

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

Par : Joseph AIDDID

Titre du mémoire : Mise en place d'un prix technique de référence pour un pilotage prospectif : application à un portefeuille Auto Professionnels

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 :

Signature :

Entreprise :

Nom : Abeille Assurances

Signature :

Directeur de mémoire en
entreprise

Membres présents du jury de
l'Institut des Actuares :

Signature :

Nom : Sébastien HALTER

Signature : 

Invité :

Nom :

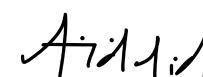
Signature :

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

Signature du responsable
entreprise :



Signature du candidat :





Remerciements

Je tiens à tout d'abord à exprimer mes remerciements et ma reconnaissance envers mes encadrants Sébastien Halter et Ngoc Long Pham pour m'avoir permis de réaliser ce mémoire, pour tout ce qu'ils m'ont appris, le temps qu'il m'ont consacré, ainsi que pour leurs nombreuses relectures et remarques avisées.

Un grand merci également à Mama Bathily pour sa constante disponibilité lorsque j'ai eu besoin de son aide, à Romain Laily pour les connaissances qu'il a pu m'apporter au début de l'élaboration de ce mémoire et à tous ceux qui ont pu y apporter leur contribution.

Je remercie également l'ensemble des membres de la direction Actuariat Pricing & Data Science d'Abeille Assurances pour leur accueil, leur bienveillance et les bons moments passés lors de mes deux stages puis de mon alternance.

Je n'oublie pas non plus Noémie Pey avec qui j'ai eu la chance de faire mon premier stage et grâce à qui j'ai pu intégrer Abeille Assurances pour la première fois.

Je souhaite également remercier Olivier Lopez, mon tuteur pédagogique, et exprimer ma reconnaissance envers l'ensemble des équipes de l'ISUP pour leur accompagnement et la qualité de leur enseignement.

Enfin je remercie mes proches pour leur soutien et leurs encouragements tout au long de la rédaction de ce mémoire.



Résumé

La direction Actuariat Pricing & Data Science d'Abeille Assurances souhaite se doter de nouveaux indicateurs de tarification et de pilotage, calculés à partir d'un prix technique de référence. Ce mémoire s'inscrit dans le cadre d'un projet pilote de mise en œuvre de ces nouveaux outils de tarification. Nous décrivons ainsi comment nous avons pu mettre en place ce prix technique de référence et les indicateurs qui en découlent sur un portefeuille automobile à destination des professionnels et des entreprises. Ce portefeuille présente un faible volume de données ainsi que certaines particularités comme la présence importante de conducteurs non désignés. Une grande partie de ce mémoire concerne le calcul d'une des principales composantes du prix technique de référence correspondant à une estimation de la sinistralité future calculée à l'ultime. Dans ce mémoire nous montrerons premièrement comment nous avons pu aboutir au calcul de ce prix technique de référence malgré les contraintes auxquelles nous avons été confrontés en raison des particularités de notre portefeuille atypique. Ensuite nous décrivons les possibles utilisations et avantages offerts par ces indicateurs ainsi que quelques cas pratiques d'application à notre portefeuille.

Mots-clés : Tarification, Pilotage, Prime Pure, GLM, Assurance automobile des professionnels.



Abstract

The Actuarial Pricing & Data Science department of Abeille Assurances wishes to acquire new pricing and monitoring indicators, calculated from a technical reference price. This dissertation is part of a pilot project to implement these new pricing tools. We describe how we were able to implement this technical reference price and the resulting indicators on a car portfolio for professionals and companies. This portfolio has a small volume of data and certain particularities such as the significant presence of non-designated drivers. A large part of this dissertation concerns the calculation of one of the main components of the technical reference price corresponding to an estimate of the future loss experience calculated at the end. In this dissertation we will first show how we were able to calculate this technical reference price despite the constraints we faced due to the particularities of our atypical portfolio. Then we will describe the possible uses and advantages offered by these indicators as well as some practical cases of application to our portfolio.

Keywords: Pricing, Piloting, Pure Premium, GLM, Professional Motor Insurance.



Synthèse

Contexte et objectifs

L'assurance de biens est un secteur très concurrentiel où il s'avère primordial pour une compagnie d'assurances de disposer d'une politique efficace de tarification, de suivi et de pilotage de ses résultats techniques. A ce titre la direction Actuariat Pricing & Data Science d'Abeille Assurances souhaite se doter de nouveaux indicateurs prospectifs de tarification et de pilotage, complémentaires, à destination de ses divers produits d'assurance. Ce mémoire s'inscrit ainsi dans le contexte d'un projet pilote de la mise en place de ces nouveaux indicateurs calculés à partir d'un prix technique de référence (PTR). Formellement, on définit cet outil de tarification de la façon suivante :

$$PTR = SUE + F_{ext} + F_{int} + C_{réa} + C_{cap} - B_{inv}.$$

Où SUE est la sinistralité ultime estimée correspondant à la meilleure estimation non actualisée et brute de réassurance des sinistres futurs calculée à l'ultime, F_{int} et F_{ext} respectivement les frais internes et externes de l'entreprise, $C_{réa}$ les coûts de réassurance, C_{cap} le coût du capital immobilisé pour un contrat et B_{inv} le retour sur investissement des capitaux provisionnés pour le règlement de sinistres.

Ce projet a été mis en œuvre dans le cadre d'un produit d'assurance automobile à destination des professionnels et des entreprises. A la différence d'un produit d'assurance automobile classique, ce produit destiné à des professionnels présente certaines particularités, comme la présence de conducteurs non désignés, et le portefeuille de contrats associé contient un relativement faible volume de données. L'objet de ce mémoire est à la fois de montrer comment nous avons pu mener à bien ce projet en tenant compte des contraintes liées à ce portefeuille atypique afin de calculer le PTR et les indicateurs qui en découlent pour les mettre en applications sur notre portefeuille.

Périmètre de la modélisation

Une grande partie de ce mémoire se concentre sur le calcul de la SUE dont la prime pure est la principale composante. On décompose cette prime pure en une prime pure des sinistres attritionnels estimée à l'aide de GLMs, une prime pure des sinistres graves pour laquelle plusieurs méthodes sont envisagées et une prime pure des sinistres liés aux événements naturels. La première étape que nous avons suivie a été de définir le périmètre sur lequel se restreint notre étude. Nous avons ainsi déterminé les garanties à modéliser et le seuil de sinistre grave à retenir lorsque cela s'avère pertinent. Ainsi nous avons choisi de modéliser les garanties Responsabilité Civile, Dommages Tous Accidents, Assistance et Bris de glace qui représentent 94% du nombre sinistres et 97% de leur charge entre 2016, date de lancement du produit, et 2021. Pour les garanties non modélisées, qui représentent une faible part de la sinistralité observée, on fait une approximation en supposant que le montant de la prime commerciale est égal au PTR. Ensuite après étude de la répartition des montants de sinistres et en adoptant une approche par seuils selon la théorie des valeurs extrêmes, il a été jugé pertinent de ne considérer que la garantie Responsabilité civile en ce qui concerne la sinistralité grave. Dans un premier temps, par manque de données et en l'absence de résultats suffisamment robustes un seuil de 76 000 €, correspondant au seuil déjà pratiqué en interne, a été retenu. Pour les sinistres climatiques qui ne représentent pas d'enjeu majeur pour ce produit, nous avons supposé que leur sinistralité était déjà prise en compte dans les primes des garanties concernées.

Base de données

Une fois que nous avons délimité le périmètre de notre étude, nous avons mis en place une base de données à destination de la modélisation des primes pures attritionnelles des garanties retenues. Cette base de données comporte des informations générales sur les assurés, les conducteurs, leur activité, leur véhicule et leur historique de sinistralité. Nous l'avons mis en place en deux temps avec d'une part une base de contrats et d'autre part une base de sinistres en vérifiant la cohérence des bases en matière de sinistralité et d'exposition par rapport au tableau de bord de suivi interne et en validant la conservation de l'information après jointure. Ensuite nous avons procédé à une présélection de variables en éliminant celles ayant trop de valeurs manquantes et pour lesquelles un retraitement n'était pas envisageable. Puis nous avons effectué des statistiques descriptives



pour mettre en avant les variables impactées et enfin nous avons observé les liens entre les variables à travers le calcul des V de Cramers et des corrélations linéaires. Enfin nous avons découpé notre base de données en une base d'apprentissage et une base test en nous assurant d'une répartition similaire des profils et de leur sinistralité dans chacune de ces bases.

Modélisation hors graves

A l'aide du logiciel EMBLEM développé par Towers Watson nous avons ensuite établi des modèles de fréquence et de sévérité attritionnelles par garantie. Par volonté d'obtenir une structure multiplicative de coefficients nous avons eu recours à une fonction de lien logarithmique et les lois utilisées pour les modèles de sévérité et de fréquence sont respectivement des distributions Gamma et de Poisson. Tout au long de la modélisation nous avons dû faire face à la contrainte d'un faible volume de données. Ainsi, lors de notre sélection de modèle effectuée selon une approche stepwise forward, selon le critère de l'AIC et du BIC, nous avons dû effectuer des regroupements de modalités dont nous nous sommes assurés de la pertinence lorsque c'était nécessaire. Nous avons ensuite validé nos modèles en nous attachant particulièrement à leur bon ajustement sur les profils représentant le cœur de notre portefeuille. Enfin on combine les modèles de fréquence et de sévérité pour obtenir des modèles de prime pure attritionnelle pour lesquels on procède à un lissage des coefficients. Le lissage consiste en l'ajustement des coefficients obtenus par rapport à la prime pure observée. En nous concentrant sur les variables les plus significatives, on ajuste ainsi les modalités les moins représentées et pour lesquelles le manque de population n'a pas pu conduire à une estimation robuste des coefficients.

Prise en compte des sinistres graves

Face au manque de données, plusieurs approches et méthodes ont été considérées pour prendre en compte les sinistres graves. Dans un premier temps nous avons tenté d'obtenir un modèle à l'aide de la théorie des valeurs extrêmes et de GLMs. Cette approche se révélant infructueuse, nous avons ensuite essayé d'obtenir des résultats à partir d'une autre base de données, issue d'un portefeuille flottes automobiles similaire et comparable à celui de notre étude. Enfin nous avons simplement expérimenté une approche consistant à répartir le poids la charge excédant le seuil de grave fixé à 76 000 € sur tous les assurés. C'est cette dernière approche que nous avons retenue car après comparaison de toutes les pistes explorées c'est celle qui s'est avérée la plus cohérente en ce qui concerne l'ordre de grandeur des résultats et la plus prudente au regard du manque de robustesse des estimations obtenues par les autres méthodes sous la contrainte d'un faible volume de données.

Passage à l'ultime

Notre modélisation s'étant établie à partir de sinistres clos et en cours, on les passe à l'ultime afin de tenir compte de leur vieillissement. Seules les garanties RC et DTA ont été passées à l'ultime. Nous avons ainsi considéré que les sinistres bris de glace et assistance, dont la grande majorité des sinistres sont clos l'année même de l'année de leur survenance, étaient déjà à l'ultime. Ainsi à l'aide de la méthode Chain Ladder nous avons calculé un passage à l'ultime des sinistres attritionnels RC et DTA. Quant aux sinistres graves, nous avons pris comme ultime leur valeur comptable tout en nous assurant de la prudence de cette approximation en utilisant la partie descriptive de la méthode Hachemeister.

Inflation et frais

Après le passage à l'ultime la dernière étape du calcul de la SUE a consisté en la prise en compte de l'inflation des coûts de sinistres entre le moment où la modélisation a été effectuée et le moment où elle est utilisée pour le calcul du PTR ainsi que des frais non allouables directement aux sinistres (comme les frais de gestion des sinistres). Nous avons ensuite détaillé les différents calculs d'inflation à prendre en compte selon les cas. En ce qui nous concerne, par manque de temps nous avons eu recours au calcul d'une inflation moyenne constatée toutes garanties confondues. Les frais sont quant à eux comptés avec les frais fixes.

Calcul du prix technique de référence

Après avoir calculé la SUE nous avons procédé au calcul général du PTR en procédant à quelques approximations. En effet, par manque de temps et dans une première approche avant d'avoir l'occasion de collaborer avec toutes les autres directions concernées par ces composantes, nous avons été contraints de modifier la formule additive



initiale en une formule un peu plus complexe. Ainsi, nous nous sommes appuyés sur un fichier général répertoriant tous les résultats du groupe de façon agrégée. Dans ce fichier les différents frais internes, externes et les FNAS sont regroupés et on ne distingue que frais fixes et frais variables. De plus toutes les composantes qu'on souhaite obtenir y sont exprimées en pourcentage des primes totales. Cela nous a conduit à la formule suivante après résolution de l'équation $PNA = PTR$.

$$PTR = \frac{SUE + \text{Frais}_{\text{fixes}}}{1 - (\% \text{Frais}_{\text{var}} + \% C_{\text{Rea}} + \% C_{\text{cap}} - \% B_{\text{inv}})}$$

Définition de nouveaux indicateurs

Avec le PTR dont nous avons un calcul approché, on définit ainsi de nouveaux indicateurs de suivi et de pilotage à commencer par le ratio d'équilibre (R_{eq}) :

$$R_{\text{eq}} = \frac{PNA}{PTR}$$

Il compare le prix technique de référence à la prime réellement émise, après éventuelles remises commerciales ou autre, et représente la rentabilité attendue estimée à partir du calcul du PTR. Un ratio d'équilibre égal à 100% signifie que le tarif proposé est en phase avec le prix technique de référence permettant d'atteindre les objectifs d'équilibre technique et financier de la compagnie. Au-dessus ou en dessous de 100% le portefeuille est soit rentable soit au contraire déficitaire. Contrairement à un ratio S/P classique qui ne donne qu'une vision rétrospective des résultats on a l'avantage d'avoir avec cet indicateur une vision prospective de la rentabilité d'un portefeuille. Ensuite pour mesurer l'écart monétaire entre primes payées, avant et après remises, par les clients et prix technique de référence ainsi que le pourcentage du PTR que cela représente on définit les indicateurs suivants :

- Avant remises :

$$RTE_{\text{avt}} = PC - PTR ; \% RTE_{\text{avt}} = \frac{(PC - PTR)}{PTR}$$

- Après remises :

$$RTE_{\text{ap}} = PNA - PTR ; \% RTE_{\text{ap}} = \frac{(PNA - PTR)}{PTR}$$

De la même façon d'autres indicateurs peuvent être définis à partir du PTR. Nous nous sommes néanmoins limités à l'usage du ratio d'équilibre et des résultats techniques estimés dans nos applications. Au contraire d'un ratio S/P classique qui ne décrit que la réalité passée et ne témoigne en rien de l'avenir, les indicateurs que nous avons construits à partir du PTR nous offrent une vision prospective de la rentabilité de notre portefeuille. On peut de plus facilement mettre à jour ou le décliner à d'autres produits si on dispose des composantes nécessaires. C'est de plus un outil facilement interprétable qui lie tous les acteurs liés à la politique tarifaire et permet d'analyser rapidement l'impact de mesures.

Surveillance de portefeuille

Enfin nous avons illustré quelques cas pratiques de l'utilisation du PTR et des indicateurs qui en découlent. Premièrement nous nous en sommes servi pour avoir une vue d'ensemble de la rentabilité attendue de notre portefeuille automobile à destination des professionnels et des entreprises. Cette vision d'ensemble nous a ainsi permis d'être alerté d'une rentabilité attendue dégradée du portefeuille. Pour identifier les populations responsables de ces dérives nous avons commencé par une vision du ratio d'équilibre par année de souscription. L'idée étant de déterminer si cette dérive est soudaine et ne concerne que la souscription récente de profils pour lesquels, dans l'ensemble, la prime payée n'est pas en adéquation avec le prix technique de référence ou si cette dérive concerne le portefeuille dans son ensemble. L'autre intérêt étant également d'observer la rentabilité attendue des affaires nouvelles qui représentent le portefeuille à venir. Ainsi nous avons pu mettre en lumière que cette dérive était causée par les affaires récentes. Nous avons donc ensuite calculé le ratio d'équilibre par variable en nous concentrant sur les parlantes d'un point de vue métier. Cela nous a permis d'identifier les profils responsables de ces dérives mais également de détecter des profils a priori sur-tarifés et qui peuvent donc représenter une niche d'opportunité. Après avoir distingué ces profils nous avons ensuite observé la distribution



de notre portefeuille en matière de résultats techniques estimés. A l'aide de cette représentation et du calcul du prix technique de référence et des résultats techniques estimés. Nous avons ainsi illustré comment on pouvait mesurer l'impact de mesures tarifaires simples sur la rentabilité du portefeuille comme la résiliation des profils les plus dégradés en matière de résultats techniques estimés. Nous avons raisonné toutes garanties confondues mais les mêmes analyses peuvent être menées par garantie. Ainsi, nous avons pu davantage préciser nos analyses de la performance du portefeuille en exhibant les garanties déficitaires et celles au contraire surtarifiées.

Affaires nouvelles

Nous avons ensuite expliqué comment nous pouvons faire évoluer progressivement le tarif vers le prix technique de référence en portant notamment une attention particulière aux affaires nouvelles qui représentent le futur du portefeuille. En pratiquant un tarif trop volatile par rapport au PTR on favoriserait l'antisélection. Or grâce au PTR nous pouvons détecter d'éventuels écarts de rentabilité entre le portefeuille et les affaires nouvelles ainsi que surveiller l'évolution de la diversité des profils de notre portefeuille et leur rentabilité associée.

Pilotage des mesures tarifaires

Après cela nous avons décrit comment le PTR pouvait nous permettre un meilleur suivi des mesures tarifaires ou offres commerciales en anticipant et en analysant immédiatement l'impact de ces mesures par leur répercussion sur le ratio d'équilibre. On mesurerait ensuite l'impact financier en termes de résultats techniques estimés.

Vision client

A cette occasion nous avons introduit la possibilité d'un pilotage de la rentabilité de la compagnie qui se ferait au niveau client et non pas uniquement au niveau contrat. En effet un assuré ayant souscrit plusieurs contrats pourrait en avoir certains qui ne seraient pas rentables au regard du PTR alors que dans leur globalité, les contrats par compensation seraient dans l'ensemble rentable. Avec un PTR sur chaque produit on pourrait donc envisager une autre méthode de pilotage qui s'intéresserait à la rentabilité d'ensemble de la compagnie et non pas uniquement par produit.

Aide à la souscription

Ensuite nous avons évoqué la possibilité d'utiliser le PTR pour la création d'un outil d'aide à la souscription qui indiquerait au souscripteur le PTR par garantie ainsi que le ratio d'équilibre associé en fonction des caractéristiques d'un prospect. Le souscripteur pourrait de plus avoir une visibilité des leviers tarifaires qu'il pourrait appliquer et serait ainsi directement informé de la rentabilité attendue des prospects et donc davantage impliqué dans le processus d'amélioration de la performance du portefeuille.

Elasticité au prix

Nous avons ensuite introduit la notion d'élasticité au prix qui mesure l'impact d'une variation du tarif sur le comportement d'un client. Cet indicateur nous permet ainsi de connaître l'impact d'une évolution de tarif sur les taux de souscription de rétention. Cet indicateur seul ne nous informe pas de la variation exacte à appliquer pour maximiser le profit de l'assureur. A l'aide du PTR on peut cependant en avoir une estimation.

Perspectives d'amélioration

Enfin nous avons évoqué les perspectives d'amélioration du PTR dont l'une des principales se situe au niveau du calcul de ses composantes comme les différents frais, les coûts de réassurance et les indicateurs financiers pour lesquels nous avons dû faire des approximations. Au niveau du calcul de la SUE sur laquelle nous nous sommes concentrés, on peut également penser à une meilleure prise en compte des sinistres graves ainsi que du risque lié aux événements naturels.



Conclusion

A travers ce mémoire nous avons ainsi pu premièrement montrer comment nous avons su faire face à toutes les contraintes liées au faible volume de données à notre disposition et les particularités de notre portefeuille pour obtenir un calcul approché du prix technique de référence tout en restant le plus cohérent et prudent possible. Enfin nous avons eu l'occasion de décrire les possibilités d'utilisation du prix technique de références et des indicateurs dérivés à l'aide notamment de cas pratiques d'application à notre portefeuille automobile à destination des professionnels et des entreprises. Nous nous sommes ainsi attachés à montrer en quoi le PTR constitue un sujet central en tarification.



Synthesis

Context and objectives

Property insurance is a highly competitive sector in which it is essential for an insurance company to have an effective pricing, monitoring and management policy for its technical results. In this respect, the Actuarial Pricing & Data Science department of Abeille Assurances wishes to acquire new, complementary, prospective pricing and monitoring indicators for its various insurance products. This thesis is therefore part of a pilot project to implement these new indicators calculated from a technical reference price (PTR). Formally, this pricing tool is defined as follows:

$$PTR = SUE + F_{ext} + F_{int} + C_{réa} + C_{cap} - B_{inv}.$$

Where SUE is the estimated ultimate loss experience corresponding to the best undiscounted and gross reinsurance estimate of future claims calculated at the ultimate, F_{int} and F_{ext} are the internal and external expenses of the company respectively, C_{rea} is the reinsurance costs, C_{cap} is the cost of capital tied up in a contract and B_{inv} is the return on investment of the capital set aside for claims settlement.

This project was implemented in the context of a motor insurance product for professionals and businesses. In contrast to a conventional motor insurance product, this product for professionals has certain peculiarities, such as the presence of unnamed drivers, and the associated portfolio of contracts contains a relatively small volume of data. The purpose of this thesis is to show how we were able to carry out this project, taking into account the constraints linked to this atypical portfolio, in order to calculate the PTR and the resulting indicators and apply them to our portfolio.

Scope of the modelling

A large part of this thesis focuses on the calculation of the SUE of which the pure premium is the main component. This pure premium is decomposed into a pure premium for attritional claims estimated using GLMs, a pure premium for severe claims for which several methods are considered, and a pure premium for claims related to natural events. The first step we took was to define the scope of our study. We thus determined the coverages to be modelled and the severe loss threshold to be retained when relevant. We therefore chose to model the Civil Liability, All Accidents, Assistance and Glass Breakdown coverages, which represent 94% of the number of claims and 97% of their cost between 2016, when the product was launched, and 2021. For the non-modelled coverages, which represent a small proportion of the observed claims experience, an approximation is made by assuming that the amount of the commercial premium is equal to the PTR. Then, after studying the distribution of claim amounts and adopting a threshold approach according to the extreme value theory, it was considered relevant to consider only the third party liability cover with regard to the serious claims experience. Initially, due to a lack of data and in the absence of sufficiently robust results, a threshold of €76,000, corresponding to the threshold already used internally, was adopted. For weather-related claims, which do not represent a major issue for this product, we assumed that their loss experience was already taken into account in the premiums for the cover concerned.

Database

Once we had defined the scope of our study, we set up a database for modelling the pure attritional premiums of the selected coverages. This database contains general information on the insured, the drivers, their activity, their vehicle and their claims history. We set it up in two stages, with a contract database and a claims database, checking the consistency of the databases in terms of claims experience and exposure in relation to the internal monitoring dashboard, and validating the retention of information after joining. We then proceeded to a pre-selection of variables by eliminating those with too many missing values and for which a reprocessing was not possible. Then we performed descriptive statistics to highlight the impacted variables and finally we observed the links between the variables through the calculation of Cramers' Vs and linear correlations. Finally, we divided our database into a learning database and a test database, ensuring that the profiles and their claims experience were similarly distributed in each of these databases.



Pure premium modelling

Using the EMBLEM software developed by Towers Watson, we then established models of attritional frequency and severity by cover. In order to obtain a multiplicative structure of coefficients we used a logarithmic link function and the distributions used for the severity and frequency models are Gamma and Poisson distributions respectively. Throughout the modelling process we were constrained by the small volume of data. Thus, during our model selection using a stepwise forward approach, according to the AIC and BIC criterion, we had to carry out groupings of modalities which we ensured were relevant when necessary. We then validated our models, paying particular attention to their good fit on the profiles representing the core of our portfolio. Finally, we combine the frequency and severity models to obtain pure attritional premium models for which we proceed to a smoothing of the coefficients. Smoothing consists of adjusting the coefficients obtained with respect to the observed pure premium. By concentrating on the most significant variables, we adjust the least represented modalities and for which the lack of population could not lead to a robust estimation of the coefficients.

Taking account of serious claims

In view of the lack of data, several approaches and methods were considered to take into account serious claims. First, we tried to obtain a model using extreme value theory and GLMs. This approach proved unsuccessful, so we then tried to obtain results from another database, from a similar and comparable car fleet portfolio to the one in our study. Finally, we simply experimented with an approach consisting of distributing the weight of the load exceeding the serious threshold set at €76,000 over all the insureds. We chose this last approach because, after comparing all the avenues explored, it proved to be the most consistent in terms of the order of magnitude of the results and the most prudent in view of the lack of robustness of the estimates obtained by the other methods under the constraint of a small volume of data.

Moving to the ultimate

As our modelling is based on closed and ongoing claims, we run them to the ultimate in order to take account of their ageing. Only the liability and DTA coverages have been run to the ultimate. We have thus considered that glass breakage and assistance claims, the vast majority of which are closed in the year in which they occur, were already in the ultimate. Thus, using the Chain Ladder method, we have calculated a transition to the ultimate for attritional liability and DTA claims. For serious claims, we have taken their book value as the ultimate, while ensuring the prudence of this approximation by using the descriptive part of the Hachemeister method.

Inflation and costs

After the final step in the calculation of the SUE was to take into account inflation in claims costs between the time the model was run and the time it is used for the calculation of the PTR, as well as costs that cannot be allocated directly to claims (such as claims handling costs). We have then detailed the different inflation calculations to be taken into account in each case. In our case, due to lack of time, we have used an average inflation calculation for all guarantees. The costs are counted with the fixed costs.

Calculation of the technical reference price

After calculating the SUE, we proceeded to the general calculation of the PTR, making some approximations. Indeed, due to lack of time and as a first approach before having the opportunity to collaborate with all the other directorates concerned by these components, we were forced to modify the initial additive formula into a slightly more complex one. Thus, we relied on a general file listing all the results of the group in an aggregated way. In this file, the various internal and external costs and the FNAS are grouped together and only fixed and variable costs are distinguished. In addition, all the components we wish to obtain are expressed as a percentage of total premiums. This leads to the following formula after solving the equation $PNA = PTR$:



$$PTR = \frac{SUE + \text{Frais}_{\text{fixes}}}{1 - (\% \text{Frais}_{\text{var}} + \% C_{\text{Rea}} + \% C_{\text{cap}} - \% B_{\text{inv}})}$$

Definition of new indicators

With the PTR, of which we have an approximate calculation, we can define new monitoring and steering indicators, starting with the balance ratio (R_eq):

$$R_{\text{eq}} = \frac{PNA}{PTR}$$

It compares the technical reference price with the premium actually issued, after any commercial or other discounts, and represents the expected profitability estimated from the calculation of the PTR. A breakeven ratio equal to 100% means that the proposed tariff is in line with the technical reference price allowing the company to reach its technical and financial breakeven objectives. Above or below 100% the portfolio is either profitable or in deficit. Contrary to a classic S/P ratio which only gives a retrospective view of the results, this indicator has the advantage of giving a prospective view of the profitability of a portfolio. Then, to measure the monetary difference between premiums paid, before and after discounts, by clients and the technical reference price, as well as the percentage of the PTR that this represents, the following indicators are defined:

- Before discounts:

$$RTE_{\text{avt}} = PC - PTR ; \% RTE_{\text{avt}} = \frac{(PC - PTR)}{PTR}$$

- After discounts:

$$RTE_{\text{ap}} = PNA - PTR ; \% RTE_{\text{ap}} = \frac{(PNA - PTR)}{PTR}$$

In the same way, other indicators can be defined from the PTR. However, we have limited ourselves to using the balance ratio and the estimated technical results in our applications. In contrast to a classical S/P ratio, which only describes the past reality and does not reflect the future, the indicators we have constructed from the PTR offer us a prospective view of the profitability of our portfolio. Moreover, it can be easily updated or applied to other products if we have the necessary components. It is also an easily interpretable tool that links all the actors involved in the tariff policy and allows the impact of measures to be analysed quickly.

Portfolio monitoring

Finally, we have illustrated some practical cases of the use of the PTR and the indicators derived from it. Firstly, we used it to obtain an overview of the expected profitability of our automobile portfolio for professionals and companies. This overview enabled us to be alerted to a deterioration in the expected profitability of the portfolio. In order to identify the populations responsible for these deviations, we began by looking at the breakeven ratio by underwriting year. The idea is to determine whether this drift is sudden and only concerns the recent underwriting of profiles for which, overall, the premium paid is not in line with the technical reference price or whether this drift concerns the portfolio as a whole. The other interest is also to observe the expected profitability of new business that represents the future portfolio. In this way we were able to highlight that this drift was caused by recent business. We then calculated the breakeven ratio by variable, focusing on the ones that speak for themselves from a business perspective. This allowed us to identify the profiles responsible for these drifts but also to detect profiles that are a priori overpriced and which may therefore represent a niche opportunity. After distinguishing these profiles we then observed the distribution of our portfolio in terms of estimated technical results. Using this representation and the calculation of the technical reference price and estimated technical results. We have thus illustrated how the impact of simple pricing measures on the profitability of the portfolio can be measured, such as the termination of the most degraded profiles in terms of estimated technical results. We have reasoned across all coverages, but the same analyses can be conducted by



coverage. In this way, we have been able to further refine our analyses of portfolio performance by showing which coverages are loss-making and which are, on the contrary, over-priced.

New business

We then explained how we could gradually move the tariff towards the technical reference price, paying particular attention to new business which represents the future of the portfolio. A tariff that is too volatile in relation to the RRP would encourage anti-selection. However, thanks to the PTR, we can detect possible profitability gaps between the portfolio and new business and monitor the evolution of the diversity of profiles in our portfolio and their associated profitability.

Monitoring of pricing measures

After that we described how the PTR could allow us to better monitor tariff measures or commercial offers by anticipating and immediately analysing the impact of these measures through their repercussion on the balance ratio. The financial impact would then be measured in terms of estimated technical results.

Customer vision

On this occasion we introduced the possibility of steering the company's profitability at the client level and not only at the contract level. Indeed, a policyholder who has taken out several contracts could have some that would not be profitable in terms of the PTR, whereas the contracts as a whole would be profitable. With a PTR for each product, we could therefore envisage another method of management which would be interested in the overall profitability of the company and not just by product.

Underwriting assistance

Next, we discussed the possibility of using the PTR to create an underwriting tool that would show the underwriter the PTR per cover and the associated balance ratio according to the characteristics of a prospect. The underwriter could also have visibility of the pricing levers that could be applied and would thus be directly informed of the expected profitability of the prospects and thus more involved in the process of improving the performance of the portfolio.

Price elasticity

We then introduced the notion of price elasticity, which measures the impact of a variation in price on a customer's behaviour. This indicator thus allows us to know the impact of a tariff change on retention subscription rates. This indicator alone does not inform us of the exact variation to be applied to maximise the insurer's profit. With the help of the PTR, however, we can estimate this.

Prospects for improvement

Finally, we mentioned the prospects for improving the PTR, one of the main ones being the calculation of its components such as the various expenses, reinsurance costs and financial indicators for which we had to make approximations. In terms of the calculation of the SUE, on which we have focused, we can also think of a better consideration of severe claims as well as the risk linked to natural events.

Conclusion

Through this thesis, we were able to show firstly how we were able to deal with all the constraints linked to the small volume of data at our disposal and the particularities of our portfolio in order to obtain an approximate calculation of the technical reference price while remaining as consistent and prudent as possible. Finally, we had the opportunity to describe the possibilities of using the technical reference price and the derived indicators with the help of practical cases of application to our automobile portfolio for professionals and companies. In this way, we have endeavoured to show how the TRP is a central topic in pricing.



Table des matières

Remerciements	2
Résumé	3
Abstract	4
Synthèse	5
Synthesis	10
Table des matières	14
Introduction.....	17
Périmètre de l'étude	17
1. Présentation de l'entreprise.....	17
2. Présentation du produit	18
3. Garanties	18
3.1. Responsabilité civile	18
3.2. Dommages tous accidents.....	18
3.3. Bris de glace.....	18
3.4. Assistance	18
3.5. Autres garanties	19
Calcul de la Sinistralité Ultime Estimée	20
1. Définitions	20
1.1. Définitions générales.....	20
1.2. Modèle linéaire généralisé	22
1.3. AIC et BIC	23
2. Périmètre de la modélisation	23
2.1. Contrats	23
2.2. Choix des garanties à modéliser.....	24
2.3. Sinistres graves	24
2.3.1. Cadre théorique.....	24
2.3.1.1. Introduction à la théorie des valeurs extrêmes	24
2.3.1.2. Méthodes du seuil	25
2.3.2. Détermination d'un seuil de grave par garantie	26
2.4. Sinistres climatiques.....	28
3. Base de modélisation	28
3.1. Mise en place de la base	28
3.2. Validation.....	29
3.3. Sélection des variables de modélisation	30



3.4.	Valeurs manquantes.....	30
3.5.	Conducteurs non désignés	31
3.6.	Coefficient de réduction majoration	32
3.7.	Analyses univariées et multivariées	33
3.8.	Corrélations linéaires et V de Cramer	35
4.	Modélisation de la prime pure	37
4.1.	Découpage en bases d'apprentissage et de test.....	37
4.2.	Test d'adéquation à une loi.....	38
4.2.1.	Modèles de fréquence.....	38
4.2.2.	Modèles de sévérité	39
5.	Sélection de modèle.....	40
5.1.	Procédure stepwise	40
5.1.1.	Critères AIC et BIC.....	40
5.1.2.	Regroupements de modalités et tests de sensibilité	42
5.2.	Critère d'équité du tarif.....	43
5.3.	Choix entre plusieurs modèles et calcul de RMSE.....	43
5.4.	Robustesse dans le temps	44
5.5.	Validation.....	45
5.5.1.	Analyse des résidus	45
5.5.2.	Validation sur la base test	47
5.6.	Combinaison de modèles et lissage des coefficients	47
6.	Choix de la méthode de prise en compte des sinistres graves	49
6.1.	Première approche.....	50
6.2.	Deuxième approche	52
6.3.	Troisième approche.....	54
7.	Passage à l'ultime	55
7.1.	Sinistres attritionnels.....	55
7.2.	Sinistres graves.....	59
8.	Inflation et frais	60
	Prix technique de référence	62
1.	Calcul du prix technique de référence	62
2.	Ratio d'équilibre	63
3.	Résultats techniques estimés.....	64
4.	Autres indicateurs	64
5.	Surveillance de portefeuille.....	65
5.1.	Suivi par génération de souscription.....	66



5.2.	Surveillance par critères tarifaires.....	66
5.3.	Exemple d'utilisation des résultats techniques estimés	68
5.4.	Simulations	69
5.5.	Surveillance par garantie	69
6.	Affaires nouvelles	70
7.	Pilotage des leviers tarifaires	71
8.	Vision client	72
9.	Aide à la souscription	72
10.	Elasticité au prix.....	72
11.	Perspectives d'améliorations	73
	Conclusion	74
	Bibliographie.....	75
	Annexes	76
	Liste des abréviations	85
	Table des figures.....	86
	Liste des tableaux	87
	Liste des annexes.....	88



Introduction

L'assurance de biens est un secteur très concurrentiel où il s'avère primordial pour une compagnie d'assurances de disposer d'une politique efficace de tarification, de suivi et de pilotage de ses résultats techniques et financiers. A ce titre la direction Actuariat Pricing & Data Science d'Abeille Assurances souhaite se doter, en complément des outils dont elle dispose déjà, de nouveaux indicateurs prospectifs construits à partir d'un prix technique de référence. Le prix technique de référence (PTR) représente la meilleure estimation du montant de prime requis par contrat pour atteindre les objectifs d'équilibre technique et financier qu'une compagnie d'assurance exige. Ce prix technique est une combinaison entre les charges de sinistres projetées, les frais internes et externes de la compagnie ainsi que du coût du capital incluant la marge de profit générée par le placement sur le marché financier des provisions qu'elle doit constituer pour faire face à ses obligations. Formellement on définit cet outil de tarification de la façon suivante :

$$PTR = SUE + F_{ext} + F_{int} + C_{réa} + C_{cap} - B_{inv}$$

Où SUE est la sinistralité ultime estimée correspondant à la meilleure estimation non actualisée et brute de réassurance des sinistres futurs calculée à l'ultime, F_{int} et F_{ext} respectivement les frais internes et externes de l'entreprise, $C_{réa}$ les coûts de réassurance, C_{cap} le coût du capital immobilisé pour un contrat et B_{inv} (bénéfices des investissements) le retour sur investissement des capitaux provisionnés pour le règlement de sinistres. Une grande partie de notre mémoire sera consacrée au calcul de la SUE. Ce calcul implique, pour des raisons d'homogénéité, les modélisations séparées de la sinistralité attritionnelle, de la sinistralité des graves et de la sinistralité due aux événements climatiques lorsqu'il y a lieu de le faire. Elle implique également la prise en compte des frais de gestion liés directement aux sinistres (frais d'expertise) ainsi que les frais de gestion non allouables directement aux sinistres (frais de gestion de sinistres). Nous détaillerons ensuite comment nous avons pris en compte les autres composantes dans le calcul du PTR. Dans le cas idéal d'un produit où toutes les garanties sont modélisées, le PTR est égal à la prime commerciale avant éventuels rabais, leviers commerciaux ou autre action sur le tarif. Malheureusement, lorsque les circonstances l'imposent, comme l'absence d'un volume de données suffisant, il peut s'avérer que le tarif d'un produit ne puisse pas être totalement modélisé. En pratique, la modélisation se fait par garantie de manière indépendante. Lorsque qu'il y a des garanties optionnelles dans un contrat, il est fréquent qu'elles ne disposent pas d'un volume de données suffisant pour faire l'objet d'une modélisation. Dans ce cas, pour ces « petites garanties » qui ne représentent qu'une faible portion de la sinistralité globale du contrat, il est assez raisonnable de considérer que le PTR de ces garanties est le juste prix du dire d'experts, c'est-à-dire le prix vendu. Cette situation est courante pour les petits portefeuilles et dans ces conditions, l'idée de disposer d'un prix technique de référence, au moins approché et que l'on peut comparer à la prime réellement payée prend tout son sens. C'est du moins ce que nous tenterons de mettre en évidence à travers notre étude et ses applications. Aussi, l'enjeu de ce mémoire est à la fois de montrer tout l'intérêt du PTR et des indicateurs qui en découlent, mais aussi de montrer comment nous avons pu le mettre en œuvre sur un petit portefeuille de contrats automobile à destination des professionnels malgré les contraintes liées à ses particularités et le relativement faible volume de données qui lui est associé. Dans un premier temps nous présenterons le périmètre de notre étude puis nous décrirons notre calcul de la SUE, à travers notamment une modélisation de prime pure et enfin nous détaillerons le calcul du PTR et ses applications possibles.

