



Mémoire présenté devant le jury de l'EURIA en vue de l'obtention du
Diplôme d'Actuaire EURIA
et de l'admission à l'Institut des Actuares

22 septembre 2021

Par : Flavien VALERI

Titre : Surveillance d'un portefeuille de contrats flottes automobiles par l'application
de la théorie de la crédibilité

Confidentialité : Oui - (Durée: 2 ans)

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

**Membre présent du jury de l'Institut
des Actuares :**

Romain LAILY

Sonia GUELOU

Signature :

Entreprise :

GMF Assurances

Signature :

Membres présents du jury de l'EURIA :

Franck VERMET

Directeur de mémoire en entreprise :

Alexandre DAUBAS & Ludovic GASCHET

Signature :

Invité :

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 remercier très sincèrement mes parents, ma sœur, mon frère et mes grands-parents pour le soutien inconditionnel qu'ils m'ont accordé tout au long de mes études et notamment lors de la réalisation de ce travail. Je vous suis également très reconnaissant d'avoir réussi à construire une famille unie et soudée. Cela restera sans aucun doute ma plus grande fierté et ma plus belle réussite.

Je tiens à remercier Alexandre DAUBAS, responsable adjoint du département de Tarification, Statistiques et Pilotage de la direction IARD de GMF Assurances, pour la confiance qu'il m'a accordée en me proposant ce sujet de mémoire au sein de son service. Je le remercie aussi d'avoir pris le temps nécessaire pour m'accompagner afin que je puisse mener à bien cette étude.

Je remercie, mon tuteur, Ludovic GASCHET, chargé d'études statistiques au sein du département de Tarification, Statistiques et Pilotage de la direction IARD de GMF Assurances, pour son soutien et son aide précieuse, tout au long de ce travail. Je le remercie également pour ses conseils techniques sur SAS toujours très avisés qui m'ont souvent facilité le travail.

Je tiens à remercier, Isabelle DEVINE, directrice du département actuariat de Péricle's, pour sa disponibilité et son accompagnement, en tant que tutrice EURIA. Je la remercie de m'avoir apporté des conseils pratiques qui se sont toujours révélés être très judicieux. J'adresse également mes remerciements à Elise CALLAC, actuaire-consultante au sein de Péricle's, pour son suivi ainsi que ses bons conseils tout au long du mémoire.

J'ai une pensée très amicale pour mes collègues de travail au sein du service Études, Statistiques et Produits de la direction IARD de GMF. Angélique TISSERAND, Catherine FAURE, Pauline PICHON, Lionel WLAZINSKI, Mathieu DEBAT et Zoraiz MOHAMMAD, merci pour vos encouragements et pour la bonne humeur que vous diffusez tous les jours au travail. J'en profite pour remercier chaleureusement ma responsable Angélique TISSERAND pour la confiance qu'elle m'a accordée en m'intégrant au sein de son équipe à la suite de mon stage en décembre 2020.

Je tiens à remercier très chaleureusement Floriane PIVETEAU, Oumar DIARRA et Sarah DIDO pour leur lecture attentive et leurs remarques pertinentes. J'ai une pensée particulière pour mon excellent ami Régis LÊ qui a beaucoup contribué à la réussite de ce travail grâce à ses nombreux encouragements.

J'ai une pensée très sincère pour Prudence HOUNKONNOU, Mathieu ETTIEN, Khalil FADIL, Alexis PERRAUD, Killian ABIVEN, Nicolas BAUMER, Noura AITOLAID et Mohamdy DAHOUD, qui m'ont soutenu et encouragé tout au long de ce mémoire. Je vous en suis très reconnaissant.

J'adresse mes remerciements à Mustapha SELLAMI, chargé d'études statistiques au sein du département de Tarification, Statistiques et Pilotage de la direction IARD de GMF Assurances, avec lequel j'ai eu le grand plaisir d'échanger régulièrement sur des sujets divers. J'ai le sentiment d'avoir acquis une meilleure compréhension du fonctionnement de l'Assurance (et notamment des sinistres) grâce à lui, je lui en suis très reconnaissant. J'ai également une pensée pour Philippe MALTAVERNE, toujours très accueillant envers moi et avec lequel j'ai eu du plaisir à discuter.

Plus globalement, j'adresse mes remerciements aux différentes personnes du département de Tarification, Statistiques et Pilotage de la direction IARD de GMF Assurances avec lesquelles c'est un plaisir de travailler au quotidien.

Avertissement

Le mémoire a trait à la revalorisation au terme du portefeuille Flottes de GMF et à la tarification d'affaires nouvelles sur le marché des Flottes automobiles en France. Le marché des flottes automobiles en France est très concurrentiel, les marges sont faibles, les données utilisées sont celles du portefeuille GMF ; ce sont les raisons pour lesquelles toutes les données numériques ont été divisées par un certain coefficient. La structure des données est néanmoins préservée un maximum.

Résumé

Les contrats d'assurance Flottes automobiles sur le marché public ont la particularité d'être, d'une part, issus d'appels d'offres publics émis par des personnes morales, et d'autre part, ils offrent la capacité de couvrir plusieurs types de véhicules à partir d'un seul contrat souscrit, sans bonus-malus, négociables et adaptables (ajout ou suppression des véhicules au sein de la Flotte chaque année).

Ces contrats ont la caractéristique d'être complexes à tarifer : informations disponibles sur l'appel d'offre peu nombreuses (répartition du parc automobile à assurer et données de sinistralité partielles) et concurrence importante. Cela explique en partie pourquoi les assureurs disposent de résultats techniques dégradés sur certaines Flottes.

L'objectif du mémoire est de mettre en place une politique annuelle de revalorisation tarifaire sur le petit portefeuille de flottes automobiles de GMF (approximativement 200 contrats pour un chiffre d'affaires annuel de 10 M€). L'idée est d'y intégrer des modèles complexes de crédibilité (Bühlmann-Straub et la crédibilité hiérarchique de Jewell) puis finalement de juger de leur pertinence en comparaison à des modèles simples basés uniquement sur les résultats historiques de la Flotte.

Mots clefs: Flottes automobiles, théorie de la crédibilité, Bühlmann-Straub, crédibilité hiérarchique, analyse en composantes principales, classification ascendante hiérarchique.

Abstract

The car fleet insurance contracts on the public market have the particularity of being, on the one hand, the result of public calls for tender issued by legal entities, and on the other hand, they offer the capacity to cover several types of vehicles from a single contract subscribed to, without bonus-malus, negotiable and adaptable (addition or deletion of vehicles within the fleet each year).

These contracts are complex to price : little information is available on the call for tenders (distribution of the fleet to be insured and partial claims data) and there is a lot of competition. This partly explains why insurers have deteriorated underwriting results for certain fleets.

The objective of the thesis is to set up an annual tariff revaluation policy on GMF's small portfolio of car fleets (approximately 200 contracts for an annual turnover of 10 M€). The idea is to integrate complex credibility models (Bühlmann-Straub and Jewell's hierarchical credibility) and finally to judge their relevance in comparison to simple models based only on the historical results of the Fleet.

Keywords: car fleets, credibility theory, Bühlmann-Straub, hierarchical credibility, principal component analysis, hierarchical ascending classification

Note de synthèse

L'assurance Flottes automobiles sur le marché public

L'étude porte sur le produit d'assurance Flottes automobiles à destination d'organismes publics, commercialisé par GMF. Ce produit a la particularité d'assurer un ensemble de véhicules (parc), potentiellement de différents types, à l'aide d'un seul et unique contrat d'assurance. Généralement, les clients assurant des flottes de véhicules importantes passent par l'intermédiaire d'un courtier. Le courtier les aide à évaluer leurs besoins, à souscrire l'assurance la plus adaptée et à gérer le quotidien du parc automobile. Ensuite, la compagnie d'assurance reverse une partie de la prime annuelle au courtier sous forme de commission.

Contrairement aux entreprises privées, les organismes publics, de par leur statut, ont l'obligation de recourir à des appels d'offres afin de mettre en concurrence les compagnies d'assurance. L'appel d'offres spécifie les conditions souhaitées par le client (période d'assurance, garanties, etc.), la nature de la flotte (titre professionnel, nombre de véhicules, types de véhicules) et indique l'historique de sinistralité de la flotte lors de la dernière période d'assurance, généralement entre 3 et 5 ans. Ce dernier critère permet aux actuaires d'évaluer le risque de la flotte et de proposer au client une cotisation annuelle. Toutefois, de par la très forte concurrence que se livrent les compagnies entre elles sur ce marché, les actuaires n'ont pas beaucoup de marge pour tarifer et réalisent régulièrement des concessions financières sur la prime afin d'optimiser leurs chances de remporter le marché. Ces éléments ainsi que le manque d'attention portée aux véhicules par les conducteurs expliquent en partie la dégradation des résultats techniques flottes toutes compagnies confondues (ratio sinistres à primes brut de 86% pour l'exercice 2019, observé à fin décembre 2019).

Objectif du produit Flottes GMF

GMF, de son côté, a toujours eu, depuis la mise en vente de son produit flotte, une stratégie plutôt prudente, consistant, la majeure partie du temps, à tarifer les appels d'offres flottes au prix "juste". L'objectif n'étant pas spécialement d'être rentable sur ce produit, mais plutôt de ne pas perdre d'argent. En effet, GMF ciblant essentiellement des agents du service public, l'idée est, pour la marque, d'être présente, dans les organismes publics (où les salariés sont la cible de la GMF pour ses produits à destination des particuliers). Le produit d'assurance Flottes est donc considéré comme une opportunité de faire de la publicité pour la marque GMF.

Portefeuille Flottes GMF

À fin décembre 2019, GMF compte 264 contrats flottes automobiles représentant approximativement 46 000 véhicules et 11 millions d'euros de primes acquises hors taxes. Le portefeuille est relativement petit, en forte progression, hétérogène et composé majoritairement de véhicules légers. De plus, le portefeuille est relativement jeune (lié à l'obligation des organismes publics à ré-émettre un nouvel appel d'offres à chaque échéance, ie tous les 5 ans généralement), composé d'une majorité de collectivités locales et localisé dans le nord de la France.

Sinistralité Flottes GMF

Quant à la sinistralité, le portefeuille est déficitaire et régulièrement soumis à des sinistres de haute intensité dégradant l'ensemble du portefeuille. Tous les indicateurs techniques sont en alerte et préconisent la nécessité de mettre en place des mesures tarifaires fortes pour redresser le portefeuille flottes GMF et minimiser les pertes.

Objectif du mémoire

Concernant la stratégie de revalorisation du portefeuille Flottes, avant juin 2020, il n'existait pas à proprement parler de surveillance, la société se contentait de majorer les contrats à hauteur de l'évolution de l'indice SRA (mesurant l'inflation du coût des réparations automobiles). Seulement, les mauvais résultats techniques des flottes et la croissance du portefeuille ont poussé GMF à revoir ses positions. L'étude prend place à fin juin 2020 et a pour objectif la mise en place d'une nouvelle politique de revalorisation tarifaire 2021 des contrats flottes GMF à l'aide, notamment, de modèles avancés de crédibilité. Ensuite, la question se posera de juger de la pertinence des modèles de surveillance basés sur la crédibilité par rapport à un modèle de surveillance simple basé uniquement sur les résultats historiques de la Flotte, sans projection mais avec les mêmes règles de majoration. Finalement, un modèle de surveillance sera retenu pour l'exercice 2021 selon la politique tarifaire souhaitée par GMF.

Constitution d'une base de données Flottes GMF

Pour cela, il est indispensable de constituer une base de données du portefeuille flotte GMF en cours à fin juin 2020. La base est construite à l'aide de SAS. En effet, les données GMF de portefeuille (contrats, clients, primes, etc.) et de sinistralité sont historisées sous SAS sous forme de tables. Les clients considérés sont les clients qui disposent d'au moins un contrat flottes automobiles GMF à fin juin 2020. Les données de sinistralité et de primes concernent les exercices 2011 à 2020, toutes observées à fin juin 2020 afin d'avoir la meilleure vision de la prime acquise et de la sinistralité du client par exercice. Les variables de localisation, de région ainsi que les informations liées au contrats en cours (préavis, date d'échéance) ont également été intégrées à la base. Toutefois, certaines informations primordiales au bon déroulement de l'étude (nombre de véhicules, répartition du parc, franchises dommages, franchises annuelles, etc.) ont été particulièrement difficiles à récupérer, car elles n'étaient, soit non accessibles à partir des infocentres SAS, soit tout simplement non documentées. C'est pourquoi certaines variables ont demandé un retraitement spécifique tandis que d'autres ont nécessité des méthodes d'estimations.

Mise en place des premiers modèles de surveillance

Une fois la base construite, les modèles de surveillance peuvent être appliqués sur les clients dont il reste au moins une échéance à payer en 2021 (date d'échéance finale supérieure ou égale au 31/12/2021). Un premier modèle de revalorisation est implémenté (revalorisation à hauteur de l'évolution de l'indice SRA entre le deuxième trimestre 2019 et le deuxième trimestre 2020, soit +4%). On obtient une revalorisation de 428 606€.

Un deuxième modèle de revalorisation tarifaire est mis en place, basé sur les résultats techniques du client et son ancienneté. Ce modèle préconise une revalorisation totale du portefeuille de l'ordre de 621 099€.

La théorie de la crédibilité

Ensuite, les modèles de crédibilité sont appliqués. Pour ce qui est de la théorie de la crédibilité en général, c'est un ensemble de techniques utilisées par les actuaires pour déterminer la prime d'un assuré au sein d'un portefeuille hétérogène. L'idée est de pondérer l'expérience individuelle du client par l'expérience historique du portefeuille.

La crédibilité de Bühlmann-Straub

Le premier modèle de crédibilité mis en place sur le jeu de données construit est celui de Bühlmann-Straub. Ce modèle consiste à attribuer à chaque observation de sinistralité des poids. Les observations doivent être des ratios (loss-ratio, fréquence de sinistralité, etc.).

D'après Hans Bühlmann et son élève Erwin Straub, la meilleure approximation linéaire non homogène de la prime de risque $\mu(\Theta_i)$ ou de $X_{i,n+1}$ est :

$$\pi_{i,n+1}^{BS} = z_i X_{i,w} + (1 - z_i) \mathbb{E}[\mu(\Theta_i)]$$

Où $z_i = \frac{w_{i,\Sigma}}{w_{i,\Sigma} + K}$, $K = \frac{s^2}{a}$

- z_i le facteur de crédibilité du contrat i ;
- $\mathbb{E}[\mu(\Theta_i)]$ la prime collective ;
- Le paramètre a symbolise l'hétérogénéité du portefeuille et le paramètre s^2 la variabilité moyenne du portefeuille.

La projection des ratios crédibilisés

Dans un premier temps, les ratios considérés sont les loss-ratio (ratio sinistres à primes) écrêté à 150k euros des individus entre 2011 et 2019 tandis que les poids considérés sont tous simplement les primes acquises de la période 2011-2019. Ce modèle projette les loss-ratio 2020 (écrêté à 150k euros). En parallèle, on projette également les fréquences de sinistralité de la garantie responsabilité civile. Cette fois, les ratios considérés sont les fréquences RC de la période 2011-2019 et les poids considérés sont les parcs assurés sur la période 2011-2019.

Définition des règles de majoration

La politique de revalorisation tarifaire est définie à partir des valeurs prises par les projections du loss-ratio écrêté 2020 et de la fréquence de sinistralité RC. Avant toute chose, une revalorisation minimale est définie, elle s'appliquera aux bons contrats et aux clients ayant intégré le portefeuille en 2020, pour lesquels on ne dispose pas encore, à fin juin 2020, d'informations suffisamment crédible. Ce montant minimal correspond à l'évolution de l'indice SRA modélisant l'inflation du coût des réparations automobiles d'une année à l'autre. Une limite maximale est également définie : on ne souhaite pas majorer les clients au-delà de 30%.

On définit un loss-ratio au-dessus duquel on juge qu'une flotte est dégradée. La valeur de 80% est retenue, représentant le SP brut d'équilibre basé sur des retours d'expérience des équipes techniques Flottes. On fixe également un seuil de crédibilité au-dessus duquel on juge l'expérience individuelle de la flotte comme suffisamment crédible. La valeur de 60% est retenue. Elle correspond aux Flottes dont la prime totale sur la période d'assurance est d'au moins 300k euros. Pour les clients vérifiant ces 2 affirmations, la différence entre le loss-ratio projeté par le modèle de Bühlmann-Straub et le seuil limite (ie 80%) est calculée. La majoration de la prime émise 2020 du client sera de l'ordre de cette différence à condition qu'elle soit supérieure à 4% et inférieure à 30%. Une surmajoration de 10% de la majoration initiale est appliquée aux flottes pour lesquelles la fréquence de sinistralité projetée est supérieure à 10%, car on estime qu'elles ont une probabilité plus importante de générer un sinistre corporel très grave. Le modèle de Bühlmann-Straub combiné à la définition des règles de majoration préconise une revalorisation totale de 901 547 €.

Limites du modèle de Bühlmann-Straub

Toutefois, le modèle de Bühlmann-Straub présente un défaut, il part du principe que, mis à part une pondération, les assurés représentent le même risque. Ce qui n'est pas vrai dans le cas de notre étude. Un modèle plus avancé de crédibilité permet de considérer des classes de risques. L'idée de ce modèle est d'abord de crédibiliser les classes puis les individus au sein des classes.

La crédibilité hiérarchique

D'après le modèle de crédibilité hiérarchique, la meilleure approximation linéaire non homogène de la prime de risque $\mu(\Theta_{k,i}, \Phi_k)$ ou de $X_{k,i,n+1}$ est donnée par les équations récursives suivantes :

$$\begin{aligned}\pi_{k,i,n+1}^J &= z_{k,i} X_{k,i,w} + (1 - z_{k,i}) \pi_k^J \\ \pi_k^J &= z_k X_{k,z,w} + (1 - z_k) m\end{aligned}$$

Où

- $z_{k,i} = \frac{w_{k,i,\Sigma}}{w_{k,i,\Sigma} + \frac{s^2}{\alpha}}$ le facteur de crédibilité du contrat i de la classe k .
- $z_k = \frac{z_{k,\Sigma}}{z_{k,\Sigma} + \frac{b}{\alpha}}$ le facteur de crédibilité de la classe de contrats k .
- Le paramètre m désigne la moyenne collective, le paramètre s^2 la variabilité moyenne du portefeuille, le paramètre a l'hétérogénéité du portefeuille ou la variabilité moyenne intraclasse, le paramètre b l'hétérogénéité interclasse.

Définition des classes de risques et des règles de majoration

Toutefois, avant d'être en mesure de pouvoir appliquer ce modèle, des classes doivent être constituées. On décide de regrouper au sein d'une même classe les 11 contrats possédant une franchise annuelle. Ensuite, il reste 117 individus à classer. le choix est fait de réaliser une classification ascendante hiérarchique à la suite d'une analyse en composantes principales du jeu de données. Après la mise en œuvre de ces étapes, 4 classes regroupent les 117 individus. Après analyse, on parvient à interpréter les classes selon le niveau de prime, la répartition du parc automobile, les niveaux de franchise dommage et les résultats techniques. On dispose donc de 5 classes au total. La crédibilité hiérarchique est appliquée en projetant encore une fois les loss-ratio et les fréquences de sinistralité de la garantie responsabilité civile. Les règles de majoration sont les mêmes que celles appliquées avec le modèle de Bühlmann-Straub. On obtient une revalorisation totale de 918 762 €, du même ordre de grandeur que le résultat obtenu avec le modèle de Bühlmann-Straub. En effet, les projections entre ces deux modèles sont assez semblables l'une de l'autre.

Application du modèle de surveillance sans projection des ratios

Ensuite, il est décidé d'appliquer les mêmes règles de majoration en considérant cette fois les ratios moyens historiques du client et non les ratios projetés avec les modèles de crédibilité. L'idée étant de juger de la pertinence d'inclure des modèles de crédibilité. Ce modèle implique une revalorisation du portefeuille de 885 603 € du même ordre de grandeur que les revalorisations obtenues en projetant les ratios grâce à des modèles de crédibilité.

Conclusion

Les modèles de crédibilité n'apportant pas de résultats significativement différents des modèles sans crédibilité, il a été décidé de ne pas en tenir compte. Pour 2021, GMF a décidé de revaloriser sensiblement (mais dans une certaine limite) son portefeuille Flottes tout en maîtrisant son évolution, c'est pourquoi le modèle de base a été retenu, d'autant plus qu'il a le mérite d'être simple à utiliser. À contrario, si la GMF souhaite développer une revalorisation tarifaire plus forte pour les flottes dans les années futures, le modèle de surveillance sans projection pourrait être parfaitement adapté, d'autant plus que les seuils sont paramétrables en fonction de la politique tarifaire souhaitée.

Summary

Fleet insurance in the public market

The study focuses on the motor fleet insurance product for public bodies, marketed by GMF. This product has the particularity of insuring a group of vehicles (fleet), potentially of different types, with a single insurance contract. Generally, customers insuring large fleets of vehicles go through a broker. The broker helps them assess their needs, purchase the most appropriate insurance and manage the day-to-day running of the fleet. The insurance company then pays a portion of the annual premium to the broker as a commission.

Unlike private companies, public bodies, by virtue of their status, are obliged to resort to calls for tender in order to put insurance companies in competition. The invitation to tender specifies the conditions desired by the client (insurance period, guarantees, etc.), the nature of the fleet (professional title, number of vehicles, types of vehicles) and indicates the claims history of the fleet during the last insurance period, generally between 3 and 5 years. This last criterion allows the actuaries to evaluate the risk of the fleet and to propose an annual premium to the client. However, due to the very strong competition between companies in this market, actuaries do not have much room for pricing and regularly make financial concessions on the premium in order to optimize their chances of winning the market. These factors, as well as the lack of attention paid to vehicles by drivers, partly explain the deterioration in fleet underwriting results for all companies combined (gross loss ratio of 86% for the 2019 financial year, observed at the end of December 2019).

GMF Fleets product objective

GMF, for its part, has always had a rather cautious strategy since the launch of its fleet product, consisting, for the most part, of pricing fleet tenders at the "right" price. The objective is not to be profitable on this product, but rather not to lose money. Indeed, as GMF targets mainly public service agents, the idea is for the brand to be present in public organizations (where employees are GMF's target for its products for individuals). The fleet insurance product is therefore seen as an opportunity to advertise the GMF brand.

GMF Fleet Portfolio

At the end of December 2019, GMF had 264 car fleet contracts representing approximately 46,000 vehicles and €11 million in earned premiums excluding taxes. The portfolio is relatively small, growing rapidly, heterogeneous and composed mainly of light vehicles. In addition, the portfolio is relatively young (due to the obligation of public bodies to reissue a new call for tenders at each expiry date, i.e. generally every 5 years), composed of a majority of local authorities and located in the north of France.

GMF Fleet Claims

As for claims, the portfolio is in deficit and is regularly subject to high intensity claims, which have a negative impact on the entire portfolio. All the technical indicators are on alert and point to the need to implement strong pricing measures to turn around the GMF fleet portfolio and minimize losses.

Objective of the thesis

Regarding the strategy for revaluing the fleet portfolio, before June 2020, there was no monitoring as such, the company was content to increase the contracts by the amount of the evolution of the SRA index (measuring the inflation of the cost of car repairs). However, the poor technical results of the fleets and the growth of the portfolio have pushed GMF to review its positions. The study takes place at the end of June 2020 and aims at setting up a new tariff revaluation policy for 2021 for GMF fleet contracts, using advanced credibility models. Then, the question will arise to judge the relevance of credibility-based monitoring models compared to a simple monitoring model based only on the Fleet's historical results, without projection but with the same mark-up rules. Finally, a monitoring model will be selected for FY 2021 based on GMF's desired rate policy.

Creation of a GMF Fleet database

For this, it is essential to build a database of the GMF fleet portfolio in progress at the end of June 2020. The database is built using SAS. Indeed, the GMF portfolio data (contracts, customers, premiums, etc.) and claims data are stored in SAS in the form of tables. The clients considered are those who have at least one GMF fleet contract at the end of June 2020. The claims and premium data concern the years 2011 to 2020, all observed at the end of June 2020 in order to have the best view of the customer's earned premium and claims per year. The variables of location, region as well as information related to current contracts (notice period, expiry date) have also been integrated into the database. However, some information that was essential to the study (number of vehicles, fleet distribution, damage deductibles, annual deductibles, etc.) was particularly difficult to retrieve, because it was either not accessible from the SAS infocenters or simply not documented. This is why some variables required specific reprocessing while others required estimation methods.

Implementation of the first monitoring models

Once the base has been built, the monitoring models can be applied to clients with at least one remaining due date in 2021 (final due date greater than or equal to 12/31/2021). A first reset model is implemented (reset at the level of the evolution of the SRA index between the second quarter of 2019 and the second quarter of 2020, that is +4%). The result is a revaluation of 428 606€.

A second pricing model has been implemented, based on the client's technical results and seniority. This model recommends a total revaluation of the portfolio of around 621 099€.

The theory of credibility

Next, credibility models are applied. In terms of credibility theory in general, it is a set of techniques used by actuaries to determine the premium for an insured within a heterogeneous portfolio. The idea is to weight the individual experience of the client by the historical experience of the portfolio.

The credibility of Bühlmann-Straub

The first credibility model implemented on the constructed dataset is the Bühlmann-Straub model. This model consists of assigning weights to each loss observation. The observations must be ratios (loss ratio, loss frequency, etc.).

According to Hans Bühlmann and his student Erwin Straub, the best non-homogeneous linear approximation of the risk premium $\mu(\Theta_i)$ ou de $X_{i,n+1}$ est :

$$\pi_{i,n+1}^{BS} = z_i X_{i,w} + (1 - z_i) \mathbb{E}[\mu(\Theta_i)]$$

Where $z_i = \frac{w_{i,\Sigma}}{w_{i,\Sigma} + K}$, $K = \frac{s^2}{a}$

- z_i the credibility factor of the contract i ;
- $\mathbb{E}[\mu(\Theta_i)]$ the collective premium ;
- The parameter a symbolizes the heterogeneity of the portfolio and the parameter s^2 the average variability of the portfolio.

The projection of credible ratios

In a first step, the ratios considered are the loss ratios (loss ratio) clipped to 150k euros of individuals between 2011 and 2019 while the weights considered are simply the earned premiums of the 2011-2019 period. This model projects the 2020 loss ratios (capped at 150k euros). At the same time, the loss frequencies for liability coverage are also projected. This time, the ratios considered are the liability frequencies for the period 2011-2019 and the weights considered are the insured fleets for the period 2011-2019.

Definition of mark-up rules

The policy of tariff revaluation is defined on the basis of the values taken by the projections of the capped loss ratio 2020 and the frequency of liability claims. First of all, a minimum revaluation is defined, which will apply to good contracts and to clients who joined the portfolio in 2020, for whom we do not yet have sufficiently credible information at the end of June 2020. This minimum amount corresponds to the evolution of the SRA index modeling the inflation of the cost of car repairs from one year to the next. A maximum limit is also defined : we do not want to increase customers beyond 30%.

A loss ratio is defined above which a fleet is considered to be degraded. The value of 80% is chosen, representing the gross equilibrium SP based on feedback from the Fleet technical teams. A credibility threshold is also set above which the individual fleet experience is judged to be sufficiently credible. The value of 60% is chosen. It corresponds to fleets whose total premium over the insurance period is at least 300k euros. For clients who meet both of these criteria, the difference between the loss ratio projected by the Bühlmann-Straub model and the threshold (i.e. 80%) is calculated. The mark-up of the client's 2020 premium will be of the order of this difference provided that it is greater than 4% and less than 30%. An additional 10% of the initial premium increase is applied to fleets with a projected loss frequency greater than 10%, as they are considered to have a higher probability of generating a very large bodily injury claim. The Bühlmann-Straub model combined with the definition of the mark-up rules recommends a total mark-up of 901 547 €.

Limitations of the Bühlmann-Straub model

However, the Bühlmann-Straub model has a shortcoming : it assumes that, apart from a weighting, the insureds represent the same risk. This is not true in the case of our study. A more advanced credibility model allows us to consider risk classes. The idea of this model is first to make the classes credible and then the individuals within the classes.

Hierarchical credibility

According to the hierarchical credibility model, the best non-homogeneous linear approximation of the risk premium $\mu(\Theta_{k,i}, \Phi_k)$ or $X_{k,i,n+1}$ is given by the following recursive equations :

$$\begin{aligned}\pi_{k,i,n+1}^J &= z_{k,i}X_{k,i,w} + (1 - z_{k,i})\pi_k^J \\ \pi_k^J &= z_kX_{k,z,w} + (1 - z_k)m\end{aligned}$$

Where :

- $z_{k,i} = \frac{w_{k,i,\Sigma}}{w_{k,i,\Sigma} + \frac{s^2}{\alpha}}$ the credibility factor of contract i of class k
- $z_k = \frac{z_{k,\Sigma}}{z_{k,\Sigma} + \frac{b}{\alpha}}$ the credibility factor of the contract class k.
- The parameter m is the collective average, the parameter s^2 the average variability of the portfolio, parameter α the portfolio heterogeneity or the average intra-class variability, parameter b the inter-class heterogeneity.

Definition of risk classes and mark-up rules

However, before this model can be applied, classes must be formed. It was decided to group the 11 contracts with an annual deductible into the same class. Then, 117 individuals remain to be classified. The choice is made to carry out a hierarchical ascending classification following a principal component analysis of the data set. After the implementation of these steps, 4 classes group the 117 individuals. After analysis, the classes can be interpreted according to the level of premium, the distribution of the vehicle fleet, the levels of damage deductible and the technical results. We therefore have a total of 5 classes. The hierarchical credibility is applied by projecting the loss ratios and the loss frequencies of the third party liability coverage once again. The rules for the mark-up are the same as those applied with the Bühlmann-Straub model. The result is a total revaluation of 918 762 €, of the same order of magnitude as the result obtained with the Bühlmann-Straub model. Indeed, the projections between these two models are quite similar to each other.

Application of the monitoring model without ratio projection

Next, it was decided to apply the same mark-up rules, this time considering the client's historical average ratios and not the ratios projected with the credibility models. The idea is to judge the relevance of including credibility models. This model implies a revaluation of the portfolio of 885 603 € of the same order of magnitude as the revaluations obtained by projecting the ratios using credibility models.

Conclusion

As the credibility models do not provide significantly different results from the models without credibility, it was decided not to take them into account. For 2021, GMF has decided to significantly (but within a certain limit) increase the value of its fleet portfolio while controlling its evolution, which is why the basic model has been retained, especially since it has the advantage of being simple to use. On the other hand, if GMF wishes to develop a stronger tariff revaluation for the fleets in the future years, the monitoring model without projection could be perfectly adapted, especially since the thresholds are parameterizable according to the desired tariff policy.

Introduction

La GMF est présente sur le marché très concurrentiel de l'assurance de flottes automobiles des organismes publics. Initialement, peu offensive sur ce secteur, la société d'assurance mutuelle a considérablement développé son activité lors des dernières années. Le faible volume de primes relativement au produit auto de GMF n'incitait pas la compagnie à mettre en place une surveillance de ce portefeuille autre qu'une majoration homogène de l'ensemble des contrats. Toutefois, la progression du portefeuille, la dégradation progressive des résultats techniques et la présence de sinistres très graves ont conduit GMF à modifier son raisonnement en juin 2020 pour la mise en place des mesures tarifaires 2021.

Après avoir défini le produit d'assurance flottes automobiles GMF, il sera question d'analyser l'évolution du portefeuille flottes, de décrire sa composition et de mettre en avant ses spécificités. Ensuite, les résultats techniques seront analysés pour comprendre l'origine de la dégradation du portefeuille. En conséquence, certains segments du portefeuille seront particulièrement ciblés et feront l'objet de revalorisations tarifaires plus sévères par la suite.

Une fois le contexte assimilé, il s'agira de constituer une base de données des clients flottes automobiles présents en portefeuille à fin juin 2020. Ces individus seront décrits par des variables propres à leur environnement (localisation, activité), par leurs contrats (nombre de véhicules assurés, types de véhicules, primes émises, franchises, etc.), par leur historique de sinistralité sur la période d'assurance (nombre de sinistres responsabilité civile, montant total des sinistres, montant des provisions, etc.) et par leurs résultats techniques (loss-ratio, fréquence de sinistralité de la garantie responsabilité civile).

La base construite, deux modèles de crédibilité seront appliqués au jeu de données. Dans un premier temps, il sera question de projeter les loss ratio et les fréquences de sinistralité responsabilité civile à l'aide du modèle de crédibilité de Bühlmann-Straub. Ensuite, selon les valeurs de ces deux indicateurs projetés et du facteur de crédibilité, des règles de majoration seront établies. Puis, dans un second temps, des classes de risques seront créées, à l'aide d'une analyse en composantes principales et d'une classification ascendante hiérarchique, afin de projeter les loss-ratio et les fréquences de sinistralité responsabilité civile selon le modèle de la crédibilité hiérarchique. Les mêmes règles de majoration seront mises en place afin de trancher quant à la pertinence des 2 modèles.

Ensuite, la question se posera de juger de la pertinence des modèles de surveillance basés sur la crédibilité par rapport à un modèle de surveillance simple basé uniquement sur les résultats historiques de la Flotte, sans projection mais avec les mêmes règles de majoration. Finalement, un modèle de surveillance sera retenu pour l'exercice 2021 selon la politique tarifaire souhaitée par GMF.

Table des matières

Résumé	3
Abstract	4
Note de synthèse	5
Summary	9
Introduction	13
I Présentation du produit d'assurance flottes automobiles	1
1 Le contexte général de l'assurance flottes automobiles en France	2
1.1 Une flotte automobile	2
1.2 Une assurance obligatoire	2
1.3 Les contrats individuels	2
1.4 Les contrats flottes	2
1.4.1 Les contrats à flotte fermée	3
1.4.2 Les contrats à flotte ouverte	3
1.5 L'assurance mission	3
1.6 Les différentes couvertures possibles	3
1.6.1 Les garanties de base	4
1.6.2 Les garanties dommages	4
1.6.3 Les garanties optionnelles	4
1.7 Franchises	4
1.8 Contrats avec conservation de sinistres	5
1.9 Les plans de prévention	5
1.10 Tarification des produits flottes	6
1.11 Modes de distribution de l'assurance flotte	6
2 L'assurance flottes automobiles en chiffres	7
2.1 Les principaux clients	7
2.2 Un marché en pleine croissance	7
2.3 Des résultats techniques dégradés	8
3 Le produit flottes automobiles GMF	9
3.1 GMF	9
3.2 L'assurance flottes automobiles GMF	9
3.3 La particularité du marché public	10
3.3.1 L'apporteur d'affaire	10
3.3.2 L'appel d'offres	11
3.3.3 La tarification des appels d'offres flottes chez GMF	12
3.3.4 La Difficulté du prix juste	13

II	Statistiques descriptives : analyse de l'évolution du portefeuille flottes GMF (production et sinistralité)	14
4	Portefeuille flottes GMF	16
4.1	Répartition du parc automobile	18
4.2	Les primes par garantie	19
4.3	Localisation du portefeuille	19
4.4	Secteur d'activité du portefeuille	20
4.5	Ancienneté des clients	20
4.6	Zoom sur les contrats avec conservation	21
5	Sinistralité flottes GMF	23
5.1	La liquidation des différentes garanties	23
5.2	Les coûts moyens observés	25
5.3	Les fréquences observées	26
5.4	Les ratios sinistres à primes toutes garanties	27
5.4.1	Écrêtement et sinistres graves	29
5.4.2	Les ratios sinistres à primes par garantie	30
5.5	Sinistralité par tranche de prime	33
5.6	Sinistralité des contrats avec conservation	33
5.7	Sinistralité par activité et par région	33
5.8	Constats	34
5.9	Les mesures tarifaires flottes	34
III	Création d'une base de données	36
6	Le périmètre	37
6.1	Le produit considéré	37
6.2	La période de temps	37
6.3	Les individus	37
6.4	Les variables	38
7	La construction de la base de données	39
7.1	Les systèmes d'informations GMF	39
7.2	Les infocentres risques spécifiques	40
7.3	Les primes	40
7.3.1	Les infocentres primes GMF	40
7.3.2	Primes émises	42
7.3.3	Primes acquises	43
7.4	Les sinistres	43
7.4.1	Les infocentres sinistres	43
7.4.2	Les charges sinistres	44
7.4.3	Le nombre de sinistres	45
7.4.3.1	Les provisions	45
7.5	Les contrats avec conservation	45
7.6	Les parcs	45
7.7	Les informations clients	45
7.8	Les données contrats	46
7.9	Les franchises	46
7.10	Le nettoyage de la base	46
7.11	Création de nouvelles variables	47
7.11.1	Ancienneté du client	47
7.11.2	Ratio sinistres à primes	47
7.11.3	Fréquence de sinistralité	48
7.12	Données manquantes	48

8	Estimations des données manquantes	49
8.1	Estimation des parcs	49
8.2	Estimation de la répartition des véhicules	52
8.3	Synthèse récupération/collecte des données	57
8.4	Synthèse des informations disponibles dans la base	57
IV	La théorie de la crédibilité avancée	58
9	Crédibilité de stabilité	59
9.1	Introduction	59
9.2	La naissance de la théorie de la crédibilité	59
9.3	La crédibilité complète	60
9.4	La crédibilité partielle	61
10	Crédibilité de précision	62
10.1	Crédibilité bayésienne	62
10.2	Crédibilité linéaire	64
10.2.1	Le modèle de Bühlmann	64
10.2.2	Le modèle de Bühlmann-Straub	66
10.2.2.1	Estimation des paramètres	67
10.2.2.2	Ordre des calculs	69
10.2.3	La crédibilité hiérarchique (Jewell)	69
V	La mise en place de la surveillance du portefeuille en appliquant la théorie de la crédibilité avancée	72
10.3	Le modèle de surveillance historique	73
10.4	Le modèle de surveillance de base	74
11	L'application du modèle de Bühlmann-Straub	76
11.1	Hétérogénéité du portefeuille	76
11.2	Ratio considéré : le loss-ratio	76
11.2.1	Estimation des paramètres de structure	77
11.2.2	Exemple numérique	77
11.2.3	Le facteur de crédibilité	77
11.2.4	La mise en place d'une stratégie de surveillance	78
11.2.4.1	Application numérique sur un exemple fictif	79
11.2.4.2	Résultats de la revalorisation tarifaire	79
11.3	Ratio considéré : la fréquence de sinistralité RC	80
11.3.1	Estimation des paramètres de structure	80
11.3.2	Le facteur de crédibilité	80
11.3.3	Suite de la mise en place de la stratégie de surveillance	80
11.3.3.1	Application numérique sur un exemple fictif	81
11.3.3.2	Résultats de la revalorisation tarifaire	81
11.4	Tests de sensibilité du modèle de Bühlmann-Straub	81
12	L'application de la crédibilité hiérarchique	83
12.1	L'analyse en composantes principales	83
12.1.1	La théorie	83
12.1.2	L'application	85
12.2	La Classification ascendante hiérarchique	93
12.2.1	La théorie	93
12.2.2	L'application	94
12.2.3	Caractérisation des classes	95
12.2.3.1	Caractérisation des classes par les variables catégorielles	95
12.2.3.2	Caractérisation des classes par les variables numériques	96
12.2.3.3	Caractérisation des classes par les axes factoriels	97
12.2.3.4	Parangons des classes	98

12.2.3.5	Caractérisation des classes par un arbre	98
12.3	La crédibilité hiérarchique	101
12.3.1	Ratio considéré : Le loss-ratio	101
12.3.1.1	Crédibilité des classes	101
12.3.1.2	Application numérique sur un exemple fictif	101
12.3.2	Ratio considéré : la fréquence de sinistralité RC	102
12.3.2.1	Crédibilité des classes	102
12.3.3	Mise en place des revalorisations tarifaires à partir du modèle de crédibilité hiérarchique	102
12.3.3.1	Résultats des revalorisations tarifaires par classe	102
12.4	L'application du modèle de surveillance sans projection des ratios	103
12.5	Résultats des différents modèles de surveillance	103
Conclusion		104
bibliographie		107

Première partie

Présentation du produit d'assurance
flottes automobiles

Chapitre 1

Le contexte général de l'assurance flottes automobiles en France

1.1 Une flotte automobile

Une flotte automobile est définie comme étant l'ensemble du parc automobile d'une entreprise. Au sein d'une flotte, on peut retrouver différents types de véhicules : des voitures de fonction, des véhicules professionnels, des deux-roues, des tracteurs, des semi-remorques, des camions, etc.

1.2 Une assurance obligatoire

Depuis la loi du 27 février 1958, tout propriétaire d'un véhicule terrestre à moteur destiné à circuler doit être assuré. L'obligation d'assurance ne concerne que la garantie responsabilité civile (assurance au tiers) destinée à indemniser les dommages causés aux tiers et aux passagers (exemples : blessures ou décès d'un piéton, d'un passager ou d'un occupant d'un autre véhicule, dégâts causés aux autres voitures, etc.). Une société utilisant des véhicules motorisés (lui appartenant ou non) dans le cadre de son activité professionnelle a donc l'obligation de souscrire au minimum une assurance au tiers pour son parc automobile.

1.3 Les contrats individuels

Une première option est d'assurer chaque véhicule indépendamment des autres (c'est-à-dire souscrire des contrats auto individuels pour chaque véhicule) mais cela implique généralement des difficultés de gestion, d'autant plus si la flotte est importante.

1.4 Les contrats flottes

Pour répondre aux besoins des entreprises, les compagnies d'assurances ont donc développé un produit d'assurance "flotte" destiné à tous les types d'activité (taxi, VTC, transporteur routier, etc.) et à tous les types de véhicules. Ce produit a la particularité d'assurer l'ensemble du parc automobile d'une entreprise (4 véhicules minimum), véhicules lourds comme véhicules légers, au sein d'un seul et même contrat d'assurance. Généralement, les contrats flottes sont plus avantageux financièrement pour l'entreprise (remises commerciales en fonction du parc, absence de bonus-malus, force de négociation) qu'une addition de contrats autos individuels. En effet, souscrire un contrat flotte plutôt que plusieurs contrats monovéhicules permet le plus souvent de bénéficier d'un tarif privilégié grâce à la mutualisation des risques. Ces produits présentent également l'intérêt de simplifier grandement la gestion (une seule police pour toutes les voitures, une seule date d'échéance, aucun conducteur désigné, renouvellement simplifié). Habituellement, le souscripteur est le chef d'entreprise, bénéficiant de la responsabilité civile à son nom. Il est tenu responsable des cotisations et de la déclaration des sinistres.

Il peut arriver que des contrats flottes soient souscrits par un représentant d'un groupement d'intérêt collectif (regroupement de collectivités, association, courtier, groupement professionnel, etc.). Ces flottes sont qualifiées de "flottes artificielles" en opposition aux "flottes naturelles" (véhicules appartenant ou loués par une entreprise).

Pour information, les contrats flottes, initialement prévus pour les entreprises, se sont développés au fil du temps, et peuvent être dorénavant souscrits par des particuliers disposant de plusieurs véhicules et souhaitant les regrouper au sein d'une même police.

Les entreprises sont généralement amenées à évoluer au fur et à mesure du temps (diversification des activités, nouveaux marchés, etc.), entraînant potentiellement des mouvements au sein de leur parc (ajout ou suppression de véhicules). Pour les accompagner au mieux, les compagnies d'assurances ont mis en place des contrats s'adaptant aux mouvements du parc : les contrats à flotte fermée et les contrats à flotte ouverte (ou parc flottant).

