

Mémoire présenté le :

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires**

Par : Benjamin MARTIALOT

Titre Exigences réglementaires et opportunités dans le cadre des meilleures estimations des sinistres Automobile, notamment corporels graves

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

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

*Membre présents du jury de l'Institut
des Actuaires* signature

Entreprise :

Nom : MAPA

Signature :

Directeur de mémoire en entreprise :

Nom : Cécile JONIOT

Signature :

Invité : Galea & Associés

Nom : Léonard FONTAINE

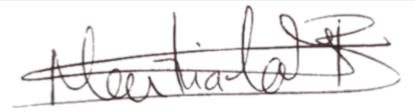
Signature :

**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



Résumé

L'évaluation des meilleures estimations dans le cadre de Solvabilité II a une influence forte sur le ratio de solvabilité de l'entreprise d'assurance, puisqu'elle joue à la fois sur le bilan prudentiel (les fonds propres), et sur les risques encourus sur l'année à venir (le capital de solvabilité requis). Cet élément central est donc bien encadré dans les différentes réglementations européennes. Mais l'Actuaire garde malgré tout une certaine souplesse technique, sur le périmètre et les méthodes utilisés, s'il est en mesure de prouver que ces derniers sont les plus adaptés.

Les principales difficultés dans cet exercice se situent généralement sur les sinistres de grande ampleur, appelés sinistres graves, qui ne présentent pas les attributs nécessaires pour pouvoir utiliser les méthodes déterministes classiques comme Chain Ladder. La première étape sensible étant tout d'abord de pouvoir définir ce qu'est un sinistre grave. La théorie des valeurs extrêmes, et les théories sur l'incertitude de l'évaluation des provisions à l'ultime permettront, respectivement, de dégager des solutions et de conforter ces résultats. Ensuite, il faut alors trouver une méthode alternative aux méthodes agrégées. La méthode exploitée est un modèle CART, permettant de faire du provisionnement ligne à ligne et d'exploiter les données individuelles liées à chaque sinistre.

Ce mémoire traitera de l'ensemble de ces sujets dans la cadre de l'assurance Automobile en France, dans une entreprise d'assurance de taille moyenne, mettant également en avant les difficultés et les solutions proposées découlant d'une base de données restreinte.

Abstract

The evaluation of best estimate liabilities under the framework of Solvency II has a strong influence on the insurance company's solvency ratio. This is due to the fact that best estimate liabilities evaluation affects both the prudential balance sheet (the own funds) and the risks incurred over the coming year (the solvency capital requirement). This central element is therefore well framed in the various European regulations. Nevertheless, Actuaries can retain a certain technical flexibility regarding the scope and the methods if they are able to prove that these variations are more suitable.

The main difficulties in the best estimate evaluations are generally on large claims. Indeed, large claims don't have the characteristics that allow the use of classic deterministic methods such as Chain Ladder. The first sensitive step is to define a major insurance claim. The extreme value theory as well as the theories on the uncertainty around the technical provisions valuation at the ultimate will allow, respectively, to find solutions and to confirm the results. Then, it is necessary to find an alternative method to the aggregated methods. We will be looking at the CART model that allows a line by line provisioning and also exploits the individual data linked to each claim.

This thesis will cover all of the above topics in the context of motor insurance in France by considering a medium-size insurance company. It will also highlight the difficulties and the proposed solutions resulting from a narrow database.

Remerciements

Je tiens tout d'abord à remercier la MAPA, et notamment Vincent LOIZEIL et Cécile JONJOT, d'avoir accepté que je me lance dans cette aventure, et de m'avoir soutenu.

Je remercie également Manon ROUSSEL, qui m'a guidé dans cette voie.

Je remercie chaleureusement mes tuteurs professionnels Nathalie RAMOS et Léonard FONTAINE, pour leur soutien, le temps qu'ils ont pris pour m'accompagner, et l'apport de leur expertise sur mes travaux.

Je remercie l'ISFA, pour son enseignement de qualité, ainsi que Aurélien COULOUMY pour son analyse critique, et ses conseils.

Je remercie toutes les personnes qui font ou qui ont fait partie de mon équipe à la MAPA, et notamment pour leur bonne humeur et l'intérêt qu'ils portaient à mon projet.

Je remercie ma famille, et mes amis, qui m'ont aidé à traverser des périodes difficiles, jonglant entre mon travail, mes études et ma vie privée.

Enfin, je remercie tout particulièrement ma femme, Élodie, sans qui, tout cela n'aurait pas été possible. Je la remercie d'avoir été si patiente.

Table des matières

Résumé.....	3
Abstract	4
Remerciements	5
Introduction.....	8
Chapitre 1 : De l'exigence réglementaire vers l'opportunité.....	10
I. Contexte général Solvabilité II.....	10
A. Les exigences quantitatives.....	10
B. Les exigences qualitatives	12
C. Les exigences en matière de transmission d'information prudentielle et de publications ..	13
D. La Qualité des données et provisions techniques.....	14
II. La MAPA (Mutuelle d'Assurance des Professionnels de l'Alimentaire).....	16
A. Présentation générale	16
B. Le rapport actuariel et lien avec la problématique	16
C. L'exemple de l'Automobile.....	17
III. La branche Automobile	18
A. Évolution du portefeuille et rentabilité.....	18
B. Réassurance et rentabilité.....	19
C. Solvabilité II et meilleures estimations	21
D. Sinistralité automobile et segmentation.....	22
Chapitre 2 : Complexité de la matière et outils mathématiques	32
I. Les fondamentaux du provisionnement	32
A. Les différentes provisions.....	32
B. Méthodes de provisionnement agrégées : Chain Ladder	35
C. Méthodes de provisionnement agrégées : Mesure de l'incertitude	38
II. Définir le seuil des graves.....	42
A. Notions sur la Théorie des Valeurs Extrêmes (TVE)	42
B. La Fonction Moyenne des Excès (Mean Excess Function)	43
III. Le provisionnement ligne à ligne.....	47
A. Arbres de décision : la méthode CART	47
B. Mesures de validation	49
Chapitre 3 : Exploration de la sinistralité Automobile	51
I. La prise de recul de l'Actuaire	51
A. Chain Ladder et vérifications des hypothèses.....	51
B. Résultats et cohérence	55

II.	Seuil des sinistres graves	58
A.	Correction de l'inflation et mise à l'ultime.....	58
B.	Analyses statistiques	59
C.	Fonction Moyenne des Excès (FME) et détermination du seuil.....	60
III.	Sinistres Automobile RC « Corporel » et Provisionnement agrégé.....	65
A.	Provision à l'ultime et incertitude	65
B.	Choix du seuil a posteriori	70
IV.	Sinistres Automobile RC « Corporel » et Provisionnement ligne à ligne	71
A.	Variables explicatives et périmètre	71
B.	Modélisations CART.....	73
V.	Répercussions sur le ratio de Solvabilité de l'exercice 2019.....	86
	Conclusion et perspectives.....	88
	Bibliographie.....	91
	Table des figures.....	92
	Table des tableaux.....	93
	Liste des sigles et abréviations	94

Introduction

La réglementation Solvabilité II a été un tournant dans le fonctionnement des entreprises d'assurance et de réassurance. Cette directive a contraint à une nouvelle organisation, notamment en termes de gouvernance, à la mise en place de nombreux processus, la rédaction d'un nombre conséquent de rapports et l'élaboration minutieuse de calculs actuariels.

C'est dans ce cadre que la MAPA, société d'assurance mutuelle de taille moyenne, a développé son Service Actuariat. Le but était de répondre aux exigences réglementaires, c'est-à-dire aux calculs et à la conception des tableaux de bords du Pilier 1, les calculs prévisionnels et scénarios de l'ORSA, et la rédaction de tout ou partie de rapports (Rapport Régulier au Contrôleur, Rapport sur la Solvabilité et la situation financière, ORSA et rapport actuariel). Malgré le principe de proportionnalité — qui devrait être renforcé dans les années à venir¹ — des attentes d'une telle ampleur engendrent un travail de longue haleine sur des temps restreints, obligeant parfois d'émettre des hypothèses, qu'il faut ensuite remettre en question.

La fonction actuarielle est l'une des fonctions clés présentes au sein de la gouvernance SII. Et, elle couvre un large périmètre : le provisionnement, notamment dans le cadre du Pilier 1, mais aussi la Qualité des Données, la tarification des produits et l'adéquation de la structure de réassurance aux risques de la société. À l'analyse des textes, les exigences sont plus vastes que les éléments notifiés des remises régulières à l'ACPR. En cas de contrôle, la société, et notamment la fonction clé Actuariat, doit être en mesure de pouvoir justifier toutes les hypothèses utilisées, d'avoir des méthodes pérennes et documentées, et une traçabilité (piste d'audit) viable.

Or, ces travaux se rajoutent aux travaux qualifiés « d'obligatoires », et demandent un investissement en termes de ressources humaines, conséquentes. Ces ressources sont d'autant plus précieuses lorsqu'elles sont initialement déjà restreintes, avec des délais toujours plus courts. Mais, ces travaux sont nécessaires. Ils permettent de conforter les choix, les hypothèses qui sont prises, et également de mieux connaître les données et pouvoir en extraire des résultats pouvant amener l'entreprise à gagner en performance.

Ce mémoire repose donc sur la recherche de l'opportunité dans le cadre des exigences réglementaires, ou comment faire d'une contrainte une opportunité ?

Certaines hypothèses du cadre de l'évaluation des meilleures estimations de sinistres sont initialement basées sur des données marchés. Par nature, ces hypothèses ne sont donc pas justifiables par la MAPA aux yeux de l'ACPR. La justification de ces hypothèses sera donc l'entrée de cette étude.

Dans le cadre des sinistres Automobile, les deux lignes d'activité prévues dans le cadre de Solvabilité II sont la ligne « RC Automobile » et « Automobile autre que RC ». Les calculs appliqués en interne dans le cadre de la meilleure estimation distinguent les sinistres RC Automobile ne provoquant que des dommages matériels et les sinistres RC Automobile comprenant des dommages corporels. De plus, les sinistres RC Automobile dits « corporels », sont de nouveaux segmentés entre les sinistres graves et les sinistres attritionnels, suivant un seuil prédéfini. Ainsi, le Service Actuariat segmente les sinistres parfois plus finement que les recommandations des directives (basées sur les lignes d'activité). Ces segmentations additionnelles semblent cohérentes, mais méritent d'être justifiées.

¹ <https://www.argusdelassurance.com/juriscopes/analyses-d-experts/solvabilite-2-la-proportionnalite-un-enjeu-pour-les-assureurs-a-taille-humaine.191522>

L'objectif réglementaire est ici de prouver que les sinistres RC Automobile « matériels » ont bien un intérêt mathématique à être séparés des sinistres RC Automobile « corporels », et de manière équivalente pour la distinction entre les sinistres graves et les sinistres attritionnels. De plus, il convient de définir de manière théorique, le montant du seuil qui sépare les sinistres graves et les sinistres attritionnels.

Les méthodes utilisées pour justifier les choix utilisés par la MAPA reposeront sur les théories du provisionnement non-vie classique et sur la théorie des valeurs extrêmes. Pour la partie provisionnement non-vie, les résultats s'appuieront sur des concepts fondamentaux comme la méthode déterministe Chain Ladder, et sur les mesures de l'incertitude via les méthodes de Mack et du Bootstrap. La théorie des valeurs extrêmes sera utilisée pour la détermination du seuil qui définit le moment où les sinistres sont considérés comme graves. Cette première analyse, a priori, sera complétée d'une seconde analyse, a posteriori, afin de conforter des résultats pouvant faire objet à interprétation.

Qu'en est-il de l'opportunité ? Les sinistres RC Automobile « corporels » graves sont, par définition, rares et leur comportement dans le processus de paiement des indemnités est long et volatile. C'est d'ailleurs pour cette raison que ces sinistres sont séparés des sinistres attritionnels. L'application des méthodes déterministes classiques n'est pas satisfaisante, et ce mémoire reviendra sur ce point, montrant les limites de ces méthodes. De ce fait, les meilleures estimations de ces sinistres se font par actualisation simple des provisions dossiers/dossiers (selon une cadence de règlement prédéfinie), permettant de rester prudent. En effet, des études liées au rapport actuariel tendent à montrer que les provisions dossiers/dossiers sont prudentes et que des bonis sont régulièrement constatés.

Même si les sinistres graves sont souvent réassurés, une diminution des meilleures estimations nettes de réassurance permettrait d'améliorer le ratio de solvabilité, puisque des fonds propres SII pourraient être dégagés par différence entre les meilleures estimations et les provisions techniques comptables (au passif et à l'actif), et les risques liés au SCR pourraient être amoindris (et notamment sur le risque de souscription Non-Vie). La difficulté étant que les méthodes de provisionnement agrégées sont limitées. L'alternative récente est d'estimer les coûts totaux des sinistres avec les informations individuelles de chaque sinistre. La méthode de provisionnement ligne à ligne utilisée fonctionne via des méthodes d'arbres, comme la méthode CART.

La difficulté principale de cette estimation par méthode CART reposera sur le volume de données, lié au portefeuille restreint de la MAPA, malgré la prise en compte d'un historique relativement long (2003-2019). Des solutions seront alors proposées, pour améliorer la qualité des modèles, au détriment du périmètre de réponse de ce dernier. Les calculs liés à la Solvabilité de l'entreprise doivent tout d'abord être fiables et relativement prudents.

Enfin, une estimation simple de l'impact de ce changement de méthode de l'estimation des meilleures estimations des sinistres RC Automobile « corporels » graves sur la solvabilité de la société sur l'exercice 2019 sera réalisée.

Ce mémoire est finalement lui-même une opportunité pour la MAPA de progresser sur plusieurs aspects, aussi bien réglementaires, que techniques.

Chapitre 1 : De l'exigence réglementaire vers l'opportunité

I. Contexte général Solvabilité II

La directive européenne Solvabilité II est entrée en vigueur en janvier 2016, suite à la crise financière de 2008 et après des années de concertations, de construction de règles, fixant un régime de solvabilité applicable aux entreprises d'assurances de l'Union Européenne.

Solvabilité II est un règlement délégué et un règlement d'exécution voté en 2015 sur des normes précises et sur des modèles à utiliser.

Le premier objectif est donc d'avoir une réglementation commune permettant de guider les entreprises d'assurances vers une gestion plus adéquate de leurs fonds propres, en intégrant la notion de risque, maintenant omniprésente, et d'intégrer à toutes les décisions les répercussions directes ou indirectes qu'elles peuvent avoir sur la solvabilité de l'entreprise. Cela permettra par ce biais de couvrir les assurés, et d'éviter des catastrophes économiques.

La directive doit s'appliquer en tenant compte **du principe de proportionnalité** qui sera fonction de la nature et de l'ampleur des risques. Ce principe porte sur les 3 piliers et est vérifié par l'ACPR.

La directive Solvabilité II est donc présentée sous forme de trois piliers :



Figure 1 : Les trois piliers sous Solvabilité II²

A. Les exigences quantitatives

Les règles quantitatives sont les règles de valorisation des actifs et des passifs, ainsi que les exigences de capital et leur mode de calcul.

² Source : <https://www.insurancespeaker-wavestone.com/2014/03/solvabilite-2-enjeux-et-contraintes/>

1. Bilan prudentiel

C'est une approche économique, où il est notamment question de bilan prudentiel et non plus de bilan comptable. Les actifs et les passifs sont valorisés à leur valeur réelle, pour avoir une vision exacte de la santé financière de l'entreprise.

Cela passe par des placements financiers en valeur de marché, des actifs incorporels nuls, du calcul d'impôts différés ou encore des engagements sociaux calculés avec les normes IFRS.

La différence comptable-prudentiel repose aussi en grande partie dans le principe d'évaluation des provisions techniques. Les provisions techniques sont vues de manière prudente en comptabilité nationale, alors qu'elles sont vues de manière juste sous Solvabilité II. Il est alors fait référence de meilleures estimations (« Best Estimate »).

La meilleure estimation est la valeur actuelle probable des flux de trésorerie futurs, calculée avec la courbe des taux sans risque. Plus précisément, il y a deux types de meilleures estimations, sur les deux pans du bilan :

- la meilleure estimation au passif (« Best Estimate Liabilities » ou BEL) :
 - meilleures estimations des provisions pour sinistres, qui représentent l'ensemble des engagements de la société d'assurance liés aux sinistres survenus et non encore clôturés avant la date de la fin de l'exercice ;
 - meilleures estimations des provisions pour primes, qui représentent les engagements futurs liés à l'exercice suivant, soit la comptabilisation de la rentabilité future (la meilleure estimation étant négative lorsque la société est rentable, positive sinon) ;
- et parallèlement, la meilleure estimation à l'actif (« Best Estimate Assets » ou BEA), avec :
 - les meilleures estimations des provisions pour sinistres cédés ;
 - les meilleures estimations des provisions pour primes cédées, qui représentent la rentabilité liée à la couverture de réassurance, généralement défavorable.

Ce rapport reviendra plus en détail sur le calcul des meilleures estimations.

À cela s'ajoute une marge de risque (« Risk Margin ») qui est le coût additionnel que représente la couverture d'obligation d'assurance sur toute la durée de vie du portefeuille. Cette couche supplémentaire est donc le coût d'immobilisation des fonds propres jusqu'à extinction des engagements.

Les fonds propres S2 ne sont autres que la différence entre l'actif et le passif du bilan prudentiel, ou dit autrement, les fonds propres S1 additionnés à la différence supplémentaire entre l'actif et le passif prudentiel par rapport à la différence comptable initiale.

2. Le Capital de Solvabilité Requis (« Solvency Capital Requirement » ou SCR)

Le SCR est le montant de fonds propres à détenir pour limiter à 0,5% la probabilité de ruine à 1 an (choc bicentenaire). Tous les risques quantifiables sont pris en compte, les principaux étant généralement le risque de Marché et les risques de souscription.

C'est le profil de risque réel de l'entreprise, la formule standard étant la plus communément appliquée pour le calculer.

B. Les exigences qualitatives

Un système est un ensemble de procédés, de pratiques organisées, destiné à assurer une fonction définie (la fonction de la gouvernance pour le système de gouvernance). Le système comprend des procédures et des animateurs.

La définition de l'ACPR est : « le système de gouvernance comprend une structure organisationnelle transparente adéquate avec une répartition claire et une séparation appropriée des responsabilités, ainsi qu'un dispositif efficace de transmission des informations afin d'assurer une administration saine de l'entreprise d'assurance ».

Trois concepts forts ressortent de ce système de gouvernance :

- l'organe d'administration, de gestion et de contrôle (« Administrative, Management or Supervisory Body » ou AMSB) : c'est le Conseil d'Administration (CA) ;

Il détermine les orientations de l'activité ainsi que la stratégie générale de l'organisme d'assurance. C'est un organisme d'impulsion et de contrôle, approuve les politiques écrites, ou encore porte l'organisation ultime du système de gestion des risques ;

- les dirigeants effectifs : avec, au minimum, le Directeur Général (DG) et un Directeur Général Délégué (DGD), avec le principe des quatre yeux ;
- les fonctions clés.

Les fonctions clés (ou fonctions importantes et critiques) se définissent comme la capacité administrative à accomplir des tâches spécifiques de gouvernance. Elles sont au nombre de quatre :

- la gestion des risques : cette fonction a pour but de mettre en place le système de gestion des risques, et est donc en lien étroit avec la fonction actuarielle ;
- la conformité : cette fonction vérifie l'adéquation du système de contrôle interne et également des autres éléments du système de gouvernance ;
- l'audit interne : cette fonction veille à conseiller l'organe d'administration sur le respect de toutes les normes législatives réglementaires et administratives nécessaires à l'activité d'assurance ;
- l'actuariat : le but de la fonction est **l'évaluation de l'exactitude des calculs des provisions techniques**. Elle vérifie qu'ils sont justes et donne un avis sur la suffisance des provisions. Pour cela elle va mesurer les impacts des garanties des contrats sur les provisions techniques. Ce calcul de provision technique doit être fait d'une année sur l'autre et être comparé et expliqué au Conseil d'Administration via le rapport actuariel.

Ainsi que deux systèmes :

- le système de gestion des risques :

Sous-élément majeur du système de gouvernance, sa fonction est de déceler, mesurer, contrôler, gérer, et déclarer les risques. Les organismes d'assurance doivent mettre en œuvre et garder opérationnel un système de gestion des risques qui comprend une stratégie globale de gestion de risque cohérente avec la stratégie générale de l'organisme. C'est de ce système que découle le

dispositif d'évaluation interne des risques et de la solvabilité (« Own Risk and Solvency Assesment » ou ORSA).

- Le système de contrôle interne :

Ce système de contrôle interne est le garant du système de gouvernance. Il a pour objectif de procéder à une revue globale des processus, des contraintes, et des risques de l'organisme d'assurance pour améliorer son efficacité, sa productivité, ses performances.

C. Les exigences en matière de transmission d'information prudentielle et de publications

Le pilier III synthétise les obligations de reporting à l'ACPR, mais aussi les publications aux assurés et à toute personne ayant de l'intérêt sur la compagnie d'assurance (ex : une personne qui navigue sur le site internet). Ces exigences sont harmonisées au niveau européen.

On distingue ainsi deux types de transmission d'informations :

- les états quantitatifs :
 - les ENS (États Nationaux Spécifiques)
 - les QRT (« Quantitative Reporting Templates »), qui se distinguent entre Solo/Groupe ou Annuel/Trimestriel. Ces états sont nombreux et permettent d'avoir une image précise de l'ensemble de la société :

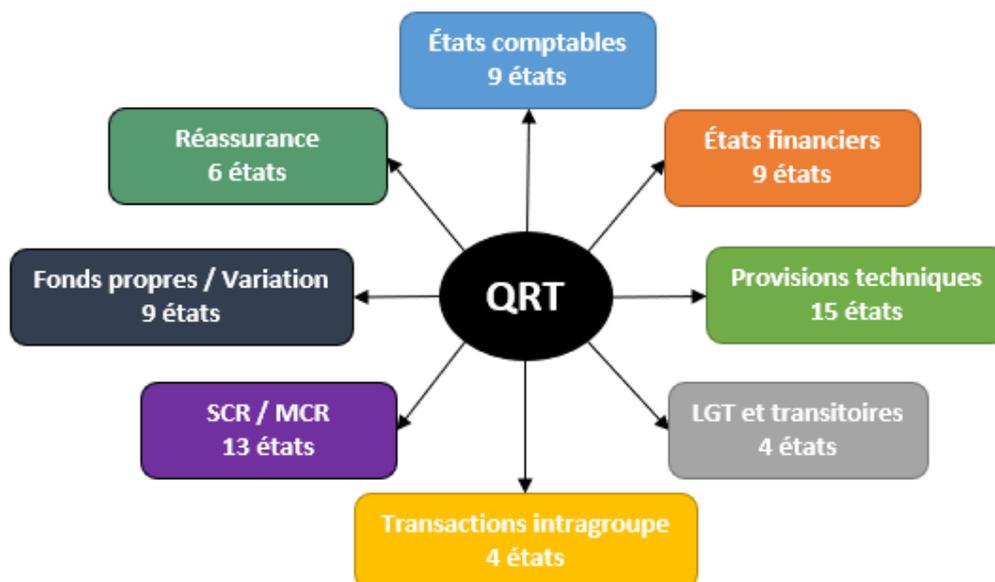


Figure 2 : Exemple de répartition des états de la remise annuelle Solo entre différents axes d'analyse

- les rapports narratifs (liste non exhaustive) :
 - le rapport sur la solvabilité et la situation financière (« Solvency and Financial Conditions Report » ou SFCR) : c'est un rapport public et annuel, qui peut être fait au niveau groupe ;
 - le rapport régulier au contrôleur (« Regular Supervisory Report » ou RSR) : rapport à destination de l'ACPR, pour chaque entité, à faire tous les trois ans. Ce rapport est plus complet que le SFCR ;

- l'ORSA : rapport annuel à destination de l'ACPR, qui peut être fait au niveau groupe ;
- le rapport actuariel : rapport annuel à disposition de l'ACPR sur demande, pour chaque entité.

D. La Qualité des données et provisions techniques

Avec des exigences à la fois sur les travaux quantitatifs et qualitatifs, Solvabilité II ne peut être une bonne mesure que si les informations exploitées ou transmises sont de bonne qualité.

Ce concept est en fait omniprésent dans le métier de l'Actuaire de manière générale, mais il est d'autant plus réglementé dans la partie Solvabilité II.

Les principaux textes liés à l'Actuariat dans le cadre de Solvabilité II sont ceux de la Directive 2009/138/CE du 25 novembre 2009 ou du règlement délégué 2015/35 du 10 octobre 2014 :

- Directive **article 48** (Fonction actuarielle), notamment aux points 1.c et 1.g :

c) « *Les entreprises d'assurance et de réassurance mettent en place une fonction actuarielle efficace afin d'apprécier la suffisance et la qualité des données utilisées dans le calcul des provisions techniques* »

g) « *Les entreprises d'assurance et de réassurance mettent en place une fonction actuarielle efficace afin d'émettre un avis sur la politique globale de souscription* »

ce dernier étant plus explicite dans le règlement délégué, dans l'**article 272** (Fonction actuarielle), notamment au point 6 :

« *En ce qui concerne la politique de souscription, l'avis que doit émettre la fonction actuarielle conformément à l'article 48, paragraphe 1, point g), de la directive 2009/138/CE contient, au minimum, des conclusions sur la **suffisance des primes à acquérir** pour couvrir les sinistres et dépenses à venir, compte tenu notamment des risques sous-jacents (y compris les risques de souscription), et l'impact des options et garanties prévues dans les contrats d'assurance et de réassurance sur la suffisance des primes ; et sur l'effet de l'inflation, du risque juridique, de **l'évolution de la composition du portefeuille** de l'entreprise et des systèmes ajustant à la hausse ou à la baisse les primes dues par les preneurs en fonction de leur historique de sinistres (systèmes de bonus-malus) ou de systèmes similaires, mis en œuvre au sein des différents **groupes de risques homogènes*** »

De part ces deux articles, la Qualité des données exigées concerne a minima :

- ✓ Les données d'évaluations des sinistres et des provisions ;
- ✓ Les données liées au profil de risque des affaires en portefeuille, au tarif, à la revalorisation ;
- ✓ Les données nécessaires au calcul du SCR et notamment des chocs bicentennaux ;
- ✓ Les données nécessaires à la gestion des risques et à l'établissement de l'ORSA.

Le périmètre réglementaire couvre donc l'ensemble des calculs Solvabilité II.

- Directive **article 82** (Qualité des données et application d'approximations, y compris par approches au cas par cas, pour les provisions techniques) :

« Les États membres veillent à ce que les entreprises d'assurance et de réassurance mettent en place des processus et procédures internes de nature à garantir le caractère **approprié, l'exhaustivité et l'exactitude des données** utilisées dans le calcul de leurs provisions techniques.

Lorsque, dans des circonstances particulières, [...], **des approximations adéquates, y compris par approches au cas par cas, peuvent être utilisées pour le calcul de la meilleure estimation.** »

Il y a donc trois critères obligatoires dans la Qualité des données, et des approximations adéquates sont possibles.

- Règlement Délégué **article 19** (Données utilisées dans le calcul des provisions techniques), dont voici des extraits ciblés :

« [...] les données incluent suffisamment d'informations **historiques** [...] »

« [...] des données sont disponibles pour chacun des groupes de **risques homogènes** [...] »

« [...] **le volume et la nature des données** sont propres à garantir que les estimations formulées sur leur fondement pour le calcul des provisions techniques ne sont pas entachées d'une erreur d'estimation importante [...] »

- Règlement Délégué **article 34** (Méthodes de calcul), dont voici un extrait ciblé :

« [...] des groupes **de risques homogènes**, qui reflètent de manière **appropriée** les risques de chacun des contrats inclus dans ces groupes [...] »

Il existe également d'autres textes sur la qualité des données ou sur l'exigence des calculs des provisions techniques, comme :

- les **Orientations sur la valorisation des provisions techniques** publiés par l'EIOPA en 2014 ;
- ou les Normes Professionnelles (NP) de l'Institut des Actuaire (IA), avec notamment la norme NPA3 « Manuel BEL Non-Vie », qui relate l'ensemble des textes Solvabilité II et préconise des méthodes de travail répondant au cadre juridique.

On notera pour terminer l'**article 84** de la Directive, qui récapitule les enjeux pour les entreprises d'appliquer ces mesures :

« Sur demande des autorités de contrôle, les entreprises d'assurance et de réassurance démontrent **le caractère approprié** du niveau de leurs provisions techniques, ainsi que **l'applicabilité et la pertinence des méthodes** qu'elles appliquent et **l'adéquation des données statistiques** sous-jacentes qu'elles utilisent. »

Il y a bien un lien fort entre la qualité des données et les méthodes de calcul pour les provisions techniques, qui utilisent la même terminologie pour décrire ce qui est attendu par l'ACPR. Des contrôles peuvent être effectués par l'Autorité de contrôle sur ces sujets, et il est de la responsabilité de la fonction clé Actuariat de prendre du recul sur les travaux effectués au sein de l'entreprise d'assurance et de procéder à une amélioration continue. Ces axes d'amélioration sont notamment relatés dans le rapport actuariel.

De plus, Solvabilité II doit être vu comme une opportunité et non une contrainte. Les entreprises d'assurance doivent savoir tirer profit de l'application des mesures.

II. La MAPA (Mutuelle d'Assurance des Professionnels de l'Alimentaire)

A. Présentation générale

La MAPA, Mutuelle d'Assurance des Professionnels de l'Alimentaire, est une société d'assurance mutuelle à cotisations variables régie par le code des assurances. Elle est spécialisée dans l'assurance non-vie des professions alimentaires. Elle propose aux artisans et commerçants indépendants et leurs salariés de cette branche des garanties d'assurance sur les risques suivants : automobile, responsabilité civile professionnelle et familiale, incendie, vol, dégâts des eaux, tempête, bris de glace, bris de machine, défense et recours, navigation de plaisance, individuelle accident, marchandises transportées, protection juridique, assistance, santé individuelle et collective.

En partenariat avec d'autres sociétés qui portent le risque d'assurance ou financier, la MAPA distribue auprès de ses sociétaires des produits d'épargne, d'assurance-vie, de prévoyance, de retraite et de prêts. Elle développe également des services avec des partenaires spécialisés inclus dans les contrats d'assurance afin de répondre au mieux aux besoins des sociétaires tels des réseaux d'opticiens, de dentistes ou de réparateurs automobiles.

La société d'assurance mutuelle dispose de 81 agences localisées en grande majorité dans les préfectures départementales. Ses produits sont vendus par un réseau de plus de 350 salariés. Son siège social est situé à Saint-Jean-d'Angély en Charente-Maritime, département où elle fut fondée en 1911, et regroupe environ 250 salariés.

Avec 171 millions d'euros de primes en 2019, la MAPA représente moins de 0,25% du marché global de l'assurance dommage des professionnels et des particuliers en France, ce qui fait d'elle un acteur de petite taille. En tant que société d'assurance mutuelle dédié aux professionnels, son profil de risque est simple. Elle dispose de 84 millions d'euros de fonds propres fin 2019.

B. Le rapport actuariel et lien avec la problématique

Depuis 2012, le Service Actuariat a été renforcé dans le but premier de répondre aux nouvelles exigences réglementaires imposées par la Directive Solvabilité II.

Le Service Actuariat a donc fait appel à un cabinet d'Actuaires pour pouvoir mettre en place ces mesures très contraignantes à l'échelle d'une société d'assurance mutuelle à taille humaine. Certains axes, ou certaines décisions ont été prises à partir de données marchés, c'est notamment le cas pour les calculs sur les provisions techniques avec :

- la distinction entre les sinistres graves et attritionnels ;
- les groupes de risques homogènes ;
- la méthode de calcul des meilleures estimations sur les sinistres graves.

Or, par manque de temps ou de moyen, ces hypothèses n'ont jamais été confrontées.

Le rapport actuariel de la MAPA, écrit par la fonction clé Actuariat, recense alors les points d'amélioration suivants sur la partie Provisions techniques :

- le calcul est réalisé par groupes de risques homogènes, parfois plus détaillés que ceux préconisés par la réglementation, pour mieux évaluer les différentes provisions, mais certains tests de sensibilité pourraient tout de même être menés pour s'assurer de la bonne adéquation des méthodes de segmentation choisies ;
- vérifier le caractère approprié du seuil et le justifier entre les sinistres graves et les sinistres attritionnels ;
- certaines estimations sont directement issues des valorisations en normes françaises sur lesquelles une surévaluation a été constatée. Une étude pourrait donc être réalisée, pour déterminer une méthodologie qui serait plus adaptée, applicable au portefeuille MAPA et permettrait une définition plus sereine du montant des provisions prudentielles, aussi bien à l'actif qu'au passif.

Ces objectifs rentrent dans le cadre réglementaire à respecter, et qui pourrait faire l'objet d'un contrôle de l'ACPR.

C. L'exemple de l'Automobile

La remise en cause et l'amélioration des calculs liés aux provisions techniques s'adaptent bien à la branche Automobile qui coche l'ensemble des problématiques. Le but de ce mémoire est de mettre au point une méthodologie applicable à l'ensemble des lignes d'activité de la MAPA.

En effet, les calculs sur les provisions techniques de la branche Automobile sont appliqués aux groupes de risques homogènes suivants :

- La ligne d'activité Responsabilité Civile Automobile (RCA), détaillée en deux sous-groupes :
 - les sinistres corporels ;
 - les sinistres matériels ;
- la ligne d'activité Automobile Autres garanties.

Ensuite, sur le groupe de risques homogène RCA des sinistres Corporels, il y a une séparation entre les sinistres graves et attritionnels, selon un seuil prédéfini de 275 000€.

La meilleure estimation des sinistres du groupe de risques homogènes RCA « Corporels » graves est évaluée en reprenant les provisions dossiers/dossiers et en les actualisant via la courbe des taux sans risque suivant une cadence de règlement défini par un des courtiers de réassurance de la MAPA.

Les thèmes d'amélioration continue de l'évaluation des provisions techniques, que sont :

- la séparation en groupe de risques homogènes ;
- la définition d'un seuil entre sinistres graves et attritionnels ;
- la méthode d'évaluation des meilleures estimations des sinistres graves.

sont donc bien représentés par la branche Automobile de la MAPA.

III. La branche Automobile

A. Évolution du portefeuille et rentabilité

Le portefeuille Automobile de la MAPA est constitué, en 2019, d'environ 100 000 véhicules, pour un chiffre d'affaires de près de 47 millions d'euros. La MAPA est donc un acteur mineur sur le marché de l'assurance Automobile puisqu'elle représente 0,2% du marché français, aussi bien en nombre de véhicules qu'en primes.

