour de deux axes : les sinistres survenus et déclarés mais pas assez provisionnés – les IBNeR – et les sinistres survenus mais non encore déclarés – les IBNyR.

Mots-clefs : Réassurance, Provisionnement, IARD, Modélisation, Traités, IBNR (IBNeR et IBNyR), Méthode de Kaplan-Meier, Régression logistique, Méthode de Schnieper, Loi de Poisson, Loi binomiale négative.

Abstract

The inversion of the production cycle is a particularity of insurance and reinsurance companies. When contracts are signed, companies do not know the amount of service they will have to pay out to policyholders in the future. Therefore, companies must set aside sufficient reserves to meet customer commitments.

Most of the reserving methods currently used by companies, so-called classical, rely on aggregated data for homogeneous risk groups. They help identify trends in claims development. This aggregation principle is based on the offsetting effect between high and low claim amounts, thus requiring a large number of claims to be considered.

In reinsurance, these approaches are less robust due to the small number of claims and the high variability of their amount. To qualify the estimates of these classical methods, it is relevant to access individual claims information to generate loss scenarios based on these pieces of information. These scenarios put the estimations of the classical methods into perspective and explain the results of these estimations.

This desire for interpretability motivates the work presented in this master thesis and carried out in the context of reinsurance. It focuses on the generation of various future claims development scenarios through an individual reserving approach. These scenarios aim to determine how the reserves calculated upstream by classical reserving methods might be consumed by the claims then developed. This study is based on two axes: Incurred But Not enough Reserved claims – IBNeR – and Incurred But Not yet Reported claims – IBNyR.

Keywords : *Reinsurance, Reserving, P&C (Property & Casualty), Modeling, Treaties, IBNR (IBNeR and IBNyR), Kaplan-Meier method, Logistic Regression, Schnieper method, Poisson distribution, Negative binomial distribution.*

Note de synthèse

Contexte

En raison de l'inversion du cycle de production dans le domaine assurantiel, les entreprises d'assurance et de réassurance ont l'obligation d'estimer des provisions afin de tenir leurs engagements auprès de leurs assurés. Dans ce cadre de provisionnement actuariel, l'équipe Reserving, appartenant au département AXA Group Ceded Reinsurance au sein de l'entreprise AXA GIE, calcule des provisions spécifiques aux sinistres IBNR (*Incurred But Not Reported* signifiant « survenus mais non déclarés »). Ces sinistres sont divisés en deux catégories :

- Les sinistres IBNeR (*Incurred But Not enough Reserved*) : ce sont les sinistres déclarés n'ayant pas encore atteint leur charge ultime et par conséquent, n'étant pas assez provisionnés.
- Les sinistres IBNyR (*Incurred But Not yet Reported*) : ce sont les sinistres tardifs et par conséquent, ceux dont aucune charge n'est connue.

L'équipe Reserving calcule ces provisions IBNR à l'aide de méthodes de provisionnement dites « classiques », sans distinction entre les provisions dites IBNeR et celles dites IBNyR. Ces méthodes s'appuient sur des données agrégées par groupes de risques homogènes et sont statistiquement stables lorsque le nombre de sinistres est élevé puisqu'elles reposent sur la mutualisation des risques. En raison du faible nombre de sinistres et de leur volatilité importante dans le cadre de la réassurance, ces méthodes s'avèrent moins robustes. L'intérêt du recours à des méthodes de provisionnement sinistre par sinistre semble alors pertinent.

L'objectif des travaux présentés dans le cadre de ce mémoire est l'interprétation des provisions IBNR calculées par l'équipe Reserving pour l'ensemble des 80 segments de provisionnement qui composent le portefeuille d'AXA SA. L'approche proposée consiste, pour chaque segment, à générer 1 000 scénarios de développement individuel des sinistres IBNeR et IBNyR par année de survenance afin de construire une distribution des provisions IBNR.

Démarche

En amont des deux modélisations distinctes des sinistres IBNeR et IBNyR, les bases de données exploitées sont retraitées afin d'établir une structure cohérente, facilitant la modéli-

sation par la suite. Dans ces bases, chaque sinistre est associé, entre autres, à une année de survenance, une année de développement, une charge FGU (c'est-à-dire au premier euro), une charge à l'acceptation d'AXA SA, une charge à la charge d'AXA SA et des conditions de traités de réassurance.

Modélisation des sinistres IBNeR

La modélisation des sinistres IBNeR proposée dans ce mémoire porte sur la simulation de leur Charge FGU ultime à travers trois méthodes dont les différentes étapes sont présentées ci-dessous. Les étapes 1 à 5 sont communes aux trois méthodes, exceptée l'étape 2 qui est spécifique à la méthode 3.

1. **Identification des sinistres IBNeR** : Une variable binaire est créée pour identifier les sinistres IBNeR. Cette variable repose en partie sur une variable existante qui détermine si un sinistre est clos ou actif. L'identification des sinistres IBNeR ne peut toutefois pas se reposer uniquement sur cette variable existante car elle s'est avérée trop peu fiable par rapport à son importance dans le modèle proposé. La nouvelle variable se base alors également sur la valeur des *link ratios* historiques des sinistres ainsi que sur la tendance du nombre d'années de développement nécessaires avant que 97,5% des sinistres n'aient atteint leur Charge FGU ultime. Plus ces sinistres ont une tendance de développement à long terme, plus le nombre d'années de développement nécessaires est élevé. Cette tendance s'appuie sur des facteurs de développement lissés. Ces derniers correspondent à la concaténation des facteurs de développement historiques – estimés par la méthode de la *Chain Ladder* – et des facteurs de queue estimés jusqu'à la 30^{ème} année de développement. Les valeurs de la nouvelle variable dépendent également du pourcentage de règlements déjà effectués par rapport à la dernière Charge FGU connue. Un ensemble de règles est établi en fonction de ces différentes variables, permettant d'identifier les sinistres IBNeR.
2. **Identification des sinistres IBNeR sujets à développement** : Parmi les sinistres identifiés comme IBNeR à l'étape précédente, certains peuvent tout de même ne plus se développer. La méthode 3 prend en compte cette possibilité en identifiant les sinistres IBNeR sujets à développement. Pour cela, elle a recours d'une part, à la méthode de Kaplan-Meier et d'autre part, à la régression logistique.

Concernant la méthode de Kaplan-Meier, le parallèle est effectué entre la durée de vie humaine et le nombre d'années de développement nécessaires avant qu'un sinistre ne soit plus sujet à développement. Les résultats de la méthode de Kaplan-Meier permettent alors, selon une loi binomiale, de simuler un nombre de sinistres IBNeR sujets à développement parmi l'ensemble des sinistres IBNeR d'une même année de survenance. Une distinction est réalisée entre les sinistres atypiques – ceux touchant la réassurance et dont la Charge FGU est supérieure à 1M€ – et les sinistres non atypiques, car leurs probabilités

d'être encore sujets à développement sont différentes.

La régression logistique permet de classer les sinistres IBNeR en fonction de leur chance d'être encore sujets à développement, en considérant que les sinistres IBNeR n'aient pas tous la même chance d'y être. Cette régression s'appuie sur leur dernière année de développement connue, leur pourcentage de règlements déjà effectués ainsi que sur le fait qu'ils aient été soumis ou non à des développements dans leurs trois dernières années de développement.

Le schéma en figure 1 ci-dessous permet de récapituler cette étape. Les pourcentages illustrent la probabilité des sinistres IBNeR de ne plus être sujets à développement. Le chiffre 3 correspond à la réalisation d'une loi binomiale à partir de la probabilité – estimée par la méthode de Kaplan-Meier – que les sinistres soient sujets à développement.

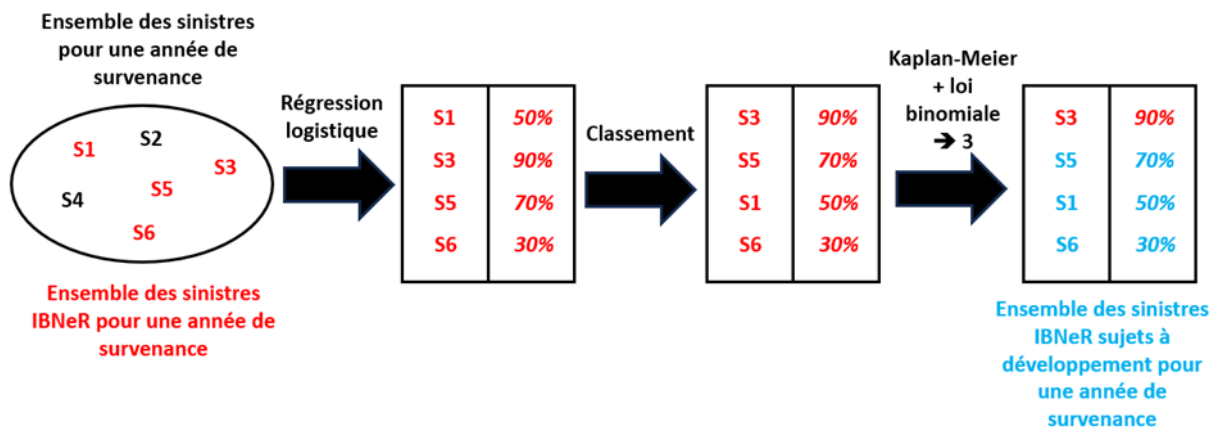


FIGURE 1 – Récapitulatif de l'identification des sinistres IBNeR sujets à développement

- 3. Construction des coefficients de développement** : Les coefficients de développement sont les coefficients appliqués aux dernières Charges FGU connues des sinistres IBNeR – ou des sinistres IBNeR sujets à développement dans le cadre de la méthode 3 – afin d'obtenir la Charge FGU ultime de ces derniers. Ces coefficients sont construits à partir des *link ratios* historiques de l'ensemble des sinistres. Lorsque les sinistres dont proviennent les *link ratios* – appelés alors « sinistres d'origine » – sont identifiés à l'étape 1 comme ayant atteint leur ultime, leur coefficient de développement correspond simplement à leur *link ratio*. Lorsque les sinistres d'origine appartiennent eux-mêmes à la catégorie des sinistres IBNeR, leur coefficient de développement est le résultat de la multiplication entre leur *link ratio* et le facteur de développement lissé – évoqué à l'étape 2 – associé à la dernière année de développement connue des sinistres.
- 4. Tirage aléatoire d'un coefficient de développement** : Un coefficient de développement est tiré aléatoirement pour chacun des sinistres IBNeR – ou des sinistres IBNeR sujets à développement dans le cadre de la méthode 3 – selon une loi uniforme parmi les

coefficients de développement construits à l'étape précédente. Les coefficients de développement pouvant être tirés aléatoirement dépendent de la méthode utilisée :

- La méthode 1 consiste à tirer un coefficient de développement parmi l'ensemble de ceux associés à l'année de développement suivant la dernière année de développement connue des sinistres IBNeR.
- La méthode 2 reprend le principe de la méthode 1 en associant toutefois aux sinistres IBNeR, des coefficients de développement issus de sinistres d'origine dont la Charge FGU est proche de celle des sinistres IBNeR.
- La méthode 3 est identique à la méthode 2, à la différence que seuls les coefficients de développement strictement supérieurs à 1 peuvent être tirés, puisqu'ils sont appliqués à des sinistres modélisés comme n'ayant pas encore atteint leur ultime.

Dans les méthodes 1 et 2, puisqu'il est possible de tirer aléatoirement un coefficient de développement égal à 1, les sinistres n'étant plus sujets à développement sont également modélisés intrinsèquement.

5. **Calcul des provisions IBNeR** : Une fois les Charges FGU des sinistres IBNeR multipliées par les coefficients de développement, les conditions des traités de réassurance – auxquels les sinistres IBNeR sont rattachés – sont appliquées. Les provisions IBNeR correspondent alors à la différence entre les charges ultimes à l'acceptation d'AXA ASA ou à la charge d'AXA SA, et celles à la dernière année de développement connue.

Modélisation des sinistres IBNyR

Les sinistres IBNyR sont les sinistres inconnus de la réassurance, tant en termes de nombre que de Charge FGU ultime. La modélisation de ces sinistres proposée dans le cadre de ce mémoire suit les deux étapes suivantes.

1. **Estimation du nombre de sinistres IBNyR** : Pour cette première étape, la méthode de Schnieper permet d'estimer l'espérance et la variance des nombres inconnus du triangle de développement de nombres incrémentaux. L'exposition prise en compte est celle reflétant le risque réel total et le triangle est composé des sinistres dont la Charge FGU est supérieure à la fois, à 75% de la priorité du traité auxquels ils sont rattachés et à 1,5M€. Les espérances et les variances sont par la suite utilisées pour déterminer les paramètres de lois – soit de Poisson, soit binomiales négatives. Ces lois permettent alors de simuler chaque nombre de sinistres IBNyR inconnu du triangle. L'addition des nombres simulés de sinistres IBNyR de chaque année de développement d'une même année de survenance permet d'obtenir le nombre total de sinistres IBNyR pour une année de survenance.
2. **Estimation de la Charge FGU ultime des sinistres IBNyR simulés** : À chaque sinistre IBNyR simulé, une Charge FGU ultime est tirée aléatoirement selon une loi

uniforme parmi des Charges FGU ultimes de sinistres historiques – alors appelés « sinistres d’origine ». Ces sinistres d’origine correspondent à ceux ayant servi à construire le triangle de nombres à l’étape précédente et non à l’ensemble des sinistres historiques. Lorsque ces sinistres d’origine appartiennent à la catégorie des sinistres IBNeR – identifiés à l’étape 1 de la modélisation des sinistres IBNeR –, leur Charge FGU ultime est considérée comme étant leur dernière Charge FGU connue multipliée par le facteur de développement lissé associé à leur dernière année de développement connue. Par ailleurs, seuls les sinistres d’origine rattachés à des traités de réassurance ayant un équivalent dans le programme auquel ils appartiennent pour l’année de survenance modélisée, sont inclus dans le tirage. Les conditions des traités de réassurance équivalents sont alors appliquées aux Charges FGU ultimes tirées à partir des sinistres d’origine. Les charges à l’acceptation d’AXA SA et à la charge d’AXA SA correspondent alors aux provisions IBNyR.

Résultats

Pour chacun des 1 000 scénarios, les provisions IBNR par année de survenance sont obtenues en additionnant les provisions IBNeR et IBNyR de chaque sinistre de l’année de survenance en question. Les distributions des provisions IBNR générées par ces 1 000 scénarios – principalement portées par les provisions IBNyR par ailleurs – sont très asymétriques quels que soient l’année de survenance et le segment de provisionnement. Le graphique en figure 2 ci-dessous illustre cette asymétrie à travers la fonction de répartition des provisions IBNR simulées d’un segment de provisionnement. La barre verticale rouge correspond aux provisions IBNR calculées par l’équipe Reserving.

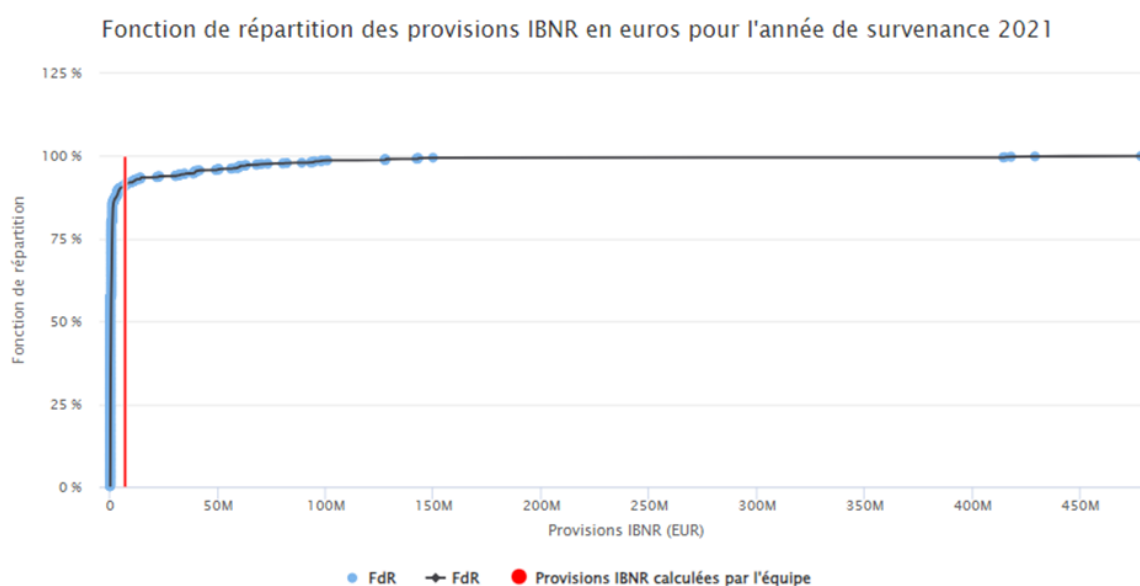


FIGURE 2 – Fonction de répartition des provisions IBNR d’un segment de provisionnement

Dans cet exemple, et comme pour la plupart des années de survenance de l'ensemble des segments de provisionnement, les provisions IBNR calculées par l'équipe Reserving sont jugées surestimées et trop prudentes d'après le modèle proposé dans le cadre de ce mémoire. Cette observation repose sur les centiles qui leur sont associés et qui atteignent généralement les 80% – 90%. Par ailleurs, la différence entre les trois méthodes d'estimation des provisions IBNeR est faible.

En outre, il est possible de s'intéresser aux scénarios générant des provisions IBNR proches de celles calculées par l'équipe, dans le but de permettre à cette dernière de mieux pouvoir justifier ses résultats.

Conclusion

L'objectif d'interprétabilité des provisions IBNR calculées par l'équipe Reserving est atteint. Il est en effet possible de retrouver le cheminement de la construction de chaque provision IBNeR et IBNyR grâce aux informations individuelles des sinistres exploitées tout au long du modèle proposé, notamment à travers l'application **RShiny** développée lors du stage. De plus, les résultats de ce modèle permettent de positionner et d'interpréter les provisions IBNR calculées par l'équipe par rapport à celles modélisées dans le cadre de ce mémoire afin d'évaluer leur niveau de prudence. Toutefois, ces résultats sont à mettre en perspective car plusieurs biais sont présents dans ce modèle. Premièrement, de nombreuses hypothèses ont été émises tout au long du modèle et leur impact serait à analyser. Deuxièmement, les reconstitutions, les AAD, les AAL et les portées et priorités indexées ne sont pas prises en compte. Troisièmement, dans le cadre de la modélisation des sinistres IBNyR, seuls des risques qui se sont déjà produits peuvent être tirés aléatoirement. Simuler les Charges FGU ultimes des sinistres IBNyR selon une loi de sévérité pourrait contribuer à l'amélioration du modèle.

À vision plus long terme, une version consolidée de ce modèle pourrait amener l'équipe Reserving à s'interroger sur la pertinence des méthodes qu'elle exploite en comparant leurs résultats avec ceux du modèle proposé dans ce mémoire et le cas échéant, ajuster ses méthodes agrégées pour mieux refléter la sinistralité.

Executive summary

Context

Due to the inversion of the production cycle in the insurance business, it is mandatory for insurance and reinsurance companies to estimate reserves in order to meet their policyholder's commitments. As part of this actuarial reserving process, the Reserving team in the AXA Group Ceded Reinsurance department within AXA GIE calculates reserves specifically for IBNR (Incurred But Not Reported) claims. These claims are divided into two categories:

- IBNeR claims (Incurred But Not enough Reserved): these are reported claims that have not yet reached their ultimate incurred loss, and are therefore not sufficiently reserved.
- IBNyR claims (Incurred But Not yet Reported): these are late claims, and consequently, those for which no incurred losses are known.

The Reserving team calculates these IBNR reserves using “classical” reserving methods, without distinguishing between IBNeR and IBNyR reserves. These methods are based on data aggregated by homogeneous risk groups and they are statistically stable when the number of claims is high, since they are based on risk pooling. However, in the case of reinsurance, these methods are less robust, due to the low number of claims and their high volatility. The use of individual reserving methods therefore seems appropriate.

The aim of the work presented in this master thesis is to interpret the IBNR reserves calculated by the Reserving team for all 80 lines of business that make up the AXA SA portfolio. The proposed approach consists, for each line of business, of generating 1,000 individual IBNeR and IBNyR claims development scenarios per accident year in order to build a distribution of IBNR reserves.

Approach

Prior to the two distinct models of IBNeR and IBNyR claims, the databases used are reprocessed to establish a consistent structure, facilitating subsequent modeling. In these databases, each claim is associated with, among other variables, an accident year, a development year, an FGU (i.e. first euro) incurred loss, an incurred loss accepted by AXA SA, an incurred loss borne by AXA SA and reinsurance treaty conditions.

IBNeR claims modeling

The modeling of IBNeR claims proposed in this master thesis focuses on the simulation of their ultimate FGU incurred loss using three methods, the different steps of which are presented below. Steps 1 to 5 are common to all three methods, except step 2 which is specific to method 3.

1. **Identification of IBNeR claims:** A binary variable is created to identify IBNeR claims.

This variable is partly based on an existing variable that determines whether a claim is closed or active. However, the identification of IBNeR claims cannot rely solely on this existing variable, as it has proved insufficiently reliable given its importance in the proposed model. The new variable is therefore also based on the value of historical claims link ratios and on the trend in the number of development years required before 95% of claims have reached their ultimate FGU incurred loss. The more these claims have a long-term development trend, the greater the number of development years required. This trend is based on smoothed development factors. These correspond to the concatenation of historical development factors – estimated by the Chain Ladder method – and tail factors estimated up to the 30th development year. The values of the new variable also depend on the percentage of paid amounts relative to the last known FGU incurred loss. Based on these different variables, a set of rules is established to identify IBNeR claims.

2. **Identification of IBNeR claims subject to development:** Among the claims identified as IBNeR in the previous step, some may nevertheless no longer develop. Method 3 takes this possibility into account by identifying IBNeR claims subject to development. To do this, it uses the Kaplan-Meier method on the one hand, and logistic regression on the other hand.

For the Kaplan-Meier method, a parallel is drawn between the human lifespan and the number of development years required before a claim is no longer subject to development. Using a binomial distribution, the results of the Kaplan-Meier method are then used to simulate the number of IBNeR claims subject to development among all IBNeR claims for the same accident year. A distinction is made between atypical claims – those involving reinsurance and with a FGU incurred loss in excess of €1M – and non-atypical claims, since they have different probabilities of still being subject to development.

Logistic regression is used to classify IBNeR claims according to their chance of still being subject to development, considering that not all IBNeR claims have the same chance of being subject to development. This regression is based on their last known development year, their percentage of paid amounts, and whether or not they were subject to development in their last three development years.

The diagram in figure 3 below summarizes this step. The percentages illustrate the probability of IBNeR claims no longer being subject to development. The number 3

corresponds to the realization of a binomial distribution based on the probability – estimated by the Kaplan-Meier method – of the claims being subject to development.

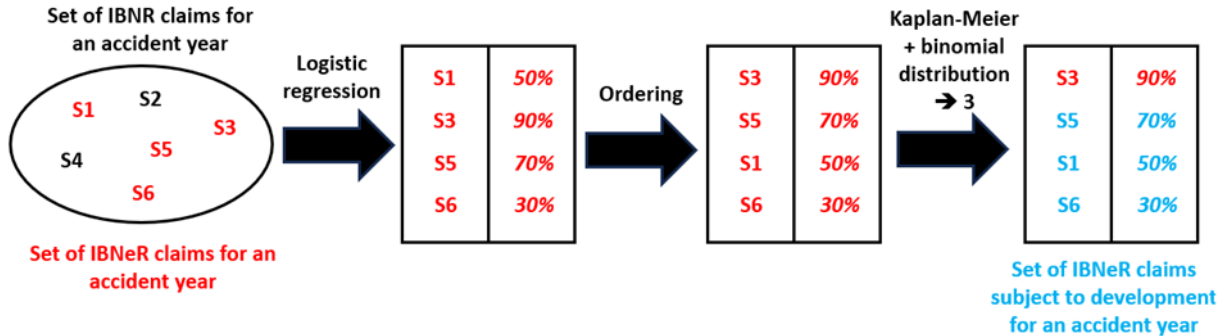


Figure 3 – Summary of the identification of IBNeR claims subject to development

3. **Construction of development coefficients:** Development coefficients are coefficients applied to the last known FGU incurred losses of IBNeR claims – or IBNeR claims subject to development under method 3 – to obtain their ultimate FGU incurred loss. These coefficients are derived from the historical link ratios for all claims. When the claims from which the link ratios are derived – referred to as “original claims” – are identified in step 1 as having reached their ultimate, their development coefficient simply corresponds to their link ratio. When the original claims themselves belong to the IBNeR claims category, their development coefficient is the result of multiplying their link ratio by the smoothed development factor – referred to in step 2 – associated with the last known development year of claims.
4. **Random drawing of a development coefficient:** A development coefficient is randomly drawn for each IBNeR claim – or for IBNeR claims subject to development under method 3 – according to a uniform distribution among the development coefficients constructed in the previous step. The development coefficients that can be drawn at random depend on the method used:
 - Method 1 consists of drawing a development coefficient from all those associated with the development year following the last known development year for IBNeR claims.
 - Method 2 uses the same principle as method 1, but associates development coefficients from original claims with FGU incurred loss close to those of IBNeR claims.
 - Method 3 is identical to method 2, with the difference that only development coefficients strictly greater than 1 can be drawn, since they are applied to claims modeled as not yet having reached their ultimate.

In methods 1 and 2, since it is possible to randomly draw a development coefficient equal to 1, claims no longer subject to development are also modeled intrinsically.

5. **Estimation of IBNeR reserves:** Once the FGU incurred losses for IBNeR claims have been multiplied by the development coefficients, the conditions of the reinsurance treaties – to which the IBNeR claims are attached – are applied. The IBNeR reserves then correspond to the difference between the ultimate incurred losses accepted by AXA ASA or borne by AXA SA and those in the last known development year.

Modeling IBNyR claims

IBNyR claims are claims unknown to reinsurance, both in terms of number and ultimate FGU incurred loss. The modeling of these claims proposed in this master thesis follows the following two steps.

1. **Estimation of the number of IBNyR claims:** For this first step, Schnieper’s method is used to estimate the expectation and variance of the unknown numbers in the incremental number development triangle. The exposure taken into account is the one reflecting the total real risk, and the triangle is made up of claims affecting the treaty priority to which they are attached and whose FGU incurred loss is greater than €2M. Expectations and variances are then used to determine the parameters of either Poisson or negative binomial distributions. These distributions are thereafter used to simulate each unknown number of IBNyR claims in the triangle. The simulated number of IBNyR claims for each development year within the same accident year is added together to obtain the total number of IBNyR claims for a given accident year.
2. **Estimation of the ultimate FGU incurred loss of simulated IBNyR claims:** For each simulated IBNyR claim, an ultimate FGU incurred loss is randomly drawn according to a uniform distribution from among the ultimate FGU incurred losses of historical claims – then called “original claims”. These original claims correspond to those used to construct the triangle of numbers in the previous step, and not to all historical claims. When these original claims belong to the IBNeR claims category – identified in step 1 of the IBNeR claims modeling – their ultimate FGU incurred loss is considered to be their last known FGU incurred loss multiplied by the smoothed development factor associated with their last known year of development. In addition, only original claims attached to reinsurance treaties with an equivalent in the program to which they belong for the accident year modeled, are included in the draw. The conditions of the equivalent reinsurance treaties are then applied to the ultimate FGU incurred losses drawn from the original claims. The incurred losses accepted by AXA SA and those borne by AXA SA then correspond to the IBNyR reserves.

Results

For each of the 1,000 scenarios, the IBNR reserves per accident year are obtained by summing the IBNeR and IBNyR reserves for each claim of the accident year in question. The distributions of IBNR reserves generated by these 1,000 scenarios – mainly driven by IBNyR reserves – are highly asymmetrical, whatever the accident year and the line of business. The graph in figure 4 below illustrates this asymmetry through the distribution function of simulated IBNR reserves for a line of business. The red vertical bar corresponds to the IBNR reserves calculated by the Reserving team.

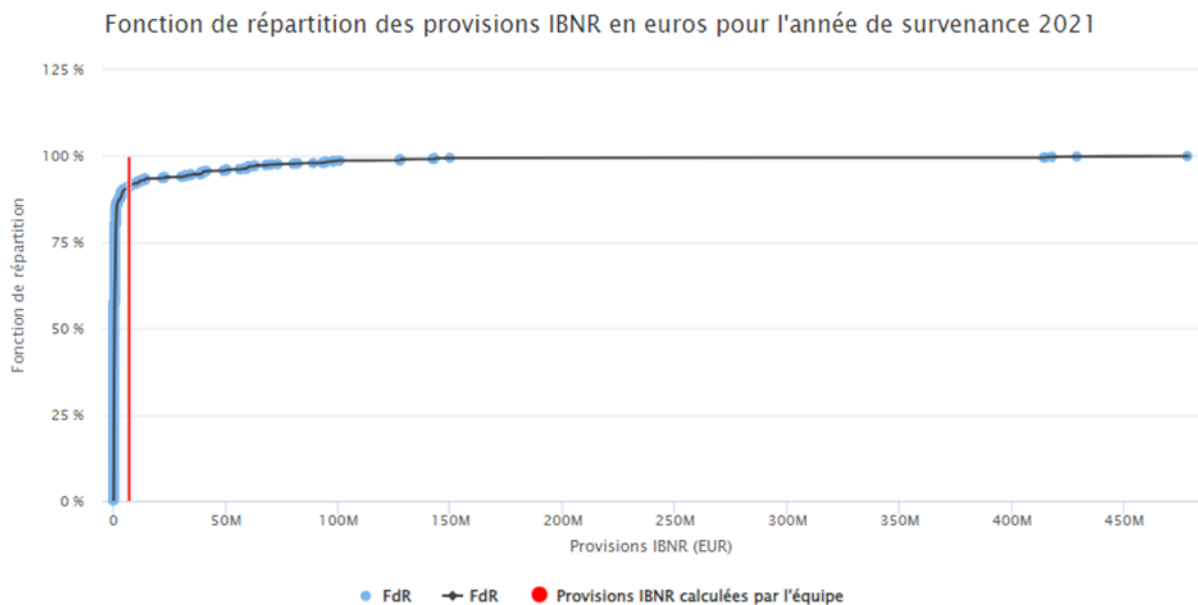


Figure 4 – A distribution function for IBNR reserves of a line of business

In this example, as in most of the accident years and most of the lines of business, the IBNR reserves calculated by the team are deemed to be overestimated and overly cautious, according to the model proposed in this master thesis. This observation is based on the associated percentiles, which generally reach 80% – 90%. Moreover, the difference between the three methods of estimating IBNeR reserves is small.

Furthermore, it is possible to look at scenarios generating IBNR reserves close to those calculated by the Reserving team, with the aim of enabling the latter to better justify its results.

Conclusion

The objective of interpretability of the IBNR reserves calculated by the Reserving team has been achieved. It is indeed possible to trace the construction of each IBNeR and IBNyR reserve

thanks to the individual claims information used throughout the proposed model, notably via the RShiny application developed during the internship. In addition, the results of this model make it possible to position and interpret the IBNR reserves calculated by the team in relation to those modeled in this master thesis, in order to assess their level of prudence. However, these results need to be put into perspective, as several biases are present in this model. Firstly, numerous assumptions were made throughout the model, and their impact would need to be analyzed. Secondly, reinstatements, AAD, AAL and indexed limits and deductible are not taken into account. Thirdly, when modeling IBNyR claims, only risks that have already occurred can be randomly drawn. Simulating ultimate FGU incurred losses of IBNyR claims according to a severity law could help improve the model.

In the longer term, a consolidated version of this model could lead the Reserving team to question the relevance of the methods it uses by comparing their results with those of the model proposed in this master thesis, and if necessary, adjust its aggregate methods to better reflect the total claims costs.

Remerciements

Je tiens tout d'abord à remercier l'entreprise AXA GIE d'avoir accepté ma candidature pour ce stage de fin d'études, ainsi que Monsieur Elie DADOUN pour l'accueil au sein de son équipe Reinsurance Actuarial Reserving faisant partie du département AXA Group Ceded Reinsurance. Je remercie à nouveau Monsieur Elie DADOUN pour la confiance qu'il m'a accordée tout au long de mon stage.

Je remercie tout particulièrement mon tuteur entreprise Monsieur Baptiste VOISIN pour le partage de ses compétences et connaissances, ses précieux conseils et sa très grande disponibilité. Son encadrement m'a été d'un grand support à chaque étape de mon stage et lors de la rédaction de mon mémoire.

Je tiens également à remercier tous ceux et toutes celles que j'ai eu l'occasion de côtoyer pendant ces six mois de stage, notamment les personnes de l'équipe Reserving et plus généralement, toutes les personnes d'AXA Group Ceded Reinsurance.

Je remercie l'ENSAE qui me permet de présenter ce mémoire à l'Institut des Actuaire. Je remercie également mon tuteur académique Monsieur Nicolas BARADEL de m'avoir suivie pendant toute la durée de mon stage et la rédaction de mon mémoire.

Je remercie évidemment ma famille et mes proches pour leur soutien précieux pendant la rédaction de ce mémoire ainsi que tout au long de mes études.

Table des matières

Note de synthèse	5
Executive summary	11
Remerciements	17
Introduction	21
1 La réassurance	23
1.1 Notions d'assurance et de réassurance	23
1.2 Les avantages de la réassurance	25
1.3 Les différentes formes de réassurance	26
1.4 Les différents types de couverture de réassurance	28
2 Problématique et revue bibliographique	37
2.1 Contexte et problématique	37
2.2 Revue de littérature	41
2.3 Apport du mémoire	47
3 Présentation et structuration des données	49
3.1 La présentation des données	49
3.2 Le nettoyage et le retraitement des données	53
3.3 La création de nouvelles variables	55
4 Modélisation des sinistres IBNeR	67
4.1 Étude des sinistres clos et actifs	68

4.2	Identification des sinistres IBNeR	74
4.3	Création de variables explicatives	90
4.4	Développement des Charges FGU à l'ultime	93
5	Modélisation des sinistres IBNyR	119
5.1	Estimation du nombre de sinistres IBNyR	120
5.2	Estimation des Charges FGU ultimes des sinistres IBNyR	125
6	Résultats de la modélisation	131
6.1	L'analyse et la comparaison des provisions IBNR	132
6.2	L'analyse des résultats de différents scénarios	136
	Conclusion	151
	Bibliographie	153
	Table des figures	155
A	Lois de probabilités	161
A.1	La loi uniforme discrète	161
A.2	La loi binomiale	161
A.3	La loi binomiale négative	162
A.4	La loi de Poisson	163
B	La méthode de la <i>Chain Ladder</i>	165
C	La méthode de Schnieper	169
D	Le critère d'information d'Akaike (AIC)	173
E	La méthode de Kaplan-Meier	175
F	La régression logistique	177
	Liste des sigles, acronymes et abréviations	179

Introduction

L'activité assurantielle se distingue des autres secteurs par l'inversion du cycle de production. Dans une entreprise industrielle, le cycle de production englobe l'ensemble des étapes requises pour transformer les matières premières en produits finis. Ce processus inclut l'approvisionnement en matières premières, la fabrication, le stockage, la distribution et la commercialisation des produits. Les coûts de production y sont établis avant la vente des produits.

Pour les secteurs de l'assurance et de la réassurance, ce cycle est inversé : les compagnies doivent provisionner des fonds avant de connaître le montant des prestations à verser à leurs clients. En effet, lors de la souscription, les compagnies demandent à leurs clients une prime d'assurance contre une prestation potentielle future qui dépend d'un événement aléatoire. Au moment où les clients achètent leur protection, les compagnies ne connaissent pas encore la hauteur des prestations qu'elles devront leur verser dans le cas où un événement, affectant les assurés, surviendrait. Les primes payées par les clients contribuent en partie à garantir les ressources nécessaires pour avoir la capacité d'honorer les engagements stipulés dans le contrat qui lie les compagnies à leurs clients. Les compagnies doivent donc estimer les coûts futurs des sinistres afin de provisionner un montant suffisant pour indemniser leurs futurs assurés sinistrés.

Pour estimer ces coûts à venir, appelés provisions, les équipes de provisionnement actuariel des compagnies font appel en grande majorité à des méthodes de provisionnement dites « classiques ». Ces méthodes utilisent des données, en l'occurrence des montants, de façon agrégée par groupes de risques homogènes. La force de telles approches réside dans la stabilité statistique offerte grâce à ces processus d'agrégation. En plus d'être faciles à mettre en œuvre et d'être visuellement déchiffrables, ces méthodes permettent d'identifier des tendances dans la sinistralité agrégée des sinistres, selon leurs années de développement.

Cependant, la réassurance a une spécificité par rapport à l'assurance : les sinistres sont en faible nombre et leurs montants peuvent varier grandement d'un sinistre à l'autre, même au sein de groupes homogènes. Certains sinistres très coûteux peuvent avoir un poids très lourd dans les méthodes classiques utilisant le processus d'agrégation et ainsi biaiser fortement les résultats de ces méthodes. Lorsqu'elles sont employées dans le cadre de la réassurance, les méthodes classiques avec l'agrégation des sinistres sont donc moins robustes et plus volatiles.

Pour nuancer les estimations, il est intéressant de s'attarder sur les informations propres à chaque sinistre. L'accès à ces informations n'est cependant pas réalisable à travers les méthodes classiques. En effet, une fois l'agrégation effectuée, les informations telles que le développement individuel des sinistres, ne sont plus récupérables. Il n'est donc pas possible de fournir une explication sinistre par sinistre de l'estimation des montants futurs calculés.

Les résultats fortement volatiles obtenus par ces méthodes classiques méritent d'être expliqués plus en détails. De surcroît, les provisions en réassurance peuvent être jugées prudentes lorsqu'elles sont calculées par ces méthodes car elles semblent excessives au regard de la sinistralité constatée à l'instant de l'estimation. Générer des scénarios de sinistralité permet alors d'interpréter le montant de ces provisions. Cela donne la possibilité aux équipes de provisionnement en réassurance de mieux communiquer et expliquer leurs résultats.

Les travaux présentés dans ce mémoire se penchent sur le sujet d'interprétabilité des résultats des méthodes de provisionnement classiques. L'approche mise en œuvre est celle du provisionnement individuel, appelé aussi provisionnement ligne à ligne ou sinistre par sinistre, afin de tirer parti des données individuelles des sinistres impliqués dans le processus de réassurance.

Parmi les différents types de provisions à estimer, les provisions pour les sinistres IBNR (*Incurred But Not Reported*) sont les seules considérées dans le cadre de ce mémoire. Ces sinistres IBNR se divisent en deux catégories :

- Les sinistres IBNeR (*Incurred But Not enough Reserved*) : ce sont les sinistres qui ont été déclarés à la réassurance mais dont le montant peut évoluer ultérieurement.
- Les sinistres IBNyR (*Incurred But Not yet Reported*) : ce sont les sinistres survenus mais non déclarés à la réassurance. Cette dernière ne connaît ni leur existence, ni leur montant.

La génération de plusieurs scénarios de développement est réalisée pour les sinistres IBNeR et IBNyR. Elle permet d'obtenir différentes projections de leurs montants et donc différentes possibilités de consommation des provisions IBNR calculées en amont de ces travaux par des méthodes de provisionnement classiques.

Ce mémoire s'articule autour de six chapitres¹. Le premier chapitre passe en revue des notions de réassurance utiles à sa compréhension. Le deuxième chapitre porte sur la problématique ainsi que sur la bibliographie disponible sur le sujet du provisionnement individuel. Le troisième chapitre présente les bases de données utilisées ainsi que la création de variables utiles à la modélisation. Les quatrième et cinquième chapitres abordent respectivement la modélisation des développements des sinistres IBNeR et celle des sinistres IBNyR. Le sixième chapitre établit le lien entre les résultats obtenus avec les méthodes de provisionnement individuel et les provisions IBNR calculées par des méthodes de provisionnement classiques.

1. À titre d'information, les chapitres sont divisés en sections, elles-mêmes subdivisées en sous-sections, pouvant à leur tour contenir des sous-sous-sections.

Chapitre 1

La réassurance

Ce premier chapitre est consacré à l'explication des principes fondamentaux de l'assurance et de la réassurance, nécessaires pour bien appréhender ce mémoire.

1.1 Notions d'assurance et de réassurance

1.1.1 L'assurance

Un assuré peut vouloir souscrire un contrat d'assurance avec un assureur afin de se couvrir en cas de survenance aléatoire d'un événement défavorable.

Le contrat d'assurance est un contrat dans lequel deux parties s'engagent : l'assuré, qui cherche à être protégé de certains risques, et l'assureur. L'assuré est alors couvert par rapport aux risques, qui font partie des garanties déterminées dans le contrat, s'ils venaient à se réaliser. L'assureur assume quant à lui les risques que l'assuré ne veut pas conserver en lui garantissant des prestations si l'événement en question survient. En contrepartie, l'assureur reçoit une somme d'argent de la part de l'assuré : la **prime d'assurance**.

L'assuré peut être une personne physique ou une personne morale, qui ne pratique pas d'activité d'assurance.

L'assureur est représenté par une entité sociale d'assurance. Cette dernière peut prendre la forme de différents organismes d'assurance :

- Les sociétés d'assurance ou les sociétés d'assurance mutuelles qui sont régies par le Code des Assurances.
- Les mutuelles qui relèvent du Code de la Mutualité.
- Les instituts de prévoyance qui sont réglementés par le Code de la Sécurité Sociale.

AXA fait partie de la catégorie des sociétés d'assurance. Le Code des Assurances est donc pris comme référence dans la suite de ce mémoire et seules les compagnies d'assurance seront mentionnées dorénavant.

Certains assureurs interviennent dans le secteur de l'assurance vie, d'autres dans le secteur de l'assurance **IARD** (Incendies, Accidents, Risques Divers). Les contrats d'assurance IARD ne dépendent pas de la durée de vie humaine, contrairement aux contrats d'assurance vie. L'assurance IARD est d'ailleurs dénommée *P&C* pour *Property and Casualty* en anglais (*Property* pour Dommages Aux Biens et *Casualty* pour Responsabilité Civile). Une entité sociale d'assurance ne peut pas exercer dans ces deux branches à la fois. Elles doivent être séparées mais peuvent tout de même appartenir à un même groupe capitalistique.

1.1.2 La réassurance

Cette deuxième sous-section est axée sur la réassurance, thème central de ce mémoire.

Un assureur peut également vouloir se couvrir contre les risques qu'il a acceptés auprès de ses assurés car ces risques restent toujours imprévisibles, même en tentant de prédire leur survenance par des méthodes mathématiques. Cet assureur peut alors faire appel à un autre assureur afin de lui céder certains de ses risques. L'assureur cédant ses risques est appelé la **cédante**. L'assureur acceptant les risques est appelé le **réassureur**. Ainsi, la réassurance peut être définie comme l'assurance des assureurs. Plus précisément, c'est l'assurance d'une partie des risques assumés par l'assureur. D'après l'article L310-1-1 du Code des Assurances, « *la réassurance est l'activité d'un organisme [...] qui consiste à accepter des risques cédés, soit par une entreprise d'assurance [ou] une autre entreprise de réassurance [...]* ». L'une des principales différences avec l'assurance réside d'ailleurs dans la nature des contreparties. Tout comme l'assuré envers l'assureur, la cédante doit moyenner une prime au réassureur, appelée **prime de réassurance**.

Certaines compagnies font uniquement de la réassurance et sont appelées compagnies de réassurance. Ces compagnies peuvent intervenir à la fois au niveau de l'assurance vie et de l'assurance IARD, contrairement aux entités sociales d'assurance. En France, le nombre de compagnies de réassurance est plus de dix fois inférieur au nombre de compagnies d'assurance, toutes branches confondues ([France Assureurs, 2024](#)).

Plusieurs assureurs peuvent mutualiser et diversifier leurs risques au sein de *Pools* d'assurance. C'est l'un des principes de la coassurance. La différence avec la réassurance repose sur le fait que les résultats des *Pools* sont partagés proportionnellement à la part de prime que chacun d'entre eux apporte. La mutualisation consiste à considérer des risques homogènes dans leur ensemble dont leurs pertes devraient se compenser selon la loi des grands nombres.

1.1.3 La rétrocession

Un réassureur peut lui aussi se réassurer auprès d'un autre réassureur. C'est le principe de rétrocession. Le réassureur cédant une partie de ses risques qu'il ne veut pas assumer est appelé la **rétrocédante**. Le réassureur acceptant les risques est appelé le **rétrocessionnaire**. Le principe de la rétrocession est le même que celui de la réassurance.

Le schéma en figure 1.1 récapitule les différents acteurs cités et les liens entre eux.

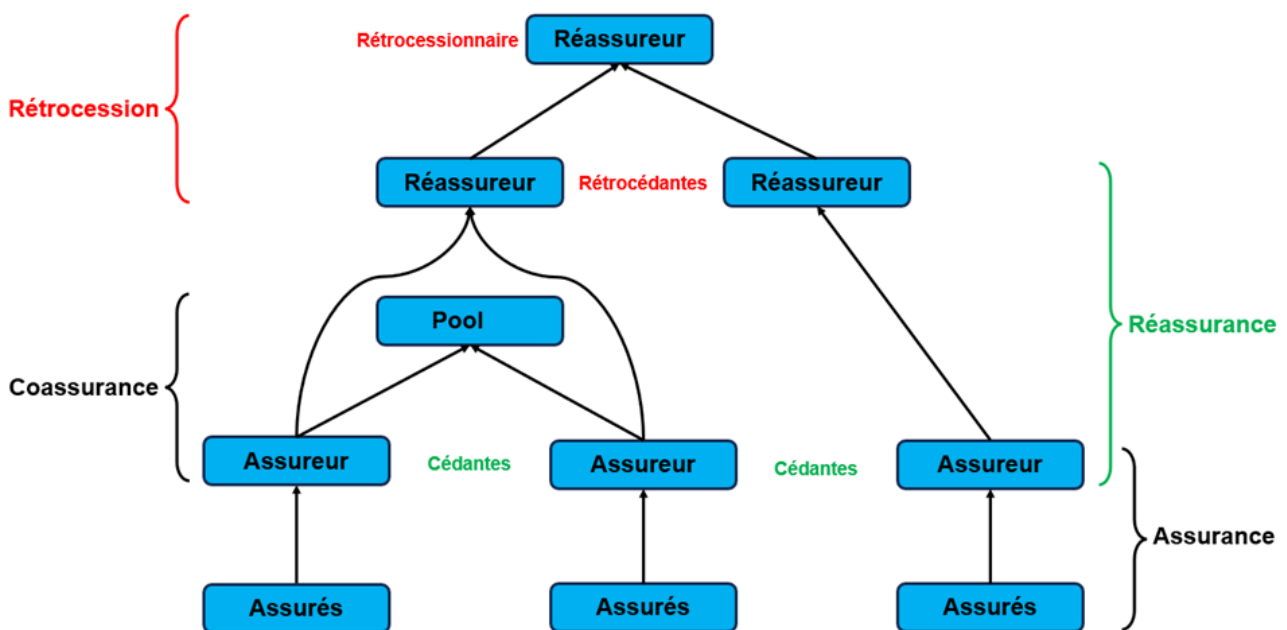


FIGURE 1.1 – Les différents acteurs de la réassurance et leurs liens

1.2 Les avantages de la réassurance

Dans cette section, l'analyse est portée sur la réassurance mais l'analogie peut être faite avec la rétrocession.

Cette section parcourt les différents avantages de faire appel à la réassurance.

L'objectif premier de la réassurance, du point de vue des cédantes, est de rester solvable en cas de sinistres de grande ampleur pour lesquels elles ont une exposition au risque élevée. Les réassureurs acceptent un large éventail de risques dont les risques les plus importants et les plus complexes.

Les compagnies de réassurance peuvent donc être considérées comme des industriels du risque qui se nourrissent de l'aléa et qui créent de la régularité à partir de chocs et d'évé-

nements extrêmes. Elles sont beaucoup plus diversifiées en termes d'activités, ont un cadre international beaucoup plus élargi et ont donc une taille beaucoup plus importante que les compagnies d'assurance.

Il existe deux grands types de risques contre lesquels les assureurs souhaitent se protéger :

- Le **risque de fréquence** : C'est le risque qu'un nombre élevé anormal de sinistres surviennent pendant la même période. Chacun de ces sinistres peut ne pas avoir un montant excessivement conséquent mais la somme de tous ces montants serait considérable.
- Le **risque de gravité** : C'est le risque qu'un sinistre unique et rare se révèle avoir un montant très élevé. Il peut par exemple s'agir d'un sinistre bicentenaire, censé survenir en moyenne tous les 200 ans et pouvant mettre en péril la cédante à lui seul.

Les cédantes peuvent aussi faire appel à des réassureurs pour d'autres raisons :

- Les engagements nets sur les risques acceptés par les cédantes sont réduits. L'exposition maximale est éliminée et l'homogénéité de leur portefeuille est améliorée, avec pour conséquence une volatilité des risques diminuée. Cela implique également la diminution des besoins en capital des cédantes. La réassurance peut donc être vue comme un outil de gestion financière du capital des assureurs permettant d'améliorer leur marge de solvabilité et de protéger leurs fonds propres. Les cédantes peuvent ainsi souscrire des risques plus importants que leur capital ne le permettrait sans réassurance. Elles peuvent même être plus compétitives sur le marché en vendant des protections à leurs clients à des prix plus faibles.
- Avoir recours à la réassurance permet aux cédantes de capter l'expertise des réassureurs. L'intérêt de capter cette expertise demeure notamment lorsque les cédantes proposent à leurs assurés de couvrir de nouveaux risques et qu'elles n'ont pas d'historique. Elles s'appuient sur les connaissances du réassureur relatives à ces risques pour appréhender au mieux ces nouveaux risques.

1.3 Les différentes formes de réassurance

Plusieurs formes de réassurance existent sur le marché : la réassurance obligatoire, la réassurance facultative et la réassurance facultative-obligatoire. Elles sont présentées dans les trois prochaines sous-sections.

1.3.1 La réassurance obligatoire

Pour cette forme de réassurance, il y a une notion d'obligation réciproque de céder et d'accepter toutes les polices qui entrent dans le cadre prédéfini. En effet, la cédante est obligée de céder tous les risques de son portefeuille qui rentrent dans les limites temporelles et géographiques et qui relèvent des types de garanties entrant dans le cadre préétabli. Le réassureur est obligé de les accepter et de les couvrir. Le contrat signé, dans lequel le réassureur s'engage à honorer ses engagements envers l'assureur, est alors appelé un **traité de réassurance**. Le traité peut également mentionner des exclusions.

Cette forme de réassurance couvre la majorité des portefeuilles, représentant environ 85% des primes de réassurance.

L'avantage de cette forme de réassurance est la simplicité de la gestion des traités grâce aux conditions définies à l'avance.

1.3.2 La réassurance facultative

L'accord entre la cédante et le réassureur s'établit pour chaque risque. La cédante peut présenter (ou non) un cas spécifique de risque qu'elle veut céder au réassureur. Ce dernier peut accepter ou décliner. La tarification se fait de manière individuelle.