1.4.1 Les contrats à flotte fermée

Les entreprises souscrivant des contrats à flotte fermée sont contraintes de déclarer à l'assureur le moindre mouvement au sein de la flotte, qu'il s'agisse d'un ajout ou d'un retrait de véhicule. Ensuite, à la fin de la durée de couverture, l'assureur détermine le prorata de prime à percevoir (ajouts de véhicules) ou au contraire, le remboursement à effectuer au client (suppressions de véhicules). Ces contrats concernent généralement des petites flottes.

1.4.2 Les contrats à flotte ouverte

À contrario, les contrats à flotte ouverte dispensent l'entreprise de déclarer des modifications du parc durant la période assurée. L'entreprise se limite à communiquer à l'assureur la composition de son parc en début et fin de période du contrat. Ces types de contrats sont plus adaptés aux entreprises possédant un parc important.

1.5 L'assurance mission

Le produit d'assurance "mission" est complémentaire (mais distinct) de l'assurance flotte et permet d'assurer (en tous risques) les salariés de l'entreprise utilisant leur véhicule personnel (4 roues < 3.5T, 2 ou 3 roues) pour les besoins professionnels. En règle générale, la souscription d'un contrat flotte implique automatiquement la souscription d'un contrat "mission". À noter que lors d'un sinistre responsable, le coefficient de réduction/majoration (récompensant les bons conducteurs par un niveau de prime plus faible que celui des conducteurs responsables d'accidents) du salarié n'est pas impacté puisque le sinistre n'est pas pris en charge sur son propre contrat.

1.6 Les différentes couvertures possibles

Pour rappel, le point fort de ce produit est sa flexibilité : tout est négociable (plafonds des garanties, franchises, prix, etc.) afin de répondre au mieux au besoin du client. D'ailleurs, les formules de garantie peuvent être différentes d'un véhicule à l'autre au sein de la même flotte. En général, les véhicules âgés de 5 ans ou plus ne sont pas assurables en tous risques, mais seulement au tiers. Les garanties des produits d'assurance flottes sont très similaires d'une compagnie d'assurance à une autre. L'assurance flotte étant un produit d'assurance auto, on y retrouve globalement les mêmes garanties. Différentes formules sont évidemment possibles (allant de l'assurance au tiers à l'assurance tous risques). La garantie responsabilité civile étant la seule obligatoire, les autres sont facultatives. Ces options et extensions de garantie concernent les dommages causés au véhicule et les biens à l'intérieur du véhicule. Pour synthétiser, l'assurance flotte comprend des garanties de base, des garanties dommages et des garanties optionnelles.

1.6.1 Les garanties de base

- Responsabilité civile : cette garantie couvre exclusivement les dommages que pourrait causer la voiture assurée à un tiers ou à un autre véhicule.
- Défense pénale et recours suite à un accident : cette garantie permet à l'assuré d'être représenté et défendu par la compagnie d'assurance dans le cadre d'une procédure de justice l'opposant à un tiers. Les honoraires de l'avocat et les frais d'expertise sont pris en charge par la compagnie d'assurance.
- Assistance : des services d'assistance en cas de panne, 24h sur 24 sont généralement proposés par les compagnies d'assurance. Le remorquage du véhicule, l'hébergement du conducteur ou encore une voiture de remplacement (le temps des réparations) sont également inclus dans la garantie assistance.
- Garantie du conducteur : Protection prenant en charge les dommages corporels subis par le conducteur assuré (frais médicaux, préjudice subi et éventuellement la perte des revenus liée à un arrêt de travail), indépendamment de sa responsabilité dans l'accident.

1.6.2 Les garanties dommages

- Bris de glace : Indemnisation des dommages subis par le pare-brise du véhicule assuré.
- Tempête : Prise en charge des dommages causés au véhicule par une tempête.
- Incendie et vol : Indemnisation du véhicule lorsque celui-ci est endommagé ou détruit par le feu, ou encore lorsqu'il est volé. Le montant de l'indemnité varie selon la police d'assurance (valeur de la voiture au moment du sinistre ou valeur conventionnelle indiquée dans le contrat).
- Dommages tous accidents (tous risques) : Prise en charge de tous les dommages subis par le véhicule, peu importe les circonstances de l'accident ou la responsabilité du conducteur.
- Catastrophes naturelles : Couverture du véhicule contre les dégâts causés par une catastrophe naturelle (inondation, tremblement de terre, avalanche, etc.). L'état de catastrophe naturelle doit faire l'objet d'un arrêté interministériel au Journal Officiel.
- Attentats : Indemnisation des dommages résultant d'actes de terrorisme et d'attentats commis sur le territoire national.

1.6.3 Les garanties optionnelles

- Objets transportés : indemnisation des dommages causés aux marchandises et aux objets transportés.
- Bris interne : Cette option concerne les entreprises dont l'activité repose sur une ou plusieurs machines et équipements industriels (engins de levage, de chantier et agricoles) dont l'arrêt pourrait mettre en danger son activité. La garantie couvre les engins et les matériels de chantier contre les dommages accidentels pouvant survenir lorsqu'ils sont en activité, en stationnement, pendant leur transport et pendant les opérations de chargement et de déchargement.

1.7 Franchises

Le montant de l'indemnité d'un sinistre dépend de la formule de garantie souscrite et des éventuelles franchises (indemnité restant à la charge de l'assuré). Les contrats flottes proposent plusieurs niveaux de franchise en fonction des garanties et de la catégorie du véhicule (2 roues, camion, engins de chantier, etc.). Encore une fois, les niveaux de franchises sont négociables. L'entreprise peut également demander un seul niveau de franchise si elle le souhaite. Les franchises permettent de jouer sur la prime : des franchises élevées impliquent une prime plus faible et inversement.

1.8 Contrats avec conservation de sinistres

De nos jours, les entreprises sont de plus en plus soucieuses de maîtriser le coût de leur assurance flottes quitte à mettre en place des montages financiers d'auto-assurance (conservation des sinistres). En pratique, ce système permet aux entreprises une optimisation de leur budget d'assurance grâce à une diminution de la prime. Le principe consiste à sortir la charge de fréquence (sinistres dont le coût unitaire est faible, mais avec une fréquence élevée) de la cotisation pour la percevoir sous forme d'un dépôt d'assurance (appelée franchise annuelle ou conservation). Cette somme (qui n'est pas une prime d'assurance) permet de payer les sinistres de fréquence au fur et à mesure de leur déclaration jusqu'à épuisement de la conservation. Ensuite, c'est la prime d'assurance qui règle les sinistres suivants. L'avantage du dépôt d'auto-assurance réside dans le fait qu'il n'est pas soumis aux taxes et contributions d'assurance (34,2% de la prime technique chargée des frais de gestion en assurance de responsabilité civile et 18% en assurance dommages au véhicule). Cependant, l'assureur prend des frais de gestion sur le montant de la conservation (gestion du montage, suivi). Approximativement, ce montage financier permettrait de réaliser une économie de l'ordre de 12%. Actuellement, 10% des flottes possèdent un contrat auto-assuré. La fiscalité étant de plus en plus conséquente, ce chiffre est sans doute amené à augmenter dans le futur.

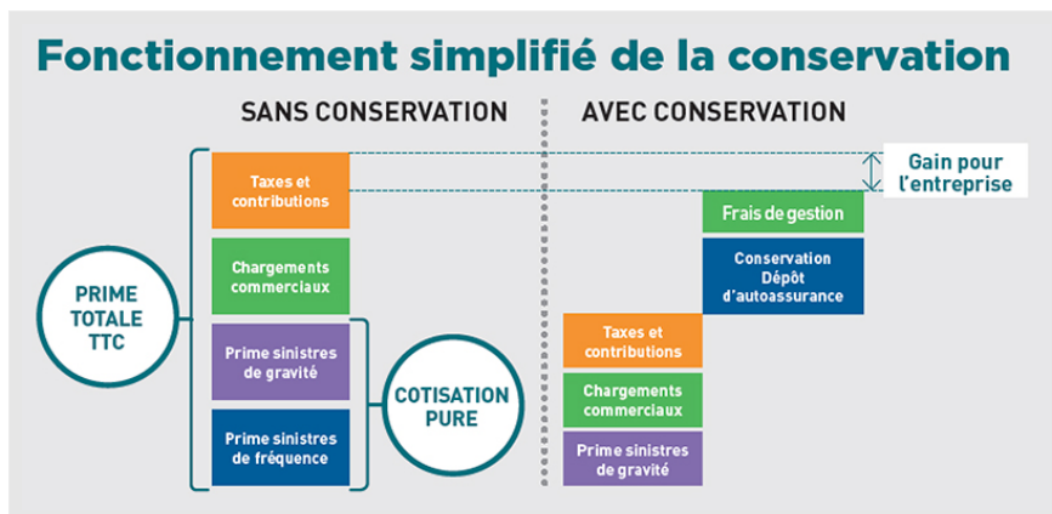


FIGURE 1.1 – Fonctionnement de la conservation (source Arts&MétiersMag)

1.9 Les plans de prévention

Les formules de garantie, les franchises et les mécanismes de conservation permettent à l'entreprise d'optimiser sa prime d'assurance. Toutefois, les assureurs recommandent de mettre en place une prévention interne auprès des conducteurs pour diminuer la fréquence des sinistres. La cotisation étant fortement impactée par l'historique de sinistralité de la flotte, il est donc dans l'intérêt de l'entreprise de la limiter. Cela permettra d'obtenir une cotisation plus abordable pour la prochaine période d'assurance, de réduire les dépenses de réparations et tout simplement de conserver les véhicules en meilleur état.

1.10 Tarification des produits flottes

Bien que la tarification de ce produit soit bien évidemment propre à chaque compagnie d'assurance, les paramètres influençant la prime d'assurance d'une flotte automobile sont sensiblement les mêmes d'une compagnie à l'autre :

- ▶ Le nombre de véhicules ;
- ▶ Leur type (véhicules légers, 2 roues, engins de chantier, etc.) ;
- ▶ Leur âge ;
- ▶ Leur usage (manutention, tourisme, transport de marchandises, etc.) ;
- ▶ La sinistralité antérieure de l'ensemble de la flotte.
- ▶ L'activité professionnelle de l'entreprise ;
- ▶ La zone de circulation ;

En règle générale, le premier réflexe de l'assureur est d'analyser la sinistralité antérieure de la flotte sur une période donnée (entre 3 et 5 ans) pour identifier le risque de l'assuré. Pour cela, il étudie le nombre de sinistres survenus, le montant de ces derniers et les garanties impactées. Il peut également calculer la fréquence et le coût moyen de chaque garantie pour affiner son analyse et détecter d'éventuelles tendances. Une attention particulière est portée sur la garantie responsabilité civile corporelle où les sinistres peuvent être particulièrement graves et donc engendrer des coûts importants. En plus de l'historique de la sinistralité, l'assureur tient également compte des autres paramètres impactant la prime (montant des franchises, conservations, frais de gestion, commissions du courtier, taxes). La particularité de ce produit d'assurance auto est l'absence de coefficient réduction-majoration pour l'assuré. En conséquence, le client dispose d'une force de négociation plus importante. Il a donc tout intérêt à comparer l'ensemble des offres disponibles sur le marché avant de faire son choix.

1.11 Modes de distribution de l'assurance flotte

Deux options sont possibles pour assurer une flotte automobile professionnelle. La première est de souscrire directement un contrat auprès d'assureurs. La deuxième est de passer par le biais d'intermédiaires d'assurance (agents généraux ou courtiers).

Les agents généraux sont majoritaires en province. Leurs portefeuilles sont composés en grande partie de flottes de taille moyenne à petite. Un agent général représente une société d'assurance dont il est le mandataire. C'est-à-dire qu'il vend des contrats à ses clients pour le compte d'une société d'assurance lui apportant en échange un soutien commercial et administratif.

Toutefois, la majorité des entreprises font appel à un courtier en assurances professionnelles pour gérer leur parc automobile. Les courtiers génèrent l'essentiel du chiffre d'affaires, certains disposent même de structures spécialisées dans les flottes. Le rôle d'un courtier est de réaliser un audit de la situation de l'entreprise, de déterminer précisément ses besoins et de proposer les solutions les plus adaptées au meilleur prix. Il est indépendant de ses compagnies partenaires et a pour unique but de satisfaire son client. Il a en charge les différentes étapes de la vie d'une flotte automobile : la recherche de la compagnie, la tarification, la mise en place du contrat, la gestion des entrées et sorties de véhicules, la gestion des sinistres et la renégociation du contrat (ou la recherche d'une autre compagnie).

Chapitre 2

L'assurance flottes automobiles en chiffres

2.1 Les principaux clients

Quatre secteurs se démarquent sur le marché des demandeurs d'assurance flotte :

- ▶ Le transport public de voyageurs ou de marchandises (33% des cotisations) ;
- ▶ Le bâtiment et travaux public (13.7%) ;
- ▶ La location (10.2%) ;
- ▶ Les collectivités locales (9.2%).

2.2 Un marché en pleine croissance

4.4 millions : c'est le nombre de véhicules assurés en France, en 2018, par un contrat flotte automobile, représentant 8% de l'ensemble du parc des véhicules assurés. La tendance est à la hausse depuis plusieurs années (3.9 millions en 2017, 3.7 millions en 2016, 4.1 millions en 2018).

En tout logique, la croissance du parc engendre une hausse des cotisations d'assurance de flotte automobile. Selon la Fédération Française de l'Assurance (FFA), en France, les cotisations flottes automobiles représenteraient 10.1% des cotisations totales du marché de l'assurance automobile et 3.9% de l'ensemble des assurances de biens et de responsabilité, soit approximativement 2.3 milliards d'euros en 2019 (en croissance de 5% sur un an).

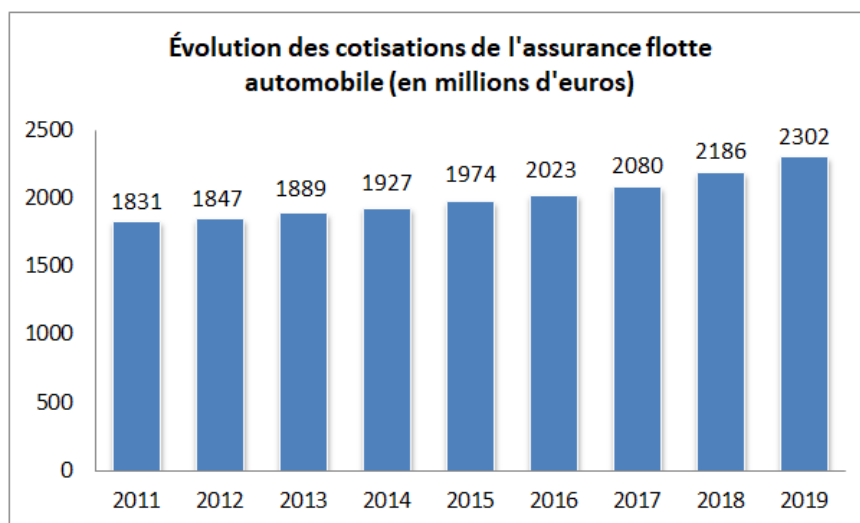


FIGURE 2.1 – Évolution des cotisations flottes en fonction du temps (source FFA)

2.3 Des résultats techniques dégradés

Bien que le marché de l'assurance flotte soit en pleine croissance, cela n'implique pas une hausse de la rentabilité. Les compagnies d'assurance estiment que le ratio d'équilibre sinistres à primes de ce produit se situe aux alentours de 70%. En effet, il est nécessaire d'avoir suffisamment de marge pour être en mesure de payer les frais annexes (commissions du courtier, frais généraux, réassurance). Au-delà, le produit est considéré comme déficitaire. Les résultats se sont sensiblement dégradés depuis 2015. En 2019, la FFA évaluait le ratio sinistres à primes à 85% (contre 73% en 2014), témoignant d'une baisse de la rentabilité de ces contrats.

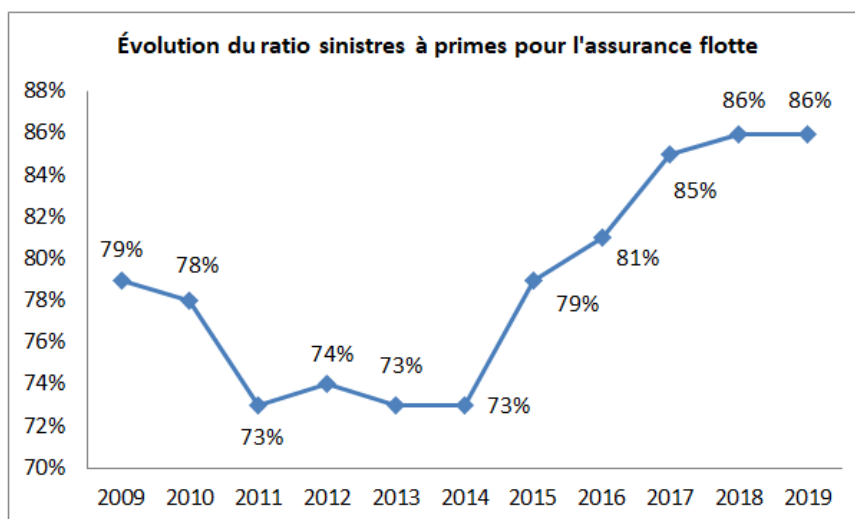


FIGURE 2.2 – Évolution du ratio sinistres à primes de l'assurance flotte en fonction du temps (source FFA)

À noter que les sinistres sont vus à fin décembre de chaque exercice. Le recul est donc faible, une grande partie des sinistres n'ont pas encore été payés à ce moment-là (ce sont des estimations) occasionnant généralement une surévaluation du SP. Toutefois, le niveau élevé du ratio sinistres à primes peut en partie s'expliquer par la forte concurrence que se livrent les compagnies d'assurance sur ce marché, le manque d'attention porté aux véhicules (qui n'appartiennent pas aux conducteurs), et l'augmentation des coûts moyens des sinistres matériels (comme on peut le voir ci-dessous).

Garantie	Fréquence 2019	Variation 19-18	Coût moyen 2019	Variation 19-18
RC matériels	7,83%	-6,5 %	1 210€	+4,0%
VOL	0,41%	+1,7%	4 915€	+2,5%
Bris de glace	7,06%	-8,8%	550€	+8,0%
Dommages tous accidents	6,69%	-2,9%	2 205€	+6,0%

TABLE 2.1 – Fréquences et coûts moyens des garanties principales flottes (source FFA)

Chapitre 3

Le produit flottes automobiles GMF

3.1 GMF

La Garantie Mutuelle des Fonctionnaires (GMF) est une société d'assurance mutuelle (établissement à but non lucratif sans actionnaire à rémunérer), régie par le Code français des Assurances. Au 31 décembre 2019, GMF Assurances comptait 3 600 000 de sociétaires pour un chiffre d'affaires de 1 758 millions d'euros. Elle appartient au groupe d'assurance mutuelle Covéa (MMA, MAAF, GMF), reconnu comme étant le premier assureur de biens et de responsabilité pour les particuliers en France en 2019 (17 milliards d'euros de primes acquises pour un résultat net de 940 millions d'euros). Comme l'indique son nom, GMF cible principalement les fonctionnaires. Elle propose des contrats d'assurance de dommages et de personnes aux particuliers et aux professionnels du service public.

3.2 L'assurance flottes automobiles GMF

GMF propose aux établissements publics souhaitant assurer leur parc automobile le choix entre un contrat à flotte fermée ou un contrat à flotte ouverte (voir définitions dans les paragraphes 1.4.1 et 1.4.2) ainsi que l'assurance mission en complément. Les flottes assurées peuvent être naturelles ou artificielles (voir paragraphe 1.4) et le client a la possibilité de détenir plusieurs contrats flottes en parallèle (subdivision de la flotte en plusieurs sous-flottes, différentes activités, etc.). À noter que l'activité flottes de GMF est relativement récente et n'en est qu'à ses débuts.

Pour information, les assurés souscrivant un contrat à la GMF mais n'étant pas éligibles au statut GMF (particuliers non fonctionnaires, risques spécifiques, **personnes morales**) souscrivent en fait un contrat à la Sauvegarde, société détenue par GMF. Les contrats flottes étant souscrits uniquement par des personnes morales, ce sont donc des contrats Sauvegarde et non GMF.

Le souscripteur doit être exclusivement :

- ▶ Une Administration ;
- ▶ Un établissement public ;
- ▶ Un CSE (comité Social et Économique) ;
- ▶ Une collectivité locale ;
- ▶ une association.

Il n'est pas possible d'assurer les artisans, les commerçants ou les entreprises privées.

Les véhicules assurables sont :

- ▶ 4 roues < 3.5 T (véhicules de tourisme, camionnette, fourgonnette) ;
- ▶ Véhicules de plus de 3.5 T (camion, tracteur routier) ;
- ▶ Engins : engins automoteurs destinés au jardinage (tondeuse autoportée), au terrassement et aux chaussées (goudronneuse) et aux autres travaux (chariot élévateur) - les grues ne sont pas assurables ;
- ▶ Autobus, autocar (du minibus au bus articulé) ;
- ▶ 2 ou 3 roues, quelle que soit la puissance ;
- ▶ Remorques et semi-remorques.

Pour rappel, aucun conducteur n'est désigné en flotte. C'est-à-dire que toute personne autorisée par le souscripteur peut conduire le véhicule, qu'elle soit novice ou expérimentée.

Le produit à flotte ouverte GMF représente 11.3 millions d'euros de primes acquises à fin décembre 2019 contre 2 millions d'euros pour le produit à flotte fermée et 0.8 million d'euros pour l'assurance mission. Les enjeux sont donc nettement plus importants sur les flottes ouvertes. D'autant plus que les résultats techniques du produit à flotte ouverte suivent la même tendance que le marché (ratio sinistres à primes de 86% à fin décembre 2019). Pour information, le produit à flotte fermée et l'assurance mission affichent de meilleurs résultats techniques (ratio sinistres à primes respectivement de 72% et 75%). En conséquence, l'étude portera uniquement sur les flottes ouvertes. Il ne sera pas non plus tenu compte de l'assurance mission.

3.3 La particularité du marché public

Lorsque des organismes publics souhaitent recourir à un assureur pour couvrir leurs risques, ils sont contraints de se conformer à des règles particulières. Ces règles ont pour but de mettre en concurrence les organismes d'assurance postulant à la couverture de ces risques :

- ▶ Liberté d'accès à la commande publique (possibilité pour toute entreprise de proposer une offre) ;
- ▶ Égalité de traitement des candidats (l'examen des offres doit s'effectuer sans parti pris et ne doit pas être discriminatoire) ;
- ▶ Transparence des procédures (justification du choix du titulaire du marché et du rejet des autres candidatures).

D'autres spécificités existent : l'organisme public a la charge de procéder à l'inventaire de ses besoins (possibilité de recourir à un conseil tiers si nécessaire) et la tacite reconduction (accord reconductible si aucune des parties ne manifeste sa volonté de rompre le contrat) est impossible.

Le produit d'assurance flotte de GMF pouvant uniquement être souscrit par des organismes publics, tout se déroule donc via des appels d'offres. En conséquence, tous les contrats présents dans le portefeuille flotte de GMF proviennent d'appels d'offres remportés.

3.3.1 L'apporteur d'affaire

GMF travaille en collaboration avec un courtier. Le courtier aide l'organisme public souhaitant assurer son parc de véhicules à définir ses besoins dans un cahier des charges puis à diffuser l'appel d'offres auprès des compagnies d'assurance. Une fois l'appel d'offres diffusé, le courtier informe GMF de l'affaire en leur proposant une prime d'assurance pour la flotte en question. Les technico-commerciaux flottes de GMF reçoivent l'appel d'offres, l'étudient et proposent à leur tour un prix (possiblement différent que celui du courtier). Puis, ils transmettent l'affaire à l'équipe tarification de GMF qui a la responsabilité de se positionner définitivement sur la prime. Leur objectif est de définir le prix le plus représentatif possible du risque de la flotte sans tenir compte des diverses propositions du courtier et des technico-commerciaux.

Ensuite, s'ils ne sont pas d'accord avec la prime initialement proposée par les technico-commerciaux, une réunion d'échange est organisée pour en discuter et aboutir à un accord. Une fois un consensus trouvé, les technico-commerciaux sont en mesure de répondre à l'appel d'offres.

3.3.2 L'appel d'offres

À chaque fin de période d'assurance, l'organisme public doit impérativement soumettre un nouvel appel d'offres. Il n'existe pas de tacite reconduction sur le marché public. L'appel d'offres peut évidemment spécifier un cahier des charges différent du précédent (période d'assurance, formules de garanties, etc.).

Voici les informations à disposition de l'assureur sur l'appel d'offres :

- Nom de l'assureur actuel ;
- Durée du nouveau contrat ;
- Date d'effet du nouveau contrat (généralement le 1er janvier) ;
- Durée du préavis ;
- Parc actuel à assurer et répartition selon le type de véhicule et la formule (-3T5 tous risques, -3T5, +3T5 tous risques, +3T5, 2 et 3 roues, engins tout tonnage, remorques) ;

Parc à assurer :	617	%/parc
- 3T5 TR	8	1,30%
-3t5	508	82,33%
+3t5	18	2,92%
+3t5 Tr		0,00%
2/3R	11	1,78%
Engins tout tonnage	27	4,38%
Remorques	45	7,29%

FIGURE 3.1 – Exemple d'un parc d'une collectivité

- Formules de garanties selon le type de véhicule ;
- Franchises selon le type de véhicule ;
- Historique de sinistralité (par garantie) de la flotte sur une période donnée (données fournies par l'assureur actuel).

SINISTRES ENREGISTRÉS ET REGLES

	SIN	RCM	COUT	RCC	COUT	DOM	COUT	VOL	COUT	INC	COUT	BDG	COUT	TOTAL COUT
2018	38	34	39 570 €			2	1 987 €	2	11 526 €					53 083 €
2019	37	30	20 304 €			3	1 290 €	2	- €			2	1 898 €	23 492 €
2020	33	29	58 530 €			2	1 145 €		- €	2	15 700 €			75 375 €
TOTAL	75	93	118 404 €	0	- €	7	4 422 €	4	11 526 €	2	15 700 €	2	1 898 €	151 950 €
Moy sur 3 ans	25	31	39 468 €	0	- €	2	1 474 €	1	3 842 €	1	5 233 €	1	633 €	50 650 €

FIGURE 3.2 – Exemple de l'historique de sinistralité du parc d'une collectivité

L'organisme public a également la possibilité, sur l'appel d'offres, de demander la tarification de plusieurs formules (par exemple : une formule sans la garantie bris de glace, une formule avec la garantie bris de glace, des niveaux de franchises différents, etc.).

À partir de ces éléments, les assureurs proposent une prime d'assurance. Ensuite, le client étudie avec l'aide du courtier, les différentes offres et sélectionnent celle répondant le plus possible aux besoins du client. Ce n'est pas forcément la proposition la plus avantageuse financièrement qui est retenue, d'autres éléments rentrent en compte (couverture, plafonds des garanties, gestion des sinistres, etc.). Le système est le suivant : des points sont attribués à chaque catégorie (prix, gestion, garanties, etc.) et l'assureur récoltant le plus de points (sur 100) remporte l'appel d'offres.

3.3.3 La tarification des appels d'offres flottes chez GMF

Les informations étant très limitées sur l'appel d'offres, les méthodes de tarification ne sont donc pas nombreuses. La majorité des assureurs fonctionne de la manière suivante :

- Calculs des coûts moyens par garantie et comparaison avec les coûts moyens de référence de chaque garantie ;
- Estimation du nombre de sinistres par garantie (moyennes ou tendances) ;
- Estimation de la sinistralité de chaque garantie ;
- Agrégation des garanties pour déterminer la sinistralité totale de la flotte ;
- Détermination d'une provision pour sinistres graves (montant fixe que l'on multiplie par le nombre de véhicules de la flotte) ;
- Détermination du coût total de la flotte (addition de la sinistralité totale et du coût des graves) ;
- Prise en compte des commissions du courtier (10% de la prime) et des frais généraux (15% de la prime) pour déterminer le tarif hors taxe de la flotte.
- Ajout des taxes pour calculer la prime toutes taxes comprises à proposer au client.

La tarification étant basée majoritairement sur l'historique de sinistralité, elle est donc propre au client. Une fois la cotisation estimée, il est coutume de calculer la prime par véhicule en divisant le montant de la prime hors taxe par le nombre de véhicules présents dans la flotte. Cet indicateur permet de se faire une idée générale du risque de la flotte. En effet, plus la prime par véhicule est élevée, plus la flotte présente un risque important (sinistralité potentiellement dégradée, formules de garanties tous risques, etc.). La réciproque n'est pas forcément vraie, car les flottes avec conservations des sinistres (franchise annuelle, voir paragraphe 1.8) paient une prime beaucoup plus faible. En conséquence, la prime par véhicule de ces flottes n'est pas très élevée. Cet indicateur est à analyser avec précaution, il serait faux d'affirmer que tous les véhicules coûtent le même prix en termes d'assurance, car ils ne portent pas tous le même risque (différents types de véhicules, différentes formules de garanties, etc.).

La tarification dépend évidemment de la stratégie de l'assureur. Certaines compagnies ont pour objectif de remporter le plus d'appels d'offres possibles pour augmenter le chiffre d'affaires. Elles sont prêtes à réaliser des concessions très importantes sur la prime pour augmenter leurs probabilités de succès. Cependant, à côté, leur stratégie de surveillance est très sévère et dès qu'une flotte présente des résultats particulièrement mauvais, le contrat est très vite résilié. GMF, de son côté, a une stratégie qu'on pourrait qualifier de prudente. L'appel d'offres est analysé en détail, des réunions d'échanges sont mises en place entre les actuaires, les technico-commerciaux et le courtier pour statuer sur l'appel d'offres. Finalement, si l'historique de sinistralité est mauvais et se dégrade significativement avec le temps, GMF n'hésite pas à se retirer de l'appel d'offres. De manière générale, bien que des concessions financières sont faites de temps en temps, le prix technique proposé par GMF reflète le risque de la flotte. Historiquement, la surveillance du portefeuille était donc réalisée "à l'entrée", lors de l'appel d'offres. Voilà pourquoi, GMF ne jugeait pas pertinent le développement d'une surveillance annuelle. Les mauvais résultats des dernières années ont changé la donne et ont montré qu'une stratégie prudente en termes de tarification n'était pas toujours synonyme de rentabilité et de profit sur ce produit.

3.3.4 La Difficulté du prix juste

Lors de la tarification d'un appel d'offre, deux situations sont possibles :

- ① GMF était l'assureur de la flotte lors de la précédente période d'assurance ;
- ② La flotte était assurée par un assureur tiers.

Ces deux situations sont très différentes et impactent grandement la tarification. Dans le premier cas, GMF est en position de "défense". Étant le précédent assureur de la flotte, elle a accès à des informations très importantes que les autres assureurs n'ont pas (primes d'assurance et donc niveaux de ratio sinistres à primes, évolutions du parc, etc.). Inversement, dans le deuxième cas, un autre assureur détenait le marché de la flotte jusqu'à présent et a donc l'avantage sur la GMF.

Partons du principe que GMF n'était pas le précédent assureur de la flotte. Dans ce cas, les informations disponibles sur l'appel d'offres sont très succinctes, rendant le travail de tarification particulièrement délicat. Il est généralement très difficile de proposer un prix représentatif du risque de la flotte pour plusieurs raisons. Tout d'abord, l'historique de sinistralité n'est pas assez précis. Les sinistres sont nets de franchises (franchise déjà appliquée) et pour chaque sinistre, il n'est pas possible de connaître le type de véhicule impliqué. Les franchises étant différentes selon les types de véhicules, il n'est donc pas possible de connaître la franchise appliquée et donc de déterminer le coût du sinistre hors franchise. Cela est très problématique si l'organisme public demande un tarif avec des franchises différentes de celles appliquées jusqu'à présent. Deuxièmement, le parc renseigné dans l'appel d'offres est le parc à assurer à l'instant "t" mais il n'est pas possible de connaître l'état du parc au cours de la vie du précédent contrat. C'est-à-dire que l'historique de sinistralité de la flotte sur lequel se basent les actuaires n'est pas nécessairement représentatif du parc actuel. Le parc peut avoir augmenté significativement depuis le début du précédent contrat. Cela peut donc conduire à des sous-tarifications. Il est donc essentiel d'être très prudent lors de l'analyse des données de sinistralité et d'essayer de déceler des indices qui pourraient faire penser que le parc a augmenté au cours du temps (augmentation du nombre de sinistres par exemple). Pour finir, l'aspect commercial est fondamental et bien que le prix ne soit pas le seul critère pour remporter un appel d'offres, il est déterminant et ne doit pas être négligé. Généralement, des concessions sont réalisées sur le prix technique afin d'optimiser les probabilités de succès. GMF étant lié à un courtier, il est essentiel de remporter suffisamment d'affaires afin qu'il soit satisfait et qu'il souhaite continuer leur collaboration.

Ces éléments sont en partie à l'origine des mauvais résultats techniques.

Deuxième partie

Statistiques descriptives : analyse de l'évolution du portefeuille flottes GMF (production et sinistralité)

L'objectif de cette partie est de comprendre l'évolution du portefeuille flottes de GMF aussi bien en termes de production qu'en termes de sinistralité. Toutes les données à l'origine des graphes qui vont suivre ont été déterminées à partir des infocentres informatiques flottes automobiles GMF. Bien qu'un tableau de bord mensuel soit généré tous les mois pour suivre le pilotage de l'activité flotte, les indicateurs indiqués n'y sont pas très nombreux en raison du faible enjeu (comparé aux produits cibles de GMF : auto et MRH). Tout a donc été recalculé à l'aide des données brutes. Un point sera fait dans la partie suivante (construction de la base de données) pour expliquer au lecteur l'organisation des bases de données informatiques GMF (contrats, sinistralité, etc.).

Chapitre 4

Portefeuille flottes GMF

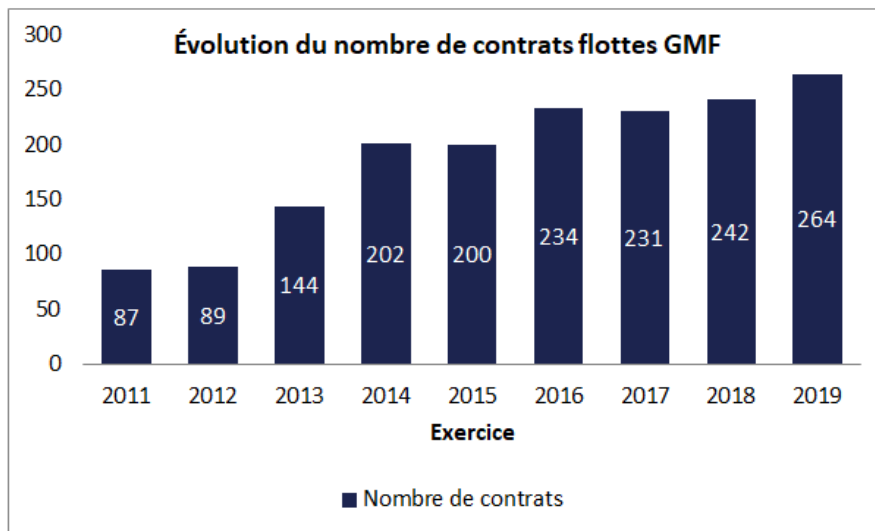


FIGURE 4.1 – Évolution du nombre de contrats flottes GMF

Le nombre de contrats est déterminé à fin décembre de chaque exercice. On constate une très nette progression de l'activité flotte GMF entre 2011 et 2019. De 87 contrats à fin décembre 2011, GMF en compte 264 à fin décembre 2019, soit +203%. On distingue 4 phases d'évolution :

- 2012-2013 : +61% ;
- 2013-2014 : +47 % ;
- 2016-2015 : +12 % ;
- 2019-2018 : +11 %.

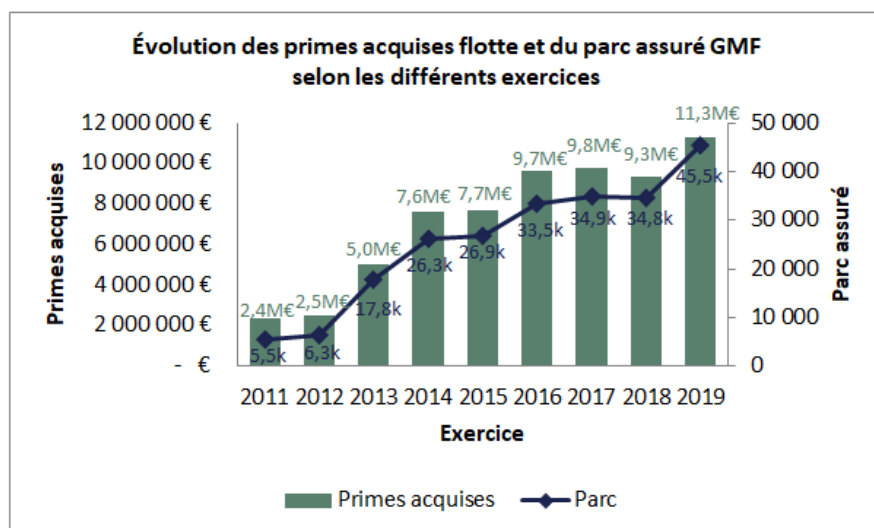


FIGURE 4.2 – Évolution des primes acquises (hors taxes et hors assistance) flotte et du parc assuré GMF selon les différents exercices

Les primes acquises (hors taxes et hors assistance) des différents exercices sont toutes vues à fin juin 2020, cela s'explique par le fait que certaines primes sont payées tardivement (en raison d'ajustements de parc au cours de l'année). Au sujet des parcs, ils sont tous vus à fin décembre de l'exercice de survenance.

Les mêmes phases d'évolution sont constatées pour les primes acquises et le parc de véhicules assurés. À fin décembre 2019, les primes acquises flottes GMF représentent 11,3 millions d'euros et concernent 46 000 véhicules. À titre de comparaison, à fin décembre 2019, les primes acquises GMF du produit auto destiné aux particuliers représentent 985 millions d'euros pour un parc total de 2,9 millions de véhicules.

La question se pose ensuite de l'homogénéité des primes au sein du portefeuille flottes de GMF. Comment sont-elles distribuées? Des contrats sont-ils plus importants que d'autres? Qu'en est-il des parcs assurés? Existe-t-il des disparités en termes de primes et de parcs assurés entre les différents contrats?

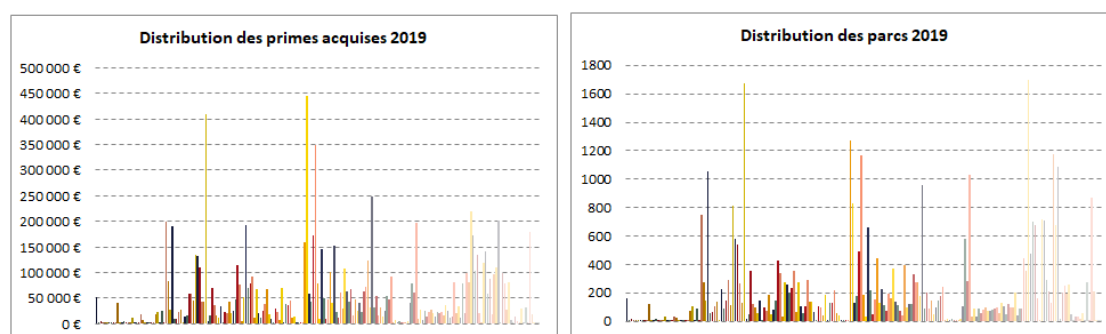


FIGURE 4.3 – Distribution des primes acquises et des parcs automobiles en 2019

En représentant la distribution des primes acquises des contrats flottes GMF en 2019, on constate une forte hétérogénéité au sein du portefeuille. Les grosses primes, peu nombreuses, semblent contraster avec les petites primes, très nombreuses. Concernant la distribution des parcs en 2019, elle présente les mêmes tendances, la prime acquise étant très fortement corrélée au nombre de véhicules assurés.

Rares exceptions pour les contrats ayant opté pour une conservation des sinistres, leur prime étant beaucoup moins élevée qu'une même flotte sans conservation de sinistres. Ce qui explique pourquoi, on ne retrouve pas exactement les mêmes pics entre les 2 distributions.

Quantiles	Primes acquises 2019	Parc 2019
0,25	7 442€	31
0,50	24 051€	102
0,75	62 785€	221

TABLE 4.1 – Quantiles des primes acquises 2019 et du nombre de véhicules assuré par contrat

La prime acquise médiane payée par contrat est de 24000€. Seulement 25% des contrats disposent d'une prime acquise supérieure à 62000€. En ce qui concerne les parcs, on note que 75% des contrats assurent une flotte composée de moins de 221 véhicules.

Tranche de primes acquises 2019	Part de contrats	Part de primes acquises
0-10k	30%	1,7%
10k-20k	13%	4,0%
20k-50k	27%	17,4%
50k-100k	17%	24,3%
100k-200k	10%	30,6%
200k-500k	3%	22,1%

TABLE 4.2 – Répartitions des contrats flottes GMF 2019 selon les tranches de primes acquises

Il apparaît que 43% des contrats (tranche 0-20000€) n'expliquent que 5,7% des primes acquises. Au contraire, les contrats associés à une prime comprise entre 100000€ et 200000€ sont très minoritaires dans le portefeuille (13%), mais représentent à eux seuls 52,7% des primes acquises.

4.1 Répartition du parc automobile

On souhaite déterminer la composition du parc automobile assuré par GMF selon le type de risque (la catégorie du véhicule).

Catégorie de véhicules	Part de la flotte totale	Part des primes acquises 2019
Véhicules 4 roues- 3 tonnes 5	79,32%	80,19%
Véhicules 4 roues + 3 tonnes 5	6,82%	14,82%
2 ou 3 roues	0,89%	0,74%
Remorques - 3 tonnes 5	5,05%	1,17%
Remorques + 3 tonnes 5	0,00%	0,01%
Engins véhicules spéciaux	7,73%	2,37%
Véhicules de transport public	0,18%	0,70%

TABLE 4.3 – Répartition du parc automobile flottes GMF 2019

On observe une surreprésentation des véhicules légers au sein du parc (globalement 80% du parc et 80% des primes acquises totales 2019). On constate également que bien que peu de poids lourds soient présents dans le portefeuille flottes GMF, ils expliquent une part non négligeable des primes. Cela s'explique par des primes plus élevées pour ces véhicules dues à leur coût moyen plus élevé. On relève que le portefeuille assure quelques remorques et quelques véhicules spéciaux (12,78% au total) bien que les primes soient très faibles (seulement 4,54% des primes acquises totales).

4.2 Les primes par garantie

Les primes sont déclinées en primes par garanties. On constate sur la figure suivante qu'à elles seules les primes RC et dommages représentent 72% de l'ensemble des primes acquises.