Périmètre de l'étude

1. Présentation de l'entreprise

Abeille assurances est une compagnie d'assurances issue du rachat d'Aviva France par le groupe mutualiste Aéma créé au mois de janvier 2021 par le rapprochement entre la Macif et Aésio mutuelle. Avec plus de 11 millions d'assurés, Aéma se classe 5^e parmi les groupes d'assurance en France. Au sein du groupe, Abeille assurances compte plus de 3 millions de clients et couvre leurs besoins à travers des produits d'épargne et de retraite, de protection et d'assurance à destination des particuliers, des professionnels et des entreprises.



2. Présentation du produit

Parmi les produits proposés par la compagnie, celui qui fait l'objet de notre étude est le produit d'assurance des véhicules individuels des entreprises. Ce produit s'adresse aux entreprises voulant assurer individuellement jusqu'à quatre véhicules motorisés de première catégorie¹ dans le cadre de leur activité. Il exclue les flottes pour lesquelles il existe un produit spécifique. Cette assurance est destinée aux professionnels dont le transport par véhicule est l'activité principale comme les taxis, les VTC, les entreprises de transport public de marchandises, les ambulances ainsi que les entreprises de transport sanitaire ou funéraire. Une des particularités de ce produit en comparaison à un produit d'assurance automobile classique est que les conducteurs peuvent être désigné ou non. Dans ce cas on ne dispose pas d'informations sur les conducteurs, comme l'âge ou l'ancienneté du permis. De plus, les contrats issus de ce produit forment un petit portefeuille de moins de 20 000 années polices² entre 2016, année de lancement de la version actuelle du produit, et 2021. Ce petit portefeuille n'a pour l'instant jamais fait l'objet d'une modélisation spécifique, celle que nous allons présenter est ainsi la première à être mise en œuvre sur ce périmètre.

3. Garanties

Ce produit couvre les assurés contre les dommages corporels, matériels et immatériels à travers de nombreuses garanties. Par anticipation avec la suite de notre étude et ne pour ne pas alourdir notre exposé, nous ne présenterons ici que les plus représentées en matière d'années polices et de sinistralité depuis 2016.

3.1. Responsabilité civile

La garantie responsabilité civile (RC) est la principale garantie de ce produit, elle est obligatoire et c'est celle qui est la plus représentée en matière de primes et de sinistralité dans notre portefeuille. Elle couvre les conséquences pécuniaires de dommages subis par des tiers et engageant la responsabilité civile de l'assuré. Le plafond d'indemnisation de la responsabilité civile matérielle est de 100 000 000 €, il n'y a par contre pas de plafond d'indemnisation pour la responsabilité civile corporelle. Cette garantie expose ainsi l'assureur au risque de devoir régler des montants de sinistres exceptionnels en particulier dans le cas de dommages corporels lourds causés à des tiers.

3.2. Dommages tous accidents

La garantie dommages tous accidents indemnise les dommages accidentels subis par le véhicule assuré résultant d'un choc, d'un acte de vandalisme, du renversement, du retournement, de la chute du véhicule ou de son immersion. Le risque climatique est aussi pris en compte dans cette garantie. En effet, elle s'applique également aux dommages causés par certains événements naturels lorsqu'ils n'ont pas été déclarés « catastrophe naturelle par les pouvoirs publics » : inondation, trombe, tornade, avalanche, éboulement, glissement ou affaissement de terrain, action de la neige.

3.3. Bris de glace

La garantie bris de glace est elle aussi une garantie de dommages aux véhicules, elle intervient dans le cas de sinistres accidentels ou fortuits concernant le pare-brise, la lunette arrière, les glaces latérales, le toit ouvrant ou le pare-soleil extérieur ainsi que l'ensemble des éléments d'éclairage du véhicule.

3.4. Assistance

L'assistance est une garantie singulière. En effet, à la différence des autres garanties que nous avons présentées, la prestation qui peut être délivrée est un service et non pas directement une indemnisation financière. Cette

¹ Les véhicules de première catégorie sont les véhicules à moteur dont le PTAC (Poids total autorisé en charge) est inférieur ou égal à 3,5T.

² Le nombre d'années polices d'un portefeuille est le nombre de contrats pondéré par leur temps, mesuré en année, passé en portefeuille. Par exemple 12 contrats n'ayant passé qu'un mois en portefeuille correspondent à une année police.



assurance propose divers services de dépannage, de remorquage, de rapatriement, de frais de taxi et d'hébergement, etc., aux assurés et à leur véhicule en cas de sinistre tel qu'une panne, un accident ou un vol. Pour évaluer un montant de sinistre on somme ainsi l'ensemble des montants de factures liées aux prestations délivrées.

3.5. Autres garanties

Un certain nombre d'autres garanties classiques peuvent être souscrites comme la garantie du conducteur, le vol, l'incendie³, la garantie dommages accidents par collision, la protection juridique, la défense pénale et recours suite à un accident (DPRSA), etc. Des options peuvent également venir enrichir les garanties de bases comme la garantie des appareils embarqués, celle des aménagements, du matériel et de l'outillage transportés. Pour toutes ces garanties nous avons fait une approximation en considérant que le prix technique de référence était égal au prix commercial pratiqué avant éventuel rabais ou geste commercial. Nous avons en effet préféré nous concentrer sur les principales garanties en matière d'exposition et de sinistralité et pour lesquelles une modélisation était envisageable.

³ La garantie incendie protège également, comme la garantie DTA, de certains risques climatiques, comme les tempêtes, les ouragans, les cyclones, la foudre, l'action du vent ou la grêle. Si cette garantie n'est pas souscrite la garantie DTA peut aussi protéger contre la grêle.



Calcul de la Sinistralité Ultime Estimée

1. Définitions

1.1. Définitions générales

Plusieurs notions utilisées dans ce mémoire nécessitent d'être explicitement définies afin d'éviter toute ambiguïté.

▪ Montant de sinistre

Le coût d'un sinistre pour l'assureur s'exprime selon cette formule :

$$\text{Montant} = \text{Règlements} + \text{Réserves} - \text{Recours}$$

Le montant d'un sinistre n'est pas définitif et varie selon la date d'observation. Il s'obtient à partir du cumul des règlements effectués à l'assuré, du montant de réserves provisionnées par l'assureur en prévision de ces règlements et enfin des éventuels recours que peut obtenir l'assureur lorsque la responsabilité d'un tiers est engagée dans le sinistre. Dans les règlements, sont également comptabilisés les honoraires d'experts éventuellement engagés dans le cadre de la procédure engagée suite à un sinistre. On parlera indifféremment de charge, de coût ou de montant de sinistre.

▪ Exposition

On définit l'exposition comme le temps passé pour un contrat dans le portefeuille, c'est-à-dire la durée durant laquelle l'assuré est couvert par ses garanties. On calcule l'exposition par exercice et on l'exprime en année⁴ :

$$\text{Exposition} = \frac{\text{Nombre de jours de présence en portefeuille}}{365,25}$$

Lorsqu'un contrat est présent toute l'année en portefeuille, son exposition est d'un an mais dans le cas d'une affaire nouvelle ou d'une résiliation en cours d'année l'exposition peut être plus faible. Aussi pour évaluer le volume et l'importance du portefeuille on préférera mesurer le nombre d'années polices, c'est-à-dire le nombre de contrats pondéré par leur exposition. En effet, un portefeuille peut avoir compté plus 100 000 contrats différents sur une année mais qui ne représentent en réalité que 60 000 années polices car il y a eu beaucoup d'affaires nouvelles en fin d'année ou des résiliations, etc.

▪ Primes émises et primes acquises

Les primes émises, $P_{\text{émise}}$, correspondent sont les primes attendues par l'assureur lors d'un exercice. Les primes sont souvent payées périodiquement et au moment où on observe le portefeuille le total des primes n'a pas encore été acquis par l'assureur. On s'intéresse ainsi aux primes réellement acquises, P_{acq} et on approxime leur montant en multipliant les primes émises par l'exposition correspondante :

$$P_{\text{acq}} = P_{\text{émise}} * \text{Exposition}$$

▪ Ratio S/P

Le ratio S/P (sinistres sur primes) est un important indicateur de rentabilité qui permet de confronter la charge de sinistre observée aux montant de primes acquises :

$$S/P = \frac{\text{Charge de sinistre}}{P_{\text{acq}}}$$

⁴ Le nombre de jours dans une année pouvant varier, on considère à chaque fois des années de 365,25 jours pour que les expositions soient à chaque fois exprimées dans la même unité et restent donc comparables d'une année à l'autre.



Plus le S/P est petit, plus le portefeuille est rentable. Cet indicateur permet uniquement de faire un constat de l'état du portefeuille en matière de rentabilité passée, il n'est pas prospectif.

- **Fréquence de survenance**

La fréquence de survenance de sinistre correspond au nombre de sinistres observés par unité de temps. On l'exprime en nombre de sinistres par période, le plus souvent d'un an :

$$\text{Fréquence} = \frac{\text{Nombre de sinistres}}{\text{Exposition}}$$

On effectue le calcul à partir de l'exposition pour éviter de sous-estimer la fréquence. La fréquence n'est ainsi par exemple pas la même pour une affaire nouvelle arrivée en fin d'année et ayant subi 2 sinistres que pour un assuré présent toute l'année en portefeuille et ayant également subi 2 sinistres. Dans le premier cas la fréquence est beaucoup plus importante.

- **Coût moyen**

Le coût moyen des sinistres est simplement la moyenne des montants de sinistres :

$$\text{Coût moyen} = \frac{\sum \text{Montants de sinistre}}{\text{Nombre de sinistres}}$$

On calcule les coûts moyens par garantie et par exercice sur l'ensemble du portefeuille.

- **Prime pure**

La prime pure, notée π , est définie en théorie comme étant l'espérance de la charge totale de sinistre :

$$S = \sum_{i=1}^N X_i$$

Où N est une variable aléatoire représentant le nombre total de sinistre et $(X_i)_{1 \leq i \leq N}$ les variables aléatoires (i.i.d)⁵ représentant les montants de sinistres. On suppose N et les $(X_i)_{1 \leq i \leq N}$ indépendantes, on obtient ainsi :

$$\begin{aligned} \mathbb{E}[S] &= \mathbb{E} \left[\sum_{i=1}^N X_i \right] = \mathbb{E} \left[\mathbb{E} \left[\sum_{i=1}^N X_i \mid N \right] \right] = \mathbb{E} \left[\mathbb{E} \left[\sum_{i=1}^N X_i \mathbb{1}_{i \leq N} \mid N \right] \right] \\ &= \mathbb{E} \left[\sum_{i=1}^N \mathbb{E}[X_i \mid N] \mathbb{1}_{i \leq N} \right], \mathbb{1}_{i \leq N} \text{ étant une variable aléatoire } N\text{-mesurable} \end{aligned}$$

$$= \mathbb{E} \left[\sum_{i=1}^N \mathbb{E}[X_i] \mathbb{1}_{i \leq N} \right] = \mathbb{E} \left[\mathbb{E}[X_1] \sum_{i=1}^N \mathbb{1}_{i \leq N} \right] = \mathbb{E}[X_1] \mathbb{E} \left[\sum_{i=1}^N \mathbb{1}_{i \leq N} \right] = \mathbb{E}[X_1] \mathbb{E}[N] \text{ par indépendance.}$$

La prime pure est donc égale au produit de l'espérance du coût des sinistres et de celle du nombre de sinistres. C'est la principale composante du calcul de la SUE, pour l'évaluer nous aurons recours au modèle linéaire généralisé (GLM pour « Generalized Linear Model » en anglais) pour obtenir d'une part un modèle de fréquence et d'autre par un modèle de sévérité.

⁵ Indépendantes et identiquement distribuées



1.2. Modèle linéaire généralisé

Pour obtenir ces modèles de fréquence et de sévérité nous aurons recours au modèle linéaire généralisé. C'est un modèle de tarification a priori permettant d'estimer la prime pure d'un assuré en prenant en compte ses caractéristiques. De manière générale, en notant Y la variable aléatoire représentant par exemple la fréquence ou la sévérité liée à un assuré et X ses caractéristiques, le principe d'une régression est, d'un point de vue statistique, de modéliser $\mathbb{E}[Y | X]$ comme une fonction des variables explicatives :

$$\mathbb{E}[Y | X] = g(X)$$

On écrit alors

$$Y = g(X) + \epsilon$$

Où ϵ est un bruit aléatoire qui représente l'écart entre Y et son espérance conditionnelle, c'est-à-dire l'erreur commise en remplaçant Y par son espérance conditionnelle. Tout l'enjeu d'une régression est ainsi de déterminer la fonction g . Le modèle le plus simple, le modèle linéaire classique, suppose que la fonction g est une fonction linéaire et que les résidus ϵ suivent une loi normale. Cependant la forme linéaire de g peut s'avérer trop restrictive et la normalité de ϵ inadaptée aux données. L'idée du modèle linéaire généralisé est de conserver la simplicité du modèle linéaire classique tout en s'autorisant une forme plus générale. Les coefficients sont ainsi estimés par la technique du maximum de vraisemblance en nous restreignant à la famille exponentielle. Un modèle statistique $(\Omega, F, (\mathbb{P}_{\theta, \phi})_{\theta \in \Theta, \phi > 0})$ est appelé famille exponentielle si les probabilités $\mathbb{P}_{\theta, \phi}$ admettent une densité f de la forme :

$$f_{\theta, \phi}(y) = c_{\phi}(y) \exp\left(\frac{y\theta - a(\theta)}{\phi}\right)$$

Les lois usuelles appartenant à la famille exponentielle sont par exemple :

- La loi normale de paramètres m et σ^2 connu $\theta = m$ et $a(\theta) = \frac{\theta^2}{2}$ et $\phi = \sigma^2$.
- La loi exponentielle ou gamma de paramètres k et λ , avec k connu et $\theta = -\frac{1}{\lambda}$, $a(\theta) = -k \log(-\theta)$ et $\phi = -1$.
- La loi de Poisson de paramètre λ , $\theta = \log \lambda$, $a(\theta) = e^{\theta}$ et $\phi = 1$.
- La loi de Bernoulli ou binomiale de paramètres n et p , avec n connu $\theta = \log\left(\frac{p}{1-p}\right)$, $a(\theta) = n \log(1 + e^{\theta})$ et $\phi = 1$.
- La loi binomiale négative de paramètres p et r , avec r connu, $\theta = \log(1 - p)$, $a(\theta) = r \log(1 - e^{\theta})$ et $\phi = 1$.

A contrario les distributions de Cauchy, de Pareto et Log-normale n'appartiennent à la famille exponentielle.

Ainsi, en reprenant les notations précédentes, un modèle est un modèle linéaire généralisé si :

- $Y | X = x \sim \mathbb{P}_{\theta, \phi}$ appartient à la famille exponentielle.
- $g(\mathbb{E}[Y|X]) = X'\beta$ pour une certaine fonction g bijective appelée fonction de lien et avec β une vecteur colonne correspondant aux paramètres du modèle.

Les fonctions de lien canoniques utilisées se les distributions utilisées sont :

- $\mu \mapsto g(\mu) = \mu$, cette fonction de lien est celle utilisée dans le modèle linéaire gaussien.
- $\mu \mapsto g(\mu) = \log \mu$, en pratique on utilisera souvent cette fonction lien afin d'obtenir une structure multiplicative de coefficients.
- $\mu \mapsto g(\mu) = \frac{1}{\mu}$
- $\mu \mapsto g(\mu) = \text{logit}(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$

Après avoir décidé d'une fonction de lien et d'une distribution appartenant à la famille exponentielle, on calcule ensuite $\hat{\beta}$ un estimateur du maximum de vraisemblance de β et on obtient enfin un estimateur $\hat{Y} = g^{-1}(X\hat{\beta})$.



Dans certains cas il peut s'avérer opportun de normaliser les données comme lorsque qu'on veut modéliser un nombre de sinistres tout en prenant en compte l'exposition, notée ici ω , telle que nous l'avons définie précédemment. On note Y le nombre de sinistre et X les caractéristiques des individus et considère par exemple un GLM avec une loi de Poisson et une fonction de lien logarithmique, on cherche ainsi un estimateur de β tel que :

- $Y | X = x$ suit une loi de Poisson
- $\log\left(\frac{\mathbb{E}[Y|X]}{\omega}\right) = X'\beta$, ce qui peut se réécrire comme $\log(\mathbb{E}[Y|X]) = \log \omega + X'\beta$
- $\log \omega$ est alors ce qu'on appelle l'offset du modèle. C'est en quelque sorte une variable supplémentaire avec un coefficient β égal à 1.

L'avantage des modèles linéaires généralisés réside principalement dans leur simplicité et leur explicabilité. En effet, une modélisation, un tarif doivent pouvoir être facilement interprétables et explicables. Néanmoins le choix des lois à ajuster se limite à la famille exponentielle, et le choix des fonctions de lien est souvent imposé par un choix canonique correspondant à la loi en question. De plus, se limiter à la famille exponentielle impose de ne pas avoir de valeurs extrêmes.

1.3. AIC et BIC

Les critères AIC et BIC sont des critères de mesure de la qualité de modèle statistique qui vont nous servir lors de nos sélections de modèles. On les définit de la façon suivante :

$$AIC = -2 \ln(L) + 2k \text{ et } BIC = -2 \ln(L) + k \ln(N)$$

Où L est la vraisemblance du modèle, k le nombre de paramètres et N le nombre d'observations. L'AIC et le BIC sont des critères de parcimonie qui pénalisent la vraisemblance des modèles en fonction du nombre de paramètres. Plus ces critères sont faibles, meilleur est le compromis entre qualité de l'ajustement et complexité des modèles. Plus l'on ajoute des variables, plus la vraisemblance du modèle augmente, ces critères permettent ainsi de limiter le surapprentissage en pénalisant le nombre de paramètres des modèles. Contrairement à l'AIC, le BIC pénalise en plus la vraisemblance par le nombre d'observations, notre procédure s'appuiera ainsi principalement sur l'AIC mais tout en restant attentif au BIC.

2. Périmètre de la modélisation

2.1. Contrats

Comme nous l'avons évoqué précédemment, les contrats de notre portefeuille sont issus de différentes versions du produit d'assurance que l'on considère. On restreint le périmètre de notre étude aux contrats correspondant à sa version actuelle. En effet, le produit actuel qui a été lancé en 2016 représente le portefeuille à venir et en cours. Sur les graphiques ci-dessous « CV2 » fait référence au produit actuel, « CV0 » et « CV1 » concernent ses anciennes versions :

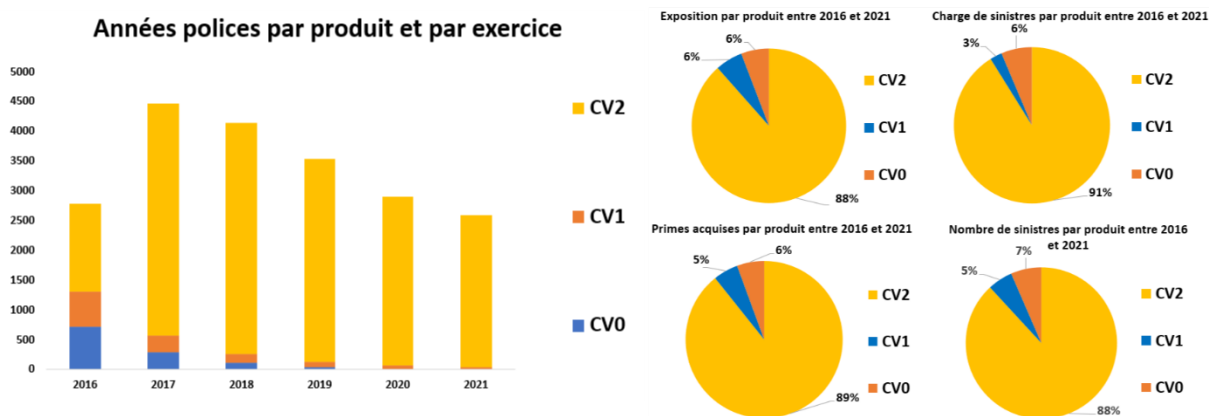


Figure 1 : Répartition de l'exposition et de la sinistralité par produit



Depuis 2016 on observe bien que notre portefeuille est très majoritairement composé de contrats du produit actuel. De même sur la période de 2016 à 2021 plus de 90% de l'exposition, des primes acquises et de la sinistralité concernent ce produit.

2.2. Choix des garanties à modéliser

Notre volume de données étant relativement restreint, notre étude se focalisera sur quelques garanties. Pour les autres garanties non modélisées on supposera le tarif commercial équivalent au tarif technique. Pour déterminer les garanties à modéliser, on observe la sinistralité de chaque garantie sur toute la période considérée dans notre étude.

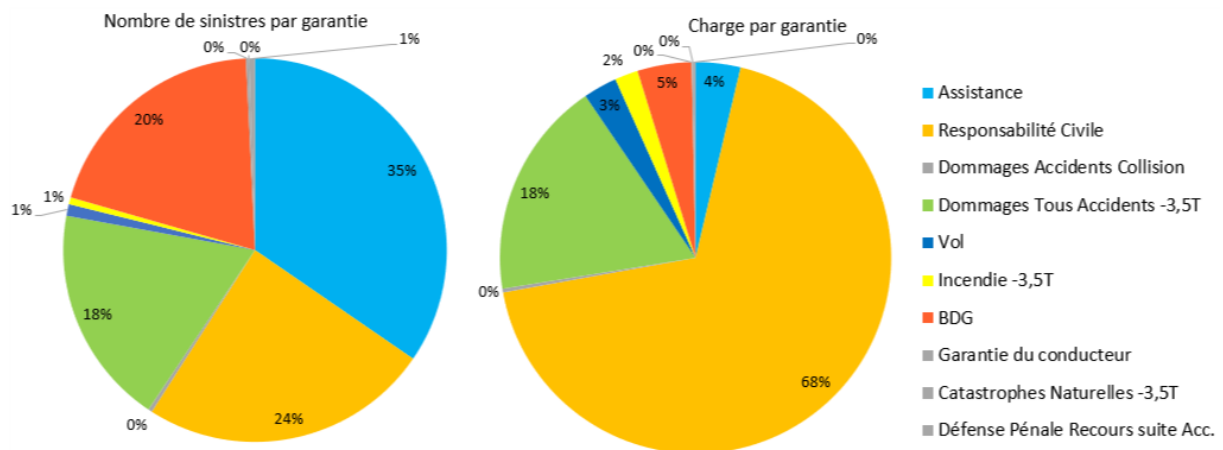


Figure 2: Sinistralité par garantie

Les garanties « assistance », « responsabilité civile », « dommages tous accidents » et « bris de glace » sont les plus représentées en ce qui concerne l'exposition, le nombre de sinistres et leur charge, ce sont les seules garanties pour lesquelles il semble y avoir assez de sinistralité observée pour envisager une modélisation. Elles concentrent plus de 94% du nombre de sinistres et 97 % de la charge de sinistres entre 2016 et 2021. Notre modélisation se restreindra donc à ces garanties. Les autres garanties non modélisées ne présentent par ailleurs pas de dérives particulières et ont des S/P proches des S/P cibles de la compagnie.

2.3. Sinistres graves

En assurance de biens, il arrive souvent que pour certaines garanties un faible nombre de sinistres soit responsable d'une grande partie de la charge totale, d'où la nécessité de modéliser séparément les sinistres d'un montant exceptionnel. On note donc la présence de sinistres à coûts élevés que l'on va considérer comme graves s'ils dépassent certains seuils que l'on va définir en utilisant la théorie des valeurs extrêmes.

2.3.1. Cadre théorique

2.3.1.1. Introduction à la théorie des valeurs extrêmes

La théorie des valeurs extrêmes s'intéresse aux valeurs extrêmes des distributions de probabilité et fournit des méthodes et outils rationnels pour estimer le comportement asymptotique des maxima de valeurs prises par des variables aléatoires. Pour rappel selon le théorème central limite si :

X_1, \dots, X_n sont n variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition F alors :



$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{\text{loi}} N(0,1), \text{ où } \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

L'idée est de disposer d'un résultat similaire pour déterminer le comportement asymptotique des maxima or si on note $M_n = \max(X_1, \dots, X_n)$ on constate que la loi limite de M_n est dégénérée :

$$\forall x \in \mathbb{R}, F_{M_n}(x) = P(M_n \leq x) = F(x)^n \text{ d'où } \forall x \in \mathbb{R}, F_n(x) \xrightarrow{n \rightarrow \infty} \begin{cases} 0 & \text{si } F(x) < 1 \\ 1 & \text{si } F(x) = 1 \end{cases}$$

On cherche ainsi par analogie avec le théorème central limite à trouver une transformation linéaire permettant d'éviter cette dégénérescence. Pour cela on commence par donner quelques définitions. On dira ainsi que F appartient au domaine d'attraction du maximum de H et on notera $F \in MDA(H)$ si H est une fonction de répartition non dégénérée et s'il existe deux suites $(a_n) > 0$ et (b_n) tel que quel que soit $x \in \mathbb{R}$, $F^n(a_n x + b_n) \xrightarrow{n \rightarrow \infty} H(x)$. De manière équivalente $\frac{M_n - b_n}{a_n} \xrightarrow{\text{loi}} Z$, où Z a comme fonction de répartition H .

On dira également que F est de même type que H s'il existe $a > 0$ et $b \in \mathbb{R}$ tel que $H(x) = F(ax + b)$. Ensuite d'après le **théorème des valeurs extrêmes** si $F \in MDA(G)$ alors G est nécessairement du même type que la distribution GEV (Generalized Extreme Value) définie par :

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

où $y_+ = \max(0, y)$. Le paramètre $\gamma \in \mathbb{R}$ détermine la nature de la queue de distribution est appelé l'indice de valeur extrême.

2.3.1.2. Méthodes du seuil

Parmi les approches les plus répandues de la théorie des valeurs extrêmes on retrouve celles basées sur le dépassement de seuils élevés. On choisit un seuil élevé $u < x_F^+ = \sup\{x : F(x) < 1\}$ puis on étudie tous les excès au-dessus de ce seuil. On définit ainsi la fonction de répartition des excès par :

$$F_u(x) = P(X - u | X > u), x \geq 0.$$

On peut ainsi énoncer le **théorème de Pickands**(1975) :

Si F appartient à un domaine d'attraction du maximum, c'est-à-dire la distribution limite de F est une GEV, alors pour u assez grand on a :

$$\lim_{u \uparrow x_F^+} \sup_{0 \leq y \leq x_F^+ - u} |F_u(y) - G_{\xi, \sigma(u)}(y)| = 0,$$

où $G_{\xi, \sigma(u)}$ est la distribution de Pareto généralisée (GPD pour Generalized Pareto distribution), $\sigma(u) > 0$, donnée par

$$G_{\xi, \sigma(u)}(y) = \begin{cases} 1 - \left(1 + \xi \frac{y}{\sigma(u)}\right)^{-\frac{1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma(u)}\right) & \text{si } \xi = 0 \end{cases}$$

où $y \geq 0$ si $\xi \geq 0$ et $0 \leq y \leq -\frac{\sigma(u)}{\xi}$ si $\xi < 0$. Pour un seuil suffisamment grand $u, \exists \sigma(u) > 0$ et ξ un nombre réel tel que $F_u(y) \approx G_{\xi, \sigma(u)}(y)$. Si X suit une GPD alors son espérance est définie pour $\xi < 1$ et est telle que : $E(X) = \frac{\sigma}{1-\xi}$.

En particulier pour $E(X^k) = \infty$ pour $k \geq 1 / \xi$. On définit à présent $e(u) = E(X - u | X > u)$ la fonction de dépassement moyen (mean excess function) de X .



Si $X - u | X - u > 0 \sim \text{GPD}(\sigma(u), \xi)$ alors on a $\forall v > u$ on a alors :

$$e(v) = E(X - v | X > v) = \frac{\sigma(u) + \xi(v-u)}{1-\xi}.$$

$E(X - v | X > v)$ est linéaire en v avec une pente égale à $\frac{\xi}{1-\xi}$. Si de plus on note $\sigma^* = \sigma(v) - \xi v$ alors σ^* ne dépend pas de v pour $v \geq u$ tel que $X - u | X - u > 0 \sim \text{GPD}(\sigma(u), \xi)$.

Cela nous permet de définir une procédure de sélection de seuil, on trace le graphique de la fonction de dépassement moyen (mean excess plot), c'est-à-dire le graphique de $e(u)$ pour plusieurs seuils u , et on observe à partir de quel seuil le graphique est approximativement linéaire. De la même manière on calcule des estimateurs du maximum de vraisemblance de σ^* et ξ et on observe à partir de quel seuil ils sont approximativement constants. Le seuil, en théorie, idéal à choisir est le seuil minimal pour lequel ces propriétés sont satisfaites. Dans la pratique, le choix d'un seuil résulte d'un compromis entre choisir un seuil relativement élevé, mais qui limite le nombre d'observations pour calculer des estimateurs, et un seuil trop bas mais qui permette de disposer de plus d'informations pour l'estimation des paramètres de la GPD mais dans ce cas l'approximation par une GPD est en générale mauvaise. Dans le premier cas on augmente la variance de l'estimateur, dans l'autre son biais. On valide enfin notre sélection de seuil en observant les diagrammes quantile-quantile des observations et des lois estimées et les fonctions de répartition théoriques et empiriques.

2.3.2. Détermination d'un seuil de grave par garantie

Après avoir rappelé quelques notions de la théorie des valeurs extrêmes nous allons à présent essayer de déterminer, lorsque cela s'avère pertinent, un seuil de montant de sinistre grave par garantie à l'aide de la procédure que nous avons définie ci-dessus. Avant de suivre cette procédure, nous observons tout d'abord les quantiles de montants de sinistres par garantie pour avoir une idée de comment se répartissent les montants, si un certain groupe de sinistres se dégage et s'ils peuvent atteindre des montants qu'on pourrait qualifier de grave au regard, par exemple, du seuil pratiqué en interne pour qualifier un sinistre de grave qui est de 76 000 €.

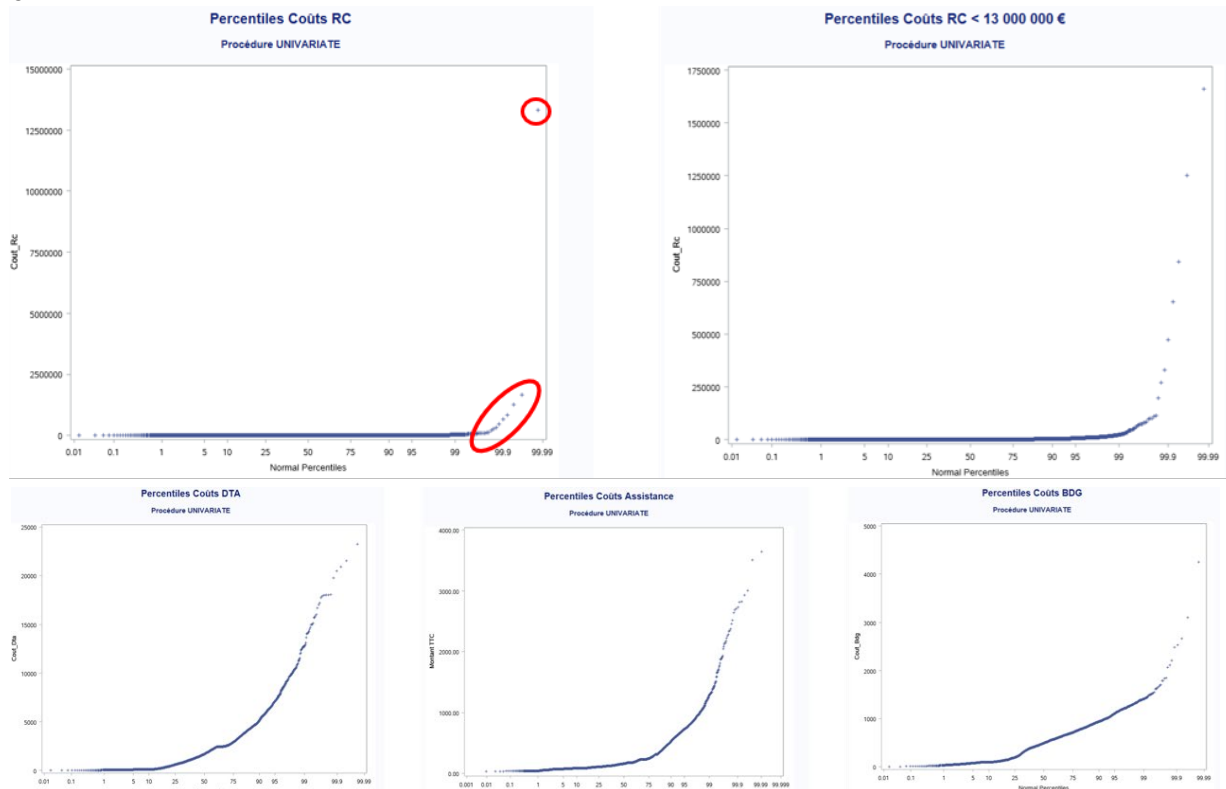


Figure 3: répartition des montants de sinistres par garantie



On s'aperçoit que sauf en RC, ou par ailleurs on constate un sinistre d'un montant exceptionnel, il n'apparaît pas pertinent de définir un seuil de grave. En effet, en BDG les montants ne dépassent pas 5000 €, en assistance ils ne dépassent pas 4000 € et en DTA où l'on n'observe pas de sinistre au-delà de 25 000€, on choisit de ne pas non plus considérer de seuil de grave bien que cela aurait pu apparaître pertinent mais l'historique de sinistres est trop restreint pour cette garantie (moins de 3600 sinistres sur la période de 2016 à 2021). On ne considérera donc que la sinistralité grave issue de la garantie RC et sans faire de distinction entre sinistres corporels et matériels car encore une fois on ne dispose pas d'un volume de données suffisant pour faire deux modélisations distinctes. La présence de ce sinistre exceptionnel et dont on ne connaît pas encore la période de retour, d'un montant supérieur à 10 000 000 € impacte très fortement la sinistralité observée. C'est pour cette raison que nous avons souhaité présenter les percentiles des montants de sinistres RC avec et sans ce sinistre car il pourrait influencer ou fausser la lecture du graphique en minimisant l'importance d'autres sinistres qu'on pourrait aussi qualifier d'extrêmes. Ainsi pour estimer la sinistralité grave en RC nous tenterons plusieurs approches pour essayer d'en adopter une qui prenne judicieusement en compte ce sinistre, sans en exagérer ou minimiser l'importance, et sous la contrainte d'un volume de données restreint. Mais avant d'aller plus loin et de détailler ces méthodes que nous décrivons ultérieurement nous devons tout d'abord convenir d'un seuil de sinistre grave pour délimiter le périmètre de notre modélisation hors graves. Pour cela nous appliquons la procédure décrite précédemment.

