



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

le 22 septembre 2016

Par : Clément BILLOREÉ

Titre : Application de l'apprentissage automatique au provisionnement ligne à ligne en assurance non-vie

Confidentialité : Non

**Membre présent du jury de l'Institut
des Actuaire**

Emmanuel DUBREUIL

Signature :

Romain NOBIS

Signature :

Membres présents du jury de l'EURIA

Pierre ALLIOT

Signature :

Entreprise

Sia Partners

Signature :

Directeur de mémoire en entreprise

Romain LAILY

Signature :

Invité

Pierre-Antoine MERLE

Signature :

**Autorisation de publication et de mise en ligne sur un site de diffusion
de documents actuariels**

Signature du responsable entreprise

Signature du candidat

Bibliothèque :

Secrétariat :

Résumé

L'émergence actuelle de nouvelles méthodes d'apprentissage statistique ainsi que l'explosion du volume de données disponibles offrent aux assureurs de nouvelles possibilités dans la gestion de leurs risques. Les assureurs non-vie, désireux de vouloir réduire les variations annuelles de leurs provisions et utilisant aujourd'hui des méthodes agrégées simples, sont particulièrement à l'écoute de ces nouvelles techniques. Dans ce contexte, ce mémoire propose une méthodologie de provisionnement ligne à ligne par l'application d'algorithmes de forêts aléatoires, en tant qu'alternatives aux modèles de places.

Sur la base des données d'un assureur français, ce mémoire s'attache dans un premier temps à développer les méthodes de provisionnement largement utilisées sur la place aujourd'hui, telles que le modèle de Mack. Après en avoir rappelé les caractéristiques, ce modèle permet de fixer une mesure de base des provisions dans le cadre de ces travaux.

Dans un second temps, le mémoire présente une méthodologie alternative : une méthode empirique projetant les sinistres individuellement, permettant ainsi de mieux prendre en compte le risque porté par l'assureur. Cette méthode vise à affiner les modèles agrégés afin de mieux prendre en compte les différents risques associés aux sinistres.

Enfin, l'algorithme d'apprentissage automatique des forêts d'arbres décisionnels est mis en place dans le cadre du provisionnement ligne à ligne, dans le but d'intégrer toutes les informations explicatives d'un sinistre dans le modèle de prédiction de sa provision. L'idée principale du mémoire est de présenter ce modèle dans le cadre des problématiques de provisionnement et d'analyser son impact ainsi que la qualité des résultats qu'il peut induire.

Mots-clés : Solvabilité 2, risque de réserve en souscription non-vie, Provisions, MSEP, Chain-Ladder, Modèle de Mack, Sinistralité atypique, Random Forest, Traité de réassurance non-proportionnel par excédent de sinistres, Apprentissage automatique.

Abstract

With the growing development of statistical methods and the improvement in data availability, insurers face new challenges in term of their risk monitoring. Interest in these technics is amplified by the insurer's willingness to reduce the volatility of their provisions where standard aggregate methods may have reached their limits. In that framework, we explore an alternative reserving claims by claim's method based on the application of a random decision forest.

Firstly we assess the existing literature and the usual technics widely used on the market as the Chain Ladder or Mack models which will be used as a benchmark applied on data from a French insurer.

Secondly, we suggest an alternative method based on empirical data and derived from aggregate approach. By taking into account more risk component, it aims to better apprehend the sources of the risk.

Finally, we put in place an algorithm of random decision forest in the framework of the reserving claims by claims method in order to integrate all the explicit claim information inside the predictive model. The main idea is then to extent the concept to the reserving approach and to analyse its effect on the stability of the provision estimation.

Keywords : Solvency 2, Non-life underwriting risk, Standard Formula, MSEF, Chain-Ladder, Mack Model, Random Forest, Excess of Loss, Machine-Learning.

Synthèse

L'explosion des données et leur collecte entraînent les assureurs à être particulièrement attentifs aux récentes technologies permettant d'en extraire de nouvelles informations pertinentes pour améliorer le suivi de leurs risques imposé par la nouvelle réglementation Solvabilité 2.