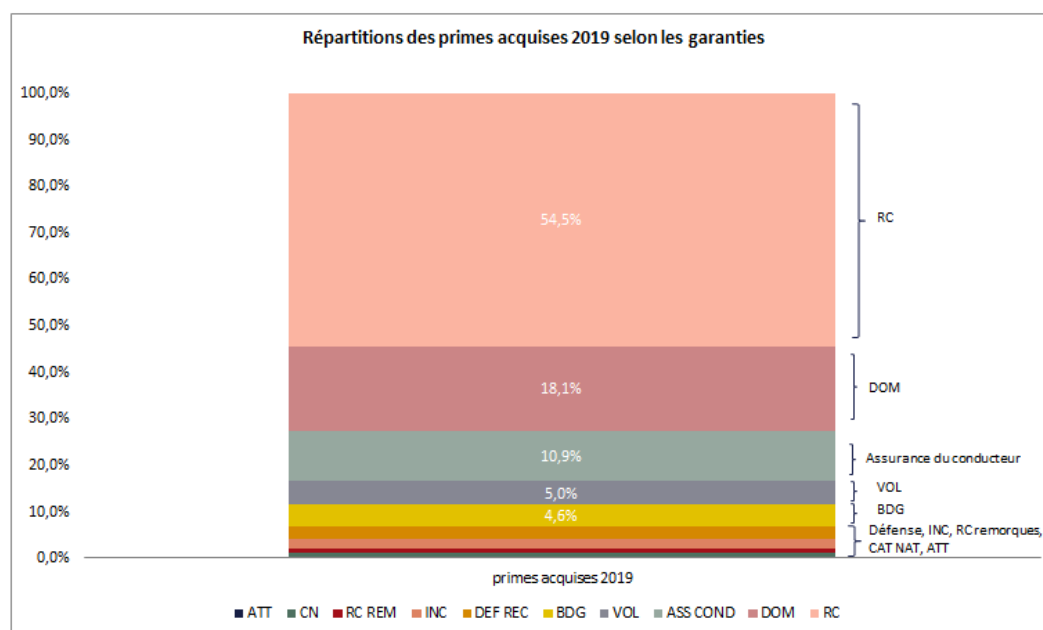


FIGURE 4.4 – Répartitions des primes acquises 2019 selon les garanties

4.3 Localisation du portefeuille

Un assureur se doit de connaître ses clients, cela se traduit par des informations géographiques (localisation de la flotte) et des informations professionnelles (secteur d'activité de la flotte).

Chaque contrat flottes est associé à un code postal, représentant le siège social de l'établissement public. Les données sont agrégées au niveau de la région. L'intérêt de cette section est d'analyser la répartition géographique du portefeuille flottes de GMF en 2019 en se basant sur deux indicateurs : le nombre de contrats et la prime acquise totale par région.

Région	Part de contrats (2019)	Part de primes acquises (2019)
Île-de-France	21,7%	26,8%
Hauts-de-France	35,4%	24,1%
Auvergne-Rhône-Alpes	6,1%	10,2%
Occitanie	4,2%	9,8%
Centre	3,3%	5,0%
Grand est	5,2%	4,8%
Provence-Alpes-Côte-d'Azur	4,7%	4,2%
Autres	19,3%	15%

TABLE 4.4 – Répartitions des primes acquises 2019 selon les régions

On constate que la répartition n'est pas homogène, certaines régions sont surreprésentées tandis que d'autres sont sous-représentées. En effet, à elles seules les régions Île-de-France et Hauts-de-France représentent 57% des contrats et 51% des primes acquises en 2019.

Certaines régions comme l'Île-de-France, l'Auvergne-Rhône-Alpes et l'Occitanie expliquent une part plus importante des primes acquises que des contrats. En conséquence, les contrats associés à ces régions sont associés à des grosses flottes. À l'inverse, la région Hauts-de-France est à l'origine de 35,4% des contrats mais seulement 24,1% des primes acquises, c'est à dire beaucoup de contrats de faibles montants.

La question peut se poser de comprendre l'origine de la concentration du portefeuille en Île-de-France et en Hauts-de-France. En effet, GMF n'a défini aucune stratégie géographique quant à la sélection des appels d'offres, tout dépend exclusivement de la sinistralité de la flotte. Comment expliquer alors cette surreprésentation de ces 2 régions ? Il est plutôt logique de trouver la région Île-de-France à cette position. En effet, de nombreux établissements publics sont situés en région parisienne et ont besoin d'assurer des parcs de véhicules conséquents. On assiste à surreprésentation des établissements publics dans cette région. Concernant les Hauts-de-France, aucune réelle explication, mais seulement des hypothèses. Les appels d'offres sont apportés à GMF par l'intermédiaire du courtier, peut-être a-t-il tendance à favoriser cette région.

4.4 Secteur d'activité du portefeuille

Pour rappel, le portefeuille est composé de contrats flottes faisant suite à des appels d'offres émis sur le marché public. Cependant, le marché public est composé d'une multitude de secteurs d'activité différents. L'idée est d'analyser la répartition des titres professionnels des flottes présentes en portefeuille en 2019.

Activité professionnelle	Part de contrats (2019)	Part de primes acquises (2019)
Commune	36,8%	38,0%
Département	5,7%	16%
Établissement public	17,9%	15,7%
Caisse publique	9,9%	10,4%
Association	10,4%	7,8%
Autres	19,3%	11,8%

TABLE 4.5 – Répartitions des primes acquises 2019 selon les secteurs d'activités professionnels

Les flottes appartenant aux communes sont surreprésentées au sein du portefeuille, elles expliquent 38% des primes acquises. Viennent ensuite les flottes liées à des départements, généralement de grande envergure (16,3% des primes acquises totales contre 5,7% de l'ensemble des contrats). On retrouve également de nombreuses flottes associées à des établissements publics ou à des caisses (santé, retraite, allocations familiales, etc.).

4.5 Ancienneté des clients

Un contrat flotte dure généralement entre 3 et 5 ans. À chaque échéance, l'organisme public a le devoir de diffuser un nouvel appel d'offres. Il arrive que certains assureurs remportent successivement le même contrat plusieurs fois de suite, mais il arrive également qu'un client opte pour une nouvelle compagnie d'assurance et quitte donc le portefeuille. Au sein du portefeuille flottes GMF, certains clients sont présents depuis longtemps tandis que d'autres sont nouveaux. Cela a des impacts sur la sinistralité, un client présent en portefeuille depuis longtemps a une expérience de sinistralité censée être plus fiable qu'un nouvel entrant. L'idée de cette sous-partie est d'observer la répartition des différentes anciennetés des clients flottes de GMF au sein du portefeuille.

Années d'expérience	Nombre de clients	Part des clients
0	54	29,7%
1	38	20,1%
2	26	14,2%
3	13	7,1%
4	11	6%
5	5	2,7%
6	21	11,5%
7	14	7,7%

TABLE 4.6 – Répartition de l'ancienneté des clients

71% des clients ont moins de 3 ans d'ancienneté, le portefeuille est relativement jeune. Toutefois, on note que 19% des clients ont plus de 6 ans d'ancienneté.

4.6 Zoom sur les contrats avec conservation

Le portefeuille flottes contient des contrats avec conservation de primes (franchises annuelles). Il est important de distinguer ces contrats des autres, car leur fonctionnement est atypique.

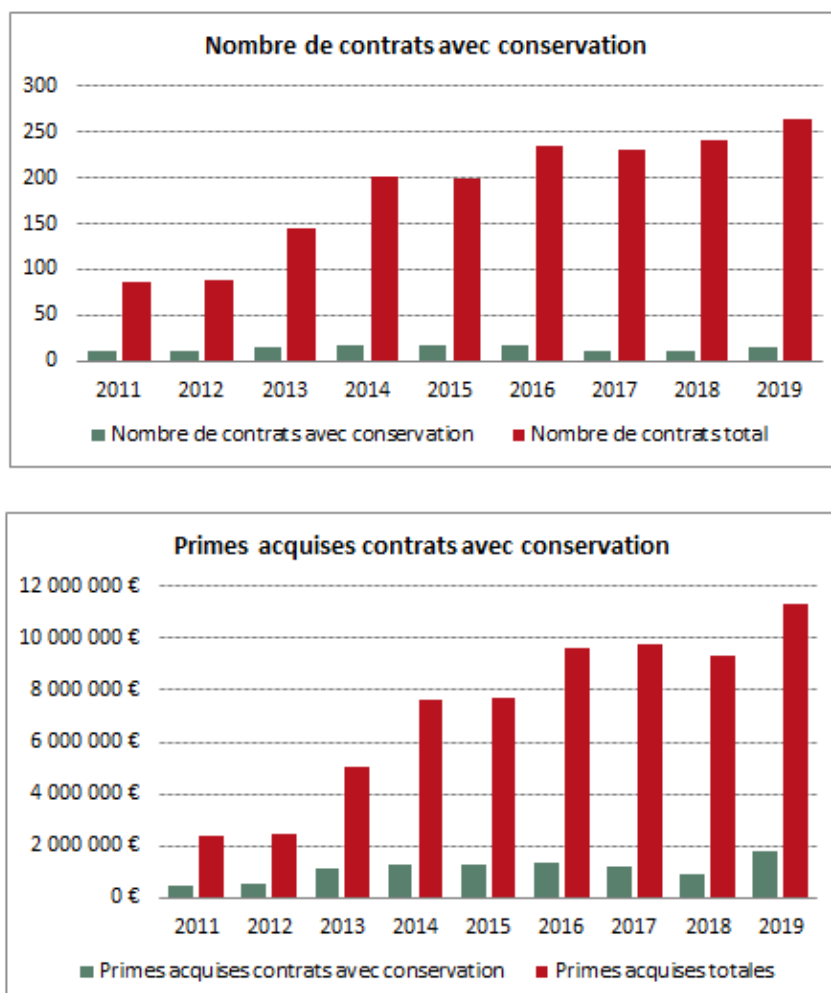


FIGURE 4.5 – Répartition des contrats avec une franchise annuelle en termes de nombre et de primes acquises

À fin décembre 2019, le portefeuille flottes GMF compte 15 contrats avec une franchise annuelle, soit 5,7% du nombre de contrats total. Ces 15 contrats représentent 1,8 million d'euros de primes acquises, soit 15,7% de l'ensemble des primes acquises. Bien que ces contrats représentent une part négligeable du nombre de contrats total, ils pèsent un poids relativement important dans les primes acquises. Les primes annuelles de ces contrats sont généralement de l'ordre de la centaine de milliers d'euros.

Chapitre 5

Sinistralité flottes GMF

Tous les sinistres sont pris en compte dans les graphes ci-dessous excepté les sinistres provenant de contrats avec conservations et les sinistres graves (supérieurs à 150000€). En ce qui concerne les graves, ces sinistres possèdent une liquidation très lente et, de par leurs poids importants, biaiserait les taux de liquidation. Les sinistres issus de contrats avec conservation ne sont pas non plus pris en compte dans les calculs, car une partie de ces sinistres ne sont pas payés par GMF.

5.1 La liquidation des différentes garanties

Pour rappel, le coût d'un sinistre se décompose en deux parties : les montants déjà versés et les provisions (montants estimés des frais qu'il reste à payer). À l'ouverture du sinistre, en fonction des circonstances, le gestionnaire provisionne un montant censé refléter le plus justement possible le coût qu'occasionnera le sinistre à sa clôture. Généralement, pour des sinistres classiques, la provision d'ouverture est égale au coût moyen de la garantie en question. Initialement, le coût d'un sinistre est donc égal à la provision. Puis, au fur et à mesure de la vie du sinistre, des montants sont versés et les provisions sont réajustées en conséquence jusqu'à devenir nulles. Un sinistre est considéré comme clos lorsque sa part de provision est nulle. Certains sinistres sont réglés plus rapidement que d'autres, cela dépend des garanties concernées et de la gravité du sinistre. Pour chaque garantie et pour chaque année d'exercice N fixée, l'objectif est de déterminer la part de provision au sein des charges à fin N , $N+1$, $N+2$ et $N+3$. L'idée derrière les graphes ci-dessous est de voir si certaines garanties sont plus sujettes à une liquidation lente des sinistres. En effet, dans le cadre de la surveillance du portefeuille, une base de données sera créée contenant les sinistres des différents clients par exercice de survenance ; la question de la date d'observation des sinistres sera amenée à être posée (à fin décembre de chaque exercice ? Avec une année de recul ? etc.). Les graphes ci-dessous permettront d'apporter des éléments de réponse.

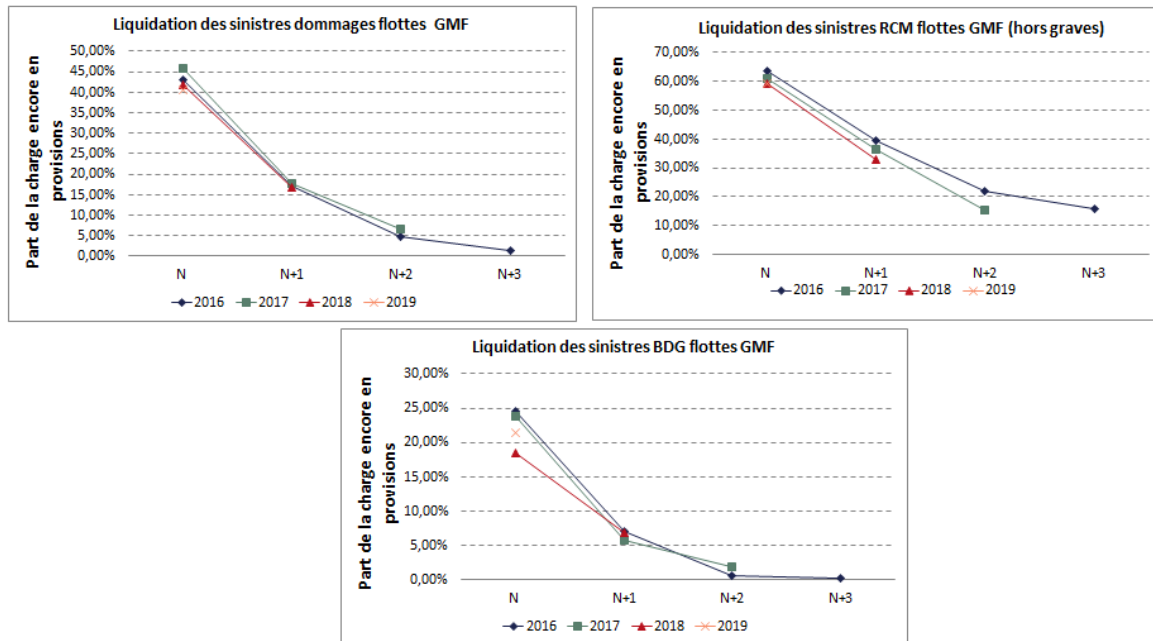


FIGURE 5.1 – Liquidation des sinistres RCM, DOM et BDG

Les sinistres flottes liées à la responsabilité matérielle semblent être ceux demandant le plus de temps à être traités. À fin décembre de l'année d'exercice, les provisions pèsent 70% de la charge totale, c'est plus que tout autre garantie. Par exemple, à fin décembre 2019, les provisions RC de l'exercice 2016 représentent encore 16% de la charge totale RC. En ce qui concerne les sinistres dommages et VOL, 3 ans sont généralement nécessaires pour régler tous les sinistres. Quant aux sinistres bris de glace et incendie, la gestion des sinistres semble être plus rapide (environ 2 ans).

Au vue des conclusions précédentes, il s'avère qu'à fin décembre de l'année d'exercice, le recul n'est pas suffisant, beaucoup de sinistres sont encore en attente d'être réglés. De ce fait, la charge de sinistralité peut être amenée à évoluer significativement. Il semble qu'un recul de 3 années soit nécessaire pour une grande partie des sinistres. Cependant, la liquidation des sinistres RC prouve que 3 années de recul peuvent ne pas être suffisantes pour disposer d'une sinistralité fiable. D'autant plus que ces graphes n'incluent pas les graves.

5.2 Les coûts moyens observés

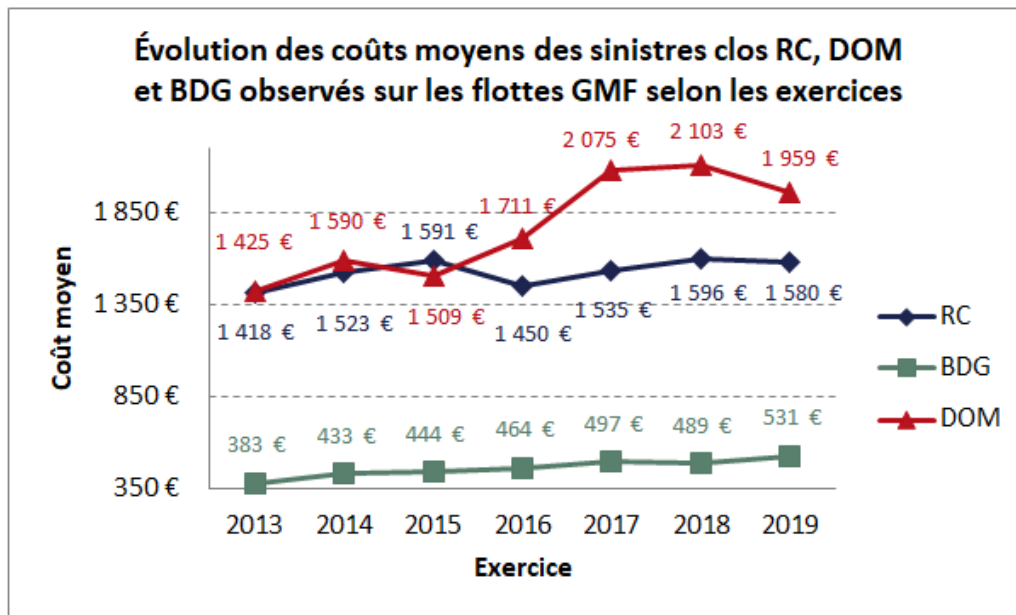


FIGURE 5.2 – Évolution des coûts moyens des sinistres clos RC, DOM et BDG observés sur les flottes GMF selon les exercices

Les coûts moyens sont calculés à fin décembre de chaque exercice en considérant exclusivement les sinistres clos flottes automobiles des 3 garanties principales : responsabilité civile (matérielle), dommages et bris de glace. Les sinistres en cours à fin décembre de chaque exercice ne sont pas pris en compte dans le calcul afin de ne pas faire intervenir les provisions qui pourraient légèrement biaiser le coût moyen. À noter que les coûts moyens sont calculés à partir de peu de données du fait du faible volume de contrats flottes présents en portefeuille. De plus, comme le portefeuille flottes GMF s'est développé au fil des années, les coûts moyens des années récentes sont plus fiables (car calculés sur plus de sinistres) que ceux des années antérieures à 2016. À fin décembre 2019, pour ces 3 garanties, les coûts moyens constatés sur le portefeuille flottes GMF sont proches de ceux constatés par la FFA sur le produit d'assurance flottes en 2019 (RC : 1210€, DOM : 2205€, BDG : 550€). Il peut être intéressant de comparer les coûts moyens observés sur les flottes avec les coûts moyens observés sur le portefeuille auto GMF, les garanties étant sensiblement similaires. À fin décembre 2019, les coûts moyens des garanties responsabilité civile matérielle, dommages et bris de glace sont respectivement de 957€, 1852€ et 496€. Cet écart s'explique par le fait que les contrats flottes assurent également des véhicules lourds, ce qui n'est pas le cas du produit auto GMF. Or, les sinistres générés par des véhicules lourds coûtent généralement plus cher, ce qui explique les coûts moyens plus élevés observés sur les flottes automobiles.

Une tendance à la hausse semble se dégager pour les garanties bris de glace et dommages (entre 2015 et 2019 : +19,6% pour le bris de glace et +29,8% pour le dommage). Cela s'explique en partie par l'augmentation du coût des réparations automobiles avec le temps. Quant au coût moyen des sinistres responsabilité civile matérielle, il semble globalement stable dans le temps. Les coûts moyens des garanties incendie et vol ne sont pas affichés sur le graphique, car les sinistres impactant ces garanties sont peu nombreux dans notre portefeuille. Par conséquent, les coûts moyens de ces 2 garanties observés à fin décembre de chaque année ne sont pas forcément représentatifs et peuvent sensiblement varier d'un exercice à un autre. À titre informatif, à fin décembre 2019, les coûts moyens incendie et vol observés sur les flottes automobiles GMF sont respectivement de 7970 euros et de 3581 euros.

5.3 Les fréquences observées

Afin de calculer les fréquences de sinistralité des différentes garanties, il est essentiel de disposer des nombres de sinistres. Les nombres de sinistres par garantie d'une année d'exercice N sont observés à la fin décembre de l'année N+1 pour tenir compte des éventuels sinistres tardifs. Tous les sinistres sont comptabilisés, c'est-à-dire les sinistres ouverts, clos et sans suite à la date d'observation.

Garantie	Sinistres 2016	Sinistres 2017	Sinistres 2018	Sinistres 2019
RC matériel	2 824	2 656	2 778	3 746
Dommages	1 868	1 785	1 736	2 204
Bris de glace	1 460	1 547	1 490	1 826
Vol	143	152	154	157
Incendie	19	35	34	28

TABLE 5.1 – Nombre de sinistres flottes des garanties RCM, DOM, BDG, VOL et INC

En assurances flottes, le nombre de sinistres touchant la garantie responsabilité civile est comparé au parc total pour déterminer la fréquence de sinistralité RC. En effet, tous les véhicules sont assurés au minimum en responsabilité civile. Ce n'est pas le cas pour les garanties bris de glace et dommages. Par exemple, les engins et remorques ne sont pas assurés en bris de glace et rarement en dommages. De plus, la majorité des contrats flottes accordent leurs garanties en fonction de l'âge des véhicules : plus le véhicule est âgé, moins il dispose de garanties. Dans le cas classique, les véhicules les plus récents de la flotte disposent de la meilleure couverture (formule tous risques), les véhicules âgés de 4 à 8 ans perdent la garantie dommages (RC, VOL, INC, BDG) et les véhicules âgés de plus de 8 ans sont assurés uniquement en RC. Afin de calculer les fréquences dommages et bris de glace, il est donc impératif de disposer du nombre de véhicules couverts par ces garanties. Les données de parc fournies par le courtier concernent uniquement la période d'assurance 2016 à 2019 et n'indiquent pas la répartition du parc par type de véhicules (-3 tonnes 5 tous risques, -3 tonnes 5, +3 tonnes 5 tous risques, +3 tonnes 5, remorques, engins, etc.). La partie suivante dans laquelle la base de données sera construite expliquera la méthode employée pour estimer au mieux les informations de parc manquantes et nécessaires à l'étude. Les nombres de véhicules couverts en bris de glace et en dommages ne sont donc que des estimations plus ou moins précises.

	Exercice 2016	Exercice 2017	Exercice 2018	Exercice 2019
Parc total	33 508	34 886	34 754	45 492
Véhicules assurés en dommages (estimation)	17 152	17 564	17 478	21 237
Véhicules assurés en bris de glace (estimation)	28 384	28 706	27 754	32 024

TABLE 5.2 – Parc assuré selon les garanties

La progression du nombre de sinistres pour les garanties principales (RC, DOM et BDG) entre 2018 et 2019 témoigne de l'évolution du portefeuille à cette période (+ 21,5% de primes acquises entre 2018 et 2019). Les sinistres touchant la garantie RC sont majoritaires. Viennent ensuite, dans l'ordre, les garanties DOM et BDG. Quant aux sinistres VOL et INC générant les coûts moyens les plus élevés (hors graves corporels) ils sont très minoritaires.

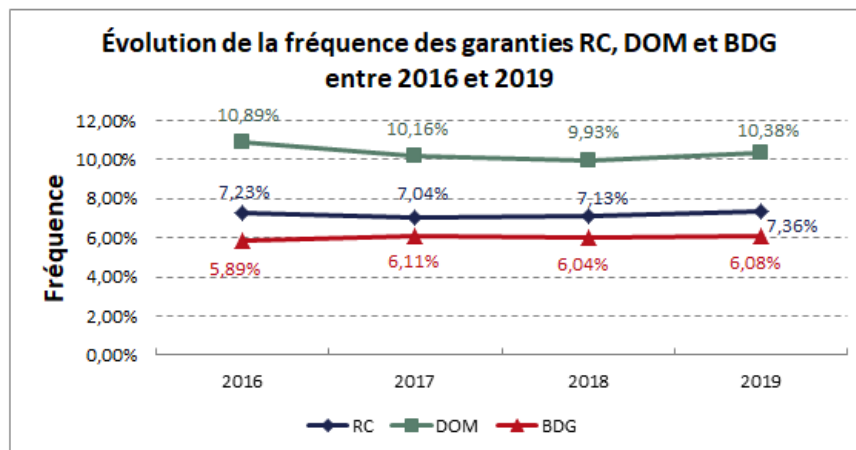


FIGURE 5.3 – Fréquences de sinistralité des garanties principales

Les fréquences des différentes garanties semblent globalement stables entre 2016 et 2019. Les fréquences relevées diffèrent légèrement de celles transmises par la FFA sur les flottes automobiles à fin décembre 2019, notamment pour la fréquence de sinistralité dommages (6,69% versus 10,38% sur le portefeuille GMF) et bris de glaces (7,06% versus 6,08% sur le portefeuille GMF). Les données fournies par la FFA sont intéressantes pour se faire une idée de la sinistralité globale des flottes mais il est essentiel de prendre du recul face à ses informations. Comme expliqué dans la première partie, il existe plusieurs produits flottes automobiles, le portefeuille flottes de GMF n'assure que des organismes publics tandis que certaines compagnies assurent des entreprises privées, la distinction n'est pas réalisée par la FFA, toutes les données flottes transmises par les acteurs de ce marché sont agrégées ensemble. Or, il est raisonnable de penser que la sinistralité peut être différente entre ces 2 domaines. Toutefois, la méthode d'estimation des véhicules couverts en bris de glace et en dommages peut également être remise en cause et pourrait potentiellement être à l'origine de ces écarts.

Les fréquences observées sur les flottes GMF sous-performent les fréquences observées sur le portefeuille auto de GMF, c'est-à-dire qu'elles sont plus élevées. Cela peut s'expliquer en partie par le manque de vigilance des conducteurs de véhicules assurés par des contrats flottes. Les véhicules ne leur appartenant pas, il est raisonnable de penser que certains sont moins prudents, d'autant plus qu'en cas de sinistre, la franchise est payée par la collectivité publique et le bonus-malus du conducteur sur son contrat auto n'est pas impacté. Ces éléments peuvent en partie expliquer des fréquences de sinistralité plus élevées.

Garantie	Fréquence flottes GMF	Fréquence auto GMF
Responsabilité civile	7,36%	5,46%
Dommages	10,38%	8,19%
Bris de glace	6,08%	5,56%
Vol	0,35%	0,27%
Incendie	0,09%	0,10%

TABLE 5.3 – Comparaison des fréquences flottes et auto GMF de l'exercice 2019 observé à fin décembre 2019

5.4 Les ratios sinistres à primes toutes garanties

Les ratios sinistres à primes sont bruts, c'est-à-dire qu'ils ne tiennent pas compte des frais de réassurance, des frais généraux et des commissions à verser au courtier. Les primes acquises sont hors taxes et hors assistance. Le vieillissement des SP permet d'avoir une meilleure vision des résultats techniques ultimes de l'exercice quelques années après. En effet, le SP calculé à fin décembre de l'exercice de survenance n'est pas très fiable, de nombreux sinistres n'ont pas encore été réglés et leurs coûts ne sont que des estimations plus ou moins précises. Le SP peut donc être amené à varier significativement les années suivant l'exercice de survenance avant de tendre vers une limite finie.

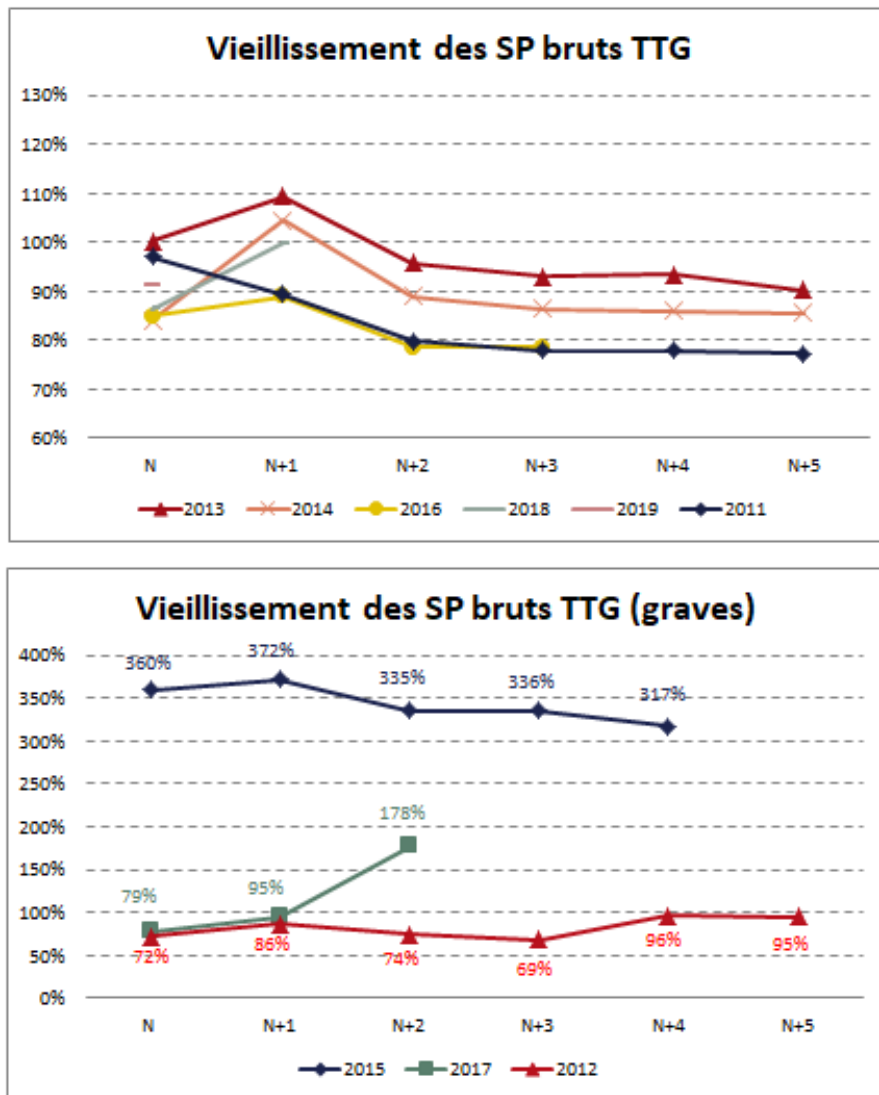


FIGURE 5.4 – Vieillessement des SP bruts

Le premier graphe représente les vieillissements des niveaux de SP des exercices pour lesquels on ne compte aucun sinistre supérieur à 1 million d'euros. Le constat est assez clair : bien que les SP soient encore particulièrement dégradés l'année suivant l'exercice de survenance, ils s'améliorent ensuite. Cette diminution du SP est sans doute liée à la liquidation d'une grande partie des sinistres 2 après leur survenance (rejoignant le constat réalisé sur les vieillissements des parts de provisions au sein des charges). Généralement, on constate une dégradation du SP entre la date d'observation N et N+1, potentiellement expliqué par la revalorisation de certaines estimations de sinistres graves. À fin décembre 2019, le SP de l'exercice 2019 affiche un SP de 91%, en progression de 5,8% par rapport à l'exercice 2018 observé avec le même recul. On observe que les SP bruts ultimes toutes garanties sont globalement tous supérieurs à 80%. Ce chiffre de référence est considéré en assurance flottes comme le seuil de rentabilité : au-delà, l'assureur perd de l'argent, en dessous, il est rentable et à 80% il est à l'équilibre. Pour des exercices typiques, sans présence de sinistres très graves (>400000 €), GMF ne semble pas être à l'équilibre. Qu'en est-il lorsque surviennent des événements particulièrement graves ?

Le deuxième graphe représente les vieillissements des niveaux de SP des exercices pour lesquels on compte au moins un sinistre supérieur à 400000 €. L'exercice 2015 est très dégradé, les niveaux de SP sont très inhabituels relativement aux autres exercices, on comprend qu'un sinistre très grave a eu lieu. En regardant de plus près, il s'agit d'un sinistre responsabilité civile estimé à 17,5 millions d'euros à fin décembre 2020 (soit 5 ans après sa survenance). L'exercice 2017 est également représenté sur ce graphe dû à la hausse soudaine de la sinistralité 2 ans après l'exercice de survenance.

En regardant de plus près, on s'aperçoit qu'un sinistre RC initialement ouvert à 50000€ a été ré-évalué 2 ans plus tard à 8,3 millions d'euros, expliquant cette hausse soudaine (+87% de SP entre fin décembre N+1 et fin décembre N+2) du ratio sinistre à prime. Le vieillissement des SP de l'exercice 2012 est également très inhabituel. Alors que le SP de l'exercice 2012 semble tendre vers 70% 3 ans après l'année de survenance, une augmentation soudaine se produit à fin décembre 2016 (+40% de SP entre fin décembre N+4 et fin décembre N+3). 2 sinistres graves sont à l'origine de cette dégradation brusque. Il s'agit encore une fois de sinistres graves (respectivement 425000€ et 250000€) initialement sous-évalués puis ré-évalués à leur juste valeur 4 ans après les faits. À titre d'information, le sinistre de 425000€ est un sinistre tardif survenu en 2012 mais seulement déclaré en 2014. Le sinistre a été évalué à l'ouverture en 2014 à 3500€ puis ré-évalué à 425000€ 2 ans plus tard. Le deuxième sinistre a été ouvert en 2012 et a été ré-évalué à la hausse tous les ans avant d'être estimé à 250000€ 4 ans après les faits. Certains sinistres semblant anodins peuvent se révéler être particulièrement difficiles à estimer à l'ouverture et évoluer significativement quelques années après les faits pour diverses raisons (sinistre corporel à priori sans danger finalement beaucoup plus grave que prévu).

Cette analyse permet de mettre en avant 2 sinistres particulièrement graves ayant touché des flottes assurés par GMF. La question se posera donc dans la partie suivante lors de la construction de la base d'écrêter les sinistres graves afin qu'ils ne dérèglent pas notre modèle de crédibilité, notamment la prime collective. En prenant légèrement du recul, on s'aperçoit encore une fois qu'une observation de la sinistralité à fin décembre de l'année d'exercice n'est pas possible, car elle ne détecterait pas ce sinistre grave.

5.4.1 Écrêtement et sinistres graves

L'analyse des ratios bruts sinistres à primes a permis de mettre en avant 4 sinistres flottes très graves. Seulement, il en existe peut-être d'autres, de moindres ampleurs, potentiellement indétectables dans la section précédente. Pour cela, l'idée est d'analyser tous les sinistres supérieurs à 120000€ déclarés entre 2011 et 2019. Il en existe 20. Seuls les 2 cités précédemment dépassent le million d'euros. Sur ces 20 sinistres, 17 sont des sinistres liés à la responsabilité civile (corporelle) et 3 à la garantie dommages. De plus, les 11 sinistres supérieurs à 250000€ sont exclusivement de l'ordre de la responsabilité civile. Ces éléments prouvent l'importance d'inclure dans les cotisations des provisions pour graves. Il est très délicat de donner des informations détaillées sur les circonstances des événements graves pour des raisons de confidentialité. Ce qu'on peut dire cependant, c'est qu'en étudiant les contrats des flottes GMF impliquées dans des événements graves, on s'aperçoit qu'il s'agit majoritairement de flotte composées d'un nombre important de véhicules et disposant d'une fréquence RC légèrement plus élevée que la moyenne. La surveillance mise en place se devra d'être particulièrement vigilante sur ces profils d'autant plus que le portefeuille flottes est en pleine croissance. Ces sinistres graves pourraient donc être plus fréquents dans les années à venir.

En assurance non vie, les valeurs d'écrêtement sont propres à chaque compagnie d'assurance et au produit considéré. Il a été décidé d'écrêter les sinistres flottes graves à 150000€, c'est-à-dire que tout sinistre supérieur à ce montant se voit attribuer la valeur d'écrêtement. Ce seuil d'écrêtement correspond au seuil historique d'écrêtement de la GMF (1 million de francs), toujours utilisé actuellement. Le portefeuille Flottes est réassuré auprès de différents réassureurs. Seuls les sinistres supérieurs à 1 million d'euros sont pris en charge par les réassureurs. D'ailleurs, les 2 sinistres Flottes très graves des dernières années ont fait augmenter les primes Flottes de réassurance que payent GMF auprès des réassureurs.

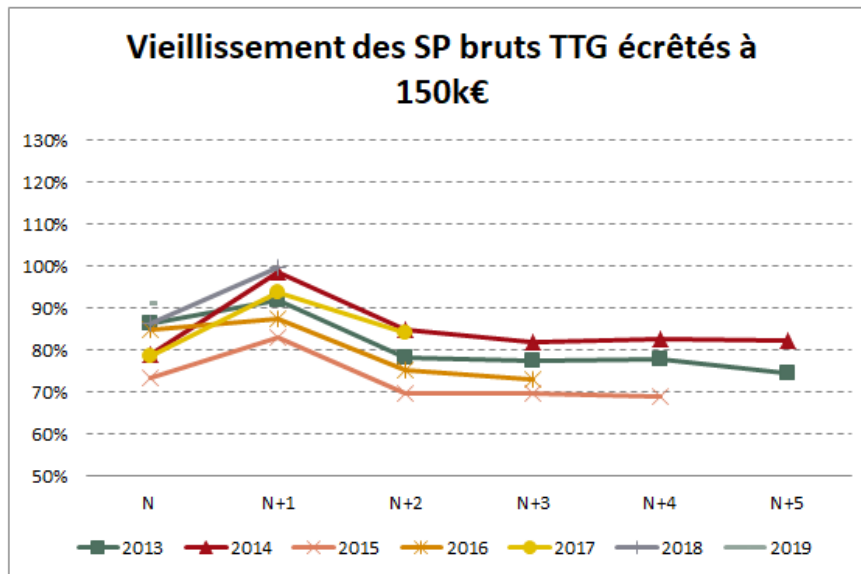


FIGURE 5.5 – Vieillessement des SP bruts écrêtés à 150000 euros

L'objectif est de lisser les résultats techniques afin qu'ils soient beaucoup moins impactés par la présence de ces événements exceptionnels. Cela permet de les rendre comparables avec les résultats techniques des autres exercices. Il semble que la tendance soit à la hausse sur ces 3 dernières années, petit à petit, on assiste à une dégradation des ratios sinistres à primes écrêtés.

5.4.2 Les ratios sinistres à primes par garantie

L'objectif de cette partie est de calculer les ratios sinistres à primes bruts par garantie afin de voir si certaines garanties ne seraient pas sous-tarifées lors des appels d'offres. Cela pourrait permettre d'adapter les primes par garantie en conséquence lors des prochains appels d'offres à tarifier.

Représenter la répartition des charges selon les garanties permet de se rendre compte du poids de chaque garantie dans la sinistralité totale.

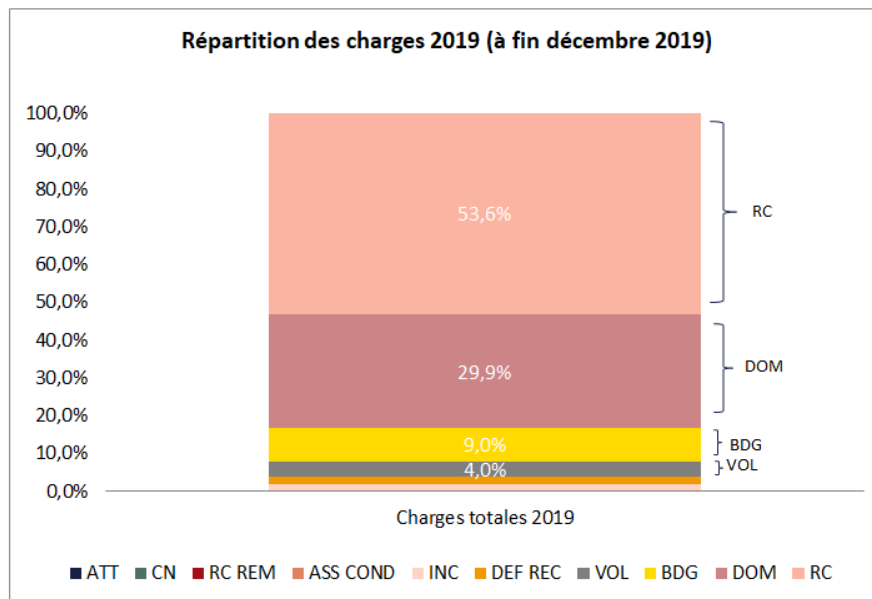


FIGURE 5.6 – Répartition des charges 2019 selon les différentes garanties

À elles seules, les garanties RC, dommages et BDG expliquent 93% de la sinistralité totale.

Ensuite, il peut être intéressant de comparer la part de chaque garantie au sein des primes et des charges en 2019 pour voir si l'équilibre est globalement respecté. L'année 2019 est choisie, car elle est relativement récente et ne comporte pas de sinistres graves, c'est une sinistralité pouvant être qualifiée de classique. Par exemple, si l'année 2015 avait été choisie (sinistre grave de l'ordre de 17 millions d'euros), la garantie RC aurait représenté plus de 75% de la sinistralité totale. En 2019, la garantie RC représente 54,5% des primes acquises et 53,6% des charges. En ce qui concerne les garanties dommages et bris de glace, la part qu'elles représentent au sein des charges sur-performe de très loin la part qu'elles représentent au sein des primes acquises (respectivement 29,9% et 9% versus 18,1% et 4,6%) laissant présager des SP particulièrement dégradés pour ces garanties et des primes sous-tarifées. Cela est cependant compensé par les très bons résultats de la garantie assurance du conducteur représentant seulement 0,1% de la sinistralité totale contre 10,9% des primes acquises.

Dans les graphiques suivants, les charges par garantie sont comparées aux primes acquises par garantie pour déterminer les SP bruts par garantie.

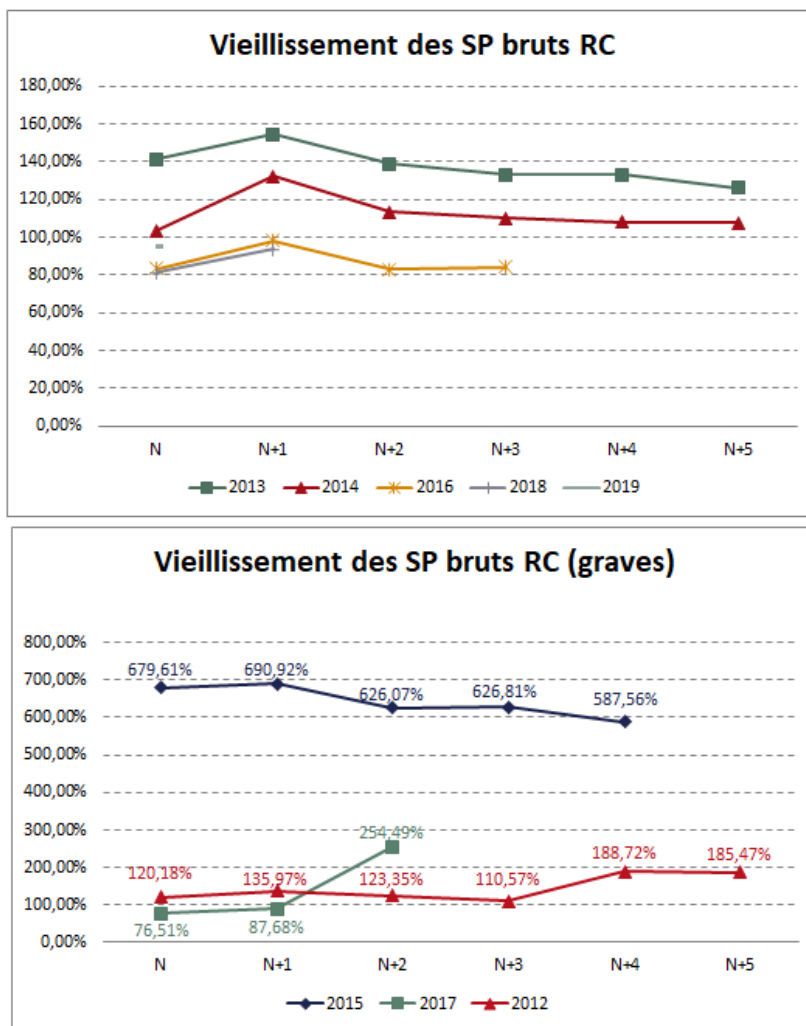


FIGURE 5.7 – Vieillessement des SP RC

La garantie RC représentant la majorité des charges et des primes, les courbes de vieillissement des SP RC sont donc similaires à celles des SP toutes garanties. À fin décembre 2019, le SP de la garantie RC affiche 95,4%, en progression de 17,5% par rapport à l'exercice 2018 vu avec le même recul. En ce qui concerne les exercices pour lesquels au moins un sinistre RC très grave est survenu, de toute évidence, les ratios sinistres à primes sont très dégradés.

