

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 15/03/2022

Par : **Axel CHARTREL**

Titre : **Classification *a priori* de sinistres
pour le provisionnement individuel**

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

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

Membres présents du jury de la filière

Entreprise : KPMG

Nom : Wissal SABBAGH

Signature : 

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

Directeur du mémoire en entreprise :

Nom : Charles-Erwin ELIACHAR

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

Secrétariat :



Bibliothèque :

Signature du candidat



Résumé

L'estimation des provisions pour sinistres à payer en assurance non-vie repose à l'heure actuelle majoritairement sur des données agrégées. Cependant, la disponibilité croissante de données de sinistres individuelles suscite un grand intérêt pour le provisionnement ligne à ligne.

Plusieurs modèles, simulant à la fois les sinistres connus et encore inconnus de l'assureur, peuvent être raffinés par l'ajout en amont d'une classification des sinistres pour constituer des groupes de sinistres au développement semblable.

Deux méthodes de classification s'appuyant sur l'apprentissage statistique sont mises en place. La première s'appuie sur l'algorithme des k-moyennes et n'est pas supervisée, tandis que la seconde l'est et repose sur un arbre de classification. Une classification manuelle proche des méthodes traditionnelles des gestionnaires est ajoutée pour fournir un point de comparaison basique. Les modèles de provisionnement individuel sont lancés séparément sur chaque classe. Les résultats sont évalués à partir de *backtesting* à horizon court, et les charges ultimes sont confrontées à des valeurs issues de méthodes classiques et de jugements d'experts, puis comparées entre elles.

Les analyses et les réflexions menées au cours du travail contribuent à approfondir la compréhension des modèles de provisionnement individuel utilisés, de leurs avantages et de leurs limites et permettent de discuter l'intérêt de procéder à une classification.

Mots-clés : Assurance non-vie, Provisionnement individuel / ligne à ligne, Classification, K-means, arbre de classification, Backtesting

Abstract

In property and casualty insurance, the estimation of provisions for claims is today mostly based on aggregated data. However, the interest in micro-reserving is increasing due to the large volume of individual claim data available.

Several models are able to simulate claims both known by the insurer and not yet reported. They can be refined by adding beforehand a classification of claims that forms groups in which the developments of claims are homogeneous.

Two classification methods based on machine learning are implemented. The first one relies on the k-means algorithm that is unsupervised. The second one is supervised and uses a tree. A handmade classification, close to what claims managers may do, is added to provide a basic comparison. Micro-reserving models are run separately on each class. Results are then evaluated with backtesting for short horizon and ultimate incurred are confronted with values obtained with classical methods and experts' judgements, then compared to one another.

The analyses and reflections led through this work contribute to enriching the comprehension of the used micro-reserving models including their advantages and limits. It allows to discuss the interest of adding a classification.

Keywords : P&C insurance, Micro-reserving, Classification, K-means, Classification tree, Backtesting

Note de synthèse

Contexte et enjeux

La constitution des provisions pour sinistres à payer par un assureur non-vie est cruciale. Cette nécessité provient de l'inversion du cycle de production : au moment de la vente du produit, l'assureur ne sait pas quelle somme il devra régler car les sinistres éventuels ne sont pas survenus, mais il doit être en capacité de payer pour respecter le contrat. Si les provisions estimées sont trop élevées, l'assureur immobilise de l'argent qu'il aurait pu placer sur des actifs risqués et se prive d'un gain potentiel. Inversement, un sous-provisionnement conduit à des réserves de liquidités insuffisantes qui peuvent mener à la faillite.

L'estimation au plus juste des provisions est un besoin historique, aussi des méthodes ont été développées pour traiter cette tâche il y a plusieurs dizaines d'années. Elles s'appuient sur des données sous forme agrégée : les sinistres sont agrégés par année de survenance et par année de développement, puisque les sinistres peuvent évoluer durant plusieurs années. Cette agrégation conduit à un triangle de développement, dont la complétion permet d'obtenir des réserves à l'ultime. La méthode de Chain-Ladder, présentée en 1938, calcule l'espérance des PSAP. La méthode de Mack, datant de 1993, fournit en plus une estimation de la volatilité. Ces méthodes, très largement utilisées et reconnues par les actuaires, ont été conçues à une époque où les capacités de calcul et de stockage de données n'avaient rien de comparable à aujourd'hui.

Les données sont maintenant collectées systématiquement avec un niveau de détail élevé, ce qui pousse la recherche actuarielle à développer de nouvelles méthodes de provisionnement s'affranchissant de l'agrégation pour traiter chaque sinistre individuellement et lui attribuer une réserve. Ce type de provisionnement est aussi appelé « ligne à ligne ».

Les travaux successifs de Riffaud [RIFFAUD 2019] et Primel [PRIMEL 2021] ont conduit à la mise en place de deux modèles permettant de simuler le développement de sinistres afin d'en tirer une distribution de réserves. La première étape consiste à générer un certain nombre de sinistres tardifs, absents de la base de données car par définition survenus mais pas encore déclarés. Cependant ces modèles ne prennent pas en compte d'autres informations disponibles. Nous cherchons donc à constituer des groupes de sinistres homogènes à partir de ces informations afin de lancer les modèles de développement sur chaque groupe séparément.

Modèles utilisés

Les deux méthodes de développement s'appuient sur une simulation de la durée de vie basée sur l'estimateur de Kaplan-Meier pour faire évoluer les sinistres chaque année jusqu'à leur clôture.

Le premier modèle, additif, voit l'ultime comme une somme de règlements (potentiellement nuls) effectués chaque année. Une notion de processus de paiement incrémental conduit à simuler le paiement associé à un sinistre pour une année de développement

en tirant uniformément dans l'historique des paiements effectués pour ce développement. L'utilisation de montants de règlements nécessite un traitement de l'inflation passée et future. A cet effet, la méthode de Verbeek est mise en place sur la base de données. Le second modèle, multiplicatif, est un pendant ligne à ligne de la méthode de Chain-Ladder : la dernière charge connue est multipliée par un facteur de développement chaque année. Les facteurs sont cette fois individuels et ont été calculés afin de constituer un historique dans lequel le tirage aléatoire a lieu. Un raffinement est ici apporté avec la prise en compte de la corrélation entre charge et facteur de développement par une copule empirique.

Deux méthodes de classification s'appuyant sur l'apprentissage statistique sont retenues. La première est l'algorithme des k -moyennes, ou *k-means*. Elle n'est pas supervisée, c'est-à-dire qu'aucune variable n'indique de « vrai » groupe pour chaque sinistre, ainsi il n'y a pas d'*a priori* sur la manière de constituer les classes. L'homogénéité entre les sinistres se traduit par la proximité entre les variables pour la distance euclidienne. Cet algorithme constitue finalement des des groupes convexes qui minimisent la variance intra-classe.

Une méthode supervisée est aussi mise en place sous la forme d'un arbre de classification. Cette méthode partitionne récursivement les données en trouvant à chaque étape le critère qui constitue deux groupes les plus homogènes pour la variable à prédire. L'arbre est ensuite élagué à partir d'un compromis entre efficacité et complexité.

Application des modèles

La base de données utilisée regroupe 75 000 sinistres de la branche Responsabilité Civile médicale survenus entre 2000 et 2019. La branche étudiée présente des spécificités comme un développement long et un grand nombre de sinistres tardifs, ce qui rend nécessaire le modèle qui les génère. Pour chaque sinistre sont fournies des informations sur le développement ainsi que des variables renseignant sur la nature du sinistre comme corporel/matériel ou litigieux/amiable.

Pour commencer, les modèles de provisionnement sont évalués en l'état par un *back-testing* à un an : les informations relatives à la dernière année sont supprimées et ce développement est simulé puis comparé avec les résultats réels. Un écart apparaît sur le modèle multiplicatif mais après analyse il est attribué à quelques développements atypiques survenus la dernière année qui ont causé des facteurs élevés. Ce modèle, comme Chain-Ladder auquel nous le comparons, ne peut anticiper ce type d'évènement. Le *back-testing* est aussi poussé à deux ans. Ce type d'évaluation a une pertinence limitée dans la mesure où les modèles visent des développements à l'ultime mais la profondeur restreinte de l'historique ne permet pas de se priver de beaucoup d'années.

Les classifications débutent ensuite. Des statistiques descriptives montrent que les sinistres se développent différemment suivant les modalités de certaines des variables définissant des catégories. Elles sont donc utilisées pour mettre en place une classification de base : les données sont simplement séparées en deux à partir d'une variable, par exemple d'un côté les sinistres corporels et de l'autre les matériels. Les modèles de développement sont ensuite lancés indépendamment sur chaque groupe et les ultimes sont regroupés. Cette manière de procéder réduit l'erreur de prédiction à un an sur la catégorie des si-

nistres les moins coûteux. A l'ultime, les effets par catégorie vont dans le sens attendu par la connaissance d'expert et les statistiques descriptives. Cette classification, bien que très élémentaire, fait apparaître des résultats encourageants mais manque de finesse.

L'étape suivante est l'utilisation du *machine learning*. Pour ces méthodes, les variables binaires de catégorie sont retenues. Deux variables quantitatives sont ajoutées : la charge estimée à l'ouverture du sinistre, qui porte l'appréciation du gestionnaire, et le délai de déclaration, qui permet de raffiner la notion de sinistre tardif. En effet, la majorité des sinistres considérés comme tardifs le sont à cause du calendrier mais n'ont pas un délai de déclaration très élevé.

Les classifications ne peuvent être entièrement évaluées seules car leur intérêt se justifie par le développement des sinistres qui les suit. Ainsi, l'analyse descriptive des classes est complétée par des méthodes d'évaluation des réserves auxquelles elles conduisent.

Pour le *k-means*, le nombre de classes est déterminé à partir de critères numériques et de manière à obtenir une stabilité des classes par rapport à des variations dans les données. Les *clusters* ainsi constitués sont interprétables en fonction des variables, il y a notamment un groupe de sinistres tardifs, un de très tardifs, un de graves et un groupe de matériels amiables.

Un *backtesting* après un an de développement avec le modèle additif montre une amélioration de la prédiction sur toutes les classes sauf celle des très graves. Ceci s'explique par le petit effectif de cette classe et la grande volatilité des paiements d'un sinistre à l'autre. Les effets sont identiques avec le modèle multiplicatif mais moins marqués, à la fois pour l'amélioration et la dégradation. A l'ultime, une évolution des montants par rapport au modèle sans classification est observée sur chaque classe. Celle-ci est cohérente avec les interprétations des classes, par exemple les ultimes associés à la classe des graves deviennent plus élevés.

A titre de comparaison, une charge ultime est calculée en développant le triangle agrégé correspondant à chaque classe par la méthode de Chain-Ladder. Les résultats sont proches de ceux obtenus par le modèle additif sauf sur les *clusters* de tardifs. En effet, des triangles avec uniquement des sinistres tardifs ont plusieurs lignes vides pour les survenances récentes ce qui aboutit à des ultimes nuls pour les survenances associées. L'effet habituel de prise en compte implicite des sinistres tardifs par Chain-Ladder ne fonctionne pas après cette classification, alors que les modèles individuels sont efficaces.

Ensuite, la variable de supervision pour l'arbre de classification est construite. La charge après un an de développement, disponible dans une bonne partie de la base et assez informative sur la vie du sinistre, est retenue. Sept classes sont constituées à partir de cette variable découpée logarithmiquement.

Compte tenu du nombre restreint de variables, la complexité atteinte par l'arbre est limitée. Une forêt aléatoire ne donne pas de meilleurs résultats et perd l'interprétabilité, ce qui conduit à conserver le choix de l'arbre. Comme attendu, la charge initiale est un très bon prédicteur de la classe. Viennent ensuite des variables comme corporel/matériel. La charge ultime globale simulée après cette classification est plus élevée que celle suivant le *k-means* ou celle sans classification. La comparaison des ultimes classe par classe montre une cohérence entre le groupe et la charge associée. Cependant, le modèle multiplicatif prédit nettement plus sur la classe la plus grave, c'est-à-dire celle ayant la charge à un an la plus élevée. Ce phénomène n'est pas attribué à la copule, mais disqualifie le modèle

multiplicatif dans ce cas. Il permet toutefois de rappeler que les modèles statistiques ne sont pas adaptés à la simulation de sinistres atypiques et que ceux-ci doivent être écartés *a priori*.

Des tests de stabilité sont effectués et permettent d'appréhender le comportement du modèle face à l'ajout d'information et de faire ressortir plusieurs caractéristiques. Pour cela, la base de données est restreinte au périmètre des sinistres survenus en 2016 ou avant et déclinée suivant quatre niveaux d'information allant de 2016 à 2019. Les modèles sont lancés sur chaque base et leurs résultats comparés. Les ultimes résultant du modèle multiplicatif en vision 2019 sont plus élevés. Ce résultat a aussi été identifié lors du premier *backtesting*, et l'ultime en vision 2017 se trouve plus élevé que celui prédit par Chain-Ladder. L'examen des facteurs de développement montre que certains facteurs des derniers développements du triangle vision 2017 sont bas. Alors que leur effet est très marqué avec la méthode de Chain-Ladder, le modèle multiplicatif en tient peu compte car peu de sinistres sont encore ouverts à ce moment. Le modèle est donc moins sensible aux variations tardives. Les classifications ont peu d'effet sur la stabilité et globalement le modèle additif est plus stable.

Afin de ne pas limiter les comparaisons à une valeur obtenue par l'application naïve de Chain-Ladder, les triangles sont fournis à des actuaires afin qu'ils les développent selon leur jugement d'expert. Les résultats des modèles de provisionnement individuel sont compris dans l'intervalle de prévision obtenu. Au delà des valeurs, les échanges auxquels l'exercice donne lieu illustrent la diversité des choix possibles mais font ressortir certaines hypothèses systématiquement prises qu'une méthode de classification ne peut pas remplacer.

Bilan

Pour finir les modèles et classifications associées sont comparés entre eux sur la base des charges ultimes qu'ils retournent. Le modèle additif est toujours inférieur au multiplicatif, quelle que soit la classification mise en place. Ceci s'explique notamment par l'utilisation des paiements au lieu des charges dans le contexte d'une branche longue et d'un historique limité. La méthode de Chain-Ladder fournit une valeur entre les modèles additifs et multiplicatifs.

Les classifications ont un effet cohérent sur chaque classe et ont tendance à augmenter la charge ultime au total. Elles ne réconcilient pas les valeurs obtenues par les deux modèles, ce qui suggère finalement que le choix d'une méthode de développement est plus important que la classification qui peut être mise en œuvre en amont.

L'arbitrage entre les modèles, compte tenu des spécificités identifiées, dépendra dans le cas général des caractéristiques de la branche étudiée. En effet, la portée des conclusions est limitée par les particularités de la branche dont elles proviennent. Une analyse à partir d'autres données permettrait d'avoir une vision plus globale.

Executive summary

Context and stakes

For a P&C insurer, it is crucial to book provisions for claims. This need arises from the reversed production cycle : when the product is sold, the insurer ignores the final amount it will have to pay since the accidents have not yet occurred. However, it must be able to pay to respect the contract. If the estimated provisions are too high, the insurer immobilises some money that could have been invested on risky assets. Inversely, underestimating provisions leads to an insufficient reserve of liquidity that may cause a bankruptcy.

Since obtaining the best estimation of provisions is a historical necessity, methods to treat this task have been developed dozens of years ago. They rely on aggregated data. That is, claims are aggregated by occurrence year and by development year, since claims may undergo a development for several years. This aggregation leads to a development triangle which is developed to obtain ultimate reserves. The Chain-Ladder method, presented in 1938, calculates the expected value of provisions. Mack's method, introduced in 1993, provides an estimation of the volatility. Although they are widely used and recognised by actuaries, those methods were conceived at a time when computational power and data storage had nothing in common with today.

Today, data is systematically collected with a high level of detail, which encourages the actuarial research to develop new ways of reserving that avoid the aggregation to treat each claim individually and calculate a reserve for it.

The successive works of Riffaud [RIFFAUD 2019] and Primel [PRIMEL 2021] led to the implementation of two models that can simulate the development of claims and derive a distribution of reserves. The first step consists in generating a number of late claims that are not yet reported in the database. These models, however, do not take into account other available information. We thus look to form homogeneous groups of claims based on that information in order to run the development models on each class separately.

Models implemented

Both development methods rely on the simulation of a claim's duration by the Kaplan-Meier estimator. Each year, the claim evolves until its closure. The first model is called "additive" and considers the ultimate amount as a sum of (potentially null) payments made every year. An incremental payment process leads to simulating an individual payment for a development year by drawing evenly a paid amount in the history for this development year. Using payments requires to take into consideration past and future inflation. To that end, the Verbeek method is implemented on the database. The second method is an individual equivalent of Chain-Ladder : the last known incurred is multiplied by a development factor every year. The individual factors are calculated to form a historical base in which the random draw is done. This model is refined by a copula that accounts for the correlation between incurred and development factor.

Two machine learning-based classification methods are selected. The first one, k-means, is unsupervised. That is, no label indicates a "true" class for the claims so there is no preconception on the classes. The homogeneity between claims translates into proximity for the Euclidean distance between variables. The resulting groups are convex and minimise within-cluster variance. A classification tree is implemented as a supervised method. It recursively partitions data by determining a locally optimal split at each step. The tree is then pruned based on a trade-off between cost and complexity.

Practical application

The database counts 75 000 claims from the medical liability branch, occurred between 2000 and 2019. This branch has several particularities such as a long development and a high number of late claims. This point makes the late claims generation necessary. For each claim, information regarding the development is available along with characteristics of the claim such as bodily injury/property or dispute/out of court.

As a first evaluation, a backtesting is done on the development models. The information concerning the last development is removed then this development is simulated and compared with the real value. A difference appears on the multiplicative model but an analysis shows that it is caused by a few atypical developments during the last year that led to high factors. Like Chain-Ladder to which it is directly compared, this model cannot predict this type of event. Backtesting is also pushed to two years. This evaluation has a limited relevance since the primary goal of the models is to simulate the ultimate, but the limited history does not allow to remove more years.

The classifications are then implemented. Descriptive statistics show that claims behave differently depending on several variables that define categories. These variables are therefore used to define a basic classification : the database is split in two parts based on a variable. For example it can be a group of bodily injury claims and a group of property ones. The development models are then run on each class before the results are regrouped. This process decreases the one year prediction error on the less costly group. The effects on the ultimate loss for each category go in the expected direction. Although elementary, this classification shows encouraging results but is not refined enough.

The next step is the use of machine learning. For these methods, binary categorical variables are retained. Two quantitative variables are added : the estimated incurred at the opening of the claim, which carries the manager's assessment, and the reporting delay, that allows to refine the notion of lateness. In fact, the majority of claims considered late are late because of the calendar but do not have a very long reporting delay. The classifications cannot be fully evaluated on their own as their interest is justified by the development of claims that follow them. Thus, the descriptive analysis of the classes is complemented by evaluating the reserves they lead to.

For the k-means, the number of classes is determined based on numerical criteria and in such a way as to obtain stability of the classes in relation to variations in the data. The resulting clusters are interpretable based on the variables. In particular, there is a group of late claims, one of very late claims and one of out of court property claims.

A backtesting after one year of development with the additive model shows an improvement in the prediction on all classes except the one of very serious claims. This is explained by the small size of this class and the high volatility of payments from one claim to another. The effects are identical with the multiplicative model but less marked, both for improvement and deterioration. For the ultimate loss, a change in the amounts compared to the model without classification is observed on each class. This is consistent with the interpretations of the classes, for example the ultimate loss associated with the severe class become higher.

For comparison, an ultimate incurred is computed by developing the aggregate triangle corresponding to each class with the Chain-Ladder method. The results are close to those obtained by the additive model except for the late triangles. Indeed, triangles with only late claims have several empty rows for recent occurrences, which results in null ultimate incurred for the associated occurrences. The usual effect of implicitly taking late claims into account by Chain-Ladder does not work after this classification, whereas the individual models are efficient.

Then, the supervision variable for the classification tree is constructed. The incurred after one year of development, which is available in a large part of the database and is quite informative about the development of the claim, is chosen. Seven classes are constituted from this logarithmically split variable.

Given the limited number of variables, the tree achieves a limited complexity. Since a random forest does not give better results and lacks the interpretability of a single tree, the tree is chosen. As expected, the initial incurred is a very good predictor of the class. It is followed by variables such as bodily/property. The global ultimate incurred simulated after this classification is higher than the one following the k-means or the one without classification. The comparison of the ultimate incurred class by class shows a consistency between the group and the associated incurred. However, the multiplicative model predicts significantly more on the most severe class, i.e. the one with the highest one-year incurred. This phenomenon is not attributed to the copula, but disqualifies the multiplicative model in this case. Nevertheless, it reminds that statistical models are not suitable for simulation atypical claims that should be excluded beforehand.

Stability tests are carried out and allow to apprehend the behaviour of the model when new information is added and to bring out several characteristics. For this purpose, the database is restricted to the perimeter of claims occurring in 2016 or earlier and four versions are made with levels of information ranging from 2016 to 2019. Models are run on each base and their results compared. The ultimate results of the multiplicative model in 2019 vision are higher. This result was also identified in the early backtesting. The ultimate incurred in 2017 vision is higher than that predicted by Chain-Ladder. Examination of the development factors shows that some factors in the last development years of the 2017 vision triangle are low. While their effect is very marked with the Chain-Ladder method, the multiplicative model takes little account of them because few claims are still open at this time. The model is therefore less sensitive to late changes. The classifications have little effect on stability and overall the additive model is more stable.

In order not to limit comparisons to a value obtained by the naive application of Chain-Ladder, actuaries are provided with the triangles and asked to develop them according to their expert judgement. The results of all micro-reserving models are included in

the prediction range. Beyond the numerical values, this exercise leads to exchanges that illustrate the variety of choices but bring out several hypotheses that are always chosen and that a classification method cannot replace.

Conclusion

Finally, the models and associated classifications are compared with one another based on the ultimate incurred they produce. The additive model is always inferior to the multiplicative, whatever the classification used. One reason for this is the use of payments instead of incurred in the context of a long branch and with a limited history. The Chain-Ladder method provides a value between the additive and multiplicative models. Classifications have a consistent effect on each class and tend to increase the ultimate load globally. They do not reconcile the values obtained by the two models, which suggests after all that the choice of a development method is more important than the classification that can be implemented beforehand.

The choice of a model, given the particularities of each one, will generally depend on the characteristics of the branch. Indeed, the reach of these conclusions is limited by the branch from where they come. An analysis based on other data would offer a broader vision.

Remerciements

Je tiens à remercier particulièrement mes encadrants, Charles-Erwin Eliachar, Damien Fabre Rudelle et Damien Primel, qui m'ont permis de travailler sur ce sujet et m'ont accordé leur confiance pour continuer ces travaux qui leur tiennent à cœur. Leur implication et leurs conseils nombreux m'ont été très précieux.

Je remercie également toute l'équipe d'actuariat non-vie de KPMG pour l'intérêt qu'ils ont porté à ce travail et pour avoir partagé leur expertise.

Enfin, j'adresse mes remerciements à l'équipe pédagogique de l'ENSAE pour le savoir qu'il m'ont transmis durant ces trois années.

Sommaire

Résumé	3
Abstract	4
Note de synthèse	5
Executive summary	9
Remerciements	13
Sommaire	15
Introduction	17
1 Contexte et enjeux du provisionnement en assurance non-vie	18
1.1 Contexte général	18
1.1.1 Principe de l'assurance non-vie	18
1.1.2 Nécessité des provisions	18
1.1.3 Les types de provisions	19
1.1.4 Le cycle de vie des sinistres	20
1.1.5 Équilibre entre mutualisation et segmentation	21
1.2 Les aspects réglementaires	21
1.2.1 La norme Solvabilité 2	22
1.2.2 La norme IFRS 17	24
1.3 Méthodes de provisionnement classiques	25
1.3.1 Représentation en triangle de développement	25
1.3.2 Les méthodes déterministes	25
1.3.3 Les méthodes stochastiques	28
1.4 Bilan et intérêt du provisionnement individuel	30
2 Présentation des modèles	32
2.1 Génération des sinistres tardifs	32
2.2 Hypothèses des modèles	32
2.2.1 État du sinistre et durée de vie	32
2.2.2 Estimation de la durée de vie	33
2.3 Le modèle additif	34
2.3.1 Présentation théorique	34
2.3.2 Pratique	34
2.4 Le modèle multiplicatif	35
2.4.1 Présentation théorique	35
2.4.2 Pratique	35
2.4.3 Prise en compte de corrélations par une copule	36
2.5 Calcul de l'inflation par la méthode de Verbeek	37
2.5.1 Intérêt du traitement de l'inflation	37
2.5.2 Principe de la méthode de Verbeek	37
2.6 Modèles de classification	39

2.6.1	La méthode des k-moyennes	39
2.6.2	Les arbres de classification	40
2.6.3	Forêt aléatoire	41
3	Application des modèles	43
3.1	Présentation des données	43
3.1.1	Branche étudiée et spécificités	43
3.1.2	Quelques statistiques descriptives	44
3.2	Backtesting	47
3.2.1	Traitement de la base	47
3.2.2	Résultats à 1 an	48
3.2.3	Résultats à 2 ans	50
3.3	Classification déterministe	51
3.3.1	Mise en œuvre pratique et méthode d'évaluation	52
3.3.2	Différentes segmentations et résultats	53
3.4	Classification par k-moyennes	57
3.4.1	Préparation des variables	58
3.4.2	Choix du nombre de classes	59
3.4.3	Analyse des classes obtenues	61
3.4.4	Stabilité des classes	63
3.4.5	Développement des sinistres classifiés et cas particulier des tardifs	64
3.4.6	Résultats du modèle	65
3.5	Arbre de classification	68
3.5.1	Choix du modèle et de la variable de supervision	69
3.5.2	Simulation à partir de la classification et résultats	73
3.6	Évaluation des résultats	74
3.6.1	Stabilité des modèles	74
3.6.2	Avis d'experts	78
3.6.3	Bilan des estimations ultimes	79
	Conclusion	81
	Bibliographie	83
	Annexes	85

Introduction

Une spécificité du secteur de l'assurance est que les contrats sont vendus et les primes encaissées avant que les éventuels sinistres ne surviennent. Les coûts ne sont donc pas connus à ce moment et l'assureur ne peut investir l'ensemble des primes perçues afin d'être capable de faire face à ses engagements et d'indemniser les victimes. Les montants qui constituent ces réserves sont les provisions pour sinistres à payer. L'évaluation au plus juste de ces provisions est un enjeu crucial pour les assureurs. En effet, une estimation trop élevée immobilise de l'argent qui pourrait être placé. A l'inverse, des réserves insuffisantes peuvent conduire à une faillite.

En assurance non-vie, les techniques de provisionnement actuellement utilisées s'appuient sur des données agrégées par année de survenance et par développement. Elles ont l'avantage d'être bien maîtrisées et faciles à mettre en œuvre. Cependant, l'agrégation entraîne la perte d'informations propres à chaque sinistre qui sont systématiquement disponibles. Elles peuvent concerner le développement, le type de sinistre ou l'assuré. Conjointement, la puissance de calcul grandissante des ordinateurs pousse à mettre au point des modèles qui prennent en compte les informations au niveau de chaque sinistre et estiment une réserve à l'échelle individuelle. Un tel provisionnement est aussi appelé « ligne à ligne ». Les avantages attendus sont nombreux et comprennent un gain de précision lié à la prise en compte de plus d'informations, mais aussi une meilleure automatisation.

Ce mémoire s'inscrit dans la continuité de travaux qui ont conduit à la mise en place de deux modèles de provisionnement individuel capables de fournir une estimation complète des réserves liées à la fois aux sinistres connus de l'assureur et aux sinistres tardifs pas encore reportés dans les données. Ces modèles en l'état n'exploitent que les informations relatives au développement des sinistres mais laissent de côté les autres. L'objectif de ce travail est d'utiliser toutes les informations disponibles pour constituer des classes de sinistres homogènes dans leur développement. Les modèles de provisionnement seront ensuite lancés séparément sur chaque classe.

Pour débiter ce mémoire, nous présenterons le secteur de l'assurance non-vie et les différents aspects du provisionnement. Ensuite, nous détaillerons les modèles de provisionnement individuel disponibles et nous expliquerons les algorithmes utilisés pour la classification. Enfin, nous appliquerons ces méthodes à une base de données réelle. Ceci permettra d'évaluer les modèles, notamment par la comparaison à des valeurs de référence. Des analyses détaillées seront menées et permettront des observations opérationnelles.

1 Contexte et enjeux du provisionnement en assurance non-vie

Dans cette partie nous décrivons le cadre de ce mémoire en introduisant les concepts essentiels de l'assurance non-vie. Ils feront apparaître la notion de provision et nous détaillerons les enjeux réglementaires qui y sont associés. Nous pourrions ensuite présenter les principales méthodes de provisionnement actuelles et la contribution que nous entendons y apporter.

1.1 Contexte général

1.1.1 Principe de l'assurance non-vie

L'assurance non-vie, aussi appelée assurance dommages ou assurance I.A.R.D (Incendie, Accidents, Risques Divers) fournit une protection face aux aléas que peuvent subir les biens d'une personne ou aux dommages que cette personne viendrait à causer à un tiers. Cette personne, l'assuré, s'acquitte d'une prime à l'assureur pour bénéficier de la couverture. En cas de survenance d'un événement, appelé sinistre, l'assureur verse éventuellement une prestation au bénéficiaire, qui peut être distinct de l'assuré.

La prestation versée en assurance non-vie est indemnitaire, c'est-à-dire qu'elle replace financièrement le bénéficiaire à son niveau avant le sinistre, sans l'enrichir. Le montant de la prestation est donc déterminé uniquement après la survenance du sinistre. Ce principe s'oppose à la prestation forfaitaire rencontrée en assurance vie, par exemple dans le cas de rentes dont le montant est fixé à l'avance contractuellement.