Les assureurs sont soucieux de réduire la volatilité liée à l'estimation de leurs provisions et de leur évolution au cours du temps. Cette réduction de la volatilité passe par une meilleure estimation de leurs provisions chaque année. C'est dans ce contexte que ce mémoire propose de développer deux méthodes de provisionnement alternatives aux méthodes classiques.

Aujourd'hui, les assureurs utilisent des méthodes de provisionnement agrégées sous forme de triangles pour estimer leurs provisions. Ces méthodes n'exploitent qu'une petite proportion de l'information et ne permettent pas d'appréhender le risque de réserve avec précision, entraînant des variations significatives de provisions d'un exercice à l'autre. Ce mémoire propose de développer un modèle exploitant toutes les données récoltées par l'assureur afin de prédire la charge des sinistres avec une meilleure estimation.

Dans un premier temps, la théorie et les spécificités des méthodes de provisionnement classiques (Chain-Ladder, Mack) sont exposées. Elles permettent de fixer une base de référence pour les travaux suivants.

Dans un second temps, ce mémoire propose une projection des sinistres ligne à ligne en intégrant l'état des sinistres à chaque année de développement. Cette méthode permet d'affiner la projection agrégée avec la méthode de Chain-Ladder et de capter d'avantage le risque.

Enfin, l'utilisation de l'algorithme des modèles automatiques des forêts aléatoires pour la projection des sinistres par une méthode ligne à ligne permet d'intégrer l'ensemble des informations à la disposition de l'assureur pour prédire la charge à l'ultime. Les données sur le traitement et les caractéristiques du sinistre complètent celles précédemment utilisées dans le modèle de projection individuelle à savoir les charges connues et les états des sinistres. Ce mémoire présente une méthodologie permettant de classifier les variables explicatives puis analyser les résultats sur la prédiction des provisions et l'impact induit.

Au préalable, le périmètre de projection est déterminé, la base de sinistre est segmentée en deux catégories : les sinistres attritionnels et les sinistres atypiques. Pour cela un seuil est déterminé en

minimisant l'erreur d'estimation de la provision par la méthode de Chain-Ladder pour les sinistres attritionnels sur les années 2006 à 2009. Ce seuil est fixé à 32 500€.

Sinistres	Nombre de sinistres	Charge estimée
Attritionnels	65 718	225 476 442
	97,7%	50,9%
Atypiques	1 530	217 298 858
	2,3%	49,1%
Total	67 248	442 775 300

TABLE 1 – Répartition des montants de charges et de nombres

La charge des sinistres atypiques représente 49,1% de la charge totale pour un montant de 442,78M€, alors qu'elle ne représente en nombre que 2,3% des sinistres, d'où l'utilité d'affiner leurs prédictions.

Une base de test est construite sur les années 2007 à 2011, en effet, la charge ultime des sinistres survenus sur ce périmètre est connue à la fin de l'année 2015. Cette base permettra de comparer les projections réalisées avec les trois méthodes développées dans la suite.

Méthode de Chain-Ladder

Le modèle de Chain-Ladder projette le montant des sinistres agrégés par année de survenance, à l'aide des facteurs de développement. Deux triangles sont construits, un triangle pour les sinistres attritionnels et un pour les sinistres atypiques.

Afin de valider ce premier modèle, les montants des provisions estimées pour les exercices 2007 à 2011 sont comparés aux montants des provisions réelles, c'est à dire le montant de provisions que l'assureur devait avoir en sa possession pour faire face à ses engagements lors de ces cinq exercices. Ce montant est constaté lors de l'année 2015. Le tableau suivant permet de visualiser l'erreur faite sur ces exercices.

	2007	2008	2009	2010	2011
Réserves réelles	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%

TABLE 2 – Ecart des estimations des provisions avec Chain-Ladder sur les exercices 2007 à 2011

Au global, Chain-Ladder aboutit à une erreur moyenne de l'estimation de la charge des sinistres atypiques de 12,9% avec un pic d'erreur à 37% pour l'année 2011.