La branche Automobile de la MAPA représente approximativement un tiers des cotisations encaissées sur l'exercice 2019. Historiquement, le parc Automobile était cependant plus important :

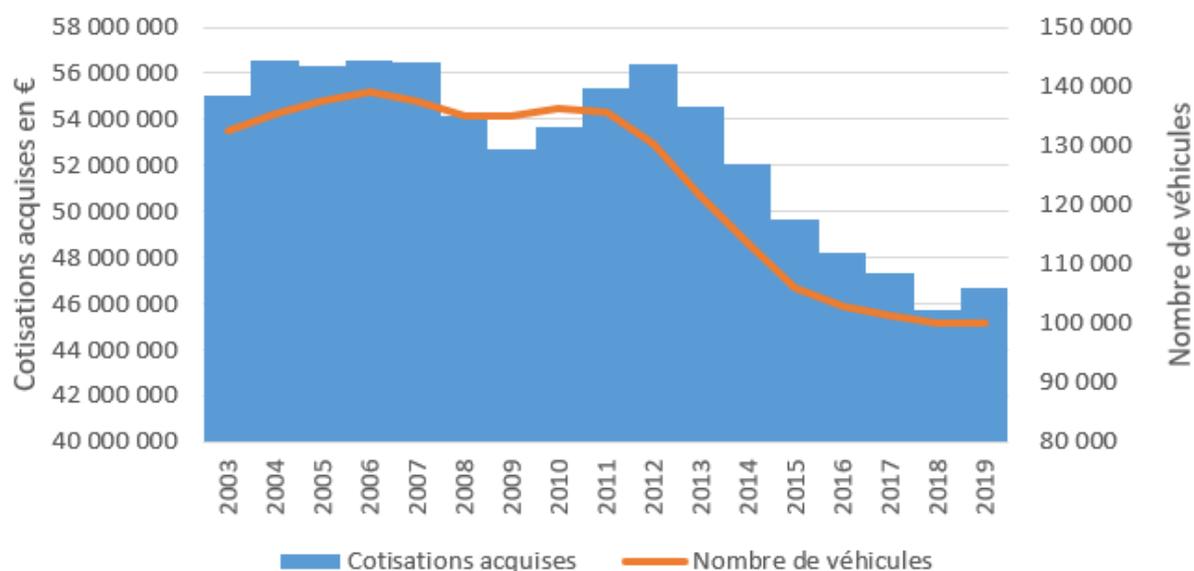


Figure 3 : Évolution de la production Automobile en cotisations et nombre de véhicules

À partir de 2012, le nombre de véhicules assurés a baissé de près de 25%, pour finalement se stabiliser en 2018-2019. Les cotisations ont, elles, diminué proportionnellement moins : -15%.

La raison est liée à la rentabilité du produit :

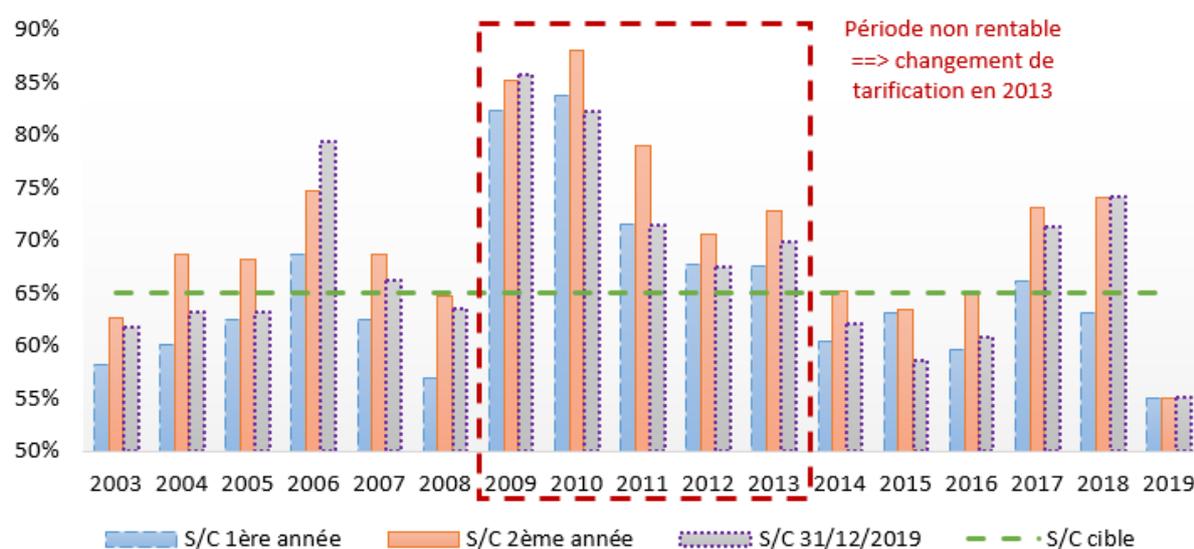


Figure 4 : Évolution du S/C en fonction de la date de vue

Le produit a longtemps été rentable (2003-2008). Mais à partir de 2009, plusieurs exercices ont successivement fragilisé la rentabilité de la société d'assurance mutuelle. De ce constat, la tarification de la branche a été complètement revue. Cela a entraîné une purge du portefeuille, écartant les risques sous-tarifés.

Cela a donc eu deux conséquences importantes :

- Le portefeuille a souffert en termes de chiffre d'affaires, le produit ayant perdu en attractivité sur un marché Automobile ultra-concurrentiel ;
- L'objectif premier a été atteint : le portefeuille a été assaini et la rentabilité est revenue au niveau souhaité.

La rentabilité est assez différente selon le recul pris sur les sinistres, et le montant des évaluations dossiers/dossiers. Il y a en moyenne +5,6 points entre le S/C vu à la fin de l'année de survenance et le S/C vu fin de l'année suivante. La MAPA, hors provisions pour tardif, enregistre de forts malis entre ces deux périodes.

À l'inverse, des bonis sont visibles sur la plupart des années de survenance entre la vision au 31/12/N+1 et la vision actuelle des sinistres. Cette vision plus ou moins lointaine, en fonction des années de survenance, est la vision la plus juste de l'évaluation des sinistres, mais elle reste incertaine.

Ce point crucial sera de nouveau traité par la suite.

Un autre aspect peut changer l'allure de la rentabilité de ce produit : la réassurance.

B. Réassurance et rentabilité

La réassurance Automobile de la MAPA a été très longtemps basée sur une couverture des sinistres graves par des « Excédent de sinistre » (XS), à des montants relativement bas. Cette approche semble tout à fait justifiée au vu de la taille du portefeuille MAPA. En effet, plus la taille du portefeuille est importante, plus une partie du risque peut être absorbée par l'entreprise d'assurance, par mutualisation avec le reste du portefeuille.

Cependant, la MAPA a totalement remis à plat son programme de réassurance en 2016, avec un système de réassurance innovant, basé sur un aggregate multi-branche, couvrant la sinistralité attritionnelle sur la Responsabilité Civile (RC) et le Dommage Aux Biens (DAB) de plusieurs branches d'assurance (Automobile, Habitation, Commerce). Ce dernier ne sera pas traité dans ce mémoire, mais il explique la rupture suivante, visible sur l'historique de réassurance de la MAPA :

Année	Priorité T1	Limite (priorité + portée)				AAD
		T1	T2	T3	T4	
2003	460 000	2 290 000	6 100 000	10 000 000 000	10 000 000 000	423 400
2004	550 000	2 290 000	6 100 000	10 000 000 000	10 000 000 000	1 137 300
2005	550 000	2 290 000	6 100 000	10 000 000 000	10 000 000 000	1 064 300
2006	550 000	2 290 000	6 100 000	10 000 000 000	10 000 000 000	1 077 400
2007	550 000	2 290 000	6 100 000	10 000 000 000	10 000 000 000	1 000 000
2008	700 000	2 290 000	6 100 000	10 000 000 000	10 000 000 000	1 000 000
2009	700 000	3 000 000	4 000 000	10 000 000 000	10 000 000 000	1 000 000
2010	1 000 000	4 000 000	8 000 000	10 000 000 000	10 000 000 000	1 000 000
2011	700 000	4 000 000	8 000 000	10 000 000 000	10 000 000 000	-
2012	700 000	4 000 000	8 000 000	10 000 000 000	10 000 000 000	-
2013	200 000	1 000 000	4 000 000	9 000 000	10 000 000 000	2 400 000
2014	200 000	1 000 000	4 000 000	9 000 000	10 000 000 000	2 400 000
2015	1 000 000	4 000 000	9 000 000	10 000 000 000	10 000 000 000	-
2016	4 000 000	9 000 000	10 000 000 000	10 000 000 000	10 000 000 000	-
2017	4 000 000	9 000 000	10 000 000 000	10 000 000 000	10 000 000 000	-
2018	4 000 000	9 000 000	10 000 000 000	10 000 000 000	10 000 000 000	-
2019	4 000 000	9 000 000	10 000 000 000	10 000 000 000	10 000 000 000	-

Tableau 1 : Programme de réassurance Automobile entre 2003 et 2019

Avant 2016, la MAPA était donc généralement couverte pour tous les sinistres dépassant les seuils de la colonne « Priorité 1 », allant du sinistre à 200K€ en 2013-2014 à 1M€ en 2010 et 2015, à hauteur de ce qui dépassait ce seuil. Il fallait ensuite prendre en compte la franchise (AAD pour *Annual Aggregate Deductible*).

La prise en compte de la sinistralité cédée, ainsi que des cotisations cédées pour ces traités de réassurance, permet d'obtenir une comparaison entre la rentabilité brute et nette de réassurance :

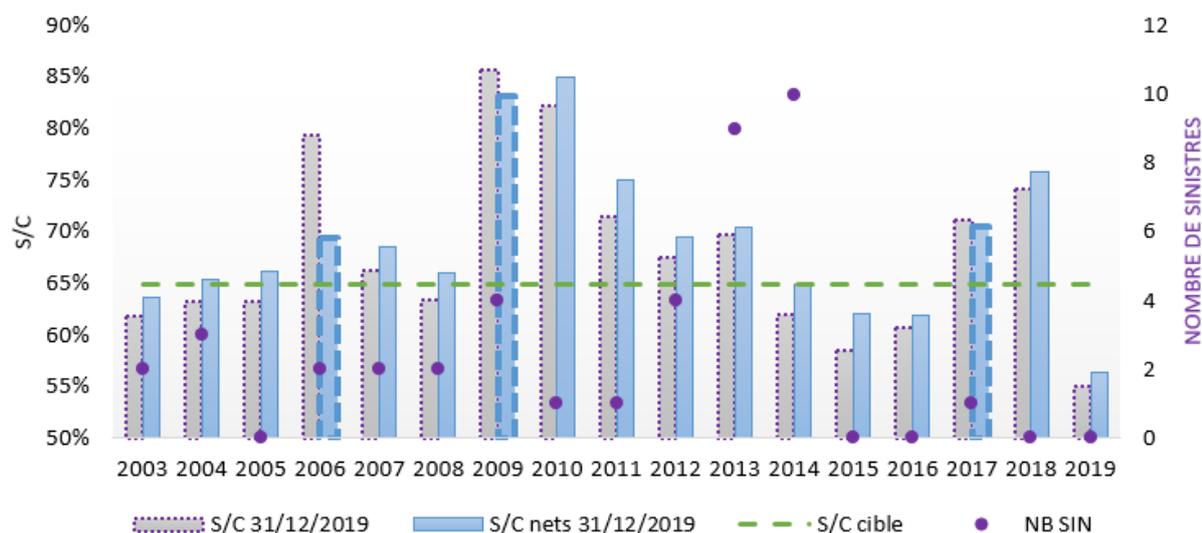


Figure 5 : Comparaison du S/C nets et bruts de réassurance entre 2003 et 2019

Les points en violet sont le nombre de sinistres entrant dans les XS de chaque année de survenance concernée, en considérant la vision actuelle de la sinistralité (31/12/2019).

Les années 2013-2014, dont la priorité des traités en XS était particulièrement basse (200K€), ont donc enregistré près de 10 sinistres cédés annuel, contre en moyenne 2 sur les années 2003-2012 et 2015. Depuis la nouvelle structure de réassurance, et la priorité fixée à 4M€, seul 1 sinistre a été enregistré en réassurance en 4 années.

Sur un historique de 17 années, seules 3 années dégagent un profit sur la couverture de réassurance (2006, 2009 et 2017), et en moyenne, la réassurance automobile en XS coûte à la MAPA plus de 1 point de S/C, soit 1,3M€ par an.

Enfin, la période non rentable identifiée en 2009-2013 reste, malgré la couverture de réassurance, une phase sur-sinistrée, à laquelle la MAPA devait réagir. La problématique de cette période était donc bien une problématique de fréquence et de mauvais risques, et non pas une problématique de sinistres de pointe, comme l'a clairement été l'année de survenance 2006.

Il est clair que la sinistralité Automobile comporte des aléas assez forts pour déstabiliser la rentabilité d'une entreprise d'assurance. Il est clair également que toutes les garanties ne sont pas concernées par ces sinistres graves. Et surtout, il est clair que le recul pris sur l'évaluation dossiers/dossiers des sinistres est déterminant pour connaître l'engagement réel de la MAPA vis-à-vis de ses assurés.

C. Solvabilité II et meilleures estimations

Le portefeuille Automobile a toujours eu un poids d'au moins un tiers du chiffre d'affaires de la MAPA. De plus, cette branche génère des sinistres qui peuvent être particulièrement importants et longs.

De ce fait, et malgré une rentabilité plutôt équilibrée sur le long terme, le poids des engagements Automobile pèsent tout particulièrement sur la société d'assurance mutuelle : près de 50% des provisions du bilan comptable normes françaises.

L'évaluation des meilleures estimations des provisions techniques sous Solvabilité II est donc un enjeu sensible, puisque les meilleures estimations jouent à la fois sur :

- le montant du capital de solvabilité requis (SCR) via notamment les risques de souscriptions ;
- le bilan prudentiel et donc sur les fonds propres SII, mais aussi sur les impôts différés, directement soustraits au SCR (capacité d'absorption des impôts différés).

Actuellement, les meilleures estimations de sinistres sur la branche Automobile, hors Rentes Automobile (risque lié à la Vie), sont calculées de la manière suivante :

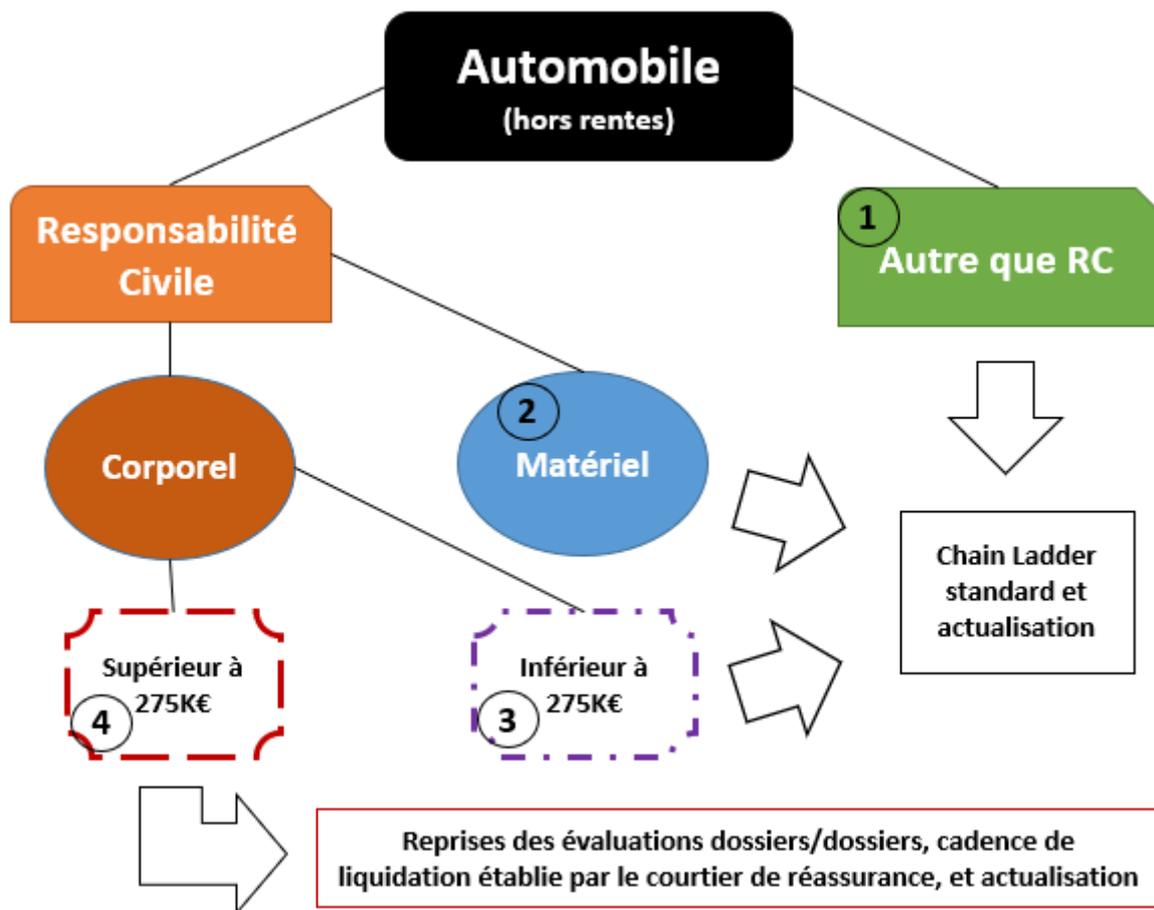


Figure 6 : Schéma des méthodes d'évaluation des provisions en fonction du périmètre

Les engagements de la branche Automobile à la MAPA sont donc évalués en cumulant les meilleures estimations de quatre segments indépendants. Sur trois d'entre eux, la méthode utilisée pour le calcul de la meilleure estimation est basée sur la méthode déterministe standard Chain Ladder. Pour chacun de ces segments, les historiques diffèrent et les années de développement sont plus ou moins longs.

Parmi ces quatre segments, seule la meilleure estimation de la Responsabilité Civile Automobile des sinistres corporels de plus de 275 000€ (4) n'est pas calculée avec Chain Ladder. Par manque de temps, il a été choisi de garder les évaluations des sinistres dossiers/dossiers considérées comme prudentes. L'actualisation se fait sur les flux de paiements futurs, qui sont déterminés grâce à une cadence de règlement fournie par le courtier en réassurance de la MAPA.

La segmentation RCA et Autre que RC est la segmentation minimale imposée par Solvabilité II sur la branche Automobile. En revanche, les autres segmentations, qui ont très certainement du sens, n'ont pas fait l'objet d'étude permettant de démontrer leur utilité. D'autres segmentations pourraient également avoir du sens pour évaluer au mieux les meilleures estimations.

D. Sinistralité automobile et segmentation

1. Sinistralité globale

Il est essentiel de bien définir les groupes de risques homogènes pour pouvoir évaluer correctement les engagements que la MAPA porte envers ses sociétaires.

Pour cela, il faut commencer par analyser statistiquement la sinistralité sur la période étudiée 2003-2019. Sur cette période, la MAPA a enregistré près de 437 454 sinistres en 17 années, soit plus de 25 000 sinistres par an.

Pour chaque sinistre, plusieurs garanties peuvent être impactées. Le cas le plus constaté en Automobile est le sinistre avec une tierce personne, qui peut faire intervenir à la fois la garantie Responsabilité Civile (RC) et la garantie Tous Risques. De ce fait, le nombre de garanties appelées, sur la même période, a été de 499 519. Il y a finalement :

- 380 505 sinistres qui ne font intervenir qu'une seule garantie, **soit 87% des sinistres** ;
- 56 959 sinistres font donc intervenir au moins 2 garanties.

À noter qu'une partie de ces sinistres sont finalement clôturés à 0€, et qu'ils représentent 18 000 sinistres environ soit un peu plus de **4%** de l'ensemble (au 31/12/2019). Ces sinistres font partie intégrante de cette étude, puisqu'ils n'ont pas toujours été nuls, et que leur évolution est à prendre en compte.

2. Détails des sinistres

Caractérisation

Solvabilité II contraint les entreprises d'assurance à distinguer en Automobile les garanties de Responsabilité Civile et les autres garanties. Ces deux catégories étant appelée Ligne d'activités (« Line of Business » ou LoB).

Pour qu'il y ait responsabilité civile, il faut que le sinistre concerne deux parties : l'assuré et un tiers. La responsabilité civile couvre les dommages consécutifs, **aussi bien matériels que corporels**, causés par une personne envers une seconde. Cela fait donc jouer la notion de **responsabilité**.

Aparté sur les bases de données :

Dans les bases de données MAPA, il y a bien la notion de « Responsabilité », mais cette notion n'implique pas forcément que la MAPA règle les sinistres dont ses sociétaires sont responsables, et qu'elle demande à la compagnie adverse de régler (recours) lorsque le sociétaire de la MAPA n'est pas responsable. En effet, cette notion est assez complexe en assurance Automobile en France depuis la loi Badinter du 5 juillet 1985, dans le sens où certaines situations imposent aux compagnies d'assurance de régler l'ensemble des dommages matériels et corporels, des deux parties, même lorsque l'assuré de ces dernières ne sont pas responsables de l'accident. C'est notamment le cas pour des accidents impliquant des piétons ou des cyclistes avec des véhicules à moteur : c'est la compagnie d'assurance de l'automobiliste qui indemniserait l'ensemble des préjudices, même si la faute (la responsabilité) de l'accident revient au piéton/cycliste (sauf exceptions, non développées dans ce mémoire). Les données peuvent donc être mal interprétées si celui qui les manipule ne prend pas le temps de bien les comprendre.

Enfin, un assuré peut se couvrir pour tout type de dommages, aussi bien matériels que corporels, qu'il soit responsable d'un accident avec un tiers, ou qu'il soit seul.

Les informations disponibles sur les sinistres sont les suivantes :

- LoB : RC et Autre que RC ;
- Sinistre seul, ou avec un tiers ;

- Sinistre comportant du Corporel ; ou intervention de la garantie « Corporel » de l'assuré (appelé ici « Garantie du conducteur »)

Type sinistre	Sinistre avec un tiers				
Nombre de sinistres	168 793				
Ligne d'activité	Automobile autre que RC			RC Automobile	
Type garantie	Garanties Matériel sans corporel	Garanties Matériel avec corporel	Garantie du conducteur	Garantie RC sans corporel	Garantie RC avec corporel
Coût total	72 433 223	29 846 694	10 255 625	93 657 784	153 714 054
Nombre	92 921	11 444	761	98 109	18 519
Coût moyen	780	2 608	13 477	955	8 300
Max	63 026	103 745	1 003 022	517 952	5 459 375
Ecart-type	1 892	4 577	76 924	2 261	82 053

Type sinistre	Sinistre sociétaire seul			Ensemble
Nombre de sinistres	268 661			437 454
Ligne d'activité	Automobile autre que RC			Automobile
Type garantie	Garanties Matériel sans corporel	Garanties Matériel avec corporel	Garantie du conducteur	
Coût total	250 175 495	311 747	574 965	610 969 587
Nombre	277 581	81	103	499 519
Coût moyen	901	3 849	5 582	1 223
Max	129 695	25 109	183 581	5 459 375
Ecart-type	2 098	4 094	22 220	16 295

Tableau 2 : Segmentation des données des sinistres Automobile, en €, de 2003 à 2019, vision à fin 2019

Rien qu'avec ces seuls éléments, il est possible de distinguer 8 segments qui pourraient potentiellement se comporter de manière différente dans l'évolution de l'évaluation des sinistres. D'ores et déjà, de grandes disparités en termes de coût moyen ou d'écart-type sont visibles, surtout entre les sinistres comportant des dommages corporels et les autres.

Il est possible de doubler ce nombre de segments en ajoutant la notion de responsabilité, entre Responsable de l'accident (même partiellement) ou non responsable. Mais comme indiqué plus haut, cela risquerait de biaiser certaines des estimations ou conclusions. En effet, la plupart des accidents où le sociétaire MAPA n'est pas responsable sont donc « indemnisés » par la compagnie adverse (via un recours), mais il arrive que la responsabilité, de par la loi Badinter, soit un élément « trompeur ».

Par prudence, et sans analyses complémentaires, la variable indiquant la responsabilité ne sera pas retenue.

D'autres éléments peuvent néanmoins être intéressants à travailler dans la bonne évaluation des sinistres automobile (Cf partie I.D.3).

Évolution du coût moyen

Après avoir analysé l'évolution de la rentabilité, il est intéressant de voir l'impact du temps sur l'évolution des coûts moyens. Cette étude comporte évidemment une limite sous-jacente au sujet de ce mémoire, puisque les coûts à l'ultime ne sont pas connus. Le choix est ici fait de prendre la dernière vision connue de la sinistralité, soit le 31/12/2019.

Les coûts moyens des sinistres Automobile évoluent de manière différente entre :

- Les préjudices corporels, qui évoluent en fonction de :
 - certains taux comme l'inflation, la courbe des taux sans risque, l'indice des prix à la consommation ;
 - la prise en charge de nouveaux postes de préjudices, de nouvelles infrastructures ;
 - l'évolution des salaires, et notamment des métiers de l'assistance à la personne.

Il est assez complexe de trouver une estimation correcte de l'évolution des préjudices corporels.

- Les indemnités matérielles, donc liées en grande partie aux véhicules (assurés et/ou adverses). L'association SRA (Sécurité et Réparation Automobiles) diffuse des évolutions des coûts liés aux réparations automobiles, dont les trois éléments majeurs sont :
 - L'évolution des coûts des pièces Automobile, qui joue en moyenne pour 50% de l'évolution du coût moyen global ;
 - L'évolution du coût de la main d'œuvre, qui joue en moyenne pour 40% de l'évolution du coût moyen global ;
 - L'évolution du coût de la peinture, qui joue en moyenne pour 10% de l'évolution du coût moyen global.

Tous ces éléments expliquent aisément que, naturellement, le coût moyen des sinistres Automobile augmente régulièrement d'une année sur l'autre. Et cela n'échappe pas à ce que constate la MAPA :

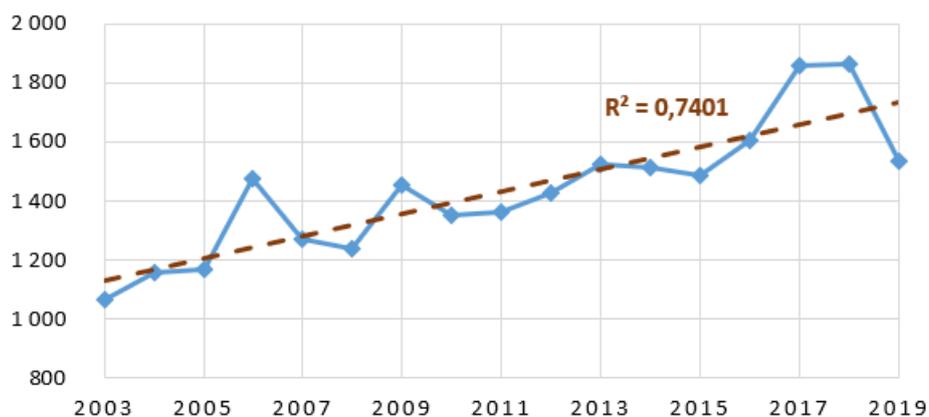


Figure 7 : Évolution du coût moyen Automobile

Depuis 2003, le coût moyen des sinistres Automobile à la MAPA augmente de manière assez régulière, en **moyenne de 2,7% par an**, passant d'environ 1 130€ par sinistre en 2003-2005 à 1 750€ par sinistre en 2017-2019.

Cette évolution peut avoir une incidence directe sur l'évaluation des sinistres dans le temps, lorsque ceux-ci ne sont pas clôturés rapidement.

Comme indiqué plus haut, les facteurs d'évolution du coût moyen des sinistres sont différents entre les préjudices corporels et les indemnités matérielles. Ceci se ressent sur l'évolution constatée par la MAPA, en prenant en compte également la notion de LoB :

- Pour les sinistres n'impliquant pas de dommages corporels

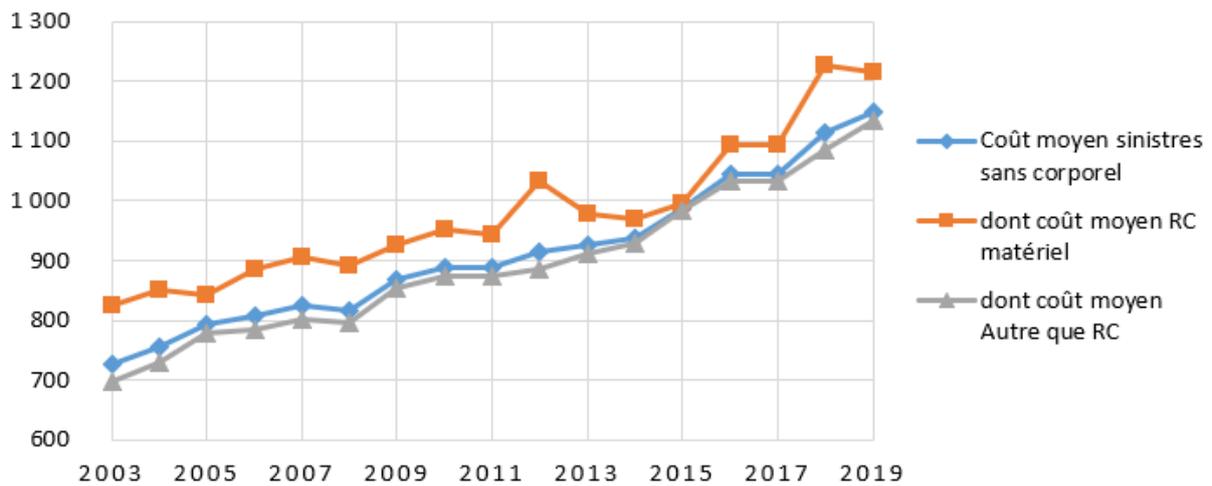


Figure 8 : Évolution du coût moyen des sinistres Automobile sans dommages corporels

Parmi les sinistres n'impliquant pas de dommages corporels, les coûts moyens et leur évolution sont très similaires. Il n'y a pas de différence significative entre ce qui est rattaché à la RC, et ce qui ne l'est pas. Il n'y a forcément que du Dommage Aux Biens (DAB), même pour la RC, et cela n'entraîne donc pas de volatilité importante. **L'évolution du coût moyen sur les sinistres n'impliquant pas de dommages corporels est de 2,7% par an en moyenne.**

- Pour les sinistres impliquant des dommages corporels : deux catégories de sinistres ressortent de manière évidente :
 - Les sinistres Autre que RC (hors Garantie du Conducteur, donc tous apparentés à du DAB)

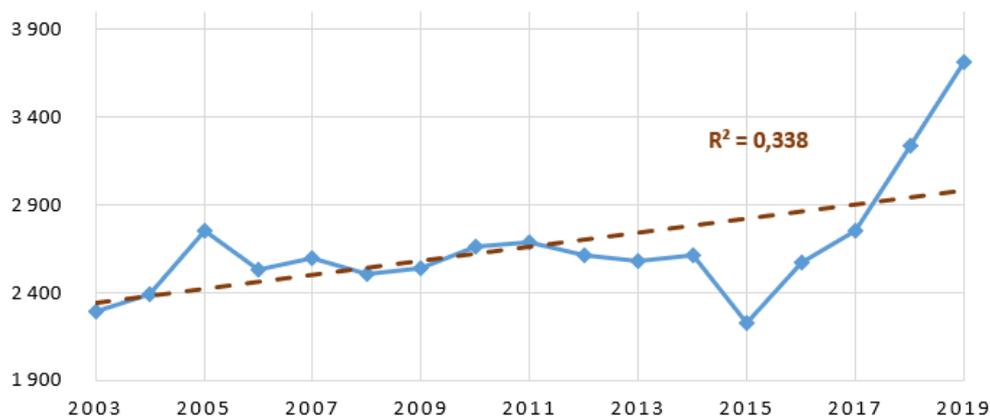


Figure 9 : Évolution des coûts moyens Automobile Autre que RC lors d'un accident impliquant du « corporel »

L'évolution du coût moyen des sinistres DAB hors RC est moins « évidente » que celle où il n'y a pas d'implications de préjudices corporels. Les coûts moyens des années 2018-2019 sont particulièrement élevés, laissant supposer que l'écoulement des sinistres est plus long, et qu'il a un impact assez fort.

En enlevant ces deux dernières années, l'évolution moyenne du coût moyen de ces sinistres est alors de **0,3% par an**, soit bien moins importante que celle des sinistres n'impliquant pas de dommages corporels.

- Les sinistres RC (auxquels sont ajoutés les sinistres Garantie du conducteur, peu nombreux, et assimilés à des préjudices corporels, comme peut l'être la RC)

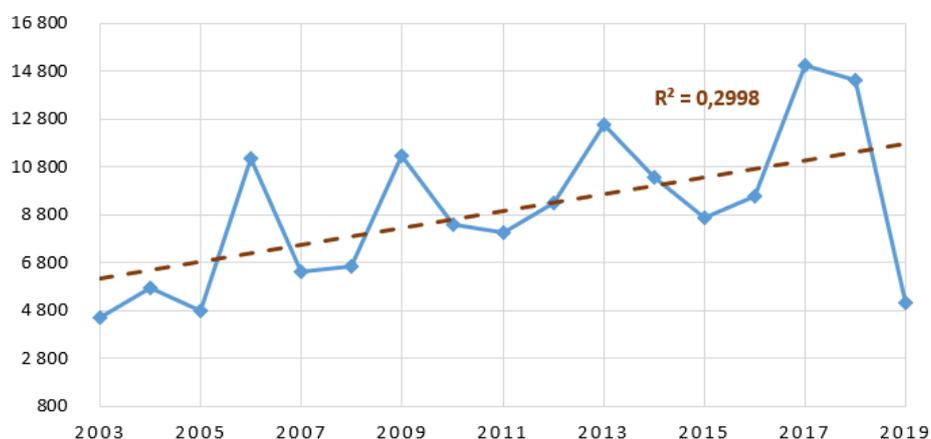


Figure 10 : Évolution des coûts moyens Automobile RC « Corporel » et Garantie du conducteur

Les sinistres RC avec des dommages corporels sont les plus volatils, mais une tendance plutôt nette se dégage de ces sinistres. L'évolution moyenne du coût moyen est **de 4,1% par an**. À noter que ces sinistres concernent à la fois des indemnités matérielles, et des indemnités de préjudices corporels.

Enfin, il est à noter que la Garantie du Conducteur, analysée seule, montre une évolution de près de **10%** en moyenne par an, mais qu'il est difficile de conclure lorsque cela repose sur moins de 100 sinistres par an.

3. Autres éléments attachés à la sinistralité

Parmi les éléments qui peuvent expliquer l'évaluation ou l'évolution du coût d'un sinistre Automobile, il y a les caractéristiques du contrat.