1.1.2 Nécessité des provisions

Une particularité cruciale de l'assurance non-vie est l'inversion du cycle de production. Celle-ci correspond à une différence, par rapport aux autres secteurs d'activité, dans la temporalité des flux échangés entre l'assureur et les assurés. Dans la majorité des entreprises, les coûts de production sont connus avant la vente du produit. Prenons l'exemple d'un menuisier : il achète le bois, puis le travaille pour obtenir un produit. A ce stade, il connaît donc son coût de revient puisqu'il a acheté la matière première, payé la main d'œuvre et stocké la marchandise. Il peut alors fixer le prix de vente afin de dégager une plus-value. Dans le cas de l'assurance, pour établir un parallèle, le coût de production correspond aux indemnités qui seront éventuellement versées par l'assureur tandis que le prix de vente est la prime encaissée. Or, la couverture de l'assuré commence au règlement de la prime mais les sinistres sont postérieurs et l'assureur ne connaît pas à ce moment les coûts qui en découleront. Ils doivent donc être estimés finement pour ajuster les tarifs et permettre une plus-value. C'est une problématique de tarification que nous n'aborderons pas. En revanche, il en découle aussi la nécessité de constituer des réserves avec les primes perçues afin de faire face aux sinistres futurs. Ces réserves nécessaires pour indemniser les contrats sur lesquels l'assureur est engagé sont appelées provisions techniques. Les primes sont investies dans des actifs sécurisés et enregistrés à l'actif du bilan pour correspondre aux provisions techniques au passif. L'estimation des provisions est donc fondamentale pour l'assureur, et généralement confiée à une équipe d'actuaire dédiée.

1.1.3 Les types de provisions

Les provisions sont séparées en catégorie selon les définitions de l’Autorité des Normes Comptables. Nous présentons ici les principales provisions pour un assureur non-vie définies par le code des assurances à l’article R343-7 [*Code des assurances 2021*] Deux groupes se distinguent : les provisions de primes et les provisions de sinistres.

Les provisions de primes

- **Provision pour prime non acquise.** L’assuré paie une prime pour être couvert pendant une certaine période. Cette période ne coïncide pas avec un exercice comptable (généralement du 1 janvier au 31 décembre) et l’assureur doit amortir la prime sur la période de couverture. Cet amortissement est fait au *pro rata temporis*. Ainsi, la provision correspond au montant de prime non acquis pendant l’exercice comptable actuel destiné à couvrir les sinistres de l’exercice comptable suivant.
- **Provision pour risques en cours.** Si le coût d’un contrat a été sous-estimé par l’assureur, la prime perçue est insuffisante. Elle sera revue à l’échéance suivante, mais entre temps l’assureur constitue une provision pour couvrir les prestations indemnitaires.

Les provisions de sinistres

- **Provision pour égalisation** elle doit permettre à l’assureur de faire face à des charges exceptionnelles liées par exemple aux catastrophes naturelles, à la responsabilité civile due à la pollution, au terrorisme...
- **Provision pour sinistres à payer (PSAP)** : c’est la provision principale en assurance non-vie. Elle sert à faire face aux engagements liés aux sinistres survenus mais pas encore entièrement payés. Elle permet de rattacher les paiements estimés à l’exercice comptable de survenance du sinistre. Elle a trois composantes :
 1. **Provision dossier-dossier.** Lorsque l’assureur apprend la survenance d’un sinistre, il estime son coût à partir des informations transmises par l’assuré et éventuellement de l’avis d’un expert envoyé pour évaluer la situation. La provision dossier-dossier sert à couvrir ce coût.
 2. **Incurred but not enough reported (IBNER).** Le coût du sinistre peut évoluer par rapport à l’estimation initiale, à la hausse ou à la baisse suivant que l’estimation dossier-dossier était trop peu ou trop prudente. La provision IBNER peut donc être positive ou négative. Un exemple d’augmentation de charge est le cas d’un accident de voiture où une victime semblait indemne mais doit être opérée quelques mois plus tard.
 3. **Incurred But Not Yet Reported (IBNYR).** Cette provision correspond aux sinistres tardifs, qui sont notifiés à l’assureur après la clôture de l’exercice comptable. C’est le cas par exemple lorsqu’une maison de vacances est sinistrée en automne et que le propriétaire le remarque seulement en y retournant l’été suivant.

Les IBNER et IBNYR sont usuellement regroupées sous le nom de IBNR. Dans la suite, nous nous concentrerons sur les provisions pour sinistres à payer. Nous illustrons dans la figure 1 la manière dont les PSAP se décomposent avec la notion de charge ultime. La charge ultime correspond au montant payé à la clôture du sinistre, une fois tous les versements effectués.

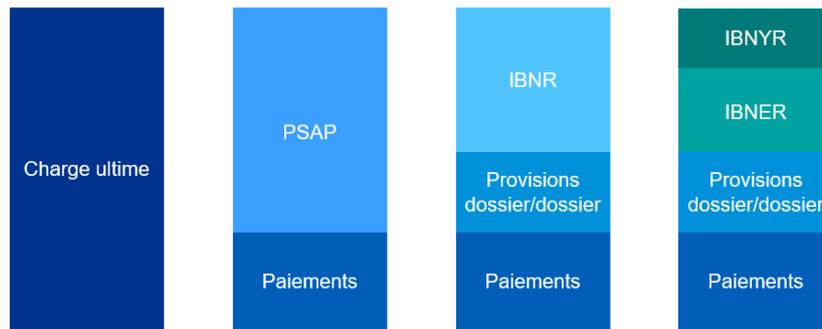


FIGURE 1 – Décomposition de la charge ultime

1.1.4 Le cycle de vie des sinistres

Nous avons mentionné la temporalité propre à l'assurance. Détaillons maintenant les étapes du cycle de vie d'un sinistre, ces notions étant nécessaires pour réfléchir à la modélisation des réserves. Elles sont placées chronologiquement sur la figure 2.

1. **Survenance.** C'est le moment où le sinistre a lieu. Elle doit se produire pendant la période de couverture
2. **Déclaration à l'assureur.** Si elle a lieu après après la clôture comptable de l'année d'exercice de survenance, le sinistre est dit « tardif ».
3. **Estimation du montant d'indemnisation.** Elle peut être revue ultérieurement.
4. **Paielement de l'indemnité à l'assuré.** Il peut s'effectuer en plusieurs fois, notamment si l'indemnité est réévaluée.
5. **Recours éventuels.** Si un tiers responsable est identifié, l'assureur peut être remboursé de l'indemnisation versée au bénéficiaire par le tiers ou son assureur.
6. **Clôture du sinistre** une fois l'indemnisation terminée.
7. **Réouverture du sinistre éventuelle.** Par exemple en cas d'aggravation.
8. **Nouveaux paielements et recours.**
9. **Clôture définitive.**

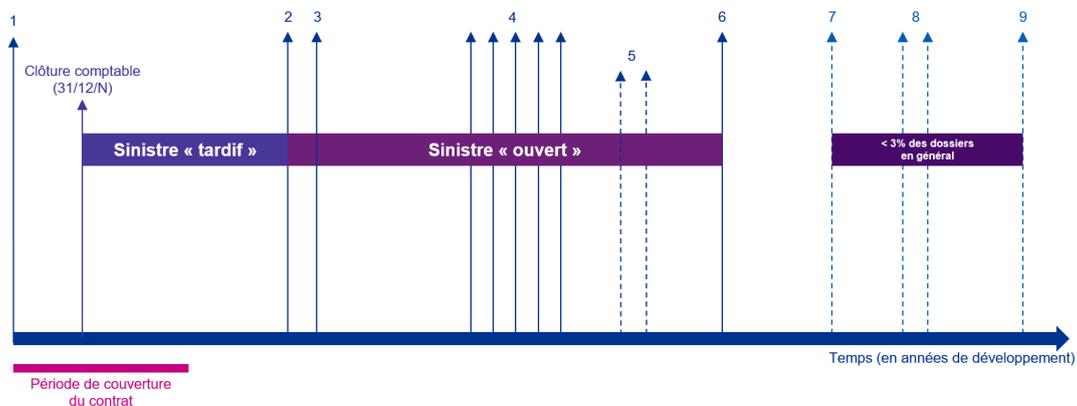


FIGURE 2 – Cycle de vie d'un sinistre

1.1.5 Équilibre entre mutualisation et segmentation

Un principe fondateur de l'assurance est la mutualisation. Les individus soumis à des risques identiques mais indépendants se regroupent et acceptent de payer pour les sinistres qui ne touchent que certains d'entre eux. Pour l'assureur, l'augmentation de la taille d'un groupe homogène conduit à une diminution de la variance des montants de sinistres. Cependant, la disponibilité croissante d'informations sur les assurés ouvre la porte à une pratique inverse : la segmentation. Elle consiste à créer des tarifs différenciés en fonction des caractéristiques de l'assuré. En théorie [CHARPENTIER, DENUIT et ELIE 2015], si un assureur segmente alors qu'un autre applique un tarif mutualisé, les profils moins risqués seront attirés par le tarif segmenté qui sera moins élevé pour eux. L'assureur qui mutualise subira alors de l'antisélection puisque les profils plus risqués se dirigeront chez lui et créeront des pertes. La concurrence pousse les assureurs à rechercher les profils moins risqués. Une conséquence logique serait que, pour éviter l'antisélection, les assureurs soient pris dans une « spirale de segmentation ». A long terme, le principe de mutualisation indispensable à l'assurance serait donc menacé. Toutefois, la segmentation a plusieurs inconvénients. D'abord, elle augmente la complexité des modèles, ce qui a un coût en termes de risque opérationnel. Mais surtout, elle conduit à avoir des groupes d'assurés de taille réduite, faisant fortement augmenter la volatilité. Or, pour garantir la solvabilité, cette variance doit être couverte par une augmentation du tarif. De plus, la part de marché occupée est réduite.

La segmentation existe donc au niveau tarifaire, mais aussi au niveau de la gestion des sinistres et du provisionnement. Citons la distinction entre les sinistres attritionnels, c'est-à-dire de coût faible mais de fréquence élevée, et les sinistres graves qui sont le cas inverse. Ils sont généralement répartis entre deux équipes de gestion distinctes, notamment parce qu'un sinistre grave peut nécessiter des évaluations plus longues et plus d'expertises. Pour le provisionnement, dans une branche donnée, il est possible de séparer les provisions (donc les triangles, notion que nous détaillerons ultérieurement) par garantie ou par produit pour une plus grande homogénéité. Les normes qui s'appliquent aux assureurs, détaillées juste après cette partie, requièrent aussi une segmentation. La norme Solvabilité 2 exige que le *Best Estimate* des provisions soit établi par groupe de risque homogène, c'est-à-dire pour des contrats dont les sinistres se développent de manière similaire. En norme IFRS 17, tous les postes du passif sauf les fonds propres sont calculés par groupes homogènes de contrats.

Ainsi, la segmentation en plusieurs classes est un procédé naturel dans plusieurs aspects de l'assurance, s'appuyant sur des informations pour former des groupes de risques plus homogènes.

1.2 Les aspects réglementaires

L'activité d'assurance, par nature aléatoire, est économiquement risquée. D'une part, le poids de l'assurance dans le domaine financier est considérable, et une faillite aurait de lourdes répercussions. D'autre part, l'assureur est engagé contractuellement envers ses clients. Il doit constituer des réserves pour faire face aux paiements qu'il estime avoir à régler dans le futur afin d'honorer ses engagements. Pour garantir un comportement prudent de la part des assureurs, le secteur est réglementé à différents niveaux. Ainsi,

un certain nombre de normes dictent des conditions pour la constitution des réserves. La norme française, établie par l'autorité des normes comptables, exige simplement une estimation « prudente ». Il suffit donc à l'assureur d'ajouter à l'espérance des réserves décrites précédemment un montant qu'il estime prudent. Cependant, deux autres normes exigent une plus grande finesse qui impose des contraintes au calcul des provisions. Nous les présentons brièvement.

1.2.1 La norme Solvabilité 2

La directive solvabilité 2 [CONSEIL DE L'UNIONS EUROPÉENNE 2009], qui s'applique depuis le 1er janvier 2016 à tous les assureurs exerçant dans l'Union Européenne, a pour objectif une meilleure protection des assurés. Il s'agit notamment de mieux limiter les risques de faillite de l'assureur en remédiant à des insuffisances de la norme Solvabilité 1. Un autre but est l'harmonisation de normes au niveau européen. Elle s'articule autour de trois piliers :

1. Exigences de quantification des risques.
2. Exigences qualitatives de gestion des risques et de gouvernance.
3. Exigences de communication financière.

Nous nous intéressons au premier pilier. Il impose aux assureurs un niveau de fonds propres appelé SCR (*Solvency Capital requirement*) permettant de faire face à une ruine à horizon 1 an dans 99,5% des cas. Il s'agit donc d'une faillite survenant au plus dans 1 cas sur 200. Le SCR peut être calculé de deux manières : par la formule standard fournie par l'Autorité européenne des assurances et des pensions professionnelles (EIOPA) ou par un modèle propre à l'assureur.

La formule standard propose une approche modulaire, c'est-à-dire qu'un SCR doit être calculé pour chaque module de risque auquel l'assureur est exposé. La cartographie complète est présentée en figure 3. Ces SCR sont ensuite agrégés selon une méthode définie, reposant sur des coefficients de corrélation fournis. Les réserves constituent un module pour l'assurance non-vie, ce qui implique la nécessité de calculer un SCR, donc de disposer d'une distribution permettant d'obtenir un quantile 99,5%.

L'intérêt principal de cette méthode est sa simplicité puisqu'elle est directement applicable et que tous les paramètres sont fournis. Cependant, son caractère générique et adaptable à tous les assureurs indépendamment du profil de risque fait perdre en précision. Ainsi, un assureur appliquant la formule standard sera amené à constituer des provisions potentiellement plus élevées, donc à immobiliser plus de capital, que s'il tenait compte des spécificités de son portefeuille et choisissait des agrégations différentes. Pour y remédier, Solvabilité 2 laisse la possibilité aux assureurs d'utiliser un modèle interne. Le développement d'un tel modèle engendre des coûts élevés liés à la recherche et développement puis à la validation. De plus, le régulateur doit approuver le modèle et demande pour cela une justification de tous les choix et paramètres retenus. Le coût de mise en place d'un modèle interne est donc très élevé. Pour cette raison, peu d'acteurs y ont recours.

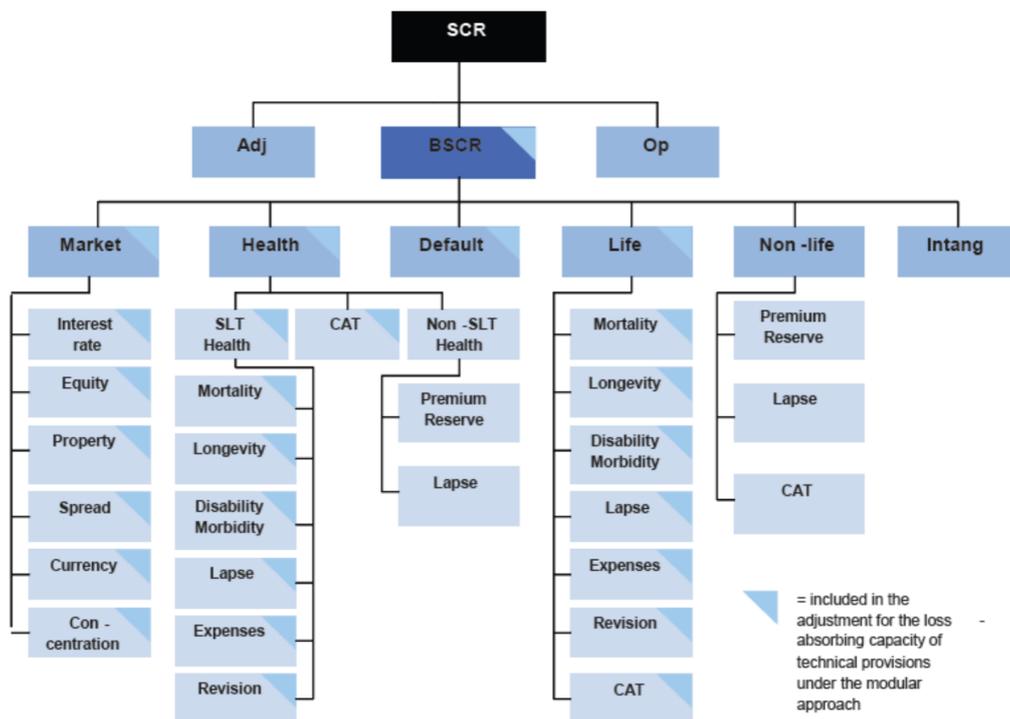


FIGURE 3 – Décomposition par module du calcul du SCR

Outre le SCR et les provisions, le passif en vision Solvabilité 2 comprend deux autres composantes :

- La marge pour risque : elle correspond au montant que demanderait un repreneur du portefeuille pour honorer les engagements déjà pris.
- Le capital excédentaire : c'est le capital disponible en plus du SCR.

Nous mentionnons ces composantes afin de décrire le bilan, schématisé en figure 4 mais elle n'ont pas d'impact sur l'estimation des provisions.

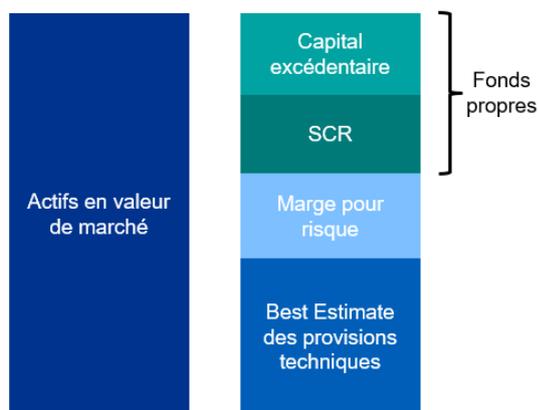


FIGURE 4 – Bilan en vision Solvabilité 2

1.2.2 La norme IFRS 17

La norme IFRS 17, publiée par l'International Accounting Standards Board (IASB), doit s'appliquer au 1er janvier 2023. Elle concerne les compagnies d'assurance qui appliquent les normes IFRS, ce qui inclut les sociétés cotées ou émettrices de dette dans l'Union Européenne. Son objectif est d'harmoniser les règles de comptabilité au niveau international, en remplaçant la précédente norme IFRS 4 qui s'appuyait sur des normes nationales. Au niveau du bilan, les actifs sont toujours évalués en valeur de marché mais les composantes du passif sont différentes et comportent :

- La valeur actuelle des flux de trésorerie futurs. Elle correspond à un *Best Estimate* des engagements de l'assureur. C'est le pendant du *Best Estimate* des provisions techniques de Solvabilité 2, qui sont des flux comptables.
- L'ajustement pour risque. C'est une marge d'erreur pour l'estimation des flux futurs. Elle doit être précise et non prudente.
- La marge de service contractuelle représente les profits futurs provenant de contrats déjà signés.
- Les fonds propres. C'est le capital investi dans des actifs non risqués.

L'ajustement pour risque remplace dans le bilan la marge pour risque de Solvabilité 2. Il est calculé de telle sorte que la somme de cet ajustement et des flux de trésorerie corresponde à un quantile de la distribution de charge ultime. Ce quantile n'est pas imposé mais doit être prudent et annoncé par l'assureur. Le bilan a alors la structure présentée en figure 7.

Finalement, compte tenu des exigences des deux normes présentées, il est fondamental de pouvoir non seulement estimer les charges ultimes, mais aussi de disposer d'une distribution de celles-ci afin de calculer des quantiles.



FIGURE 5 – Bilan en vision IFRS 17

1.3 Méthodes de provisionnement classiques

Le provisionnement constituant un défi majeur de l'assurance, les actuaires ont développé une batterie de méthodes pour le réaliser. Nous introduisons les plus classiques afin de présenter leurs avantages et inconvénients, qui permettront de mieux cerner l'intérêt du modèle individuel.

1.3.1 Représentation en triangle de développement

La majorité des méthodes de provisionnement reposent sur des agrégations de sinistres. Elles sont faites à partir de deux variables : l'année de survenance et l'année de développement. Cette dernière correspond au nombre d'années écoulées depuis la survenance. Ainsi, un sinistre survenu en 2018 et toujours ouvert en 2021 aura connu 4 développements. Ceci permet de former un triangle dont chaque case est l'agrégation des sinistres pour un couple d'année de survenance et d'année de développement. Nous notons $i \in \{0, 1, \dots, n\}$ l'année de survenance, $j \in \{0, 1, \dots, n\}$ l'année de développement et $C_{i,j}$ la charge cumulée pour en année j pour les sinistres survenus l'année i . Le triangle a alors l'allure de la table 1 :

Année de survenance	Année de développement				
	0	1	...	$n - 1$	n
0	$C_{0,0}$	$C_{0,1}$...	$C_{0,n-1}$	$C_{0,n}$
1	$C_{1,0}$	$C_{1,1}$...	$C_{1,n-1}$	
...		
$n - 1$	$C_{n-1,0}$	$C_{n-1,1}$			
n	$C_{n,0}$				

TABLE 1 – Triangle de développement sur $n + 1$ années de survenance et $n + 1$ années de développement

Le triangle est construit à partir de toute l'information connue en année n . La partie inférieure correspond aux développements futurs. Tout l'enjeu est de la compléter pour obtenir un montant de charge ultime et donc de provisions.

Une simplification importante sous-tend les méthodes qui suivent : nous supposons que tous les sinistres sont totalement développés au plus tard après n années. Ainsi, la charge ultime pour une année de survenance i sera dans le triangle le montant $C_{i,n}$. Notons que le triangle peut être constitué des paiements au lieu des charges. Par principe ils doivent être égaux à l'ultime.

1.3.2 Les méthodes déterministes

Chain-Ladder

C'est la méthode la plus connue et probablement la plus utilisée. Elle a l'avantage d'être simple et de demander peu d'hypothèses. Elle suppose simplement que les sinistres de toutes les années se développent suivant le même schéma. Alors, des facteurs de développement sont calculés pour passer d'un développement au suivant. Ces facteurs sont définis par :

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

Et l'hypothèse se formule comme suit :

Hypothèse : Les facteurs de développement sont indépendants de l'année de survenance.

Ainsi :

$$\forall i \in \{0, 1, \dots, n\}, \forall j \in \{0, 1, \dots, n-1\}, f_{i,j} = f_j = \frac{C_{0,j+1}}{C_{0,j}} = \dots = \frac{C_{n,j+1}}{C_{n,j}}$$

Il est alors possible d'estimer le facteur de développement j en s'appuyant sur toutes les années qui ont connu un tel développement pour $j \in \{0, 1, \dots, n-1\}$

$$\widehat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}$$

La charge cumulée pour un développement futur j issue d'une survenance i s'estime comme suit, pour $i \in \{0, 1, \dots, n\}$ et $j \in \{n-i, \dots, n\}$

$$\widehat{C}_{i,j} = \prod_{k=n-i}^{j-1} \widehat{f}_k \times C_{i,n-i}$$

Ceci permet de compléter le triangle :

Année de survenance	Année de développement				
	0	1	...	$n-1$	n
0	$C_{0,0}$	$C_{0,1}$...	$C_{0,n-1}$	$\widehat{C}_{0,n}$
1	$C_{1,0}$	$C_{1,1}$...	$C_{1,n-1}$	$\widehat{C}_{1,n}$
...
$n-1$	$C_{n-1,0}$	$C_{n-1,1}$...	$\widehat{C}_{n-1,n-1}$	$\widehat{C}_{n-1,n}$
n	$C_{n,0}$	$\widehat{C}_{n,1}$...	$\widehat{C}_{n,n-1}$	$\widehat{C}_{n,n}$

TABLE 2 – Triangle de développement complété

Les réserves se déduisent ensuite à partir des charges ultimes estimées. En notant R_i la réserve correspondant à l'année de survenance i :

$$\widehat{R}_i = \widehat{C}_{i,n} - C_{i,n-i}$$

Enfin, la provision pour sinistres à payer s'estime comme la somme des réserves par année :

$$PSAP_{ultime} = \sum_{i=1}^n \widehat{R}_i$$

La simplicité de la méthode de Chain-Ladder explique sa popularité dans les compagnies encore aujourd'hui. Cependant, elle souffre d'un certain nombre de limites.

- L'unique hypothèse d'indépendance entre les facteurs de développement et l'année de survenance implique une stabilité du portefeuille d'une année sur l'autre. Or, plusieurs facteurs peuvent invalider cette hypothèse : une rupture de tendance dans le développement des sinistres due notamment à un changement dans le mode de gestion, une volatilité causée par des sinistres graves ou un historique insuffisant.

- Les derniers facteurs de développement sont calculés à partir d'une petite quantité de données, ce qui nuit à leur fiabilité. De plus, ces facteurs affectent les prédictions de charges pour de nombreuses années de survénance, entraînant un réel risque d'instabilité.
- La méthode étant déterministe, elle ne permet pas d'obtenir un intervalle de confiance autour du résultat.
- Le résultat fourni est unique, or les réglementations requièrent un quantile calculé à partir d'une distribution.

Ainsi, la qualité des données est primordiale. Une analyse en amont devra notamment exclure les valeurs aberrantes liées à un sinistre exceptionnel ou une erreur de saisie, qui biaiserait le modèle. Malgré la simplicité du modèle, un jugement d'expert est donc nécessaire pour son application.

Des variations du modèle de Chain-Ladder ont été proposées au fil du temps, dont la méthode de London Chain [BENJAMIN et EAGLES 1986] que nous évoquons brièvement. Elle propose une relation affine entre les charges cumulées, c'est-à-dire :

$$\forall i \in \{0, 1, \dots, n\}, \forall j \in \{0, 1, \dots, n-1\}, C_{i,j+1} = f_j C_{i,j} + a_j$$

Les coefficients f_j et a_j peuvent être estimés par une régression linéaire, puis conduisent au développement du triangle. Ce modèle est pertinent si l'on suppose que le lien entre les charges cumulées n'est pas proportionnel. Cette hypothèse est vérifiable graphiquement en traçant les points $(C_{i,j}, C_{i,j+1})$ pour $i \in \llbracket 0, n-j+1 \rrbracket$ qui doivent s'aligner sur une droite ne passant pas par l'origine.

Bornhuetter-Ferguson

La méthode de Bornhuetter-Ferguson [BORNHUETTER et FERGUSON 1972] est une alternative répandue à la méthode de Chain-Ladder qui s'appuie aussi sur les triangles de développement. En revanche, elle requiert une estimation externe de la charge ultime, qui sera notée α_i . Au lieu des facteurs de développement, les cadences sont utilisées. Elles représentent la proportion de la charge ultime payée à une année de développement. Formellement, la cadence pour une année de survénance i au développement j , notée $\gamma_{i,j}$, est définie par :

$$\forall (i, j) \in \llbracket 0, n \rrbracket^2, \gamma_{i,j} = \frac{C_{i,j}}{C_{i,n}}$$

Elle vérifie une hypothèse similaire à celle de Chain-Ladder :

Hypothèse : Les cadences de développement sont indépendantes de l'année de survénance.

Il en découle :

$$\forall (i, j) \in \llbracket 0, n \rrbracket^2, \gamma_{i,j} = \gamma_j = \frac{C_{i,j}}{C_{i,n}}$$

Pour l'estimation de ces coefficients, plusieurs méthodes existent mais la plus courante s'appuie sur les estimateurs de Chain-Ladder :

$$\forall j \in \llbracket 0, n-1 \rrbracket, \hat{\gamma}_j = \prod_{k=j}^{n-1} \frac{1}{\hat{f}_k}$$

A partir des estimations de charges ultimes $\hat{\alpha}_i$ et de la dernière diagonale du triangle, il est possible de compléter les charges :

$$\forall i \in \llbracket 1, n \rrbracket, \forall j \in \llbracket n + 1 - i, n \rrbracket, \widehat{C}_{i,j} = C_{i,i} + \hat{\alpha}_i \times (\hat{\gamma}_j - \hat{\gamma}_i)$$

Un avantage de cette méthode par rapport à Chain-Ladder est la moindre dépendance aux premières charges des estimations pour les années de développement élevées, ainsi qu'une meilleure stabilité. L'avantage principal (simplicité et rapidité) et les autres inconvénients (pas de distribution ni d'estimation de l'erreur, hypothèse d'indépendance) sont communs aux deux méthodes. Cette méthode nécessite de plus de disposer d'une estimation des charges ultimes. Mentionnons brièvement un moyen de les obtenir. Le volume des primes pour l'année i , P_i est connu. Si l'on évalue les loss ratios LR_i attendus, par exemple à partir des données historiques, la charge ultime peut se déduire simplement :

$$\forall i \in \llbracket 1, n \rrbracket, \hat{\alpha}_i = LR_i \times P_i$$

1.3.3 Les méthodes stochastiques

Ces méthodes considèrent les observations $C_{i,j}$ comme des réalisations de variables aléatoires, ce qui permet d'obtenir plus de résultats.

Le modèle de Mack

Proposé par Mack [MACK 1993], ce modèle est une version stochastique de la méthode de Chain-Ladder. Ainsi, outre un montant de PSAP, il fournit une estimation de l'erreur quadratique moyenne de prédiction. L'hypothèse d'indépendance de Chain-Ladder fait place aux trois hypothèses suivantes :

Hypothèse 1 : Les années de survenance sont indépendantes entre elles. Pour deux années de survenance i et i' , les vecteurs $(C_{i,0}, \dots, C_{i,n})$ et $(C_{i',0}, \dots, C_{i',n})$ sont indépendants.