La méthode de Chain-Ladder est une méthode de provisionnement simple d'application et justifiable auprès du régulateur. Cependant, l'estimation des provisions n'est pas stable au cours du temps, la

projection surestime le montant des provisions en 2007, 2008 et 2011 et les sous-estime en 2009 et 2010.

Projection individuelle

Alors que la méthode de Chain-Ladder s'attache à agréger les sinistres, le modèle présenté par la suite permet de simuler des trajectoires pour chaque sinistre de manière individuelle, en tenant compte de leur état, afin d'améliorer l'estimation des provisions.

La méthode de projection développée est dite hybride car la projection des sinistres attritionnels est toujours réalisée avec la méthode de Chain-Ladder tandis que les sinistres atypiques sont projetés avec le modèle de projection individuelle.

Dans un premier temps, pour apporter d'avantage d'informations à la projection, les états de chaque sinistre (ouvert, clos, réouvert et annulé) sont projetés à l'aide de matrices de transition. La construction de ces matrices est réalisée en calculant la fréquence observée sur les changements d'état pour chaque année de développement.

Une fois les états des sinistres connus jusqu'à l'ultime, il convient de calculer les facteurs de développement pour chacun des quatre états disponibles afin de pouvoir développer les charges de chaque sinistre jusqu'à l'ultime.

Les charges des sinistres sont ensuite projetées à l'aide des facteurs de développement individuels conditionnés à l'état du sinistre.

50 000 simulations de la matrice des charges sont réalisées pour obtenir le montant de la charge ultime globale.

Afin de valider le modèle, une étude de l'erreur sur le même périmètre 2007-2011 est menée, cette étude permet de valider la performance du modèle et également de le comparer avec les estimations de la méthode de Chain-Ladder précédente.

	2007	2008	2009	2010	2011
Réserves réelles	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%
Projection individuelle	49 841	48 990	63 127	77 479	83 030
	15,5%	7,1%	3,5%	8,8%	10,4%

TABLE 3 – Ecart entre la provision réelle et l'estimation faite avec les deux modèles

L'erreur moyenne calculée sur les exercices 2007 à 2011 est de 9,1% contre 12,1% pour la méthode de Chain-Ladder. De plus, la projection individuelle permet de diminuer la volatilité et l'erreur d'estimation des provisions.

Modèle automatique

Afin de prendre en compte les informations collectées par l'assureur, un modèle utilisant l'apprentissage automatique par construction d'arbre de décision est développé.

L'algorithme d'apprentissage automatique des forêts aléatoires est une méthode ensembliste agrégeant plusieurs arbres de décisions décorrélés. Chaque arbre découpe l'espace des variables explicatives en groupes homogènes. Dans le cadre du provisionnement non-vie, les forêts aléatoires permettent de prendre en compte les informations disponibles de chaque sinistre, à travers une modélisation ligne à ligne.

Dans un premier temps les variables caractérisant les sinistres sont sélectionnées, puis classifiées avec l'algorithme des forêts aléatoires. Chaque arbre réalise une classification descendante hiérarchique permettant de calculer l'importance de chaque variable dans le modèle. Les variables ainsi sélectionnées sont ensuite utilisées pour réaliser la régression et prédire la charge ultime de chaque sinistre. Un modèle de projection est construit pour chaque année de survenance. Chaque modèle exploite l'intégralité des charges connues du sinistres à la fin de chaque année de développement. Le modèle est ensuite calibré sur une base d'apprentissage contenant les sinistres survenus antérieurement aux sinistres à projeter. Le schéma suivant présente un exemple de base d'apprentissage pour la projection des sinistres survenus en 2007.