Cette forme de réassurance est réservée aux cas particuliers, en termes de montant ou de nature, exclus du cadre de la réassurance obligatoire.

Ses principaux désavantages sont la lourdeur de la gestion et le coût de souscription plus élevé puisque tout est géré au cas par cas.

1.3.3 La réassurance facultative-obligatoire

Cette forme de réassurance, surnommée « Fac-Ob », est très minoritaire sur le marché car il y a un déséquilibre en défaveur du réassureur. En effet, la cession des risques est facultative pour l'assureur mais l'acceptation des risques est obligatoire pour le réassureur, selon des conditions spécifiées au préalable. L'offre est limitée à des cas où le réassureur a pleine confiance en la souscription et la cession de l'assureur. Elle permet cependant de simplifier la gestion.

Le schéma en figure 1.2 ci-dessous récapitule le fonctionnement de ces trois formes de réassurance.

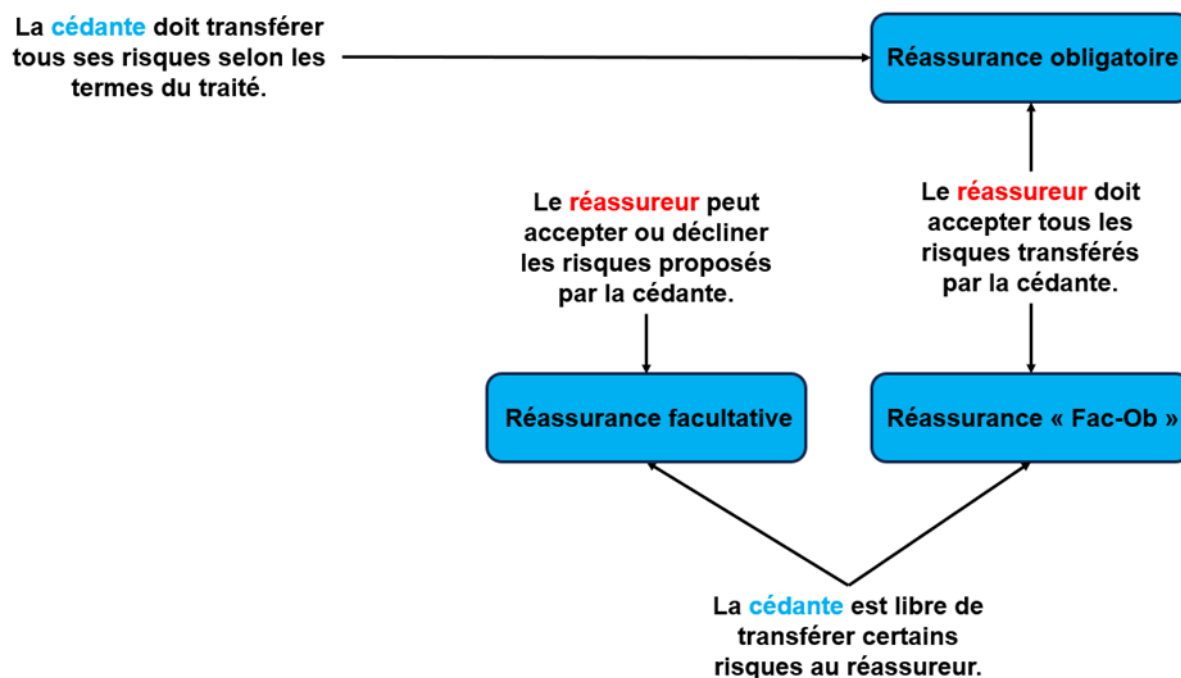


FIGURE 1.2 – Le fonctionnement des trois formes de réassurance

1.4 Les différents types de couverture de réassurance

Les risques peuvent être cédés de différentes manières selon la couverture définie dans les traités de réassurance. Il existe deux types de couvertures de réassurance : les couvertures proportionnelles et celles non proportionnelles. Leur différence réside dans la manière de céder le risque. Les trois sous-sections suivantes les détaillent.

La différence notoire est que les couvertures proportionnelles sont basées sur une répartition des primes et des sinistres entre l'assureur et le réassureur, suivant un même facteur de proportionnalité, tandis que les couvertures non proportionnelles sont basées sur des seuils et des limites de sinistralité avec une prime de réassurance fixée. En termes de vocabulaire, le montant du sinistre conservé par la cédante pour son compte propre est appelé la **réfaction**. Ce terme sera utilisé à plusieurs reprises dans les prochaines sous-sections.

1.4.1 Les couvertures proportionnelles

Dans ce type de couverture, une proportion de l'exposition, donc du risque, transférée au réassureur est prédéfinie. La cédante et le réassureur partagent proportionnellement les pertes si un sinistre survient. En contrepartie, la même proportion des primes revient au réassureur.

Ainsi :

$$\frac{\text{Prime de réassurance}}{\text{Primes totales reçues par la cédante}} = \frac{\text{Montant des sinistres à la charge du réassureur}}{\text{Montant brut des sinistres à la charge de la cédante}}$$

Lorsqu'il est mentionné « brut », cela signifie que le montant est **brut de réassurance**. C'est le montant du sinistre avant l'intervention de la réassurance, c'est-à-dire le montant à la hauteur duquel le sinistré sera indemnisé. Le montant **net de réassurance** est quant à lui le montant restant à la charge de la cédante après l'intervention de la réassurance.

Le tableau en figure 1.3 récapitule ce qui reste du côté de la cédante (la rétention) et ce qu'elle cède (la cession) avec α , la proportion définie dans le traité.

	Rétention	Cession
Primes	$\alpha \sum \text{Primes}$	$(1 - \alpha) \sum \text{Primes}$
Sinistres	$\alpha \sum \text{Montants des sinistres}$	$(1 - \alpha) \sum \text{Montants des sinistres}$

FIGURE 1.3 – Tableau des primes et montants de sinistres dans les cas de la rétention et de la cession

Il existe deux mécanismes principaux de couvertures proportionnelles. Leur différence réside dans le taux de cession des risques.

Pour compenser les frais de gestion des dossiers de la cédante, le réassureur verse *a posteriori* à cette dernière une commission de réassurance, qui est un pourcentage des primes cédées, soit fixe, soit variable en fonction de la sinistralité. Cette commission doit respecter au global une proportionnalité du résultat technique.

Le Quote-Part (QP)

La proportion du risque et de la prime transférés est définie au niveau du portefeuille global de la cédante. Le taux de cession est identique pour tous les risques.

La convention veut que l'appellation « Quote-Part $x\%$ » soit utilisée lorsque la cédante transfère $x\%$ de tous ses risques afférents au traité en question.

Cette couverture intervient dès qu'un montant de sinistre à indemniser est positif, autrement dit, dès le premier euro. Ce montant total du sinistre, après franchise mais avant toute rétention ou réassurance, se nomme **FGU**, pour « *From Ground Up* », expression anglaise signifiant « à partir de zéro ».

Une telle couverture ne modifie pas la structure de l'exposition aux risques de la cédante, seulement la hauteur de ses engagements. Aussi bien les risques faibles que les risques importants

sont transférés au réassureur.

Une couverture en QP a l'avantage d'être simple en termes de gestion. En contrepartie, une grande partie des primes peut être cédée au réassureur, dans le cas où la cédante souhaite être bien protégée, limitant alors sa rentabilité. De plus, ce type de couverture ne permet pas à l'assureur de plafonner ses pertes maximales.

Le schéma en figure 1.4 représente les engagements portés par la cédante et le réassureur dans le cadre d'un traité QP 60%.

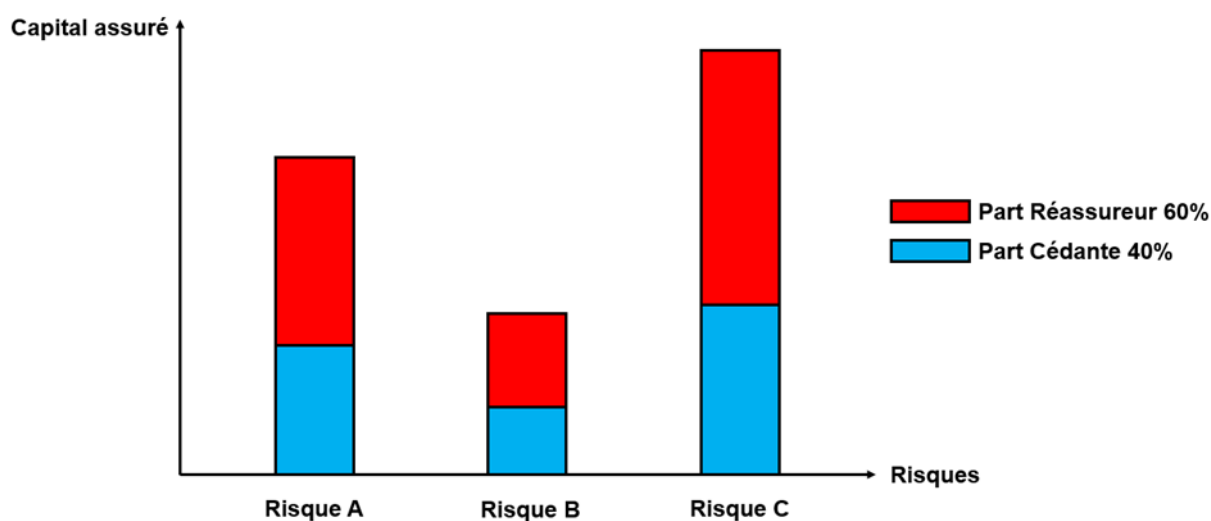


FIGURE 1.4 – Exemple d'un traité de réassurance QP

L'Excédent de Plein (EP)

La proportion du risque et de la prime transférés est définie au niveau de la police d'assurance lors de la souscription, sur la base d'un montant de capitaux maximum conservés par la cédante. Le taux de cession varie d'un risque à l'autre, selon entre autres, le capital assuré.

Le montant de capitaux maximum conservés par la cédante est le niveau à partir duquel intervient le réassureur, appelé **plein de rétention**. Quel que soit le risque, ce montant reste constant. Il correspond donc à l'exposition maximale que l'assureur conserve sur chaque risque. La partie excédant le plein de rétention est transférée au réassureur. Tous les risques inférieurs au plein de rétention ne sont donc pas réassurés.

Cependant, dans le traité signé par les deux parties, peut être mentionnée une **capacité** maximale cédée au réassureur. Dans ce cas-là, les pertes maximales de ce dernier lui sont connues et la cédante conserve une exposition plus élevée que son plein de rétention défini au traité. La plupart du temps, cette capacité est un multiple du plein de rétention.

Dans le cadre de l'EP, le taux de cession EP α_{EP} est calculé individuellement comme suit :

- Avec une capacité :

$$\alpha_{EP} = \frac{\max\{0; \min(\text{Capitaux assurés}; \text{Plein de rétention} + \text{Capacité}) - \text{Plein de rétention}\}}{\text{Capitaux assurés}}$$

- Sans capacité :

$$\alpha_{EP} = \frac{\max\{0; \text{Capitaux assurés} - \text{Plein de rétention}\}}{\text{Capitaux assurés}}$$

Pour un traité ayant une couverture en EP, le partage du risque entre la cédante et le réassureur n'est pas total puisque le réassureur intervient seulement lorsque les capitaux assurés dépassent le plein de rétention. De plus, ces traités ne protègent pas la cédante contre les risques de fréquence.

Le schéma en figure 1.5 représente un exemple de traité en EP avec les taux de cession pour chaque risque assuré.

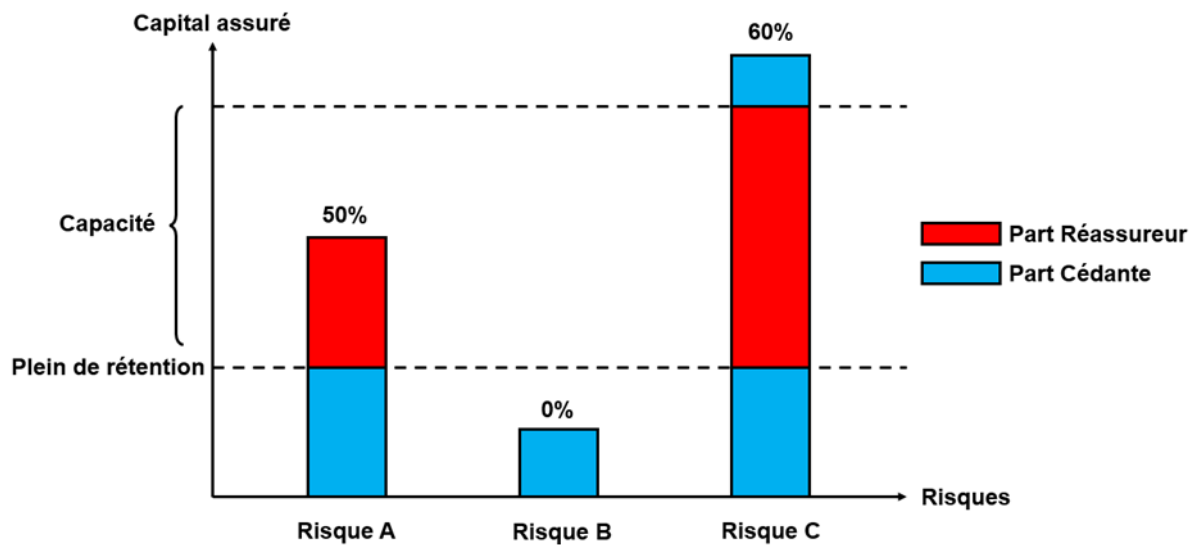


FIGURE 1.5 – Exemple de traité de réassurance en EP

1.4.2 Les couvertures non proportionnelles

Le critère des couvertures non proportionnelles repose sur la répartition de la sinistralité entre celle conservée et celle cédée par l'assureur. Le montant cédé à la réassurance est fonction du montant du sinistre et non de l'exposition de la cédante. Il n'y a d'ailleurs aucune relation

entre le montant payé par le réassureur et les primes de réassurance qu'il a perçues de la cédante. Les primes sont cependant généralement moins élevées que dans le cadre d'une couverture proportionnelle. Le nom de ces couvertures vient du fait que les engagements de la cédante et du réassureur ne sont pas proportionnels.

Dans les traités ayant recours à ces couvertures, le réassureur n'intervient que lorsque le montant du sinistre, exprimé en termes de pertes brutes totales supportées par l'assureur (c'est-à-dire en vision *From Ground Up* (FGU), au premier euro), dépasse un certain seuil, appelé **priorité**. La cédante doit donc supporter le sinistre en deçà de cette priorité. Si le sinistre est inférieur à la priorité, alors il n'est pas réassuré.

L'engagement maximum du réassureur sur chaque sinistre est appelé la **portée**. Elle peut être limitée ou illimitée. Ce deuxième cas est très rare car le réassureur s'expose potentiellement à un sinistre de grande ampleur.

La somme de la priorité et de la portée est appelée la **limite**. Au-delà de cette limite, la cédante n'est plus réassurée.

La priorité et la portée caractérisent les traités à couverture non proportionnelle. La sinistralité cédée $S_{cédée}$ est définie comme suit :

$$S_{cédée} = \min [\max \{ Perte - Priorité ; 0 \} ; Portée]$$

Le schéma en figure 1.6 représente un exemple de traité à couverture non proportionnelle.

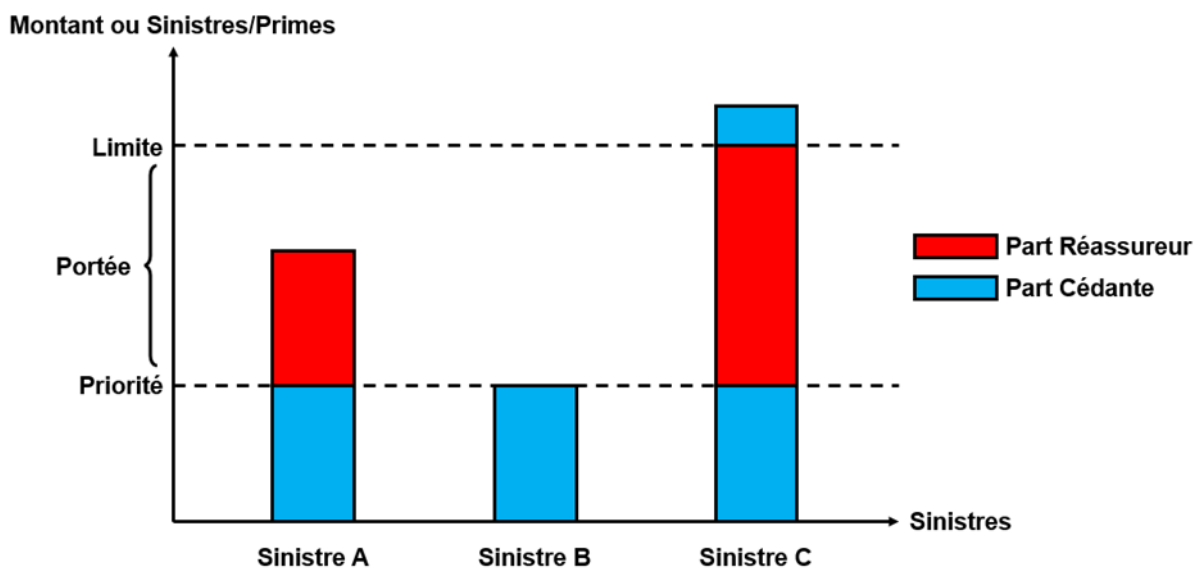


FIGURE 1.6 – Exemple de traité de réassurance à couverture non proportionnelle

Parfois, plusieurs traités peuvent être rassemblés au sein d'un programme de réassurance. Chaque traité représente alors une **tranche** d'un **programme de réassurance**. La plupart du temps, la limite d'une tranche coïncide avec la priorité d'une autre, comme dans le schéma du programme de réassurance en figure 1.7 ci-dessous. La « portée totale » est la plus grande limite de ces tranches, diminuée de la rétention. Les primes de réassurance des tranches diminuent à mesure que la priorité augmente. Ceci est dû au fait que les sinistres touchent de moins en moins les tranches au fur et à mesure que la priorité de ces dernières augmente. « **Toucher** » signifie avoir un montant de sinistre supérieur ou égal à la priorité définie pour une tranche de réassurance. Un sinistre qui ne dépasse pas la priorité de la tranche est considéré comme ne touchant pas la tranche. L'intérêt d'un programme est de pouvoir minimiser les coûts de réassurance en choisissant le meilleur réassureur pour chacune de ses tranches, après analyse de leur offre en termes de primes proposées.

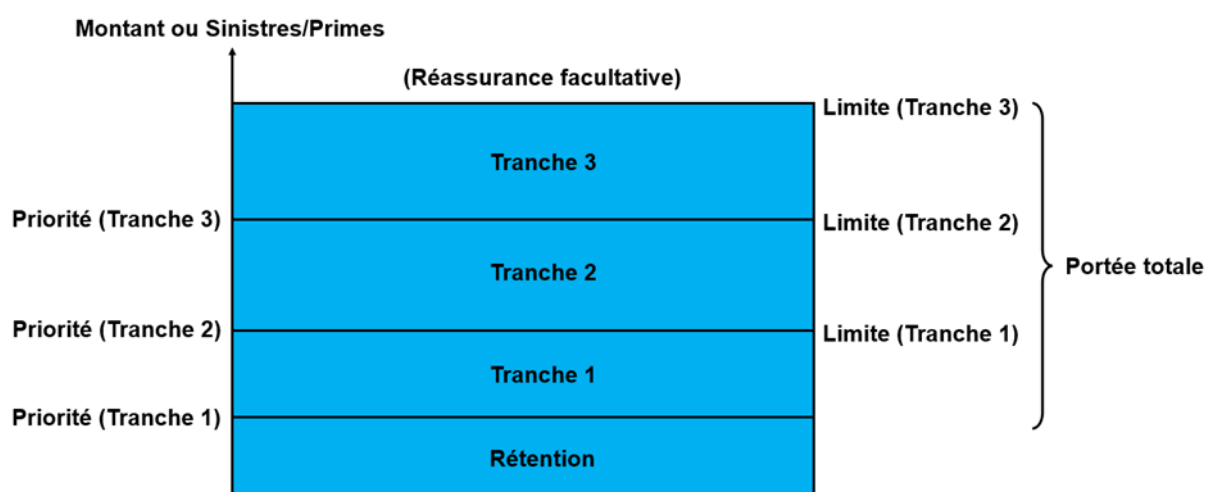


FIGURE 1.7 – Exemple d'un programme de réassurance

Il existe trois différents types de traités non proportionnels : les traités en Excédent de Sinistres par risque, les traités en Excédent de Sinistres par événement et les traités en Excédent de perte annuelle.

Excédent de Sinistres par risque

Le critère est par risque. Chaque risque correspond à un objet physique, un bâtiment ou un ensemble de bâtiments très proches les uns des autres par exemple. Ce type de couverture permet de limiter les pertes de l'assureur sur chaque sinistre dépassant la priorité, en cédant une partie de ses pertes à la réassurance jusqu'à la limite. La cédante est alors protégée des sinistres importants, donc du risque de gravité, tout en conservant une part plus élevée de

primes. L'inconvénient est qu'elle n'est pas couverte contre une accumulation de petits sinistres qui ne touchent pas le traité, ni de sinistres très conséquents.

Excédent de Sinistres par événement

Le critère est par événement. Un événement correspond à un fait générateur provoquant plusieurs sinistres au titre duquel on agrège tous ces derniers. Il peut se rapporter à une catastrophe naturelle ou à une catastrophe générée par l'Homme, dite *Man-Made*. Un événement est défini au traité dans sa nature, dans l'espace et dans le temps. C'est le montant de l'agrégation qui compte pour savoir si le traité est touché ou non. La cédante est alors protégée contre une accumulation de petits sinistres tous liés au même événement, c'est-à-dire du risque de fréquence.

Le schéma en figure 1.8 ci-dessous explique le fonctionnement du traité en Excédent de Sinistres par événement. Les sinistres A, B et C ont été causés par le même événement. Si le traité était un Excédent de Sinistres par risque, seule la partie jaune du sinistre A à gauche aurait été réassurée. Avec une couverture en Excédent de Sinistres par événement, la part réassurée est beaucoup plus importante.

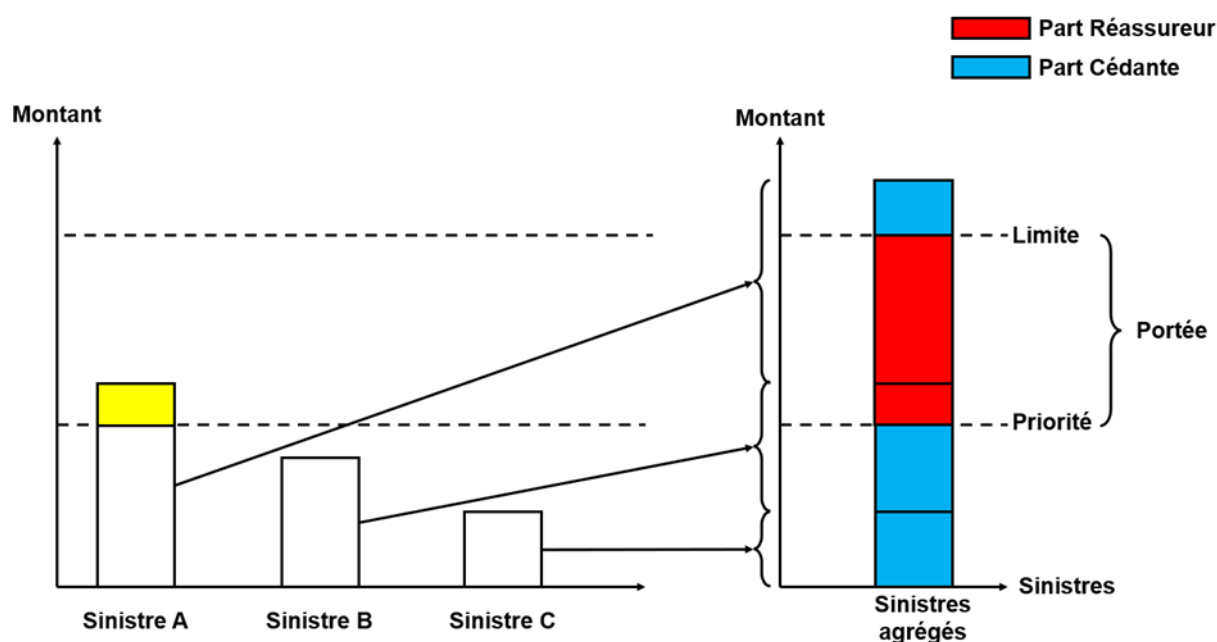


FIGURE 1.8 – Exemple d'un traité de réassurance en Excédent de Sinistres par événement

La notation habituelle pour la définition de ces couvertures par risque et par événement est $a \text{ XS } b$ avec :

- a : la portée ;

- b : la priorité ;
- XS : le nom de la couverture dans le cadre du traité, soit Excédent de Sinistres (XS).

Là encore, la cédante peut structurer sa couverture avec plusieurs tranches différentes.

Excédent de perte annuelle ou *Stop-Loss* (SL)

Le critère est par sinistralité annuelle. Dans la couverture en SL, la priorité et la portée ne sont pas définies en montant de sinistres mais en pourcentage d'assiette de primes des portefeuilles réassurés. Le pourcentage, associé à la cédante souhaitant se protéger avec cette couverture en année t , correspond au rapport S/P suivant :

$$S/P_t = \frac{\sum_{j=1}^{m_t} S_{j,t}}{\sum_{k=1}^{n_t} P_{k,t}}$$

Avec :

- $S_{j,t}$: le montant du $j^{\text{ème}}$ sinistre de la cédante pour l'année t ;
- $P_{k,t}$: la $k^{\text{ème}}$ prime de la cédante pour l'année t ;
- m_t : le nombre total de sinistres pour la cédante pendant l'année t ;
- n_t : le nombre total de primes pour la cédante pendant l'année t .

La réassurance intervient de même, lorsque le rapport S/P dépasse la priorité.

Ce type de couverture permet de garantir le résultat global de l'assureur même si une importante sinistralité annuelle advient.

Les traités en Excédent de perte annuelle couvrent d'abord un risque de fréquence mais peuvent aussi couvrir un risque de gravité.

Le traité SL peut intervenir avant toute couverture de réassurance ou après plusieurs couvertures de réassurance.

Les clauses particulières

Il existe des clauses qui peuvent être mentionnées dans un traité non proportionnel :

- La limite annuelle agrégée (*Annual Aggregate Limit* (AAL)) : Afin de limiter son engagement, le réassureur peut déterminer un montant maximum de sinistres cumulés sur une année au-delà duquel il n'interviendra plus du tout.
- La franchise annuelle agrégée (*Annual Aggregate Deductible* (AAD)) : Sur une année, le réassureur interviendra auprès de la cédante seulement lorsque la somme des sinistres à

réassurer sera supérieure à la franchise déterminée au préalable. Cela permet de réduire la prime de réassurance, en diminuant la charge couverte par la réassurance.

- Les reconstitutions : Il s’agit d’un principe de reconduction-renouvellement de la portée unitaire lorsqu’elle a été consommée dans son entièreté. En effet, la portée est une garantie annuelle. Une fois consommée entièrement, de potentielles futures pertes ne seraient alors pas couvertes par ce traité. Il existe cependant des clauses de reconstitutions permettant de reconstituer cette portée. Ces reconstitutions peuvent être conditionnées par le paiement d’une prime additionnelle, appelée **prime de reconstitution**. Cette dernière est notée P_{rec} et est calculée comme suit :

$$P_{rec} = P_{init} \times \beta_{Portée} \times \alpha_{rec}$$

Avec :

- P_{init} : la prime de réassurance initiale ;
- $\beta_{Portée}$: la proportion de consommation de la portée ;
- α_{rec} : le taux de reconstitution prédéfini.

Cette prime est déclenchée dès que la portée est impactée, même si elle n’est pas consommée entièrement. Le taux de reconstitution prédéfini est souvent de 100% et le nombre de primes de reconstitutions, donc de reconstitutions possibles, varie généralement entre 1 et 2. Il existe le lien suivant entre les reconstitutions et l’AAL :

$$AAL = (1 + n_{rec}) \times Portée$$

Avec n_{rec} : le nombre de reconstitutions.

Chapitre 2

Problématique et revue bibliographique associée

Ce chapitre présente le contexte dans lequel a été élaboré le sujet ainsi que la problématique qui en découle. Un état de l’art d’articles scientifiques et de mémoires d’actuariat traitant du provisionnement ligne à ligne, est exposé par la suite avec les principales idées de ces travaux.

2.1 Contexte et problématique

D’après le *Guide de provisionnement des sinistres en assurance non-vie* de l’Institut des Actuaires ([Groupe de Travail sur le Provisionnement des Sinistres Non-Vie, 2023](#)), « les **provisions techniques** sont l’ensemble des provisions et des réserves constituées par l’assureur pour faire face à ses engagements vis-à-vis des assurés et des bénéficiaires à la date d’inventaire. Elles sont inscrites au passif du bilan de la compagnie. »

Ces provisions doivent être estimées le plus justement possible pour répondre à certaines normes et réglementations de plusieurs organismes ou textes, tels que l’Autorité de Contrôle Prudentiel et de Résolution (ACPR) ou le Code des Assurances. Des audits externes mais aussi internes à AXA sont effectués régulièrement. D’après l’article 84 de la directive cadre de l’Union Européenne ([Parlement Européen et Conseil de l’Union Européenne, 2009](#)) : « Article 84 : Caractère approprié du niveau des provisions techniques – Sur demande des autorités de contrôle, les entreprises d’assurance et de réassurance démontrent le caractère approprié du niveau de leurs provisions techniques, ainsi que l’applicabilité et la pertinence des méthodes qu’elles appliquent et l’adéquation des données statistiques sous-jacentes qu’elles utilisent. »

Il existe plusieurs types de provisions techniques, dont les Provisions pour Sinistres À Payer (PSAP). Le Code des Assurances, dans son article R343-7, définit les PSAP comme

la « *valeur estimative des dépenses en principal et en frais, tant internes qu’externes, nécessaires au règlement de tous les sinistres survenus et non payés [...]* ». Les PSAP sont donc élaborées pour prendre en considération la possibilité que l’indemnisation de certains sinistres puisse s’échelonner sur plusieurs années, notamment lorsque les montants sont importants. Les PSAP sont composées des provisions dossiers/dossiers, appelées **provisions D/D**, et des provisions pour sinistres IBNR (*Incurring But Not Reported*, signifiant survenus mais non déclarés), appelées **provisions IBNR**. Ces deux sous-catégories de provisions sont estimées par des équipes actuarielles de provisionnement, comme l’équipe Reinsurance Actuarial Reserving que j’ai intégrée lors de mon stage de 6 mois au sein d’AXA GIE. Cette équipe, dénommée plus simplement « Reserving », fait partie du département AXA Group Ceded Reinsurance, dénommé quant à lui « Ceded Re » par la suite et qui est le réassureur interne du Groupe AXA. La réassurance est particulièrement sujette à ces PSAP car elle couvre surtout des sinistres à montants très élevés.

D’après le *Guide de provisionnement des sinistres en assurance non-vie* ([Groupe de Travail sur le Provisionnement des Sinistres Non-Vie, 2023](#)), les provisions D/D « *correspondent aux charges à prévoir pour faire face à la sinistralité déclarée mais non encore réglée et à la sinistralité non encore déclarée mais prévisible des contrats en cours* ». Ces provisions sont établies pour chaque sinistre déclaré. Elles représentent le montant qui n’a pas encore été indemnisé à l’assuré par rapport au total de l’indemnisation qui lui est due. Dans le cadre de ce mémoire, les provisions D/D sont considérées comme connues et ne nécessitent pas d’estimation.

Les estimations des provisions IBNR sont quant à elles le sujet principal de ce mémoire. Elles sont constituées de deux types de provisions :

- Les provisions IBNeR (*Incurring But Not enough Reserved*) : Elles se réfèrent aux provisions pour les sinistres IBNeR, c’est-à-dire les « *sinistres survenus et déclarés à la date de clôture des états financiers* » ([Groupe de Travail sur le Provisionnement des Sinistres Non-Vie, 2023](#)) mais qui ne sont pas suffisamment provisionnés.
- Les provisions IBNyR (*Incurring But Not yet Reported*) : Elles correspondent au « *coût ultime des sinistres survenus mais non encore déclarés à la date de clôture* » ([Groupe de Travail sur le Provisionnement des Sinistres Non-Vie, 2023](#)). Ces derniers sont appelés sinistres IBNyR et sont les « vrais » sinistres IBNR.

Les provisions IBNR sont calculées par des méthodes de provisionnement classiques qui s’appuient, pour la majorité, sur le principe d’agrégation de données homogènes. Ces données homogènes sont réunies dans des groupes, appelés *segments de provisionnement*. Ces regroupements sont d’ailleurs obligatoires et réglementés par l’article 80 de la Directive 2009/138 ([Parlement Européen et Conseil de l’Union Européenne, 2009](#)) : « *Article 80 : Segmentation – Lorsqu’elles calculent leurs provisions techniques, les entreprises d’assurance et de réassurance*

segmentent leurs engagements d'assurance et de réassurance en groupes de risques homogènes et, au minimum, par ligne d'activité. ». En effet, le comportement des sinistres d'un segment à l'autre peut différer.

Ces méthodes classiques permettent d'identifier les tendances dans l'évolution du montant des sinistres selon leur année de développement de manière très stable et de manière très visuelle grâce à l'utilisation de triangles de développement. L'équipe Reserving a recours à ces méthodes, telles que celle de la *Chain Ladder*, de Bornhuetter-Ferguson ou encore de *Loss Ratio*, pour prédire ces provisions IBNR. Toutes ces méthodes sont déterministes.

Cependant, dans le cadre de la réassurance, contrairement à celui de l'assurance, le nombre de sinistres est beaucoup plus faible. D'une part, tous les sinistres ne sont pas réassurés, notamment ceux très fréquents, très communs, peu sévères et bien appréhendés par les cédantes. D'autre part, conséquence de la raison citée ci-avant, les sinistres réassurés sont en majorité des sinistres peu fréquents et sévères.

Avec un nombre faible, voire très faible de sinistres dans le cadre de la réassurance, la stabilité statistique des méthodes de provisionnement classiques est diminuée. De plus, les montants de ces sinistres réassurés sont très variables. Le processus d'agrégation des sinistres, sur lequel reposent les méthodes classiques, est donc moins robuste et plus volatile.

En effet, certains sinistres dont le montant est vraiment important exercent une forte influence dans les résultats des méthodes classiques. Les provisions sont déterminées en prenant en compte ces sinistres de grande ampleur. Elles peuvent donc être jugées prudentes, souvent *a posteriori*, car ces sinistres sont rares et les provisions IBNR ne sont pas toujours consommées dans leur intégralité. Il est alors pertinent d'accéder aux informations individuelles des sinistres pour apporter des précisions sur les estimations des provisions données par les méthodes classiques et interpréter ces estimations. Pour y parvenir, l'idée est de générer des scénarios de développement individuel des sinistres IBNeR et IBNyR puis de sommer les provisions nécessaires pour chaque sinistre après les avoir calculées. Cela permettrait de mieux communiquer et de mieux justifier les résultats obtenus par l'équipe Reserving en ayant des scénarios plus ou moins probables de montants totaux de provisions IBNR.

Plus précisément, le modèle proposé dans ce mémoire consiste en 1 000 scénarios de provisions IBNeR et IBNyR, générés respectivement pour chaque sinistre IBNeR et chaque sinistre IBNyR, puis mis en commun dans l'optique d'obtenir 1 000 scénarios de provisions IBNR. La génération des scénarios des provisions IBNeR vise principalement à simuler le développement futur du montant des sinistres IBNeR. Quant à la génération des scénarios des provisions IBNyR, elle consiste notamment à simuler le nombre de sinistres IBNyR dans un premier temps, puis le montant de ces nouveaux sinistres dans un second temps. Une fois les simulations réalisées, il est nécessaire d'appliquer les conditions des traités aux montants simulés afin

d'obtenir les provisions de chaque sinistre IBNeR et IBNyR. À ce stade, les provisions peuvent être additionnées par scénario pour obtenir une distribution des provisions IBNR totales.

Ces scénarios sont à mettre en perspective avec les provisions IBNR préalablement calculées par l'équipe Reserving. Ces provisions sont globales et ne font pas la distinction entre les sinistres IBNeR et IBNyR. Cette confrontation permet de comprendre la façon dont ces montants IBNR pourraient être consommés par les différents sinistres IBNeR et IBNyR.

La problématique de ce mémoire est alors définie par la question suivante :

Quels sont les scénarios de développement individuel des sinistres IBNeR et IBNyR dont les provisions agrégées par année de survenance sont les plus proches des provisions IBNR calculées par l'équipe Reserving ?

Les prochains chapitres de ce mémoire retracent les travaux effectués pour tenter de répondre à cette problématique de provisionnement individuel grâce à l'utilisation de données propres aux sinistres. Trois segments de provisionnement constituent le fil rouge de ce mémoire, parmi les 80 modélisés.

Les segments de provisionnement correspondent à trois lignes d'activité : $\mathcal{CT}$, $\mathcal{MT}$ et $\mathcal{LT}$ ². La ligne $\mathcal{CT}$ – Court Terme – rassemble les traités couvrant les Dommages Matériels. Le segment $\mathcal{MT}$ – Moyen Terme – fait référence aux traités couvrant le transport des marchandises. La ligne $\mathcal{LT}$ – Long Terme – correspond quant à elle aux traités couvrant la Responsabilité Civile Professionnelle.

Plus généralement, les traités exploités concernent principalement la branche IARD de la réassurance de Ceded Re. Le portefeuille de ce dernier comprend également des traités Vie (dénommés ainsi par simplification) mais en très faible proportion, d'où le choix de ne pas les analyser dans ce mémoire.

En outre, les travaux sont centrés sur les traités de réassurance obligatoire et à couvertures non proportionnelles. En effet, dans la plupart des cas, les sinistres ne sont connus de la réassurance que lorsqu'ils touchent le traité. Du point de vue du réassureur, ces sinistres peuvent donc être inconnus pendant plusieurs années – si leur montant FGU est en deçà de la priorité – puis avoir un montant correspondant à plusieurs millions d'euros. Cependant, les entités du Groupe AXA tiennent parfois informé Ceded Re de certains sinistres survenus même s'ils ne touchent pas encore le traité mais pour lesquels ce sera potentiellement le cas dans le futur. Ainsi Ceded Re peut les prendre en compte dans ses bases et mieux les appréhender.

2. Les noms des lignes d'activités ont été modifiés afin de préserver la confidentialité des données.

2.2 Revue de littérature

Le provisionnement individuel occupe une place de plus en plus importante dans les recherches actuarielles, tant dans des papiers que dans de précédents mémoires. En effet, comme le provisionnement individuel nécessite des bases de données très volumineuses, ce type de provisionnement se développe au fur et à mesure que la capacité des ordinateurs grandit, tant au niveau stockage que puissance. La gestion de ces bases, de plus en plus facile et rapide, amène les actuaires à approfondir ce sujet.

Une revue non exhaustive des méthodes de provisionnement ligne à ligne analysées dans les papiers consultés est proposée ci-dessous. Dans une deuxième sous-section, trois mémoires d'actuariat sont davantage détaillés.

2.2.1 Les papiers actuariels

Cette sous-section trace un aperçu de différents papiers actuariels publiés au XXI^{ème} siècle. Tout d'abord, l'un des premiers articles à introduire le provisionnement sinistre par sinistre a été publié en 2012 (*Stochastic Claims Reserving Methods in Insurance*, 2012). Les méthodes stochastiques, encore assez peu répandues à cette époque, en sont au cœur avec notamment les processus de Poisson et les modèles linéaires généralisés.

De manière plus générale, la principale raison de l'intérêt porté sur le provisionnement individuel vient du fait qu'il est reproché aux méthodes basées sur le principe d'agrégation des sinistres de nécessiter d'hypothèses fortes trop peu souvent vérifiées, pourtant essentielles à la stabilité des résultats (Gabrielli, 2021).

Les problématiques de ces travaux restent globalement équivalentes : comment prédire le montant à l'ultime des sinistres IBNeR et IBNyR ? Les différentes solutions proposées exploitent toutes, *a minima*, des données historiques. Parmi ces solutions, on retrouve des méthodes de *machine learning* et des méthodes paramétriques ou semi-paramétriques, souvent utilisées de façon complémentaire. De plus, dans tous les papiers parcourus, la prédiction des développements des sinistres IBNeR et IBNyR est effectuée de façon séparée. En effet, leur traitement doit être adapté selon leur type car les sinistres IBNyR n'existent pas dans les bases alors que les sinistres IBNeR y sont présents.

Concernant les méthodes de *machine learning*, celles fréquemment citées sont les arbres de régression et de décision dont les algorithmes de classification binaire CART (« *Classification And Regression Trees* »), le *Gradient Boosting Machine* et les réseaux de neurones (Gabrielli, 2021).

Un article s'est même penché sur les principes du *deep learning* avec un algorithme LSTM

(*Long Short-Term Memory* en anglais) qui intègre une notion de dépendance temporelle ([Chaoubi et al., 2023](#)). Il est de surcroît agrémenté d’une distribution de Pareto généralisée pour les sinistres dont les montants sont extrêmes, faisant intervenir la théorie des valeurs extrêmes. Ce type de réseau de neurones récurrents fonctionne en deux temps :

- La probabilité de paiement au cours d’une période donnée est déterminée par des méthodes de classification, en prenant en compte le fait que les montants les plus extrêmes sont ceux dont les paiements sont les plus tardifs.
- Le montant de chaque paiement non nul est prédit grâce à des méthodes de régression.

L’article ([Antonio et al., 2016](#)) exploite lui aussi la théorie des valeurs extrêmes. Il présente un cadre multi-états permettant de modéliser les montants des sinistres IBNeR en fonction de variables explicatives. Les états sont la date de survenance des sinistres, leur date de déclaration, leurs paiements déjà effectués, ainsi que leur date de clôture. C’est une approche en temps discret dont le but est de modéliser les transitions entre les différents états au cours du temps. Chaque état doit avoir sa propre prédiction. Dans ce modèle, les paiements appartenant au corps sont ajustés grâce à un modèle paramétrique : le modèle additif généralisé pour la localisation, l’échelle et la forme (*Generalized Additive Model for Location, Scale and Shape* (GAMLSS)). Ceux appartenant à la queue de distribution sont ajustés par une distribution de Pareto généralisée. Le principe des méthodes paramétriques est donc utilisé dans le cadre de l’ajustement de lois entre la date de survenance du sinistre et ses délais de déclaration et paiements, de même que dans les articles ([Gabrielli, 2021](#); [Chaoubi et al., 2023](#); [Bücher and Rosenstock, 2022](#)). L’article ([Lopez et al., 2018](#)) mentionne aussi l’emploi de modèles non paramétriques.

D’autres papiers exploitent aussi des dépendances entre certaines variables, que ce soit, de même que le GAMLSS, entre les paiements d’un même sinistre ([Charpentier and Pigeon, 2016](#)) et entre le montant à l’ultime d’un sinistre et son nombre de développements, avec l’intervention de copules ([Lopez, 2019](#)). Dans le papier ([Charpentier and Pigeon, 2016](#)), des modèles économétriques sont abordés tels que la régression de Poisson (avec ou sans effet aléatoire) et de quasi-Poisson.

Concernant d’autres méthodes stochastiques, la méthode de Monte-Carlo est appliquée afin d’obtenir des distributions des provisions estimées pour les sinistres IBNeR ([Jin, 2013](#)).

Dans certains articles, une composante d’analyse de survie est ajoutée afin de prendre en compte les données censurées. Ces dernières font référence aux sinistres qui n’ont pas encore atteint leur ultime à la date de l’étude. Ces données censurées sont pourtant très importantes car de manière générale, plus les sinistres mettent du temps à se développer, plus leur ultime est élevé. Pour réduire le biais induit par cette censure et ne pas sous-estimer les montants futurs des sinistres, la méthode de Kaplan-Meier est implémentée. Les variables de durée correspondent

alors au temps écoulé entre l'ouverture des sinistres et leur clôture ([Lopez, 2019](#)). Certains articles faisant appel aux algorithmes CART les adaptent ainsi aux données censurées ([Lopez et al., 2016, 2018](#)).

Les avantages du provisionnement ligne à ligne cités dans tous les travaux sont similaires. Ce type de provisionnement émergent permet de prendre en compte des informations propres aux sinistres comme le délai de déclaration à l'assureur, l'année voire la date de survenance, les paiements déjà effectués, les dates de clôture des sinistres si elles sont déjà connues, les programmes de réassurance qui leur sont associés, les informations sur les sinistrés, l'exposition. Plus généralement, toutes les caractéristiques des sinistres, dès lors qu'elles sont disponibles, peuvent être utiles aux prédictions. Le papier ([Lopez et al., 2016](#)) mentionne justement une segmentation des sinistres en fonction de leurs caractéristiques. En outre, le provisionnement sinistre par sinistre permet de capter les changements dans les structures d'assurance ou de produits, dans les conditions des contrats ou des traités de réassurance, dans la réglementation et la régulation et dans l'environnement financier, économique et juridique.

La conclusion qui se dégage dans la grande majorité des papiers cités ci-dessus est que le provisionnement individuel est une avancée significative dans le domaine des sinistres dont le développement prend plusieurs années et qui sont visiblement souvent sous-estimés dans les méthodes classiques comme celle la Chain Ladder. Ces sinistres sont pour la plupart du temps ceux qui ont les montants les plus importants et ne sont donc pas à négliger dans les prédictions des provisions. En revanche, la différence n'est pas notable entre les méthodes classiques et le provisionnement ligne à ligne pour les sinistres dont le nombre d'années de développement est faible.

2.2.2 Les mémoires d'actuariat

L'attention est désormais portée sur trois mémoires d'actuariat présentés à l'Institut des Actuaires dans les dix dernières années. Leur ligne directrice est analogue. De même que pour les papiers actuariels, c'est à nouveau celle de l'estimation des provisions pour chaque sinistre IBNeR et IBNyR. Ces mémoires ont permis d'obtenir une idée globale de l'orientation à adopter pour le provisionnement individuel dans le cadre de ce mémoire.

Le premier mémoire ([Primel, 2021](#)) consiste en deux approches différentes : l'une dite additive et l'autre dite multiplicative. Elles sont appliquées à des données réelles de Responsabilité Civile Médicale dans le cadre de l'assurance. Les sinistres de cette ligne d'activité tendent à atteindre leur montant définitif seulement après de nombreuses années suivant leur survenance.

Pour ces deux méthodes, les sinistres tardifs – les IBNyR – sont ajoutés en amont de la base des sinistres utilisée. Le travail d'estimation du nombre de sinistres IBNyR est fait en amont.

1. Dans le modèle additif, deux grandes étapes se succèdent :

- (a) La première étape est l'estimation du nombre d'années de développement avant qu'un sinistre ne soit considéré comme clos, c'est-à-dire au moment auquel le sinistre a été complètement payé. Elle consiste à calculer des probabilités de survie grâce à l'estimateur de Kaplan-Meier afin de tenir compte de la censure. Ces probabilités sont celles qu'un sinistre continue à se développer ou non et sont propres à chaque année de développement. La réalisation d'une loi uniforme comprise entre 0 et 1 est attribuée à chaque année de développement future d'un sinistre, si elle existe. Une comparaison est ensuite faite entre cette réalisation et l'estimateur de Kaplan-Meier afin de déterminer si le sinistre est considéré encore ouvert ou déjà clos à l'année de développement en question. L'estimation du nombre d'années de développement avant qu'un sinistre soit considéré comme clos est alors la différence entre l'année à laquelle il est la première fois considéré clos et celle à laquelle il est survenu.
- (b) La deuxième étape est la simulation d'un montant de paiement pour chaque année de développement future des sinistres, c'est-à-dire jusqu'à leur année de clôture estimée dans la première étape. Pour cela, un montant de paiement historique pour chacune de ces années de développement est tiré aléatoirement. Une disjonction de cas est faite à ce moment-là entre les sinistres IBNeR et les sinistres IBNyR. Pour les paiements associés aux sinistres IBNeR, le tirage aléatoire se fait parmi tous les montants historiques de paiements des sinistres à l'année de développement en question. Pour les paiements associés aux sinistres IBNyR, le tirage est réduit aux montants historiques des paiements liés à des sinistres IBNyR, que ce soit pour la première année de développement ou les suivantes.

L'addition de tous les montants de paiements estimés correspond au montant de provisions pour tous les sinistres IBNR.

2. Dans le modèle multiplicatif, la première étape est identique à celle décrite dans le modèle additif. La différence réside dans la deuxième étape. Dans ce modèle, le but est d'associer un *link ratio* à chacune des années de développement futures de chaque sinistre. Un *link ratio* est le rapport entre deux montants cumulés adjacents. Aussi bien pour les sinistres IBNeR qu'IBNyR, un *link ratio* est tiré aléatoirement pour chacune des années de développement futures des sinistres parmi tous les *link ratios* historiques associés à l'année de développement en question. La différence entre les sinistres IBNeR et les sinistres IBNyR intervient à ce moment-là. Concernant les sinistres IBNeR, le produit de ces *link ratios* est appliqué directement au montant du paiement de la dernière année de développement connue du sinistre afin d'estimer le montant d'un sinistre à sa clôture. Or, pour les sinistres IBNyR, seulement leur nombre et leur année de déclaration ont été

simulés préalablement (grâce à une variante de la méthode du *Bootstrap* appliquée sur des triangles de nombres tirée du mémoire présenté ci-après (Riffaud, 2020)). À ce stade, ils n'ont donc pas de montant qui leur est associé et les *link ratios*, par conséquent, ne peuvent pas être exploités. C'est pourquoi le montant devant être multiplié par le produit des *link ratios* est tiré aléatoirement parmi tous les premiers montants déclarés des sinistres IBNyR historiques. Le montant total de provisions IBNR est donc la différence entre la somme des montants cumulés des sinistres à leur ultime et la somme des montants cumulés à la dernière année de développement connue des sinistres.

Le deuxième mémoire (Riffaud, 2020), duquel s'est inspiré le mémoire présenté ci-dessus (Primel, 2021) pour l'estimation du nombre de sinistres IBNyR – appelés sinistres « tardifs » – et leur année de développement associée, porte sur une méthode de fréquence/sévérité. Cette méthode suppose l'indépendance entre la fréquence – le nombre de sinistres survenus mais non encore déclarés – et la sévérité – la charge à l'ultime pour chacun des sinistres prédits. La prédiction de la fréquence et celle de la sévérité sont distinctes. Différents modèles de fréquence et de sévérité sont présentés dans le mémoire et sont récapitulés sommairement ici. Deux bases de données assurantielles sont utilisées afin de mettre en pratique cette méthode. L'une contient des données de Responsabilité Civile Médicale – comme dans le mémoire présenté ci-dessus – tandis que l'autre comporte des données Automobile, une ligne d'activité où les sinistres atteignent généralement leur montant définitif en moins d'années que ceux de la Responsabilité Civile Médicale.