Hypothèse 2 : L'espérance de la charge sachant l'information de tous les développements passés dépend uniquement du développement précédent. Formellement :

$$\forall i \in \llbracket 0, n \rrbracket, \forall j \in \llbracket 0, n - 1 \rrbracket, \exists f_j \in \mathbb{R}, \mathbb{E}[C_{i,j+1} | C_{i,0}, \dots, C_{i,j}] = f_j C_{i,j}$$

Hypothèse 3 : La variance de la charge sachant l'information de tous les développements passés dépend uniquement du développement précédent. Formellement :

$$\forall i \in \llbracket 0, n \rrbracket, \forall j \in \llbracket 0, n - 1 \rrbracket, \exists \sigma_j \in \mathbb{R}^+, \mathbb{V}[C_{i,j+1} | C_{i,0}, \dots, C_{i,j}] = \sigma_j^2 C_{i,j}$$

S'appuyant sur ces hypothèses, Mack prouve que les estimateurs de Chain-Ladder \hat{f}_j sont des estimateurs sans biais des f_j et qu'ils sont de plus décorrélés. Formellement :

$$\forall j \in \llbracket 0, n - 1 \rrbracket, \mathbb{E}[\hat{f}_j] = f_j$$

$$\forall (j, k) \in \llbracket 0, n - 1 \rrbracket^2, j \neq k \Rightarrow \text{Cov}(\hat{f}_j, \hat{f}_k) = 0$$

Il en découle un estimateur non biaisé de la charge ultime $\mathbb{E}[C_{i,j+1} | C_{i,0}, \dots, C_{i,j}]$ avec la formule de Chain-Ladder :

$$\forall i \in \llbracket 0, n \rrbracket, \forall j \in \llbracket n + 1 - j, n \rrbracket, \widehat{C}_{i,n} = \prod_{k=n-i}^{j-1} \hat{f}_k \times C_{i,n-i}$$

Une possibilité pour prouver l'absence de biais est d'utiliser la non corrélation des facteurs de développement. Finalement, la réserve pour la survenance i est toujours :

$$\widehat{R}_i = \widehat{C}_{i,n} - C_{i,n-i}$$

A présent, considérons les erreurs de prédiction. Mack propose un estimateur sans biais des σ_j^2

$$\forall j \in \llbracket 0, n-2 \rrbracket, \widehat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_{i,j} \right)$$

$$\widehat{\sigma}_{n-1}^2 = \min \left(\frac{\widehat{\sigma}_{n-2}^4}{\widehat{\sigma}_{n-3}^2}, \min(\widehat{\sigma}_{n-3}^2, \widehat{\sigma}_{n-2}^2) \right)$$

ou $\widehat{\sigma}_j^2 = 0$ si $\widehat{f}_j = 0$. L'apport principal de cette méthode est une quantification de l'erreur de prédiction par l'erreur quadratique moyenne (MSEP, mean square error prediction) sachant l'information disponible que nous notons D . Alors :

$$MSEP(\widehat{R}_i) = \mathbb{E}[(\widehat{R}_i - R_i)^2 | D] = \mathbb{E}[(\widehat{C}_{i,n} - C_{i,n})^2 | D] = \mathbb{V}[\widehat{C}_{i,n} | D] + (\mathbb{E}[\widehat{C}_{i,n} | D] - \widehat{C}_{i,n})^2$$

Ce résultat fait apparaître une décomposition de l'erreur quadratique en deux termes. Le premier est purement stochastique. Le second correspond à l'erreur d'estimation des paramètres f_j par les \widehat{f}_j . A partir de cette formule et des estimateurs de σ_j^2 , Mack fournit un estimateur de l'erreur quadratique de prédiction pour la survenance i :

$$\forall i \in \llbracket 1, n \rrbracket, \widehat{MSEP}(\widehat{R}_i) = \widehat{C}_{i,n}^2 \sum_{j=n-i}^{n-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2} \left(\frac{1}{\widehat{C}_{i,j}} + \frac{1}{\sum_{k=0}^{n-j-1} C_{k,j}} \right)$$

Ceci conduit à l'estimateur suivant pour les réserves totales :

$$MSEP(\widehat{R}_{ultime}) = \sum_{i=1}^n \left(MSEP(\widehat{R}_i) + \widehat{C}_{i,n} \sum_{k=i+1}^n \widehat{C}_{k,n} \sum_{l=n-i}^{n-1} \frac{2\widehat{\sigma}_l^2}{\widehat{f}_l^2 \sum_{m=0}^{n-l-1} C_{m,l}} \right)$$

Ce modèle stochastique a l'avantage principal de fournir une mesure de l'erreur de prédiction. Cependant, la distribution des réserves reste inconnue, à moins de supposer que celles-ci suivent une certaine loi et d'en estimer les paramètres avec la méthode des moments. De plus, il est toujours nécessaire de vérifier si les données respectent les hypothèses du modèle.

Ré-échantillonnage Bootstrap

Le modèle de Mack apporte une mesure de la variance, mais la distribution complète reste inconnue, empêchant notamment le calcul d'un quantile. La méthode bootstrap consiste à répéter un grand nombre G de fois un échantillonnage et une simulation afin d'obtenir une distribution [ENGLAND et VERRALL 1999]. Le triangle utilisé cette fois n'est plus constitué de montants cumulés mais de montants incrémentaux. Ces montants, notés $I_{i,j}$ sont liés aux montants cumulés par la relation : $I_{i,j} = C_{i,j} - C_{i,j-1}$, et $I_{i,0} = C_{i,0}$. La méthode classique s'appuie sur les résidus de Pearson définis par :

$$r_{i,j} = \frac{I_{i,j} - \mathbb{E}[\widehat{I}_{i,j}]}{\sqrt{\mathbb{V}[\widehat{I}_{i,j}]}}$$

L'utilisation de montants incrémentaux permet de poser l'hypothèse suivante sur les résidus :

Hypothèse : Les résidus sont des réalisations de variables aléatoires indépendantes et identiquement distribuées.

La méthode bootstrap se déroule selon le schéma suivant :

1. Calculer à partir du modèle choisi les résidus dans le triangle supérieur.
2. Répéter G fois les étapes suivantes :
 - Ré-échantillonner avec remise les résidus dans le triangle supérieur afin d'obtenir un nouveau triangle de résidus $r'_{i,j}$
 - En inversant la définition des résidus, obtenir un nouveau triangle de montants incrémentaux $I'_{i,j}$
 - Cumuler le triangle précédent et le développer pour obtenir une estimation des réserves
3. A partir des G montants de réserve, estimer une distribution.

Parmi les méthodes présentées, seule celle-ci fournit une distribution complète, ce qui constitue son intérêt, notamment pour des questions réglementaires. Cependant, elle a comme limite la nécessité de faire une hypothèse de loi sur les montants incrémentaux. Un choix courant est la loi de Poisson, qui conduit aux mêmes résultats que Chain-Ladder.

1.4 Bilan et intérêt du provisionnement individuel

Encore aujourd'hui, les méthodes de provisionnement agrégées prédominent largement dans les compagnies d'assurance. Cette hégémonie s'explique par leurs nombreux avantages. Elles ont fait leurs preuves, sont connues par des générations d'actuaire et sont largement documentées. Elles sont simples, à la fois au niveau de la mise en place, du fait de la faible quantité de données requise, et au niveau de la puissance de calcul utilisée. Ce dernier argument, historique, n'est plus d'actualité.

En effet, les capacités des ordinateurs et le nombre de variables disponibles pour chaque contrat ne sont plus du même ordre. Ceci fait ressortir un inconvénient majeur de ces méthodes qui est la perte d'information, dont la plupart sont pourtant systématiquement collectées. De plus, calculer une valeur unique pour un groupe plutôt qu'une multitude de valeurs individuelles réduit le nombre de points disponibles pour un calcul d'incertitude ou de quantile, requis par les réglementations récentes. Ces réglementations exigent aussi une segmentation pour le calcul des réserves. Elle est possible avec des méthodes agrégées mais complexe. Enfin, ajoutons que la réassurance non proportionnelle n'est pas compatible avec une agrégation puisqu'elle fonctionne à l'échelle du sinistre.

Face à ces limites et profitant de l'amélioration des données, la recherche actuarielle s'est penchée sur le sujet du provisionnement individuel. Norberg [NORBERG 1993] a proposé puis raffiné [NORBERG 1999] une approche où le développement des sinistres suit un processus de Poisson marqué, inspirant de nombreux modèles ultérieurs. Plus récemment, grâce à l'augmentation de la puissance disponible, des méthodes de *machine learning* ont été développées ou adaptées pour le provisionnement, avec des approches variées : arbres de régression [BAUDRY et ROBERT 2019], [LOPEZ, MILHAUD et THÉRON 2019], boosting [DUVAL et PIGEON 2019], réseau de neurones profond [KUO 2019]. L'objectif dans ces cas n'est pas de calculer directement une réserve globale, mais de d'obtenir une réserve

individuelle pour chaque sinistre.

En majorité, les mémoires d'actuariat s'intéressant à ce sujet évaluent des réserves pour les sinistres déjà ouverts, mais laissent de côté les IBNYR. Riffaud [RIFFAUD 2019] a proposé une méthode pour traiter cet aspect. Son intégration à des modèles individuels pour prédire les IBNR donne de bons résultats [PRIMEL 2021]. Cependant, ces modèles ne prennent pas en compte toute l'information disponible sur les sinistres. Ainsi, une méthode utilisée développe les sinistres individuellement, mais les développe tous avec la même loi. Les intérêts de la segmentation dans l'assurance ont déjà été mentionnés. Ici, utiliser d'autres variables pour classifier les sinistres *a priori* permet de constituer des groupes homogènes avec des caractéristiques de développement distinctes et d'améliorer les prédictions.

2 Présentation des modèles

Dans cette partie, les modèles servant au développement individuel des sinistres sont présentés. Nous détaillons d'abord la théorie à laquelle ils s'adossent puis la manière dont elle se traduit en pratique pour la mise en place. Après cela, nous expliquerons les algorithmes de classification.

2.1 Génération des sinistres tardifs

Les sinistres tardifs sont par définition absents de la base, bien qu'ils soient déjà survenus. Notre objectif étant d'aboutir à des provisions complètes, nous devons simuler ces sinistres. Nous reprenons la méthode de Riffaud [RIFFAUD 2019]. Il s'agit de simuler un nombre de sinistres tardifs pour chaque année de survenance. Ensuite, ils seront développés avec le même modèle que les autres sinistres. Ceci correspond à une approche fréquence-sévérité, courante en actuariat.

La simulation de la fréquence s'appuie sur un triangle de nombres de sinistres, similaire aux triangles de charges présentés précédemment. Dans les travaux évoqués, plusieurs méthodes sont mises en œuvre, dont le calibrage de lois de fréquence. Celle qui est retenue et que nous utilisons est un bootstrap non paramétrique. Dans cette variante de la méthode présentée en partie 1, les résidus sont ré-échantillonnés dans l'intégralité du tableau et non dans le triangle supérieur. Elle fournit donc G valeurs de sinistres survenus et non ouverts, représentés dans des triangles.

2.2 Hypothèses des modèles

Quelle que soit la manière dont nous le développerons, le sinistre est représenté par un quadruplet (n_i, I_i, Ch_i, D_i) où :

- $n_i \in \mathbb{N}$ est l'**année de survenance** du sinistre.
- I_i est le **processus de règlements incrémentaux**. C'est une suite de variables aléatoires $(I_{i,j})_{j \in \mathbb{N}}$ qui représente, pour le sinistre i , le paiement réalisé en année de développement j .
- Ch_i est le **processus de charges**. C'est une suite de variables aléatoires $(Ch_{i,j})_{j \in \mathbb{N}}$ qui représente, pour le sinistre i , la somme des paiements effectués et des réserves estimées par le gestionnaire de sinistres en fin d'année de développement j .
- D_i est le **processus d'état du sinistre** C'est une suite de variables aléatoires de Bernoulli $(D_{i,j})_{j \in \mathbb{N}}$ qui indique si le sinistre est toujours ouvert. Ainsi en année de développement j , $D_{i,j} = 1$ si le sinistre est ouvert et 0 s'il est clos.

2.2.1 État du sinistre et durée de vie

Le processus d'état du sinistre est naturellement lié à sa durée de vie. Cette notion est fondamentale pour simuler le développement des sinistres. Présentons la manière dont il sera estimé.

Dans la littérature, deux hypothèses sont courantes à propos du processus d'état afin de le simplifier.

Hypothèse 1 : les sinistres sont ouverts en année de développement 0. Formellement, $\mathbb{P}(D_{i,0}) = 1$.

Hypothèse 2 : il n'y a pas de réouverture. Autrement dit, l'état 0 est absorbant : $\forall j \geq 0, \forall j' \geq j, \mathbb{P}(D_{i,j'} = 1 | D_{i,j} = 0) = 0$.

La première hypothèse est naturelle par rapport à la notion d'année de développement. Il convient simplement de la poser. La seconde hypothèse apporte une nette simplification en évitant un processus à plusieurs états successifs difficile à traiter. En revanche, elle est nettement plus forte. Nous verrons toutefois que le nombre de sinistres réouverts est très faible, ce qui nous permet de conserver l'hypothèse.

2.2.2 Estimation de la durée de vie

La durée de vie du sinistre est représentée par la variable aléatoire Γ_i . Elle correspond à la dernière année de développement avant la clôture.

Mathématiquement : $\Gamma_i = \inf(j | D_{i,j+1} = 0)$

Pour l'estimation, nous allons travailler avec la fonction de survie. Elle est définie par :

$$S_{\Gamma_i}(j) = \mathbb{P}(\Gamma_i > j) = F_{\Gamma_i}(j)$$

Nous disposons de sinistres clos, dont la durée de vie est connue, et de sinistre encore ouverts, qui sont des observations censurées. Un estimateur classique de la fonction de survie dans ce cas est l'estimateur de Kaplan-Meier [KAPLAN et MEIER 1958].

En notant :

- r_j le nombre de sinistres encore ouverts en année de développement $j - 1$, donc qui peuvent fermer sur l'intervalle $]j - 1, j]$. De plus, $r_0 = 1$.
- d_j le nombre de sinistres clos au développement j qui étaient ouverts au développement $j - 1$. De plus, $d_0 = 0$

Il s'appuie sur un découpage discret du temps, dans notre cas à l'échelle de l'année. Un sinistre est encore ouvert en année j s'il était ouvert en année $j - 1$ et n'a pas été clos entre $j - 1$ et j . Formellement, avec les notations introduites :

$$\begin{aligned} \mathbb{P}(D_{i,2} = 1) &= \mathbb{P}(D_{i,2} = 1 | D_{i,1} = 1) \mathbb{P}(D_{i,1} = 1) \\ &= \mathbb{P}(D_{i,2} = 1 | D_{i,1} = 1) \mathbb{P}(D_{i,1} = 1 | D_{i,0=1}) \mathbb{P}(D_{i,0} = 1) \end{aligned}$$

L'estimateur est :

$$\widehat{S}_{\Gamma_i}(t) = \prod_{j=0}^t \left(1 - \frac{d_j}{r_j}\right)$$

Le lien avec les probabilités conditionnelles apparaît car chaque terme du produit correspond pour un sinistre à la probabilité conditionnelle de rester ouvert une année j sachant qu'il est encore ouvert l'année précédente, formellement :

$$\forall j > 0, \mathbb{P}(D_{i,j} = 1 | D_{i,j-1} = 1) = 1 - \frac{d_j}{r_j}$$

En pratique dans les simulations, la durée de vie du sinistre n'est pas déterminée *a priori*. Nous simulons les années de développement itérativement : chaque année, si le sinistre

est encore ouvert, nous procédons à un tirage aléatoire pour déterminer s'il reste ouvert. Avec U suivant une loi uniforme sur $[0, 1]$, le sinistre ouvert en $j - 1$ est clos en j si $U > \mathbb{P}(D_{i,j} = 1 | D_{i,j-1} = 1)$.

2.3 Le modèle additif

2.3.1 Présentation théorique

Dans ce modèle, nous calculons la charge ultime à partir des paiements incrémentaux réalisés chaque année. La relation qui les lie est la suivante :

$$CU_i = \sum_{j=0}^{\infty} I_{i,j}$$

L'enjeu du modèle est donc de simuler les paiements incrémentaux chaque année. Ensuite, ils seront sommés pour obtenir la charge et en déduire les provisions.

Nous pouvons lier la relation précédente aux processus introduits auparavant et inclure les informations connues, c'est-à-dire la partie de développement du sinistre qui est déjà écoulée. Nous notons $C_{0,i}$ les montants déjà réglés au cours des k développements connus du sinistre. Il est nécessaire que les sinistres aient une durée de vie finie. Ceci est assuré par l'hypothèse liée aux triangles de développement stipulant que les sinistres sont complètement développés après n années. Nous avons donc : $\forall i, \forall j > n, I_{i,j} = 0$ et $\forall i, \forall j > \Gamma_i, I_{i,j} = 0$. La charge ultime se réécrit alors :

$$CU_i = C_{i,0} + \sum_{j=k}^{\Gamma_i} I_{i,j}$$

Et la réserve estimée est :

$$\widehat{R}_i = \sum_{j=k}^{\Gamma_i} \widehat{I}_{i,j}$$

2.3.2 Pratique

Le modèle étant défini, précisons les aspects relatifs à sa mise en place. Nous avons déjà présenté l'évaluation et la simulation de la durée de vie. La valeur de Γ_i déterminée, il ne reste qu'à simuler les $\Gamma_i - k + 1$ paiements.

Pour une année de développement j , nous isolons tous les sinistres, clos ou non, qui présentent un développement j et extrayons les montants payés cette année. Puis nous tirons dans cet échantillon selon une loi uniforme pour obtenir le paiement simulé. Ainsi, nos simulations de $I_{i,j}$ suivent bien la loi empirique des données. Nous répétons ensuite cette étape jusqu'à la clôture du sinistre, ce qui nous fournit la charge ultime du sinistre. Répéter cette simulation pour tous les sinistres permet d'aboutir à la charge ultime, puis aux réserves.

2.4 Le modèle multiplicatif

2.4.1 Présentation théorique

Le second modèle est multiplicatif et s'appuie sur des facteurs de développement, comme Chain-Ladder, mais cette fois au niveau individuel et non plus agrégé. Nous n'étudions plus le processus de paiement incrémental, mais de paiement cumulé. Alternativement l'étude de la charge cumulée est possible. Le processus cumulé pour un sinistre i est noté C_i . Ainsi, $C_{i,j}$ représente la somme des paiements versés jusqu'à l'année j . L'idée d'étudier les facteurs de développement individuels a été proposée en 2006 par McLennan et Murphy [MURPHY et FIAA 2006]. Elle correspond à une version individuelle et stochastique de la méthode de Chain-Ladder. Cette fois, les facteurs de développement correspondent à un sinistre et plus à une agrégation par année de survenance. Ces facteurs sont extraits directement de la base.

Le facteur de développement j pour le sinistre i est le réel vérifiant $C_{i,j+1} = f_{i,j}C_{i,j}$ pour un processus de paiements cumulés ou de charges cumulées. Contrairement au modèle de Chain-Ladder, ces coefficients ne sont définis que jusqu'à la clôture du sinistre et pas pour les n années de développement (techniquement ils vaudraient 1 puisque le montant n'évolue plus, mais nous ne les conservons pas).

En notant $C_{0,i}$ le premier paiement réalisé, la charge ultime s'obtient avec :

$$CU_i = C_{0,i} \prod_{j=0}^{\Gamma_i-1} f_{i,j}$$

Notons que cette relation se maintient pour le processus de charges cumulées puisqu'elles sont égales aux paiements à la clôture du sinistre. La réserve associée au sinistre est similaire à celle fournie par la méthode de Chain-Ladder :

$$\widehat{R}_i = \widehat{CU}_i - C_{0,i}$$

2.4.2 Pratique

La mise en place est similaire à celle du modèle additif. Encore une fois, la durée de vie est déjà déterminée. Pour un sinistre i , tous les sinistres ayant un développement j sont isolés et les facteurs $f_{i,j}$ correspondant sont extraits. Ensuite, un de ces facteurs est sélectionné de manière équiprobable. Cette étape correspond au modèle mais a été raffinée comme nous le précisons dans la section suivante. Les facteurs sont tirés jusqu'à la clôture du sinistre, ce qui conduit à la contrepartie empirique de la formule précédente pour la charge ultime :

$$\widehat{CU}_i = C_{0,i} \prod_{j=0}^{\Gamma_i-1} \widehat{f}_{i,j}$$

Cette procédure est répétée pour chaque sinistre de manière à obtenir un montant total de charge ultime et à en tirer la PSAP. L'ensemble de la simulation est reproduit un grand nombre de fois pour aboutir à une distribution de provisions.

La mise en place de ce modèle nécessite de traiter un certain nombre de cas particuliers qui sont détaillés par [PRIMEL 2021]. Nous conservons tous ces choix. Mentionnons

à titre d'exemple le cas répandu des sinistres dont la dernière charge connue est nulle. Il est inutile d'appliquer des facteurs de développement, aussi le parti est pris de tirer aléatoirement une charge ultime dans la base et de l'affecter au sinistre.

Notons que dans ce modèle comme dans le modèle additif, les possibilités de simulations pour un sinistre sont techniquement finies puisque seuls les facteurs de développement, ou les paiements incrémentaux, présents dans la base peuvent contribuer au développement. Nous disposons d'un grand nombre de sinistres et de 20 ans d'historique, ce qui fournit un échantillon large pour les premiers développements, souvent présents. En revanche, peu de sinistres atteignent des développements longs, ce qui réduit la diversité des facteurs possibles dans ces cas.

2.4.3 Prise en compte de corrélations par une copule

Lors de sa première implémentation, le modèle multiplicatif fournissait des résultats aberrants, trop grands d'un facteur 100. Après analyse, cet effet a été imputé à la dépendance entre le montant d'une charge et le facteur de développement suivant, éléments implicitement supposés indépendants. Concrètement, les sinistres graves avec des charges de plusieurs millions ont au cours de leur évolution des facteurs proches de 1, car même un léger écart se traduit par un montant élevé. En revanche, les petits sinistres ont fréquemment des facteurs de plusieurs dizaines, voire centaines. Or, lors des simulations, les facteurs élevés de ces sinistres sont parfois attribués aux charges graves, ce qui aboutit à des montants colossaux.

Pour limiter ce phénomène, une copule a été mise en place par [PRIMEL 2021]. Nous ne détaillons pas la théorie de cet outil, mais précisons simplement que c'est une fonction qui permet, pour un couple de variables aléatoires, d'obtenir la loi jointe à partir des lois marginales. Il est courant de calibrer des copules connues sur les données afin de modéliser les dépendances, mais ici c'est une copule empirique qui est retenue. En pratique, un paramètre de largeur de copule de $2 \times l$ détermine l'intensité de la corrélation à appliquer. Ensuite, le tirage d'un facteur de développement $f_{i,j}$ à appliquer à une charge $C_{i,j}$, suit les étapes suivantes :

1. Les charges j de l'historique sont extraites et ordonnées. Un intervalle est constitué avec les l charges les plus proches de $C_{i,j}$ par défaut et les l charges les plus proches de $C_{i,j}$ par excès.
2. Les facteurs correspondant à ces charges sont extraits et le facteur $f_{i,j}$ est tiré uniformément dans cet échantillon.

Cela revient à supposer qu'un sinistre va se développer de manière similaire, à la largeur de l'intervalle près, à ceux qui ont une charge proche.

Grâce à cette correction, les résultats du modèle multiplicatif sont cohérents. Le modèle que nous utilisons par la suite a donc toujours cette copule intégrée.

2.5 Calcul de l'inflation par la méthode de Verbeek

2.5.1 Intérêt du traitement de l'inflation

L'historique de sinistres dont nous disposons couvre une période de 20 ans. Au cours de cette période, les montants réglés pour l'indemnisation ont subi une inflation, de sorte qu'un euro versé en 2000 et un euro versé en 2019 ne sont pas comparables directement. Les modèles doivent tenir compte de ce phénomène.

L'inflation est compatible avec le modèle multiplicatif. En effet, pour deux sinistres identiques survenant à des périodes différentes, la charge initiale sera différente à cause de l'inflation, mais cette donnée est connue. En revanche, l'évolution, qui est simulée en facteurs de développement, sera la même. Les facteurs correspondent à l'aggravation relative du sinistre : si le premier sinistre voit sa charge doubler la deuxième année, le deuxième aura un facteur de développement identique. Ainsi, l'utilisation des facteurs de développement individuels permet de ne pas traiter l'inflation à part.

Au contraire, avec le modèle additif, l'aggravation n'est pas en facteur mais en montant. Une opération médicale facturée 3000 € en 2002 aura un coût nettement plus élevé des années plus tard. Il est donc nécessaire de prendre en compte l'inflation. Précisément, par rapport à notre date d'observation, nous distinguons deux périodes. D'une part, il convient de d'actualiser les montants passés afin que tous les règlements de l'historique soient comparables. D'autre part, les montants à verser dans le futur doivent être actualisés puisque l'estimation des provisions est réalisée pour l'année actuelle.

Plusieurs indices d'inflation existent mais ils concernent des secteurs spécifiques, ou sont au contraire généraux. Or, notre base a une structure particulière et une inflation associée qui ne sera pas forcément reflétée par un indice extérieur. Nous choisissons donc de calculer l'inflation à partir de nos données en utilisant la méthode de Verbeek.

2.5.2 Principe de la méthode de Verbeek

Cette méthode, initialement appliquée à un nombre de sinistres [VERBEEK 1972] a été adaptée pour isoler l'inflation de montants de règlements [TAYLOR 1977]. Elle s'applique sur un triangle agrégé et vise à séparer deux effets :

1. L'effet de l'année de développement, lié aux colonnes du triangle.
2. L'effet de l'année calendaire, lié aux diagonales du triangle.

L'inflation que nous cherchons à calculer correspond au second effet. Nous faisons l'hypothèse que ces deux effets sont indépendants.

Le triangle que nous utilisons est constitué des montants incrémentaux moyens. Ils sont définis par $M_{i,j} = \frac{I_{i,j}}{n_{i,j}}$ pour $0 \leq i, j \leq n$ où $n_{i,j}$ correspond au nombre de règlements qui constituent $I_{i,j}$, c'est-à-dire le nombre de sinistres survenus en année i et encore ouverts au développement j . L'hypothèse suivante va introduire formellement les deux effets.

Hypothèse : en notant r_i un facteur dépendant uniquement de l'année de développement et λ_{i+j} un facteur dépendant de l'année calendaire, qui comporte l'inflation, nous avons la décomposition :

$$\forall i, j \in \llbracket 0, n \rrbracket, M_{i,j} = r_j \lambda_{i+j}$$

Et nous posons la contrainte d'identification $\sum_{j=0}^n r_j = 1$.
 Nous obtenons donc le triangle 3.

Année de survenance	Année de développement				
	0	1	...	$n-1$	n
0	$r_0\lambda_0$	$r_1\lambda_1$...	$r_{n-1}\lambda_{n-1}$	$r_n\lambda_n$
1	$r_0\lambda_1$	$r_1\lambda_2$...	$r_{n-1}\lambda_n$	
...		
$n-1$	$r_0\lambda_{n-1}$	$r_1\lambda_n$			
n	$r_0\lambda_n$				

TABLE 3 – Triangle de développement avec les hypothèses de la méthode de Verbeek

Notons v_j la somme des termes de la colonne j du triangle et D_k la somme des termes d'une diagonale ou sous-diagonale $\sum_{i+j=k} M_{i,j}$. A partir de ces termes connus, nous pouvons estimer itérativement les paramètres du modèle comme suit :

1. A partir de la dernière diagonale et avec la contrainte d'identification :

$$D_n = \lambda_n(r_0 + r_1 + \dots + r_n) = \lambda_n$$

Il vient un estimateur naturel de λ_n :

$$\widehat{\lambda}_n = D_n$$

Vient ensuite un estimateur de r_n à partir de la dernière colonne du triangle (ici juste une case) :

$$\widehat{r}_n = \frac{v_n}{\widehat{\lambda}_n}$$

2. Nous passons à la diagonale précédente :

$$D_{n-1} = \lambda_{n-1}(r_0 + r_1 + \dots + r_{n-1}) = \lambda_{n-1}(1 - r_n)$$

Les estimateurs de la première étape permettent de calculer λ_{n-1} :

$$\widehat{\lambda}_{n-1} = \frac{D_{n-1}}{1 - \widehat{r}_n}$$

3. Ensuite nous poursuivons de proche en proche de sorte qu'à la $k^{\text{ème}}$ diagonale les estimateurs sont :

$$\widehat{\lambda}_k = \frac{D_k}{1 - \widehat{r}_n - \widehat{r}_{n-1} - \dots - \widehat{r}_{k+1}}$$

$$\widehat{r}_k = \frac{v_k}{\widehat{\lambda}_k + \widehat{\lambda}_{k+1} - \dots - \widehat{\lambda}_n}$$

A l'issue de ces calculs, nous connaissons les λ_i qui correspondent à l'inflation passée estimée à partir de données connues. C'est la première moitié du travail, qui permet de mettre les paiements historiques en vision actuelle. Il reste ensuite à estimer l'inflation

future. Aucune méthode n'est préconisée dans la littérature, aussi nous recourons simplement à une régression linéaire pour projeter les coefficients.

Enfin, nous plaçons notre vision à la dernière année de la base, 2019, en divisant tous les λ_i par le coefficient de 2019.

Le cadre théorique étant posé, soulignons que nous disposons en réalité de plus de finesse dans notre base car les paiements sont décomposés entre les frais et le principal. Nous pouvons donc constituer deux triangles distincts et calculer deux indices d'inflation. Ainsi nous pouvons traiter l'inflation dans le modèle additif.

2.6 Modèles de classification

Les modèles présentés simulent le développement de chaque sinistre à partir d'un historique de développement. En l'absence de raffinement, cet historique est la base complète, ce qui signifie que tous les sinistres suivent les mêmes lois. Le but de notre travail est de constituer des classes afin d'effectuer les simulations sur des groupes de sinistres plus homogènes. Nous introduisons dans cette partie des algorithmes qui permettront de constituer de telles classes.

Nous pouvons distinguer deux grandes familles de méthodes en fonction de l'étiquetage des données :

- **L'apprentissage supervisé** correspond au cas où les données sont étiquetées. Les algorithmes apprennent alors à partir d'exemples pour lesquels il existe une « vraie classe ». Cela permet d'employer une large variété de méthodes, mais nécessite évidemment de disposer d'une « vraie classe » pour nos données.
- **L'apprentissage non supervisé** fonctionne lui sans étiquette et doit découvrir une structure sous-jacente dans les données. Dans notre cas, les classes *a priori* ne sont pas indiquées par une variable particulière. Une telle méthode permet donc de faire moins d'hypothèses quant aux classes à constituer.