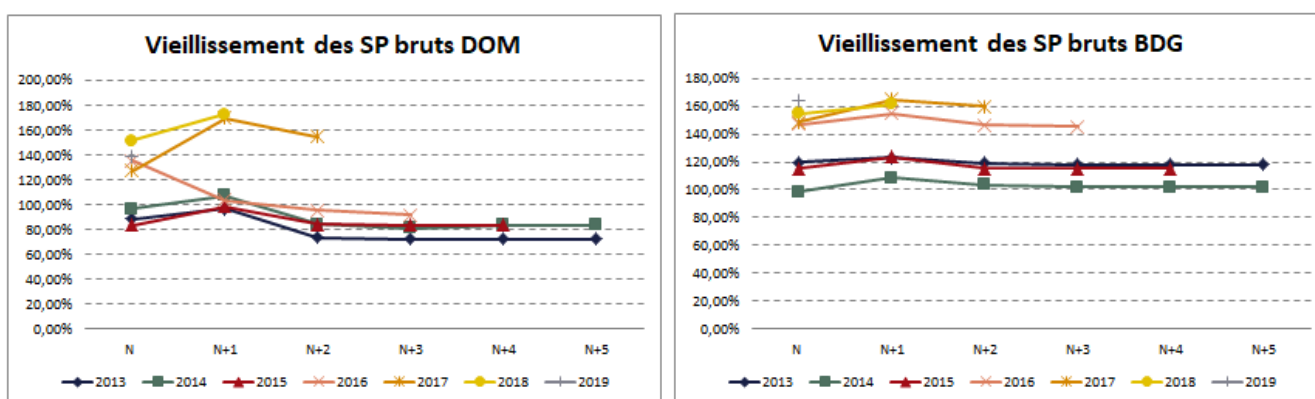


FIGURE 5.8 – Vieillessement des SP DOM et BDG

Pour ces deux garanties, on constate une très nette dégradation des ratios sinistres à primes au fil des années et notamment entre la période 2013-2015 et la période 2016-2019. L'année 2016 est l'année de rupture et elle correspond au moment où le portefeuille flottes a évolué significativement (+25% de primes acquises). Le SP dom de l'exercice 2016 observé à fin décembre 2016 s'élève à 135%, en progression de 62% par rapport à l'exercice 2015 observé à fin décembre 2015. Sur la même période, la hausse du SP BDG s'élève à +27,1% (146,9% à fin décembre 2016 versus 115,5% à fin décembre 2015). Les résultats techniques de la garantie BDG sont cependant à relativiser, car elle représente beaucoup moins de poids par rapport à la garantie dommages (0,7 million d'euros de charges pour l'exercice BDG 2019 observé à fin décembre 2019 contre 2,8 millions d'euros pour l'exercice DOM 2019 observé à fin décembre 2019), l'impact est donc moindre sur les résultats techniques toutes garanties. Toutefois, le constat est clair, ces deux garanties sont sous-tarifées lors des appels d'offres.

5.5 Sinistralité par tranche de prime

On a pu voir précédemment qu'une grande partie des primes acquises était expliquée par les gros contrats en portefeuille. Il s'agit, ici, d'observer les résultats techniques selon le niveau de la prime émise annuelle. Pour cela, on se base sur l'exercice 2019 observé à fin décembre 2019.

Tranche	Part de primes acquises	SP
<10 000 €	1,7%	0,60
10 000-20 000 €	4,0%	1,05
20 000-50 000 €	17,4%	0,88
50 000-100 000 €	24,3 %	0,97
100 000-200 000 €	30,6%	0,94
>200 000 €	22,1%	1,07

TABLE 5.4 – Ratios sinistres à primes bruts de l'exercice 2019 (observé à fin décembre 2019) selon le niveau de prime

Comme on pouvait s'y attendre, on constate que les résultats techniques sont d'autant plus dégradés que la prime est importante. Les gros contrats semblent sous-tarifés. Ici, seul l'exercice 2019 est représenté, mais cette tendance est stable dans le temps. Cela signifie que ces contrats sont délicats à tarifer lors de l'appel d'offres. D'où l'importance de mettre en place une surveillance efficace pour détecter d'éventuelles erreurs de tarification et les corriger, année après année, à l'échéance annuelle du contrat. La revalorisation tarifaire du portefeuille aura donc pour objectif de cibler ces contrats en priorité.

5.6 Sinistralité des contrats avec conservation

À fin décembre 2019, on compte 11 contrats avec conservation des sinistres. Pour rappel, ces contrats représentent une part non négligeable des primes acquises. Le ratio sinistres à primes écrêté de l'exercice 2019 (observé à fin décembre 2019) de ce segment est évalué à 111%. La tendance est la même pour les autres exercices. Ce segment va donc demander des mesures plus ciblées.

5.7 Sinistralité par activité et par région

Il est très raisonnable de penser que la sinistralité d'une flotte peut être impactée par sa localisation ou encore son activité professionnelle. Toutefois, le portefeuille étant relativement peu volumineux, il est difficile de tirer des conclusions. D'autant plus que les données liées à l'activité professionnelle de la flotte ne sont pas toujours très fiables.

5.8 Constats

Cette partie a permis de mettre en avant quelques spécificités du portefeuille flottes automobiles de GMF (contrats avec conservation, périodes d'ancienneté différentes, forte hétérogénéité des contrats, concentration des contrats dans certaines régions, surreprésentation de certaines activités, etc.) et de constater la sinistralité effective du portefeuille. Une surveillance doit être mise en place pour les raisons suivantes :

- ▶ Forte évolution du portefeuille en 2019 en termes de primes acquises.
- ▶ Dégradation des SP bruts sur les 3 dernières années (seuil d'équilibre dépassé).
- ▶ Sous-tarification de certaines garanties (dommages, BDG).
- ▶ Présence de sinistres très graves au sein du portefeuille.

5.9 Les mesures tarifaires flottes

En assurance, la surveillance du portefeuille consiste à réviser les conditions de garanties accordées aux assurés présentant ultérieurement un surrisque. L'objectif de la surveillance est de maîtriser les résultats du portefeuille et de faire en sorte qu'ils ne se détériorent pas ou qu'ils s'améliorent. Généralement les mesures tarifaires de l'année $N+1$ sont mises en place lors de l'été de l'année N en se basant sur les résultats techniques des années J (avec $J \leq N - 1$).

En assurance auto, tous les ans, en plus de la mise à jour des coefficients bonus-malus personnalisés pour chaque contrat, des mesures de revalorisations tarifaires plus générales sont appliquées au portefeuille. Une cible est déterminée (par exemple 3%) dans le but d'augmenter la cotisation moyenne de l'ordre de la cible. Ensuite, l'idée est de répartir cette augmentation sur les garanties ou les critères présentant des résultats dégradés sur les années passées. En pratique, cela consiste à majorer certaines primes de base (par exemple la prime de base bris de vitre) ainsi que certains coefficients s'appliquant sur ces primes de base (par exemple les coefficients concernant l'âge des conducteurs). En assurance flottes, rien de cela n'est possible, car la tarification est essentiellement basée sur l'historique de sinistralité de la flotte. Il n'existe donc ni bonus-malus, ni primes de base, ni coefficients sur lesquels jouer pour équilibrer le portefeuille. Les revalorisations sont donc soit individuelles (selon la sinistralité du contrat) soit collectives (+3% par exemple).

Jusqu'à juin 2020, il n'existait pas de réelle surveillance du portefeuille flotte ouverte GMF. Les seules (et rares) majorations réalisées étaient de l'ordre de l'indice SRA (Sécurité et Réparations Automobiles). Cet indice correspond à la moyenne de trois indices (pièces de rechange, main d'oeuvre et peinture) et représente l'inflation du coût des réparations automobiles. Il est très utilisé en assurance auto et explique une partie des évolutions tarifaires des contrats auto d'une année à l'autre.

La dégradation des résultats techniques et la croissance significative du portefeuille flotte ces dernières années ont conduit GMF à mettre en place une surveillance active du portefeuille en 2020 pour les mesures tarifaires 2021. Le portefeuille de contrats flottes étant peu volumineux, il aurait pu être envisageable d'analyser chaque année les résultats techniques de chaque contrat. Évidemment, cette option, coûteuse en termes de temps, a vite été écartée au profit d'un processus de surveillance du portefeuille automatisé, efficace, robuste, compréhensible et réutilisable chaque année lors des mesures tarifaires. L'objectif est donc double : repérer les contrats dégradés et déterminer le montant des majorations selon l'expérience du client. Par exemple, les clients présents dans le portefeuille flotte GMF seulement depuis 2020 seront exemptés de mesures tarifaires, quels que soient leurs résultats. En effet, il est difficile de majorer un contrat avec une seule année d'expérience d'autant plus que lors des mesures tarifaires, à fin juin, les données de sinistralité de l'année en cours sont basées sur seulement 6 mois. Une flotte peut présenter des résultats dégradés à fin juin et de bons résultats à fin décembre (et réciproquement).

Le nouvel outil a pour but d'aider à la décision en préconisant une mesure tarifaire pour chaque contrat présentant des mauvais résultats. Il se base uniquement sur des indicateurs techniques. Cependant, la mesure tarifaire finale, celle soumise au client, peut évidemment être différente de la mesure recommandée par l'outil pour diverses raisons (clauses empêchant les majorations au-delà d'un certain seuil, raisons commerciales, etc.). En effet, les actuaires ont des compétences techniques que les technico-commerciaux n'ont pas, mais réciproquement, ils ne disposent pas de la connaissance client et commerciale des technico-commerciaux.

Le coefficient de réduction-majoration n'existant pas pour l'assurance flottes, le seul moyen d'optimiser les résultats techniques est de majorer la prime d'assurance. Encore faut-il repérer les contrats particulièrement dégradés. En assurance auto, les critères pertinents d'identification des contrats à traiter en priorité sont généralement la fréquence des sinistres (exemple : 2 accidents ou plus en 24 mois) et la nature (ou l'importance) du sinistre (exemple : dommage corporel). Ces critères ne sont pas facilement transposables à l'assurance flottes (ensemble de véhicules, types de véhicules différents, etc.). Une adaptation de ces critères ou la création de nouveaux critères seront donc nécessaires.

Bien que des majorations de primes soient évidemment possibles (en fonction du contrat et à un certain niveau), il est très difficile (voire impossible) pour un assureur de résilier le contrat flotte d'un organisme public parce qu'il présente de mauvais résultats. Cela pourrait ternir son image vis-à-vis de l'organisme public et détériorer ses relations avec le courtier. Si le client refuse la majoration de son contrat, le contrat est résilié à l'échéance suivante et l'organisme public doit soumettre un nouvel appel d'offres. Or, un appel d'offres est coûteux en temps et en argent donc ce n'est pas tellement dans l'intérêt du client de refuser la majoration.

Troisième partie

Création d'une base de données

Chapitre 6

Le périmètre

Mettre en place une surveillance d'un produit d'assurance requiert une base de données contenant suffisamment d'informations sur les contrats présents en portefeuille. En conséquence, une grande partie du travail préliminaire a consisté à créer la base de données. La première étape a été de définir le périmètre et le cadre de l'étude (type de produit, période de temps considérée, individus et variables).

6.1 Le produit considéré

Le produit considéré est le produit d'assurance flottes automobiles (ouvertes) de GMF. Bien que les autres marques Covéa (MAAF et MMA) proposent également des assurances flottes automobiles, il n'est pas possible d'agréger les données (d'autant plus que les produits et la clientèle ciblée peuvent être très différents : GMF assure uniquement des organismes publics contrairement à la MAAF qui dispose d'un portefeuille de flottes professionnelles).

6.2 La période de temps

La période de temps considérée est la période janvier 2011-juin 2020. L'année de départ 2011 a été choisie par défaut, car il n'était pas possible de remonter plus loin dans les données informatiques GMF. L'objectif était de remonter le plus loin possible pour détenir une quantité importante d'informations sur les clients présents en portefeuille flottes depuis longtemps. Plus l'expérience d'un client est grande, plus sa surveillance est aisée. En effet, il est plus simple d'estimer sa sinistralité future grâce aux nombreuses données antérieures de sinistralité à disposition. La période de fin, juin 2020, est liée au moment où l'étude a été réalisée et à l'objectif de l'étude, c'est-à-dire la mise en place des mesures tarifaires 2021. Pour informations, les préavis des contrats flottes sont généralement de 6 mois et les dates d'échéance fixées au 31 décembre. Par conséquent, les mesures tarifaires du premier janvier 2021 devaient être définies plus de six mois à l'avance, c'est-à-dire en mai-juin 2020. Les données 2020 sont incluses dans la base de données mais n'auront pas de grande utilité pour les mesures tarifaires 2021. En effet, à cette période-là (mai-juin 2020), les données de sinistralité de l'année en cours (2020) ne sont pas suffisantes (5 mois seulement) pour se faire une idée précise des résultats techniques à fin décembre 2020. D'autant plus que le début de l'année 2020 présente des résultats très atypiques dus au confinement déclaré en mars 2020. Effectivement, le confinement a fait reculer de manière très significative la sinistralité auto. De ce fait, de nombreux assureurs (dont Covéa) ont pris la décision de geler les primes 2021 des contrats auto, c'est-à-dire ne pas augmenter les tarifs des cotisations. Cela n'a pas été le cas en assurance flottes.

6.3 Les individus

Les individus de la base de données sont des organismes publics ayant au moins un contrat flotte ouverte GMF à fin juin 2020. Les contrats des clients de la base proviennent tous d'un appel d'offres émis sur le marché public. À chaque ligne correspond un client, mais un client ne correspond pas nécessairement à un seul contrat. Plusieurs cas sont possibles pour un client dans notre base :

- Disposer d'un seul contrat à juin 2020 (cas le plus répandu).

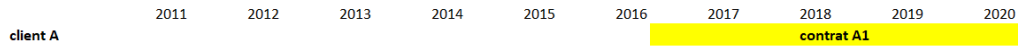


FIGURE 6.1 – Exemple d’une situation contractuelle d’un client (1)

- Disposer d’un contrat en cours étant la suite d’un précédent contrat.

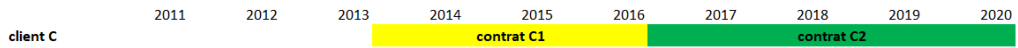


FIGURE 6.2 – Exemple d’une situation contractuelle d’un client (2)

- Disposer de plusieurs contrats en cours.

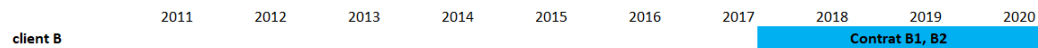


FIGURE 6.3 – Exemple d’une situation contractuelle d’un client (3)

- Disposer d’années creuses entre 2 contrats d’un même client (exemple : contrat assuré par GMF entre janvier 2011 et décembre 2013, ensuite par un autre assureur entre janvier 2014 et décembre 2016 puis par GMF entre janvier 2017 et janvier 2021). Encore une fois, c’est le même parc, GMF n’a cependant pas toujours remporté l’appel d’offres.

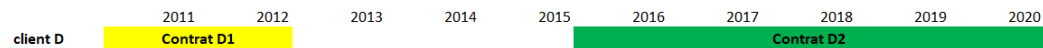


FIGURE 6.4 – Exemple d’une situation contractuelle d’un client (4)

Il aurait pu être possible de travailler à une maille plus fine en considérant chaque contrat flottes en tant qu’individu de la base de données. Les clients avec plusieurs contrats seraient donc apparus plusieurs fois dans la base de données, sur plusieurs lignes. Cependant, en fonctionnant de la sorte, cela ne permet pas de considérer l’expérience du client dans son ensemble. L’idée est plutôt de repérer les "mauvais clients" que les "mauvais contrats" bien que les deux soient généralement liés. De plus, si un contrat est dégradé, mais que les autres contrats du client présentent de bons résultats, un effet de compensation est possible et ce serait dans l’intérêt du client. En travaillant à la maille client, l’idée est d’optimiser la connaissance client en capitalisant sur les clients possédant plusieurs contrats (en cours ou terminés) et (ou) présents dans le portefeuille depuis longtemps. La décision a donc été prise d’afficher les informations par client en regroupant tous les contrats d’un même client selon la période d’assurance sur une seule ligne. Les exemples dans les paragraphes suivants permettront d’y voir plus clair.

6.4 Les variables

L’objectif de la base de données est d’associer à chaque organisme public le plus d’informations possibles (connaissance client, flottes, primes, sinistralité) sur la période de temps considérée (janvier 2011-juin 2020). Plus les informations sur les clients seront nombreuses, plus la surveillance sera optimale. Il n’existait pas de base de données flottes préexistante sur laquelle s’appuyer pour gagner du temps. Il a donc été nécessaire de construire la base dans son intégralité sur SAS.

Chapitre 7

La construction de la base de données

L'objectif de cette partie est de donner au lecteur un bref aperçu des méthodes mises en place pour construire la base de données SAS et d'expliquer le périmètre de chaque variable présente dans la base de données. La construction de cette base s'est avérée être plus compliquée que prévue. En effet, des informations primordiales et spécifiques du produit flottes étaient difficilement accessibles.

7.1 Les systèmes d'informations GMF

La qualité des données étant essentielle à la bonne réalisation de l'étude actuarielle, il est donc pertinent d'expliquer brièvement au lecteur le fonctionnel général des systèmes d'information GMF.

Tous les actes de gestion (création d'un nouveau contrat, saisie de nouvelles informations clients, déclaration de sinistres, etc.) sont réalisés via des interactions hommes-machines (site WEB, Multiples Applicatifs Internes GMF), appelées également IHM. Ces interactions alimentent des bases de gestion centralisées et les mettent à jour instantanément. Ces bases de gestion font partie du système d'information opérationnel et sont régies par des systèmes de gestion de base de données DB2 (SGBD). Elles sont segmentées en plusieurs groupes (données clients, contrats IARD, contrats VIE, sinistres IARD, sinistres VIE, etc.). Chacun de ces groupes contient plusieurs bases. Par exemple, au sein des bases contrats IARD, on retrouve une base pour les garanties, une autre pour les risques, etc. Ces bases DB2 utilisent le langage SQL et possèdent donc des clés de jointure (attributs contenant des valeurs communes à plusieurs jeux de données) permettant d'agrèger les bases de données entre elles. Il existe plusieurs clés possibles : référence interne du client, du contrat, etc. La centralisation des données fait la grande force de GMF : toutes les informations sont contenues dans approximativement 8000 tables DB2.

De plus, à chaque acte de gestion, un message est transmis ("modification du conducteur du contrat A1", "ajout d'un véhicule pour le contrat B2", etc.) sous forme de code. Ces messages sont stockés encore une fois dans des tables DB2 et à chaque fin de mois, tous les messages du mois en question sont lus et interprétés par une équipe dont c'est la spécialité, afin de créer le système d'information décisionnel, appelé également infocentres. Ce sont des bases de données permettant de regrouper et d'agrèger au sein d'un même endroit des données provenant de sources différentes. À GMF, les infocentres sont divisés en 4 grands groupes :

- Les données contrats (portefeuille, avenants, affaires nouvelles, chutes) ;
- Les devis ;
- Les projets d'avenants ;
- Les sinistres.

À noter que les informations ne sont pas les mêmes entre les deux systèmes d'information. Les données SID et SIO n'ont pas le même but :

- ▶ SIO (opérationnel) : gestion des contrats, des clients, des sinistres ;
- ▶ SID (décisionnel) : réaliser des études, suivre les S/P, les garanties, les sinistres, les primes acquises, etc.

Le décisionnel a pour objectif de mettre en forme et de faire le tri en conservant uniquement les données utiles potentiellement exploitables lors d'études et d'assurer le pilotage technique de la Direction IARD. Le système d'information décisionnel n'est pas instantané comme le SIO, il est mis à jour à chaque fin de mois. Cette vision "fin de mois" est disponible à chaque début de mois suivant. Les tables sont historisées (fin décembre 2018, fin janvier 2019, etc.) pour capturer les données à chaque évolution de ces dernières. Par exemple, la table sinistres auto à fin décembre 2017 permet de voir l'état des sinistres des 5 dernières années à fin décembre 2017. Le système d'information décisionnel (SID) est accessible à tous les collaborateurs GMF (ce qui n'est pas le cas du système d'information opérationnel), via SAS. Les infocentres sont des tables de données SAS présentes dans des bibliothèques SAS. Chaque bibliothèque SAS a sa propre spécificité (sinistres, devis, etc.) et est triée selon les différents historiques.

7.2 Les infocentres risques spécifiques

Pour rappel, les flottes automobiles GMF font partie des risques spécifiques (les assurés sont des personnes morales, collectivités ou associations principalement). Les contrats risques spécifiques sont des contrats à destination principalement d'associations, de collectivités avec des conditions de souscription très différentes des contrats classiques. On y retrouve également des produits liés aux besoins de la fonction publique (élèves infirmiers, assurance péronnelle de l' élu, véhicule de l'administration, etc.) ou encore des risques ne pouvant pas être assurés par GMF (bâtiments classés, nombre de pièces supérieur à 7, etc.).

Sur SAS GMF, plusieurs librairies concernent les risques spécifiques. Ce sont uniquement les tables de données présentes dans ces bibliothèques qui seront utilisées pour collecter les données propres aux flottes automobiles. Toutes les données étant sous SAS, le plus simple a donc été de créer la base sous SAS sous forme d'un projet SAS. Le projet est décomposé en plusieurs programmes et chaque programme a sa spécificité. Un programme récupère les primes, un autre les sinistres, etc. À chaque nouvelle étape (une étape=un programme), les nouvelles informations sont jointes (grâce au numéro client) à la base de données vue à la fin du programme précédent. La clé de jointure, c'est-à-dire le numéro de client, est une variable numérique portée par 6 caractères.

Comme les risques spécifiques regroupent plusieurs produits différents, chaque produit a un code unique. Le produit Multirisques Automobile (flottes automobiles ouvertes) est caractérisé par le code produit 04 dans les infocentres risques spécifiques.

7.3 Les primes

7.3.1 Les infocentres primes GMF

La première grande étape a été de récupérer les primes des clients flottes entre janvier 2011 et juin 2020. Les infocentres liées aux primes sont construites de la manière suivante :

- ▶ Périodicité mensuelle : le fichier d'un mois donné reprend toutes les factures saisies, annulées ou contre-passées entre le premier janvier de l'année en question et la fin du dit mois ;

Description d'un enregistrement	Date d'arrêt du fichier			
	31/01/2010	28/02/2010	31/03/2010	30/04/2010
Facture émise le 25/02/2010 de période d'assurance 01/01/2010 au 31/12/2010	rien	3 lignes 1 par garantie 1 pour le total	3 lignes 1 par garantie 1 pour le total	3 lignes 1 par garantie 1 pour le total

FIGURE 7.1 – Conséquence dans les fichiers de pilotage d'une prime émise sur 2 garanties

- Archivage des versions du fichier, à une date donnée mm/aaaa, sont accessibles les versions suivantes :
 - Le fichier arrêté à la fin du mois dernier (mois m-1);
 - Les 11 fichiers mensuels antérieurs (mois m-2, m-3, etc.);

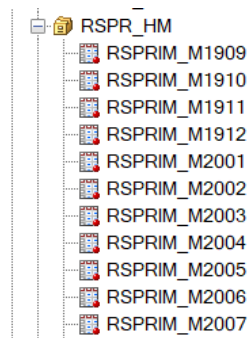


FIGURE 7.2 – Fichiers mensuels des primes

- Les fichiers arrêtés au 30 juin et au 31 décembre des 10 dernières années.

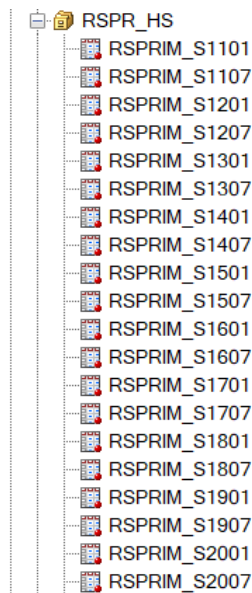


FIGURE 7.3 – Fichiers semestriels et annuels des primes

L'objectif est de collecter les primes émises et les primes acquises hors taxes (et hors assistance) au 31 décembre des années concernées (à l'exception de l'année 2020) en se basant sur les infocentres primes GMF. Les primes acquises sont calculées à partir des primes émises et de la période d'assurance. C'est la partie de la prime à retenir pour la période de temps durant laquelle une police est en vigueur. Elle représente le montant de prime en relation avec la proportion de couverture d'assurance procurée à l'assuré. On peut également la définir comme une mesure d'exposition au risque. Pour une certaine période de temps, la prime acquise est égale à :

$$\text{Prime acquise} = (\% \text{ de couverture écoulee}) * \text{Prime émise}$$

Prenons l'exemple d'un client avec une date d'échéance à fin juin. Le client paie, début juillet, la totalité de sa prime pour la période 01/07/N-30/06/N+1. À fin décembre N, la prime acquise pour la période d'assurance 01/07/N-31/12/N est donc égale à la moitié de la prime émise. L'autre moitié de la prime couvrira la période d'assurance 01/01/N+1-30/06/N+1.

Les primes acquises sont essentielles, car elles permettront de calculer par la suite les ratios sinistres à primes.

7.3.2 Primes émises

Pour récupérer les primes émises à la fin de l'année N, il suffit de travailler sur la table SAS RSPRIM_S(N+1)01 (RSPRIM_S1201 pour l'année 2011, etc.). Concernant l'année 2020, la table RSPRIM_M2007 permet de récupérer toutes les primes émises par les clients flottes entre janvier 2020 et fin juin 2020. Un client a la possibilité de payer sa prime en plusieurs fois. Chaque paiement est associé à une facture. Chaque facture est décomposée en primes par garanties. Dans ces bases de données, pour un contrat, il existe autant de lignes qu'il y a de factures dans l'année et de garanties dans le contrat. Une ligne affiche le montant total de la facture toutes garanties confondues en tenant compte des frais de gestion. En effet, les frais de gestion des contrats flottes GMF sont intégrés dans la prime et payés par le client. Prenons l'exemple fictif d'un client ayant assuré sa flotte en responsabilité civile, dommages et assistance et payant sa prime semestriellement. Voici la retranscription de ces informations dans les infocentres primes à fin décembre 2011.

client	code produit	numéro contrat	date	Période d'assurance	numéro de facture	garantie	frais de gestion	Prime émise HT
A	4	C1	janv-11	01/01/11-30/06/11	1	RC	0	200
A	4	C1	janv-11	01/01/11-30/06/11	1	DOM	0	350
A	4	C1	janv-11	01/01/11-30/06/11	1	ASS	0	50
A	4	C1	janv-11	01/01/11-30/06/11	1	TTG	15	615
A	4	C1	juil-11	01/07/11-31/12/11	2	RC	0	120
A	4	C1	juil-11	01/07/11-31/12/11	2	DOM	0	250
A	4	C1	juil-11	01/07/11-31/12/11	2	ASS	0	30
A	4	C1	juil-11	01/07/11-31/12/11	2	TTG	7	407

FIGURE 7.4 – Exemple fictif d'un client dans les infocentres primes

La garantie assistance est incluse dans le montant des factures toutes garanties or l'assistance n'est pas gérée par GMF mais par une de ses filiales. Il faut donc ne pas en tenir compte lors de l'étude. L'idée est de synthétiser toutes les informations en une seule ligne dans notre nouvelle base.

client	code produit	année	Prime émise HT (hors ASS)
A	4	2011	942

FIGURE 7.5 – Exemple fictif de l'information à récupérer dans la nouvelle base

Comme expliqué précédemment, la base de donnée finale doit contenir une ligne par client et non une ligne par contrat. L'objectif est de connaître le montant total payé par le client pour son assurances flotte à la fin de chaque année, tous contrats confondus. Dans la majorité des cas, un client ne dispose que d'un contrat à l'année, mais il existe quelques cas particuliers. Voici comment ils sont traités :

- Plusieurs contrats de mêmes périodes d'assurance (donc une flotte pour chaque contrat).

client	code produit	numéro contrat	année	période assurance	Prime émise HT (hors ASS)
C	4	C1	2011	01/01/11-31/12/11	20 000
C	4	C2	2011	01/01/11-31/12/11	30 000
C	4	C3	2011	01/01/11-31/12/11	10 000

↓

client	code produit	année	Prime HT (hors ASS)
C	4	2011	60 000

FIGURE 7.6 – Calcul de la prime annuelle d'un client possédant plusieurs contrats par année

- Disposer de contrats avec des périodes d'assurance se chevauchant dans la même année (exemple : premier contrat entre juillet 2009 et juin 2011 et deuxième contrat entre juillet 2011 et juillet 2013).

client	code produit	numéro contrat	année	période assurance	Prime émise HT (hors ASS)
B		4 B1	2011	01/01/11-30/06/11	942
B		4 B2	2011	01/07/11-31/12/11	1500

↓

client	code produit	année	Prime émise HT (hors ASS)
B		4 2011	2442

FIGURE 7.7 – Calcul de la prime annuelle d'un client lors d'un chevauchement

La collecte des primes émises de chaque client est réalisée pour chaque année entre janvier 2011 et juin 2020. Ensuite, une jointure SQL selon le numéro de client permet de récupérer l'information dans une seule base de données.

7.3.3 Primes acquises

Bien que la démarche globale soit sensiblement la même, le calcul des primes acquises est légèrement différent, car il dépend de la période d'assurance. Pour calculer les primes acquises de l'année N, il est fondamental de disposer des primes des contrats avec une période d'assurance chevauchant l'année N (N-1/N ou N/N+1). Si on souhaite calculer les primes acquises à la fin de l'année N, il est impératif de disposer des infocentres RSPRIM_S(N)01 pour les primes payées an année N-1, RSPRIM_S(N+1)01 pour les primes payées an année N et RSPRIM_S(N+2)01 pour les primes payées an année N+1. Cependant, en assurance flottes, il arrive très fréquemment que des clients émettent une prime en année J avec une période d'assurance I (avec $I < J$). Cela s'explique par des mouvements au sein du parc (ajouts de véhicules durant l'année) qu'il est nécessaire de régulariser par la suite, c'est une des particularités de ce produit. C'est pourquoi il est pertinent de travailler avec des historiques supplémentaires (RSPRIM_S(N+3)01, RSPRIM_S(N+4)01, etc.) pour disposer de la prime acquise la plus représentative possible du parc.

L'idée est de calculer pour chaque garantie de chaque facture la prime acquise vue à fin décembre de chaque année (et à fin juin 2020) puis d'agréger les lignes correspondantes à un même client entre elles. Les étapes PROC MEANS sur SAS permettent de réaliser ces calculs sur les sous-populations définies par les modalités d'une variable (typiquement la somme des primes acquises ou émises par numéro de client).

Évidemment les primes émises et acquises sont des variables quantitatives. À noter que si un client n'est pas assuré sur la totalité de la période janvier 2011-juin 2020, des chaînes de caractères vides sont renseignées pour les primes des années en question.

7.4 Les sinistres

7.4.1 Les infocentres sinistres

Les informations de sinistralité de notre nouvelle base sont collectées via des requêtes sur les infocentres sinistres GMF. Les règles de prise en compte des sinistres dans ces infocentres sont les suivantes :

- Prise en compte des sinistres de l'année en cours quelque soit leur état (pour le fichier de fin février 2009, tous les sinistres survenus en 2009) ;
- Les sinistres des 5 années antérieures à l'année en cours quelque soit leur état (pour le fichier de fin février 2009, tous les sinistres survenus en 2008, 2007, 2006, 2005 et 2004) ;
- Les sinistres antérieurs aux 5 dernières années ouverts ou clos dans le mois (pour le fichier de fin février 2009, tous les sinistres survenus avant le premier janvier 2004 encore ouverts au premier février).

La figure suivante donne des exemples concrets de sinistres afin de bien assimiler les différentes règles.

Description du sinistre	Date d'arrêt du fichier			
	31/10/2009	30/11/2009	31/12/2009	31/01/2010
Sinistre survenu le 25/10/2009 déclaré le 12/11/2009 clos le 20/12/2009	Absent	Présent Ouvert	Présent Clos	Présent Clos
Sinistre survenu le 5/01/1997 déclaré le 7/01/1997 clos le 15/11/2009	Présent Ouvert	Présent clos	Absent	Absent
Sinistre survenu le 5/01/1997 déclaré le 7/01/1997 clos le 15/11/1998	Absent	Absent	Absent	Absent
Sinistre survenu le 5/01/2004 déclaré le 7/01/2004 clos le 15/01/2004	Présent clos	Présent Clos	Présent Clos	absent

FIGURE 7.8 – Règles de prise en compte des sinistres dans les infocentres

Les infocentres sinistres GMF sont constitués d'une ligne par garantie mise en jeu par les sinistres concernés. C'est-à-dire que si un sinistre impacte 3 garanties, 3 lignes sont créées dans les infocentres.

7.4.2 Les charges sinistres

La nouvelle base contient les charges sinistres par client pour chaque année comprise entre 2011 et 2020 (selon sa période d'assurance). Les charges sinistres sont égalés à la somme des décaissements (indemnités déjà versées) et des provisions pour sinistres à payer (coût total estimé des sinistres déclarés et non encore indemnisés). Cela revient, pour chaque année, à additionner toutes les charges sinistres générées par le client (tous contrats confondus, s'il en possède plusieurs de même période d'assurance). Tous les sinistres sont concernés pour déterminer les charges annuelles des clients, y compris les sinistres négatifs issus de recours. Ces montants négatifs sont la conséquence de la mise en place de la convention IRSA (Convention d'indemnisation directe de l'assuré et de Recours entre sociétés d'assurance automobile), signée par la plupart des sociétés d'assurance en France, et destinée à faciliter l'indemnisation des dommages matériels en cas d'accidents de la circulation. En cas de sinistre non responsable, l'assuré est indemnisé par son propre assureur qui ensuite effectuera un recours conventionnel auprès de (ou des) assureur(s) adverse(s) pour recevoir une indemnisation forfaitaire ou réelle. Si l'indemnisation forfaitaire reçue par l'assureur est supérieure au montant du sinistre, cela génère des gains, c'est-à-dire des sinistres négatifs. Cependant, si la charge sinistre d'un client venait à être négative pour une année en particulier, on partira du principe que la charge est nulle. Cela pourrait, en effet, générer des incohérences dans les modèles de crédibilité que nous appliquerons à la base ensuite.

Tous les sinistres sont vus à fin juin 2020. En conséquence, les exercices les plus anciens (2011, 2012) sont observés avec beaucoup plus de recul que les exercices récents (2018, 2019, 2020). En ce qui concerne les exercices anciens, la majeure partie des sinistres ont été payés, les provisions sont généralement anodines (excepté pour les sinistres graves dont la liquidation est beaucoup plus lente). Les charges de sinistralité sont donc plutôt fiables. À contrario, pour les années récentes, de nombreux sinistres n'ont pas encore été réglés, les charges sinistres sont donc estimées et peuvent être amenées à varier dans les années à venir. Pour les années récentes, il ne sera pas possible de faire mieux que la vision à fin juin 2020, aucun outil d'estimation des charges ultimes n'a été développé pour les flottes automobiles. Il aurait pu être envisageable d'observer tous les exercices à la même date (par exemple les sinistres de l'exercice N vus à fin décembre N+1). Toutefois, un an de recul n'est généralement pas suffisant pour avoir une idée claire de la charge sinistre. Par exemple, un sinistre peut se révéler être beaucoup plus grave que prévu (ou inversement) par la suite. L'idée est vraiment d'avoir la meilleure image possible de la sinistralité des clients au moment des mesures tarifaires pour faire en sorte de leur appliquer une surveillance juste et adaptée.

Au vu de certains sinistres particulièrement graves (millions d'euros) la décision a été prise d'écrêter (de plafonner) les sinistres à 150 000 euros. La surcrête, c'est-à-dire la différence entre le montant brut du sinistre et la valeur d'écrêtement (150 000), est également intégrée à notre base de données afin de conserver cette information. Cet écrêtement est également réalisé en anticipation de l'application de la théorie de la crédibilité. En effet, un sinistre très grave (plusieurs millions d'euros) pourrait détériorer significativement la prime individuelle du client sinistré ainsi que la prime collective du portefeuille. Ce point sera détaillé par la suite.

7.4.3 Le nombre de sinistres

En plus des charges sinistres, la nouvelle base affiche également le nombre de sinistres total (toutes garanties confondues) et le nombre de sinistres impactant la garantie responsabilité civile, de chaque client par année d'expérience. On expliquera par la suite l'intérêt de disposer du nombre de sinistres responsabilité civile. À noter que pour le calcul du nombre total de sinistres, un sinistre impactant plusieurs garanties différentes (RC et dommages par exemple) représente bien un seul sinistre.

7.4.3.1 Les provisions

Tout comme les sinistres, les provisions de chaque exercice sont observées à fin juin 2020.

7.5 Les contrats avec conservation

Pour des raisons financières, certains organismes publics optent pour une assurance flottes avec conservation des sinistres (voir paragraphe 1.8). Pour rappel, une conservation équivaut à une franchise annuelle, c'est-à-dire que l'assureur indemnise les sinistres seulement une fois que la somme cumulée des premiers sinistres a dépassé le montant de la franchise annuelle. Tant que ce n'est pas le cas, les sinistres sont à la charge du client. En pratique, le client verse la franchise annuelle à la compagnie d'assurance qui règle les premiers sinistres avec. Les infocentres flottes ne contiennent aucune information vis-à-vis d'une potentielle conservation et de son montant. Toutefois, tous les sinistres (même ceux n'étant pas à la charge de GMF) sont enregistrés dans les infocentres sinistres. En pratique, un sinistre négatif du montant de la conservation est généré dans les infocentres sinistres pour simuler la conservation. Il suffit donc de réaliser une requête sur des sinistres négatifs inférieurs à un certain montant pour obtenir les contrats avec conservation. Ensuite, il est possible d'aller vérifier, contrat par contrat, en production, dans le système de client standard, si ces contrats contiennent bien une clause de conservation. Dans la nouvelle base de données, une variable binaire indique si oui ou non le contrat dispose d'une franchise annuelle. Le cas échéant, une autre variable indique le montant de la franchise.

Il est important de repérer ces contrats en vue de la surveillance du portefeuille, car ils ont des particularités. Ils seront donc traités à part.

7.6 Les parcs

Dans les infocentres risques spécifiques, aucune variable ne renseigne le nombre de véhicules associés aux contrats flottes automobiles. Le courtier, affilié à GMF sur le produit flottes, tous les ans, envoie à GMF un fichier Excel indiquant le parc automobile de chaque contrat à fin décembre de l'année en cours. Malheureusement, ces fichiers Excel ne sont pas disponibles avant 2016. À ce stade, on dispose donc seulement des parcs assurés entre 2016 et 2020. On expliquera par la suite la méthode employée pour estimer les parcs automobiles des années antérieures.

Ces fichiers Excel sont exportés sous SAS afin d'extraire l'information par client et non par contrat puis de la joindre avec la base en construction. Lors de la jointure, il apparaît que certains clients ne se voient affecter aucun parc. Les données fournies par le courtier sont en partie incomplètes.

7.7 Les informations clients

Les informations clients de notre nouvelle base de données sont collectées à partir des infocentres clients des risques spécifiques. Les infocentres clients sont en phase avec les autres fichiers des risques spécifiques (même périodicité, même archivage, etc.). Il existe une ligne par client et à chaque client sont associés différents enregistrements le caractérisant. Les informations suivantes sont collectées au sein de notre nouvelle base : nom du client, raison sociale, son code postal, catégorie socioprofessionnelle de la flotte, etc.

7.8 Les données contrats

Le délai du préavis et la date d'échéance finale sont récupérés dans les infocentres contrats auto risques spécifiques. Ces informations seront utiles lors de la mise en place de la surveillance afin d'adapter les mesures tarifaires à la période d'assurance du contrat et de respecter le délai de préavis. Si un client possède plusieurs contrats en cours à fin juin 2020, la date d'échéance retenue est la plus lointaine. De même pour le délai du préavis. Ces cas sont très rares. Généralement, si un client dispose de plusieurs contrats en cours, les contrats ont la même date d'échéance et le même délai de préavis.

7.9 Les franchises

Malheureusement, dans les infocentres risques spécifiques GMF, il n'existe aucune variable indiquant les montants des différentes franchises des contrats flottes. Pour rappel, en assurance flottes, les franchises dépendent des garanties et des types de véhicules et peuvent varier significativement d'un contrat à un autre. Toutefois, ces informations sont disponibles en production dans les systèmes de gestion client utilisés par les gestionnaires. Cet outil de gestion contient toutes les informations possibles sur les clients, les contrats et les sinistres. Il est utilisé par les gestionnaires pour interroger, consulter, créer ou modifier des informations relatives à un contrat. Par exemple, il est possible d'y réaliser des devis, de créer un contrat, un client, etc. Évidemment, sur cet outil, les recherches peuvent être effectuées uniquement contrat par contrat. Le portefeuille flottes automobiles GMF à fin juin 2020 étant relativement petit (environ 200 contrats), il était donc envisageable de récupérer, grâce à l'outil de gestion, pour chaque contrat, les différentes franchises souscrites en fonction des garanties (dommages, vol, incendie, bris de glace) et des types de véhicules (moins de 3 tonnes 5, plus de 3 tonnes 5, remorques, 2 roues, etc.).

7.10 Le nettoyage de la base

Les données récupérées étant brutes, certains retraitements étaient nécessaires pour améliorer la lisibilité de la base, pour anticiper la création de futures variables (ratio sinistres à primes, fréquences, etc.) et en vue de l'application de la théorie de la crédibilité sur la base.

Par exemple, initialement, si un client n'avait aucun sinistre lors d'une année N (incluse dans la période d'assurance du contrat), des chaînes de caractères vides " " étaient renseignées dans les variables "charge totale", "nombre de sinistres" et "nombre de sinistres responsabilité civile". Comme ce client n'avait aucun sinistre cette année-là, il n'existait aucune information de sinistralité pour ce client dans les infocentres de l'année en question, d'où la chaîne de caractère vide.

	Prime acq 2017	Prime acq 2018	Prime acq 2019	Prime acq 2020	charge 2017	charge 2018	charge 2019	charge 2020
client B	4000	4300	5000	2500	2000	" "	3400	500

FIGURE 7.9 – Problème des chaînes de caractères vides

Or, une chaîne de caractère vide était également indiquée dans le cas où un contrat n'était pas assuré sur ces années-là. Cela pouvait donc porter à confusion. Il a été décidé d'affecter un 0 aux variables de sinistralité si le client était bien assuré (une prime était payée) mais qu'il n'avait tout simplement aucun sinistre à déplorer et de conserver la chaîne de caractère vide " " pour les années où le contrat n'était pas assuré. Cela permettra également de calculer plus facilement le ratio sinistres à primes et de n'avoir aucune anomalie dans les modèles de crédibilité à venir. En effet, dans un des modèles, les primes joueront le rôle de poids. Elles doivent donc être associées à une charge de sinistralité positive (≥ 0) et non à une chaîne de caractère vide.