Une des principales particularités de la MAPA par rapport au marché de l'assurance, c'est sa spécialisation dans l'assurance pour les professionnels. De ce fait, le portefeuille professionnel Automobile de la MAPA représente près du tiers du nombre de véhicules total assuré. La question peut alors se poser sur la différence qu'occasionne cette séparation entre les véhicules privés et professionnels sur l'évaluation ou l'évolution de la sinistralité. Le premier élément à observer est que les coûts moyens des sinistres des contrats pour les professionnels sont 10% plus élevés, que pour les sinistres sur contrats privés.

On peut également ajouter les variables classiques de tarification Automobile :

- L'âge du conducteur principal et/ou l'ancienneté de permis (qui sont des variables extrêmement corrélées) ;

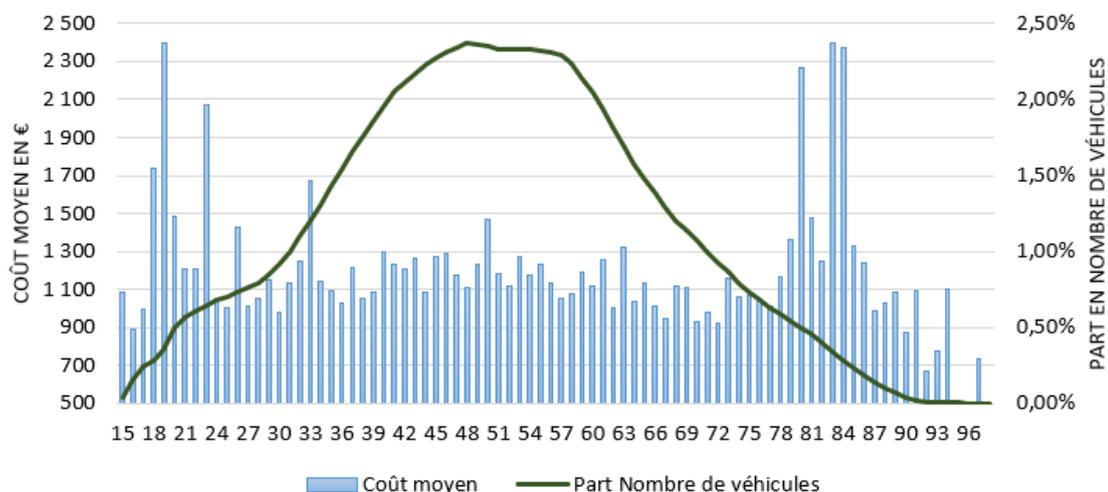


Figure 11 : Part du portefeuille et coût moyen en fonction de l'âge du conducteur principal

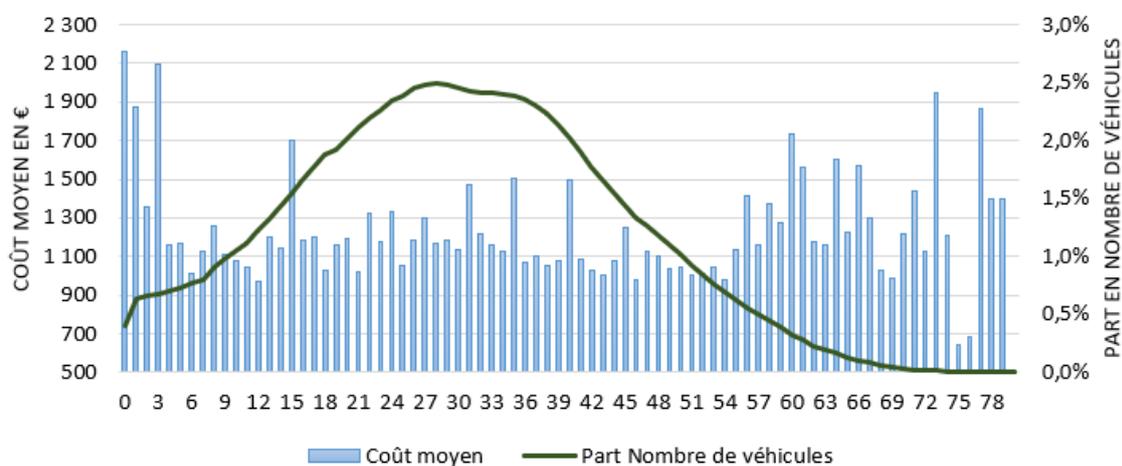


Figure 12 : Part du portefeuille et coût moyen en fonction de l'ancienneté de permis du conducteur

Il est difficile d'affirmer que les coûts moyens des sinistres sont différents en fonction de l'âge ou de l'ancienneté de permis du conducteur. Il est tout de même possible de voir une certaine instabilité sur les deux extrêmes, qui fait tout de même penser que les comportements, ou les véhicules, pour les assurés de moins de 25 ans et de plus de 55 ans entraînent des coûts moyens plus importants.

- Le genre de véhicule :

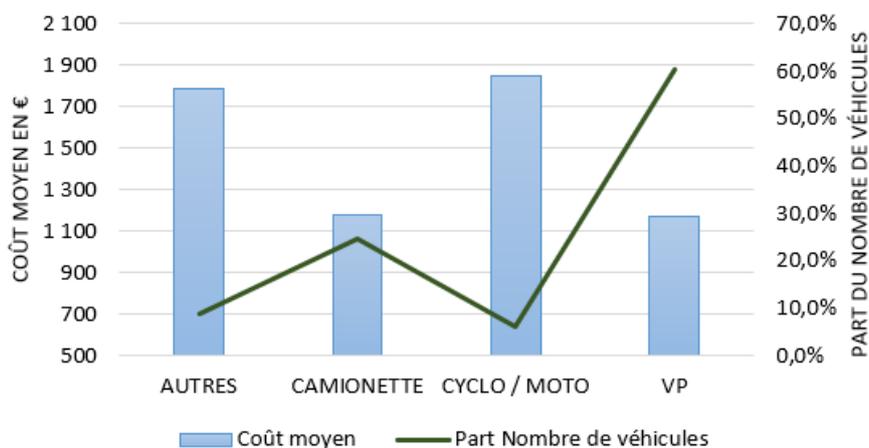


Figure 13 : Part du portefeuille et coût moyen en fonction du genre du véhicule

Les Véhicules de Plaisance (VP) représentent près de 60% du portefeuille, et le coût moyen des sinistres qui y sont liés ne sont pas réellement différents de celui des camionnettes (25% du portefeuille).

Les cyclomoteurs, les motocyclettes, et les autres véhicules (camions, remorques et engins) ont quant à eux des coûts moyens bien plus importants.

- L'ancienneté du véhicule ;

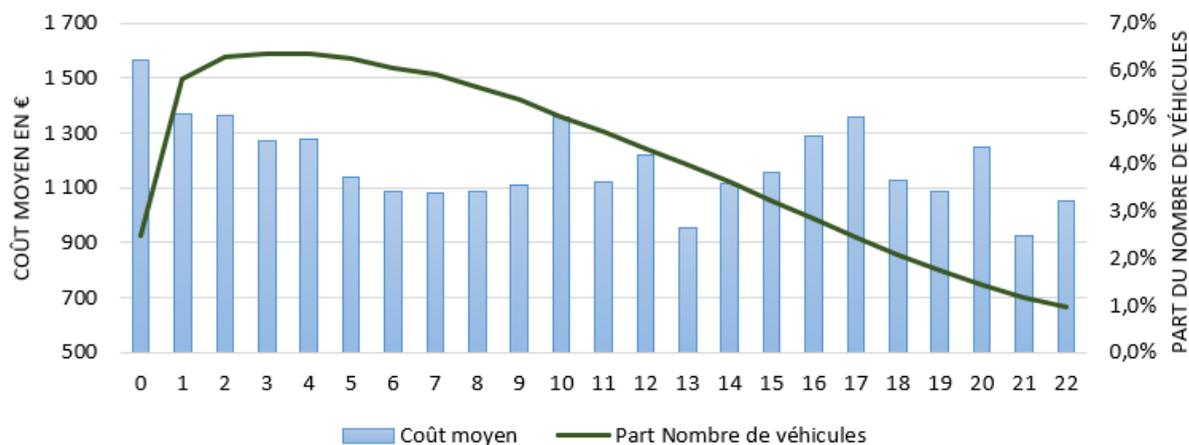


Figure 14 : Part du portefeuille et coût moyen en fonction de l'âge du véhicule

L'ancienneté du véhicule joue sur le coût moyen mais pas de manière linéaire. Une tendance claire se distingue sur les 7 premières années de vie du véhicule, mais le coût moyen ne baisse plus, voir est un peu plus élevé sur les années suivantes.

- La puissance du véhicule ;

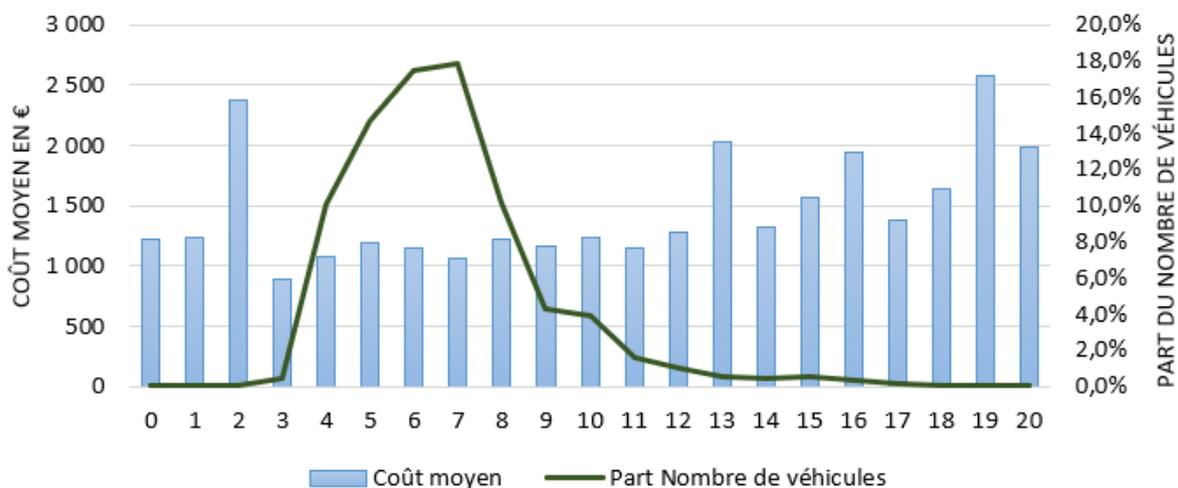


Figure 15 : Part du portefeuille et coût moyen en fonction de la puissance du véhicule

La puissance du véhicule est une variable qui joue sur le coût moyen, avec un coût moyen qui augmente avec la puissance à partir de 3 chevaux fiscaux. Avant les 3 chevaux fiscaux, les véhicules concernés ne sont pas des voitures au sens classique du terme, mais des voiturettes, des engins de type transpalette, ou des remorques.

- Les groupes et classes prix SRA

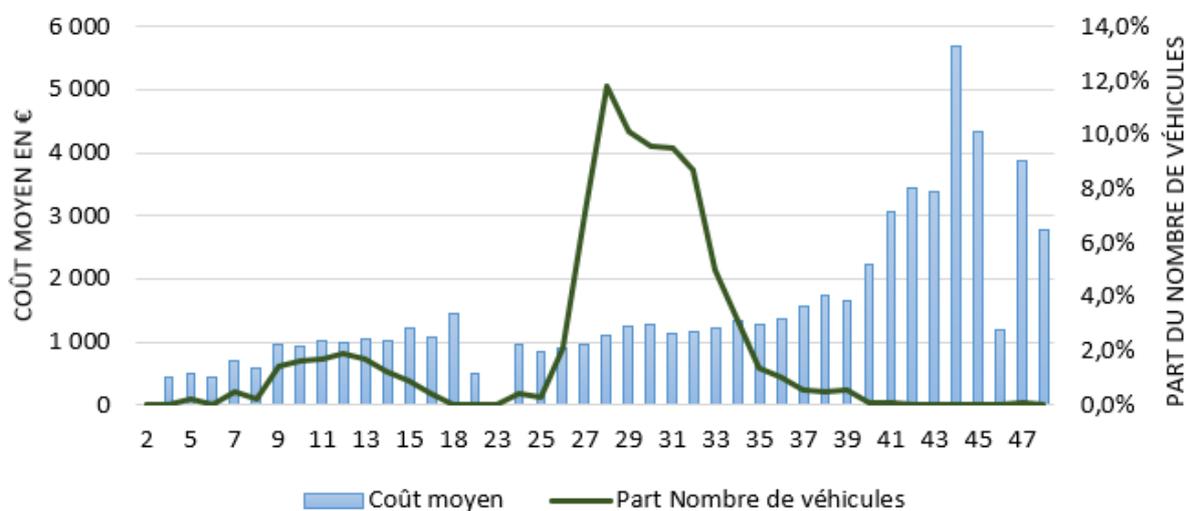


Figure 16 : Part du portefeuille et coût moyen en fonction du groupe SRA

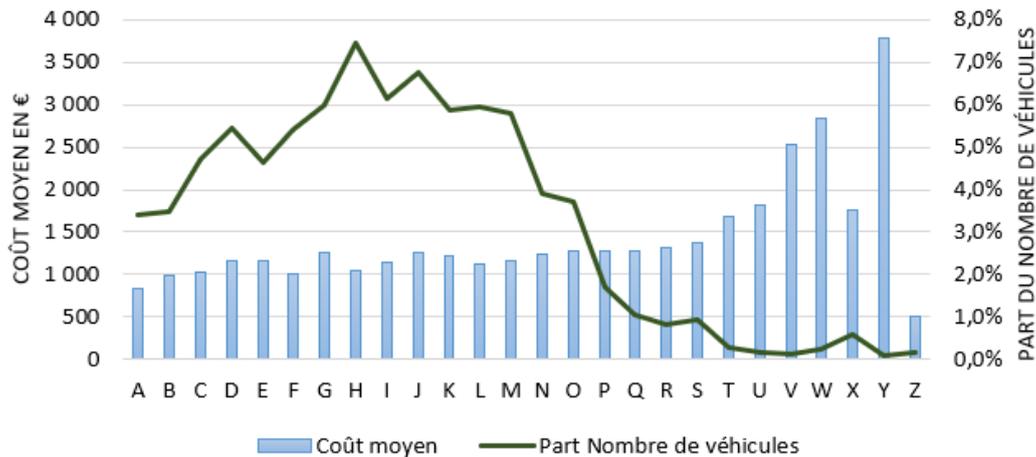


Figure 17 : Part du portefeuille et coût moyen en fonction de la classe prix SRA

Deux variables influençant clairement le coût moyen sont le groupe et la classe prix SRA. Comme indiqué dans le paragraphe III.2.b, l'association SRA publie des statistiques mais également des classements de véhicules en fonction :

- de la dangerosité intrinsèque du véhicule, appelée Groupe SRA : plus le véhicule est dangereux, plus le numéro attribué au Groupe SRA s'approche de 50 ;
- de la tranche de prix du véhicule à neuf (en TTC), appelée Classe Prix SRA : plus le véhicule est cher à l'achat, plus la lettre attribuée à la classe prix SRA est haute.

Le coût moyen d'un sinistre Automobile est clairement influencé par la dangerosité et le prix à neuf d'un véhicule.

Solvabilité II impose aux entreprises d'assurances une certaine dichotomie des risques, et encourage celles-ci à les affiner. Cette réglementation doit être vue comme une opportunité de mieux connaître et maîtriser ses risques, permettant un pilotage plus efficace de la société d'assurance.

La MAPA a aujourd'hui l'occasion de répondre aux remarques exprimées dans son rapport actuariel sur l'amélioration et/ou la justification de certains procédés utilisés dans le cadre de l'évaluation des meilleures estimations.

L'Automobile est un axe riche en enseignement puisqu'il couvre les sujets des lignes d'activité, de l'étendue disparate des sinistres, et de l'influence de la réassurance. Ces différences se ressentent dès l'analyse simple de la sinistralité, et notamment du coût moyen. C'est également une branche où les informations collectées par les organismes d'assurance sont nombreuses, ce qui pourraient avoir leur importance en fonction des méthodes de provisionnement utilisées.

Chapitre 2 : Complexité de la matière et outils mathématiques

I. Les fondamentaux du provisionnement

Le monde assurantiel a pour particularité unique de fonctionner avec un cycle de production inversé : les primes sont perçues avant même que l'entreprise d'assurance ait délivré les prestations aux assurés, prestations se présentant, en assurance non-vie, sous forme d'indemnités. Ces indemnités sont la conséquence de la survenance de sinistres affectant le patrimoine économique et/ou les préjudices corporels des assurés, permettant ainsi de limiter cette perte.

L'entreprise d'assurance doit être avant tout capable d'appeler les cotisations nécessaires permettant d'assurer ses engagements futurs, ainsi que le coût de fonctionnement de l'entité.

Cette notion d'engagements futurs peut être plus ou moins complexe en fonction du sinistre indemnisé. Un sinistre Automobile sur un pare-brise, faisant jouer la garantie « Bris de glace », est généralement indemnisé dans le mois suivant la déclaration. En revanche, un sinistre Automobile grave, impliquant des dommages corporels aux victimes peut engendrer une forte incertitude quant au préjudice réellement subi, et l'indemnisation peut s'étaler sur des dizaines d'années, et donc, bien au-delà de l'année d'exercice de la perception des primes ou de la déclaration du sinistre.

Ces engagements futurs, notamment liés à la complexité des dossiers, sont inscrits dans le passif du bilan de l'entreprise d'assurance, sous forme de provisions. Ces provisions représentent la part la plus importante et la plus incertaine du passif, et sont un enjeu majeur pour l'entreprise d'assurance.

A. Les différentes provisions

Les provisions en non-vie permettent de prendre en compte les spécificités du déroulement des sinistres, il sera donc ici question de décrire cette réalité, et de comprendre ce que cela implique.

Cette partie définira les grands concepts du provisionnement, abordera les notions d'IBNER, d'IBNYR et de meilleure estimation, et restreindra le périmètre d'étude.

Il existe évidemment bien d'autres provisions spécifiques, qui n'auront pas leur place dans ce mémoire.

1. La vie d'un sinistre et la provision dossiers/dossiers

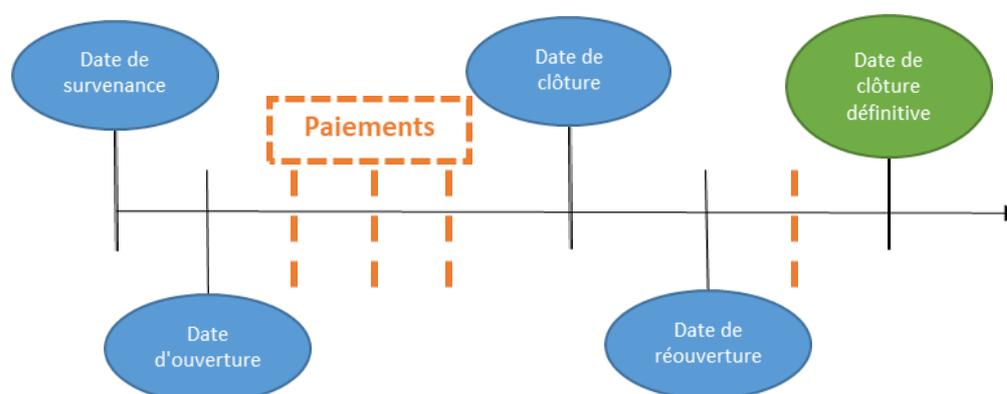


Figure 18 : Schéma de la vie d'un sinistre

La vie d'un sinistre peut se résumer en quelques étapes clés :

1) La survenance du sinistre

L'assuré est touché par un sinistre, il n'a pas encore prévenu l'entreprise d'assurance dans laquelle il a souscrit un contrat couvrant ce risque, durant la période établie.

2) L'ouverture du sinistre

L'assuré contacte son entreprise d'assurance pour commencer les démarches d'indemnisations. Cette déclaration peut être faite le même jour que la date de survenance, ou peut prendre plusieurs années.

3) La période d'évaluation et d'indemnisation

C'est à cette période que le dossier est évalué, suite à une expertise. Cette période est plus ou moins longue, en fonction de plusieurs critères :

- la gravité du sinistre ;
- les recours ou les procédures judiciaires ;
- la complexité du dossier.

Une première évaluation du dossier dans son ensemble est estimée : c'est la provision dossiers/dossiers. Cette dernière évolue à chaque nouvelle information significative entrant dans le dossier, permettant d'affiner cette estimation. Chaque paiement octroyé à l'assuré diminue d'autant le montant de cette provision.

4) La clôture du sinistre

C'est le moment où la dernière indemnisation délivrée à l'assuré correspond à ce qu'il restait de provision, et donc le moment où le montant de la provision tombe à 0€. L'assuré a été indemnisé de son préjudice à hauteur de ce qui était prévu dans les conditions générales et particulières de son contrat.

5) La réouverture du dossier

Cette étape, loin d'être systématique, arrive lorsque toutes les informations liées au sinistre n'étaient pas connues, ce qui a généralement pour effet d'augmenter le préjudice.

Une nouvelle phase d'évaluation et d'indemnisations commence alors.

6) La clôture définitive du dossier

On comprendra que toute la difficulté est d'évaluer au plus juste la sinistralité alors même que la compagnie d'assurance n'a qu'une partie, parfois ténue, de l'information.

La provision dossiers/dossiers est ici une provision individuelle, évaluée par les gestionnaires sinistres à partir des éléments fournis par le sociétaire, par des experts, etc. Le gestionnaire sinistre essaiera, dans la mesure du possible, de toujours évaluer au mieux cette provision, avec l'aide de consignes propres à chaque société d'assurance.

2. La provision pour sinistres tardifs (IBNYR)

Cette provision, généralement calculée par l'Actuaire, permet de prendre en compte une des caractéristiques de la vie du sinistre : l'assureur n'a pas connaissance immédiatement de la survenance du sinistre (IBNYR : Incurred But Not Yet Reported).

Ce décalage pose en réalité un problème uniquement lorsque le sinistre est déclaré sur une autre année que l'année de survenance du sinistre. La société d'assurance a en fait un engagement sur l'année comptable, dont elle n'a pas encore la connaissance.

Sur les entités non-vie, les distinctions suivantes sont courantes :

- les cas « simples » : les provisions qui concernent les branches où le décalage est faible entre la date de survenance et la date d'ouverture, qui sont généralement engendrées par les sinistres de fin d'année, déclarées en début d'année suivante, comme en Automobile ;
- les cas « complexes » : les provisions qui concernent les branches où le décalage est fort entre la date de survenance et la date d'ouverture, c'est notamment le cas des sinistres Catastrophes Naturelles Sécheresse, dont la déclaration dépend d'un arrêté ministériel pouvant être publié jusqu'à deux ans après la sécheresse en question.

Ces deux cas ne sont pas traités exactement de la même manière, mais ils sont gérés avec des méthodes de provisionnement agrégées. En effet, avant la déclaration la compagnie d'assurance n'a aucune information sur ces sinistres, il n'y a donc aucun moyen d'évaluer leur coût individuel. Ce qui n'est pas le cas des provisions couvrant l'insuffisance potentielle de provisionnement dossiers/dossiers.

3. La provision pour insuffisance potentielle (IBNER)

Cette provision, généralement calculée par l'Actuaire, permet de prendre en compte une des caractéristiques de la vie du sinistre : le gestionnaire sinistre n'a pas les informations nécessaires pour connaître précisément le montant total du sinistre, et peut parfois sous-évaluer cette valeur (IBNYR : Incurred But Not Enough Reserved).

Cette provision est un enjeu de prudence comptable. Elle permet d'ajouter une provision supplémentaire aux provisions dossiers/dossiers lorsque celles-ci paraissent insuffisantes. Le terme de mali est alors utilisé.

Il est également bon de savoir qu'il existe une taxe sur les bonis. Les bonis que dégagent les entreprises d'assurance d'une année sur l'autre sont taxés. Les sociétés d'assurance doivent donc s'assurer de la suffisance de ses provisions, sans les surpondérer.

4. La meilleure estimation des provisions pour sinistres (Best Estimate)

Sous Solvabilité II, les provisions sont à calculer comme la meilleure estimation ou « Best Estimate », défini dans le Règlement Délégué comme étant « *la moyenne pondérée par leur probabilité des flux de trésorerie futurs* », dont les flux sont ensuite actualisés via une courbe des taux.

Cette définition exige des entreprises d'assurance d'évaluer au mieux leur provision, sans pouvoir y intégrer une marge de prudence.

Ce mémoire a pour objectif d'évaluer cette meilleure estimation, d'obtenir les provisions techniques à l'ultime pour des sinistres Automobile. Une partie de cette meilleure estimation, et notamment pour l'année de survenance la plus récente, intègre des provisions pour sinistres tardifs (INBYR). Cette provision n'est pas l'objet de ce mémoire, puisqu'elle est assez « simple » à identifier. Cette partie de

la meilleure estimation ne sera donc pas intégrée. Enfin, l'actualisation ne sera pas ou peu abordée, pour des raisons similaires.

B. Méthodes de provisionnement agrégées : Chain Ladder

1. Notations

Chaque sinistre connu par les entreprises d'assurance a un déroulé de vie qui lui est propre, cela dépend à la fois de la complexité du sinistre, de la réactivité de l'assuré, de la surcharge de travail des gestionnaires sinistre, et de nombreux autres paramètres.

Mais, en regroupant des sinistres similaires, il est possible de distinguer une tendance, un comportement qui leur est propre, et que l'Actuaire utilise pour évaluer la charge ultime.

Pour rappel, la vie d'un sinistre peut se résumer à une évaluation dossiers/dossiers qui est actualisé au fur à mesure des informations perçues, et qui aboutissent à des règlements plus ou moins fractionnés dans le temps en fonction de la complexité du sinistre. Le coût total réel est connu à la clôture (définitive) du sinistre, et est la somme des règlements effectués à l'assuré entre l'ouverture et la clôture du dossier. Il est également important de rattacher un sinistre à son année de survenance, et non d'ouverture, en raison de la validité des termes du contrat reliant l'assuré et l'assureur.

De cette description de la sinistralité en ressort la représentation suivante, sous forme de triangle, qui est la plus communément utilisée :

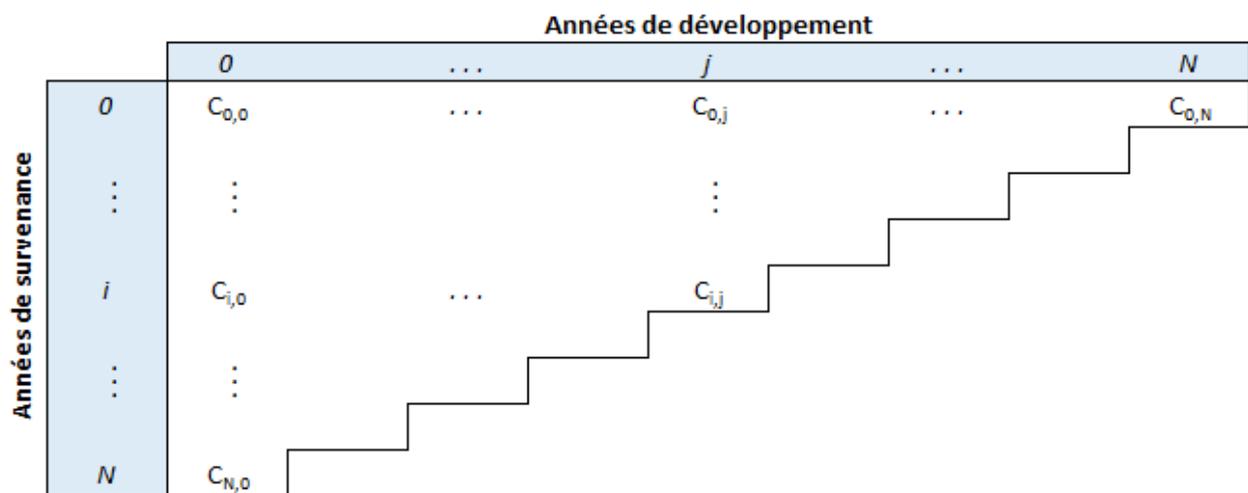


Figure 19 : Triangle standard et notations

Les sinistres de l'année de survenance i , ont eu des règlements à différentes dates, parfois sur des années différentes, appelées années de développement.

Ainsi, les notations suivantes expriment :

- $C_{i,j} = \sum_{h=0}^j X_{i,h}$ est le montant cumulé des paiements ayant eu lieu entre les années de développement 0 et j , pour les sinistres dont la survenance concernait l'année i ;
- $X_{i,j} = C_{i,j} - C_{i,j-1}$ est le montant payé durant l'année de développement j (pour $j \geq 1$), au titre des sinistres dont la survenance concernait l'année i ;

- $C_{i,0} = X_{i,0}$ puisque le cumul des montants payés à la fin de la première année est bien égal au montant payé la première année.

Les $X_{i,j}$ sont des variables aléatoires, représentant les paiements effectués lors d'une année de développement j , pour l'année de survénance i . À noter que la dernière diagonale correspond au moment où l'information est la plus complète, soit la dernière date de vue des sinistres.

L'objectif est d'évaluer le montant des paiements qu'il reste à payer, soit la partie basse (et inconnue) du triangle. Pour ce faire, le moyen le plus standard de parvenir à cet objectif est d'utiliser la méthode déterministe Chain Ladder.

2. Fonctionnement

Chain Ladder est une technique d'évaluation des provisions techniques qui repose sur les cadences de règlements issues du triangle des paiements cumulés. Autrement dit, la façon dont les sinistres ont été réglés sur les années de survénance antérieures est représentatif de ce qui va se passer sur les années de survénance à venir.

Pour cela, la méthode Chain Ladder propose de calculer des coefficients de développement qui représentent la variation moyenne entre l'année de développement j et $j + 1$ sur l'ensemble des années de survénance antérieures à l'année i considérée. Cela se traduit par la formule suivante :

$$\hat{f}_j^{CL} = \frac{\sum_{i=0}^{N-j-1} C_{i,j+1}}{\sum_{i=0}^{N-j-1} C_{i,j}} \text{ pour } j = 1, \dots, n$$

Avec ces facteurs de développement, il est alors possible de compléter le triangle inférieur, et donc d'estimer le coût total à l'ultime de la sinistralité par année de survénance, et d'en déduire une provision.

Chaque paiement cumulé pour une année de développement donné peut être calculé à partir du dernier paiement cumulé connu :

$$\hat{C}_{i,j+1} = \hat{f}_j^{CL} \times \hat{C}_{i,j} \quad \text{pour } i + j > N$$

ou, à partir de n'importe quel paiement cumulé antérieur :

$$\hat{C}_{i,j+1} = \left(\prod_{k=N-i}^j \hat{f}_k^{CL} \right) \times C_{i,N-i}$$

La connaissance du triangle inférieur permet alors d'estimer les paiements restants à payer (Réserves) pour chaque année de survénance, par différence entre le montant cumulé sur la dernière année de développement et le dernier montant cumulé connu :

$$\hat{R}_i = \hat{C}_{i,N} - C_{i,N-i}$$

La somme de ces réserves pour chaque année de survénance permet d'avoir la provision totale estimée à l'ultime :

$$\hat{R} = \sum_{i=1}^N \hat{R}_i$$

Chain Ladder est une méthode particulièrement simple d'utilisation, et, dans de nombreux cas, c'est une méthode très efficace.

Pour pouvoir mieux comprendre les avantages et inconvénients de cette méthode, étayons les hypothèses qui lui sont sous-jacentes :

- ✓ le déroulement de la vie des sinistres est indépendant de l'année de survenance ;

Cette hypothèse est relativement forte et déterminante, et il est donc du devoir de l'Actuaire de faire en sorte que la réalité s'en approche au mieux, notamment par des méthodes graphiques (ce point sera développé dans la partie consacrée à la méthode de Mack). Hormis les analyses graphiques, les principaux points d'attention sont les suivants :

- **segmenter** : c'est un excellent moyen de réduire la volatilité et d'avoir des cadencements proches d'une année sur l'autre ;
- corriger l'inflation ;
- retirer les années ou cadences atypiques ;
- et de manière plus générale, pour le Responsable de la fonction clé :
 - analyser la déformation du portefeuille ;
 - connaître les potentiels changements dans la gestion des sinistres.
- ✓ le modèle suppose que le déroulé des sinistres est (au moins) terminé à la dernière année de développement, puisque l'objectif est de déterminer les charges à l'ultime. Cette hypothèse est simple à respecter sur les sinistres des branches dites « courtes » (généralement clos avant la 5^{ème} année de développement), et beaucoup moins pour les branches dites longues. Cela pose le problème de l'historique disponible.

Pour pallier cette problématique sur les sinistres de longs termes, où les triangles n'auraient pas un historique suffisamment profond pour pouvoir estimer sans sous-évaluer, une queue de développement du triangle sera estimée (appelée « *Tail factors* »).

Pour estimer les « *Tail factors* », dans les modèles reposant sur les hypothèses de Chain Ladder, la méthode consiste à extrapoler les facteurs de développement sur les années qui vont au-delà du triangle initial, par une fonction comme la fonction exponentielle ou la fonction puissance (progression lente des coefficients de développement vers l'unité).

- ✓ enfin, même si cela est une évidence, les données, aussi bien en termes de quantité, qu'en termes de qualité, jouent un rôle prépondérant dans l'estimation de ces provisions. Pour rappel, le Responsable de la fonction clé Actuariat a des engagements forts dans la détermination des provisions techniques sous Solvabilité II.

Voici quelques exemples où Chain Ladder est donc difficilement applicable :

- ✗ les branches longues (Responsabilité Civile « corporel », assurance construction, ...)
- ✗ les sinistres graves, par définition peu nombreux, et qui sont dépendants de facteurs trop aléatoires (analyses médicales, tribunaux, recours, ...) pour assurer une stabilité ;
- ✗ les années de survenance récentes de la garantie sécheresse (sinistres tardifs).

Dans le cadre de la recherche de la meilleure estimation des provisions en Automobile, cette première approche semble déjà déterminante sur la manière d'aborder le sujet. Il faut pouvoir distinguer les sinistres pouvant être traités avec Chain Ladder, de ceux qui sont peu adaptés à la méthode.

Les bases de données permettent déjà de distinguer les sinistres RC « Corporel », des autres sinistres Automobile qui semblent tous pouvoir être traités avec Chain Ladder (ce qui sera vérifié avec un exemple).

En revanche, il est plus compliqué de définir le seuil à partir duquel un sinistre est considéré comme « grave ». Le sujet de la segmentation des sinistres graves et non graves sera abordé dans la 2^{ème} partie de ce chapitre.

Enfin, une des limites importantes du modèle Chain Ladder est qu'il n'y a pas de mesure de l'incertitude. Cette appréhension du risque peut être très intéressante, et notamment en termes de meilleures estimations.

La prochaine sous-partie sera consacrée à ce sujet, via les méthodes de Mack et du Bootstrap.