Concernant les modèles de fréquence, trois méthodes sont décrites :

- **Un *Bootstrap*** : Il est réalisé sur des triangles de nombres de sinistres, avec trois rééchantillonnages différents testés (sur la partie supérieure du triangle, la partie inférieure et les deux).
- **Une méthode de calibrage de la loi de fréquence** : Il est considéré que le nombre de sinistres pour une même année de développement, indépendamment des années de survenance, est la réalisation d'une même variable aléatoire. Le but est de calibrer de façon paramétrique cette variable aléatoire grâce au nombre de sinistres historiques par année de développement. Les nombres du triangle inférieur sont alors tirés aléatoirement selon la loi de la variable associée à l'année de développement en question.
- **Une méthode supposant la dépendance entre les nombres des différentes années de développement** : Par exemple, si pour une année de développement donnée le nombre de sinistres est faible, il est fort probable que le nombre de sinistres de l'année de développement suivante soit plus élevé. Pour estimer le nombre de sinistres selon cette méthode, deux approches ont été testées : celle du modèle linéaire et celle du modèle linéaire généralisé.

Concernant les différents modèles de sévérité, deux approches sont présentées :

- Le montant à l’ultime pour chaque sinistre tardif estimé est tiré aléatoirement parmi tous les montants à l’ultime des sinistres clos auxquels sont appliqués des coefficients d’inflation afin de prendre en compte cette dernière.
- Une loi de sévérité est calibrée sur la distribution empirique des montants à l’ultime des sinistres clos. Le montant à l’ultime de chaque sinistre tardif estimé est tiré aléatoirement selon la loi de sévérité calibrée.

Le troisième mémoire ([Giuly, 2018](#)) traite également du provisionnement individuel sur différentes lignes d’activités IARD mais dans le cadre de la réassurance, contrairement aux deux autres mémoires. Il semblait donc essentiel qu’il soit mentionné. Il présente une démarche permettant d’estimer le plus précisément possible un montant de provisions pour chaque traité de réassurance. Deux modèles sont exposés : un pour les sinistres IBNeR et un pour les sinistres IBNyR.

1. Le modèle sur les sinistres IBNeR cherche à prédire la charge ultime de ces sinistres à partir de données historiques. Deux approches sont considérées :
 - (a) La première approche consiste à développer les sinistres en leur appliquant une moyenne pondérée des *link ratios* – associés à l’année de développement en question – de tous les sinistres déclarés et ce, pour chaque année de développement du sinistre jusqu’à sa clôture. La détermination de l’année de clôture n’est cependant pas traitée.
 - (b) La deuxième approche consiste à développer les sinistres directement à leur ultime en appliquant une moyenne pondérée des facteurs de développement. Cette fois-ci, ces facteurs sont calculés uniquement à partir des sinistres clos. Chaque facteur correspond au montant du sinistre lorsqu’il est clos, divisé par son montant au moment de l’année de développement en question des sinistres à développer.

Dans les deux approches, la pondération se fait en fonction de la proximité des sinistres. Des sinistres sont considérés comme proches lorsqu’ils appartiennent à la même sous-branche, sont rattachés à la même cédante, sont survenus la même année et ont eu le même délai de déclaration. Ces caractéristiques expliquent au mieux les développements d’un sinistre, d’après le résultat de l’algorithme d’apprentissage supervisé *Gradient Boosting Machine* qui classe les variables selon leur ordre d’influence relative.

2. Le travail sur le modèle des sinistres IBNyR a été notamment effectué sur la fréquence, c’est-à-dire sur le nombre de sinistres qui vont toucher un traité au cours de leur vie. À nouveau, plusieurs approches ont été envisagées :
 - (a) La distribution du nombre de sinistres à l’ultime est simulée *a priori*, c’est-à-dire sans utilisation de données historiques, en prenant seulement en compte les caractéristiques des traités. Trois modèles sont testés :

- i. Utilisation d'un arbre de régression ;
 - ii. Ajustement d'une loi de fréquence ;
 - iii. Utilisation des processus de renouvellement avec calibration de loi sur les temps d'inter-arrivées – supposés indépendants – des sinistres.
- (b) Le nombre de sinistres à l'ultime est estimé *a posteriori*, à partir du nombre de sinistres déclarés et de leur cadence de déclaration.
- (c) Une méthode inspirée de celle de Bornhuetter-Ferguson combinant les estimations des approches *a priori* avec l'arbre de régression et *a posteriori* est étudiée.

2.3 Apport du mémoire

Tous les papiers et mémoires cités ci-dessus traitent du sujet du provisionnement individuel dans l'optique de prédire les provisions à l'ultime engendrées par les sinistres IBNeR et IBNyR. Ils ont fourni une vision d'ensemble de l'état de l'art du provisionnement individuel dans la recherche actuarielle.

Leur motivation est d'analyser les résultats d'une ou de différentes méthodes de provisionnement ligne à ligne en les comparant aux résultats des méthodes ayant recours à l'agrégation des données. Ils espèrent qu'en exploitant plus de données propres aux sinistres, les résultats seront plus précis et plus fiables que ceux des méthodes classiques de provisionnement.

La problématique de ce mémoire est différente dans le sens où, bien que le montant à l'ultime de chaque sinistre soit également calculé, ce n'est pas à titre de comparaison avec les méthodes classiques, mais au contraire pour expliquer, analyser et interpréter les résultats obtenus avec ces dites méthodes classiques. Les provisions IBNR calculées en amont doivent donc être analysées sinistre par sinistre. Cette analyse repose sur la génération des scénarios de développement des sinistres IBNeR et IBNyR par l'intermédiaire du provisionnement individuel. Ce dernier est donc un outil permettant d'atteindre un objectif d'explication tout en simulant une distribution des provisions IBNeR et IBNyR.

De plus, ce mémoire traite à la fois le sujet des sinistres IBNeR et celui des sinistres IBNyR, ce qui n'est pas le cas dans la plupart des travaux cités ci-dessus.

En outre, peu de travaux ont été effectués dans le cadre de la réassurance, qui est moins facile à appréhender du fait de sinistres plus volatiles et moins nombreux. La plupart d'entre eux sont effectivement réalisés dans le cadre de l'assurance et notamment testés sur les sinistres issus de la Responsabilité Civile, qui est une branche connue pour être à développement long. Dans les travaux réalisés au cours du stage, toutes les lignes d'activité ont été prises en compte, soit environ 80. En revanche, seulement trois lignes sont exposées dans ce mémoire. Il s'agit des

lignes nommées $\mathcal{CT}$, $\mathcal{MT}$ et $\mathcal{LT}$.

Ces différentes visions du provisionnement individuel ont tout de même participé à l'élaboration du cheminement pour la résolution de la problématique de ce mémoire.

Chapitre 3

Présentation et structuration des données

Les principaux termes et notions ayant été définis précédemment, l'intérêt se porte maintenant sur les données exploitées afin de répondre à la problématique de ce mémoire.

Tout d'abord, un rappel de cette problématique s'avère nécessaire. L'un des buts est de générer des scénarios de développement des sinistres IBNeR et IBNyR afin d'estimer quelles pourraient être les différentes consommations des provisions IBNR par ces sinistres IBNeR et IBNyR, les provisions IBNR étant calculées en amont de ces travaux par l'équipe Reserving.

Il est donc essentiel que les bases de données utilisées dans le cadre de ce mémoire soient les mêmes que celles employées dans le calcul des provisions IBNR par l'équipe, afin de pouvoir justifier ces dernières de manière cohérente.

Ce chapitre met l'accent sur les différentes bases de données qui sont utiles dans le cadre global de ce mémoire. Dans une première section, les bases de données sont présentées ainsi que les différentes variables intéressantes vis-à-vis de la problématique. La deuxième section traite d'une partie du nettoyage et du retraitement des données. La troisième section aborde quant à elle la création de nouvelles variables.

3.1 La présentation des données

Pour le développement des travaux, trois bases de données sont nécessaires : *Traités*, *Sinistres* et *Sinistres en vision comptable*. Elles portent uniquement sur la réassurance au sein de Ceded Re.³

3. Tous les montants sont supposés en euros dans ce mémoire, les taux de change étant considérés comme étant déjà appliqués.

3.1.1 La base de données *Traités*

Les traités renseignés dans la base de données *Traités* sont en majorité des traités à couverture non proportionnelle, c'est-à-dire des couvertures en XS ou en SL. Il existe tout de même certains traités à couverture proportionnelle en EP ou en QP. Parfois, les traités de réassurance font partie de programmes de réassurance. Un programme contient plusieurs traités. Il peut donc être amené à y avoir différents types de couvertures dans un même programme selon les traités qui le constituent. Un programme de réassurance peut ainsi apparaître plusieurs fois dans cette base, bien que ce genre de cas reste rare.

Les variables renseignées dans cette base et servant pour l'avancement du mémoire sont présentées dans le tableau en figure 3.1 ci-dessous. Les noms entre parenthèses représentent les noms des variables utilisées dans la suite du mémoire, sauf indication contraire.

Variable	Descriptif
Numéro du traité	Ce numéro permet d'identifier de façon unique un traité.
Nom du programme (<i>P</i>)	C'est le nom d'un programme de réassurance.
Identifiant du traité (<i>T</i>)	C'est le code pour identifier le traité. C'est la granularité la plus fine de la base.
Année d'exercice	C'est l'année d'application du traité. Pour qu'un sinistre touche un traité, en plus que son montant FGU dépasse la priorité, il faut qu'il soit survenu pendant l'année d'exercice du traité, même s'il est déclaré plus tard à l'assureur et/ou que son montant évolue ultérieurement. Dans le cadre de ce mémoire, les traités de réassurance en <i>Risk Attaching</i> ne sont pas considérés. Il s'agit de traités couvrant des sinistres sur plusieurs années d'accident. Ces traités en <i>Risk Attaching</i> existent notamment dans le cas de garanties décennales.
Segment de provisionnement	C'est la ligne d'activité à laquelle est rattaché le traité.
Nom du <i>Pool</i>	C'est le nom du <i>Pool</i> auquel est rattaché le traité.
Part signée du traité (<i>Part signée_T</i>)	C'est la part signée du traité. C'est le pourcentage accepté par Ceded Re du montant du sinistre à charge du traité. Cette proportion du montant est à l'acceptation d'AXA SA. Elle est comprise entre 0 et 100%. Si elle vaut 100%, le montant est à l'acceptation d'AXA SA dans sa totalité. Si elle vaut 50%, cela signifie qu'AXA SA signe pour indemniser seulement la moitié du montant à charge du traité et qu'elle reçoit uniquement la moitié des primes de la part des cédantes. Montant du sinistre à charge du traité à l'acceptation d'AXA SA = Montant total du sinistre à charge du traité $\times$ <i>Part signée</i>

Variable	Descriptif
Part nette du traité (<i>Part nette_T</i>)	C'est le pourcentage du montant du sinistre à charge du traité mis au sein du <i>Pool</i> auquel est rattaché le traité en question. C'est le montant à la charge d'AXA SA. Cette part nette existe en raison de mécanismes de rétrocession d'AXA SA envers la réassurance externe qui sont mis en place pour la plupart des traités. La part nette peut donc être différente de la part signée du fait de cette rétrocession mais elle est également comprise entre 0 et 100%. Si elle vaut 60%, seulement trois cinquièmes de ce montant sont à la charge d'AXA SA. Montant du sinistre à charge du traité à la charge d'AXA SA = Montant total du sinistre à charge du traité $\times$ <i>Part nette</i>
Prime acquise	C'est la prime payée par la cédante. Pour l'année courante, soit l'année actuelle, la prime acquise est proportionnelle au temps passé depuis le début de l'année. Si la moitié de l'année courante est passée, le montant de cette prime doit correspondre à la moitié de la totalité de la prime pour l'année entière.

FIGURE 3.1 – Présentation des variables de la base *Traités*

3.1.2 La base de données *Sinistres*

La base de données *Sinistres* contient de nombreuses informations concernant les sinistres déclarés à Ceded Re de la part des entités. La plupart d'entre eux touchent au moins un traité, parfois plus, voire tous les traités d'un même programme de réassurance. Mais il existe aussi des sinistres renseignés dans cette base dont la charge FGU – la charge totale de l'indemnisation du sinistre vis-à-vis de l'assuré – est en deçà de la priorité minimale, toutes tranches du programme confondues. En effet, les sinistres sont notifiés à Ceded Re lorsqu'ils dépassent un seuil qui est généralement un pourcentage de la priorité. Ainsi, Ceded Re est informé de leur survenance et peut se préparer à une éventuelle hausse de leur montant FGU. Cette hausse pourrait amener certains sinistres à toucher un ou plusieurs traités dans le cas où les conditions seraient réunies pour qu'ils soient réassurés, ce qui impacterait les provisions de Ceded Re.

Les variables les plus pertinentes pour le sujet du mémoire et présentes dans cette base sont celles du tableau de la figure 3.2 ci-dessous. Les noms entre parenthèses représentent à nouveau les noms des variables utilisées dans la suite du mémoire, sauf indication contraire.

Variable	Descriptif
Numéro de sinistre (<i>S</i>)	Ce numéro permet d'identifier de façon unique un sinistre.

Variable	Descriptif
Année de survenance (i)	C'est l'année à laquelle est survenu le sinistre.
Montant payé FGU ($Payé$)	C'est le montant qui a déjà été payé par la cédante à l'assuré pour son sinistre dans sa globalité.
Charge FGU ($Charge_{FGU}$)	C'est le montant total du sinistre en vision FGU. Elle correspond donc au montant brut du sinistre vis-à-vis de la cédante.
Provision D/D FGU ($Provision_{D/D}$)	C'est le montant devant être mis de côté pour indemniser le sinistre. La cédante doit encore verser ce montant à l'assuré, en plus de ce qu'elle a déjà payé. Le montant payé, la provision D/D et la charge en vision FGU sont liés de la façon suivante : $Charge_{FGU} = Payé + Provision_{D/D}$
Statut	C'est une variable qui prend deux valeurs. Elle vaut « Clos » si le sinistre n'attend plus de prochaines évaluations – il a alors atteint son ultime. L' ultime est la valeur finale arrêtée de la Charge FGU du sinistre. La variable vaut « Actif » si la Charge FGU peut encore potentiellement être réévaluée dans le futur – le sinistre n'a alors pas encore atteint son ultime et pourrait continuer à se développer. Un sinistre qui se développe a de nouvelles évaluations qui modifient ses montants antérieurs.
Identifiant du traité (T)	C'est le code pour identifier le traité auquel est rattaché le sinistre. Grâce à cet identifiant, une correspondance peut être faite entre la base <i>Traités</i> et la base <i>Sinistres</i> . Cette variable est nécessaire pour connaître le segment de provisionnement auquel est rattaché un sinistre. S'il y a plusieurs identifiants de traités pour un même sinistre, cela signifie que le sinistre a touché plusieurs traités.
Priorité du traité ($Priorité_T$)	C'est la priorité du traité. Elle peut ne pas être renseignée dans certains cas, comme dans celui d'une couverture proportionnelle n'ayant pas de rétention définie dans les conditions au traité.
Portée du traité ($Portée_T$)	C'est la portée du traité. Elle peut valoir l'infini si elle n'a pas été spécifiée dans le traité.
Type de traité	C'est le type du traité de réassurance auquel est rattaché le sinistre. Il existe des traités en EP, en QP, en XS et en SL.
Taux de cession QP (α_{QP})	C'est le taux de cession dans le cadre des traités en Quote-Part.
Taux de cession EP (α_{EP})	C'est le taux de cession dans le cadre des traités en Excédent de Plein.

FIGURE 3.2 – Présentation des variables de la base *Sinistres*

Les priorités et les portées peuvent évoluer dans le temps afin de suivre le cours de l'inflation. Elles sont alors appelées **portées indexées** et **priorités indexées**. Ces dernières sont les

valeurs des priorités et des portées définies aux traités sur lesquelles des coefficients d'inflation sont appliqués. Cette démarche permet d'éviter de considérer qu'un sinistre évalué après sa date de survenance touche un traité seulement dû à un effet d'inflation. Dans le cas où les priorités et portées contractuelles ne seraient pas ajustées, le risque deviendrait injustement réparti entre le réassureur et les cédantes puisque le réassureur devrait indemniser des sinistres qui n'auraient pas touché le traité s'ils avaient été évalués durant leur année de survenance. Actuellement, dans les bases de données, ces priorités et portées indexées ne sont pas disponibles. Dans toute la suite, le concept d'indexation n'est pas abordé, ce qui constitue une limite au modèle. Cependant, ce problème est relatif à la structure des données. Ainsi, une fois que les informations des variables indexées seront accessibles, il devrait être aisé de mettre à jour le modèle pour prendre en compte ces nouvelles valeurs. Les développements réalisés présentent donc actuellement des limites en termes de résultats, mais ce problème n'impacte pas la démarche globale.

3.1.3 La base *Sinistres en vision comptable*

Dans ce mémoire, la base de données *Sinistres en vision comptable* fait uniquement office de table de vérification. Elle est créée par l'équipe Comptable au sein de Ceded Re. Dans la section 3.3 ci-après, des calculs sont effectués sur la base *Sinistres* en vue de préparer le modèle établi dans le cadre de ce mémoire. La base *Sinistres en vision comptable* sert à contrôler l'exactitude de ces calculs, car les résultats, déjà déterminés par l'équipe Comptable, y sont répertoriés.

3.2 Le nettoyage et le retraitement des données

Comme dans tout travail ayant recours à des bases de données, cette partie de nettoyage et de retraitement des données n'est pas à négliger car elle peut influencer tout le reste de la modélisation.

Cette section ne développe pas l'entièreté du nettoyage et des retraitements effectués sur les bases citées ci-dessus. Elle se focalise sur ceux ayant un impact sur les résultats de la modélisation et plus particulièrement sur ceux de la base *Sinistres*.

Premièrement, tous les sinistres renseignés dans cette base ne sont pas significatifs car certains sinistres ont une Charge FGU très faible comparée à la priorité minimale du programme auquel ils sont rattachés. Cela est dû au fait que les cédantes envoient les informations des sinistres même s'ils ne touchent pas encore la priorité. Les sinistres qui ne sont pas significatifs ont peu de chances de dépasser la priorité d'un traité lors d'une nouvelle évaluation. Les supprimer permet alors d'alléger la base de données.

Dans ces travaux, un sinistre S est considéré significatif lorsque :

$$\frac{\max\{Charge_{FGU,S}\}}{\min\{Priorité_P\}} > Seuil$$

Avec :

- $Charge_{FGU,S}$: la Charge FGU du sinistre S ;
- $Priorité_P$: la priorité d'une tranche du programme P .

Le seuil a été choisi à 5% par des experts. Cela signifie qu'un sinistre est considéré comme ne pouvant pas atteindre plus de 20 fois sa Charge FGU maximale recensée historiquement lorsqu'il est en deçà de la priorité minimale du programme auquel il est rattaché.

Comme exemple, le tableau en figure 3.3 présente quelques colonnes de la base *Sinistres*.

Numéro de sinistre	Numéro du traité	Année d'évaluation	Charge FGU	Priorité	Portée
S300	T560	2010	1 000	10 000 000	20 000 000
S300	T561	2010	1 000	20 000 000	10 000 000
S300	T560	2011	500	10 000 000	20 000 000
S300	T561	2011	500	20 000 000	10 000 000
S452	T532	2010	400 000	5 000 000	20 000 000
S671	T789	2010	1 000 000	600 000	1 500 000

FIGURE 3.3 – Extrait de la base de données *Sinistres*

Dans cet exemple théorique :

- Le sinistre S300 est supprimé de la base car :

$$\frac{\max\{1\,000 ; 500\}}{\min\{10\,000\,000 ; 20\,000\,000\}} = \frac{1\,000}{10\,000\,000} = 0,0001 < 0,05.$$

- Le sinistre S452 sera conservé dans la base car :

$$\frac{400\,000}{5\,000\,000} = 0,08 > 0,05.$$

- Le sinistre S671 touche la priorité et est donc conservé. Le calcul le confirme :

$$\frac{1\,000\,000}{600\,000} = 1,66 > 0,05.$$

Deuxièmement, pour construire les triangles de développement présentés dans la section suivante et, plus généralement, pour élaborer le modèle, il est fait en sorte que chaque sinistre

soit renseigné dans la base *Sinistres* exactement une fois pour chaque année de développement, ce qui n'est pas le cas dans la base non retraitée.

3.3 La création de nouvelles variables

Cette section fournit des explications sur les nouvelles variables, leur processus de création et les raisons de leur ajout.

3.3.1 Années de développement et triangles de développement

Bien que le sujet du mémoire concerne le provisionnement individuel, les **triangles de développement** exploités par les méthodes classiques ne sont pas à proscrire. En effet, ils offrent une visualisation globale importante de la sinistralité de chaque segment de provisionnement d'une manière assez directe. Dans ces triangles, les années de survenance sont en ligne et les années de développement sont en colonne.

L'**année de développement** j d'un sinistre correspond au nombre d'années écoulées depuis l'année de survenance de ce sinistre. Les années de développement n'étant pas renseignées dans la base *Sinistres*, le choix de la convention pour les déterminer afin de les ajouter à la base est la suivante :

$$\text{Année de développement } j = \text{Année d'évaluation} - \text{Année de survenance } i + 1$$

L'année de survenance i d'un sinistre est alors considérée comme sa première année de développement. Un sinistre est à sa $j^{\text{ème}}$ année de développement lorsqu'il est survenu il y a $j - 1$ années. Plus l'année de survenance est récente, plus le nombre d'années de développement historiques du sinistre est faible. L'année maximale de développement connue d'un sinistre S est notée J et est celle répondant au calcul suivant :

$$\text{Dernière année de développement connue } J = \text{Année courante}^4 - \text{Année de survenance } i + 1$$

La **profondeur historique** d'un segment de provisionnement, notée D , est le nombre maximal d'années de développement au sein des sinistres du segment. Elle est définie comme suit :

$$D = \max_{S \in \mathcal{S}} \{\text{Dernière année de développement connue du sinistre } S\}$$

Avec $\mathcal{S}$: l'ensemble des sinistres du segment de provisionnement.

4. Dans l'ensemble de ce mémoire, l'année courante est considérée comme étant 2024.

Les profondeurs historiques des trois segments de provisionnement sont présentées dans le tableau en figure 3.4.

Segment de provisionnement	$\mathcal{CT}$	$\mathcal{MT}$	$\mathcal{LT}$
Profondeur historique	41	35	42

FIGURE 3.4 – Profondeur historique des trois segments de provisionnement

L'un des enjeux du provisionnement est d'estimer les montants associés aux années de développement futures des sinistres qui ne sont pas clos, afin d'en déduire les provisions IBNR nécessaires pour les indemniser.

Plus le nombre d'années de développement avant la clôture d'un sinistre est élevé – respectivement faible –, plus le sinistre est considéré à développement long – respectivement court. Généralement, une tendance se retrouve dans le nombre d'années de développement avant la clôture des sinistres d'un même segment de provisionnement. Chaque segment est alors considéré comme étant à développement plutôt court ou plutôt long. Les trois lignes d'activités qui constituent le fil de rouge de ce mémoire ont chacune une tendance différente d'après les experts, ce qui est vérifié dans la partie modélisation. La ligne $\mathcal{CT}$ est à développement court terme, la $\mathcal{MT}$ à développement moyen terme et la $\mathcal{LT}$ à développement long terme.

Pour revenir aux triangles de provisionnement, cinq types sont exploités dans ce mémoire :

- Les triangles de montants cumulés : Les montants⁵ renseignés dans le triangle correspondent à l'agrégation de tous les montants des années de développement par année de survenance, c'est-à-dire :

$$M_{agreg,i,j} = \sum_{S \in \mathcal{S}_i} M_{S,i,j}$$

Avec :

- $\mathcal{S}_i$: l'ensemble des sinistres survenus en année i ;
- $M_{S,i,j}$: le montant du sinistre S survenu en année i à l'année de développement j .

Le tableau en figure 3.5 ci-dessous présente les variables des sinistres utiles à la construction du triangle de montants cumulés en figure 3.6 ci-après, avec les couleurs correspondantes.

5. À chaque fois que les termes « montant » ou « montants » sont mentionnés dans cette section, ils peuvent aussi bien se référer aux primes qu'aux montants associés aux sinistres comme les Charges FGU, Provisions D/D FGU ou Montants payés FGU.

Numéro du sinistre	Année de survenance	Année de développement	Charge FGU
S211	2020	1	10
S211	2020	2	15
S211	2020	3	20
S211	2020	4	20
S211	2020	5	20
S349	2021	1	15
S349	2021	2	20
S349	2021	3	25
S349	2021	4	22
S774	2021	1	30
S774	2021	2	20
S774	2021	3	10
S774	2021	4	20
S159	2022	1	35
S159	2022	2	45
S159	2022	3	55
S318	2023	1	15
S318	2023	2	80
S933	2023	1	5
S933	2023	2	0
S104	2024	1	20

FIGURE 3.5 – Différentes données de sinistres

		Année de développement				
		1	2	3	4	5
Année de survenance	2020	10	15	20	20	20
	2021	45	40	35	42	
	2022	35	45	55		
	2023	20	80			
	2024	20				

FIGURE 3.6 – Triangle de montants cumulés

Dans l'exemple en figure 3.6 ci-dessus, la somme des montants des sinistres survenus en 2021 à leur troisième année de développement est de 35 et correspond à l'addition des

Charges FGU des sinistres S349 (25) et S774 (10).

Avec cette visualisation, le triangle supérieur est connu tandis que triangle inférieur est quant à lui inconnu puisqu'il représente les futures années de développement des sinistres.

- Les triangles de *link ratios* incrémentaux : Les *link ratios* incrémentaux sont les coefficients qui permettent de passer du montant cumulé d'une année de développement à celui de la suivante. Ils sont définis comme suit :

$$Link\ ratio_{incr,i,j} = \begin{cases} \frac{M_{cumul,i,j}}{M_{cumul,i,j-1}} & \text{si } M_{cumul,i,j-1} \neq 0, \\ 1 & \text{si } M_{cumul,i,j-1} = 0 \text{ et } M_{cumul,i,j} \neq 0, \\ \text{NA} & \text{sinon.} \end{cases}$$

Avec $M_{cumul,i,k}$: le montant cumulé de l'année de survenance i pour l'année de développement k ($k \neq 1$).

Le triangle de *link ratio* incrémentaux associé au triangle des montants cumulés de l'exemple précédent est en figure 3.7.

		Année de développement				
		1	2	3	4	5
Année de survenance	2020	NA	1,5	1,333	1	1
	2021	NA	0,889	0,875	1,2	
	2022	NA	1,289	1,222		
	2023	NA	4			
	2024	NA				

FIGURE 3.7 – Triangle de *link ratios* incrémentaux

Les triangles de développement peuvent également être complétés avec ces *link ratios* incrémentaux. Il n'existe cependant pas de *link ratios* incrémentaux pour la première année de développement de chaque année de survenance.

Les *link ratios* incrémentaux peuvent être classés en trois catégories :

- $Link\ ratio_{incr,i,j} > 1$ si $M_{cumul,i,j} > M_{cumul,i,j-1}$, c'est-à-dire que le montant cumulé augmente de l'année de développement $j - 1$ à l'année de développement j pour l'année de survenance i ;
- $Link\ ratio_{incr,i,j} = 1$ si $M_{cumul,i,j} = M_{cumul,i,j-1}$, c'est-à-dire que le montant cumulé reste constant de l'année de développement $j - 1$ à l'année de développement j pour l'année de survenance i ;
- $Link\ ratio_{incr,i,j} < 1$ si $M_{cumul,i,j} < M_{cumul,i,j-1}$, c'est-à-dire que le montant cumulé

diminue de l'année de développement $j - 1$ à l'année de développement j pour l'année de survénance i .

- Les triangles de *link ratios* cumulés : Les *link ratios* cumulés sont les produits des *link ratios* incrémentaux. Ce sont les coefficients qui permettent de passer du montant cumulé d'une année de développement au montant cumulé de la dernière année de développement. Dans le cadre de ce mémoire, les *link ratios* cumulés valent 1 au minimum afin de ne pas prendre en compte les éventuelles baisses de montants. Ce choix de modélisation permet de justifier une certaine prudence dans les estimations des provisions IBNR. Ils sont alors calculés comme suit :

$$Link\ ratio_{cumul,i,j} = \begin{cases} \max\left(1 ; \prod_{k \geq j} Link\ ratio_{incr,i,k}\right) & \text{si } Link\ ratio_{incr,i,k} \neq NA \forall k \geq j \\ NA & \text{si } Link\ ratio_{incr,i,k} = NA \forall k \geq j \end{cases}$$

Avec :

- $Link\ ratio_{cumul,i,j}$: le *link ratio* cumulé de l'année de développement j ($j \neq 1$) pour l'année de survénance i ;
- $Link\ ratio_{incr,i,k}$: le *link ratio* incrémental de l'année de développement k ($k \neq 1$) pour l'année de survénance i .

Le triangle de *link ratios* cumulés associé au triangle des *link ratios* incrémentaux de l'exemple précédent est en figure 3.8 ci-dessous. À nouveau, les *link ratios* cumulés ne sont pas définis pour la première année de développement de chaque année de survénance.

		Année de développement				
		1	2	3	4	5
Année de survénance	2020	NA	$\max\{1 ; 1,5 \times 1,333 \times 1 \times 1\}$ $= 2$	1,333	1	1
	2021	NA	$\max\{1 ; 0,889 \times 0,875 \times 1,2\}$ $= \max\{1 ; 0,993\}$ $= 1$	$\max\{1 ; 0,875 \times 1,2\}$ $= 1,05$	1,2	
	2022	NA	1,575	1,222		
	2023	NA	4			
	2024	NA				

FIGURE 3.8 – Triangle de *link ratios* cumulés

Les *link ratios* cumulés peuvent être quant à eux classés en deux catégories dans le cadre de ce mémoire :

- $Link\ ratio_{cumul,i,j} > 1$ si $M_{cumul,i,J} > M_{cumul,i,j-1}$, c'est-à-dire que le montant cumulé augmente de l'année de développement $j - 1$ à la dernière année de développement J pour l'année de survenance i ;
- $Link\ ratio_{cumul,i,j} = 1$ si $M_{cumul,i,J} \leq M_{cumul,i,j-1}$, c'est-à-dire que le montant cumulé reste constant ou diminue de l'année de développement $j - 1$ à la dernière année de développement J pour l'année de survenance i .

Il est important de noter que, pour un sinistre, un *link ratio* cumulé supérieur à 1 n'implique pas que tous ses *link ratios* incrémentaux soient supérieurs à 1, comme pour le *link ratio* cumulé rouge. Cela ne signifie pas que le montant de ce sinistre croît forcément d'une année de développement à l'autre, simplement que son montant est plus élevé à sa dernière année de développement qu'à son année de développement $j - 1$. Un *link ratio* cumulé égal à 1 ne suppose pas non plus que le montant reste constant ou décroît d'une année de développement à l'autre. Il indique uniquement que son montant à sa dernière année de développement est au plus égal à celui de son année de développement $j - 1$.

- Les triangles de nombres incrémentaux : Les triangles de nombres exploités dans le cadre de ce mémoire sont des triangles incrémentaux. Ils recensent le nombre de sinistres dont les charges sont supérieures à 0 pour la première fois à l'année de développement j :

$$N_{i,j} = \sum_{S \in \mathcal{S}_i} \mathbf{1}_{\{Charge_{S,i,j} > 0 \text{ et } Charge_{S,i,j'} = 0 \ \forall j' < j\}}$$

Avec :

- $\mathcal{S}_i$: l'ensemble des sinistres survenus en année i ;
- $Charge_{S,i,k}$: la charge du sinistre S survenu en année i à l'année de développement k ;
- $N_{i,j}$: le nombre de sinistres survenus en année i et présents à l'année de développement j avec une charge positive.

Il est également possible de construire des triangles de nombres cumulés mais, comme ils ne sont pas abordés dans ce mémoire, ils ne sont pas détaillés ici.

- Les triangles de développement des sinistres : Ces triangles de développement sont différents de ceux présentés ci-dessus. En plus des années de survenance en ligne, les numéros de sinistres s'y adjoignent. Ces triangles n'ont donc pas une forme de triangle mais l'idée initiale des triangles de développement reste la même : ils permettent de constater l'évolution des montants des sinistres selon leurs années de développement. Les montants de chaque sinistre sont explicités un par un pour chaque année de développement.

Le tableau en figure 3.9 ci-dessous est un exemple d'un triangle de développement de différents sinistres dans le cas de la vision cumulée des Charges FGU.

Numéro de sinistre	Année de survenance	Année de développement				
		1	2	3	4	5
S202	2020	50	60	60	70	100
S899	2020	30	40	45	60	60
S340	2021	35	45	55	80	
S765	2021	20	80	100	150	
S912	2021	20	20	20	20	

FIGURE 3.9 – Triangle de développement des sinistres

3.3.2 Le pourcentage de la provision D/D

La variable « Pourcentage de la provision D/D » reflète le pourcentage de la provision D/D par rapport à la Charge FGU et est calculée comme suit :

$$\text{Pourcentage de la provision } D/D_{S,j} = \frac{\text{Provision}_{D/D,S,j}}{\text{Charge}_{FGU,S,j}} \times 100$$

Avec :

- $\text{Provision}_{D/D,S,j}$: le montant de la provision D/D en vision FGU pour l'année de développement j du sinistre S ;
- $\text{Pourcentage de la provision } D/D_{S,j}$: le pourcentage du montant de la provision D/D en vision FGU pour l'année de développement j du sinistre S .

Le tableau en figure 3.10 est un exemple de pourcentages de provisions D/D calculés pour différentes années de développement de deux sinistres fictifs.

Numéro du sinistre	Année de développement	Provision D/D FGU	Charge FGU	Pourcentage de la provision D/D
S154	1	2 000 000	5 000 000	40%
S154	2	1 000 000	5 000 000	20%
S154	3	1 000 000	5 000 000	20%
S293	1	500 000	1 000 000	50%
S293	2	1 000 000	2 000 000	50%
S293	3	600 000	2 000 000	30%
S293	4	0	2 000 000	0%

FIGURE 3.10 – Tableau contenant des pourcentages de provisions D/D

3.3.3 Fusion des bases de données *Sinistres* et *Traités*

Dans cette sous-section, la description porte sur la fusion de la base *Sinistres* et de certaines variables de la base *Traités* dont le segment de provisionnement.

L'équipe Reserving calcule les provisions des sinistres IBNR par segment de provisionnement. La même démarche doit donc être suivie dans le cadre de ce mémoire. Il n'existe pas moins de 80 segments de provisionnement dont les lignes $\mathcal{CT}$, $\mathcal{MT}$ et $\mathcal{LT}$. Il est donc primordial d'associer chaque sinistre à un segment de provisionnement car toute la modélisation porte sur cette séparation des sinistres.

Cependant, dans la base *Sinistres*, les segments de provisionnement ne sont pas renseignés. Afin d'associer chaque sinistre à ces derniers, une fusion entre les tables *Sinistres* et *Traités* est indispensable. La clé de jointure est la variable « Identifiant du traité ». Grâce à cet identifiant qui est la granularité la plus fine de la base *Traités*, chaque sinistre est attribué à sa ligne d'activité. Il permet également d'attribuer à chaque sinistre les variables *Part signée* et *Part nette*, utiles aux calculs des différents types de montants à charge des traités dans la sous-section suivante.

Les tables des figures 3.11 et 3.12 ci-dessous sont respectivement des exemples des bases *Traités* et *Sinistres* avant la fusion. La nouvelle table *Sinistres* résultante de cette fusion est en figure 3.13 ci-après.

Identifiant du traité	Segment de provisionnement	Part signée	Part nette
T56	$\mathcal{CT}$	80%	0%
T33	$\mathcal{MT}$	55%	25%

FIGURE 3.11 – Extrait de la base *Traités*

Numéro du sinistre	Identifiant du traité
S403	T56
S805	T33
S678	T56

FIGURE 3.12 – Extrait de la base *Sinistres*

Numéro du sinistre	Identifiant du traité	Segment de provisionnement	Part signée	Part nette
S403	T56	$\mathcal{CT}$	80%	0%
S805	T33	$\mathcal{MT}$	55%	25%
S678	T56	$\mathcal{CT}$	80%	0%

FIGURE 3.13 – Extrait de la base *Sinistres* après la fusion avec la base *Traités*

3.3.4 Calcul de la charge des sinistres à charge des traités

Au sein de cette sous-section, les charges des sinistres à charge des traités, c'est-à-dire celles pertinentes dans le cadre de la réassurance, sont calculées selon deux perspectives différentes pour chaque année de développement. La première vision est celle à l'acceptation d'AXA SA, la deuxième vision est celle à la charge d'AXA SA.

Pour information, dans la suite de ce mémoire, les différentes charges du sinistre S à l'année de développement j sont désignées par les notations suivantes :

- $Charge_{100,S,j}$: la Charge 100% – introduite juste après ;
- $Charge_{FGU,S,j}$: la Charge FGU ;
- $Charge_{ACC,S,j}$: la Charge à l'acceptation d'AXA SA ;
- $Charge_{ASA,S,j}$: la Charge à la charge d'AXA SA.

Pour obtenir les charges selon les deux perspectives mentionnées ci-dessus, il est nécessaire de calculer au préalable la charge du sinistre touchant le traité. Cette charge est appelée « **Charge 100%** » et est définie différemment selon le type de couverture.

- Si le traité de réassurance est un EP ou un QP, le calcul est :

$$Charge_{100,S,j} = Charge_{FGU,S,j} \times \alpha$$

Avec α : le taux de cession EP ou QP selon le traité de réassurance.

- Si le traité de réassurance est un Excédent de Sinistres (que ce soit par risque ou par événement) ou un SL, la formule suivante est appliquée :

$$Charge_{100,S,j} = \min [\max \{ Charge_{FGU,S,j} \times (1 - \alpha_{QP} - \alpha_{EP}) - Priorité_T ; 0 \} ; Portée_T]$$

Lorsqu'il n'y a pas de portée renseignée, alors il suffit de faire :

$$Charge_{100,S,j} = \max \{ Charge_{FGU,S,j} \times (1 - \alpha_{QP} - \alpha_{EP}) - Priorité_T ; 0 \}$$

Une fois la Charge 100% calculée pour chaque traité touché par un sinistre, il est possible de calculer les charges à l'acceptation d'AXA SA et celles à la charge d'AXA SA.

La charge du sinistre touchant le traité T , à l'acceptation d'AXA SA, est appelée « **Charge ACC** ». Elle est définie de la façon suivante :

$$Charge_{ACC,S,j} = Charge_{100,S,j} \times Part \text{ signée}_T$$

La charge du sinistre touchant le traité T , à la charge d'AXA SA, est appelée « **Charge**

ASA ». Elle est définie de la façon suivante :

$$Charge_{ASA,S,j} = Charge_{100,S,j} \times Part\ nette_T$$

Le tableau en figure 3.14 ci-dessous est l'exemple récapitulatif d'un sinistre fictif S347 survenu en 2020 regroupant la plupart des notions mentionnées ci-dessus. La part signée du programme de réassurance est 30% et la part nette du programme est 20%. Le taux de cession QP α_{QP} est 40% et le taux de cession EP α_{EP} est 0%.

Identifiant du traité	Année de développement	Type de traité	Priorité du traité	Portée du traité	Charge FGU	Incréments Charge FGU
T173	1	QP	NA	NA	460 000	460 000
T173	2	QP	NA	NA	460 000	0
T173	3	QP	NA	NA	552 000	92 000
T173	4	QP	NA	NA	644 000	92 000
T173	5	QP	NA	NA	644 000	0
T281	1	XS	92 000	1 012 000	460 000	460 000
T281	2	XS	92 000	1 012 000	460 000	0
T281	3	XS	92 000	1 012 000	552 000	92 000
T281	4	XS	92 000	1 012 000	644 000	92 000
T281	5	XS	92 000	1 012 000	644 000	0
Identifiant du traité	Charge 100%	Incréments Charge 100%	Charge ACC	Incréments Charge ACC	Charge ASA	Incréments Charge ASA
T173	184 000	184 000	55 200	55 200	36 800	36 800
T173	184 000	0	55 200	0	36 800	0
T173	220 800	36 800	66 240	11 040	44 160	7 360
T173	257 600	36 800	77 280	11 040	51 520	7 360
T173	257 600	0	77 280	0	51 520	0
T281	184 000	184 000	55 200	55 200	36 800	36 800
T281	184 000	0	55 200	0	36 800	0
T281	239 200	55 200	71 760	16 560	47 840	11 040
T281	294 400	55 200	88 350	16 590	58 900	18 390
T281	294 400	0	88 350	0	58 900	0

FIGURE 3.14 – Tableau récapitulatif du calcul des différentes charges

Afin de connaître la Charge 100%, la Charge ACC et la Charge ASA d'un sinistre dans sa globalité pour une année de développement donnée, il suffit de sommer respectivement chacune de ces charges pour chaque traité que ce sinistre touche.

Dans l'exemple en figure 3.14 ci-dessus, les différentes charges totales en euros du sinistre S347 pour la cinquième et dernière année de développement sont les suivantes (avec les couleurs correspondantes dans le tableau) :

- Charge totale 100% = 257 600 + 294 400 = 552 000 ;
- Charge totale ACC = 77 280 + 88 350 = 165 630 ;
- Charge totale ASA = 51 520 + 58 900 = 110 420.

En contrôlant la valeur des Charges ACC et ASA calculées à partir de la base *Sinistres* avec les montants correspondants de la base *Sinistres en vision comptable*, les résultats sont identiques pour la plupart des sinistres. Lorsqu'il y a un écart, celui-ci est identifiable et s'explique par l'absence de prise en compte de certaines conditions des traités dont les AAD, les AAL et les priorités et portées indexées. Dans ces cas-là, les montants calculés à charge des traités sont surestimés mais la raison est connue.

Un premier axe d'amélioration pour le calcul des charges dans ce mémoire serait donc la prise en compte des montants indexés (déjà abordés dans la sous-section 3.1.2), des AAD, des AAL et des reconstitutions pour chaque traité. Les Charges ACC et ASA seraient diminuées grâce à ces options.

Les charges ACC et ASA trouvent leur utilité dans la visualisation des sinistres et de leurs développements mais aussi dans la finalité de ce mémoire. En effet, les travaux visent à estimer la somme de la différence entre la charge à l'ultime (ACC ou ASA) et la dernière charge connue (ACC ou ASA) pour chaque sinistre IBNeR et la somme de la charge à l'ultime (ACC ou ASA) de chaque sinistre IBNyR. L'addition de ces sommes correspond alors aux provisions IBNR (ACC ou ASA) estimées dans le cadre de ce mémoire :

$$\begin{aligned}
 Provisions_{IBNR,i} &= Provisions_{IBNeR,i} + Provisions_{IBNyR,i} \\
 &= \sum_{S \in \mathcal{S}_{IBNeR,i}} \sum_{T \in \mathcal{T}} Charge_{T,ultime,S} - \sum_{S \in \mathcal{S}_{IBNeR,i}} \sum_{T \in \mathcal{T}} Charge_{T,J_i,S} \\
 &\quad + \sum_{S \in \mathcal{S}_{IBNyR,i}} \sum_{T \in \mathcal{T}} Charge_{T,ultime,S}
 \end{aligned} \tag{3.1}$$

- J_i : la dernière année de développement connue pour les sinistres survenus en année i ;
- $\mathcal{S}_{IBNeR,i}$ ou $\mathcal{S}_{IBNyR,i}$: l'ensemble des sinistres IBNeR ou IBNyR survenus en année i ;
- $\mathcal{T}$: l'ensemble des traités associés aux sinistres appartenant à l'ensemble $\mathcal{S}_{IBNeR,i}$ ou $\mathcal{S}_{IBNyR,i}$;
- $Provisions_{IBNR,i}$, $Provisions_{IBNeR,i}$ ou $Provisions_{IBNyR,i}$: les provisions IBNR, IBNeR

- ou IBNyR pour l'année de survenance i ;
- $Charge_{T,ultime,S}$ ou $Charge_{T,J_i,S}$: la Charge ACC ou ASA à l'ultime ou en année de développement J_i pour le traité T touché par le sinistre S .

Ces provisions IBNR estimées dans le cadre de ce mémoire sont alors comparées aux provisions IBNR à l'acceptation et à la charge d'AXA SA calculées par l'équipe Reserving avec des méthodes agrégées. Cette comparaison permet ainsi d'interpréter les provisions IBNR calculées par l'équipe avec des charges individuelles de sinistres estimées dans ce mémoire.

Conclusion du chapitre

Afin de faciliter la mise en place du modèle de provisionnement individuel pour les IBNeR et les IBNyR, une structure uniforme pour chaque segment de provisionnement est créée à partir des différentes bases de données, agrémentées des nouvelles variables présentées dans ce chapitre. Cette structure est affichée sur une application **RShiny**, qui est un package de R permettant de faire des visualisations. Cette application a été développée au cours du stage pendant lequel ce mémoire a été rédigé. Elle permet de retracer avec transparence les calculs des Charges 100%, ACC et ASA sinistre par sinistre en affichant entre autres les variables nécessaires, telles que la portée, la priorité, la part signée et la part nette. L'un des intérêts de l'application **RShiny**, complétée avec la génération des scénarios de développement des sinistres IBNeR et IBNyR par la suite, réside dans l'explication des montants présents dans les triangles de développement car toutes les étapes de calculs sont présentes. De plus, cette application conçue au cours du stage permet de vérifier de manière efficace la concordance des montants des tables avec ceux de la base *Sinistres en vision comptable*, qui est affichée sinistre par sinistre. L'équipe Reserving l'a déjà intégrée dans sa propre application **RShiny**.

Chapitre 4

Génération de scénarios de développement des sinistres IBNeR

Ce quatrième chapitre se penche sur la modélisation des développements des sinistres IBNeR afin d'estimer leurs provisions pour les années de développement futures.

Les sinistres IBNeR – *Incurred But Not enough Reserved* – font partie des sinistres présents dans la base *Sinistres* introduite dans le chapitre 3. Ces sinistres IBNeR n'ont potentiellement pas encore atteint leur Charge FGU ultime à leur dernière année de développement connue. Cela signifie qu'ils peuvent ne pas être encore entièrement développés et peuvent être réévalués au cours de leurs années de développement futures. Ces sinistres peuvent donc faire l'objet de provisions IBNeR à mettre de côté par AXA SA pour qu'elle puisse honorer ses engagements en cas de besoin d'indemnisation.

Pour estimer le montant des provisions pour chaque sinistre IBNeR, plusieurs étapes sont nécessaires. Tout d'abord, il convient de déterminer pour chaque sinistre s'il relève ou non de la catégorie des sinistres IBNeR, c'est-à-dire des sinistres dont la Charge FGU est supposée pouvoir encore se développer. Ensuite, pour chaque sinistre identifié comme IBNeR, sa Charge FGU ultime est simulée 1 000 fois, générant ainsi 1 000 scénarios de développement pour chaque sinistre IBNeR. Enfin, à chaque simulation, la provision pour chaque sinistre IBNeR est calculée en appliquant les conditions des traités de réassurance à la Charge FGU ultime simulée. Ces nombreuses simulations permettent de construire la distribution des provisions IBNeR pour un ensemble de sinistres.

Dans les sections suivantes, chaque étape est expliquée plus en détail.

4.1 Étude préliminaire des sinistres clos et actifs

La première section porte sur l'identification des sinistres IBNeR. Cette démarche est très importante puisqu'elle conditionne tout le reste de la modélisation, que ce soit celle du développement des sinistres IBNeR ou de la Charge FGU estimée des sinistres IBNyR. Dans le contexte de ce chapitre, plus le nombre de sinistres IBNeR identifiés augmente, plus il y a de sinistres susceptibles d'être développés, entraînant potentiellement une hausse des provisions IBNeR. En considérant que la variable Statut de la base de données *Sinistres* indique les sinistres qui n'ont potentiellement pas encore atteint leur ultime si elle vaut « Actif » et ceux l'ayant déjà atteint si elle vaut « Clos »⁶, les sinistres actifs correspondraient donc exactement aux sinistres IBNeR.

Il paraît primordial de tester la fiabilité de cette variable Statut, compte-tenu de son importance dans le modèle. L'analyse effectuée permet de vérifier si les sinistres clos ont effectivement atteint leur ultime et si les sinistres actifs attendent encore potentiellement un développement.

Cette étude consiste à comparer le pourcentage de sinistres actifs et clos pour chaque année de survenance de chaque segment de provisionnement.