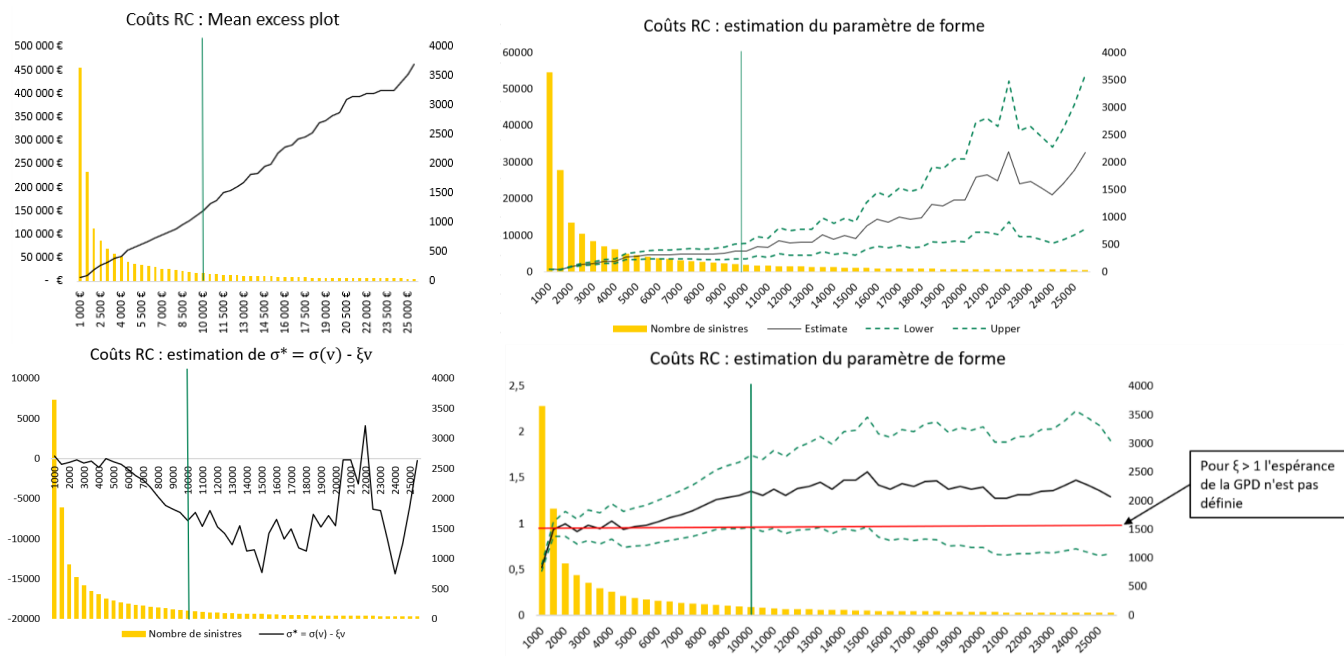


Figure 4 : Détermination d'un seuil de graves RC

En observant les graphiques ci-dessus, on constate qu'aucun seuil ne semble réellement pertinent pour estimer notre sinistralité grave, il nous faudrait un volume de données plus conséquent pour en déterminer un qui semble suffisamment crédible. D'un point de vue purement statistique on aurait pu choisir, par exemple, un seuil autour de 2000 €. Pour ce seuil, le mean excess plot est linéaire, les paramètres de la GPD sont bien estimés et le graphique de σ^* tel que défini précédemment semble constant mais on ne s'attend pas à un seuil de sinistre grave de cet ordre de grandeur lorsque l'on cherche à en déterminer un pour cette garantie où certains sinistres, notamment corporels, peuvent atteindre des montants bien plus élevés. On cherche en effet un seuil de grave qui soit cohérent par rapport à la survenance de ces sinistres. Or sous le seuil actuellement pratiqué, le coût moyen d'un sinistre attritionnel est égal à environ 2100 € et plus de 25% des sinistres observés excèdent 2000 €. C'est pourquoi ce seuil n'est pas pertinent et dans une première approche, nous adopterons le seuil de sinistre grave actuellement pratiqué qui est égal à 76 000 €. Pour aller plus loin, tel que nous le verrons dans notre partie sur l'estimation de sinistralité grave, nous explorerons aussi la possibilité de prendre un seuil de grave égal à 10 000 €, 3% des sinistres excèdent ce seuil et représentent 54% de la charge observée entre 2016 et 2021. Nous jugerons de la pertinence de l'usage de ce seuil, qui respecte en partie notre procédure de sélection de seuil, lorsque nous aborderons l'estimation de la sinistralité grave.



2.4. Sinistres climatiques

Au même titre que les sinistres graves, les sinistres climatiques ont un comportement en matière de fréquence de survenance et de sévérité qui nécessite une prise en compte spécifique dans le calcul final de notre prime pure. Dans notre produit, ces sinistres peuvent être issus des garanties dommages tous accidents, incendie et catastrophes naturelles. Notre historique compte trop peu de sinistres de cette nature pour envisager une modélisation. Cette sinistralité ne présente pas d'enjeu majeur dans notre périmètre on fait donc l'hypothèse qu'elle est déjà prise en compte dans la prime Incendie non modélisée et dans la prime pure DTA que nous allons modéliser. On peut d'ailleurs voir que dans les garanties DTA et INC, ces sinistres sont minoritaires :

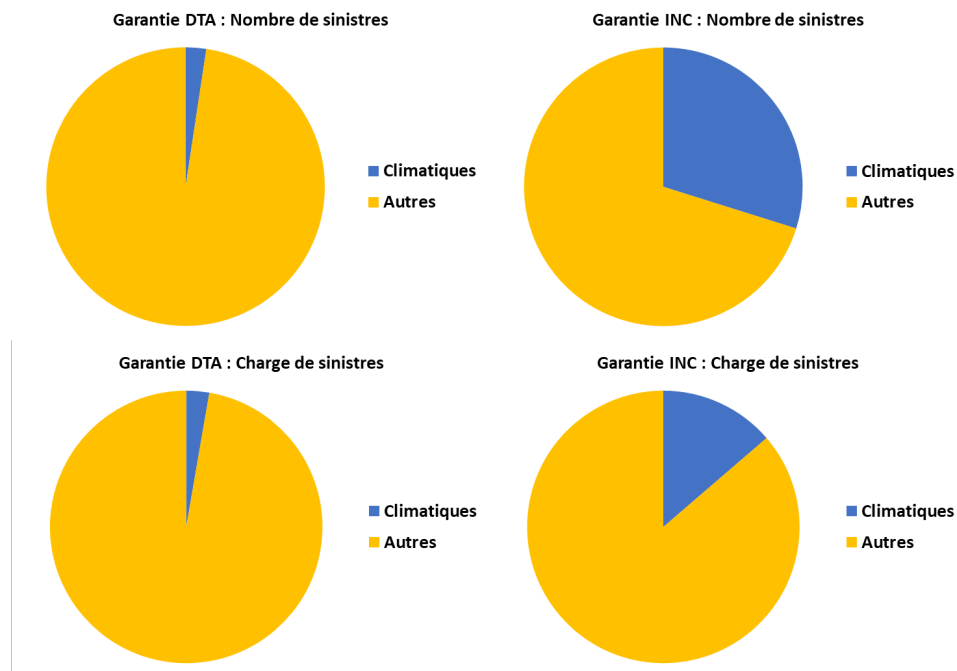


Figure 5: Poids de la sinistralité climatique par garantie

3. Base de modélisation

3.1 Mise en place de la base

La base de données utilisée pour la modélisation de la prime pure attritionnelle est constituée de données internes comprenant toutes les situations par exercice des contrats du produit actuel. Elle comporte des informations générales sur les assurés, les conducteurs, leur activité, leur véhicule et leur historique de sinistralité entre 2016 et 2021. Une situation pour un contrat et un exercice donné, c'est-à-dire une ligne de notre base, est identifiée par une date de début et une date de fin d'observation. Tout changement de situation d'un assuré par un avenant, une mise à jour ou une modification d'une information, introduit en effet une nouvelle ligne dans notre base de données. Cette nouvelle ligne correspond à une photographie, une image, du contrat sur la période délimitée par les dates de début et de fin de situation. Les données sont mises à jour mensuellement. Pour obtenir notre base de données finale nous avons dans un premier temps établi deux bases de données distinctes, une base « contrats » contenant toutes les situations des assurés et les informations correspondantes, et une base « sinistres » contenant la dernière vision observée fin 2021 de chaque sinistre



survenu entre 2016 et 2021. On peut voir dans le schéma ci-dessous les grandes catégories de variables présentes dans nos bases :

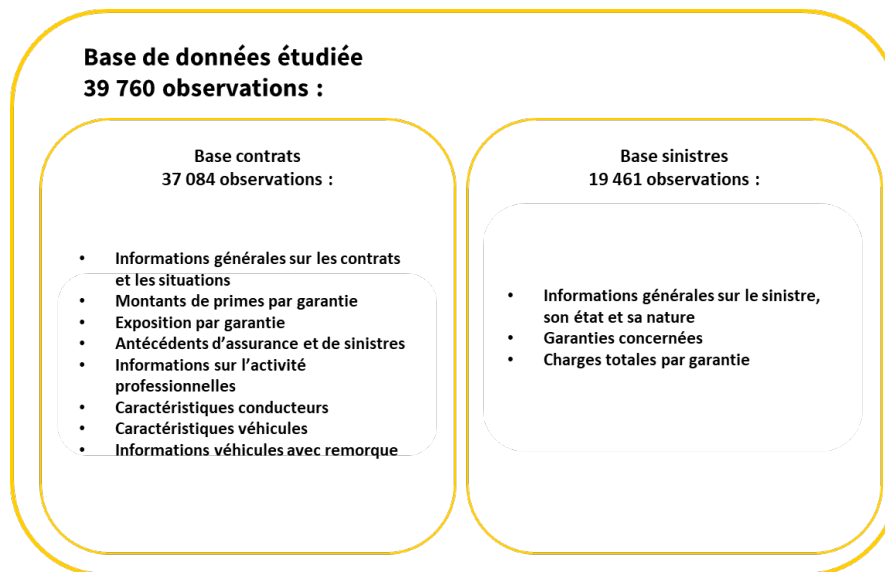


Figure 6: Schéma de la structure de la base de modélisation

On a ensuite joint ces deux tables en rattachant chaque sinistre d'un assuré à la situation concernée, et donc au risque correspondant, selon sa date de survenance. Or on souhaite qu'un sinistre soit rattaché à une seule et même situation. Ainsi lorsque plusieurs sinistres sont survenus lors de la même période délimitant une situation, on la divise de sorte que chaque sinistre ne soit associé qu'à une seule et même situation. On adapte ainsi les dates de début et de fin de ces situations dérivées à partir des dates de survenance des sinistres :

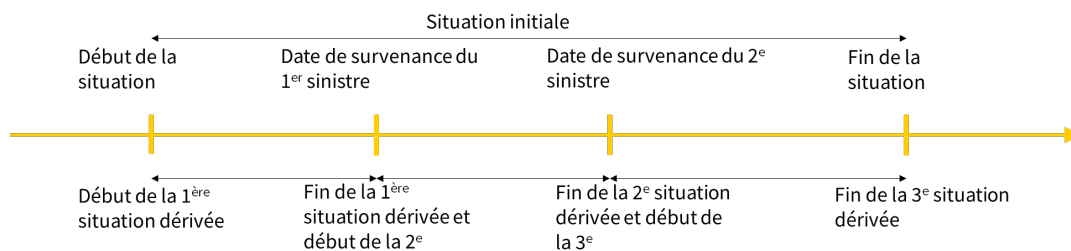


Figure 7: Découpage des situations selon la survenance de sinistres

La base de données finale contient ainsi plus d'observations que la base contrats initiale à laquelle on a joint les sinistres correspondants.

3.2. Validation

Avant toute chose on contrôle la mise en place de notre base de données. Nos outils de suivi incluent l'ensemble des contrats sans différencier le produit actuel et ses précédentes versions, nous avons donc intégré dans un premier temps ces anciens contrats à notre base de données. On vérifie ainsi qu'avant jointure entre nos bases sinistres et contrats on retrouve l'exposition et la sinistralité observées dans nos tableaux de bord de suivi interne. On s'assure ensuite qu'après jointure entre nos bases on n'a pas perdu d'information et qu'enfin après extraction des contrats du produit actuel qui constituent le périmètre de notre modélisation on retrouve les proportions de départ lorsque nous avons évalué le poids de ces contrats en matière d'exposition et de sinistralité :



Validation et bilan des écarts		Responsabilité Civile	Dommages Tous Accidents - 3,5T	BDG	Assistance	Total y compris autres garanties
Conservation des données après jointure entre les bases sinistres et contrats	Exposition	100%	100%	100%	100%	100%
	Primes acquises	100%	100%	100%	100%	100%
	Nombre de sinistres	100%	100%	100%	99%	99%
	Montants de sinistres	99%	100%	100%	99%	99%
Correspondance avec le tableau de bord avant jointure	Exposition	100%	100%	100%	100%	100%
	Primes acquises	100%	100%	100%	100%	100%
	Nombre de sinistres	100%	100%	100%	100%	100%
	Montants de sinistres	100%	100%	100%	100%	100%
Correspondances avec le tableau de bord après jointure	Exposition	100%	100%	100%	100%	100%
	Primes acquises	100%	100%	100%	100%	100%
	Nombre de sinistres	99%	100%	100%	99%	99%
	Montants de sinistres	100%	100%	100%	100%	100%
Correspondances avec le dashboard après extraction CV2	Exposition	88%	90%	88%	88%	88%
	Primes acquises	87%	87%	90%	98%	89%
	Nombre de sinistres	88%	92%	88%	86%	88%
	Montants de sinistres	93%	92%	90%	85%	92%

Tableau 1 : Validation de la mise en place de la base de données

Après extraction des contrats du produit actuel on conserve près de 90% de l'exposition et de la sinistralité observées. Ainsi on obtient bien les mêmes écarts que ceux présentés précédemment lors de notre choix de nous restreindre au produit actuel. Les écarts après jointure entre les bases sinistres et contrats, de l'ordre de 1% du nombre de sinistre et de 1% de la charge de sinistre, sont dus à des sinistres survenus hors d'une période de couverture. Nous ne les prenons pas en compte dans notre étude. A ce stade on a ainsi pu contrôler la bonne mise en place de notre base de données, on peut donc procéder à la prochaine étape précédant notre modélisation.

3.3. Sélection des variables de modélisation

Une autre étape importante avant de débiter notre modélisation est de faire une présélection de variables parmi celles présentes dans notre base de données. Cette étape est indispensable pour réduire le nombre de variables à tester dans le modèle par la suite. Les catégories de variables pouvant figurer dans nos modèles sont celles concernant les antécédents de sinistres et d'assurance, l'activité professionnelle, les caractéristiques des conducteurs et de leur véhicule. On trouvera en annexe (cf. annexe 1) un tableau récapitulatif de la liste des variables des variables de départ avec des indications sur leur présélection ou leur exclusion.

3.4. Valeurs manquantes

On élimine dans un premier temps les variables ayant une part très importante de valeurs manquantes. Cette étape nous permet déjà de faire un premier nettoyage de notre base de données en éliminant par exemple toutes celles ayant 100 % ou presque de valeurs manquantes. Pour celles-ci l'information n'existe tout simplement pas ou n'est pas renseignée et donc aucun retraitement n'est envisageable. Pour cette raison nous avons exclu les variables décrivant les caractéristiques des remorques de nos assurés. Comme la quasi-totalité n'en possède pas toutes les informations décrivant leurs caractéristiques sont manquantes :



Procédure FREQ

remorque O/N				
REMORQUE	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
N	39287	98.93	39287	98.93
O	424	1.07	39711	100.00
Fréquence manquante = 49				

Variable	Libellé	Valeurs manquantes
REMORQUE	Remorque O/N	0,12%
GAR_COMPLETE_RMQ	Garantie complète	98,93%
GAR_BASE_RMQ	Garantie de base	98,93%
GENRE_RMQ	Genre	98,93%
MARQUE_RMQ	Marque	98,93%
MEC_RMQ	Date mise en circulation remorque	98,93%
MODELE_RMQ	Modèle	98,93%
POIDS_RMQ	PTAC	98,93%
USAGEPRO_RMQ	Usage professionnel	98,93%

Tableau 2 : Valeurs manquantes par variables liées aux remorques

En ce qui concerne les autres variables on peut choisir selon les cas de les exclure ou de les retraiter. A ce titre nous avons choisi d'exclure les variables d'antécédents d'assurance et de sinistre avant la souscription qui présentent un grand nombre de valeurs manquantes :

Variable	Libellé	Valeurs manquantes
RESILIATIONANT	Antécédent de résiliation	24%
motifresilant	Motif antécédent de résiliation	94%
MOTIFDUREEASS	Motif Duree assurance moins de 12 mois	100%
SINISTREANT	Antécédents Sinistres	19%
dureeassurance	Durée assurance	19%
NBANTSIN_AN	Nombre de sinistres moins de 24 mois	60%
ant_bdg	Antecedents BDG moins de 24 mois	60%
ant_inc	Antecedents INC moins de 24 mois	60%
ant_resc	Antecedents resp. corporels moins de 24 mois	60%
ant_resm	Antecedents resp. materiels moins de 24 mois	60%
ant_rest	Antecedents non resp. avec tiers. moins de 24 mois	60%
ant_srec	Antecedents semi-resp. mat. moins de 24 mois	60%
ant_srem	Antecedents semi-resp. corp. moins de 24 mois	60%
ant_vol	Antecedents vol moins de 24 mois	60%

Tableau 3 : Valeurs manquantes des variables d'antécédents d'assurance et de sinistre

En effet ces variables correspondent à des questions posées aux assurés lors de leur souscription et donc ces informations ne sont pas directement renseignées dans ces variables pour tous les clients déjà en portefeuille avant le déploiement du produit actuel. De plus les informations qu'elles contiennent ne sont pas mises à jour, en effet lorsqu'un client est en portefeuille depuis plusieurs années et que sa situation a évolué, les informations contenues dans ces variables renvoient toujours à celles renseignées lors de la souscription. Nous aurions pu les retraiter en fonction de l'expérience que nous avons acquis sur nos assurés mais dans une première approche nous avons décidé d'exclure ces variables qui par ailleurs peuvent poser question quant à leur fiabilité dans le cadre de notre modélisation.

3.5. Conducteurs non désignés

En effet, une particularité importante de notre portefeuille que nous avons déjà évoquée est la présence de conducteurs non désignés. Dans ce cas toutes les informations d'antécédents d'assurance et de sinistre à la souscription sont biaisées car elles ne correspondent pas aux conducteurs mais à leur entreprise ou aux souscripteurs. Or, comme on le verra dans la prochaine section, les conducteurs non désignés représentent plus d'un tiers de notre base de données on ne peut donc pas négliger ce biais. Les valeurs manquantes des autres variables de notre base de données sont par ailleurs la conséquence de la présence de ces conducteurs non désignés. C'est le cas des variables correspondant à l'âge des conducteurs et à l'ancienneté de leur permis de conduire. Lorsque le conducteur est non désigné, ces informations sont par conséquent non renseignées. Or, comme on souhaite spécialement prendre en compte à la fois cette particularité de notre portefeuille dans notre modélisation et ne pas nous priver de ces informations pour les conducteurs désignés, on décide de retraiter ces variables. On les croise ainsi avec la variable « Typconducteur » qui a pour modalités « Désigné » et « Non



désigné » et qui nous permet d'identifier les conducteurs désignés ou non désignés. Ainsi on remplace leurs valeurs manquantes dues à un conducteur non désigné par la modalité « Non désigné ». On réduit ainsi le nombre de valeurs manquantes et on obtient deux nouvelles variables nous permettant de prendre en compte le caractère non désigné des conducteurs :

Variable	Libellé	Valeurs manquantes
age_cond	Age du conducteur avant retraitement	34%
age_cond2	Age du conducteur après retraitement	0%
anc_permis	Ancienneté du permis avant retraitement	34%
anc_permis2	Ancienneté du permis après retraitement	0%

Tableau 4: Age du conducteur – valeurs manquantes avant et après retraitement

3.6. Coefficient de réduction majoration

La présence importante de conducteurs non désignés nous contraint par ailleurs à ne pas retenir le CRM (coefficient de réduction-majoration ou coefficient bonus-malus) parmi la liste des variables candidates à la modélisation. Le CRM est pourtant une variable incontournable en assurance automobile. C'est un système d'ajustement de la prime d'assurance en fonction du comportement de l'assuré. Cet ajustement intervient à chaque échéance annuelle, sur la base des sinistres impliquant la responsabilité de l'assuré. Le fonctionnement du CRM est encadré par le code des assurances, chaque assuré débute avec un CRM égal à 1 puis chaque année sans sinistre responsable le CRM décroît de 5%, on le multiplie par 0,95, tandis que pour chaque sinistre totalement responsable il est multiplié par 1,25 et par 1,125 pour chaque sinistre à responsabilité partielle. Il ne peut ni descendre en dessous de 0,5 ni excéder 3,5. En cas de changement d'assurance, l'assuré conserve son CRM. L'assureur est tenu d'appliquer ces règles de fonctionnement du CRM par la réglementation. Cette variable intervient particulièrement pour la segmentation des portefeuilles automobiles voire dans les modèles. Dans le cas de conducteurs non désignés, c'est le CRM du souscripteur du contrat qui est renseigné alors que c'est le conducteur qui est assuré. On peut ainsi avoir un contrat avec un CRM égal à 0,5, ce qui nécessite au moins 14 ans d'assurance sans sinistre responsable, alors que le conducteur est en réalité un jeune conducteur voire un conducteur avec des antécédents de sinistres responsables. Pour cette raison nous n'avons pas retenu le CRM comme variable de modélisation. De plus, le CRM, par son fonctionnement, est déjà en quelque sorte une estimation, un score, mesurant le comportement de l'assuré. On ne peut donc pas, par exemple, l'inclure tel quel à nos modèles de fréquences. Les assureurs étant tenus de prendre en compte malgré tout ce coefficient dans leur calcul de prime, il est courant dans le cas de modèles GLM de le mettre en offset, comme nous l'avons décrit précédemment pour l'exposition. En ce qui nous concerne nous avons fait le choix de modéliser sans le CRM et de l'appliquer ensuite à notre estimation finale de prime pure. Si le CRM n'est pas présent dans la modélisation, il pourrait néanmoins intervenir à d'autres niveaux comme dans le cadre de mise en place de nouvelles règles de souscription. Ce n'est cependant pas le sujet que nous avons souhaité aborder ici.



3.7. Analyses univariées et multivariées

Après avoir traité le cas des valeurs manquantes et éliminé certaines variables, on cherche ensuite à identifier les variables discriminantes en réalisant quelques statistiques descriptives. On étudie ainsi comment se répartissent les individus de notre base de données selon les valeurs prises par les variables. Cette étape nous permet d'emblée de détecter les variables où une seule modalité est représentée comme on peut le voir sur cet exemple :

Carburant				
CARBURANT	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
Autre	21	0.05	21	0.05
Biocarburant	4	0.01	25	0.06
Diesel	1048	2.64	1073	2.70
Diesel Turbo	34228	86.19	35301	88.89
Electrique	234	0.59	35535	89.48
Essence	319	0.80	35854	90.29
Hybride	3857	9.71	39711	100.00
Fréquence manquante = 49				

site aeroportuaire O/N				
SITEAEROPORTUAIRE	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
N	39699	99.97	39699	99.97
O	12	0.03	39711	100.00
Fréquence manquante = 49				

Tableau 5 : Analyses univariées des variable « carburant » et « siteaeroportuaire »

Les variables représentées ci-dessus concernent le type de carburant des véhicules et l'exercice ou non d'une activité sur un site aéroportuaire. Elles affichent toutes les deux une modalité surreprésentée voire quasi unique, elles ne peuvent donc pas discriminer les individus de notre base de données. Toutes les variables dans le même cas que celles-ci sont ainsi éliminées de notre base de modélisation. Si on avait davantage de données on aurait pu les prendre en compte et estimer des coefficients pour ces modalités sous représentés mais en l'absence de données on risque du surapprentissage. Il est ainsi plus prudent d'éliminer ces variables pour une première modélisation. A l'avenir on pourra toutefois surveiller ces variables et leurs modalités, en particulier la variable « Carburant » vue la constante progression de l'apparition des véhicules électriques ou hybrides. On fait cependant une exception pour la variable « typconducteur » que nous avons déjà évoquée :

Type de conducteur				
TYPCONDUCTEUR	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
Désigné	26352	66.36	26352	66.36
Non désigné	13359	33.64	39711	100.00
Fréquence manquante = 49				

Tableau 6 : Répartition des conducteurs désignés et non désignés



Il n'y a que deux modalités, environ deux tiers des individus partagent la première et le tiers restant la seconde. C'est un déséquilibre d'une ampleur plus raisonnable que ce qu'on a pu voir précédemment dans nos exemples mais qui n'est toutefois pas négligeable. En suivant notre logique précédente nous aurions également supprimé cette variable. Néanmoins, elle nous renseigne sur une caractéristique très importante de notre portefeuille et qu'on désire particulièrement prendre en compte dans notre modélisation. Aussi on décide malgré tout de la conserver. Ensuite on supprime les variables qualitatives ayant un trop grand nombre de modalités par rapport au nombre d'individus de notre base de données comme les marques et modèles de véhicules :

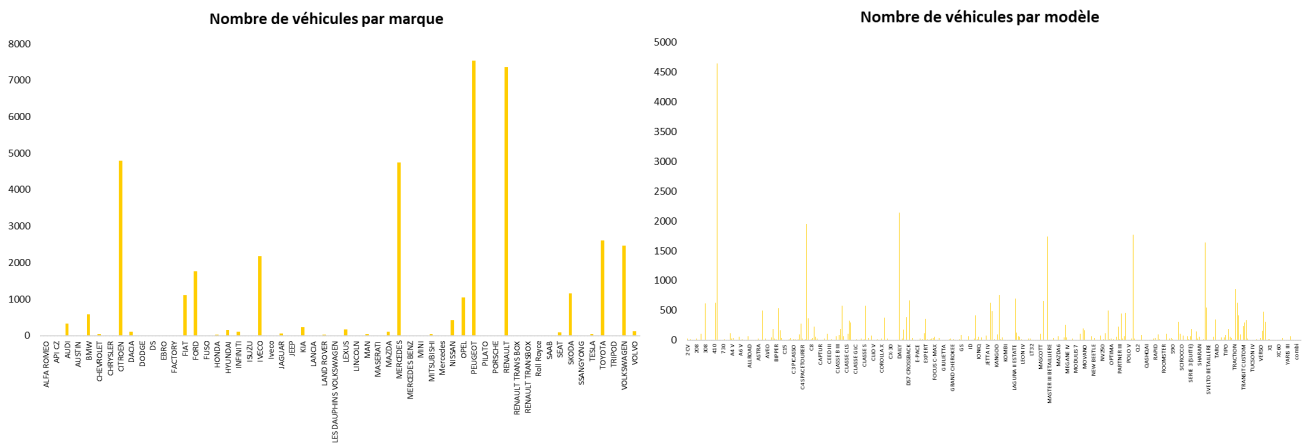


Figure 8: Nombre de véhicules par marque et par modèle

Si certaines marques ou certains modèles semblent très représentés, on observe aussi qu'il y a de très nombreuses autres modalités avec peu d'individus :

Variable	Nombre de modalités			
	Total	Moins de 500 individus	Moins de 100 individus	Moins de 10 individus
Marque	61	49	40	28
Modèle	418	399	351	212

Tableau 7: Nombre d'individus par marque et modèle de véhicule

Près de la moitié des marques et des modèles sont par exemple représentés moins de 10 fois. On ne peut donc pas inclure telles quelles ces variables à notre modélisation et prendre en compte le risque sous-jacent aux véhicules via leur marque ou leur modèle sans risquer du surapprentissage. Il nous faudrait établir des regroupements ou classifications voire établir un véhiculier pour répartir chaque marque ou modèle dans des classes de risque homogènes. Ce n'est pas l'objet de ce mémoire, on décide ici de nous appuyer sur les autres caractéristiques des véhicules pour prendre en compte leur risque. Enfin après avoir étudié les modalités de nos variables et leurs effectifs, on observe comment se comportent nos grandeurs cibles, les coûts moyens et fréquences de survenance de sinistre par variable et par garantie. On montre un exemple avec l'activité professionnelle :

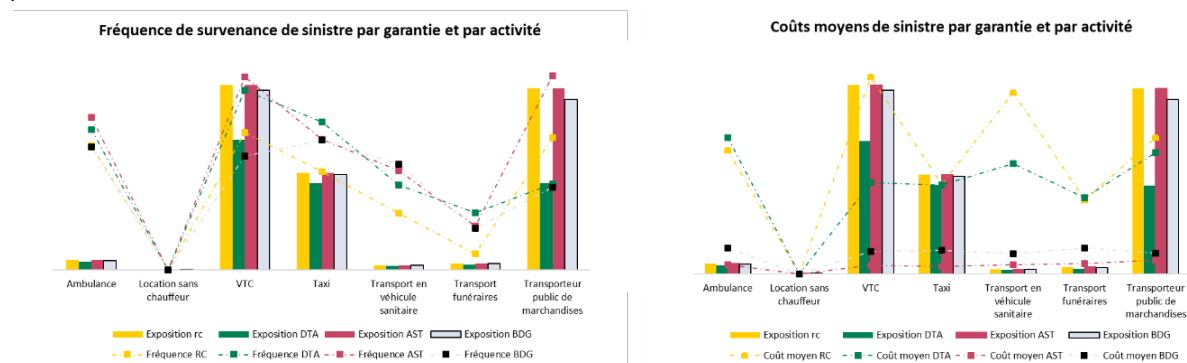




Figure 9: Fréquence et coûts moyens par activité⁶

Ces vérifications sont indispensables car elles nous permettent de vérifier que nos variables sont bien discriminantes. Sur ces graphiques, on peut en effet voir le poids en matière d'exposition, la fréquence et le coût moyen associés par garantie. Si on constate que ces grandeurs ne varient pas en fonction des valeurs prises par une variable, on peut alors éliminer cette dernière car elle n'a a priori pas d'impact sur ce qu'on veut modéliser.

3.8. Corrélations linéaires et V de Cramer

A ce stade on a déjà éliminé un certain nombre de variables or le modèle linéaire généralisé que nous souhaitons utiliser pour établir nos modèles suppose l'indépendance des variables explicatives. Il est néanmoins difficile de vérifier totalement cette hypothèse en pratique. Pour avoir cependant une idée des liens entre nos variables on calcule deux à deux leur V de Cramer, pour les variables qualitatives et leur corrélations linéaires pour les variables quantitatives. On définit le V de Cramer entre deux variables aléatoires X et Y de la façon suivante :

$$V = \sqrt{\frac{\chi^2 / n}{\min(k - 1, r - 1)}}$$

Où χ^2 est la statistique du test d'indépendance du khi-deux entre X et Y , n le nombre d'observations et r et k respectivement le nombre de lignes et de colonnes du tableau de contingence associé au test. Le V de Cramer varie entre 0, correspondant à une totale indépendance, et 1 à une association complète. Pour ce qui est de la corrélation linéaire entre deux variables quantitatives X et Y on la définit de cette façon :

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Où n est le nombre d'observations, $(x_i)_{1 \leq i \leq n}$ et $(y_i)_{1 \leq i \leq n}$ les valeurs prises par X et Y , \bar{x} et \bar{y} les moyennes empiriques de ces séries de valeurs. Voici ce qu'on obtient :

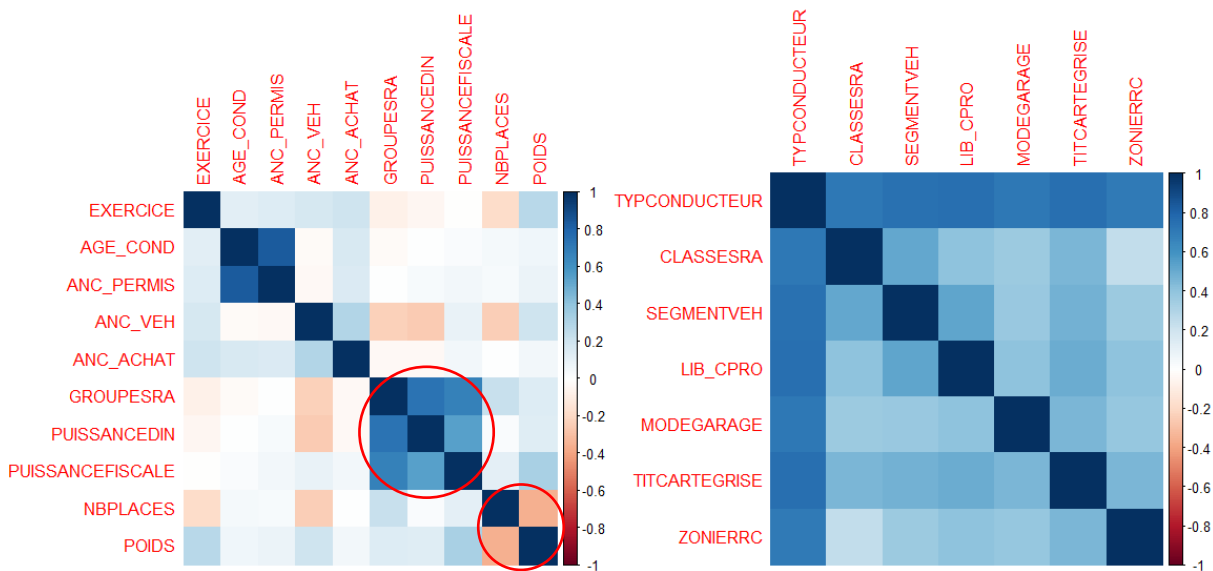


Figure 10: Corrélations et V de Cramer

Les deux graphiques ci-dessus représentent, à gauche, les corrélations entre variables numériques et à droite les V de Cramer entre variables qualitatives. Les variables concernant la puissance Din et la puissance fiscale, deux informations renseignées sur la carte grise des véhicules, sont fortement corrélées à leur groupe SRA qui note le

⁶ Pour des raisons de confidentialité, les axes des graphiques ont été volontairement supprimés



niveau de puissance et de dangerosité des véhicules. C'est une information fournie par l'association SRA (Sécurité et Réparation Automobiles), un organisme professionnel auquel adhèrent toutes les compagnies d'assurances automobiles en France. Afin de limiter le nombre de variables quantitatives dont les valeurs devront être discrétisées on conserve le groupe SRA, dont les valeurs sont des nombres entiers compris entre 20 et 80, au détriment de la puissance fiscale et de la puissance DIN qu'on retire de la base de données. De la même façon on élimine la variable « NBPLACES » qui est très corrélée à la variable « POIDS » qu'on conserve à la place. Ce sont deux informations déclarées sur la carte grise des véhicules elles indiquent le nombre de places pouvant être occupées et le poids total autorisé en charge. En ce qui concerne l'observation des V de Cramer, on constate sans surprise que la variable « tyconducteur » que nous avons décidé de maintenir dans notre base, est très liée aux autres. Elle n'a que deux modalités dont une qui est partagée par plus deux tiers des individus de notre base de données, d'où sa forte association avec les autres variables. Aussi pour rendre plus lisible le graphique on lui ôte cette variable sans toutefois la supprimer de notre base de données. On s'appuiera enfin sur ces deux graphiques pour éviter d'inclure en même temps deux variables fortement corrélées dans nos modèles comme l'âge des conducteurs et l'ancienneté de leur permis de conduire :

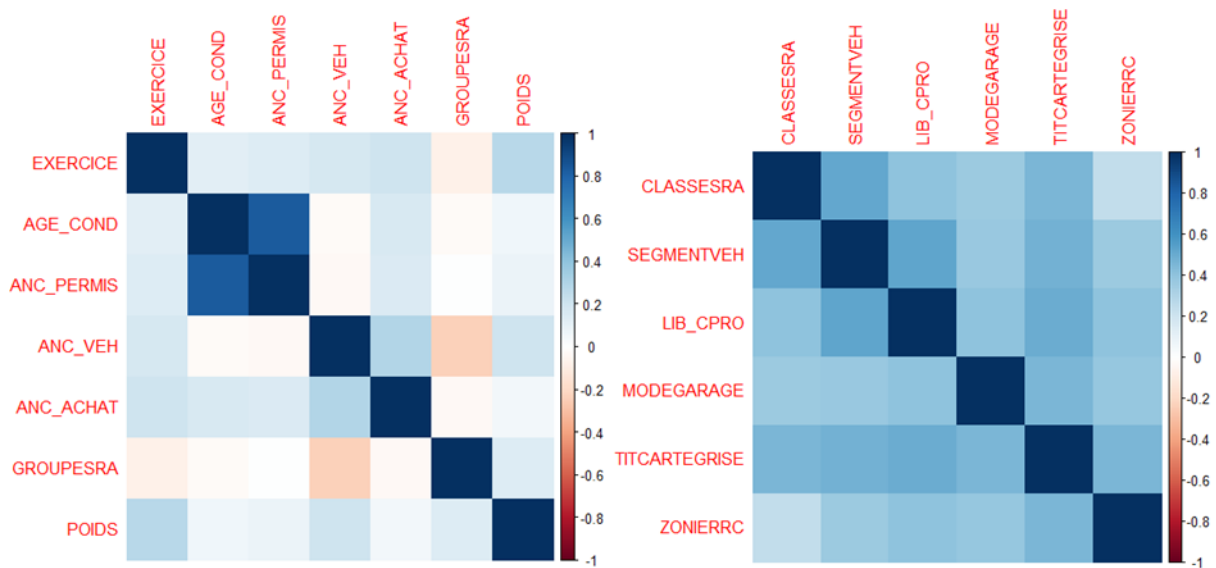


Figure 11: Corrélations et V de Cramer des variables présélectionnées

Les variables enfin retenues et leur libellé sont les suivantes :

Variable	Libellé
EXERCICE	Exercice
AGE_COND2	Age du conducteur
ANC_PERMIS2	Anciennete du permis
TYPCONDUCTEUR	Type de conducteur
CLASSESRA	Classe sra
ANC_VEH	Anciennete du vehicule
ANC_ACHAT	Anciennete achat du veh.
GROUPE_SRA	groupe sra
POIDS	Poids total autorisé en charge
SEGMENTVEH	Segment
LIB_CPRO	Libellé activite
MODEGARAGE	Mode garage
TITCARTEGRISE	Titulaire carte grise
ZONIERRC	Zonier RC

Tableau 8: Liste des variables de modélisation présélectionnées



4. Modélisation de la prime pure

On modélise pour chaque garantie les coûts moyens et fréquences de survenance en utilisant les modèles linéaires généralisés. On s'appuie pour cela sur le logiciel statistique EMBLEM, développé par Towers Watson. Il s'agit d'un outil de modélisation, utilisé pour la tarification et permettant d'adapter et d'ajuster rapidement des modèles linéaires généralisés sur des bases de données volumineuses. Pour chaque modèle, on justifiera la loi retenue pour appliquer nos GLM. On se restreint aux sinistres dont les montants n'excèdent pas le seuil de 76 000 € précédemment retenu. Dans nos modèles de fréquences on exclut ces sinistres et dans nos modèles de sévérité, on les écrête pour qu'ils n'excèdent pas le seuil fixé.