	Prime acq 2017	Prime acq 2018	Prime acq 2019	Prime acq 2020	charge 2017	charge 2018	charge 2019	charge 2020
client B	4000	4300	5000	2500	2000	0	3400	500

FIGURE 7.10 – Traitement des 0

De plus, après quelques vérifications, les données de parc fournies par le courtier ont révélé quelques incohérences et ont nécessité un travail de retraitement. Dans ces fichiers Excel (un par année entre 2016 et 2020), à chaque contrat flotte GMF est associé le parc automobile à fin décembre de l'année en question. Cependant, il arrivait de temps en temps qu'un contrat débutant en année N+1 soit inclus dans le fichier de l'année N. En conséquence, dans notre base, des clients ne payant pas de primes l'année N étaient associés à un parc cette année-là. A-delà des soucis de cohérence, cela risquait de poser des problèmes notamment pour les calculs de fréquence et de primes par véhicules. Autre cas, certains clients pour lesquels les contrats se chevauchaient (un contrat se terminant par exemple le 31/12/N et un autre commençant le 01/01/N+1) pouvaient se voir attribuer la même année 2 parcs automobiles au lieu d'un seul, il fallait donc faire attention aux potentiels doublons. La décision a été prise d'analyser les évolutions de parcs de chaque client pour y déceler de potentielles anomalies et le cas échéant d'aller regarder en détail dans les infocentres primes si le contrat en question était bien assuré sur cette période.

Quelques incohérences ont également été remarquées en ce qui concerne l'affectation de la catégorie socioprofessionnelle du client. Il pouvait arriver que des clients de même catégorie socioprofessionnelle se voient affecter une modalité différente. Par exemple, certaines flottes SDIS (service départemental d'incendie et de secours) étaient affectées à la modalité "Caserne", d'autres à la modalité "Service", d'autres à la modalité "Association", etc. En fait, il existe beaucoup de modalités possibles concernant les titres professionnels des flottes, certaines se ressemblent et peuvent porter à confusion. Il arrive donc que le gestionnaire du contrat se trompe au moment de saisir cette information dans l'interface client. Cette information n'est donc pas très fiable, elle est à manipuler avec précaution. Certains retraitements ont pu être réalisés "à la main" grâce au nom du client. Typiquement les clients SDIS ont été regroupés ensemble au sein d'une même modalité, de même pour les mairies et les communes. Cependant, il n'a pas été possible de réaliser ce travail sur tous les clients, car certains noms n'étaient pas toujours explicites quant à leur activité professionnelle.

Les données géographiques sont collectées à l'échelle du code postal. Le portefeuille flottes GMF étant relativement petit, cette information n'est pas spécialement pertinente. Cependant, il est important de disposer d'une information sur la localisation de la flotte car c'est une variable potentiellement tarifaire. Le compromis a été fait de travailler à l'échelle régionale selon le nouveau découpage des régions (14 au lieu de 22). La nouvelle variable géographique comporte donc 14 modalités.

7.11 Création de nouvelles variables

Une fois le retraitement réalisé, des nouvelles variables sont déterminées à partir des variables existantes dans notre base.

7.11.1 Ancienneté du client

L'ancienneté du client est exprimée en années et calculée à partir de l'année 2011 en fonction des primes acquises.

7.11.2 Ratio sinistres à primes

Le ratio sinistres à primes, appelé couramment S/P, est un indicateur utilisé par les assureurs pour estimer le rendement de leurs produits d'assurance ainsi que pour calibrer le meilleur équilibre tarifaire en fonction des sinistres constatés. Ce ratio met en rapport le coût des sinistres avérés (déjà payés) et évalués (sinistres déclarés, mais pas encore indemnisés) et le montant total des primes acquises.

$$S/P_{i,j} = \frac{\text{Charge}_{i,j}}{\text{Prime acquise}_{i,j}}$$

Avec :

- Soit i un client du portefeuille 2011-2020 ;
- Soit j une année comprise entre 2016 et 2020 ;
- Primes acquises hors assistance, hors taxes ;
- Charges de sinistralité écartées à 150 000 euros de l'exercice j, hors réassurance, hors frais généraux, observées à fin décembre j+1.

7.11.3 Fréquence de sinistralité

En plus du montant des accidents, la sinistralité en assurance automobile se mesure également en termes de fréquence.

$$\text{Fréquence}_{i,j} = \frac{\text{Nombre de sinistres}_{i,j}}{\text{Parc}_{i,j}}$$

Avec :

- Soit i un client du portefeuille 2011-2020 ;
- Soit j une année comprise entre 2016 et 2020 ;
- Nombre de sinistres de l'exercice j observés à fin décembre de l'année $j+1$;
- Parc du client i à fin décembre de l'année j .

Les assureurs flottes attachent également beaucoup d'importance à la fréquence de sinistralité de la garantie responsabilité civile car ces sinistres peuvent potentiellement engendrer des coûts très importants (sinistres corporels).

$$\text{Fréquence RC}_{i,j} = \frac{\text{Nombre de sinistres RC}_{i,j}}{\text{Parc}_{i,j}}$$

Avec :

- Soit i un client du portefeuille 2011-2020 ;
- Soit j une année comprise entre 2016 et 2020 ;
- Nombre de sinistres impactant la garantie responsabilité civile de l'exercice j observés à fin décembre de l'année $j+1$;
- Parc du client i à fin décembre de l'année j .

7.12 Données manquantes

Les données manquantes concernent majoritairement les parcs automobiles associés aux contrats. Premièrement, comme expliqué précédemment, pour certains clients, on ne connaît pas le nombre de véhicules assurés. Deuxièmement, les parcs fournis par le courtier ne concernent que les années 2016 à 2019 (à fin juin 2020, on ne dispose pas des parcs de l'année 2020 car ils sont transmis à fin décembre 2020) ; or l'étude porte sur la période de temps 2011-2020. Troisièmement, dans les fichiers transmis par le courtier, la répartition du parc (-3 tonnes 5, + 3 tonnes 5, 2 roues, remorques, etc.) par contrat n'est pas spécifiée. Cette information est cruciale en assurance flottes. Il n'était pas envisageable de joindre le courtier directement pour obtenir ces renseignements, car le délai de réponse et de transmission des données aurait été particulièrement long. Il fallait donc envisager d'autres méthodes pour estimer au mieux ces données.

Chapitre 8

Estimations des données manquantes

8.1 Estimation des parcs

Au sein de notre base, pour chaque année de la période d'assurance d'un client, un parc doit lui être associé. C'est important en vue de l'application de la théorie de la crédibilité sur notre base. En effet, dans un des modèles, les parcs seront assimilés à des poids, il est donc indispensable d'avoir cette information. Nous y reviendrons par la suite. Les parcs étaient seulement disponibles entre 2016 et 2019. Cela signifiait que si un client était assuré depuis 2014, les parcs 2014 et 2015 étaient manquants. Un exemple fictif de contrat est proposé pour bien comprendre le mécanisme de l'estimation des parcs manquants.

	Prime acq 2014	Prime acq 2015	Prime acq 2016	Prime acq 2017	Prime acq 2018	Parc 2014	Parc 2015	Parc 2016	Parc 2017	Parc 2018
client B	3800	4000	4300	5000	4700			11	13	12

FIGURE 8.1 – Exemple fictif de données manquantes de parcs automobiles

Pour rappel, les parcs sont amenés à évoluer d'une année à l'autre (ajouts ou suppressions de véhicules au sein de la flotte). Il fallait mettre en place une méthode pour déterminer les parcs à partir des informations disponibles dans notre nouvelle base.

La première étape est de calculer les primes acquises par véhicule pour chaque année pour laquelle on disposait des parcs automobiles (à partir de 2016).

$$\text{Prime acquise par véhicule}_{i,j} = \frac{\text{Prime acquise}_{i,j}}{\text{Parc}_{i,j}}$$

Avec :

- Soit i un client du portefeuille 2011-2020 ;
- Soit j une année comprise entre 2016 et 2020 ;
- Primes acquises hors assistance, hors taxes ;
- Parc du client i à fin décembre de l'année j .

	Prime acq par véhicule 2014	Prime acq par véhicule 2015	Prime acq par véhicule 2016	Prime acq par véhicule 2017	Prime acq par véhicule 2018
client B			390,9	384,6	391,7

FIGURE 8.2 – Exemple fictif du calcul des primes acquises par véhicule

Ensuite, l'idée est d'estimer les primes acquises par véhicules des années antérieures à 2016 en se basant sur la prime acquise par véhicule en 2016 et l'inflation du coût des réparations automobiles (indice SRA) ces années-là.

L'indice SRA correspond à la moyenne de trois indices (pièces de rechange, main d'oeuvre et peinture). Les assureurs s'intéressent particulièrement à l'évolution de cet indice qui explique en grande partie l'augmentation des coûts moyens. Les contrats auto sont généralement indicés par l'indice SRA, c'est-à-dire que l'augmentation des cotisations est corrélée à l'augmentation de l'indice. Il en est de même pour les contrats flottes. Pour rappel, avant juin 2020, les contrats flottes GMF ne subissaient aucune augmentation tarifaire autre que celle relevant de l'évolution de cet indice.

Les indices SRA sont communiqués trimestriellement. Ces données sont publiques et accessibles via internet sur le site de l'association SRA.

		INDICES SRA		
		Pièces	MO	Peinture
2015	T1	99,8	99,8	99,8
2015	T2	100,2	100,1	99,8
2015	T3	100,2	100,1	100,2
2015	T4	99,6	100	100,2
2016	T1	101,1	102,4	103,5
2016	T2	101,3	102,4	103,5
2016	T3	101,2	102,6	103,6
2016	T4	100,5	102,6	103,6

FIGURE 8.3 – Indices SRA de l'année 2015 et 2016

À partir des indices des différentes années, il est possible de calculer les évolutions d'une année à l'autre. L'objectif est de calculer toutes les évolutions entre 2011 et 2016.

Le calcul de l'évolution de l'indice SRA entre j et $j+1$ se fait en déterminant l'évolution de la moyenne des 3 indices entre le trimestre 4 de l'année N et le trimestre 4 de l'année $j+1$.

$$\text{Évolution SRA}_{j,j+1} = \frac{\text{indice}_{j+1,T4} - \text{indice}_{j,T4}}{\text{indice}_{j,T4}}$$

Avec :

- Soit j une année ;
- Soit T4 le trimestre 4 de l'année j ;
- $\text{indice}_{j,T4} = \frac{\text{indice Pièces}_{j,T4} + \text{indice Main d'oeuvre}_{j,T4} + \text{indice Peinture}_{j,T4}}{3}$.

		INDICES SRA					
		Pièces	MO	Peinture	moyenne Indice		
2015	T1	99,8	99,8	99,8	99,80		
2015	T2	100,2	100,1	99,8	100,03		
2015	T3	100,2	100,1	100,2	100,17		
2015	T4	99,6	100	100,2	99,93		
2016	T1	101,1	102,4	103,5	102,33		
2016	T2	101,3	102,4	103,5	102,40		
2016	T3	101,2	102,6	103,6	102,47	evo 16-15	
2016	T4	100,5	102,6	103,6	102,23	2,30%	

FIGURE 8.4 – Évolution de l'indice SRA entre 2015 et 2016

Une fois que les évolutions SRA des années 11-12, 12-13, 13-14, 14-15, 15-16 sont calculées de manière récursive, l'idée est d'appliquer l'évolution SRA 15-16 à la prime acquise par véhicule 2016 pour estimer la prime acquise par véhicule 2015 puis ensuite d'appliquer l'évolution SRA 14-15 à la prime par véhicule 2015 pour estimer la prime acquise par véhicule 2014 et ainsi de suite.

$$\text{Estimation Prime acquise par véhicule}_{i,2015} = \frac{\text{Prime acquise par véhicule}_{i,2016}}{1 + \text{Évolution SRA}_{2015,2016}}$$

$$\text{Estimation Prime acquise par véhicule}_{i,j} = \frac{\text{Estimation Prime acquise par véhicule}_{i,j+1}}{1 + \text{Évolution SRA}_{j,j+1}}$$

Avec :

- Soit i un client pour lequel on dispose au moins d'une information de parc entre 2016 et 2020 ;
- Soit j une année comprise entre 2011 et 2014 ;

	Prime acq par véhicule 2014	Prime acq par véhicule 2015	Prime acq par véhicule 2016	Prime acq par véhicule 2017	Prime acq par véhicule 2018
client B	?	?	390,9	384,6	391,7
evol SRA 15-16			2,30%		
	Prime acq par véhicule 2014	Prime acq par véhicule 2015	Prime acq par véhicule 2016	Prime acq par véhicule 2017	Prime acq par véhicule 2018
client B		382,12	390,9	384,6	391,7
evol SRA 14-15			2,82%		
	Prime acq par véhicule 2014	Prime acq par véhicule 2015	Prime acq par véhicule 2016	Prime acq par véhicule 2017	Prime acq par véhicule 2018
client B	371,6	382,12	390,9	384,6	391,7

FIGURE 8.5 – Exemple de l'estimation des primes acquises par véhicules

En appliquant cette évolution d'indice à la prime acquise, cela permet de modéliser les mesures tarifaires indicelles réalisées par GMF chaque année. Pour chaque contrat, à parc constant, la prime par véhicule est censée être croissante avec le temps, car la prime augmente chaque année selon l'indice SRA (indice croissant en fonction du temps). En pratique, le parc est toujours en mouvement et dans ce cas, la phrase précédente est fautive. En effet, la suppression de véhicules lourds (prime élevée) au profit de véhicules légers (primes faibles) aura tendance à diminuer la prime par véhicule.

Une fois les primes acquises par véhicule déterminées pour les années qui posent problème, il ne reste plus qu'à estimer le nombre de véhicules en divisant la prime totale acquise par la prime acquise par véhicule et en arrondissant à l'entier près.

$$\text{Estimation Parc}_{i,j} = \frac{\text{Prime acquise}_{i,j}}{\text{Estimation Prime acquise par véhicule}_{i,j}}$$

Avec :

- Soit i un client pour lequel on dispose au moins d'une information de parc entre 2016 et 2020 ;
- Soit j une année comprise entre 2011 et 2015 ;

	Prime acq 2014	Prime acq 2015	Prime acq 2016	Prime acq 2017	Prime acq 2018	Parc 2014 estimé	Parc 2015 estimé	Parc 2016	Parc 2017	Parc 2018
client B	3800	4000	4300	5000	4700	10	10	11	13	12

FIGURE 8.6 – Exemple de l'estimation des parcs

Cette estimation n'est évidemment pas parfaite, mais généralement la prime acquise par véhicule, d'un contrat, d'une année à l'autre, n'évolue pas significativement, car les parcs sont généralement relativement stables. Il est donc plutôt pertinent de modéliser les évolutions de la prime acquise par véhicule uniquement avec l'évolution de l'indice SRA. De cette façon, les ordres de grandeur des parcs sont respectés. Cette modélisation a également l'avantage d'être rendue possible pour les contrats avec conservation. Ces contrats, de par leur montage financier, ont pour conséquence des primes par véhicule beaucoup plus faibles.

8.2 Estimation de la répartition des véhicules

Malheureusement, il n'est pas possible de récupérer d'une quelconque façon la répartition des véhicules (proportion de 2 roues, remorques, véhicules lourds, etc.) au sein des contrats, que ce soit via les infocentres risques spécifiques ou le système de gestion client. Cependant, avec les infocentres auto risques spécifiques, il est possible de récupérer à chaque fin d'année (à partir de 2016 seulement), par contrat, les primes émises par type de véhicule selon les différentes garanties souscrites dans le contrat.

Client	Contrat	catégorie de risque	Prime RC	Prime DOM	Prime VOL	Prime totale
A	A1	"- 3 tonnes 5"	12000	6000	1000	19000
A	A1	"+ 3 tonnes 5"	4000	0	0	4000
A	A1	"2 roues"	300	0	150	450
A	A1	"remorques"	1000	0	0	1000

FIGURE 8.7 – Exemple fictif des primes émises d'un contrat par catégorie de risque

Il suffit donc maintenant d'avoir une idée des primes par véhicule par catégorie de risque pour déterminer le nombre de véhicules par catégorie. Cependant, ce n'est pas tout, nous aimerions pouvoir distinguer le nombre de véhicules -3 tonnes 5 tous risques versus le nombre de véhicules -3 tonnes 5 "hors tous risques". De même avec les +3 tonnes 5 et les + 3 tonnes 5 tous risques. En effet, il n'existe pas de formules de garanties pour tous les véhicules, les formules de garantie dépendent de l'âge du véhicule, du type de véhicule et du choix du client.

Le problème réside dans le fait qu'aucune distinction n'est faite dans les infocentres entre ces 2 catégories différentes. Pour cela, il est donc indispensable d'estimer au mieux les primes totales payées par les -3 tonnes 5 tous risques des primes totales payées par les -3 tonnes 5 hors tous risques. De même avec les +3 tonnes 5 et les +3 tonnes 5 tous risques. À partir des primes payées par garantie par les -3 tonnes 5 ou les +3 tonnes 5, l'objectif est de calculer les primes payées par garantie par les véhicules tous risques d'une part et par les véhicules "hors tous risques" d'autre part pour pouvoir finalement déterminer la prime totale payée par les tous risques versus la prime totale payée par les "hors tous risques" et d'estimer le nombre de ces véhicules (-3 tonnes 5 tous risques, -3 tonnes 5, +3 tonnes 5 tous risques, +3 tonnes 5). Les véhicules n'étant pas assurés en tous risques disposent généralement des mêmes garanties que les tous risques (incendie, vol, etc.) hormis la garantie dommages évidemment. La seule information dont on dispose est donc la prime dommage payée par les véhicules tous risques. Hormis cette évidence, la répartition des primes par garanties entre les véhicules tous risques et les autres véhicules n'est pas connue.

catris	RC	DOM	VOL	Prime_tot
"- 3 tonnes 5"	12000	6000	3000	21000

catris	RC	DOM	VOL	Prime_tot
"- 3 tonnes 5" TR	?	6000	?	?
"- 3 tonnes 5"	?	0	?	?

FIGURE 8.8 – Problème de distinction des véhicules tous risques des véhicules hors tous risques

À partir de la prime dommages, il va falloir estimer les autres primes par garantie payées par les véhicules tous risques. L'idée est de se dire que si un client paie une prime x en dommages alors il paie $a \cdot x$ pour la garantie RC, $b \cdot x$ pour la garantie VOL, $d \cdot x$ pour la garantie INC, etc. Cela revient donc à déterminer ces coefficients. Ensuite, il ne restera plus qu'à soustraire, pour chaque garantie, la prime totale payée pour les 2 catégories (tous risques et hors tous risques) par la prime payée estimée pour les tous risques. Puis d'additionner ces primes par garantie par catégorie pour déterminer les primes totales estimées pour les tous risques et les "hors tous risques".

Pour estimer les coefficients, l'idée a été d'utiliser les données des contrats auto GMF pour lesquels le volume est bien plus important. Le travail s'est déroulé en plusieurs étapes. Tout d'abord, la sélection des contrats auto (pour les particuliers) avec une formule tous risques, à fin décembre 2019. Ensuite la récupération des primes puis le calcul des coefficients.

$$\text{Coefficient}_b = \frac{\sum_{k=1}^N \text{Prime}_{k,b}}{\sum_{k=1}^N \text{Prime dommages}_k}$$

Avec :

- Soit b une garantie incluse dans la sélection suivante : responsabilité civile, vol, incendie, bris de glace, défense-recours, conducteur ;
- Soit k un contrat auto assuré en tous risques ;
- Soit N le nombre de contrats auto assurés en tous risques à fin décembre 2019.
- $\text{Prime}_{k,b}$ la prime payée pour la garantie b par le contrat k .

Garantie	Coefficient
Dommages	1
Responsabilité civile	0,73
Bris de glace	0,2
Vol	0,18
Assurance conducteur	0,16
Défense-recours	0,06
Incendie	0,03

TABLE 8.1 – Tableau des coefficients des primes en fonction des primes dommages

Ce tableau indique tout simplement que si un assuré paie sa prime dommages 100 euros, alors sa prime responsabilité civile vaut approximativement 73€, sa prime bris de glace approximativement 20 euros, etc. Grâce à ce tableau, il est donc maintenant possible de déterminer, à partir de la prime payée pour la garantie dommage, les primes payées pour les autres garanties en multipliant la prime dommage par le coefficient de la garantie pour laquelle on souhaite déterminer la prime. Revenons à notre exemple. Ensuite, la prime totale de la catégorie "tous risques" est calculée en faisant la somme des primes par garantie.

catris	RC	DOM	VOL	Prime_tot
"- 3 tonnes 5"	12000	6000	3000	21000

catris	RC	DOM	VOL	Prime_tot
"- 3 tonnes 5" TR	4380	6000	1080	11460
"- 3 tonnes 5"	?	0	?	?

	coef
DOM->RC	0,73
DOM->Vol	0,18

FIGURE 8.9 – Exemple fictif du calcul des primes tous risques

Encore une fois, cette méthode présente quelques défauts, les coefficients sont calculés à partir du produit auto GMF et non du produit flottes. Toutefois, on peut faire l'hypothèse que les ordres de grandeur sont les mêmes entre ces deux produits.

Par soustraction, on peut donc maintenant déterminer les primes par garantie payées par la catégorie "hors tous risques" ainsi que la prime totale.

catris	RC	DOM	VOL	Prime_tot
"- 3 tonnes 5"	12000	6000	3000	21000

catris	RC	DOM	VOL	Prime_tot
"- 3 tonnes 5" TR	4380	6000	1080	11460
"- 3 tonnes 5"	7620	0	1920	9540

FIGURE 8.10 – Exemple fictif du calcul des primes hors tous risques

La segmentation des véhicules "tous risques" et des véhicules "hors tous risques" a donc été réalisée.

Client	Contrat	catégorie de risque	Prime totale
A	A1	"- 3 tonnes 5"	21000
A	A1	"+ 3 tonnes 5"	4000
A	A1	"2 roues"	450
A	A1	"remorques"	1000

↓

Client	Contrat	catégorie de risque	Prime totale
A	A1	"- 3 tonnes 5 TR"	11460
A	A1	"- 3 tonnes 5"	9540
A	A1	"+ 3 tonnes 5"	4000
A	A1	"2 roues"	450
A	A1	"remorques"	1000

FIGURE 8.11 – Exemple fictif de la segmentation des véhicules tous risques des véhicules hors tous risques

Ensuite, il reste à disposer d'estimations des primes par véhicule par catégorie de risque pour pouvoir déterminer le nombre de véhicules par catégorie. Encore une fois, l'idée est de travailler sur les produits d'assurance des particuliers GMF (4 roues, 2 roues) pour estimer ces primes de manière précise. Les contrats GMF "véhicules terrestres à moteur" sont vendus en grande quantité, on dispose donc d'informations fiables. Pour le produit GMF 4 roues, les primes tous risques et hors tous risques sont déterminées. Concernant le produit d'assurance 2 roues, la distinction n'est pas faite. Le calcul est tout simple, à partir des infocentres GMF (auto ou 2 roues), il consiste à faire une simple moyenne des primes annuelles 4 roues (tous risques et hors tous risques) et des primes annuelles 2 roues (en exceptant les contrats 2 roues temporaires, assurés seulement une partie de l'année), à fin 2019.

Catégorie de risque	Prime par véhicule
Moins 3 tonnes 5	220€
Moins 3 tonnes 5 tous risques	390€
2 roues	210€

TABLE 8.2 – Primes par véhicules : 4 roues et 2 roues

En moyenne, une couverture tous risques annuelle en auto coûte 390€.

Les remorques (-3 tonnes 5, + 3 tonnes 5), les engins véhicules spéciaux (EVS), les véhicules transport public de voyageurs (TPV) et les poids lourds (+ 3 tonnes 5) sont des véhicules ne pouvant pas être assurés par un contrat GMF classique. Ils concernent les risques spécifiques GMF et sont donc assurés par la filiale de GMF, appelée Sauvegarde. Pour chacune de ces catégories, on calcule la prime moyenne à partir des infocentres auto risques spécifiques en faisant la distinction "tous risques" et "hors tous risques" pour les poids lourds.

Catégorie de risque	Prime par véhicule
Plus 3 tonnes 5	500€
Plus 3 tonnes 5 tous risques	1 300€
Remorques moins 3 tonnes 5	70€
Remorques plus 3 tonnes 5	100€
Engins	1 200€
Véhicules de transport public	300€

TABLE 8.3 – Primes par véhicules : Poids lourds, remorques, engins

Toutefois, bien que les ordres de grandeur semblent être respectés, les primes moyennes déterminées pour les poids lourds, les remorques, les engins et les véhicules transport public sont basées sur peu de données. En effet, ces portefeuilles représentent peu de volume par rapport aux portefeuilles 4 roues et 2 roues de GMF.

Il est donc maintenant possible de déterminer le nombre de véhicules par type de risque pour les années 2016 à 2019.

$$\text{Estimation véhicules}_{i,j,c} = \frac{\text{Prime}_{i,j,c}}{\text{Estimation Prime par véhicule}_c}$$

Avec :

- Soit c un type de risque (- 3 tonnes 5 tous risques, -3 tonnes 5, 2 roues, etc.) ;
- Soit i un client disposant d'une période d'assurance incluse entre 2016 et 2020 ;
- Soit j une année comprise entre 2016 et 2019
- $\text{Prime}_{i,j,c}$ la prime totale émise par le client i pour l'exercice j (observé à fin décembre de l'année j) pour les véhicules de catégorie c .

Ensuite, il ne reste qu'à calculer la proportion de chaque risque au sein de la flotte pour avoir une idée de la répartition des véhicules. Voici ce que cela donne sur notre exemple fictif.

Client	Contrat	Période assurance	catégorie de risque	Prime totale	catégorie de risque	prime par véhicule
A	A1	01/01/17-31/12/18	"- 3 tonnes 5 TR"	11460	"moins 3 tonnes 5 TR"	390 €
A	A1	01/01/17-31/12/18	"- 3 tonnes 5"	9540	"moins 3 tonnes 5"	220 €
A	A1	01/01/17-31/12/18	"+ 3 tonnes 5"	4000	"plus 3 tonnes 5"	500 €
A	A1	01/01/17-31/12/18	"2 roues"	450	"2 roues"	210 €
A	A1	01/01/17-31/12/18	"remorques -3 tonnes 5"	1000	"remorque - 3 tonnes 5"	70 €

Client	Contrat	catégorie de risque	catégorie de risque	Prime totale	Estimation véhicules	Proportion
A	A1	"- 3 tonnes 5 TR"	"- 3 tonnes 5 TR"	11 460 €	29	30%
A	A1	"- 3 tonnes 5"	"- 3 tonnes 5"	9 540 €	43	45%
A	A1	"+ 3 tonnes 5"	"+ 3 tonnes 5"	4 000 €	8	8%
A	A1	"2 roues"	"2 roues"	450 €	2	2%
A	A1	"remorques -3 tonnes 5"	"remorques -3 tonnes 5"	1 000 €	14	15%

FIGURE 8.12 – Exemple fictif de l'estimation du nombre de véhicules par type de risque

Cette méthode permet d'estimer pour tous les contrats flottes présents dans le portefeuille entre 2016 et 2019 la répartition de leur parc ces années-là. On ne cherchera pas à estimer la répartition des parcs des années antérieures à 2016, car cela n'apporterait pas une plus-value importante en vue de l'application de nos modèles de crédibilité.

On pourrait remettre en question la méthode de calcul des primes par type de véhicule. Effectivement, les produits sur lesquels on se base pour faire ces calculs ne sont pas des produits flottes. Par exemple, la prime par véhicule d'une voiture assurée en tous risques au sein d'une flotte est légèrement différente de la prime par véhicule d'une voiture assurée en tous risques avec un contrat auto GMF. Cependant, l'hypothèse de considérer que ces primes sont relativement semblables n'est pas aberrante d'autant plus que les formules de garantie sont sensiblement les mêmes.

8.3 Synthèse récupération/collecte des données

Les données des contrats flottes ont été particulièrement difficiles à récupérer et ont demandé beaucoup d'investissement, d'adaptations et de réflexions. Toutes les sources possibles de données (infocentres, systèmes de gestion, etc.) ont été envisagées et utilisées dans le but d'avoir à disposition le meilleur support possible pour y développer les modèles de crédibilité. L'activité flotte n'en est encore qu'à ses débuts à GMF, elle se développe, mais peu d'études avaient été réalisées sur ce produit jusqu'à présent. Ceci explique pourquoi certaines données ne sont pas encore disponibles dans le décisionnel, elles n'ont pas fait l'objet de demandes spécifiques sous forme de cahier des charges à l'informatique. Toutefois, le problème majeur concernait les informations de parc transmises par le courtier, des informations sont manquantes, d'autres sont approximatives et d'autres sont fausses. En addition, ces fichiers ne contiennent aucune information concernant la répartition du parc et le nombre de véhicules par risque. Or, il n'était pas envisageable de ne pas avoir les informations concernant le parc automobile des différentes flottes présentes en portefeuille. En conséquence, le seul recours possible a été d'estimer ces données au mieux. La création de la base s'est retrouvée confrontée à tous les problèmes possibles de collecte de données : données brutes non retraitées, données incohérentes liées à des erreurs de saisie, manque de liaison des données, données manquantes à estimer, sources multiples, etc. Le faible volume du portefeuille flotte de GMF facilite les différents contrôles "à la main". En effet, comme la base contient des données estimées, il a été très important de s'assurer des bons ordres de grandeur et de réaliser des vérifications simples à la main pour détecter de potentielles anomalies.

8.4 Synthèse des informations disponibles dans la base

En synthèse, la base de données contient les informations suivantes des clients flottes automobiles GMF sur la période janvier 2011-juin 2020 :

- ▶ Les données client (nom, région, activité professionnelle, ancienneté) ;
- ▶ Les données contrats (conservation, délai de préavis, date d'échéance de la période d'assurance, parc assuré, répartition du parc, franchises, primes acquises).
- ▶ Les données sinistres (charges de sinistralité, nombre de sinistres TTG et RC) ;
- ▶ Les résultats (fréquences de sinistralité TTG et RC, Ratio sinistres à primes).

Quatrième partie

La théorie de la crédibilité avancée

Chapitre 9

Crédibilité de stabilité

9.1 Introduction

La théorie de la crédibilité regroupe un ensemble de techniques utilisées par les actuaires pour déterminer la prime d'un assuré dans un portefeuille hétérogène. La tarification doit répondre à deux objectifs principaux :

- Les primes doivent être suffisantes pour payer les sinistres ;
- Les primes doivent être distribuées équitablement entre les assurés. Si le portefeuille est hétérogène et que l'assureur applique la même prime à tout le monde, il risque de perdre ses bons contrats qui auraient le sentiment de payer trop cher (risque d'anti-sélection) et au contraire attirer les mauvais risques.
- Le risque doit être mutualisé, il n'est pas envisageable d'individualiser totalement la prime.

La tarification basée sur l'expérience ("Experience rating") a pour objectif d'affecter à chaque risque sa prime juste et équitable. Cette méthode de tarification requiert un volume d'expérience important. C'est pourquoi elle est utilisée en assurance automobile ou en accidents du travail. À contrario, cette tarification n'est pas envisageable en assurance vie ou en assurance habitation où les fréquences sont plus faibles.

La théorie de la crédibilité est segmentée en 2 courants de pensée :

- La crédibilité de stabilité américaine ("limited fluctuations") consistant à tenir compte de l'expérience individuelle seulement si celle-ci est suffisamment stable dans le temps.
- La crédibilité de précision européenne ("greatest accuracy") consistant à tenir compte de l'expérience individuelle de façon à obtenir la meilleure estimation de l'expérience future. Dans ce cas, le poids de l'expérience individuelle augmente avec l'hétérogénéité du portefeuille.

9.2 La naissance de la théorie de la crédibilité

La théorie de la crédibilité serait née au début du 20^e siècle. À cette période, le groupe d'assurance Allstate assure des constructeurs automobiles, dont la grande multinationale General Motors et le petit constructeur indépendant Tucker, contre les accidents du travail. Chaque assuré paie une prime d'assurance tout à fait classique selon sa taille et la sinistralité historique du portefeuille (un taux moyen, appliqué à chaque constructeur, est calculé à partir de l'expérience de l'ensemble du portefeuille). De son côté, General Motors, émet l'hypothèse que sa sinistralité serait meilleure que celles des autres constructeurs. La compagnie se met donc à calculer son propre taux selon sa propre expérience de sinistralité et s'aperçoit, année après année, que son taux est inférieur au taux moyen qui lui est appliqué. Elle demande donc à Allstate de lui appliquer son propre taux en prétextant que son expérience de sinistralité est suffisamment stable du fait de son nombre d'employés important. Dans un premier temps, les actuaires d'Allstate sont intuitivement d'accord avec cet argumentaire et semblent prêts à accepter. La question ne se pose pas d'en faire de même pour le modeste constructeur Tucker dont l'expérience n'est pas jugée fiable.

Toutefois, les actuaires hésitent, un problème se pose : quel est le nombre minimal d'employés à considérer pour juger de la stabilité de l'expérience d'un client ?

9.3 La crédibilité complète

Selon Vincent GOULET, professeur titulaire de l'Université de Laval, Mowbray aurait été le premier à apporter des éléments de réponse en 1914. Pour cela, il définit la crédibilité complète d'ordre (k, p) , c'est-à-dire une prime pure "fiable" comme "une prime pour laquelle la probabilité est forte qu'elle ne diffère pas de la vraie prime par plus d'une limite arbitraire".

Mathématiquement, cela s'écrit de la manière suivante :

$$\mathbb{P}\left((1 - k)E[S] \leq S \leq (1 + k)E[S]\right) \geq p.$$

Où :

- k est petit (généralement 5%) ;
- p est proche de 1 (généralement 0,9 0,95 ou 0,99) ;
- S représente l'expérience d'un contrat, sous une forme ou une autre. Dans le cadre d'Allstate, S est associé au nombre d'accidents par année. $S \sim \mathcal{B}(n, \theta)$ avec n le nombre d'employés et θ la probabilité d'accident (connue).

Il ne reste ensuite qu'à isoler n (le nombre d'employés) en fonction de k et θ dans l'équation précédente pour déterminer le nombre d'employés minimal pour considérer l'expérience d'un employeur pleinement crédible.

En crédibilité complète, si l'expérience d'un contrat d'assurance est stable alors il est considéré comme crédible. L'expérience d'un contrat fait appel à sa taille. Cette taille peut être exprimée à l'aide de différents indicateurs : le nombre d'employés (exemple d'Allstate), le nombre de sinistres, le volume de prime, la masse salariale, le nombre d'années d'expérience, etc.

Ce premier modèle de crédibilité complète présente un défaut majeur. En définissant un seuil au-dessus duquel une pleine crédibilité peut être accordée, cela implique que des clients pour lesquels les valeurs sont proches du seuil, mais légèrement en dessous ne se verront accorder aucune crédibilité, le modèle est binaire : l'expérience du client est fiable et il est possible de s'y référer totalement ou alors elle ne l'est pas et la sinistralité historique du portefeuille est appliquée. Cette approche semble excessive pour certains détracteurs. Pendant les années suivantes, des mathématiciens vont chercher à optimiser le modèle de Mowbray en y intégrant une nuance entre la sinistralité individuelle et collective. C'est la naissance de la crédibilité partielle.

9.4 La crédibilité partielle

La crédibilité partielle permet de tenir compte de l'expérience individuelle d'un contrat se trouvant sous le seuil de crédibilité. En 1918, Whitney est le premier à proposer de pondérer l'expérience individuelle et la prime collective par un facteur de crédibilité. La prime est de la forme :

$$\pi = zS + (1 - z)m.$$

Où :

- π la prime prédite pour un contrat ;
- S l'expérience individuelle du contrat ;
- m la prime collective du portefeuille ;
- z le facteur de crédibilité (variant entre 0 et 1) lié au contrat

Reste encore à calculer ce facteur de crédibilité. À cette époque, plusieurs formules ont été proposées.

La formule de Whitney :

$$z = \frac{n}{n + K'}.$$

Où :

- n la taille du contrat (nombre d'employés, volume de primes, etc.) ;
- K' une constante déterminée de façon à limiter les fluctuations dans la prime d'une année à l'autre.

D'autres formules populaires pour déterminer le facteur de crédibilité :

$$z = \min \left\{ \sqrt{\frac{n}{n_0}}, 1 \right\}.$$

$$z = \min \left\{ \left(\frac{n}{n_0} \right)^{\frac{2}{3}}, 1 \right\}.$$

Où :

- n_0 le seuil de crédibilité complète ;
- n la taille du contrat (nombre d'employés, volume de primes, etc.).

Cette approche a le mérite de pondérer l'expérience individuelle du contrat par l'expérience collective, mais la distribution des primes est basée uniquement sur la taille des assurés. En conséquence, la tarification n'est pas nécessairement précise ni équitable. Des travaux d'amélioration ont continué à être menés par la suite et ont donné lieu à la crédibilité bayésienne.

Chapitre 10

Crédibilité de précision

10.1 Crédibilité bayésienne

On suppose que chaque contrat a un "profil de risque" différent. On note θ_i le profil du risque du contrat numéro i . C'est une quantité qui n'est pas observable (variable "cachée" ou "latente"). On suppose que $(\theta_1, \dots, \theta_n)$ sont des variables aléatoires indépendantes et identiquement distribuées et on note μ la densité de la loi. On suppose que les sinistres dépendent du profil de risque. On note f_{θ_i} la densité de la loi $X_{i,j}$ conditionnellement au profil de risque θ_i . On suppose que les sinistres $X_{i,1}, \dots, X_{i,n}$ sont conditionnellement indépendants connaissant θ_i .

L'objectif est d'estimer la prime pure pour l'année $n+1$. En termes mathématiques :

$$\mu_{i,n+1} = \mathbb{E}[X_{i,n+1} | X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}].$$

Pour cela, il est nécessaire de déterminer la loi jointe des sinistres afin d'en déduire la loi a posteriori.

D'après la formule des probabilités totales :

$$\mathbb{P}(X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}) = \int_{-\infty}^{+\infty} \mathbb{P}(X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}, \Theta_i = \theta_i) d\theta_i.$$

$$\mathbb{P}(X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}) = \int_{-\infty}^{+\infty} \mathbb{P}(X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n} | \Theta_i = \theta_i) \mathbb{P}(\Theta_i = \theta_i) d\theta_i.$$

Les sinistres sont indépendants connaissant θ_i , d'où

$$\mathbb{P}(X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}) = \int_{-\infty}^{+\infty} \left(\prod_{j=1}^n \mathbb{P}(X_{i,j} = x_{i,j} | \Theta_i = \theta_i) \right) \mathbb{P}(\Theta_i = \theta_i) d\theta_i$$

On a noté :

$$\mathbb{P}(X_{i,j} = x_{i,j} | \Theta_i = \theta_i) = f_{\theta_i}(x_{i,j}) \text{ et } \mathbb{P}(\Theta_i = \theta_i) = \mu(\theta_i).$$

D'où la formule de la loi jointe des sinistres :

$$\mathbb{P}(X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}) = \int_{-\infty}^{+\infty} \left(\prod_{j=1}^n f_{\theta_i}(x_{i,j}) \right) \mu(\theta_i) d\theta_i$$

On peut donc en déduire la loi à priori :

$$\mathbb{P}(X_{i,j} = x_{i,j}) = \int_{-\infty}^{+\infty} f_{\theta_i}(x_{i,j}) \mu(\theta_i) d\theta_i.$$

Ainsi que la prime à priori (prime collective) $\mathbb{E}[X_{i,j}]$.

De la loi jointe des sinistres, on peut également en déduire la loi a posteriori :

$$\mathbb{P}(X_{i,n+1} = x_{i,n+1} | X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}) = \frac{\mathbb{P}(X_{i,1}=x_{i,1}, \dots, X_{i,n}=x_{i,n}, X_{i,n+1}=x_{i,n+1})}{\mathbb{P}(X_{i,1}=x_{i,1}, \dots, X_{i,n}=x_{i,n})}.$$

Finalement, la prime a posteriori $\mu_{i,n+1}$ est égale à :

$$\mathbb{E}[X_{i,n+1} | X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}] = \int_0^{+\infty} x_{i,n+1} \mathbb{P}(X_{i,n+1} = x_{i,n+1} | X_{i,1} = x_{i,1}, \dots, X_{i,n} = x_{i,n}).$$

Le modèle Exponentiel-Gamma et le modèle Binomial-négatif sont les plus connus au sein de la crédibilité bayésienne. Le modèle Exponentiel-Gamma est généralement utilisé lorsqu'on cherche à prédire le montant des sinistres tandis que le modèle Binomial-négatif est généralement utilisé lorsqu'on cherche à prédire le nombre de sinistres.

Le modèle Exponentiel-Gamma :

$$\begin{aligned} \Theta_i &\sim \text{Gamma}(\alpha, \beta) \\ X_{i,j} | \Theta_i &\sim \text{Exp}(\Theta_i) \end{aligned}$$

Le modèle Binomial-négatif :

$$\begin{aligned} \Theta_i &\sim \text{Gamma}(\alpha, \beta) \\ X_{i,j} | \Theta_i &\sim \text{Poisson}(\Theta_i) \end{aligned}$$

Ces deux modèles ont la particularité d'aboutir sur des lois à priori et a posteriori classiques. Il est donc aisé d'en calculer l'espérance. Dans ces deux cas, la prime bayésienne a posteriori peut se mettre sous la forme suivante :

$$\mu_{i,n+1} = (1 - \lambda)\mathbb{E}[X_{i,j}] + \lambda\bar{x}_i$$

Où :

- \bar{x}_i l'expérience de sinistralité du contrat ;
- λ le facteur de crédibilité dépendant des paramètres α et β .

Cependant, ce n'est pas toujours le cas. La prime bayésienne n'est pas toujours une prime de crédibilité. De ce fait, le calcul peut donc être particulièrement difficile à mener. De plus, la prime repose sur des hypothèses paramétriques sur la loi des sinistres (pour les distributions de θ_i et de $X_{i,j}$). Ces raisons font que la crédibilité bayésienne ne sera pas appliquée au portefeuille.

Les mathématiciens cherchent à pallier les défauts de la crédibilité bayésienne. En 1965, à un congrès à Lucerne, Bühlmann propose un nouveau modèle de crédibilité. Ce modèle deviendra le modèle de base de la théorie de la crédibilité.

10.2 Crédibilité linéaire

10.2.1 Le modèle de Bühlmann

Hans Bühlmann décide de restreindre l'approximation de la prime de risque aux seules fonctions linéaires des observations. On considère que les primes s'écrivent sous la forme :

$$\pi_{i,n+1} = c_0 + \sum_{j=1}^n c_j S_{i,j}$$

Où

- $P_{i,n+1}$ la prime de l'assuré i l'année $n+1$;
- $S_{i,j}$ la sinistralité de l'assuré i pour l'année j ;
- c_0, \dots, c_j des éléments de \mathbb{R} .

L'avantage du modèle de Bühlmann est qu'il dispose de moins d'hypothèses contraignantes que le modèle Bayésien bien que le modèle de portefeuille soit similaire à celui utilisé en crédibilité bayésienne. De plus, il privilégie la modélisation non paramétrique.

On considère un portefeuille de I contrats. Comme la crédibilité bayésienne, chaque contrat est caractérisé par un profil de risque θ_i (non observable), réalisation d'une variable aléatoire Θ_i et par des observations de sinistralité $(s_{i,1}, \dots, s_{i,n})$, réalisation d'une variable aléatoire $S_i = (S_{i,1}, \dots, S_{i,n})$. Les hypothèses sont les suivantes :

- Les contrats (Θ_i, S_i) , $i = 1, \dots, I$ sont indépendants ;
- Les variables aléatoires $\Theta_1, \dots, \Theta_I$ sont identiquement distribuées ;
- Les variables aléatoires $S_{i,j}$ ont une variance finie ;
- Les variables aléatoires $S_{i,j}$ vérifient les conditions suivantes :

$$\begin{aligned} \mathbb{E}[S_{i,j}|\Theta_i] &= \mu(\Theta_i) \quad i = 1, \dots, I \\ \text{Cov}(S_{i,t}, S_{i,u}|\Theta_i) &= \delta_{t,u} \sigma^2(\Theta_i) \quad t, u = 1, \dots, n. \end{aligned}$$

En résumé, le modèle demande une indépendance intercontrats, une homogénéité temporelle ($\mu(\Theta_i)$ constante dans le temps) et une indépendance intracontrats (observations conditionnellement non corrélées).