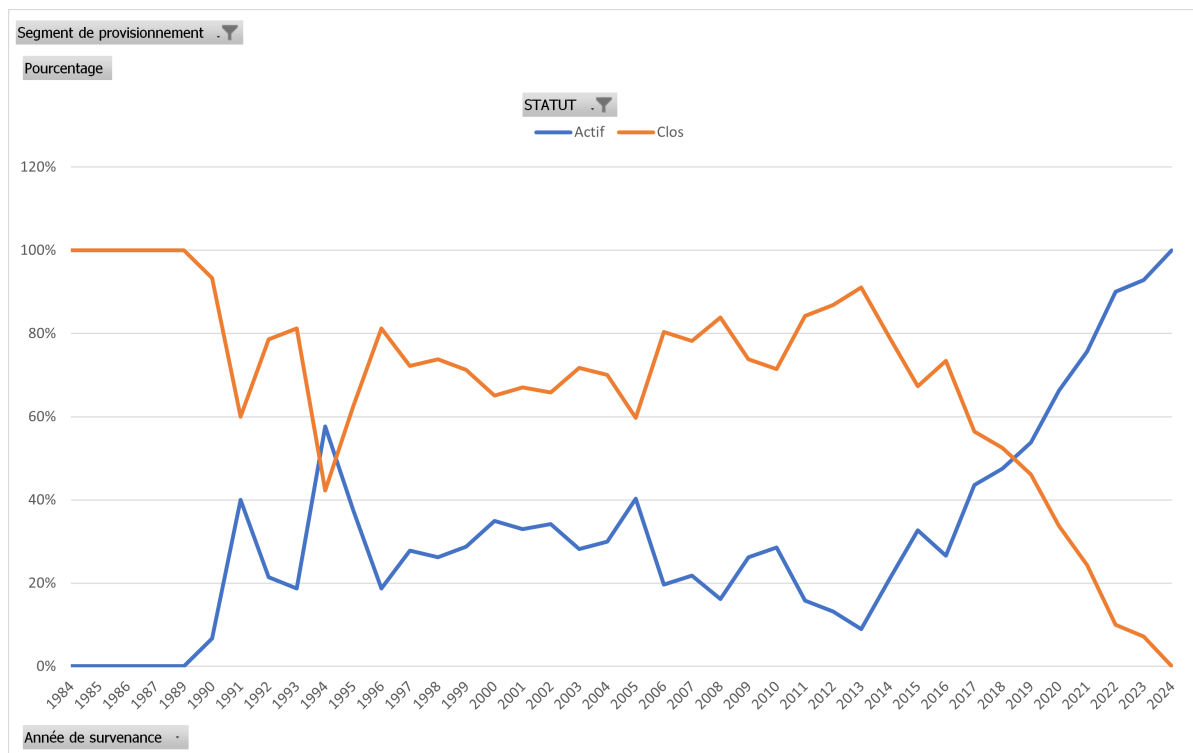
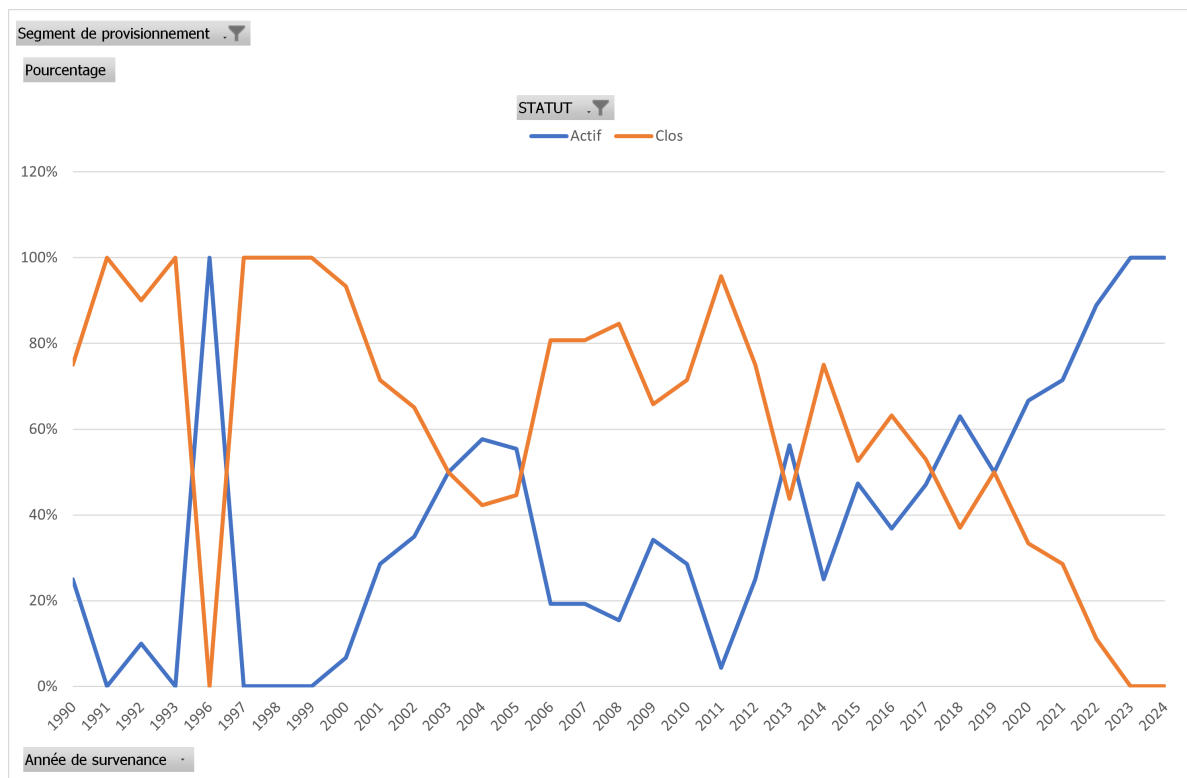
Tout d'abord, un tableau recensant le nombre de sinistres par année de survenance et par valeur de la variable Statut – tel que celui en figure 4.1 – est construit pour chaque segment de provisionnement.

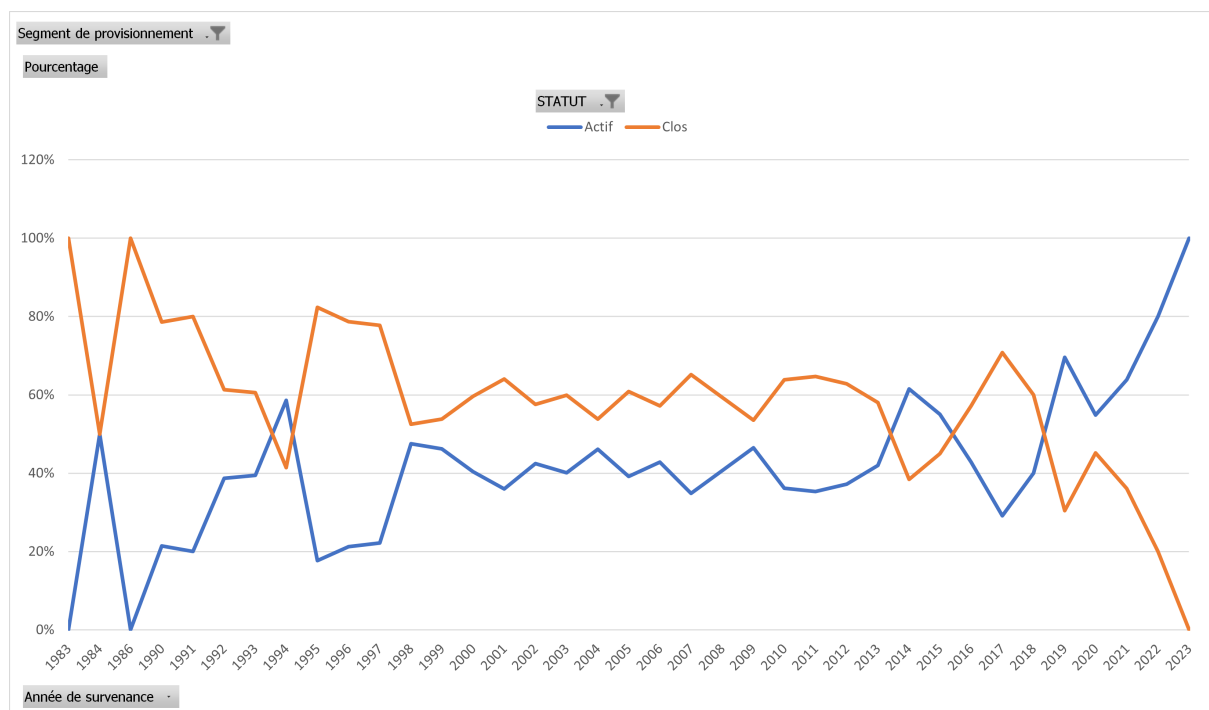
Année de survenance	Statut	Nombre de sinistres	Pourcentage
2007	Clos	3	$\frac{3}{4} \times 100 = 75\%$
2007	Actif	1	25%
2010	Clos	1	50%
2010	Actif	1	50%
2014	Actif	1	100%
2022	Actif	1	100%

FIGURE 4.1 – Nombre de sinistres et pourcentage selon la variable Statut

Les trois graphiques en figures 4.2, 4.3 et 4.4 ci-dessous représentent les pourcentages de sinistres actifs et clos par année de survenance pour les trois segments de provisionnement.

6. Par la suite, les sinistres sont directement désignés par actifs ou clos, sans référence explicite à la variable Statut.

FIGURE 4.2 – Pourcentage de sinistres actifs et clos pour le segment $\mathcal{CT}$ FIGURE 4.3 – Pourcentage de sinistres actifs et clos pour le segment $\mathcal{MT}$

FIGURE 4.4 – Pourcentage de sinistres actifs et clos pour le segment $\mathcal{LT}$

Les courbes bleues correspondent aux pourcentages de sinistres actifs et les courbes orange à ceux des sinistres clos. De prime abord, le pourcentage de sinistres actifs est parfois assez élevé dans les années de survenance les plus anciennes. Cela implique qu'un nombre significatif de sinistres survenus lors de ces anciennes années sont encore actifs, en comparaison du nombre de sinistres clos. Pourtant, plus l'année de survenance est ancienne, plus le pourcentage de sinistres actifs devrait intuitivement avoir une tendance à la baisse. Cette observation entraîne un premier doute sur la fiabilité de la variable Statut, justifiant l'analyse suivante sur les trois segments de provisionnement. Cette analyse examine certains sinistres anciens encore actifs afin de comprendre pourquoi leur pourcentage reste important et ne décroît plus. Le tableau en figure 4.5 détaille les pourcentages des plus anciennes années de survenance du segment $\mathcal{LT}$.

Année de survenance	Statut	Nombre de sinistres	Pourcentage
1983	Clos	1	100%
1984	Actif	2	50%
1984	Clos	2	50%
1986	Clos	1	100%
1990	Actif	3	21,43%
1990	Clos	11	78,57%

FIGURE 4.5 – Nombre de sinistres actifs et clos du segment $\mathcal{LT}$

Ce tableau indique que 3 sinistres sur 14 survenus en 1990 ne sont pas encore clos. L'application RShiny, construite dans le cadre de la préparation du modèle, permet de retrouver facilement ces sinistres, le développement de leur Charge FGU et le pourcentage de leur provision D/D. Il s'agit des sinistres S10035, S2539 et S2544. Les tableaux en figures 4.6, 4.7 et 4.8 présentent les variables associées à ces trois sinistres.

Année de développement	Charge FGU	Pourcentage de la provision D/D
1	595 777	100%
2	639 565	100%
...	...	...
5	639 565	100%
6	760 955	0%
7	760 955	0%
8	760 955	0%
9	760 978	0%
10	760 978	0%
11	800 502	0%
...	...	...
35	800 502	0%

FIGURE 4.6 – Données associées au sinistre actif S2539

Année de développement	Charge FGU	Pourcentage de la provision D/D
1	0	NA
2	1 006 944	100%
3	660 389	100%
...	...	...
9	730 198	0%
...	...	...
15	511 769	54%
16	534 655	53%
17	660 505	59%
18	606 556	54%
...	...	...
35	606 556	54%

FIGURE 4.7 – Données associées au sinistre actif S2544

Année de développement	Charge FGU	Pourcentage de la provision D/D
1	0	NA
...	...	...
13	0	NA
14	100 000	100%
15	100 000	100%
16	100 000	100%
...	...	...
31	100 000	100%
32	100 000	100%
33	100 000	100%
34	100 000	100%
35	100 000	100%

FIGURE 4.8 – Données associées au sinistre actif S10035

Ces trois sinistres n'ont pas été réévalués depuis plus de 15 ans. En effet, les Charges FGU de couleur bleue sont identiques pour un même sinistre. Même si le segment $\mathcal{LT}$ est considéré par les experts comme étant à développement long terme, un délai de plus d'une dizaine d'années sans réévaluation peut paraître excessif. Il convient donc de vérifier si ces sinistres devraient être considérés clos et non encore actifs dans la modélisation des sinistres IBNeR.

Il paraît alors judicieux de regarder la valeur de la variable « Pourcentage de la provision D/D » construite à la section 3.3.2. D'après des experts, plus ce pourcentage est faible pour un sinistre, moins il est probable que ce sinistre soit soumis à des réévaluations dans ses années de développement futures. Parmi ces trois sinistres, seul le sinistre S2539 affiche des pourcentages de provisions D/D de 0% lors de ses dernières années de développement, indiquant alors que

l'indemnisation de ce sinistre envers l'assuré est achevée. Il serait alors pertinent de considérer le sinistre S2539 comme ayant atteint son ultime. Pour les deux autres sinistres, ce pourcentage dépasse 50%. Il vaut même 100% pour le sinistre S10035, ce qui signifie que l'assuré n'a pas encore été indemnisé, pas même partiellement. Bien que ces sinistres S10035 et S2544 n'aient pas évolué depuis plusieurs dizaines d'années, il paraît cohérent de les considérer encore actifs au vu de leur pourcentage de provisions D/D, comme l'indique effectivement la variable Statut.

Le tableau en figure 4.9 concerne le segment de provisionnement $\mathcal{CT}$. C'est un extrait du tableau dans lequel sont calculés les pourcentages de sinistre actifs et clos pour chaque année de survenance d'un segment de provisionnement.

Année de survenance	Statut	Segment de provisionnement	Nombre de sinistres	Pourcentage
1984	Clos	$\mathcal{CT}$	4	100%
1985	Clos	$\mathcal{CT}$	5	100%
1986	Clos	$\mathcal{CT}$	3	100%
1987	Clos	$\mathcal{CT}$	2	100%
1989	Clos	$\mathcal{CT}$	7	100%
1990	Actif	$\mathcal{CT}$	1	6,67%
1990	Clos	$\mathcal{CT}$	14	93,33%

FIGURE 4.9 – Nombre de sinistres actifs et clos du segment $\mathcal{CT}$

Ce tableau indique qu'**1** sinistre survenu en 1990 est encore actif. L'intérêt se porte alors sur celui-ci. Il s'agit du sinistre S5575. Sa Charge FGU n'a pas connu de développement depuis presque 30 ans. Le tableau en figure 4.10 indique les développements historiques de ce sinistre ainsi que ses pourcentages de provisions D/D.

Année de développement	Charge FGU	Pourcentage de la provision D/D
1	0	NA
...	...	...
6	0	NA
7	100 000	0%
...	...	...
35	100 000	100%

FIGURE 4.10 – Données associées au sinistre actif S5575

Le sinistre S5575, survenu en 1990 mais présent dans la base *Sinistres* seulement depuis 1996, n'a donné lieu à aucune réévaluation et a été entièrement indemnisé dès son apparition

dans la base en 1996. Il paraît alors approprié de le considérer clos pour la modélisation des sinistres IBNeR.

Concernant le segment de provisionnement $\mathcal{MT}$, 25% de ses sinistres survenus en 1990 sont encore actifs, comme l'indique le tableau en figure 4.11.

Année de sur-venance	Statut	Segment de provisionnement	Nombre de sinistres	Pourcentage
1990	Actif	$\mathcal{MT}$	2	25%
1990	Clos	$\mathcal{MT}$	6	75%

FIGURE 4.11 – Nombre de sinistres actifs et clos du segment $\mathcal{MT}$

Les deux sinistres actifs, représentant les 25% des sinistres, sont les S34670 et S39633. Le développement de leur Charge FGU, selon les années de développement, et leurs pourcentages de provisions D/D sont présentés dans les deux tableaux en figures 4.12 et 4.13.

Année de développement	Charge FGU	Pourcentage de provisions D/D
1	0	NA
...	...	...
11	0	NA
12	515 752	90%
13	520 816	90%
14	520 816	90%
15	520 816	91%
16	589 402	90%
17	589 402	90%
18	81 066	15%
...	...	...
35	81 066	15%

FIGURE 4.12 – Données associées au sinistre actif S34670

Année de développement	Charge FGU	Pourcentage de provisions D/D
1	0	NA
...	...	...
11	0	NA
12	226 301	98%
13	232 913	93%
14	235 540	92%
...	...	...
17	235 540	92%
18	25 540	23%
...	...	...
34	25 540	23%
35	25 540	23%

FIGURE 4.13 – Données associées au sinistre actif S39633

Il est intéressant de remarquer que les Charges FGU de ces deux sinistres n'ont pas évolué

depuis une quinzaine d'années, ces charges constantes étant de couleur bleue. Pourtant, leurs pourcentages de provisions D/D diffèrent. La conclusion sur ces deux sinistres concernant la variable Statut est donc différente :

- Le sinistre S39633 semble effectivement actif puisque son dernier pourcentage de la provision D/D connu est encore plutôt élevé.
- Le sinistre S34670 est censé être clos, contrairement à ce qui est renseigné dans la base *Sinistres*, au vu de ses données individuelles.

Après ces diverses observations, la variable Statut apparaît comme étant erronée pour certains sinistres. Cette variable est renseignée manuellement, ce qui renforce le fait que des incohérences peuvent se glisser dans la base de données. Pour parer l'inexactitude de cette variable et donc un potentiel biais dans les estimations des provisions IBNeR, il est raisonnable de créer une autre variable afin d'apporter une correction à cette variable Statut. Cette démarche fait l'objet de la section suivante. Les six sinistres mentionnés précédemment sont ultérieurement examinés à nouveau afin de vérifier que la nouvelle variable créée respecte bien les attentes définies pour caractériser un sinistre ayant atteint l'ultime.

4.2 Identification des sinistres IBNeR

Dans la section précédente, il s'est avéré que la variable Statut n'était pas assez fiable pour distinguer les sinistres qui n'avaient pas atteint leur ultime de ceux qui l'avaient atteint.

Initialement, il était envisagé d'identifier les sinistres actifs comme les sinistres IBNeR puisque les deux se rapportent à des sinistres n'ayant potentiellement pas atteint leur ultime. Cependant, après l'analyse présentée dans la section précédente, une telle équivalence devient aberrante à effectuer.

Dans le cadre de la modélisation des sinistres IBNeR, ceux-ci doivent donc être identifiés de manière plus précise que ce que fournissent les valeurs en l'état de la variable Statut. Cette nouvelle spécification résulte de l'application de conditions correctives sur la variable Statut. En effet, l'information fournie par la variable Statut n'est pas entièrement à écarter.

Le résultat de cette nouvelle identification est déterminé pour chaque année de développement de tous les sinistres présents dans la base *Sinistres*. Ce résultat est assigné à une nouvelle variable appelée « Ultime Atteint ». Cette variable prend deux valeurs :

$$Ultime\ Atteint_{S,j} = \begin{cases} 1 & \text{si } Charge_{FGU,S,j} = Charge_{FGU,S,ultime} \\ 0 & \text{si } Charge_{FGU,S,j} \neq Charge_{FGU,S,ultime} \end{cases}$$

Si $Ultime\ Atteint_{S,j}$ vaut 1, cela signifie que le sinistre S a atteint son ultime à son année

de développement j ou antérieurement. Si cette variable vaut 0, cela signifie que le sinistre n'a pas encore atteint son ultime à son année de développement j .

Ainsi, en plus d'apporter des corrections par rapport à la variable Statut sur le fait qu'un sinistre ait atteint son ultime ou non, la variable Ultime Atteint – également binaire – donne une information supplémentaire. Cette indication complémentaire est celle de la première année de développement pour laquelle le sinistre a atteint son ultime.

Pour prendre un exemple, le tableau en figure 4.14 recense deux sinistres fictifs et leur valeur pour la variable Ultime Atteint pour chacune de leurs années de développement.

Numéro de sinistre	Année de surveillance	Année de développement	Ultime Atteint
S462	2021	1	0
S462	2021	2	0
S462	2021	3	1
S462	2021	4	1
S111	2022	1	0
S111	2022	2	0
S111	2022	3	0

FIGURE 4.14 – Variable Ultime Atteint associée aux sinistres

La valeur de la variable Ultime Atteint à la dernière année de développement indique d'une manière générale si le sinistre a déjà atteint son ultime – si elle vaut 1 – ou non – si elle vaut 0. Dans cet exemple, le sinistre S111 attend encore de potentiels développements tandis que le sinistre S462 a déjà atteint son ultime. Il est même possible de savoir qu'il l'a atteint à sa troisième année de développement.

Pour identifier les sinistres IBNeR, il est suffisant de connaître la valeur de la variable Ultime Atteint à la dernière année de développement de chacun des sinistres de la base *Sinistres*. Les sinistres IBNeR sont ceux dont la variable Ultime Atteint vaut 0.

Les sinistres actifs sont donc assimilés aux sinistres pour lesquels la variable Ultime Atteint est égale à 0 à leur dernière année de développement connue. Les sinistres clos sont ceux pour laquelle elle vaut 1.

La règle d'attribution de la valeur 0 ou 1 pour la variable Ultime Atteint du sinistre S à l'année de développement j est la suivante :

Si :

- *Le link ratio cumulé du sinistre S vaut 1 pour l'année de développement j et pour toutes les suivantes ;*

Et :

- La variable *Statut* du sinistre S vaut « Clos » ;

Ou :

- La variable *Statut* du sinistre S vaut « Actif » ;
- L'année de développement la plus récente est supérieure au seuil de développement du segment de provisionnement ;
- Le pourcentage de la provision D/D du sinistre S est inférieur à 20% pour l'année de développement j ;

Alors :

- La variable ***Ultime Atteint*** vaut **1** pour l'année de développement j du sinistre S .

Sinon, ou si :

- La variable *Statut* du sinistre S vaut « Clos » ;
- Le link ratio cumulé du sinistre S est différent de 1 pour l'année de développement j ;

Ou :

- La variable *Statut* du sinistre S vaut « Actif » ;
- L'année de développement la plus récente est inférieure au seuil de développement du segment de provisionnement ;

Alors :

- La variable ***Ultime Atteint*** vaut **0** pour l'année de développement j du sinistre S .

La définition et la construction du seuil de développement sont abordées dans les prochaines sous-sections.

4.2.1 Définition du seuil de développement

La variable « Seuil de développement », ou plus simplement le seuil de développement, est propre à chaque segment de provisionnement. Cette variable correspond à l'année de développement à partir de laquelle les Charges FGU historiques de tous les sinistres sont développées. Sa construction est détaillée par la suite.

Le seuil de développement est défini à partir des Charges FGU agrégées par année de survenance et par année de développement, c'est-à-dire :

$$Charge_{FGU,agreg,i,j} = \sum_{S \in \mathcal{S}_i} Charge_{FGU,S,i,j}$$

Avec $\mathcal{S}_i$: l'ensemble des sinistres survenus en année i .

Pour déterminer le seuil de développement, il est nécessaire de calculer au préalable les cadences de développement des Charges FGU pour chaque année de développement historique de chaque segment de provisionnement. La cadence de développement en année de développement j est définie comme suit :

$$Cadence_j = \frac{1}{Facteur\ lissé_j}$$

Avec $Facteur\ lissé_j$: le facteur de développement lissé des Charges FGU à l'année de développement j .

Les facteurs de développement lissés de chaque année de développement j sont déterminés à partir des facteurs de développement historiques des Charges FGU de la même année de développement j ⁷. Les deux prochaines sous-sections se focalisent sur la construction de ces facteurs de développement historiques et lissés.

4.2.2 Calcul des facteurs de développement historiques

Les facteurs de développement historiques associés aux années de développement j , notés f_j , sont calculés d'après la méthode de la *Chain Ladder*. Cette dernière fait partie des méthodes classiques de provisionnement et est développée plus précisément en annexe B.

Le principe de la méthode de la *Chain Ladder* pour estimer les montants futurs des sinistres repose en partie sur une étape de calcul de facteurs de développement. C'est sur cette étape que se base la détermination des f_j dans le cadre de ce mémoire. Les autres étapes de la méthode de la *Chain Ladder* ne sont pas exploitées ici pour les raisons exposées ci-après. Les trois hypothèses, énoncées en annexe B, sont nécessaires à la stabilité des résultats de la méthode de la *Chain Ladder*. Elles sont exprimées sous forme textuelle dans les points suivants ([Baradel, 2024](#)) :

1. Les années de survenance des sinistres sont indépendantes.
2. La variation du montant, en moyenne pour l'année de survenance i lors de l'année de développement $j + 1$, est proportionnelle à l'année de développement j .
3. Pour l'année de survenance i , la variance du montant lors de l'année de développement $j + 1$ est proportionnelle au montant de l'année de développement j , ce qui revient implicitement à avoir un grand portefeuille diversifié.

Les deux dernières hypothèses sont rarement pleinement satisfaites en réassurance puisque le nombre de sinistres est faible. La méthode de la *Chain Ladder* devient alors moins robuste et plus volatile, comme c'est le cas pour toutes les méthodes classiques. C'est l'une des raisons pour

7. Dans toute la suite, le terme « Charges FGU » est sous-entendu lorsqu'il est question de facteur(s) de développement et n'est plus explicitement mentionné.

lesquelles la modélisation des provisions IBNeR et IBNyR dans le cadre de ce mémoire ne repose pas exclusivement sur cette méthode. Un autre argument en faveur de l'utilisation d'outils complémentaires est que cette méthode s'appuie sur des données agrégées, ne permettant pas une interprétation individuelle des provisions IBNR calculées par l'équipe Reserving.

La méthode de la *Chain Ladder* est tout de même exploitée pour le calcul des f_j dans ce mémoire puisqu'elle offre une première vision globale de la sinistralité pour chaque segment de provisionnement, utile à la modélisation du seuil de développement. Cette vision globale s'obtient à travers ces facteurs de développement historiques reflétant la tendance de développement des sinistres. Un ajustement introduit dans le cadre de ce mémoire est que la valeur minimale des facteurs de développement est de 1, afin de ne pas considérer de baisse de sinistralité.

Le tableau en figure 4.15 ci-dessous recense les facteurs de développement historiques f_j calculés selon la méthode de la *Chain Ladder* pour le segment de provisionnement $\mathcal{CT}$. Le nombre de facteurs de développement est égal à la profondeur historique du segment.

Année de développement	Facteur de développement historique	Année de développement	Facteur de développement historique	Année de développement	Facteur de développement historique
1	1,364	9	1,006	18	1,000
2	1,096	10	1,016	19	1,000
3	1,090	11	1,001	20	1,012
4	1,019	12	1,000	21	1,000
5	1,000	13	1,020	22	1,000
6	1,000	14	1,000	23	1,002
7	1,000	15	1,000	24	1,000
8	1,000	16	1,000	...	...
		17	1,000	41	1,000

FIGURE 4.15 – Facteurs de développement historiques

Les facteurs f_j sont calculés uniquement pour les années de développement des sinistres historiques du segment. La profondeur historique des segments de provisionnement ne permet pas de calculer les facteurs de développement au-delà de cette profondeur. Pour reprendre l'exemple du segment $\mathcal{CT}$, cette dernière est de 41 années puisque ses sinistres les plus anciens sont survenus en 1984. Pourtant, les sinistres les plus anciens du segment de provisionnement peuvent ne pas encore avoir atteint leur ultime. Ainsi, il est important d'estimer des facteurs de développement futurs au-delà des facteurs de développement historiques. Ces facteurs futurs sont appelés **facteurs de queue**. Leur estimation fait l'objet de la sous-section suivante.

4.2.3 Estimation des facteurs de queue

Les facteurs de queue sont estimés à partir d'une courbe de paramétrisation construite sur la base des facteurs de développement historiques cumulés.

Avant d'aborder la notion de paramétrisation, il est essentiel de définir les facteurs de développement historiques cumulés, notés $f_{j,cumul}$. Ils correspondent aux produits des facteurs de développement historiques f_j issus de la méthode de la *Chain Ladder*. Ils sont calculés comme suit, avec D la profondeur historique :

$$f_{j,cumul} = \prod_{k=j}^D f_k$$

Concernant la paramétrisation, le principe est de modéliser une relation entre un ensemble de données empiriques $(X, Y) = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ à l'aide d'une fonction paramétrique dépendant de paramètres $\Theta = \{\theta_1, \theta_2, \dots, \theta_m\}$. Les paramètres Θ sont déterminés à partir d'une méthode d'optimisation de manière à ce que cette fonction s'ajuste le mieux aux données empiriques. Cet ajustement est donc effectué à partir de ces dernières. La courbe de paramétrisation est représentée par cette fonction paramétrique f telle que :

$$Y \approx f(X, \Theta)$$

À partir de cette fonction paramétrique f , il est possible de prédire des valeurs de données non observées. Dans le contexte de l'estimation des facteurs de queue, l'ensemble (X, Y) représente chaque facteur de développement historique f_j associé à son année de développement j . Le tableau récapitulatif est en figure 4.16.

Dans le cadre théorique	Dans le cadre du mémoire
X	L'ensemble des années de développement historiques
Y	L'ensemble des facteurs de développement historiques associés à leur année de développement
$(X, Y) = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$	$\{(1, f_1), (2, f_2), \dots, (D, f_D)\}$ avec D le nombre d'années maximales de développement historiques, c'est-à-dire la profondeur historique
Θ	$\{\theta_1, \theta_2\}$
Données non observées	Facteurs de développement futurs associés à leur année de développement

FIGURE 4.16 – Comparaison de la paramétrisation entre le cadre théorique et celui du mémoire

Afin d'obtenir la meilleure fonction paramétrique – en termes d'ajustement aux données empiriques – pour l'ensemble des données concernant les facteurs de développement historiques f_j , quatre courbes de paramétrisation sont testées parmi les suivantes :

- Fonction de type puissance inverse : $f(j; \theta_1; \theta_2) = 1 + \frac{\theta_1}{j^{\theta_2}}$
- Fonction de type exponentielle : $f(j; \theta_1; \theta_2) = 1 + \theta_1 \cdot \exp(-\theta_2 \cdot j)$
- Fonction de type puissance : $f(j; \theta_1; \theta_2) = \theta_1^j$
- Fonction de type Weibull : $f(j; \theta_1; \theta_2) = \frac{1}{1 - \exp(-\theta_1 \cdot j^{\theta_2})}$

Les valeurs optimales de θ_1 et θ_2 sont déterminées pour chacune de ces courbes de paramétrisation. Parmi ces courbes ajustées, la fonction paramétrique retenue est celle dont le R^2 est le plus élevé et idéalement proche de 1.

Le R^2 est un indicateur statistique de la qualité de l'ajustement de la fonction paramétrique aux données empiriques. Il correspond au pourcentage de variation de la valeur prédite par le modèle de paramétrisation par rapport aux données empiriques. Sa valeur maximale est 1. Plus R^2 est proche de 1, meilleur est l'ajustement. Il se calcule comme suit :

$$R^2 = \frac{\text{Variation}_{\text{expliquée}}}{\text{Variation}_{\text{totale}}} = 1 - \frac{SS_{\text{résiduelle}}}{SS_{\text{totale}}}$$

Avec :

- $SS_{\text{résiduelle}}$: la somme des carrés des écarts entre les valeurs observées y_i et les valeurs prédites par le modèle $\hat{y}_i$, c'est-à-dire $SS_{\text{résiduelle}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$;
- SS_{totale} : la somme des carrés des écarts entre les valeurs observées y_i et leur moyenne $\bar{y}$, c'est-à-dire $SS_{\text{totale}} = \sum_{i=1}^n (y_i - \bar{y})^2$.

Le tableau en figure 4.17 recense les valeurs du R^2 pour les trois segments.

Segment de provisionnement	Fonction de type puissance inverse	Fonction de type exponentielle	Fonction de type puissance	Fonction de type Weibull
$\mathcal{CT}$	98%	88%	65%	92%
$\mathcal{MT}$	98%	90%	9%	96%
$\mathcal{LT}$	99%	82%	3%	95%

FIGURE 4.17 – Comparaison des R^2 entre les différentes fonctions paramétriques

Comme le R^2 est plus élevé – tout en étant également proche de 1 – dans le cas des fonctions paramétriques ajustées de type puissance inverse, c'est la fonction paramétrique retenue pour chacun des trois segments de provisionnement. La courbe de paramétrisation permet alors d'estimer les facteurs de queue jusqu'à l'année de développement choisie. Pour ce modèle, elle

est fixée à 30. Cela signifie que si un segment de provisionnement a une profondeur historique de 20 années de développement, 10 facteurs de queue sont estimés. Les facteurs de développement futurs au-delà de la 30^{ème} année de développement sont considérés comme étant 1.

Par ailleurs, les valeurs de l'application de la fonction paramétrique aux années de développement historiques sont également conservées, lissant ainsi les facteurs de développement historiques. Pour des questions de simplicité, tous les facteurs – les facteurs de développement lissés et les facteurs de queue – sont dénommés facteurs de développement lissés par la suite. Les facteurs de développement lissés dorénavant estimés, il est possible de calculer les cadences de développement pour chacun d'entre eux.

Les facteurs de développement lissés ainsi que les cadences de développement en pourcentage des trois segments de provisionnement sont en figures 4.18, 4.19 et 4.20.

Année de dév.	Facteur lissé	Cadence
1	1,792	55,8%
2	1,199	83,4%
3	1,096	91,23%
4	1,057	94,5%
5	1,040	96,1%
6	1,029	97,1%
7	1,023	97,8%
8	1,018	98,2%
9	1,015	98,6%
10	1,012	98,8%
11	1,010	99,0%
12	1,008	99,2%
13	1,007	99,3%
14	1,006	99,4%
15	1,005	99,5%
16	1,004	99,5%
17	1,004	99,6%
18	1,003	99,7%
19	1,003	99,7%
20	1,002	99,8%
...	...	...
41	1,000	100%

FIGURE 4.18 – Facteurs lissés et cadences du segment $\mathcal{CT}$

Année de dév.	Facteur lissé	Cadence
1	1,747	57,2%
2	1,281	78,0%
3	1,167	85,7%
4	1,117	89,5%
5	1,089	91,9%
6	1,070	93,4%
7	1,057	94,6%
8	1,048	95,4%
9	1,040	96,1%
10	1,035	96,6%
11	1,030	97,1%
12	1,026	97,4%
13	1,023	97,8%
14	1,020	98,0%
15	1,017	98,3%
16	1,015	98,5%
17	1,013	98,7%
18	1,012	98,8%
19	1,010	99,0%
20	1,009	99,1%
...	...	...
35	1,000	100%

FIGURE 4.19 – Facteurs lissés et cadences du segment $\mathcal{MT}$

Année de dév.	Facteur lissé	Cadence
1	2,193	45,6%
2	1,501	66,6%
3	1,315	76,0%
4	1,228	81,4%
5	1,177	84,9%
6	1,144	87,4%
7	1,120	89,3%
8	1,101	90,8%
9	1,087	92,0%
10	1,075	93,0%
11	1,066	93,8%
12	1,058	94,5%
13	1,051	95,2%
14	1,045	95,7%
15	1,039	96,2%
16	1,035	96,6%
17	1,030	97,0%
18	1,027	97,4%
19	1,023	97,7%
20	1,020	98,0%
...	...	...
42	1,000	100%

FIGURE 4.20 – Facteurs lissés et cadences du segment $\mathcal{LT}$

D'après les pourcentages de couleur rouge, les cadences de développement en cinquième année de développement informent qu'en majorité, les Charges FGU de tous les sinistres du segment $\mathcal{CT}$ sont développées à plus de 96% tandis que celles du segment $\mathcal{MT}$ sont développées à plus de 91% et celles du segment $\mathcal{LT}$, à seulement 84% environ. Cette différence appuie déjà la tendance de développement plus ou moins longue des segments de provisionnement.

Par ailleurs, ces trois segments sont des cas particuliers pour cette étape de paramétrisation car leur profondeur historique est de plus de 30 ans. Cette partie de la modélisation est donc intéressante uniquement pour lisser les facteurs de développement historiques. Aucun facteur de queue n'est estimé.

De plus, les cadences de développement historiques de l'année de développement j peuvent être calculées pareillement aux cadences de développement, en remplaçant les facteurs de développement lissés par les facteurs de développement historiques cumulés, comme suit :

$$Cadence_{historique,j} = \frac{1}{f_j}.$$

4.2.4 Construction du seuil de développement

Le seuil de développement d'un segment de provisionnement est déterminé à partir des cadences de développement issues des facteurs de développement lissés. Il correspond à l'année de développement à partir de laquelle les sinistres du segment de provisionnement ne se développent plus. Les sinistres sont considérés comme ne se développant plus lorsque la cadence de développement est supérieure à $1 - x$. Ainsi, le seuil de développement est défini tel que :

$$Seuil\ de\ développement = j \quad \text{avec } j = \min\{k \mid Cadence_k \geq 1 - x\}$$

Dans le cadre de ce mémoire, x est fixé à 2,5%. Ainsi, le seuil de développement d'un segment de provisionnement est la première année de développement pour laquelle la cadence de développement de ce segment est supérieure à 97,5%.

Les trois segments de provisionnement ont chacun un seuil de développement différent d'après cette modélisation. Chaque seuil est mis en évidence avec la couleur bleue dans les trois tableaux en figures 4.18, 4.19 et 4.20 ci-dessus. Les valeurs correspondantes sont récapitulées dans le tableau en figure 4.21.

Segment de provisionnement	$\mathcal{CT}$	$\mathcal{MT}$	$\mathcal{LT}$
Seuil de développement	7	13	19

FIGURE 4.21 – Seuils de développement

Les cadences de développement des facteurs de développement lissés et celles des facteurs de développement historiques cumulés sont affichées dans les graphiques en figures 4.22, 4.23 et 4.24 ci-dessous. Ils représentent ces cadences en fonction des années de développement. Sous cette forme, il est plus facile de les comparer, de les visualiser et de les interpréter.

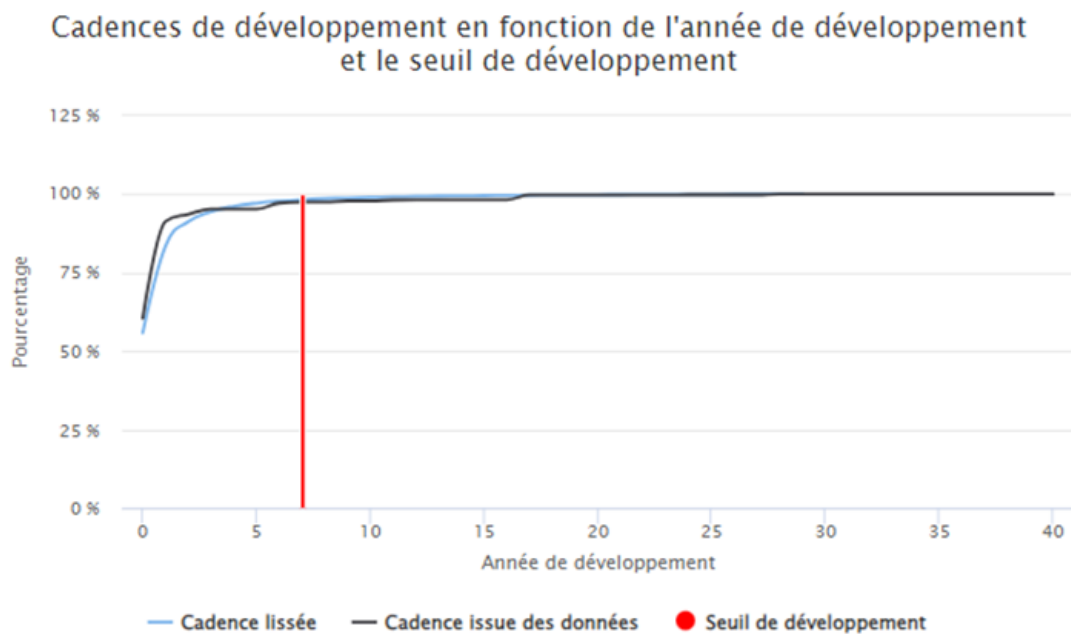


FIGURE 4.22 – Cadences et seuil de développement du segment $\mathcal{CT}$

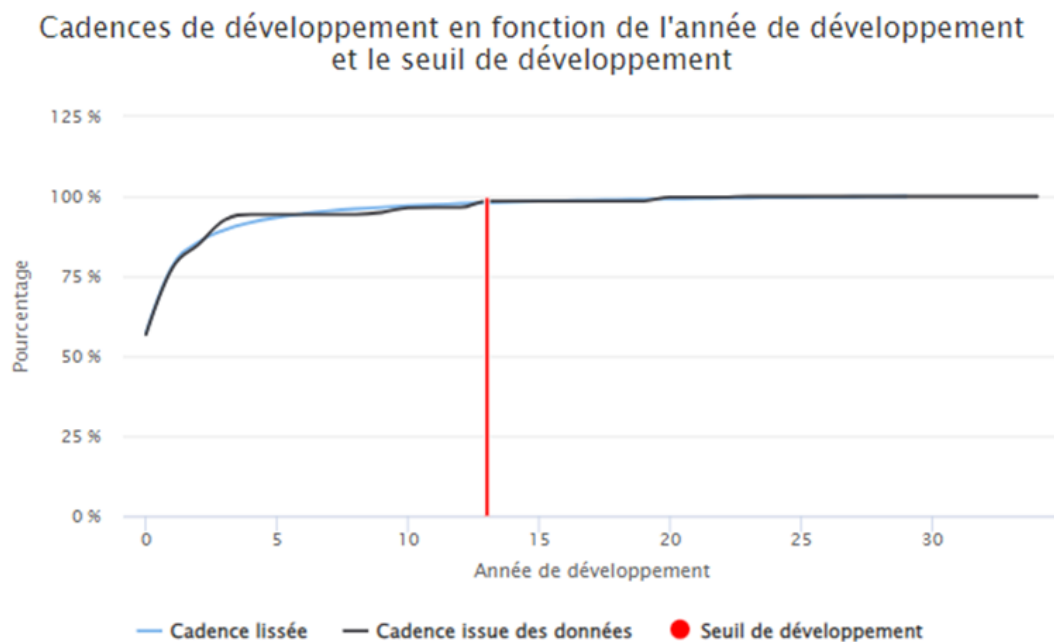
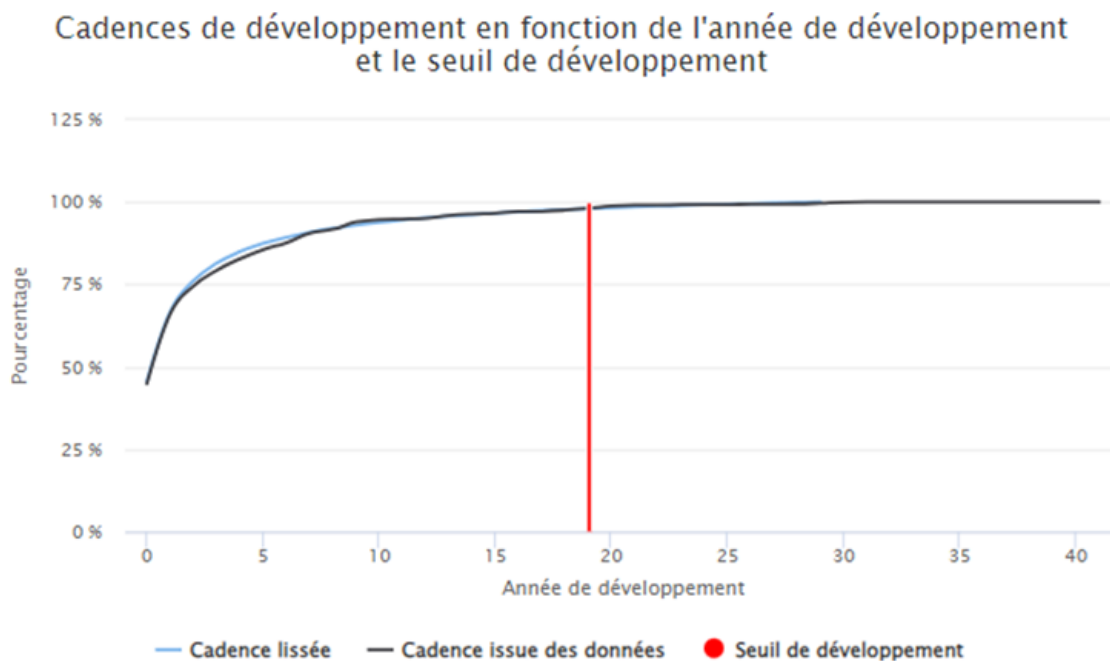


FIGURE 4.23 – Cadences et seuil de développement du segment $\mathcal{MT}$

FIGURE 4.24 – Cadences et seuil de développement du segment $\mathcal{LT}$

Les courbes noires correspondent aux cadences de développement issues des facteurs développement historiques cumulés. Les courbes bleues correspondent quant à elles aux cadences de développement des facteurs de développement lissés. Les barres verticales rouges correspondent au seuil de développement de chaque segment de provisionnement.

Ces résultats sont à nouveau cohérents avec le dire d'experts sur la tendance du nombre d'années de développement avant la clôture des sinistres de ces segments de provisionnement, à savoir court terme pour le segment $\mathcal{CT}$, moyen terme pour le $\mathcal{MT}$ et long terme pour le $\mathcal{LT}$.

4.2.5 Détermination de la variable indiquant si un sinistre a atteint son ultime

La règle permettant de déterminer la valeur de la variable Ultime Atteint, variable indiquant si un sinistre a atteint son ultime pour chaque année de développement historique des sinistres, est énoncée au début de la section 4.2.

Pour rappel, cette variable vaut 1 si le sinistre a déjà atteint son ultime et 0 s'il ne l'a pas atteint, pour une année de développement j donnée. Sa caractérisation repose tout d'abord sur le fait qu'un sinistre soit clos ou actif. S'il est clos, la variable Ultime Atteint vaut 1, sauf si le *link ratio* cumulé est supérieur à 1 à l'année de développement j . Si le sinistre est actif, cette variable vaut 1 à trois conditions :

1. La dernière année de développement connue du sinistre est supérieure au seuil de développement du segment de provisionnement auquel il est associé.
2. Les *link ratios* cumulés du sinistre à partir de l'année de développement j jusqu'à sa dernière année de développement connue sont égaux à 1.
3. Le pourcentage de la provision D/D du sinistre à l'année de développement j est inférieur à 20%.

Les paragraphes suivants expliquent le choix de ces trois conditions une à une.

- Le seuil de développement est l'année de développement au-delà de laquelle les sinistres ne tendent plus à se développer. Tous les sinistres, dont la dernière année de développement connue est supérieure à ce seuil de développement, sont susceptibles d'avoir atteint leur ultime bien qu'ils soient actifs d'après la variable Statut. Deux autres conditions sont néanmoins nécessaires pour que la valeur 1 soit attribuée à un sinistre. Pour les sinistres actifs ne répondant pas à cette première condition, indépendamment des deux autres, le doute persiste quant à la possibilité qu'ils n'aient pas encore atteint leur ultime.
- Concernant la deuxième condition, à première vue, il pourrait paraître approprié de simplement vérifier si la valeur du *link ratio* cumulé à l'année de développement j est égale à 1. Toutefois, ce n'est pas suffisant puisqu'un *link ratio* cumulé valant 1 à l'année de développement j n'implique pas forcément que la Charge FGU du sinistre reste constante à partir de l'année de développement $j - 1$. Il y a trois raisons possibles pour lesquelles un *link ratio* cumulé vaut 1 :
 - Tous les *link ratios* incrémentaux des années de développement à partir de l'année de développement j valent 1. La Charge FGU du sinistre n'évolue alors effectivement plus.
 - Un ou plusieurs *link ratios* incrémentaux à partir de l'année de développement j sont inférieurs à 1. Les autres valent 1. La Charge FGU évolue alors encore à la baisse.
 - Certains *link ratios* incrémentaux sont inférieurs 1, d'autres sont supérieurs à 1. La Charge FGU évolue alors encore à la baisse et à la hausse.

Dans les deux derniers points, le *link ratio* cumulé vaut 1 en raison de sa définition vue en sous-section 3.3.1, qui stipule qu'il ne peut être inférieur à cette valeur. Dans le cadre de ce mémoire, il est opté que la variable Ultime Atteint puisse être égale à 1 – à condition que les deux autres conditions soient satisfaites – uniquement lorsque les *link ratios* cumulés valant 1 répondent au premier point. Cela permet d'éviter de confondre les diminutions des Charges FGU et les sinistres n'ayant plus d'évaluations. Le tableau en figure 4.25 ci-dessous montre un exemple fictif de l'évolution de la valeur de la variable Ultime Atteint d'un sinistre du segment $\mathcal{CT}$ selon ses différentes années de développement.

Numéro de sinistre	Année de développement	Charge FGU cumulée	<i>Link ratio</i> incrémental	<i>Link ratio</i> cumulé	Pourcentage de la provision D/D	Ultime Atteint
S560	1	0	NA	NA	NA	0
S560	2	3 000 000	1	1,166	96%	0
S560	3	4 000 000	1,333	1,166	97%	0
S560	4	2 500 000	0,625	1	1%	0
S560	5	3 500 000	1,4	1,4	0%	0
S560	6	3 500 000	1	1	0%	1
S560	7	3 500 000	1	1	0%	1
S560	8	3 500 000	1	1	0%	1

FIGURE 4.25 – Valeur de la variable Ultime Atteint d'un sinistre selon ses années de développement

Dans cet exemple, le sinistre satisfait la première condition puisque le seuil de développement du segment $\mathcal{CT}$ est 7 et que le sinistre s'étend sur 8 années de développement historiques. À partir de l'année de développement 4, chaque pourcentage de la provision D/D est inférieur à 20%, répondant ainsi à la troisième condition. L'attention est à présent portée sur la valeur des *link ratios* cumulés. Le premier *link ratio* cumulé égal à 1 est rattaché à la quatrième année de développement. Si la deuxième condition était de simplement vérifier qu'un *link ratio* cumulé valait 1, alors la valeur 1 serait attribuée à la variable Ultime Atteint pour cette année de développement. Or, il est possible de constater que la Charge FGU varie encore par la suite, indiquant donc que le sinistre n'a pas atteint son ultime en année de développement 4. Ne pas ajouter de conditions sur les *link ratios* cumulés suivants fausserait alors les valeurs obtenues pour la variable Ultime Atteint. Celle-ci est utile non seulement pour catégoriser un sinistre ayant atteint son ultime, mais aussi pour connaître la première année de développement à laquelle il l'a atteint. Cette première année de développement est répertoriée dans une nouvelle variable. Elle est essentielle à la modélisation et elle conduirait à des provisions IBNeR faussées si elle était erronée. Sa construction est abordée dans la sous-section suivante.

- Concernant la troisième et dernière condition, la barre des 20% pour le pourcentage de la provision D/D est choisie par des experts. Ils considèrent qu'il est fort probable qu'un sinistre ne se développe plus dans ses années de développement futures si son pourcentage de la provision D/D est inférieur à 20%. Dans le cas où ce sinistre a un pourcentage de plus de 20%, cela suggère que sa part restante de Charge FGU à payer est trop importante

pour qu'il soit probable qu'il cesse d'évoluer dans les prochaines années.

Pour comparer les valeurs de la variable Statut et celles de cette nouvelle variable Ultime Atteint, il est intéressant d'analyser les matrices de contingence de chaque segment de provisionnement. Si la variable Statut est fiable pour catégoriser si un sinistre a atteint son ultime ou non d'après les règles établies dans cette modélisation, les sinistres actifs et clos doivent être respectivement assignés aux valeurs 0 et 1 de la variable Ultime Atteint. La correspondance est effectuée uniquement pour la dernière année de développement connue des sinistres. En effet, la variable Statut est indiquée seulement pour l'évaluation la plus récente du sinistre.

Les matrices de contingence pour chaque segment sont en figures 4.26, 4.27 et 4.28.

	0	1
Actif	648	415
Clos	2	2 216

FIGURE 4.26 – Matrice de contingence du segment $\mathcal{CT}$

	0	1
Actif	218	91
Clos	1	498

FIGURE 4.27 – Matrice de contingence du segment $\mathcal{MT}$

	0	1
Actif	1 007	323
Clos	19	1 893

FIGURE 4.28 – Matrice de contingence du segment $\mathcal{LT}$

Les nombres figurant sur la diagonale de chacune des matrices de contingence représentent les nombres de sinistres appartenant à la même catégorie au regard des variables Statut et Ultime Atteint, à savoir :

- La catégorie des sinistres n'ayant pas atteint leur ultime avec les valeurs « Actif » et 0 ;
- La catégorie des sinistres ayant atteint leur ultime avec les valeurs « Clos » et 1.

Le nombre total de sinistres d'une même catégorie, rapporté au nombre total de sinistres du segment, est présenté sous forme de pourcentages dans le tableau en figure 4.29.