4.1. Découpage en bases d'apprentissage et de test

Afin de valider ensuite nos modèles de façon impartiale et d'éviter le surapprentissage on réalise un découpage de notre base de données en une base d'apprentissage (« train »), dont on se servira pour ajuster nos modèles, et une base de test constituée de données inconnues lors de l'ajustement des modèles et qui nous servira donc à tester et valider la qualité de nos modèles. Cette étape est indispensable pour maximiser la confiance que l'on pourra avoir en nos modèles et leurs prédictions. A l'aide de la procédure « surveiselect » sur SAS, on divise ainsi notre base de données en deux avec 80% des données dans la base d'apprentissage et les 20% restant dans la base de test. On réalise ce découpage en s'assurant que pour chaque garantie on ait environ les mêmes coûts moyens et fréquences de survenance de sinistres dans les bases d'apprentissage et de test. On représente ainsi les écarts de fréquences et de coûts moyens, par garantie, entre les bases d'apprentissage et de test :

		Train	Test	Écart	Écart %
RC	Coûts moyens	2 450 €	2 278 €	172 €	8%
	Fréquence	25%	26%	1%	3%
DTA	Coûts moyens	2 211 €	2 164 €	47 €	2%
	Fréquence	29%	31%	1%	4%
AST	Coûts moyens	245 €	231 €	14 €	6%
	Fréquence	37%	36%	1%	2%
BDG	Coûts moyens	425 €	420 €	6 €	1%
	Fréquence	22%	21%	1%	7%

Tableau 9 : Ecart de fréquence et de coûts moyens entre les bases d'apprentissage et de test

On remarque des écarts entre 1% et 8% sur certains indicateurs. On les accepte, étant donné la petite taille de notre portefeuille. Nous nous sommes également assuré que nos profils étaient répartis de la même manière sur les variables les plus significatives d'un point de vue « métier » de notre base de données comme l'activité professionnelle.

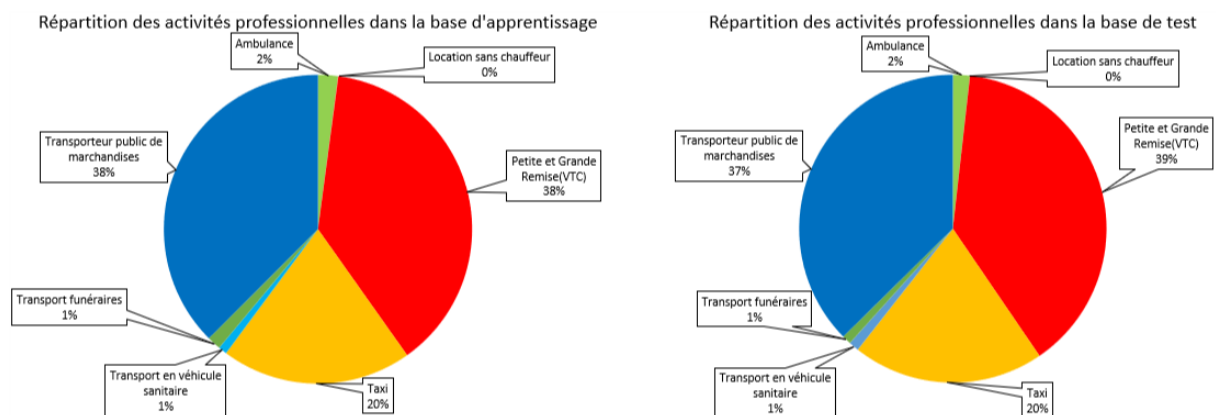


Figure 12: Répartition des activités dans les bases d'apprentissage et de test



On vérifie par ailleurs que tous les profils représentés dans la base de test sont également présents dans la base d'apprentissage, sinon on intègre ces profils à la base d'apprentissage. Sans cette précaution, on risquerait de ne pas pouvoir prédire de valeurs de fréquences et de coûts moyens pour ces profils lors de la validation sur la base de test. En présence d'un volume de données plus important, on aurait également souhaité faire un découpage qui réduise au minimum les écarts par variable et par modalité en ce qui concerne les fréquences et coûts moyens par garantie. Nous avons néanmoins pris soin de les minimiser sur les modalités les plus importantes comme pour celles de la variable « activité professionnelle » :

Ecart par activité professionnelle		VTC	Taxi	Transporteur public de marchandises	Ambulance	Location sans chauffeur	Transport en véhicule sanitaire	Transport funéraires	TOTAL	
									Ecart absolu	Ecart en %
Fréquence	RC	0%	2%	3%	9%	0%	7%	4%	1%	3%
	DTA	4%	2%	1%	8%	0%	9%	11%	1%	4%
	AST	2%	2%	1%	5%	0%	3%	7%	1%	2%
	BDG	3%	0%	0%	4%	0%	5%	3%	2%	7%
Coûts Moyens	RC	14%	19%	2%	27%	0%	12%	0%	173 €	8%
	DTA	4%	8%	5%	22%	0%	6%	72%	47 €	2%
	AST	1%	8%	9%	30%	0%	34%	41%	14 €	6%
	BDG	9%	2%	11%	28%	0%	29%	78%	5 €	1%

Tableau 10 : Ecart par garantie et par activités de fréquence et de coûts moyens entre les bases d'apprentissage et de test

Nous nous sommes surtout attachés à minimiser les écarts pour les activités les plus représentées en matière d'exposition, c'est-à-dire les taxis, les VTC et les transporteurs publics de marchandises, pour les autres activités on accepte les écarts car ils sont dus à notre faible volume de données.

4.2. Test d'adéquation à une loi

Le modèle linéaire généralisé contient plusieurs composantes telles que la loi utilisée pour la structure d'erreur et la fonction de lien. Le logiciel EMBLEM permet l'utilisation de plusieurs lois et fonctions de lien. Quel que soit le modèle on choisit la fonction logarithme comme fonction de lien car on souhaite avoir un tarif multiplicatif. En pratique on utilise souvent la loi de Poisson ou la loi binomiale négative pour les modèles de fréquences et la loi Gamma pour les modèles de sévérité. Pour les modèles de fréquences nous réaliserons quelques études, ci-dessous, nous permettant de déterminer la loi la plus appropriée entre la loi de Poisson et la loi binomiale négative. En ce qui concerne les modèles de sévérité le logiciel EMBLEM ne nous permet pas de recourir à une loi plus appropriée que la loi Gamma parmi les lois de la famille exponentielle loi nous nous restreindrons à l'usage de cette loi tout en nous assurant néanmoins de la pertinence de cette approche.

4.2.1. Modèles de fréquence

Pour déterminer la meilleure approche à considérer entre la loi de Poisson et la loi binomiale négative pour ajuster nos modèles de fréquences de survenance de sinistres, on analyse les résidus obtenus selon l'usage de chacune des deux lois lorsque l'on intègre toutes les variables présélectionnées à nos modèles. On observe les résidus des modèles de fréquences de façon groupée car il peut y avoir beaucoup d'observations sans sinistres et donc avec une fréquence nulle. Les graphiques ci-dessous montrent par exemple les résidus de déviance normalisés obtenus pour la garantie RC selon la loi sélectionnée. On définit ces résidus pour une observation i de la façon suivante :

$$r_i^* = \frac{\text{signe}(y_i - \hat{y}_i)\sqrt{d_i}}{\sqrt{1 - h_i}}$$

Où y_i est l'observation i , \hat{y}_i l'estimation correspondante, h_i le levier associé à l'observation i et d_i la contribution de l'observation i à la déviance D définie par :

$$D = -2(L - L_{sat})$$



où L est la vraisemblance du modèle et L_{sat} la vraisemblance du modèle saturé, c'est-à-dire le modèle possédant autant de paramètres que d'observations et estimant donc exactement les données.

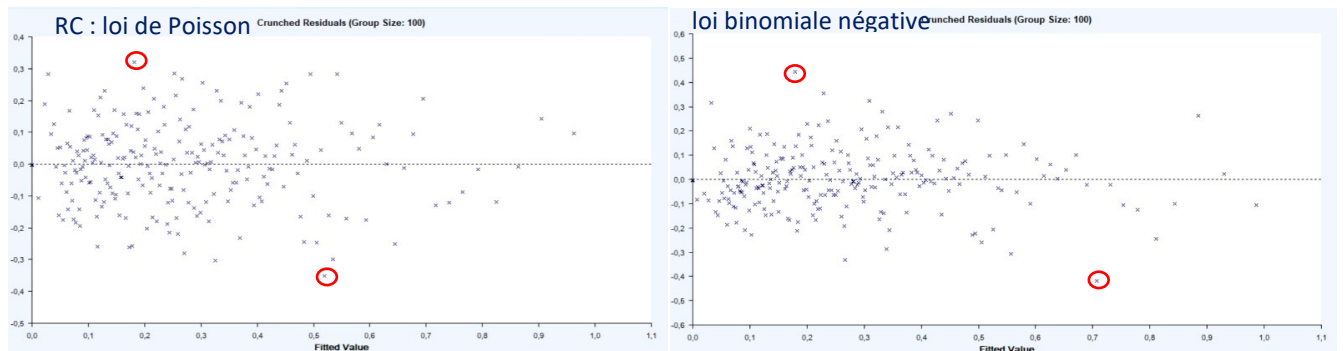


Figure 13: Modèle de fréquence et choix entre une loi de Poisson et une loi binomiale négative

On observe que les résidus sont légèrement plus regroupés autour de 0 avec la loi de Poisson, environ dans l'intervalle $[-0.35, 0.35]$, qu'avec la loi binomiale négative où les résidus sont plutôt compris dans l'intervalle $[-0.45, 0.45]$. Pour faire notre choix on compare ensuite les écarts entre observations et prédictions par exercice comme on peut le voir sur le tableau ci-dessous qui concerne aussi la garantie RC.

Loi de poisson

Exercice	Observations	Exposition	Fréquence		Ecart : pred vs obs
			observée	prédite	
2016	2 787	1159,36	32,76%	32,69%	0,21%
2017	5 134	3133,84	29,63%	29,61%	0,07%
2018	5 095	3112,04	23,82%	23,78%	0,17%
2019	4 331	2751,73	25,91%	25,91%	0,00%
2020	3 386	2273,65	18,30%	18,30%	0,00%
2021	3 239	2058,81	24,82%	24,82%	0,00%

Loi binomiale négative

Exercice	Observations	Exposition	Fréquence		Ecart : pred vs obs
			observée	prédite	
2016	2 787	1159,36	33,04%	32,69%	1,07%
2017	5 134	3133,84	29,65%	29,61%	0,14%
2018	5 095	3112,04	24,06%	23,78%	1,18%
2019	4 331	2751,73	26,50%	25,91%	2,28%
2020	3 386	2273,65	17,62%	18,30%	-3,72%
2021	3 239	2058,81	25,83%	24,82%	4,07%

Tableau 10: Modèles de fréquence - écarts de prédiction par exercice selon la loi utilisée

On procède de la même façon pour les autres garanties et voici ce qu'on obtient :

Exercice	Fréquences moyennes : écarts entre prédictions et observations par garantie							
	RC		DTA		AST		BDG	
	Poisson	Binomiale négative	Poisson	Binomiale négative	Poisson	Binomiale négative	Poisson	Binomiale négative
2016	0,21%	1,07%	0,00%	0,55%	0,00%	0,53%	0,04%	14,70%
2017	0,07%	0,14%	0,00%	1,50%	0,02%	0,55%	0,00%	0,08%
2018	0,17%	1,18%	0,00%	-1,07%	0,00%	-0,27%	0,00%	-0,38%
2019	0,00%	2,28%	0,00%	0,29%	0,00%	0,67%	0,00%	0,32%
2020	0,00%	-3,72%	0,00%	0,12%	0,00%	-1,27%	0,00%	0,97%
2021	0,00%	4,07%	0,00%	-0,09%	0,00%	2,38%	0,00%	1,51%

Garantie	Etendue des résidus	
	Poisson	Binomiale négative
RC	$[-0.35, 0.35]$	$[-0.45, 0.45]$
DTA	$[-0.30, 0.40]$	$[-0.30, 0.35]$
AST	$[-0.40, 0.50]$	$[-0.40, 0.60]$
BDG	$[-0.30, 0.35]$	$[-0.35, 0.35]$

Tableau 11: Modèles de fréquence - écarts de prédiction par exercice et par garantie selon la loi utilisée

On constate que lorsqu'on utilise la loi de Poisson, on obtient à chaque fois des écarts moins importants entre prédictions et observations. Ces résultats nous incitent donc à préférer cette loi pour ajuster nos modèles de fréquences.

4.2.2. Modèles de sévérité

Pour ajuster un modèle de sévérité hors graves à l'aide des modèles linéaires généralisés, il est courant d'utiliser une loi gamma ou une loi log-normale, ou plutôt un modèle gaussien sur les logarithmes des coûts car la loi log-normale ne fait pas partie de la famille exponentielle. Nous avons fait le choix de nous tourner vers une loi Gamma avec une fonction de lien logarithmique, bien que la fonction de lien canonique de la loi Gamma soit la



fonction inverse, car on souhaite obtenir une structure multiplicative de coefficients. Pour valider la pertinence de cette approche nous avons tout d'abord effectué quelques contrôles. Par garantie on calcule par la méthode du maximum de vraisemblance les estimateurs des paramètres de la loi gamma qui correspondent le plus possible, selon cette méthode, à la distribution de nos montants de sinistres. On observe ensuite les résultats graphiquement comme on peut le voir ci-dessous pour la garantie RC :

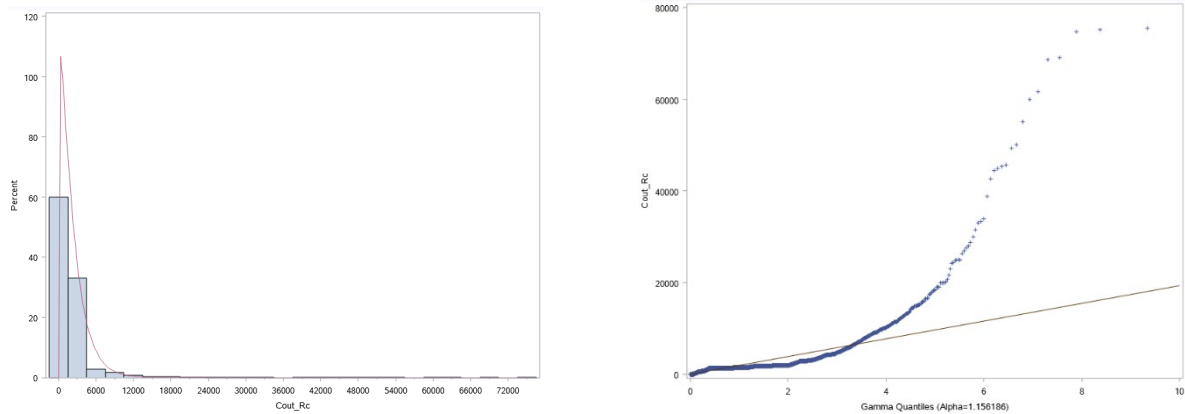


Figure 14: Modèle de sévérité RC - QQ-Plot de l'ajustement d'une loi gamma

Le qq-plot est d'abord linéaire puis diverge à cause du manque de données pour les montants plus élevés. Cela coïncide avec l'histogramme des charges RC très ramassé à gauche. Pour d'autres garanties, comme la DTA, ou le spread des montants est plus large les qq-plot sont plus linéaires :

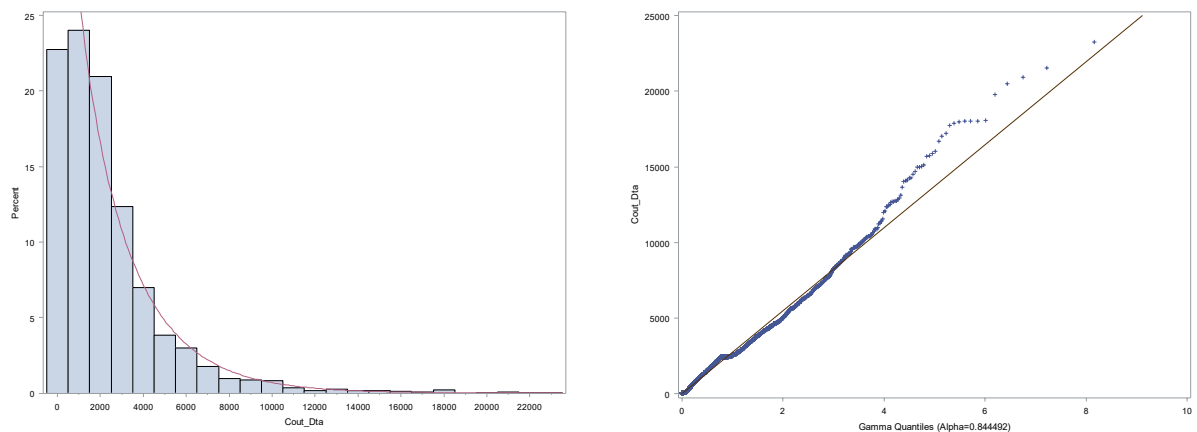


Figure 15 : Modèle de sévérité DTA - QQ-Plot de l'ajustement d'une loi gamma

5. Sélection de modèle

On décrit ici notre méthode de sélection de modèle en l'illustrant avec la garantie RC. Nous nous restreignons à la description de notre démarche avec cette garantie afin d'éviter d'alourdir notre exposé.

5.1. Procédure stepwise

5.1.1. Critères AIC et BIC

Ayant choisi nos structures d'erreurs on peut désormais commencer à modéliser nos fréquences et coûts moyens de sinistres par garantie. Pour décider des variables à inclure dans nos modèles on réalise une procédure stepwise forward selon les critères de l'AIC et du BIC. Cette procédure consiste à inclure au fur et à mesure chaque variable en vérifiant si elle améliore ou au contraire détériore l'AIC ou le BIC du modèle. Une sélection selon ces critères vise à obtenir des modèles à la fois les plus vraisemblables et les plus parcimonieux possible.



Par modèle on ajoute ainsi une par une les variables en s'assurant à chaque fois, à l'aide de nos études préliminaires, qu'on n'ait jamais en même temps deux variables corrélées ensemble. Si une variable n'améliore pas significativement nos critères en les diminuant on ne la retient pas. On a peu de données et on veut en effet éviter de surcharger nos modèles en paramètres et ainsi détériorer la qualité de leur ajustement ou introduire du surapprentissage. La première variable que l'on ajoute à tous nos modèles est l' « Exercice », qui comme on le rappelle est l'année relative à chaque observation. En ajoutant cette variable l'objectif est de pouvoir tester ensuite la robustesse de nos modèles au cours du temps. Certaines variables comme les variables continues ont des modalités qui comptent peu d'observations dans notre base de données. Lorsqu'on exécute notre procédure stepwise on effectue ainsi divers regroupements de modalités. Par exemple dans notre modèle de fréquence RC, on ajoute la variable « ancienneté du permis » sans retraitement et on constate qu'elle détériore dans un premier temps l'AIC et le BIC du modèle :

Modèle de fréquence RC	AIC	BIC
Sans l'ancienneté du permis	12930	12978
Avec l'ancienneté du permis	12974	13499

Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
Ancienneté du permis (0)	2,6826	0,74629	27,8		0	0,0	14,6224
Ancienneté du permis (1)	0,5439	0,72697	133,6		6	0,0	1,7228
Ancienneté du permis (2)	0,3079	0,45410	147,5		20	0,1	1,3606
Ancienneté du permis (3)	0,4269	0,26881	63,0		57	0,4	1,5325
Ancienneté du permis (4)	0,1630	0,25056	153,7		87	0,6	1,1771
Ancienneté du permis (5)	0,1613	0,20114	124,7		159	1,1	1,1750
Ancienneté du permis (6)	0,1712	0,18344	107,1		213	1,5	1,1868
Ancienneté du permis (7)	0,2326	0,17471	75,1		237	1,6	1,2618
Ancienneté du permis (8)	0,0346	0,17770	513,2		275	1,9	1,0352
Ancienneté du permis (9)	0,0990	0,16693	168,6		321	2,2	1,1041
Ancienneté du permis (10)	-0,0450	0,17178	382,0		337	2,3	0,9560
Ancienneté du permis (11)	-0,0081	0,16167	1 993,2		406	2,8	0,9919
Ancienneté du permis (12)					430	3,0	
Ancienneté du permis (13)	-0,1317	0,16925	128,5		388	2,7	0,8786
Ancienneté du permis (14)	-0,1252	0,17022	135,9		380	2,6	0,8823
Ancienneté du permis (15)	-0,1643	0,17576	107,0		352	2,4	0,8485
Ancienneté du permis (16)	-0,1714	0,17519	102,2		359	2,5	0,8425
Ancienneté du permis (17)	-0,1620	0,17230	106,4		379	2,6	0,8504
Ancienneté du permis (18)	-0,2589	0,18035	69,7		358	2,5	0,7719
Ancienneté du permis (19)	-0,3088	0,18406	59,6		352	2,4	0,7344
Ancienneté du permis (20)	-0,4648	0,19726	42,4		335	2,3	0,6283
Ancienneté du permis (21)	-0,0480	0,17835	371,6		307	2,1	0,9531
Ancienneté du permis (22)	-0,3795	0,19848	52,3		301	2,1	0,6842
Ancienneté du permis (23)	-0,1713	0,19616	114,5		257	1,8	0,8426
Ancienneté du permis (24)	-0,5488	0,22881	41,7		228	1,6	0,5778
Ancienneté du permis (25)	-0,3908	0,22458	57,5		214	1,5	0,6765
Ancienneté du permis (26)	-0,1906	0,20333	106,7		233	1,6	0,8264
Ancienneté du permis (27)	-0,6043	0,24409	40,4		218	1,5	0,5464

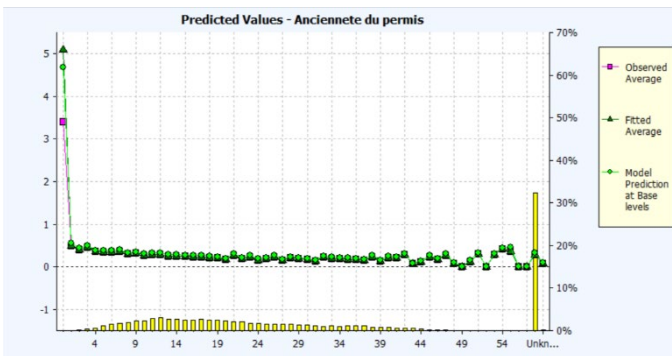
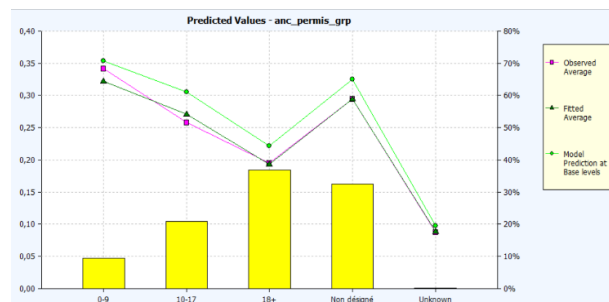


Figure 16: Modèle de fréquence RC - ajout de la variable « ancienneté du permis » avant retraitement

Ce graphique illustre les observations et prédictions moyennes, on remarque qu'à ce stade on prédit assez bien en moyenne malgré le grand nombre de modalités et le peu de poids sur chacune d'elles. On peut d'ailleurs suspecter la présence de surapprentissage. Il nous suffit de consulter les erreurs types de prédiction et le poids de chaque modalité, comme on peut le voir sur le tableau ci-dessus pour quelques valeurs, pour nous rendre compte de la mauvaise qualité de cet ajustement. Nous allons ainsi essayer d'améliorer l'AIC et le BIC du modèle et réduire les erreurs types de prédictions en effectuant des regroupements de modalités :



Modèle de fréquence RC	AIC	BIC
Sans l'ancienneté du permis	12930	12978
Avec l'ancienneté du permis	12974	13499
Avec l'ancienneté du permis retraitée	12878	12950

Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
anc_permiss_grp (0-9)	0,2527	0,07277	28,8		1 375	9,5	1,2875
anc_permiss_grp (10-17)					3 030	20,9	
anc_permiss_grp (18+)	-0,2636	0,05901	22,4		5 352	36,9	0,7683
anc_permiss_grp (Non désigné)	0,1103	0,05595	50,7		4 698	32,4	1,1186
anc_permiss_grp (Unknown)	-1,0933	0,69242	63,3		34	0,2	0,3351

Figure 17: Modèle de fréquence RC - ajout de la variable « ancienneté du permis » après retraitement

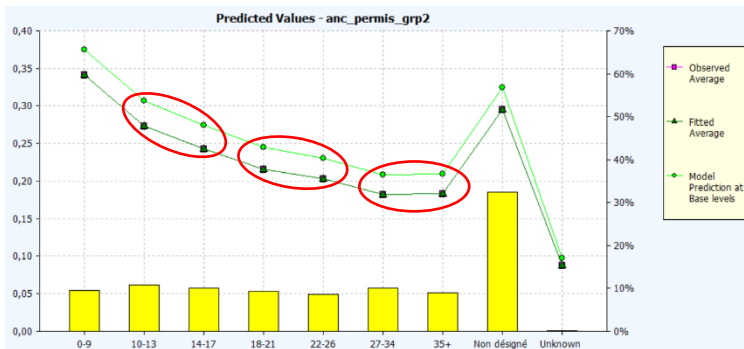
On a regroupé les modalités qui avaient des valeurs observées similaires et on améliore ainsi significativement l'AIC. Ensuite, sur le graphique on constate qu'en moyenne on prédit proche de l'observé et de plus on a considérablement réduit les erreurs types de prédiction. En établissant ce regroupement nous nous sommes par ailleurs assurés qu'il respectait la tendance de nos observations avec une décroissance de la fréquence en fonction de l'ancienneté du permis des conducteurs désignés. Bien entendu nous avons laissé la modalité « Non désigné » à part pour bien prendre en compte dans notre modèle le caractère désigné ou non du conducteur. Ce retraitement nous a permis d'améliorer l'AIC et le BIC de notre modèle, on a contrôlé que le modèle s'ajustait bien en moyenne et avec de faibles erreurs types de prédiction, on choisit donc d'inclure cette variable retraitée



à notre modèle. On poursuit ensuite notre procédure stepwise de la même façon pour chaque variable et en effectuant toujours les mêmes contrôles.

5.1.2. Regroupements de modalités et tests de sensibilité

Pour faire les retraitements que nous avons décrits dans la partie précédente, nous avons testé des regroupements successifs de modalités. Par exemple, toujours dans le cas de l'ancienneté du permis et du modèle de fréquence RC, ce regroupement donne des prédictions moyennes qui se superposent bien aux observations et améliore significativement l'AIC :



Modèle de fréquence RC : avec sous sans l'ancienneté du permis	AIC	BIC
Sans l'ancienneté du permis	12930	12978
Avec l'ancienneté du permis	12974	13499
Avec un retraitement intermédiaire	12881	12994
Avec le retraitement final	12878	12950

Figure 18: Modèle de fréquence RC – regroupement intermédiaire de la variable « ancienneté du permis »

On améliore aussi significativement les erreurs types de prédictions :

Name	Value	Standard Error	Standard Error (%)	Alias Indicator (%)	Weight	Weight (%)	Exp(Value)
Mean	-1,2985	0,06054	4,7		14 489	100,0	0,2729
anc_permis_grp2 (0-9)	0,2251	0,08356	37,1		1 375	9,5	1,2525
anc_permis_grp2 (10-13)					1 561	10,8	
anc_permis_grp2 (14-17)	-0,1163	0,08965	77,1		1 469	10,1	0,8902
anc_permis_grp2 (18-21)	-0,2372	0,09502	40,1		1 352	9,3	0,7888
anc_permis_grp2 (22-26)	-0,2973	0,09929	33,4		1 243	8,6	0,7428
anc_permis_grp2 (27-34)	-0,4048	0,09760	24,1		1 461	10,1	0,6671
anc_permis_grp2 (35+)	-0,4013	0,10121	25,2		1 297	9,0	0,6695
anc_permis_grp2 (Non désigné)	0,0771	0,06922	89,8		4 698	32,4	1,0801
anc_permis_grp2 (Unknown)	-1,1300	0,69814	61,8		34	0,2	0,3230

Tableau 12 : Erreurs de prédiction après regroupement intermédiaire des modalités d'ancienneté du permis

Certaines modalités importantes, comme la modalité « Non désigné », présentent toutefois des erreurs types de prédictions importantes. Et par ailleurs, malgré ce premier regroupement, le nombre de modalités demeure toujours trop important par rapport à notre volume de données. Le graphique des prédictions et observations moyennes nous montre du reste que plusieurs de nos nouvelles modalités présentent des valeurs similaires. Ce sont ces constatations qui nous permettent d'aboutir à notre regroupement final qui améliore sensiblement l'AIC du modèle de départ, qui s'ajuste bien en moyenne et qui réduit grandement les erreurs types de prédictions. Bien entendu, chaque fois que nous avons établi un regroupement, nous avons pris soin de n'associer que des modalités avec des profils similaires. D'autant plus si deux modalités n'ont pas le même poids dans la base de données et qu'elles correspondent à des observations de la variable cible totalement différentes. Les regrouper nous conduirait alors à un modèle qu'on pourrait peut-être valider statistiquement mais qui poserait, en pratique, des problèmes de sens et d'interprétabilité. On peut s'en rendre compte ci-dessous :

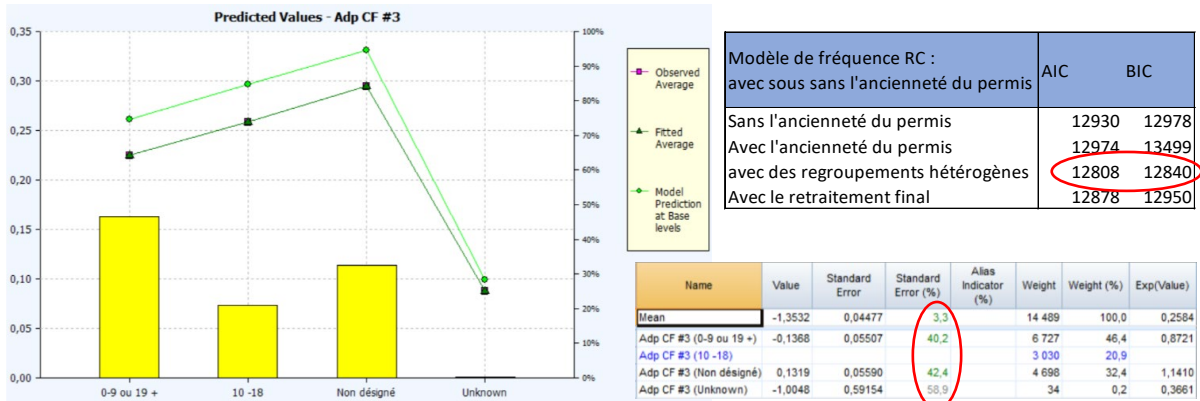


Figure 19: Modèle de fréquence RC – regroupement alternatif de la variable « ancienneté du permis »

On a regroupé ensemble les conducteurs ayant moins de 9 ans d'ancienneté de permis et ceux en ayant au moins 19. On se rend compte que d'un point de vue théorique on pourrait conserver ce regroupement car on a très significativement amélioré l'AIC et le BIC, bien plus qu'avec le regroupement retenu, et que notre modèle s'ajuste très bien en moyenne avec de faibles erreurs types de prédiction. Or en associant des conducteurs très expérimentés et des conducteurs dont l'obtention du permis est bien plus récente, on met dans le même panier des profils de risque très différents. Cela pénalise à la fois les profils moins à risque et au contraire, dans ce cas précis, avantage considérablement les profils les plus risqués. En moyenne, sur un grand volume de données, on prédirait bien mais ligne à ligne nos prédictions auraient peu de sens. Ainsi, chaque variable regroupée qui figure dans nos modèles définitifs est issue de tests successifs de sensibilité à la qualité de l'ajustement de nos modèles. Ces tests sont indispensables car ils permettent d'une part d'évaluer la robustesse de nos modèles et d'autre part ils nous permettent de renforcer notre confiance en leur prédictibilité.

5.2. Critère d'équité du tarif

On vérifie ensuite que les coefficients de nos variables ne présentent pas d'écarts significatifs entre elles. Cela pourrait engendrer un problème d'équité entre les clients. En effet notre modèle étant multiplicatif, une grande amplitude des coefficients sur plusieurs variables pourrait grandement avantager ou pénaliser des clients qui auraient des profils cumulant ces modalités aggravantes. On étudie ainsi par modèle, l'étendue des coefficients de nos variables ainsi que leur maximum et leur minimum.

Coefficients du modèle de fréquence RC	min	max	mediane	étendue	écart-type
Ancienneté du permis	0,86	1,25	1,1	0,39	0,18
Ancienneté achat du véhicule	0,8	1	0,9	0,2	0,14
Zonier RC	0,57	1	0,82	0,43	0,19
Mode garage	1	1,27	1,11	0,27	0,14
Groupe SRA	0,69	1	0,96	0,31	0,16

Tableau 13: Etendue des coefficients du modèle de fréquence RC

Comme on peut le voir sur l'exemple ci-dessus de notre modèle de fréquence RC, nous avons bien pris soin de ne pas avoir d'écarts significatifs entre nos coefficients. Le modèle ainsi construit respecte bien ce principe d'équité. Ce processus a été effectué sur l'ensemble des sous-modèles de chaque garantie.