On introduit les quantités suivantes :

La prime collective m comme étant la moyenne de toutes les primes de risques pour chacun des contrats :

$$m = \mathbb{E}[\mu(\Theta)]$$

La variance "inter" (between) du portefeuille mesurant son hétérogénéité notée a :

$$a = \text{Var}(\mathbb{E}[S_{i,j}|\Theta_i]) = \text{Var}(\mu(\Theta_i))$$

La variance "intra" (within) du portefeuille, notée s^2

$$s^2 = \mathbb{E}[\text{Var}(S_{i,j}|\Theta_i)] = \mathbb{E}[\sigma^2(\Theta_i)]$$

Théorème 1 (Bühlmann-1967) *Pour un portefeuille tel qu'illustré précédemment et sous les hypothèses citées, la meilleure approximation linéaire non homogène de la prime de risque $\mu(\Theta_i)$ est :*

$$\pi_{i,n+1} = z\bar{S}_i + (1-z)m$$

$$\bar{S}_i = \frac{1}{n} \sum_{j=1}^n S_{i,j}$$

$$z = \frac{n}{n+K}, \quad K = \frac{s^2}{a}$$

Quelques remarques :

- La prime de crédibilité est sans biais : $\mathbb{E}[\pi_{i,n+1}] = m$. Cela signifie qu'en moyenne, l'assureur perçoit suffisamment de primes pour payer les sinistres. Cela permet également que les primes reçues par l'assureur ne soient pas impactées négativement par une mauvaise estimation du facteur de crédibilité.
- $\lim_{n \rightarrow +\infty} \pi_{i,n+1} = \mu(\Theta_i)$
- La prime de crédibilité peut également se mettre sous la forme $\pi_{i,n+1} = m + z(\bar{S}_i - m)$.
- Le facteur de crédibilité z est d'autant plus important dans les cas suivants :
 - Le nombre d'années d'expérience est élevé. En effet, à long terme, l'expérience d'un contrat représente exactement son niveau de risque.
 - Le paramètre s^2 est petit. Cela signifie que l'expérience est stable dans le temps. Les moyennes \bar{S}_i représentent bien les niveaux de risque des contrats, réduisant l'utilité de la prime collective.
 - Le paramètre a est grand. Dans ce cas-là, les moyennes individuelles diffèrent beaucoup d'un contrat à un autre. Les moyennes individuelles sont donc de meilleures approximations des primes de risque que la prime collective.

Ensuite, il reste à estimer les paramètres de structure du portefeuille :

- $m = \mathbb{E}[\mu(\Theta)]$, moyenne du portefeuille ;
- $s^2 = \mathbb{E}[\sigma^2(\Theta)]$, variabilité moyenne du portefeuille, homogénéité temporelle.
- $a = \text{Var}[\mu(\Theta)]$, variance entre les moyennes des contrats, homogénéité du portefeuille.

L'objectif est de faire en sorte que les estimateurs des paramètres soient sans biais.

L'estimateur sans biais de m est assez intuitif :

$$\hat{m} = \bar{S} = \frac{1}{In} \sum_{i=1}^I \sum_{j=1}^n S_{i,j}$$

Concernant l'estimateur s^2 , l'estimateur sans biais est tout simplement la moyenne de l'estimateur sans biais de la variance de chaque contrat i :

$$s^2 = \frac{1}{I(n-1)} \sum_{i=1}^I \sum_{j=1}^n (S_{i,j} - \bar{S}_i)^2$$

En ce qui concerne l'estimateur a , un estimateur intuitif aurait été le suivant :

$$\frac{1}{I-1} \sum_{i=1}^I (\bar{S}_i - \hat{m})^2$$

. Cependant, cet estimateur est biaisé. L'estimateur sans biais de a est le suivant :

$$\hat{a} = \frac{1}{I-1} \sum_{i=1}^I (\bar{S}_i - \hat{m})^2 - \frac{1}{n} \hat{s}^2$$

Le modèle de Bühlmann présente le défaut de ne pas tenir compte de l'exposition au risque des contrats. Cela peut notamment poser problème dans des situations où la taille des contrats varie beaucoup. Par exemple, l'exposition au risque d'une flotte de 1000 véhicules est beaucoup plus grande que celle d'une flotte avec 10 véhicules. De ce fait, si ce modèle était appliqué sur le portefeuille flotte de GMF, un client avec une petite flotte (moins de 10 véhicules) risquerait de payer très cher du fait que le portefeuille est également composé de très grosses flottes (plus de 1000 véhicules), tirant donc la prime collective vers le haut. C'est pourquoi ce modèle ne sera pas appliqué.

10.2.2 Le modèle de Bühlmann-Straub

En 1970, Hans Bühlmann et Erwin Straub proposent une généralisation du modèle de Bühlmann en tenant compte, cette fois, de l'exposition au risque des contrats. De nos jours, le modèle de Bühlmann-Straub est considéré comme étant le modèle de crédibilité le plus utilisé en pratique que ce soit en assurance non Vie, en assurance Vie ou encore en Réassurance.

L'idée de cette modélisation est de faire en sorte que l'expérience d'un gros contrat soit plus stable dans le temps que celle d'un petit contrat. Dans le modèle de Bühlmann-Straub, un poids est associé à chaque observation. Les contrats ne sont donc plus considérés comme égaux comme c'était le cas avec le modèle de Bühlmann.

On considère un portefeuille de I contrats. Chaque contrat est caractérisé par un profil de risque Θ_i (non observable). De plus, à chaque observation $X_{i,j}$ est affecté un poids $w_{i,j}$. Il est essentiel de noter que dans ce modèle, les observations $X_{i,j}$ doivent impérativement être des ratios. Généralement, $X_{i,j}$ est défini de la manière suivante :

$$X_{i,j} = \frac{S_{i,j}}{w_{i,j}}$$

Plusieurs ratios sont possibles, mais les deux suivants sont les plus fréquents :

- Loss ratio : $S_{i,j}$ est le montant total des sinistres et $w_{i,j}$ est la prime totale payée ;
- Fréquence de sinistralité : $S_{i,j}$ est le nombre d'accidents dans une flotte de véhicules $w_{i,j}$ est le nombre de véhicules ;

Voici les hypothèses du modèle de Bühlmann-Straub :

- (i) Les contrats (Θ_i, X_i) , $i = 1, \dots, I$ sont indépendants, les variables aléatoires $\Theta_1, \dots, \Theta_I$ sont identiquement distribuées et les variables aléatoires $X_{i,j}$ ont une variance finie.
- (ii) Les variables aléatoires $X_{i,j}$ sont telles que

$$\mathbb{E}[X_{i,j} | \Theta_i] = \mu(\Theta_i) \quad i = 1, \dots, I$$

$$Cov(X_{i,t}, X_{i,u} | \Theta_i) = \begin{cases} \frac{\sigma^2(\Theta_i)}{w_{i,t}} & \text{si } t = u \\ 0 & \text{sinon.} \end{cases}$$

$$\text{D'où } Var[X_{i,j} | \Theta_i] = \frac{\sigma^2(\Theta_i)}{w_{i,j}}$$

D'après la condition (ii), la quantité $Var[X_{i,j} | \Theta_i]$ est inversement proportionnelle au poids $w_{i,j}$. En conséquence, plus le poids associé à un contrat est élevé, moins le ratio est variable. Les poids sont donc un critère de stabilité.

On introduit les notations suivantes :

- $w_{i,\Sigma} = \sum_{j=1}^n w_{i,j}$ Le poids total du contrat i
- $w_{\Sigma,\Sigma} = \sum_{i=1}^I w_{i,\Sigma}$ Le poids total du portefeuille
- $z_{\Sigma} = \sum_{i=1}^I z_i$ La crédibilité totale du portefeuille
- $X_{i,w} = \sum_{j=1}^n \frac{w_{i,j}}{w_{i,\Sigma}} X_{i,j}$ Le ratio moyen (pondéré) du contrat i
- $X_{w,w} = \sum_{i=1}^I \frac{w_{i,\Sigma}}{w_{\Sigma,\Sigma}} X_{i,w}$ Le ratio moyen (pondéré) du portefeuille
- $X_{z,w} = \sum_{i=1}^I \frac{z_i}{z_{\Sigma}} X_{i,w}$

L'objectif est d'obtenir la meilleure approximation linéaire de la prime de risque d'un contrat.

Théorème 2 (Bühlmann Straub-1970) *Pour un portefeuille tel qu'illustré précédemment et sous les hypothèses (i) et (ii), la meilleure approximation linéaire non homogène de la prime de risque $\mu(\Theta_i)$ ou de $X_{i,n+1}$ est :*

$$\pi_{i,n+1}^{BS} = z_i X_{i,w} + (1 - z_i) \mathbb{E}[\mu(\Theta_i)]$$

Où

- z_i le facteur de crédibilité du contrat i ;
- $\mathbb{E}[\mu(\Theta_i)]$ la prime collective ;
- $z_i = \frac{w_{i,\Sigma}}{w_{i,\Sigma} + K}$, $K = \frac{s^2}{a}$. Les paramètres a et s^2 ont les mêmes significations que dans le modèle initial de Bühlmann (sans les poids). Le paramètre a symbolise l'hétérogénéité du portefeuille, c'est-à-dire la variance entre les moyennes de sinistralité des contrats. En ce qui concerne le paramètre s^2 , il caractérise la variabilité moyenne du portefeuille, c'est-à-dire la moyenne des variances de la sinistralité des contrats.

Les remarques énoncées pour le modèle initial de Bühlmann sont également valables pour le modèle de Bühlmann-Straub.

10.2.2.1 Estimation des paramètres

Tout comme pour le modèle de Bühlmann, les paramètres de structure du modèle (m , s^2 et a) doivent être estimés.

Pour la prime collective $m = \mathbb{E}[\mu(\Theta)]$, l'estimateur sans biais intuitif est le suivant :

$$\hat{m} = X_{w,w} = \sum_{i=1}^I \frac{w_{i,\Sigma}}{w_{\Sigma,\Sigma}} X_{i,w} = \sum_{i=1}^I \sum_{j=1}^n \frac{w_{i,j}}{w_{\Sigma,\Sigma}} X_{i,j}$$

Or, il est possible de démontrer que l'estimateur linéaire à variance minimale est le suivant :

$$\tilde{m} = X_{z,w} = \sum_{i=1}^I \frac{z_i}{z_{\Sigma}} X_{i,w} \quad (10.1)$$

Cependant, cet estimateur n'en est pas réellement un, car il dépend de z_i qui dépend des paramètres inconnus s^2 et a . Ces types d'estimateurs sont appelés "pseudo-estimateurs".

L'estimateur sans biais du paramètre s^2 (variance intra) du modèle de Bühlmann-Straub est tout simplement la généralisation de l'estimateur sans biais obtenu dans le modèle de Bühlmann.

$$\hat{s}^2 = \frac{1}{I(n-1)} \sum_{i=1}^I \sum_{j=1}^n w_{i,j} (X_{i,j} - X_{i,w})^2 \quad (10.2)$$

Un estimateur sans biais du paramètre a (variance inter) est :

$$\hat{a} = \frac{w_{\Sigma, \Sigma}}{w_{\Sigma, \Sigma}^2 - \sum_{i=1}^I w_{i, \Sigma}^2} \left(\overbrace{\sum_{i=1}^I w_{i, \Sigma} (X_{i, w} - X_{w, w})^2}^{\text{estimateur intuitif mais biaisé}} - (I-1) \hat{s}^2 \right) \quad (10.3)$$

Le problème de cet estimateur est qu'il peut prendre des valeurs négatives. Afin de résoudre ce problème, il est possible d'utiliser l'estimateur suivant :

$$\hat{a}' = \max(\hat{a}, 0) \quad (10.4)$$

Toutefois, cet estimateur est biaisé. En réponse à ce problème, en 1981, Vylder a défini un pseudo-estimateur (appelé estimateur de Bichel-Straub) lorsque les facteurs de crédibilité sont connus. Il est sans biais et toujours positif.

$$\tilde{a} = \frac{1}{I-1} \sum_{i=1}^I z_i (X_{i, w} - X_{z, w})^2 \quad (10.5)$$

À noter que dans le cadre de l'étude sur les flottes, les clients possèdent des périodes d'assurance différentes. Les contrats ne prennent pas effet au même moment et n'ont pas tous la même durée. Les clients ne sont donc pas tous nécessairement assurés sur les n années de l'étude. Il en résulte un nombre d'observations différent entre les contrats.

Les données et les poids sont donc disponibles pour $i = 1, \dots, I$ et $j = 1, \dots, n_i$. La formule indiquant le poids total du contrat i et celle indiquant le ratio moyen (pondéré) de chaque contrat sont donc légèrement différentes :

- $w_{i, \Sigma} = \sum_{j=1}^{n_i} w_{i, j}$
- $X_{i, w} = \sum_{j=1}^{n_i} \frac{w_{i, j}}{w_{i, \Sigma}} X_{i, j}$

Le seul estimateur de paramètre affecté par ce changement est \hat{s}^2

$$\hat{s}^2 = \frac{1}{I(n_i - 1)} \sum_{i=1}^I \sum_{j=1}^{n_i} w_{i, j} (X_{i, j} - X_{i, w})^2 \quad (10.6)$$

10.2.2.2 Ordre des calculs

Les calculs sont réalisés dans un certain ordre bien précis (dû à la présence des pseudos estimateurs).

1. Calculs des $w_{i,\Sigma}$, $i = 1, \dots, I$ et $w_{\Sigma,\Sigma}$.
2. Calcul des $X_{i,w}$, $i = 1, \dots, I$ et $X_{w,w}$.
3. Calcul de \hat{s}^2 .
4. Calcul de \hat{a} .
5. Calcul de $\hat{z}_i = \frac{w_{i,\Sigma}}{w_{i,\Sigma} + \frac{\hat{s}^2}{a}}$.
 - (a) Calcul de $\tilde{m} = X_{z,w} = \sum_{i=1}^I \frac{\hat{z}_i}{\hat{z}_\Sigma} X_{i,w}$.
 - (b) Calcul de $\tilde{a} = \frac{1}{I-1} \sum_{i=1}^I \hat{z}_i (X_{i,w} - X_{z,w})^2$.
 - (c) Calcul de $\tilde{z}_i = \frac{w_{i,\Sigma}}{w_{i,\Sigma} + \frac{\tilde{s}^2}{a}}$.
6. Calcul de la prime de crédibilité $\widehat{\pi}_{i,n+1}^{BS} = \tilde{z}_i X_{i,w} + (1 - \tilde{z}_i) \tilde{m}$

Dans le cadre de l'étude, il ne sera pas question de déterminer la prime de crédibilité, mais plutôt le ratio sinistre à prime crédibilisé 2020 et la fréquence de sinistralité de la garantie responsabilité civile crédibilisée 2020. Ensuite, en fonction de la valeur de ces deux indicateurs pour chaque contrat du portefeuille en cours à fin juin 2020, des majorations personnalisées seront appliquées.

10.2.3 La crédibilité hiérarchique (Jewell)

En 1975, Jewell s'inspire du modèle de Bühlmann-Straub pour développer son modèle de crédibilité hiérarchique. Dans ce modèle, une classe est attribuée à chaque contrat. L'idée est, dans un premier temps, de crédibiliser les classes, puis, dans un second temps, de crédibiliser les contrats à l'intérieur des classes. Ceci est d'autant plus conseillé que le portefeuille est hétérogène.

Les observations sont notées $X_{k,i,j}$ où l'indice $k = 1, \dots, K$ identifie la classe (ou cohorte), l'indice $i = 1, \dots, I_k$ identifie le contrat et l'indice $j = 1, \dots, n_{k,i}$ identifie l'année d'observation. Tout comme le modèle de Bühlmann-Straub, les observations sont des ratios (sinistres à primes, fréquence de sinistralité) associés à des poids $w_{k,i,j}$. À chaque classe et à chaque contrat sont associés des niveaux de risque (non mesurables). La variable aléatoire Φ_k représente le niveau de risque de la classe k tandis que la variable aléatoire $\Theta_{k,i}$ représente le niveau de risque du contrat i présent dans la classe k . Les hypothèses du modèle sont les suivantes :

- (i) Les variables aléatoires Φ_1, \dots, Φ_K sont indépendantes et identiquement distribuées.
- (ii) Les variables aléatoires $\Theta_{k,1}, \dots, \Theta_{k,I_k}$ sont conditionnellement indépendantes et identiquement distribuées sachant Θ_k , $k = 1, \dots, K$.
- (iii) Les variables aléatoires $X_{k,i,1}, \dots, X_{k,i,n_{k,i}}$ sont conditionnellement indépendantes et identiquement distribuées sachant $\Theta_{k,i}$ (et Φ_k), $i = 1, \dots, I_k$. Leur variance est finie.
- (iv) Pour tout $t, u = 1, \dots, n_{k,i}$

$$E[X_{k,i,j} | \Theta_{k,i}, \Phi_k] = \mu(\Theta_{k,i}, \Phi_k)$$

$$Cov(X_{k,i,t}, X_{k,i,u} | \Theta_{k,i}, \Phi_k) = \begin{cases} \frac{\sigma^2(\Theta_{k,i}, \Phi_k)}{w_{k,i,t}} & \text{si } t = u \\ 0 & \text{sinon.} \end{cases}$$

$$D'où \text{Var}[X_{k,i,j} | \Theta_{k,i}, \Phi_k] = \frac{\sigma^2(\Theta_{k,i}, \Phi_k)}{w_{k,i,j}}$$

On définit les paramètres de structure suivants :

- $m = \mathbb{E}[\mathbb{E}[\mu(\Theta_{k,i}, \Phi_k)] | \Phi_k]$ La moyenne collective (moyenne des moyennes des contrats).
- $s^2 = \mathbb{E}[\mathbb{E}[\sigma^2(\Theta_{k,i}, \Phi_k)] | \Phi_k]$ La variance intra-assuré moyenne (moyenne des variances des contrats).
- $a = \mathbb{E}[Var[\mu(\Theta_{k,i}, \Phi_k)] | \Phi_k]$ La variance inter-assuré ou la variance intra-classe moyenne (moyenne des variances des classes).
- $b = Var[\mathbb{E}[\mu(\Theta_{k,i}, \Phi_k)] | \Phi_k]$ La variance inter-classe (variance des moyennes des classes).

L'objectif est d'estimer les primes de risque des contrats $\mu(\Theta_{k,i}, \Phi_k)$. La prime de crédibilité du modèle hiérarchique s'écrit sous la même forme que celle du modèle de Bühlmann-Straub à la différence que le complément de crédibilité $(1 - z_i)$ n'est plus attribué à la prime collective du portefeuille, mais à la prime collective de la classe dans laquelle se trouve le contrat. Cette prime de crédibilité de la classe est une moyenne pondérée de l'expérience de la classe et de celle du portefeuille total.

Théorème 3 () *Pour un portefeuille tel qu'illustré précédemment et sous les hypothèses (i), (ii), (iii) et (iv), la meilleure approximation linéaire non homogène de la prime de risque $\mu(\Theta_{k,i}, \Phi_k)$ ou de $X_{k,i,n+1}$ est donnée par les équations récursives suivantes :*

$$\begin{aligned} \pi_{k,i,n+1}^J &= z_{k,i} X_{k,i,w} + (1 - z_{k,i}) \pi_k^J \\ \pi_k^J &= z_k X_{k,z,w} + (1 - z_k) m \end{aligned} \quad (10.7)$$

Où

- $w_{k,i,\Sigma} = \sum_{j=1}^{n_{k,i}} w_{k,i,j}$
- $w_{k,\Sigma,\Sigma} = \sum_{i=1}^{I_k} w_{k,i,\Sigma}$
- $z_{k,i} = \frac{w_{k,i,\Sigma}}{w_{k,i,\Sigma} + \frac{s^2}{a}}$ le facteur de crédibilité du contrat i de la classe k.
- $w_{k,i,\Sigma} = \sum_{j=1}^{n_{k,i}} w_{k,i,j}$ le poids total du contrat i de la classe k.
- $z_k = \frac{z_{k,\Sigma}}{z_{k,\Sigma} + \frac{b}{\theta}}$ le facteur de crédibilité de la classe de contrats k.
- $z_{\Sigma} = \sum_{k=1}^K z_k$
- $z_{k,\Sigma} = \sum_{i=1}^{I_k} z_{k,i}$
- $z_{\Sigma,\Sigma} = \sum_{k=1}^K z_{k,\Sigma}$
- $X_{k,i,w} = \sum_{j=1}^{n_{k,i}} \frac{w_{k,i,j}}{w_{k,i,\Sigma}} X_{k,i,j}$ La moyenne pondérée individuelle
- $X_{k,z,w} = \sum_{i=1}^{I_k} \frac{z_{k,i}}{z_{k,\Sigma}} X_{k,i,w}$ La moyenne pondérée de la classe

4 paramètres de structure sont à estimer pour modèle hiérarchique à 2 niveaux. Le pseudo-estimateur de la moyenne collective est

$$\hat{m} = X_{z,z,w} = \sum_{k=1}^K \frac{z_k}{z_{\Sigma}} X_{k,z,w} \quad (10.8)$$

L'estimateur sans biais de la dispersion intra-assuré est le suivant :

$$\hat{s}^2 = \frac{1}{\sum_{k=1}^K \sum_{i=1}^{I_k} (n_{k,i} - 1)} \sum_{k=1}^K \sum_{i=1}^{I_k} \sum_{j=1}^{n_{k,i}} w_{k,i,j} (X_{k,i,j} - X_{k,i,w})^2 \quad (10.9)$$

L'estimateur sans biais de Ohlsson pour la variance intra-classe est :

$$\hat{a} = \frac{\sum_{k=1}^K A_k}{\sum_{k=1}^K c_k} \quad (10.10)$$

Avec

$$A_k = \sum_{i=1}^{I_k} w_{k,i,\Sigma} (X_{k,i,w} - X_{k,w,w})^2 - (I_k - 1)s^2$$

$$c_k = w_{k,\Sigma,\Sigma} - \sum_{i=1}^{I_k} \frac{w_{k,i,\Sigma}^2}{w_{k,\Sigma,\Sigma}}$$

L'estimateur sans biais de Ohlsson pour la variance inter-classe est :

$$\hat{b} = \frac{B}{d} \quad (10.11)$$

Avec

$$B = \sum_{k=1}^K z_{k,\Sigma} (X_{k,z,w} - \bar{X}_{z,z,w})^2 - (K - 1)a$$

$$\bar{X}_{z,z,w} = \sum_{k=1}^K \frac{z_{k,\Sigma}}{z_{\Sigma,\Sigma}} X_{k,z,w}$$

$$d = z_{\Sigma,\Sigma} - \sum_{k=1}^K \frac{z_{i,\Sigma}^2}{z_{\Sigma,\Sigma}}$$

De même qu'avec le modèle de Bühlmann-Straub, il est possible de calculer des pseudos-estimateurs plus précis pour les paramètres a et b.

Le pseudo-estimateur sans biais itératif de a est le suivant :

$$\tilde{a} = \frac{1}{\sum_{k=1}^K (I_k - 1)} \sum_{k=1}^K \sum_{i=1}^{I_k} z_{k,i} (X_{k,i,w} - X_{k,z,w})^2 \quad (10.12)$$

Le pseudo-estimateur sans biais itératif de b est le suivant :

$$\tilde{b} = \frac{1}{K - 1} \sum_{k=1}^K z_k (X_{k,z,w} - X_{z,z,w})^2 \quad (10.13)$$

Cinquième partie

La mise en place de la surveillance du portefeuille en appliquant la théorie de la crédibilité avancée

L'objectif de cette partie est de mettre en place une surveillance basée sur un modèle de crédibilité (Bühlmann-Straub ou Jewell). Les modèles de Bühlmann-Straub et de Jewell (crédibilité hiérarchique) sont appliqués au portefeuille flottes GMF à fin juin 2020. Un modèle simple de surveillance est également mis en place pour servir de comparaison. L'idée est d'analyser si un modèle complexe de crédibilité apporte des résultats sensiblement différents d'un modèle très basique.

L'étude prend place à fin juin 2020 correspondant à la période des mesures tarifaires flottes GMF pour l'année 2021. Les clients considérés sont les clients présents en portefeuille à fin juin 2020 pour lesquels il reste au moins une échéance à payer en 2021 (date d'échéance finale supérieure ou égale au 31 décembre 2021). Cela représente 182 clients. Pour rappel, la quasi-totalité des contrats flottes disposent d'une échéance annuelle au premier janvier. Toutes les durées de préavis sont considérées. Généralement, elles s'échelonnent entre 2 mois et 6 mois. Les contrats avec un préavis de 2 mois sont moins prioritaires, mais il est décidé de tout traiter en même temps. L'évolution de l'indice SRA entre le deuxième trimestre 2020 et le deuxième trimestre 2019 est évaluée à +4%. Tous les clients flottes vont se

voir proposer une mesure tarifaire d'au moins +4% pour leur échéance 2021. Cette hausse tarifaire sera d'autant plus élevée que les résultats techniques du client seront dégradés. Certains clients disposent de plusieurs contrats. Dans la base de données, la sinistralité est vue par client tous contrats confondus. À la suite de l'analyse des résultats du client, une majoration est définie puis elle est répartie équitablement sur les différents contrats du client.

Il s'avère qu'aucun des clients considérés n'était présent en portefeuille avant 2013. Certains ont intégré le portefeuille en 2013 tandis que d'autres sont assurés seulement depuis 2018, c'est très disparate. Une variable renseigne le nombre d'années d'expérience du client au sein du portefeuille flottes GMF. Les clients ayant intégré le portefeuille au début de l'année 2020 se voient affecter une expérience égale à 0. À fin juin 2020, l'exercice 2020 n'est pas terminé, les données de sinistralité sont donc difficilement exploitables d'autant plus que le confinement lié à la pandémie de Covid-19 biaise significativement les résultats techniques. Les primes des clients, dont l'expérience est nulle, sont majorées uniquement par l'indice SRA (+4%), représentant l'inflation des réparations automobiles. 54 clients sont concernés pour un total de 3,2 millions d'euros de primes émises en 2020. Les mesures tarifaires sur ce segment représentent donc 127 806 euros (4% de 3,2 millions d'euros).

Années d'expérience	Nombre de clients	Part des clients
0	54	29,7%
1	38	20,1%
2	26	14,2%
3	13	7,1%
4	11	6%
5	5	2,7%
6	21	11,5%
7	14	7,7%

TABLE 10.1 – Répartition de l'ancienneté des clients

71% des clients ont moins de 3 ans d'ancienneté, le portefeuille est relativement jeune. Toutefois, on note que 19% des clients ont plus de 6 ans d'ancienneté. Dans la suite, on ne tient plus compte des nouveaux clients que l'on traitera à part.

10.3 Le modèle de surveillance historique

Historiquement, la GMF avait pour habitude de revaloriser les contrats à hauteur de l'évolution de l'indice SRA entre le deuxième trimestre de l'année N et le deuxième trimestre de l'année N-1. Si on applique cette revalorisation à notre portefeuille (nouveaux clients comme anciens clients) on obtient une revalorisation totale de l'ordre de 428 606 euros.

10.4 Le modèle de surveillance de base

Ce modèle initial est basé sur des indicateurs de SP et de SP moyen. Le premier travail est de repérer les contrats particulièrement dégradés parmi les 128 clients (182-54 nouveaux clients de 2020). Pour cela les SP moyens (écrêtés à 150 000 euros) 2013-2019 et 2017-2019 sont calculés.

$$\text{SP écrêté Moyen}_{a,b} = \frac{\sum_{j=a}^b \text{Sinistralité}_j}{\sum_{j=a}^b \text{Prime acquise}_j}$$

Où

- La sinistralité est vue à fin juin 2020 et les sinistres sont écrêtés à 150 000 euros.
- a-b la période de temps (par exemple 2013-2019).
- La prime acquise est hors taxe et hors assistance.

Le SP moyen de la période 2013-2019 permet de se donner une idée de la sinistralité historique du client tandis que le SP de la période 2017-2019 donne un aperçu récent et a pour objectif de déceler d'éventuelles dégradations au cours des dernières années. Il est décidé que les contrats dont au moins l'un des deux SP moyens est supérieur à 80% se verront attribuer une majoration supérieure à +4%. 90 contrats sur 128 (soit 70%) sont concernés. Les primes des 38 bons contrats sont majorées de la hausse de base +4%. Ces contrats représentent 1,7 million d'euros de primes émises en 2020. Les mesures tarifaires de ce segment sont donc évaluées à 68 000 euros (4% de 1,7 million d'euros). Il reste à définir les mesures tarifaires des contrats considérés comme dégradés.

En ce qui concerne les contrats détériorés, l'idée est de les regrouper selon différentes classes de résultats techniques allant d'une dégradation légère à une dégradation sévère. Ces classes sont définies selon les valeurs des SP moyens écrêtés, les valeurs de SP annuel écrêtés et le nombre d'années d'expérience du client. À chaque classe est associée une majoration spécifique.

Dégradation	Condition	Majoration
1	SP moyen 2013-2019 et 2017-2019 \leq 100% ou moins de 2 ans d'expérience	+6%
2	SP moyen 2013-2019 ou 2017-2019 dégradé mais à cause d'une ou deux années	+8%
3	SP moyen 2013-2019 et 2017-2019 $>$ 100 et au moins 3 années avec un SP $>$ 100 %	+15%
4	SP moyen 2013-2019 et 2017-2019 $>$ 120 et au moins 3 années avec un SP $>$ 120	+30%

TABLE 10.2 – Classes de dégradation

L'objectif est de faire en sorte que les classes de dégradation 3 ou 4 soient exceptionnelles. Pour rappel, GMF est une société d'assurance mutuelle, elle est donc considérée à but non lucratif. Voilà pourquoi, il paraît un peu délicat de majorer très sévèrement un client. Les majorations sont à comparer avec la majoration de base (+4%). Par exemple, une augmentation de 6% représente +2 points par rapport à la mesure de base.

Dégradation	Nombre de clients	Primes émises 2020	Majoration	Mesures tarifaires
1	60	3,94M €	+6%	196 987 €
2	19	1,31M €	+8%	104 486 €
3	7	0,31M €	+15%	46 867 €
4	4	0,26M €	+30%	76 951 €

TABLE 10.3 – Répartition des contrats dégradés selon les niveaux de dégradation

À noter que pour déterminer la mesure tarifaire 2021, la majoration est bien réalisée sur les primes émises 2020 et non sur les primes émises 2019. En effet, bien que seuls les résultats techniques de l'année 2019 soient analysés, les primes émises en 2020 ne sont pas égales aux primes émises en 2019 dû à la majoration selon l'évolution de l'indice SRA en 2019.

Toutes classes de dégradation confondues, les mesures tarifaires 2021 des clients dégradés représentent 425 293 euros. Lorsqu'on y agrège les mesures tarifaires appliquées aux nouveaux clients ayant intégré le portefeuille en 2020 et aux bons clients déjà présents en portefeuille, on obtient 621 099 euros.

Ce modèle a le mérite d'être très simple et facilement compréhensible. Les résultats serviront de base de comparaison avec les modèles de crédibilité détaillés par la suite.

Chapitre 11

L'application du modèle de Bühlmann-Straub

11.1 Hétérogénéité du portefeuille

Comme on a pu le voir dans la partie 2, le portefeuille est très hétérogène en termes de primes et de sinistralité. On y retrouve aussi bien des contrats assurant une flotte d'une dizaine de véhicules que des contrats assurant des flottes de centaines de véhicules. Pour s'en assurer, il suffit de tester l'hétérogénéité du portefeuille en réalisant un test de comparaison des moyennes de sinistralité (test d'analyse de la variance à un facteur).

On suppose que les montants de sinistralité de chaque contrat i sont des réalisations d'une variable aléatoire $X_{i,j}$.

$$X_{i,j} \sim \mathcal{N}(\mu_i, \sigma^2)$$

Avec :

- j l'année d'observation.
- μ_i l'espérance du contrat $i \in \{1, \dots, n\}$.
- $X_{i,j}, X_{i',j'}$ indépendantes pour $i \neq i'$ ou $j \neq j'$.

On peut alors tester l'hypothèse suivante (test de Fisher par exemple) :

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_n$$

$$H_1 : \exists(i, j) \text{ tel que } m_i \neq m_j$$

L'application du test de Fisher sur R donne une pvalue égale à $4,6 \cdot 10^{-4}$, très inférieure à $\alpha = 5\%$, l'hypothèse H_0 est refusée. Le portefeuille est bel et bien hétérogène.

11.2 Ratio considéré : le loss-ratio

Le modèle de Bühlmann-Straub est appliqué au portefeuille de 128 contrats (ne tenant pas compte des nouveaux contrats présents seulement depuis 2020). Pour rappel, le modèle de Bühlmann-Straub exige que les observations soient des ratios. Des poids (associé à chaque observation) doivent donc être définis. Dans un premier temps, les ratios considérés sont les ratios sinistres à primes entre 2013 et 2019 et les poids associés sont les primes acquises entre 2013 et 2019. L'objectif n'est donc pas de prédire une prime, mais de projeter le ratio sinistre à prime de l'année suivante en tenant compte de la sinistralité du client et de celle du portefeuille.

11.2.1 Estimation des paramètres de structure

L'application du modèle de Bühlmann-Straub se fait avec la fonction `cm` du package actuar de R. Ce package, très complet au sujet de la théorie de la crédibilité, a été développé en 2005 par Vincent GOULET. On spécifie à la fonction `cm` qu'on souhaite utiliser la méthode itérative permettant d'estimer les pseudos-estimateurs. Les paramètres de structure du portefeuille obtenus sont les suivants :

Prime collective \tilde{m}	Variance intra portefeuille \hat{s}^2	Variance inter portefeuille \tilde{a}	$\hat{K} = \frac{\hat{s}^2}{\tilde{a}}$
1,02	10 124	0.052	194 692

TABLE 11.1 – Estimation des paramètres du modèle de Bühlmann-Straub sur les ratios sinistres à primes

La prime collective représente le SP moyen écrêté du portefeuille entre 2013 et 2019. Cette prime est supérieure à 1, d'où l'importance d'agir et de mettre en place des mesures afin que la tendance soit à la baisse dans les années à venir.

11.2.2 Exemple numérique

Pour bien assimiler la méthode de calcul, prenons l'exemple du client fictif suivant en supposant qu'il est assuré par un contrat flottes GMF depuis début janvier 2018. On souhaite projeter son SP en 2020 avec le modèle de Bühlmann-Straub.

client	Prime acquise HT 2018	Prime acquise HT 2019	Sinistralité 2018	Sinistralité 2019
A	16 00 €	18 000 €	13 000 €	21 000 €

TABLE 11.2 – Calcul du loss ratio projeté d'un client fictif

- $w_{A,2018} = 16000, w_{A,2019} = 18000, w_{A,\Sigma} = 34000$
- $X_{A,2018} = \frac{13000}{16000}, X_{A,2019} = \frac{21000}{18000}$
- $X_{A,w} = \sum_{j=2018}^{2019} \frac{w_{A,j}}{w_{A,\Sigma}} X_{A,j} = 1,000$
- $\hat{z}_A = \frac{w_{A,\Sigma}}{w_{A,\Sigma} + \hat{K}} = 0,148$
- $\pi_{A,2020}^{BS} = z_A X_{A,w} + (1 - z_A) \tilde{m} = 1,019$

11.2.3 Le facteur de crédibilité

Pour rappel le facteur de crédibilité \hat{z}_i est de la forme suivante : $\hat{z}_i = \frac{w_{i,\Sigma}}{w_{i,\Sigma} + \hat{K}}$

On cherche le poids total $w_{i,\Sigma,50}$ (somme des primes du client entre 2013 et 2019) correspondant à un facteur de crédibilité de 50%. Après calcul, il vient :

$$w_{i,\Sigma,50} = \hat{K}$$

En résumé, un facteur de crédibilité de 50% est appliqué à un client si et seulement si le poids total de ses primes est supérieur à 194 692 euros. De manière générale, on peut représenter le facteur de crédibilité selon le poids total du client :

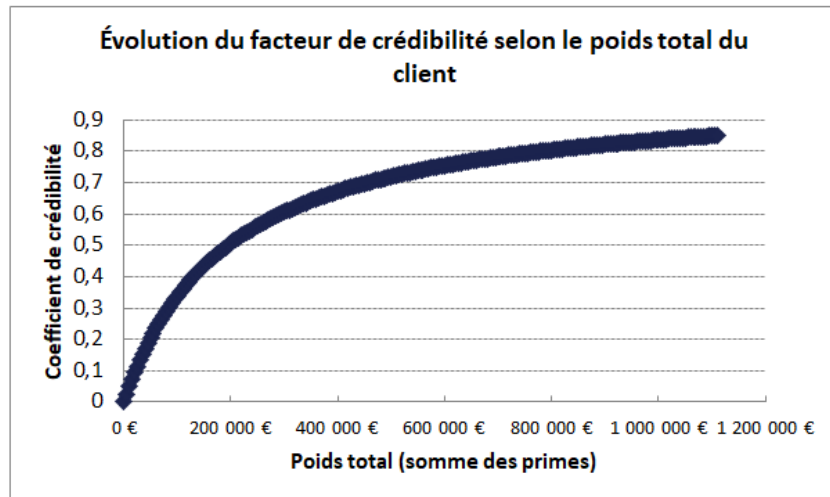


FIGURE 11.1 – Évolution du nombre de clients et de contrats flottes GMF

Plus un client a un poids important, plus on peut estimer sa sinistralité crédible. Ensuite, on observe la répartition des clients selon leur facteur de crédibilité.

Facteur de crédibilité	Nombre de clients	Part des primes acquises 2019
0-0,25	55	11%
0,25-0,5	43	37%
0,5-0,75	19	22%
0,75-1	11	31%

TABLE 11.3 – Répartition des facteurs de crédibilité associés au loss-ratio

La répartition est assez logique. Comme vu dans la partie 2, le portefeuille est constitué majoritairement de petits contrats, associés à des petites primes. En conséquence, les poids totaux de ces contrats sont faibles, expliquant ainsi le nombre élevé de clients disposant d'un facteur de crédibilité inférieur à 0,25.

11.2.4 La mise en place d'une stratégie de surveillance

Une fois le loss ratio prédit pour l'année 2020 avec le modèle de Bühlmann-Straub, il reste à mettre en place la stratégie de surveillance. Pour cela, on considère qu'un contrat est dégradé au-delà d'un SP projeté supérieur à 80%. Ce seuil étant considéré comme le seuil d'équilibre. Au-delà, le contrat est déficitaire pour GMF car la prime reçue ne permet pas de payer la totalité des charges (sinistres, commissions courtier, frais généraux, frais de réassurance). Les règles suivantes sont établies :

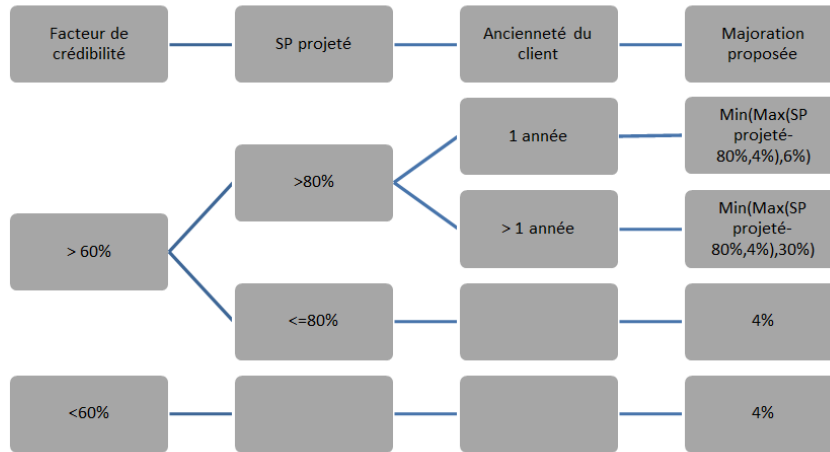


FIGURE 11.2 – Règles de surveillance définies sous forme d’arbre

La partie 2 avait mis en avant que les flottes conséquentes présentaient des résultats particulièrement dégradés. L’objectif est donc de cibler ces clients et de les majorer en conséquence. Le facteur de crédibilité étant d’autant plus important que le poids de la prime est grand, cet indicateur sera utilisé pour repérer ces contrats. On estime qu’un facteur de crédibilité de 60% est suffisant. Cela correspond à un poids total d’environ 300k euros de primes sur la période d’assurance du client. Les clients pour lesquels la sinistralité n’est pas jugée suffisamment crédible sont majorés de la hausse de base, c’est-à-dire 4%. Ces contrats sont généralement des petites primes et n’impactent que très peu les résultats techniques du portefeuille. L’enjeu est donc moindre. D’autant plus que, du fait de leur petit poids, les petits contrats se voient attribuer un coefficient de crédibilité proche de 0 et donc le SP projeté est sensiblement similaire à la prime collective, ie le loss-ratio moyen du portefeuille (102%). Cela peut sembler un peu aberrant si le contrat présente de bons résultats comme c’est le cas pour ce genre de contrats généralement. Ensuite, si le client est jugé crédible, son ancienneté et son loss-ratio projeté avec le modèle de Bühlmann-Straub permettent de définir la majoration de son contrat. Si le loss-ratio projeté est inférieur à 80%, on estime que le client est un bon risque. En conséquence, la majoration est la plus faible possible, c’est-à-dire 4%. Pour rappel, le SP brut d’équilibre est défini à 80% selon un retour d’expérience des équipes Flottes. C’est pourquoi si le client présente un loss-ratio projeté supérieur à 80%, des mesures supplémentaires sont prises. Dans ce cas, la hausse est corrélée avec la différence entre le loss ratio projeté et le loss ratio d’équilibre (défini à 80%). Si le client est récent dans le portefeuille, la majoration ne peut excéder 6%. Si ce n’est pas le cas, la hausse peut être bien plus importante (majorée à 30%).

11.2.4.1 Application numérique sur un exemple fictif

Un exemple fictif permettra de mieux assimiler les règles de revalorisation tarifaire. Prenons l’exemple d’un client A, présent en portefeuille depuis 3 ans, émettant une prime annuelle 2020 de 95 000 € et pour lequel on estime un facteur de crédibilité de 65% et un loss ratio crédibilisé de 95%. On calcule la différence entre son loss ratio projeté et le supposé ratio d’équilibre : $95\% - 80\% = 15\%$. On calcule le maximum entre cette valeur et la revalorisation de base du portefeuille (4%). On obtient 15%. Ensuite, on détermine le minimum entre cette valeur et 30% afin de ne pas revaloriser un contrat de plus de 30%. Cela donne 15%. On applique ces 15% à la prime émise en 2020 par le client 1, $15 * 95000 = 109250$ pour déterminer la proposition de revalorisation à faire au client en respectant le délai du préavis. S’il refuse. Le client a le droit de refuser et de résilier son contrat. S’il accepte, la prime émise 2021 sera de 109 250 €, soit une augmentation de 14 250 €.

11.2.4.2 Résultats de la revalorisation tarifaire

En appliquant cette stratégie tarifaire, la revalorisation totale des contrats représente 752 070 €. Pour rappel, les loss-ratio des clients ayant rejoint le portefeuille en 2020 ne sont pas projetés avec le modèle de Bühlmann-Straub car on ne dispose pas d’au moins une année de sinistralité complète. Ces contrats sont revalorisés à hauteur de 4%. En les agrégeant aux contrats déjà présents dans le portefeuille en 2019, on obtient une revalorisation totale de 879 876 €.