	Variables explicatives										Variable à expliquer	
	Charge année dév. 0	Charge année dév. 1	Charge année dév. 2	Charge année dév. 3	Charge année dév. 4	Charge année dév. 5	Charge année dév. 6	Charge année dév. 7	Ancienneté permis	...	Existence PV	Charge Ultime
	Base d'apprentissage	C _{2000,0}	C _{2000,1}	C _{2000,2}	C _{2000,3}	C _{2000,4}	C _{2000,5}	C _{2000,6}	C _{2000,7}			
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮		...		⋮
C _{2006,0}		C _{2006,1}	C _{2006,2}	C _{2006,3}	C _{2006,4}	C _{2006,5}	C _{2006,6}	C _{2006,7}				C _{2006,8}

TABLE 4 – Base d'apprentissage pour la projection des charges des sinistres de 2007

Cette base d'apprentissage permet de développer la forêt aléatoire qui servira à projeter les charges des sinistres de l'année 2007 à l'ultime. Le tableau suivant présente les sinistres de 2007 à projeter.

	Variables explicatives										Variable à expliquer	
	Charge année dév. 0	Charge année dév. 1	Charge année dév. 2	Charge année dév. 3	Charge année dév. 4	Charge année dév. 5	Charge année dév. 6	Charge année dév. 7	Ancienneté permis	...	Existence PV	Charge Ultime
	Base de projection	C _{2007,0}	C _{2007,1}	C _{2007,2}	C _{2007,3}	C _{2007,4}	C _{2007,5}	C _{2007,6}	C _{2007,7}		...	

TABLE 5 – Base des sinistres survenus en 2007

Comme pour les deux méthodes précédentes, l'écart entre les provisions estimées et celles réellement observées en 2015 sur les exercices 2007 à 2011 est calculé.

	2007	2008	2009	2010	2011
Réserves réelles	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%
Projection individuelle	49 841	48 990	63 127	77 479	83 030
	15,5%	7,1%	3,5%	8,8%	10,4%
Random Forest	41 366	44 246	56 479	65 966	69 902
	-4,1%	-3,3%	-7,4%	-7,3%	-7,0%

TABLE 6 – Ecart entre les trois projections et la provision réelle

À noter que l'estimation des provisions avec l'algorithme d'apprentissage statistique sous-estime les provisions à constituer sur tout le périmètre de test 2007 à 2011.

L'erreur moyenne sur cette période est de 5,8% ce qui améliore encore l'estimation des provisions. L'erreur est réduite de 3 points par rapport à la projection individuelle. Le modèle d'apprentissage automatique améliore l'estimation des provisions et réduit la volatilité.

Le tableau suivant présente l'évolution des provisions sur les exercices 2007 à 2011 :

	2007 -> 2008	2008 -> 2009	2009 -> 2010	2010 -> 2011
Réserves réelles	6,09%	33,23%	16,77%	5,61%
Chain-Ladder	4,57%	11,74%	28,17%	45,28%
Projection individuelle	-1,71%	28,86%	22,74%	7,16%
Random Forest	6,96%	27,65%	16,80%	5,97%

TABLE 7 – Evolution des trois projections et de la provision réelle

Alors que les provisions réelles progressent de 72% sur cette période, la projection avec l'apprentissage statistique prédit une augmentation de la provision de 77%, là où l'évolution des provisions est de 148% avec la méthode de Chain-Ladder et de 84% avec la projection individuelle.

Afin de visualiser au mieux la tendance, le graphique suivant présente les trois projections sur le périmètre de test 2007-2011.