Segment de provisionnement	$\mathcal{CT}$	$\mathcal{MT}$	$\mathcal{LT}$
Pourcentage de sinistres appartenant à la même catégorie	87,3%	88,6%	89,4%

FIGURE 4.29 – Tableau des pourcentages de sinistres appartenant à la même catégorie

Les différences de catégorisation résident principalement dans le fait que de nombreux sinistres étant actifs d'après la variable Statut se voient attribuer une valeur de 1 pour la variable Ultime Atteint. À l'inverse, une valeur de 0 pour cette variable est attribuée à très peu de sinistres clos.

Ces matrices de contingence étayaient l'idée que certains sinistres actifs sont peu susceptibles d'évoluer dans leurs prochaines années de développement et qu'il est pertinent de les considérer comme ayant atteint leur ultime. Les inclure dans la modélisation des sinistres IBNeR pourrait alors augmenter considérablement les provisions IBNeR. L'explication des retraitements

visant à éviter au maximum cette situation – celle de développer des sinistres qui ont déjà vraisemblablement atteint leur ultime – est détaillée dans les sous-sections précédentes.

Pour conclure, tous les sinistres IBNeR pris en compte dans le cadre de la modélisation des sinistres IBNeR sont ceux pour lesquels la variable **Ultime Atteint** n'est jamais égale à 1.

4.2.6 Détermination de la variable indiquant la première année d'atteinte à l'ultime

Une variable découlant de la variable **Ultime Atteint** est alors créée : **Première Année Ultime**. Elle correspond à la première année de développement d'un sinistre pour laquelle **Ultime Atteint** vaut 1. C'est la raison pour laquelle il est nécessaire de calculer la variable **Ultime Atteint** pour chaque année de développement de chacun des sinistres. Elle est utilisée ultérieurement dans la modélisation. Lorsque la variable **Ultime Atteint** est égale à 0 pour toutes les années de développement d'un sinistre, alors la variable **Première Année Ultime** correspond au nombre d'années de développement historiques de ce sinistre. Plus formellement :

$$Première\ Année\ Ultime = \begin{cases} \min\{j \in \mathcal{J} \mid Ultime\ Atteint_j = 1\} & \text{si } \exists j \in \mathcal{J} \mid Ultime\ Atteint_j = 1 \\ \max\{\mathcal{J}\} & \text{sinon} \end{cases}$$

Avec $\mathcal{J}$: l'ensemble des années de développement historiques du sinistre.

Les résultats pour les variables **Ultime Atteint** et **Première Année Ultime** sont exposés dans les tableaux en figures 4.30 à 4.35 pour les sinistres cités et étudiés dans la section précédente.

Année de développement	Ultime Atteint	Première Année Ultime
1	0	12
...	...	...
11	0	12
12	1	12
...	...	...
35	1	12

FIGURE 4.30 – Valeurs des variables pour le sinistre S2539

Année de développement	Ultime Atteint	Première Année Ultime
1	0	35
...	...	...
32	0	35
33	0	35
34	0	35
35	0	35

FIGURE 4.31 – Valeurs des variables pour le sinistre S10035

Année de développement	Ultime Atteint	Première Année Ultime
1	0	35
...	...	...
32	0	35
33	0	35
34	0	35
35	0	35

FIGURE 4.32 – Valeurs des variables pour le sinistre S2544

Année de développement	Ultime Atteint	Première Année Ultime
1	0	7
...	...	...
6	0	7
7	1	7
...	...	...
35	1	7

FIGURE 4.33 – Valeurs des variables pour le sinistre S5575

Année de développement	Ultime Atteint	Première Année Ultime
1	0	18
...	...	...
17	0	18
18	1	18
...	...	...
35	1	18

FIGURE 4.34 – Valeurs des variables pour le sinistre S34670

Année de développement	Ultime Atteint	Première Année Ultime
1	0	35
...	...	...
32	0	35
33	0	35
34	0	35
35	0	35

FIGURE 4.35 – Valeurs des variables pour le sinistre S39633

Le tableau en figure 4.36 ci-dessous rappelle les résultats souhaités concernant l'atteinte à l'ultime des sinistres et présente les résultats obtenus pour les variables Ultime Atteint et Première Année Ultime, avec J la dernière année de développement connue du sinistre S .

Numéro de sinistre	S2539	S10035	S2544	S5575	S34670	S39633
Statut	Actif	Actif	Actif	Actif	Actif	Actif
Devrait être considéré comme ayant atteint l'ultime	Oui	Non	Non	Oui	Oui	Non
Valeur attendue pour Ultime Atteint _{S,J}	1	0	0	1	1	0
Valeur calculée pour Ultime Atteint _{S,J}	1	0	0	1	1	0
Valeur calculée de Première Année Ultime _S	12	35	35	7	18	35

FIGURE 4.36 – Tableau récapitulatif des sinistres étudiés

Les valeurs d'Ultime Atteint_{S,J} répondent bien aux attentes fixées. Quant à la variable Première Année Ultime, elle prend la valeur 35 pour les sinistres S10035, S2544 et S39633 car ceux-ci n'ont pas encore atteint leur ultime au cours de leurs années de développement connues et sont survenus en 1990, soit il y a 35 ans.

Ces vérifications fiabilisent la mise en place correcte des deux variables ainsi que leurs résultats. La suite de la modélisation peut donc être envisagée avec davantage de confiance.

4.3 Création de variables explicatives

Pour les différents besoins de la génération des scénarios de développement d'IBNeR, deux autres variables explicatives sont nécessaires : Top Atypique et Top Dvpt Historique. Elles sont décrites dans les deux sous-sections de cette section.

4.3.1 La variable indiquant si un sinistre est atypique

La variable Top Atypique renseigne sur le caractère atypique d'un sinistre. Cette variable binaire est définie comme suit :

$$Top\ Atypique = \begin{cases} 1 & \text{si le sinistre est atypique} \\ 0 & \text{sinon.} \end{cases}$$

Dans le cadre de ce mémoire, un sinistre est considéré comme atypique si, pour au moins une année de développement, sa Charge FGU est supérieure ou égale à la priorité du traité auquel ce sinistre est rattaché. Cependant, lorsque cette priorité est inférieure à 1M€, une règle différente s'applique : le sinistre est considéré comme atypique uniquement si, pour au moins une année de développement, sa Charge FGU dépasse ou égale 1M€. Plus formellement :

$$Top\ Atypique_S = \begin{cases} 1 & \text{si } \max_j \{Charge_{FGU,S,j}\} \geq \max\{1M€, Priorité_T\} \\ 0 & \text{sinon.} \end{cases}$$

Cette catégorisation repose sur la Charge FGU des sinistres par rapport à la priorité des traités, plutôt que sur un seuil fixé. Ce choix s'explique par le fait que les sinistres qui touchent un traité font l'objet d'un traitement spécifique de la part des cédantes.

Pour donner un exemple, considérons le tableau théorique en figure 4.37.

Numéro du sinistre	Identifiant du traité	Priorité du traité	Charge FGU maximum connue	Top Atypique
S103	T56	3 000 000	5 000 000	1
S432	T21	10 000 000	3 000 000	0
S941	T90	500 000	700 000	0
S109	T22	800 000	1 000 000	1

FIGURE 4.37 – Identification des sinistres atypiques et des sinistres non atypiques

Le sinistre S103 est considéré comme atypique car le maximum de sa Charge FGU est supérieur à la priorité du traité, priorité elle-même supérieure à 1M€..

Lorsque la Charge FGU d'un sinistre est inférieure à la priorité du traité auquel ce sinistre est rattaché, comme c'est le cas du sinistre S432, le sinistre n'est pas considéré comme atypique peu importe le montant de la priorité. C'est pourquoi le sinistre S432 n'est pas considéré comme atypique dans le cadre du choix de la modélisation.

Le sinistre S941 n'est pas non plus considéré comme un sinistre atypique car sa Charge FGU est inférieure à 1M€. En effet, peu importe la priorité du traité auquel un sinistre est rattaché, si la Charge FGU de ce sinistre est inférieure à 1M€, alors il sera considéré comme n'étant pas atypique.

Le sinistre S109 est considéré comme atypique puisque sa Charge FGU maximum équivaut à 1M€ et que la priorité du traité auquel il est rattaché est inférieure à 1M€. Cependant, si sa Charge FGU avait été inférieure 1M€ et supérieure à la priorité du traité – 800k€ dans cet exemple –, il n'aurait pas été considéré comme atypique selon la règle énoncée ci-dessus.

4.3.2 La variable indiquant si un sinistre a connu des développements

La variable Top Dvpt Historique, indiquant si un sinistre a connu des développements, est également binaire. Elle renseigne sur les évaluations passées des Charges FGU des sinistres. Plus précisément, pour une année de développement j donnée, elle indique si la Charge FGU d'un sinistre s'est développée de manière significative au moins une fois au cours de ses y années de développement précédant son année de développement j . La variable Top Dvpt Historique vaut 1 si la Charge FGU s'est développée, 0 sinon. Elle est définie formellement comme suit :

$$Top\ Dvpt\ Historique_{S,y,j} = \begin{cases} 1 & \text{si } j \leq y \text{ ou si } \exists d \in \llbracket j - y + 1; j \rrbracket \mid Link\ ratio_{cumul,S,y,d} \geq 1 + x \\ 0 & \text{sinon} \end{cases}$$

Avec :

- y : le nombre fixé de précédentes années de développement à étudier ;
- $x \in \mathbb{R}^+$ et $y \in \mathbb{N}^+$: des constantes ;
- $Top\ Dvpt\ Historique_{S,y,j}$: la valeur de la variable Top Dvpt Historique associée à l'année de développement j du sinistre S en étudiant ses y années de développement précédentes ;
- $Link\ ratio_{cumul,S,y,d}$: le *link ratio* cumulé des y derniers *link ratios* incrémentaux par

rapport à l'année de développement d du sinistre S , défini comme :

$$Link\ ratio_{cumul,S,y,d} = \begin{cases} \max \left\{ 1 ; \prod_{k=d-y+1}^d Link\ ratio_{incr,S,k} \right\} & \text{si } \forall k \in \llbracket d-y+1; d \rrbracket \mid \\ & Link\ ratio_{incr,S,k} \neq \text{NA} \\ \text{NA} & \text{si } \exists k \in \llbracket d-y+1; d \rrbracket \mid \\ & Link\ ratio_{incr,S,k} = \text{NA} \end{cases}$$

Avec $Link\ ratio_{incr,S,k}$: le *link ratio* incrémental de l'année de développement k du sinistre S .

L'introduction du paramètre x permet de considérer exclusivement les développements significatifs des sinistres. Dans le cadre de ce mémoire, ce paramètre est considéré comme étant 0,025, soit 2,5%, tout comme lors de la création de la variable Ultime Atteint dans la section 4.2. Cela signifie qu'un sinistre est considéré comme ayant connu un développement significatif au cours de ses y précédentes années de développement si l'un des *link ratios* cumulés associés à ces années est supérieur à 1,025 ($1 + x = 1 + 0,025$).

La variable Top Dvpt Historique est calculée pour chaque année de développement connue de chaque sinistre pour différents y fixés : 1, 2, 3 et 5.

Le tableau en figure 4.38 est un exemple des valeurs de la variable Top Dvpt Historique pour chaque année de développement d'un sinistre, en prenant $y = 3$ et $x = 0,025$.

Numéro de sinistre	Année de développement	<i>Link ratio</i> cumulé sur 3 années	Top Dvpt Historique
S459	1	NA	1
S459	2	NA	1
S459	3	NA	1
S459	4	1,5	1
S459	5	1	1
S459	6	1,01	1
S459	7	1	0
S459	8	1	0

FIGURE 4.38 – Variable Top Dvpt Historique associée à un sinistre

La variable Top Dvpt Historique vaut 1 pour les trois premières années de développement car ces années de développement sont inférieures ou égales à y , qui vaut 3. Elle vaut également 1 pour les années de développement 4, 5 et 6 car le sinistre s'est développé en année de développement 4. En revanche, même si le *link ratio* associé à l'année de développement 6

est inférieur à 1,025, la Charge FGU du sinistre S459 n'est pas considérée comme s'étant développée significativement entre les années de développement 5 et 6. Ainsi, la variable Top Dvpt Historique prend la valeur 0 pour les années de développement 7 et 8 car le sinistre ne s'est pas développé de manière significative au cours des 3 dernières années de développement précédant respectivement les années de développement 7 et 8.

4.4 Développement des Charges FGU à l'ultime

Dans la section 4.2, les sinistres appartenant à la catégorie des IBNeR sont déterminés et identifiés par l'intermédiaire de la variable Ultime Atteint. Tous les sinistres pour lesquels cette variable vaut 0 sont potentiellement soumis à un développement de leur Charge FGU dans leurs futures années de développement et peuvent donc éventuellement donner lieu à une constitution de provisions IBNeR le cas échéant. Dans le cadre de ce mémoire, la Charge FGU des sinistres IBNeR est modélisée à l'ultime. Le but de cette section est d'explicitier trois différentes méthodes utilisées afin d'obtenir les Charges FGU ultimes des sinistres. Les résultats de ces différentes méthodes sont comparés dans le chapitre 6.

Tout d'abord, il est important de noter qu'il est impossible de connaître avec certitude la valeur à l'ultime d'un sinistre IBNeR. Souhaitant préserver cet aspect aléatoire, des méthodes stochastiques ont été privilégiées aux méthodes déterministes dans le cadre de ce mémoire, que ce soit pour la modélisation des sinistres IBNeR ou des sinistres IBNyR. La méthode pour développer ces derniers est présentée dans le chapitre 5 suivant.

Pour revenir sur les sinistres IBNeR, 1 000 scénarios de développement de leur Charge FGU sont générés pour chacun d'entre eux et pour chacune des trois méthodes. Pour une méthode donnée, un scénario de développement correspond à la simulation des Charges FGU ultimes de tous les sinistres IBNeR étant survenus la même année. L'utilisation de méthodes stochastiques permet de générer plusieurs scénarios offrant chacun une vision différente de l'évolution des sinistres IBNeR. Les Charges FGU ultimes simulées permettent alors de calculer les provisions IBNeR en appliquant les conditions des traités de réassurance associés aux sinistres correspondants.

Le principe clef pour simuler la Charge FGU à l'ultime d'un sinistre IBNeR est l'application d'un coefficient de développement à la dernière Charge FGU connue de ce sinistre. Les coefficients de développement sont principalement basés sur des données historiques. Pour chacune de ses 1 000 simulations, un coefficient de développement est tiré aléatoirement parmi une liste de coefficients, qui est propre à chaque segment de provisionnement. Cependant, il arrive parfois que le coefficient appliqué ne soit pas aléatoire dans certains cas spécifiques abordés dans la suite.

Les coefficients de développement présents dans la liste mentionnée ci-dessus et associés aux années de développement j des sinistres S sont construits de la manière suivante :

$$\text{Coefficient de développement}_{S,j(j \neq 1)} = \begin{cases} \text{Link ratio}_{cumul,S,j} & \text{si } \text{Ultime Atteint}_{S,J} = 1 \\ \text{Link ratio}_{cumul,S,j} \times \text{Facteur lissé}_{S,J} & \text{sinon} \end{cases}$$

Avec :

- $\text{Ultime Atteint}_{S,J}$: la valeur de la variable Ultime Atteint pour la dernière année de développement connue J du sinistre S ;
- $\text{Link ratio}_{cumul,S,j}$: le *link ratio* cumulé associé à l'année de développement j du sinistre S ;
- $\text{Facteur lissé}_{S,J}$: le facteur de développement lissé associé à la dernière année de développement connue J du sinistre S (voir la définition de ces facteurs en section 4.2) ;

Les sinistres à partir desquels sont construits les coefficients de développement sont appelés les **sinistres d'origine**.

Les *link ratios* cumulés existent uniquement pour les années de développement historiques des sinistres d'origine, hormis pour la première année de développement. Plus un sinistre d'origine est ancien, plus son nombre de *link ratios* cumulés est élevé, tout comme le nombre de coefficients de développement qui lui sont associés, et inversement.

Lorsque le sinistre d'origine n'a pas atteint son ultime à sa dernière année de développement connue, c'est-à-dire qu'il fait partie des sinistres IBNeR, ses *link ratios* cumulés sont multipliés par le facteur de développement lissé associé à la dernière année de développement connue de ce sinistre d'origine. Considérer implicitement que la dernière Charge FGU connue est plus élevée que celle factuelle – en multipliant la valeur des *link ratios* cumulés par des nombres supérieurs à 1 – permet de pallier en partie le fait que les *link ratios* cumulés du sinistre d'origine sont potentiellement sous-estimés. Dans le cas où cette multiplication par des facteurs de développement lissés ne serait pas effectuée, les Charges FGU ultimes seraient sous-estimées puisque qu'elles seraient multipliées par des *link ratios* cumulés sous-estimés.

Cette approche permet d'inclure tous les sinistres dans la modélisation, y compris les sinistres IBNeR qui n'ont pas fini de se développer. Ainsi, la liste des coefficients de développement intègre plus de coefficients par rapport à des méthodes qui se limiteraient à tirer aléatoirement parmi des *link ratios* cumulés issus uniquement de sinistres ayant atteint leur ultime.

Le tableau en figure 4.39 ci-dessous reprend le processus d'obtention des coefficients de développement à travers quelques exemples de sinistres d'origine, en considérant que le segment de provisionnement étudié est le segment $\mathcal{CT}$.

Coefficient de développement	Numéro de sinistre d'origine	Année de survenance du sinistre d'origine	Année de développement du sinistre d'origine	Dernière année de dév. connue du sinistre d'origine	Ultime Atteint	Link ratio cumulé du sinistre d'origine	Facteur lissé associé à l'année de développement J
1,351	S561	2000	15	25	1	1,351	1,001
1,563 × 1,005 = 1,571	S888	2010	2	15	0	1,563	1,005
1	S123	2012	10	13	1	1	1,040

FIGURE 4.39 – Obtention des coefficients de développement

L'année de développement du sinistre d'origine correspond à l'année de développement à laquelle est associé le coefficient de développement.

Dans l'exemple ci-dessus, les sinistres S561 et S123 ont atteint leur ultime. Leur coefficient de développement rattaché correspond donc à leur *link ratio* cumulé associé à l'année de développement pour laquelle est calculé ce coefficient. Concernant le sinistre S888, il n'a pas atteint son ultime. Le coefficient de développement qu'il génère, associé à sa deuxième année de développement dans cet exemple, est donc le résultat de la multiplication entre son *link ratio* cumulé à cette même année de développement et le facteur de développement lissé, considéré quant à lui à la dernière année de développement connue du sinistre d'origine.

Lorsqu'un coefficient de développement est tiré aléatoirement afin d'être appliqué à une Charge FGU, les mêmes informations que celles dans le tableau ci-dessus sont disponibles et affichées explicitement dans l'application RShiny. Il s'agit d'informations à propos du sinistre d'origine (son numéro, s'il a atteint son ultime, l'année de développement pour laquelle le coefficient est calculé) et de l'information sur le facteur de développement lissé associé à cette même année de développement. Ces différentes informations offrent une traçabilité des étapes ayant conduit à la valeur de chaque coefficient de développement. Elles contribuent donc à l'interprétabilité et l'explicabilité des résultats, qui sont les objectifs principaux de ce mémoire.

La liste dans laquelle sont recensés tous les coefficients de développement, c'est-à-dire ceux de chaque année de développement historique de chaque sinistre d'un segment de provisionnement, est la suivante :

$$\mathcal{C} = \{ \text{Coefficient de développement}_{S,j} \mid S \in \mathcal{S} \text{ et } j \in \mathcal{J}_S \text{ avec } j \neq 1 \}$$

Avec :

- $\mathcal{S}$: l'ensemble des sinistres existants pour un segment de provisionnement donné ;
- $\mathcal{J}_S$: l'ensemble des années de développement connues du sinistre S .

Cependant, afin de garantir la cohérence de la valeur à l'ultime des sinistres IBNeR, les coefficients de développement appliqués à ces sinistres ne peuvent pas être tirés aléatoirement parmi l'intégralité de cette liste. Pour obtenir les Charges FGU ultimes des sinistres IBNeR, ces derniers sont développés à partir de leur dernière année de développement connue j^* . Par conséquent, les coefficients de développement tirés aléatoirement doivent être exclusivement associés à des années de développement $j^* + 1$. En effet, ces coefficients reflètent le développement souhaité des Charges FGU, c'est-à-dire celui de la Charge FGU d'un sinistre d'origine entre l'année de développement j^* et son ultime. Les coefficients de développement appliqués aux sinistres IBNeR sont donc associés à des sinistres d'origine survenus nécessairement antérieurement à ces sinistres IBNeR à développer et pour lesquels il existe donc un coefficient associé à une année de développement $j^* + 1$. Par exemple, les coefficients de développement appliqués à des sinistres IBNeR survenus en 2020 proviennent uniquement de sinistres d'origine survenus au plus tard en 2019.

La liste contenant tous les coefficients de développement pouvant être appliqués aux sinistres IBNeR dont la dernière année de développement connue est j^* se trouve réduite par rapport à la liste $\mathcal{C}$ contenant l'ensemble des coefficients. Cette liste restreinte est la suivante :

$$\mathcal{C}_{j^*} = \{ \text{Coefficient de développement}_{S,j} \mid j = j^* + 1 \text{ et } S \in \mathcal{S}_{j^*+1} \}$$

Avec :

- $\text{Coefficient de développement}_{S,j}$: le coefficient de développement du sinistre S associé à son année de développement j ;
- $\mathcal{S}_{j^*+1}$: l'ensemble des sinistres pour lesquels l'année de développement $j^* + 1$ existe, défini par :

$$\mathcal{S}_{j^*+1} = \{ S \in \mathcal{S} \mid j^* + 1 \in \mathcal{J}_S \}$$

Avec $\mathcal{J}_S$: l'ensemble des années de développement connues du sinistre S .

L'utilisation de données historiques implique toutefois que les sinistres IBNeR les plus anciens d'un segment de provisionnement ne peuvent pas se voir attribuer un coefficient de développement, par manque de sinistres antérieurs. Par exemple, la profondeur historique du segment de provisionnement $\mathcal{MT}$ est 35. Les sinistres appartenant à ce segment qui sont survenus en 1990 se voient toujours attribuer arbitrairement un coefficient de développement puisqu'il n'existe pas de coefficient de développement associé à l'année de développement 36. Les coefficients de développement attribués aux sinistres IBNeR étant survenus en premier dans

le segment de provisionnement sont alors définis comme suit :

$$\text{Coefficient de développement}_{S,j} = \begin{cases} \text{Facteur lissé}_j & \text{si } \text{Facteur lissé}_j \text{ existe} \\ 1 & \text{sinon.} \end{cases}$$

Dans une telle situation, le sinistre d'origine est alors renseigné comme « Créé » puisque ce coefficient appliqué ne provient pas de données issues d'un sinistre historique.

Les deux tableaux en figures 4.40 et 4.41 ci-dessous servent d'exemple simplifié dans le cadre du segment de provisionnement $\mathcal{CT}$. Le premier recense les sinistres IBNeR et leur dernière année de développement connue, c'est-à-dire celle à partir de laquelle leur Charge FGU doit être développée jusqu'à l'ultime. Le second tableau répertorie les coefficients de développement et quelques variables liées à leur sinistre d'origine. Dans cet exemple, seuls deux sinistres d'origine sont pris en compte par simplification, dont l'un fait partie des sinistres IBNeR à développer. Dans une situation concrète, tous les sinistres IBNeR sont présents dans le deuxième tableau.

Numéro de sinistre	Année de survenance	Dernière année de développement connue	Charge FGU
S877	2015	10	7 600 000
S801	2019	6	2 870 000
S511	2020	5	10 430 000
S442	2020	5	5 460 000
S541	2023	2	5 350 000
S493	2024	1	10 000 000

FIGURE 4.40 – Table des sinistres IBNeR

Coefficient de développement	Sinistre d'origine	Année de survenance	Année de développement	Ultime Atteint	Facteur lissé	Link ratio cumulé du sinistre d'origine
3,521	S567	2014	1	1	1,008	3,521
2,843	S567	2014	2	1	1,008	2,843
2,555	S567	2014	3	1	1,008	2,555
2,403	S567	2014	4	1	1,008	2,403
2,193	S567	2014	5	1	1,008	2,193
2	S567	2014	6	1	1,008	2
1,876	S567	2014	7	1	1,008	1,876
1,702	S567	2014	8	1	1,008	1,702

Coefficient de développement	Sinistre d'origine	Année de survenance	Année de développement	Ultime Atteint	Facteur lissé	<i>Link ratio</i> cumulé du sinistre d'origine
1,696	S567	2014	9	1	1,008	1,696
1,311	S567	2014	10	1	1,008	1,311
1	S567	2014	11	1	1,008	1
1	S567	2014	12	1	1,008	1
3,079	S801	2019	1	0	1,029	2,992
2,641	S801	2019	2	0	1,029	2,567
2,204	S801	2019	3	0	1,029	2,142
1,609	S801	2019	4	0	1,029	1,564
1,718	S801	2019	5	0	1,029	1,670
1,380	S801	2019	6	0	1,029	1,341

FIGURE 4.41 – Table des coefficients de développement et leur sinistre d'origine

Pour le sinistre d'origine S567, les colonnes « Coefficient de développement » et « *Link ratio* cumulé du sinistre d'origine » sont identiques car ce sinistre a atteint son ultime. Ce n'est pas le cas du sinistre d'origine S801 qui fait partie des sinistres IBNeR dont les *link ratios* cumulés sont donc multipliés par un facteur de développement lissé pour obtenir les coefficients de développement associés à ce sinistre.

Les lignes d'une certaine couleur du premier tableau en figure 4.40 ci-dessus, sont associées aux lignes du deuxième tableau en figure 4.41 de ladite même couleur. Dans le premier tableau, les mêmes couleurs sont attribuées aux lignes des sinistres IBNeR ayant la même année de survenance et donc pour lesquels la liste $\mathcal{C}_j^*$ des coefficients de développement est identique. Chacune de ces listes est composée des coefficients du deuxième tableau ayant la même couleur que les lignes des sinistres du premier tableau. Par exemple, les coefficients de développement pouvant être tirés aléatoirement afin d'être appliqués à la dernière Charge FGU connue du sinistre S541 sont les coefficients 2,555 et 2,204, associés respectivement aux sinistres d'origine S567 et S801, ce dernier étant lui-même un sinistre IBNeR (voir les lignes bleues).

Il est intéressant de souligner que, plus les sinistres IBNeR à développer sont récents, plus le nombre de coefficients de développement dans lesquels il est possible de tirer aléatoirement est élevé, en raison du plus grand nombre de sinistres historiques pouvant servir de référence, et inversement.

De plus, lorsqu'il y a plus de 20 coefficients de développement présents dans une liste, les 5% d'entre eux les plus élevés sont supprimés de la liste, afin de ne pas tirer de coefficients

de développement aberrants conduisant à des Charges FGU ultimes aberrantes. En effet, ce choix fait suite à l'observation de coefficients de développement de plusieurs dizaines, voire de plusieurs centaines. C'est par ailleurs une hypothèse importante du modèle. Les nombres 20 et 5 retenus sont choisis par avis d'experts. La liste réduite est alors la suivante :

$$\mathcal{C}_{j^*, \text{réduit}} = \begin{cases} \{ \text{Coefficient de développement} \in \mathcal{C}_{j^*} \mid \text{Coefficient de développement} \leq q_{0,95}(\mathcal{C}_{j^*}) \} & \text{si } |\mathcal{C}_{j^*}| > 20 \\ \mathcal{C}_{j^*} & \text{sinon} \end{cases}$$

Avec :

- $|\mathcal{C}_{j^*}|$: le nombre d'éléments de la liste $\mathcal{C}_{j^*}$;
- $q_{0,95}(\mathcal{C}_{j^*})$: le quantile 95% des coefficients de développement de la liste $\mathcal{C}_{j^*}$.

Il est maintenant temps de décrire les trois méthodes mentionnées au début de cette section dans les trois sous-sections suivantes. Leurs différences se situent dans la liste des coefficients de développement applicables aux sinistres IBNeR ainsi que dans les sinistres IBNeR qui sont à développer. Selon la méthode, la liste peut ne pas correspondre exactement à $\mathcal{C}_{j^*, \text{réduit}}$, comme c'est le cas des méthodes 2 et 3 pour lesquelles la liste est encore plus restreinte.

4.4.1 Méthode 1 : Tirage aléatoire d'un coefficient de développement de la liste réduite

La méthode 1 consiste à attribuer, pour chaque sinistre IBNeR à développer, un coefficient de développement tiré aléatoirement parmi ceux de la liste $\mathcal{C}_{j^*, \text{réduit}}$. Cette liste contient l'ensemble des coefficients de développement applicables aux Charges FGU des sinistres IBNeR à développer. Elle dépend de l'année de survenance de ces derniers.

La liste qui recense les coefficients possibles pour la méthode 1 est la suivante :

$$\mathcal{C}_{j^*, 1} = \mathcal{C}_{j^*, \text{réduit}}.$$

Pour chaque sinistre IBNeR identifié, un coefficient de développement est tiré selon une loi uniforme discrète $\mathcal{U}$ définie sur $\mathcal{C}_{j^*, 1}$ (voir l'annexe A.1). Ce tirage uniforme s'énonce mathématiquement de la façon suivante :

$$c_{h,1} \sim \mathcal{U}(\mathcal{C}_{j^*, 1}) \quad \forall h \in \mathcal{H}.$$

Par conséquent, $c_{h,1}$ représente une réalisation de la variable uniforme discrète définie sur

la liste $\mathcal{C}_{j^*,1}$ associée au sinistre h , avec $h \in \mathcal{H}$, l'ensemble des sinistres IBNeR identifiés.

La situation est identique à l'exemple décrit ci-dessus avec les deux tableaux en figures 4.40 et 4.41.

4.4.2 Méthode 2 : Tirage aléatoire d'un coefficient de développement parmi ceux étant associés à des sinistres d'origine ayant des Charges FGU proches

Tout comme pour les trois méthodes, la méthode 2 consiste à attribuer pour chaque sinistre IBNeR à développer, un coefficient de développement. Ce coefficient est tiré aléatoirement parmi ceux de la liste $\mathcal{C}_{j^*,réduit}$ associés à des sinistres d'origine dont **la Charge FGU à l'année de développement j^* est proche de celle de la Charge FGU du sinistre IBNeR à développer** à sa dernière année de développement connue.

La liste qui recense les coefficients possibles pour la méthode 2 dépend des sinistres IBNeR à développer, à la fois par leur année de survenance mais aussi par leur dernière Charge FGU connue. Cette liste est définie comme suit :

$$\mathcal{C}_{h,j^*,2} = \left\{ \begin{array}{l} \text{Coefficient de développement}_{S,j} \in \mathcal{C}_{j^*,réduit} \mid \\ \frac{\text{Charge}_{FGU,h,j^*} - \text{Charge}_{FGU,S,j^*}}{\text{Charge}_{FGU,h,j^*}} \leq \text{tolérance}, \quad \forall h \in \mathcal{H} \end{array} \right.$$

Avec :

- $\mathcal{H}$: l'ensemble des sinistres IBNeR identifiés ;
- $\text{Charge}_{FGU,h,j^*}$: la Charge FGU du sinistre IBNeR à développer h à l'année de développement j^* qui est sa dernière année de développement connue ;
- $\text{Charge}_{FGU,S,j^*}$: la Charge FGU du sinistre d'origine S à son année de développement j^* ;
- tolérance (tolérance $\in \mathbb{R}^+$) : le seuil permettant de définir si deux Charges FGU sont considérées comme étant proches.

Lorsque le nombre de coefficients de développement présents dans la liste $\mathcal{C}_{h,j^*,2}$ est inférieur à un certain nombre, la tolérance est augmentée. Dans le cadre de ce mémoire, la tolérance est initialement fixée à 0,5 puis est augmentée par paliers de 0,2 si la liste contient moins de 5 coefficients de développement. La tolérance maximum considérée est 10. Ces choix numériques proviennent de dire d'experts.

Pour chaque sinistre IBNeR h , un coefficient de développement est tiré selon une loi uniforme discrète $\mathcal{U}$ définie sur $\mathcal{C}_{h,j^*,2}$. Ce tirage uniforme s'énonce mathématiquement de la façon

suivante :

$$c_{h,2} \sim \mathcal{U}(\mathcal{C}_{h,j^*,2}).$$

Par conséquent, $c_{h,2}$ représente une réalisation de la variable uniforme définie sur la liste $\mathcal{C}_{h,j^*,2}$ associée au sinistre h .

Dans ces méthodes 1 et 2, le coefficient de développement peut valoir 1. Dans le cas où un tel coefficient est appliqué à un sinistre IBNeR, la Charge FGU de ce dernier reste inchangée à l'ultime par rapport à sa dernière année de développement connue. Ce sinistre n'implique donc pas une provision IBNeR.

Ces situations révèlent que tous les sinistres catégorisés IBNeR ne sont pas sujets à de nouvelles évolutions au cours de leurs prochaines années de développement. Cette catégorisation en tant que « sinistres IBNeR » indique seulement que ces sinistres n'ont potentiellement pas atteint leur ultime, sans pour autant garantir qu'ils se développeront de manière certaine dans le futur.

Cette précision permet d'amorcer l'explication de la méthode 3 dans la sous-section suivante.

4.4.3 Méthode 3 : Tirage aléatoire d'un coefficient de développement parmi ceux étant associés à des sinistres d'origine ayant des Charges FGU proches, uniquement pour les sinistres IBNeR sujets à une évolution

Dans les deux premières méthodes décrites, c'est le coefficient de développement tiré aléatoirement qui détermine si un sinistre IBNeR évolue ou non. En revanche, dans cette méthode 3, les sinistres IBNeR sujets à une évolution sont identifiés en amont du tirage aléatoire des coefficients. Seuls ces sinistres se voient attribuer un coefficient de développement tiré aléatoirement. Cette identification se déroule en deux étapes. Ces dernières font l'objet des deux sous-sous-sections de cette présente sous-section.

La liste des coefficients de développement utilisée dans cette dernière méthode reprend tous les principes de celle de la méthode 2, en conservant cependant uniquement les coefficients strictement supérieurs à 1. Cette restriction est nécessaire puisque les sinistres IBNeR sur lesquels sont appliqués les coefficients tirés aléatoirement sont sujets à une évolution certaine.

Tout comme dans le cadre de la méthode 2, la liste qui recense les coefficients possibles pour la méthode 3 dépend des sinistres IBNeR sujets à une évolution, de leur année de survenance et de leur dernière Charge FGU connue. La liste est alors définie de la manière suivante :

$$\mathcal{C}_{k,j^*,3} = \{ \text{Coefficient de développement} \in \mathcal{C}_{h,j^*,2} \mid \text{Coefficient de développement} > 1 \}, \quad \forall k \in \mathcal{K}$$

Avec $\mathcal{K}$: l'ensemble des sinistres IBNeR sujets à une évolution.

Pour chaque sinistre IBNeR k , un coefficient est tiré selon une loi uniforme discrète $\mathcal{U}$ définie sur $\mathcal{C}_{k,j^*,3}$. Ce tirage uniforme s'énonce mathématiquement de la façon suivante :

$$c_{k,3} \sim \mathcal{U}(\mathcal{C}_{k,j^*,3}).$$

Par conséquent, $c_{k,3}$ représente une réalisation de la variable uniforme définie sur la liste $\mathcal{C}_{k,j^*,3}$ associée au sinistre k .

La méthode 3 reprenant le principe de la méthode 2, seul un exemple de la méthode 3 est développé dans la suite, expliquant ainsi les deux méthodes.

L'exemple de la méthode 3 est présenté à travers les deux tableaux en figures 4.42 et 4.43 ci-dessous dans le cadre du segment de provisionnement $\mathcal{CT}$. Le premier tableau recense deux sinistres IBNeR sujets à une évolution, étant survenus la même année mais ayant des Charges FGU différentes à leur dernière année de développement connue. Le second tableau est identique au tableau en figure 4.41 ci-dessus. Il répertorie des coefficients de développement et quelques variables liées à leur sinistre d'origine.

Numéro de sinistre	Année de surveillance	Dernière année de développement connue	Charge FGU
2020	5	S511	1 000 000
2020	5	S442	1 200 000

FIGURE 4.42 – Table des sinistres IBNeR sujets à évolution

Coefficient de développement	Numéro du sinistre d'origine	Année de surveillance	Année de développement	Charge FGU	Ultime Atteint	Facteur lissé	Link ratio cumulé du sinistre d'origine
3,521	S567	2014	1	1 300 000	1	1,008	3,521
2,843	S567	2014	2	1 300 000	1	1,008	2,843
2,555	S567	2014	3	1 300 000	1	1,008	2,555
2,403	S567	2014	4	1 300 000	1	1,008	2,403
2,193	S567	2014	5	1 300 000	1	1,008	2,193
2	S567	2014	6	1 300 000	1	1,008	2
1,876	S567	2014	7	1 300 000	1	1,008	1,876
1,702	S567	2014	8	1 300 000	1	1,008	1,702
1,696	S567	2014	9	1 300 000	1	1,008	1,696

Coefficient de développement	Numéro du sinistre d'origine	Année de surveillance	Année de développement	Charge FGU	Ultime Atteint	Facteur lissé	Link ratio cumulé du sinistre d'origine
1,311	S567	2014	10	1 300 000	1	1,008	1,311
1	S567	2014	11	1 300 000	1	1,008	1
1	S567	2014	12	1 300 000	1	1,008	1
3,079	S801	2019	1	1 600 000	0	1,029	2,992
2,641	S801	2019	2	1 600 000	0	1,029	2,567
2,204	S801	2019	3	1 600 000	0	1,029	2,142
1,609	S801	2019	4	1 600 000	0	1,029	1,564
1,718	S801	2019	5	1 600 000	0	1,029	1,670
1,380	S801	2019	6	1 600 000	0	1,029	1,341

FIGURE 4.43 – Table des coefficients de développement et leur sinistre d'origine

Les deux sinistres du tableau en figure 4.42 sont survenus en 2020 et en sont à leur cinquième année de développement. Les seuls coefficients de développement pouvant appartenir à la liste $\mathcal{C}_{k,j^*,3}$ sont donc les coefficients 2 et 1,380 puisqu'ils sont associés à l'année de développement 6 ($5 + 1$). Ils répondent également au critère de stricte supériorité à 1. En revanche, il convient de vérifier la proximité des Charges FGU entre les sinistres IBNeR sujets à évolution du premier tableau et les sinistres d'origine. Ces calculs sont réalisés uniquement sur les Charges FGU des sinistres d'origine associés aux coefficients de développement déjà présélectionnés :

– Pour le sinistre S511 :

- Proximité avec la Charge FGU du sinistre S567 : $\frac{|1\,000\,000 - 1\,300\,000|}{1\,000\,000} = 0,3 \leq 0,5$;
- Proximité avec la Charge FGU du sinistre S801 : $\frac{|1\,000\,000 - 1\,600\,000|}{1\,000\,000} = 0,6 > 0,5$.

– Pour le sinistre S442 :

- Proximité avec la Charge FGU du sinistre S567 : $\frac{|1\,200\,000 - 1\,300\,000|}{1\,200\,000} \simeq 0,08 \leq 0,5$;
- Proximité avec la Charge FGU du sinistre S801 : $\frac{|1\,200\,000 - 1\,600\,000|}{1\,200\,000} \simeq 0,33 \leq 0,5$.

Pour le sinistre S511, seul le coefficient de développement 2, associé au sinistre S567, peut être tiré aléatoirement, car c'est le seul qui respecte la condition de proximité parmi les deux coefficients (ligne surlignée en jaune). En revanche, pour le sinistre S411, les deux coefficients de développement font partie de la liste puisqu'ils satisfont les deux à la condition de proximité (lignes bleues).

Dans cet exemple simplifié, la condition des 5 coefficients de développement minimum n'est pas remplie (voir la méthode 2). Si cette condition était respectée, ces calculs seraient repris en

appliquant une tolérance de 0,7.

Au cœur de ces trois méthodes, l'utilisation d'une loi uniforme discrète définie sur la liste finie des coefficients de développement permet d'avoir une probabilité de réalisation identique pour chaque coefficient de développement (voir l'annexe A.1).

Le tableau en figure 4.44 résume les coefficients de développement présents dans les différentes listes selon la méthode utilisée.

	Ensemble des coefficients de développement dépendant de l'année de survenance des sinistres IBNeR à développer	Ensemble des coefficients de développement provenant d'un sinistre d'origine dont la Charge FGU est proche de la Charge FGU du sinistre IBNeR à développer	Ensemble des coefficients de développement strictement supérieurs à 1
Méthode 1	Oui	Non	Non
Méthode 2	Oui	Oui	Non
Méthode 3	Oui	Oui	Oui

FIGURE 4.44 – Récapitulatif du tirage aléatoire des coefficients dans les trois méthodes

Dans les trois méthodes présentées ci-dessus, il est possible qu'une liste de coefficients de développement soit vide, lorsqu'aucun coefficient de développement ne répond aux conditions pour y être intégré. Différents cas possibles sont cités dans la suite, sans que la liste ne soit toutefois exhaustive. Tout d'abord, en ce qui concerne les méthodes 2 et 3, cette situation se présente lorsqu'aucun sinistre IBNeR n'est proche d'un sinistre, au sens de leur Charge FGU, malgré l'augmentation de la tolérance. De plus, l'appui sur des données historiques implique que les sinistres IBNeR les plus anciens d'un segment de provisionnement ne peuvent pas se voir attribuer un coefficient de développement, par manque de sinistres antérieurs. Par exemple, la profondeur historique du segment de provisionnement $\mathcal{MT}$ est 35. La liste de coefficients de développement élaborée pour les sinistres appartenant à ce segment et survenus en 1990 est donc vide, puisqu'il n'existe pas de coefficient de développement associé à l'année de développement 36. Les sinistres relevant de ces circonstances se voient alors toujours attribuer arbitrairement un coefficient de développement défini comme suit :

$$\text{Coefficient de développement}_{S,j} = \begin{cases} \text{Facteur lissé}_j & \text{si Facteur lissé}_j \text{ existe} \\ 1 & \text{sinon.} \end{cases}$$

Dans de telles situations, le sinistre d'origine est alors renseigné comme « Créé » puisque ce coefficient appliqué ne provient pas de données issues d'un sinistre historique.

Pour résumer, voici les différentes étapes qui mènent aux listes de coefficients de développement construites, différentes selon les méthodes :

1. **Création de la liste $\mathcal{C}$** : liste contenant l'ensemble des coefficients historiques ;
2. **Élaboration de la liste $\mathcal{C}_{j^*}$** : sous-liste de $\mathcal{C}$ obtenue en conservant uniquement les coefficients de développement applicables aux sinistres IBNeR à développer d'une année de développement donnée j^* ;
3. **Élaboration de la liste $\mathcal{C}_{j^*, \text{réduit}}$** : sous-liste de $\mathcal{C}_{j^*}$ obtenue en retirant les coefficients les plus élevés – dernière étape de la construction de la liste $\mathcal{C}_{j^*, 1}$ pour la méthode 1 ;
4. **Élaboration de la liste finale** : sous-liste de $\mathcal{C}_{j^*, \text{réduit}}$ obtenue en deux étapes :
 - En conservant uniquement les coefficients de développement associés à des sinistres d'origine proches – en termes de Charges FGU – des sinistres IBNeR à développer – dernière étape de la construction de la liste $\mathcal{C}_{h, j^*, 2}$ pour la méthode 2 ;
 - En supprimant les coefficients de développement égaux à 1 – dernière étape de la construction de la liste $\mathcal{C}_{k, j^*, 3}$ pour la méthode 3.

Les deux prochaines sous-sous-sections portent sur l'identification des sinistres IBNeR sujets à évolution. Cette identification repose sur les méthodes de Kaplan-Meier et de régression logistique. D'une perspective globale, l'estimateur de Kaplan-Meier est employé afin de déterminer le nombre de sinistres IBNeR sujets à évolution pour une année de survéance d'un segment de provisionnement. Cependant, cet estimateur n'indique pas les sinistres en question – ceux sujets à évolution – parmi l'ensemble des sinistres IBNeR d'une année de survéance. Les sinistres IBNeR sujets à évolution sont clairement identifiés lors de l'étape de la régression logistique. Cette identification fait suite au tri de tous les sinistres IBNeR d'une année de survéance selon leur probabilité de faire l'objet d'au moins une nouvelle évaluation de leur Charge FGU. Pour illustrer le principe de cette démarche, considérons un ensemble de 10 sinistres IBNeR survenus en 2020 numérotés de S1 à S10. À titre d'exemple, si l'estimateur de Kaplan-Meier indique que 3 sinistres IBNeR sont aléatoirement sujets à évolution sur les 10, la régression logistique permet quant à elle de déterminer leur numéro par la suite – S3, S4 et S7 par exemple.

La méthode de Kaplan-Meier

La méthode de Kaplan-Meier est une approche non-paramétrique d'estimation de la distribution d'une variable aléatoire censurée (voir l'annexe E). Cette méthode est principalement utilisée en **analyse de survie** qui est l'étude de la période de temps s'écoulant avant la survéance d'un événement d'intérêt. Elle permet notamment de définir la probabilité que cet événement se produise après un certain temps. Dans le cas particulier de l'étude de la durée

de vie humaine, la méthode de Kaplan-Meier est par exemple exploitée lors de la construction de tables de mortalité. La suite s'inscrit dans ce contexte lié à la durée de vie humaine.

Le principal outil mathématique exploité pour l'analyse de survie est la **fonction de survie**. Cette dernière représente la probabilité qu'un événement d'intérêt – le décès d'un individu – ne se produise pas avant un certain temps t . De même que la fonction de répartition, cette probabilité définit de manière unique la distribution de la variable aléatoire T représentant la durée avant cet événement d'intérêt. Ladite variable T est appelée **variable de durée** et représente la durée de vie résiduelle d'un individu.

Cependant, les observations supposées comme étant des réalisations de la variable de durée T peuvent être censurées. Les observations correspondent aux informations des individus prenant part à l'étude et permettant d'estimer la distribution de T . La **censure** désigne quant à elle un phénomène parasite perturbant l'étude des données, conduisant à des observations incomplètes et altérant ainsi l'estimation de T . Ces observations sont tout de même des données pertinentes à ne pas écarter, au risque de sous-estimer la valeur des réalisations de T , puisqu'elles indiquent que l'événement d'intérêt n'est pas encore survenu au temps t . Dans la méthode de Kaplan-Meier, ces informations censurées sont prises en compte afin d'éviter au maximum le biais qu'elles peuvent engendrer. La censure constitue l'une des singularités de l'analyse de survie en raison de sa structure temporelle. Dans le cadre de la durée de vie, ce manque d'informations est principalement dû au fait que l'individu ait quitté l'étude avant son décès ou bien qu'il ne soit pas décédé avant la fin de la période d'observation. Ces deux exemples illustrent des cas de **censure à droite**. La variable de censure C intervient alors. Elle correspond à la période d'observation jusqu'à censure des informations. Un individu dont la durée de vie n'est pas censurée est une personne décédée pendant la période d'observation. Plus formellement, deux possibilités existent concernant le phénomène de censure :

- Si $T \geq C$, alors l'observation est incomplète et il y a censure.
- Si $T < C$, alors l'observation est complète.

La fonction de répartition empirique ne peut pas être calculée car les observations ne sont pas toutes complètes. C'est pourquoi il est nécessaire de faire appel à la méthode de Kaplan-Meier pour estimer la distribution de T .

L'hypothèse fondamentale de cette méthode repose sur l'indépendance entre les variables T et C . En d'autres termes, la probabilité qu'une observation soit censurée doit être indépendante de la durée de vie résiduelle des individus.

L'estimateur de Kaplan-Meier fournit l'estimation de la probabilité que l'individu soit encore vivant et ce, pour chaque âge. Ces différentes probabilités peuvent être représentées graphiquement sous forme de courbe en escalier – appelée **courbe de survie** – permettant de visualiser l'évolution de la fonction de survie. La façon dont est construit cet estimateur

implique que les probabilités ne sont pas croissantes avec l'âge mais qu'elles décroissent ou restent constantes.

Dans le cadre de ce mémoire, le but d'avoir recours à la méthode de Kaplan-Meier n'est pas l'estimation de la distribution de la durée de vie humaine mais la distribution du nombre d'années de développement avant qu'un sinistre n'atteigne son ultime pour la première fois. Ce nombre d'années correspond à la variable Première Année Ultime introduite dans la sous-section 4.2.6. Il s'agit en quelque sorte de l'analyse de la durée de vie des sinistres selon leur segment de provisionnement. Il est intéressant de noter que la notion de « durée de vie résiduelle » est encore existante puisque l'étude porte sur le développement futur des sinistres. La censure est également présente avec les sinistres IBNeR puisqu'ils n'ont pas encore atteint leur ultime lors du gel des bases.

Le tableau en figure 4.45 récapitule alors le parallèle effectué entre la durée de vie humaine et l'atteinte à l'ultime des sinistres.

	Kaplan-Meier dans le cadre de la durée de vie humaine	Kaplan-Meier dans le cadre de ce mémoire
Sujets de l'étude	Individus	Sinistres d'un même segment de provisionnement
Événement d'intérêt	Décès	Atteinte de l'ultime
Temps T jusqu'à l'événement d'intérêt	Durée de vie résiduelle	Nombre d'années de développement restantes avant que les sinistres n'atteignent leur ultime
Censure	- Individus étant encore vivants à la fin de la période d'observation - Individus étant sortis de l'étude pour une autre cause que leur décès	Sinistres n'ayant pas encore atteint leur ultime à la fin de la période d'observation
Indépendance entre les variables T et C	Indépendance entre la probabilité qu'une observation soit censurée et la durée de vie résiduelle des individus	Indépendance entre la probabilité qu'une observation soit censurée et le nombre d'années de développement restantes avant que le sinistre n'atteigne son ultime
But de l'étude	Déterminer la probabilité qu'un individu survive selon son âge	Déterminer la probabilité qu'un sinistre atteigne son ultime selon son année de développement

FIGURE 4.45 – Comparaisons de l'utilisation de la méthode de Kaplan-Meier entre le cadre de la durée de vie et celui de l'atteinte à l'ultime des sinistres

Lors de la construction de l'estimateur de Kaplan-Meier dans le cadre de ce mémoire, une disjonction de cas est faite entre les sinistres atypiques et les sinistres non atypiques afin d'être plus précis dans les probabilités estimées. Cette distinction implique qu'il existe deux probabilités de survie des sinistres pour chaque année de développement selon le caractère atypique ou non des sinistres.

Les trois graphiques en figures 4.46, 4.47 et 4.48, sont les courbes de survie des sinistres pour les trois segments de provisionnement constituant le fil rouge de ce mémoire.