C. Méthodes de provisionnement agrégées : Mesure de l'incertitude

La mesure de l'incertitude, autour de l'estimation du montant de la provision, peut être utile pour deux objectifs :

- avoir une notion probabiliste dans l'évaluation de la meilleure estimation (notion essentielle à Solvabilité II) ;
- définir un seuil entre les sinistres graves et les sinistres attritionnels, en minimisant cette incertitude.

Les deux méthodes présentées ci-dessous répondent à ces problématiques.

1. Méthode de Mack

La méthode de Mack est une méthode dite « stochastique simple », car elle ne comporte pas de caractère aléatoire. Le principe de l'évaluation des provisions est strictement le même que pour la méthode déterministe Chain Ladder, mais elle permet, en plus, d'estimer une erreur de prédiction liée à ces provisions.

Pour cela, la méthode s'appuie sur trois hypothèses (dont les deux premières sont déjà présentes dans la méthode Chain Ladder) :

- l'existence des coefficients de passage entre chaque année de développement, tel que :

$$\mathbb{E}[C_{i,j+1}|C_{i,j} \dots C_{i,1}] = \mathbb{E}[C_{i,j+1}|C_{i,j}] = f_j C_{i,j}$$

L'évolution des paiements est donc un processus Markovien.

- l'indépendance des années d'origine ;
- l'existence des paramètres inconnus $(\sigma_0, \dots, \sigma_{N-1})$ désignant la volatilité interne du triangle tels que :

$$Var[C_{i,j+1}|C_{i,j} \dots C_{i,1}] = Var[C_{i,j+1}|C_{i,j}] = \sigma_j^2 C_{i,j}$$

Sous ces hypothèses, les \hat{f}_j^{CL} sont des estimateurs sans biais et non corrélés des f_j^{CL} .

Ces hypothèses peuvent être vérifiées, notamment par des analyses graphiques :

- ✓ un alignement par rapport à l'origine des points $(C_{i,j}, C_{i,j+1})$, notamment pour les premières années de développement ;
- ✓ une absence de tendance sur les résidus de Pearson (écart normalisé) $\frac{C_{i,j+1} - \hat{f}_j^{CL} C_{i,j}}{\sqrt{C_{i,j}}}$ mis sur un graphique.

Afin de définir l'erreur assimilée à l'estimation des provisions, il convient donc de définir l'estimation de la volatilité interne du triangle. La proposition de Mack est la suivante :

$$\begin{cases} \hat{\sigma}_j^2 = \frac{1}{N-j-1} \sum_{i=0}^{N-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j^{CL} \right)^2, & 0 \leq j \leq N-2 \\ \hat{\sigma}_{N-1}^2 = \min\left(\frac{\hat{\sigma}_{N-2}^4}{\hat{\sigma}_{N-3}^2}, \min(\hat{\sigma}_{N-3}^2, \hat{\sigma}_{N-2}^2)\right) \end{cases}$$

À partir de cet estimateur, il est alors possible de définir l'erreur quadratique moyenne de prédiction (Mean Squared Error of Prediction) de la provision de l'année de survénance i de la manière suivante :

$$\widehat{MSEP}(\hat{R}_i) = \hat{C}_{i,N}^2 \sum_{j=N-i+1}^{N-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{N-j} C_{k,j}} \right)$$

L'erreur autour de l'estimation est donc expliquée à la fois par l'erreur de processus, soit la variabilité autour de la moyenne ; et l'erreur d'estimation puisque la valeur réelle des facteurs de développement n'est pas connue.

Enfin, l'erreur quadratique moyenne de l'estimation de la provision totale prend en compte la corrélation qui existe entre les différentes erreurs quadratiques individuelles puisqu'elles dépendent des mêmes facteurs de développement. L'erreur totale est donc la suivante :

$$\widehat{MSEP}(\hat{R}) = \sum_{i=1}^N \left[\widehat{MSEP}(\hat{R}_i) + \hat{C}_{i,N} \left(\sum_{k=i+1}^N \hat{C}_{k,N} \right) \left(\sum_{j=N-i}^{N-1} \frac{2\hat{\sigma}_j^2}{\hat{f}_j^2 \sum_{l=0}^{N-j} C_{l,j}} \right) \right]$$

$$\text{Erreur standard relative} = \frac{\sqrt{\widehat{MSEP}(\hat{R})}}{\hat{R}}$$

L'obtention de cette erreur quadratique moyenne de prédiction permet donc d'estimer un intervalle de confiance autour de l'estimation de la provision. Le modèle de Mack étant non paramétrique, il est donc nécessaire de supposer la distribution de cette provision.

Généralement, les lois utilisées sont la loi Normale et la loi Log-Normale. La seconde ayant l'avantage d'estimer un intervalle de confiance strictement positif :

- Intervalle de confiance loi Normale :

$$IC_{95\%}^N = [\hat{R} - q_{95\%} se(\hat{R}); \hat{R} + q_{95\%} se(\hat{R})] \quad \text{avec } se(\hat{R}) = \sqrt{\widehat{MSEP}(\hat{R})}$$

- Intervalle de confiance loi Log-Normale :

$$IC_{95\%}^{LN} = \left[\hat{R} \exp\left\{\frac{-\sigma^2}{2} - q_{95\%}\sigma\right\}; \hat{R} \exp\left\{\frac{-\sigma^2}{2} + q_{95\%}\sigma\right\} \right] \quad \text{avec } \sigma = \ln\left(1 + \left(\frac{se(\hat{R})}{\hat{R}}\right)^2\right)$$

L'inconvénient de cette méthode se trouve sur l'instabilité de la variance, très sensible aux coefficients de passage individuels atypiques. L'application de cette méthode d'estimation d'un intervalle de confiance nécessite, encore plus que pour l'estimation de la provision elle-même, un retraitement du triangle en amont.

2. Méthode du Bootstrap

La méthode du Bootstrap possède deux avantages majeurs :

- ✓ elle permet de s'affranchir d'hypothèses contraignantes sur une famille de lois de probabilité³ ;
- ✓ elle permet d'obtenir une distribution probabiliste.

Son fonctionnement repose sur le rééchantillonnage, et les simulations de Monte-Carlo. Tout comme pour Mack, c'est une méthode non-paramétrique, qui est donc peu contraignante.

Les variables utilisées doivent être indépendantes et identiquement distribuées (iid). C'est la principale hypothèse de cette méthode. Cette hypothèse n'est généralement pas vérifiée avec les montants de charge des sinistres. Le Bootstrap repose donc sur le rééchantillonnage des résidus de Pearson calculés à partir des incréments de paiements, notamment ceux dans le cas d'une régression de Poisson, qui sont plus proches d'être iid.

Le principe est donc de faire B simulations, à partir de ces résidus, afin d'obtenir B provisions \hat{R} , et ce, afin d'extraire la volatilité autour de cette estimation. Les étapes sont les suivantes :

1. Définir les facteurs de développement \hat{f}_j^{CL} .
2. Reconstruire la partie haute du triangle (en cumulée, puis en décumulée), en partant de la diagonale grâce aux facteurs de développement de Chain Ladder.
3. Calculer les résidus de Pearson (distribution de Poisson) :

$$r_{i,j}^P = \frac{C_{i,j} - \hat{C}_{i,j}}{\sqrt{\hat{C}_{i,j}}}$$

Ces résidus sont ensuite corrigés, afin de prendre en compte le nombre de paramètres dans le modèle, et corriger le biais des estimations du Bootstrap :

$$r_{i,j}^{adj} = \sqrt{\frac{M}{M-P}} \times r_{i,j}^P$$

Avec $M = \frac{N(N+1)}{2}$ qui correspond au nombre de données et $P = 2N - 1$, soit le nombre de degrés de liberté.

4. Rééchantillonnage avec remise des résidus sous une loi uniforme
5. Reconstruction de la partie haute du triangle (en décumulée, puis en cumulée), grâce au triangle théorique et aux résidus bootstrapés $r_{i,j}^*$:

³ Charpentier A., Denuit M. (2004), Mathématiques de l'assurance non-vie, tome II – Chapitre 15.3.1

$$\hat{C}_{i,j}^* = r_{i,j}^* \times \sqrt{\hat{C}_{i,j} + \hat{C}_{i,j}}$$

6. Estimation des facteurs de développement bootstrapés \hat{f}_j^{CL} et de la provision totale à l'ultime \hat{R}^* .
7. Les points 4, 5 et 6 sont donc à réitérer B fois.

Cette méthode permet d'extraire une moyenne, un écart-type et de nombreuses mesures de risque, notamment utiles sous Solvabilité II (comme la *Value at Risk* ou *VaR*).

L'inconvénient principal de cette méthode c'est qu'elle n'est pas adaptée aux triangles dont les montants cumulés ne sont pas croissants. Elle est donc difficilement applicable aux triangles de règlements nets de recours ou aux triangles de coûts totaux (présence de bonis de liquidation). Une solution (en gardant la même méthode) est de forcer les résidus concernés à 0, ou de ne pas considérer ces résidus comme faisant partie du tirage. Mais cela apporte forcément un biais à l'estimation.

Chain Ladder est un outil intéressant dans l'estimation de la charge ultime des sinistres, lorsque les hypothèses précédemment citées sont respectées⁴ (Cf 2.1.B). Évidemment, cette méthode n'est pas sans défaut, mais elle reste la principale méthode utilisée aujourd'hui.

Le Bootstrapping et la méthode de Mack vont permettre de connaître l'incertitude autour de la moyenne, dont l'estimation est basée sur la méthode déterministe de Chain Ladder. Il sera alors possible de :

- définir un intervalle de confiance, une borne limite à partir de laquelle la société d'assurance aura la quasi-certitude d'avoir les provisions nécessaires pour faire face à ses engagements ;
- comparer l'incertitude, afin de la minimiser, ce qui pourrait notamment être utile dans la séparation entre les sinistres attrionnels et les sinistres graves.

L'objectif de ce mémoire étant également de définir le seuil séparant la gravité des sinistres, ces méthodes seront utilisées pour appuyer les solutions trouvées par les bases de la théorie des valeurs extrêmes.

En revanche, elles présentent leurs limites, et notamment sur les données atypiques, ou sur les sinistres de long terme. La MAPA possède un portefeuille Automobile conséquent, mais ne lui permettant pas d'atteindre la stabilité des plus grands assureurs de la branche. La problématique du mémoire concerne le provisionnement de la RC Automobile et notamment « corporel », de ce fait, les triangles auront donc les deux aspects pouvant avoir des effets néfastes sur l'incertitude.

Il sera alors proposé dans ce mémoire d'utiliser une approche différente des méthodes classiques de provisionnement par des données agrégées, en étudiant les sinistres un à un, considérant les caractéristiques individuelles et disponibles du sinistre.

⁴ Charpentier A., Denuit M. (2004), Mathématiques de l'assurance non-vie, tome II – Chapitre 13.3

II. Définir le seuil des graves

Comme vu dans la partie précédente, les méthodes de provisionnement agrégées sont d'autant plus efficaces lorsque les sinistres représentent le même type de sinistralité, lorsque ces derniers se comportent relativement de la même manière dans le temps, permettant ainsi de réduire l'incertitude autour de l'évaluation de la provision.

Une des causes principales de l'instabilité des triangles de coûts est les sinistres graves. En effet, et notamment en RC Automobile « Corporel », ces sinistres ont une phase de première évaluation assez longue, suivie d'ajustements en fonction de l'information reçue, avec une moyenne avant clôture bien plus élevée que sur les sinistres attritionnels. Il est donc généralement recommandé de séparer les sinistres graves, des autres sinistres.

Pour essayer de définir au mieux la limite entre les sinistres graves, et les autres sinistres de manière théorique, des notions de la Théorie des Valeurs Extrêmes (TVE) seront exploitées. La TVE est le plus souvent utilisée en réassurance. Elle permet notamment de définir par des lois les queues de distribution des sinistres, et ainsi connaître, par exemple, la valeur maximale pouvant être atteinte. Pour répondre à ce genre de problématique, il faut tout d'abord identifier les sinistres faisant partie de la queue, des autres sinistres. C'est en cela que la TVE peut répondre à la problématique de seuil.

A. Notions sur la Théorie des Valeurs Extrêmes (TVE)

L'objectif initial de la Théorie des Valeurs Extrêmes (TVE) est donc de pouvoir estimer les pertes maximales que pourraient subir une entreprise dans le cas où un événement extrême se produirait, ce qui sous-entend que l'événement en question n'a jamais eu lieu. Cette théorie est très utilisée dans le cadre de la réassurance.

Cette théorie permet de construire des modèles statistiques sur des événements qui paraissent dénués de règles.

Pour cela, la TVE cherche à identifier les queues de distribution de F (inconnue), une distribution de variables aléatoires et identiquement distribuées X_1, \dots, X_n , dont les valeurs connues sont, pour la plupart, autour de la moyenne. Or, les valeurs qui permettent le mieux d'identifier les queues de distributions sont rares.

Cette approche s'intéresse tout particulièrement à la loi du maximum :

$$M_n = \max(X_1, \dots, X_n)$$

Alors :

$$\begin{aligned}\Pr(M_n \leq x) &= \Pr(X_1 \leq x, \dots, X_n \leq x) \\ \Pr(M_n \leq x) &= \Pr(X_1 \leq x) \dots \Pr(X_n \leq x) \\ \Pr(M_n \leq x) &= [F(x)]^n\end{aligned}$$

Des approximations asymptotiques et simples de M_n sont souhaitées. Or, la distribution asymptotique de M_n est dégénérée. Il faut donc passer par une normalisation. C'est ce qui est d'ailleurs fait par le Théorème Centrale Limite (TCL), où la normalisation linéaire dépend des deux premiers moments de F .

Théorème Fisher-Tippett : S'il existe des suites de réels $c_n > 0$ et d_n , telles que, quand $n \rightarrow \infty$

$$\Pr\left(\frac{M_n - d_n}{c_n} \leq x\right) \rightarrow \mathcal{G}(x)$$

pour une distribution non-dégénérée \mathcal{G} , alors \mathcal{G} est du même type que l'une des trois distributions qui peuvent être unifiées par une seule distribution appelée distribution *GEV* (distributions des extrêmes généralisées). On a $GEV(\mu, \sigma, \xi)$:

$$\mathcal{G}(x) = \begin{cases} \exp\left(-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]_+^{-\frac{1}{\xi}}\right) & \text{si } \xi \neq 0 \\ \exp\left(-\exp\left(-\left(\frac{x - \mu}{\sigma}\right)\right)\right) & \text{si } \xi = 0 \end{cases}$$

où $x_+ = \max(x, 0)$ et $\sigma > 0$. μ est un paramètre de position, σ est un paramètre d'échelle et ξ est un paramètre de forme. C'est ce paramètre de forme qui détermine plus précisément la correspondance avec l'une des trois distributions suivantes :

- la distribution de Gumbel pour $\xi = 0$, correspondant à une queue de distribution intermédiaire comme les lois Exponentielle, Normale et Log-Normale ;
- la distribution de Fréchet pour $\xi > 0$, correspondant à une queue de distribution épaisse comme la loi de Pareto ;
- la distribution de Weibull pour $\xi < 0$, correspondant à une queue de distribution fine comme la loi Uniforme.

Maintenant, plutôt que de s'intéresser au maximum, il est également possible de s'intéresser au nombre de valeurs dépassant des seuils u_n suffisamment grand, permettant ainsi de conserver une information plus importante. C'est la méthode des dépassements de seuil ou méthode des excès (ou *Peak Over Threshold*).

On s'intéresse donc aux $N_n = \sum_{i=1}^n \mathbb{I}_{\{X_i \geq u_n\}}$ dépassements de seuil u_n , c'est-à-dire aux observations $(X_i - u_n)_+$ qui sont strictement positives. Elles sont caractérisées par une la Distribution de Pareto Généralisée $GPD(\beta, \xi)$ définie par :

$$\mathcal{G}_{\xi, \beta}^p(x) = \begin{cases} 1 - \left[1 + \xi\left(\frac{x}{\beta}\right)\right]_+^{-1/\xi} & \text{si } \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right) & \text{si } \xi = 0 \end{cases}$$

où $x \geq 0$ si $\xi \geq 0$; et $0 \leq x \leq -\frac{\beta}{\xi}$ si $\xi < 0$.

La difficulté repose sur le choix des seuils. Il doit être suffisamment grand pour que la loi des excès puisse converger vers une GPD, mais un seuil trop élevé, amène forcément à un nombre de données plus restreint, et à une volatilité accrue, baissant la précision de l'estimation.

B. La Fonction Moyenne des Excès (Mean Excess Function)

Afin de déterminer le seuil à partir duquel il est possible de distinguer les sinistres dits « attritionnels » des sinistres graves, le choix d'utiliser un outil graphique est pris : la **fonction moyenne des excès**. Cet outil est souvent utilisé en réassurance, notamment pour la tarification. En effet, la fonction moyenne des excès est définie par :

$$e(u) = \frac{1}{\mathbb{P}[X > u]} \int_u^\infty \mathbb{P}[X > t] dt$$

ce qui représente le coût moyen (perte moyenne espérée), pour un traité en excédent par exemple, sachant la franchise u , le seuil en question. Cette fonction peut être approchée par l'estimateur empirique suivant :

$$\hat{e}(u) = \frac{1}{\sum_{i=1}^n \mathbb{I}_{X_i > u}} \sum_{i=1}^n [X_i - u]_+$$

soit la somme des montants de sinistre dépassant le seuil u , divisé par le nombre de sinistres remplissant la même condition. Cette fonction est donc utile en réassurance pour la tarification d'un produit dont le seuil est connu, mais il peut aussi l'être pour identifier ce seuil. Pour cela, il faut s'appuyer sur les résultats de la TVE, et notamment celui sur la distribution au-delà d'un seuil assez grand : la Distribution de Pareto Généralisée (GPD). Or, la GPD possède la propriété de stabilité par seuil :

$$\text{Si } X \sim GPD(\beta, \xi) \text{ alors } e(u) = \frac{\beta}{1-\xi} + \frac{\xi}{1-\xi} u$$

Cette propriété est connue depuis 2005, et elle signifie que si l'approximation de la GPD est valide pour un seuil donné, alors elle est valide pour tout seuil supérieur. L'objectif est donc de trouver un seuil u , le plus petit possible, permettant à l'estimateur empirique de l'espérance résiduelle des excès au-delà de ce seuil d'être approximativement linéaire.

Plus précisément, trois cas de figure sont possibles :

- à partir du seuil u , la pente de la fonction des excès est **linéaire et positive**, exactement dans le cas présenté ci-dessus, avec des queues lourdes de distribution des sinistres, où les observations suivent une GPD, ou alors une distribution de Pareto ;
- à partir du seuil u , la fonction des excès est **constante**, c'est-à-dire en présence d'une queue intermédiaire de distribution des sinistres, où les observations suivent une distribution exponentielle ;
- à partir du seuil u , la pente de la fonction des excès est **linéaire et négative**, c'est-à-dire en présence d'une queue fine de distribution des sinistres.

Cet outil permet alors de distinguer le seuil à partir duquel les sinistres sont considérés comme graves, mais également de connaître la lourdeur de la queue de distribution.

L'inconvénient, c'est que la limite n'est pas toujours claire, et qu'elle se base sur une interprétation graphique. La zone où se trouve le seuil u est approximativement visible, mais plusieurs points sont possibles. Afin d'affiner les critères de sélection de ce seuil u , le choix est ici fait d'exploiter les outils généralement dédiés à la tarification de produits de réassurance.

Pour les produits de type Excédent de sinistre (XS), les sinistres dits « graves » sont basiquement décrits comme :

$$S^{Grave}(t) = \sum_{i=1}^{N(t)} X_i$$

avec :

- $\{N(t): t \geq 0\}$ un processus de Poisson de paramètre λ , pour approximer la fréquence ;
- $\{X_i: i \geq 1\}$ i.i.d. $\sim Pareto(A, \alpha)$ indépendant de $\{N(t): t \geq 0\}$, pour approximer la sévérité.

Rappel et propriétés :

Poisson : $\mathbb{P}[N(t) = k] = e^{-\lambda} \frac{\lambda^k}{k!}$

Propriétés : nombre entier non-négatif, s'adapte parfaitement pour un nombre de sinistres

Pareto : $\mathbb{P}[X \leq x] = \begin{cases} 1 - \left(\frac{x}{A}\right)^{-\alpha} & \text{si } x > A \\ 0 & \text{si } x \leq A \end{cases}$ (*shifted distribution*)

Propriétés : distribution à queue lourde

Nota Bene : par la suite A sera appelé u, le seuil entre les sinistres attritionnels et graves

Ces deux distributions sont des candidates idéales pour définir les sinistres au-delà du seuil u.

Afin de conforter le choix qui est fait à partir du graphique de dépassement moyen, plusieurs contrôles, basés sur la définition précédemment faite des sinistres graves (Poisson/Pareto), seront réalisés :

- ✓ Stabilité du paramètre α de la loi de Pareto

L'estimateur du maximum de vraisemblance de α vérifie :

$$\frac{1}{\hat{\alpha}} = \frac{1}{n} \sum_{i=1}^n \log\left(\frac{X_i}{u}\right)$$

Cet estimateur sera utilisé pour déterminer $\hat{\alpha}$, et ce, pour tout u appartenant à X_i . Il sera alors possible, à l'instar du graphique de dépassement moyen, de faire un graphique de l'estimateur $\hat{\alpha}$.

L'objectif étant de trouver un u assez grand à partir duquel l'estimateur $\hat{\alpha}$ est assez élevé pour représenter correctement les sinistres graves, c'est-à-dire lorsque les $\hat{\alpha}$ deviennent relativement constants en fonction de u.

- ✓ Acceptation de l'hypothèse d'égalité des lois

Enfin, l'adéquation sera testée entre les distribution empiriques et théoriques des sinistres graves, et ce, pour le nombre de sinistres (Poisson) et le montant des sinistres (Pareto), une fois le seuil u identifié.

Pour cela, la statistique de Kolmogorov-Smirnov sera utilisée :

$$k = \sup_{x \in \mathbb{R}} |F(x) - F_n(x)| = \max_{i=1}^n \left(F(x_{(i)}) - \frac{i-1}{n} \right) \vee \left(\frac{i}{n} - F(x_{(i)}) \right)$$

où $(x_{(i)})_{i=1}^n$ est un échantillon ordonné, et $a \vee b = \max(a, b)$.

On accepte l'hypothèse d'égalité des lois si $k \leq k_{n,1-\alpha}$, pour un seuil α donné. $k_{n,1-\alpha}$ est donné.

Si le test rejette l'égalité des lois, que ce soit pour l'adéquation à la loi de Poisson pour le nombre de sinistres, ou à la loi de Pareto pour leur montant, alors il sera considéré que le seuil u choisi n'est pas approprié.

L'ensemble de ces bases de TVE fournissent des procédures rationnelles et scientifiques d'identification d'un seuil séparant les sinistres graves et les sinistres attritionnels. Malgré un ajout

théorique pour compléter la méthodologie du graphique de dépassement moyen pour l'identification d'un seuil par des moyens statistiques, il n'en reste pas moins que cette méthodologie est sensible, et l'analyse graphique parfois difficilement interprétable.

De ce fait, les résultats obtenus seront également mis en parallèle des conclusions sur la stabilité des estimations des provisions à l'ultime.

III. Le provisionnement ligne à ligne

Comme mentionné dans la première partie de ce deuxième chapitre, le provisionnement agrégé comporte des limites importantes, notamment sur l'évaluation des sinistres de long terme. Une alternative aux méthodes agrégées est le provisionnement ligne à ligne (ou *Micro-level reserving*).

Le provisionnement ligne à ligne permet d'exploiter toute l'information disponible liée à un sinistre et aux conditions contractuelles au moment de la survenance de ce dernier. Cette méthode est donc capable de gérer l'hétérogénéité des données.

Elle permet également de se détacher de la problématique du seuil entre les sinistres attritionnels et graves, même si, elle génère une problématique de quantité de données (sujet abordé par la suite).

Car en effet, pour évaluer les provisions à l'ultime individuelles de chaque sinistre, les outils utilisés sont ceux de la data science, comme la méthode CART, méthode d'autant plus efficace dans le cadre des Big data.

A. Arbres de décision : la méthode CART

La méthode CART (Arbres de Régression et de Classification ou *Classification And Regression Trees*) est une méthode d'apprentissage statistique. Ici, plus particulièrement, il est question d'arbres de régression, la variable à expliquer, ou variable d'intérêt, étant une variable quantitative (les provisions à l'ultime de chaque sinistre). Afin de pouvoir estimer au mieux cette variable d'intérêt, il faut implémenter aux modèles les différentes caractéristiques des sinistres disponibles, et extraire les informations qui ont une influence significative sur le montant des provisions à l'ultime.

La méthode consiste alors à construire un arbre constitué :

- d'une racine, qui est le point de départ de l'arbre, la première segmentation de la base ;
- de nœuds, qui représentent les différentes segmentations de la base (règle de division), toujours en fonction de modalités d'une variable explicative (facteurs de risque ou covariance) des sinistres ;
- de feuilles, qui représentent la quantité des sinistres dont les caractéristiques amènent à une évaluation similaire de la variable d'intérêt (sous-populations homogènes). Chaque sinistre n'appartient qu'à une seule et unique feuille (population disjointe).

Afin d'affiner la manière dont est construit l'arbre, posons les notations suivantes :

- $i \in \llbracket 1, n \rrbracket$ identifiant des n sinistres ;
- $j \in \llbracket 1, k \rrbracket$ identifiant des k facteurs de risque ;
- Y_i réponse observée du $i^{\text{ème}}$ individu/sinistre (coût total du sinistre) ;
- $X_i = (X_{i1}, \dots, X_{ik})$ est le vecteur des facteurs de risque du $i^{\text{ème}}$ sinistre.

L'objectif d'une régression classique est d'estimer la quantité d'intérêt :

$$\pi_0(x) = \mathbb{E}[Y|X = x]$$

Or, il est difficile dans le cadre des arbres de régression de procéder à l'évaluation de la valeur optimale des estimateurs via l'erreur quadratique (Moindre Carré Ordinaire ou MCO) en raison du temps de calcul que cela engendrerait. L'algorithme permettant la résolution de ce problème repose donc sur

une procédure récursive visant à minimiser l'erreur quadratique moyenne, sans avoir à estimer l'ensemble des estimateurs possibles.

Pour chaque nœud, l'algorithme cherche alors les deux sous-ensembles les plus homogènes (qui minimisent donc la variance intra-nœuds) et ce, en testant l'ensemble des segments de chaque variable explicative. Les feuilles qui en résultent ont pour estimateur la moyenne empirique des Y_i des observations restantes.

L'arbre grandit alors jusqu'à atteindre sa taille maximale. Cette taille est atteinte lorsque chaque feuille ne comporte plus :

- qu'une seule observation ;
- ou que toutes les observations de la feuille aient les mêmes caractéristiques.

L'arbre maximale n'est évidemment pas exploitable. En effet, attribuer une estimation à chaque sinistre, à un niveau aussi fin, entraîne nécessairement un phénomène de surapprentissage (les estimations des Y_i ne seraient pas adaptées à un autre jeu de données). C'est d'autant plus vrai sur une base de données où la volumétrie est relativement restreinte.

L'arbre doit donc être élagué, c'est-à-dire réduit afin de limiter le nombre de divisions qui n'entraînerait pas d'apport significatif à l'estimation de la variable d'intérêt. La complexité de cette étape repose donc sur l'équilibre entre les bonnes capacités prédictives de l'arbre et le risque de surapprentissage.

Deux méthodes sont alors possibles :

- ✓ un élagage a posteriori

Dans ce premier cas, la méthode consiste à partir de l'arbre maximal, et d'élaguer les feuilles/nœuds. Le critère étant de minimiser le **coût de la complexité**. Ce coût est constitué des variances de chaque sous-arbre et de la taille de ces derniers. Il est représenté par l'indicateur $\alpha \geq 0$.

L'arbre est maximal lorsque que $\alpha = 0$, et il est minimal lorsque α est grand. L'objectif étant de trouver un α faisant le compromis entre la qualité des estimations et la taille de l'arbre.

- ✓ un élagage a priori

Dans ce deuxième cas, la méthode consiste à construire l'arbre maximal, comme présenté ci-dessus, mais en changeant le critère de non-division qu'était l'unicité de l'individu dans une feuille. Le principe est donc de définir le seuil à partir duquel le nombre d'individus minimal dans une feuille est considéré comme étant trop faible pour être pertinent.

Dans ce mémoire, les deux formes d'élagage seront analysées, mais la seconde méthode semble plus adaptée sur des jeux de données dont le volume est restreint.

La méthode CART a aussi le net avantage d'être visuelle et les résultats sont facilement interprétables. Il n'en reste pas moins que, si aucun regroupement préalable sur les modalités des variables explicatives n'est fait en amont et si l'arbre n'est pas très élagué, il est difficile de dégager des profils « simples » qui puissent être facilement communiqués. Le provisionnement n'ayant pas pour but d'avoir une utilité commerciale comme la tarification ou le profiling, le choix sera ici fait de privilégier la précision des résultats plutôt que la simplification de la communication de ces derniers.

Hormis le surapprentissage, l'autre inconvénient de cette méthode est sa sensibilité aux optimums locaux. Ces derniers influent sur les divisions, généralement dans les premiers nœuds. L'algorithme n'étant pas fait pour réétudier les critères déjà utilisés, il est possible que la construction même de l'arbre soit différente sans certains optimums (et peut-être meilleur). L'algorithme ne calcule pas toutes les combinaisons possibles d'arbres, ce qui le rend forcément perfectible, mais aussi pour cette raison que les temps de calculs sont si intéressants.

Un moyen d'améliorer ces arbres est de passer par la méthode de régression non-paramétrique des forêts aléatoires ou *Random forest*. Cette méthode permet de générer une multitude d'arbre en tirant aléatoirement les observations, et en faisant pivoter les critères à disposition. Les arbres seront individuellement de moins bonne qualité mais ils sont ensuite agrégés (généralement par régression) pour former un arbre moyen.

Cette méthode est flexible et plus robuste, et permet de limiter le problème des optimums locaux. En revanche, elle est opérationnellement moins simple à mettre en œuvre, notamment pour les interprétations, avec des temps de calculs plus longs.

Le choix sera fait de ne pas appliquer cette méthode dans ce mémoire, mais elle reste une solution alternative qui pourrait potentiellement être travaillée dans le futur.

B. Mesures de validation

Afin de conforter les résultats de cette méthode d'apprentissage statistique, les mesures suivantes seront systématiquement prises :

- ✓ Découpage entre base d'apprentissage et base de validation

La base de données sera séparée en deux bases :

- une base d'apprentissage représentant 80% de la base totale
- une base de validation

La base d'apprentissage sera utilisée pour construire le modèle. Celui-ci sera appliqué à la base de validation pour mesurer sa pertinence à l'aide des indicateurs ci-dessous.

La validation du modèle sur la base de validation permet d'éviter le problème de surapprentissage

- ✓ Indicateurs

Afin de tester la performance du modèle, sur les estimations des coûts totaux, des indicateurs classiques de distance seront utilisés :

- La racine carrée de l'erreur quadratique moyenne (RMSE ou *Root Mean Squared Error*)

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

- L'erreur absolue moyenne (MAE ou *Mean Absolute Error*)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

Tous ces outils mathématiques sont en capacité de pouvoir répondre aux problématiques de provisionnement des sinistres de long terme comme peuvent l'être les sinistres corporels en Automobile. Ils permettront de reprendre les fondamentaux et de mesurer l'incertitude, notamment sur les sinistres graves. Ces derniers auront pu être identifiés par des notions de valeurs extrêmes. Une proposition alternative aux méthodes classiques sera alors explorée, via le provisionnement ligne à ligne.

Chapitre 3 : Exploration de la sinistralité Automobile

I. La prise de recul de l'Actuaire

Les méthodes et les calculs font parties intégrantes du métier de l'Actuaire. Il se base sur des fondamentaux mathématiques pour extraire des données l'information recherchée. Il se doit ensuite de prendre du recul sur les résultats, et de restituer ces derniers en s'adaptant au mieux à son public.

Dans cette première partie, la problématique de l'évaluation des charges à l'ultime des sinistres corporels en Automobile sera réintroduite. Pour cela, le parallèle sera fait entre les résultats des méthodes agrégées sur le provisionnement Non-Vie sur les sinistres Automobile Autre que RC, n'impliquant aucun dommage corporel ; et les sinistres RC impliquant des dommages corporels.

Un exemple sera mis en avant pour montrer que la prise de recul doit être omniprésente dans le métier de l'Actuaire pour ne pas mettre en péril une société d'assurance.

A. Chain Ladder et vérifications des hypothèses

Dans cette partie, deux types de sinistre Automobile seront comparés :

- les sinistres Autre que Responsabilité Civile dans un accident n'incluant pas de dommage corporel ; c'est classiquement le cas des Bris de Glace, des Vols, ou des sinistres avec ou sans véhicule tiers, généralement sans gravité, impliquant la garantie Tous Risques (petits chocs, parkings, ...)
- les sinistres Responsabilité Civile suite à un accident ayant causé des dommages corporels, quel que soit la gravité.

Ces deux types de sinistres sont diamétralement différents :

Sur la période 2003 - 2019	Autre que RC n'impliquant pas de dommage corporel	RC impliquant des dommages corporels
Nombre de sinistres annuel moyen	21 794	1 140
Coût moyen au 31/12/2019	1 090 €	10 454 €
Temps de gestion moyen du sinistre en nombre de jours (sinistres clos)	105	522
Montant du sinistre maximum	129 695 €	5 459 375 €

Tableau 3 : Comparaison des caractéristiques des sinistres Automobile

Les sinistres Autre que RC sans dommage corporel sont des sinistres courants et donc nombreux, qui coûtent relativement peu cher à l'entreprise d'assurance individuellement (environ 1 000€) et qui sont rapidement traités (en 3 mois en moyenne). Surtout, ces sinistres n'ont qu'une ampleur maximum très limitée, et sur 17 années de survenance, le sinistre maximum atteint est de 130K€. Le sinistre maximum possible de ce type de sinistre ne serait pas beaucoup plus élevé.

Les sinistres RC « Corporel » sont en tout point différents de ces derniers. Les sinistres sont plutôt rares, individuellement coûteux, et mettent presque un an et demi à être traités en moyenne. La

différence la plus marquante se situe sur le montant maximum constaté sur la période : 5,5M€, 42 fois plus, sachant que le sinistre maximum possible pourrait très bien dépasser les 100M€.