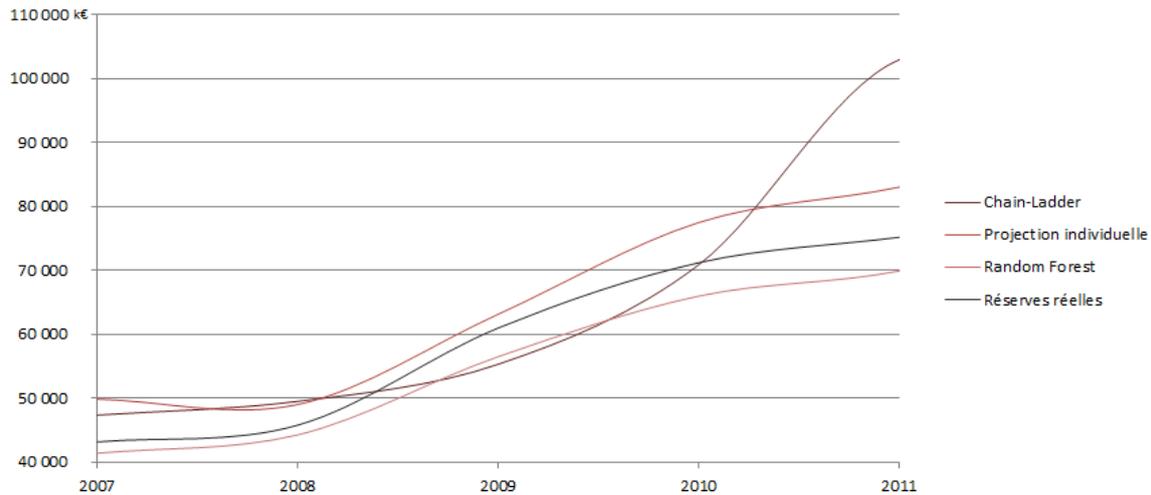


FIGURE 1 – Comparaison de l'estimation des provisions entre les trois méthodes

L'algorithme d'apprentissage statistique reproduit au mieux la tendance de la variation des provisions sur le périmètre temporel 2007-2011 contrairement à la projection individuelle ou la méthode de Chain-Ladder.

L'application d'un traité de réassurance par excédent de sinistre accentue les résultats précédents, c'est-à-dire que les variations des provisions sont plus importantes et que le modèle d'apprentissage statistique s'adapte mieux. De plus, les deux méthodes projettent les sinistres individuellement, cela permet de simplifier la gestion de ce type de traité et sa prise en compte dans l'estimation des provisions.

Les informations sous-jacentes de chaque sinistre contribuent à améliorer l'estimation de son coût final. Une projection ligne à ligne permet de prendre en compte ces informations et ainsi de mieux capter le risque de réserve dans le cadre de la sinistralité atypique. De plus, l'ajout des variables caractérisant la sinistralité lors de la projection, à l'aide du modèle d'apprentissage statistique, continue d'améliorer l'estimation des provisions.

Cependant le manque de traçabilité ne permet pas à l'assureur d'utiliser le modèle d'apprentissage automatique comme unique outil de provisionnement mais plus comme un atout permettant un suivi de son risque de réserve.

Executive summary

As insurers have more and more data to handle, recent technologies that provide solutions for data handling have caught their attention. Solvency II requires that insurers follow more closely and more accurately their risk. In this framework we propose to develop two reserving methods as an alternative to standard methods based on aggregate data.

Nowadays, most of the insurer on the market use aggregate reserving method in order to estimate the reserves. These methods are often criticized as they don't integrate all the available information leading to potential reserves variation between two estimations. By adding, inside the model, more explanatory parameters, we aim to better predict the best estimate of the reserves.

Firstly, we present standard reserving methods as Chain Ladder and Mack that will be used as a benchmark through the essay.

Secondly, we project each claim individually by adding the status of the claim at each stage of their development. The risk is then better captured.

Finally by applying an algorithm of random decision forest on the claim by claims projection, we can incorporate all the available information for the estimation. The treatment of the data and the characteristics of the claim supply the model in addition to the status of the claim. This essay presents a classification methodology for explanatory variables and then analyses the result regarding the prediction of the reserves.

Claims	Number of claims	Estimated load
Attritionals	65 718	225 476 442
	97,7%	50,9%
Atypicals	1 530	217 298 858
	2,3%	49,1%
Sum	67 248	442 775 300

TABLE 8 – Sharing Claims

Beforehand, the scope is defined and the claims database is segmented between attritional claims and atypical claims. A threshold is determined by minimizing the estimation error of the provision of the Chain Ladder method for the attritional part from 2006 to 2009. This threshold is €32,500. The incurred for atypical claims accounts for 49.1% of the total incurred (€442,78M), which corresponds to 2.3% of the claims number.