2.6.1 La méthode des k-moyennes

La méthode des k-moyennes est un algorithme d'apprentissage non supervisé qui, à partir d'un nombre de classes k donné, constitue k groupes de manière à minimiser la variance intra-classe. Cela revient à obtenir les classes les plus homogènes possibles ou, de manière équivalente, les classes les plus distinctes les unes des autres. La variance est généralement calculée à partir de la distance euclidienne, et nous faisons ce choix ici.

L'algorithme est le suivant :

1. Initialiser en choisissant k centres de *clusters*.
2. Répéter jusqu'à convergence :
 - Affecter chaque point au *cluster* dont il est le plus proche du centre.
 - Calculer le barycentre de chaque *cluster* et le définir comme nouveau centre.

Un tel algorithme converge, mais est très sensible à l'initialisation, qui consiste ici à choisir des centres de *clusters*. Le minimum atteint est donc local. Une solution usuelle à ce problème consiste à répéter un grand nombre de fois l'algorithme avec une initialisation différente et à conserver le résultat qui aboutit à la plus faible variance intra-classe.

Puisque la méthode repose sur un calcul de distance, les variables utilisables sont uniquement quantitatives. De plus, il convient de centrer et réduire les variables afin qu'elles aient toutes un effet du même ordre sur la distance.

Un point fort de cette méthode est sa simplicité qui lui confère une grande rapidité d'exécution. Cependant, la nécessité de la lancer un grand nombre de fois contrebalance nettement ce gain. Un obstacle à une mise en place naïve est la nécessité pour l'utilisateur de déterminer le nombre de classes. Plusieurs indicateurs ont été développés à cette fin mais la méthode la plus usuelle reste de tracer la variance intra-classe en fonction du nombre de classes et de localiser un coude dans la courbe. Enfin, notons que par construction les *clusters* obtenus sont convexes ce qui rend la méthode inadaptée dans le cas de structures de données complexes.

2.6.2 Les arbres de classification

L'arbre binaire de classification est une méthode supervisée qui s'appuie sur une succession de règles de décision simples. La détermination d'un arbre optimal est un problème NP-complet, mais l'algorithme CART (*Classification And Regression Tree*) [L. BREIMAN et al. 1984] est couramment utilisé pour fournir un bon arbre. C'est celui auquel nous aurons recours. Il procède récursivement avec des décisions localement optimales : l'ensemble des données est d'abord divisé en deux sous-ensembles, les nœuds, les plus homogènes pour la variable à prédire, puis l'opération est répétée sur les sous-ensembles récursivement jusqu'à un critère d'arrêt. Lorsqu'un nœud ne se divise plus, il est appelé une feuille et la valeur la plus représentée dans les données de ce nœud lui est affectée.

Formalisons le problème. A un nœud m nous avons un ensemble E_m de N_m données étiquetées $(x_i, y_i)_{i=1, \dots, N_m}$ avec x_i composé de p variables $x_i = (x_i^1, \dots, x_i^p)$. Afin de séparer notre ensemble de données nous cherchons un critère de séparation $\theta = (j, s)$ où $j \in \{1, \dots, p\}$ indique la variable sur laquelle sera faite la séparation et s est le seuil tel que :

- Si la variable j est continue, $s \in \mathbb{R}$ et le premier sous-ensemble sera

$$E_m^{gauche}(\theta) = \{(x_i, y_i) \in E_m, x_i^j \leq s\}$$

- Si la variable j est discrète, s est un sous-ensemble des modalités de x^j et le premier sous-ensemble de la séparation sera

$$E_m^{gauche}(\theta) = \{(x_i, y_i) \in E_m, x_i^j \in s\}$$

Le sous-ensemble de droite est ensuite le complémentaire de celui de gauche :

$$E_m^{droite}(\theta) = E_m \setminus E_m^{gauche}(\theta)$$

A partir d'une fonction de perte L , le critère de séparation est celui qui minimise l'impureté :

$$\theta^* = \operatorname{argmin}_{\theta} \frac{N_m^{gauche}}{N_m} L(E_m^{gauche}(\theta)) + \frac{N_m^{droite}}{N_m} L(E_m^{droite}(\theta))$$

Cette étape est ensuite itérée jusqu'à atteindre la profondeur maximale, par exemple avec le critère $N_m = 1$. Lorsqu'un nœud est terminal, la classe prédite est la classe majoritaire dans ce nœud.

Précisons maintenant la fonction de perte L . Notons à ce propos que c'est là que se situe la différence entre un arbre de régression et de classification. Nous utiliserons l'indice Gini, choix par défaut de la fonction `rpart`. En définissant la proportion d'observations appartenant à la classe k au nœud m , $P_m^k = |\{(x_i, y_i) \in E_m, y_i = k\}|/N_m$, l'indice de Gini pour ce nœud sera :

$$L(E_m) = \sum_k P_m^k (1 - P_m^k)$$

A ce niveau, nous avons un arbre de profondeur maximal puisque toutes les branches ont été développées au maximum. Une première solution pour réduire la complexité est de donner un critère plus fin pour limiter la croissance de l'arbre. Deux critères sont classiques :

- Fixer un effectif minimal dans les deux sous-ensembles pour procéder à la séparation. Les inconvénients de cette condition sont le fait qu'elle ne tienne pas compte de l'information apportée par une éventuelle séparation et son caractère arbitraire.
- Fixer une amélioration minimale de l'impureté. Cependant, ce critère peut conduire à rejeter une séparation qui aurait ensuite conduit à un nœud très pertinent.

Nous choisissons donc une méthode alternative qui consiste à développer l'arbre au maximum puis à l'élaguer. C'est la *cost-complexity pruning*, qui repose sur un compromis entre l'erreur et la complexité. Nous notons pour un arbre A : la perte pondérée par le poids des nœuds $R(A)$, qui dépend de la fonction de perte L , et $|A|$ le nombre de nœuds terminaux. Alors, pour un paramètre α fixé, cette méthode trouve l'arbre élagué A qui minimise le critère de coût-complexité :

$$C_\alpha(A) = R(A) + \alpha|A|$$

Le paramètre α fait donc pencher le compromis vers un arbre de petite taille lorsque sa valeur est élevée.

La théorie étant posée, évoquons les principaux avantages et inconvénients de l'arbre de classification :

- Ce modèle est facilement compréhensible dans son fonctionnement et même représentable visuellement puisqu'il procède par division de l'espace.
- L'arbre peut prendre en compte à la fois des variables quantitatives et des variables qualitatives.
- Un arbre peut se limiter à des *optima* locaux. En effet, puisque les critères de séparation ne sont pas réutilisés, le fait qu'un critère intervienne à un nœud condamne son utilisation pour tous les nœuds qui en descendent alors qu'il aurait pu être meilleur à une plus grande profondeur.
- Un arbre est très sensible aux données : un changement dans le premier nœud va modifier toute la structure de l'arbre qui en découle. Ce modèle manque donc de stabilité.

Ces inconvénients peuvent cependant être limités par l'agrégation d'arbres en une forêt.

2.6.3 Forêt aléatoire

Une des grandes méthodes d'agrégation d'arbres est le *bagging* (*bootstrap aggregation*). Présenté par [BREIMAN 2001], il consiste à utiliser une moyenne de prédicteurs indépendants afin de réduire la variance.

Pour la formalisation, nous reprenons les notations pour les variables et notons A un prédicteur, dans notre cas un arbre binaire de classification. Il est irréaliste d'espérer disposer d'échantillons indépendants par rapport à la quantité de données. A la place, sont tirés K échantillons *bootstrap* $(B_k)_{1 \leq k \leq K}$. Puis les K arbres associés $(A_{B_k})_{1 \leq k \leq K}$ sont construits. La classe prédite par le modèle de *bagging* est le résultat du vote majoritaire de ces arbres, c'est-à-dire la modalité qui a été le plus prédite.

La forêt aléatoire apporte un raffinement supplémentaire au *bagging* en décorrélant les arbres par l'ajout d'une composante aléatoire lors de leur construction. Ainsi, à chaque feuille de l'arbre, le critère de séparation n'est pas déterminé sur toutes les variables mais sur un sous-échantillon aléatoire de taille fixée.

Cette agrégation permet de réduire la variance et de limiter les inconvénients mentionnés pour les arbres. Toutefois, ceci se fait au prix d'une perte d'interprétabilité, notamment graphique.

Pour notre problème, les méthodes non supervisées et supervisées ont chacune des raisons d'être employées. Dans le premier cas, nous n'apportons pas d'*a priori* et laissons l'algorithme découvrir une structure. Dans le second, un choix est nécessaire mais la structure qui en découle a un lien certain avec notre objectif. Ayant présenté de façon théorique des modèles appartenant à ces catégories, nous allons les mettre en application.

3 Application des modèles

Dans cette partie, les modèles que nous venons de présenter, à la fois pour la simulation du développement des sinistres et pour la classification, sont mis en œuvre. Leurs résultats sont analysés de différentes manières et comparés à des valeurs de référence. Avant cela, nous présentons les données utilisées.

3.1 Présentation des données

3.1.1 Branche étudiée et spécificités

Afin de mettre en pratique nos modèles nous disposons d'une base de données réelle. Celle-ci a été fournie par un grand assureur français et couvre la branche Responsabilité Civile Médicale. Cette assurance est souscrite obligatoirement par les professionnels de santé libéraux et les établissements de santé, et garantit leur responsabilité en cas de dommages causés à des tiers dans le cadre de leur activité.

La base est une version mise à jour de celle utilisée pour les travaux de [RIFFAUD 2019]. Nous reprenons les traitements déjà implémentés en les mettant à jour. Ils incluent des vérifications sur les valeurs manquantes, des contrôles de cohérence entre les différentes dates liées au déroulement de chaque sinistre, aux montants de charges, paiements et réserves. Ils nous conduisent à exclure 0,23% des sinistres, représentant 0,29% de la charge totale.

La base finale compte 75 000 lignes, chacune correspondant à un sinistre survenu entre 2000 et 2019. Les variables sont des informations associées à chaque sinistre. Elles comportent plusieurs variables qui catégorisent le sinistre et sur lesquelles nous présentons des statistiques dans la section suivante. Les variables de montants sont renseignées pour chaque année de survenance :

- Le règlement des frais
- Le règlement du principal
- La réserve estimées par le gestionnaire de sinistres (provision au dossier)
- La charge

Nous disposons aussi de plusieurs variables qui ne semblent pas pertinentes dans le cadre de notre étude, comme le code postal de survenance.

Enfin, plusieurs dates nous renseignent sur la chronologie de chaque sinistre : survenance, ouverture, clôture si elle a eu lieu avant 2019, éventuelle réouverture. Notons que par rapport à la figure 2 de la partie 1, la date de déclaration par l'assuré n'est pas présente. Nous la considérons égale à la date d'ouverture par le gestionnaire.

Nous appelons « tardifs » les sinistres déclarés lors d'un exercice comptable postérieur à celui de la survenance. Dans la branche RC médicale, il peut exister des sinistres dont la date de survenance et la date de manifestation (pour nous, indifféremment, d'ouverture) sont très distantes. Un exemple fréquemment cité est celui d'une compresse oubliée dans le corps d'un patient lors d'une opération. Les conséquences de cette erreur peuvent apparaître plusieurs mois voire plusieurs années après l'opération. Ainsi, le nombre de sinistres tardifs dans la branche n'est pas négligeable.

De plus, compte tenu du très faible nombre de réouvertures, nous considérons que toute clôture est définitive. Ce choix est cohérent avec l'hypothèse posée en partie 2 sur le processus d'état du sinistre.

Par ailleurs, une spécificité réglementaire de la branche doit être prise en compte pour l'exploitation des dates. Il s'agit de la transition entre la base « réclamation » et la base « fait générateur ». La loi dite About du 30 décembre 2002 (loi n° 2002-1577) a modifié le régime juridique pour ces contrats d'assurance. Auparavant, le principe de « base fait générateur » s'appliquait. Les assureurs étaient tenus de réparer un sinistre si le fait générateur survenait pendant la période de couverture. Depuis cette loi, le principe de « base réclamation » s'applique : c'est une réclamation effectuée pendant la période de couverture qui déclenche l'intervention de l'assureur.

S'ajoute une garantie subséquente qui vise à garantir la continuité de couverture. Elle est par exemple de 10 ans en cas de départ à la retraite du praticien.

En conséquence, les sinistres de notre base dont l'ouverture est antérieure à 2002 ne sont pas du même type que les autres. Nous choisissons toutefois de les conserver en première approche, car le but premier est d'analyser le comportement des modèles. De plus, cela nous priverait d'une partie non négligeable de notre historique alors que, comme nous le montrons dans la section suivante, la branche est longue.

3.1.2 Quelques statistiques descriptives

Nous présentons maintenant quelques statistiques afin de montrer des caractéristiques de la base et de dégager quelques tendances qui guideront la suite de notre travail.

Longueur des déroulés

La RC médicale est considérée comme une branche longue du fait des déroulés de sinistres qui durent souvent plusieurs années, voire une vingtaine. En effet, le coût définitif lié à une réclamation s'établit à partir de deux facteurs : le *quantum* et la part de responsabilité.

- Le quantum correspond au coût total à payer pour une responsabilité de 100%. Il ne peut être déterminé qu'une fois l'état de la victime stabilisé. C'est la principe de consolidation. Cette consolidation nécessite souvent plusieurs années. De plus, pour des victimes mineures, il est courant qu'elle n'intervienne qu'après la majorité. Un juge pourra par exemple définir des indemnisations durant l'enfance mais fixer la somme définitive après la majorité.
- La part de responsabilité à attribuer à chaque intervenant d'un acte médical ayant causé un sinistre donne souvent lieu à débat, ce qui cause entraîne un délai pour sa détermination.

La longueur importante des déroulés est confirmée dans notre base par l'observation du nombre de sinistres ouverts en 2000 et encore ouverts à chaque année de développement, présenté dans la figure 6.

Dans une branche comme le dommage automobile, presque aucun sinistre ne serait ouvert après 8 ans ; nous en avons ici un nombre non négligeable. Nous observons que tous les sinistres sont clos à la fin de notre période d'observation, mais l'historique est tout juste suffisant pour couvrir leur durée de vie entière. Or, la mise en place des modèle nous impose de supposer une durée de vie maximale pour nos sinistres, égale à la profondeur de

l'historique. C'est aussi une contrainte de modèles classiques comme Chain-Ladder, mais il convient de garder cette limite à l'esprit. En particulier, la suppression des sinistres antérieurs à 2003 à cause de la loi About évoquée précédemment rendrait souhaitable l'utilisation d'un facteur de queue.

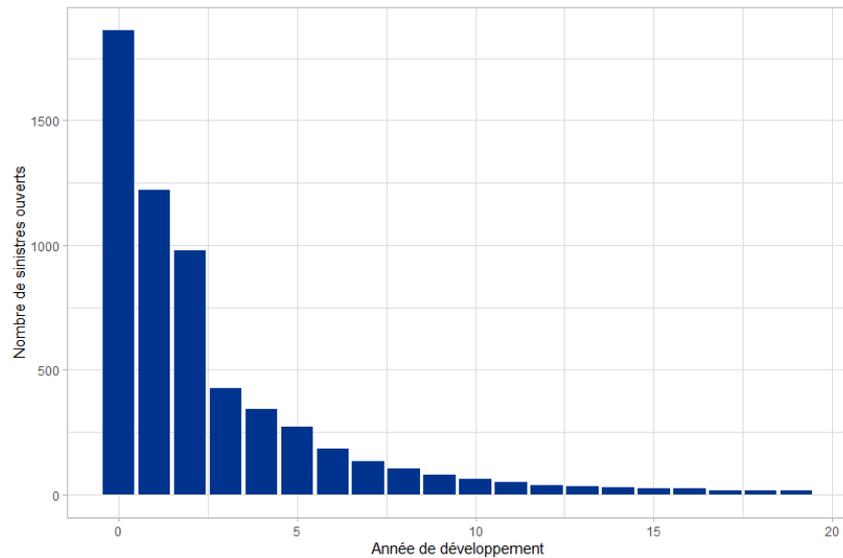


FIGURE 6 – Nombre de sinistres survenus en 2000 encore ouverts par développement

Homogénéité temporelle

L'homogénéité du nombre de sinistres survenus par année est un point important pour la simulation des sinistres tardifs. Comme évoqué, nous avons recours à une méthode de bootstrap sur un triangle de nombre, qui nécessite une certaine stabilité des survenances. Afin de la valider, nous présentons le nombre de sinistres survenus par année calendaire dans la figure 7.

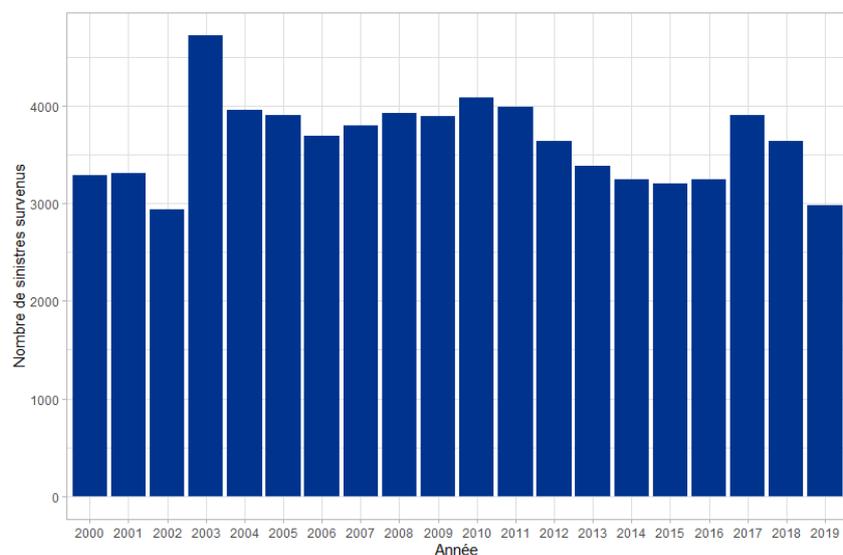


FIGURE 7 – Nombre de sinistres survenus par année calendaire

L'homogénéité est visuellement acceptable. Notons que l'écart des années 2000 à 2002, expliqué par le changement de réglementation, n'est pas significatif et nous pousse à conserver ces survenances. Le nombre de survenances reportées en 2019 inférieur aux années précédents à cause des sinistres tardifs.

Variables de catégorie

Plusieurs variables précisent le type de sinistre :

- Corporel/matériel, indique si le sinistre concerne des biens matériels
- Amiable/litigieux, indique si le sinistre a été jugé devant un tribunal
- Tardif/non tardif, qui correspond à la définition que nous avons donnée
- Courtier/autre, décrit le mode de distribution du contrat
- pooled/other/[modalité vide] qui indique un type de contrat, mais dont la définition précise n'est pas présente.

Ces variables sont remplies pour chaque sinistre et permettent de dresser une première typologie. Nous pouvons supposer que certaines vont influencer le développement des sinistres. Afin de vérifier ceci, nous étudions les sinistres clos de chaque catégorie et nous relevons les moyennes de deux indicateurs clés : les paiements versés et la durée de vie du sinistre. Pour préserver la confidentialité des données, les montants ont été multipliés par un facteur, proche de 1 pour conserver les ordres de grandeur et faciliter l'interprétation. Les montants seront donc exprimés en unités monétaires (UM). Sur la totalité des sinistres clos, soit 89% de la base, le paiement moyen est de 6 035,65 UM pour une durée de vie de 2,17 ans. Les statistiques par catégorie sont présentées dans le tableau 4.

	Corporel	Matériel	Litigieux	Amiable	Tardif	Non tardif	Courtier	Autre
Proportion tous sinistres	0,63	0,37	0,13	0,87	0,17	0,83	0,66	0,34
Proportion clos	0,61	0,39	0,09	0,91	0,17	0,83	0,65	0,35
Charge finale (UM)	9 113	1 273	25 405	4 147	6 264	5 988	6 704	4 793
Durée de vie (années)	2,90	1,07	4,36	1,97	2,09	2,20	2,49	1,61

TABLE 4 – Statistiques descriptives par catégorie

En première ligne est indiquée la proportion que cette catégorie représente parmi toute la base, et en deuxième ligne parmi les sinistres clos, à partir desquels nous calculons les statistiques suivantes. Rappelons que les modalités des variables présentées ici vont par paires : chaque sinistre est à la fois corporel ou matériel et litigieux ou amiable et tardif ou non tardif, etc. La comparaison des sinistres corporels et matériels révèle une nette différence des montants moyens réglés, ce qui est compréhensible, et une durée de vie plus longue pour les corporels. Cela se comprend dans la mesure où la gravité d'une atteinte corporelle n'est pas toujours évaluée rapidement, et l'état de la victime peut évoluer. Le même type de différence se retrouve entre les sinistres litigieux et amiables : logiquement, la durée de vie des premiers est plus élevée à cause de la durée d'une procédure judiciaire. Les montants réglés sont aussi supérieurs, pour deux raisons :

- Le passage au tribunal nécessite l'emploi d'avocats et la réalisation d'expertises, ce qui augmente les frais.
- Le tribunal fixe souvent un montant à régler à la victime plus élevé que ce que l'assureur proposait à l'amiable (si le montant fixé par la justice était au contraire plus bas, l'assureur saisirait systématiquement le tribunal), ce qui augmente le principal.

Le mode de souscription conduit aussi à une différence mais elle est moins intuitivement explicable.

Enfin, attardons-nous sur les sinistres tardifs. Notons d’abord leur proportion : dans une branche comme l’assurance auto, ils représentent une part assez faible, de l’ordre de 5%. Nous en avons ici 7 fois plus dans la base. De surcroît, la proportion réelle est sous-estimée puisque par définition, les sinistres tardifs survenus en 2019 ne sont pas encore reportés dans la base, et certains plus tardifs non plus. Cette statistique confirme une particularité de la branche RC médicale et souligne la nécessité de simuler les sinistres tardifs pour obtenir une estimation correcte de la charge ultime. Par ailleurs, ces sinistres ont aussi des règlements plus élevés et une durée de vie légèrement plus longue, ce qui avait conduit [RIFFAUD 2019] à les simuler à part. Cependant, cette différence est moins marquée qu’avec les catégories corporel/matériel et litigieux/amiable.

Nous avons ici limité notre analyse comparative à des groupes définis par une seule variable. Cela permet de constater une typologie de sinistre qui influe sur le développement, ce qui souligne la pertinence d’effectuer une classification au lieu de les simuler tous identiquement. Dans la suite de ce mémoire, lorsque nous lancerons un modèle pour développer les sinistres, 1000 simulations seront toujours effectuées. Ainsi, une courbe correspondra à la distribution obtenue à partir de 1000 points et une valeur d’ultime seule sera une moyenne sur les résultats de toutes les simulations.

3.2 Backtesting

Avant de perfectionner les modèles de développement, nous cherchons à évaluer ceux dont nous disposons. Nous les reprenons tels quels, aussi un contrôle permet de situer notre point de départ. L’objectif de ces modèles est de fournir une simulation des charges à l’ultime, mais nous ne connaissons évidemment pas la vraie valeur de cette charge. Elle ne sera disponible que lorsque tous les sinistres de la base auront achevé leur développement, c’est-à-dire 20 ans après la dernière année de la base, d’après notre hypothèse. En revanche, il est possible d’évaluer les prédictions à un ou deux ans, en simulant des années déjà connues. C’est le principe du *backtesting*. Plusieurs principes développés ici seront réutilisés par la suite.

3.2.1 Traitement de la base

Une méthode courante pour cette analyse est de reculer la vision de la base. Pour nous, cela correspond à la mettre en vision 2018, autrement dit enlever toutes les informations postérieures à 2018. Ensuite, en simulant un an de développement, nous pouvons comparer la simulation à ce qui s’est réellement produit en 2019, que nous connaissons.

Précisément, la mise en vision 2018 de la base se fait en plusieurs points :

- Retrait des colonnes décrivant les développements de 2019 (paiements incrémentaux, facteurs de développement).
- Retrait des sinistres déclarés en 2019. Cela correspond d’une part aux IBNER survenus et ouverts en 2019, et d’autre part aux sinistres tardifs survenus avant. La capacité de simulation des IBNYR du modèle sera donc elle aussi mise à l’épreuve.
- Adaptation du triangle de nombres qui sert à la génération des sinistres tardifs. La ligne correspondant à l’année 2019 et la colonne du dernier développement sont

supprimées.

- Modification de l'inflation. Le référentiel change et il convient de se placer en vision 2018 en recalculant les coefficient d'inflation. Pour cela, nous appliquons la méthode de Verbeek évoquée précédemment sur la base en vision 2018.

Notons qu'avec ces modifications, le modèle ne pourrait plus simuler les sinistres avec 20 développements mais seulement 19. Toutefois, ceci n'intervient pas puisque nous nous limitons à un an. Nous réaliserons ensuite un *backtesting* à 2 ans et le processus pour mettre la base en vision 2017 sera le même.

3.2.2 Résultats à 1 an

A partir de cette base adaptée, nous simulons un développement des sinistres, y compris l'apparition de tardifs, pendant un an seulement. Nous obtenons donc une distribution de paiements (respectivement de charges pour le modèle multiplicatif) que nous pouvons comparer aux paiement réellement versés en 2019 (respectivement à l'évaluation des charges en 2019).

Ce type de *backtesting* avait déjà été réalisé par [PRIMEL 2021] sur la base 2018 mise en visions 2017 et 2016 avec les mêmes modèles. Une des conclusions était que les montants réels étaient associés à une probabilité relativement élevée dans la distribution simulée, à la fois pour le modèle additif et pour le modèle multiplicatif. Nous avons maintenant un historique plus profond d'un an est une base légèrement différente.

Présentons les résultats des simulations à un an.

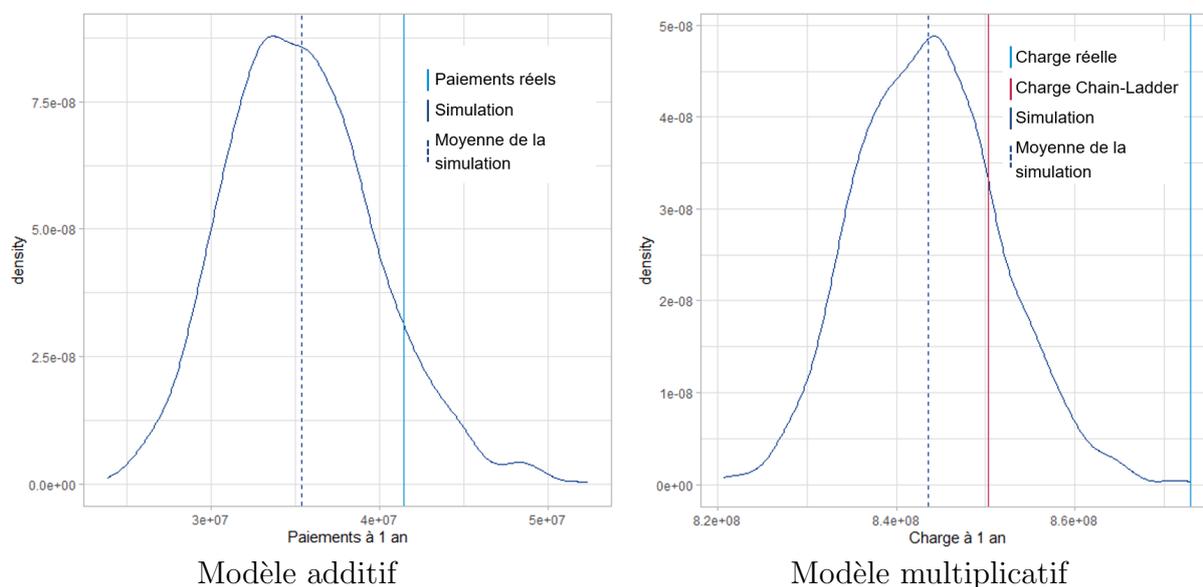


FIGURE 8 – Résultats du backtesting à 1 an

Pour le modèle additif, nous constatons que le paiement réalisé est dans une zone assez probable de la distribution simulée, ce qui montre la pertinence du modèle à un an.

Pour le modèle multiplicatif, la distribution de charges à 1 an prédite à partir de la base en vision 2018 est très inférieure à la charge effectivement enregistrée dans la base 2019. La vraie valeur tombe même dans une zone de probabilité nulle. Le *backtesting* de ce modèle effectué sur des données allant jusqu'à 2018 avait montré une tendance du modèle à prédire légèrement au dessus de la réalité, tout en gardant la charge réelle dans une zone de probabilité non nulle. Notre résultat est donc très surprenant. Pour comparer, nous faisons la même démarche avec le modèle de Chain-Ladder : un triangle agrégé de charges en vision 2018 est développé un an. En effet, cette méthode est le pendant agrégé du modèle. Comme l'écart n'est pas observé sur le modèle additif, nous pouvons supposer qu'il est lié à un effet multiplicatif. Le résultat est porté sur le graphe en figure 8. Nous constatons que le résultat de Chain-Ladder est vraisemblable dans la distribution que nous simulons, mais qu'il est aussi éloigné de la charge réelle. Le fait que ces deux modèles soient proches mais éloignés de la réalité ne nous oriente pas vers une anomalie du modèle multiplicatif, mais plutôt vers un développement particulier en 2019. Nous cherchons donc à étudier cette piste, pour confirmer ou invalider cette supposition.

En première approche nous repartons de la méthode de Chain-Ladder à un an. Ceci correspond donc à ajouter une sous-diagonale au triangle. Nous comparons, élément par élément, la diagonale obtenue avec la diagonale réelle, issue de la base 2019. Ceci fait apparaître une sous-estimation importante des charges sur plusieurs années de développement, les deux montants les plus élevés étant 5,2 et 4 millions d'UM. Dans certains cas, la charge est au contraire surestimée mais au total, Chain Ladder donne 22,6 millions de moins que les observations.