Le montant des provisions dossiers/dossiers par année de survenance montre également que la problématique d'estimation globale à l'ultime est très différente entre ces deux types de sinistre :

Provisions dossiers/dossiers au 31/12/2019	Autre que RC n'impliquant pas de dommage corporel	RC impliquant des dommages corporels
2003	-20 €	400 701 €
2004	0 €	37 240 €
2005	0 €	18 328 €
2006	0 €	4 638 142 €
2007	0 €	39 122 €
2008	0 €	323 640 €
2009	-23 254 €	1 118 310 €
2010	-1 351 €	52 181 €
2011	-30 682 €	1 068 828 €
2012	520 €	3 391 482 €
2013	-10 350 €	3 086 752 €
2014	384 €	3 532 211 €
2015	-17 079 €	2 658 876 €
2016	-38 937 €	3 547 362 €
2017	67 827 €	8 238 419 €
2018	271 794 €	7 977 225 €
2019	2 421 735 €	2 812 202 €
ENSEMBLE :	2 640 587 €	42 941 021 €

Tableau 4 : Comparaison des montants des provisions dossiers/dossiers

Hormis quelques dossiers avec des recours, il ne reste que peu de dossiers encore ouverts sur les sinistres Automobile Autre que RC sans dommages corporels avec une ancienneté de plus de 3 années. L'incertitude liée à l'évaluation des provisions à l'ultime se situe essentiellement sur les 3 années de survenance les plus récentes.

A contrario, il reste presque 3 fois plus de sinistres ouverts sur les sinistres Automobile RC « Corporel », que sur les sinistres Autre que RC, sur les années de survenance 2003 à 2016 (alors qu'il y a près de 20 fois plus de sinistre total par an en moyenne sur les sinistres Autre que RC). Ces sinistres ont, de plus, des montants de provisions dossiers/dossiers importants. À noter que le montant de 43M€ de provisions RC « Corporel » représente un quart des provisions totales de la MAPA, toutes branches confondues.

Revenons maintenant sur l'application de la méthode Chain Ladder. Les choix suivants ont été faits :

- ✓ l'application de la méthode se fait sur les paiements nets de recours ;
- ✓ le développement des années récentes est généralement le plus significatif dans l'évaluation de la provision totale à l'ultime, une attention particulière sera donc accordée sur les premiers facteurs de développement, et, pour une année de développement donnée (pour j fixé), les éléments suivants seront pris en compte :
 - les facteurs de développement individuels récents ;

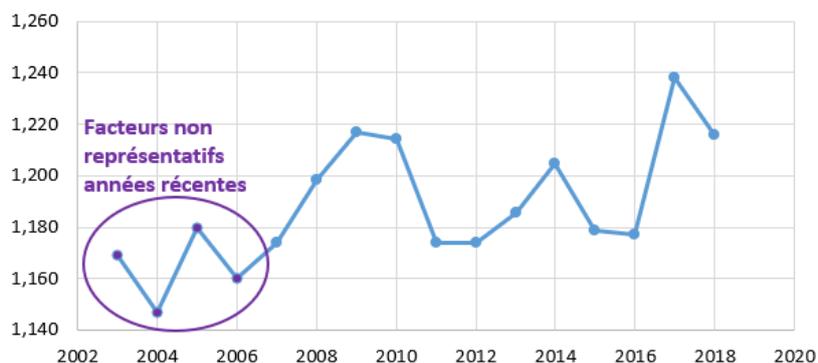


Figure 20 : Facteurs de développement individuel de la 1^{ère} année de développement Automobile Matériel

- les tendances sur les facteurs de développement individuels ;
 - le retrait des facteurs de développement individuels atypiques ;
- Un facteur de développement individuel est considéré comme atypique, pour j fixé, lorsque :

$$\begin{cases} \hat{f}_{ji}^{CL} < \frac{1}{N-j} \sum_{i=0}^{N-j} \hat{f}_{ji}^{CL} - 2 \times \sqrt{\frac{\sum_{i=0}^{N-j} \left(\frac{1}{N-j} \sum_{i=0}^{N-j} \hat{f}_{ji}^{CL} - \hat{f}_{ji}^{CL} \right)^2}{N-j-1}} \\ \hat{f}_{ji}^{CL} > \frac{1}{N-j} \sum_{i=0}^{N-j} \hat{f}_{ji}^{CL} + 2 \times \sqrt{\frac{\sum_{i=0}^{N-j} \left(\frac{1}{N-j} \sum_{i=0}^{N-j} \hat{f}_{ji}^{CL} - \hat{f}_{ji}^{CL} \right)^2}{N-j-1}} \end{cases}$$

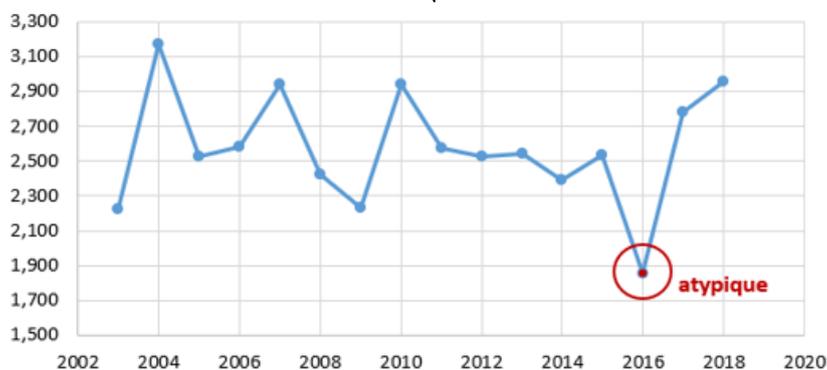


Figure 21 : Exemple de facteur de développement individuel considéré comme atypique

- des liens peuvent également être fait entre cette partie et la validation des hypothèses afférentes à la méthode Chain Ladder.

La méthode étant décrite ci-dessus, les détails sur les choix qui ont pu être faits dans le calcul des facteurs de développement ne seront pas abordés.

Les facteurs de développement sur les deux périmètres étudiés sont finalement les suivants :

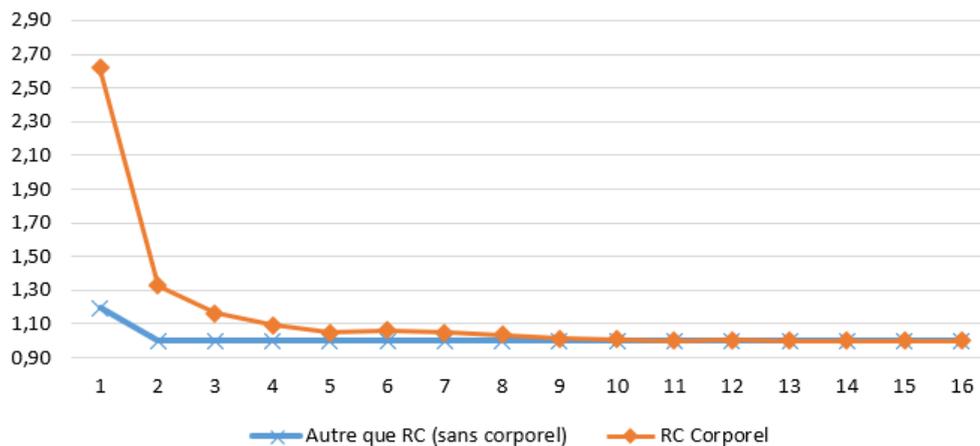


Figure 22 : Comparaison des facteurs de développement RC « Corporel » / Autre que RC

Les facteurs de développement de la partie Autre que RC montrent bien le fait que ce type de sinistre est assez rapide à gérer, et, hormis entre la toute première année et la seconde, il n’y a que très de peu de règlements, la majeure partie des sinistres sont déjà clos. Ce constat coïncide parfaitement avec la répartition des provisions dossiers/dossiers où 90% des provisions totales se situe sur la dernière année de survenance.

Pour ce qui est des sinistres RC « Corporel », les cadences courent jusqu’à la 10^{ème} année de développement. Pourtant, des provisions dossiers/dossiers importantes sont présentes sur les années de survenance les plus anciennes.

Avant de faire le comparatif entre les provisions dossiers/dossiers et les provisions à l’ultime calculées via la méthode Chain Ladder, vérifions les hypothèses relatives à l’utilisation de cette méthode.

Les hypothèses sont, là aussi, essentiellement vérifiées sur les premiers facteurs de développement, au vu de leur importance dans le modèle.

Seule la validation sur la première année de développement sera ici présentée, le principe étant ensuite répété.

Vérification de l’hypothèse sur les facteurs de développement :

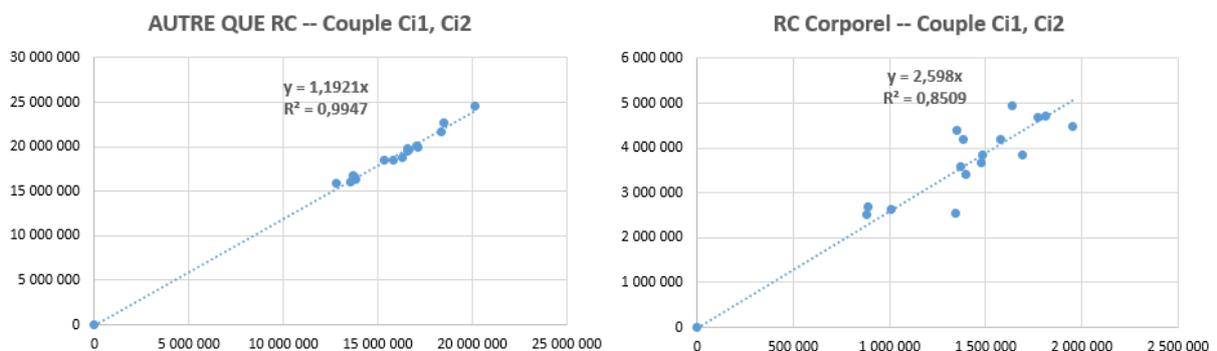


Figure 23 : Vérifications des hypothèses sur les facteurs de développement

Les paiements cumulés individuels entre la première et la seconde année de développement sont bien alignés avec à l'origine, ce qui vérifie que l'évolution des paiements suit un processus Markovien.

On remarquera cependant que cette hypothèse est bien plus fiable sur les sinistres Autre que RC que sur les sinistres RC « Corporel ».

Vérification de l'hypothèse sur la volatilité interne du triangle :

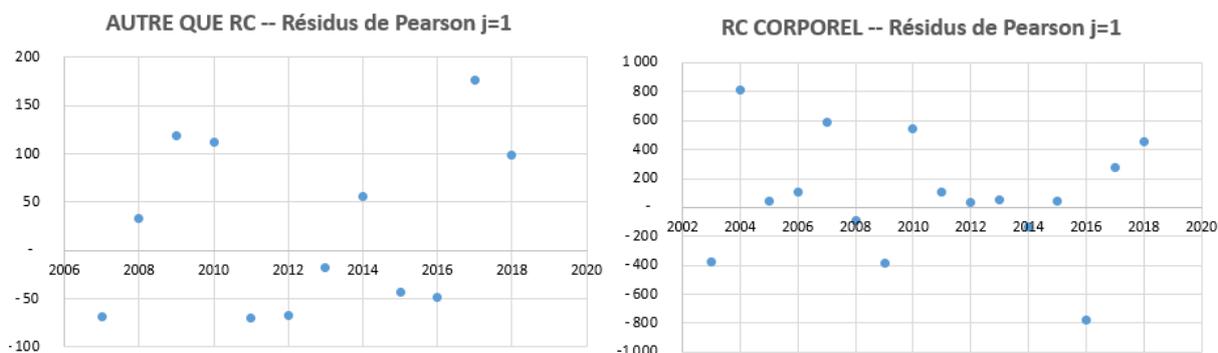


Figure 24 : Vérifications des hypothèses sur la volatilité interne des triangles

Il y a bien une absence de tendance sur les résidus de Pearson sur la première année de développement, quel que soit le type de sinistre. À noter tout de même qu'une légère tendance était visible sur les sinistres Autre que RC, en raison des résidus des survenances 2003-2006, qui ont donc été retirés. Cela coïncidait avec l'analyse des facteurs de développement individuels.

B. Résultats et cohérence

Les hypothèses étant vérifiées, analysons les résultats suivants :

AUTRE QUE RC	PROV_D/D	PROV_ULTIME	Différence	RC CORPOREL	PROV_D/D	PROV_ULTIME	Différence
2003	- 20	-	20	2003	400 701	-	- 400 701
2004	-	25	25	2004	37 240	4 262	- 32 978
2005	-	27	27	2005	18 328	6 731	- 11 597
2006	-	43	43	2006	4 638 142	7 585	- 4 630 557
2007	-	88	88	2007	39 122	9 379	- 29 743
2008	-	117	117	2008	323 640	48 560	- 275 080
2009	- 23 254	172	23 082	2009	1 118 310	93 857	- 1 024 453
2010	- 1 351	6 367	5 016	2010	52 181	192 973	140 792
2011	- 30 682	5 833	24 849	2011	1 068 828	284 376	- 784 452
2012	520	13 869	14 389	2012	3 391 482	459 468	- 2 932 014
2013	- 10 350	14 826	4 476	2013	3 086 752	1 033 656	- 2 053 096
2014	384	10 743	11 127	2014	3 532 211	1 087 614	- 2 444 597
2015	- 17 079	10 308	6 771	2015	2 658 876	1 094 421	- 1 564 455
2016	- 38 937	11 930	27 007	2016	3 547 362	1 436 852	- 2 110 510
2017	67 827	5 248	73 075	2017	8 238 419	2 163 693	- 6 074 726
2018	271 794	14 103	257 691	2018	7 977 225	3 007 870	- 4 969 355
2019	2 421 735	2 810 630	388 895	2019	2 812 202	2 796 213	- 15 989
ENSEMBLE :	2 640 587	2 745 137	104 550	ENSEMBLE :	42 941 021	13 727 510	-29 213 511
			4,0%				-68,0%

Tableau 5 : Comparaison des résultats de l'estimation des provisions à l'ultime entre Automobile Autre que RC et Automobile RC « Corporel » par la méthode classique Chain Ladder

Pour ce qui est des sinistres Autre que RC, la provision à l'ultime de Chain Ladder est supérieure de 4% à celle des provisions dossiers/dossiers. Cette différence est assez faible, et est cohérente. En effet, cette provision à l'ultime comprend, en plus de l'évaluation à l'ultime des sinistres encore ouverts au 31/12/2019, l'évaluation des sinistres non encore déclarés (IBNYR), n'étant pas intégrée aux provisions dossiers/dossiers. Chain Ladder montre ici toute sa pertinence et son efficacité.

Ce n'est en revanche pas le cas des sinistres RC « Corporel ». Hormis l'année de survenance 2019, aucune provision à l'ultime par année de survenance ne semble adaptée aux provisions dossiers/dossiers. La provision à l'ultime sur l'ensemble des années de survenance semble complètement déconnectée des provisions totales dossiers/dossiers (-68%).

Une autre manière de valider, ou non, ces résultats est de faire une analyse des bonis/malis. Pour rappel, les bonis/malis représentent la différence entre les paiements effectués sur une année de développement et la variation de provision dossiers/dossiers entre cette année de développement et la précédente. Cette mesure permet d'apprécier la manière dont l'entreprise gère l'évaluation de ses sinistres.

Sur la partie Autre que RC, la première année de développement est marquée par des montants importants de malis, avec près d'1,5M€ par an en moyenne (sachant que le coût total moyen à la fin de l'année de développement 0 est de 18M€ sur les 17 années observées, soit +9%). Une analyse complémentaire sur cette première année de développement montre que ces malis proviennent pour près de 50% d'entre eux des sinistres dits tardifs (IBNYR).

Bonis / Malis										
AUTRE QUE RC	B/M_0	B/M_1	B/M_2	B/M_3	B/M_4	B/M_5	B/M_6	B/M_7	B/M_8	B/M_9
2003	-	1 260 226	- 407 807	- 76 967	5 112	- 1 789	2 851	- 49	- 87	- 40
2004	-	1 326 934	- 292 300	- 42 745	- 7 619	- 2 930	- 1 326	1	- 588	750
2005	-	1 665 159	- 374 970	- 75 422	5 154	- 1 166	201	20 089	- 3 467	-
2006	-	1 650 731	- 332 611	- 109 379	706	- 83 312	3 980	- 4 426	4 871	- 4
2007	-	1 775 167	- 334 833	- 41 517	- 14 386	3 257	- 2 701	- 3 537	- 400	-
2008	-	1 589 691	- 314 109	- 74 064	- 617	- 2 504	3 389	- 1 805	-	-
2009	-	2 032 183	- 414 074	- 91 982	19 120	81 544	27 119	5 319	- 76 147	- 3 760
2010	-	2 341 006	- 404 332	- 159 392	15 128	- 25 151	647	- 4 411	- 1 064	363
2011	-	1 725 564	- 414 491	- 27 535	4 733	7 641	2 311	3 233	- 29 466	-
2012	-	1 622 036	- 343 455	- 81 074	- 804	- 4 811	- 1 485	5 863	-	-
2013	-	1 582 074	- 308 542	- 64 350	- 66 297	-	3 698	- 15 823	-	-
2014	-	1 208 317	- 331 420	- 57 126	- 1 400	- 1 092	-	-	-	-
2015	-	1 193 307	- 226 282	- 118 482	1 417	-	-	-	-	-
2016	-	1 136 787	- 257 275	- 48 780	-	-	-	-	-	-
2017	-	1 512 752	- 286 129	-	-	-	-	-	-	-
2018	-	1 530 013	-	-	-	-	-	-	-	-
2019	-	-	-	-	-	-	-	-	-	-
Moyenne		1 571 997	- 336 175	- 76 344	- 3 058	- 2 218	1 706	2 028	- 11 816	- 336

Tableau 6 : Triangle des bonis/malis sur les sinistres Automobile Autre que RC

Pour les autres années de développement, la MAPA dégage clairement des bonis, notamment sur la 2^{ème} et la 3^{ème} année de développement. Ensuite, tout comme les facteurs de développement, les variations sont faibles.

Ce regard sur les bonis/malis permet de conforter les résultats obtenus via la méthode Chain Ladder sur le périmètre des sinistres Automobile Autre que RC sans dommages corporels.

On remarquera également la stabilité des bonis/malis par année de survenance.

Pour ce qui est des sinistres Automobile RC « Corporel », le triangle des bonis/malis est le suivant :

Bonis / Malis										
RC CORPOREL	B/M_0	B/M_1	B/M_2	B/M_3	B/M_4	B/M_5	B/M_6	B/M_7	B/M_8	B/M_9
2003	-	2 354 916	175 431	641 486	20 721	400 829	89 290	49 792	583 245	- 145 235
2004	-	4 947 830	3 144 108	- 1 006 533	533 063	- 2 856 128	1 160 250	124 818	- 881 009	1 497 724
2005	-	2 884 279	222 802	- 150 166	- 25 942	- 69 098	- 266 834	- 48 983	- 53 596	- 91 246
2006	-	914 934	314 813	450 185	478 297	- 848 369	987 794	51 353	- 229 659	838 781
2007	-	2 474 419	948 336	942 306	- 1 110 453	25 351	515 279	257 656	217 518	- 124 976
2008	-	2 882 355	690 204	723 688	885 201	197 623	147 738	327 000	- 1 398 599	- 1 944
2009	-	- 498 157	2 433 463	184 320	3 517 201	- 390 152	334 658	248 707	- 2 668 886	150 986
2010	-	44 865	975 833	54 903	7 629	- 635 818	- 181 857	102 119	- 248 832	- 721 147
2011	-	2 573 779	237 228	- 164 583	- 1 125 594	363 188	- 45 112	- 632 526	- 1 124 181	
2012	-	56 811	309 519	- 124 224	45 020	294 553	- 32 206	31 803		
2013	-	1 090 155	99 818	- 688 674	176 152	- 230 219	478 325			
2014	-	1 248 869	- 1 416 266	399 536	509 139	- 168 583				
2015	-	- 1 164 895	- 369 148	- 112 392	- 1 168 479					
2016	-	1 111 840	- 350 355	- 930 741						
2017	-	1 156 857	- 205 843							
2018	-	2 807 985								
2019	-									
Moyenne		1 555 428	480 663	15 651	210 920	- 326 402	289 757	51 174	- 644 889	175 368

Tableau 7 : Triangle des bonis/malis sur les sinistres Automobile RC « Corporel »

Contrairement aux sinistres Autres que RC, l'instabilité de cette mesure est évidente, quel que soit l'année de développement considérée. En moyenne, sur la première année de développement, les sinistres RC « Corporel » engendrent des malis (+15% de coût total par rapport à celui de l'année de développement 0), pour ensuite alterner entre bonis et malis.

Les montants restent significatifs jusqu'au bout du triangle (Annexes), ce qui semble logique puisque des sinistres sont encore ouverts sur les années les plus anciennes avec des montants de provisions dossiers/dossiers importants.

L'exercice a été fait d'appliquer un Chain Ladder sur le triangle des coûts totaux, afin d'estimer les bonis/malis à l'ultime : la MAPA dégagerait alors approximativement 5M€ de bonis, toutes années de survenance confondues.

Même s'il sera toujours préféré d'évaluer la provision à l'ultime via un triangle de règlements, cette approche basée sur le coût total permet néanmoins de confirmer le ressenti sur le manque de cohérence d'une évaluation de la provision à l'ultime plus faible de 68% par rapport aux provisions dossiers/dossiers.

Cela s'explique par le manque d'historique sur ces sinistres, et notamment sur les années de développement.

Afin d'affiner les résultats de cette estimation, une solution consiste à segmenter ces sinistres RC « Corporel », entre les sinistres graves et attritionnels, ce qui permettra :

- d'avoir une évaluation de qualité sur les sinistres attritionnels, à l'instar des sinistres Autre que RC (un comparatif sera alors fait entre les deux populations) ;
- d'essayer de pallier le problème d'historique sur les sinistres graves avec les tails factors, ce qui n'était pas possible sur l'ensemble des sinistres RC « Corporel » en raison du poids des attritionnels ;
- de faire l'analyse de l'incertitude autour des différentes évaluations.

II. Seuil des sinistres graves

Les provisions à l'ultime des sinistres Automobile Autre que RC peuvent être correctement évaluées avec la méthode déterministe Chain Ladder, sans qu'il y ait à affiner la segmentation (ce qui ne veut pas dire qu'une segmentation plus fine ne permettrait pas d'avoir des résultats meilleurs). En revanche, ce n'est pas le cas des sinistres RC « Corporel », qui dénombrent plus de 200 sinistres dont les coûts totaux sont bien supérieurs au montant maximum observé sur les sinistres Autre que RC.

A. Correction de l'inflation et mise à l'ultime

L'objectif de cette partie est de déterminer le seuil à partir duquel un sinistre est considéré comme grave. Pour se faire, il est donc nécessaire de pouvoir comparer les sinistres. Or, comme le montre la partie D.2.b du chapitre 1, le coût moyen des sinistres n'a cessé d'augmenter. Cela s'explique à la fois par la montée en gamme des véhicules, de mieux en mieux équipés, notamment de dispositifs d'aide à la conduite ; mais aussi de l'évolution générale du coût de la vie et de l'amélioration de la prise en charge des sinistres corporels, qui ont un impact sur les indemnisations liées aux préjudices corporels.

La première étape avant de pouvoir comparer les sinistres de 17 années de survenance différentes entre eux est donc de corriger l'inflation et de faire en sorte que tous les sinistres soient « comme si » (« As-If ») ils avaient eu lieu au cours de l'année de survenance 2019.

Le choix s'est porté sur l'application ce qui est observé sur le portefeuille de la MAPA, en fonction des différents axes proposés dans le chapitre 1, soit :

- +2,7% par an en moyenne sur les sinistres n'impliquant que du matériel, que les garanties soient de la RC ou des garanties Autre que RC ;
- +0,3% par an en moyenne pour les sinistres impliquant des dommages corporels, et qui concernent les garanties Autre que RC ;
- +4,1% par an en moyenne sur les sinistres impliquant des dommages corporels, en RC.

Cette correction de l'inflation est appliquée aux paiements, en fonction de l'année de survenance du sinistre et du moment (année de développement) où le paiement a été effectué. Le niveau des provisions à fin 2019 est considéré comme déjà corrigé. Les sinistres corrigés de l'inflation sont donc de la forme suivante :

$$\text{Coût}_{total}^{corrigé} = \sum_{j=0}^N \left(\text{Paiement}_j \times (1 + Tx_{inflation})^{(2019 - (\text{Année}_{surv} + j))} \right) + \text{Provisions}_{2019}$$

avec j les années de développement (*pour* $j = 0, \dots, N$) et $Tx_{inflation}$ le taux d'inflation correspondant au type de sinistre.

Pour compléter cette approche, le coût total corrigé de l'inflation a été mis à l'ultime, pour que les sinistres soient comme s'ils avaient le même nombre d'années de développement. Le but ici est de faire une correction « simple » puisque l'étude des seuils doit permettre de faire des estimations des provisions à l'ultime, et non l'inverse.

Cette mise à l'ultime a été faite à partir de l'ensemble des sinistres Automobile, sans distinction, à partir d'un triangle de coût total, avec la méthode Chain Ladder. La formule est la suivante :

$$Coût_total_i^{Corrigé_Ultime} = Coût_total_i^{Corrigé} \times \prod_{j=N-i+1}^N \hat{f}_j^{CL}$$

avec i les années de survenance (pour $i = 0, \dots, N$)

Les résultats obtenus sont les suivants, et démontrent l'importance de cette étape lorsque des sinistres de survenances différentes sont comparés :

Coûts totaux au ...	31/12/2019	31/12/2019 avec règlements corrigés de l'inflation	Différence en %	Coûts totaux au ...	31/12/2019 avec règlements corrigés de l'inflation	31/12/2019 avec inflation et mise à l'ultime	Différence en %
2003	33 999 753	52 046 831	53,1%	2003	52 046 831	52 046 831	0,0%
2004	35 789 618	53 624 803	49,8%	2004	53 624 803	53 621 367	0,0%
2005	35 591 767	51 940 765	45,9%	2005	51 940 765	51 937 289	0,0%
2006	40 472 878	55 250 468	36,5%	2006	55 250 468	55 139 216	-0,2%
2007	37 418 315	51 492 103	37,6%	2007	51 492 103	51 103 507	-0,8%
2008	34 361 732	45 631 610	32,8%	2008	45 631 610	45 280 729	-0,8%
2009	45 199 475	58 230 579	28,8%	2009	58 230 579	57 797 958	-0,7%
2010	44 103 878	55 507 609	25,9%	2010	55 507 609	55 030 873	-0,9%
2011	39 540 484	48 397 301	22,4%	2011	48 397 301	48 141 458	-0,5%
2012	38 045 237	44 734 847	17,6%	2012	44 734 847	43 950 755	-1,8%
2013	38 067 537	43 692 567	14,8%	2013	43 692 567	42 970 099	-1,7%
2014	32 304 215	36 093 387	11,7%	2014	36 093 387	35 704 970	-1,1%
2015	29 064 285	31 761 743	9,3%	2015	31 761 743	31 213 489	-1,7%
2016	29 306 271	31 220 400	6,5%	2016	31 220 400	30 763 810	-1,5%
2017	33 725 341	34 848 095	3,3%	2017	34 848 095	34 240 178	-1,7%
2018	33 905 004	34 387 114	1,4%	2018	34 387 114	33 659 036	-2,1%
2019	25 682 787	25 682 787	0,0%	2019	25 682 787	28 233 559	9,9%

Tableau 8 : Impact de la prise en compte de l'inflation et la mise en ultime sur les sinistres Automobile

B. Analyses statistiques

Il est assez évident en s'intéressant aux fonctions de répartition des sinistres que les sinistres Automobile RC « Corporel » doivent être traités différemment des autres sinistres Automobile :

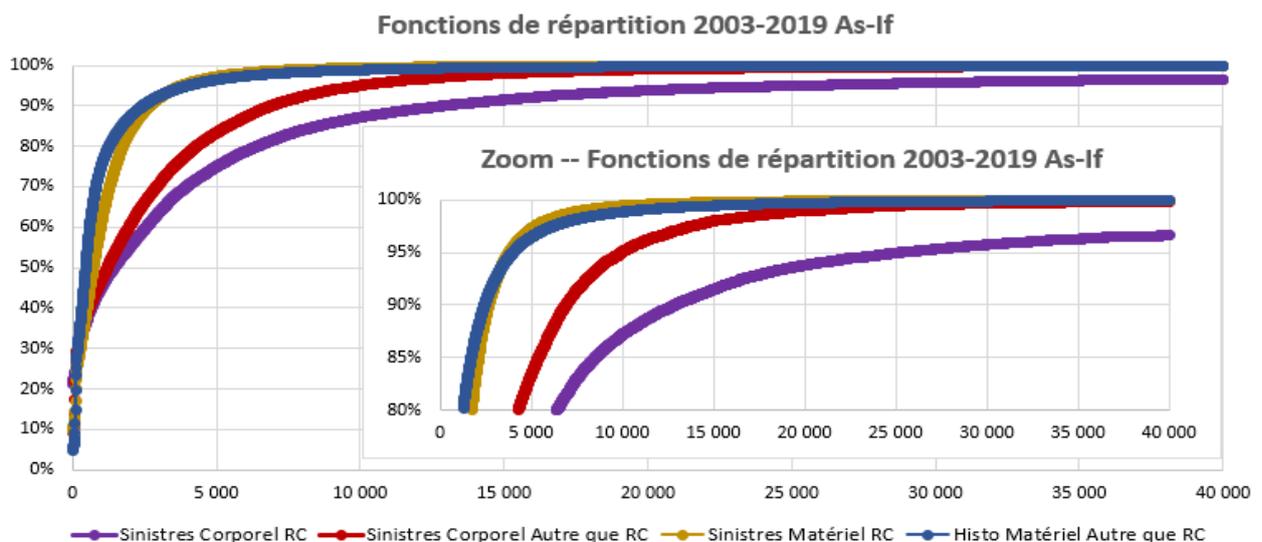


Figure 25 : Fonctions de répartition des différentes catégories de sinistre Automobile mis en As-If

Les sinistres Automobile qui n'impliquent pas de dommages Corporels sont relativement similaires en termes de répartition des sinistres entre la garantie RC et les autres garanties. Ces deux types de sinistres atteignent très vite les centiles élevés.

Les sinistres Automobile Corporels se comportent différemment des sinistres purement Matériel, mais aussi différemment entre les sinistres RC et Autre que RC. Les sinistres RC « Corporel » ont une queue de distribution bien plus épaisse :

Sinistres As-If à l'ultime	RC dans un accident corporel	Autre que RC dans un accident corporel	RC dans un accident matériel	Autre que RC dans un accident matériel
Nombre de sinistres	19 381	11 525	98 109	370 502
Nombre <= 0	4 121	1 052	9 371	18 101
MAX	5 758 797	103 626	511 967	149 680
Q99,9	1 053 557	50 481	19 599	30 895
Q99	149 415	20 653	7 777	10 752
Q98	73 069	14 969	5 791	7 093
Q90	12 883	6 926	2 678	2 447
Q75	5 015	3 531	1 504	998
Moyenne	10 380	2 683	1 189	1 085
Q50	1 443	1 213	734	460
Q40	659	581	509	337
Q30	224	187	279	183
Q20	0	100	125	126
MIN	-10 847	-7 851	-9 582	-10 993

Tableau 9 : Statistiques des différentes catégories de sinistre Automobile mis en As-If

La différence en termes de quantiles est flagrante entre les sinistres RC « Corporel » et les autres types de sinistre, notamment à partir du quantile à 90%. De plus, la proportion de sinistres négatifs ou nuls est bien plus importante (supérieur à 20%, contre moins de 10% pour les autres types de sinistre).

Cette analyse simple de la sinistralité permet déjà de confirmer que les sinistres RC « Corporel » sont les seuls sinistres à avoir une problématique de sinistres graves (le quantile à 99,9% des autres types de sinistre étant de maximum 50K€).

Une première approche sur le seuil de sinistre grave en RC « Corporel » est de prendre le quantile à 99%, soit les sinistres supérieurs à 150K€.

C. Fonction Moyenne des Excès (FME) et détermination du seuil

L'étude des sinistres graves se fera donc uniquement sur les sinistres RC « Corporel ». Pour rappel, l'objectif de la fonction moyenne des excès est de trouver un seuil, le plus petit possible, permettant à l'estimateur empirique de l'espérance résiduelle des excès au-delà de ce seuil d'être approximativement linéaire.

Dans la pratique, les sinistres suivants ont été retirés :

- les sinistres dont les montants sont inférieurs à 10€, puisqu'ils n'apportent que peu d'informations et alourdissent le processus. Ces sinistres ont été totalement retirés ;
- les sinistres les plus graves, c'est-à-dire dont le montant total dépasse les 2M€ (5 sinistres), puisque la fonction moyenne des excès est très sensible à ces valeurs, qui ne sont finalement pas réellement interprétables. Ces sinistres ont été intégrés au calcul des excès moyen de tous les sinistres, mais ne sont pas représentés graphiquement.

Fonction Moyenne des Excès des sinistres RC Corporel

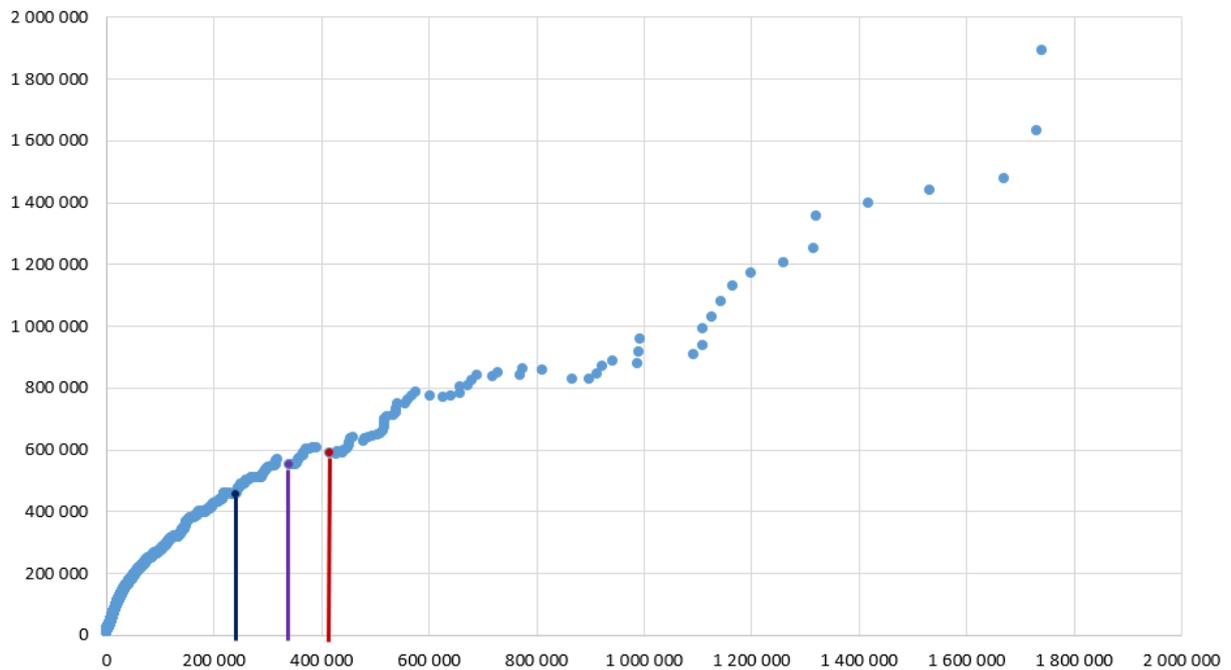


Figure 26 : Fonction moyenne des Excès des sinistres RC « Corporel »

La première chose que dégage ce graphique est le type de queue de distribution des sinistres RC « Corporel » : sans surprise, **la queue de distribution de ces sinistres est lourde de type Pareto ou GPD**. En effet, et sans connaître le point de rupture entre les sinistres graves et les sinistres attritionnels, quel que soit ce seuil, la pente de la fonction des excès serait positive.