11.3 Ratio considéré : la fréquence de sinistralité RC

On projette également la fréquence de sinistralité de la garantie responsabilité civile avec le modèle de Bühlmann-Straub. En effet, dans la partie 2, on a observé que les sinistres graves étaient essentiellement des sinistres liés à la responsabilité civile. Bien qu'une provision pour sinistres graves soit incluse dans la cotisation des contrats, on souhaite majorer plus fermement les contrats en portefeuille qui, en plus d'un loss-ratio projeté supérieur à 80%, présente une fréquence de sinistralité RC sensiblement dégradée, laissant penser que la probabilité de survenance d'un sinistre grave est bien plus importante. Dans ce modèle, les ratios considérés sont les fréquences de sinistralité RC entre 2013 et 2019 et les poids associés sont les parcs de véhicules entre 2013 et 2019.

11.3.1 Estimation des paramètres de structure

Prime collective \tilde{m}	Variance intra portefeuille \hat{s}^2	Variance inter portefeuille \tilde{a}	$\hat{K} = \frac{\hat{s}^2}{\tilde{a}}$
0,0678	0,0903	0,0004	225,75

TABLE 11.4 – Estimation des paramètres du modèles de Bühlmann-Straub sur les fréquences de sinistralité RC

En faisant le même raisonnement que pour le loss-ratio, on en déduit qu'un poids total (somme des parcs assurés) de 223 véhicules est nécessaire pour accorder au moins 50% de crédibilité à l'expérience du client. La prime collective, c'est-à-dire la fréquence de sinistralité RC moyenne du portefeuille entre 2013 et 2019 est de l'ordre de 6,7%. La variance inter est relativement petite signifiant que les individus ont une fréquence de sinistralité RC très semblable. De même, la variance intra n'est pas très élevée, signifiant qu'elle ne varie pas beaucoup d'une année à l'autre pour un client considéré. Cela signifie qu'on peut donc plus facilement faire confiance à l'expérience d'un client. Pour la projection des loss-ratio, on avait vu qu'une crédibilité de 50% était attribuée à un client disposant d'un poids total de primes de 194 692 €. Si on fait l'approximation qu'un véhicule représente en moyenne une prime de 350 €, on obtient 550 véhicules. Cela signifie qu'un client doit disposer d'un poids total d'au moins 550 véhicules sur sa période d'assurance pour être affecté d'une crédibilité supérieure à 50% pour son loss-ratio historique. Or, on observe que pour la fréquence de sinistralité, un poids total de 226 véhicules est suffisant pour qu'un client soit crédité de 50% de crédibilité pour sa fréquence de sinistralité RC historique. À noter qu'un client assurant un parc de 25 véhicules entre 2016 et 2019 aura le même poids ($25 \times 4 = 100$), et donc la même crédibilité, qu'un client assurant une flotte de 100 véhicules depuis 2019 ($100 \times 1 = 100$).

11.3.2 Le facteur de crédibilité

Facteur de crédibilité	Nombre de clients	Part des primes acquises 2019
0-0,25	22	1,3%
0,25-0,50	27	9%
0,50-0,75	32	21%
0,75-1,00	47	68%

TABLE 11.5 – Répartition des facteurs de crédibilité associés à la fréquence de sinistralité

Au vu de la répartition, très différente de celle des SP, on en déduit que la fréquence de sinistralité RC est beaucoup plus stable que le SP. On a plus facilement tendance à faire confiance à la fréquence de sinistralité historique du client même si sa flotte est moyenne.

11.3.3 Suite de la mise en place de la stratégie de surveillance

La fréquence de sinistralité projetée de la garantie RC est intégrée à l'arbre de décision des mesures tarifaire. Il est décidé de sanctionner les clients présentant des fréquences de sinistralité supérieures à 10% en leur appliquant une surmajoration de 10%. Ce seuil de 10% correspond au centile 90% des fréquences RC observées sur le portefeuille.

À noter qu'on ne teste pas le facteur de crédibilité accordé à l'historique de la fréquence de la flotte car si celui accordé au loss-ratio est déjà supérieur à 60% alors celui de la fréquence sera généralement supérieur à 80%. Ce n'est donc pas nécessaire.

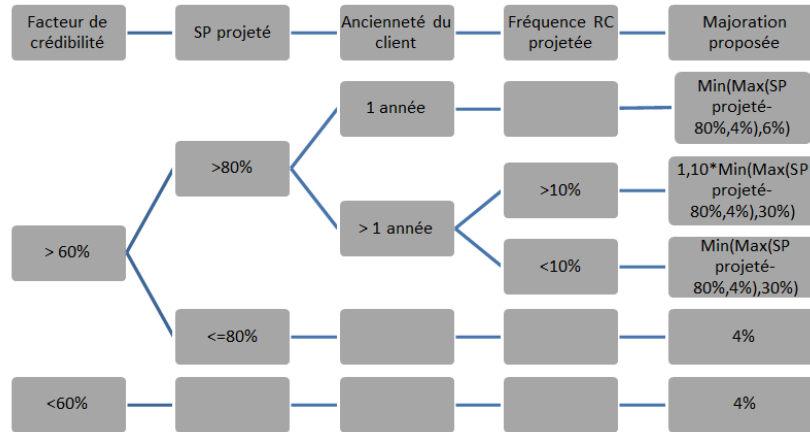


FIGURE 11.3 – Mis à jour des règles de surveillance en y intégrant la fréquence RC projetée

11.3.3.1 Application numérique sur un exemple fictif

Reprenons l'exemple fictif précédent et supposons maintenant que la fréquence de sinistralité RC projetée du client est de 11,5%. Cette valeur est supérieure au seuil 10%. En conséquence, une majoration de 10% s'applique à la majoration initiale $1,10 \times 10,15\% = 16,5\%$. La revalorisation de ce contrat est donc dorénavant de 15 675 € soit une prime émise 2021 de 110 675 €.

11.3.3.2 Résultats de la revalorisation tarifaire

L'intégration de la fréquence RC projetée à la stratégie tarifaire ajoute 21 671 € de revalorisation supplémentaire, soit un total de 901 547 €.

11.4 Tests de sensibilité du modèle de Bühlmann-Straub

	SP et Freq RC projetés avec Bühlmann-Straub				Revalorisation totale	Ecart scénario central
	H1 : facteur de crédibilité	H2 : SP	H3 : fréquence RC			
Scénario central	60%	80%	10,0%		901 547 €	0%
Test de Sensibilité H1 : +5% H1	63%	80%	10,0%		895 010 €	-0,7%
Test de Sensibilité H1 : -5% H1	57%	80%	10,0%		911 413 €	1,1%
Test de Sensibilité H1 : +5% H2	60%	84%	10,0%		832 335 €	-7,7%
Test de Sensibilité H1 : -5% H2	60%	76%	10,0%		969 827 €	7,6%
Test de Sensibilité H1 : +5% H3	60%	80%	10,5%		889 702 €	-1,3%
Test de Sensibilité H1 : -5% H3	60%	80%	9,5%		908 783 €	0,8%

FIGURE 11.4 – Tests de sensibilité du modèle de Bühlmann-Straub

Pour rappel, le modèle de revalorisation mis en place à partir de la crédibilité de Bühlmann-Straub repose sur 3 seuils (scénario central) : un seuil de crédibilité (fixé à 60%), un seuil de SP projeté (fixé à 80%) et un seuil de fréquence RC projetée (fixé à 10%). L'objectif des tests de sensibilité est de faire varier chaque hypothèse (de plus ou moins 5%) indépendamment des autres pour s'assurer de la relative stabilité des résultats. Pour cela, on compare la nouvelle revalorisation tarifaire avec celle obtenue par le scénario central. Si on varie le facteur de crédibilité de plus ou moins 5%, cela implique une évolution de revalorisation comprise approximativement entre -1% et 1%. Même constat lorsqu'on joue sur le seuil de la fréquence RC. Au contraire, faire varier de plus ou moins 5% le seuil de SP projeté à partir duquel on décide de majorer plus sévèrement (ie le SP d'équilibre) impacte fortement le résultat (7% en valeur absolue). Ce constat était prévisible car la majoration projetée est une différence entre le SP projeté et le SP d'équilibre, donc si on diminue le SP d'équilibre de 5%, on augmente naturellement la majoration (et réciproquement si on augmente le SP d'équilibre de 5%). Prenons l'exemple d'un client avec un facteur de crédibilité de 70% et un SP projeté de 88%. Le facteur de crédibilité est supérieur à 60% donc il subit une majoration égale à la différence entre son SP projeté et le SP d'équilibre. Si le SP d'équilibre est de 80%, la majoration est de l'ordre de 8%. Cependant, si on baisse de 5% le SP d'équilibre (76% au lieu de 80%), la majoration n'est plus de 8% mais de 12%. L'hypothèse 2 est donc particulièrement sensible.

Chapitre 12

L'application de la crédibilité hiérarchique

La théorie de la crédibilité hiérarchique repose sur des classes de risques. Il est donc nécessaire, au préalable, de réaliser des regroupements homogènes de clients selon leur niveau de risque. Les variables considérées pour le regroupement sont les suivantes :

- Les variables décrivant le client (région, activité professionnelle de la flotte, ancienneté).
- Les variables liées au risque de la flotte propres à 2019 (nombre de véhicules, primes acquises, répartition du parc selon les différentes catégories de véhicules, niveaux de franchises dommages, conservation ou non des sinistres).
- Les variables de sinistralité propres à 2019 (Sinistralité écrêtée, SP écrêté, montant des provisions, nombre de sinistres RC, fréquence RC).

Les 11 clients possédant un contrat avec conservation des sinistres sont d'office regroupés au sein d'une classe. En effet, ces contrats sont très atypiques et très différents, de par leur fonctionnement, des contrats flottes classiques. Ils ne sont donc pas pris en compte dans l'ACP et dans la CAH. Cela fait donc au total 22 variables et 117 individus. L'idée est de réaliser dans un premier temps une analyse en composantes principales afin de réduire le nombre de dimensions et de mieux visualiser les données. Ensuite, dans un second temps, il sera question, à partir des résultats de l'ACP, de réaliser une classification ascendante hiérarchique pour déterminer les classes. Finalement, la crédibilité hiérarchique pourra être mise en place.

12.1 L'analyse en composantes principales

12.1.1 La théorie

L'analyse en composantes principales (ACP) est une méthode factorielle, consistant à transformer des variables corrélées en nouvelles variables décorrélées les unes des autres. Cela permet notamment de réduire le nombre de variables d'un jeu de données afin de mieux l'analyser et de mieux le visualiser. En ce qui concerne l'analyse, son objectif est double. En effet, l'ACP permet d'obtenir des informations aussi bien sur les variables (potentiels liens entre elles) que sur les individus (ressemblances, dissemblance). Les nouvelles variables transformées, non corrélées, sont appelées composantes principales (ou axes principaux), elles synthétisent l'information du jeu de données initial. Ces nouvelles variables correspondent à une combinaison linéaire des variables d'origine.

Soit un jeu de données avec n variables ($n > 3$). Les individus sont donc représentés dans un espace à n dimensions rendant la visualisation des données particulièrement difficile. En pratique, l'ACP projette les individus dans un espace de dimension restreint (un ou plusieurs plans de projection). Bien qu'au cours de la projection, des informations soient perdues, parmi toutes les projections possibles, certaines permettent de perdre moins d'informations que d'autres.

Soit un échantillon de K réalisations de N variables aléatoires X_1, \dots, X_N , représenté sous la forme d'une matrice M disposant de K lignes et N colonnes. Pour chaque variable aléatoire X_n (de moyenne \bar{X}_n et d'écart-type σ_{X_n}), les réalisations $X_{1,n}, \dots, X_{K,n}$ sont indépendantes.

$$M = \begin{pmatrix} X_{1,1} & \cdots & X_{1,N} \\ \vdots & \ddots & \vdots \\ X_{K,1} & \cdots & X_{K,N} \end{pmatrix}$$

La notion de distance est utilisée pour analyser le nuage. Généralement, la distance euclidienne classique est choisie pour calculer la distance entre un individu i et k .

$$d_{i,k}^2 = \sum_{j=1}^p (x_{i,j} - x_{k,j})^2 \quad (12.1)$$

Des poids différents p_i peuvent être associés aux observations. Si les poids sont uniformes alors chaque réalisation a le même poids $\frac{1}{K}$

La matrice M est généralement transformée en centrant le nuage de points sur le centre de gravité. Elle peut également être réduite.

$$\widetilde{M} = \begin{pmatrix} \frac{X_{1,1} - \bar{X}_1}{\sigma(X_1)} & \cdots & \frac{X_{1,N} - \bar{X}_N}{\sigma(X_N)} \\ \vdots & \ddots & \vdots \\ \frac{X_{K,1} - \bar{X}_1}{\sigma(X_1)} & \cdots & \frac{X_{K,N} - \bar{X}_N}{\sigma(X_N)} \end{pmatrix}$$

Si les poids sont uniformes et que M est réduite, il suffit de multiplier M par sa transposée pour obtenir la matrice de corrélation. Cette matrice est tout simplement la matrice des coefficients de corrélation statistiques calculés sur plusieurs variables prises deux à deux. Généralement, les coefficients de Pearson sont utilisés.

Le coefficient de Pearson entre deux variables aléatoires réelles X et Y ayant chacun une variance finie est défini par :

$$r = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad (12.2)$$

où

- $Cov(X, Y)$ désigne la covariance des variables X et Y ;
- σ_X et σ_Y désignent leurs écarts types (égaux à 1 comme la matrice est réduite).

Le terme générique de la matrice de corrélation est donné par :

$$r_{i,j} = Cor(X_i, X_j) = Cov(X_i, X_j) \quad (12.3)$$

Cette matrice est carrée (taille N), symétrique et réelle. D'après le théorème spectral, elle est donc diagonalisable dans une base orthonormée.

L'inertie totale (appelée également variance) représente l'information contenue dans un jeu de données. L'ACP a pour but d'identifier les directions (axes principaux) le long desquelles la variation des données est maximale. Pour déterminer la première composante principale, l'ACP recherche la combinaison linéaire des variables d'origine de variance maximale. Pour cela, l'idée est de déterminer le vecteur u tel que la projection du nuage (la matrice \widetilde{M} centrée réduite) sur u ait une variance maximale. La projection de \widetilde{M} sur u s'écrit :

$$\pi_u(\widetilde{M}) = \widetilde{M}u \quad (12.4)$$

La recherche du deuxième axe w fonctionne sur le même principe en imposant qu'il soit orthogonal à u et ainsi de suite.

Enfin, la recherche des composantes principales revient à calculer les valeurs propres et les vecteurs propres de la matrice de corrélation. La variance expliquée par un vecteur propre est égale à sa valeur propre. Le vecteur expliquant le plus d'inertie du nuage est le premier vecteur propre, celui associé à la valeur propre la plus élevée. Finalement, d'un ensemble de variables plus ou moins corrélées est né un nouvel ensemble de variables (axes principaux) non corrélées, d'importance décroissante.

12.1.2 L'application

Du fait du faible volume de données, on anticipe que les variables catégorielles (région et activité professionnelle de la flotte) doivent être retraitées pour contenir le moins de modalités possible.

Il est donc décidé de regrouper les régions en fonction de leur géographie et du poids qu'elle représente en termes de primes acquises 2019 dans le portefeuille :

- Ouest : Bretagne, Normandie, Pays de la Loire
- Nord : Hauts-de-France
- Ile-de-France : Ile-de-France
- Centre-Est : Centre, Bourgogne-Franche-Comté, Grand-Est
- Sud : Nouvelle Aquitaine, Occitanie, Auvergne-Rhône-Alpes, Provence-Alpes-Côte d'azur

En ce qui concerne les activités professionnelles, seuls les titres Communes, Établissements, Départements, Caisses et Associations sont conservés. Les autres activités sont associées à "Autres".

Dans le cadre de l'ACP, ces variables qualitatives sont des variables supplémentaires. Elles ne servent pas à calculer les distances entre individus, mais elles peuvent être utilisées en complément pour aider à interpréter les axes principaux.

L'analyse en composantes principales est appliquée sur R à l'aide des package FactoMineR et factoextra. À noter que la fonction PCA du package FactoMineR centre et réduit les données automatiquement.

La première étape est de déterminer le nombre de composantes principales à prendre en considération.

Pour cela, on extrait les valeurs propres qui permettent de mesurer la quantité de variance expliquée par chaque axe principal.

Axe principal	Valeur propre	Quantité de variance expliquée	Cumul de variance expliquée
Dim 1	5,37	26,89%	26,89%
Dim 2	2,65	13,23%	40,12%
Dim 3	2,00	10,01%	50,13%
Dim 4	1,62	8,11%	58,24%
Dim 5	1,51	7,54%	65,78%
Dim 6	1,43	7,16%	72,94%
Dim 7	1,20	6,00%	78,94%

TABLE 12.1 – Variance expliquée par les 7 premiers axes

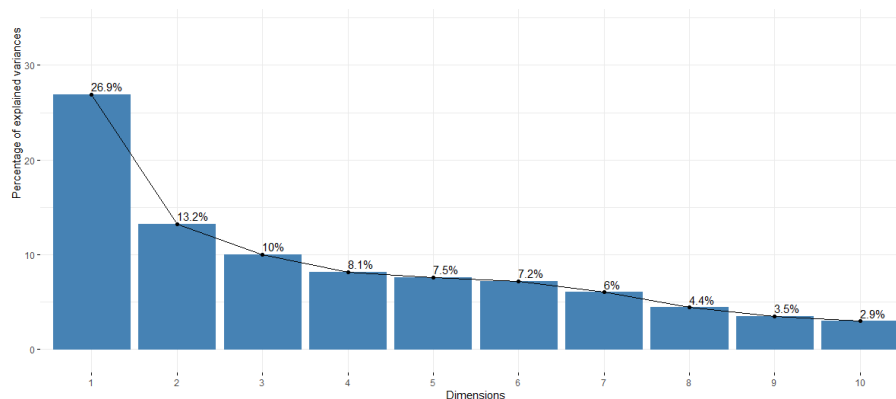


FIGURE 12.1 – Pourcentage de variance expliquée par les axes principaux

Aucune méthode objective n'existe quant au nombre précis d'axes principaux à considérer. Toutefois, le critère de Kaiser (1961) préconise de ne conserver que les axes pour lesquels la valeur propre est supérieure à 1. Cela signifie que la composante principale en question représente plus de variance par rapport à une seule variable d'origine. Le critère de Kaiser est retenu. 7 axes principaux sont considérés. Ces 7 axes principaux expliquent approximativement 80% de la variance totale. 4 axes principaux sont conservés. On constate que les 2 premiers axes principaux représentent 40,12 % de la variance totale. Il peut être intéressant de comparer ce résultat avec le quantile à 95% du pourcentage d'inertie des 2 premières dimensions de 10 000 ACP de 117 individus et de 22 variables indépendantes. Une table indique que sur 10 000 ACP construites avec 100 individus et 20 variables indépendantes, 95% des premiers plans de l'ACP ont un pourcentage d'inertie inférieur à 19,3%. Ce chiffre est donc légèrement plus faible si on considère 17 individus et 2 variables en plus. Si le pourcentage d'inertie sur le premier plan d'une ACP à 117 individus et 22 variables était inférieur à 19%, cela signifierait que le plan d'ACP n'explique pas plus que ce qu'on aurait eu avec des variables indépendantes et que le résumé n'est pas synthétique. Dans notre exemple, le premier plan explique 40% de la variance totale, supérieur à 19%. On en déduit donc que le plan de l'ACP résume bien l'information. D'où l'existence de liaisons fortes entre les variables.

Voici la signification des variables présentes dans les graphes qui vont suivre :

- ANC : ancienneté du client.
- VEH : nombre de véhicules, PRI : primes acquises 2019.
- M3T5 : proportion de véhicules - 3 tonnes 5 non assurés en tous risques dans la flotte. M3T5TR : proportion de véhicules - 3 tonnes 5 assurés en tous risques dans la flotte. P3T5 : proportion de véhicules + 3 tonnes 5 non assurés en tous risques dans la flotte. P3T5TR : proportion de véhicules + 3 tonnes 5 assurés en tous risques dans la flotte. _2R : proportion de véhicules 2 roues dans la flotte. R3T5 : proportion de remorques -3 tonnes 5 dans la flotte. RP3T5 : proportion de remorques +3 tonnes 5 dans la flotte. EVS : proportion d'engins véhicules spéciaux dans la flotte. TPV : proportion de véhicules de transport public dans la flotte.
- FRM3T5 : niveau de franchises dommages des véhicules - 3 tonnes 5 de la flotte. FRP3T5 : niveau de franchises dommages des véhicules + 3 tonnes 5 de la flotte. FR2_3R : niveau de franchises dommages des véhicules 2 (ou 3) roues de la flotte.
- S : sinistralité écrêtée 2019. SP : ratio sinistre à primes écrêté moyen de la flotte sur sa période d'assurance. PRO : montant des provisions 2019. SINRC : nombre de sinistres RC 2019. FR_RC : fréquence RC moyenne de la flotte sur sa période d'assurance.

Il a été décidé de se limiter à l'année 2019 pour l'ACP. Il aurait très bien pu être possible de considérer plus d'années ou encore d'inclure des variables moyennes (sinistralité, primes). Toutefois, le SP moyen et la fréquence RC moyenne de la flotte permettent d'avoir des informations historiques sur les résultats techniques de la flotte.

Ensuite, on cherche à visualiser toutes les variables du jeu de données. Elles sont représentées sur le cercle de corrélation entre les variables et les 2 premières composantes principales. La coordonnée de la variable sur un axe principal est la corrélation entre la variable et la composante principale.

Avant d'analyser le cercle des corrélations, il est important de se poser la question de la qualité de représentation des variables sur le plan principal (constitué des 2 premiers axes principaux). La qualité de représentation d'une variable est mesurée à l'aide du \cos^2 de l'angle entre la variable et sa projection. Un \cos^2 élevé indique une bonne représentation de la variable sur les axes principaux en question. Dans ce cas, la variable se trouve proche de la circonférence du cercle de corrélation. Au contraire, un faible \cos^2 indique une mauvaise représentation de la variable sur les axes principaux. Dans ce cas, la variable se situe proche du centre, elle est moins importante pour les premières composantes. Pour ces variables, 2 axes ne suffisent pas à représenter parfaitement les données. La figure suivante indique la qualité de représentation des variables sur le plan principal :

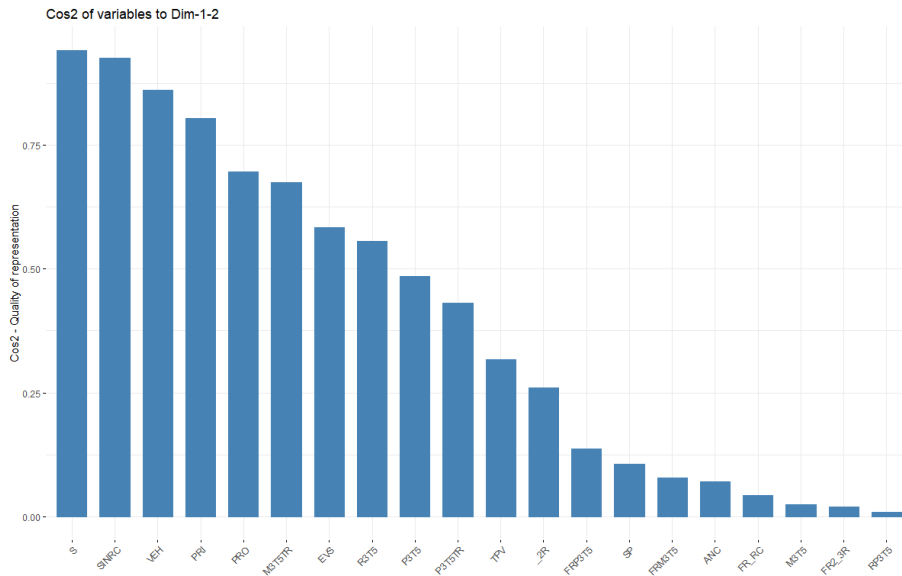


FIGURE 12.2 – Qualité de représentation des variables sur le plan principal

On s’aperçoit que les variables de sinistralité et de répartition du parc sont plutôt bien représentées sur le premier plan. Toutefois, ce n’est pas le cas de la variable ancienneté, des variables propres aux franchises et des variables propres aux résultats techniques (SP et fréquence). Pour ces variables, il ne sera donc pas possible de tirer des conclusions à partir du plan principal. On souhaite maintenant déterminer les variables contribuant le plus à la construction des 2 premiers axes principaux.

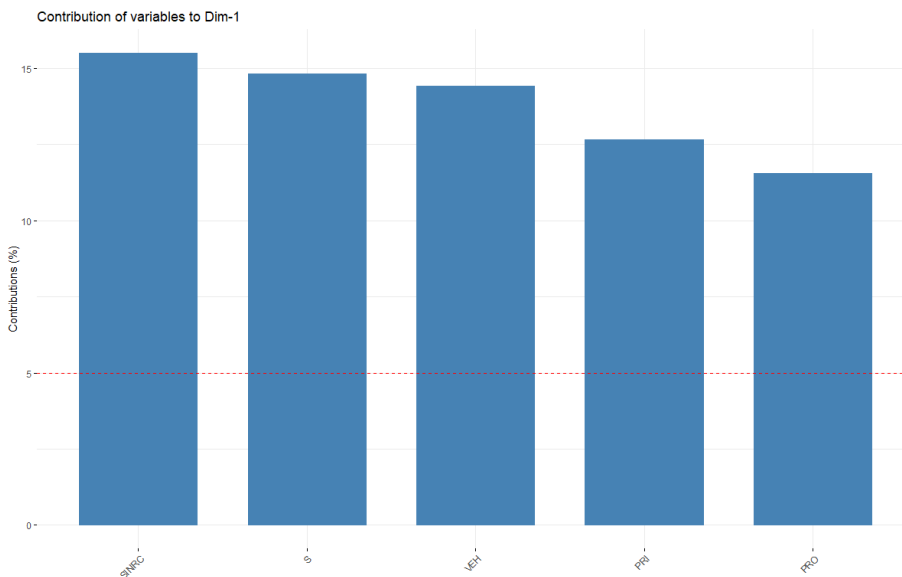


FIGURE 12.3 – Contribution des variables d’origine au premier axe

Les variables contribuant le plus à la constitution du premier axe sont la prime acquise 2019, le nombre de véhicules, le montant écrêté de sinistralité 2019, les provisions 2019 et le nombre de sinistres RC 2019. Ces variables donnent une information sur l’importance de la flotte.

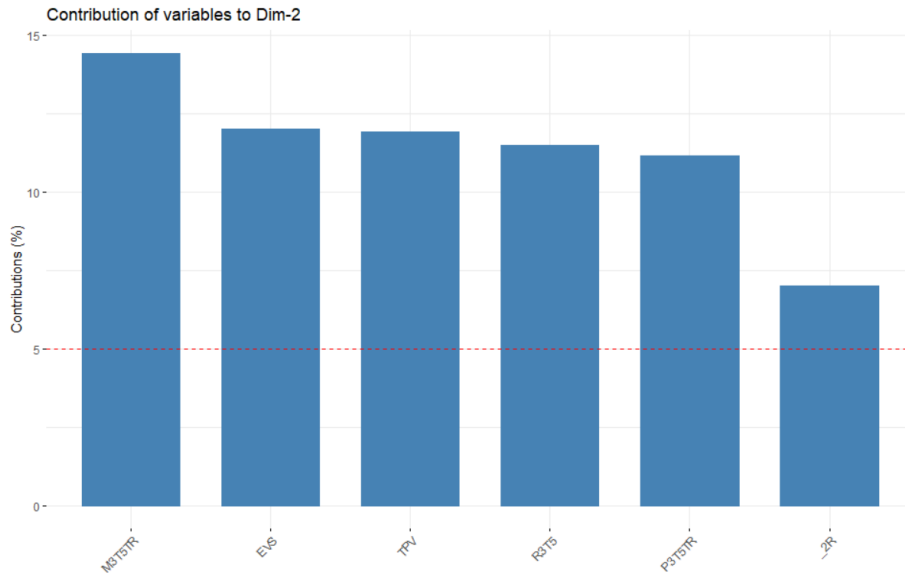


FIGURE 12.4 – Contribution des variables d’origine au deuxième axe

Les variables contribuant le plus à la constitution du deuxième axe sont les variables représentant la proportion au sein de la flotte des véhicules légers assurés en tous risques, engins, remorques, poids lourds assurés en tous risques, transport public et deux roues. Ces variables donnent une information sur les types de véhicules présents au sein de la flotte.

La figure suivante représente le graphique de corrélation entre les variables et les composantes principales selon la valeur de \cos^2 .

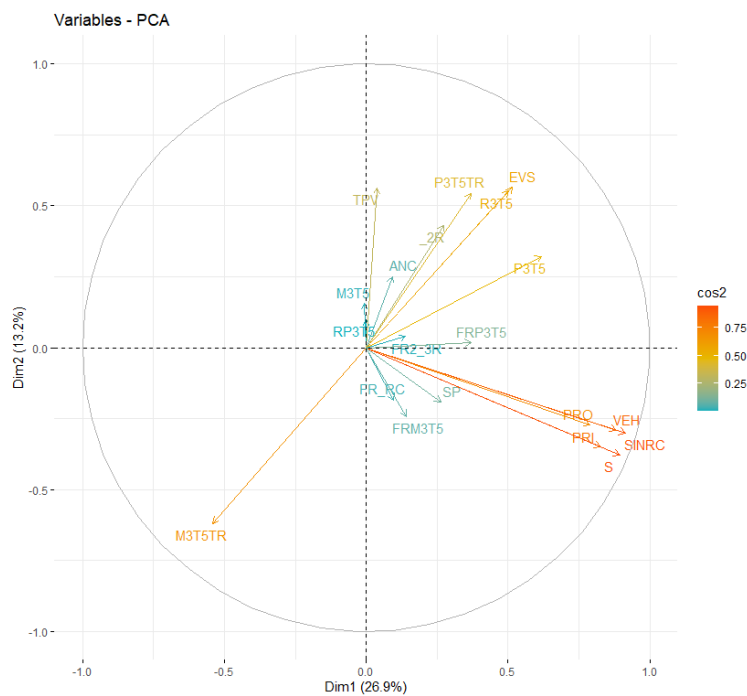


FIGURE 12.5 – Cercle de corrélation

D'après le graphe des corrélations, on constate que l'axe principal, c'est-à-dire la dimension 1, est très corrélé positivement avec la prime acquise 2019, le nombre de véhicules, le montant écrêté de sinistralité 2019, les provisions 2019 et le nombre de sinistres RC 2019. Comme on pouvait s'y attendre, ces variables sont très corrélées entre elles : plus le contrat assure un nombre de véhicules élevé, plus le risque est important (plus de sinistres RC, sinistralité plus élevée, plus de provisions, etc.).

En ce qui concerne le deuxième axe principal, il est corrélé avec les variables de répartition du parc selon la catégorie de véhicules. On y relève une corrélation positive avec les variables représentant la proportion au sein de la flotte des véhicules engins, remorques, poids lourds, transport public et deux roues. À contrario, le deuxième axe principal est corrélé très négativement avec la variable représentant la proportion de véhicules moins 3 tonnes 5 assurés en tous risques.

Principaux facteurs de variabilité :

1. Grosses flottes et petites flottes.
2. À flotte égale : la répartition des types de véhicules au sein de la flotte.

D'après la représentation des variables, les gros contrats seront situés à la droite du graphe, les contrats moyens au centre et les petits contrats à gauche du graphe. En bas du graphe, on retrouvera des contrats dont la flotte est constituée majoritairement de véhicules légers assurés en tous risques. À contrario, en haut du graphe seront situés des contrats constitués d'une part importante de poids lourds, de deux roues et d'engins.

Les individus (identifiés selon leur numéro de client à 6 chiffres) sont projetés sur le plan principal afin de les visualiser. La meilleure représentation graphique des individus est la suivante :

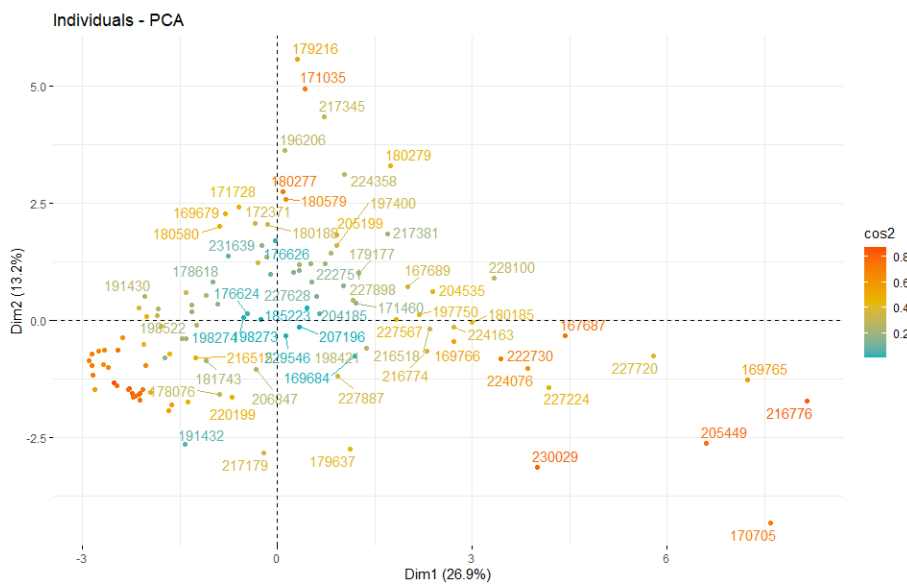


FIGURE 12.6 – Projection des individus sur le plan principal

On attache plus d'importance aux individus pour lesquels la qualité de la projection sur le plan principal (\cos^2) est jugée suffisante. Certains individus ne sont pas bien projetés sur les deux premiers axes (ceux au centre) mais le seraient sur les axes principaux suivants. Ces individus se distinguent selon d'autres critères (comme le niveau de franchise ou les résultats techniques par exemple).

Il est toujours important ensuite de vérifier sur quelques exemples la qualité de la représentation. Par exemple, le client 170705 tout en bas à droite du graphique émet des primes annuelles d'environ 400 000 € pour un parc composé de 88% de véhicules légers dont 40% de véhicules légers assurés en tous risques. À contrario, le client 180279 situé légèrement en haut à droite du graphe émet des primes annuelles plus faibles, de l'ordre de 30 000 € et assure un parc constitué de 30% de poids lourds (dont 6% en tous risques, ce qui est assez conséquent).

Les modalités des variables catégorielles peuvent être également représentées (barycentre des individus prenant la modalité).

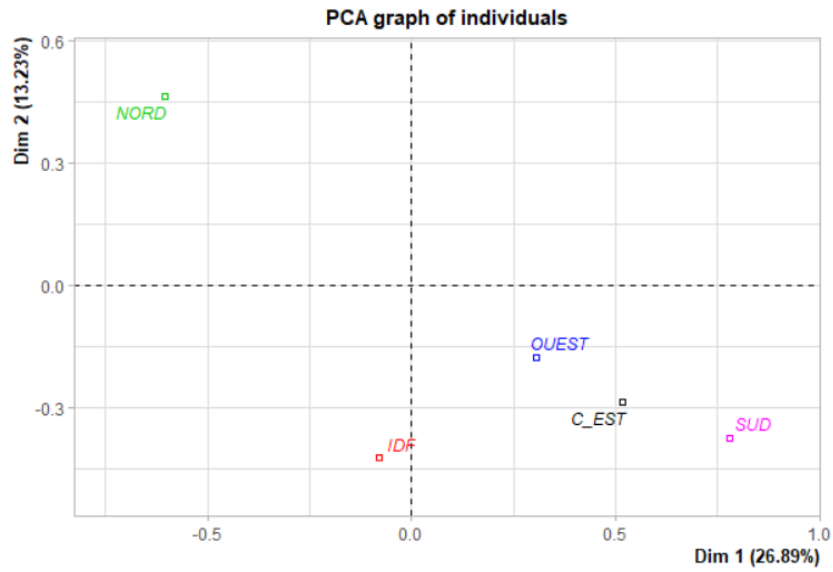


FIGURE 12.7 – Barycentres des modalités de la variable catégorielle région

On observe que les régions ouest, centre-est et sud sont situés dans la partie inférieure droite du graphe, signifiant que les flottes appartenant à ces régions sont généralement conséquentes avec une part importante de véhicules légers assurés en tous risques. À contrario, les flottes enregistrées dans le nord de la France sont localisées dans la partie supérieure gauche du graphe, signifiant que ce sont majoritairement des petites flottes assurant une part non négligeable de véhicules autres que des véhicules légers (poids lourds, engins, remorques, etc.). Les flottes parisiennes semblent être des contrats relativement moyens constitués d’une part très importante de véhicules légers tous risques.

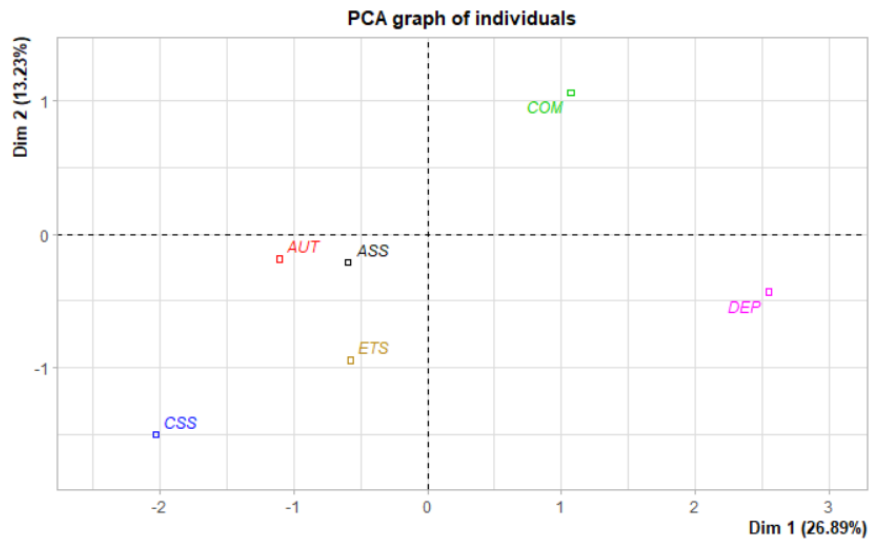


FIGURE 12.8 – Barycentres des modalités de la variable catégorielle activité professionnelle

Concernant les régions, les barycentres semblent assez éloignés les uns des autres. On constate que les flottes assurés par des départements sont des grosses flottes. Les communes assurent également des parcs conséquents avec une proportion importante de poids lourds (+ tonnes 5), d'engins et de remorques. Les flottes liées à des associations semblent être des parcs de moyenne ampleur, composés d'une répartition homogène des différents types de véhicules. Les flottes appartenant aux établissements publics sont également de taille moyenne mais assurant une quantité plus importante de véhicules légers tous risques. Finalement, les flottes appartenant à des caisses publiques sont des petites flottes composées presque essentiellement de véhicules légers.

Les axes principaux 3 et 4 expliquent 18% de la variance totale, il est donc intéressant de les conserver et d'analyser brièvement les variables originelles permettant de les constituer.

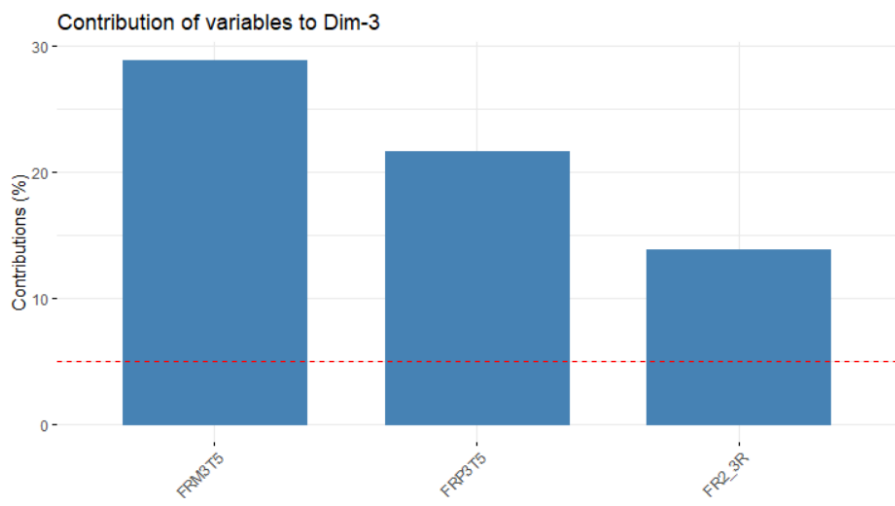


FIGURE 12.9 – Contribution des variables d'origine au troisième axe

Les variables indiquant les franchises dommages selon la catégorie du véhicule (véhicule léger, poids lourds ou 2 roues) contribuent significativement à la création du troisième axe principal.

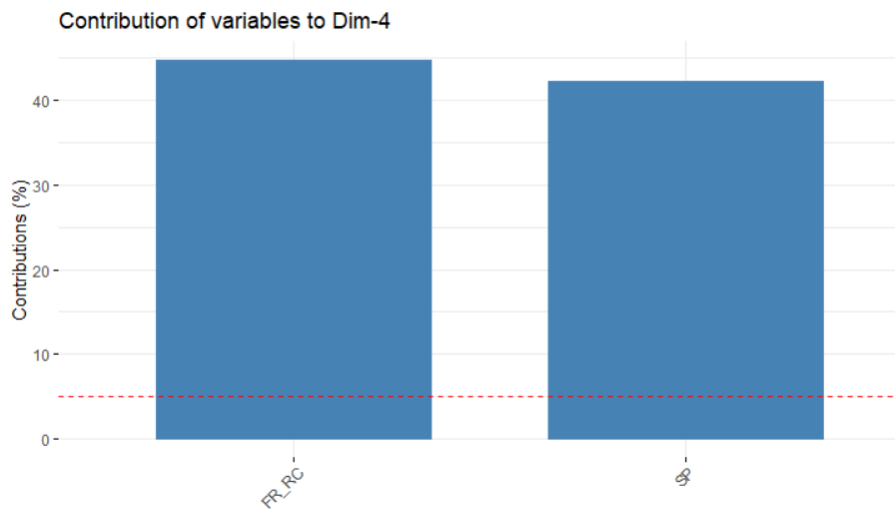


FIGURE 12.10 – Contribution des variables d'origine au quatrième axe

Le quatrième axe principal est expliqué en grande partie par les variables des résultats techniques des flottes (SP écrêté et fréquence de sinistralité RC).

Pour résumer l'ACP, voici les 4 principaux facteurs de variabilité :

1. Grosses flottes et petites flottes.
2. À flotte égale : la répartition des types de véhicules au sein de la flotte.
3. À flotte et répartition égale : les niveaux de franchises dommages selon les types de véhicules.
4. À flotte, répartition et franchise égale : les résultats techniques de la flotte (SP et fréquence de sinistralité RC).

12.2 La Classification ascendante hiérarchique

12.2.1 La théorie

La classification ascendante hiérarchique (CAH) est une des nombreuses techniques statistiques de partitionnement d'une population en différents sous-groupes possédant des traits de caractère communs. En d'autres termes, une partition est bonne si les individus d'une même classe sont proches et si les individus de 2 classes différentes sont éloignés. Mathématiquement, cela revient à maximiser l'hétérogénéité inter-classe (variabilité inter-classe grande) tout en optimisant l'homogénéité intra-classe (variabilité intra-classe petite).