A test database is built for 2007 to 2011, as the incurred claims already known has a view date at the end of the year 2015. This database will be used to compare the projections made with the three methods previously mentioned.

Method of Chain Ladder

The model of Chain Ladder is based on a projection of the incurred through development factors where claims are aggregated by occurrence year. The projection of the reserves is made by Chain Ladder. Two triangles are built, one for the attritional part and one for the atypical part. To determine the accuracy of the Chain Ladder model, the amount of reserves estimated for 2007 to 2011 are compared to the real amounts, that is the amount that the insurer should have at the time to face his commitment during these 5 years. This information is available in 2015. The table below shows the prediction error for these 5 years :

	2007	2008	2009	2010	2011
Actual Reserves	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%

TABLE 9 – Gap estimation with Chain-Ladder (2007 to 2011)

Globally, Chain Ladder has an average error of the atypical estimation part of 12.9% with a highest point at 37% for 2011. The CL method is easy to implement and to justify. However, the estimation is not stable, the projection overestimates the provision for 2007, 2008 and 2011 and underestimates the real provision for 2009 and 2010.

As opposed to CL which considers aggregate data, the method that we present allows a simulation claims by claims and takes into account the status of the claims and so aims to improve the estimation of the reserves.

Individual projection

The projection method developed can be viewed as a hybrid method as the projection of the attritional claims will be done through a classical aggregate method whereas the atypical claims will be projected on a claim by claims basis.

First, in order to bring more information into the model, the status of the claim (opened, closed, cancelled or reopened) is projected with the utilization of transition matrixes. To build these matrixes we use the observed frequency on the status change at each development period. Once we know all the status until the ultimate, we have to calculate the development factors for each of the possible state of the claims in order to develop each claim to get the ultimate position.

Then the incurred are being projected with those individual development factors depending on the

claim status.

5000 simulations of the incurred matrix have been processed to get the global ultimate amount.

In order to validate the model, we use the prediction error for 2007-2011 similarly to the methodology applied on the CL estimation.

	2007	2008	2009	2010	2011
Actual Reserves	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%
Individual Projection	49 841	48 990	63 127	77 479	83 030
	15,5%	7,1%	3,5%	8,8%	10,4%

TABLE 10 – Gap estimation with two models

The mean error for 2007-2011 is 9.1% as compared to 12.1% with CL. On this data, one can see that the individual projection always overestimates the provision for each year. Besides, the individual projection decreases the volatility and the prediction error of the reserves.

In order to take into account the data collected by the insurers, a model based on machine learning has been developed.

Unsupervised model

The algorithm of random decision forest is an ensemblist method aggregating multiple uncorrelated decision trees. For each tree, the space of explanatory variables is separated in homogenous groups. In the framework of the non life reserving, random forests allow to take into account the all available information for each claim through a claim by claims method.

First the variables are selected and then classified with the algorithm of random decision forest. Each tree make a hierarchical descending classification which can determine the importance of each variable in the model. Those selected variables are then used in the regression to predict the ultimate position for each claim.

A projection model is built for each occurrence year. Each model is then calibrate on a learning base including the claims already reported. The next figure shows the learning base for the projection of the claims reported in 2007.

	Explanatory variables									Variables to explain	
	Load dev. year 0	Load dev. year 1	Load dev. year 2	Load dev. year 3	Load dev. year 4	Load dev. year 5	Load dev. year 6	Load dev. year 7	Load dev. year 8	Existence of fine	Ultimate load
Learning basis	C _{2000,0}	C _{2000,1}	C _{2000,2}	C _{2000,3}	C _{2000,4}	C _{2000,5}	C _{2000,6}	C _{2000,7}	...		C _{2000,14}
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...		⋮
	C _{2006,0}	C _{2006,1}	C _{2006,2}	C _{2006,3}	C _{2006,4}	C _{2006,5}	C _{2006,6}	C _{2006,7}			C _{2006,8}

TABLE 11 – Learning basis

This learning base can develop the random forest which will be used to project the claims from 2007 to the ultimate. The table below shows the 2007 claims to be projected.