L'objectif de l'analyse graphique est ici de trouver le point à partir duquel les excès moyens se retrouvent sur une même pente linéaire. Tout d'abord, la fonction observée est relativement concave sur les sinistres allant de 10€ jusqu'aux environs des 200K€. Ensuite, trois points de rupture sont visibles, avant que les points ne soient plus épars :

1. le seuil à 240K€ en **bleu foncé** ;
2. le seuil à 325K€ en **violet** ;
3. le seuil à 400K€ en **rouge**.

D'autres points supérieurs à ces trois seuils sont sûrement possibles, mais l'objectif est bien de trouver le plus petit afin d'obtenir un nombre de sinistres le plus grand possible.

L'analyse des pentes ne permet pas de faire un choix sur le seuil à choisir :

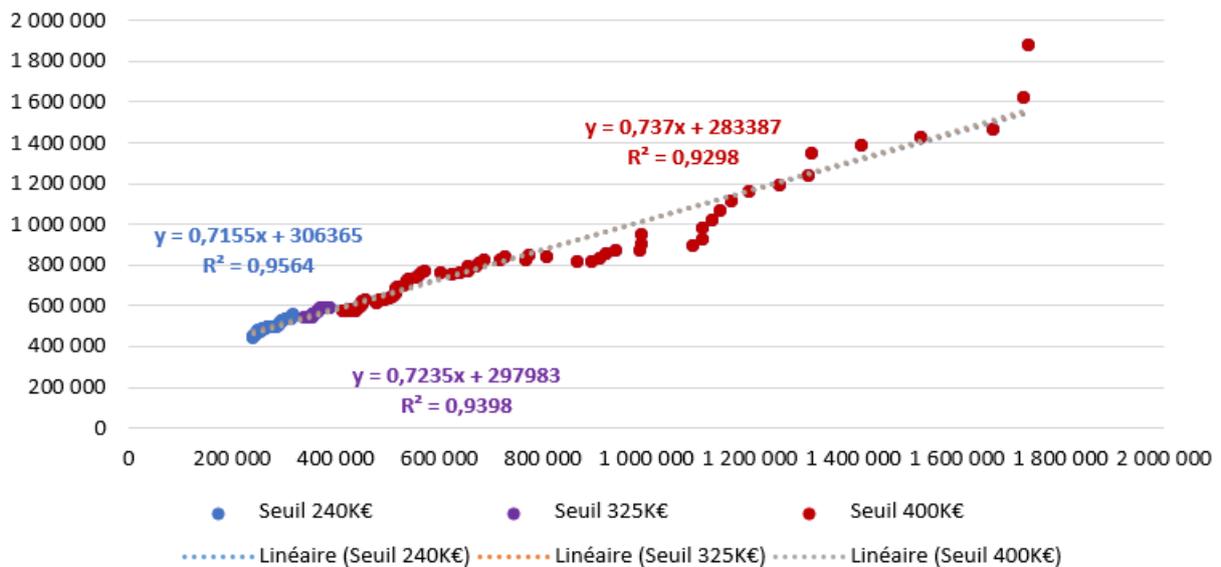


Figure 27 : Pentes Excès moyen en fonction du seuil choisi

Elles sont toutes acceptables, en considérant que le seul critère de sélection est la linéarité.

Comme évoqué dans le chapitre 2, dans le cas où la queue de distribution serait de type GPD ou Pareto, il est alors possible de s'appuyer sur l'analyse de l'estimateur $\hat{\alpha}$. Ces derniers doivent être relativement constants à partir du seuil où les sinistres ont une distribution de leur coût de type Pareto, largement utilisée pour approcher les sinistres graves. L'estimateur varie de la manière suivante :

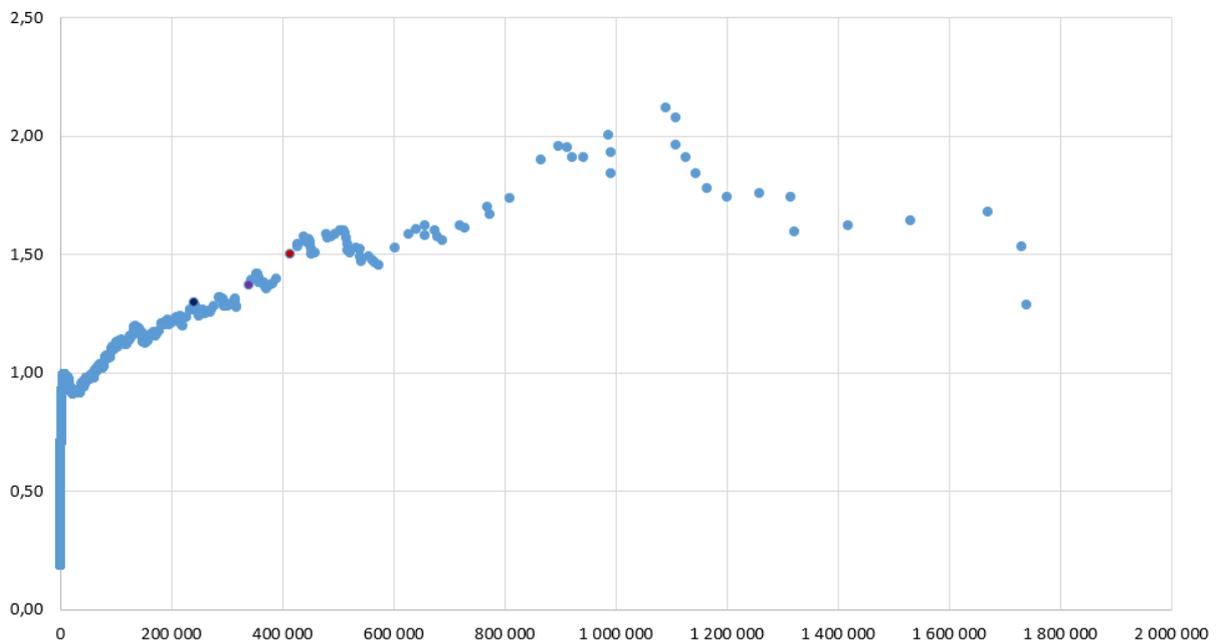


Figure 28 : Estimateurs par maximum de vraisemblance de α

Les trois points précédemment cités ont été reportés sur ce graphique. Les trois points correspondent également à des ruptures sur cet estimateur. En zoomant, il est observable que l'estimateur pourrait être viable pour les trois seuils choisis :

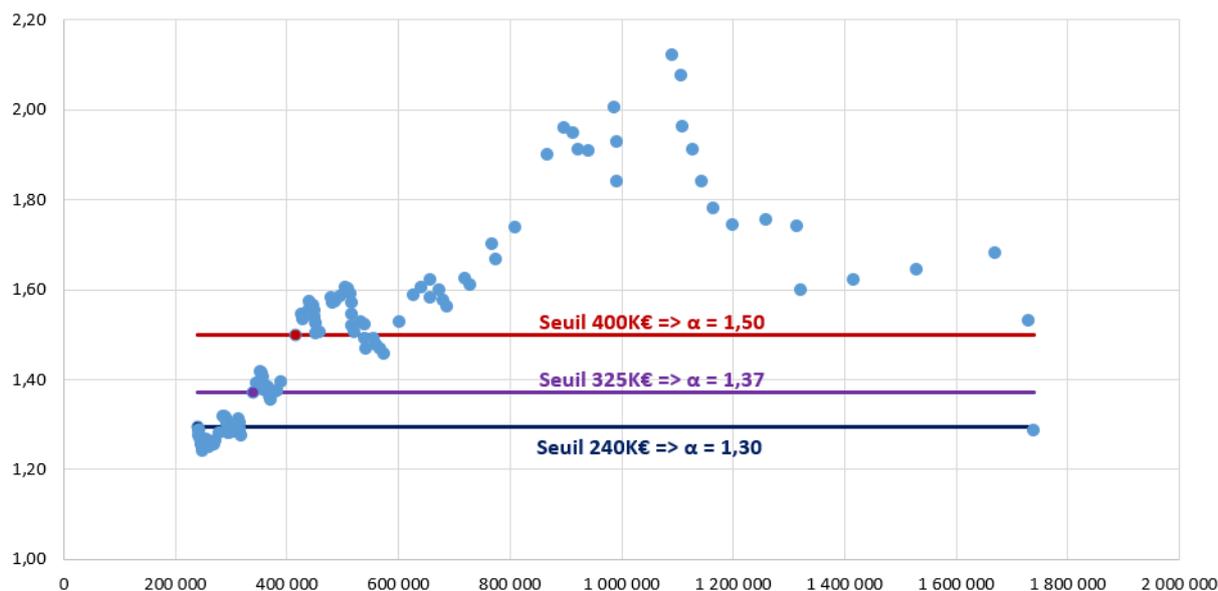


Figure 29 : Estimateur α en fonction du seuil choisi

Le seuil le plus adapté semble tout de même être le seuil à 400K€, mais l'estimateur $\hat{\alpha}$ pour la dernière valeur affichée étant de 1.30, les deux autres seuils restent possibles.

Les statistiques des sinistres dépassant les différents seuils sont les suivantes :

RC Corporel 2003-2019	Seuil à 240K€	Seuil à 325K€	Seuil à 400K€
Nombre de sinistres	126	86	71
Coût total	87 764 136	76 788 746	71 338 485
Coût moyen	696 541	892 892	1 004 767
Écart-type	803 818	909 396	965 088
Nombre moyen de sinistres graves par année de survenance	7,4	5,1	4,2
Nombre de sinistres maximum sur une année de survenance	12	8	8

Tableau 10 : Caractéristiques des sinistres au-dessus du seuil α choisi

Pour finaliser cette approche, les résultats doivent confirmer que les sinistres au-delà de ces seuils suivent bien :

- une loi de Poisson pour ce qui est de la fréquence

Cette loi utilisera le nombre de sinistre moyen par année de survenance comme paramètre λ ;

- une loi de Pareto pour la gravité des sinistres, de paramètre $\hat{\alpha}$

Pour cela, le test de Kolmogorov-Smirnov (K-S) sera utilisé, pour un niveau de confiance à 95% :

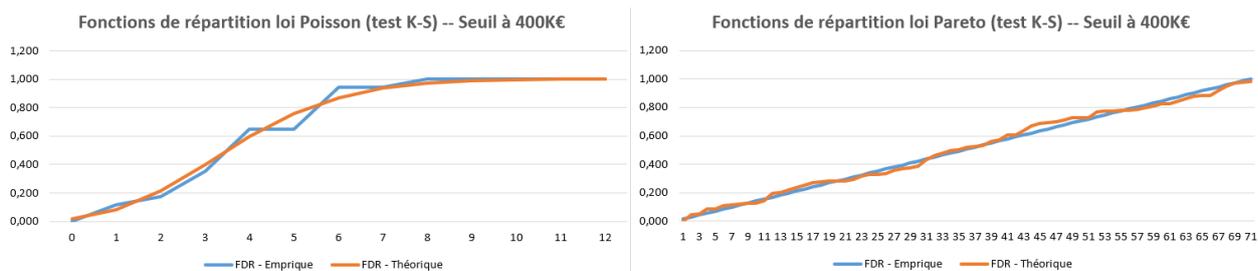


Figure 30 : Représentations graphiques du test de Kolmogorov-Smirnov

Niveau confiance : 95%		Statistique K-S	Statistique observée	Conclusion
Seuil à 240K€	Loi de Poisson	0,3614	0,1099	On accepte l'hypothèse d'égalité des lois
	Loi de Pareto	0,1586	0,0523	
Seuil à 325K€	Loi de Poisson	0,3376	0,1392	
	Loi de Pareto	0,1444	0,0794	
Seuil à 400K€	Loi de Poisson	0,3014	0,1336	
	Loi de Pareto	0,1210	0,0738	

Tableau 11 : Résultats des tests de Kolmogorov-Smirnov en fonction du seuil choisi

Les tests de Kolmogorov-Smirnov amènent unanimement à la même conclusion : les lois testées sont en adéquation avec ce qui est observé, quel que soit le seuil à partir duquel les sinistres sont considérés comme graves.

Il semble graphiquement difficile de retenir un seuil plus faible que celui des 240K€, notamment vis-à-vis de l'estimateur $\hat{\alpha}$. Les contrôles supplémentaires effectués sur les lois des sinistres graves en fonction des seuils permettent de consolider les choix pris dans ce mémoire. Étant donné que le meilleur résultat théorique serait le plus petit seuil permettant à la fois d'avoir une loi des excès convergeant vers une GPD, et assez de données pour avoir une estimation précise, peu volatile, le seuil de 240K€ semble être la meilleure solution.

Par la suite, le seuil des sinistres graves sera donc établi à 240K€, mais d'autres tests seront réalisés sur les deux autres seuils, notamment dans l'exercice d'estimation des provisions à l'ultime et mesure de l'incertitude, pour conforter ce résultat.

III. Sinistres Automobile RC « Corporel » et Provisionnement agrégé

Pour les sinistres Automobile RC « Corporel », il semble inadapté d'utiliser la méthode déterministe Chain Ladder sans prendre en compte les spécificités de ces sinistres.

Maintenant qu'un seuil a été établi entre les sinistres attritionnels et les sinistres graves, revenons sur l'estimation de la provision à l'ultime.

À noter, en termes de périmètre que :

- l'évaluation de la provision à l'ultime se fait sur les sinistres vus au 31/12/2019, même si, sans que cela ne soit analysé dans le mémoire, une comparaison sera faite avec les sinistres corrigés de l'inflation ;
- un sinistre est considéré comme « graves » lorsqu'il a dépassé le seuil fixé, quel que soit le moment au cours de sa vie. Un sinistre peut donc être grave même lorsque, à l'ultime, son coût total est nul.

A. Provision à l'ultime et incertitude

L'évaluation de la provision à l'ultime sur le segment Automobile RC « Corporel » se fera, d'une part, par l'évaluation des provisions à l'ultime des sinistres attritionnels, d'autre part, par celle des sinistres graves. Le seuil choisi pour séparer ces deux types de sinistre est 240K€.

1. Sinistres attritionnels

Les sinistres attritionnels en RC « Corporel » représentent 99% des sinistres de ce segment, pour seulement 50% du coût total dossiers/dossiers. Les provisions sur les années de survenance allant de 2003 à 2007 représentent moins de 80K€, pour 10 sinistres encore ouverts (soit 0,05% du nombre de sinistres attritionnels total).

La séparation des sinistres graves a permis d'extraire un triangle qui semble avoir la profondeur adéquate pour pouvoir utiliser le modèle déterministe classique, tout en ayant un nombre de sinistres suffisamment important pour garantir la stabilité du modèle. Les facteurs de développement représentent bien cette situation :

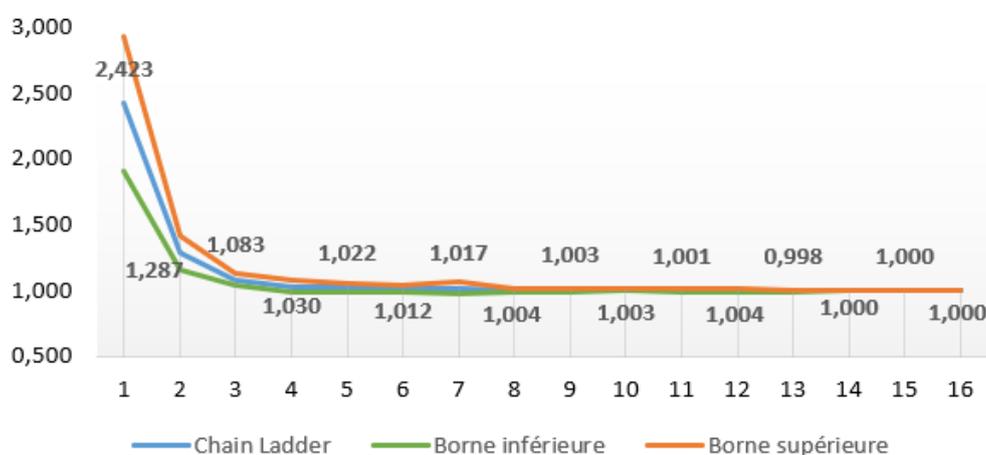


Figure 31 : Facteurs de développement RC « Corporel » attritionnel et intervalle de confiance à 95%

À partir de la 8^{ème} année de développement, les facteurs Chain Ladder sont inférieurs à 1,005, et ils sont inférieurs ou égaux à 1 à partir de la 13^{ème} année de développement. La variation autour de la moyenne est également très faible à partir de cette 8^{ème} année. Les hypothèses sous-jacentes à l'utilisation de la méthode Chain Ladder sont également vérifiées.

L'évaluation à l'ultime atteint les 4,4M€, soit 46% de la provision dossiers/dossiers :

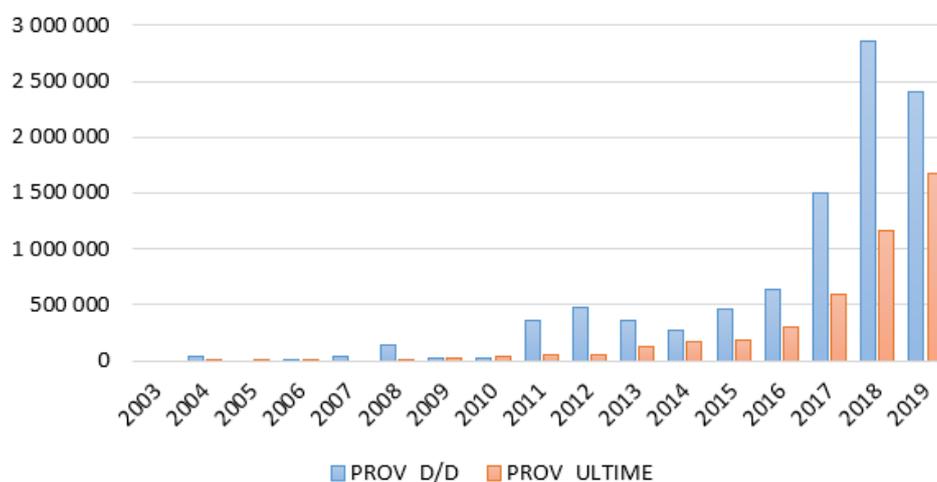


Figure 32 : Comparaison des provisions dossiers/dossiers et de l'évaluation à l'ultime par Chain Ladder

Les gestionnaires Sinistre restent extrêmement prudents lorsqu'il s'agit de dossiers avec des dommages corporels, ce qui se constate pleinement sur les sinistres attritionnels.

C'est aussi ce qui est constaté sur l'analyse des bonis/malis de liquidation. Hormis la première année de développement qui présente des malis (connaissance du dossier et sinistres tardifs), des bonis sont dégagés sur toutes les autres années de développement, quelque soit l'année de survenance.

Pour ce qui est de la mesure de l'incertitude, les éléments qui ressortent des méthodes de Mack et du Bootstrap sont similaires à ce qui est constatable sur les sinistres Automobile Matériel Autre que RC :

- Mack – Erreur relative
 - Autre que RC Matériel : 14,6%
 - RC « Corporel » attritionnel : 15,0%
- Bootstrap – Rapport écart-type sur moyenne
 - Autre que RC Matériel : 4,7%
 - RC « Corporel » attritionnel : 13,3%

Les sinistres Autre que RC Matériel sont une référence pour comparer l'incertitude de l'évaluation à l'ultime d'une population de sinistre à une autre. Il est difficile d'être plus stable que ce segment, sachant le poids du portefeuille Automobile de la MAPA.

À noter que, comme mentionné au chapitre 2, la méthode du Bootstrap ne supporte pas les valeurs négatives. Les facteurs de développement ont tous été forcés à 1 lorsqu'ils étaient inférieurs à 1.

La distribution de l'évaluation à l'ultime des sinistres RC « Corporel » attritionnels est alors la suivante :

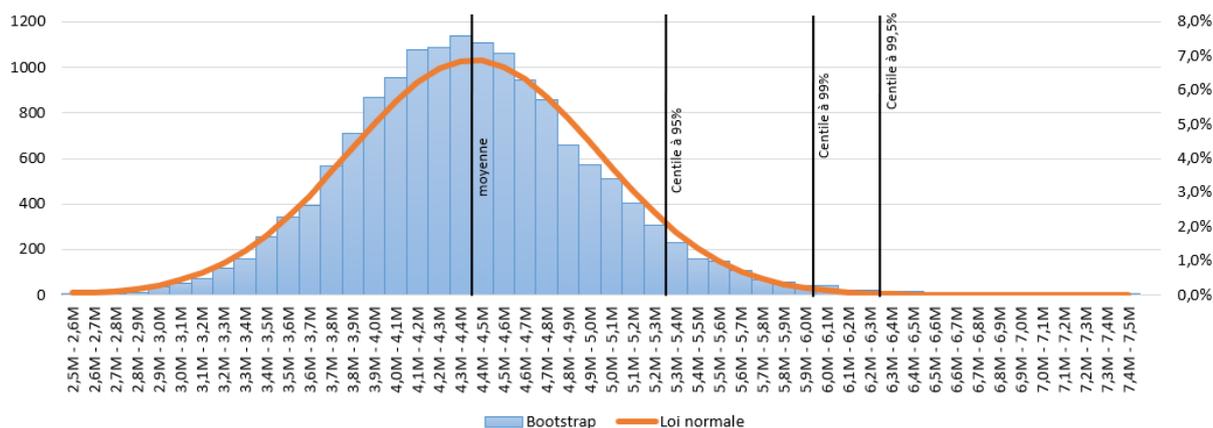


Figure 33 : Fonction de répartition des évaluations à l'ultime des sinistres RC « Corporel » attritionnels

La distribution du Bootstrap est très proche de la loi normale générée avec la moyenne et l'écart-type, elle est donc tout à fait exploitable.

La moyenne est presque identique à l'évaluation à l'ultime via Chain Ladder, et la convergence du Bootstrap se fait relativement vite (moins de 3000 itérations), ce qui montre la stabilité du triangle. Les trois limites habituellement observées sont les suivantes :

- Centile à 95% : 5,3M€ (Bootstrap ou Loi normale) ;
- Centile à 99% : 5,7M€ pour la loi normale et 6,0M€ pour le Bootstrap ;
- Centile à 99,5% : 5,9M€ pour la loi normale et 6,3M€ pour le Bootstrap.

Enfin, les mesures relatives à la méthode de Mack sont identiques à celles de la méthode Bootstrap normalisée.

L'évaluation à l'ultime de la RC « Corporel » des sinistres attritionnels par la méthode Chain Ladder est donc fiable et exploitable.

2. Sinistres graves

Les sinistres graves sont au nombre de 175 sur les 17 années étudiées. L'année avec le moins de sinistres graves est 2019, avec seulement 1 sinistre ; l'année la plus sinistrée est 2013, avec 19 sinistres graves. Les sinistres encore ouverts représentent 39% des sinistres et il n'y a pas plus d'un sinistre ouvert entre 2003 et 2008. Les provisions dossiers/dossiers représentent 33,4M€, pour 45,5M€ de prestations payées. Les sinistres RC « Corporel » graves ont un coût total moyen de 450K€.

Contrairement aux sinistres attritionnels, le volume de données est très limité, et les sinistres les plus anciens et les plus graves ne sont pas encore clos. Cela sera démontré, mais ces éléments mettent d'ores et déjà en avant les limites d'une application style Chain Ladder sur un tel périmètre.

Pour compenser le manque de recul sur le triangle des sinistres graves, la méthode des « tails factors » sera appliquée. La régression utilisée sera de la forme puissance :

$$y = 1 + b \times (c + x)^a$$

Le process appliqué est le suivant :

- ✓ Détermination des paramètres par minimisation des résidus (Solveur Excel) de l'ensemble des facteurs de développement ;

- ✓ Arrêt des facteurs de queue à la 20^{ème} année de développement ;
- ✓ Par prudence, remplacement des facteurs de développement Chain Ladder par les facteurs de queue à partir du moment où tous les facteurs de queue sont supérieurs aux facteurs Chain Ladder.

Les facteurs de développement de la méthode Chain Ladder sont les suivants, ainsi que la volatilité autour de ces facteurs :

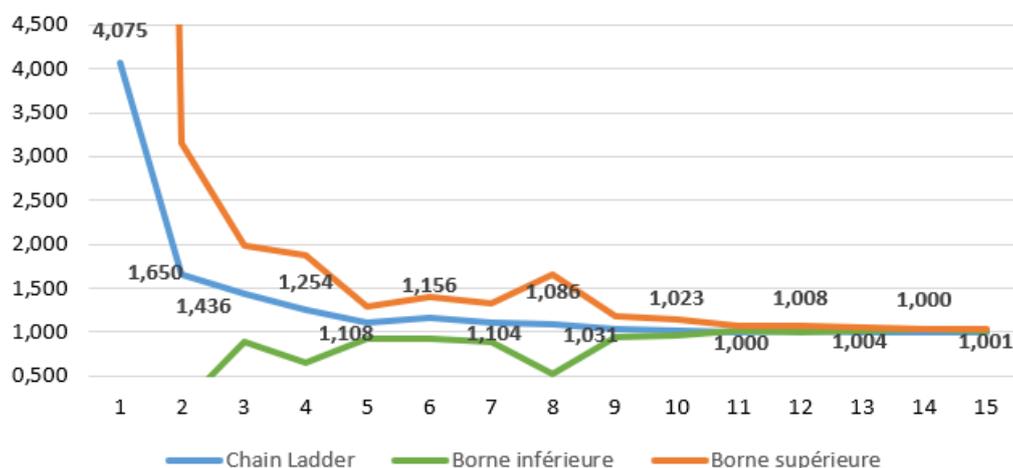


Figure 34 : Facteurs de développement RC « Corporel » grave et intervalle de confiance à 95%

Les facteurs de développement restent significatifs jusqu'à la 10^{ème} année. Les premiers facteurs de développement sont très volatiles d'une année de survenance à l'autre, l'incertitude est a priori très forte. De plus, l'hypothèse selon laquelle l'évolution des paiements suivrait un processus Markovien n'est pas respectée pour le couple (Ci1, Ci2).

Finalement, les facteurs retenus sont les suivants (hors 1^{er} facteur, pour une question de lisibilité) :

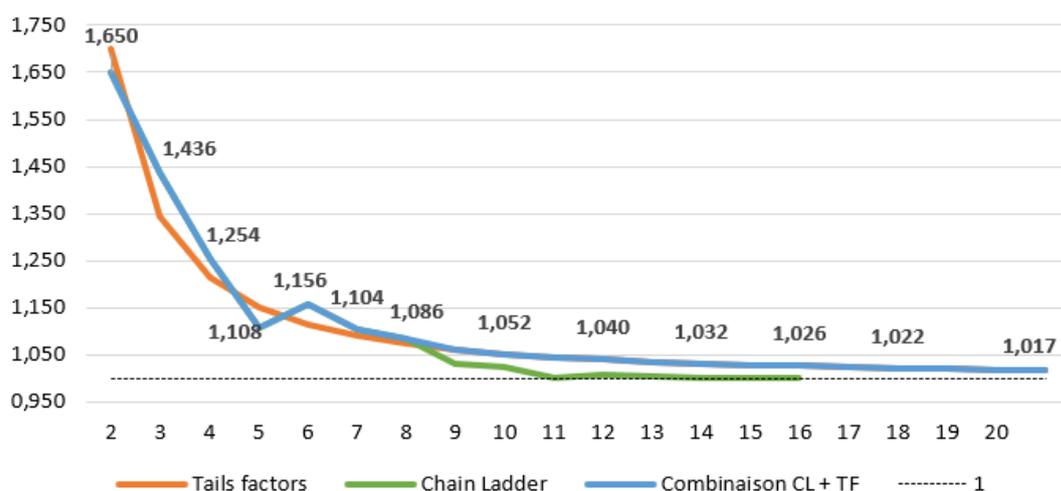


Figure 35 : Facteurs de développement RC « Corporel » grave et « Tails factors »

La différence entre la méthode Chain Ladder classique et celle des « Tails factors » est très importante sur l'estimation des provisions à l'ultime : elles s'élèvent respectivement à 8,3M€ et 26,2M€, soit presque 18M€ d'écart. Les provisions dossiers/dossiers sont, elles, dressées à 33,6M€. La seconde

méthode semble donc bien plus adaptée. Cependant, l'estimation par année de survenance est plus contrastée :

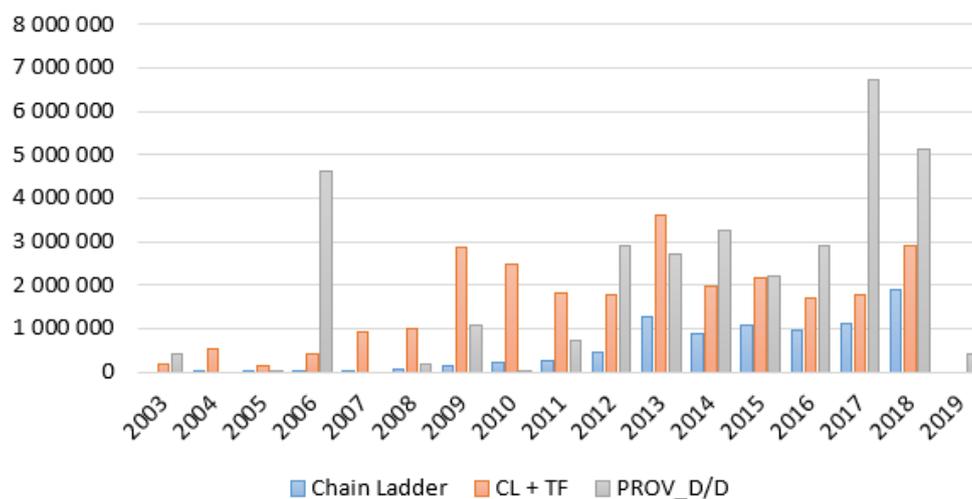


Figure 36 : Figure 31 : Comparaison des provisions dossiers/dossiers et de l'évaluation à l'ultime par Chain Ladder + « Tails factors »

Contrairement aux sinistres attritionnels, où l'écart entre provisions dossiers/dossiers et provisions à l'ultime était presque toujours du même signe, et dans des proportions ressemblantes ; ici les provisions à l'ultime avec les « Tails factors » sont :

- ✗ Sur-évalués entre 2003 et 2011 (sauf 2006) ;
- ✗ Sous-évalués en 2017 et 2018 ;
- ✗ Relativement proche entre 2012 et 2016.

L'évolution de la sinistralité des sinistres corporels graves en RC est difficilement généralisable à l'ensemble des sinistres. C'est aussi visible sur l'étude des bonis/malis où les résultats sont très différents entre les années de survenance, contrairement à l'étude sur les sinistres attritionnels.

Les résultats sur les méthodes de Mack et Bootstrap démontrent également une incertitude beaucoup plus importante :

- Mack – Erreur relative
 - Autre que RC Matériel : 14,6%
 - RC « Corporel » grave : 34,0%
- Bootstrap – Rapport écart-type sur moyenne
 - Autre que RC Matériel : 4,7%
 - RC « Corporel » grave : 43,3%

Il est important de noter que l'année 2019 ne comporte qu'un seul sinistre grave, sans règlement en 2019, ce qui implique une provision à l'ultime nulle. Or, c'est sur le 1^{er} facteur de développement que l'incertitude est la plus forte. Ces résultats sont donc certainement sous-évalués.

Les méthodes classiques s'adaptent difficilement à ces situations où le volume de données est faible, avec des sinistres très spécifiques.

Le provisionnement ligne à ligne semble, a priori, bien plus adapté, car il permettrait d'identifier les sinistres qui se ressemblent, et d'adapter au mieux la provision associée.

B. Choix du seuil a posteriori

Afin de conforter le choix du seuil à 240K€, une évaluation de l'incertitude sera présentée, liée à chacun des seuils définis dans la partie précédente, soit : 240K€, 325K€ et 400K€.

L'un des objectifs de segmentation optimale des sinistres en provisionnement est de diminuer au maximum l'incertitude autour de l'estimation de la provision à l'ultime. L'incertitude peut alors être comparée d'un seuil à un autre permettant l'aide à la décision.

L'incertitude sera ici calculée par les méthodes de Mack et du Bootstrap, et ce, à la fois sur les sinistres attritionnels et sur les sinistres graves, à chacun des seuils :

Loi normale 95%	Méthode Mack		Bootstrap	
CORPOREL RC	Erreur Relative	Borne Supérieure	Erreur Relative	Borne Supérieure
Sans Seuil	20,5%	18 367 268	18,0%	17 799 608
Attritionnels <240K€	15,0%	5 300 819	13,3%	5 313 274
Attritionnels <325K€	14,7%	5 878 605	13,1%	6 136 019
Attritionnels <400K€	16,1%	6 979 074	14,2%	6 945 030
Graves >240K€	34,0%	12 951 696	43,3%	14 525 400
Graves >325K€	35,1%	12 602 671	32,0%	12 604 394
Graves >400K€	34,4%	13 068 381	45,4%	15 195 035

Tableau 12 : Comparaison de l'incertitude en fonction du seuil choisi

Il est tout de suite possible de voir l'intérêt d'avoir segmenté entre sinistres graves et attritionnels : l'erreur relative est bien meilleure sur les sinistres attritionnels, par rapport à une évaluation de l'incertitude sur l'ensemble des sinistres. Les résultats sont d'ailleurs équivalents entre la méthode de Mack et le Bootstrap.

En revanche, et en toute logique, l'incertitude s'est décalée sur les sinistres graves, où l'erreur relative est maintenant beaucoup plus élevée. Ce résultat confirme les conclusions de la partie précédente : il est très compliqué de pouvoir faire une estimation fiable des provisions à l'ultime sur les sinistres graves en RC « Corporel » Automobile, via une méthode déterministe classique aussi élaborée soit elle.

Pour ce qui est de la comparaison entre les différents seuils, la minimisation de l'erreur relative est observée sur le seuil à 325K€ pour les sinistres attritionnels, quel que soit la méthode ; et sur les seuils à 240K€ et 325K€ pour les sinistres graves, en fonction de la méthode.