Nous cherchons ensuite à savoir si ces montants plus élevés sont le fait de quelques sinistres qui ont connu un développement exceptionnel ou le résultat d'un effet global, comme une modification de la gestion. Nous nous concentrons sur les trois développements qui ont le plus d'écart avec Chain-Ladder pour examiner les sinistres graves. Il s'agit des développements 7, 12 et 14. Pour chacun des développements j , nous retenons les 10% (ou 5% en fonction de l'effectif) de sinistres ayant la charge $j - 1$ la plus élevée. L'objectif est de voir si ces charges restent stables ou augmentent fortement. Nous relevons donc les facteurs de développement $f_{i,j}$ associés et nous les comparons à la distribution historique (de 2000 à 2018) des facteurs $f_{i,j}$. Le résultat est présenté dans la figure 9. Les facteurs correspondant aux sinistres les plus graves en 2019 pour ces développements sont représentés par des points et surmontés de la charge $j - 1$ en milliers d'UM.

La plupart des sinistres les plus graves ont un facteur proche de 1, comme le montre l'accumulation de points à cet endroit. Toutefois, nous constatons que pour ces trois développements, des sinistres de charge élevée (plus d'un million d'UM) ont subi des aggravations très fortes avec des facteurs de plus de 1,5. Ces facteurs sont très improbables dans la distribution historique. Ceci indique donc que l'année 2019 a vu des sinistres graves se développer de manière atypique.

Afin de vérifier si les sinistres communs se sont développés de manière inhabituelle, nous comparons, pour chaque année de développement, la distribution historique et la distribution 2019. Il ressort qu'elles sont semblables avec parfois légèrement plus de facteurs élevés pour 2019. Ainsi, le niveau élevé de charges observé en 2019 peut être imputé principalement au développement exceptionnel de quelques sinistres graves.

Une limite du modèle multiplicatif est ici mise en évidence : puisqu'il s'appuie sur l'historique, il ne peut pas prendre en compte des tendances nouvelles qui sont absentes de

notre base.

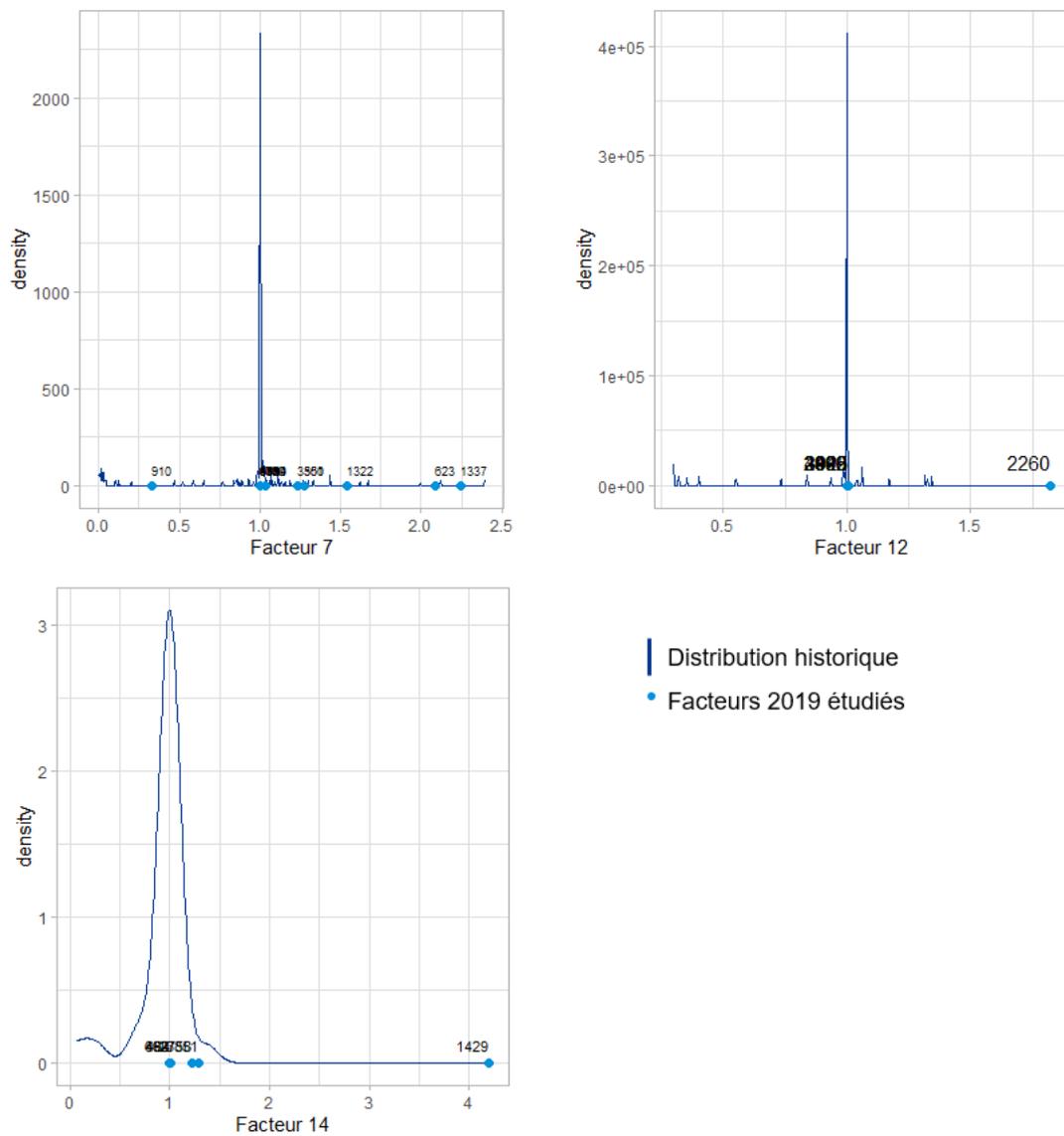


FIGURE 9 – Analyse du développement de sinistres graves

3.2.3 Résultats à 2 ans

L'évaluation à un an est limitée pour un modèle de prédiction de charges ultimes. Cependant, nous sommes contraints dans la profondeur du *backtesting* puisque les années que nous voulons simuler sont retirées de la base, donc retirées de l'historique. Le choix d'une profondeur de deux ans semble laisser un historique suffisant. De plus, nous avons vu que l'année 2019 a un développement atypique. Une simulation sur deux ans pourrait permettre de mitiger cette année particulière avec une année plus classique, et de voir l'impact d'une année exceptionnelle sur une simulation à plus long terme.

Nous plaçons la base en vision 2017 et simulons jusqu'à 2019. Les résultats sont présentés en figure 10.

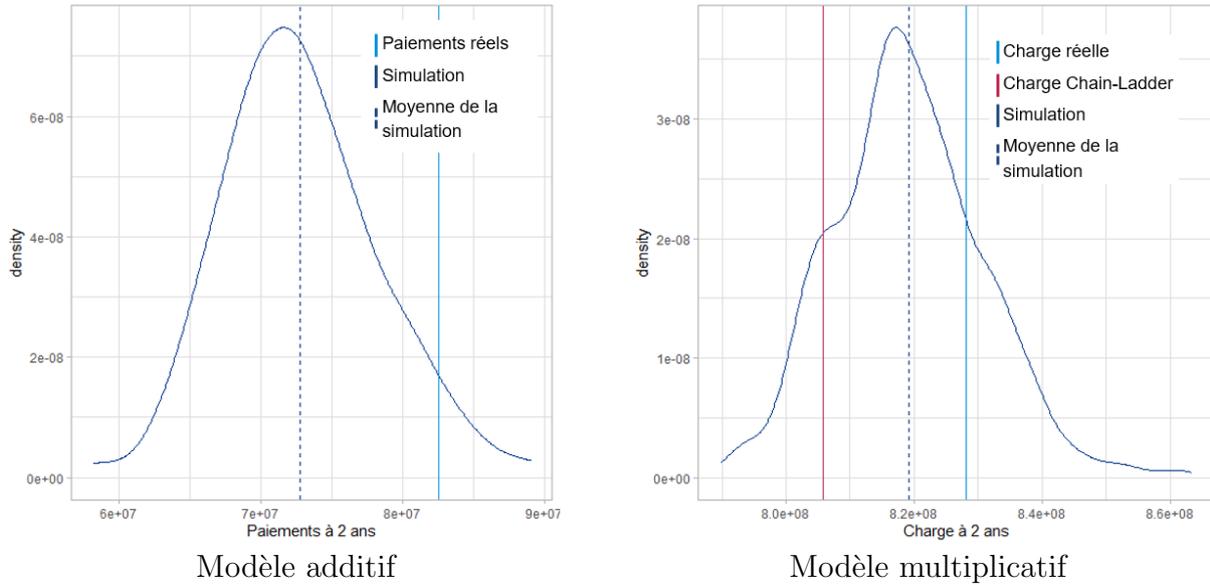


FIGURE 10 – Résultats du backtesting à 2 ans

Encore une fois, le modèle additif est cohérent et la valeur réelle est associée à une probabilité non nulle. Le modèle multiplicatif se rapproche nettement de la charge réelle, au point de lui attribuer une probabilité élevée. [PRIMEL 2021] avait observé pour ce modèle une tendance à surestimer la charge ultime par rapport à Chain-Ladder. Cette tendance se retrouve ici et contribue probablement à rapprocher l’estimation de la charge élevée de 2019. L’écart dû au développement exceptionnel s’est donc résorbé.

Nous tirons plusieurs conclusions de ces *backtestings*. D’abord, le modèle additif fournit une bonne estimation à 1 an, puisque la valeur réalisée a une probabilité non nulle dans nos simulations. Ensuite, le modèle multiplicatif peut fournir une bonne estimation mais est sensible aux développements extrêmes de sinistres graves. En particulier, les modèles, par leur structure, ne peuvent simuler un effet nouveau qui serait absent de l’historique. Toutefois, si une seule année est atypique, l’écart est rattrapé en développant sur plusieurs années.

3.3 Classification déterministe

Dans cette section nous présentons une segmentation qui consiste simplement à constituer manuellement des groupes de sinistres à partir d’une variable et à les développer indépendamment. Nous disposons dans notre base de variables de catégorie (comme corporel/matériel) qui ont un sens du point de vue métier. En envisageant une classification, il est intuitif d’utiliser ces catégories en première approche. De plus, les statistiques descriptives présentées en début de partie ont montré que ces variables définissent des groupes de sinistres aux développements distincts, mais nous avons jusqu’alors simulé tous les sinistres ensemble. Rappelons qu’une segmentation manuelle est déjà couramment mise en place dans les compagnies d’assurance pour la gestion des sinistres et le provisionnement. Une des étapes élémentaires est la séparation des sinistres attritionnels, qui peuvent être développés via la méthode de Chain-Ladder, et les sinistres graves, dont le faible nombre et les montants élevés rendent le développement par des triangles plus difficile.

Cette classification a pour avantage de ne nécessiter aucun algorithme, donc aucun choix humain si ce n'est la variable avec laquelle séparer. Ceci est de plus totalement interprétable, et une connaissance métier peut anticiper l'effet de la classification pour juger les résultats de notre simulation. Outre la connaissance métier, nous pouvons justifier la différenciation des simulations par les statistiques descriptives par catégorie.

3.3.1 Mise en œuvre pratique et méthode d'évaluation

La mise en place de cette méthode de classification est très simple puisqu'elle est faite manuellement. Un fois les classes déterminées, la base est simplement séparée en groupes, puis le modèle de simulation est lancé sur chacun des groupes indépendamment.

Nous introduisons ici une nouvelle manière d'évaluer les simulations. Comme nous l'avons mentionné, il nous est impossible de comparer la charge ultime à une vraie valeur. En revanche, nous pouvons reprendre le principe du *backtesting*, c'est-à-dire simuler un développement en réalité connu. Les analyses précédentes étaient principalement graphiques, mais nous cherchons ici à quantifier les erreurs afin de les comparer. De plus, en cohérence avec l'idée du développement individuel, nous cherchons à évaluer l'erreur au niveau individuel, pas seulement sur la somme des réserves. Notre but est en effet d'être précis sur chaque groupe de sinistre, ce qui n'est pas forcément le cas avec un modèle bon en moyenne.

A cette fin, nous mettons en place un calcul d'indicateurs d'erreur absolue. Nous employons le RMSE (*Root Mean Square Error*) et le MAE (*Mean Absolute Error*), deux mesures très classiques. En notant y_i une valeur et \hat{y}_i sa prédiction, ces indicateurs sont définis respectivement par :

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

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

A cause de la convexité du carré, le RMSE est plus sensible aux grandes erreurs, contrairement au MAE.

Le calcul de ces métriques nécessite de pouvoir relier la simulation du développement d'un sinistre à l'évolution réelle de ce sinistre en 2019. Il faut donc que le sinistre soit dans la base, ce qui exclut les tardifs. Pour cette raison, dans une optique d'évaluation à un an nous nous limitons à la simulation des IBNER.

3.3.2 Différentes segmentations et résultats

Séparation corporel/matériel

Nous commençons par la variable qui d'après nos statistiques descriptives causait le plus de différence dans le développement des sinistres. Nous partons de la base en vision 2018 et nous la séparons en une base regroupant les sinistres corporels et une base regroupant les sinistres matériels. Nous simulons ensuite un an de développement et calculons des indicateurs d'erreur à partir du réel 2019 que nous connaissons. Ceci nous fournit donc des mesures d'erreur sur les deux catégories de sinistre envisagées, et nous pouvons calculer une moyenne pondérée afin de connaître l'erreur du modèle sur la base globale. Pour constituer un point de comparaison, nous lançons le modèle sur la base globale (modèle sans séparation) et nous agrégeons les erreurs par catégorie. Ceci nous conduit au tableau 5.

	Réserves : moyenne (kUM)	Réserves : écart-type (kUM)	RMSE (UM)	MAE (UM)
Modèle sans séparation	34 666	4 267	73 548	10 352
Corporel	31 693	4 201	82 007	12 443
Matériel	2 972	588	11 805	2 101
Modèle avec séparation	34 814	4 586	74 105	10 331
Corporel	33 497	4 536	82 682	12 658
Matériel	1 316	546	8 600	1 148

TABLE 5 – Résultats à 1 an de la classification corporel/matériel, modèle additif

Nous constatons que les montants de réserve prédits par les modèles avec et sans séparation sont proches au total. Cependant, le modèle avec séparation prédit comparativement plus sur les sinistres corporels et moins sur les matériels. Ainsi, il se rapproche de la tendance intuitive observée dans les statistiques descriptives, à savoir un coût plus élevé pour les corporels. D'un point de vue d'expert, cet effet est très positif. La variance est aussi réduite pour les matériels, ce qui est cohérent avec des valeurs plus basses, tandis que l'effet est inversé sur l'autre classe.

Ensuite, les indicateurs d'erreur MAE et RMSE nous signalent que le modèle avec séparation n'est globalement pas meilleur à un an. Ceci s'explique par un gain de précision sur la modalité matérielle mais une dégradation sur l'autre modalité.

	Charges : moyenne (kUM)	Charges : écart-type (kUM)	RMSE (UM)	MAE (UM)
Modèle sans séparation	835 047	7 254	148 681	24 236
Corporel	801 549	7 227	166 203	29 504
Matériel	33 498	701	14 890	3 447
Modèle avec séparation	836 604	7 846	152 356	24 333
Corporel	805 223	7 852	170 443	29 969
Matériel	31 380	485	8 342	2 079

TABLE 6 – Résultats à 1 an de la classification corporel/matériel, modèle multiplicatif

Nous répétons la même procédure avec le modèle multiplicatif afin de comparer les modèles entre eux. Les résultats sont présentés dans le tableau 6. Les observations sont similaires : l'effet global de la classification sur les charges à un an est négligeable, mais la répartition entre les catégories change en se rapprochant de ce qui est attendu compte tenu des statistiques descriptives. Au niveau des indicateurs d'erreur, le RMSE augmente très

légèrement avec la classification alors que le MAE est stable. Le premier indicateur donne plus d'importance aux grandes erreurs, ce qui suggère que celles-ci ont augmenté alors que les petites erreurs ont diminué, laissant le MAE stable. En regardant par catégorie, nous constatons que les deux indicateurs baissent sur les sinistres matériels. En revanche, le RMSE augmente sur les sinistres corporels, révélant des erreurs plus grandes en valeur, en moyenne. Dans cette catégorie se trouve une grande amplitude de charges, avec des montants très élevés mais aussi des montants faibles, ce qui explique qu'elle concentre l'erreur.

A ce stade, il nous paraît nécessaire de rappeler une limite de notre analyse : pour le calcul des indicateurs d'erreur, les sinistres sont seulement développés un an. Or, notre modèle a vocation à fournir une charge à l'ultime. Nous n'évaluons donc ici qu'une portion des résultats. En particulier, l'importance de l'estimation de la durée de vie est occultée. Les sinistres corporels par exemple ont une durée de vie moyenne de 3 ans, et n'évoluent pas tous les ans. Une analyse à un an est donc insuffisante pour capter pleinement les distinctions qu'apporte la classification. Cependant, la profondeur de notre historique est limitée en regard des durées de vie rencontrées dans la branche RC médicale, il est donc difficile de se priver de plusieurs années d'historique pour effectuer un *backtesting* à plus long terme.

Cette première classification simple a produit un effet marqué entre les deux catégories. Nous allons naturellement vérifier s'il en est de même pour les autres variables. Avant cela, examinons un problème liée à la séparation des simulations.

Comparaison du calibrage et de la simulation

Nous avons montré qu'une classification produit des différences marquées dans la simulation, mais se pose toujours la question de la capacité de notre modèle à produire des résultats cohérents. Les valeurs de réserves et de charges ultimes obtenues sont cohérentes, et nous le vérifierons ultérieurement en les comparant à des prévisions d'experts. Cependant, les comparaisons à notre historique sur le principe du *backtesting* sont limitées, et seule la moyenne des simulations est examinée. En effet, la valeur des paiements effectivement réalisés une année est unique, alors que notre modèle simule toute une distribution.

Nous pouvons en revanche examiner à partir des données réelles une distribution des règlements ou des facteurs de développement individuels. Nous menons cette analyse année de développement par année de développement, en incluant à chaque fois toutes les années de survenance. Nous extrayons de la base tous les paiements réalisés, lorsqu'un paiement non nul a eu lieu, pour une année de développement donnée, ce qui fournit une distribution. Ensuite, nous lançons 1000 simulations de notre modèle en extrayant les paiements et nous comparons les deux distributions. Les résultats pour les développements 1 à 4 sont présentés en figure 11. En raison de l'étendue de la plage de paiements, avec un petit nombre de valeurs élevées, nous passons à une échelle logarithmique pour l'axe des abscisses.

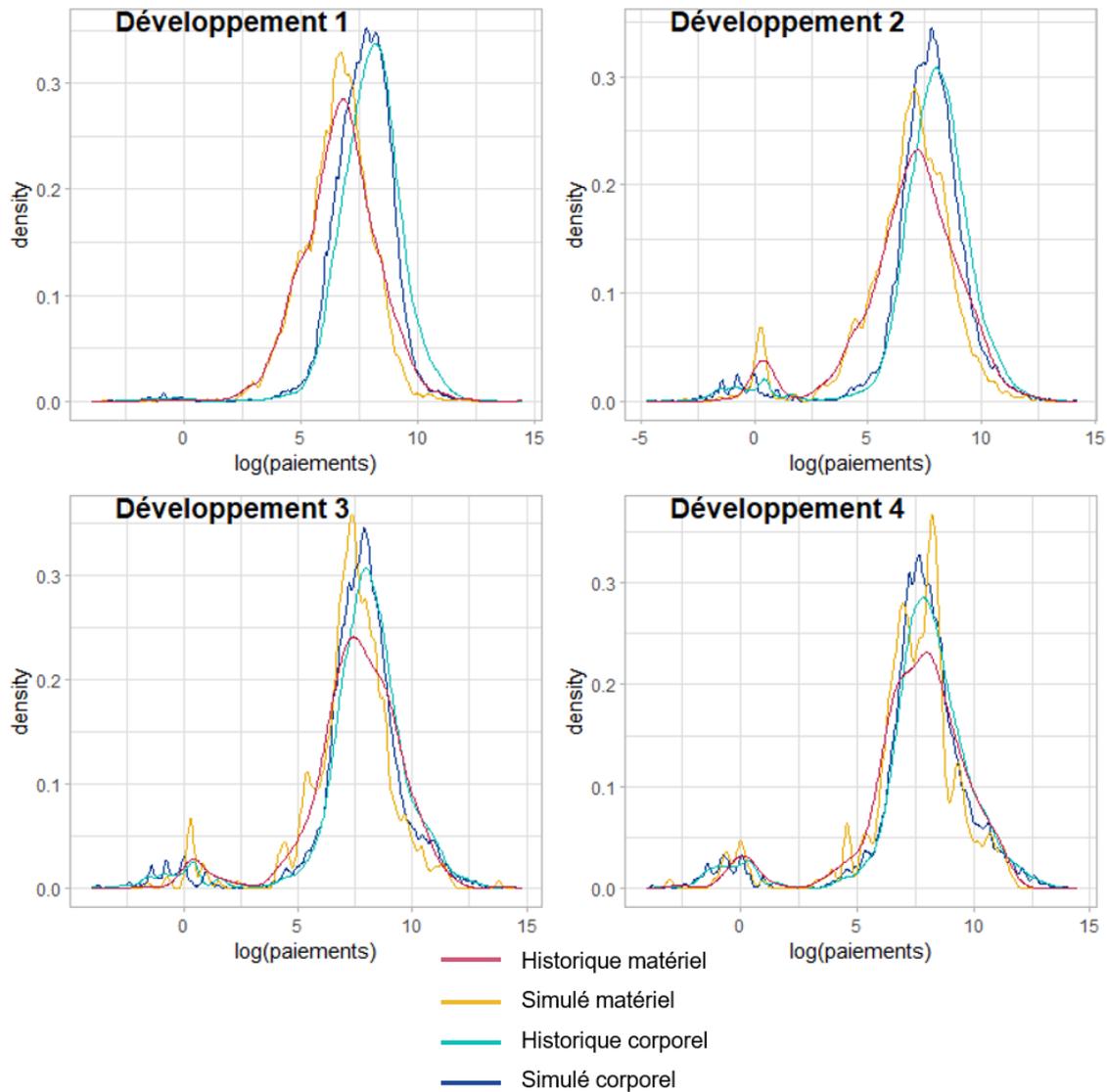


FIGURE 11 – Comparaison des distributions de paiements historiques et simulés

Dans un premier temps, observons seulement les courbes historiques. Nous constatons en année 1 une nette différence entre les deux catégories avec une distribution de paiements pour les sinistres matériels globalement décalée vers la gauche. Aux développements suivants, les courbes se superposent quasiment. Ainsi, la différence de développement entre les deux types de sinistres est due au tout début du développement. Il convient aussi de garder à l'esprit, bien que ce ne soit pas visible ici, que le nombre de sinistres matériels ouverts chaque année diminue beaucoup plus vite que pour les corporels. Cela contribue notamment à rendre les courbes moins lisses. Cela signifie aussi que les sinistres matériels encore ouverts au développement 2, dont nous voyons qu'ils ont des règlements plus élevés, sont la partie la plus grave de cette catégorie, et qu'une durée de vie plus longue est associée à des paiements plus élevés. Aux développements suivants il ne reste presque plus de sinistres matériels et peu de corporels.

Dans un second temps, comparons les courbes des distributions historiques et simulées. Elles se superposent pour les sinistres corporels. Pour les matériels, quelques écarts sont observés à partir du deuxième développement, liés à l'effectif plus réduit, mais les deux courbes restent proches. Cela montre que, bien que nous nous concentrons sur la moyenne,

notre modèle additif reproduit les distributions réelles de règlements. De plus, la séparation par catégories conduit à reproduire les différences réellement observées.

Nous pouvons mener la même étude avec la distribution de facteurs du modèle multiplicatif, mais nous savons [PRIMEL 2021] qu’il ne suffit pas de reproduire la distribution pour avoir un résultat cohérent, il est aussi important d’attribuer le bon facteur au bon sinistre, d’où la présence de la copule.

Autres séparations

Nous avons détaillé la mise en place et l’évaluation de notre modèle à classification déterministe dans le cas de la séparation entre sinistres corporels et matériels. Cette séparation peut se faire à partir des autres variables de catégorie précédemment listées. Nous présentons les résultats associés plus brièvement. Les tableaux détaillés sont en annexe.

Les statistiques descriptives avaient indiqué une différence marquée entre les sinistres litigieux et amiables. En effet, la tenue d’un procès rallonge la durée de vie du sinistre et le tribunal choisit souvent un montant plus élevé que ce que l’assureur proposait. Il est donc naturel de distinguer ces deux types de sinistres pour une simulation. Les résultats obtenus, présentés dans le tableau 7, sont similaires à ceux de la séparation corporel/matériel : les réserves globales sont proches de celles prédites par le modèle sans séparation, mais sont réparties différemment entre les catégories. Le montant de réserves augmente pour les sinistres litigieux et diminue pour les amiables, ce qui est conforme à ce que nous attendons. L’erreur globale est semblable entre les deux modèles mais la classification fait gagner de la précision sur la catégorie dont les sinistres ont une plus petite charge, c’est-à-dire celle des sinistres amiables.

	Réserves : moyenne (kUM)	Réserves : écart-type (kUM)	RMSE (UM)	MAE (UM)
Modèle sans séparation	34 842	4 545	73 087	10 374
Amiable	15 198	2 535	38 598	5 652
Litigieux	19 644	3757	105 331	17 247
Modèle avec séparation	40 509	4 710	76 545	10 864
Amiable	9 949	1 679	28 554	4 636
Litigieux	30 560	44 196	114 031	19 928

TABLE 7 – Résultats à 1 an de la classification amiable/litigieux, modèle additif

Les séparation sur d’autres catégories, en revanche, ne conduisent pas à des différences de montant de réserve ni à des différence d’erreur marquées entre les deux modèles, au niveau global ou sur une catégorie. Cependant, puisque nous disposons de deux classifications efficaces, une idée naturelle est de les combiner.

Classification à deux niveaux

Lorsque nous séparons les sinistres corporels et matériels, l’erreur ne diminue pas sur la plus grande classe, celle des corporels. L’effectif conséquent de la classe suggère de la diviser à nouveau à partir de la seconde variable de classification pertinente : amiable/-litigieux. L’effet sur la classe des matériels est satisfaisant, nous ne la divisons donc pas. Cela correspond à un modèle à 3 classes, dont les résultats à 1 an sont présentés dans le tableau 8. La classe des sinistres matériels n’a pas changé par rapport à notre premier

	Réserves : moyenne (kUM)	Réserves : écart-type (kUM)	RMSE (UM)	MAE (UM)
Modèle sans séparation	34 638	4 318	73 808	10 352
Corporel litigieux	19 406	3 546	105 625	17 326
Corporel amiable	12 282	2 433	45 702	7 467
Matériel	2 950	582	11 530	2 085
Modèle avec séparation	40 280	4 606	76 567	10 814
Corporel litigieux	30 368	4 301	113 861	20 013
Corporel amiable	8 596	1 800	35 454	6 376
Matériel	1 316	546	8 600	1 148

TABLE 8 – Résultats à 1 an de la classification à 2 niveaux, modèle additif

modèle et nous la laissons de côté dans l’analyse. Nous constatons que le RMSE global augmente très légèrement sur le modèle avec séparation alors que le MAE est stable, ce qui indique quelques erreurs de montant élevé. Encore une fois, une classe porte les augmentations d’erreurs (les corporels litigieux) alors que l’autre s’améliore. Cependant, cette fois l’effectif ne semble pas jouer puisqu’il est similaire entre les deux. Au niveau des réserves, nous observons la redistribution attendue entre les deux classes, avec la catégorie amiable qui reçoit des réserves plus faibles.

Le bilan de cette classification déterministe est donc mitigé. Nous parvenons à réduire l’erreur à un an sur certaines classes, au prix d’une dégradation sur d’autres. Les modèles peuvent être améliorés en ajoutant un étage à la séparation en scindant à nouveau la classe la plus peuplée. Ce traitement pourrait être encore répété afin de gagner en granularité, mais il nécessite une intervention humaine et des choix basés uniquement sur nos statistiques descriptives. Par ailleurs, nous nous sommes appuyés uniquement sur les variables de catégories identifiées comme telles dans la base, mais elles sont limitées. Par exemple, elles ne permettent pas d’identifier les sinistres graves alors que ceux-ci sont souvent traités à part par les actuaires.

La méthode de classification déterministe est proche de ce qui est déjà mis en place par les équipes de provisionnement, bien que le critère retenu soit souvent uniquement un seuil de charge. La suite de notre travail présente de nouvelles méthodes de classification automatiques, plus fines et efficaces, qui sont adaptées à un provisionnement individuel.

3.4 Classification par k-moyennes

Nous avons exploré en première approche une classification manuelle. Nous allons maintenant étudier des classifications à partir de méthodes d’apprentissage statistique. Dans cette section, nous mettons en place la méthode des k-moyennes. Cette méthode, que nous avons présentée théoriquement auparavant, est non supervisée. Nous n’avons pas d’opinion *a priori* sur les groupes qui devraient être constitués, d’où le choix de ce type de classification. De plus, une variable « réponse » pour la supervision serait construite à partir d’informations sur le développement du sinistre. En nous passant de celle-ci, nous pouvons exploiter tous les sinistres de la base, même ceux qui n’ont connu aucun développement.

3.4.1 Préparation des variables

Avant de lancer un quelconque algorithme, il convient de sélectionner et retraiter les variables à utiliser. Nous disposons de certaines variables dans la base déjà présentées comme des « catégories », évoquées dans la section précédente. Nous les conservons toutes, sauf la variable indiquant si un sinistre est tardif que nous remplaçons par un autre indicateur détaillé ultérieurement.