Projection basis	Explanatory variables										Variables to explain
	Load dev. year 0	Load dev. year 1	Load dev. year 2	Load dev. year 3	Load dev. year 4	Load dev. year 5	Load dev. year 6	Load dev. year 7	Seniority of the license	Existence of fine	Ultimate load
	C _{2007,0}	C _{2007,1}	C _{2007,2}	C _{2007,3}	C _{2007,4}	C _{2007,5}	C _{2007,6}	C _{2007,7}	C _{2007,14}

TABLE 12 – Projection basis

As for the two previously explained methods, the gap between real and estimated provision observed in 2015 for the period 2007-2011 is calculated.

	2007	2008	2009	2010	2011
Actual Reserves	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%
Individual Projection	49 841	48 990	63 127	77 479	83 030
	15,5%	7,1%	3,5%	8,8%	10,4%
Random Forest	41 366	44 246	56 479	65 966	69 902
	-4,1%	-3,3%	-7,4%	-7,3%	-7,0%

TABLE 13 – Gap estimations with three methods

The estimation of the reserves with the algorithm of random forest underestimates the provision required. The mean error is 5.8% so the error is reduced. Indeed, the error is 3% less than the individual projection. This unsupervised model, based on this data, improve the estimation of the provision and decrease the volatility.

The next table shows the evolution of the reserves between 2007 and 2011 :

	2007 -> 2008	2008 -> 2009	2009 -> 2010	2010 -> 2011
Actual Reserves	6,09%	33,23%	16,77%	5,61%
Chain-Ladder	4,57%	11,74%	28,17%	45,28%
Individual Projection	-1,71%	28,86%	22,74%	7,16%
Random Forest	6,96%	27,65%	16,80%	5,97%

TABLE 14 – Reserves variations

The real reserves go up to 72% during this period. The projection with the algorithm predicts an increase of 77% where the reserves estimation is 148% with Chain Ladder and 84% with the individual projection.

In order to enlighten the trend, the next figure shows the three different projections :

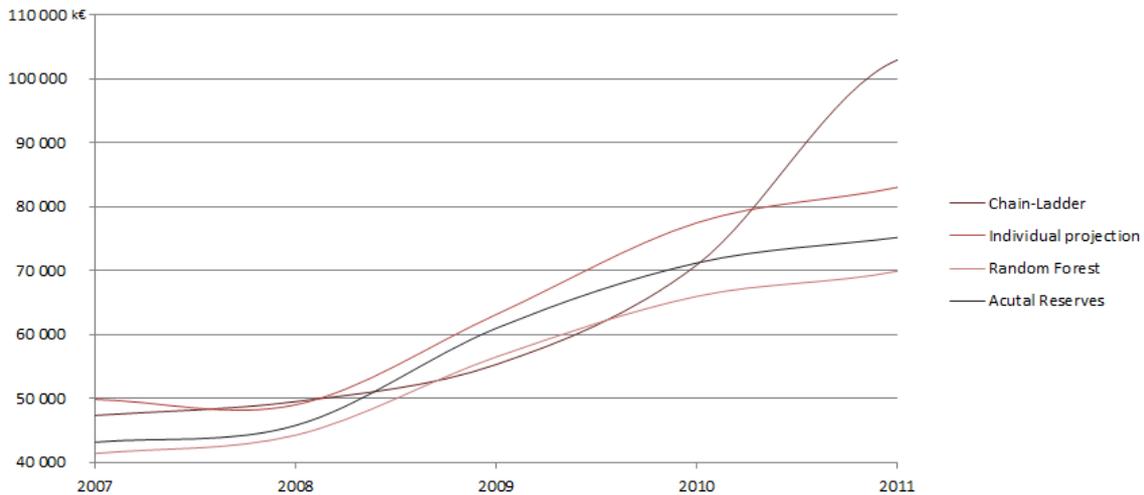


TABLE 15 – Comparison of estimating reserves

The algorithm method replicates better the trend of the reserves variation between 2007 and 2011 as compared to CL or a simple individual projection. The application of an XS reinsurance treaty increases the power of the alternative method proposed in this essay. As the individual method, this type of treaty is easier to handle as the conditions can really vary from one claim to another.