Les individus sont définis par des variables quantitatives. Un critère de ressemblance, exprimé sous forme de distance, est défini pour permettre de rassembler les individus entre eux. La distance entre deux individus est d'autant plus importante que les individus sont différents. Plusieurs définitions de distance sont possibles. La distance la plus naturelle est la distance euclidienne.

la distance euclidienne entre un individu i et k est :

$$d_{i,k}^2 = \sum_{j=1}^p (x_{i,j} - x_{k,j})^2$$

Une fois que la distance entre deux individus est définie, il est nécessaire de définir une distance entre 2 groupes d'individus. La classification est évidemment dépendante de la distance choisie. Par exemple, la distance appelée "saut minimum" correspond au minimum des distances entre les individus de deux groupes. Quant au saut maximum, il correspond au maximum des distances entre les individus de deux groupes. Le critère de Ward est également un critère de ressemblance. La méthode de Ward tend à regrouper des individus ou groupes d'individus tels que leur regroupement minimise la baisse de l'inertie inter-classe.

D'après le théorème de Huygens :

$$\underbrace{\sum_{q=1}^Q \sum_{i=1}^I d^2(x_{i,q} - \bar{x})}_{\text{Inertie totale}} = \underbrace{\sum_{q=1}^Q \sum_{i=1}^I d^2(x_{i,q} - \bar{x}_q)}_{\text{Inertie intra-classe}} + \underbrace{\sum_{q=1}^Q d^2(\bar{x}_q - \bar{x})}_{\text{Inertie inter-classe}}$$

Avec

- \bar{x} le centre de gravité total
- \bar{x}_q le centre de gravité de la classe q

L'inertie totale étant constante, minimiser l'inertie intra revient à maximiser l'inertie inter.

La CAH est ascendante, car les observations individuelles sont le point de départ et elle est hiérarchique de par sa création de classe de plus en plus grande. La méthode est basée sur un processus itératif. Dans un premier temps, les individus les plus proches sont regroupés. Puis, dans un second temps, des sous-ensembles d'individus sont également regroupés pour finalement former des "clusters".

La méthode de Ward :

- Initialisation : 1 classe=1 individu : inertie inter=1
- À chaque étape : agréger les classes a et b minimisant la diminution de l'énergie inter.

Généralement, la CAH est complémentaire de l'ACP. L'ACP permet de réduire les dimensions du jeu de données, facilitant la construction de classes homogènes.

L'objectif final est double : produire une arborescence permettant de mettre en avant les liens hiérarchiques entre individus ou groupes d'individus et mettre en avant un découpage naturel de l'arbre.

12.2.2 L'application

La classification ascendante hiérarchique est appliquée sur R à l'aide de la fonction **HCPC** du package FactoMineR. Cette fonction nécessite que l'objet en argument soit le résultat d'une analyse factorielle. Les classes sont donc construites à partir des résultats de l'ACP obtenus dans la section précédente. Pour rappel, il a été décidé de ne conserver que 7 axes principaux. Le critère de Ward est appliqué pour construire les classes. Pour rappel, l'objectif est de réaliser des regroupements d'individus ou de sous-groupes d'individus en minimisant la perte d'inertie inter-classe (ou en minimisant le gain d'inertie intra-classe). Le choix du nombre de classes est déterminant, car construire une partition avec trop peu de classes risque de créer des classes non homogènes. Au contraire, trop de classes risque de créer des classes peu différenciantes difficiles à analyser. Les classes doivent pouvoir être interprétables. Généralement, lorsqu'on dispose de peu d'individus, il est recommandé de ne pas constituer beaucoup de classes à moins que les individus soient statistiquement très différents.

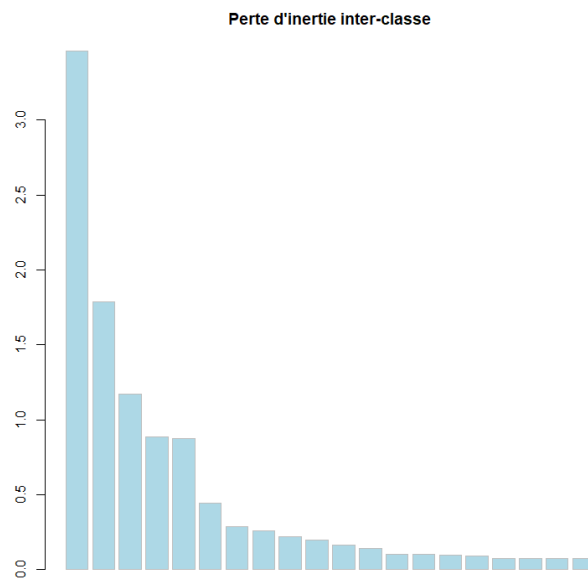


FIGURE 12.11 – Perte d'inertie inter-classe selon le nombre de classes

Le graphique à barres ci-dessus représente l'évolution de l'inertie pour différentes partitions, c'est-à-dire la perte d'inertie inter-classe lorsqu'on passe de N classes à $N - 1$ classes. Par exemple, la première barre représente la perte d'inertie inter lorsqu'on passe de 2 classes à 1 classe. La deuxième, lorsqu'on passe de 3 classes à 2 classes et ainsi de suite. La perte d'inertie inter au passage d'une partition à N classes à une partition à $N - 1$ classes est d'autant moins importante que la N est grand. L'objectif est de déterminer un nombre de classes N , adapté au jeu de données, pour lequel la perte d'inertie inter entre $N + 1$ et N est relativement faible, sous-entendant qu'une partition en $N + 1$ classes ne permettrait pas de mieux différencier les individus qu'une partition en N classes. Au vu du graphique et du faible nombre d'individus présents dans le jeu de données (117), on se satisfait d'une partition en 4 classes.

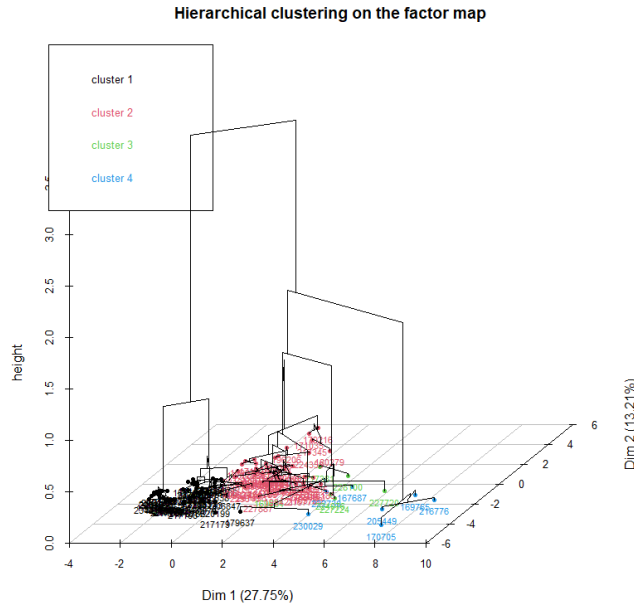


FIGURE 12.12 – Partitionnement des individus en 4 classes

Sur le graphe ci-dessus, on dispose de 3 informations : le plan de l’ACP avec les deux premières dimensions, l’arbre hiérarchique et l’appartenance aux classes (selon les couleurs). À noter que les informations sont différentes, le plan de l’ACP résume les informations principales sur uniquement 2 dimensions tandis que l’arbre hiérarchique se base sur 7 dimensions.

Classe	Nombre d’individus
1	52
2	52
3	6
4	7

TABLE 12.2 – Partitionnement des individus selon 4 classes

On constate que la répartition des individus au sein des classes n’est pas homogène. Cela n’est pas surprenant au vu de l’hétérogénéité du portefeuille constitué majoritairement de petits contrats, contrastant avec la présence de quelques très gros contrats. Ce n’est pas un problème, l’important est que les individus d’une même classe soient homogènes afin d’améliorer l’efficacité du modèle de crédibilité hiérarchique reposant sur des classes de même nature de risque.

12.2.3 Caractérisation des classes

Une fois les classes définies, il est important de les caractériser selon les variables d’origines afin de pouvoir mieux les interpréter et de comprendre comment elles ont été constituées.

12.2.3.1 Caractérisation des classes par les variables catégorielles

Tout d’abord, on regarde s’il est possible de caractériser les classes selon les variables catégorielles (région et activité). Pour ces variables, un test du Chi2 est réalisé pour voir si la variable de classe est liée aux variables qualitatives.

Variable catégorielle	P.value
Activité professionnelle de la flotte	$1,65.10^{-11}$

TABLE 12.3 – Test du Chi2 pour la variable catégorielle activité

La probabilité critique est inférieure à 5% , le test du Chi2 est significatif. On en déduit donc que la variable activité de la flotte intervient dans la constitution des classes. Ensuite, en regardant modalité par modalité et classe par classe, on peut tirer quelques enseignements.

- Sur-représentation des modalités "caisses publiques" et "établissements publics" dans la classe 1 :
 - 100 % des flottes liées à des caisses publiques sont dans la classe 1.
 - 83% des établissements publics sont dans la classe 1
 - 60% des flottes de la classe 1 appartiennent à des caisses publiques ou des établissements publics.
- Sur-représentation de la modalité commune "dans" la classe 2 :
 - 83 % des flottes appartenant aux communes sont dans la classe 2.
 - 75% des flottes de la classe 2 appartiennent à des communes.

En ce qui concerne la variable catégorielle liée à la localisation de la flotte, elle ne permet pas de décrire les classes.

12.2.3.2 Caractérisation des classes par les variables numériques

Le rapport de corrélation est une mesure utilisée en analyse de variance pour déterminer le lien entre une variable quantitative et une variable qualitative. Dans le cadre de l'application de classification ascendante hiérarchique avec le package FactoMineR, ce rapport de corrélation est calculé entre chaque variable quantitative et la variable de classe (qualitative) afin de mettre en avant les variables participant le plus à la constitution des classes.

Variable quantitative	Rapport de corrélation avec la variable de classe
Nombre de sinistre RC 2019	0,71
Nombre de véhicules	0,67
Montant total écrêté des sinistres 2019	0,65
Prime acquise 2019	0,57
Provisions 2019	0,56
Part d'engins dans la flotte	0,54
Part de véhicules légers assurés en tous risques	0,49
Part de remorques	0,48
Part de poids lourds non assuré en tous risques	0,45
Niveau de franchise dommage poids lourds	0,40
Niveau de franchise dommage véhicules légers	0,35

TABLE 12.4 – Rapports de corrélation entre les variables quantitatives et la variable de classe

On observe que les variables relatant de l'importance de la flotte (sinistralité, primes, parc) sont très liées à la variable de classe.

Ici, l'idée est de caractériser chaque classe en fonction des valeurs numériques prises par les individus au sein de la classe en question.

- Classe 1 : surreprésentation des petites flottes assurées tous risques présentant des bons résultats techniques.
 - Les individus de la classe 1 émettent des primes annuelles beaucoup plus faibles que la moyenne (20 767 € versus 46 910 €). Le constat est similaire pour la sinistralité, les provisions, le parc, etc.
 - Les individus de la classe 1 ont une part plus importante de véhicules légers assurés en tous risque (40% contre 26% en moyenne) et inversement une part beaucoup plus faible des autres types de véhicules.

- Les individus de la classe 1 ont de meilleurs résultats techniques que la moyenne (SP : 0,81 versus 0,97 en moyenne. Fréquence RC : 4,2% versus 7,5% en moyenne).
- Classe 2 : sur-représentation des flottes assurant une proportion non négligeable d’engins, de 2 roues et de poids lourds.
 - Les individus de la classe 2 ont une part plus importante d’engins au sein de leurs flottes (10% contre 6% en moyenne). Même constat pour les remorques, les 2 roues et les poids lourds. Inversement, la part de véhicules légers assurés en tous risques est plus faible (14% versus 26% en moyenne).
 - Les individus de la classe 2 ont des niveaux de franchise dommage plus faible pour les véhicules légers (201 € versus 290€ en moyenne).
- Classe 3 : sur-représentation des grosses flottes (contenant notamment une part non négligeable de poids lourds), avec des niveaux élevés de franchise dommage et présentant des résultats techniques dégradés.
 - Les individus de la classe 3 présentent des niveaux de franchise dommage particulièrement élevés relativement à la moyenne (véhicules légers : 930€ vs 280€ en moyenne. poids lourds : 1 710€ versus 390€ en moyenne).
 - Les individus de la classe 3 assurent des parcs de véhicules beaucoup plus importants que la moyenne (430 vs 200 en moyenne). Même constat pour la sinistralité et les primes.
 - Les résultats techniques des individus de la classe 3 sont plus dégradés que la moyenne (fréquence RC : 9,5% versus 7,5% en moyenne)
 - Les individus de la classe 3 ont une part plus importante de poids lourds que la moyenne (17 % versus 4% en moyenne).
- Classe 4 : sur-représentation des très grosses flottes.
 - Les individus de la classe 4 affichent un nombre de sinistres RC 2019 beaucoup plus important que la moyenne (60 vs 11 en moyenne) expliqué par le nombre de véhicules assurés (935 vs 200 en moyenne).
 - Les individus de la classe 4 ont une part plus importante d’engins que la moyenne (10% vs 6% en moyenne).

Les classes de risques 3 et 4 contenant les grosses flottes du portefeuille GMF sont celles pour lesquelles l’enjeu est le plus fort en termes de revalorisation tarifaire.

12.2.3.3 Caractérisation des classes par les axes factoriels

Il est également possible de caractériser les classes grâce aux axes factoriels de l’ACP. Le même calcul du rapport de corrélation avec la variable de classe est réalisé.

Axe factoriel	Rapport de corrélation avec la variable de classe
1	0,80
2	0,48
3	0,42
4	0,11

TABLE 12.5 – Rapports de corrélation entre les axes factoriels de l’ACP et la variable de classe

On relève une liaison très forte entre le premier axe principal et la variable de classe, une liaison modérée entre les axes principaux 2,3 et la variable de classe et pour finir une liaison très faible entre le quatrième axe principal et la variable de classe.

Quelques constats :

- Les individus de la classe 1 ont des coordonnées significativement beaucoup moins élevées que la moyenne sur les deux premiers axes principaux.
- Les individus de la classe 2 ont des coordonnées significativement plus élevées que la moyenne sur le deuxième axe principal.
- Les individus de la classe 3 ont des coordonnées significativement plus élevées que la moyenne sur le premier et le troisième axe principal.
- Les individus de la classe 4 ont des coordonnées significativement beaucoup plus élevées que la moyenne sur le premier axe principal.

12.2.3.4 Parangons des classes

La fonction **HCPC** du package FactoMineR indique également pour chaque classe les parangons. Ce sont les individus les plus représentatifs de la classe. Soit G_i le parangon de la classe i . Voici un aperçu des valeurs prises par certaines variables

Parangon	Activité	Parc	Prime	SP	Fréq RC	Vehic. légers TR	Engins	Vehic. lourds	Franch. VL
G1	Établ. public	104	40226	78%	4,8%	35%	3%	0%	450
G2	Commune	131	37852	73%	7,6%	8%	16%	6%	400
G3	Commune	258	80783	129%	14%	10%	5%	28%	900
G4	Département	955	249228	95%	8,7%	10%	11%	23%	750

TABLE 12.6 – Parangons des différentes classes

12.2.3.5 Caractérisation des classes par un arbre

Le package **rpart** sur R propose une implémentation des méthodes de construction d'arbres de décision inspirée de l'approche CART. Ces méthodes ne seront pas détaillées dans ce mémoire. La fonction **rpart** du package permet de visualiser les classes sous forme d'arbre.

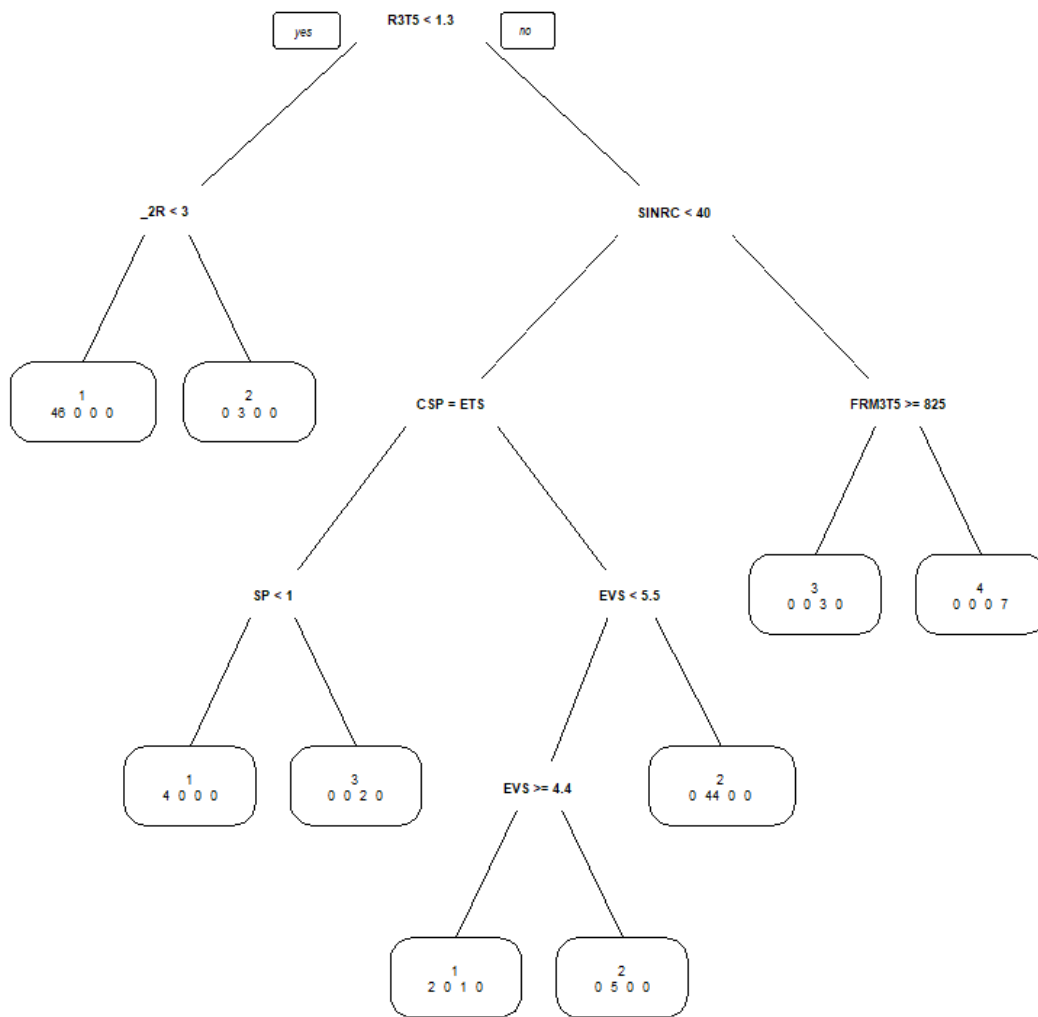


FIGURE 12.13 – description des classes sous forme d’un arbre de décision

En premier lieu, la proportion de remorques légères (moins 3 tonnes 5) est testée. Si elle est inférieure à 1,3 % alors la part de véhicules 2 roues est testée. Si cette valeur est inférieure à 3% alors le client appartient à la classe 1 (46). Sinon, il appartient à la classe 2 (3). Inversement, si la part de remorques est supérieure à 1,3%, c’est le nombre de sinistres RC qui distingue les classes. Si ce nombre est supérieur à 40 alors le niveau de franchise dommage des véhicules légers est testé. S’il est supérieur à 825 € alors le client est dans la classe 3 (3). Sinon il est dans la classe 4 (7). Réciproquement, si le nombre de sinistres RC est inférieur à 40, l’activité professionnelle du client est testée. Si le client est un établissement alors le SP intervient. S’il est inférieur à 1, le client appartient à la classe 1 (4). Sinon, il est dans la classe 3 (2). Réciproquement, si le client n’est pas un établissement public alors c’est la part d’engins qui distingue les classes. Si la flotte contient plus de 5,5% d’engins alors le client appartient à la classe 2 (44). Sinon si la proportion est comprise entre 4,4% et 5,5% alors le client est dans la classe 2 (2) ou 3 (1). Sinon il est dans la classe 2 (5).

Cet arbre de décision a l'avantage de faire intervenir beaucoup de variables :

- L'importance du parc (le nombre de sinistres RC, plus il est important plus la flotte est grande généralement).
- La répartition de la flotte selon les différentes catégories de véhicules (remorques, engins, 2 roues).
- L'activité du client.
- Les résultats techniques (SP).
- Les niveaux de franchise (véhicules légers).

Au total, les 128 contrats sont répartis en 5 classes : 117 contrats répartis dans les 4 classes citées plus haut et 11 clients, possédant un contrat avec conservation des sinistres, dans la classe 5. À noter que les contrats avec conservation sont généralement des gros contrats présentant des résultats assez dégradés.

12.3 La crédibilité hiérarchique

Tout comme le modèle de Bühlmann-Straub, le modèle de crédibilité hiérarchique est appliqué à l'aide de la fonction `cm` du package `actuar`. À la différence que, cette fois, le jeu de données contient les classes. L'objectif est de mettre en place le même processus de revalorisation du portefeuille que celui implémenté avec le modèle de Bühlmann-Straub. Le loss-ratio et la fréquence de sinistralité RC seront donc projetées à l'aide du modèle de crédibilité hiérarchique.

12.3.1 Ratio considéré : Le loss-ratio

Prime collective \tilde{m}	Variance intra portefeuille \hat{s}^2	Variance inter portefeuille \tilde{a}	$\hat{K} = \frac{\hat{s}^2}{\tilde{a}}$	Variance inter-classe \tilde{b}
1,0454	10 124	0.0497	203 701	0,0071

TABLE 12.7 – Estimation des paramètres du modèles de crédibilité hiérarchique sur les ratios sinistres à primes

12.3.1.1 Crédibilité des classes

Classe	Moyenne	Poids	Facteur de crédibilité	Loss ratio crédibilisé π_k^J
1	0,815	2,139	0,234	0,987
2	0,944	9,355	0,571	0,985
3	1,046	6,588	0,484	1,043
4	1,066	18,736	0,728	1,059
5	1,137	4,871	0,410	1,080

TABLE 12.8 – Crédibilité des classes

L'écrêtement des sinistres prend ici tout son sens. Si les ratios sinistres à primes bruts avaient été considérés, les sinistres très graves du portefeuille auraient impacté très fortement les loss ratios crédibilisés de certaines classes (notamment avec le sinistre à 18 millions d'euros). Toutefois, il aurait pu être envisageable de conserver les ratios sinistres à primes bruts en se disant que les hausses engendrées sur la prime collective et les primes par classe permettraient de modéliser la répartition du coût des sinistres graves sur l'ensemble du portefeuille. C'est une idée qui se défend, mais elle n'a pas été retenue. D'autant plus que le montant total de la revalorisation tarifaire obtenu en écrétant les sinistres est jugé suffisamment élevé.

12.3.1.2 Application numérique sur un exemple fictif

Pour bien assimiler la méthode de calcul, prenons l'exemple du client fictif suivant en supposant qu'il est assuré par un contrat flottes GMF depuis début janvier 2018. On souhaite projeter son SP en 2020 avec le modèle de la crédibilité hiérarchique.

Pour rappel, les loss-ratios crédibilisés vérifient l'équation suivante :

$$\begin{aligned}\pi_{k,i,n+1}^J &= z_{k,i} X_{k,i,w} + (1 - z_{k,i}) \pi_k^J \\ \pi_k^J &= z_k X_{k,z,w} + (1 - z_k) m\end{aligned}$$

client	Classe	Prime acquise HT 2018	Prime acquise HT 2019	Sinistralité 2018	Sinistralité 2019
B	3	95 000 €	102 050 €	106 000 €	115 000 €

TABLE 12.9 – Calcul du loss ratio projeté d'un client fictif avec le modèle de la crédibilité hiérarchique

- $w_{3,B,2018} = 95000$, $w_{3,B,2019} = 102050$, $w_{3,B,\Sigma} = 197050$
- $X_{3,B,2018} = \frac{106000}{95000}$, $X_{3,B,2019} = \frac{115000}{102050}$
- $X_{3,B,w} = \sum_{j=2018}^{2019} \frac{w_{3,B,j}}{w_{3,B,\Sigma}} X_{3,B,j} = 1,122$
- $\widehat{z}_{3,B} = \frac{w_{3,B,\Sigma}}{w_{3,B,\Sigma} + \hat{K}} = 0,49$

$$\bullet \pi_{3,B,2020}^J = z_{3,B} X_{3,B,w} + (1 - z_{3,B}) \pi_3^J = 1,082$$

12.3.2 Ratio considéré : la fréquence de sinistralité RC

Prime collective \tilde{m}	Variance intra portefeuille \hat{s}^2	Variance inter portefeuille \tilde{a}	$\hat{K} = \frac{\hat{s}^2}{\tilde{a}}$	Variance inter-classe \tilde{b}
0,0673	0,0902	0,00034	265	0,00017

TABLE 12.10 – Estimation des paramètres du modèles de crédibilité hiérarchique sur les fréquences de sinistralité RC

12.3.2.1 Crédibilité des classes

Classe	Moyenne	Poids	Facteur de crédibilité	Fréquence RC crédibilisée
1	0,063	3,873	0,663	0,0644
2	0,069	16,753	0,895	0,0687
3	0,071	9,228	0,824	0,070
4	0,061	32,648	0,943	0,061
5	0,075	6,077	0,755	0,073

TABLE 12.11 – Crédibilité des classes

12.3.3 Mise en place des revalorisations tarifaires à partir du modèle de crédibilité hiérarchique

Les mêmes stratégies tarifaires que celles mises en place sur le modèle de crédibilité de Bühlmann-Straub sont appliquées aux clients. Finalement, après surmajoration des clients présentant une fréquence de sinistralité RC supérieure à 10% et en tenant compte des nouveaux contrats présents en portefeuille, non concernés par le modèle de crédibilité, on obtient un montant total de revalorisation de l'ordre de 918 762 €.

Il s'avère que la revalorisation tarifaire totale obtenue à la suite de l'application de la crédibilité hiérarchique (918 762 €) est du même ordre de grandeur que celle obtenue avec l'application du modèle de crédibilité de Bühlmann-Straub (901 547 €). En comparant les facteurs de crédibilité et les projections obtenues, que ce soit pour le loss-ratio ou la fréquence de sinistralité RC, entre le modèle de Bühlmann-Straub et le modèle de crédibilité hiérarchique, on s'aperçoit que les valeurs sont très proches, les différences ne semblent pas être significatives.

12.3.3.1 Résultats des revalorisations tarifaires par classe

Le tableau ci-dessous représente le montant de revalorisation obtenu pour chaque classe. Les classes 3,4 et 5 composées majoritairement de très grosses flottes et présentant des résultats techniques dégradés représentent 75% de la revalorisation totale.

Classe	Nombre de clients	Revalorisation tarifaire
1	52	47 931€
2	52	131 954 €
3	6	64 904 €
4	7	253 538 €
5	11	270 445€

TABLE 12.12 – Crédibilité des classes

Les primes émises flottes à fin décembre 2020 représentent 14,1 millions d'euros. Si on y intègre les mesures tarifaires préconisées par les modèles de crédibilité (et en partant du principe qu'elles soient acceptées par les clients), on aboutit à 15 millions d'euros, soit une évolution de +6,4%. En fait, ces 6,4% ont été répartis de manière hétérogène entre les clients. Pour certains, la majoration est inférieure à 6,4% (par exemple 4% pour les bons contrats) tandis que pour d'autres, elle est supérieure à 6% (tout en étant majoré par 30%). En pratique, certains contrats sont plus difficilement majorables que d'autres de par la présence de clauses empêchant des majorations au-delà d'une certaine valeur. Voici pourquoi une fois l'étude mise en place, il est important de discuter avec les technico-commerciaux, disposant de la connaissance client pour savoir ce qu'il est possible de faire ou non.

12.4 L'application du modèle de surveillance sans projection des ratios

L'idée de cette partie est d'appliquer le même algorithme de surveillance que celui appliqué avec les modèles de crédibilité, la différence étant qu'au lieu de considérer les ratios projetés on considère les ratios historiques moyens des flottes. Cela permet de juger de l'intérêt ou non de la crédibilité. Le seuil de 300k euros concernant le poids total des primes du client remplace le seuil de 60% du facteur de crédibilité associé au loss-ratio projeté, ces deux informations sont équivalentes.

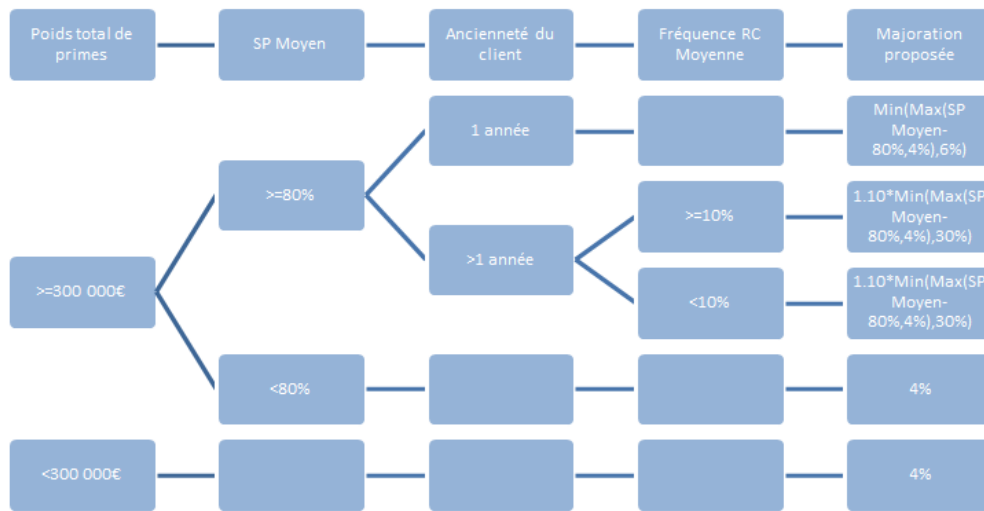


FIGURE 12.14 – modèle de surveillance sans projection des ratios

En appliquant le modèle de surveillance ci-dessus, on obtient une revalorisation totale de 885 603 euros. L'ordre de grandeur est le même que les revalorisations obtenues avec les modèles de crédibilité. On peut donc en déduire que les modèles de surveillance intégrant la théorie de la crédibilité ne semblent pas apporter de résultats significativement différents de modèles simples basés sur l'historique des flottes.

12.5 Résultats des différents modèles de surveillance

Modèle de surveillance 2021	Revalorisation du portefeuille
Modèle SRA	428 606€
Modèle de base	621 099€
Modèle de surveillance de Bühlmann-Straub	901 547€
Modèle de surveillance crédibilité hiérarchique	918 762 €
Modèle de surveillance sans projection des ratios	885 603€

TABLE 12.13 – Crédibilité des classes

La revalorisation totale obtenue avec le modèle de base est très différente de celles obtenues avec les autres modèles de surveillance car les règles de majoration définies sont totalement différentes.

Conclusion

Finalement, les modèles de crédibilité ne semblent pas apporter des résultats significativement différents d'un modèle se basant simplement sur les résultats historiques de la Flotte (sans projection). Il a donc été décidé de ne pas inclure de modèles de crédibilité dans la surveillance. Le modèle de surveillance de base, basé sur l'historique du client (SP annuels, SP moyens, ancienneté au sein du portefeuille) a été conservé au détriment du modèle de surveillance sans projection des ratios. Ce modèle a l'avantage d'être simple à implémenter, compréhensible par tous et n'implique pas une revalorisation trop brutale du portefeuille.

Toutefois, il pourrait être envisageable, dans les années futures, d'utiliser un des modèles de crédibilité présenté dans ce mémoire (Bühlmann-Straub ou la crédibilité hiérarchique) tel quel ou de continuer à les développer. Ces méthodes de crédibilité ont l'avantage de proposer une solution à la dégradation des résultats techniques flottes automobiles de GMF. Ces méthodes permettent de projeter les résultats techniques du portefeuille en pondérant l'expérience individuelle du client par la sinistralité historique du portefeuille. Ensuite, une fois, ces résultats techniques projetés, des règles de majoration sont mises en place. Elles ont pour objectif de cibler en particulier les grosses flottes (poids total des primes supérieur à 300k euros sur la période d'assurance). Ces règles sont basées sur le facteur de crédibilité accordé au loss-ratio individuel du client, sur les projections du loss-ratio écrêté à 150k euros et de la fréquence de sinistralité responsabilité civile et sur l'ancienneté du client. Elles dépendent également fortement des seuils définis à dire d'expert (80% pour le SP d'équilibre et 10% pour la fréquence RC). Ces seuils ont l'avantage d'être paramétrables. C'est-à-dire que si une année la GMF décidait de majorer plus fermement les contrats flottes, il suffirait de diminuer les seuils. Au contraire, si la GMF souhaite une politique de revalorisation légère, il suffirait d'augmenter les seuils. De même, si la GMF souhaite des mesures beaucoup plus ciblées sur les flottes présentant des fréquences de sinistralité RC relativement importante, il serait possible de baisser ce seuil tout en conservant le seuil défini pour le loss-ratio. Les règles pourraient être étoffées en considérant peut-être plus de variables ou alors en majorant plus sévèrement les flottes dont les résultats techniques se dégradent au fil du temps. Le loss ratio projeté avec les modèles de crédibilité est écrêté à 150k euros, il est également tout à fait possible de projeter le SP avec un écrêtement différent (plus grand si on souhaite majorer plus sévèrement ou plus petit si on souhaite une revalorisation légère du portefeuille).

Quant au modèle de crédibilité hiérarchique, il repose sur 5 classes, il serait tout à fait envisageable d'en créer plus ou d'appliquer d'autres méthodes de classification. Toutefois, les modèles de classifications risquent de se heurter au problème du faible volume de données. Le modèle de crédibilité hiérarchique gagnerait également en précision s'il était possible d'avoir à disposition les valeurs exactes des parcs et de leurs répartitions et non seulement des estimations plus ou moins précises. Cela permettrait d'optimiser les regroupements de clients dans des classes plus homogènes.

Dans tous les cas, modèles de crédibilité ou non, il ne faut pas oublier que l'objectif des modèles est seulement de préconiser pour chaque client une mesure. Il est probable qu'une grande partie des préconisations ne soient pas acceptées ou alors minorées. En fait, il tient d'en discuter avec les technico-commerciaux et le courtier pour savoir si les mesures sont réalisables et si elles ne mettraient pas en péril la potentielle résiliation du contrat par l'assuré (sauf si la GMF souhaite résilier la Flotte en question). Des paramètres autres que les résultats techniques de la flotte rentrent en jeu ; négociations, clauses, stratégie commerciale, etc.

Dans tous les cas, cela ne remet pas en question la véritable surveillance du portefeuille qui consiste à sélectionner les risques à l'entrée, c'est-à-dire lors de l'analyse de l'appel d'offres. Afin de redresser le portefeuille, au-delà de la mise en place de la nouvelle stratégie de revalorisation, il a été décidé de ralentir la croissance du portefeuille et de le stabiliser afin de le redresser. D'où une attention encore plus forte à venir sur la tarification des prochains appels d'offres.

Table des figures

1.1	Fonctionnement de la conservation (source Arts&MétiersMag)	5
2.1	Évolution des cotisations flottes en fonction du temps (source FFA)	7
2.2	Évolution du ratio sinistres à primes de l'assurance flotte en fonction du temps (source FFA)	8
3.1	Exemple d'un parc d'une collectivité	11
3.2	Exemple de l'historique de sinistralité du parc d'une collectivité	11
4.1	Évolution du nombre de contrats flottes GMF	16
4.2	Évolution des primes acquises (hors taxes et hors assistance) flotte et du parc assuré GMF selon les différents exercices	17
4.3	Distribution des primes acquises et des parcs automobiles en 2019	17
4.4	Répartitions des primes acquises 2019 selon les garanties	19
4.5	Répartition des contrats avec une franchise annuelle en termes de nombre et de primes acquises	21
5.1	Liquidation des sinistres RCM, DOM et BDG	24
5.2	Évolution des coûts moyens des sinistres clos RC, DOM et BDG observés sur les flottes GMF selon les exercices	25
5.3	Fréquences de sinistralité des garanties principales	27
5.4	Vieillessement des SP bruts	28
5.5	Vieillessement des SP bruts écrêtés à 150000 euros	30
5.6	Répartition des charges 2019 selon les différentes garanties	31
5.7	Vieillessement des SP RC	32
5.8	Vieillessement des SP DOM et BDG	32
6.1	Exemple d'une situation contractuelle d'un client (1)	38
6.2	Exemple d'une situation contractuelle d'un client (2)	38
6.3	Exemple d'une situation contractuelle d'un client (3)	38
6.4	Exemple d'une situation contractuelle d'un client (4)	38
7.1	Conséquence dans les fichiers de pilotage d'une prime émise sur 2 garanties	40
7.2	Fichiers mensuels des primes	41
7.3	Fichiers semestriels et annuels des primes	41
7.4	Exemple fictif d'un client dans les infocentres primes	42
7.5	Exemple fictif de l'information à récupérer dans la nouvelle base	42
7.6	Calcul de la prime annuelle d'un client possédant plusieurs contrats par année	42
7.7	Calcul de la prime annuelle d'un client lors d'un chevauchement	43
7.8	Règles de prise en compte des sinistres dans les infocentres	44
7.9	Problème des chaînes de caractères vides	46
7.10	Traitement des 0	46
8.1	Exemple fictif de données manquantes de parcs automobiles	49
8.2	Exemple fictif du calcul des primes acquises par véhicule	49
8.3	Indices SRA de l'année 2015 et 2016	50
8.4	Évolution de l'indice SRA entre 2015 et 2016	51
8.5	Exemple de l'estimation des primes acquises par véhicules	51
8.6	Exemple de l'estimation des parcs	52

8.7	Exemple fictif des primes émises d'un contrat par catégorie de risque	52
8.8	Problème de distinction des véhicules tous risques des véhicules hors tous risques	53
8.9	Exemple fictif du calcul des primes tous risques	54
8.10	Exemple fictif du calcul des primes hors tous risques	54
8.11	Exemple fictif de la segmentation des véhicules tous risques des véhicules hors tous risques	55
8.12	Exemple fictif de l'estimation du nombre de véhicules par type de risque	56
11.1	Évolution du nombre de clients et de contrats flottes GMF	78
11.2	Règles de surveillance définies sous forme d'arbre	79
11.3	Mis à jour des règles de surveillance en y intégrant la fréquence RC projetée	81
11.4	Tests de sensibilité du modèle de Bühlmann-Straub	81
12.1	Pourcentage de variance expliquée par les axes principaux	85
12.2	Qualité de représentation des variables sur le plan principal	87
12.3	Contribution des variables d'origine au premier axe	87
12.4	Contribution des variables d'origine au deuxième axe	88
12.5	Cercle de corrélation	88
12.6	Projection des individus sur le plan principal	89
12.7	Barycentres des modalités de la variable catégorielle région	90
12.8	Barycentres des modalités de la variable catégorielle activité professionnelle	90
12.9	Contribution des variables d'origine au troisième axe	91
12.10	Contribution des variables d'origine au quatrième axe	91
12.11	Perte d'inertie inter-classe selon le nombre de classes	94
12.12	Partitionnement des individus en 4 classes	95
12.13	description des classes sous forme d'un arbre de décision	99
12.14	modèle de surveillance sans projection des ratios	103

Liste des tableaux

2.1	Fréquences et coûts moyens des garanties principales flottes (source FFA)	8
4.1	Quantiles des primes acquises 2019 et du nombre de véhicules assuré par contrat	18
4.2	Répartitions des contrats flottes GMF 2019 selon les tranches de primes acquises	18
4.3	Répartition du parc automobile flottes GMF 2019	18
4.4	Répartitions des primes acquises 2019 selon les régions	19
4.5	Répartitions des primes acquises 2019 selon les secteurs d'activités professionnels	20
4.6	Répartition de l'ancienneté des clients	21
5.1	Nombre de sinistres flottes des garanties RCM, DOM, BDG, VOL et INC	26
5.2	Parc assuré selon les garanties	26
5.3	Comparaison des fréquences flottes et auto GMF de l'exercice 2019 observé à fin décembre 2019	27
5.4	Ratios sinistres à primes bruts de l'exercice 2019 (observé à fin décembre 2019) selon le niveau de prime	33
8.1	Tableau des coefficients des primes en fonction des primes dommages	54
8.2	Primes par véhicules : 4 roues et 2 roues	55
8.3	Primes par véhicules : Poids lourds, remorques, engins	56
10.1	Répartition de l'ancienneté des clients	73
10.2	Classes de dégradation	74
10.3	Répartition des contrats dégradés selon les niveaux de dégradation	74
11.1	Estimation des paramètres du modèle de Bühlmann-Straub sur les ratios sinistres à primes	77
11.2	Calcul du loss ratio projeté d'un client fictif	77
11.3	Répartition des facteurs de crédibilité associés au loss-ratio	78
11.4	Estimation des paramètres du modèles de Bühlmann-Straub sur les fréquences de sinistralité RC	80
11.5	Répartition des facteurs de crédibilité associés à la fréquence de sinistralité	80
12.1	Variance expliquée par les 7 premiers axes	85
12.2	Partitionnement des individus selon 4 classes	95
12.3	Test du Chi2 pour la variable catégorielle activité	96
12.4	Rapports de corrélation entre les variables quantitatives et la variable de classe	96
12.5	Rapports de corrélation entre les axes factoriels de l'ACP et la variable de classe	97
12.6	Parengons des différentes classes	98
12.7	Estimation des paramètres du modèles de crédibilité hiérarchique sur les ratios sinistres à primes	101
12.8	Crédibilité des classes	101
12.9	Calcul du loss ratio projeté d'un client fictif avec le modèle de la crédibilité hiérarchique	101
12.10	Estimation des paramètres du modèles de crédibilité hiérarchique sur les fréquences de sinistralité RC	102
12.11	Crédibilité des classes	102
12.12	Crédibilité des classes	102
12.13	Crédibilité des classes	103

Bibliographie

- BOUKROUTE, A. (2017). L'application de la théorie de la crédibilité dans le cadre du renouvellement en prévoyance collective. *Mémoire d'actuariat*.
- BÜHLMANN, H. (1970). Experience rating and credibility. *ASTIN Bulletin*.
- BÜHLMANN, H. et STRAUB, E. (1970). Glaubwürdigkeit für Schadensätze. *Bulletin of the Swiss Association of Actuaries*.
- CHARPENTIER, A. (2016). Actuariat de l'assurance non-vie.
- FINK, J.-B. (2019). Mise en place d'un nouveau pilotage technique pour les risques spécifiques d'une compagnie d'assurances. *Mémoire d'économétrie et de statistiques*.
- GOOVAERTS, M. et HOOGSTAD, W. (1987). Credibility theory. surveys of actuarial studies. *ASTIN Bulletin*.
- GOULET, V. (2008). *Mathématiques actuarielles IARD*.
- LANDON, D. (2018). Construction d'une échelle bonus-malus pour la tarification des flottes de véhicules. *Mémoire d'actuariat*.
- LAVAL, E. (2017). Impact de la réforme 100% santé sur les contrats de frais de santé par un modèle de crédibilité. *Mémoire d'actuariat*.
- MOWBRAY, A. H. (1914). How extensive a payroll exposure is necessary to give a dependable pure premium? *Proceedings of the Casualty Actuarial Society*.
- STÉPHANE KOLASA, J. P. (2008). Tarification en iard et nouvelles contraintes de rentabilité : Étude d'un produit flotte automobile. *Mémoire d'actuariat*.
- THÉRON, P. (2017). Théorie de la crédibilité.