5.3. Choix entre plusieurs modèles et calcul de RMSE

En suivant toutes les étapes et contrôles que nous avons décrits dans les parties précédentes, il peut arriver qu'on hésite entre plusieurs modèles. C'est notamment le cas lorsque deux variables, prises séparément, améliorent à peu près autant l'AIC et le BIC du modèle. Dans ce cas on départage ces modèles en calculant respectivement leur « RMSE » pour « Root Mean Square Error », en français la racine carrée de l'erreur quadratique moyenne. On la définit de la façon suivante :



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

Où n est le nombre d'observations de notre base de données, $(\hat{y}_i)_{1 \leq i \leq n}$ les prédictions et $(y_i)_{1 \leq i \leq n}$ les observations. C'est une mesure de la qualité prédictive d'un modèle, plus la RMSE est petite plus le modèle est prédictif. On calcule ainsi sur la base test la RMSE de chacun de nos modèles candidats pour les départager :

Modèle de fréquence RC	Variabes	AIC	BIC	RMSE
Retenu	Exercice, Ancienneté du permis, Ancienneté achat du véhicule, Zonier RC, Modegarage, Groupe SRA	12 840,37	12 985,89	12,2307
Alternatif	Exercice, Ancienneté du permis, Ancienneté achat du véhicule, activité	12 803,57	12 908,67	12,2319

Tableau 14 : RMSE par modèle

Les deux modèles ont des RMSE très proches, il y a également peu de différences entre leur AIC et leur BIC. Néanmoins, le modèle retenu prend en compte le zonier qui est une variable que l'on a jugée importante d'un point de vue métier. L'activité, présente dans le second modèle et qui est également une variable très importante dans tous nos modèles, est déjà présente dans notre modèle de sévérité RC (cf. annexe 2), elle est donc de toute façon prise en compte dans la prime pure si l'on retient le premier modèle.

5.4. Robustesse dans le temps

Comme indiqué précédemment, nous vérifions que les estimations par année pour les variables incluses dans les modèles sont cohérentes dans le temps, c'est-à-dire qu'elles suivent les mêmes tendances d'année en année comme avec l'activité dans notre modèle de sévérité RC :

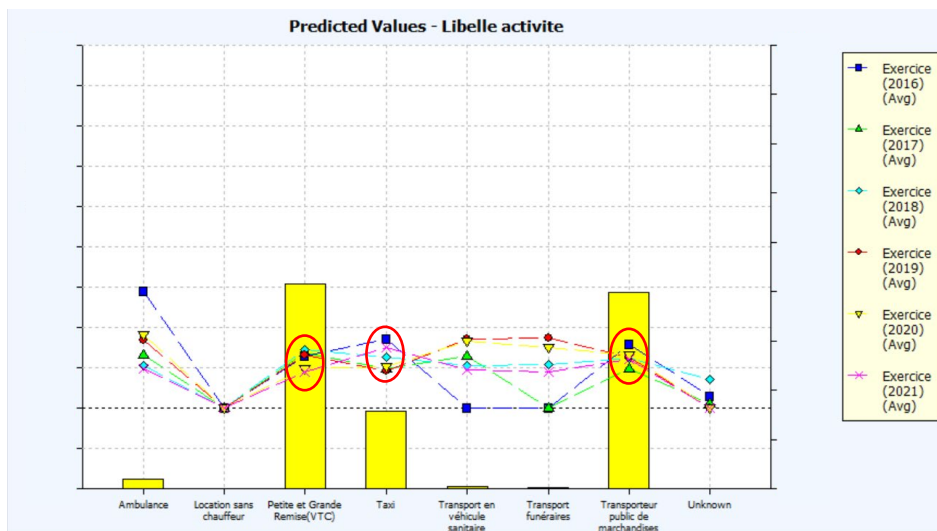


Figure 20: Estimations des modalités de la variable activité par exercice⁷

On remarque qu'on retrouve bien les mêmes tendances quelle que soit l'année, en particulier pour les activités les plus représentées de notre base de données comme les taxis, les VTC et les transporteurs publics de marchandises. Nous faisons cette vérification tout au long de notre procédure stepwise et éliminons les variables que ne sont pas robustes dans le temps.

⁷ Pour des raisons de confidentialité les valeurs des axes du graphiques ont été supprimées



5.5. Validation

Une fois que nous avons établi nos modèles en effectuant tous les contrôles décrits précédemment, on procède à quelques dernières étapes de validation.

5.5.1. Analyse des résidus

La première de ces étapes est l'analyse des résidus. On observe ainsi les graphiques des résidus de déviance normalisés. On illustre cela avec la garantie RC :

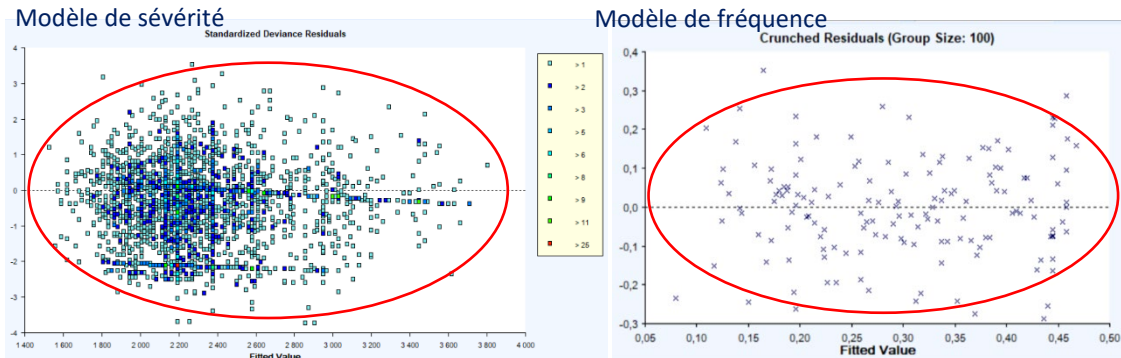


Figure 21: Résidus des modèles de sévérité et de fréquence RC

Les résidus pour les modèles de sévérité et de fréquences présentent une absence significative de tendance ou forme particulière avec une équidispersion des points autour de l'ordonnée 0. Les points sont répartis symétriquement dans une ellipse. On peut ensuite s'intéresser à la dispersion des résidus par variable, notamment pour les variables les plus significatives d'un point de vue métier comme l'activité professionnelle :

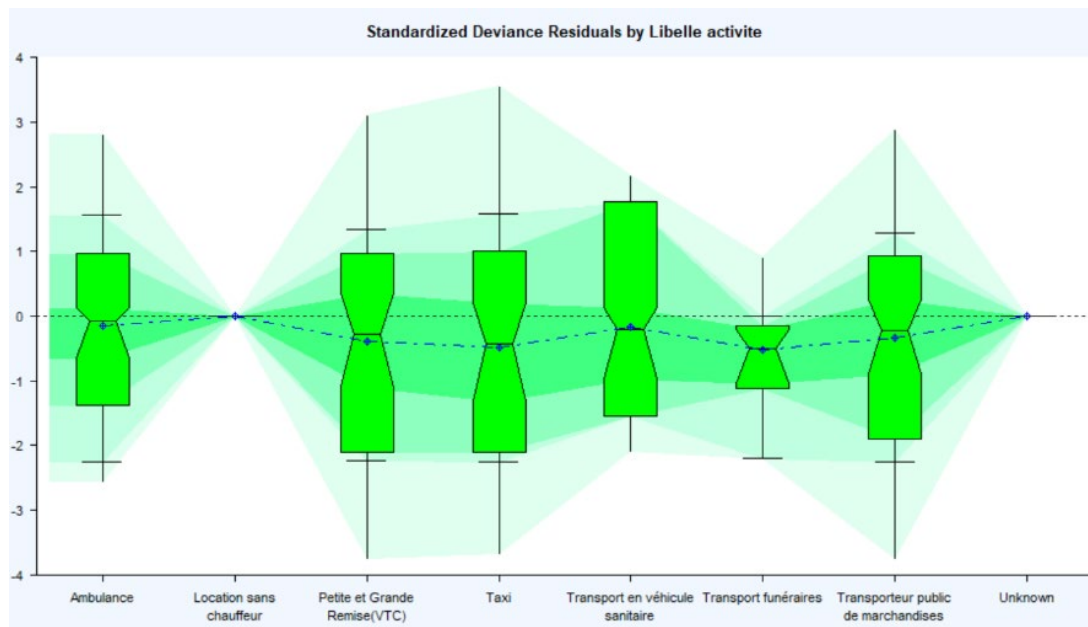


Figure 22: Etendue des résidus par activité

On s'intéresse tout particulièrement aux activités les plus représentées dans notre base de données, les taxis, les VTC et les transporteurs publics de marchandises. Leurs résidus sont approximativement centrés autour de zéro et compris au maximum entre -2 et 2, on observe également des résultats satisfaisants pour les activités plus minoritaires de notre portefeuille. On observe ensuite les distances de Cook de nos données. La distance de Cook



mesure l'effet de la suppression d'une donnée. Celles avec d'importants résidus ou un fort effet de levier pourraient fausser les résultats et la précision de nos modèles. On définit la distance de Cook par :

$$D_i = \frac{1}{p} (\beta - \beta_{(i)})' (XWX)^{-1} (\beta - \beta_{(i)})$$

Où β contient les paramètres estimés avec toutes les observations, $\beta_{(i)}$ sans l'observation i , p le nombre de paramètres du modèle, X contient les variables explicatives et W les poids de nos modèles, l'exposition pour nos modèles de fréquence et le nombre de sinistres dans nos modèles de sévérité. Ainsi plus la distance de Cook d'une observation est élevée plus elle influence l'estimation de nos paramètres. Pour chaque modèle on s'assure qu'aucune de nos observations n'a une distance de Cook trop importante car on ne souhaite pas la présence d'observations de cette nature. On ne veut pas biaiser nos modèles. Dans notre cas on jugera une distance de Cook trop élevée si elle excède 1, ce seuil est très répandu dans la littérature. On représente ainsi à titre d'exemple les graphiques des distances de Cook de nos modèles de sévérité. Nous avons procédé aux mêmes observations pour les modèles de fréquences.

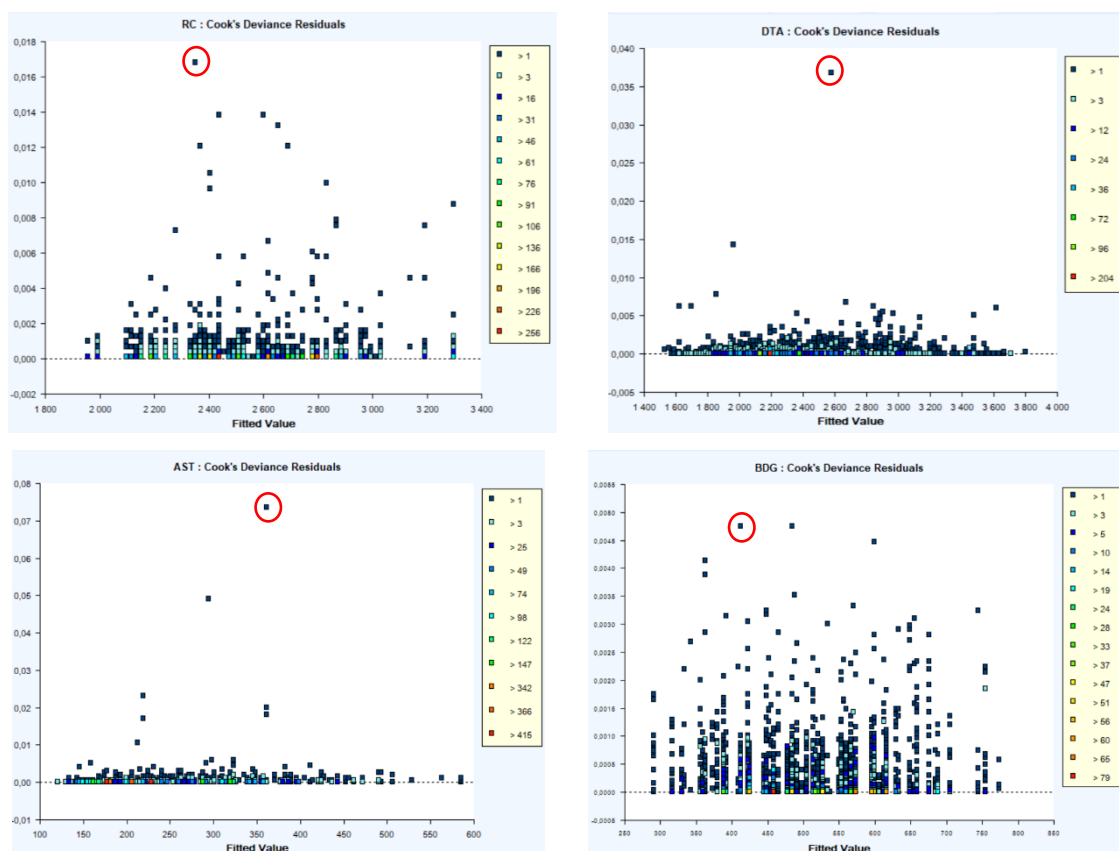


Figure 23: Distances de Cook des modèles de sévérité

On observe que quelques points sont éloignés des autres, sans être non plus aberrants dans la mesure ou leur distance de Cook est, quelle que soit la garantie, assez éloignée du seuil théorique que nous avons retenu. L'objectif de ces analyses est de valider le pouvoir explicatif de nos modèles. Elles nous permettent d'être en mesure de détecter d'éventuelles défaillances. Si par exemple on avait observé une structure particulière des résidus, cela nous aurait alerté sur la présence d'une information présente dans les résidus et dont on n'aurait pas tenu compte dans notre modélisation. De même des individus avec de grandes distances de Cook, supérieures à 1, nous auraient alerté sur la présence d'individus atypiques dans nos données. Dans ce cas nous aurions réalisé des études complémentaires pour comprendre leur nature et juger de la pertinence de les conserver dans le périmètre de notre modélisation.



5.5.2. Validation sur la base test

Après avoir analysé les résidus, la prochaine étape de notre validation est l'application de nos modèles à la base de données de test. On veut vérifier pour chaque garantie que nos modèles s'ajustent également de façon satisfaisante à cette base de données constituées de données inconnues lors de leur ajustement. On calcule ainsi les écarts entre prédictions et observations moyennes par exercice pour vérifier une nouvelle fois la robustesse de nos modèles au cours du temps. On représente ainsi ci-dessous les écarts moyens obtenus entre les prédictions et observations sur les bases d'apprentissage (Train) et de test :

Ecart entre prédiction et observations		RC	DTA	AST	BDG
Test	Fréquences	0,15%	-0,04%	0,07%	-1,74%
	Coûts moyens	0,34%	-3,74%	1,45%	0,61%
Train	Fréquences	0,00%	0,01%	0,00%	0,02%
	Coûts moyens	0,31%	-0,18%	0,03%	-0,04%

Tableau 15 : Ecart de prédictions par garantie pour les bases d'apprentissage et de test

Nous nous sommes ainsi assurés qu'en moyenne, par modèle, il n'y avait pas d'écart trop important entre observations et prédictions.

5.6. Combinaison de modèles et lissage des coefficients

Une fois qu'on a validé nos modèles de fréquences et de sévérité par garantie, on peut enfin les combiner pour obtenir des modèles de prime pure hors sinistres graves. A l'aide du logiciel Radar développé comme Emblem par l'entreprise Towers Watson, on combine nos modèles et on observe dans un premier temps l'écart moyen entre les prédictions et les observations par garantie. On prête ensuite attention aux écarts par variable, en particulier les plus significatives d'un point de vue métier, et on vérifie que les modalités les plus représentées sont bien estimées. On montre un exemple avec la garantie RC :

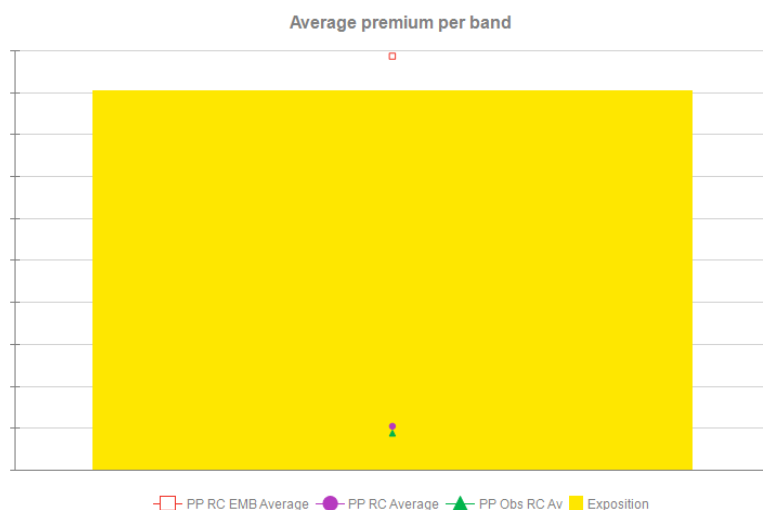


Figure 24 : Ecart moyen entre prédictions et observations de primes pures⁸

En légende « PP RC EMB Average » concerne la prime pure issue de notre modèle, « PP RC Average » est la prime pure après lissage et enfin « PP Obs RC Av » la prime pure observée. L'écart entre la prime pure moyenne estimée et celle observée est d'environ 45 €. On a ainsi réduit cet écart en procédant à un lissage des coefficients par variable en ajustant manuellement les coefficients des modalités les moins bien estimées car très peu représentées. On a surtout agi sur les modalités ayant très peu de poids dans notre base de données. Avec un volume de données plus conséquent ces modalités auraient pu être mieux prédites. Mais dans notre situation,

⁸ Pour des raisons de confidentialité les valeurs des axes ont été supprimées.



avec notre petit portefeuille, nous sommes contraints d'ajuster nos prédictions pour ces modalités en fonction de l'observé. Au contraire on ne retouche pas les variables et les coefficients lorsque le modèle s'ajuste bien. A titre d'exemple avec la variable « activité », on a à peine retouché les coefficients concernant les modalités les plus représentatives (les VTC, les taxis et les transporteurs publics de marchandises). Cependant on a ramené le niveau des ambulances et des activités de transport en véhicule sanitaire au niveau de l'observé :

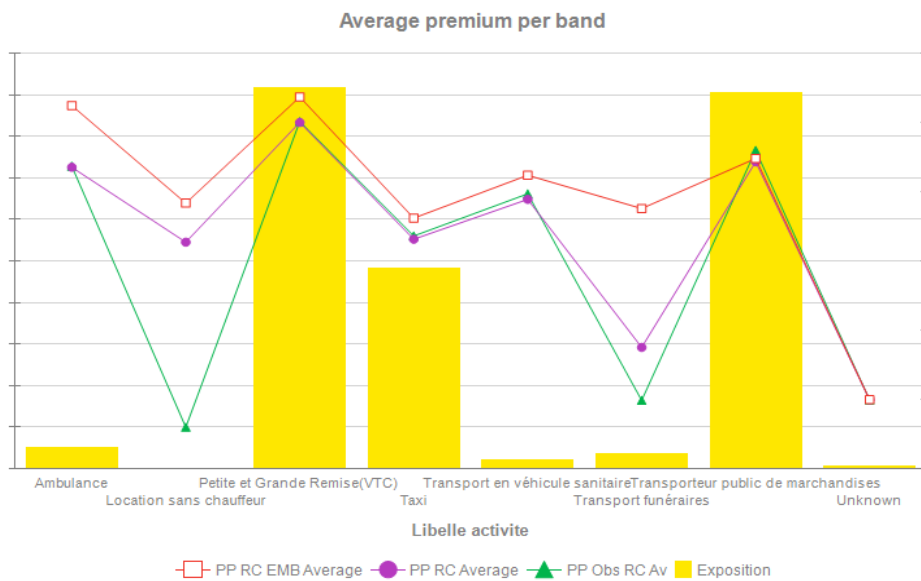


Figure 25: Prime pure par activité⁹

Sur les variables continues telles que le groupe SRA, on a seulement vérifié que sur les modalités où on a du volume, le modèle s'ajuste bien à l'observé :

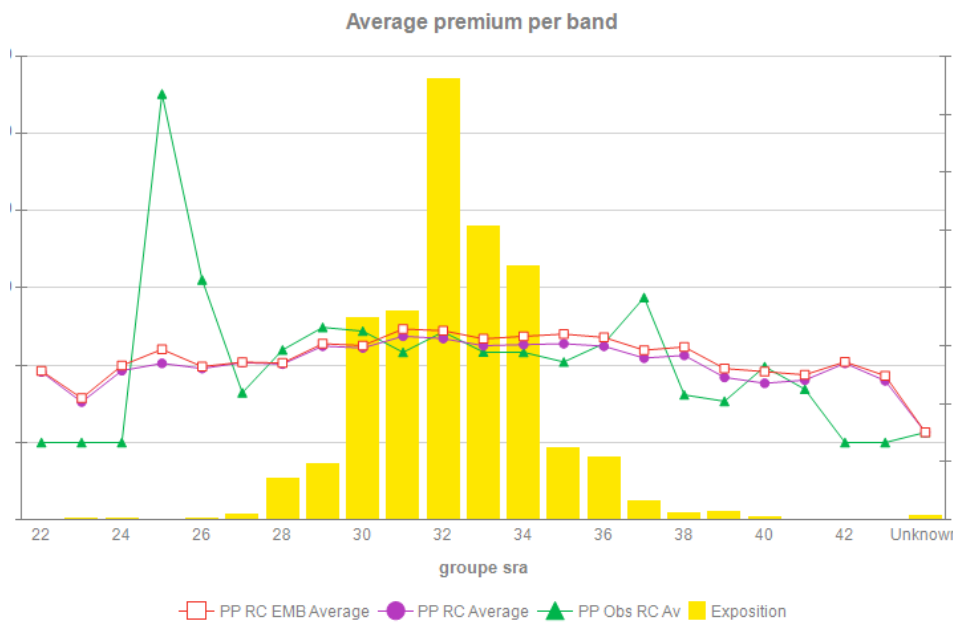


Figure 26: Prime pure par groupe SRA

⁹ Pour des raisons de confidentialité les valeurs des axes des figures 26 et 27 ont été supprimées.



On vérifie enfin, qu'au global, par exercice, on n'a pas trop dénaturé le comportement du modèle :



Figure 27: Prime pure par exercice¹⁰

On remarque, par ailleurs, que l'année 2016 dans un contexte de lancement du produit et avec une faible population est plus dégradée. A l'inverse on observe une très faible sinistralité en 2020 due au confinement. Le fait d'ajuster les coefficients des modalités de faible exposition au niveau de l'observé ne remet pas en cause la prédictivité du modèle sur ce qui fait le cœur du portefeuille. Pour ces modalités avec très peu de population on n'est ainsi pas moins prédictif que le modèle dont on peut légitimement douter de la robustesse des coefficients.

6. Choix de la méthode de prise en compte des sinistres graves

Maintenant que nous avons achevé notre modélisation de prime pure hors sinistres grave il nous reste à définir une méthode pour prendre en compte la sinistralité grave dans notre calcul final de prime pure. Pour cela nous rencontrons plusieurs difficultés, la première, toujours la même, nous n'avons à disposition qu'une petite population, peu de données et donc peu d'historique de sinistres. Ce qui obscurcit déjà grandement nos perspectives de modélisation des sinistres graves. L'autre difficulté importante que nous rencontrons est la présence d'un sinistre RC exceptionnel d'un montant dépassant 10 000 000 € et qui représente à lui seul 41% de la charge RC et 28% de la charge totale observée sur l'ensemble de l'historique de sinistre de notre périmètre. Nous nous sommes interrogés sur comment prendre en compte ce sinistre dont nous ne connaissons pas la période de retour et qui pourrait nous conduire à surestimer la sinistralité grave voire à la sous-estimer si on décidait d'exclure ce sinistre de notre étude. Ainsi pour faire face à ces difficultés nous tenterons plusieurs approches que nous comparerons pour prendre en compte notre sinistralité grave. Nos deux premières approches suivent ce schéma :

¹⁰ Pour des raisons de confidentialité les valeurs des axes ont été supprimées.

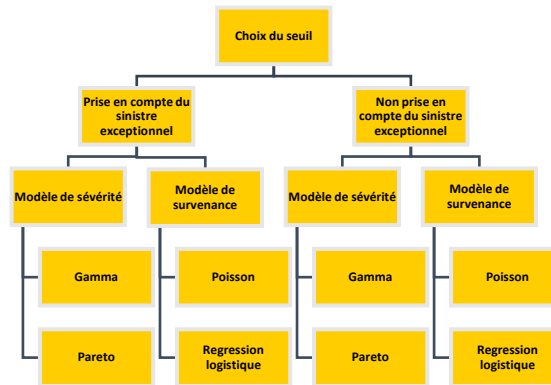


Figure 28 : Schéma des méthodes d'estimation de la sinistralité grave

On détermine un seuil de sinistre grave puis on décide ou non d'inclure au périmètre de l'étude notre sinistre exceptionnel. Ensuite pour nos modèles de sévérité, on réalise un modèle GLM, avec une fonction de lien logarithme et une distribution Gamma, et d'autre part on estime le coût moyen de nos sinistres graves en les approchant par une loi de Pareto généralisée dont on aura estimé les paramètres par la méthode du maximum de vraisemblance. On détermine ensuite l'espérance de cette loi pour obtenir un estimateur du coût moyen de nos sinistres au-delà du seuil considéré. En ce qui concerne les modèles de survénance on teste à la fois un modèle GLM, avec une fonction de lien logarithme et une distribution de Poisson, et un modèle de propension à partir d'une régression logistique. L'objectif d'une régression logistique est de prédire une variable binaire, c'est un cas particulier du modèle linéaire généralisé où la loi utilisée est une loi de Bernoulli et où la fonction de lien est la fonction logit définie sur $]0,1[$ et à valeurs dans \mathbb{R} par :

$$\text{logit} : p \mapsto \text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

Ce modèle nous permet d'estimer la probabilité p_{graves} qu'un sinistre excède le seuil considéré sachant qu'il est survenu. Dans ce cas on modélise la survénance d'un sinistre grave en calculant :

$$Freq * p_{graves}$$

Où $Freq$ est la fréquence issue de notre modèle de fréquence. Par simplification et par manque de temps nous avons pour cela utilisé nos modèles de fréquence attritionnelle. En effet, pour un seuil à 76 000 € nous avons vu qu'il n'y avait que 3% de sinistres graves, les fréquences totales et attritionnelles sont ainsi quasi identiques et de plus, quel que soit le seuil retenu, pourvu qu'il soit suffisamment cohérent et élevé alors c'est aussi le cas. On établit ensuite chacun de nos modèles GLM de la même façon que précédemment à la différence qu'on ne découpe pas notre base de données en une base d'apprentissage et une base test. Nous ne disposons pas d'assez de données pour cela.

6.1. Première approche

La première approche que nous avons expérimentée a consisté à essayer d'appliquer les modèles du schéma que nous avons décrit aux sinistres de notre périmètre. Tout d'abord on fixe un seuil de grave à 10 000 €. En effet, comme on peut le voir sur les graphiques ci-dessous que nous avons introduit précédemment, au-delà de 10 000 €, le mean excess plot est linéaire et les paramètres sont bien estimés, c'est-à-dire que la largeur de leur intervalle de confiance semble raisonnable. On constate de plus qu'avec la loi estimée, on obtient un diagramme quantile-quantile à peu près linéaire :

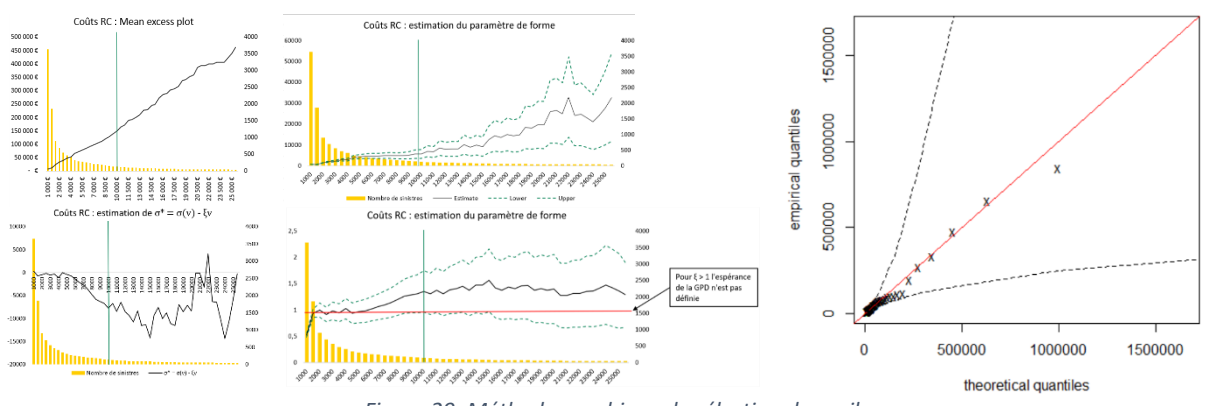


Figure 29: Méthode graphique de sélection de seuil

De la même manière on observe les graphiques des courbes de survie qui représentent la fonction de survie de la loi estimée et la fonction de survie empirique de nos données. On définit la fonction de survie pour une variable aléatoire X de la façon suivante :

$$S(x) = P(X > x)$$

On définit ensuite la fonction de survie empirique de nos données notées $x_1 \leq \dots \leq x_n$ par :

$$\tilde{S}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x > x_i}$$

Comme les fonctions de répartition, les fonctions de survie caractérisent une loi et on remarque sur les graphiques ci-dessous que nos données semblent bien coïncider avec la loi estimée sauf, légèrement moins, en queue de distribution .

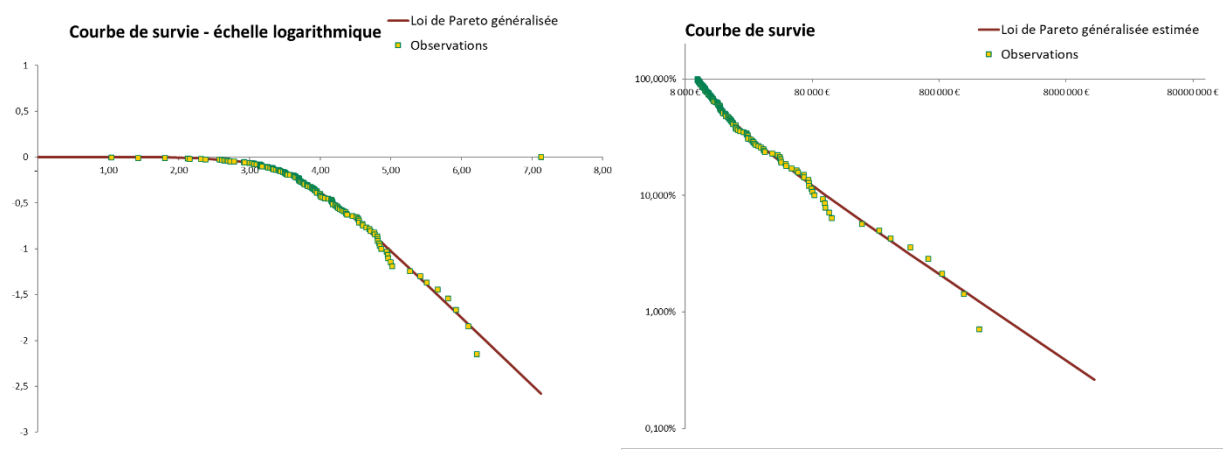


Figure 30: Ajustement d'une loi de Pareto et courbes de survie empirique et théorique

L'inconvénient est que la loi que nous avons estimée a un paramètre de forme supérieur à 1 or dans ce cas l'espérance de la GPD est infinie. Comme on peut le voir sur le graphique ci-dessous, au-delà de 1500 € l'estimation du paramètre de forme oscille légèrement autour de 1 jusqu'environ 5000 € puis dépasse définitivement 1.

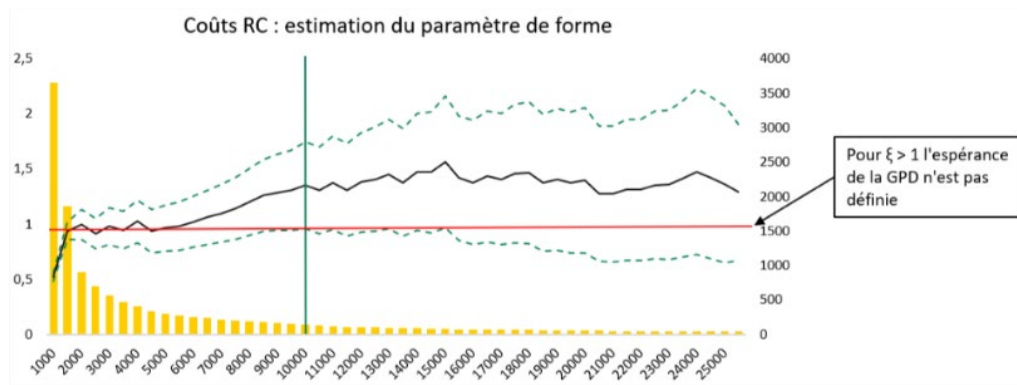


Figure 31: Estimation du paramètre de forme de la GPD

Pour les raisons que nous avons évoquées précédemment, définir un seuil en dessous de 1500 € ou même 5000 € n'est pas pertinent dans le cadre de notre étude. De même qu'estimer une sinistralité infinie. Nous nous restreignons donc à une estimation de la sévérité selon notre deuxième méthode avec un modèle linéaire généralisé et une loi gamma. On modélise dans un premier temps en conservant dans notre base de données, notre sinistre exceptionnel et on observe ensuite les distances de Cook pour constater s'il a exercé une influence sur nos estimations.

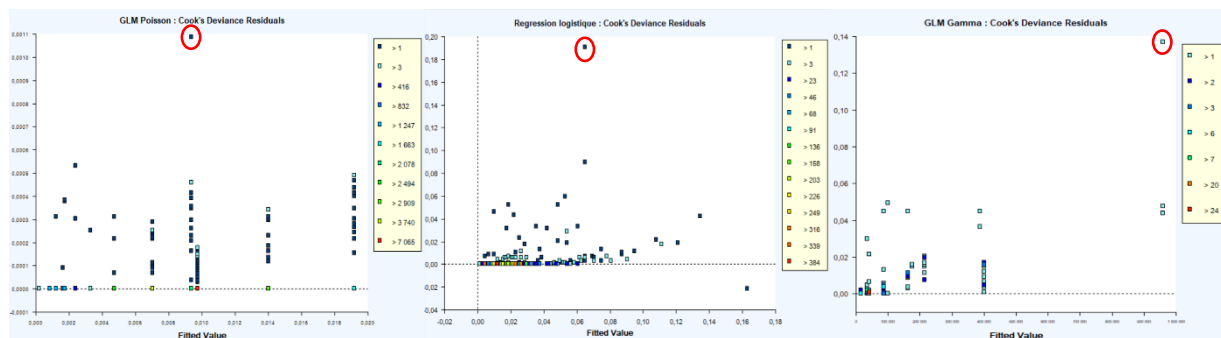


Figure 32: Distances de Cook par modèle

On constate sur chacun de ces graphiques un point plus éloigné des autres. Ainsi, pour écarter tout doute sur un éventuel effet levier de notre sinistre exceptionnel nous réalisons des modélisations alternatives sans ce sinistre. En ce qui concerne la survenance, on suit la même méthode en établissant un modèle de fréquence selon une loi de poisson et une régression logistique pour notre modèle de propension. Avant de conclure sur la méthode la plus pertinente, notamment pour la façon de modéliser la survenance de ces sinistres, on explore une autre approche pour estimer notre sinistralité grave.