À choisir un seuil sur le seul critère de la minimisation de l'erreur relative amène à choisir le seuil de 325K€, d'autant plus que l'axe prédominant est celui des sinistres attritionnels qui seront réellement traités par les méthodes de provisionnement agrégés, contrairement aux sinistres graves.

Cependant, l'objectif est également de dégager un maximum de sinistres graves pour améliorer par la suite les estimations des provisions à l'ultime sur les sinistres graves. Le choix se porte toujours sur les seuils à 240K€ puisque, même si les meilleurs résultats sont au niveau du seuil à 325K€, les écarts entre les erreurs relatives de ces deux seuils sont similaires (0,3 point d'écart avec la méthode Mack, 0,2 point d'écart avec le Bootstrap). L'incertitude n'est pas significativement différente, et la préférence se dirige donc sur le volume des sinistres graves.

Cela montre également que le seuil de 240K€ est a priori bien une limite, comme le montrait les résultats de la fonction moyenne des excès, et qu'il n'aurait pas été possible de descendre au-delà.

IV. Sinistres Automobile RC « Corporel » et Provisionnement ligne à ligne

Le provisionnement agrégé est problématique, notamment pour les sinistres Corporels graves en RC. Les données sont trop peu nombreuses, et les comportements peuvent largement différés d'un sinistre à l'autre.

L'objectif de cette partie est donc d'analyser le sinistre ligne à ligne, avec l'ensemble des caractéristiques qui le composent afin de définir le plus finement possible la provision à l'ultime.

A. Variables explicatives et périmètre

Le périmètre reste celui des sinistres Automobile RC avec des dommages corporels, entre 2003 et 2019. Cette fois-ci, le lien est évidemment fait entre la partie sinistre et la partie liée au contrat souscrit au moment du sinistre. Les informations concernant le sinistre sont alors les suivantes :

- Caractéristiques du véhicule au moment du sinistre
 - Groupe SRA, représentant la dangerosité du véhicule ;
 - Classe prix SRA, représentant le prix du véhicule neuf à l'achat ;
 - Age du véhicule ;
 - Genre du véhicule (Voiture / Camionnette / Moto / Autres) ;
 - Utilisation (Privée ou professionnelle) ;
 - Puissance en chevaux fiscaux.
- Caractéristiques du conducteur au moment du sinistre
 - Bonus / Malus du conducteur ;
 - Âge du conducteur ;
 - Ancienneté du permis de conduire.
- Autres caractéristiques du contrat
 - Flotte automobile (OUI / NON) ;
 - Utilisation (Privée / Professionnelle) ;
 - Code postal du risque.
- Caractéristiques du sinistre
 - Date de survenance, dont le mois de survenance ;
 - Date d'ouverture ;
 - Date de clôture ;
 - Différence entre :
 - La date de survenance et la date d'ouverture (temps de déclaration) ;
 - La date de clôture et la date de survenance (temps de traitement) ;
 - Le 31/12/2019 et la date de survenance (temps de vie des sinistres encore ouverts) ;
 - Responsabilité au sens Badinter (RESPONSABLE ou NON RESPONSABLE) ;
 - État du sinistre (CLOS ou OUVERT) ;
 - Montant total des paiements, provisions ou Coût totaux vu à n'importe qu'elle date de développement, lorsque cela est possible.

D'autres variables auraient pu être très intéressantes à traiter, comme le taux d'AIPP (Atteinte à l'Intégrité Physique ou Psychique), mais cette information n'est disponible que dans les dossiers papier de chaque sinistre. Des évolutions devraient être demandées en ce sens pour affiner les exercices actuariels dans les années à venir.

La question qui se pose maintenant est la manière dont la modélisation doit être faite pour estimer le coût total des sinistres à partir des caractéristiques de celui-ci. Plusieurs axes sont à traiter :

➤ **Correction de l'inflation ?**

La correction de l'inflation, contrairement aux méthodes agrégées, semble être incontournable, puisque chaque caractéristique sera étudiée, et que sans correction de l'inflation, des segments peuvent alors être sous ou sur-estimés en termes de coût d'un dossier sinistre.

La correction de l'inflation se fera de la même manière que dans la partie traitant du seuil entre sinistres graves et attritionnels.

➤ **Sinistres clos et/ou ouverts ?**

Deux choix s'offrent à cette étude : ne traiter que les sinistres clos pour l'estimation des paramètres de la modélisation, puisque le coût total et définitif des dossiers est connu, et que l'intégration des sinistres ouverts apporterait un biais important.

Cependant, les sinistres encore ouverts et anciens (plus de 10 ans), sont généralement des sinistres d'autant plus graves et coûteux. Ne pas les intégrer pourrait sous-estimer l'estimation des coûts totaux des dossiers. Cela implique évidemment une censure à droite qu'il faut alors gérer dans les modèles. Des solutions existent, notamment en surpondérant les sinistres clos (estimateur de Kaplan-Meier).

Le choix a été fait de ne prendre que les sinistres clos. En effet, plusieurs raisons motivent cette décision :

- La volumétrie des données de la MAPA ne permet pas de complexifier les modèles sous-jacents ;
- Le nombre de sinistres encore ouverts sur les années de survenance plus ancienne n'est pas très important (seulement 5 sinistres ouverts de plus de 240K€ sur les années antérieures à 2009) ;

➤ **Sur l'ensemble des sinistres ?**

La problématique de l'évaluation à l'ultime des sinistres Automobile concernant essentiellement les sinistres RC « Corporel » graves (>240K€). Mais, il est évidemment impossible de faire du provisionnement ligne à ligne avec seulement 107 sinistres clos.

Il faut donc élargir ce périmètre. Voici la répartition du nombre de sinistres Corporels RC par seuil :

RC Corporelle inflation corrigée

Seuil	Nombre de sinistres	Part de l'ensemble	Nombre de sinistres clos	Nombre de sinistres ouverts
ENSEMBLE	19 381	100,0%	18 080	1 301
240	16 940	87,4%	15 745	1 195
2 400	10 399	53,7%	9 538	861
5 000	6 419	33,1%	5 797	622
10 000	3 242	16,7%	2 849	393
24 000	1 363	7,0%	1 129	234
50 000	727	3,8%	570	157
100 000	417	2,2%	305	112
240 000	193	1,0%	124	69

Tableau 13 : Répartition des sinistres RC « Corporel » Automobile, corrigés de l'inflation

À noter que, tout comme pour les triangles agrégés, les sinistres sont considérés comme dépassant un seuil dès qu'ils ont dépassé ce seuil au moins une fois au cours de leur vie, et non en fonction de ce qu'ils coûtent à la clôture ou à la dernière vision connue. Cela permet de prendre en compte, et notamment pour le provisionnement ligne à ligne, l'évolution des coûts totaux d'un sinistre, aussi bien sur les bonis que sur les malis.

Faut-il élargir à l'ensemble des sinistres ? Il est possible que non. L'objectif étant principalement orienté vers l'estimation du coût des sinistres graves, intégrer plus de 20% de sinistre dont le coût est nul, ou intégrer tous les sinistres clôturés sur une période très courte, apportera un biais et une sous-estimation des estimations.

La contrainte est double :

- Maximiser le nombre de sinistres pour tendre vers une modélisation cohérente et fiable ;
- Minimiser l'influence des sinistres « communs », de court terme.

Le choix s'est donc porté sur les sinistres supérieurs à 2400€, représentant donc 1% du seuil des sinistres graves (240K€). Cela permet de capter plus de 50% des sinistres corporels RC, d'avoir moins de 10% de sinistres encore ouverts, et d'avoir un seuil assez important pour ne pas trop sous-estimer les sinistres graves.

La modélisation se fera donc sur 9 538 sinistres clos, en Automobile RC « Corporel », dont les montants des règlements ont été corrigés de l'inflation.

B. Modélisations CART

Le périmètre étant défini, et les bases préparées (travail sur la Qualité des données), la modélisation des coûts totaux de ces sinistres peut alors commencer. Trois approches de cette modélisation peuvent être discernées :

- ✓ Les variables explicatives se limitent aux caractéristiques du contrat au moment du sinistre, et à la notion de responsabilité ;
- ✓ Les variables explicatives mentionnées précédemment, mais également l'évolution au cours des premières années de développement des règlements des sinistres.

- ✓ Les variables explicatives mentionnées précédemment, mais également l'évolution au cours des premières années de développement des règlements, et des provisions des sinistres.

Chacune de ces approches ayant leurs propres avantages et inconvénients.

1. Méthodologie

Avant d'analyser les résultats de ces deux approches, une définition de la méthodologie à appliquer a été adoptée lors de cette étude.

➤ Segmentation des variables explicatives

L'un des grands avantages de la méthode CART est qu'elle permet d'optimiser les segmentations, et de regrouper les modalités avec une certaine souplesse. Le revers étant que la succession importante de nœuds rend la communication moins lisible, moins imagé.

Le choix ici fait est d'accepter cette difficulté d'interprétation, et de laisser les modalités des variables explicatives à leur niveau le plus fin (que les variables soient qualitatives ou quantitatives). Des exceptions seront faites pour les modalités regroupant moins de 5 occurrences. Elles seront alors regroupées avec les modalités qui se rapprochent le plus en termes de cohérence (les véhicules âgés de plus de 40 ans sont rassemblés en une seule modalité « 40 ans et plus »). Cette exception permet d'éviter d'avoir des modalités absentes dans la base d'apprentissage, alors qu'elles existent dans la base de validation (ou même dans la base des sinistres ouverts), empêchant ainsi des erreurs applicatives.

➤ Mode opératoire

Les modélisations sont faites sous R, avec la librairie « rpart ».

Les mêmes étapes seront réalisées pour chaque approche :

À partir de la base d'apprentissage :

- 1) Construction de l'arbre maximal, sans contrainte ;
- 2) Étude de la table des cp, et notamment la localisation du cp qui minimisera l'erreur de prédiction ;
- 3) Répétition avec contraintes, et notamment en contraignant le nombre de sinistres minimum par feuille (dans un premier temps, **1% de la volumétrie de la base d'apprentissage**) ;
- 4) Choix de l'arbre ;
- 5) Application à la base de validation ;
- 6) Analyse des erreurs sur les deux bases.

Cette limite de 1% est un moyen de se protéger du sur-apprentissage, mais il est plutôt contraignant vis-à-vis des sinistres graves, qui ont, justement, des coûts « non courants ». L'analyse des trois approches sera faite avec la même contrainte, pour pouvoir les comparer, mais cette contrainte sera ensuite adaptée.

L'application aux sinistres ouverts se fera dans un second temps, lorsque l'approche la plus convaincante sera identifiée.

La base d'apprentissage est tirée aléatoirement et représente 7 580 sinistres, soit approximativement 80% des sinistres clos. Les autres sinistres sont attribués à la base de validation.

2. 1^{ère} approche : sans intégration des éléments des premières années de développement

Le modèle CART est donc composé des éléments suivants :

Variables Qualitatives	Variables Quantitatives
Responsabilité	Age du véhicule
Groupe SRA	Puissance
Classe SRA	Bonus/Malus
Genre du véhicule	Age du conducteur
Utilisation	Ancienneté de permis
Flotte	Nombre de jours avant l'ouverture du sinistre
Mois de survenance	
Département	
Variable à expliquer :	Coût total à l'ultime inflation corrigée

Tableau 14 : Variables explicatives potentielles de l'estimation du coût total des sinistres RC « Corporel »

L'arbre maximale est évidemment illisible, et ne peut pas être exploité, en raison du sur-apprentissage :



Figure 37 : Arbre maximal du 1er modèle CART

L'objectif est donc d'identifier le cp (Complexity Parameter) qui minimise l'erreur de modèle. Or, celui-ci n'est autre que le nœud racine (la réponse au modèle est alors la moyenne de l'ensemble des sinistres, quel que soit le sinistre). Le modèle n'est donc pas exploitable.

Dans le but d'expliquer ce phénomène, et d'en apprendre davantage sur les données, la démarche d'estimation est néanmoins prolongée. L'arbre est alors élagué de telle sorte qu'il ait au moins 75

sinistres par feuille (paramètre par défaut représentant 1% de la volumétrie). L'arbre ainsi obtenu est alors plus lisible, mais encore complexe :

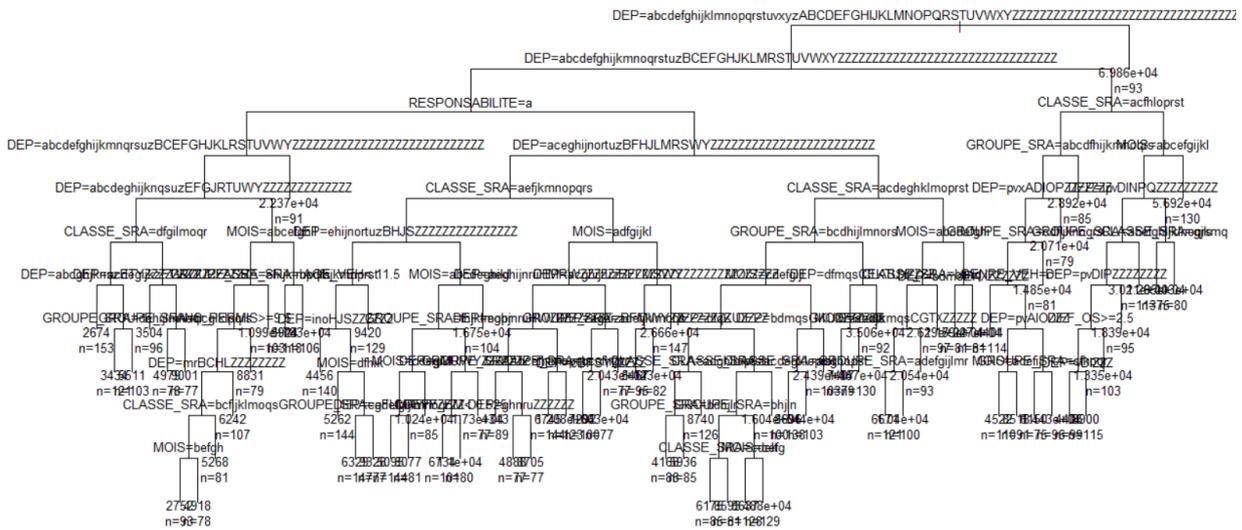


Figure 38 : Arbre élagué en fonction du nombre de sinistres minimum par feuille du 1er modèle CART

Cette complexité n'est pas réellement un souci puisque cette simulation se limite à des fins techniques, et non, commerciales. Il ressort de ce modèle que les principales variables explicatives sont : la localisation du risque, les critères de coût et de dangerosité du véhicule, le mois de survenance et la responsabilité du conducteur.

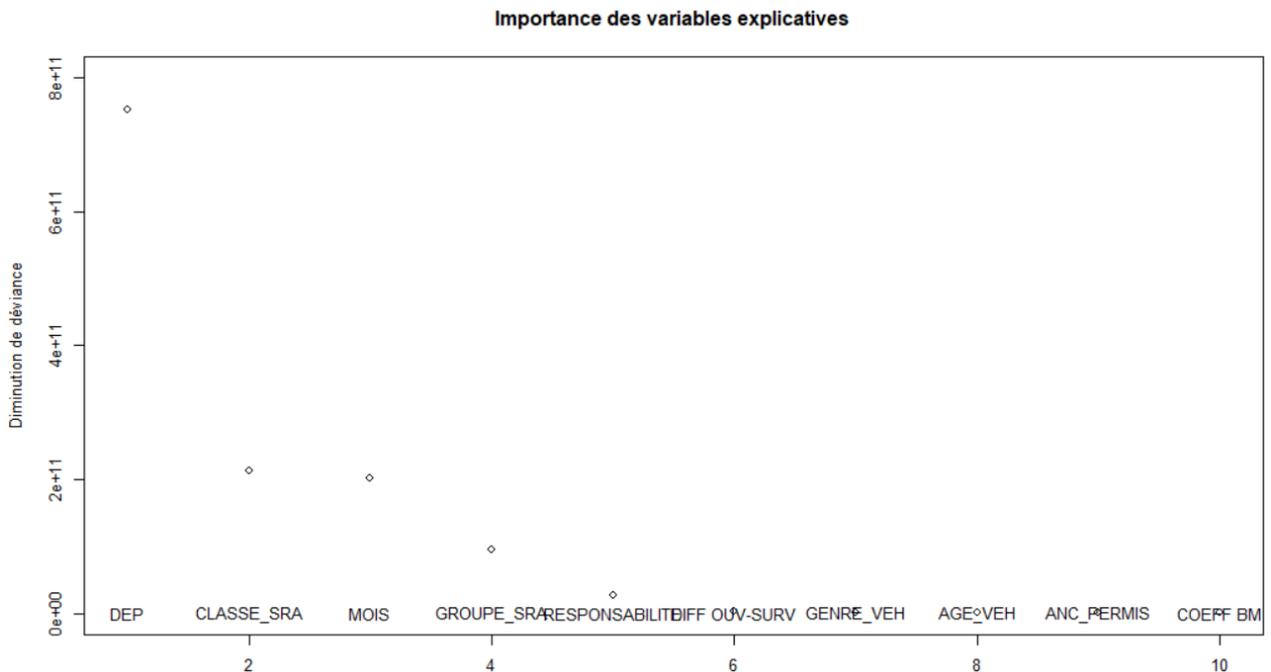


Figure 39 : Variables explicatives significatives et ordre d'importance (1er modèle CART)

Il est difficile de comprendre que la variable « Département » puisse avoir l'influence la plus significative sur le coût total à l'ultime. Après avoir observé les résultats de prédiction de ce modèle par département, on s'aperçoit que les départements avec les coûts moyens estimés les plus élevés sont tous situés en zone rurale ou montagneuse (dans l'ordre décroissant : Haute-Vienne ; Creuse ;

Pyrénées-Orientales ; Dordogne ; Doubs ; Puy-de-Dôme). Mais cela n'est pas vrai pour les autres départements du même type. Le portefeuille n'est pas assez étoffé et cette variable provoquera certainement un biais dans nos estimations, elle sera donc retirée des simulations finales. À noter que le retrait de cette variable ne change que peu le classement ci-dessus.

L'analyse des comparaisons entre l'estimation et l'observé confirme les craintes sur la pertinence du modèle : le montant maximum prédit est de 70K€, alors que les sinistres clos les plus graves dépassent aisément le million d'euros. À noter qu'il y a 76 modalités de prédiction distinctes, soit la limite fixée sur le minimum de sinistres par feuille.

Pourquoi les prédictions sont si mauvaises, et notamment sur les sinistres graves ? Quatre raisons peuvent l'expliquer :

- ✗ La volumétrie de la base de données ;
- ✗ La représentation des graves parmi les attritionnels ;
- ✗ La contrainte du nombre de sinistres par feuille : le sujet sera de nouveau abordé une fois la meilleure approche déterminée ;
- ✗ Le manque de variables explicatives pouvant faire la distinction entre les graves et les attritionnels.

L'étude a été réalisée en connaissance de cause, sur les deux premiers points. En revanche, il aurait pu être possible de pouvoir faire la distinction entre graves et attritionnels avec des variables telles que la Responsabilité, la dangerosité du véhicule ou la puissance, l'âge du conducteur. Il s'avère que ces dernières ne sont finalement pas suffisantes pour avoir une estimation correcte du coût des sinistres.

Les deux premiers points sont difficilement ajustables : si la volumétrie est augmentée, le déséquilibre se fera plus grand sur le rapport graves/attritionnels, et l'ensemble des sinistres disponibles a été sélectionné (avec un historique de plus de 15 ans).

En revanche, il est encore possible de trouver des variables explicatives, en contrepartie d'un morceau de la volumétrie des données.

3. 2^{ème} modèle : intégration des éléments des premières années de développement

Comme évoqué dans le chapitre 1, les sinistres graves sont de moins en moins nombreux dans le portefeuille de la MAPA, et ce, pour plusieurs raisons :

- Une sélection des risques beaucoup plus rigoureuse depuis 2014 ;
- En lien direct avec le point précédent, une baisse du nombre de véhicules assurés, stabilisée depuis 2018 ;
- Des moyens de prévention mis en place, notamment pour les jeunes conducteurs.

De ce fait, et alors même que les sinistres survenus récemment sont ceux qui ont la probabilité la plus élevée d'être encore ouverts, le nombre de sinistres graves encore ouverts des trois dernières années de survenance est de 16 (moins d'un quart des sinistres encore ouverts parmi les graves en corporel RC).

De plus, le temps de règlement des sinistres, sur les sinistres clos utilisés pour la modélisation CART, sont observables :

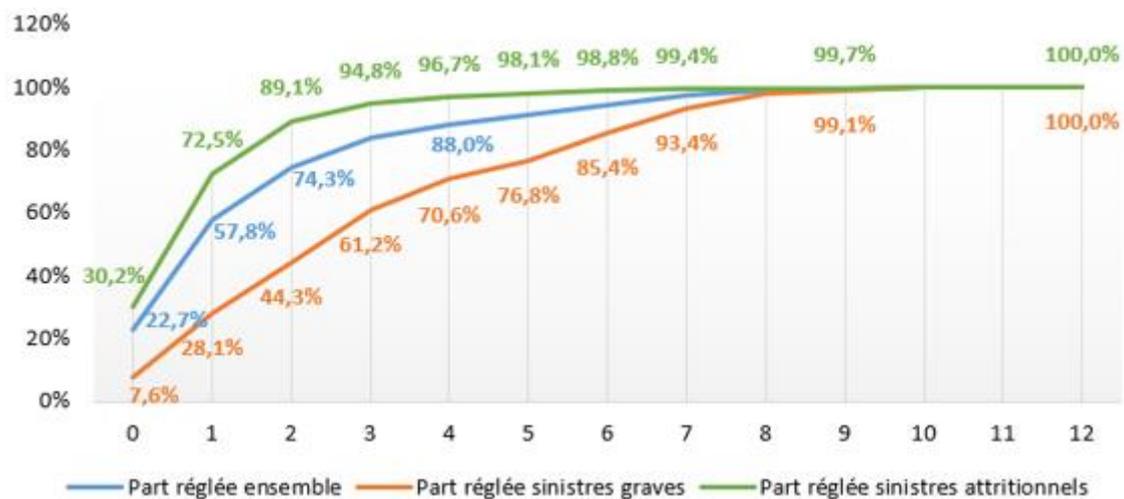


Figure 40 : Sinistres Corporels RC clos > 2400€ et Répartition des règlements dans le temps

Au bout des trois premières années de développement, près de 75% des règlements ont été effectués en moyenne. Ce chiffre monte à presque 90% pour les sinistres attritionnels, et n'atteint en revanche pas les 50% pour les sinistres graves. Ce qui montre encore une fois la large différence qu'il peut y avoir entre sinistres graves et attritionnels, et l'importance de les séparer dans le cadre de l'estimation des provisions à l'ultime.

Malgré les 44% pour les sinistres graves, ces trois premières années de développement sont donc une source importante d'information, qui pourrait permettre d'estimer le coût total à l'ultime de ces sinistres.

L'idée est alors d'observer les premiers règlements de la vie du sinistre, en cumulés et en décumulés, afin d'en déduire, en plus des autres variables explicatives, le coût total du dossier.

L'inconvénient de ces nouvelles variables explicatives c'est qu'elles ne sont applicables qu'aux sinistres qui ont au moins vécu ces trois années de développement. Les sinistres des survenances 2017, 2018 et 2019 ne pourront donc pas être estimés via cette nouvelle approche.

La nouvelle base de données pour la modélisation CART est donc de 8 926 sinistres, 7 136 seront attribués à la base d'apprentissage et 1 790 pour la base de validation.

La modélisation CART ressort un niveau de cp optimal différent de la racine :

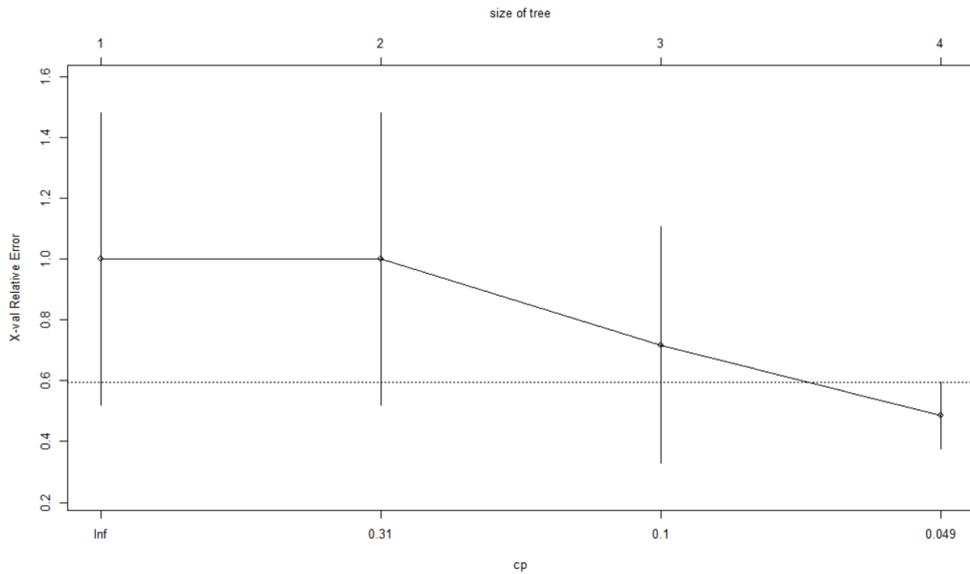


Figure 41 : 2ème modèle, paramètre de complexité et optimisation

Trois nœuds sont identifiés et permettent de minimiser l'erreur de modèle. Cependant, le modélisation CART se fait dans un premier temps sans contrainte, et certaines feuilles ne sont pas exploitables :

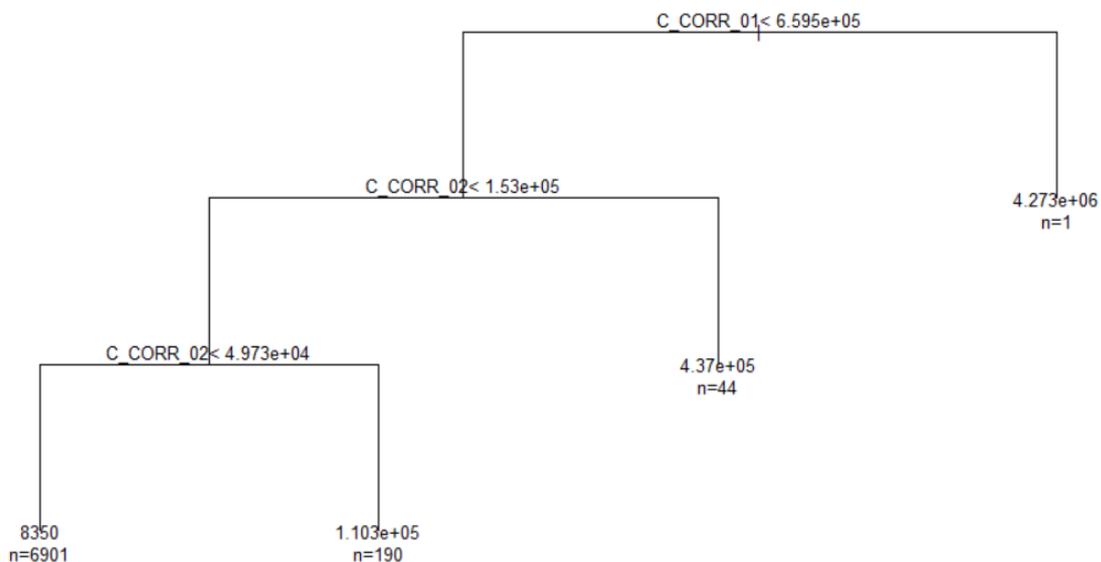


Figure 42 : 2ème modèle, arbre optimal

La feuille de droite a une volumétrie bien trop faible pour être exploitée, qui pourrait être un problème de surapprentissage.

À noter également que la principale variable discriminante de ce premier arbre semble être le coût des paiements cumulés de la troisième année de développement, ce qui n'est pas incongrue.

Le modèle, même s'il est a priori plus robuste que le précédent (tableaux 15 et 16), reste encore assez instable puisque les résultats peuvent encore varier d'un calcul à l'autre de façon importante.

Tout comme pour le modèle précédent, le nombre minimum de sinistre par feuille sera le critère qui permettra d'élaguer l'arbre.

La variable explicative principale est toujours les paiements cumulés de la troisième année de développement, puis, avec des pouvoirs explicatifs moindre, le département, les paiements cumulés de la deuxième année de développement, la classe SRA, les paiements non cumulés de la troisième année de développement et le mois de survenance.

Les prévisions sont déjà bien plus pertinentes : la contrainte sur le nombre de sinistres par feuille limite forcément la valeur prédite la plus importante, mais elle atteint 368K€, ce qui semble bien plus adapté que les 70K€ de la première approche. En effet, les 78 plus gros sinistres de la base d'apprentissage ont un coût total moyen de 400K€. En revanche, les sinistres graves dont les paiements sont tardifs (au-delà des trois premières années de développement), ne sont alors pas captés par le modèle, et des montants compris entre 10K€ et 100K€ leur sont affectés.

La même chose est constatée sur les sinistres de la base de validation.

4. 3^{ème} modèle : intégration des éléments des premières années de développement, y compris les provisions

L'inconvénient d'ajouter les provisions dossiers/dossiers des premières années de développement du sinistre dans la modélisation repose sur le fait d'utiliser une estimation pour estimer.

Si le gestionnaire sinistre a trop provisionné, s'il ne croyait pas qu'un recours était possible, s'il n'avait pas toutes les informations ... tout cela entre donc dans le modèle. Mais, il est possible que le modèle puisse très bien, par le biais des autres informations dont il dispose, détecter les cas où le gestionnaire était trop prudent ou pas assez.

Ce modèle permet d'avoir un arbre plus complexe que le précédent, la modélisation CART ressortant un niveau de cp optimal plus faible :

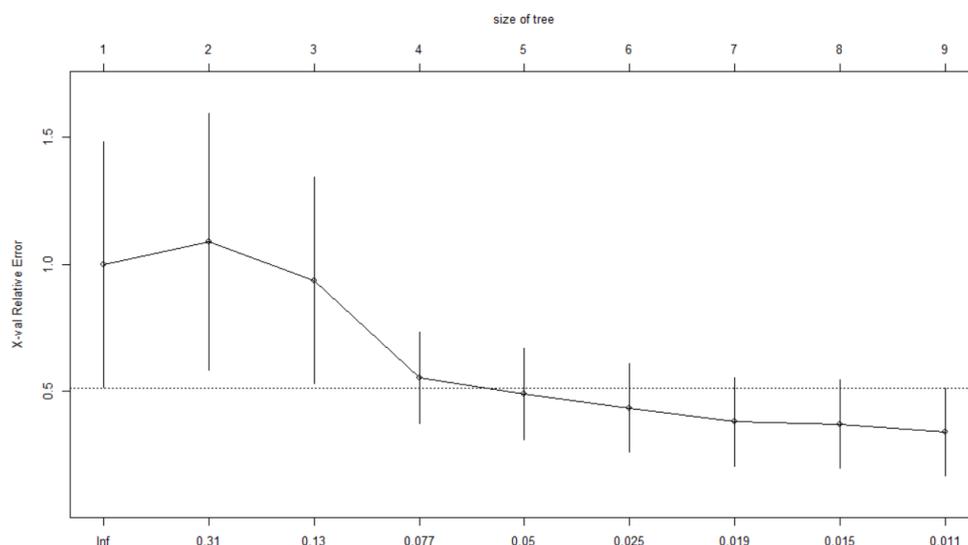


Figure 43 : 3^{ème} modèle, paramètre de complexité et optimisation

Le niveau de cp optimal pourrait être revue à la baisse puisque les cp suivants, avec des nœuds supplémentaires, sont du même ordre de grandeur :

Erreur en fonction de la complexité du modèle

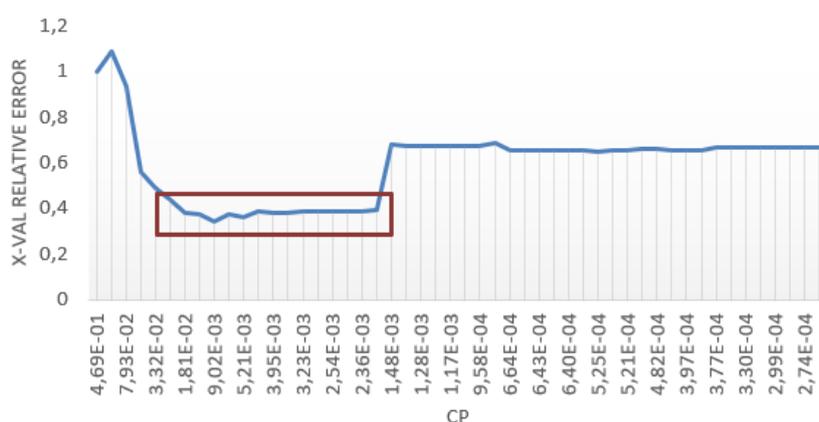


Figure 44 : Erreur en fonction de la complexité du modèle

Comme dans les autres modèles, le nombre de sinistres minimum par feuille est imposé, pour qu’une comparaison soit faite *in fine*.

Les variables explicatives qui ressortent le plus sont ici toutes liées soit aux provisions, soit aux règlements.

5. Choix du modèle et optimisation

Pour faire un choix sur le modèle à exploiter pour l’estimation des provisions à l’ultime par le modèle CART ligne à ligne, plusieurs critères seront comparés :

- La complexité du modèle relative à l’erreur minimale : l’arbre doit être assez complexe pour pouvoir modéliser les niveaux de coûts totaux assez fins ;
- Les indicateurs sur l’erreur de prédiction sur le modèle de validation ;
- La comparaison des indicateurs sur l’erreur de prédiction entre le modèle d’apprentissage et le modèle de validation, afin de déceler un éventuel surapprentissage.

Pour ce qui concerne le modèle d’apprentissage, les résultats sont les suivants :

Base d'apprentissage	Complexité optimale	RMSE	MAE
Modèle 1 : sans éléments des 1ères années de dév.	Racine	68 066	15 629
Modèle 2 : avec les paiements des 1ères années de dév.	3 nœuds	61 217	7 413
Modèle 3 : avec les paiements et les provisions des 1ères années de dév.	8 nœuds	60 187	6 204

Tableau 15 : Comparaison des modèles CART sur les bases d’apprentissage

Quel que soit le critère, le modèle qui présente les meilleurs résultats est le 3^{ème} modèle : le modèle CART optimal est le plus complexe (8 nœuds), et les erreurs prédictives sont les plus faibles.

À noter que la différence entre les modèles des erreurs prédictives est relativement plus importante entre le modèle 1 et le modèle 2, qu’entre le modèle 2 et le modèle 3. L’ajout des paiements des

premières années de développement a donc ajouté un pouvoir explicatif plus important que l'ajout des provisions des premières années de développement lorsque les paiements des premières années de développement sont déjà intégrés.