Les algorithmes que nous souhaitons mettre en place requièrent des variables numériques, aussi une transformation est nécessaire. Nous procédons à un encodage binaire, ou *one-hot encoding*. Le principe de cet encodage est que pour une variable à n modalités x_1, x_2, \dots, x_n , nous créons n colonnes x_1, x_2, \dots, x_n valant 1 si la variable vaut x_i et 0 sinon. Le principe est illustré dans le tableau 9 :

	variable	x_1	x_2	x_3
1	x_1	1	0	0
2	x_3	0	0	1
3	x_2	0	1	0

TABLE 9 – Principe de l’encodage binaire

Ensuite, nous retirons la colonne n de manière à avoir $n - 1$ colonnes linéairement indépendantes. Avec le principe de l’encodage, nous ne perdons pas d’informations.

Charge initiale

Nous ajoutons pour la classification la charge initiale. Cette variable est disponible pour tous les sinistres de la base, et dans une perspective d’application réelle, elle est disponible dès qu’un sinistre a été signalé et évalué par un gestionnaire. C’est naturellement un bon indicateur de la gravité du sinistre, et elle permet d’intégrer l’avis d’expert du gestionnaire.

La base contient des charges initiales renseignées entre 2000 et 2019, il faudrait donc les inflater pour les rendre comparables. Toutefois, l’inflation des charges est un problème nettement plus délicat que l’inflation des paiements. Les paiements sont réellement des flux de trésorerie, effectués une année donnée. La méthode de Verbeek s’applique bien sur les paiements incrémentaux, comme nous l’avons vu. En revanche, une charge prend déjà en compte une anticipation de l’étalement temporel des paiements, et ne peut pas être inflatée simplement. Les niveaux d’inflation calculés précédemment pour les paiements sont modérés, et il est vraisemblable que l’inflation des charges est du même ordre. De plus, notre base couvre une période de 20 ans. Ainsi, même sans inflation, les ordres de grandeur des charges seront conservés, et la distinction de gravité sera toujours raisonnablement valable. Nous choisissons donc de ne pas modifier les valeurs.

Délai de déclaration

Les sinistres sont qualifiés de tardifs si l’année de survenance est différente de l’année d’ouverture. L’idée sous-jacente est que le délai de déclaration est plus long que pour un sinistre non tardif, mais la définition englobe aussi bien des sinistres déclarés après 3 ans que des sinistres survenus en décembre et déclarés deux mois plus tard. Ainsi, un

certain nombre de sinistres considérés comme tardifs ont été déclarés aussi rapidement que d'autres sinistres, mais sont survenus en fin d'année d'exercice. Cette supposition se vérifie en étudiant la répartition des délais de déclaration présentée dans la figure 12.

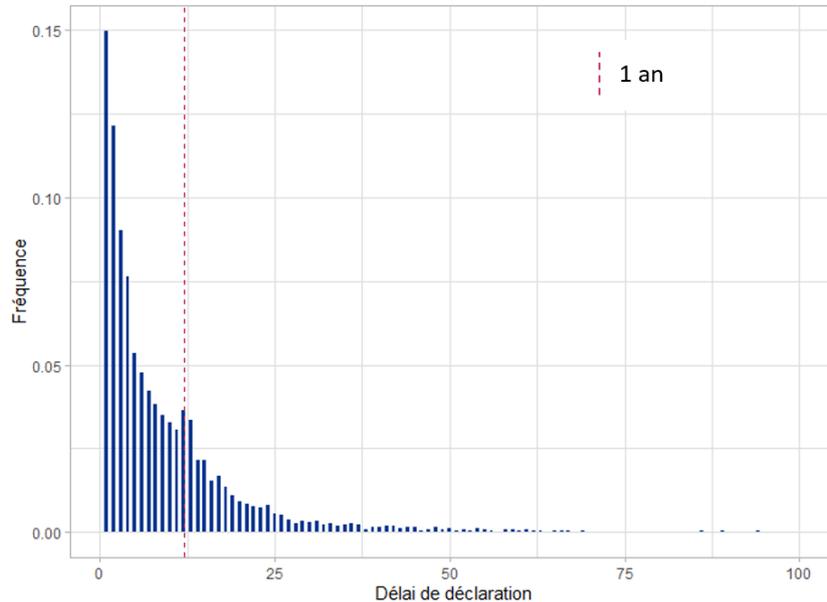


FIGURE 12 – Délai de déclaration des sinistres tardifs, en mois

La majorité des sinistres dits tardifs sont déclarés rapidement, et presque tous sont déclarés en moins d'un an (ligne pointillée verticale).

La séparation indiquée par la variable binaire tardif/non tardif est donc peu représentative de ces sinistres. Nous la remplaçons par la variable du délai de déclaration en mois, qui permettra plus de finesse, notamment en séparant les sinistres peu tardifs des très tardifs. Cette variable, tout comme la charge initiale, est centrée et réduite. A l'issue de cette phase, notre base peut être fournie en entrée aux modèles de classification.

3.4.2 Choix du nombre de classes

Avant de chercher à constituer des classes, il convient de vérifier que les données peuvent être partitionnées. Nous avons auparavant utilisé une seule variable pour une classification déterministe, mais nous considérons à présent toutes les variables à la fois. Nous utilisons la statistique de Hopkins qui indique si les données sont éloignées d'une distribution uniforme [LAWSON et JURIS 1990]. Elle se calcule comme suit : notons X l'ensemble des données et Y un ensemble de points simulés selon une loi uniforme ayant la même variance que les points de X . Nous tirons n points de X , p_1, \dots, p_n . A chacun de ces points est associée x_i la distance à son plus proche voisin dans X et y_i la distance à son plus proche voisin dans Y . Alors, la statistique de Hopkins est :

$$H = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n y_i + \sum_{i=1}^n x_i}$$

Si les données de X suivent une loi uniforme, $\sum_{i=1}^n y_i$ et $\sum_{i=1}^n x_i$ seront proches, et H sera proche de 0,5. Dans le cas contraire, la distance aux points uniformément simulés augmentera, donc $\sum_{i=1}^n y_i$ augmentera et H tendra vers 1.

La statistique de Hopkins vaut 0,996, il est donc clair que notre jeu de données peut être partitionné. Notons que nous pouvions supposer ce résultat puisque la classification déterministe a montré des améliorations de prédiction, mais nous avons cette fois un argument *a priori*.

Un point essentiel à déterminer pour une classification par *k-means* est le nombre de classes. Celui-ci doit être fixé dès le début. Il existe plusieurs méthodes pour choisir un nombre de classes adapté. Nous présentons une première approche qui n'utilise pas de modèle, puis dans un second temps nous utiliserons une approche qui évalue des résultats de *k-means*.

Estimation visuelle sans modèle : VAT

Cette méthode d'estimation du nombre de *clusters* s'appuie uniquement sur les données et est donc indépendante de tout modèle. Proposée par [BEZDEK et HATHAWAY 2002], la VAT, pour *Visual Assessment of cluster Tendency* est une matrice de dissimilarité entre les observations, ordonnée. L'ordre est calculé de manière à regrouper le long de l'axe les points proches. Ainsi, des carrés de couleur apparaissent le long de la diagonale pour des groupes de points proches. Compter le nombre de carrés donne un nombre de *clusters*. Sur un sous-échantillon de nos données, nous obtenons la matrice en figure 13.

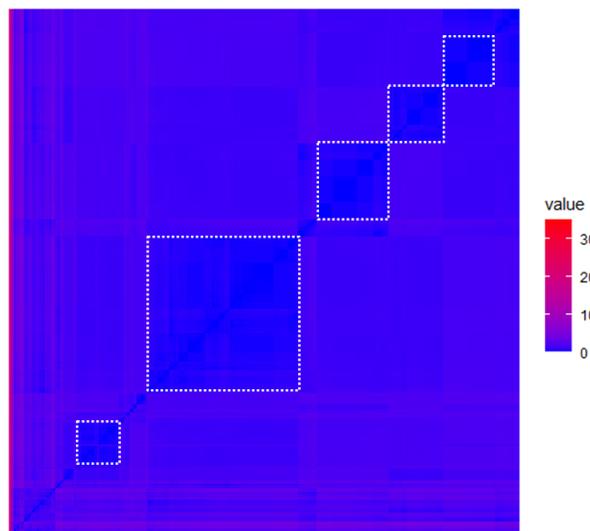


FIGURE 13 – Matrice VAT

En repérant les carrés foncés, signes d'une faible dissimilarité, nous comptons 5 gros groupes, et quelques plus petits. La délimitation des petits carrés n'est pas franche, ce qui rend leur comptage problématique. De plus, ils correspondent à des classes de très petit effectif, inadaptées à nos méthodes de développement qui reposent sur des tirages dans ces classes. Le nombre de 5 *clusters* est donc retenu par cette méthode.

Evaluation *a posteriori* du modèle

Puisque nous comptons appliquer un algorithme de k-moyennes, une idée naturelle est d'appliquer cet algorithme avec différentes valeurs de k et de comparer les résultats. Le but de l'algorithme est de constituer des groupes homogènes, autrement dit de minimiser

la variance intra-classe. Nous utilisons cette quantité comme indicateur de qualité de la classification. Mécaniquement, elle sera basse avec un très grand nombre de *clusters*, jusqu'à devenir nulle lorsque chaque point sera seul dans sa classe. Le choix repose donc sur la très usuelle méthode du coude : nous sélectionnons un nombre tel qu'ajouter une classe améliore peu la fonction objectif, ce qui correspond à un coude dans la courbe. La courbe obtenue, figure 14, conduit à choisir $k=8$.

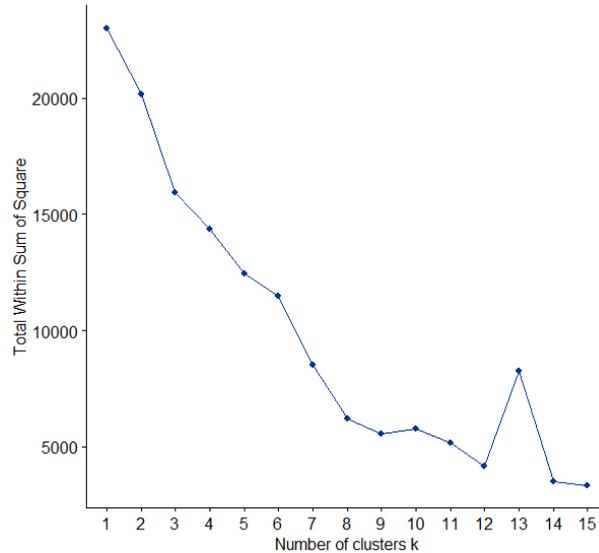


FIGURE 14 – Variance intra-classe pour l’algorithme de k-moyennes en fonction de k

Chacune des deux méthodes est pertinente, aussi les deux valeurs de k retenues, 5 et 8, seront essayées pour la classification par k-moyennes.

La difficulté dans l’évaluation des classes qui se pose à nous tout au long de ce travail est que la classification n’est pas une fin en soi. Les méthodes classiques d’évaluation des modèles sont limitées puisque notre but est de prédire des ultimes, la classification n’étant qu’un moyen. Aussi, nous allons mettre en œuvre d’autres méthodes pour juger les classes, en lien avec les problèmes opérationnels.

3.4.3 Analyse des classes obtenues

A ce stade nous retenons deux valeurs possibles de l’hyperparamètre k. Nous commençons avec 5 classes. Puisque les k-moyennes sont sensibles à l’initialisation, nous en lançons 1000 avec des initialisations différentes et retenons le meilleur, c’est-à-dire celui avec la plus faible variance intra-classe. Ceci prend une dizaine de minutes sur notre ordinateur de bureau, une durée très raisonnable.

Une fois les classes obtenues par l’algorithme des k-moyennes, nous effectuons quelques analyses afin de voir s’il est possible de caractériser les classes et de les interpréter. Le tableau 10 présente des statistiques par *cluster* et sur la base complète pour la comparaison. Les lignes 2 à 6 correspondent à des variables binaires. Le nombre indiqué est donc la proportion de sinistres pour laquelle cette variable est vraie. Les deux dernières lignes correspondent à la moyenne de la variable dans chaque classe.

	Base complète	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Effectif	72 723	55	350	25 128	42 622	4 568
Corporel	0,631	0,945	0,874	0	1	0,641
Courtier	0,663	0,873	0,454	0,538	0,745	0,593
Litigieux	0,132	0,691	0,226	0,008	0,208	0,099
"other"	0,943	1	0,969	0,965	0,927	0,972
"pooled"	0,03	0	0,029	0,019	0,038	0,019
Délai de déclaration (mois)	2,891	3,236	65,434	1,981	1,54	15,702
Charge initiale (UM)	14 979	2 297 921	19 369	1 934	20 134	10 814

TABLE 10 – Statistiques descriptives des 5 clusters

Nous constatons d’abord que la variable du type de contrat (*other* / *pooled* / ...) n’a pas d’influence sur la classification puisque les proportions dans chaque groupe sont sensiblement les mêmes. Ensuite, nous pouvons commenter les classes :

1 - La charge initiale moyenne est très élevée, 100 fois plus que la seconde charge la plus élevée. Il s’agit clairement de sinistres graves.

2 - Ici le délai de déclaration moyen est haut : plus de 5 ans. Nous avons donc une classe de sinistres très tardifs.

5 - Avec un délai de déclaration moyen d’un an et 3 mois, les sinistres sont moyennement tardifs.

3 - Cette classe n’a aucun sinistre corporel, et presque aucun litigieux. Elle comprend donc des sinistres matériels amiables. Réciproquement, des statistiques sur la base nous indiquent que presque tous les sinistres matériels amiables sont dans cette classe. Notons que la charge initiale est moins élevée que dans les autres groupes, ce qui est attendu compte tenu des caractéristiques de ces sinistres.

4 - Enfin cette classe semble regrouper les sinistres qui n’ont pas de caractéristique marquée.

Nous constatons donc que plusieurs de ces classes, bien que produites par un modèle entièrement automatique, sont interprétables en termes de combinaisons de variables de classification. Toutefois, apparaît aussi la limite du faible effectif de certains groupes comme le numéro 1. Cela pourrait fortement limiter la variété des tirages aléatoires à partir de ce groupe dans nos simulations à venir.

Les groupes étant interprétés à partir des variables de classification, l’analyse qui suit naturellement est celle de l’effet sur la variable que nous cherchons à simuler, c’est-à-dire la charge ultime. Cette fois, les statistiques ne sont plus calculées sur l’ensemble des individus de chaque classe mais seulement sur les sinistres déjà clos, afin de connaître la charge ultime et la durée de vie. Les résultats sont présentés dans le tableau 11.

	Base complète	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Durée de vie (années)	2,17	5,44	3,02	1,03	2,86	2,60
Charge ultime (UM)	5 287	781 421	9 975	902	7 464	6 215

TABLE 11 – Variables d’intérêt pour les 5 clusters

Nous vérifions que le groupe désigné comme contenant les sinistres graves a bien une charge ultime moyenne nettement plus élevée que celle des autres groupes. La durée de vie est aussi supérieure, ce qui est attendu. Les sinistres matériels, en classe 3, ont une charge moindre, ce qui est aussi attendu. Ensuite, nous notons que les sinistres tardifs sont légèrement plus chers que les deux autres groupes restants, mais ces trois groupes ont des charges ultimes proches.

Pour la classification à 8 groupes, nous obtenons aussi des classes facilement interprétables à partir des variables du modèle. Les statistiques sont regroupées dans le tableau 12. Nous distinguons cette fois trois niveaux de sinistres tardifs, correspondant aux catégories 2, 4 et 5. Nous retrouvons un groupe de très graves, le 7. Les sinistres corporels issus de contrats distribués par un courtier sont répartis en deux groupes (6 et 8) en fonction de leur gravité. Enfin nous avons un groupe à prédominance matérielle et regroupant exactement les sinistres matériels distribués par un courtier, le 3. L'analyse de la charge ultime conforte cette interprétation par rapport à la gravité des sinistres. Nous n'indiquons pas la variable du type de contrat qui, comme pour le cas à 5 classes, n'a pas d'effet.

	Base complète	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Effectif	72723	22132	5029	13248	672	76	31 324	26	216
Corporel	0,631	0,481	0,612	0	0,823	0,921	1	0,962	0,981
Courtier	0,663	0	0,616	1	0,435	0,632	0,999	0,885	0,833
Litigieux	0,132	0,095	0,088	0,009	0,162	0,408	0,213	0,808	0,671
Délai de déclar. (mois)	2,891	1,474	13,147	2,132	40,445	113,447	1,499	5,308	1,718
Charge initiale (UM)	14 979	7 195	9 601	2 206	15 260	26 895	18 900	3 173 562	767 250
Durée de vie (années)	2,179	1,485	2,516	1,33	3,204	2,062	2,988	6,714	5,712
Charge ultime (UM)	5 287	3 140	5 543	1 020	8 912	3 768	7 522	521 654	219 771

TABLE 12 – Statistiques descriptives des 8 clusters

3.4.4 Stabilité des classes

Nous avons à ce stade deux possibilités quant au choix du nombre de classes. Un élément à prendre en compte est la stabilité de ces classes dans le temps, lorsque la base évolue. Intuitivement, si le nombre de sinistre augmente en préservant la structure de la base, il est souhaitable que les classes obtenues par un *k-means* sur cette nouvelle base soient proches des anciennes au niveau de l'interprétation. Afin de réaliser un *backtesting*, nous devons répéter le travail de classification sur la base en vision 2018. A cette occasion, nous étudions la stabilité des classes.

Les indicateurs utilisés pour le choix de k ont sensiblement la même allure, ce qui nous conduit encore à deux classifications : 5 et 8 classes. Nous calculons les mêmes statistiques descriptives que précédemment pour interpréter les résultats. Les détails figurent en annexe. La classification à 5 groupes fournit la même typologie qu'avant : un *cluster* de graves, deux de tardifs, un de sinistres matériels amiables. En revanche, le modèle à 8 classes ne produit plus les mêmes groupes. Nous avons cette fois un groupe de sinistres grave et un de très graves, deux de tardifs au lieu de trois et d'autres groupes moins interprétables. La robustesse des classes à de légères variations dans la structure de la base est une qualité intéressante pour une application dans un cadre réel, qui pourrait voir ce traitement lancé plusieurs années de suite. La stabilité dans le temps est donc un fort avantage pour le modèle à 5 *clusters*.

3.4.5 Développement des sinistres classifiés et cas particulier des tardifs

Le développement des IBNER par classe est similaire à ce qui a été fait dans le cas de la classification déterministe : tout le programme est lancé groupe par groupe au lieu d'être lancé sur la base complète. En particulier, l'estimateur de Kaplan-Meier de la durée de vie est calculé sur chaque groupe.

La simulation des IBNYR est modifiée. Détaillons les trois premières étapes : la simulation du délai de déclaration, l'attribution à un *cluster* et le tirage d'une catégorie.

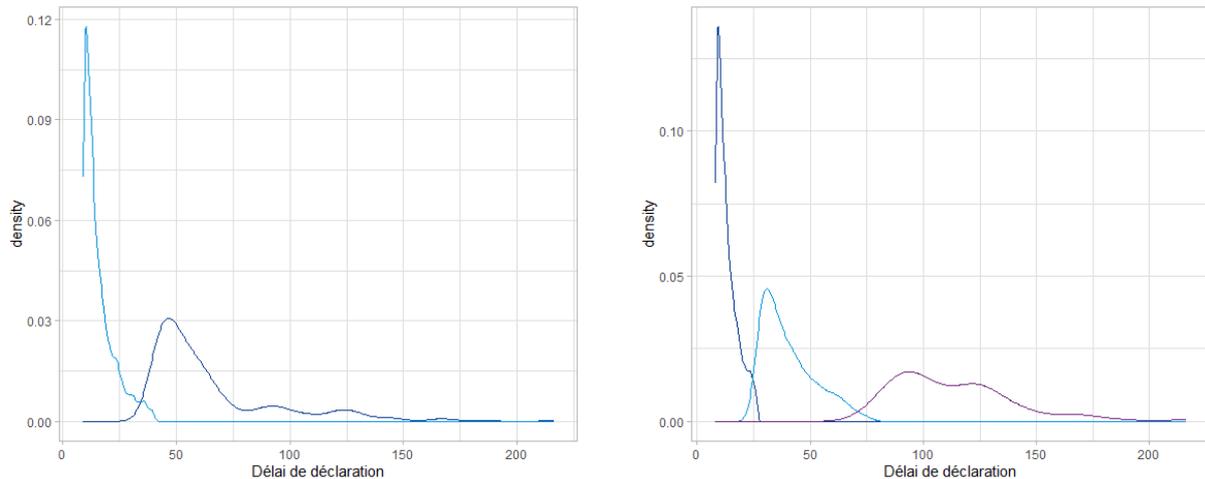
Simulation du délai de déclaration

Nous conservons la méthode de simulation du nombre de sinistre tardifs par bootstrap sur un triangle de nombre. Cette méthode fournit pour chaque sinistre une année d'ouverture et une année de survenance, mais nous avons maintenant besoin de plus de finesse pour obtenir un délai de déclaration en mois. Simplement convertir cet écart d'années en mois donnerait des délais supérieurs à 12 mois, or les analyses précédentes ont montré qu'un grand nombre de sinistres tardifs n'ont que deux ou trois mois de délai de déclaration. Nous procédons donc à un tirage du mois de survenance et du mois de déclaration.

Détaillons la méthode pour la survenance, et elle sera identique pour la déclaration. Les sinistres tardifs surviennent majoritairement en fin d'année (début d'année pour la déclaration) à cause de la clôture d'exercice en décembre que nous avons déjà évoquée. Pour chaque sinistre tardif de notre base, nous ajoutons son mois de survenance à une liste. Ainsi, cette liste correspond à la distribution des mois de survenance et il suffit d'y tirer uniformément pour simuler les dates des tardifs simulés. Grâce à cette méthode, nous disposons d'un délai de déclaration en mois.

Affectation du sinistre à un *cluster*

Lors de la phase de classification, nous avons identifié des *clusters* de tardifs. L'analyse de la distribution des délais de déclaration est présentée en figure 15. Notons d'abord que dans les deux cas, le groupe le moins tardif comporte une proportion conséquente de sinistres de moins d'un an, ce qui justifie l'utilité de la simulation d'un délai de déclaration en mois. Il apparaît ensuite que ces *clusters* sont bien disjoints, ce qui permet de leur affecter des sinistres sans ambiguïté. Ainsi, les abscisses des intersections des courbes constituent les limites des *clusters*. Nous attribuons à chaque sinistre le *cluster* dans lequel son délai de déclaration le place.



5 clusters, 2 clusters tardifs

8 clusters, 3 clusters tardifs

FIGURE 15 – Répartition du délai de déclaration dans les clusters tardifs

Attribution d'une catégorie au sinistre

Dans le modèle général, la catégorie du sinistre était tirée uniformément dans les catégories de l'historique. Or, les catégories associées aux sinistres tardifs ne sont pas distribuées comme celles du reste de la base, comme l'indique le tableau de statistiques descriptives par *cluster*. Pour tenir compte de cette dépendance, nous tirons la catégorie conditionnellement au *cluster*, c'est-à-dire en réalisant un tirage dans les catégories du *cluster* correspondant au sinistre.

3.4.6 Résultats du modèle

A présent, nous lançons les modèles afin d'analyser les résultats. Nous allons mener une étude à un an avec un calcul d'indicateurs d'erreurs, et une analyse à l'ultime. Le *k-means* à 5 classes produit des *clusters* stables lorsque la base change (vision 2018 ou vision 2019), c'est donc celui que nous retenons. Dans chaque cas, nous lancerons le modèle additif et le multiplicatif. Les résultats les plus pertinents seront présentés en détails.

Analyse à horizon un an

Pour le *backtesting*, nous partons de la base en vision 2018, d'abord divisée en 5 *clusters*. Nous simulons un an de développement pour les IBNER et calculons les indicateurs d'erreur pour chaque classe. Afin de comparer, nous lançons aussi une simulation sur la base complète (modèle sans classification) pour laquelle nous agrégeons les indicateurs d'erreur par classe. Les résultats sont présentés dans le tableau 13 et nous ajoutons l'interprétation des classes que nous avons faite à partir des statistiques descriptives.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Global
Interprétation	matériel amiable	-	tardif	grave	très tardif	-
Sinistres développés	1 604	6 122	375	31	49	8 181
RMSE	4 539	70 538	29 777	424 461	431 729	76 310
MAE	970,75	11 790	7 978	169 641	69 613	10 438
RMSE sans classif.	11 013	71 767	50 500	181 741	432 284	73 109
MAE sans classif.	2 022	11 655	9 680	89 985	71 500	10 331

TABLE 13 – Indicateurs d’erreur sur les réserves (UM), modèle additif

Le *cluster 2*, qui regroupe les sinistres les plus communs, sans caractéristique particulière, ne présente pas d’amélioration ni de dégradation. Cette classe a l’effectif le plus élevé : 6 122 sinistres non clos y ont été développés. Comme avec la classification déterministe, nous n’observons pas de gain sur la classe majoritaire, bien que beaucoup de sinistres aux caractéristiques marquées en soient exclus, puisqu’ils sont dans les autres classes. En revanche, le gain est net sur les classes 1, 3 et 5 correspondant respectivement aux sinistres matériels amiables, tardifs et très tardifs. Le RMSE et le MAE diminuent fortement par rapport à ceux obtenus avec un modèle sans classification, ce qui est très satisfaisant.

Cependant, nous trouvons une très forte dégradation sur le groupe des sinistres graves. Celle-ci peut toutefois s’expliquer : la classe contient peu de sinistres et les paiements pour les graves sont très volatils. Certaines années de développement n’enregistrent aucun paiement alors que d’autres donnent lieu à des règlements de plusieurs centaines de milliers, voire millions d’euros. Si notre modèle tire un règlement très élevé pour un sinistre qui n’a en réalité pas occasionné de règlements en 2019, l’erreur qui en résulte est très importante. Nous voyons ici une des limites de l’évaluation de l’erreur ligne à ligne dans un modèle pour lequel l’effet de moyenne compte. Or, une conséquence de notre classification est d’isoler dans une petite classe les sinistres qui donnent lieu au plus d’erreur avec ce modèle. En pratique dans une compagnie d’assurance, ces sinistres que nous avons identifiés n’auraient pas été développés avec un modèle, mais seulement examinés par des gestionnaires pour un traitement au cas par cas. Précisément, il est possible de retenir simplement la charge prévue par les gestionnaires pour l’ultime. La charge est utilisée dans notre second modèle, le multiplicatif.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Global
Interprétation	matériel amiable	-	tardif	grave	très tardif	-
RMSE	9 804	148 669	138 592	1 189 614	133 495	151 630
MAE	2 124	26 784	29 159	642 146	41 655	24 479
RMSE sans classif.	11 640	148 703	126 347	1 116 463	135 615	149 280
MAE sans classif.	3 311	26 699	26 293	602 544	34 210	24 322

TABLE 14 – Indicateurs d’erreur sur les charges, modèle multiplicatif

Nous présentons les résultats de la même étude sur le modèle multiplicatif, qui utilise les charges et non les règlements, dans le tableau 14. Les observations sont qualitativement les mêmes, mais quantitativement moins marquées. L’erreur est quasiment identique sur la classe majoritaire. Les sinistres matériels amiables, tardifs et très tardifs voient leur er-

reur diminuer par rapport au modèle sans classification. Cependant, le changement relatif est plus faible qu’avec le modèle additif.

Du côté des sinistres graves, l’erreur augmente d’un facteur environ 1,07, alors qu’elle avait plus que doublé avec le modèle additif. L’erreur dans la simulation des règlements survenait lorsque le modèle simulait un règlement n’ayant aucun rapport avec celui réellement survenu. Pour les charges, ce phénomène est limité de deux manières. D’abord, il y a une plus grande stabilité temporelle de la charge au niveau du sinistre. Celle-ci reste constante tant qu’aucune réévaluation n’a lieu, alors que les règlements peuvent être nuls pendant plusieurs années, et ponctuellement importants. Ainsi, nous avons beaucoup de facteurs de développement à 1 parmi ceux que nous tirons, et parmi ceux réellement observés en 2017. Ensuite, nous avons une copule qui doit éviter le tirage de facteurs aberrants, mais nous reviendrons sur ce point dans l’analyse à l’ultime.

Pour les deux modèles, nous constatons donc une amélioration sur 3 classes et une stagnation sur la classe majoritaire, contrebalancées par une dégradation de l’erreur sur les sinistres graves. Au niveau global, les erreurs à un an sont donc proches entre les modèles avec et sans classification. Cependant, nous avons déjà pointé les limites de l’analyse à un an, nous complétons donc avec le développement à l’ultime.

Analyse à l’ultime

Cette fois nous partons de la base en vision 2019 et simulons le développement à l’ultime des IBNER et IBNYR. Encore une fois nous avons d’une part les modèles classifiés qui retournent des résultats par *cluster* et d’autre part les modèles non classifiés pour lesquels nous agrégeons les résultats par *cluster* pour permettre la comparaison. Nous ajoutons aussi un modèle de Chain-Ladder par classe. Les charges ultimes sont présentées dans le tableau 15. Pour le modèle additif, la charge est la somme des réserves et des règlements déjà effectués et connus dans la base.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	total
Interprétation	grave	très tardif	matériel amiable	-	tardif	-
Effectif	55	350	25 128	42 622	4 568	72 723
Additif avec classif.	51 528	12 340	29 468	695 610	44 208	833 156
Additif sans classif.	34 268	8 269	51 332	675 413	50 382	819 667
Chain-Ladder paiements	50 810	6 926	29 154	691 130	42 340	820 363
Multiplicatif avec classif.	119 322	11 132	27 836	788 822	55 138	1 002 252
Multiplicatif sans classif.	112 996	10 360	41 268	771 264	54 293	990 182

TABLE 15 – Charges ultimes, en kUM, par classe

A la fois pour le modèle additif et le multiplicatif, nous observons que la classification modifie les réserves, par rapport au modèle sans classification, dans le sens suggéré par notre interprétation du *cluster* : pour les sinistres graves les ultimes sont plus élevés, pour les matériels amiables ils sont plus faibles, etc. Sur le *cluster* 4, qui comporte les sinistres communs, la différence est faible, avec toutefois une légère augmentation pour la classification. Ceci peut s’expliquer par le fait que nous avons isolé un bon nombre de sinistres de faible montant dans la classe 3, ce qui rehausse mécaniquement les montants tirés dans la classe 4. Toutefois, d’un point de vue d’expert, il est difficile de déterminer si cette augmentation est une amélioration. Sur les autres classes en revanche, le modèle semble proposer des montants plus cohérents et conformes à l’attendu.