6.2. Deuxième approche

Sur le périmètre de notre modélisation et le produit qui le constitue, l'assurance automobile des professionnels ayant au maximum quatre véhicules, nous n'avons pas pu fixer un seuil de sinistres grave d'un ordre de grandeur supérieur à 10 000 €. On explore ainsi la possibilité de nous tourner vers une autre base de données d'un autre portefeuille, mais comparable au nôtre, pour tenter d'obtenir une estimation de la sinistralité grave au-dessus du seuil de 76 000 € qui est actuellement pratiqué et qui nous semble plus raisonnable. Nous nous tournons ainsi vers les flottes automobiles, c'est-à-dire l'assurance automobile des professionnels et des entreprises ayant au moins 5 véhicules à assurer. Nous nous restreignons aux véhicules de première catégorie et on vérifie que les activités professionnelles sur ce périmètre sont comparables à celles de notre périmètre initial :



Flottes auto : Répartition des activités entre 2013 et 2021

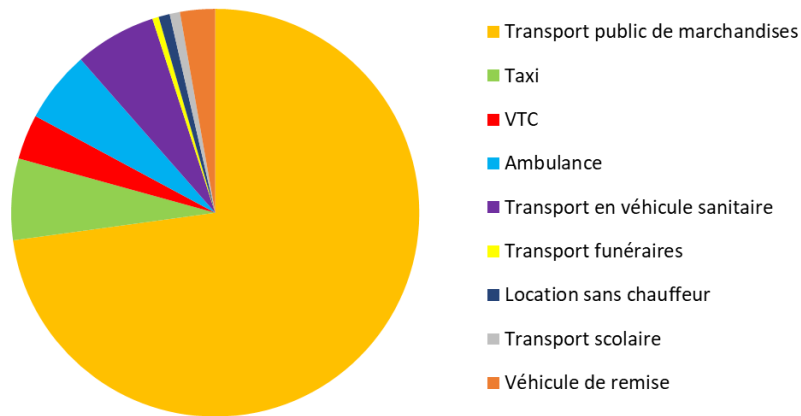


Figure 33: Répartition des activités du produit flottes auto retenues

Les entreprises avec d'autres activités que celles présentées, qui ne coïncident pas avec celles de notre périmètre initial, ont été exclues. Une des principales différences avec notre première base de données est qu'ici tous les conducteurs sont non désignés. Avec ce portefeuille nous avons à disposition un historique de sinistre RC s'étendant de 2013 à 2021. On y applique ensuite les mêmes méthodes que précédemment. On commence ainsi par observer les graphiques correspondants :

Flottes : Coûts RC : Mean excess plot

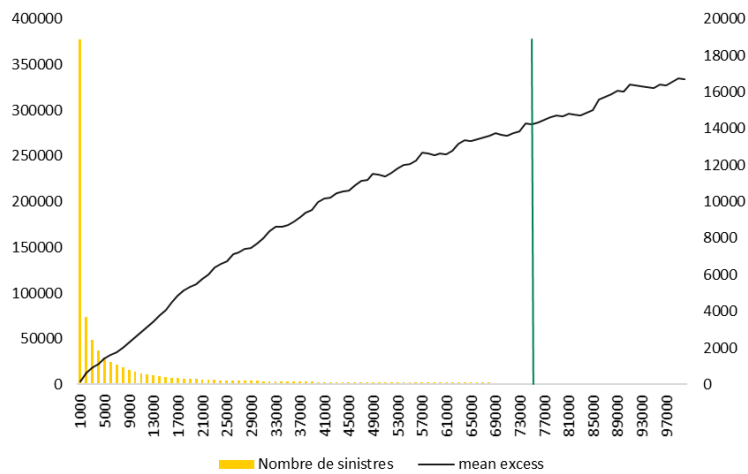


Figure 34: Flottes auto - mean excess plot

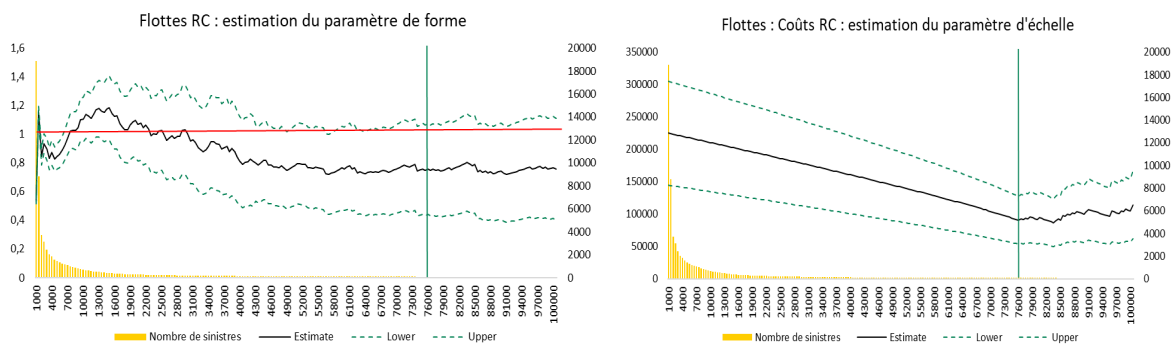


Figure 35: Flottes auto - estimations des paramètres de forme et d'échelle



On constate, en observant les différents graphiques que le seuil que nous avons imposé, 76 000 €, ne semble pas optimal et de plus très peu de sinistres excèdent ce seuil. On vérifie tout de même graphiquement que la loi estimée s’ajuste à la distribution empirique de nos sinistres :

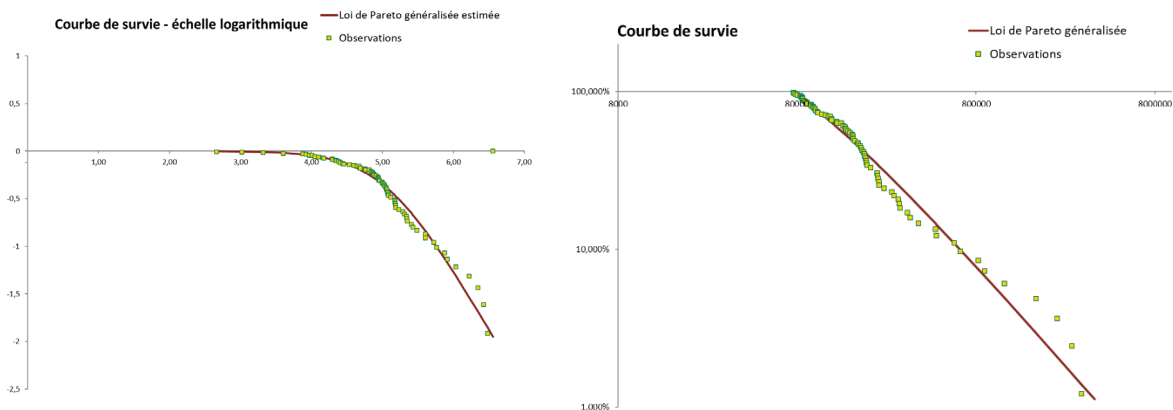


Figure 36: Flottes auto - ajustement d’une loi de Pareto et courbes de survie empirique et théorique

La loi estimée semble bien s’ajuster à nos données, pour le seuil considéré, le paramètre de forme est strictement inférieur à un, on peut ainsi calculer l’espérance de la GPD correspondante :

$$\mathbb{E}(X) = \frac{\sigma}{1 - \xi}$$

On en déduit que le coût moyen de la charge d’un sinistre en surcôte au-dessus de ce seuil est égal à 395 890€. Avec seulement 82 sinistres il n’est pas envisageable d’établir une estimation alternative de la charge sur ce périmètre en utilisant un GLM et une loi Gamma comme nous l’avons fait dans notre première approche. De même on ne peut pas établir de modèle de propension ou estimer de fréquence. Ces difficultés couplées à notre estimation de la sévérité, qui au regard des différents graphiques que nous avons observés semble manquer de robustesse, nous contraignent à éliminer cette approche comme solution candidate à la prise en compte des sinistres graves dans notre calcul de prime pure.

6.3. Troisième approche

Notre troisième et dernière approche consiste à conserver notre seuil de grave à 76 000 € et à simplement répartir le poids de la charge RC au-delà de ce seuil sur tous les contrats. On calcule le coût moyen par contrat de la charge de sinistre résiduelle au-delà de notre seuil de sinistre. L’assiette sur laquelle répartir toute notre charge et notamment notre sinistre exceptionnel, qu’on ne souhaite pas exclure de notre calcul mais dont encore une fois on ignore la période de retour, étant trop réduite on calcule une moyenne pondérée par l’exposition de chaque contrat que l’on note PP_{gr} :

$$PP_{gr} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

Où x_i représente la charge résiduelle au-delà de 76 000 €, et w_i l’exposition. Sans cette pondération on risquerait de surestimer la sinistralité grave par contrat. On obtient ainsi une approximation de la prime pure totale incluant les sinistres graves en multipliant la prime pure attritionnelle par un coefficient :

$$c_{graves} = \left(1 - \frac{PP_{gr}}{PP_{moy,attri,RC} + PP_{gr}} \right)^{-1}$$



Où $PP_{\text{moy,attri,RC}}$ est la prime pure RC attritionnelle moyenne. Cette approximation revient à charger chaque prime pure attritionnelle de la même façon, c'est-à-dire en proportion du poids de PP_{gr} , la moyenne pondérée de la charge des graves au-dessus du seuil, dans sa somme avec la prime pure moyenne attritionnelle.

Pour rendre cette méthode comparable à notre première approche avec un seuil à 10 000 €, nous avons premièrement établi une modélisation alternative de la prime pure hors grave sous le seuil de 10 000 €, selon la même méthode que celle exposée précédemment, puis nous avons comparé les RMSE, les moyennes et les écarts-types de chaque estimation de prime pure globale, incluant les sinistres graves et attritionnels. On confronte enfin ces résultats à la prime RC moyenne observée.

Modèle de prime pure des sinistres graves				Prime pure totale		
Sévérité	Survenance	Seuil	Prise en compte du sinistre exceptionnel	RMSE	Moyenne	Écart-type
Glm loi Gamma	Glm loi de Poisson	10 000 €	Oui	223947	792 €	1 094 €
Glm loi Gamma	Glm loi de Poisson	10 000 €	Non	223964	761 €	997 €
Glm loi Gamma	Propension	10 000 €	Oui	223965	745 €	902 €
Glm loi Gamma	Propension	10 000 €	Non	223968	751 €	886 €
Calcul du coût moyen pondéré de la charge résiduelle		76 000 €	Oui	223966	713 €	390 €
Prime RC observée					758 €	423 €

Tableau 16 : Comparaison des différentes méthodes de prise en compte des sinistres graves.

Il y a peu d'écart entre les RMSE de chacune de nos estimations, on ne peut donc pas décider de la méthode à retenir avec cet indicateur. Si par contre on observe les moyennes et écart-types relatifs à chaque méthode on constate que les estimations issues de notre première approche ont en particulier des écarts-types d'un ordre de grandeur trop supérieur à celui de la prime RC observée, plus de 900 € voire 1000 € contre environ 400 € pour la prime observée. En suivant la première approche on risquerait ainsi de trop souvent surestimer ou sous-estimer la prime pure. Ces écarts-types, trop importants, confirment le manque de robustesse et l'imprédictibilité de ces modèles. La méthode finalement retenue est celle issue de notre troisième approche et qui consiste simplement à répartir le poids de la charge des graves sur tous les contrats. C'est, d'une part, celle en laquelle on peut avoir le plus confiance au vu des résultats des modèles et, d'autre part, celle qui est la plus commercialement cohérente en matière de prime moyenne et d'écart-type en comparaison à la prime RC observée. Une méthode de chargement des graves qui alourdirait de manière trop importante la prime pure serait certes prudente mais inappropriée dans un contexte de forte concurrence entre les tarifs des assureurs.

7. Passage à l'ultime

Dans l'idée de capter le plus d'informations possibles, notre modélisation a été faite à partir de sinistres clos et en cours. Aussi pour qu'elle tienne compte du vieillissement de ces sinistres toujours en cours, on les passe à l'ultime. On cherche ainsi à calculer un coefficient de passage à l'ultime pour la prime pure attritionnelle et en ce qui concerne la prime pure des sinistres graves nous faisons le choix, que nous justifierons, de considérer leur valeur ultime égale à leur valeur comptable.

7.1. Sinistres attritionnels

Pour ajuster notre modélisation de prime pure attritionnelle à l'ultime nous nous sommes uniquement concentrés sur les garanties RC et DTA. Nous nous sommes permis cette simplification car les autres garanties BDG et AST sont des branches très courtes et leurs sinistres sont quasiment tous clos l'année même de leur survenance. On passe ainsi à l'ultime les sinistres des garanties RC et DTA en utilisant une technique très répandue appelée « Chain Ladder ». C'est une méthode qui repose sur des données agrégées et des triangles de développement. On peut considérer des triangles de développement des charges ou de règlement de sinistres. Nous avons de notre côté opté pour des triangles de règlements car nous avons assez d'historique de sinistres. C'est en effet la méthode la plus juste dans ce cas car, pour les anciennes survenances on sait exactement ce qui a été payé et on peut donc s'affranchir de la partie provision estimée par le gestionnaire de sinistre.



On considère ainsi dans notre cas les triangles de règlements cumulés qui se présentent ainsi :

Année d'origine	Délai de règlement								
	1	2	...	j	...	n-i+1	...	n-1	n
1	C_{11}	C_{12}	...	C_{1j}	$C_{1,n-1}$	C_{1n}
2	C_{21}	C_{22}	...	C_{2j}	$C_{2,n-1}$	
...	
i	C_{ij}	...	$C_{i,n-i+1}$	
...	
n-j+1	$C_{n-j+1,j}$	
...	
n-1	$C_{n-1,1}$	$C_{n-1,2}$	
n	C_{n1}								

Figure 37: Schéma des triangles de développement

Où $C_{i,j}$ est le montant cumulé des règlements depuis j années des sinistres survenus l'année i . La diagonale encadrée correspond au cumul des règlements effectués lors de l'année calendaire n , celle où l'étude est menée. C'est la dernière observation du cumul des règlements des sinistres selon leur année de survenance. A l'aide de ces tableaux on a ainsi une vision de comment évoluent les montants de règlement depuis leur année de survenance jusqu'à l'année d'observation. On peut ensuite calculer les facteurs de développement des règlements entre deux années consécutives de la manière suivante :

$$f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$$

L'hypothèse principale de la méthode Chain Ladder est que pour un délai de développement donné, toutes les années de survenance d'origine se développent de la même façon. C'est-à-dire que pour un facteur de développement $f_{i,j}$ l'espérance ne dépend pas de i , l'année de survenance, mais uniquement de j , l'année de développement, et toutes les années se développent donc selon un facteur commun f_j entre les années j et $j + 1$. Un moyen de valider cette hypothèse est de donc de vérifier l'alignement des couples $(C_{i,j}, C_{i,j+1})$ pour j fixé, on trouvera ainsi en annexes 3 et 4 les graphiques des couples $(C_{i,j}, C_{i,j+1})$ pour $j = 0$ pour les garanties RC et DTA. L'approche la plus courante pour estimer les f_j est de calculer la moyenne des facteurs de développement pondérée par les montants cumulés associés :

$$f_j = \frac{\sum_{i=1}^{n-j} f_{i,j} C_{i,j}}{\sum_{i=1}^{n-j} C_{i,j}}$$

Etant donné que $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ on a également $f_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}$. On préférera cependant nous appuyer sur la première formule et calculer un à un les $f_{i,j}$ pour mieux contrôler la présence d'anomalies dans les triangles de développements. On passe donc ensuite à l'ultime de la façon suivante :



Année d'origine	Délai de règlement								
	1	2	...	j	...	n-i+1	...	n-1	n
1	C ₁₁	C ₁₂	...	C _{1j}	C _{1,n-1}	C _{1n}
2	C ₂₁	C ₂₂	...	C _{2j}	C _{2,n-1}	...
...
i	C _{ij}	...	C _{i,n-i+1}
...
n-j+1	C _{n-j+1,j}
...
n-1	C _{n-1,1}	C _{n-1,2}
n	C _{n1}

f_1 f_2 f_j f_{n-i+1} f_{n-1}

Figure 38: Triangles de développement et passage à l'ultime

Quelle que soit l'année de survenance on a $C_{i,n} = C_{i,n-i+1} \prod_{j=n-i+1}^{n-1} f_j$. On obtient enfin le coefficient de passage à l'ultime recherché que l'on calcule en divisant la colonne des ultimes obtenus par la dernière diagonale :

$$c_{u,attri} = \frac{\sum_{i=1}^n C_{i,n}}{\sum_{i=1}^n C_{n-i+1,i}}$$

Dans la pratique, on utilise parfois un trapèze plutôt qu'un triangle pour calculer les facteurs de développement :

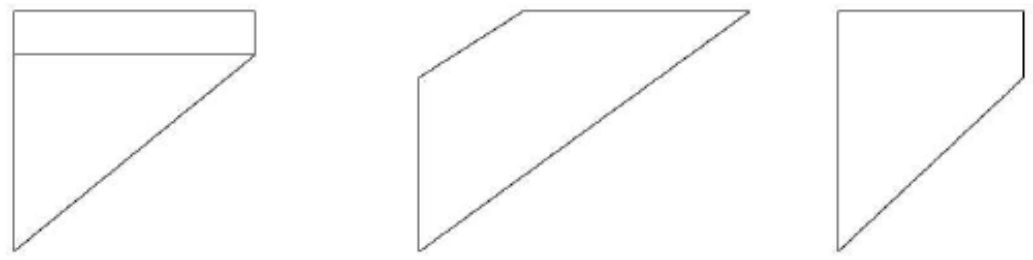


Figure 39: Formes possible de trapèzes de liquidation

Cette situation survient, par exemple, lorsqu'on estime plus pertinent de nous intéresser au développement des années récentes. Si notre triangle est constitué d'un large historique, on peut alors en effet juger plus intéressant de nous restreindre à considérer ses dernières diagonales et omettre les plus anciennes qui pourraient concerner d'anciens processus et modes de gestion des règlements de sinistres. Dans notre cas nos triangles de développement des sinistres attritionnels RC et DTA entourent la période de 2000 à 2021 et on ne s'appuiera que sur leurs cinq dernières diagonales pour calculer nos coefficients de passage à l'ultime. Contrairement à l'historique retenu pour notre base de modélisation ici nous nous sommes autorisés à en prendre un beaucoup plus large et comprenant les sinistres liés aux anciennes versions du produit étudié. Si nous avons pu prendre cette liberté c'est que dans ce cas précis il ne s'agit pas d'étudier un risque qui a pu évoluer, il ne s'agit que de déterminer comment les sinistres se développent à partir de l'évolution des montants sur les années récentes. Pour illustrer notre démarche, on s'appuie sur les résultats obtenus avec la garantie RC (cf. annexe 5 pour la garantie DTA). On obtient ainsi les coefficients de passage à l'ultime en calculant dans un premier temps les facteurs de développement :



7.2. Sinistres graves

Nous ne disposons pas assez de sinistres graves pour constituer des triangles de liquidation. Puisque nous avons décidé de prendre comme ultime des graves leur valeur comptable, nous voulons nous assurer que cette approche est prudente. Nous allons ainsi nous inspirer de la partie descriptive de la méthode Hachemeister pour comprendre le développement de nos sinistres. Contrairement aux méthodes basées sur les triangles de liquidation qui sont des méthodes « agrégées », cette méthode descend au niveau des sinistres individuels. Elle permet d'identifier de manière très visuelle les tranches de coûts, préalablement définies, d'où proviennent des éventuels bonis et malis. La première étape consiste à définir une période d'observation. On choisit ensuite les années de survenance de sinistre que l'on souhaite observer, on définit ensuite des tranches de coût, puis on classe en quatre catégories les sinistres selon qu'ils étaient ouverts ou clos en début et en fin de période : ouverts/ouverts, ouverts/clos, clos/clos et clos/ouverts. Enfin pour chacune de ces catégories on construit des tableaux de nombre de sinistres et de boni mali en croisant les tranches de coût en début et fin de période :

En début de période	En fin de période			Total
	Tranche 1	Tranche 2	...	
Tranche 1	(1,1)	(1,2)		
Tranche 2	(2,1)	(2,2)		
...				
Total				

Figure 42: Structure des tableaux Hachemeister

On peut ainsi observer par catégorie le nombre de sinistres ayant changé de tranche de coût et les bonis et malis que ces variations ont engendrés. Pour une période donnée on a donc au total huit tableaux. C'est une méthode coûteuse en matière de traitement de données mais qui s'avère très précieuse pour avoir une idée générale des catégories de sinistres responsables de boni ou mali importants. En ce qui nous concerne nous avons choisi d'étudier deux périodes, les exercices comptables 2020 et 2021, sur lesquels nous avons considéré les survenances récentes de sinistres RC depuis 2016. Pour définir nos tranches de coûts nous avons isolé les sinistres graves, supérieurs à 76 000€, les sinistres sans suite en fin de période et de montant nul et pour le reste nous nous sommes appuyés sur les quartiles des montants de sinistres inférieurs au seuil de grave. Pour illustrer cela on montre un exemple avec les tableaux des ouverts/ouverts correspondants à la période de l'année 2020 :

Ouvert ouvert

Nombre de sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période							Total
Ouverts en début de période	Nb sinistres	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	
0	1	0	0	0	1	0	0	1
]0,1150]	79	0	43	1	1	23	0	68
]1150,1350]	28	0	0	25	0	2	0	27
]1350,1750]	259	0	2	0	246	4	0	252
]1750,76000]	351	0	2	0	4	338	0	344
]76000,+]	10	0	0	0	0	2	8	10
Total	728	0	47	26	252	369	8	702

Tableau 17: Sinistres ouverts/ouverts nombres de sinistres par tranche de coût



Ouvert / ouvert								
Boni/mali des sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Ouverts en début de période	CM sinistres	- €	702 €	1 235 €	1 508 €	4 575 €	551 063 €	15 043 €
0	- €	- €	- €	- €	- €	- €	- €	- €
]0,1150]	770 €	- €	89 €	- €	- 550 €	- 21 503 €	- €	- 21 964 €
]1150,1350]	1 243 €	- €	- €	- €	- €	- €	- €	- €
]1350,1750]	1 501 €	- €	1 800 €	- €	- 87 €	- 3 775 €	- €	- 2 062 €
]1750,76000]	5 125 €	- €	67 940 €	10 002 €	299 €	179 938 €	- €	258 179 €
]76000,+]	481 315 €	- €	- €	- €	- €	198 483 €	- 13 423 €	185 060 €
Total	15 570 €	- €	69 829 €	10 002 €	- 338 €	353 143 €	- 13 423 €	419 213 €

Tableau 18: Sinistres ouverts/ouverts bonis/malis par tranche de coût

Par exemple, sur ces tableaux on peut observer qu'en 2020, 702 des 728 sinistres ouverts en début de période étaient toujours ouverts en fin de période et qu'ils ont dégagé au total plus de 400 000 € de boni. On remarquera que les sinistres ont en majorité plutôt eu tendance à rester dans leur tranche de coût initial. Ainsi, de la même façon, à partir de l'analyse de tous les tableaux (cf. annexes 6 à 19) correspondant aux deux périodes considérées, on peut faire plusieurs observations au sujet des sinistres graves. Premièrement on constate que sur chacune des deux périodes considérées nos sinistres graves finissent en boni. Le montant s'élève à plus de 1 843 242€ en 2020 et à plus de 286 156€ en 2021. De plus, même en allant dans le détail et en regardant catégorie par catégorie (ouverts/ouverts, ouverts clos, etc.), ils dégagent à chaque fois des bonis bien que la majorité d'entre eux demeure toujours dans la tranche des graves en fin de période. En nous appuyant sur ces observations, on décide de laisser telle quelle notre prime pure des sinistres graves en considérant qu'elle est déjà à l'ultime. Compte tenu de nos observations, on estime cette hypothèse suffisamment prudente. Dans l'idéal nous aurions pu approfondir nos études et par exemple poursuivre l'application de la méthode Hachemeister dans son aspect prédictif. Cependant dans une première approche et par manque de temps nous nous autorisons cette simplification au demeurant prudente.

8. Inflation et frais

Entre le moment où nous avons effectué notre modélisation et celui où l'on souhaite calculer le PTR, les coûts moyens de sinistres ont pu subir une inflation. Pour en tenir compte dans notre calcul on calcule ainsi deux types d'inflation :

- L'inflation moyenne entre l'année N - d et N

$$\hat{i}_{N-d \rightarrow N} = \left(\sqrt[d]{\frac{S_N}{S_{N-d}}} - 1 \right)_+$$

Où S_N est la charge de sinistre cumulée sur l'année calendaire N.

- L'inflation réelle entre l'année n et n + 1 que l'on calcule grâce aux coefficients des triangles de développement tel que nous les avons définis précédemment :

$$i_{n \rightarrow n+1} = \frac{C_{n+1}}{C_n} * \frac{f_{n+1}}{f_n} - 1$$

Pour prendre en compte l'inflation on distinguera ainsi deux cas de figure. Si le calcul du PTR est effectué la même année que celle de la dernière mise à jour de la modélisation de la prime pure on appliquera l'inflation moyenne et on obtient par garantie (gar):

$$SUE(gar) = (PP_{gar}) * (1 + c_{ult}(gar)) * [(1 + \hat{i}_{N-1 \rightarrow N}(gar))] + FNAS(gar)$$



Avec PP_{gar} la prime pure par garantie, $c_{ult}(gar)$ le coefficient de passage à l'ultime et FNAS les frais non allouables directement aux sinistres (coûts de gestion d'indemnisation, commissions, frais généraux).

Si au contraire l'année de calcul du PTR en est éloignée alors on applique l'inflation moyenne à l'année en cours et l'inflation réelle aux années achevées :

$$SUE(gar) = PP_{gar} \left[(1 + \hat{i}_{N-1 \rightarrow N}(gar)) \prod_{n=k}^{N-1} (1 + i_{n \rightarrow n+1}(gar)) \right. \\ \left. + c_{ult}(gar)(1 + \hat{i}_{N-1 \rightarrow N})(gar) \prod_{n=j}^{N-1} (1 + i_{n \rightarrow n+1})(gar) \right] + FNAS(gar)$$

Où k est l'année de la dernière mise à jour de la modélisation de prime pure, N l'année du calcul du PTR et j la dernière année du triangle de liquidation utilisée). Les définitions sont équivalentes pour les sinistres graves ou climatiques. Dans notre cas, notre modélisation a été faite à partir de données vues à fin 2021 et on souhaite calculer le PTR sur le portefeuille 2022, il nous faut donc appliquer l'inflation moyenne calculée à partir de triangles de liquidation. Or par manque de temps nous avons été contraints de faire une approximation. Nous avons ainsi décidé d'approcher l'inflation par l'augmentation moyenne constatée, toutes garanties confondues, des coûts moyens de sinistre $\hat{i}_{2021 \rightarrow 2022}$. Pour ce qui est des frais non allouables aux sinistres, dans notre cas ils sont comptés avec les frais fixes. Nous verrons plus en détail dans la partie suivante comment nous les avons pris en compte.



Prix technique de référence

1. Calcul du prix technique de référence

Avant d'aller plus et de pouvoir montrer comment on peut utiliser le PTR à des fins de tarification et de pilotage, on doit d'abord disposer de toutes ses composantes. Or pour le moment nous nous sommes concentrés sur le calcul de la SUE, la sinistralité ultime estimée. Pour rappel la formule du PTR est la suivante :

$$\text{PTR} = \text{SUE} + F_{\text{ext}} + F_{\text{int}} + C_{\text{réa}} + C_{\text{cap}} - B_{\text{inv}}$$

Avec F_{int} et F_{ext} , les frais internes et externes qui sont les commissions d'administration et d'acquisition et qui nous sont fournis par la direction comptable. Au sujet des coûts de réassurance, $C_{\text{réa}}$, on ne considère que la réassurance non proportionnelle. Pour calculer cette composante, on pourrait par exemple récupérer auprès de la direction de réassurance l'historique de données sur les résultats de la réassurance non proportionnelle et ne conserver que les pertes (coûts négatifs) qui représentent le coût qu'on veut faire porter aux clients. Dans l'idéal on peut ensuite calibrer un modèle statistique pour estimer l'espérance du coût moyen. En ce qui concerne C_{cap} , au sens large le coût du capital immobilisé pour un contrat, se calcule selon la formule suivante :

$$\frac{x\% * \text{VA}(\text{RC})}{1 - \text{Tx}_{\text{Imp}}}$$

Où $x\%$ est l'objectif fixé par l'entreprise de rendement des capitaux propres, $\text{VA}(\text{RC})$ est la valeur actuelle du risque capital (RC) c'est-à-dire la valeur actuelle du capital immobilisé pour un contrat donné et Tx_{Imp} est le taux d'imposition de la société. Cette valeur actuelle découle elle-même d'autres calculs et d'estimations que nous ne détaillerons pas ici, notre intention étant de récupérer ces informations auprès de la direction financière, au même titre par ailleurs que la dernière composante B_{inv} les bénéfices des investissements représentant les retours sur investissement des capitaux provisionnés pour les règlements de sinistres. Cependant par manque de temps et dans une première approche avant d'avoir l'occasion de collaborer avec toutes les autres directions concernées par ces composantes, nous avons procédé à quelques simplifications. A l'aide d'un fichier général répertoriant tous les résultats du groupe de façon agrégée nous avons pu approximer ces composantes. Dans ce fichier les différents frais internes, externes et les FNAS sont agrégés et on ne distingue que frais fixes et frais variables. De plus toutes les composantes qu'on souhaite obtenir y sont exprimées en pourcentage des primes totales, on obtient ainsi :

$$\text{PTR} = \text{SUE} + \text{Frais}_{\text{fixes}} + \% \text{Frais}_{\text{var}} * \text{PNA} + \% C_{\text{réa}} * \text{PNA} + \% C_{\text{cap}} * \text{PNA} - \% B_{\text{inv}} * \text{PNA}$$

Où $\text{Frais}_{\text{fixes}}$ correspond aux frais fixes par contrats, $\% \text{Frais}_{\text{var}}$, $\% C_{\text{réa}}$, $\% C_{\text{cap}}$ et $\% B_{\text{inv}}$ qui représentent respectivement les pourcentages des primes (PNA) associés à $\text{Frais}_{\text{var}}$, $C_{\text{réa}}$, C_{cap} et B_{inv} . Et en résolvant l'équation $\text{PTR} = \text{PNA}$ on a obtenu la formule suivante :

$$\text{PTR} = \frac{\text{SUE} + \text{Frais}_{\text{fixes}}}{1 - (\% \text{Frais}_{\text{var}} + \% C_{\text{Rea}} + \% C_{\text{cap}} - \% B_{\text{inv}})}$$

Pour exprimer cette formule par contrat et par garantie, on doit premièrement ventiler toutes les composantes de ce calcul par garantie. Or dans notre fichier d'où proviennent ces composantes, les résultats sont agrégés toutes garanties confondues. On les retraite ainsi selon le poids de la prime de chaque garantie dans la prime totale. On obtient ainsi, en particulier, pour les frais fixes ($\text{Frais}_{\text{fixes}}(\text{gar})$) :

$$\text{Frais}_{\text{fixes}}(\text{gar}) = \frac{\text{Frais}_{\text{fixes}}}{\text{nb}_{\text{contrats}}} * \frac{\text{PNA}_{\text{gar}}}{\text{PNA}}$$

Où $\text{nb}_{\text{contrats}}$ est le nombre total de contrats en portefeuille, $\frac{\text{PNA}_{\text{gar}}}{\text{PNA}}$ le poids de la prime totale de la garantie considérée (PNA_{gar}) dans la prime totale (PNA). En retraitant de la même façon les différents pourcentages que nous avons introduit on a plus généralement :



$$\text{PTR}(\text{gar}) = \begin{cases} \frac{\text{PP}_{\text{gar}}(1 + c_{\text{ult}}(\text{gar}))(1 + \hat{i}_{2021 \rightarrow 2022})(\mathbb{1}_{\text{gar} \neq \text{RC}} + c_{\text{graves}}\mathbb{1}_{\text{gar} = \text{RC}}) + \text{Frais}_{\text{fixes}}(\text{gar})}{1 - \frac{\text{PNA}_{\text{gar}}}{\text{PNA}}(\% \text{Frais}_{\text{var}} + \% C_{\text{Rea}} + \% C_{\text{cap}} - \% B_{\text{inv}})} \\ \text{PNA}_{\text{gar}} \text{ si la garantie est non modélisée} \end{cases}$$

Avec :

- PP_{gar} , la prime pure par garantie issue de nos modèles.
- $c_{\text{ult}}(\text{gar})$ le coefficient de passage à l'ultime par garantie qu'on a supposé nul pour les garanties BDG et AST.
- $\hat{i}_{2021 \rightarrow 2022}$ notre approximation de l'inflation moyenne.
- c_{graves} le coefficient de prise en compte de la sinistralité grave dans la prime pure RC.
- $\mathbb{1}_{\text{gar} = \text{RC}}$ la fonction indicatrice qui nous permet d'identifier la garantie RC pour la prise en compte ou non des sinistres graves.

En pratique, notre calcul du PTR n'a cependant pas pu être aussi exhaustif que dans ces formules dans la mesure ou n'avons pas pu prendre en compte les composantes C_{cap} et B_{inv} . Ces informations ne nous étant pas parvenues, nous avons dû, dans une première approche nous restreindre à ses autres composantes. Le PTR que l'on calcule est n'inclue donc pas de marge de bénéfice. Il est purement technique.

2. Ratio d'équilibre

Grâce au prix technique de référence dont nous avons désormais un calcul approché, nous allons pouvoir définir de nouveaux indicateurs de suivi et de pilotage dont nous montrerons quelques applications dans la partie suivante. On définit ainsi pour une population donnée :

- Le ratio d'équilibre R_{eq} :

$$R_{\text{eq}} = \frac{\text{PNA}}{\text{PTR}}$$

Il compare le prix technique de référence à la prime réellement émise, après éventuelles remises commerciales ou autre, et représente la rentabilité attendue estimée à partir du calcul du PTR. Contrairement à un ratio S/P ou à un ratio combiné classiques on a ainsi l'avantage avec cet indicateur de bénéficier d'une vision prospective de la rentabilité. On peut le calculer au niveau contrat, toutes garanties confondues mais également par garantie pour les garanties modélisées et dont on peut donc calculer le PTR correspondant. Il s'interprète facilement, un ratio d'équilibre inférieur à un prévoit un déficit entre la prime réellement émise et le prix technique de référence, on est alors sous-tarifé. Lorsque R_{eq} est égal à un, cela indique une tarification parfaitement ajustée au prix technique de référence qui, on le rappelle est la meilleure estimation du prix à faire payer pour atteindre les objectifs d'équilibre technique et financier. Un ratio d'équilibre supérieur à un prévoit lui des bénéfices mais peut aussi alerter sur une sur-tarifation lorsqu'il est trop élevé. Pour résumer cet indicateur nous permet de savoir si ce qui est vendu l'est au « bon prix » c'est-à-dire le PTR. Grâce à cet indicateur nous allons ainsi pouvoir définir des zones pour lesquelles on considère que notre portefeuille est plus ou moins correctement tarifé :

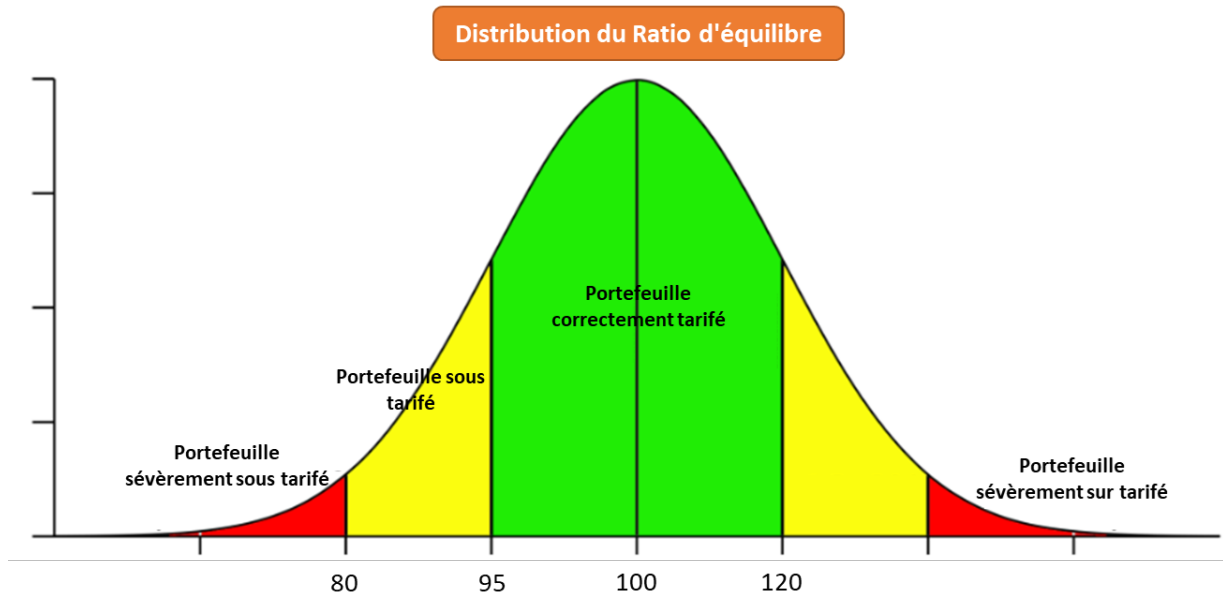


Figure 43: Rentabilité du portefeuille en fonction du ratio d'équilibre

A 100% le portefeuille est correctement tarifé et le tarif coïncide parfaitement avec le PTR. Si l'objectif de la compagnie est de faire 5% de marge alors à 95% le portefeuille n'est ni bénéficiaire ni déficitaire, en dessous le portefeuille est sous-tarifé. A l'inverse au-dessus de 120% on peut raisonnablement penser que le portefeuille est sur-tarifé et que la compagnie s'expose à des résiliations.