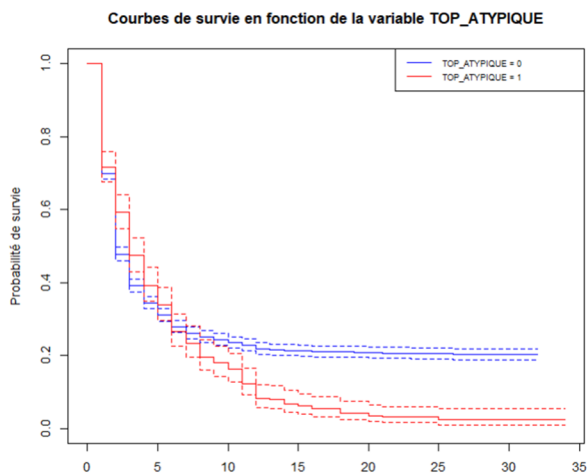


FIGURE 4.46 – Courbes de survie du segment $\mathcal{CT}$

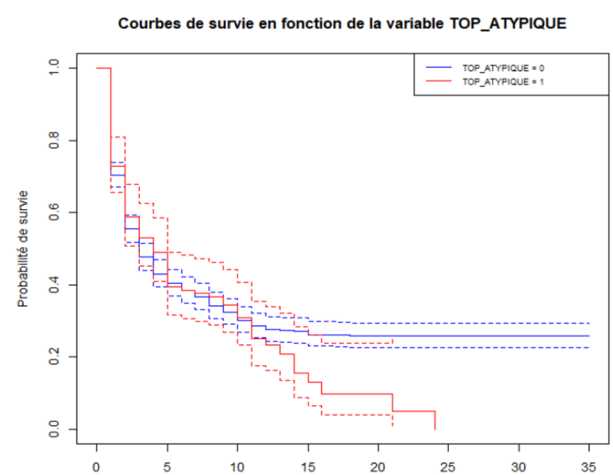


FIGURE 4.47 – Courbes de survie du segment $\mathcal{MT}$

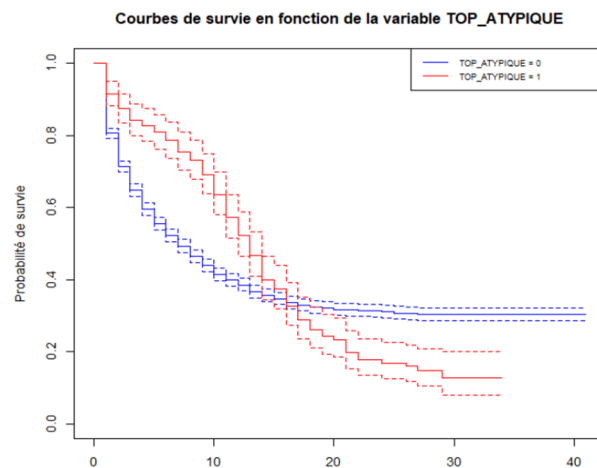


FIGURE 4.48 – Courbes de survie du segment $\mathcal{LT}$

Les années de développement sont sur l'axe des abscisses. Les courbes rouges représentent les fonctions de survie des sinistres atypiques tandis que les courbes bleues représentent celles des

sinistres non atypiques. Les courbes pleines correspondent aux probabilités de survie estimées par la méthode de Kaplan-Meier. Les courbes en pointillés représentent quant à elles les bornes inférieures et supérieures des intervalles de confiance à 95%.

Les courbes pleines rouges et bleues des segments de provisionnement $\mathcal{CT}$ et $\mathcal{MT}$ sont très similaires pour les premières années de développement, jusqu'aux environs de la 8^{ème} année pour le segment $\mathcal{CT}$ et de la 13^{ème} pour le segment $\mathcal{MT}$. Des courbes similaires impliquent des probabilités de survie proches. Cela indique que l'impact du caractère atypique ou non des sinistres est très faible sur leur probabilité de survie sur les premières années de développement des segments de provisionnement court terme.

Quant au segment de provisionnement $\mathcal{LT}$, les deux courbes pleines évoluent différemment jusqu'à la 15^{ème} année de développement environ. En effet, la courbe pleine rouge est en-dessous de la courbe bleue jusqu'à la 15^{ème} année de développement. Les probabilités de survie des sinistres atypiques sont donc plus élevées que celles des sinistres non atypiques. Pour les segments long terme, la distinction entre les sinistres atypiques et ceux non atypiques est donc très pertinente. D'après les experts, les sinistres atypiques sont plus lents à se développer que ceux qui ne le sont pas, en raison de leur Charge FGU plus élevée. Il paraît donc cohérent que leur probabilité de survie soit plus élevée que celle des sinistres non atypiques.

Concernant les années de développement les plus élevées, le comportement des courbes pleines bleues semble anormal. En effet, les probabilités de survie devraient converger vers 0 quel que soit le caractère atypique des sinistres. Cette convergence est observée pour les courbes rouges mais n'est pas constatée pour les courbes bleues qui ont tendance à rester constantes à partir d'une certaine année de développement. Cette stabilité est d'ailleurs la raison pour laquelle les courbes bleues se situent au-dessus des courbes rouges à partir d'une quinzaine d'années de développement pour le segment de provisionnement $\mathcal{LT}$. L'une des explications de cette constance des probabilités de survie des sinistres non atypiques réside dans le fait que ces sinistres ne sont plus actualisés dans les bases de données. En effet, lorsque les gestionnaires savent qu'un sinistre non atypique ne touchera jamais un traité de réassurance, ses informations peuvent ne plus être mises à jour dans les bases de données remontées à Ceded Re. Par conséquent, leur pourcentage de provisions D/D peut ne plus être mis à jour. Ce pourcentage jouant un rôle dans l'identification des sinistres IBNeR, un manque d'actualisation peut donc conduire à une identification erronée. Si ce pourcentage reste supérieur à 20% pour un sinistre actif, ce dernier ne sera jamais considéré comme ayant atteint son ultime. Cette observation permet, d'une part, de constater la possible présence de biais dans l'identification des sinistres IBNeR et d'autre part, de mettre en évidence l'impact de ce biais sur l'estimateur de Kaplan-Meier. Un axe d'amélioration serait de corriger ce biais humain afin d'affiner l'identification des sinistres ayant atteint leur ultime, permettant également d'obtenir un estimateur de Kaplan-

Meier moins biaisé sur les années de développement les plus élevées.

Grâce à ces probabilités de survie des sinistres selon leur année de développement, il est possible d'estimer le nombre de sinistres IBNeR sujets à évolution pour chaque dernière année de développement connue.

L'explication à suivre s'appuie sur un exemple fictif de sinistres appartenant au segment de provisionnement $\mathcal{LT}$ et survenus en 2015. Considérons 50 sinistres survenus en 2015 pour ce segment, parmi lesquels 20 d'entre eux sont identifiés comme ayant atteint leur ultime. Ainsi, 30 sinistres dont la dernière année de développement connue est l'année 10 correspondent à des sinistres IBNeR. 10 d'entre eux sont des sinistres atypiques, impliquant que 20 sinistres IBNeR sont non atypiques.

L'application de la méthode de Kaplan-Meier sur l'ensemble des sinistres de ce segment $\mathcal{LT}$ donne les résultats concrets suivants :

- La probabilité que les sinistres atypiques à l'année de développement 10 n'aient pas atteint leur ultime est de 64%.
- La probabilité que les sinistres non atypiques à l'année de développement 10 n'aient pas atteint leur ultime est de 42%.

Il est donc plus probable que les sinistres atypiques n'aient pas atteint leur ultime en année de développement 10, comparativement à ceux non atypiques.

À partir de ces différents nombres, il est possible de simuler le nombre de sinistres IBNeR sujets à évolution pour cette année de survenance à travers une loi binomiale (voir l'annexe A.2). Le tableau en figure 4.49 présente les données nécessaires à la simulation, en différenciant les sinistres IBNeR en fonction de leur caractère atypique.

	Sinistres IBNeR non atypiques	Sinistres IBNeR atypiques
Nombre de simulations	1 000	1 000
Nombre de sinistres IBNeR pouvant être sujets à évolution	20	10
Probabilité qu'un sinistre IBNeR n'ait pas atteint son ultime à sa dixième année de développement	42%	64%

FIGURE 4.49 – Données nécessaires à la simulation des sinistres IBNeR sujets à développement

1 000 simulations sont réalisées afin d'obtenir des nombres différents de sinistres IBNeR sujets à évolution pour chacun des 1 000 scénarios, permettant de renforcer la dimension stochastique de ce modèle.

Les résultats de la simulation de cette loi binomiale se présentent sous forme de deux vecteurs contenant chacun 1 000 nombres entiers. Les valeurs du vecteur associé aux sinistres IBNeR atypiques sont comprises entre 0 et 10 et celles du vecteur associé aux sinistres IBNeR non atypiques varient entre 0 et 20.

Pour un scénario en particulier, par exemple le cinquième, les résultats obtenus peuvent être les suivants :

- 14 sinistres IBNeR non atypiques ont atteint leur ultime parmi les 20. Il y en a donc 6 encore sujets à évolution.
- 5 sinistres IBNeR atypiques ont atteint leur ultime parmi les 10. Les 5 autres sont donc encore sujets à évolution.

Désormais, la prochaine étape est de déterminer précisément les sinistres n'ayant pas encore atteint leur ultime grâce à l'utilisation de la régression logistique.

La régression logistique

La régression logistique est une méthode statistique. Elle constitue un cas particulier des modèles linéaires généralisés par l'utilisation d'une fonction logistique comme fonction de lien (voir l'annexe F pour plus de détails).

Elle est employée afin d'expliquer une variable binaire Y dépendante d'une ou plusieurs variables explicatives. Plus précisément, elle permet de modéliser la probabilité que l'événement d'intérêt se produise – équivalent à $Y = 1$ – en fonction des valeurs de ces différentes variables explicatives. Y vaut 0 lorsque cet événement ne se produit pas.

Dans le cadre de ce mémoire, la régression logistique est introduite afin d'induire une dépendance entre les sinistres IBNeR d'une même année de survenance et les variables explicatives Pourcentage de la provision D/D et Top Dvpt Historique, définies respectivement à la sous-section 3.3.2 du chapitre précédent et à la sous-section 4.3.2 du présent chapitre. L'idée développée à travers cette régression est que les sinistres d'une même année de survenance n'ont pas tous la même chance de se développer. Par exemple, la régression exploite l'intuition qu'un sinistre avec un pourcentage de provision D/D élevé a plus de chances de se développer qu'un sinistre pour lequel ce pourcentage est faible – d'où l'intégration de la variable « Pourcentage de la provision D/D » en variable explicative du modèle. Cette régression permet donc par la suite de trier les sinistres IBNeR selon leur probabilité d'être sujets à évolution après leur dernière année de développement connue. À nouveau, les segments de provisionnement sont étudiés indépendamment les uns des autres.

La variable binaire Y correspond alors à la variable Ultime Atteint. Les variables retenues pour expliquer Y sont les trois suivantes :

- Année de développement ;
- Top Dvpt Historique avec y égal à 3 ;
- Pourcentage de la provision D/D.

Puisque plusieurs variables explicatives entrent en jeu, il s'agit d'une régression logistique multiple. La formule utilisée pour la régression logistique est la combinaison linéaire suivante :

$$\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = a + bX + cW + d(X \cdot W) + eZ$$

Avec :

- p : la probabilité que $Y = 1$ sachant X , W et Z , autrement dit, p correspond à la probabilité $\mathbb{P}(Y = 1 | (X, W, Z))$;
- a , b , c , d et e : des constantes réelles correspondant aux coefficients de la régression logistique ;
- X : la variable explicative correspondant à l'année de développement ;
- W : la variable explicative correspondant à Top Dvpt Historique avec $y = 3$;
- Z : la variable explicative correspondant au Pourcentage de la provision D/D ;
- bX : l'effet de la variable X sur le *logit* de la probabilité p ;
- cW : l'effet de la variable W sur le *logit* de la probabilité p ;
- $d(X \cdot W)$: l'effet d'interaction conjoint des variables X et W sur le *logit* de la probabilité p ;
- eZ : l'effet de la variable Z sur le *logit* de la probabilité p .

Cette régression logistique est effectuée et appliquée pour chaque groupe de sinistres IBNeR ayant la même dernière année de développement connue.

Concernant la variable Top Dvpt Historique, différentes valeurs de y ont été testées parmi lesquelles 1, 2, 3 et 5. Celle retenue est la valeur qui minimisait l'AIC (*Akaike Information Criterion*, le Critère d'Information d'Akaike en français – voir l'annexe D). Pour les trois segments de provisionnement constituant le fil rouge de ce mémoire, la valeur de y retenue est égale à 3.

Concernant les deux variables Année de développement et Top Dvpt Historique, leur effet conjoint a été examiné au préalable afin de confirmer l'incidence de leur interaction sur la variable Ultime Atteint. Le tableau en figure 4.50 ci-dessous, correspondant au segment de provisionnement $\mathcal{CT}$, est un tableau croisé permettant d'analyser l'influence conjointe de ces deux variables sur la valeur de la variable Ultime Atteint.

Année de développement	Top Dvpt Historique			
	0		1	0 - 1
1			48%	3 222
2			73%	3 222
3			81%	3 221
4	87%	2 765	80%	454
5	90%	2 919	78%	291
6	93%	2 996	76%	185
7	94%	3 025	80%	132
8	95%	2 979	80%	97
9	96%	2 929	85%	68
10	96%	2 883	84%	50
11	97%	2 841	88%	40
12	98%	2 737	84%	32
13	98%	2 692	86%	21
14	98%	2 661	93%	14
15	99%	2 626	91%	11
16	99%	2 553	86%	14
17	99%	2 495	91%	11
18	99%	2 433	90%	10
19	99%	2 359	100%	4
20	99%	2 303	100%	4

Année de développement	Top Dvpt Historique			
	0		1	0 - 1
21	99%	2 240	100%	5
22	99%	2 103	100%	7
23	99%	1 932	100%	7
24	99%	1 744	100%	3
25	100%	1 259	100%	3
26	100%	757	100%	2
27	99%	393	100%	2
28	99%	155	100%	1
29	99%	134	100%	1
30	100%	118	100%	1
31	100%	102	100%	1
32	100%	77		
33	100%	60		
34	100%	46		
35	100%	36		
36	100%	21		
37	100%	14		
38	100%	14		
39	100%	12		
40	100%	9		
41	100%	4		

FIGURE 4.50 – Tableau d'interaction des variables Année de développement et Top Dvpt Historique

Dans ce tableau, la première colonne correspond à l'année de développement. Les autres colonnes sont toutes liées à la variable Top Dvpt Historique.

Les deuxième et quatrième colonnes correspondent au pourcentage de sinistres ayant atteint leur ultime à chaque année de développement :

- Soit parmi tous les sinistres pour lesquels un développement a été observé pour les 3 dernières années de développement (deuxième colonne) ;
- Soit parmi ceux pour lesquels aucun développement significatif n'a été observé pour les 3 dernières années de développement (quatrième colonne).

Dans ces deux colonnes, les pourcentages augmentent de manière significative au fur et à mesure des années de développement. Cette observation implique alors la présence d'influence entre l'année de développement et le nombre de sinistres ayant atteint leur ultime. Il est donc pertinent de faire un croisement entre les années de développement et la variable Top Dvpt Historique. Dans le cas où les pourcentages seraient similaires pour chaque année de développement, les deux variables ne seraient pas conjointement liées et il serait superflu d'ajouter à la régression logistique une composante reflétant leur interaction.

Les troisième et cinquième colonnes représentent le nombre de sinistres associé au pourcentage à leur gauche. De manière générale, il est à noter que ce nombre diminue avec l'année de développement. Cela signifie qu'il y a de moins en moins de sinistres survenus dans les

plus anciennes années. L'addition des nombres de ces deux colonnes renseigne le nombre total de sinistres ayant atteint l'année de développement associée à la ligne. Pour donner un exemple, considérons l'année de développement 7. 94% des sinistres n'ayant pas eu récemment de développement, ont atteint leur ultime à cette année de développement tandis que c'est le cas de seulement 80% des sinistres ayant eu un développement. Les 94% couvrent 3 025 sinistres et les 80% en couvrent quant à eux 132. Il y a donc 3 157 sinistres survenus en 2018 ou antérieurement qui ont atteint leur année de développement 7. La plupart d'entre eux ont déjà atteint leur ultime. La dernière colonne est le résultat de la différence entre les pourcentages des deuxième et quatrième colonnes. Entre les années de développement 4 et 18, la différence se situe entre 6 et 17%. À partir de l'année de développement 19, cette différence tend vers 0. Ce changement est dû au fait que cette année de développement correspond au seuil de développement du segment de provisionnement $\mathcal{LT}$ et que de nombreux sinistres sont désormais considérés comme ayant atteint leur ultime. En effet, tant que les sinistres ouverts n'ont pas atteint le seuil de développement du segment, ils sont considérés comme n'ayant pas atteint leur ultime même si leur pourcentage de provision D/D est inférieur à 20% et que leur *link ratio* cumulé reste constant à 1. Dès que ces sinistres atteignent le seuil de développement, ils sont considérés comme ayant atteint leur ultime puisqu'ils répondent à toutes les conditions.

Pour revenir sur les années de développement 4 à 18, la différence de pourcentage entre les sinistres n'ayant pas eu de développement au cours des 3 dernières années et ceux en ayant eu, est significative puisqu'elle dépasse 5% pour toutes les années de développement. Elle met en évidence l'impact de la valeur de la variable Top Dvpt Historique sur la variable Ultime Atteint. Cette influence conforte le fait d'ajouter l'effet de la variable Top Dvpt Historique à la régression logistique.

Ce tableau croisé ci-dessus permet d'illustrer les nombreuses informations apportées par les variables Année de développement et Top Dvpt Historique, ainsi que par leur interaction puisque les pourcentages diffèrent en fonction des lignes et des colonnes. Cette étude complémentaire accentue ainsi l'intérêt d'intégrer le croisement de ces deux variables dans le modèle de régression logistique.

Une fois que ce modèle de régression s'est entraîné sur des données historiques, les probabilités d'atteinte à l'ultime sont associées aux sinistres IBNeR. Ces probabilités dépendent de la valeur des trois variables explicatives à la dernière année de développement connue de ces sinistres. Il est à noter que ces probabilités sont déterministes. Par conséquent, chaque sinistre se voit systématiquement attribuer la même probabilité quel que soit le scénario. Ces probabilités correspondent à la variable Proba Logit.

Suite à cette attribution, une valeur entre 0 et 1 est générée aléatoirement pour chaque sinistre IBNeR selon une loi uniforme afin d'introduire une composante stochastique au déter-

minisme de la régression logistique. L'ensemble de ces réalisations aléatoires constitue la variable Simul Logit. Puis, une nouvelle variable dépendant de la variable Simul Logit est créée. Il s'agit de la variable Simul Logit Evt et elle est définie comme suit :

$$Simul\ Logit\ Evt = \begin{cases} 1 & \text{si } Proba\ Logit \geq Simul\ Logit \\ 0 & \text{si } Proba\ Logit < Simul\ Logit. \end{cases}$$

Les variables Simul Logit et Simul Logit Evt sont quant à elles propres à chaque scénario puisque leur construction dépend d'une valeur générée aléatoirement.

C'est à partir de la variable intermédiaire Simul Logit Evt que les sinistres IBNeR d'une même année de survenance et d'un même segment de provisionnement peuvent être ordonnés selon leur susceptibilité d'avoir déjà atteint leur ultime. La démarche de classement est décrite ci-dessous.

Les sinistres pour lesquels la variable Simul Logit Evt vaut 1 sont considérés comme étant plus susceptibles d'avoir atteint leur ultime que les sinistres IBNeR pour lesquels cette variable vaut 0. Parmi les sinistres pour lesquels la variable Simul Logit Evt prend la même valeur, les plus susceptibles d'avoir atteint leur ultime sont ceux pour lesquels les valeurs de la variable Proba Logit sont les plus élevées.

Un ordre des sinistres peut alors être établi suivant ces conditions, permettant ainsi de sélectionner les n sinistres IBNeR les plus susceptibles d'être sujets à évolution, avec n le nombre de sinistres estimés par Kaplan-Meier comme n'ayant pas atteint leur ultime.

Il est essentiel de rappeler le fait que l'estimateur de Kaplan-Meier est calculé distinctement pour les sinistres atypiques et non atypiques. Ainsi, l'ordre des sinistres doit également être établi séparément afin de garantir une attribution correcte des sinistres IBNeR sujets à évolution à leur catégorie respective.

Les sinistres figurant en haut de chaque catégorie sont ceux qui sont les plus susceptibles d'avoir atteint leur ultime. Ceux figurant en bas sont donc ceux qui sont le moins susceptibles d'avoir atteint leur ultime, c'est-à-dire ceux qui sont le plus susceptibles d'être sujets à évolution.

Le tableau en figure 4.51 ci-dessous est un exemple de l'ordonnancement des sinistres décrit ci-dessus. Tous les sinistres présents dans ce tableau sont issus d'un même segment de provisionnement, d'une même année de survenance – car ils ont la même dernière année de développement connue – et d'un même scénario.

Numéro de sinistre	Charge FGU	Top Atypique	Proba Lo-git	Simul Logit	Simul Logit Evt
S100	8 000 000	0	0,616	0,606	1
S213	300 000	0	0,593	0,133	1
S003	5 000 000	0	0,457	0,021	1
S043	2 000 000	0	0,457	0,888	0
S967	3 000 000	1	0,842	0,335	1
S403	6 000 000	1	0,457	0,513	0

FIGURE 4.51 – Classement des sinistres IBNeR par l'intermédiaire de la régression logistique

D'après ce tableau, le classement des sinistres IBNeR des plus aux moins susceptibles d'avoir atteint leur ultime est le suivant :

- Pour les sinistres non atypiques : $S100 > S213 > S003 > S043$;
- Pour les sinistres atypiques : $S967 > S403$.

Par conséquent, les sinistres IBNeR des plus ou moins susceptibles d'être sujets à évolution sont ordonnés de la manière suivante :

- Pour les sinistres non atypiques : $S043 > S003 > S213 > S100$;
- Pour les sinistres atypiques : $S403 > S967$.

Le nombre de sinistres sujets à évolution parmi ces 6 sinistres IBNeR est estimé par la méthode de Kaplan-Meier. Considérons que l'estimateur de Kaplan-Meier fournisse les résultats suivants :

- Un sinistre n'a pas encore atteint son ultime parmi les 4 sinistres IBNeR non atypiques ;
- Aucun sinistre n'a encore atteint son ultime parmi les 2 sinistres IBNeR atypiques.

La régression logistique permet d'identifier explicitement les sinistres IBNeR sujets à évolution en tenant compte de ces classements. Dans cet exemple, il s'agit des sinistres :

- S043 pour les sinistres non atypiques ;
- S403 et S967 pour les sinistres atypiques.

Un coefficient de développement tiré aléatoirement parmi la liste construite suivant la méthode 3 – différente pour chaque sinistre – est alors appliqué à ces trois sinistres puisqu'il est considéré que leur dernière Charge FGU connue augmentera. Pour les deux autres sinistres, leur Charge FGU à leur dernière année de développement connue est égale à leur Charge FGU à l'ultime.

Deux remarques importantes sont à mentionner :

- La variable Simul Logit Evt n'indique pas si un sinistre a atteint son ultime ou pas. Elle permet uniquement de créer un classement des sinistres IBNeR selon leur susceptibilité

d'être sujets à évolution. Ainsi, des sinistres IBNeR sujets à évolution peuvent être associés à la valeur 1, tout comme il est possible que des sinistres associés à la valeur 0 ne soient pas sujets à évolution.

- L'ordre des sinistres IBNeR et le nombre d'entre eux ayant atteint leur ultime sont différents à chaque scénario, bien que les sinistres IBNeR en question soient les mêmes. Cette remarque reflète le caractère stochastique de la méthodologie pour identifier les sinistres IBNeR sujets à évolution.

Les sinistres IBNeR considérés comme ayant atteint leur ultime ne font l'objet d'aucune provision IBNeR puisque leur dernière Charge FGU équivaut à leur charge ultime. Dans cet exemple, il s'agit des sinistres S100, S213 et S003. Les sinistres IBNeR sujets à évolution peuvent quant à eux être à l'origine de provisions IBNeR, sans que ce ne soit pour autant systématique.

De manière générale, les provisions IBNeR correspondent à la différence entre les Charges FGU à l'ultime et les dernières Charges FGU connues. Dans le cadre de ce mémoire, les baisses de Charges FGU n'étant pas considérées, les provisions IBNeR ne peuvent pas être négatives. Il existe des provisions IBNeR ACC – à l'acceptation d'AXA SA – et des provisions IBNeR ASA – à la charge d'AXA SA – puisque les charges se déclinent elles-mêmes en Charges ACC et ASA. Les calculs pour obtenir ces différentes provisions IBNeR sont les suivants :

$$Provision_{IBNeR,ACC,S} = Charge_{ACC,ultime,S} - Charge_{ACC,J,S}$$

$$Provision_{IBNeR,ASA,S} = Charge_{ASA,ultime,S} - Charge_{ASA,J,S}$$

Avec :

- S : le sinistre IBNeR dont la provision IBNeR est calculée ;
- $Provision_{IBNeR,ACC,S}$ ou $Provision_{IBNeR,ASA,S}$: la provision IBNeR ACC ou la provision IBNeR ASA associée au sinistre S ;
- $Charge_{ACC,ultime,S}$ ou $Charge_{ASA,ultime,S}$: la Charge ACC ou la Charge ASA à l'ultime du sinistre S ;
- $Charge_{ACC,J,S}$ ou $Charge_{ASA,J,S}$: la Charge ACC ou la Charge ASA à la dernière année de développement connue J du sinistre S .

Les Charges ACC et ASA à l'ultime des sinistres IBNeR sont déterminées de la même façon que dans la section 3.3, c'est-à-dire en appliquant les conditions des traités de réassurance auxquels sont rattachés les sinistres.

Le tableau en figure 4.52 ci-dessous est la suite de l'exemple ci-dessus. Les deux types de provisions IBNeR y sont présentes.

Nu- méro de si- nistre	Charge FGU	Charge ACC	Charge ASA	Coef- ficient de dév.	Charge FGU ultime	Charge ACC ultime	Char- ge ASA ultime	Provi- sion IBNeR ACC	Provi- sion IB- NeR ACC
S100	8 000 000	0	0	NA	8 000 000	0	0	0	0
S213	300 000	0	0	NA	300 000	0	0	0	0
S003	5 000 000	1 000 000	700 000	NA	5 000 000	1 000 000	700 000	0	0
S043	2 000 000	200 000	0	3	6 000 000	600 000	500 000	400 000	500 000
S967	6 000 000	2 000 000	500 000	1,5	9 000 000	2 000 000	500 000	0	0
S403	3 000 000	1 000 000	1 000 000	2,5	7 000 000	2 500 000	1 000 000	1 500 000	0

FIGURE 4.52 – Détermination des provisions IBNeR

Parmi les 6 sinistres IBNeR de cet exemple, seule la moitié d’entre eux sont sujets à évolution. Parmi ces 3 sinistres, tous ne donnent pas lieu à des provisions IBNeR ACC et ASA une fois leur Charge FGU ultime passée dans les conditions des traités. En effet :

- Le sinistre S043 nécessite des provisions IBNeR ACC et ASA ;
- Le sinistre S967 ne nécessite pas de provisions IBNeR ;
- Le sinistre S403 nécessite uniquement une provision IBNeR ACC.

À travers cet exemple, il peut être constaté que, malgré l’augmentation de la Charge FGU de certains sinistres, leurs Charges ACC et ASA n’augmentent pas systématiquement entre leur dernière année de développement connue et leur ultime. Cela dépend des conditions des traités auxquels sont rattachés les sinistres.

Conclusion du chapitre

Toute cette modélisation est effectuée pour chaque sinistre de chaque segment de provisionnement et pour chacune des 1 000 simulations. En additionnant les provisions IBNeR de tous les sinistres ayant la même année de survenance, il est possible d’obtenir la distribution de ces provisions ainsi que certaines de leurs statistiques descriptives. Ces différents outils statistiques sont analysés et discutés dans le chapitre 6.

Chapitre 5

Génération de scénarios de développement de sinistres IBNyR

Ce cinquième chapitre se concentre sur la modélisation des sinistres IBNyR – *Incurring But Not yet Reported* – ainsi que sur leurs provisions associées. Il s’agit de sinistres déjà survenus mais qui n’ont pas encore été déclarés. Ces sinistres sont donc inconnus mais leur montant se révèle souvent être important.

Dans le cadre de la réassurance interne d’AXA, deux facteurs peuvent être à l’origine de sinistres non déclarés. D’une part, les assurés peuvent ne pas encore avoir notifié leur sinistre à leur assureur, n’étant eux-mêmes éventuellement pas au courant de leur survenance. D’autre part, les cédantes peuvent ne pas avoir informé Ceded Re de sinistres dont elles ont connaissance mais ne touchant pas encore la réassurance. Quelle que soit la situation, les sinistres ne sont pas présents dans la base de données *Sinistres* et aucune information ne leur est donc associée. Ces sinistres encore inconnus font partie de la catégorie des sinistres IBNyR.

La modélisation des provisions pour les sinistres IBNyR consiste en deux modèles distincts. D’une part, un premier modèle permet d’estimer, pour chaque année de survenance, le nombre de sinistres IBNyR qui seront notifiés dans les futures années de développement. D’autre part, le deuxième modèle permet d’associer une Charge FGU ultime à chacun de ces sinistres simulés. Afin d’obtenir les provisions IBNyR ACC et ASA, les conditions des traités de réassurance sont appliquées à ces Charges FGU estimées. Tout comme pour les sinistres IBNeR, la modélisation est souhaitée stochastique afin d’obtenir une distribution des provisions IBNyR estimées selon 1 000 scénarios. Les différents segments de provisionnement sont également traités indépendamment les uns des autres.

Les deux sections de ce chapitre portent respectivement sur le premier et le deuxième modèle évoqués ci-dessus.

5.1 Estimation du nombre de sinistres IBNyR

Dans cette section, la méthodologie pour estimer le nombre de sinistres IBNyR de chaque année de survenance est expliquée. Elle repose principalement sur la méthode de Schnieper (voir l'annexe C) dont le principal avantage par rapport à la méthode de la Chain Ladder est la prise en compte de l'exposition. L'exposition est une mesure de la quantité de risque. Plus elle est élevée en valeur relative, plus la probabilité qu'un sinistre survienne est élevée. Le niveau d'exposition fluctuant d'une année sur l'autre, pondérer les risques en fonction du niveau d'exposition permet alors d'éviter des biais.

De manière générale, la méthode de Schnieper – également déterministe par ailleurs – est semblable à celle de la *Chain Ladder* en termes de concepts. La différence réside principalement dans le fait que les sinistres IBNyR sont supposés arriver en proportion de l'exposition de l'année de survenance dans la méthode de Schnieper tandis que, pour celle de la *Chain Ladder*, il est supposé que la charge cumulée augmente de manière multiplicative au fur et à mesure des années de développement.

Dans le cadre de ce mémoire, l'exposition exploitée dans la méthode de Schnieper est définie comme suit :

$$Exposition = \frac{Prime\ acquise}{Part\ signée}$$

Le fait de rapporter la prime acquise à la part signée permet de refléter le risque réel total, indépendamment de la part signée par AXA SA. En effet, AXA SA ne conserve pas la totalité des primes qu'elle reçoit de la part des entités, puisqu'elle cède certains des risques qui lui sont transférés par ces mêmes entités.

La méthode Schnieper s'appuie, tout comme la méthode de la *Chain Ladder*, sur des triangles de développement qui peuvent être soit de montants, soit de nombres. Pour modéliser le nombre de sinistres IBNyR pour chaque année de survenance dans le cadre de ce mémoire, cette méthode est appliquée sur des triangles de développement incrémentaux de nombres⁸ (voir la sous-section 3.3.1). Les nombres correspondent aux sinistres dont la Charge FGU est positive pour la première fois à l'année de développement associée. Cependant, tous les sinistres ne sont pas considérés. Ne sont comptabilisés que ceux dont la dernière Charge FGU connue est supérieure ou égale à 75% – par avis d'experts – de la priorité du traité auquel ces sinistres sont rattachés. Lorsque cette priorité est inférieure à 2M€, une règle différente s'applique : le sinistre est pris en compte uniquement si sa Charge FGU dépasse ou égale 75% de 2M€. Plus formellement, les sinistres pris en compte sont les suivants :

8. Dans la suite de ce chapitre, les triangles de développement incrémentaux de nombres sont désignés simplement par « triangles de nombres ».

$$\{S \mid Charge_{FGU,S,J} \geq 75\% \times Seuil_T\}$$

Avec :

- $Charge_{FGU,S,J}$: le montant de la charge du sinistre S à sa dernière année de développement connue J en vision FGU ;

–

$$Seuil_T = \begin{cases} Priorité_T \text{ si } Priorité_T \geq 2M\text{€} \\ 2M\text{€} \text{ si } Priorité_T < 2M\text{€}. \end{cases}$$

Le tableau en figure 5.1 est un exemple dans lequel il est déterminé si les sinistres historiques sont pris en compte ou non dans le triangle de nombres.

Numéro du sinistre	Année de surveillance	Première année de développement avec Charge FGU > 0	Dernière Charge FGU connue	Priorité du traité	$Seuil_T$	Sinistre pris en compte ?
S452	2021	3	1 000 000	500 000	2 000 000	Non car $0,75 \times 2\,000\,000 = 1\,500\,000 > 1\,000\,000$
S420	2021	1	2 500 000	3 000 000	3 000 000	Oui car $0,75 \times 3\,000\,000 = 2\,250\,000 < 2\,500\,000$
S501	2022	3	500 000	5 000 000	5 000 000	Non
S740	2023	2	10 000 000	7 000 000	7 000 000	Oui
S583	2023	1	100 000	400 000	2 000 000	Non
S942	2024	1	2 000 000	4 000 000	4 000 000	Non

FIGURE 5.1 – Détermination des sinistres historiques pour le triangle de nombres

Cet ensemble de sinistres utilisés pour construire le triangle de nombres sert également de base globale à partir de laquelle sont tirés aléatoirement les sinistres historiques pour être associés aux sinistres IBNyR estimés. Prendre en compte uniquement les sinistres dont les Charges FGU atteignent au moins 75% de la priorité, permet que certains d'entre eux puissent toucher le traité lors de l'application des conditions des traités, engendrant ainsi des provisions

IBNyR. La méthodologie pour estimer les Charges FGU des sinistres IBNyR modélisés dans cette section est présentée dans la section suivante.

Dans la suite de cette section, l'explication de la méthodologie d'estimation des nombres de sinistres IBNyR s'appuie sur un exemple de segment de provisionnement dont les plus anciens sinistres sont survenus en 2021.

Tout d'abord, les expositions des 4 années de survénance pour cet exemple sont représentées dans le tableau en figure 5.2 ci-dessous. Les années de survénance 2022 et 2023 sont les plus exposées aux risques puisqu'elles sont associées aux primes les plus élevées.

Année de survénance	Exposition (en M€)
2021	6
2022	10,1
2023	13,8
2024	8,5

FIGURE 5.2 – Expositions par année de survénance

De plus, le triangle incrémental de nombres associé à cet exemple est considéré comme étant celui en figure 5.3.

		Année de développement			
		1	2	3	4
Année de survénance	2021	17	4	2	0
	2022	8	1	0	?
	2023	6	2	?	?
	2024	0	?	?	?

FIGURE 5.3 – Triangle de nombres

Le nombre total de sinistres historiques comptabilisés dans ce triangle est de 40 ($17 + 4 + 2 + 8 + 1 + 6 + 2$). Le tableau présenté précédemment en figure 5.1 est considéré comme étant un extrait de la table à partir de laquelle proviennent les sinistres historiques comptabilisés dans le triangle de nombres en figure 5.3 ci-dessus. Les sinistres S420 et S740 comptent alors chacun pour un sinistre, respectivement dans les cases (2021 ; 1) et (2023 ; 2) de ce triangle. Les 4 autres sinistres historiques présentés (les S452, S501, S583 et S942) n'y sont pas comptabilisés puisqu'ils ne répondent pas à la règle pour faire partie de sinistres pris en compte.

Il est intéressant de noter que les sinistres comptabilisés dans les années de développement strictement supérieures à 1 sont des anciens sinistres IBNyR.

Afin de projeter la partie inférieure du triangle – celle avec les $?$ – pour obtenir le nombre de sinistres IBNyR auquel s'attendre pour chaque année de survenance, la méthode de Schnieper est utilisée en premier lieu. Elle permet d'obtenir l'espérance et la variance des nombres manquants – respectivement $\mathbb{E}(N_{i,j})$ et $\mathbb{V}(N_{i,j})$ – grâce aux deux formules suivantes :

$$\mathbb{E}(N_{i,j}) = \lambda_j E_i \quad \text{et} \quad \mathbb{V}(N_{i,j}) = \sigma_j^2 E_i.$$

Avec :

- $N_{i,j}$: le nombre de sinistres à l'année de développement j de l'année de survenance i ;
- E_i : la $i^{\text{ème}}$ valeur du vecteur d'exposition correspondant à l'année de survenance i ;
- λ_j et σ_j^2 : la $j^{\text{ème}}$ valeur des vecteurs λ et σ^2 .

Les vecteurs λ et σ^2 calculés d'après les formules de la méthode de Schnieper en annexe C sont les suivants dans l'exemple :

$$\lambda = (0,807 ; 0,234 ; 0,124 ; 0,000) \quad \text{et} \quad \sigma^2 = (10,695 ; 0,708 ; 0,418 ; 0,000)$$

Les triangles en figures 5.4 et 5.5 correspondent aux triangles de nombres de sinistres complétés respectivement par l'espérance et la variance du nombre de sinistres IBNyR à prévoir, estimées par la méthode de Schnieper.

		Année de développement			
		1	2	3	4
Année de survenance	2021	17	4	2	0
	2022	8	1	0	0
	2023	6	2	1,7	0
	2024	0	2,0	1,1	0

		Année de développement			
		1	2	3	4
Année de survenance	2021	17	4	2	0
	2022	8	1	0	0
	2023	6	2	5,8	0
	2024	0	6,0	3,5	0

FIGURE 5.4 – Triangle de nombres complété par l'espérance du nombre de sinistres IBNyR

FIGURE 5.5 – Triangle de nombres complété par la variance du nombre de sinistres IBNyR

Les espérances et les variances sont calculées de manière déterministe. Comme depuis le début de la modélisation présentée dans ce mémoire, l'objectif est de développer une méthode intégrant une composante stochastique afin de générer différents scénarios de développement. Afin d'y parvenir, ces espérances et ces variances sont considérées comme des paramètres de lois. En fonction de leur valeur, elles sont associées soit au paramètre d'une loi de Poisson, soit aux paramètres d'une loi binomiale négative (voir l'annexe A). Ces deux types de lois sont couramment employés pour simuler la fréquence de sinistres. Il semble donc raisonnable d'y recourir pour les simulations des nombres de sinistres IBNyR.

Les règles établies dans le cadre de ce mémoire pour déterminer les lois utilisées afin de simuler le nombre de sinistres IBNyR à l'année de développement j de l'année de survenance i sont les suivantes :

- Si $\mathbb{E}(N_{i,j}) \geq \mathbb{V}(N_{i,j})$ et $\mathbb{E}(N_{i,j}) \neq 0$, la simulation du nombre est réalisée selon une loi de Poisson $\mathcal{P}(\ell)$ de paramètre $\ell = \mathbb{E}(N_{i,j})$;
- Si $\mathbb{V}(N_{i,j}) > \mathbb{E}(N_{i,j})$ et $\mathbb{V}(N_{i,j}) \neq 0$ et $\mathbb{E}(N_{i,j}) \neq 0$, la simulation du nombre est réalisée selon une loi binomiale négative $\mathcal{BN}(r; q)$ de paramètres :

$$r = \frac{\mathbb{E}(N_{i,j})}{\mathbb{V}(N_{i,j})} \quad \text{et} \quad q = \frac{\mathbb{E}^2(N_{i,j})}{\mathbb{V}(N_{i,j}) - \mathbb{E}(N_{i,j})}.$$

Lorsque les valeurs de l'espérance et de la variance ne répondent à aucune de ces conditions pour une année de développement j d'une année de survenance i , le nombre de sinistres IBNyR est fixé arbitrairement à 0.

Dans l'exemple qui constitue le fil rouge de cette section, les lois et leurs paramètres d'après lesquelles sont simulés les nombres de sinistres IBNyR dans les cases correspondantes du triangle sont en figure 5.6.

		Année de développement			
		1	2	3	4
Année de survenance	2021	17	4	2	0
	2022	8	1	0	0
	2023	6	2	$\mathcal{BN}(0, 3; 0, 7)$	0
	2024	0	$\mathcal{BN}(0, 3; 1, 0)$	$\mathcal{BN}(0, 3; 0, 4)$	0

FIGURE 5.6 – Triangle de nombres complété par les lois selon lesquelles sont simulés les nombres de sinistres IBNyR

Dans cet exemple, les nombres simulés seront donc soit des réalisations d'une loi binomiale négative, soit fixés de manière arbitraire à 0 pour la colonne de la 4^{ème} année de développement. À chaque année de développement de chaque année de survenance, les lois et leurs paramètres restent identiques quel que soit le numéro du scénario puisque les espérances et variances sont déterministes. En revanche, la réalisation de ces lois est différente d'un scénario à l'autre car une simulation est effectuée pour chacun des 1 000 scénarios.

Un exemple de simulation de nombres de sinistres IBNyR pour chacune des trois différentes lois binomiales négatives du tableau en figure 5.6 ci-dessus, est en figure 5.7 ci-dessous.

		Année de développement			
		1	2	3	4
Année de survenance	2021	17	4	2	0
	2022	8	1	0	0
	2023	6	2	3	0
	2024	0	1	0	0

FIGURE 5.7 – Triangle de nombres complété par les nombres simulés de sinistres IBNyR

Il est donc désormais possible de connaître le nombre de sinistres IBNyR à prévoir pour chaque année de survenance, en additionnant les chiffres en rouge de chaque ligne. Ainsi, les nombres de sinistres IBNyR simulés pour un tel scénario sont les suivants :

- Pour l’année de développement 2022 : 0 ;
- Pour l’année de développement 2023 : $3 + 0 = 3$;
- Pour l’année de développement 2024 : $1 + 0 + 0 = 1$.

La prochaine section explique la façon dont sont attribuées les Charges FGU à chacun de ces sinistres IBNyR simulés.

5.2 Estimation des Charges FGU à l’ultime des sinistres IBNyR simulés

Le nombre de sinistres IBNyR à prévoir est désormais simulé pour chaque année de survenance. Il reste à estimer la Charge FGU de chacun de ces sinistres IBNyR simulés avant de pouvoir calculer les provisions IBNyR engendrées par ces sinistres.

Les Charges FGU associées à ces sinistres IBNyR simulés sont issues des dernières Charges FGU connues des sinistres historiques utilisés pour construire le triangle de nombres de sinistres IBNyR dans la section précédente. Ces sinistres historiques sont, comme dans le chapitre 4, désignés par « sinistres d’origine ».

Pour chaque sinistre IBNyR simulé, un tirage aléatoire selon une loi uniforme est réalisé parmi une liste de sinistres d’origine pour qu’une Charge FGU à l’ultime soit attribuée à ce sinistre IBNyR. Or, les sinistres d’origine peuvent ne pas avoir atteint leur ultime. C’est le cas lorsque la variable Ultime Atteint vaut 0 à la dernière année de développement connue de ces sinistres (voir la section 4.2). Pour obtenir la Charge FGU ultime des sinistres d’origine, il est nécessaire de multiplier les dernières Charges FGU connues des sinistres n’ayant pas atteint leur ultime par le facteur de développement lissé associé à la dernière année de développement connue de ces sinistres.

Plus formellement :

$$Charge_{FGU,S_O,ultime} = \begin{cases} Charge_{FGU,S_O,J_O} & \text{si } Ultime\ Atteint_{S_O,J_O} = 1 \\ Charge_{FGU,S_O,J_O} \times Facteur\ liss\acute{e}_{J_O} & \text{si } Ultime\ Atteint_{S_O,J_O} = 0 \end{cases}$$

Avec :

- $Charge_{FGU,S_O,ultime}$: la Charge FGU à l'ultime estimée pour le sinistre d'origine S_O ;
- $Charge_{FGU,S_O,J_O}$: la Charge FGU du sinistre d'origine S_O à sa dernière année de développement connue J_O ;
- $Ultime\ Atteint_{S_O,J_O}$: la valeur de la variable Ultime Atteint pour la dernière année de développement connue J_O du sinistre d'origine S_O ;
- $Facteur\ liss\acute{e}_{J_O}$: le facteur de développement lissé associé à l'année de développement J_O .

Comme mentionné ci-dessus, les sinistres d'origine sont les sinistres historiques ayant été comptabilisés dans les tableaux de nombres de sinistres IBNyR. Dans le cas des sinistres historiques du tableau en figure 5.1 de la section précédente, les sinistres S420 et S740 font donc partie des sinistres d'origine. Pour des questions de simplicité, au sein de cette section et à des fins d'illustration, seuls ces deux sinistres d'origine sont considérés, et non les 40 sinistres historiques comptabilisés au total dans le triangle de nombres. Les variables conduisant aux Charges FGU ultimes des sinistres d'origine considérés sont présentes dans le tableau en figure 5.8 ci-dessous qui contient également les Charges FGU ultimes estimées. Il est supposé que ces sinistres appartiennent au segment de provisionnement $\mathcal{MT}$.

Numéro du sinistre d'origine	Dernière année de développement connue J_O	Dernière Charge FGU connue	Ultime Atteint	Facteur lissé pour l'année de développement J_O	Charge FGU à l'ultime
S420	4	2 500 000	0	1,117	2 792 000
S740	2	10 000 000	1	NA	10 000 000

FIGURE 5.8 – Extrait de données des sinistres d'origine

Cette multiplication par le facteur de développement lissé permet de ne pas sous-estimer les Charges FGU à l'ultime des sinistres IBNyR simulés. Lorsqu'un sinistre d'origine est tiré aléatoirement pour être associé à un sinistre IBNyR simulé, c'est la Charge FGU à l'ultime estimée du sinistre d'origine qui est attribuée au sinistre IBNyR.

Concernant l'élaboration de la liste des sinistres d'origine mentionnée au début de cette

section, cette liste dépend de l'année de survenance pour laquelle les sinistres IBNyR sont simulés. Elle est notée $\mathcal{L}_i$, i correspondant à l'année de survenance. Pour chaque sinistre IBNyR k simulé pour l'année de survenance i , un sinistre d'origine est tiré selon une loi uniforme discrète $\mathcal{U}$ définie sur $\mathcal{L}_i$. Ce tirage uniforme s'énonce mathématiquement de la façon suivante :

$$S_{i,k} \sim \mathcal{U}(\mathcal{L}_i) \quad \forall k \in \{0 ; \text{Nombre de sinistres IBNyR simulés pour l'année de survenance } i\}$$

Par conséquent, $S_{i,k}$ représente une réalisation de la variable uniforme définie sur la liste $\mathcal{L}_i$ associée au sinistre k de l'année de survenance i . Cette réalisation comprend l'ensemble des données associées au sinistre d'origine. Les plus importantes dans la suite de la modélisation sont la portée, la priorité, le type de traité, la part signée, la part nette et les taux de cession QP et EP associés au traité auquel le sinistre d'origine est rattaché car ces données permettent le calcul des provisions IBNyR ACC et ASA.

Lorsque le nombre de sinistres d'origine constituant la liste $\mathcal{L}_i$ est inférieur au nombre de sinistres IBNyR simulés dont la Charge FGU doit être modélisée pour l'année de survenance i , le tirage aléatoire s'effectue avec remise. Dans le cas contraire, le tirage aléatoire s'effectue sans remise.

En outre, la liste de sinistres d'origine pour une année de survenance peut être vide. Cette situation se présente lorsqu'aucun sinistre d'origine ne correspond aux conditions pour faire partie de cette liste. La Charge FGU des sinistres IBNyR associés à cette année de survenance est alors arbitrairement fixée à 0, entraînant systématiquement des provisions IBNyR ACC et ASA nulles pour cette année de survenance. Le sinistre d'origine est quant à lui considéré comme « Absent ».

Deux points qu'il convient de considérer sont à noter et sont les suivants :

- Les conditions des traités de réassurance peuvent changer d'une année à l'autre.
- Les programmes de réassurance peuvent disparaître d'une année à l'autre.

Ainsi, il convient de prendre en compte ces changements potentiels de la structure des traités de réassurance dans la liste de sinistres d'origine. Cela se concrétise par le fait de conserver uniquement les sinistres pour lesquels le programme de réassurance est toujours actif sur l'année de survenance des sinistres IBNyR projetés et de mettre à jour les conditions des traités.

En effet, contrairement aux sinistres IBNeR pour lesquels les traités auxquels ils sont rattachés sont connus, aucun traité n'est rattaché aux sinistres IBNyR puisque ces sinistres n'existent pas dans la base *Sinistres*. Néanmoins, il est nécessaire que ces sinistres IBNyR simulés soient rattachés à un traité de réassurance afin de calculer leurs provisions IBNyR. Les sinistres d'origine étant rattachés à des traités, la première idée serait de considérer les traités des sinistres d'origine comme les traités auxquels sont également rattachés les sinistres IBNyR.

Cependant, les sinistres d'origine ne sont pas forcément survenus la même année que les sinistres IBNyR pour lesquels les sinistres d'origine sont tirés. Les années d'exercice des traités auxquels sont rattachés les sinistres d'origine sont donc potentiellement différentes des années d'exercice à considérer pour les traités auxquels sont rattachés les sinistres IBNyR. Les traités des sinistres d'origine ne peuvent donc pas être considérés comme les traités auxquels sont également rattachés les sinistres IBNyR. En revanche, les traités de ces sinistres d'origine peuvent avoir des équivalents pour les années de survenance des sinistres IBNyR. Des traités sont considérés comme « équivalents » lorsqu'ils font partie d'un même programme de réassurance et qu'ils sont associés à des années d'exercice différentes. Il est important de noter que les conditions peuvent être différentes d'une année d'exercice à l'autre pour ces traités équivalents. Ainsi, afin d'avoir des conditions de traités cohérentes avec les années de survenance des sinistres IBNyR, les listes contiennent uniquement les sinistres d'origine dont les traités auxquels ils sont rattachés ont un équivalent aux années de survenance des sinistres IBNyR simulés.

Revenons sur le tableau en figure 5.8 ci-dessus afin de donner un exemple de la construction des listes de sinistres d'origine. Considérons que le sinistre S420 est rattaché au traité T101 et que le sinistre S740 est rattaché au traité T691. Le tableau en figure 5.9 ci-dessous illustre les conditions de ces traités et de leurs équivalents pour l'année d'exercice 2021 – lorsqu'ils existent.

Programme de réassurance	Identifiant du traité	Année d'exercice	Type de traité	Priorité	Portée	Part signée	Part nette
P1	T101	2021	XS	1 000 000	3 000 000	70%	60%
P1	T102	2022	XS	2 000 000	4 000 000	70%	60%
P1	T103	2023	XS	2 000 000	4 000 000	70%	60%
P2	T691	2023	XS	5 000 000	5 000 000	50%	30%

FIGURE 5.9 – Traités de réassurance et leurs équivalents pour l'année d'exercice 2021

Le programme P1 – auquel appartient le traité T101 – est composé de traités pour différentes années d'exercice dont les conditions sont différentes. Le programme P2 contient quant à lui uniquement le traité de réassurance T691.