Revenons brièvement sur la copule du modèle multiplicatif. Elle avait été introduit pour empêcher des sinistres de montant élevés, qui ont normalement des facteurs de développement proches de 1, de recevoir des facteurs élevés parfois rencontrés sur les petits sinistres. Or, notre classification pourrait contribuer à limiter ce phénomène. Nous lançons donc un modèle multiplicatif sans copule sur la base classifiée, et nous constatons que le problème est toujours là. Son intensité est moindre qu'en l'absence de classification mais les résultats restent aberrants. La séparation en 5 classes n'est pas assez fine pour empêcher des sinistres de montant assez élevés de rencontrer des facteurs de développement élevés. Cela n'était pas le but initial de notre classification mais aurait pu être un effet collatéral satisfaisant. Nous continuons alors d'utiliser la copule.

La classification étant réalisée, la question de l'intérêt d'un modèle de développement individuel en aval pourrait se poser. En effet, peut-être qu'un modèle classique, comme Chain-Ladder, lancé sur chaque classe, suffirait à bénéficier des effets de la classification sans complexifier la démarche ensuite. Afin de vérifier ceci, nous avons développé un triangle de paiements pour chaque classe avec la méthode de Chain-Ladder. Les résultats sont intégrés dans le tableau 15.

Les ultimes obtenus sont proches de ceux de notre modèle additif pour toutes les classes sauf celle des sinistres très tardifs. En effet, ces sinistres ont plus de 3 ans de délai de déclaration. Alors que notre modèle simule leur existence, le triangle agrégé correspondant a ses 3 premières lignes vides, ce qui crée un déficit de charge. La capacité de Chain-Ladder à prendre en compte implicitement les tardifs est ici mise en défaut. En revanche, sur une classe nombreuse avec des sinistres communs, comme la quatrième, notre modèle de provisionnement individuel n'apporte pas un avantage marqué pour la simulation de la moyenne (mais rappelons qu'il fournit néanmoins une distribution complète des ultimes).

Observons maintenant les résultats globaux. La classification fait augmenter la charge ultime prédite par le modèle, qu'il soit additif ou multiplicatif. L'augmentation est toutefois plus importante, en valeur et en proportion, sur l'additif. C'est un effet intéressant puisque les analyses menées par [PRIMEL 2021] ont montré que ce modèle (sans classification) sous-estimait les ultimes. Toutefois, cela ne suffit pas pour le réconcilier avec le multiplicatif. Ainsi, cette classification par *k-means* a plusieurs points positifs au niveau de l'interprétabilité des classes et du *backtesting*. A l'ultime, ses effets peuvent apporter un avantage dans une optique de provisionnement, mais le choix du modèle, additif ou multiplicatif, semble rester déterminant.

3.5 Arbre de classification

Dans la partie précédente nous avons utilisé une méthode de classification non supervisée. L'avantage principal de ce choix est qu'il ne nécessite pas de trouver une variable de supervision adaptée. Cependant, l'inconvénient correspondant est que les groupes obtenus pourraient n'avoir aucun lien avec le développement des sinistres, puisque celui-ci n'intervient pas dans la classification. Nous avons constaté que ce n'est pas le cas, d'une part en reliant la composition des groupes aux statistiques descriptives, et d'autre part en observant les résultats des modèles. Toutefois, une classification supervisée garantirait la pertinence des classes par rapport à notre problème, sous réserve de choisir une bonne variable. Nous allons donc exposer dans cette partie la mise en place de la deuxième grande

méthode de classification et l'analyser.

3.5.1 Choix du modèle et de la variable de supervision

Choix de la variable

La variable que nous retiendrons comme variable à prédire par le modèle de classification doit répondre à deux critères : être disponible pour un nombre d'observations suffisant et avoir une influence sur le développement des sinistres.

Le premier critère assure que la base sur laquelle nous entraînons le modèle de classification sera être de taille conséquente. Or, plus nous souhaitons avoir une information précise sur le développement des sinistres, plus nous devons nous restreindre à des sinistres développés plusieurs années. Il faut donc choisir une variable rapidement observable. Les variables disponibles uniquement à la clôture du sinistre comme la charge ultime ou la durée de vie sont disponibles pour une portion restreinte de la base. De plus, comme les sinistres longs sont mécaniquement moins clos, cette portion est biaisée en faveur des sinistres courts. La charge après un an de développement est un très bon candidat. puisqu'elle est renseignée dans 79% de la base.

Elle satisfait aussi le second critère puisque malgré la longueur de la branche, beaucoup de sinistres ont connu une évolution significative après un an, et certains sont même clos (la durée de vie moyenne des matériels est d'à peine plus d'un an). La question de superviser plutôt par la charge ultime peut se poser à ce stade. Cependant, le but de notre classification n'est pas de la prédire puisque c'est le rôle des modèles de développement. Faire ce choix conduirait à une redondance et sortirait de notre cadre de travail. Nous retenons donc la charge à un an. Comme nous connaissons la charge initiale, l'évolution de la charge après un an est un bon indicateur d'un sinistre qui va se dérouler rapidement et sans évolution (si la charge à un an est la même que l'initiale) ou d'un sinistre qui sera potentiellement plus compliqué car il a subi une réévaluation dans sa première année.

La variable étant déterminée, il faut ensuite la traiter pour en faire des classes, c'est-à-dire définir des plages de valeur. Il est couramment supposé en actuariat que les charges de sinistres suivent une loi log-normale. Nous ne cherchons pas à appliquer cette hypothèse sur nos données, mais effectivement l'histogramme du logarithme de la charge à un an dessine approximativement une cloche. Cela suggère un découpage logarithmique. Précisons que ne nous ne cherchons pas à avoir des classes d'effectif égal, ce qui conduirait à des intervalles petits autour des valeurs les plus fréquentes, mais à constituer des classes de sinistre semblables. Nous isolons par ailleurs les sinistres de charge nulle dans une classe à part, et obtenons le découpage présenté en tableau 16. Un grand nombre de sinistres

Classe	0]0, 100[[100, 1000[[1000, 10 ⁴ [[10 ⁴ , 10 ⁵ [[10 ⁵ , 10 ⁶ [[10 ⁶ , 10 ⁷ [
Effectif	15081	429	2187	17963	20908	1071	68

TABLE 16 – Classes retenues

ont une charge à 1 an nulle, ce qui justifie d'en faire une classe à part. Pour ceux-ci, les gestionnaires estiment que l'assurance n'aura pas d'indemnisation à verser, par exemple parce que son assuré n'est pas responsable. De l'autre côté, la classe des sinistres avec la plus grosse charge compte un effectif assez petit, mais leur gravité extrême nécessite de les séparer. En particulier, dans un cadre de provisionnement classique, ces sinistres sont

toujours isolés par les actuaires. Nous disposons maintenant d'une variable de supervision exploitable.

Pour les variables explicatives de notre modèle, nous reprenons celles utilisées avec le *k-means* qui conservent toute leur pertinence ici.

Mise en place du premier modèle

Le premier modèle que nous implémentons est un arbre de classification. Il a l'avantage de la simplicité en première approche, c'est-à-dire sans rentrer dans le détail des paramètres, et de l'interprétabilité. Nous procédons en deux temps.

D'abord, nous entraînons un arbre avec une très grande profondeur sur toute la base. Les arbres de la bibliothèque *Rpart* [THERNEAU et ATKINSON 2019] possèdent une fonction de validation croisée intégrée qui à chaque étape de la croissance de l'arbre calcule une erreur relative. Précisons cette notion. L'évaluation d'un modèle de *machine learning* se fait toujours sur un ensemble de test, à partir duquel le modèle n'a pas été entraîné. La validation croisée à *k* blocs (*k-fold cross validation*) est une méthode d'évaluation des modèles d'apprentissage automatique qui divise les données en *k* groupes. Sur chaque groupe, le modèle est testé après avoir été entraîné sur les *k-1* groupes restants. Ceci permet d'avoir plusieurs ensembles de test différents à partir d'une quantité de données limitée. Ensuite, une méthode courante pour le choix de la complexité des modèles est de regarder l'erreur de test en fonction de la complexité. Au début l'erreur du modèle sur les données qui n'ont pas servi à l'entraînement diminue avec la complexité. A partir d'un seuil cependant, le modèle devient trop spécifique aux données d'entraînement (surapprentissage) et sa performance se détériore sur les données inconnues, qui sont différentes. Puisque le modèle doit avoir une bonne performance sur les données inconnues, le niveau de complexité retenu est ce seuil. Le graphe correspondant pour notre premier arbre est présenté en figure 16. L'axe horizontal du bas indique le paramètre de complexité correspondant au paramètre α du *pruning*. Nous raisonnerons cependant avec la profondeur qui est plus directement reliée à la structure de l'arbre.

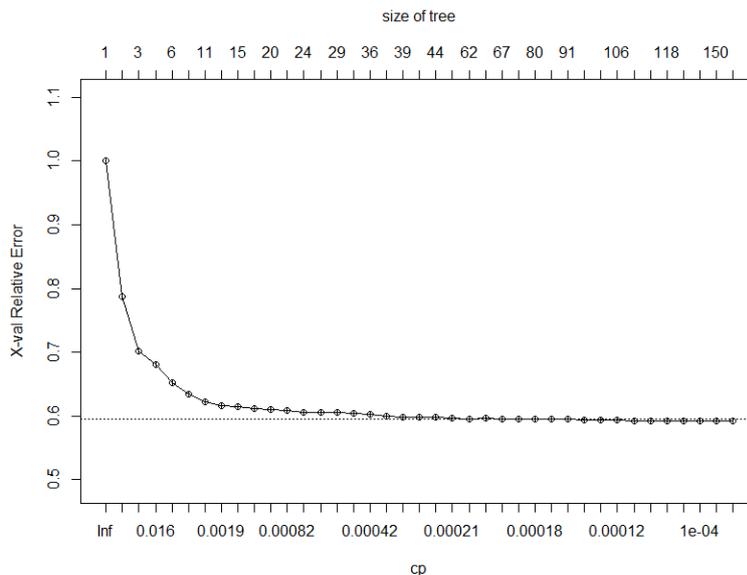


FIGURE 16 – Erreur de validation croisée en fonction de la profondeur de l'arbre

Comme nous venons de l'expliquer, nous attendions une remontée de la courbe à partir d'un certain seuil, or elle reste ici constante. Ceci peut s'expliquer par le nombre assez restreint de variables que nous utilisons, dont plusieurs binaires, qui ne permet pas au modèle d'atteindre une grande complexité. Ainsi, un apprentissage limité restreint mécaniquement le surapprentissage. La charge à l'ouverture du sinistres est un très bon prédicteur de la charge à 1 an mais il est difficile d'ajouter beaucoup de complexité avec les autres variables.

Dans un deuxième temps, nous entraînons un nouvel arbre en limitant la profondeur compte tenu du graphe précédent. Toutefois, nous composons avec une seconde contrainte puisque nous voulons que notre arbre prédise toutes les classes. Elles n'apparaissent toutes qu'à partir d'une profondeur de 5. Nous retenons donc cette valeur qui fournit de bons résultats avec une complexité permettant encore l'interprétation. Nous obtenons l'arbre de la figure 17.

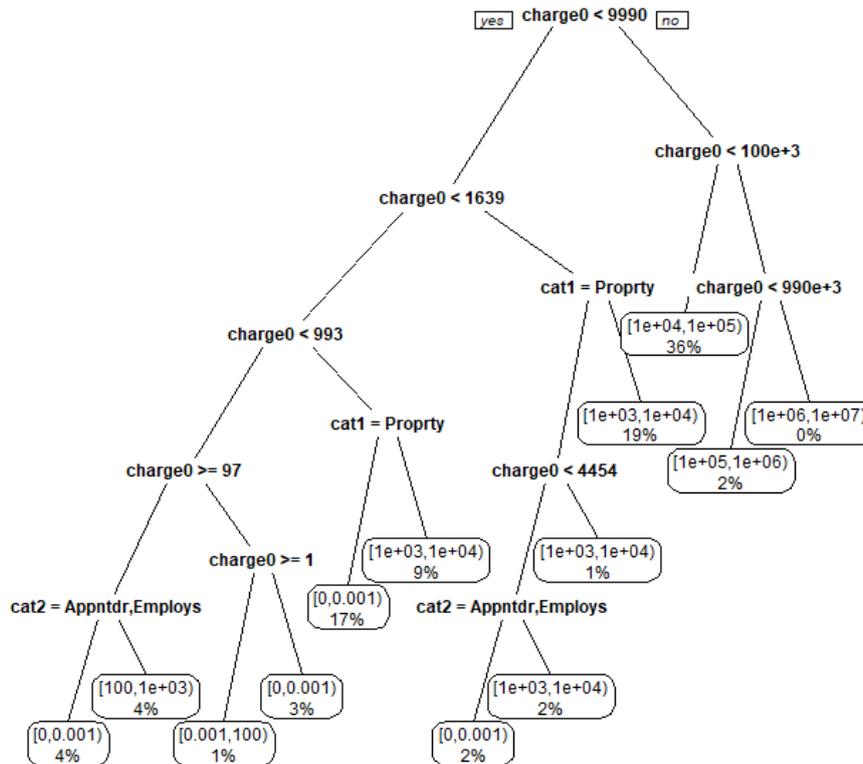


FIGURE 17 – Arbre de classification

A chaque nœud est présentée une condition logique. La branche de gauche vérifie cette condition et celle de droite ne la vérifie pas. Pour le nom des variables : **charge0** indique la charge initiale, **cat1 = Property** signifie que le sinistre est matériel, **cat2 = Appntdr, Employs** signifie que le contrat n'a pas été souscrit auprès d'un courtier. Chaque branche aboutit à une classe présentée dans un cadre avec le pourcentage des observations totales qui tombent dans cette case.

Nous remarquons immédiatement l'omniprésence dans l'arbre de la charge initiale, ce qui confirme notre supposition qu'elle prédit très bien la charge à un an. La première autre

variable à faire une distinction est matériel/corporel, ce qui marque aussi son importance, que nous avons déjà notée lors de la classification déterministe. Quant à la souscription par un courtier, nous n'avons pas noté de différence significative causée par cette variable, et son effet n'est pas immédiatement explicable, mais elle pourrait être un proxy pour une information inconnue dans la base comme des clauses négociées par le courtier.

Second modèle et comparaison

Après avoir ajusté un arbre de classification sur nos données, se pose la question de passer à une forêt aléatoire. L'avantage principal de ce type de modèle, par rapport à un arbre seul, est une plus grande précision et une meilleure stabilité, au prix de la perte de la visualisation et de l'interprétation associée.

Nous optimisons les hyperparamètres avec une validation croisée et obtenons finalement une forêt à 100 arbres. Maintenant, nous cherchons à comparer ce modèle à l'arbre précédent à l'aide d'une métrique. En classification, une métrique usuelle est l'AUC (*area under the curve*) qui mesure l'aire sous la courbe ROC (*receiver operating characteristic*). Son avantage principal est qu'elle ne nécessite pas de définir le coût de chaque type d'erreur de classification. Pour illustrer ce problème avec notre cas, imaginons un sinistre dont la charge 1 est 1500 €. Nous aurons une erreur plus grande sur la prédiction de l'ultime en le plaçant dans la classe $[10^5, 10^6[$ que dans la classe $[10^4, 10^5[$. Pourtant, aucune de ces deux classes n'est bonne. Cependant, quantifier ces erreurs de classification est délicat. L'AUC étant initialement définie pour les problèmes à deux classes, nous utilisons une extension multi-classe [HAND et TILL 2004] implémentée dans la bibliothèque pROC [ROBIN et al. 2011]. L'AUC est de 0,84 pour l'arbre seul et de 0,85 pour la forêt aléatoire. Ce second modèle apporte donc un gain, mais assez réduit. Comme il s'accompagne de la perte de l'interprétabilité par un graphique, le gain ne nous semble pas justifier le choix du second modèle et nous conservons finalement l'arbre de classification. Notons par ailleurs que le score obtenu est assez bon (un classificateur aléatoire a une AUC de 0,5).

Comme nous l'avons fait avec le *k-means*, nous calculons des statistiques descriptives par classe sur la proportion de chaque variable, le délai de déclaration et la charge initiale moyenne. Le détail est présenté en annexe. La charge initiale, bon prédicteur de la charge à un an, est évidemment cohérente avec l'intervalle de charge à un an définissant chaque classe. La proportion de sinistres litigieux et la proportion de corporels augmentent avec la charge, ce qui est un effet attendu pour ces deux variables. Le taux de courtiers augmente aussi légèrement. Nous n'observons pas de prédominance nette de variables dans certaines classe par rapport à d'autres. Cet effet est bien sûr attendu puisque ce modèle ne repose pas sur la proximité des variables. En cela, ce type de classification s'éloigne plus de ce qu'une équipe de provisionnement pourrait mettre en place que le *k-means*. Ici en effet, un actuaire distinguerait probablement les sinistres graves, puis constituerait probablement un triangle de sinistres corporels et un triangle de matériels. Notre arbre n'utilise pas directement ces critères.

3.5.2 Simulation à partir de la classification et résultats

Avec la classification par k-moyennes nous avons simulé les IBNER en lançant le modèle de simulation des IBNER sur chaque classe séparément. En revanche, une démarche spécifique pour les IBNYR avait été mise en place car certaines classes étaient dédiées uniquement aux sinistres tardifs. Ici encore, se pose la question de l'affectation des sinistres non encore présents dans la base à une classe. Les variables qui permettent la classification ne sont pas renseignées pour ces sinistres. Nous pourrions les simuler toutes puis faire passer le sinistre dans l'arbre pour obtenir sa place. Nous choisissons la solution inverse, plus simple, d'affecter directement des classes aux sinistres. Chaque sinistre tardif est initialement simulé avec une date de déclaration et une date d'ouverture. Comme les délais de déclaration moyens sont presque les mêmes dans chaque classe, nous ne tenons pas compte de ces dates pour l'affectation à une classe. L'attribution se fait par un tirage uniforme dans la distribution des classes des sinistres tardifs de notre historique. Ensuite, le développement se fait séparément sur chaque classe. En particulier, la charge initiale d'un sinistre est tirée dans l'historique de sa classe.

Présentons maintenant les résultats par catégorie, dans le tableau 17. Les charges pour le modèle additif correspondent à la somme des paiements déjà effectués et des réserves. Pour nous affranchir de la taille des groupes et comparer avec la charge à un an qui définit la classe, nous ajoutons aussi les charges moyennes, c'est-à-dire divisées par l'effectif de la classe. Les sinistres ayant une charge à 1 an nulle sont particuliers, mais il est naturel que la charge ultime ne soit pas nulle. Pour les autres classes, la charge ultime moyenne est en adéquation avec la charge à 1 an prédite, mais pas toujours dans les bornes.

	0]0, 100[[100, 1000[[1000, 10 ⁴ [[10 ⁴ , 10 ⁵ [[10 ⁵ , 10 ⁶ [[10 ⁶ , 10 ⁷ [total
Additif	29 935	574	5 385	137 620	514 355	161 212	61 837	910 918
Multiplicatif	42 358	1 011	4 571	145 375	490 612	187 831	153 148	1 024 906
Additif moyenne	1,267	0,298	1,365	6,880	23,199	147,945	883,381	-
Multiplicatif moyenne	1,793	0,526	1,181	7,279	22,129	172,480	2187,830	-

TABLE 17 – Charges ultimes par classe (kUM)

Nous pouvons ensuite comparer les deux modèles de développement entre eux. Le modèle multiplicatif prédit plus que l'additif sur la majorité des classes, ce qui rejoint nos observations précédentes. Cependant, l'écart est très marqué sur les sinistres les plus graves. Cette classe ayant un petit effectif, il est facile d'observer son historique à la main pour tenter d'expliquer l'écart entre les modèles. Rappelons le principe de la copule pour le tirage des facteurs de développement. Le sinistre à simuler a une charge et la copule tire dans l'historique des facteurs associés à des charges proches de cette charge. Ainsi, si nous avons dans notre historique un sinistre de charge déjà élevée qui a connu un fort développement, avec un facteur élevé, la copule aura tendance lors des simulations à affecter ce facteur important à des sinistres de charge déjà élevée. Un tel phénomène expliquerait les prévisions ultimes très hautes. Cependant, ce n'est pas le cas : les facteurs les plus élevés correspondent à des charges inférieures à 100 000 UM, ce qui est petit dans cette classe. La copule ne provoque donc pas d'explosion des charges. Réciproquement, les paiements incrémentaux les plus élevés (1 700 000 UM au premier développement, 1 300 000 UM au deuxième...) sont associés à des facteurs proches de 1. Ils correspondent donc à des versements anticipés dans la charge. Encore une fois, la simulation des ultimes

par un modèle pour les sinistres les plus graves n'est pas forcément adaptée. Ceci souligne d'autant plus l'importance de les isoler par une classification *a priori*.

Toutefois, afin de faire un choix entre nos deux modèles sur cette classe la plus grave, nous calculons sur les sinistres clos de l'historique de cette classe la charge ultime moyenne et constatons qu'elle est très proche de celle obtenue par le modèle additif, ce qui semble disqualifier le modèle multiplicatif dans ce cas de figure précis.

Pour la classification supervisée, nous avons donc retenu un arbre. Malgré une précision légèrement inférieure, nous préférons conserver la simplicité et l'interprétabilité que n'a pas la forêt aléatoire, puisque nous cherchons un modèle qui soit réellement utilisable et pleinement justifiable. Quant aux simulations d'ultime qui en découlent, elles sont conformes à ce que nous pouvons attendre sur chaque classe sauf celle des sinistres les plus graves. Au niveau global, cette classification ne rapproche pas les deux modèles.

3.6 Évaluation des résultats

Nous avons présenté les différentes méthodes de classification mises en place et le développement des sinistres par les modèles adaptés. Nous évaluons maintenant certains points qui restent à examiner et comparons à cette occasion nos modèles.

3.6.1 Stabilité des modèles

Un point essentiel à examiner est la réaction des modèles de développement face à l'ajout d'information supplémentaire sur les mêmes sinistres. Si les résultats des modèles varient peu lorsque nous ajoutons des données, cela peut s'interpréter comme une anticipation correcte de la part du modèle. À l'inverse, une trop grande stabilité correspondrait à une insensibilité aux évolutions des données.

Pour évaluer ce point, nous utilisons un procédé proche du *backtesting*. Nous partons d'une base en vision 2016. Nous appliquons nos méthodes de classification puis lançons les modèles de développement afin d'obtenir une distribution d'ultimes. Ensuite, sur le même périmètre de sinistres, c'est-à-dire ceux ouverts en 2016 ou avant, nous ajoutons l'information de 2017, c'est-à-dire un an de développement, et relançons la simulation des ultimes. Nous faisons de même avec l'information de 2018 et 2019 et comparons les distributions.

Modèles sans classification

Une analyse de ce type avait été conduite par [PRIMEL 2021] sur le modèle additif uniquement. Elle avait montré une bonne stabilité du modèle, et notamment une sensibilité aux sinistres extrêmes plus faible que pour le modèle de Chain-Ladder. Afin de fournir une comparaison avec les modèles précédés de classification, nous reproduisons l'exercice sur notre base de données.

Nous menons ensuite l'examen, sur le modèle multiplicatif. Nous incluons aussi les résultats du développement de triangles de charges par le modèle de Chain-Ladder. Dans ce cas, l'ajout d'information sur une année de développement revient simplement à ajouter une diagonale au triangle. Les résultats sont présentés en figure 18.

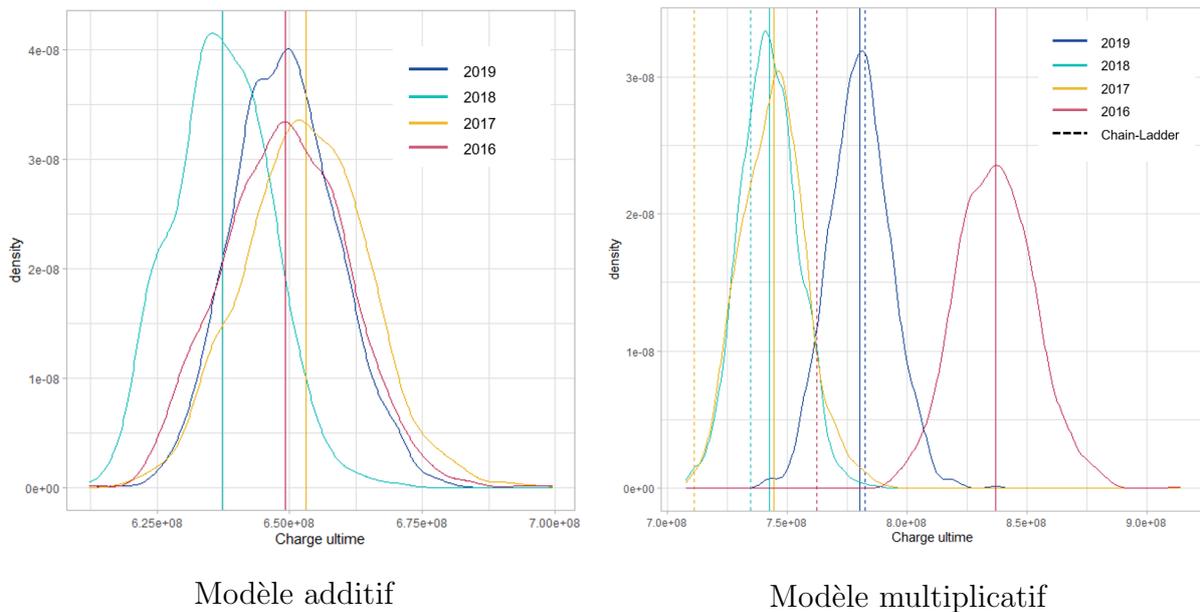


FIGURE 18 – Stabilité des modèles, charges ultimes (UM)

Les lignes verticales en pointillées correspondent à l’ultime prédit par Chain-Ladder pour l’année indiquée par la couleur, et les traits pleins verticaux représentent la moyenne de la distribution correspondante. Les *best-estimates* s’étendent sur une plage d’environ 100 millions d’UM, ce qui est conséquent. Les ultimes ne sont pas ordonnés par rapport aux années d’information, il ne semble donc pas y avoir un effet structurel du modèle multiplicatif qui prédirait plus ou moins de charge simplement en fonction de l’avancement des informations, ce qui équivaldrait à une sous-estimation ou sur-estimation systématique des charges. Le modèle additif apparaît plus stable, avec des charges moyennes prédites très groupées.

Quelques spécificités du modèle multiplicatif peuvent être analysées. Nous notons que les prédictions pour l’année 2019 sont sensiblement plus élevées, d’environ 40 millions d’UM. Nous avons déjà observé cette année en détail et constaté un certain nombre de sinistres importants qui affectent le modèle individuel et le triangle agrégé. Plus globalement, en observant le triangle de coefficients de développement, une partie de la diagonale 2019 est plus élevée. Cela contribue à augmenter la valeur prédite à la fois par notre modèle et par Chain-Ladder. En revanche, avec l’information de 2017, la charge Chain-Ladder est nettement inférieure alors que notre charge n’est que très légèrement décalée. Encore une fois, nous nous reportons au triangle de coefficient de développement, qui permet d’expliquer le résultat de la méthode agrégée. Il apparaît que la diagonale associée à 2017 ne présente pas de tendance nette par rapport aux autres, mais nous notons des facteurs de développement plus bas, particulièrement par rapport aux années 2018 et 2019. L’extrait du triangle de facteurs est reproduit en figure 18 avec la diagonale 2017 en gras.

Année de survenance	Année de développement			
	10	11	12	13
2003	0,989	1,035	0,983	0,976
2004	0,980	0,990	0,993	0,993
2005	0,974	0,985	0,882	1,118
2006	0,941	0,996	0,998	
2007	1,016	1,073		
2008	0,997			

TABLE 18 – Extrait du triangle de coefficients de développement

Un point notable est que ces facteurs correspondent à des développements assez tardifs, ce qui a deux conséquences opposées suivant les modèles.

- Pour Chain-Ladder, dans une colonne sur la droite du triangle, beaucoup de lignes sont encore à développer. Ainsi, un facteur situé à cet endroit aura un impact fort puisqu’il affectera de nombreuses années de survenance. En nous plaçant en vision 2017, le triangle du tableau 18 s’arrête à la diagonale en gras. Dans ce cas, le facteur n°13 affecte toutes les survenances postérieures à 2005, soit 12 survenances sur un total de 17 (nous sommes sur le périmètre 2016). Ces quelques facteurs ont donc un effet à la baisse significatif sur le modèle de Chain-Ladder.
- Pour le modèle multiplicatif individuel, les sinistres ont une chance d’être affectés par un facteur de développement individuel seulement s’ils sont encore ouverts à ce niveau de développement. Rappelons que les facteurs individuels sont distincts des facteurs agrégés présentés ci-dessus, mais un facteur agrégé plus bas indique plusieurs facteurs individuels bas. Or, puisque ces facteurs bas correspondent à des développements tardifs, peu de sinistres les atteignent. L’effet des facteurs tardifs est donc au contraire limité.

Nous avons donc mis en évidence une force du modèle multiplicatif individuel : sa moindre sensibilité aux développements atypiques tardifs. Cependant, comme nous l’avons montré avec l’effet de l’année 2019, ce modèle prend en compte les tendances globales sur une année.