3. Résultats techniques estimés

- Les résultats techniques estimés avant et après remises commerciales :
 - Avant remises :

$$RTE_{avt} = PC - PTR ; \%RTE_{avt} = \frac{(PC - PTR)}{PTR}$$

Où PC est la prime commerciale issue du tarif avant remises.

- Après remises :

$$RTE_{ap} = PNA - PTR ; \%RTE_{ap} = \frac{(PNA - PTR)}{PTR}$$

Ces indicateurs, en complément du ratio d'équilibre évaluent l'écart monétaire entre primes commerciales ou réelles et prix technique de référence ainsi que le pourcentage du PTR que cela représente. Par exemple pour une PNA égale à 100 € et un prix technique de 120 € on a $RTE_{ap} = - 20$ € et $\%RTE_{ap} = - 20$ %. Comme le ratio d'équilibre ces indicateurs peuvent se calculer au niveau contrat ou par garantie. Ils vont nous permettre de classer les contrats par tranche de $\%RTE_{ap}$ correspondant et de calculer ainsi par tranche les pertes ou gains associés à l'aide de RTE_{ap} et RTE_{avt} . Nous pouvons les calculer avant ou après applications d'éventuels rabais. Cette particularité sera traitée dans notre la partie concernant les leviers commerciaux.

4. Autres indicateurs

D'autres indicateurs peuvent être calculés à partir du PTR, on peut citer par exemple le ratio S/P de référence qui représente le ratio S/P attendu à partir de la sinistralité ultime estimée :

$$SP_R = \frac{SUE}{PNA}$$



On peut définir également les ratios combinés prospectifs bruts et nets de réassurance :

$$RCP_{\text{Brut}} = \frac{(SUE + F_{\text{Ext}} + F_{\text{Int}} + C_{\text{Réa}})}{PNA}, \quad RCR_{\text{net}} = \frac{(SUE + F_{\text{Ext}} + F_{\text{Int}})}{PNA}$$

Comme pour le ratio S/P de référence, le RCP_{Brut} et le RCR_{net} sont en quelque sorte une version prospective des ratios combinés calculés à partir de données comptables. On peut également définir le ratio combiné de référence :

$$RCR = \frac{PTR}{PNA}$$

Cet indicateur est totalement équivalent au ratio d'équilibre, on pourra préférer l'un ou l'autre selon les goûts. On remarquera, par ailleurs, que comme le PTR que nous avons calculé n'inclue pas les composantes C_{cap} et B_{inv} alors on a $RCR = RCP_{\text{Brut}}$. Il y a ainsi une multitude d'indicateurs qu'il est possible de définir à partir du PTR selon ce qui nous semble pertinent à prendre en compte dans une optique de suivi et de pilotage des résultats. En ce qui nous concerne nous ferons principalement usage du ratio d'équilibre R_{eq} et des résultats techniques estimés, avant et après remises, dans les applications que nous présenterons.

Au contraire d'un ratio S/P classique qui ne décrit que la réalité passée et ne témoigne en rien de l'avenir, les indicateurs que nous avons construits à partir du PTR nous offrent une vision prospective de la rentabilité de notre portefeuille. On peut de plus facilement le mettre à jour. Si on change les modèles il nous suffit de modifier la prime pure dans le PTR. Si un quelconque évènement a une influence sur l'augmentation de la fréquence ou du coût moyen, imaginons qu'une évolution de la jurisprudence conduit à redéfinir la responsabilité dans certains types de sinistres, ce qui ferait croire la fréquence de sinistres RC, ou qu'une entreprise prestataire, dans le contexte de la garantie assistance, augmente le tarif de ses prestations alors on peut ensuite répercuter cela sur le PTR en ajustant la fréquence ou le coût moyen dans la prime pure. C'est un outil facilement interprétable, qui lie tous les acteurs que ce soient les actuaires, les différents intermédiaires ou la direction et qui permet de prendre des décisions rapidement. On peut de plus le décliner à tous les produits pour lesquels on dispose de modèles, le calcul se ferait ainsi de la même manière. Le PTR constitue un sujet central, comme nous allons tenter de l'illustrer, toutes les problématiques de tarifications alimentent le PTR et réciproquement toutes les problématiques de tarification peuvent être vues sous l'angle du PTR.

5. Surveillance de portefeuille¹¹

Une première application que nous avons pu mettre en œuvre sur notre portefeuille est la surveillance de portefeuille. A l'aide du PTR nous allons pouvoir anticiper des dérives de portefeuille, identifier les populations responsables de ces dérives et détecter des niches d'opportunité. On commence ainsi par observer nos indicateurs calculés sur l'ensemble du portefeuille pour avoir une vision globale de sa performance et de sa rentabilité.

Ratio d'équilibre	91%
Ratio combiné de référence	109%
Ratio Combiné prospectif brut de réass	109%
Ratio S/P de référence	77%

Tableau 19: Rentabilité attendue du portefeuille

On s'intéresse particulièrement au ratio d'équilibre $R_{\text{eq}} = PTR/PNA$. Dans notre cas, il est égal à 91% ce qui peut nous alerter d'emblée sur un portefeuille qui dans l'ensemble est sous-tarifé. Nous allons ainsi réaliser des études complémentaires pour comprendre d'où vient cette dérive attendue.

¹¹ Pour des raisons de confidentialité tous les chiffres dans cette partie ont été modifiés.



5.1. Suivi par génération de souscription

La première maille à travers laquelle il est intéressant de regarder pour comprendre l'origine de nos résultats d'ensemble est celle des générations de souscriptions. On observe ainsi nos indicateurs par année de souscription pour comprendre d'une part si nos résultats, qu'ils soient bons ou mauvais, sont influencés par la présence de contrats souscrits lors d'années antérieures et qui peuvent soit tirer vers le haut la rentabilité attendue de notre portefeuille soit au contraire la dégrader si les mesures de pilotage prises dans le passé n'ont pas été suffisantes. D'autre part on cherche particulièrement à observer la rentabilité attendue des affaires nouvelles, les contrats souscrits lors de l'année en cours, qui représentent pour ainsi dire le portefeuille de demain :

Année de souscription	Ratio d'équilibre	Ratio combiné de référence
2016	114%	88%
2017	111%	90%
2018	119%	84%
2019	107%	94%
2020	96%	104%
2021	83%	120%
2022	77%	130%

Tableau 20: Rentabilité attendue par génération de souscription

On constate sur le tableau ci-dessus que la rentabilité attendue des affaires récentes en 2021 et en particulier les affaires nouvelles en 2022 est très dégradée. Les affaires anciennes toujours en portefeuille affichent, elles, une rentabilité attendue assez satisfaisante avec un R_{eq} supérieur à 100% sauf en 2020. Le ratio d'équilibre R_{eq} nous a permis de mettre en lumière une dérive des résultats attendus des affaires récentes et notamment des affaires nouvelles où il apparaît que dans l'ensemble elles ont été sous-tarifées par rapport à leur PTR. C'est un point d'attention très important car il révèle que sans action immédiate, on s'attend à ce que notre portefeuille futur constitué de ces affaires n'ait pas la rentabilité souhaitée.

5.2. Surveillance par critères tarifaires

Pour identifier plus précisément l'origine de la rentabilité dégradée prédite par nos indicateurs, notamment sur les affaires récentes, on calcule nos indicateurs par variable. On montre ici un exemple avec l'activité professionnelle, le critère tarifaire que l'on juge le plus pertinent d'un point de vue métier pour suivre notre portefeuille. Nous nous restreignons aux principales activités :



Exercice/Activité	RCR	R_eq
2016, Petite et grande remise VTC	83%	121%
2017, Petite et grande remise VTC	79%	127%
2018, Petite et grande remise VTC	73%	138%
2019, Petite et grande remise VTC	81%	124%
2020, Petite et grande remise VTC	97%	103%
2021, Petite et grande remise VTC	66%	152%
2022, Petite et grande remise VTC	68%	147%
2016, Taxi	78%	128%
2017, Taxi	79%	126%
2018, Taxi	84%	119%
2019, Taxi	82%	122%
2020, Taxi	82%	122%
2021, Taxi	94%	106%
2022, Taxi	87%	116%
2016, Transporteur public de marchandises	92%	108%
2017, Transporteur public de marchandises	101%	99%
2018, Transporteur public de marchandises	87%	114%
2019, Transporteur public de marchandises	99%	101%
2020, Transporteur public de marchandises	108%	93%
2021, Transporteur public de marchandises	125%	80%
2022, Transporteur public de marchandises	139%	72%

Tableau 21: Rentabilité attendue par activité et par génération de souscription

Sur l'exemple ci-dessus, les activités sont croisées avec les générations de souscriptions. On peut ainsi observer par générations de souscriptions les activités pour lesquelles le tarif pratiqué n'est pas approprié. On constate que parmi les affaires récentes qui sont dans l'ensemble sous-tarifées, les transporteurs publics de marchandises constituent la seule activité à être particulièrement sous-tarifées dans notre portefeuille. On peut donc en déduire que cette activité a beaucoup d'influence sur la rentabilité globale de notre portefeuille. Pour plus d'informations on étudie l'évolution du poids de chaque activité par exercice :

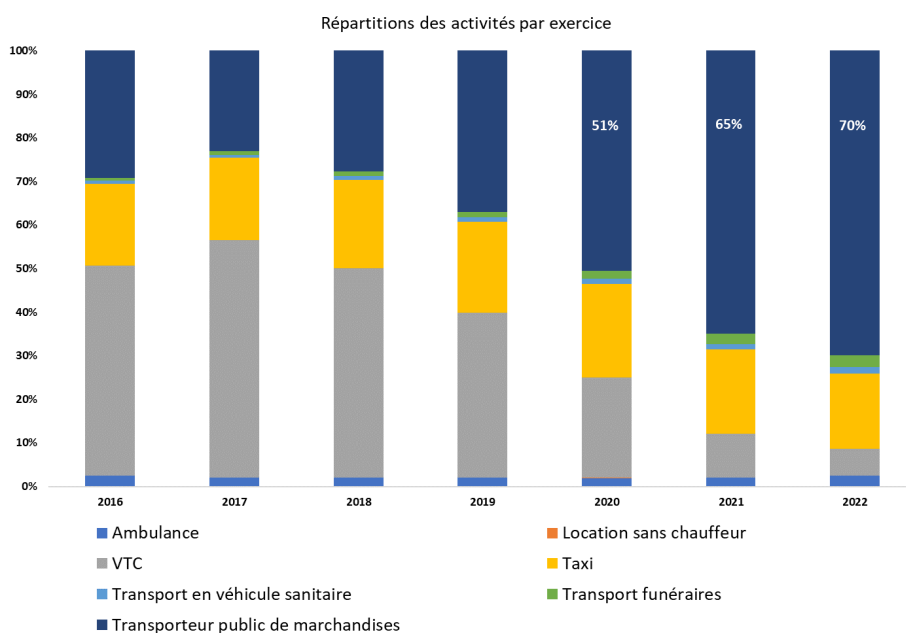


Figure 44: Evolution de la répartition des activités entre 2016 et 2021



Le graphique ci-dessus nous permet de constater qu'entre 2016 et 2021 la part de VTC dans notre portefeuille a fondu au profit de celle des transporteurs publics de marchandises qui représentent aujourd'hui 70 % du portefeuille. Or on peut constater que les VTC qui sont aujourd'hui très minoritaires sont particulièrement sur-tarifés tandis que les TPM qui les ont remplacés sont très sous-tarifés. Sans entrer dans les détails par soucis de confidentialité, on peut néanmoins révéler que par le passé les VTC ont fait l'objet de mesures de majorations spécialement sévères pour faire face aux mauvais résultats dont ils faisaient l'objet. Les préjugés sur les VTC résultant de leur historique ont ainsi conduit à surestimer leur risque sous-jacent et à sous-estimer celui des TPM qui avaient historiquement de bons résultats. D'où l'avantage, une fois de plus, d'avoir une vision prospective des résultats, qui ne se base pas uniquement sur des résultats constatés, et qui donc nous prémunit contre tout mauvais a priori sur la rentabilité des profils de notre portefeuille. Grâce au R_{eq} on a ainsi une vision immédiate de l'impact de l'évolution des profils de notre portefeuille. De la même façon que nous avons croisé les activités et les générations de souscriptions, on peut ensuite croiser autant qu'on le souhaite les variables pour cibler encore plus précisément les profils responsables de dérives. Le R_{eq} par croisement est obtenu aisément étant donné qu'il nous suffit de sommer les PTR et PNA correspondants.

5.3. Exemple d'utilisation des résultats techniques estimés

Après avoir identifié les profils impactant la rentabilité attendue de notre portefeuille, on peut s'intéresser à la distribution des écarts entre prix réellement vendu et prix technique de référence. Pour ce faire on s'appuie en particulier sur l'observation des %RTE. Comme nous l'avons défini précédemment cet indicateur exprime le pourcentage du PTR que représente l'écart entre le PTR et la PNA. Si par exemple %RTE est négatif et inférieur à - 50 % alors cela indique que la perte due à une PNA inférieure au PTR représente plus de 50 % du PTR. On découpe ainsi notre portefeuille en tranche de %RTE :

% RTE	RCR	Req	Nb contrats	RTE	PC	PNA
<= -50	227%	44%	207	- 210 356 €	376 107 €	165 751 €
>-50 et <= -45	191%	52%	130	- 112 470 €	236 132 €	123 662 €
>-45 et <= -40	174%	57%	158	- 114 673 €	268 764 €	154 091 €
>-40 et <= -35	160%	62%	150	- 105 221 €	279 341 €	174 120 €
>-35 et <= -30	148%	68%	163	- 91 960 €	283 854 €	191 894 €
>-30 et <= -25	138%	73%	129	- 57 815 €	210 397 €	152 582 €
>-25 et <= -20	129%	77%	123	- 43 645 €	193 214 €	149 569 €
>-20 et <= -15	121%	82%	129	- 36 650 €	208 856 €	172 206 €
>-15 et <= -10	115%	87%	130	- 25 862 €	201 129 €	175 267 €
>-10 et <= -5	108%	93%	139	- 15 610 €	211 270 €	195 660 €
>-5 et <= -0	102%	97%	102	- 3 403 €	141 899 €	138 496 €
>0 et <=5	97%	103%	126	4 801 €	182 903 €	187 704 €
>5 et <=10	93%	108%	93	10 605 €	140 645 €	151 250 €
>10 et <=15	89%	113%	102	17 274 €	137 823 €	155 097 €
>15 et <=20	85%	117%	88	19 709 €	114 878 €	134 587 €
>20 et <=25	82%	122%	88	26 513 €	119 669 €	146 182 €
>25 et <=30	78%	128%	71	26 104 €	94 576 €	120 680 €
>30 et <=35	75%	133%	61	26 419 €	80 953 €	107 372 €
>35 et <=40	73%	137%	60	29 956 €	80 179 €	110 135 €
>40 et <=45	70%	143%	52	29 333 €	68 831 €	98 164 €
>45 et <=45	68%	147%	27	19 323 €	40 711 €	60 034 €
>50	55%	180%	230	263 333 €	327 351 €	590 684 €

Tableau 22: Distribution de portefeuille par tranche de résultats techniques estimés

Sur le tableau ci-dessus, on voit pour chaque tranche le nombre de contrats associé et l'écart entre PNA et PTR (RTE). On peut ainsi savoir comment se répartissent les contrats selon ces tranches et avoir un aperçu du poids



en matière de gains ou de pertes sur chacune d’entre elles. L’objectif étant d’avoir un portefeuille qui se situe le moins possible dans les tranches extrêmes avec un R_{eq} correspondant entre par exemple 95% et 120% (cf. figure 45). En dessous les contrats sont trop sous-tarifés et on peut craindre de mauvais résultats, alors qu’au-dessus les contrats sont sur-tarifés et plus propices à subir l’action de la concurrence et à être résiliés. En ce qui concerne notre portefeuille on constate que nos contrats se répartissent en majorité sur les tranches correspondant à des R_{eq} inférieurs à 100 %. On remarque également que les deux tranches aux extrémités sont celles regroupant le plus de contrats. De la même façon que précédemment on peut ensuite se focaliser sur une de ces tranches pour déterminer les profils qui la compose. On montre un exemple avec la tranche des %RTE inférieurs à -50 % :

<=-50	2016	2017	2018	2019	2020	2021	2022	Total	% du total
Transporteur public de marchandises	5	4	1	4	10	70	86	180	87%
Taxi	2	2				3		7	3%
VTC	3		2					5	2%
Autres	2		1	2	4	4	2	15	7%
Total	12	6	4	6	14	77	88	207	100%

Tableau 23: Répartition des activités dans la tranche de %RTE <= -50

On observe ainsi encore une fois que ce sont les TPM souscrits lors des années récentes qui composent en majorité cette tranche.

5.4. Simulations

Une fois qu’on a fait apparaître la distribution de nos contrats en matière de %RTE on peut simuler simplement à quel point on peut améliorer notre rentabilité attendue en prenant des mesures simples comme résilier les contrats situés sur les tranches extrêmes et dégradant le plus le ratio d’équilibre d’ensemble. Il nous suffit de recalculer le ratio d’équilibre privé du PTR et de la PNA de ces tranches :

Résiliation	RCR d'origine	RCR	Req	Gain(pt)
contrats <=-50	109%	104%	96	5

Tableau 24: Rentabilité attendue après résiliation de la tranche de %RTE <= -50%

Sur l’exemple ci-dessus on constate qu’en résiliant les 207 contrats de la tranche de %RTE inférieurs à -50%, on améliore de cinq points notre ratio d’équilibre. De la même façon on observe l’effet de la résiliation des contrats de la tranche adjacente :

Résiliation	RCR d'origine	RCR	Req	Gain(pt)
contrats <=-45	109%	101%	99	9

Tableau 25: Rentabilité attendue après résiliation de la tranche de %RTE <= -45%

En résiliant les contrats de cette tranche on améliore cette fois de neuf points notre ratio. On ne prend d’ailleurs pas beaucoup de risques en proposant de résilier ces contrats pour lesquels le tarif pratiqué n’est pas adapté. En effet sur ces deux tranches le ratio d’équilibre correspondant est au moins inférieur à 55% et donc l’écart entre PNA et PTR est inférieur à -45% du PTR (i.e. un %RTE). Les résiliations sont néanmoins des mesures extrêmes, d’autres mesures tarifaires plus ciblées peuvent être entreprises. On a ainsi montré la facilité que nous apporte la vision par tranche de %RTE pour en mesurer les impacts en recalculant simplement les PNA et PTR après actions. Grâce à ces indicateurs on a par ailleurs une vision immédiate des effets de ces mesures.

5.5. Surveillance par garantie

Jusqu’à présent nous n’avons montré que les avantages offerts par les indicateurs issus du PTR au niveau contrat, c’est-à-dire toutes garanties confondues. Or nous avons montré précédemment que le PTR pouvait aussi être



calculé par garantie pour les garanties modélisées. Aussi, pour faire un suivi plus précis de notre portefeuille et prendre des mesures plus ciblées on peut réaliser les mêmes études par garantie :

Garantie	Total	RC	DTA	BDG	AST
Ratio d'équilibre	91%	92%	82%	189%	96%
Ratio combiné de référence	109%	109%	122%	53%	104%
Ratio Combiné prospectif brut de réass.	109%	109%	122%	53%	104%
Ratio S/P de référence	77%	77%	92%	36%	80%

Tableau 26: Rentabilité attendue par garantie

On constate ainsi que les garanties RC et DTA sont fortement sous-tarifées, la garantie AST est également sous-tarifée tandis que la garantie BDG est très fortement sur-tarifée. De plus le ratio d'équilibre de la garantie RC est sensiblement le même que celui d'ensemble. Les résultats de cette garantie ont un fort impact sur la rentabilité générale du portefeuille. Pour éviter toutes redondances nous ne reproduisons pas les mêmes exemples que ceux décrits précédemment mais on peut néanmoins, pour conclure cette partie, nous allons nous intéresser finalement à la rentabilité attendue de ces garanties par génération de souscription :

Année de souscription	RC		DTA		AST		BDG	
	RCR	Req	RCR	Req	RCR	Req	RCR	Req
2016	88%	114%	92%	109%	85%	118%	52%	192%
2017	87%	115%	98%	102%	79%	127%	57%	175%
2018	78%	128%	97%	103%	80%	126%	52%	192%
2019	93%	107%	99%	101%	98%	103%	52%	194%
2020	104%	97%	113%	89%	113%	88%	53%	190%
2021	120%	83%	141%	71%	109%	92%	52%	191%
2022	129%	77%	157%	64%	115%	87%	54%	185%

Tableau 27: Rentabilité attendue par garantie et par génération de souscription

En dehors de la garantie BDG dont le phénomène de sur-tarifcation concerne toutes les générations de souscription, on s'aperçoit que ce sont une nouvelle fois les affaires récentes qui au regard de leur ratio d'équilibre sont sous-tarifées. L'idée générale de ces visions par génération de souscription, par variable, par garantie, etc., est de pouvoir isoler et cibler les profils nécessitant des actions sur le portefeuille. En suivant la même approche de surveillance par générations de souscription, par profils, par garantie, etc., on pourrait également imaginer la mise en place d'un suivi par zones géographiques voire par intermédiaire d'assurance. On pourrait ainsi anticiper toute dérive de portefeuille par intermédiaire, identifier la nature de ces dérives, déterminer si elles sont liées à l'abus de leviers tarifaires, à de l'antisélection, etc., puis mettre en place des actions pour rétablir la rentabilité de leur portefeuille.

6. Affaires nouvelles

Grâce au prix technique de référence nous allons pouvoir faire évoluer le tarif. On peut soit procéder à des ajustements de la structure tarifaire, ajouter ou enlever des questions dans le questionnaire de souscription voire faire abstraction de l'existant et mettre en place un tout nouveau produit. On ne peut cependant pas faire évoluer trop brutalement le tarif. On doit le faire évoluer progressivement vers le prix technique de référence tout en conservant une certaine cohérence. En effet, on ne peut pas changer brutalement le tarif des contrats en portefeuille arrivés au terme. D'un point de vue éthique on ne peut pas par exemple faire admettre à un client qui n'aurait jamais eu de sinistres le doublement de sa prime pour l'ajuster au PTR. Ces brusques variations pourraient aussi engendrer des résiliations voire nuire à l'image de la compagnie si elle venait à adopter la réputation d'avoir un tarif instable qui peut changer subitement. En ce qui concerne les affaires nouvelles on ne peut pas non plus leur proposer un tarif qui soit trop éloigné de celui des contrats déjà en portefeuille. Certains seraient ainsi pénalisés par leur ancienneté. Avec le PTR nous allons donc pouvoir faire évoluer progressivement



le portefeuille vers le niveau de rentabilité et de performance souhaité en ajustant le tarif portefeuille et affaires nouvelles d'année en année :

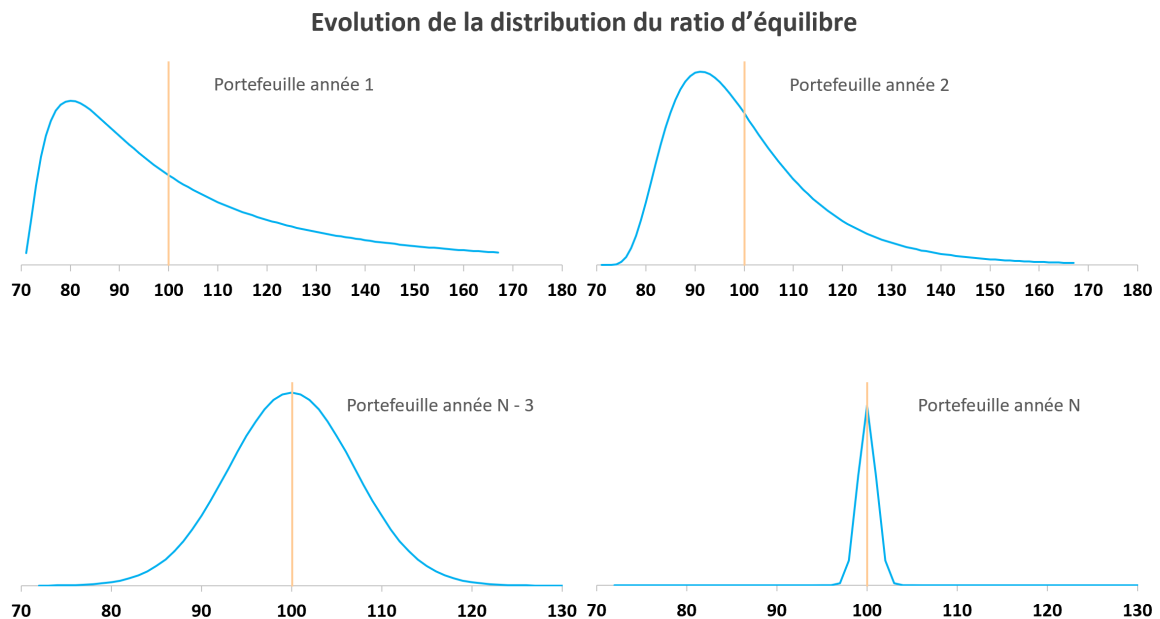


Figure 45: Evolution de la distribution du ratio d'équilibre en fonction du temps

On peut imaginer, comme on peut le voir sur les graphiques ci-dessus, que la première année une part non négligeable du portefeuille n'est pas rentable puis que petit à petit chaque année, jusqu'en année N sur le graphique, le tarif s'ajuste au PTR et que le portefeuille soit donc centré autour d'un ratio d'équilibre égal à 100%. Pour cela une attention particulière doit être portée aux affaires nouvelles car elles représentent le portefeuille futur. Sans prendre trop risque on peut supposer qu'un prospect voulant s'assurer cherche le prix le moins cher possible. En ayant un tarif affaires nouvelles trop volatile par rapport au prix technique de référence c'est-à-dire trop souvent supérieur ou inférieur au prix technique de référence on favoriserait l'antisélection en attirant les profils pour lesquels le tarif est sous-estimé. Comme nous l'avons vu dans la partie précédente le PTR va nous permettre de détecter et de mesurer d'éventuels écarts de rentabilité entre le portefeuille et les affaires nouvelles. On doit par ailleurs surveiller l'évolution de la diversité des profils de notre portefeuille. En effet, si les affaires nouvelles ne sont en majorité que constituées que des profils les moins rentables, l'évolution de la composition du portefeuille risque de s'accompagner d'une dégradation de la rentabilité. Bien que l'objectif final soit d'avoir un tarif adapté à tous types de profils, des aspects commerciaux ainsi que la prise en compte du marché peuvent conduire dans certains cas à disposer malgré tout d'un tarif inférieur au prix technique de référence. Par exemple, si pour certaines raisons la concurrence décide d'avoir un tarif très avantageux sur certains profils au risque de ne pas être rentable, mais dans l'idée de les attirer et de leur faire souscrire d'autres contrats, la compagnie peut également choisir de suivre cette tendance.

7. Pilotage des leviers tarifaires

Une autre application possible du calcul du prix technique de référence est le suivi des leviers tarifaires. Ces leviers tarifaires peuvent être de plusieurs natures. Les leviers les plus fréquents sont les leviers commerciaux. En effet pour attirer de nouveaux clients les souscripteurs peuvent proposer certains rabais. Ces rabais peuvent par exemple provenir d'une campagne commerciale mise en place par la compagnie pour attirer des profils cibles. On peut ainsi penser, dans le cadre de l'assurance de professionnels, à des réductions proposées aux créateurs d'entreprise pour encourager l'entrepreneuriat ou alors à des rabais offerts à des entreprises ayant des activités en lien avec la politique RSE (Responsabilité Sociétale de l'entreprise). S'il juge cela opportun le souscripteur peut aussi proposer des rabais à un client lors de l'acquisition d'un nouveau contrat. Il peut également dans certains cas juger nécessaire d'augmenter la prime proposée par le tarif s'il estime qu'elle n'est pas en adéquation avec le profil de risque présenté par un prospect. En effet, il se peut que certaines spécificités n'aient pas été prises en compte par le questionnaire de souscription ou par le prix technique de référence comme les antécédents d'assurance ou de sinistres. De même, comme nous avons dû exclure toutes les variables



liées aux remarques lors de notre présélection de variables, le PTR ne les intègre pas et est le même quelle que soit la présence ou non d'une remorque à assurer. Sans actions de la part du souscripteur, le PTR, non prévu pour cette situation, risquerait d'être sous-estimé. En calculant le prix technique de référence puis ses indicateurs dérivés comme le ratio d'équilibre ou les résultats techniques estimés avant et après applications de leviers nous disposons ainsi d'outils nous permettant de mesurer leur impact sur la rentabilité globale du portefeuille. Lorsque l'on détecte une dérive on peut donc savoir si elle est spécifiquement due à l'application de leviers ou à alors certains profils risqués. L'autre avantage du PTR est qu'il peut également nous aider à simuler l'impact de mesures tarifaires. Il nous suffit pour cela de simuler le ratio d'équilibre et les résultats techniques estimés après applications de ses rabais. On peut donc se servir du PTR pour mettre en place des leviers commerciaux dont on aura anticipé les effets puis de nouveau nous en servir pour en suivre les impacts.

8. Vision client

Enfin en ce qui concerne les leviers tarifaires, nous pouvons aborder le cas des clients multi-détenteurs. Ces clients possèdent plusieurs contrats et bénéficient ainsi souvent de leviers commerciaux sur certains d'entre eux. Dans ce cas on peut se retrouver dans la situation paradoxale ou dans le cadre du pilotage de notre produit, un contrat soit identifié comme déficitaire alors que le client est rentable si l'on considère l'ensemble de ces contrats. Cette situation particulière nous donne l'occasion d'introduire l'idée d'un pilotage qui se ferait au niveau client plutôt qu'au niveau contrat. En disposant du prix technique de référence sur l'ensemble des produits on pourrait ainsi connaître la rentabilité attendue de chaque client et donc pouvoir par exemple leur proposer des rabais plus ciblés par contrat. En allant plus loin on peut également imaginer la perspective d'un pilotage qui ne prenne pas uniquement en compte la performance des produits un par un mais qui s'attache plutôt à garantir la rentabilité d'ensemble de l'entreprise en maximisant le ratio d'équilibre général calculé à partir de chaque PTR sur chaque produit. On pourrait ainsi mieux mesurer l'impact de mesures comme disposer d'un produit d'appel dont on accepterait la moindre performance au profit d'autres produits plus performant vers lesquels on pourrait attirer les clients. De la même façon le PTR nous aiderait à mettre en place ces mesures, à en anticiper puis mesurer les effets.

9. Aide à la souscription

Le prix technique de référence est un outil de partage qui permet à tous les acteurs que ce soient les actuaires, les membres de la direction ou les divers souscripteurs d'avoir un indicateur commun auquel se référer. Pour le souscripteur, il peut permettre la création d'un outil d'aide à la souscription. Cet outil fournirait le prix technique de référence par garantie ainsi que le ratio d'équilibre associé en fonction des caractéristiques d'un prospect. Le souscripteur serait ainsi directement informé de la rentabilité attendue des prospects, avant ou après application de leviers tarifaires, et donc placé en première ligne du processus de préservation de la performance du portefeuille. Des commentaires viendraient en complément informer le souscripteur des aspects du profil de risque du prospect rendant le prix technique de référence associé plus ou moins éloigné du tarif proposé par la compagnie. Lorsqu'à terme le tarif pratiqué serait ajusté par rapport au prix technique de référence, le souscripteur pourrait en plus avoir une visibilité des leviers qu'il pourrait appliquer aux affaires nouvelles. En effet, en nous appuyant sur le PTR, on aurait la possibilité d'adopter une stratégie selon laquelle on accepterait d'avoir un ratio d'équilibre plus bas pour les nouvelles afin d'augmenter le taux de souscription et d'être le plus compétitif possible par rapport à la concurrence puis d'augmenter progressivement le tarif, comme nous l'avons évoqué précédemment, pour atteindre un ratio d'équilibre situé dans une zone cible (cf. figure 45).

10. Elasticité au prix

En faisant l'hypothèse forte que le risque est le même quelle que soit la compagnie, on peut supposer que chaque compagnie a à peu près le même prix technique de référence mais pas le même tarif. C'est-à-dire qu'en supposant que le risque est le même partout, alors le prix technique de référence, en quelque sorte le juste prix, ne devrait pas varier selon les compagnies. La seule chose qui pourrait ainsi varier c'est l'ajustement de leur tarif qui peut notamment varier en fonction de la précision de leur évaluation du risque. En effet, selon que les compagnies leur éventuelle stratégie commerciale ou selon qu'elles aient pu disposer ou non d'un volume de données suffisant pour construire des modèles robustes, leurs tarifs peuvent diverger car différentes



modélisations peuvent amener des différences de tarif plus ou moins importantes. Si on considère trois compagnies « A », « B » et « C » avec différents niveaux de précision dans leur tarification on peut donc avoir pour le même PTR trois distributions de ratios d'équilibres différentes :

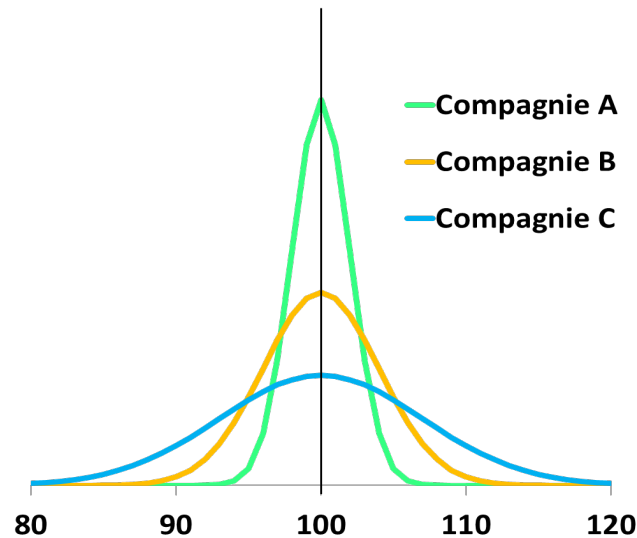


Figure 46 : Distributions du ratio d'équilibre selon trois compagnies avec niveaux de tarification différents

La compagnie « A » a ainsi le tarif le plus ajusté au PTR et les deux autres présentent plus de volatilité dans leur tarification. On peut donc imaginer que si on ne propose pas de rabais aux clients ayant une prime trop élevée par rapport au PTR alors on risque le départ de ces clients vers une compagnie, comme la compagnie « A » fictive, ayant un tarif plus proche de celui-ci. A l'inverse, pour les clients ayant une prime trop inférieure au PTR on suppose qu'on peut augmenter le tarif sans risque puisque par hypothèse ces clients ne trouveraient pas moins cher ailleurs. En suivant cette hypothèse il nous suffirait alors d'observer la distribution des ratios d'équilibre puis d'identifier les contrats aux extrémités pour ensuite mettre en place des actions tarifaires selon les cas. En ce qui concerne les contrats sous-tarifés, même si après augmentation leur tarif reste malgré tout compétitif et inférieur au PTR, si l'effort demandé est trop important alors on peut craindre des résiliations. En effet, on peut imaginer qu'un client sévèrement sous-tarifé qui verrait sa prime augmenter de 50% soit susceptible de résilier en réaction à cette majoration brutale. D'où l'intérêt de prendre en compte l'élasticité au prix par rapport au prix technique de référence. On définit l'élasticité au prix p par :

$$\epsilon(p) = \frac{\partial \tau(p) / \tau(p)}{\partial (p) / p}$$

Où $\tau(p)$ est, selon le cas, le taux de rétention ou de souscription lié au prix p . L'élasticité au prix est un indicateur qui permet de mesurer l'impact d'une variation de tarif sur le comportement d'un client. Elle ne permet cependant pas de connaître la variation exacte permettant de maximiser le profit de l'assureur. Toutefois, en nous basant sur le PTR on pourrait avoir une estimation de cette variation à appliquer.