La profondeur historique du segment de l'exemple étant de 4 années, les listes de sinistres d'origine construites d'après cet exemple sont au nombre de 4. Elles se présentent comme suit :

- Pour les années de survenance 2021 et 2022, la liste $\mathcal{L}_i$ est composée uniquement du sinistre rattaché au traité qui fait partie du programme 1 puisque ce traité possède un équivalent pour ces années de survenance. Ainsi, $\mathcal{L}_{2021} = \mathcal{L}_{2022} = \{S420\}$.

- Pour l'année de survenance 2023, la liste $\mathcal{L}_{2023}$ est composée des 2 sinistres rattachés aux programmes 1 et 2, c'est-à-dire les sinistres S420 et S740. Ainsi, $L_{2023} = \{S420 ; S740\}$.
- Pour l'année de survenance 2024, la liste $\mathcal{L}_i$ est vide. Ainsi, $\mathcal{L}_{2024} = \{\}$.

D'après le triangle en figure 5.7 dans la section précédente, 3 sinistres IBNyR sont simulés pour l'année de survenance 2023. Ces 3 sinistres IBNyR sont tirés aléatoirement parmi les sinistres présents dans la liste de sinistres d'origine $\mathcal{L}_{2023}$. Il est considéré que le résultat des 3 simulations est le suivant :

- Le sinistre S420 est tiré 2 fois, une fois pour le sinistre $S_{2023,1}$ et une fois pour le sinistre $S_{2023,3}$.
- Le sinistre S740 est tiré 1 fois, pour le sinistre $S_{2023,2}$.

Le tableau en figure 5.10 récapitule les données des traités attribuées aux sinistres IBNyR simulés pour l'année de survenance 2023, selon le sinistre d'origine qui leur est affecté.

Sinistre IBNyR	Numéro du sinistre d'origine	Identifiant du traité	Type de traité	Priorité	Portée	Part signée	Part nette
$S_{2023,1}$	S420	T103	XS	2 000 000	4 000 000	70%	60%
$S_{2023,2}$	S740	T691	XS	5 000 000	5 000 000	50%	30%
$S_{2023,3}$	S420	T103	XS	2 000 000	4 000 000	70%	60%

FIGURE 5.10 – Traités de réassurance pour les sinistres IBNyR simulés à l'année de survenance 2023

Il est intéressant de mentionner le fait que les conditions associées aux sinistres IBNyR $S_{2023,1}$ et $S_{2023,3}$ auraient été différentes si le traité auquel ils avaient été rattachés avait été le T101.

Dans le tableau en figure 5.11 ci-dessous, les différentes charges associées à ces 3 sinistres sont présentées. Les taux de cession EP et QP – α_{EP} et α_{QP} – sont de 0% pour tous les traités.

Sinistre IBNyR	Charge FGU ultime	Charge 100%	Charge ACC = Provision IBNyR ACC	Charge ASA = Provision IBNyR ASA
$S_{2023,1}$	2 792 500	792 500	554 750	475 500
$S_{2023,2}$	10 000 000	5 000 000	2 500 000	1 500 000
$S_{2023,3}$	2 792 500	792 500	554 750	475 500

FIGURE 5.11 – Tableau des différentes charges associées aux trois sinistres

Dans le cadre de la modélisation des sinistres IBNeR, une soustraction doit être effectuée entre les charges à l'ultime et les dernières charges connues de chaque sinistre IBNeR afin d'obtenir les provisions IBNeR. Dans le cadre de la modélisation des sinistres IBNyR, les charges des sinistres IBNyR sont quant à elles équivalentes aux provisions IBNyR puisqu'aucune Charge ACC ni ASA n'est associée à ces sinistres.

Les provisions IBNyR ACC et ASA estimées pour l'année de survenance 2023 dans le cadre de l'exemple ci-dessus sont donc les suivantes :

- Provisions IBNyR ACC : $554\,750 + 2\,500\,000 + 554\,750 = 3\,609\,500$;
- Provisions IBNyR ASA : $475\,500 + 1\,500\,000 + 475\,500 = 2\,451\,000$.

Il est important de mentionner que les sinistres IBNyR peuvent ne pas engendrer de provisions IBNyR, que ce soient les provisions ACC ou ASA. Cela peut être dû aux parts signées et nettes, mais aussi au fait que certains sinistres IBNyR à l'ultime ne touchent pas le traité. Cette potentielle situation vient du fait que tous les sinistres d'origine ne touchent pas les traités auxquels ils sont rattachés, même après que leur Charge FGU soit développée à l'ultime.

Ces différentes étapes sont effectuées pour chacun des 1 000 scénarios de développement et pour chaque segment de provisionnement indépendamment des autres.

Conclusion du chapitre

Grâce à la méthodologie présentée dans ce chapitre, il est possible de retracer le chemin menant aux provisions IBNyR estimées, ce qui est l'objectif recherché dans le cadre de ce mémoire. Par ailleurs, les différents choix de modélisation de la Charge FGU des sinistres IBNyR simulés entraînent des biais. D'une part, s'appuyer uniquement sur des sinistres historiques empêche la possibilité de modéliser des Charges FGU très élevées qui n'ont jamais été observées mais qui pourraient exister dans le futur. Pour pallier ce problème, il faudrait utiliser des lois de sévérité afin de simuler les Charges FGU à l'ultime des sinistres IBNyR. D'autre part, les listes des sinistres d'origine sont de petite taille, impliquant la possibilité que des sinistres extrêmes très peu fréquents en termes de sévérité soient tirés plusieurs fois, entraînant des provisions IBNyR anormalement élevées.

Chapitre 6

Exploitation des résultats des modélisations des sinistres IBNeR et IBNyR

Ce dernier chapitre rassemble les résultats obtenus suite à la génération des scénarios de développement des sinistres IBNeR et des scénarios de développement des sinistres IBNyR. Dans le cadre de ce mémoire, les provisions IBNeR sont différenciées des provisions IBNyR puisqu'une modélisation est effectuée pour les sinistres IBNeR et une autre est réalisée pour les sinistres IBNyR. L'équipe Reserving calcule quant à elle les provisions IBNR dans leur globalité, sans distinction entre les provisions IBNeR et IBNyR.

Pour comparer les résultats obtenus dans le cadre de ce mémoire et ceux obtenus par l'équipe, il est donc indispensable d'additionner les provisions IBNeR et les provisions IBNyR générées à chaque scénario. Un scénario correspond à une estimation des provisions IBNeR et IBNyR pour chaque année de survenance d'un segment de provisionnement.

Les provisions IBNeR d'une année de survenance associées à un scénario correspondent à la somme des provisions IBNeR simulées de chaque sinistre survenu cette année-là pour le même scénario. Les provisions IBNyR d'une année de survenance associées à un scénario s'obtiennent quant à elles, en sommant les provisions IBNyR estimées de chaque sinistre IBNyR simulé comme étant survenu cette année-là dans le cadre de ce même scénario. Les provisions IBNR d'une année de survenance associées à un scénario représentent alors l'addition des provisions IBNeR et IBNyR de la même année de survenance qui correspondent au même numéro de scénario et à la même méthode de modélisation des sinistres IBNR. Pour donner un exemple, les provisions IBNeR pour l'année de survenance 2020 du 500^{ème} scénario dans le cadre de la méthode 3 sont additionnées aux provisions IBNyR pour la même année de survenance du même scénario. Cette addition correspond aux provisions IBNR associées à l'année de survenance 2020

du 500^{ème} scénario dans le cadre de la méthode 3 de la modélisation des sinistres IBNeR.

Ce chapitre s'appuie entièrement sur les tableaux et graphiques affichés dans l'application RShiny élaborée au cours du stage ayant conduit à ce mémoire.

Dans ce chapitre, la première section compare les provisions IBNR obtenues dans le cadre de ce mémoire et celles calculées par l'équipe Reserving. La seconde section s'intéresse à la consommation des provisions IBNR, d'un côté par les provisions IBNeR et de l'autre par les provisions IBNyR, pour certains scénarios de l'un des trois segments de provisionnement.

Pour information, seules les provisions IBNR ACC, IBNeR ACC et IBNyR ACC sont analysées. Pour rappel, les Charges ACC correspondent aux Charges IBNR à l'acceptation d'AXA SA, c'est-à-dire avant la prise en compte d'une potentielle rétrocession ou de Couvertures Groupe, ces dernières protégeant les risques retenus par AXA SA. Lorsque la rétrocession et les Couvertures Groupe sont prises en compte, on parle de Charges ASA. La part nette des segments de provisionnement $\mathcal{CT}$, $\mathcal{MT}$ et $\mathcal{LT}$ étant de 0%, les provisions IBNeR et IBNyR ASA sont donc nulles puisque les Charges ASA le sont. L'équipe Reserving considère également des provisions IBNR ASA égales à 0. Les provisions IBNR, IBNeR et IBNyR ACC sont simplement dénommées provisions IBNR, IBNeR et IBNyR par la suite.

6.1 L'analyse et la comparaison des provisions IBNR

Dans cette section, les différents scénarios de provisions IBNR sont comparés avec les provisions IBNR obtenues par l'équipe Reserving à l'aide de méthodes agrégées.

Grâce aux différents scénarios modélisés, il est possible de connaître le centile associé aux provisions IBNR calculées par l'équipe par rapport à la distribution des provisions IBNR obtenue à partir de ces différents scénarios. Cette information est très importante dans le contexte de ce mémoire car elle permet de déterminer si les provisions calculées par l'équipe correspondent environ à des montants de provisions IBNR fréquemment ou peu observés dans les scénarios. Si les provisions IBNR calculées par l'équipe correspondent à un centile trop faible, ces dernières sont considérées comme probables, voire très probables. En effet, il est alors possible que les provisions IBNR calculées par l'équipe ne suffisent pas pour tenir les engagements auprès des assurés sinistrés puisque les provisions IBNR d'un nombre non négligeable de scénarios sont plus élevées que celles calculées par l'équipe. En revanche, si les provisions IBNR calculées par l'équipe se situent dans les centiles les plus élevés de la distribution, alors elles sont jugées prudentes, voire excessivement prudentes, puisque la plupart des scénarios modélisent des provisions IBNR plus faibles. Afin d'avoir un équilibre entre ces deux situations, le centile optimal dans le cadre de ce mémoire se situe entre 70% et 80% d'après l'avis d'experts. Ce niveau peut paraître élevé mais il est important de souligner que la réassurance couvre des risques

avec des sévérités très élevées et des fréquences plutôt faibles, générant ainsi une distribution très asymétrique des provisions. Cette première analyse permet ainsi de qualifier le niveau de prudence ou de risque associé aux provisions établies par l'équipe Reserving.

De plus, il est possible d'obtenir des statistiques sur la distribution des provisions IBNR obtenue à partir des 1 000 scénarios telles que la moyenne, la médiane, le minimum, le maximum et la moyenne des provisions IBNeR et IBNyR par scénario. Ces statistiques sont calculées suite à l'addition des provisions IBNeR et IBNyR de chaque scénario et ce, pour chaque année de survenance. Les provisions IBNeR et IBNyR de chaque année de survenance s'obtiennent en sommant les provisions IBNeR et IBNyR de chaque sinistre IBNeR et IBNyR. Le centile correspondant aux provisions IBNR calculées par l'équipe Reserving pour une année de survenance est quant à lui obtenu à partir de la distribution des provisions IBNR issues de la somme des provisions IBNeR et IBNyR de cette année de survenance dans les 1 000 scénarios.

Les deux prochaines sous-sections se concentrent sur les résultats des provisions IBNR simulées avec la méthode 3 de la modélisation des sinistres IBNeR pour la première et sur les résultats des méthodes 1 et 2 pour la seconde.

6.1.1 Les résultats des provisions IBNR simulées avec la méthode 3 de la modélisation des sinistres IBNeR

Les trois tableaux ci-dessous, en figures 6.1, 6.2 et 6.3, présentent différentes statistiques pour chacun des trois segments de provisionnement. Les provisions IBNeR sont issues du recours à la méthode 3. Pour information, seules les années de survenance dont les provisions IBNR calculées par l'équipe sont strictement supérieures à 0 sont extraites du tableau présent dans l'application RShiny, par souci de simplification.

Année de survenance	Provisions IBNR calculées par l'équipe	Cen-tiles	Moyenne	Médiane	Mini-mum	Maximum	Moyenne des provisions IBNeR	Moyenne des provisions IBNyR
2019	276 105	80%	6 906 464	8 917	125	463 827 960	34 183	6 872 281
2020	2 621 688	81%	11 422 657	123 743	193	480 552 684	100 123	11 322 534
2021	6 875 337	91%	6 917 795	380 623	3 476	478 571 187	415 356	6 502 439
2022	65 997 083	97%	7 610 943	588 657	40	807 737 806	1 271 650	6 339 293
2023	121 809 590	98%	8 948 279	1 406 560	0	288 654 131	1 377 435	7 570 844

FIGURE 6.1 – Statistiques des provisions IBNR simulées pour le segment $\mathcal{CT}$

Année de sur-ve-nance	Provisions IBNR calculées par l'équipe	Cen-tiles	Moyenne	Médiane	Mini-mum	Maximum	Moyenne des pro-visions IBNeR	Moyenne des pro-visions IBNyR
2016	2 556 319	94%	919 014	32 271	0	33 541 090	33 621	885 393
2017	19 345 575	99%	1 167 642	38 819	0	23 477 405	44 917	1 122 725
2018	1 828 019	99%	417 963	221 588	0	22 615 983	306 181	111 782
2019	318 964	68%	341 343	3 260	0	5 173 585	341 343	0
2020	940 890	93%	1 303 912	3 666	0	34 448 045	5 988	1 297 924
2021	672 775	95%	794 168	39 416	0	19 508 365	62 705	731 463
2022	4 763 228	89%	1 229 013	99 257	0	27 858 945	111 517	1 117 496
2023	7 004 697	86%	2 498 705	350 371	0	27 993 994	302 797	2 195 908

FIGURE 6.2 – Statistiques des provisions IBNR simulées pour le segment $\mathcal{MT}$

Année de sur-ve-nance	Provisions IBNR calculées par l'équipe	Cen-tiles	Moyenne	Médiane	Mini-mum	Maxi-mum	Moyenne des pro-visions IBNeR	Moyenne des pro-visions IBNyR
1992	175	100%	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
1997	423 019	100%	9 918	4 377	0	49 986	9 918	0
...	...	...	...	...	...	...	...	...
2016	3 590 451	68%	1 288 259	11 955	0	4 087 109	9 213	1 279 046
2017	7 728 360	100%	1 752 184	644 187	0	4 508 083	200 473	1 551 711
2018	2 767 572	97%	466 863	2 948	0	6 487 593	8 943	457 920
2019	23 392 406	100%	8 012 349	8 297 782	2 170	13 928 990	820 628	7 191 721
2020	15 269 097	100%	6 469 834	5 003 634	569	14 448 769	195 010	6 274 824
2021	5 512 251	81%	3 158 639	1 547 711	12 943	17 657 471	1 282 194	1 876 445
2022	6 362 000	94%	1 342 199	211 122	2 563	6 919 608	159 384	1 182 815
2023	10 681 396	100%	128 526	59 953	0	897 049	128 526	0

FIGURE 6.3 – Statistiques des provisions IBNR simulées pour le segment $\mathcal{LT}$

Premièrement, il est intéressant de remarquer le fait que, plus le segment de provisionnement est long terme, plus les premières années de survenance pour lesquelles les provisions IBNR calculées par l'équipe sont strictement supérieures à 0, sont anciennes. En effet, les sinistres appartenant à des segments de provisionnement long terme nécessitent plus d'années de développement avant leur clôture et peuvent donc engendrer des provisions IBNR plus tardivement.

Deuxièmement, les centiles associés aux provisions IBNR calculées par l'équipe sont élevés dans l'ensemble. Ils se situent principalement au-dessus de 80%. Cela signifie que, d'après la modélisation effectuée dans le cadre de ce mémoire, les provisions IBNR calculées par l'équipe Reserving sont jugées très prudentes. Cependant, si les scénarios les plus extrêmes se produisent, les provisions IBNR calculées par l'équipe sont trop faibles. En effet, les provisions IBNR maximales simulées dépassent les provisions IBNR calculées par l'équipe. Cependant, au vu de la rareté que de telles provisions IBNR soient réellement nécessaires pour indemniser les sinistres IBNR, considérer un tel montant de provisions serait excessivement prudent sans réelle justification.

Pour rebondir sur cette deuxième remarque, en s'intéressant aux provisions moyennes IBNeR et IBNyR des trois segments sur l'ensemble des années de survenance, les provisions IBNR sont principalement composées de provisions IBNyR. Il existe cependant quelques exceptions. En effet, les provisions IBNR des années de survenance 2018 et 2019 du segment $\mathcal{MT}$ et de l'année de survenance 2023 du segment $\mathcal{LT}$ sont constituées en grande partie – voire uniquement – de provisions IBNeR. Pour l'année de survenance 2019, les provisions IBNyR sont nulles en raison de l'absence de traités équivalents pour les sinistres IBNyR simulés pour cette année de survenance, impliquant des Charges FGU ultimes estimées égales à 0 quel que soit le scénario. Par ailleurs, les provisions IBNR de l'année de survenance 2024 – qui est l'année de survenance courante dans le cadre de ce mémoire – ne sont pas pertinentes à analyser. En effet, le gel des bases ayant été réalisé courant 2024, tous les sinistres dont l'année de survenance est 2024 ne sont pas encore forcément recensés, impliquant des résultats de modélisation biaisés.

De surcroît, les moyennes des provisions IBNeR des scénarios par année de survenance sont beaucoup plus élevées que les médianes. Cette constatation implique une très forte asymétrie dans les distributions simulées.

En outre, le modèle d'estimation des provisions IBNR développé dans le cadre de ce mémoire associe des provisions IBNR à toutes les années de survenance pour lesquelles les provisions IBNR calculées par l'équipe sont supérieures à 0, excepté pour les années de survenance 1992 à 1996 du segment $\mathcal{LT}$. Cependant, les provisions IBNR calculées par l'équipe de ces cinq années sont très faibles comparées à celles des autres années de survenance. Avoir estimé que les provisions IBNR pour ces années de survenance sont nulles ne paraît donc pas incohérent.

6.1.2 Les résultats des provisions IBNR simulées avec les méthodes 1 et 2 de la modélisation des sinistres IBNeR

Les statistiques des provisions IBNR sont également produites lorsque les méthodes 1 et 2 sont utilisées pour la modélisation des sinistres IBNeR. Les différences dans les statistiques

ainsi que dans les centiles associés aux provisions IBNR calculées par l'équipe sont très peu importantes par rapport aux résultats obtenus avec l'utilisation de la méthode 3. Ces différences sont de quelques milliers d'euros – voire de centaines d'euros – pour les statistiques et sont inexistantes pour les centiles, excepté celui de 2019. Le tableau récapitulatif en figure 6.4 est celui associé au segment de provisionnement $\mathcal{MT}$ et dont les provisions IBNeR sont issues de la méthode 1.

Année de survenance	Provisions IBNR calculées par l'équipe	Centiles	Moyenne	Médiane	Minimum	Maximum	Moyenne des provisions IBNeR	Moyenne des provisions IBNyR
2016	2 556 319	94%	914 977	26 040	0	33 513 390	29 584	885 393
2017	19 345 575	99%	1 142 047	1 043	0	23 439 633	19 322	1 122 725
2018	1 828 019	99%	193 419	38 569	0	22 667 062	81 637	111 782
2019	318 964	82%	201 423	0	0	4 722 221	201 423	0
2020	940 890	93%	1 300 302	0	0	34 438 838	2 378	1 297 924
2021	672 775	95%	770 097	23 803	0	19 504 726	38 634	731 463
2022	4 763 228	89%	1 160 066	18 184	0	27 880 382	42 570	1 117 496
2023	7 004 697	86%	2 300 605	72 689	0	27 770 773	104 697	2 195 908

FIGURE 6.4 – Statistiques des provisions IBNR simulées pour le segment $\mathcal{MT}$ dans le cadre de la méthode 1

Une différence est cependant plus importante sur les statistiques des années de survenance 2018 et 2019, lorsque la moyenne des provisions IBNyR est très faible. En effet, les provisions IBNeR représentent une part beaucoup plus importante des provisions IBNR : presque la moitié pour l'année de survenance 2018 et la totalité pour l'année de survenance 2019. Le centile passe quant à lui de 68 à 82% pour l'année de survenance 2019 car les provisions IBNeR sont plus faibles en moyenne qu'avec la méthode 3, avec des provisions IBNyR toujours égales à 0 – car leur moyenne est nulle. Les différentes méthodes influencent donc effectivement les provisions IBNeR mais cet impact reste peu perceptible au niveau global des provisions IBNR puisque la majorité des provisions IBNR provient des provisions IBNyR.

6.2 L'analyse des résultats de différents scénarios

Dans cette section, l'intérêt est porté sur les provisions IBNR de l'année de survenance 2021 du segment de provisionnement $\mathcal{CT}$. Les scénarios associés à cette année de survenance parmi

ceux les plus « proches » des provisions IBNR calculées par l'équipe, ainsi que le « pire » scénario associé à cette même année de survenance, sont analysés. Les différentes études réalisées dans cette section permettent de rendre compte de l'explicabilité du modèle mis en place dans le cadre de ce mémoire. En effet, ces études sont réalisables sur l'ensemble des scénarios de toutes les années de survenance de chaque segment de provisionnement.

Une analyse globale des provisions IBNR est réalisée avant d'examiner plus en détail et de manière distincte les provisions IBNyR et IBNeR. Concernant la modélisation des sinistres IBNeR, l'analyse se concentre sur la méthode 3, qui représente l'approche la plus aboutie en termes de modélisation, sans que cela ne soit toujours rappelé.

6.2.1 L'analyse globale des provisions de scénarios proches des provisions calculées par l'équipe Reserving

L'objectif de cette sous-section est d'appréhender les provisions IBNR de l'année de survenance 2021 avant d'effectuer une analyse plus poussée des provisions IBNeR et IBNyR du scénario choisi ultérieurement dans cette sous-section.

Les données présentées relatives aux provisions IBNR de l'année de survenance 2021 du segment $\mathcal{CT}$ dans la section précédente sont rappelées dans le tableau en figure 6.5.

An- née de sur- ve- nance	Provi- sions IBNR calcu- lées par l'équipe	Cen- tiles	Moyenne	Mé- diane	Centile 80%	Mini- mum	Maximum	Moy- enne des provi- sions IB- NeR	Moy- enne des provi- sions IBNyR
2021	6 875 337	91%	6 917 795	380 623	1 049 744	3 476	478 571 187	415 356	6 502 439

FIGURE 6.5 – Statistiques de l'année de survenance 2021 des provisions IBNR simulées pour le segment $\mathcal{CT}$

Tout d'abord, il est intéressant de constater que la moyenne des provisions IBNR simulées est très proche des provisions IBNR calculées par l'équipe – moins de 50k€. En revanche, la médiane et le centile 80% – dont les montants associés devaient être les plus proches des montants des provisions IBNR calculées par l'équipe – sont beaucoup plus faibles. C'est le cas pour la plupart des années de survenance de l'ensemble des segments de provisionnement.

La fonction de répartition en figure 6.6 ci-dessous représente la distribution des provisions IBNR estimées dans le cadre de ce mémoire pour cette année de survenance. La barre verticale rouge correspond au montant des provisions IBNR calculées par l'équipe Reserving.

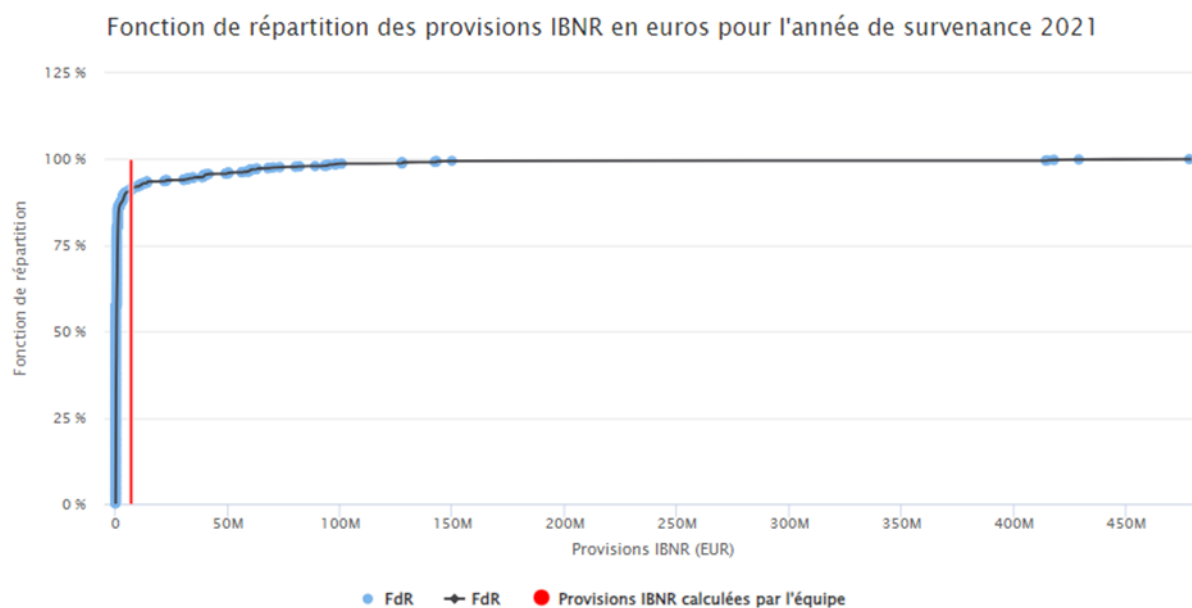


FIGURE 6.6 – Fonction de répartition des provisions IBNR de l'année de survenance 2021 pour le segment $\mathcal{CT}$

Cette distribution permet de constater que la grande majorité des provisions IBNR est estimée à moins de 6,8M€, qui est le chiffre représentant les provisions IBNR calculées par l'équipe et symbolisé par la barre verticale rouge. Le point d'intersection entre la courbe noire et bleue et la barre verticale rouge correspond au centile présent dans le tableau précédent en figure 6.5. Il est de 91% pour cette année de survenance.

En plus des statistiques mentionnées dans la section précédente, les quatre statistiques suivantes sont également calculées pour chaque année de survenance :

- Le nombre simulé moyen de sinistres IBNeR et de sinistres IBNyR ;
- Le nombre simulé moyen de sinistres IBNeR et de sinistres IBNyR qui engendrent, respectivement, des provisions IBNeR et des provisions IBNyR strictement positives.

Ces statistiques sont présentées pour l'année de survenance 2021 du segment $\mathcal{CT}$ dans le tableau en figure 6.7.

Nombre simulé moyen de sinistres IBNeR	Nombre simulé moyen de sinistres IBNeR engendrant des provisions IBNeR	Nombre simulé moyen de sinistres IBNyR	Nombre simulé moyen de sinistres IBNyR engendrant des provisions IBNyR
10,6	9,6	0,4	0,2

FIGURE 6.7 – Statistiques des nombres de sinistres de la survenance 2021 du segment $\mathcal{CT}$

D'après le tableau précédent en figure 6.5, les provisions IBNyR représentent, pour l'année de survenance 2021, la part la plus importante des provisions IBNR en termes de montants,

avec plus de 6M€ en moyenne contre moins d'1M€ pour les provisions IBNeR. En revanche, le nombre de sinistres IBNyR qui engendrent ces provisions IBNyR est bien plus faible, en moyenne, que le nombre de sinistres IBNeR engendrant les provisions IBNeR. Le tableau en figure 6.7 ci-dessus permet également de constater que la plupart des sinistres IBNeR, respectivement IBNyR, simulés engendrent des provisions IBNeR, respectivement IBNyR.

Le tableau en figure 6.8 indique les 10 scénarios pour lesquels les provisions IBNR simulées pour l'année de survenance 2021 sont les plus proches des provisions IBNR calculées par l'équipe Reserving, qui sont de 6 875 337€.

Numéro de simulation	Provisions IB-NeR	Provisions IB-NyR	Provisions IBNR	Centile
14	261 780	6 710 754	6 972 534	91%
206	498 394	6 274 791	6 773 185	91%
59	200 738	6 989 985	7 190 723	91%
425	169 451	6 274 791	6 444 242	91%
510	157 623	6 201 385	6 359 008	91%
769	1 138 186	6 274 791	7 412 977	91%
543	171 585	7 247 784	7 419 369	91%
977	741 003	5 424 356	6 165 359	91%
746	620 727	6 989 985	7 610 712	92%
816	372 867	7 247 784	7 620 651	92%

FIGURE 6.8 – Les 10 scénarios les plus proches des provisions IBNR calculées par l'équipe

Pour ces 10 scénarios, leurs provisions IBNR simulées sont comprises entre 6M€ et 8M€. Elles se situent donc parfois en-dessous et parfois au-dessus de celles calculées par l'équipe. La différence entre les provisions simulées et les provisions IBNR calculées par l'équipe est de moins de 100k€ pour le scénario le plus proche tandis qu'il est d'environ 750k€ pour le 10^{ème} scénario le plus proche. Ces chiffres correspondent à une différence entre 1 et 11% environ. De plus, ces 10 scénarios génèrent des provisions IBNyR entre 5 fois – pour le scénario 769 – et 43 fois – pour le scénario 543 – plus élevées que des provisions IBNeR, qui sont toutes situées entre 100k€ et 1,1M€.

Une autre remarque intéressante sur le tableau en figure 6.8 ci-dessus est le fait que les provisions IBNeR sont toutes différentes d'un scénario à l'autre. Les provisions IBNyR sont parfois quant à elle identiques pour plusieurs scénarios. Cela vient du fait que d'un côté, le nombre simulé de sinistres IBNyR est faible pour la plupart des scénarios et que d'un autre côté, le nombre de sinistres d'origine présents dans la liste $\mathcal{L}_i$ pouvant être associés à un sinistre IBNyR est faible. Les possibilités d'obtenir différentes provisions sont donc réduites et les mêmes

sinistres d'origine peuvent donc être tirés pour plusieurs scénarios, ce qui constitue par ailleurs une limite au modèle de simulation des sinistres IBNyR. Concernant la modélisation des sinistres IBNeR avec l'utilisation de la méthode 3, il existe un nombre de possibilités élevé de provisions IBNeR simulées. En effet, d'une part, les sinistres IBNeR ne sont pas les mêmes pour chaque scénario et d'autre part, les coefficients de développement pouvant être tirés aléatoirement afin d'être appliqués aux Charges FGU des sinistres IBNeR sont nombreux.

D'après le modèle de simulation développé, les provisions IBNR sont ainsi principalement portées par les sinistres IBNyR à hauteur de 80% - 90%.

6.2.2 L'analyse des provisions IBNR d'un scénario extrême

Grâce à la fonction de répartition des provisions IBNR en figure 6.6 précédente, il peut être remarqué que quelques provisions IBNR avoisinent les 400M€. La suite de cette sous-section s'intéresse au cheminement permettant de simuler une telle provision pour un scénario.

L'analyse globale des provisions IBNR

Il est intéressant de regarder les 10 scénarios simulant les provisions IBNR les plus élevées et qui constituent la queue de la distribution. Par ailleurs, c'est une fonctionnalité qui a été conçue dans l'application RShiny. Ces 10 scénarios sont présentés dans le tableau en figure 6.9, avec les scénarios dont les provisions IBNR les plus élevées sont en haut.

Numéro de simulation	Provisions IBNeR	Provisions IBNyR	Provisions IBNR
177	515 048	478 056 139	478 571 187
550	1 479 766	427 935 295	429 415 061
447	707 158	417 588 013	418 295 171
291	458 597	414 606 014	415 064 611
364	243 045	414 606 014	414 849 059
590	552 222	149 376 094	149 928 316
151	299 070	142 665 340	142 964 410
134	187 501	142 665 340	142 852 841
89	46 066	142 665 340	142 711 406
815	659 173	127 406 110	128 065 283

FIGURE 6.9 – Les 10 scénarios les plus extrêmes de la survenance 2021 pour le segment $\mathcal{CT}$

La tendance de la répartition des provisions IBNeR et IBNyR vue à la section précédente

indique que les provisions IBNyR sont beaucoup plus élevées que les provisions IBNeR, en moyenne, pour la plupart des scénarios des différentes années de survenance. Cela se confirme avec les scénarios affichés dans le tableau en figure 6.9 ci-dessus avec des provisions IBNyR entre 500 et 1 000 fois plus élevées que les provisions IBNeR.

Il est également intéressant de remarquer que ce sont les 5 premiers scénarios qui engendrent des provisions IBNyR de plus de 400M€ alors que les 5 suivants en engendrent seulement un peu plus de 100M€. Cette différence peut être due au fait qu'un sinistre d'origine dont la Charge FGU ultime est très élevée, ait été tiré aléatoirement pour ces 5 premiers scénarios, engendrant ainsi la majeure partie des provisions IBNR. Le nombre moyen de sinistres IBNyR élevés étant faible, il est important de comprendre que la majorité des provisions IBNyR ne provient que de quelques sinistres IBNyR. Lorsque ces provisions IBNyR sont élevées, elles sont la conséquence de quelques sinistres d'origine tirés ayant une Charge FGU ultime élevée plutôt que de nombreux sinistres d'origine tirés ayant une Charge FGU ultime faible ou moyenne.

Les provisions IBNeR et IBNyR du scénario avec les provisions IBNR les plus élevées – c'est-à-dire le scénario 177 – sont analysées dans la suite de la sous-section, après cette étude des provisions IBNeR et IBNyR globales pour l'année de survenance 2021.

L'analyse plus détaillée des provisions IBNR les plus proches de celles calculées par l'équipe suit exactement le même cheminement présenté dans cette sous-section.

L'analyse spécifique des provisions IBNyR dans le cadre du scénario 177

Dans cette sous-sous-section, l'intérêt est porté sur les provisions IBNyR de l'année de survenance 2021 du segment de provisionnement $\mathcal{CT}$.

Concernant cette année de survenance, le développement du nombre de sinistres construit à partir des règles énoncées dans la section 5.1 est en figure 6.10.

		Année de développement				
		1	2	3	4	...
Année de survenance	...	...	...	...	...	
	2021	15	8	0	0	
	...	...	...	...	...	
	...	...	...	...	...	

FIGURE 6.10 – Triangle incrémental de nombres des sinistres IBNyR du segment $\mathcal{CT}$

Pour information, plus de 300 sinistres historiques – étant également les sinistres d'origine dans le cadre de ce mémoire – sont comptabilisés dans le triangle de nombres de sinistres

IBNyR du segment $\mathcal{CT}$. Pour la suite, il est également intéressant de préciser que l'année de développement la plus élevée comptabilisant au moins 1 sinistre est la 15^{ème} année de développement de l'année de survenance 1997. La majorité des sinistres comptabilisés se trouve aux années de développement 1 et 2.

Le graphique en figure 6.11 ci-dessous représente la fonction de répartition des provisions IBNyR estimées dans le cadre de ce mémoire pour l'année de survenance 2021. La barre verticale rouge correspond au montant des provisions IBNR calculées par l'équipe – qui sont de 6 875 337€ – puisqu'il n'existe pas d'estimation séparée des provisions IBNeR et IBNyR.

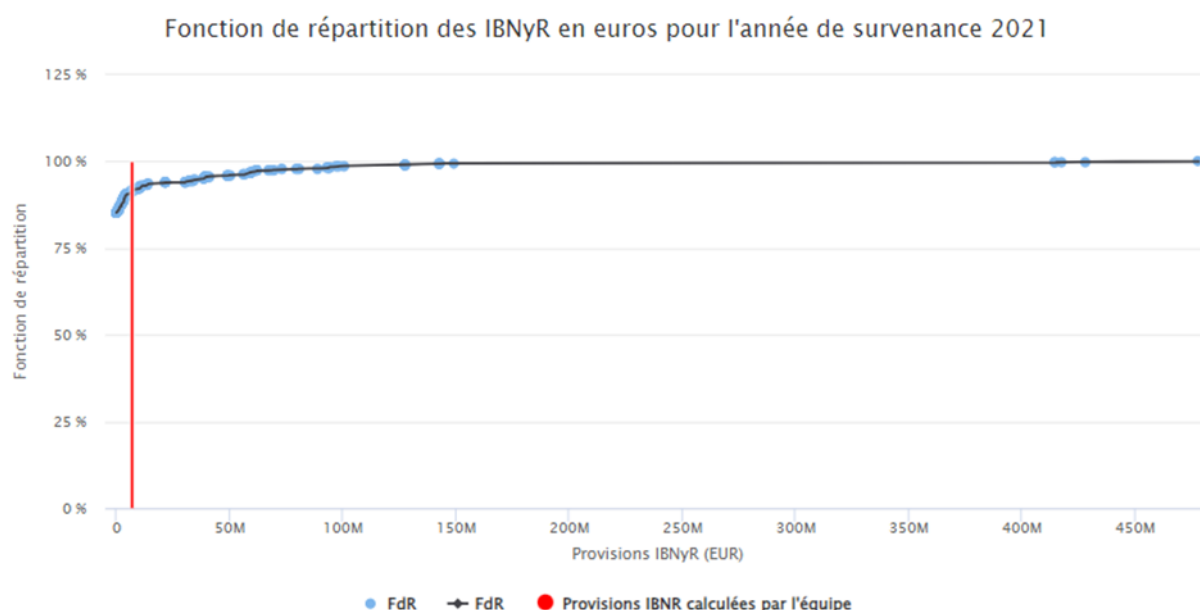


FIGURE 6.11 – Fonction de répartition des provisions IBNyR pour l'année de survenance 2021

Cette distribution permet de constater qu'une trentaine de provisions IBNyR sur les 1 000 estimées dépassent les provisions IBNR calculées par l'équipe, sans même y avoir ajouté les provisions IBNeR simulées.

Les statistiques des provisions IBNyR et du nombre de sinistres IBNyR des 1 000 scénarios sont étudiées pour l'année de survenance 2021 et sont en figures 6.12 et 6.13 ci-dessous.

Année de survenance	Provisions IBNR calculées par l'équipe	Cen-tilés	Moyenne	Mé-diane	Centile 80%	Mini-mum	Maximum
2021	6 875 337	91%	6 502 439	0	0	0	478 056 139

FIGURE 6.12 – Statistiques de l'année de survenance 2021 des provisions IBNyR simulées

Année de survenance	Moyenne	Médiane	Centile 90%	Minimum	Maximum
2021	0,4	0	0	1	6

FIGURE 6.13 – Statistiques de la survenance 2021 des nombres simulés de sinistres IBNyR

Ces statistiques confirment l'hypothèse que les provisions IBNyR du scénario 177 – étant de 478 056 139€ comme l'indique le tableau en figure 6.9 de la sous-sous-section précédente – sont les provisions IBNyR les plus élevées des différents scénarios de l'année de survenance 2021.

Il est également possible de retrouver l'information que 0,4 sinistres IBNyR sont simulés en moyenne pour cette année de survenance. Cette moyenne est d'ailleurs bien égale à l'espérance estimée par la méthode de Schnieper – plus exactement, à l'addition des espérances estimées de toutes les années de développement de l'année de survenance 2021. Le nombre maximal simulé de sinistres IBNyR est quant à lui de 6. Il se peut que le scénario 177 fasse partie des scénarios pour lesquels 6 sinistres IBNyR sont simulés, tout comme il est possible qu'il y ait moins de sinistres simulés mais que les Charges FGU qui leur soient attribuées soient très élevées.

Il est possible de connaître le nombre simulé de sinistres IBNyR grâce au tableau en figure 6.14 ci-dessous. En effet, ce tableau indique le nombre simulé de sinistres IBNyR pour chaque année de développement ainsi que la loi à partir de laquelle ce nombre est simulé en fonction des paramètres de lois renseignés. Si les paramètres r et q sont renseignés, la réalisation de nombre de sinistres IBNyR est celle d'une loi $\mathcal{BN}(r; q)$. Si le paramètre ℓ est renseigné, la réalisation du nombre de sinistres IBNyR est quant à elle celle d'une loi $\mathcal{P}(\ell)$. Si aucun de ces 3 paramètres n'est renseigné, cela signifie que les espérances et les variances de la méthode de Schnieper ne répondent à aucune des conditions pour qu'ils soient associés à une loi. Dans ce cas, le nombre de sinistres IBNyR est fixé arbitrairement à 0. Pour information, seule l'année de développement 15 est affichée car elle seule simule des sinistres IBNyR.

Année de survenance	Année de développement	Espérance	Variance	r	q	ℓ	Nombre de sinistres IBNyR simulés
2021	15	0,09	0,12	0,28	0,75	NA	4

FIGURE 6.14 – Réalisations des lois pour les nombres de sinistres IBNyR

Dans le cadre du scénario 177, 4 sinistres IBNyR sont simulés pour l'année de développement 15. Il est intéressant de noter qu'une loi est associée à cette année de développement en raison du sinistre survenu en 1997 mentionné au début de la présente sous-sous-section. Les réalisations des lois – toutes binomiales négatives par ailleurs – des autres années de développement donnent 0. L'ensemble des sinistres IBNyR estimés pour l'année de survenance 2021 provient donc d'une

simulation de la loi binomiale négative de l'année de développement 15. Les provisions IBNyR de plus de 400M€ de ce scénario sont donc issues de 4 sinistres seulement. Malgré le fait que ce nombre de sinistres IBNyR ne soit pas le maximum simulé pour cette année de survenance, il est tout de même supérieur au centile 90%, qui est de 1 sinistre simulé. Cela implique que plus de 90% des nombres simulés de sinistres IBNyR sont inférieurs à 4 pour l'année de survenance 2021. Il suffit que la Charge FGU des sinistres d'origine tirés aléatoirement soit élevée pour que les provisions IBNyR soient estimées à des centaines de millions d'euros.

Le tableau en figure 6.15 indique les sinistres d'origine attribués aux sinistres IBNyR simulés de l'année de survenance 2021 pour le scénario 177.

Numéro de sinistre d'origine	Année de survenance	Ultime Atteint	Charge FGU ultime	Numéro de traité	Année d'exercice du programme	Type de traité	Charge IBNyR ACC = Provision IBNyR ACC
S60015	2011	1	905 866 866	T100	2021	XS	414 606 014
S94617	2021	0	101 442 514	T200	2021	XS	56 834 971
S83106	2020	0	13 761 285	T300	2021	XS	4 663 502
S66616	2013	1	6 730 429	T400	2021	XS	1 951 652

FIGURE 6.15 – Simulation des provisions ACC de sinistres IBNyR

Les 400M€ du scénario 177 proviennent donc en majorité d'un seul sinistre d'origine – le sinistre S60015 – dont la Charge FGU ultime est très élevée. Ce sinistre a par ailleurs déjà atteint son ultime et n'est donc pas développé. Pour les 4 autres scénarios avec les provisions IBNR les plus élevées, c'est-à-dire les scénarios 550, 447, 291 et 364, ce sinistre d'origine est également tiré. Pour les scénarios 291 et 364, c'est même le seul sinistre d'origine qui est tiré aléatoirement. Ce sinistre S60015 entraîne donc des Charges FGU très élevées qui constituent la queue de distribution. Il est important de souligner le fait que ce sinistre ne fait pas partie des sinistres d'origine avec les Charges FGU ultimes les plus élevées parmi les 300 sinistres d'origine. Cette observation est due au fait que les traités auxquels sont rattachés les sinistres d'origine n'ont pas d'équivalence pour l'année de survenance 2021. Si de tels sinistres étaient tirés, les provisions IBNyR seraient encore plus élevées.

L'analyse spécifique des provisions IBNeR dans le cadre du scénario 177

Cette dernière sous-sous-section présente quelques résultats des provisions IBNeR de l'année de survenance 2021 du segment $\mathcal{CT}$.

Les statistiques des provisions IBNeR des 1 000 scénarios pour l'année de survenance 2021 sont exposées dans le tableau en figure 6.16 ci-dessous.

Année de surveillance	Provisions IBNR calculées par l'équipe	Cen-tilles	Moyenne	Médiane	Centile 90%	Mini-mum	Maxi-mum
2021	6 875 337	100%	415 356	310 010	887 192	3 476	1 995 287

FIGURE 6.16 – Statistiques de l'année de surveillance 2021 des provisions IBNeR simulées

Les provisions IBNeR maximales sont plus de 200 fois plus faibles que les provisions IBNyR maximales pour la même année de surveillance étudiée, ces dernières étant de plus de 400M€. Cette différence peut s'expliquer par le fait que les provisions IBNyR correspondent à la totalité des Charges FGU ultimes des sinistres IBNyR tandis que les provisions IBNeR correspondent aux Charges FGU qui restent à développer jusqu'à l'ultime des sinistres IBNeR sujets à développement. Les provisions IBNeR sont donc notamment influencées par le nombre de sinistres IBNeR modélisés comme étant sujets à développement.

Les provisions IBNeR estimées dans le scénario 177 sont d'environ 500k€ d'après le tableau en figure 6.9 dans la sous-sous-section précédente. Ce ne sont donc pas les provisions IBNeR les plus élevées puisque ces dernières sont d'environ 1,6M€. Ces provisions du scénario 177 sont situées entre la médiane et le troisième quartile. Le scénario avec les provisions IBNR les plus élevées possibles au regard des provisions IBNeR et IBNyR maximales n'a donc pas été simulé.

Pour rebondir sur la remarque de la sous-section 6.1.2, les distributions des provisions IBNeR sont similaires entre les trois méthodes utilisées, comme le montre le graphique en figure 6.17.

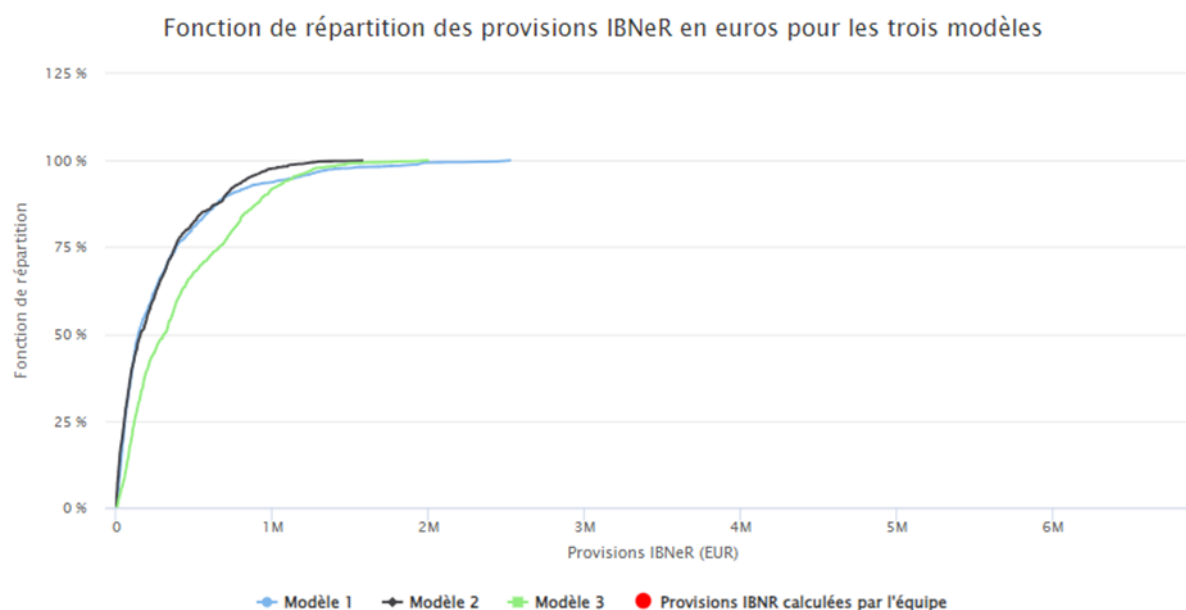


FIGURE 6.17 – Fonctions de répartition des provisions IBNeR pour l'année de surveillance 2021

En effet, la fonction de répartition de la méthode 3 est légèrement en-dessous de celles des méthodes 1 et 2. Cela signifie que les provisions IBNeR simulées par la méthode 3 sont plus nombreuses en proportion que celles simulées par les deux autres méthodes. Plus précisément, la probabilité que les provisions IBNeR soient inférieures à un nombre, est plus élevée avec les méthodes 1 et 2. En revanche, la méthode 1 engendre les provisions IBNeR maximales.

La barre verticale rouge représente les provisions IBNR de 6,8M€ calculées par l'équipe. Contrairement à la fonction de répartition des provisions IBNyR en figure 6.11 de la sous-sous-section précédente, les provisions IBNR calculées par l'équipe dépassent les provisions IBNeR les plus élevées de la distribution. Cela confirme le fait que les provisions IBNyR ont un impact très important dans les provisions IBNR.

Pour information, il existe 37 sinistres appartenant au segment $\mathcal{CT}$ recensés dans la base *Sinistres* pour cette année de survenance 2021. En s'intéressant plus particulièrement aux sinistres IBNeR, ils sont au nombre de 28. Ils représentent donc environ trois quarts des sinistres connus du segment pour cette année de survenance. Certains d'entre eux sont représentés dans le tableau en figure 6.18 ci-dessous. D'autres variables relatives à ces sinistres y sont également présentes, comme leur Charge FGU à leurs différentes années de développement connues.

Numéro de sinistre	Dernière Charge FGU connue	Top Atypique	Année de développement 1	Année de développement 2	Année de développement 3	Année de développement 4
S97957	246 617 218	0	0	217 775 857	217 775 857	246 617 218
S89874	9 161 350	1	9 492 511	9 124 962	9 161 350	9 161 350
S92454	4 322 912	1	6 071 606	4 322 912	4 322 912	4 322 912
S91994	1 212 330	0	461 839	1 212 330	1 212 330	1 212 330
S93655	603 294	0	603 294	603 294	603 294	603 294
S91778	370 433	0	370 433	370 433	370 433	370 433
S94497	106 397	0	0	106 397	106 397	106 397

FIGURE 6.18 – Sinistres IBNeR de l'année de survenance 2021 du segment $\mathcal{CT}$

Il est intéressant de noter que le sinistre S97957, ayant la dernière Charge FGU connue la plus élevée, n'est pas un sinistre atypique.

Parmi ces 28 sinistres IBNeR, tous ne sont pas considérés comme sujets à développement. Comme expliqué dans la sous-section 4.4.3, les sinistres IBNeR sujets à développement sont déterminés grâce à l'utilisation des méthodes de Kaplan-Meier et de la régression logistique.

Concernant la méthode de Kaplan-Meier, les sinistres IBNeR ayant atteint leur année de développement 4 – ce qui est le cas pour ceux survenus en 2021 – ont une probabilité de

0,34 d'être encore sujets à développement pour les sinistres non atypiques. Cette probabilité augmente à 0,39 pour les sinistres atypiques (se référer à la courbe de survie de la méthode Kaplan-Meier à la sous-sous-section 4.4.3).