Modèles avec classification

Nous pouvons maintenant évaluer la stabilité des modèles lorsqu’ils sont précédés d’une classification, avec le même procédé. Précisons que la classification est faite à partir de la base en vision 2016 et qu’elle ne change pas lorsque l’information disponible augmente. En effet, puisque les classes sont établies à partir d’informations disponibles à l’ouverture du sinistre, ajouter des renseignements sur les développements ne peut pas les modifier. Les résultats sont présentés pour la classification *k-means* en figure 19 et pour l’arbre de classification en figure 20. Nous faisons encore figurer l’ultime prédit par Chain-Ladder sur un triangle de charge à titre de comparaison. En revanche, nous n’ajoutons pas l’équivalent sur un triangle de paiements du fait de sa pertinence très limitée, point que nous développerons ultérieurement.

Tout d’abord, nous remarquons la grande concentration des modèles additifs. Quelle que soit la classification, les ultimes moyens prédits ont moins de 30 millions d’UM d’écart alors que le modèle multiplicatif présente presque le triple. Dans le cas additif, les classi-

fications font dériver l'ultime issu de la vision 2017 et dans le cas multiplicatif toutes les visions sont globalement plus éloignées les unes des autres, mais l'ordre est conservé par rapport aux résultats sans classification. Ainsi, il semble que la constitution de groupes de sinistres amplifie les disparités entre les années de vision, tout en les maintenant du même ordre.

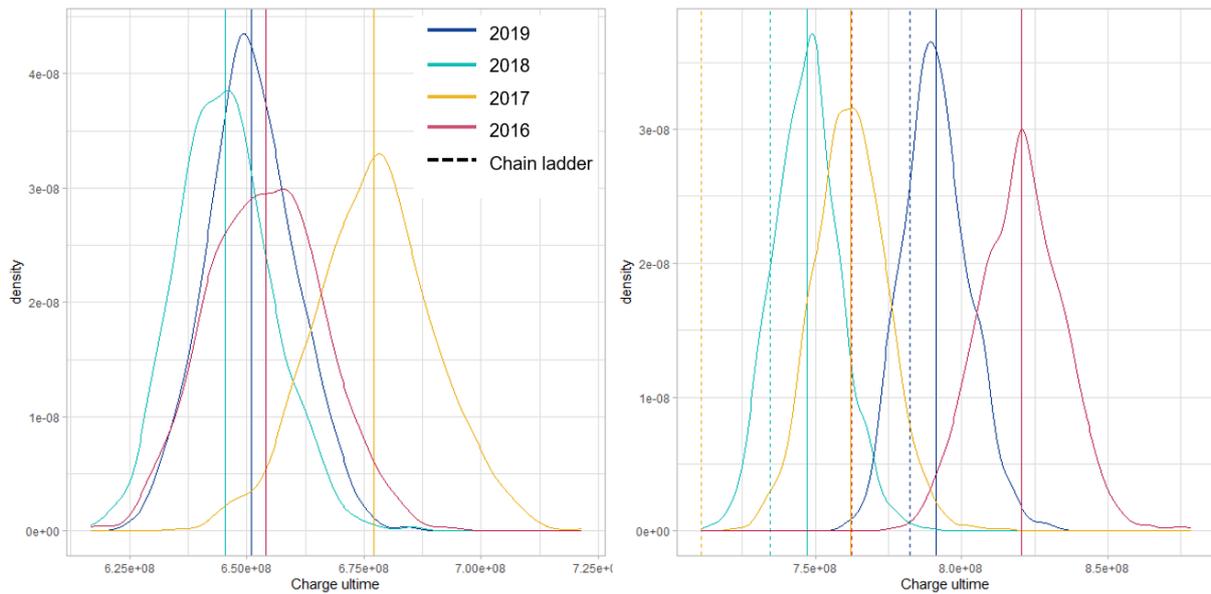


FIGURE 19 – Stabilité avec classification par k-moyennes, charges ultimes en UM (additif à gauche, multiplicatif à droite)

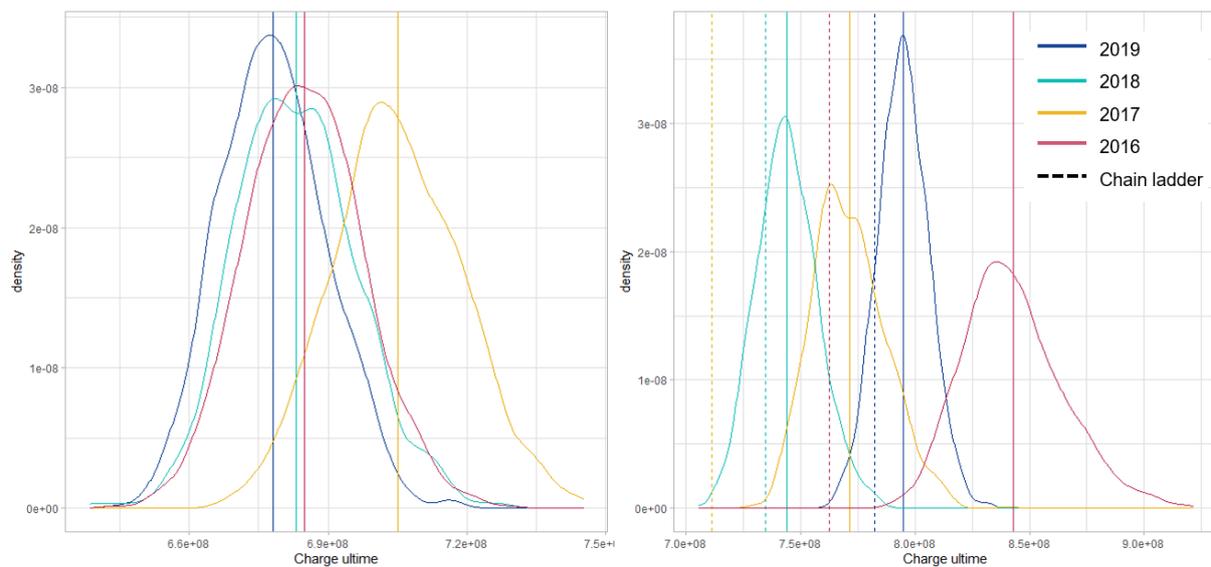


FIGURE 20 – Stabilité avec classification par arbre, charges ultimes en UM (additif à gauche, multiplicatif à droite)

Sur le plan de la stabilité, l'effet de la classification n'améliore pas nos modèles. Cependant, cette étude montre que la sensibilité à l'ajout d'information est un élément inévitable, puisque chaque année a ses spécificités. Même Chain-Ladder, le modèle de référence, produit des estimations à 71 millions d'UM d'écart en fonction des visions considérées. Cette variabilité est actée et justifie la pratique fréquente lors des audits d'analyses *boni/mali* qui consistent à comparer les charges ultimes estimées sur un périmètre fixe à deux visions différentes.

3.6.2 Avis d'experts

Nous avons fréquemment comparé les valeurs obtenues par nos modèles à des prévisions issues du développement de triangles agrégés par la méthode de Chain-Ladder, toujours appliquée sans aucun retraitement. Cependant, dans la mesure où nous cherchons un *best estimate*, il serait insuffisant de s'en contenter, d'une part parce que qu'aucun actuaire n'appliquerait réellement cette méthode naïvement et d'autre part parce qu'une multiplicité de valeurs serait plus informative pour placer nos résultats. Nous avons donc sollicité les actuaires provisionnement de KPMG afin qu'ils proposent chacun une valeur d'ultime à partir de leur jugement d'expert. Nous leur avons fourni un triangle de charges, un triangle de règlements sans inflation et un triangle de nombres issus de notre base, et avons spécifié la branche dont provenaient les données, c'est-à-dire la RC médicale. En ne leur imposant aucune contrainte, nous espérions, en plus des simples estimations ultimes, comparer les choix de modélisation individuels.

Nous collectons 14 résultats différents auprès des actuaires. Avant de regarder les valeurs, attardons-nous sur les choix méthodologiques retenus. Une fois toutes les projections faites, ceux-ci ont fait l'objet d'un échange entre les participants où chacun a pu défendre ses choix, et plusieurs similitudes sont ressorties :

- Les premières années de développement ont été écartées. Par la connaissance de la loi About d'une part, et l'examen des courbes de développement par survenance d'autre part, les actuaires ont jugé bon d'exclure les 3 voire 4 premières années.
- Un facteur de queue a presque toujours été ajouté compte tenu de la longueur de la branche et de la faible profondeur de l'historique, en particulier après la suppression des premières survenances.
- Tous les actuaires sauf un ont utilisé le triangle de charges. Les règlements sont en effet mal adaptés dans le cas d'une branche longue et volatile. Le dernier actuaire a choisi une méthode nombre \times coût moyen, compte tenu de la volatilité, ce qui utilise indirectement le triangle de charges.
- Les diagonales les plus récentes ont été privilégiées, en l'absence d'information sur la politique de gestion, pour avoir une vision proche des des conditions actuelles. Cependant la dernière diagonale, dont nous avons déjà pointé la particularité, a parfois été exclue. Ceci conforte nos observations.

Un certain nombre de points communs apparaissent donc. Toutefois, les résultats finaux diffèrent, ce qui illustre la difficulté de réaliser un *best estimate*. Alors que la prudence d'estimation donne des principes généraux, l'estimation au plus juste reste plus subjective. Au cours de la discussion cependant, certains actuaires ont changé d'avis sur leurs hypothèses de modélisation. Ainsi, l'échange de points de vue au cours du travail présente un réel intérêt. C'est d'ailleurs ce qui a lieu dans un processus de provisionnement classique et lors de l'audit des provisions.

Les résultats sont synthétisés dans le tableau 19 et détaillés en annexe.

Minimum	Médiane	Maximum	Moyenne
824 038	902 445	1 175 408	946 475

TABLE 19 – Synthèse des ultimes proposés par les experts (kUM)

Cet échange sur les hypothèses retenues par les actuaires permet de discuter la place du jugement d’expert dans notre modèle. Parmi les points susmentionnés, un ne peut être évité : le choix de travailler avec les charges ou avec les règlements. Dans notre cas, ce choix est équivalent au choix du modèle. Bien que les charges semblent plus adaptées à la branche dont proviennent nos données, nous avons pointé quelques faiblesses du modèle multiplicatif. La mise à l’écart des premières années de survénance est faisable manuellement en les excluant de la base de données, à condition de rajouter à la fin les charges correspondantes. Les modèles ne sont toutefois pas prévus pour cela, et l’avis d’expert est indispensable. Les différentes classifications sont aussi incapables d’isoler des années de survénance à partir des variables que nous avons utilisées qui ne décrivent pas le développement. Quant à la dernière diagonale, le développement de 2019, nous avons déjà mesuré son effet lors du *backtesting*. La position de nos résultats par rapport à cette distribution sera présentée dans le tableau récapitulatif.

3.6.3 Bilan des estimations ultimes

Pour finir, nous présentons les résultats de tous les modèles implémentés. Nous avons mené au fur et à mesure des analyses sur divers points, mais la comparaison globale des ultimes reste à faire. Sont indiqués pour chaque modèle la moyenne des charges ultimes et le quantile 75%, souvent utilisé. Nous ajoutons aussi les résultats agrégés selon les catégories corporel et matériel, retenues pour la classification déterministe. Le tout est reporté dans le tableau récapitulatif 20.

Modèle	Moyenne corporel	Moyenne matériel	Moyenne total	Q75%
Additif	766	62	828	840
Additif, séparation déterministe	796	37	833	846
Additif, k-means	817	43	860	860
Additif, arbre	866	45	911	927
Additif, charges	-	-	971	987
Multiplicatif	955	54	1 010	1 021
Multiplicatif, séparation déterministe	974	34	1 008	1 020
Multiplicatif, k-means	965	53	1 018	1 031
Multiplicatif, arbre	974	51	1 025	1 041
Chain-Ladder charges (par segment)	931	29	960	-
Chain-Ladder charges (total réalloué)	925	33	959	-
Avis d’experts	[824, 1 175]		957	1 043

TABLE 20 – Charges ultimes pour les différents modèles (millions d’UM)

Comparons d’abord les deux modèles de provisionnement individuel utilisés. Les modèles additifs s’appuyant sur des paiements incrémentaux prédisent toujours des ultimes inférieurs à ceux des modèles multiplicatifs. Une raison à cela est l’insuffisance des paiements sur les branches longues dans le cas d’un historique limité. Alors qu’une charge a vocation à estimer l’ultime, les informations sur les règlements n’incluent pas d’estimation du futur.

De ce fait, les développements très longs peuvent être sous-estimés par les modèles. C'est notamment pour cette raison que les experts à qui nous avons soumis les triangles ont tous eu recours aux charges.

Pour évaluer le modèle additif en s'affranchissant des limites des paiements, nous l'avons adapté à une utilisation avec des charges. En particulier, l'inflation a été neutralisée. Les ultimes obtenus sont bien supérieurs à tous ceux des modèles à base de paiements mais restent inférieurs à ceux des modèles multiplicatifs. Ceci montre qu'à la fois le choix des facteurs ou des incréments et des paiements ou des charges ont un impact.

Nous avons déjà décrit l'effet de la classification déterministe sur la répartition des charges entre sinistres corporels et matériels, à savoir qu'elles augmentent sur les premiers tout en diminuant sur les seconds. Nous constatons que les classification s'appuyant sur l'apprentissage statistique ont le même effet. Ceci pouvait être attendu avec le *k-means* puisque nous avons pointé l'effet de la variable corporel/matériel dans la caractérisation d'une classe. En revanche, il était moins évident de retrouver ce phénomène avec l'arbre de classification, bien que cette variables apparaisse à plusieurs nœuds de l'arbre. Au niveau global, ces classifications font augmenter les ultimes, à un niveau plus élevé à la fois en valeur et en proportion sur le modèle additif.

En ce qui concerne les modèles agrégés, le modèle de référence naïf, Chain-Ladder sur un triangle de charges, prédit une valeur entre nos modèles additifs et multiplicatifs. Nous avons ici deux versions : le premier, par segments, est le résultat du développement d'un triangle de sinistres corporels et d'un triangle de matériels, tandis que le second est un triangle global dont les ultimes ont été réalloués par catégorie au *prorata* de la charge courante. Quant aux avis d'experts, ils recouvrent une plage assez large mais qui inclut les résultats de tous nos modèles.

Enfin, nous avons toujours utilisé dans nos modèles tout l'historique à notre disposition malgré les différences des années de survenance 2000 à 2002 dues à la loi About. Afin d'estimer la pertinence de ce choix, nous relançons tous les modèles en excluant ces années. Nous ajoutons ensuite à l'ultime obtenu la charge courante de ces années, logiquement proche de la charge ultime puisque très peu de sinistres survenus en 2002 ou avant sont encore ouverts. Les résultats détaillés sont présentés en annexe. La classification a encore tendance à faire augmenter les résultats, sauf par rapport au modèle multiplicatif non classifié. Les résultats des modèles additifs sont à nouveau plus bas que ceux des modèles multiplicatifs, mais l'écart avec les modèles lancés sur la base complète est plus élevé pour les méthodes incrémentales. Il est donc justifié de conserver ces années pour gagner en profondeur d'historique.

Conclusion

Les grandes quantités de données disponibles et les capacités de traitement accrues des ordinateurs poussent la recherche actuarielle à développer de nouvelles techniques de provisionnement, individuelles. Là où les méthodes agrégées produisent une réserve par année de survenance, les méthodes individuelles visent à estimer une réserve à l'échelle du sinistre. Un traitement ligne par ligne permet théoriquement la prise en compte de toutes les informations disponibles pour un sinistre qui sont perdues lors d'une agrégation. Les bénéfices attendus d'une telle méthode sont un gain de précision, éventuellement des fonctionnalités qui manquent à des modèles traditionnels comme l'estimation d'une distribution d'ultimes, ou encore une meilleure automatisation.

Ce mémoire s'inscrit dans la suite de travaux ayant conduit à l'élaboration de deux modèles de provisionnement individuel. Ceux-ci peuvent simuler le développement année par année des sinistres connus de l'assureur, et aussi prendre en compte les sinistres tardifs, survenus mais non encore déclarés. Ces deux composantes conduisent à une estimation de charge ultime. Le premier modèle, dit « additif », s'appuie sur un processus de paiements cumulés se traduisant par le tirage d'un montant de règlement par sinistre à chaque année de développement. Le second modèle, « multiplicatif », fait évoluer la charge du sinistre avec des facteurs de développement, comme un pendant individuel de Chain-Ladder. La répétition de ces simulations conduit à l'obtention d'une charge ultime.

Une limite majeure de ces modèles est que, bien qu'ils fournissent une estimation de la charge ligne à ligne, ils n'utilisent pas les informations supplémentaires disponibles, ce qui revient à simuler tous les sinistres de la même manière. Le but de ce mémoire était de pallier cette lacune qui réduit sensiblement l'intérêt de procéder à des estimations individuelles. Les modèles de provisionnement utilisés s'appuient sur un historique de données dans lequel sont tirés les montants de règlements ou les facteurs de développement. Pour garder ce fonctionnement en le raffinant, l'objectif était de procéder à une classification des sinistres pour constituer des groupes homogènes, puis de lancer les modèles indépendamment sur chaque groupe. Une analyse fine des effets de la classification et d'autres caractéristiques des modèles, pas encore évaluées systématiquement, devait permettre d'évaluer l'applicabilité opérationnelle de ces modèles.

Une classification basique consistant à séparer les sinistres en deux catégories montre déjà des effets allant dans le sens attendu. Une segmentation de ce type est parfois réalisée par les gestionnaires et conduit les actuaires à développer plusieurs triangles. Elle montre toutefois des limites en finesse, et laisse encore des variables de côté. Deux méthodes d'apprentissage statistique pour la classification sont ensuite mises en place. La première, non supervisée, est l'algorithme des k-moyennes, fonctionnant par proximité entre les variables. Elle produit des classes interprétables. Pour la seconde, supervisée, nous avons défini des classes à partir de la charge après un an de développement. La classification s'appuie ensuite sur un arbre qui partitionne récursivement les données à partir d'un critère localement optimal.

Les simulations de développement des sinistres qui découlent de ces modèles, avec ou sans classification, sont évaluées de plusieurs façons : avec un *backtesting* à un an, par la comparaison à la référence Chain-Ladder, par rapport à des jugements d'experts. En revanche, il est impossible de comparer la charge ultime prédite à une valeur réelle par définition inconnue. La nécessité de comparer nos résultats à des valeurs de référence lors des analyses est un premier point notable de nos modèles. Nous avons identifié plusieurs phénomènes à l'origine d'écarts entre des résultats. Ceci nous a montré que nos modèles étaient affectés par les développements atypiques de certaines années ou qu'ils étaient moins sensibles que Chain-Ladder aux facteurs tardifs. Toutefois, sans point de comparaison, le calcul d'un *best estimate* est un exercice délicat. Nos modèles apportent dans ce cadre des capacités complémentaires aux méthodes agrégées.

Les ultimes prédits diffèrent en fonction du modèle choisi. Les méthodes à base de paiements ont une tendance à la sous-estimation liée à la longueur de la branche et la profondeur limitée de l'historique. Aussi, une analyse équivalente sur d'autres données pourrait apporter d'autres informations. La classification choisie a aussi un impact sur les résultats. D'une part, elle ne permet pas de rapprocher suffisamment les deux modèles. D'autre part, l'effet du choix de la classification est moindre que celui du choix du modèle additif ou multiplicatif. Cependant, les effets par classe sont satisfaisants, mais certaines exceptions limitent l'amélioration globale.

Les choix retenus pour la classification sont toutefois liés aux données considérées. Par exemple, il apparaît pertinent de regrouper ensemble les sinistres tardifs sur notre branche qui en compte de nombreux. D'autres variables de supervision auraient pu être retenues pour l'arbre, comme la durée de vie, mais les variables explicatives disponibles restent limitées.

Chaque modèle a ses points forts. Le multiplicatif, reposant sur les charges, est adapté à la longueur de la branche. L'additif est plus stable et moins sensible aux développements atypiques. Il ne surestime pas l'ultime sur la classe des sinistres grave. En effet, un groupe de sinistres graves est ressorti des deux algorithmes de classification, et il apparaît plus pertinent de ne pas utiliser un modèle pour les traiter. Par ailleurs, les hypothèses retenues par les actuaires pour le provisionnement n'ont pas toujours d'équivalence dans nos modèles, ce qui souligne que le jugement d'expert reste nécessaire.

Les travaux réalisés permettent donc de mieux appréhender les modèles. La constitution de groupes homogènes de sinistres constitue en grande partie l'intérêt des modèles individuels, bien que l'effet de la classification sur les ultimes soit secondaire. La confrontation de ces méthodes à des données issues d'autres branches permettrait d'étendre ou de nuancer les conclusions tirées ici.

Bibliographie

- BAUDRY, Maximilien et Christian Y. ROBERT (2019). “A machine learning approach for individual claims reserving in insurance”. In : *Applied Stochastic Models in Business and Industry* 35.5, p. 1127-1155.
- BENJAMIN, S. et L. M. EAGLES (1986). “Reserves in Lloyd’s and the London market”. In : *Journal of the Institute of Actuaries* 113.2, p. 197-256.
- BEZDEK, James et R.J. HATHAWAY (fév. 2002). “VAT: A tool for visual assessment of (cluster) tendency”. In : t. 3, p. 2225-2230. ISBN : 0-7803-7278-6. DOI : 10.1109/IJCNN.2002.1007487.
- BORNHUETTER, R. et R. FERGUSON (1972). “The actuary and ibnr”. In : *LIX*. Proceedings of the Casualty Actuarial Society, p. 181-195.
- BREIMAN, L (oct. 2001). “Random Forests”. In : *Machine Learning* 45, p. 5-32. DOI : 10.1023/A:1010950718922.
- BREIMAN, L. et al. (1984). *Classification and Regression Trees*. Monterey, CA : Wadsworth et Brooks.
- CHARPENTIER, Arthur, Michel DENUIT et Romuald ELIE (jan. 2015). “Segmentation et Mutualisation, les deux faces d’une même pièce”. In : *Risques* 103. *Code des assurances* (2021).
- CONSEIL DE L’UNIONS EUROPÉENNE, Parlement européen et (2009). “DIRECTIVE 2009/138/CE sur l’accès aux activités de l’assurance et de la réassurance et leur exercice (solvabilité II)”. In : *JO L* 335.
- DUVAL, Francis et Mathieu PIGEON (juill. 2019). “Individual Loss Reserving Using a Gradient Boosting-Based Approach”. In : *Risks* 7.3, p. 1-18.
- ENGLAND, Peter et Richard VERRALL (1999). “Analytic and bootstrap estimates of prediction errors in claims reserving”. In : *Insurance: Mathematics and Economics* 25.3, p. 281-293. ISSN : 0167-6687.
- HAND, David J. et Robert J. TILL (2004). “A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems”. In : *Machine Learning* 45, p. 171-186.
- KAPLAN, E. L. et Paul MEIER (1958). “Nonparametric Estimation from Incomplete Observations”. In : *Journal of the American Statistical Association* 53.282, p. 457-481.
- KUO, Kevin (sept. 2019). “DeepTriangle: A Deep Learning Approach to Loss Reserving”. In : *Risks* 7.3, p. 97. ISSN : 2227-9091.
- LAWSON, Richard G. et Peter C. JURIS (1990). “New index for clustering tendency and its application to chemical problems”. In : *Journal of Chemical Information and Computer Sciences* 30.1, p. 36-41. DOI : 10.1021/ci00065a010. eprint : <https://doi.org/10.1021/ci00065a010>. URL : <https://doi.org/10.1021/ci00065a010>.
- LOPEZ, Olivier, Xavier MILHAUD et Pierre-E. THÉRON (2019). “A tree-based algorithm adapted to microlevel reserving and long development claims”. In : *ASTIN Bulletin* 49.3, p. 741-762.
- MACK, Thomas (1993). “Distribution-free Calculation of the Standard Error of Chain Ladder Reserve Estimates”. In : *ASTIN Bulletin* 23.2, p. 213-225.
- MURPHY, Karl et Andrew McLennan FIAA (2006). “A Method For Projecting Individual Large Claims”. In.

- NORBERG, Ragnar (1993). “Prediction of Outstanding Liabilities in Non-Life Insurance”.
In : *ASTIN Bulletin* 23.1, p. 95-115.
- (1999). “Prediction of Outstanding Liabilities II. Model Variations and Extensions”.
In : *ASTIN Bulletin* 29.1, p. 5-25.
- PRIMEL, D (2021). *Comparaison de différents modèles de provisionnement ligne à ligne et discussions dans le cadre d’une application en non-vie.*
- RIFFAUD, J (2019). *Modélisation de la charge IBNYR dans le cadre d’un modèle de provisionnement individuel en assurance non-vie.*
- ROBIN, Xavier et al. (2011). “pROC: an open-source package for R and S+ to analyze and compare ROC curves”. In : *BMC Bioinformatics* 12, p. 77.
- TAYLOR, G. C. (1977). “Separation of Inflation and other Effects from the Distribution of Non-Life Insurance Claim Delays”. In : *ASTIN Bulletin* 9.1-2, p. 219-230. DOI : 10.1017/S0515036100011533.
- THERNEAU, Terry et Beth ATKINSON (2019). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1-15. URL : <https://CRAN.R-project.org/package=rpart>.
- VERBEEK, H. G. (1972). “An approach to the analysis of claims experience in motor liability excess of loss reinsurance”. In : *ASTIN Bulletin* 6.3, p. 195-202. DOI : 10.1017/S0515036100010989.

Annexes

Classification courtier/autre

	Réserves : moyenne (kUM)	Réserves : écart-type (kUM)	RMSE (UM)	MAE (UM)
Modèle sans séparation	34 406	4 308	73 529	10 347
Courtier	26 566	3 776	70 083	10 363
Autres	7 840	1 901	82 953	10 295
Modèle avec séparation	37 318	4 848	77 969	10 678
Courtier	29 182	4 326	75 986	10 743
Autres	8 137	2 058	82 844	10 459

TABLE 21 – Résultats à 1 an de la classification courtier/autre, modèle additif

Classification tardif/non tardif

	Réserves : moyenne (kUM)	Réserves : écart-type (kUM)	RMSE (UM)	MAE (UM)
Modèle sans séparation	34 662	4 640	73 536	10 355
Non tardif	26 645	4 211	56 943	9 523
Tardif	5 017	1 675	134 527	15 364
Modèle avec séparation	35 012	4 467	73 782	10 393
Non tardif	30 365	4 379	59 245	9 614
Tardif	4 646	1 060	129 394	15 086

TABLE 22 – Résultats à 1 an de la classification tardif/non tardif, modèle additif

Stabilité des *clusters* des k-moyennes

	Base complète	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Effectif	69184	23644	40683	4433	51	373
Corporel	0,634	0	1	0,64	0,941	0,858
Courtier	0,659	0,526	0,745	0,596	0,863	0,434
Litigieux	0,119	0,008	0,185	0,089	0,686	0,18
”other”	0,953	0,97	0,941	0,972	1	0,973
”pooled”	0,031	0,02	0,039	0,023	0	0,027
Délai de déclaration (mois)	2,918	2,026	1,551	15,403	3,196	60,099
Charge initiale (UM)	15 031	1 978	20 109	10 616	2 374 790	18 347

TABLE 23 – Statistiques descriptives des 5 clusters obtenus par k-means sur la base en vision 2018

	Base complète	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Effectif	69184	19308	5951	7737	68	156	35183	769	12
Corporel	0,634	0,438	0,55	0,974	0,912	0,974	0,676	0,797	1
Courtier	0,659	0	0,671	0,761	0,559	0,853	1	0,432	0,917
Litigieux	0,119	0	0,065	0,98	0,294	0,705	0	0,142	0,833
"other"	0,953	0,998	0,975	0,94	0,956	0,987	0,927	0,97	1
"ppooled"	0,031	0	0,019	0,01	0,044	0	0,056	0,029	0
Délai de déclar. (mois)	2,918	1,467	11,756	1,103	107	2,276	1,658	37,769	2,667
Charge initiale (UM)	15031	5290	8536	37640	27020	1028718	10527	15235	4275637

TABLE 24 – Statistiques descriptives des 8 clusters obtenus par k-means sur la base en vision 2018

Description des classes de l'arbre de classification

	Base complète	0]0, 100[[100, 1000[[1000, 10 ⁴ [[10 ⁴ , 10 ⁵ [[10 ⁵ , 10 ⁶ [[10 ⁶ , 10 ⁷ [
Effectif	57707	15081	429	2187	17963	20908	1071	68
Corporel arbre	0,714	0,139	0,534	0,656	0,912	0,953	0,979	0,956
Courtier	0,713	0,53	0,58	1	0,773	0,759	0,824	0,868
Litigieux	0,148	0,024	0,04	0,021	0,1	0,275	0,518	0,662
Délai de déclaration (mois)	2,991	3,203	4,033	2,496	3,143	2,751	2,782	2,618
Charge initiale (UM)	17 941	1 061	15	426	3 456	26 847	238 503	2 052 417

TABLE 25 – Statistiques descriptives par classe pour l'arbre de classification

Détail des prévisions d'experts

1	2	3	4	5	6	7	8	9	10	11	12	13	14
831 274	877 189	873 250	1 066 063	921 671	992 460	875 645	1 074 339	1 175 409	883 220	1 043 971	824 038	965 527	874 601

TABLE 26 – Ultimes prédits par les experts (kUM)

Exclusion des années 2000-2002

Modèle	Moyenne corporel	Moyenne matériel	Moyenne total	Q75%
Additif	691,2	53,6	744,8	751,7
Additif, séparation déterministe	710,6	33,3	743,9	750,7
Additif, k-means	710,1	36,2	746,4	752,8
Additif, arbre	707,7	39,1	746,8	753,7
Multiplicatif	905,4	61,1	966,5	979,4
Multiplicatif, séparation déterministe	917,6	32,2	949,8	961,8
Multiplicatif, k-means	925,7	37,2	962,9	975,2
Multiplicatif, arbre	973,9	49,5	1 023,5	1 041,9
Chain-Ladder charges -	-	-	889,9	-

TABLE 27 – Charges ultimes en excluant les survenances 2000-2002 (millions d'UM)