11. Perspectives d'améliorations

En ce qui nous concerne, la principale perspective d'amélioration du calcul du PTR et des utilisations qui en découlent consiste en une meilleure prise en compte de ses composantes, des indicateurs financiers, de l'inflation, etc., à travers notamment une collaboration avec les services concernés. Disposer de ces composantes de manière plus précise nous permettrait également d'étendre les possibilités d'applications du PTR. Les coûts de réassurance sont, par exemple, pour l'instant mutualisés sur chaque contrat. En modélisant les coûts de réassurance non proportionnelle on pourrait ainsi mieux les répartir et les faire davantage peser sur les profils plus consommateurs de réassurance que sur les autres. En procédant de façon similaire avec les autres composantes on aurait ainsi un prix technique de référence plus précis par contrat. Enfin, dans notre cas on pourrait envisager, dans une prochaine version du calcul du PTR, une meilleure prise en compte des sinistres graves et sinistres liés aux événements naturels.



Conclusion

L'objectif de ce mémoire était de décrire la mise en place d'indicateurs de rentabilité et de performance calculés à partir d'un prix technique de référence puis d'en montrer les avantages et les applications possibles. Tout au long des travaux qui ont permis d'aboutir au calcul de ces indicateurs, nous avons eu à faire face à de nombreuses contraintes et difficultés liées en partie au périmètre de notre étude. Nous avons peu de données à disposition et notre portefeuille comportait des particularités comme la présence de conducteurs non désignés. Nous n'avions pas non plus à disposition toutes les informations nécessaires pour calculer de manière optimale toutes les composantes du prix technique de référence. Cette situation nous a ainsi permis de montrer comment nous avons su surmonter ces obstacles pour obtenir un calcul approché du prix technique de référence tout en restant le plus cohérent et prudent possible. Nous n'avions pas assez de données pour modéliser toutes les garanties mais nous avons montré qu'en nous restreignant à quatre garanties on conservait près de 95 % de la sinistralité. Ensuite, lors de la modélisation de prime pure attritionnelle nous nous sommes attachés à ce que nos modèles s'ajustent correctement sur les profils représentant le cœur de notre portefeuille en assumant que les modalités avec une faible population, et ne représentant pas d'enjeu particulier, face l'objet d'un lissage par rapport à l'observé. La même démarche a été employée pour l'estimation de la sinistralité grave où par manque de données nous avons jugé plus prudent de la prendre en compte à partir d'un calcul de répartition de la charge observée sur tous les contrats. Idem pour le passage à l'ultime, où pour les sinistres graves nous avons décidé de retenir leur valeur comptable. Chacune de ses alternatives, à l'application de la théorie dans un cadre idéal, a ainsi été choisie de manière à nous garantir la plus grande confiance possible en nos calculs et estimations. Enfin, une fois la mise en place du prix technique de référence, nous avons décrit les perspectives d'utilisation des indicateurs qui en découlent à travers des cas pratiques et exemples d'applications possibles. Nous avons ainsi pu montrer en quoi le PTR nous offrait la possibilité d'adopter une méthode prospective de tarification et de pilotage à partir d'indicateurs facilement interprétables pour anticiper d'éventuelles dérives de portefeuille, détecter des niches d'opportunité, mesurer un éventuel écart de rentabilité entre les affaires nouvelles et le portefeuille, avoir une vision directe de la répartition des profils de notre portefeuille et leur rentabilité associée, mesurer rapidement l'impact de leviers tarifaires voire adopter un pilotage au niveau client, etc. On peut de plus facilement le mettre à jour en ajustant ses composantes si besoin et c'est un outil de partage et de communication entre les différents acteurs concernés par le processus de tarification ainsi que la politique tarifaire de la compagnie. Tout ce qui concerne la tarification est liée au PTR, chaque sujet peut être envisagé à la lumière du PTR et réciproquement le PTR se complète et s'alimente de chacun de ces sujets, il est central en tarification. C'est du moins ce que nous avons cherché à montrer.



Bibliographie

[1] : Article 1240 du code civil sur la responsabilité civile : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032041571/

[2] : Article A 121-1 du code des assurances sur le CRM : https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000036754143/

[3] : Articles R312-1 à R312-9 du code de la route sur le PTAC : https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006074228/LEGISCTA000006177084/#LEGISCTA000006177084

[4] : Groupes et Classes SRA : <https://www.sra.asso.fr/informations-vehicules/automobiles/A-propos>

[5] : Cours Maud Thomas, ISUP 2021 : Econométrie de l'assurance de non-vie

[6] : Yi Rong (2013), mémoire d'actuariat : Modélisation de prime pure en assurance multirisque habitation

[7] : Arthur Charpentier. Statistique de l'assurance. 3rd cycle. Université de Rennes 1 et Université de Montréal, 2010, pp.133. ffcel-00550583f

[8] : Cours Marie Kratz, ISUP 2021 : Extreme Value Theory - Theory and Application to Risk Management

[9] : Cours Christian Y. Robert, ISFA - Université Lyon 1 2016 : Théorie des valeurs extrêmes

[10] : Guillaume Gonnet (2010), mémoire d'actuariat : Etude de la tarification et de la segmentation en assurance automobile

[11] : Cook R.D. (1977), « Detection of influential observations in linear regression », Technometrics, vol. 19, p 15-18

[12] : Modou Ndiaye (2017), mémoire d'actuariat : Provisionnement non-vie en responsabilité civile des personnes morales de droit public : Deux nouvelles approches de modélisations, le modèle de Hachemeister et celui basé sur les processus de transitions Markoviens

[13] : Mahon, J. B. (2005). Transition matrix theory and individual claim loss development. In CAS Forum (pp. 115-170)

[14] : Thomas Bouché (2014), mémoire d'actuariat : Modèle de propension des assurés par rapport aux risques de sinistres corporels graves en assurance automobile

[15] : Yu Tang (2017), mémoire d'actuariat : Création d'un modèle dynamique pour l'optimisation tarifaire en assurance MRH



Annexes

Annexe 1 : Tableau récapitulatif de la présélection de variables de modélisation

Variable	Libellé	Top exclure	Top exclure après étude valeurs manquantes	Top exclure après stat descriptives	Top exclure après étude corrélations
EXERCICE	Exercice	0	0	0	0
RESILIATIONANT	Resiliation ant	1	1	0	0
MOTIFRESILANT	Motifresilant	1	1	0	0
MOTIFDUREEASS	Motif Duree assurance moins de 12 mois	1	1	0	0
SINISTREANT	Antecedents Sinistres	1	1	0	0
DUREEASSURANCE	Duree assurance	1	1	0	0
TOP_PJ	Top garantie pj.	1	0	1	0
TOP_AME	Top garantie ame.	1	0	1	0
TOP_ANXB	Top garantie anxb.	1	0	1	0
TOP_DET	Top garantie det.	1	0	1	0
TOP_FRAIS	Top garantie frais.	1	0	1	0
TOP_IMMO	Top garantie immo.	1	0	1	0
TOP_OP	Top garantie op.	1	0	1	0
TOP_PEINT	Top garantie peint.	1	0	1	0
TOP_PF	Top garantie pf.	1	0	1	0
TOP_PPE	Top garantie ppe.	1	0	1	0
TOP_SECURT	Top garantie p_securt.	1	0	1	0
ANT	CONDUCTEUR : 2 ANS D ANTECEDENT	1	0	1	0
EXP	CONDUCTEUR : EXPERIMENTE	1	0	1	0
AJC	CONDUCTEUR : ANNEE JEUNE CONDUCTEUR	1	1	0	0
CRM	Coefficient reduction majoration	1	0	1	0
AGE_COND	Age du conducteur	1	1	0	0
AGE_COND2	Age du conducteur	0	0	0	0
ANC_PERMIS	Anciennete du permis	1	1	0	0
ANC_PERMIS2	Anciennete du permis	0	0	0	0
NBRECRM50	nbre annee CRM 50	1	0	1	0
SITUFAMILLE	Situation familiale	1	0	1	0
TYPCONDUCTEUR	Type de conducteur	0	0	0	0
NBANTSIN_AN	Nombre de sinistres	1	1	0	0
ANT_BDG	Antecedents BDG	1	1	0	0
ANT_INC	Antecedents INC	1	1	0	0
ANT_RESC	Antécédents resp. corporels	1	1	0	0
ANT_RESM	Antecedents resp. materiels	1	1	0	0
ANT_REST	Antecedents non resp. avec tiers.	1	1	0	0



ANT_SREC	Antecedents semi-resp. mat.	1	1	0	0
ANT_SREM	Antecedents semi-resp. corp.	1	1	0	0
ANT_VOL	Antecedents vol	1	1	0	0
CARBURANT	Carburant	1	0	1	0
CARROSSERIE	Carrosserie	1	0	1	0
CLASSESRA	Classe sra	0	0	0	0
ANC_VEH	Anciennete du vehicule	0	0	0	0
ANC_ACHAT	Anciennete achat du veh.	0	0	0	0
GEN	Genre	1	0	1	0
GENRE	genre	1	0	1	0
GROUPE SRA	groupe sra	0	0	0	0
NBPLACES	nbre place	1	0	0	1
POIDS	Poids total autorisé en charge	0	0	0	0
PUISSANCEDIN	puissance reelle (din)	1	0	0	1
PUISSANCEFISCALE	puissance fiscale (cv)	1	0	0	1
SEGMENTVEH	Segment	0	0	0	0
TYPECARROSSERIE	type carrosserie	1	0	1	0
VALEURNEUF	valeur a neuf	1	1	0	0
VEHCLASSE	vehicule classe O/N	1	0	1	0
VEHICULEPROPRE	vehicule propre O/N	1	0	1	0
BATTERIECOMPRISE	batterie comprise	1	1	0	0
CLASSEREPARATION	classe reparation	1	1	0	0
CYLINDREE	cylindree	1	1	0	0
ENERGIE	Energie	1	1	0	0
LIBMARQUE	libelle marque	1	0	1	0
MARQUE_GRP		1	0	1	0
LIBMODELE	libelle modele	1	0	1	0
MARQUE	marque	1	0	1	0
MODELE	modele	1	0	1	0
PRIXBATTERIE	prix batterie	1	0	1	0
PRIXBATTERIECONNU	prix batterie connu O/N	1	1	0	0
RECHNGC	Véhicule recherché NGC	1	1	0	0
SUPP9PLACES	supp 9 places	1	0	1	0
TYPEMINES	type mines	1	1	0	0
GAR_COMPLETE_RMQ	garantie complete	1	1	0	0
GAR_BASE_RMQ	garantie de base	1	1	0	0
GENRE_RMQ	genre	1	1	0	0
MARQUE_RMQ	marque	1	1	0	0
MEC_RMQ	date mise en circulation remorque	1	1	0	0
MODELE_RMQ	modele	1	1	0	0
POIDS_RMQ	poids	1	1	0	0
USAGPRO_RMQ	usage pro	1	1	0	0
CPR	CODE POSTAL RISQUE SITUE	1	1	0	0



CPS	CODE POSTAL DU SOUSCRIPTEUR	1	0	1	0
CLASSESINISTRALITE	classe sinistralite	1	0	1	0
CODENAF	Codenaf	1	0	1	0
LIB_CPRO	Libellé activite	0	0	0	0
CPRO	CODE PROFESSIONNEL	1	0	1	0
CODPOSTALGARAGE	code postal garage	1	0	1	0
COMMUNEINSEEGARAGE	commune insee	1	0	1	0
IMMATETRANGER	Immat étranger O/N	1	0	1	0
LEASING	leasing O/N	1	0	1	0
MODEGARAGE	Mode garage	0	0	0	0
NBRVEHASSURE	nbre vehicule assure	1	0	1	0
REMORQUE	remorque O/N	1	0	1	0
TITCARTEGRISE	Titulaire carte grise	0	0	0	0
USAGEVEH	Usage du veh.	1	0	1	0
VEHSUPP	vehicule supplemen O/N	1	0	1	0
ZONIERRC	Zonier RC	0	0	0	0
ZONIERVOL	zone vol	1	0	1	0
CHGVEHICULE	changement de vehicule O/N	1	1	0	0
SITEAEROPORTUAIRE	site aeroportuaire O/N	1	0	1	0
VEHTEMPORAIRE	vehicule temporaire	1	1	0	0
VILLEGARAGE	ville garage	1	0	1	0

Annexe 2: Liste des variables retenues par garantie

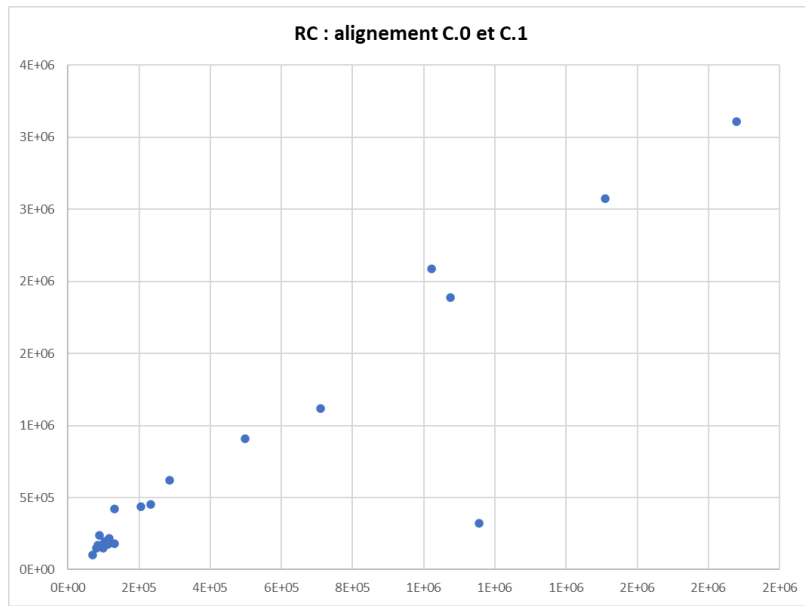
Le tableau ci-dessous récapitule la liste des variables par modèle et indique si elles font partie du tarif actuel.

Variables	Variable tarifaire	RC		DTA		AST		BDG	
		FREQ	CM	FREQ	CM	FREQ	CM	FREQ	CM
Variable 1	x	✓	✓	✓	✓	✓	✓	✓	✓
Variable 2	✓	x	x	x	x	✓	x	x	x
Variable 3	✓	✓	x	x	x	x	x	x	x
Variable 4	✓	x	x	x	x	x	x	x	x
Type de conducteur	✓	x	✓	x	x	x	✓	x	x
Variable 6	✓	x	x	x	x	x	x	✓	✓
Variable 7	✓	x	x	x	✓	x	x	x	x
Activité	✓	x	✓	✓	x	x	✓	✓	x
Variable 9	✓	x	x	x	x	x	✓	✓	x
Variable 10	x	✓	x	✓	x	x	x	x	x
Variable 11	x	x	x	x	x	✓	✓	x	x
Anciennete du permis	✓	✓	x	✓	✓	x	x	x	x
Variable 13	✓	✓	✓	x	x	✓	x	x	✓
Zonier RC	✓	✓	x	x	x	x	x	x	x

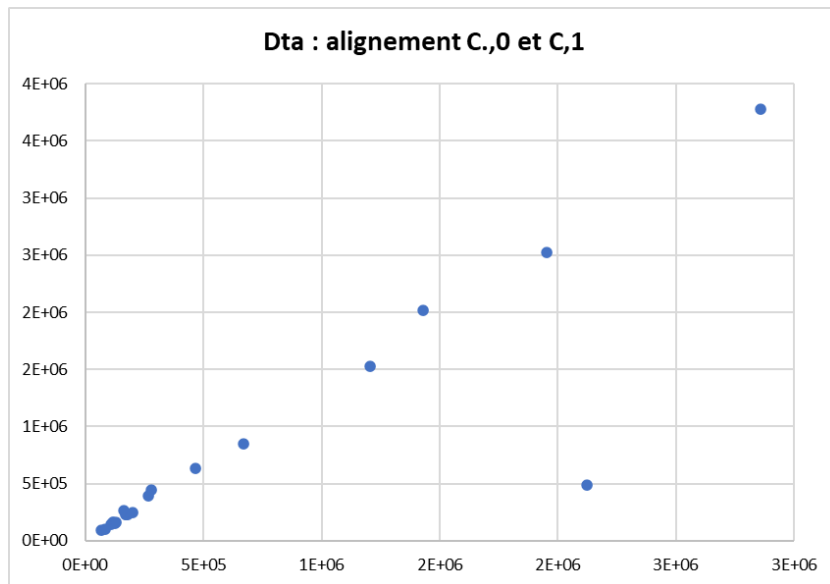
Pour des raisons de confidentialité l'ordre ainsi que les noms des variables ont été modifiés par rapport au tableau 8.



Annexe 3: Garantie RC alignement des couples $(C_{i,j}, C_{i,j+1})$ pour $j = 0$



Annexe 4 : Garantie DTA alignement des couples $(C_{i,j}, C_{i,j+1})$ pour $j = 0$





Annexe 8: Tableau Hachemeister 2021 des nombres de sinistres ouverts en début de période et clos en fin de période

Ouvert / clos

Nombre de sinistres ayant changé de tranche								
Tranche	Clos en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Ouverts en début de période	Nb sinistres							
0	0	0	0	0	0	0	0	0
]0,1150]	54	4	2	0	0	0	0	6
]1150,1350]	10	0	0	0	0	0	0	0
]1350,1750]	63	0	0	0	2	0	0	2
]1750,76000]	360	0	2	0	1	2	0	5
]76000,+]	12	0	0	0	0	0	0	0
Total	499	4	4	0	3	2	0	13

Annexe 9: Tableau Hachemeister 2021 des bonis et mali de sinistres ouverts en début de période et clos en fin de période

Ouvert / clos

Boni/mali des sinistres ayant changé de tranche								
Tranche	Clos en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Ouverts en début de période	CM sinistres	- €	318 €	- €	1 499 €	4 297 €	- €	1 105 €
0	- €	- €	- €	- €	- €	- €	- €	- €
]0,1150]	770 €	4 368 €	1 944 €	- €	- €	- €	- €	6 312 €
]1150,1350]	1 243 €	- €	- €	- €	- €	- €	- €	- €
]1350,1750]	1 501 €	- €	- €	- €	- €	- €	- €	- €
]1750,76000]	5 125 €	- €	5 962 €	- €	10 001 €	2 701 €	- €	18 664 €
]76000,+]	481 315 €	- €	- €	- €	- €	- €	- €	- €
Total	15 570 €	4 368 €	7 906 €	- €	10 001 €	2 701 €	- €	24 976 €

Annexe 10: Tableau Hachemeister 2021 des nombres de sinistres clos en début et en fin de période

Clos / clos

Boni/mali des sinistres ayant changé de tranche								
Tranche	Clos en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Clos en début de période	CM sinistres	- €	517 €	1 254 €	1 433 €	5 224 €	326 312 €	1 215 €
0	- €	- €	- €	- €	- €	- €	- €	- €
]0,1150]	520 €	1 641 €	166 €	- €	- €	7 497 €	- €	5 690 €
]1150,1350]	1 252 €	- €	- €	- €	- €	- €	- €	- €
]1350,1750]	1 439 €	4 618 €	1 606 €	- €	- €	- €	- €	6 224 €
]1750,76000]	5 437 €	7 689 €	- €	- €	- €	26 116 €	- €	33 805 €
]76000,+]	270 009 €	- €	- €	- €	- €	- €	- €	- €
Total	1 246 €	13 948 €	1 772 €	- €	- €	18 619 €	- €	34 339 €

Annexe 11: Tableau Hachemeister 2021 des nombres de sinistres clos en début de période et ouverts en fin de période

Clos / ouvert

Nombre de sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Clos en début de période	Nb sinistres							
0	3227	1	35	15	124	153	0	328
]0,1150]	1215	0	27	2	19	62	0	110
]1150,1350]	83	0	2	2	0	4	0	8
]1350,1750]	2062	0	1	1	120	81	0	203
]1750,76000]	819	0	5	0	1	83	0	89
]76000,+]	4	0	1	0	0	0	0	1
Total	7410	1	71	20	264	383	0	739



Annexe 12: Tableau Hachemeister 2021 des bonis et malis de sinistres clos en début de période et ouverts en fin de période

Clos / ouvert

Boni/mali des sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Clos en début de période	CM sinistres	- €	680 €	1 177 €	1 688 €	3 767 €	- €	2 652 €
0	- €	- €	- 25 504 €	- 17 460 €	- 214 324 €	- 369 686 €	- €	-626 973 €
]0,1150]	520 €	- €	- 4 370 €	- 882 €	- 21 862 €	- 207 771 €	- €	-234 886 €
]1150,1350]	1 252 €	- €	1 250 €	- 196 €	- €	- 7 325 €	- €	- 6 272 €
]1350,1750]	1 439 €	- €	571 €	397 €	- 23 399 €	- 66 152 €	- €	- 88 583 €
]1750,76000]	5 437 €	- €	29 791 €	- €	1 480 €	- 40 289 €	- €	- 9 018 €
]76000,+]	270 009 €	- €	101 099 €	- €	- €	- €	- €	101 099 €
Total	1 246 €	- €	102 837 €	- 18 141 €	-258 105 €	- 691 223 €	- €	-864 632 €

Annexe 13 : Tableau Hachemeister 2020 des nombres de sinistres ouverts en début de période et en fin de période

Ouvert ouvert

Nombre de sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Ouverts en début de période	Nb sinistres							
0	1	0	0	0	1	0	0	1
]0,1150]	79	0	43	1	1	23	0	68
]1150,1350]	28	0	0	25	0	2	0	27
]1350,1750]	259	0	2	0	246	4	0	252
]1750,76000]	351	0	2	0	4	338	0	344
]76000,+]	10	0	0	0	0	2	8	10
Total	728	0	47	26	252	369	8	702

Annexe 14 : Tableau Hachemeister 2020 des bonis et malis de sinistres ouverts en début et en fin de période

Ouvert ouvert

Bonis/malis des sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Ouverts en début de période	CM sinistres	- €	527 €	1 182 €	1 710 €	5 412 €	495 461 €	9 184 €
0	- €	- €	- €	- €	- 1 740 €	- €	- €	- 1 740 €
]0,1150]	770 €	- €	423 €	- 1 100 €	- 1 606 €	- 51 408 €	- €	- 53 691 €
]1150,1350]	1 243 €	- €	- €	- €	- €	- 1 533 €	- €	- 1 533 €
]1350,1750]	1 501 €	- €	1 734 €	- €	- 57 €	- 2 979 €	- €	- 1 302 €
]1750,76000]	5 125 €	- €	12 777 €	- €	8 423 €	163 742 €	- €	184 942 €
]76000,+]	481 315 €	- €	- €	- €	- €	1 055 466 €	461 473 €	1 516 939 €
Total	15 570 €	- €	14 934 €	- 1 100 €	5 019 €	1 163 288 €	461 473 €	1 643 615 €



Annexe 15 : Tableau Hachemeister 2020 des nombres de sinistres ouverts en début de période et clos en fin de période

Ouvert clos

Nombre de sinistres ayant changé de tranche								
Tranche	Clos en fin de période							Total
Ouverts en début de période	Nb sinistres	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	
0	1	0	0	0	0	0	0	0
]0,1150]	79	6	5	0	0	0	0	11
]1150,1350]	28	0	0	1	0	0	0	1
]1350,1750]	259	2	0	0	5	0	0	7
]1750,76000]	351	1	0	1	0	5	0	7
]76000,+]	10	0	0	0	0	0	0	0
Total	728	9	5	2	5	5	0	26

Annexe 16 : Tableau Hachemeister 2020 des bonis et malis des sinistres ouverts en début de période et clos en fin de période

Ouvert clos

Bonis/malis des sinistres ayant changé de tranche								
Tranche	Clos en fin de période							Total
Ouverts en début de période	CM sinistres	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	
0	- €	- €	- €	- €	- €	- €	- €	- €
]0,1150]	770 €	5 351 €	1 319 €	- €	- €	- €	- €	6 670 €
]1150,1350]	1 243 €	- €	- €	- €	- €	- €	- €	- €
]1350,1750]	1 501 €	2 964 €	- €	- €	191 €	- €	- €	3 155 €
]1750,76000]	5 125 €	4 250 €	- €	2 810 €	- €	8 701 €	- €	15 762 €
]76000,+]	481 315 €	- €	- €	- €	- €	- €	- €	- €
Total	15 570 €	12 565 €	1 319 €	2 810 €	191 €	8 701 €	- €	25 586 €

Annexe 17 : Tableau Hachemeister 2020 des nombres de sinistres clos en début et en fin de période

Clos clos

Nombre de sinistres ayant changé de tranche								
Tranche	Clos en fin de période							Total
Clos en début de période	Nb sinistres	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	
0	2795	2385	0	0	0	0	0	2385
]0,1150]	1045	6	919	0	0	0	0	925
]1150,1350]	65	0	0	56	0	0	0	56
]1350,1750]	1757	2	0	1	1446	0	0	1449
]1750,76000]	692	0	0	0	2	576	0	578
]76000,+]	3	0	0	0	0	0	2	2
Total	6357	2393	919	57	1448	576	2	5395



Annexe 18 : Tableau Hachemeister 2020 des bonis et malis des sinistres clos en début et en fin de période

Clos clos

Bonis/malis des sinistres ayant changé de tranche								
Tranche	Clos en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Clos en début de période	CM sinistres	- €	509 €	1 246 €	1 422 €	4 919 €	96 217 €	1 042 €
0	- €	- €	- €	- €	- €	- €	- €	- €
]0,1150]	520 €	3 855 €	1 633 €	- €	- €	- €	- €	5 487 €
]1150,1350]	1 252 €	- €	- €	- €	- €	- €	- €	- €
]1350,1750]	1 439 €	2 962 €	- €	218 €	- €	- €	- €	3 180 €
]1750,76000]	5 437 €	- €	- €	- €	2 960 €	5 327 €	- €	8 287 €
]76000,+]	270 009 €	- €	- €	- €	- €	- €	- €	- €
Total	1 246 €	6 817 €	1 633 €	218 €	2 960 €	5 327 €	- €	16 955 €

Annexe 19 : Tableau Hachemeister 2020 des nombres de sinistres clos en début de période et ouverts en fin de période

Clos ouvert

Nombre de sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période							Total
Clos en début de période	Nb sinistres	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	
0	2795	0	47	9	210	144	0	410
]0,1150]	1045	0	18	5	35	62	0	120
]1150,1350]	65	0	0	1	3	5	0	9
]1350,1750]	1757	0	2	2	185	119	0	308
]1750,76000]	692	0	4	2	2	105	1	114
]76000,+]	3	0	0	0	0	0	1	1
Total	6357	0	71	19	435	435	2	962

Annexe 20 : Tableau Hachemeister 2020 des bonis et malis des sinistres clos en début de période et ouverts en fin de période

Clos ouvert

Bonis/malis des sinistres ayant changé de tranche								
Tranche	Ouverts en fin de période	0]0,1150]]1150,1350]]1350,1750]]1750,76000]]76000,+]	Total
Clos en début de période	CM sinistres	- €	679 €	1 172 €	1 613 €	3 661 €	276 350 €	3 033 €
0	- €	- €	- 33 289 €	- 10 476 €	- 347 243 €	- 402 860 €	- €	- 793 867 €
]0,1150]	520 €	- €	- 2 087 €	- 3 112 €	- 41 896 €	- 122 177 €	- €	- 169 272 €
]1150,1350]	1 252 €	- €	- €	- €	- 1 058 €	- 5 256 €	- €	- 6 313 €
]1350,1750]	1 439 €	- €	2 165 €	786 €	- 24 357 €	- 117 270 €	- €	- 138 675 €
]1750,76000]	5 437 €	- €	18 320 €	2 459 €	7 696 €	- 37 204 €	- 46 870 €	- 55 600 €
]76000,+]	270 009 €	- €	- €	- €	- €	- €	326 303 €	326 303 €
Total	1 246 €	- €	- 14 891 €	- 10 343 €	- 406 857 €	- 684 767 €	279 433 €	- 837 425 €



Liste des abréviations

- PTR : Prix technique de référence
- R_{eq} : Ratio d'équilibre
- RCR : Ratio combiné de référence
- RTE_{avt} : Résultat technique estimé avant rabais
- RTE_{ap} : Résultat technique estimé après rabais
- SP_R : Ratio Sinistres sur Primes de Référence
- SUE : Sinistralité ultime estimée
- F_{ext} : Frais externes
- F_{int} : Frais internes
- $C_{Réa}$: Coût de réassurance
- C_{cap} : Coût du capital immobilisé pour un contrat
- B_{inv} : Bénéfices des investissements
- GLM : Generalized Linear Model
- GPD : Generalized Pareto distribution
- GEV : Generalized Extreme Value distribution
- RC : Responsabilité civile
- DTA : Dommages tous accidents
- BDG : Bris de glaces
- DPRSA : Défense pénale
- INC : Incendie
- AST : Assistance
- FNAS : Frais non allouables aux sinistres
- PNA : Prime nette annuelle
- PC : Prime commerciale
- S/P : Ratio Sinistres sur Primes
- CRM : Coefficient réduction-majoration
- RMSE : root mean square error



Table des figures

Figure 1 : Répartition de l'exposition et de la sinistralité par produit	23
Figure 2: Sinistralité par garantie	24
Figure 3: répartition des montants de sinistres par garantie.....	26
Figure 4 : Détermination d'un seuil de graves RC	27
Figure 5: Poids de la sinistralité climatique par garantie	28
Figure 6: Schéma de la structure de la base de modélisation.....	29
Figure 7: Découpage des situations selon la survenance de sinistres.....	29
Figure 8: Nombre de véhicules par marque et par modèle	34
Figure 9: Fréquence et coûts moyens par activité	35
Figure 10: Corrélations et V de Cramer.....	35
Figure 11: Corrélations et V de Cramer des variables présélectionnées	36
Figure 12: Répartition des activités dans les bases d'apprentissage et de test	37
Figure 13: Modèle de fréquence et choix entre une loi de Poisson et une loi binomiale négative.....	39
Figure 14: Modèle de sévérité RC - QQ-Plot de l'ajustement d'une loi gamma.....	40
Figure 15 : Modèle de sévérité DTA - QQ-Plot de l'ajustement d'une loi gamma	40
Figure 16: Modèle de fréquence RC - ajout de la variable « ancienneté du permis » avant retraitement	41
Figure 17 : Modèle de fréquence RC - ajout de la variable « ancienneté du permis » après retraitement.....	41
Figure 18: Modèle de fréquence RC – regroupement intermédiaire de la variable « ancienneté du permis » ...	42
Figure 19: Modèle de fréquence RC – regroupement alternatif de la variable « ancienneté du permis »	43
Figure 20: Estimations des modalités de la variable activité par exercice.....	44
Figure 21: Résidus des modèles de sévérité et de fréquence RC.....	45
Figure 22: Etendue des résidus par activité	45
Figure 23: Distances de Cook des modèles de sévérité	46
Figure 24 : Ecart moyens entre prédictions et observations de primes pures	47
Figure 25: Prime pure par activité.....	48
Figure 26: Prime pure par groupe SRA.....	48
Figure 27: Prime pure par exercice	49
Figure 28 : Schéma des méthodes d'estimation de la sinistralité grave	50
Figure 29: Méthode graphique de sélection de seuil.....	51
Figure 30: Ajustement d'une loi de Pareto et courbes de survie empirique et théorique	51
Figure 31: Estimation du paramètre de forme de la GPD	52
Figure 32: Distances de Cook par modèle.....	52
Figure 33: Répartition des activités du produit flottes auto retenues.....	53
Figure 34: Flottes auto - mean excess plot	53
Figure 35: Flottes auto - estimations des paramètres de forme et d'échelle	53
Figure 36: Flottes auto - ajustement d'une loi de Pareto et courbes de survie empirique et théorique	54
Figure 37: Schéma des triangles de développement	56
Figure 38: Triangles de développement et passage à l'ultime	57
Figure 39: Formes possible de trapèzes de liquidation.....	57
Figure 40: Retraitement de coefficients dans les triangles de développement	58
Figure 41: Triangle de développement après retraitement.....	58
Figure 42: Structure des tableaux Hachemeister	59
Figure 43: Rentabilité du portefeuille en fonction du ratio d'équilibre	64
Figure 44: Evolution de la répartition des activités entre 2016 et 2021.....	67
Figure 45: Evolution de la distribution du ratio d'équilibre en fonction du temps.....	71
Figure 46 : Distributions du ratio d'équilibre selon trois compagnies avec niveaux de tarification différents.....	73



Liste des tableaux

Tableau 1 : Validation de la mise en place de la base de données	30
Tableau 2 : Valeurs manquantes par variables liées aux remorques.....	31
Tableau 3 : Valeurs manquantes des variables d'antécédents d'assurance et de sinistre	31
Tableau 4: Age du conducteur – valeurs manquantes avant et après retraitement	32
Tableau 5 : Analyses univariées des variable « carburant » et « siteaeroportuaire »	33
Tableau 6: Répartition des conducteurs désignés et non désignés	33
Tableau 7: Nombre d'individus par marque et modèle de véhicule.....	34
Tableau 8: Liste des variables de modélisation présélectionnées	36
Tableau 9 : Ecart de fréquence et de coûts moyens entre les bases d'apprentissage et de test.....	37
Tableau 10: Modèles de fréquence - écarts de prédiction par exercice selon la loi utilisée	39
Tableau 11: Modèles de fréquence - écarts de prédiction par exercice et par garantie selon la loi utilisée.....	39
Tableau 12 : Erreurs de prédiction après regroupement intermédiaire des modalités d'ancienneté du permis	42
Tableau 13: Etendue des coefficients du modèle de fréquence RC.....	43
Tableau 14 : RMSE par modèle	44
Tableau 15 : Ecart de prédictions par garantie pour les bases d'apprentissage et de test	47
Tableau 16 : Comparaison des différentes méthodes de prise en compte des sinistres graves.	55
Tableau 17: Sinistres ouverts/ouverts nombres de sinistres par tranche de coût.....	59
Tableau 18: Sinistres ouverts/ouverts bonis/malis par tranche de coût	60
Tableau 19: Rentabilité attendue du portefeuille.....	65
Tableau 20: Rentabilité attendue par génération de souscription	66
Tableau 21: Rentabilité attendue par activité et par génération de souscription	67
Tableau 22: Distribution de portefeuille par tranche de résultats techniques estimés	68
Tableau 23: Répartition des activités dans la tranche de %RTE <= -50.....	69
Tableau 24: Rentabilité attendue après résiliation de la tranche de %RTE <= -50%	69
Tableau 25: Rentabilité attendue après résiliation de la tranche de %RTE <= -45%	69
Tableau 26: Rentabilité attendue par garantie.....	70
Tableau 27: Rentabilité attendue par garantie et par génération de souscription	70



Liste des annexes

Annexe 1 : Tableau récapitulatif de la présélection de variables de modélisation.....	76
Annexe 2: Liste des variables retenues par garantie	78
Annexe 3: Garantie RC alignement des couples $(C_{i,j}, C_{i,j+1})$ pour $j = 0$	79
Annexe 4 : Garantie DTA alignement des couples $(C_{i,j}, C_{i,j+1})$ pour $j = 0$	79
Annexe 5 : Triangle de passage à l'ultime DTA	80
Annexe 6 : Tableau Hachemeister 2021 des nombres de sinistres ouverts en début et fin de période.	80
Annexe 7 : Tableau Hachemeister 2021 des bonis et malis des sinistres ouverts en début et fin de période	80
Annexe 8: Tableau Hachemeister 2021 des nombres de sinistres ouverts en début de période et clos en fin de période	81
Annexe 9: Tableau Hachemeister 2021 des bonis et mali de sinistres ouverts en début de période et clos en fin de période	81
Annexe 10: Tableau Hachemeister 2021 des nombres de sinistres clos en début et en fin de période	81
Annexe 11: Tableau Hachemeister 2021 des nombres de sinistres clos en début de période et ouverts en fin de période	81
Annexe 12: Tableau Hachemeister 2021 des bonis et malis de sinistres clos en début de période et ouverts en fin de période.....	82
Annexe 13 : Tableau Hachemeister 2020 des nombres de sinistres ouverts en début de période et en fin de période	82
Annexe 14 : Tableau Hachemeister 2020 des bonis et malis de sinistres ouverts en début et en fin de période	82
Annexe 15 : Tableau Hachemeister 2020 des nombres de sinistres ouverts en début de période et clos en fin de période	83
Annexe 16 : Tableau Hachemeister 2020 des bonis et malis des sinistres ouverts en début de période et clos en fin de période.....	83
Annexe 17 : Tableau Hachemeister 2020 des nombres de sinistres clos en début et en fin de période	83
Annexe 18 : Tableau Hachemeister 2020 des bonis et malis des sinistres clos en début et en fin de période	84
Annexe 19 : Tableau Hachemeister 2020 des nombres de sinistres clos en début de période et ouverts en fin de période	84
Annexe 20 : Tableau Hachemeister 2020 des bonis et malis des sinistres clos en début de période et ouverts en fin de période.....	84