Le tableau en figure 6.19 ci-dessous présente quelques résultats de la régression logistique pour l'année de développement 4. Plus précisément, il met en avant les 3 sinistres IBNeR ayant les probabilités Proba Logit les plus élevées issues de la régression logistique, ainsi que les 3 sinistres ayant les probabilités Proba Logit les plus faibles.

Numéro de simulation	Proba Logit	Pourcentage de la provision D/D	Top Dvpt Historique
S91778	93%	0%	0
S92454	93 %	0%	0
S89874	93 %	0%	0
S93655	52%	100%	0
S94497	52%	100%	0
S91994	31%	100%	1

FIGURE 6.19 – Probabilités d'avoir atteint l'ultime des sinistres IBNeR issues de la régression logistique

Les probabilités logistiques représentent les probabilités qu'un sinistre IBNeR ne soit plus sujet à développement. Elles sont identiques lorsque les variables Pourcentage de la provision D/D et Top Dvpt Historique ont les mêmes valeurs. En analysant les résultats de la régression logistique pour ces 6 sinistres, il s'agit effectivement de ceux attendus lorsque le modèle a été développé. En effet, les sinistres atypiques ayant connu des développements récemment – comme le S91994 car Top Dvpt Historique vaut 1 – ont plus de chances de poursuivre leur développement que les sinistres non atypiques. À l'inverse, les sinistres présentant des pourcentages de provisions D/D faibles ont bien moins de chances de se développer.

Par ailleurs, il est intéressant de mentionner le fait que, pour la plupart des années de survenance dont l'année 2021, toutes les variables utilisées pour la régression logistique (Année de développement, Top Dvpt Historique, Pourcentage de la provision D/D et Année de développement $\times$ Top Dvpt Historique) présentent une *p-value* inférieure à 0,05, visible dans l'application `RShiny`. Cette valeur traduit une bonne adéquation du modèle de régression logistique aux données.

La suite de cette sous-sous-section permet d'analyser les différentes provisions simulées pour ces sinistres IBNeR.

Le tableau en figure 6.20 ci-dessous représente les statistiques de l'ensemble des coefficients

de développement associés à l'année de développement 2021 et pouvant être tirés aléatoirement afin d'être appliqués aux dernières charges connues des sinistres IBNeR. Dans ces statistiques, les 5% des coefficients les plus élevés sont considérés, bien qu'ils soient écartés lors de la modélisation des sinistres IBNeR (voir la section 4.4).

Moyenne	Médiane	Minimum	Maximum	Centile 95%
1,141	1,000	1,000	261,156	1,040

FIGURE 6.20 – Statistiques des coefficients de développement pour l'année de survenance 2021 du segment $\mathcal{CT}$

Il est intéressant de constater que le coefficient maximal de développement pour l'année de survenance 2021 est supérieur à 200. L'appliquer sur une Charge FGU de quelques millions d'euros amènerait à des provisions IBNeR aberrantes. C'est la raison pour laquelle les 5% des coefficients de développement les plus élevés ne sont pas conservés dans le modèle proposé. Ainsi, les coefficients de développement pouvant être tirés aléatoirement sont au maximum de 1,040, ce qui correspond au centile 95% affiché dans le tableau en figure 6.20 ci-dessus.

Le tableau en figure 6.21 ci-dessous présente les sinistres d'origine à partir desquels sont issus quelques coefficients de développement. Ce tableau est disponible dans l'application RShiny pour l'ensemble des coefficients de développement pour chaque année de survenance.

Numéro de sinistre	Année de survenance	Dernière Charge FGU connue	Charge FGU en année de développement 4	Ultime Atteint	<i>Link ratio</i>	Facteur de développement lissé	Coefficient de développement
S51936	2008	141 529	542	1	261,156	1,004	261,156
S44344	1999	387 295	106 440	0	5,297	1,001	5,301
S57223	2010	6 389 699	6 277 857	1	1,016	1,005	1,016

FIGURE 6.21 – Coefficients de développement et leur sinistre d'origine

Le coefficient de développement maximal provient donc d'un sinistre à très faible Charge FGU qui n'est pas spécialement représentatif des Charges FGU des sinistres IBNeR. Cela amène à confirmer le fait que le choix de tirer aléatoirement parmi des coefficients issus de sinistres d'origine de Charges FGU similaires est pertinent.

Concernant le coefficient 1,040 associé au centile 95%, il s'agit de la valeur du facteur de développement lissé pour l'année de développement 5. La raison de ce phénomène est la suivante. La plupart des sinistres survenus en 2020 sont associés à un *link ratio* égal à 1 et ne sont pas

encore identifiés comme ayant atteint leur ultime. Ainsi, le coefficient de développement qui leur est associé est le produit de leur *link ratio* et du facteur de développement lissé de l'année de développement 5. C'est pourquoi de nombreux sinistres d'origine sont associés à des coefficients de développement égaux à 1,040.

Le tableau en figure 6.22 correspond aux sinistres IBNeR modélisés comme étant sujets à développement dans le cadre du scénario 177 ainsi que les provisions IBNeR qu'ils engendrent.

Numéro de sinistre	Dernière Charge FGU connue	Coefficient de développement tiré aléatoirement	Numéro du sinistre d'origine	Charge FGU ultime	Provisions IBNeR
S94577	11 659 022	1,016	S57223	11 845 566	220 636
S94617	11 442 514	1,009	S50822	11 545 497	123 815
S94679	15 496 135	1,005	S74897	15 573 616	79 536
S95137	2 537 424	1,017	S12853	2 580 560	50 843
S94277	2 081 040	1,023	S78319	2 128 904	27 873
S93475	1 783 856	1,003	S76694	1 789 208	6 328
S94117	2 030 400	1,001	S72124	2 032 430	2 142
S92695	24 996 696	1,001	S31550	25 021 693	1 891
S94437	1 363 969	1,001	S33433	1 365 333	1 371
S91994	1 212 330	1,001	S24730	1 213 542	583

FIGURE 6.22 – Provisions IBNeR des sinistres IBNeR sujets à développement

Parmi les 28 sinistres considérés comme étant IBNeR, 10 sont modélisés comme n'ayant pas encore atteint leur ultime et pour lesquels un coefficient de développement doit donc leur être appliqué.

Lorsque les provisions IBNeR des sinistres sujets à développement sont additionnées pour l'année de survenance 2021, les provisions IBNeR totales du scénario 177 sont égales à 515 048, ce qui correspond à la valeur des provisions IBNeR dans le tableau précédent en figure 6.9.

Dans l'application **RShiny**, lorsqu'un sinistre IBNeR est sélectionné, tous les coefficients de développement et les sinistres dont ils sont issus et pouvant être tirés aléatoirement pour le sinistre sélectionné, sont affichés. Cette liste de coefficients diffère selon la méthode de modélisation des sinistres IBNeR choisie.

Conclusion du chapitre

Les provisions IBNR sont engendrées en grande partie par les provisions IBNyR, que ce soit dans les scénarios pour lesquels les provisions IBNR simulées sont les plus proches de celles calculées par l'équipe ou dans les « pires » scénarios. Les provisions IBNyR sont conséquentes pour chaque sinistre IBNyR simulé puisqu'elles correspondent à leur Charge FGU ultime. Les Charges FGU ultimes simulées exercent donc une influence très importante dans le montant de provisions IBNyR et mériteraient une méthodologie plus approfondie pour leur simulation. Les provisions IBNeR sont quant à elles plus faibles pour chaque sinistre IBNeR sujet à développement. Cependant, le nombre de ces derniers peut être élevé. Les montants des provisions IBNeR sont donc quant à eux plutôt influencés par le nombre de sinistres IBNeR simulés comme étant sujets à développement, que par leur coefficient de développement qui est appliqué à leur dernière Charge FGU connue. La démarche de l'analyse dans ce chapitre est applicable pour tous les segments de provisionnement.

Conclusion

Au sein des compagnies d'assurance et de réassurance, le provisionnement constitue un enjeu très important puisqu'il permet aux compagnies d'honorer leurs engagements envers leurs clients. Dans le cadre de la réassurance, ce mémoire s'intéresse à un type spécifique de provisions : les provisions IBNR (*Incurred But Not Reported*). Ces dernières sont celles engendrées par des sinistres dont la charge ultime n'est pas encore connue. Deux types de sinistres composent les sinistres IBNR : les sinistres IBNeR et les sinistres IBNyR dont la somme de leurs provisions constituent les provisions IBNR. Les sinistres IBNeR correspondent aux sinistres déjà déclarés mais dont les charges n'ont pas encore atteint leur ultime. Les sinistres IBNyR désignent quant à eux les sinistres tardifs.

L'équipe Reserving du département AXA Group Ceded Reinsurance calcule ces provisions IBNR à l'aide de méthodes de provisionnement agrégées, sans distinction entre les provisions IBNeR et IBNyR. Cependant, ces méthodes, pouvant conduire à certains biais en raison du faible nombre de sinistres en réassurance et de leur volatilité accrue, soulèvent la question de la manière dont ces provisions IBNR pourraient être consommées par les différents sinistres.

Après une structuration des bases de données de sinistres et de traités, ce mémoire propose une approche stochastique pour répondre à la problématique d'interprétabilité des provisions IBNR. Il s'appuie sur le provisionnement individuel qui consiste à simuler la charge de chaque sinistre IBNR à l'ultime à partir de laquelle les provisions IBNR sont estimées. Les charges ultimes des sinistres sont modélisées de façon stochastique afin de refléter le caractère incertain du futur. Par conséquent, les provisions IBNR simulées présentent également ce caractère stochastique. Pour tirer parti de ce comportement stochastique, 1 000 scénarios de développement sont générés afin d'obtenir une distribution de provisions IBNR pour chaque année de survenance des 80 segments de provisionnement. Cette distribution permet de positionner les provisions IBNR calculées par l'équipe par rapport à chacun des 1 000 scénarios simulés.

Dans le cadre de ce mémoire, les modélisations des sinistres IBNeR et des sinistres IBNyR sont traitées séparément afin de tenir compte de leurs spécificités respectives. D'une part, les sinistres IBNeR sont développés jusqu'à l'ultime après avoir été identifiés comme tels. D'autre part, la modélisation des sinistres IBNyR s'articule autour de deux étapes : en premier lieu, le

nombre de sinistres IBNyR est estimé et en second lieu, une charge ultime est attribuée à chacun de ces sinistres IBNyR simulés. Une fois les charges de ces sinistres développées à l'ultime, les provisions IBNeR et IBNyR sont déterminées avant d'être additionnées afin de permettre une comparaison avec les provisions IBNR calculées par l'équipe Reserving.

Les résultats qui découlent de cette comparaison supposent que les provisions IBNR calculées par l'équipe Reserving sont surestimées pour la plupart des années de développement, ce qui établirait la prudence avérée des calculs de l'équipe. Une analyse plus approfondie révèle que les provisions IBNyR sont plus volatiles et plus élevées en moyenne que les provisions IBNeR. En effet, les provisions IBNyR peuvent prendre des valeurs très élevées en raison de la modélisation proposée des charges ultimes des sinistres IBNyR, tout comme elle peuvent être nulles. Les nombres de sinistres simulés sont quant à eux faibles. La génération de toutes les provisions IBNeR et IBNyR est transparente quel que soit le segment de provisionnement, notamment à travers l'application *RShiny* développée et déjà exploitée par l'équipe Reserving. Cela permet ainsi de répondre à la demande d'interprétabilité des provisions IBNR calculées par l'équipe en s'appuyant sur celles estimées par le modèle mis en œuvre dans le cadre de ce mémoire.

Le modèle élaboré dans ce mémoire offre plusieurs avantages pour l'équipe Reserving. D'une part, il permet de justifier les provisions IBNR calculées par l'équipe en leur associant des scénarios générant des provisions IBNR qui en sont proches. D'autre part, ce modèle permet d'identifier des écarts entre les provisions IBNR qu'il génère et celles calculées par l'équipe.

Cependant, les résultats du modèle proposé sont à nuancer. Tout d'abord, plusieurs hypothèses sont émises dans la modélisation. Leur impact mériterait d'être approfondi. De plus, une grande partie du modèle s'appuie sur des données historiques. Dans le cadre de la modélisation des sinistres IBNyR, ce choix implique que tous les risques simulés – peu nombreux par ailleurs – se sont déjà réalisés, empêchant la considération de nouveaux risques qui ne se sont encore jamais matérialisés. Le caractère aléatoire de la modélisation des sinistres IBNyR et donc de leurs provisions, se trouve ainsi atténué. Pour pallier ce problème, une simulation des charges ultimes des sinistres IBNyR selon une loi de sévérité serait à envisager lors de prochains travaux.

En outre, les reconstitutions, les AAD, les AAL et les priorités et portées indexées ne sont pas encore prises en compte dans le modèle. Considérer ces données pourrait améliorer l'estimation des provisions IBNR. Toutefois, les portées et priorités indexées pourront aisément être ajoutées au modèle lorsqu'elles seront disponibles grâce à la structure des données.

Dans une perspective à plus long terme, lorsque le modèle sera consolidé, l'analyse de la cause des différences significatives entre les provisions IBNR simulées et celles calculées par l'équipe Reserving pourrait ainsi mener cette dernière à s'interroger sur les provisions IBNR qu'elle calcule et à évaluer la pertinence du modèle proposé dans ce mémoire. Le cas échéant, l'équipe pourrait ajuster ses méthodes agrégées afin de refléter plus justement la sinistralité.

Bibliographie

- Antonio, K., Godecharle, E. and Oirbeek, R. V. (2016), ‘A multi-state approach and flexible payment distributions for micro-level reserving in general insurance’, *SSRN Electronic Journal* .
- Baradel, N. (2024), ‘Actuariat de l’assurance non-vie’.
- Bücher, A. and Rosenstock, A. (2022), ‘Micro-level prediction of outstanding claim counts based on novel mixture models and neural networks’, *European Actuarial Journal* .
- Chaoubi, I., Besse, C., Cossette, H. and Côté, M.-P. (2023), ‘Micro-level reserving for general insurance claims using a long short-term memory network’, *Applied Stochastic Models in Business and Industry* **39**(3), 382–407.
- Charpentier, A. and Pigeon, M. (2016), ‘Macro vs. micro methods in non-life claims reserving (an econometric perspective)’ . hal-01280033.
- France Assureurs (2024), L’assurance française : Données clés 2023, Technical report.
- Gabrielli, A. (2021), ‘An individual claims reserving model for reported claims’, *European Actuarial Journal* .
- Giuly, A. (2018), Approches de provisionnement individuel en réassurance non-vie, Master’s thesis, Université de Paris Dauphine, Paris.
- Groupe de Travail sur le Provisionnement des Sinistres Non-Vie (2023), Guide de provisionnement des sinistres en assurance non-vie, Technical report, Institut des Actuaire.
- Jin, X. (2013), Comparing micro- and macro-level loss reserving models.
- Lopez, O. (2019), ‘A censored copula model for micro-level claim reserving’, *Insurance : Mathematics and Economics* pp. 1–14. hal-01706935.
- Lopez, O. (2023), ‘Duration models and survival analysis’.

- Lopez, O., Milhaud, X. and Thérond, P.-E. (2016), ‘Tree-based censored regression with applications in insurance’, *Electronic Journal of Statistics* **10**(2), 2685–2716.
- Lopez, O., Milhaud, X. and Thérond, P.-E. (2018), ‘Micro-level vs macro-level reserving in non-life insurance : why and when ?’. hal-01868744v1.
- Marceau, E. (2013), *Modélisation et évaluation quantitative des risques en actuariat*, Springer.
- Parlement Européen et Conseil de l’Union Européenne (2009), ‘Directive 2009/138/ce du 25 novembre 2009 sur l’accès aux activités de l’assurance et de la réassurance et leur exercice (solvabilité II)’.
- Primel, D. (2021), Comparaison de différents modèles de provisionnement ligne à ligne et discussions dans le cadre d’une application en non-vie, Master’s thesis, Université de Paris Dauphine, Paris.
- Riffaud, J. (2020), Modélisation de la charge ibnyr dans le cadre d’un modèle de provisionnement individuel en assurance non-vie, Master’s thesis, Université de Paris Dauphine, Paris.
- Stochastic Claims Reserving Methods in Insurance* (2012), John Wiley & Sons, Ltd, chapter Front Matter, pp. i–xiii.

Table des figures

1	Récapitulatif de l'identification des sinistres IBNeR sujets à développement . . .	7
2	Fonction de répartition des provisions IBNR d'un segment de provisionnement .	9
3	Summary of the identification of IBNeR claims subject to development	13
4	A distribution function for IBNR reserves of a line of business	15
1.1	Les différents acteurs de la réassurance et leurs liens	25
1.2	Le fonctionnement des trois formes de réassurance	28
1.3	Tableau des primes et montants de sinistres dans les cas de la rétention et de la cession	29
1.4	Exemple d'un traité de réassurance QP	30
1.5	Exemple de traité de réassurance en EP	31
1.6	Exemple de traité de réassurance à couverture non proportionnelle	32
1.7	Exemple d'un programme de réassurance	33
1.8	Exemple d'un traité de réassurance en Excédent de Sinistres par événement . . .	34
3.1	Présentation des variables de la base <i>Traités</i>	51
3.2	Présentation des variables de la base <i>Sinistres</i>	52
3.3	Extrait de la base de données <i>Sinistres</i>	54
3.4	Profondeur historique des trois segments de provisionnement	56
3.5	Différentes données de sinistres	57
3.6	Triangle de montants cumulés	57
3.7	Triangle de <i>link ratios</i> incrémentaux	58
3.8	Triangle de <i>link ratios</i> cumulés	59
3.9	Triangle de développement des sinistres	61

3.10	Tableau contenant des pourcentages de provisions D/D	61
3.11	Extrait de la base <i>Traités</i>	62
3.12	Extrait de la base <i>Sinistres</i>	62
3.13	Extrait de la base <i>Sinistres</i> après la fusion avec la base <i>Traités</i>	62
3.14	Tableau récapitulatif du calcul des différentes charges	64
4.1	Nombre de sinistres et pourcentage selon la variable Statut	68
4.2	Pourcentage de sinistres actifs et clos pour le segment $\mathcal{CT}$	69
4.3	Pourcentage de sinistres actifs et clos pour le segment $\mathcal{MT}$	69
4.4	Pourcentage de sinistres actifs et clos pour le segment $\mathcal{LT}$	70
4.5	Nombre de sinistres actifs et clos du segment $\mathcal{LT}$	70
4.6	Données associées au sinistre actif S2539	71
4.7	Données associées au sinistre actif S2544	71
4.8	Données associées au sinistre actif S10035	71
4.9	Nombre de sinistres actifs et clos du segment $\mathcal{CT}$	72
4.10	Données associées au sinistre actif S5575	72
4.11	Nombre de sinistres actifs et clos du segment $\mathcal{MT}$	73
4.12	Données associées au sinistre actif S34670	73
4.13	Données associées au sinistre actif S39633	73
4.14	Variable Ultime Atteint associée aux sinistres	75
4.15	Facteurs de développement historiques	78
4.16	Comparaison de la paramétrisation entre le cadre théorique et celui du mémoire	79
4.17	Comparaison des R^2 entre les différentes fonctions paramétriques	80
4.18	Facteurs lissés et cadences du segment $\mathcal{CT}$	81
4.19	Facteurs lissés et cadences du segment $\mathcal{MT}$	81
4.20	Facteurs lissés et cadences du segment $\mathcal{LT}$	81
4.21	Seuils de développement	82
4.22	Cadences et seuil de développement du segment $\mathcal{CT}$	83
4.23	Cadences et seuil de développement du segment $\mathcal{MT}$	83
4.24	Cadences et seuil de développement du segment $\mathcal{LT}$	84

4.25	Valeur de la variable Ultime Atteint d'un sinistre selon ses années de développement	86
4.26	Matrice de contigence du segment $\mathcal{CT}$	87
4.27	Matrice de contigence du segment $\mathcal{MT}$	87
4.28	Matrice de contigence du segment $\mathcal{LT}$	87
4.29	Tableau des pourcentages de sinistres appartenant à la même catégorie	87
4.30	Valeurs des variables pour le sinistre S2539	88
4.31	Valeurs des variables pour le sinistre S10035	88
4.32	Valeurs des variables pour le sinistre S2544	88
4.33	Valeurs des variables pour le sinistre S5575	89
4.34	Valeurs des variables pour le sinistre S34670	89
4.35	Valeurs des variables pour le sinistre S39633	89
4.36	Tableau récapitulatif des sinistres étudiés	89
4.37	Identification des sinistres atypiques et des sinistres non atypiques	90
4.38	Variable Top Dvpt Historique associée à un sinistre	92
4.39	Obtention des coefficients de développement	95
4.40	Table des sinistres IBNeR	97
4.41	Table des coefficients de développement et leur sinistre d'origine	98
4.42	Table des sinistres IBNeR sujets à évolution	102
4.43	Table des coefficients de développement et leur sinistre d'origine	103
4.44	Récapitulatif du tirage aléatoire des coefficients dans les trois méthodes	104
4.45	Comparaisons de l'utilisation de la méthode de Kaplan-Meier entre le cadre de la durée de vie et celui de l'atteinte à l'ultime des sinistres	107
4.46	Courbes de survie du segment $\mathcal{CT}$	108
4.47	Courbes de survie du segment $\mathcal{MT}$	108
4.48	Courbes de survie du segment $\mathcal{LT}$	108
4.49	Données nécessaires à la simulation des sinistres IBNeR sujets à développement	110
4.50	Tableau d'interaction des variables Année de développement et Top Dvpt Historique	113
4.51	Classement des sinistres IBNeR par l'intermédiaire de la régression logistique	116
4.52	Détermination des provisions IBNeR	118

5.1	Détermination des sinistres historiques pour le triangle de nombres	121
5.2	Expositions par année de survenance	122
5.3	Triangle de nombres	122
5.4	Triangle de nombres complété par l'espérance du nombre de sinistres IBNyR . .	123
5.5	Triangle de nombres complété par la variance du nombre de sinistres IBNyR . .	123
5.6	Triangle de nombres complété par les lois selon lesquelles sont simulés les nombres de sinistres IBNyR	124
5.7	Triangle de nombres complété par les nombres simulés de sinistres IBNyR . . .	125
5.8	Extrait de données des sinistres d'origine	126
5.9	Traités de réassurance et leurs équivalents pour l'année d'exercice 2021	128
5.10	Traités de réassurance pour les sinistres IBNyR simulés à l'année de survenance 2023	129
5.11	Tableau des différentes charges associées aux trois sinistres	129
6.1	Statistiques des provisions IBNR simulées pour le segment $\mathcal{CT}$	133
6.2	Statistiques des provisions IBNR simulées pour le segment $\mathcal{MT}$	134
6.3	Statistiques des provisions IBNR simulées pour le segment $\mathcal{LT}$	134
6.4	Statistiques des provisions IBNR simulées pour le segment $\mathcal{MT}$ dans le cadre de la méthode 1	136
6.5	Statistiques de l'année de survenance 2021 des provisions IBNR simulées pour le segment $\mathcal{CT}$	137
6.6	Fonction de répartition des provisions IBNR de l'année de survenance 2021 pour le segment $\mathcal{CT}$	138
6.7	Statistiques des nombres de sinistres de la survenance 2021 du segment $\mathcal{CT}$. .	138
6.8	Les 10 scénarios les plus proches des provisions IBNR calculées par l'équipe . .	139
6.9	Les 10 scénarios les plus extrêmes de la survenance 2021 pour le segment $\mathcal{CT}$.	140
6.10	Triangle incrémental de nombres des sinistres IBNyR du segment $\mathcal{CT}$	141
6.11	Fonction de répartition des provisions IBNyR pour l'année de survenance 2021 .	142
6.12	Statistiques de l'année de survenance 2021 des provisions IBNyR simulées . . .	142
6.13	Statistiques de la survenance 2021 des nombres simulés de sinistres IBNyR . .	143
6.14	Réalisations des lois pour les nombres de sinistres IBNyR	143

6.15	Simulation des provisions ACC de sinistres IBNyR	144
6.16	Statistiques de l'année de survenance 2021 des provisions IBNeR simulées	145
6.17	Fonctions de répartition des provisions IBNeR pour l'année de survenance 2021 .	145
6.18	Sinistres IBNeR de l'année de survenance 2021 du segment $\mathcal{CT}$	146
6.19	Probabilités d'avoir atteint l'ultime des sinistres IBNeR issues de la régression logistique	147
6.20	Statistiques des coefficients de développement pour l'année de survenance 2021 du segment $\mathcal{CT}$	148
6.21	Coefficients de développement et leur sinistre d'origine	148
6.22	Provisions IBNeR des sinistres IBNeR sujets à développement	149

Annexe A

Lois de probabilités ²

A.1 La loi uniforme discrète

Soit $\{x_1, \dots, x_n\}$ une partie de $\mathbb{R}$ à n éléments, avec $n \in \mathbb{N}^*$.

Une variable aléatoire $X : \Omega \rightarrow \mathbb{R}$ suit une loi uniforme discrète sur $\{x_1, \dots, x_n\}$, notée $X \sim \mathcal{U}(\{x_1, \dots, x_n\})$, si :

- $X(\Omega) = \{x_1, \dots, x_n\}$;
- $\forall i \in \{1, \dots, n\}, \mathbb{P}(X = x_i) = \frac{1}{n}$.

L'espérance de la variable aléatoire X est :

$$\mathbb{E}(X) = \frac{\sum_{i=1}^n x_i}{n}.$$

La variance de la variable aléatoire X est :

$$\mathbb{V}(X) = \frac{\sum_{i=1}^n x_i^2}{n} - \mathbb{E}^2(X).$$

La particularité de cette loi réside dans le fait que toutes les valeurs de $\{x_1, \dots, x_n\}$ sont équiprobables.

A.2 La loi binomiale

Soit $n \in \mathbb{N}$ et $q \in (0; 1)$.

2. ([Marceau, 2013](#))

Une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ suit une loi binomiale de paramètres n et q , notée $X \sim \mathcal{B}(n; q)$, si :

- $X(\Omega) = \{0; \dots; n\}$;
- $\forall k \in \{0; \dots; n\}, P(X = k) = \binom{n}{k} q^k (1 - q)^{n-k}$.

L'espérance de la variable aléatoire X est :

$$\mathbb{E}(X) = nq.$$

La variance de la variable aléatoire X est :

$$\mathbb{V}(X) = nq(1 - q).$$

A.3 La loi binomiale négative

Soit $r \in \mathbb{R}^+$ et $q \in (0; 1)$.

Une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ suit la loi binomiale négative de paramètres r et q , notée $X \sim \mathcal{BN}(r; q)$, si :

- $X(\Omega) = \mathbb{N}$;
- $\forall k \in \mathbb{N}, P(X = k) = \binom{r+k-1}{k} q^r (1 - q)^k$.

L'espérance de la variable aléatoire X est :

$$\mathbb{E}(X) = r \frac{1 - q}{q}.$$

La variance de la variable aléatoire X est :

$$\mathbb{V}(X) = r \frac{1 - q}{q^2}.$$

À partir de ces formules d'espérance et de variance, il est possible de retrouver les paramètres r et q en fonction de la variance et de l'espérance :

$$r = \frac{\mathbb{E}(X)}{\mathbb{V}(X)} \quad \text{et} \quad q = \frac{\mathbb{E}^2(X)}{\mathbb{V}(X) - \mathbb{E}(X)}.$$

A.4 La loi de Poisson

Soit $\lambda > 0$.

Une variable aléatoire $X : \Omega \rightarrow \mathbb{N}$ suit une loi de Poisson de paramètre λ , notée $X \sim \mathcal{P}(\lambda)$, si :

- $X(\Omega) = \mathbb{N}$;
- $\forall k \in \mathbb{N}, P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

L'espérance et la variance de la variable aléatoire X sont identiques et sont égales à λ :

$$\mathbb{E}(X) = \mathbb{V}(X) = \lambda.$$

Annexe B

La méthode de la *Chain Ladder*²

Définissons, pour $1 \leq i, j \leq n$, $C_{i,j}$ le coût total des sinistres survenus l'année i et observés jusqu'à l'année de développement j . $C_{i,j}$ est $\mathcal{F}_k$ -mesurable si $i + j \leq k + 1$. Si l'on suppose que les sinistres se développent en un maximum de n années, alors $C_{i,n}$ est le coût total des sinistres survenus pendant l'année i . L'objectif principal de la méthode de la *Chain Ladder* est alors d'estimer la partie inconnue du triangle de développement associé à ces coûts et d'en déduire une approximation du montant restant à payer pour les contrats passés. Cela revient à estimer $(C_{i,n})_{2 \leq i \leq n}$.

Il s'agit de la méthode la plus utilisée en assurance. Elle repose sur les trois hypothèses suivantes :

H1 Les variables aléatoires $(C_{i_1,j})_{1 \leq j \leq n}$ et $(C_{i_2,j})_{1 \leq j \leq n}$ sont indépendantes pour $i_1 \neq i_2$.

H2 Pour $1 \leq j \leq n - 1$, il existe $f_j \geq 0$ tel que

$$\mathbb{E}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) = f_j C_{i,j}, \quad 1 \leq i \leq n.$$

H3 Pour $1 \leq j \leq n - 1$, il existe $\sigma_j \geq 0$ tel que

$$\mathbb{V}(C_{i,j+1} \mid \mathcal{F}_{i+j-1}) = \sigma_j^2 C_{i,j}, \quad 1 \leq i \leq n.$$

Pour $1 \leq k \leq n$, introduisons la filtration :

$$\mathcal{B}_k := \sigma(C_{i,j} \mid i + j \leq n + 1, j \leq k) \subset \mathcal{F}_n.$$

2. ([Baradel, 2024](#))

Sous l'hypothèse $H1$, les hypothèses $H2$ et $H3$ vérifient alors également :

$$\mathbb{E}(C_{i,j+1} \mid \mathcal{B}_j) = f_j C_{i,j} \quad \text{et} \quad \mathbb{V}(C_{i,j+1} \mid \mathcal{B}_j) = \sigma_j^2 C_{i,j}.$$

Posons $\widehat{f}_{i,j} := \frac{C_{i,j+1}}{C_{i,j}}$ pour $i + j + 1 \leq n + 1$.

Avec des poids positifs $p := (p_{i,j})$ non tous nuls, définissons un estimateur de f_j :

$$\widehat{f}_j := \frac{\sum_{i=1}^{n-j} p_{i,j} \widehat{f}_{i,j}}{\sum_{i=1}^{n-j} p_{i,j}}.$$

Lemme 1. *Pour tout $1 \leq j \leq n - 1$:*

$$\mathbb{E}(\widehat{f}_{i,j} \mid \mathcal{B}_j) = f_j.$$

Proposition 1. *Sous l'hypothèse $H2$, pour tout $1 \leq j \leq n - 1$, pour tous les poids p qui sont $\mathcal{B}_j$ -mesurables, l'estimateur $\widehat{f}_j$ est un estimateur sans biais de f_j conditionnellement à l'information disponible $\mathcal{B}_j$, c'est-à-dire :*

$$\mathbb{E}(\widehat{f}_j \mid \mathcal{B}_j) = f_j.$$

En particulier, il est sans biais :

$$\mathbb{E}(\widehat{f}_j) = f_j.$$

Proposition 2. *Sous les hypothèses $H1$, $H2$ et $H3$, le meilleur estimateur linéaire de f_j en $(\widehat{f}_{i,j})_{1 \leq i \leq n-j}$ pour l'erreur quadratique moyenne conditionnellement à $\mathcal{B}_j$ est :*

$$\widehat{f}_j := \frac{\sum_{i=1}^{n-j} C_{i,j} \widehat{f}_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}.$$

L'objectif est maintenant d'estimer la partie inconnue du triangle.

Lemme 2. *Soient $1 \leq i \leq n$ et $n + 1 - i \leq j \leq n$. On a :*

$$\mathbb{E}(C_{i,j} \mid \mathcal{F}_n) = \left(\prod_{k=n+1-i}^{j-1} f_k \right) C_{i,n+1-i}.$$

Dans le lemme 2 ci-dessus, il est nécessaire de remplacer les f_j par leurs estimateurs. Le lemme 3 ci-dessous implique en particulier que les estimateurs $\widehat{f}_j$ ne sont pas corrélés entre eux.

Lemme 3. Soient $1 \leq k \leq n-1$ et $j_1 < j_2 < \dots < j_k$, on a :

$$\mathbb{E} \left(\prod_{i=1}^k \widehat{f}_{j_i} \mid \mathcal{B}_{j_1} \right) = \prod_{i=1}^k \mathbb{E}(\widehat{f}_{j_i}) = \prod_{i=1}^k f_{j_i}.$$

Introduisons maintenant l'estimateur des $C_{i,j}$ de la partie inconnue du triangle.

Proposition 3. Soit $1 \leq i \leq n$ et $n+1-i \leq j \leq n$. L'estimateur

$$\widehat{C}_{i,j} := \left(\prod_{k=n+1-i}^{j-1} \widehat{f}_k \right) C_{i,n+1-i}$$

est un estimateur sans biais de $C_{i,j}$ conditionnellement à $\mathcal{B}_{n+1-i}$.

Le résultat précédent permet d'en déduire un estimateur sans biais des provisions associées à chaque année de survénance qui est le suivant :

$$\widehat{R}_i := \widehat{C}_{i,n} - C_{i,n+1-i}.$$

Et enfin, un estimateur du besoin en provisions totales est :

$$\widehat{R} := \sum_{i=2}^n \widehat{R}_i.$$

Annexe C

La méthode de Schnieper ²

La méthode de Schnieper a recours à l'exposition, contrairement à la méthode de la *Chain Ladder*. Cette notion importante correspond à une mesure du risque du portefeuille. L'exposition de l'année de survenance i est notée $E_i \in \mathbb{R}^+$. Le montant total des primes est un exemple de mesure de l'exposition qui tient compte de la différence des risques des différents contrats.

Définissons quelques variables :

- Les variables aléatoires $(C_{i,j})_{1 \leq i,j \leq n}$ représentent le coût total cumulé des sinistres.
- Les variables aléatoires $(N_{i,j})_{1 \leq i,j \leq n}$ représentent le montant des nouveaux sinistres survenus en année i et déclarés en année de développement j .
- Les variables aléatoires $(D_{i,j})_{1 \leq i,j \leq n}$ représentent la baisse du coût estimé des sinistres déclarés les années précédentes.

Par construction, $D_{i,1} = 0$ pour tout $1 \leq i \leq n$ et $D_{i,j} \leq C_{i,j-1}$ pour $1 \leq i \leq n$ et $2 \leq j \leq n$.

Le coût cumulé des sinistres s'écrit :

$$C_{i,1} = N_{i,1}, \quad 1 \leq i \leq n,$$

$$C_{i,j+1} = C_{i,j} + N_{i,j+1} - D_{i,j+1}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq n-1.$$

Parmi les variables aléatoires C , N et D , il suffit d'en connaître deux pour en déduire la troisième car l'un d'entre eux est redondant linéairement.

Introduisons la filtration :

$$\mathcal{F}_k := \sigma(N_{i,j}, D_{i,j} \mid i+j \leq k+1).$$

Nous avons donc deux triangles de données, dont l'agrégation donne la variable aléatoire C .

2. ([Baradel, 2024](#))

Le modèle repose sur les trois hypothèses suivantes :

H1 Les variables aléatoires $(N_{i_1,j}, D_{i_1,j})_{1 \leq j \leq n}$ et $(N_{i_2,j}, D_{i_2,j})_{1 \leq j \leq n}$ sont indépendantes pour $i_1 \neq i_2$.

H2 Pour $1 \leq j \leq n$, il existe $\lambda_j \geq 0$ et pour $1 \leq j \leq n-1$, il existe $\delta_j \leq 1$ tels que

$$\mathbb{E}(N_{i,j} \mid \mathcal{F}_{i+j-2}) = \lambda_j E_i \quad \text{et} \quad \mathbb{E}(D_{i,j+1} \mid \mathcal{F}_{i+j-1}) = \delta_j C_{i,j}.$$

H3 Pour $1 \leq j \leq n-1$, il existe $\sigma_j^2 \geq 0$ et $\tau_j^2 \geq 0$ tels que

$$\mathbb{V}(N_{i,j} \mid \mathcal{F}_{i+j-2}) = \sigma_j^2 E_i \quad \text{et} \quad \mathbb{V}(D_{i,j+1} \mid \mathcal{F}_{i+j-1}) = \tau_j^2 C_{i,j}.$$

Remarque 1. Par la formule des espérances conditionnelles emboîtées et de la variance totale, en combinant les hypothèses *H2* et *H3*, on a également :

$$\mathbb{E}(N_{i,j}) = \lambda_j E_i \quad \text{et} \quad \mathbb{V}(N_{i,j}) = \sigma_j^2 E_i.$$

Pour $1 \leq k \leq n$, introduisons la filtration :

$$\mathcal{B}_k = \sigma(N_{i,j}, D_{i,j} \mid i+j \leq n+1, j \leq k) \subset \mathcal{F}_n.$$

Remarque 2. Sous l'hypothèse *H1*, les hypothèses *H2* et *H3* vérifient également :

$$\mathbb{E}(N_{i,j} \mid \mathcal{B}_{j-1}) = \lambda_j E_i \quad \text{et} \quad \mathbb{E}(D_{i,j+1} \mid \mathcal{B}_j) = \delta_j C_j,$$

$$\mathbb{V}(N_{i,j} \mid \mathcal{B}_{j-1}) = \sigma_j^2 E_i \quad \text{et} \quad \mathbb{V}(D_{i,j+1} \mid \mathcal{B}_j) = \tau_j^2 C_j.$$

Posons $\hat{\lambda}_{i,j} := \frac{N_{i,j}}{E_i}$ pour $i+j \leq n+1$ et $\hat{\delta}_{i,j} := \frac{D_{i,j+1}}{C_{i,j}}$ pour $i+j+1 \leq n+1$.

En introduisant des poids, un estimateur de λ_j et un estimateur de δ_j sont :

$$\hat{\lambda}_j := \frac{\sum_{i=1}^{n-j+1} p_{i,j} \hat{\lambda}_{i,j}}{\sum_{i=1}^{n-j+1} p_{i,j}}, \quad 1 \leq j \leq n,$$

$$\hat{\delta}_j := \frac{\sum_{i=1}^{n-j} p'_{i,j} \hat{\delta}_{i,j}}{\sum_{i=1}^{n-j} p'_{i,j}}, \quad 1 \leq j \leq n-1.$$

Lemme 4. On a :

$$\mathbb{E}(\hat{\lambda}_{i,j} \mid \mathcal{B}_{j-1}) = \lambda_j, \quad 1 \leq j \leq n,$$

$$\mathbb{E}(\hat{\delta}_{i,j} \mid \mathcal{B}_j) = \delta_j, \quad 1 \leq j \leq n-1.$$

Proposition 4. Sous l'hypothèse *H2*, pour tout $1 \leq j \leq n$, pour tous les poids $\mathcal{B}_{j-1}$ -mesurables

pour l'estimateur $\widehat{\lambda}_j$, et pour tout $1 \leq j \leq n-1$ pour tous les poids $\mathcal{B}_j$ -mesurables pour l'estimateur $\widehat{\delta}_j$, on a :

$$\mathbb{E}(\widehat{\lambda}_j \mid \mathcal{B}_{j-1}) = \lambda_j \quad \text{et} \quad \mathbb{E}(\widehat{\delta}_j \mid \mathcal{B}_j) = \delta_j.$$

En particulier, ils sont sans biais :

$$\mathbb{E}(\widehat{\lambda}_j) = \lambda_j \quad \text{et} \quad \mathbb{E}(\widehat{\delta}_j) = \delta_j.$$

Proposition 5. *Sous les hypothèses H1, H2 et H3, le meilleur estimateur linéaire de λ_j en $(\widehat{\lambda}_{i,j})_{1 \leq i \leq n-j+1}$ conditionnellement à $\mathcal{B}_{j-1}$ et le meilleur estimateur linéaire de δ_j en $(\widehat{\delta}_{i,j})_{1 \leq i \leq n-j}$ conditionnellement à $\mathcal{B}_j$ pour l'erreur quadratique moyenne sont :*

$$\widehat{\lambda}_j := \frac{\sum_{i=1}^{n-j+1} E_i \widehat{\lambda}_{i,j}}{\sum_{i=1}^{n-j+1} E_i} = \frac{\sum_{i=1}^{n-j+1} N_{i,j}}{\sum_{i=1}^{n-j+1} E_i} \quad \text{et} \quad \widehat{\delta}_j := \frac{\sum_{i=1}^{n-j} C_{i,j} \widehat{\delta}_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}} = \frac{\sum_{i=1}^{n-j} D_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}.$$

Lemme 5. *On a :*

$$\mathbb{V}(\widehat{\lambda}_j \mid \mathcal{B}_{j-1}) = \frac{\sigma_j^2}{\sum_{i=1}^{n-j+1} E_i} \quad \text{et} \quad \mathbb{V}(\widehat{\delta}_j \mid \mathcal{B}_j) = \frac{\tau_j^2}{\sum_{i=1}^{n-j} C_{i,j}}.$$

L'objectif est maintenant d'estimer la partie inconnue du triangle.

Lemme 6. *Soient $1 \leq i \leq n$ et $n-i+1 \leq j \leq n$, on a :*

$$\mathbb{E}(C_{i,j} \mid \mathcal{F}_n) = \left(\prod_{k=n+1-i}^{j-1} (1 - \delta_k) \right) C_{i,n+1-i} + E_i \sum_{k=n+2-i}^j \lambda_k \left(\prod_{\ell=k}^{j-1} (1 - \delta_\ell) \right).$$

Il est nécessaire de remplacer les λ_j et δ_j par leurs estimateurs. Le résultat du lemme suivant implique une absence de corrélation entre les estimateurs impliqués, ce qui permet d'établir un estimateur sans biais de l'estimation des provisions.

Lemme 7. *Soient $1 \leq k \leq n-1$ et $1 \leq j_1 < j_2 < \dots < j_k$, on a :*

$$\mathbb{E} \left(\prod_{i=1}^k (1 - \widehat{\delta}_{j_i}) \mid \mathcal{B}_{j_1} \right) = \prod_{i=1}^k \mathbb{E}(1 - \widehat{\delta}_{j_i}) = \prod_{i=1}^k (1 - \delta_{j_i}).$$

De plus, on a :

$$\mathbb{E} \left(\widehat{\lambda}_{j_1} \prod_{i=1}^k (1 - \widehat{\delta}_{j_i}) \mid \mathcal{B}_{j_1-1} \right) = \mathbb{E}(\widehat{\lambda}_{j_1}) \prod_{i=1}^k \mathbb{E}(1 - \widehat{\delta}_{j_i}) = \lambda_{j_1} \prod_{i=1}^k (1 - \delta_{j_i}).$$

Proposition 6. *Soient $1 \leq i \leq n$ et $n - i + 1 \leq j \leq n$. L'estimateur*

$$\widehat{C}_{i,j} := \left(\prod_{k=n+1-i}^{j-1} (1 - \widehat{\delta}_k) \right) C_{i,n+1-i} + E_i \sum_{k=n+2-i}^j \widehat{\lambda}_k \left(\prod_{\ell=k}^{j-1} (1 - \widehat{\delta}_\ell) \right)$$

est un estimateur sans biais de $C_{i,j}$ conditionnellement à $\mathcal{B}_{n+1-i}$.

Annexe D

Le critère d'information d'Akaike (AIC)

Le critère d'information d'Akaike – AIC pour *Akaike Information Criterion* en anglais – est une méthode mathématique permettant de mesurer la qualité d'ajustement d'un modèle statistique prédictif aux données. Il permet ainsi de comparer différents modèles et de déterminer celui qui s'ajuste le mieux aux données tout en pénalisant le nombre de variables indépendantes. Il s'applique plus particulièrement aux modèles estimés par une méthode de maximum de vraisemblance. Il est défini par la formule suivante :

$$AIC = -2 \ln \tilde{L} + 2k$$

Avec :

- $\tilde{L}$: l'estimation de la vraisemblance maximale du modèle ;
- k : le nombre de paramètres, c'est-à-dire le nombre de variables indépendantes, à estimer dans le modèle.

Avec ce critère, la déviance du modèle $-2 \ln \tilde{L}$ est pénalisée du double du nombre de paramètres k . L'AIC mesure donc un compromis entre deux objectifs opposés : celui de réduire le biais – ce qui passe par l'augmentation du nombre de paramètres – et celui de disposer d'un modèle simple, utilisant le moins de paramètres pour décrire les données. L'AIC est d'autant plus faible que la vraisemblance est élevée et que le nombre de paramètres est restreint.

Le modèle optimal au sens de l'AIC correspond à celui minimisant l'AIC.

Annexe E

La méthode de Kaplan-Meier²

L'estimateur construit grâce à la méthode de Kaplan-Meier – appelé estimateur de Kaplan-Meier – permet de réaliser une estimation statistique non-paramétrique de la distribution de durées.

Dans un cadre statistique « classique », lorsque le vecteur $(T_1, \dots, T_n)$ est observé, la fonction de répartition $F(t) = \mathbb{P}(T \leq t)$ peut être estimée de façon non paramétrique par la fonction de répartition empirique des données d'après la formule suivante :

$$\hat{F}_{\text{emp}}(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{T_i \leq t}.$$

Cependant, la fonction de répartition empirique des données n'est pas disponible dans le cadre de l'analyse de survie, notamment en raison de la présence de censure et de troncature. D'autres fonctions de répartitions peuvent toutefois être estimées telles que :

$$\begin{aligned}\hat{H}_1(t) &= \frac{1}{n} \sum_{i=1}^n \delta_i \mathbf{1}_{Y_i \leq t}, \\ \hat{H}_0(t) &= \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \mathbf{1}_{Y_i \leq t}, \\ \hat{H}(t) &= \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i \leq t} = \hat{H}_0(t) + \hat{H}_1(t).\end{aligned}$$

Considérons des données pouvant être uniquement aléatoirement censurées à droite, alors

2. ([Lopez, 2023](#))

les limites de ces estimateurs sont les suivantes :

$$H_1(t) = \mathbb{P}(Y \leq t, \delta = 1),$$

$$H_0(t) = \mathbb{P}(Y \leq t, \delta = 0),$$

$$H(t) = \mathbb{P}(Y \leq t).$$

Le but de la méthode de Kaplan-Meier est de chercher un estimateur de la fonction de répartition $\hat{F}$ en présence de données censurées, à partir de ces estimateurs sous-jacents. Bien que $\hat{F}_{\text{emp}}$ définisse une distribution discrète sur $\{T_1, \dots, T_n\}$, la distribution discrète est estimée sur $\{Y_1, \dots, Y_n\}$.

Soit μ , le taux de risque de T et $\hat{\mu}$, le taux de risque de $\hat{F}$.

Dans le cas d'une distribution discrète :

$$\hat{F}(t) = 1 - \prod_{Y_i \leq t} (1 - \hat{\mu}(Y_i)).$$

Les différentes étapes de la construction de l'estimateur de Kaplan-Meier sont les suivantes :

1. Exprimer μ en fonction de H_0 , H_1 et H .
2. Remplacer H_0 , H_1 et H par leurs versions empiriques afin d'obtenir $\hat{\mu}$.
3. Calculer $\hat{F}$ grâce à l'estimation de $\hat{\mu}$.

En temps discret, on a :

$$\mu(t) = \frac{\mathbb{P}(T = 1)}{1 - F(t^-)} = \frac{dF(t)}{1 - F(t^-)}.$$

Théorème 1. *Sous absence de troncature et sous l'hypothèse fondamentale que T soit indépendante de C , la variable de censure, on a :*

$$\frac{dF(t)}{1 - F(t^-)} = \frac{dH_1(t)}{1 - H(t^-)}.$$

Définissons :

$$\hat{\mu}(t) = \frac{d\hat{H}_1(t)}{1 - \hat{H}(t^-)} = \frac{\#\{i : Y_i = t, \delta_i = 1\}}{\#\{i : Y_i \geq t\}}.$$

L'estimateur de Kaplan-Meier est alors défini comme suit :

$$\hat{F}(t) = 1 - \prod_{i: Y_i \leq t} \left(1 - \frac{\delta_i}{\sum_{j=1}^n 1_{Y_i \leq Y_j}} \right).$$

Annexe F

La régression logistique

La régression logistique est un cas particulier de modèle linéaire généralisé. C'est une méthode statistique principalement utilisée en classification binaire. En effet, elle permet de prédire une variable binaire dépendante Y en fonction d'un ensemble de variables explicatives indépendantes $X = (x_1, \dots, x_n)$. La régression logistique est également utilisée pour trouver une relation mathématique entre Y et les différentes variables explicatives.

Y est la variable à prédire et prend deux valeurs possibles : 0 ou 1. Considérons que lorsqu'elle vaut 1, l'événement se réalise. Les X_i sont quant à elles les variables prédictives et peuvent être quantitatives ou qualitatives.

La probabilité que Y prenne la valeur 1 – $\mathbb{P}(Y = 1)$ – correspond à la probabilité que l'événement se réalise. Elle est donnée par la fonction de lien logistique suivante et est notée p :

$$p = \mathbb{P}(Y = 1) = \frac{1}{1 + \exp(-(b_0 + b_1x_1 + b_2x_2 + \dots + b_nX_n))}.$$

Le logit de p est exprimé comme une combinaison linéaire qui est définie de la manière suivante :

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1x_1 + \dots + \beta_kx_k.$$

Les coefficients $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ représentent l'effet des différentes variables x_i sur le *logit* de p . Ils sont habituellement estimés par la méthode du maximum de vraisemblance dont la fonction de vraisemblance est la suivante :

$$L(\beta) = \prod_{i=1}^n p(x_i)^{y_i} \cdot (1 - p(x_i))^{1-y_i}.$$

L'objectif est alors de maximiser $\ln L(\beta)$ afin d'obtenir les paramètres β qui maximisent la

fonction de vraisemblance. Ces paramètres correspondent alors aux estimateurs du maximum de vraisemblance de la régression logistique.

Par la suite, en fonction de la valeur de p fournie par la régression logistique et sur un seuil fixé en amont, il est alors possible de déterminer si l'événement associé à la variable Y s'est réalisé ou non.

Liste des sigles, acronymes et abréviations

AAD *Annual Aggregate Deductible* – Franchise annuelle agrégée

AAL *Annual Aggregate Limit* – Limite annuelle agrégée

ACC À l'acceptation d'AXA SA

ACPR Autorité de Contrôle Prudentiel et de Résolution

AIC *Akaike Information Criterion* – Critère d'Information d'Akaike

ASA À la charge d'AXA SA

CART *Classification And Regression Trees*

D/D Dossier/Dossier

EP Excédent de Plein

FGU *From Ground Up* – Terme définissant le montant total du sinistre, après franchise mais avant toute rétention ou réassurance

GAMLSS *Generalized Additive Model for Location, Scale and Shape*

IARD Incendies, Accidents et Risques Divers

IBNeR *Incurred But Not enough Reserved*

IBNR *Incurred But Not Reported*

IBNyR *Incurred But Not yet Reported*

k Milliers (d'euros)

Mds Milliards (d'euros)

M Millions (d'euros)

PSAP Provisions pour Sinistres À Payer

QP Quote-Part

SL *Stop-Loss*

XS Excédent de Sinistres