Pour ce qui concerne le modèle d'apprentissage, les résultats sont les suivants :

Base de validation	RMSE	MAE
Modèle 1 : sans éléments des 1ères années de dév.	71 225	17 929
Modèle 2 : avec les paiements des 1ères années de dév.	51 600	8 144
Modèle 3 : avec les paiements et les provisions des 1ères années de dév.	50 346	5 852

Tableau 16 : Comparaison des modèles CART sur les bases de validation

Les erreurs de prédictions de la base de validation confortent les résultats de la base d'apprentissage : le 3^{ème} modèle semble effectivement le meilleur modèle.

Les erreurs prédictives entre base d'apprentissage et base de validation sont du même ordre de grandeur, ce qui permet d'écarter le risque de surapprentissage.

Le choix a été fait de conserver ce 3^{ème} modèle, qui, pour rappel, à tout de même les contraintes suivantes :

- Ne pas pouvoir faire l'estimation des sinistres dont l'ancienneté est de moins de 4 ans ;
- Qui utilise les provisions dossiers/dossiers des premières années de développement comme variable explicative.

À cette étape, le modèle n'est pour le moment pas achevé : le nombre de sinistre à chaque feuille a été fixé à un minimum de 1% du nombre de sinistres total, pour pouvoir comparer les différents modèles, sans risquer le phénomène de sur-apprentissage.

Le modèle étant choisi, ce nombre de sinistres minimum par feuille doit être redéfini, à un niveau plus faible. L'objectif étant, pour rappel, de pouvoir faire une estimation correcte des sinistres graves, qui sont par définition rares, ce chiffre devra être bien plus faible.

Un minimum de 4 sinistres par feuille a été fixé de manière arbitraire. Ce chiffre a été obtenu en comptabilisant le nombre de sinistres supérieur à 1M€ parmi les sinistres clos, et en prenant en compte que le modèle est calibré selon une base partielle (base d'apprentissage).

Le résultat du modèle final, avec cette nouvelle contrainte est le suivant :

Modèle final	RMSE	MAE
Base d'apprentissage	35 666	2 330
Base de validation	32 189	3 638

Tableau 17 : Erreur constatée sur le modèle CART choisi et optimisé

Les résultats avec un minimum de 4 sinistres par feuille, sans autre contrainte sur la complexité, donnent des prédictions de bien meilleure qualité. De plus, la base de validation enregistre les mêmes ordres de grandeur en termes d'erreur, ce qui tend à prouver la bonne application du modèle.

Les principales variables explicatives sont finalement les suivantes :

Variables significatives dans le modèle final et ordre d'importance			
1	Provisions en N+2	11	Différence entre la date d'ouv. et la date de surv.
2	Paielements cumulés en N+2	12	Paielements non cumulés en N+1
3	Classe SRA du véhicule	13	Ancienneté de permis du conducteur
4	Paielements non cumulés en N+2	14	Flotte automobile
5	Mois de survenance du sinistre	15	Age du conducteur
6	Puissance du véhicule	16	Utilisation (Privé / Pro)
7	Age du véhicule	17	Provisions en N+1
8	Provisions en N+0	18	Responsabilité de l'assuré
9	Groupe SRA du véhicule	19	Genre du véhicule
10	Paielements en N+0	20	Paielements cumulés en N+1

Bleu : Développement du coût du sinistre ; Orange : Caractéristiques du véhicule ; Vert : Caractéristiques du conducteur

Tableau 18 : Variables explicatives significatives du modèle final et ordre d'importance

Près du tiers des variables explicatives sont les variables de développement du coût du sinistre, et notamment la provision en N+2. Ensuite, les variables les plus discriminantes sont les variables qui concernent le véhicule.

Maintenant, il est également intéressant de zoomer sur les sinistres graves. Pour rappel, un sinistre grave est considéré comme, tout sinistre qui a été, à un moment de sa vie, supérieur à 240K€. Cela permet notamment de prendre en compte les effets de bonis/malis sur les triangles de liquidation.

Les prédictions sont relativement de bonne qualité pour la base d'apprentissage, sur les sinistres inférieurs à 1M€ :

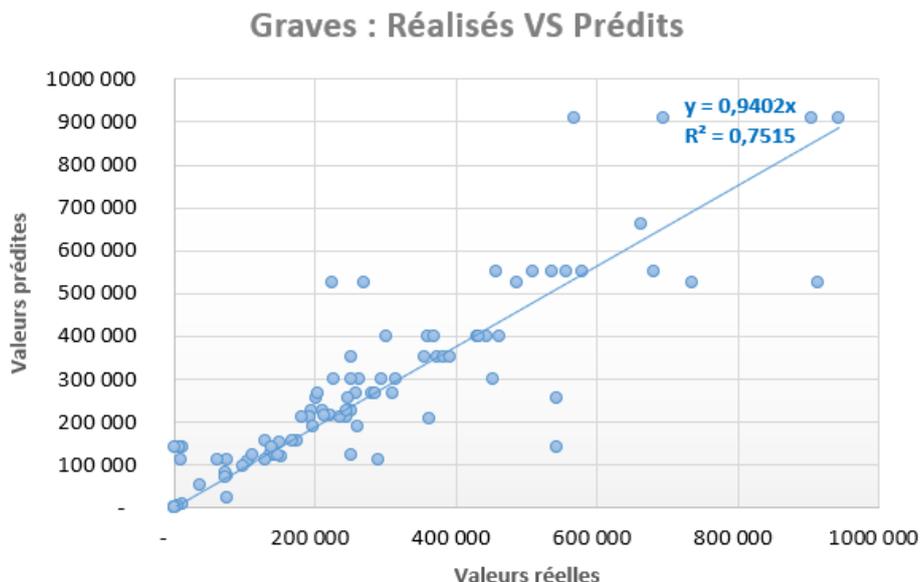


Figure 45 : Comparaison des valeurs réalisées et prédites (CART) des sinistres graves sur la base d'apprentissage

Les points sont majoritairement placés sur la bissectrice, et la régression est très proche de $y=x$. Les écarts se ressentent essentiellement sur les sinistres les plus graves.

À noter que les sinistres à forts bonis, et dont les provisions sont parfois très élevées en N+2 (le variable provision en N+2 étant celle qui a le plus d'influence), ont un coût total prédit proche du réel, c'est-à-dire beaucoup plus faible que l'évaluation dossiers/dossiers en N+2. Cela montre que le modèle a la capacité de distinguer des sinistres justement évalués, des sinistres évalués prudemment.

Ajoutons ensuite les sinistres prédits de la base de validation :

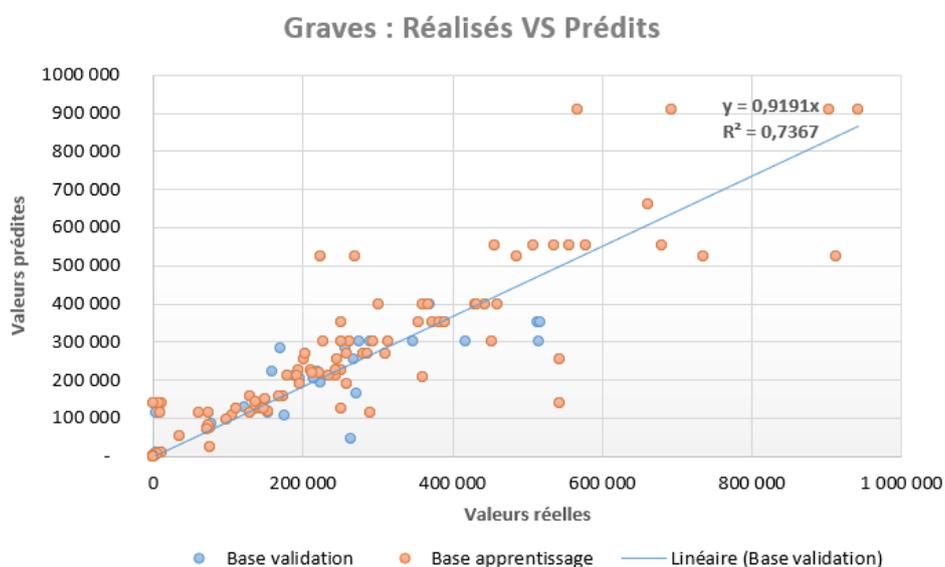


Figure 46 : Comparaison des valeurs réalisées et prédites (CART) des sinistres graves sur les bases d'apprentissage et de validation

Les prédictions de la base de validation ont également tendance à sous-estimer les coûts totaux à l'ultime, de surcroit, de manière plus marquée. La droite de régression s'est donc légèrement éloignée de la bissectrice. Les points de sous-évaluations sont généralement des cas de malis, que le modèle distingue mal.

Pour les sinistres supérieurs à 1M€, le modèle a bien réussi à les capter. En revanche, l'amplitude des sinistres supérieurs à 1M€ est importante : de 1M€ à 4,3M€. La valeur retenue par le modèle est de 1,9M€ (soit la moyenne des 4 sinistres identifiés comme étant les plus graves). L'application de cette valeur prédite est donc limitée.

La solution retenue est alors la suivante :

- Un sinistre dont la provision dossiers/dossiers est supérieur à 1M€, et dont la prédiction du modèle atteint la valeur maximale, sera alors évalué à sa provision dossiers/dossiers ;
- Un sinistre dont la provision dossiers/dossiers est supérieur à 1M€, et dont la prédiction du modèle n'atteint pas la valeur maximale, sera alors évalué à sa prédiction de modèle ;
- Un sinistre dont la provision dossiers/dossiers est inférieur à 1M€, et dont la prédiction du modèle atteint la valeur maximale, sera alors évalué à sa prédiction de modèle.

Une légère sous-estimation est visible sur l'évaluation de la provision à l'ultime :

Modèle final	Montant	Erreur relative
Provisions à estimer en N+2	24 299 502	
Provisions dossiers/dossiers en N+2	32 265 811	32,8%
Provisions estimées en N+2	23 533 151	-3,2%

Tableau 19 : Comparaison entre les coûts totaux réels, les provisions dossiers/dossiers en N+2 et les estimations CART

Alors que les provisions dossiers/dossiers des sinistres graves surestiment de 8M€ la provision à l'ultime (soit +32,8% du montant réel), la provision à l'ultime estimée par le modèle CART est sous-estimée de 0,8M€ (soit -3,2% du montant réel). Cette dernière est toute fois plus juste que la somme des provisions dossiers/dossiers, mais peu prudente.

Ces éléments ne remettent donc pas en cause la validité du modèle. Les sinistres graves ouverts depuis plus de 3 ans seront donc utilisés dans ce dernier.

6. Application aux sinistres graves ouverts depuis plus de 3 ans

À présent, il est possible de faire une estimation à l'ultime des sinistres graves ouverts, sauf pour les années de survenance 2017, 2018 et 2019. Pour ces derniers, par prudence, la provision dossiers/dossiers sera reprise.

Les sinistres graves ouverts de 2003 à 2016 représentent 21,1M€ (contre 33,4M€ sur le périmètre total 2003-2019), et 52 sinistres (contre 68 sinistres sur le périmètre total).

L'application du modèle sur ces sinistres donne les résultats suivants :

- Les provisions à l'ultime sont estimés à 15,3M€ soit 5,8M€ de moins que les provisions dossiers/dossiers (-27,7%) ;
- Si une couche de prudence de 10% est ajoutée à cette estimation pour les sinistres estimés avec la méthode CART (hors sinistres où la provision dossiers/dossiers est directement reprise), afin de prendre en compte la légère sous-estimation constatée sur les provisions des sinistres clos, le résultat obtenu est de 16,0M€ de provisions à l'ultime (-24,4% par rapport aux provisions dossiers/dossiers).

La répartition des provisions à l'ultime (avec la couche de prudence) est la suivante :

Comparaison des provisions dossiers/dossiers et des estimations par la méthode CART

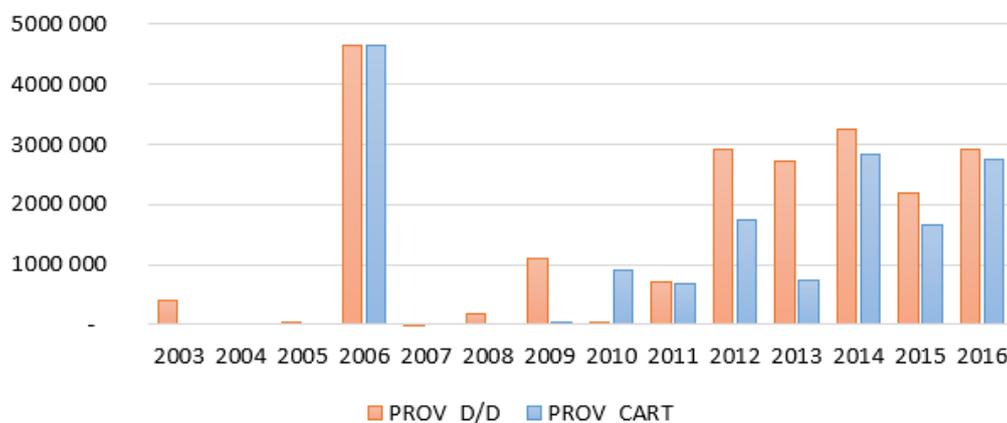


Figure 47 : Comparaison des sinistres ouverts en RC « Corporel » Automobile à fin décembre 2019 et évaluation des provisions à l'ultime par la méthode CART

Les principaux sinistres pouvant faire l'objet de bonis sont survenus en 2009, 2012 et 2013. Pour les autres années, les écarts sont finalement assez minimes avec les provisions dossiers/dossiers.

Contrairement au graphique de la partie III.A.2 du chapitre 3, avec l'estimation des provisions à l'ultime par une méthode de triangle classique, il n'y a pas d'incohérence particulière.

V. Répercussions sur le ratio de Solvabilité de l'exercice 2019

Maintenant que les estimations des provisions à l'ultime des sinistres Automobile ont été déterminées, notamment ceux des sinistres Corporels graves, il est possible d'évaluer l'impact quantitatif sur le ratio de Solvabilité de la MAPA.

Impact sur les meilleures estimations des provisions pour sinistre

La délimitation obtenue dans ce mémoire entre les sinistres graves et les sinistres attritionnels est sensiblement la même que celle qui était appliquée dans le cadre des calculs de la meilleure estimation. De ce fait, les impacts quantitatifs sur les meilleures estimations à l'ultime des provisions techniques des sinistres attritionnels ne sont pas significatifs.

Pour ce qui est de l'évaluation des meilleures estimations des sinistres graves, la différence est plus importante. En effet, les sinistres graves sont actuellement évalués de manière prudente, en reprenant les provisions dossiers/dossiers et en les actualisant suivant un cadencement prédéfini. Les résultats obtenus via la méthode CART diminue l'évaluation des provisions à l'ultime de 5,1M€.

Dans le même temps, cette nouvelle évaluation des sinistres graves a évidemment un impact sur la réassurance. Certains bonis/malis n'ont aucun impact sur la réassurance, soit parce que la priorité du traité concerné est plus haute, soit parce que la franchise (AAD) n'a pas été remplie. Au contraire, lorsque les conditions étaient déjà remplies, la part des provisions cédées à la réassurance diminue donc avec la diminution des provisions à l'ultime.

Enfin, parmi les 5,1M€ de bonis dégagés sur les sinistres graves, 3,0M€ sont en fait réassurés. Les bonis nets de réassurance sur les sinistres graves se portent donc à 2,1M€.

Par simplification, l'impact de la courbe des taux ne sera pas traité, ni même l'impact potentiel sur les meilleures estimations des provisions pour prime.

Impact sur le bilan prudentiel

Les changements effectués sur le bilan prudentiel sont les suivants :

- -5,1M€ sur les meilleures estimations au passif ;
- -3,0M€ sur les meilleures estimations à l'actif ;
- -0,3M€ sur la marge de risque Non-Vie.

Ces changements ont pour conséquences d'augmenter le montant des impôts différés au passif (considérés comme un tout) de 0,7M€, et d'augmenter les fonds propres SII de 1,7M€.

Impact sur le SCR

L'impact sur le SCR est principalement porté sur les trois éléments suivants :

- ✓ Le SCR Risque de souscription Non-Vie

Les bonis nets de réassurance étant de 2,1M€, le SCR de risque de Prime et de Provisionnement diminue donc légèrement, le risque passant de 31 253K€ à 30 892K€.

- ✓ Le SCR de Défaut de Contreparties

Les cessions auprès des réassureurs de la MAPA étant moins importantes, le risque de défaut pourrait se retrouver diminué. Mais l'impact n'est ici pas significatif, les notations des réassureurs de la MAPA étant de très bonne qualité.

- ✓ La capacité d'absorption des impôts différés

Comme vu dans la partie bilan prudentiel, les impôts différés ont augmenté de 0,7M€. Ce montant sera retranché au BSCR, pour calculer le SCR.

Impact sur le ratio de Solvabilité

Enfin, le ratio de solvabilité de la MAPA pourrait **gagner 6 points**, par le double effet de l'augmentation des fonds propres SII et la diminution du SCR, sur l'exercice 2019.

Conclusion et perspectives

Les exigences liées à la directive Solvabilité II, sont vastes, et notamment pour la fonction clé Actuariat. De telles exigences sont complexes à respecter dès l'application de la directive. Cela concerne parfois des travaux de fonds, qu'il faut parvenir à mettre en place, en synergie avec la stratégie de l'entreprise. L'objectif initial étant la sécurité économique des sociétés d'assurances vis-à-vis des assurés.

Pour les entreprises d'assurances, il faut alors voir les contraintes comme des opportunités ; des travaux chronophages comme une nécessité. Pour prendre l'exemple de la Qualité des données, la mise en place d'un tel dispositif, d'une telle architecture, peut prendre plusieurs années. Mais, n'est-elle pas nécessaire dans le monde de l'assurance, hyper concurrentiel, et de plus en plus connecté ? Les données sont évidemment un moyen de se distinguer.

C'est dans cet esprit que ce mémoire a été conçu. Il permet de répondre à des exigences réglementaires, notamment sur la justification des hypothèses qui sous-tendent la conception des meilleures estimations de sinistre. Il permet également, de mieux connaître les données, de prendre le temps de les manipuler, et, si cela est possible, d'optimiser des résultats, qui pourraient être bénéfiques à l'entreprise.

Pour ce qui est des exigences réglementaires, plusieurs points liés à la Qualité des données et au rôle de la fonction actuariel ont pu être soulevés.

La réglementation impose certaines segmentations dans le traitement des meilleures estimations, appelées ligne d'activité. Mais celles-ci peuvent elles-mêmes être segmentées, comme l'entend l'entreprise, si cela se justifie pour améliorer les estimations, ou diminuer la volatilité. La MAPA distinguait déjà les sinistres RC issus d'accident comprenant des dommages corporels, de ceux qui n'étaient que strictement matériel. L'analyse qui a été faite dans ce mémoire montre que ce choix est judicieux et qu'il permet de distinguer des sinistres dont les comportements indemnitaires diffèrent totalement.

Une seconde segmentation est appliquée à la MAPA, uniquement sur les sinistres RC comprenant des dommages corporels, en distinguant les sinistres graves, des sinistres dits « attritionnels ». Initialement, la limite entre les deux catégories de sinistres a été définie aux prémices de la directive Solvabilité II, sur la base de données marchés à 275K€. Mais, il n'était pas possible pour la MAPA de justifier concrètement un tel choix. Pour se faire, les travaux de ce mémoire se sont appuyés sur des outils de la théorie des valeurs extrêmes. À noter que certains des résultats reposent sur des analyses graphiques, parfois difficilement interprétables. Ces dernières ont donc été complétées par des méthodes pragmatiques, reposant sur des pratiques courantes du monde de la réassurance. Afin d'étoffer les éléments concordants, une deuxième étude, a posteriori, a été réalisée, basée sur l'analyse de l'incertitude dans le cadre du provisionnement. Ici encore, la distinction entre les deux populations est justifiée par de grandes différences, notamment sur les cadencements des paiements, ou la volatilité des résultats.

Ces travaux ont permis de consolider les acquis de la MAPA, et d'avoir une base solide pour optimiser les meilleures estimations de sinistre, notamment sur les sinistres graves. Les résultats de l'estimation des provisions à l'ultime sur ces sinistres par la méthode déterministe Chain Ladder laissait voir toutes les limites d'un tel modèle sur des données instables. C'est d'ailleurs pour cette raison que le Service Actuariat faisait le choix de reprendre la valeur des provisions dossiers/dossiers, les estimations des sinistres graves par Chain Ladder ne pouvant être exploitées. Mais les analyses de bonis/malis

montrent que ces sinistres graves sont globalement surestimés et laissent apercevoir qu'une optimisation est possible.

Des méthodes récentes d'évaluation des provisions à l'ultime utilisent les sinistres ligne à ligne plutôt que des triangles agrégés, permettant d'exploiter toutes les informations liées à un sinistre pour pouvoir en estimer le coût total. Cela repose notamment sur le modèle CART. Ces modèles sont assez sensibles à la volumétrie des données, qui a été une des contraintes les plus importantes de ce mémoire. La MAPA est une société d'assurance mutuelle avec un portefeuille Automobile de 100 000 véhicules, loin des volumes de certaines entreprises concurrentes. Elle doit donc faire avec cette contrainte pour établir des tarifs, faire des analyses de portefeuille ou travailler sur les provisions techniques. Le travail sur les données doit d'autant plus être exigeant, et le recul sur les résultats est indispensable. Les modèles utilisés dans ce mémoire ne sont donc pas les plus complexes ou élaborés existant dans le domaine de l'Actuariat. Mais les contraintes de données ne permettent pas d'essayer une multitude de concepts mathématiques, mais d'essayer de mettre au point le modèle le plus adapté, et de trouver des moyens pour avoir des résultats fiables.

Plusieurs modèles ont été comparés, en modifiant notamment les variables explicatives, au détriment de pouvoir faire l'estimation de tous les sinistres graves. Le choix a ici été fait de conserver le meilleur modèle, et de limiter l'ampleur de son périmètre.

En termes de méthodologie, une deuxième approche, plus complexe, était possible : faire le CART sur l'ensemble des sinistres, clos et ouverts. Les sinistres encore ouverts auraient alors été sous-représentés, et censurés à droite. L'objectif aurait alors été d'estimer la durée de vie restante des sinistres encore ouverts, et de mettre des poids à ces sinistres. Mais le modèle utilisé dans ce mémoire ne permettait déjà pas de choisir l'arbre optimal (trop élagué), alors un modèle plus complexe, avec des estimations d'estimations aurait certainement eu des résultats trop instables.

Finalement, l'approche choisie a permis de créer un modèle qui semble viable, et qui estime les provisions à l'ultime des sinistres graves, sinistre par sinistre, pour peu qu'ils ne soient pas « trop graves » (>1M€) ou trop récents (moins de 3 ans d'ancienneté).

La même étude globale pourrait être réalisée sur les lignes d'activité Dommage aux Biens ou RC Générale, avec la distinction entre les sinistres attritionnels et graves, et l'évaluation ligne à ligne. Il est d'ailleurs possible de traiter les sinistres graves RC Générale avec dommages corporels, peu nombreux, avec les sinistres graves RC Automobile avec dommages corporels, si leurs caractéristiques le permettent.

Il faut aussi relativiser le gain avec le temps passé à l'obtenir : cette nouvelle estimation permet de gagner approximativement 6 points de ratio de solvabilité. En effet, les sinistres graves sont généralement réassurés. Avoir une estimation au plus juste de ces sinistres n'implique pas nécessairement un dégagement important de fonds propres, malgré le postulat initial de l'étude sur les bonis/malis. Et, cette approche ne pourra pas être appliquée immédiatement après ce mémoire : elle devra être réitérée et confirmée avec un nouveau jeu de données pour s'assurer de la robustesse des différents modèles.

Même si toute la partie recherche et tous les codes sont, de fait, existants et réutilisables, c'est une étude qui peut être longue par rapport aux ressources du Service Actuariat de la MAPA, par rapport au gain espéré. Il faut également ajouter à cela le temps nécessaire à la validation d'approches comme

les modèles CART ligne à ligne, en comparaison avec du provisionnement agrégé via Chain Ladder, dans l'exercice du rapport de la fonction clé Actuariat.

Enfin, concernant les résultats indirects, ce mémoire a permis plusieurs axes de progression sur la Qualité des données. Le premier est la construction d'une base de données assez sophistiquées, qui pourrait avoir son utilité dans d'autres travaux actuariels, et notamment sur des problématiques de réassurance. Le second concerne les connaissances acquises de ces bases de données, et notamment sur la gestion des sinistres Corporels. Les échanges avec les gestionnaires et leur responsable ont permis de mieux comprendre, et permettront de mieux diriger nos travaux futurs. Le troisième concerne les outils créés dans le cadre de ce mémoire : Mack, Bootstrap ou *Mean Excess Function*, tous intégrés dans des tableurs, utilisables par tous les membres de l'équipe. Ils seront exploités, et permettront d'enrichir certains de nos travaux.

Et, du point de vu de la fonction clé Actuariat, les résultats de ce mémoire enrichiront le rapport actuariel de la MAPA, qui pourra conforter le Conseil d'Administration de la société d'assurance mutuelle que les meilleures estimations des provisions sont évaluées de manière adéquate et que les hypothèses liées aux modèles sont justifiées.

Bibliographie

Beirlant J., Goegebeur Y., Segers J., Teugels J. (2004), *Statistics of Extremes : Theory and applications*, Wiley Series in Probability and Statistics, Chichester.

Benlagha N., Grun-Réhomme M., Vasechko O. (2009), Les sinistres graves en assurance Automobile : Une nouvelle approche par la théorie des valeurs extrêmes, *Revue MODULAD* n°39.

Breiman L., Friedman J., Olshen R. A., Stone C. J. (1984), *Classification and regression trees*, Chapman and Hall.

Charpentier A., Denuit M. (2004), *Mathématiques de l'assurance non-vie*, tome II, Economica.

Deprey E., Godzinski A. (2007), Problématique de seuil dans la modélisation de la sinistralité en Réassurance Non Vie, *Mémoire Institut des Actuaire*s.

EIOPA (2014). Orientations sur les valorisations des provisions techniques.

Fabre D. (2018), Apport des méthodes d'apprentissage statistique pour le provisionnement individuel en assurance non-vie, *Mémoire Institut des Actuaire*s.

Franquet, S. (2018), Modélisation de la fréquence des sinistres graves en assurance Automobile : apports et interprétabilité des méthodes d'apprentissage statistique, *Mémoire Institut des Actuaire*s.

Gey S., Nédélec E. (2005), Model selection for CART Regression Trees, *IEEE* 51 (2), pp.658 – 670.

Journal officiel de l'Union européenne (2009). DIRECTIVE 2009/138/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL, sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (Solvabilité II).

Journal officiel de l'Union européenne (2014). RÉGLEMENT DÉLÉGUÉ (UE) 2015/35 DE LA COMMISSION, complétant la directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (Solvabilité II).

Mack T. (1993) Distribution-free calculation of the standard error of chain ladder reserve estimates, *Astin Bulletin* 23 (2), pp.213 – 225.

Orozco Penalosa S. (2015), Mise en place d'une couverture en réassurance rétrospective sur les dommages corporels de la branche automobile, *Mémoire Institut des Actuaire*s.

Partrat C., Lecoœur E., Nessi J.-M., Nisiparu E., Reiz O. (2007), Provisionnement technique en assurance non-vie, Economica.

TAILLIEU F. - Responsable du groupe de travail (2016). Groupe de travail « Best Estimate Liabilities Non-vie ». Normes Professionnelles de l'Institut des Actuaires (NPA 3).

URL : <https://www.institutdesactuaire.com/decouvrir-l-institut/textes/normes-professionnelles-25>

Tosetti A., Behar T., Fromenteau M., Menart S. (2002), Assurance, comptabilité Réglementation, Actuariat, Economica.

Zougagh F.-Z. (2018), Tarification Automobile à l'aide de modèles de Machine Learning et apport des données télématiques, *Mémoire Institut des Actuaire*s.

Table des figures

Figure 1 : Les trois piliers sous Solvabilité II	10
Figure 2 : Exemple de répartition des états de la remise annuelle Solo entre différents axes d'analyse	13
Figure 3 : Évolution de la production Automobile en cotisations et nombre de véhicules	18
Figure 4 : Évolution du S/C en fonction de la date de vue	18
Figure 5 : Comparaison du S/C nets et bruts de réassurance entre 2003 et 2019	20
Figure 6 : Schéma des méthodes d'évaluation des provisions en fonction du périmètre	22
Figure 7 : Évolution du coût moyen Automobile	25
Figure 8 : Évolution du coût moyen des sinistres Automobile sans dommages corporels	26
Figure 9 : Évolution des coûts moyens Automobile Autre que RC lors d'un accident impliquant du « corporel »	26
Figure 10 : Évolution des coûts moyens Automobile RC « Corporel » et Garantie du conducteur	27
Figure 11 : Part du portefeuille et coût moyen en fonction de l'âge du conducteur principal.....	28
Figure 12 : Part du portefeuille et coût moyen en fonction de l'ancienneté de permis du conducteur	28
Figure 13 : Part du portefeuille et coût moyen en fonction du genre du véhicule.....	29
Figure 14 : Part du portefeuille et coût moyen en fonction de l'âge du véhicule	29
Figure 15 : Part du portefeuille et coût moyen en fonction de la puissance du véhicule.....	30
Figure 16 : Part du portefeuille et coût moyen en fonction du groupe SRA.....	30
Figure 17 : Part du portefeuille et coût moyen en fonction de la classe prix SRA	31
Figure 18 : Schéma de la vie d'un sinistre	32
Figure 19 : Triangle standard et notations	35
Figure 20 : Facteurs de développement individuel de la 1 ^{ère} année de développement Automobile Matériel	53
Figure 21 : Exemple de facteur de développement individuel considéré comme atypique.....	53
Figure 22 : Comparaison des facteurs de développement RC « Corporel » / Autre que RC.....	54
Figure 23 : Vérifications des hypothèses sur les facteurs de développement.....	54
Figure 24 : Vérifications des hypothèses sur la volatilité interne des triangles.....	55
Figure 25 : Fonctions de répartition des différentes catégories de sinistre Automobile mis en As-If..	59
Figure 26 : Fonction moyenne des Excès des sinistres RC « Corporel ».....	61
Figure 27 : Pentas Excès moyen en fonction du seuil choisi	62
Figure 28 : Estimateurs par maximum de vraisemblance de α	62
Figure 29 : Estimateur α en fonction du seuil choisi.....	63
Figure 30 : Représentations graphiques du test de Kolmogorov-Smirnov	64
Figure 31 : Facteurs de développement RC « Corporel » attritionnel et intervalle de confiance à 95%	65
Figure 32 : Comparaison des provisions dossiers/dossiers et de l'évaluation à l'ultime par Chain Ladder.....	66
Figure 33 : Fonction de répartition des évaluations à l'ultime des sinistres RC « Corporel » attritionnels	67
Figure 34 : Facteurs de développement RC « Corporel » grave et intervalle de confiance à 95%	68
Figure 35 : Facteurs de développement RC « Corporel » grave et « Tails factors »	68
Figure 36 : Figure 31 : Comparaison des provisions dossiers/dossiers et de l'évaluation à l'ultime par Chain Ladder + « Tails factors ».....	69

Figure 37 : Arbre maximal du 1er modèle CART	75
Figure 38 : Arbre élagué en fonction du nombre de sinistres minimum par feuille du 1er modèle CART	76
Figure 39 : Variables explicatives significatives et ordre d'importance (1er modèle CART)	76
Figure 40 : Sinistres Corporels RC clos > 2400€ et Répartition des règlements dans le temps	78
Figure 41 : 2ème modèle, paramètre de complexité et optimisation	79
Figure 42 : 2ème modèle, arbre optimal.....	79
Figure 43 : 3ème modèle, paramètre de complexité et optimisation	80
Figure 44 : Erreur en fonction de la complexité du modèle.....	81
Figure 45 : Comparaison des valeurs réalisées et prédites (CART) des sinistres graves sur la base d'apprentissage	83
Figure 46 : Comparaison des valeurs réalisées et prédites (CART) des sinistres graves sur les bases d'apprentissage et de validation	84
Figure 47 : Comparaison des sinistres ouverts en RC « Corporel » Automobile à fin décembre 2019 et évaluation des provisions à l'ultime par la méthode CART.....	86

Table des tableaux

Tableau 1 : Programme de réassurance Automobile entre 2003 et 2019	20
Tableau 2 : Segmentation des données des sinistres Automobile, en €, de 2003 à 2019, vision à fin 2019.....	24
Tableau 3 : Comparaison des caractéristiques des sinistres Automobile	51
Tableau 4 : Comparaison des montants des provisions dossiers/dossiers	52
Tableau 5 : Comparaison des résultats de l'estimation des provisions à l'ultime entre Automobile Autre que RC et Automobile RC « Corporel » par la méthode classique Chain Ladder.....	55
Tableau 6 : Triangle des bonis/malis sur les sinistres Automobile Autre que RC	56
Tableau 7 : Triangle des bonis/malis sur les sinistres Automobile RC « Corporel »	57
Tableau 8 : Impact de la prise en compte de l'inflation et la mise en ultime sur les sinistres Automobile.....	59
Tableau 9 : Statistiques des différentes catégories de sinistre Automobile mis en As-If	60
Tableau 10 : Caractéristiques des sinistres au-dessus du seuil α choisi.....	63
Tableau 11 : Résultats des tests de Kolmogorov-Smirnov en fonction du seuil choisi	64
Tableau 12 : Comparaison de l'incertitude en fonction du seuil choisi	70
Tableau 13 : Répartition des sinistres RC « Corporel » Automobile, corrigés de l'inflation	73
Tableau 14 : Variables explicatives potentielles de l'estimation du coût total des sinistres RC « Corporel »	75
Tableau 15 : Comparaison des modèles CART sur les bases d'apprentissage	81
Tableau 16 : Comparaison des modèles CART sur les bases de validation	82
Tableau 17 : Erreur constatée sur le modèle CART choisi et optimisé	82
Tableau 18 : Variables explicatives significatives du modèle final et ordre d'importance	83
Tableau 19 : Comparaison entre les coûts totaux réels, les provisions dossiers/dossiers en N+2 et les estimations CART.....	85

Liste des sigles et abréviations

ACPR : Autorité de Contrôle Prudentiel et de Résolution

BE : Best Estimate (Meilleures Estimations)

CART : Classification And Regression Trees

EIOPA : European Insurance and Occupational Pensions Authority

IFRS : International Financial Reporting Standards

MAPA : Mutuelle d'Assurance des Professionnels de l'Alimentaire

MEF : Mean Excess Function

ORSA : Own Risk and Solvency Assessment

RC : Responsabilité Civile

S2 : Solvabilité 2

SCR : Solvency Capital Requirement

SRA : Sécurité et Réparation Automobiles

TVE : Théorie des Valeurs Extrêmes

XS : Excédent de sinistre