By taken into account of all the information at a claim level, the estimation of the provision is improved. This method is particularly interesting for atypical claims characterized by a high severity for a low frequency. However, we can argue that the algorithm method doesn't allow the same transparency than the CL approach for example and can be less easily justified for an insurer. Then this type of method can be more seen as a complementary tool for the insurer.

Remerciements

Ce mémoire est le fruit d'un travail qui a eu lieu lors d'un stage de fin d'études dans le cabinet de conseil Sia Partners. Je tiens à remercier plusieurs personnes qui ont contribué au bon déroulement de ce travail.

En premier lieu mon tuteur de stage M. Romain LAILY, sans qui ce stage n'aurait pas été possible, je le remercie vivement pour son écoute, son aide, sa disponibilité et tout son soutien au cours de ce stage.

Je remercie M. Michael DONIO, directeur du service Actuariat, ainsi que M. Pierre-Antoine MERLE, responsable du département IARD et Data Science, pour leurs conseils et leurs disponibilités.

De plus, je tiens à remercier tout le service Actuariat / Assurance de Sia Partners et en particulier Amélie, Benjamin, Giacomo, Adeline, Arnaud, Louis et Thibault pour leur aide tout au long de ce travail. J'adresse également ma sympathie à l'ensemble des consultants de Sia Partners pour leur bonne humeur et leur accueil amical.

J'adresse enfin mes plus sincères remerciements à l'ensemble des enseignants de l'EURIA, représentés par M. Marc QUINCAMPOIX, directeur de l'EURIA et M. Franck VERMET, directeur des études, pour la qualité de leurs enseignements au cours de ces trois années de formation. Je tiens également à remercier le collège de direction actuariaire et particulièrement Mme Marine HABART qui m'a suivi tout au long de ce travail.

Table des matières

Résumé	1
Abstract	2
Synthèse	3
Executive summary	10
Remerciements	15
Introduction	19
I Contexte et enjeux	20
1 Contexte réglementaire	21
1.1 Introduction à Solvabilité 2	21
1.1.1 Approche sous la directive Solvabilité 2	21
1.1.2 Le calcul de SCR	22
1.1.3 Le risque de réserve en souscription non-vie	22
1.2 Le bilan prudentiel sous Solvabilité 2	23
1.3 Les méthodes de provisionnement en souscription non-vie	24
1.3.1 La méthode de Chain-Ladder	24
1.3.1.1 Présentation de la méthode	25
1.3.1.2 Validation des hypothèses	26
1.3.2 Le modèle récursif de Mack	27
1.3.2.1 Présentation du modèle	27
1.3.2.2 Validation du modèle	29
1.3.3 Mise en oeuvre dans le cadre du provisionnement	29
1.4 Traité de réassurance non proportionnel : l'excédent de sinistre	29
2 Problématique du mémoire	31
2.1 Objectif du mémoire	31
2.2 Étude préalable de la base de données	31

3	Estimation des provisions par méthode agrégée	36
3.1	Choix du seuil de gravité	36
3.2	Application des méthodes agrégées	39
3.2.1	Application de la Méthode de Chain-Ladder	40
3.2.1.1	Vérification des hypothèses	40
3.2.1.2	Application aux triangles de liquidation	42
3.3	Evolution de la sinistralité	44
3.4	Evolution de l'estimation des provisions	45
3.5	Calcul de l'écart d'estimation sur le périmètre 2007-2011	46
3.6	Application de la Méthode de Mack	47
3.6.1	Vérification des hypothèses	47
3.6.2	Calcul des volatilités	47
3.6.3	Calcul du Capital Réglementaire	48
3.7	Application d'un traité de réassurance non proportionnel : Excédent de sinistre	49
3.7.1	Calcul des provisions à constituer	50
3.7.2	Evolution de l'estimations des provisions	52
3.7.3	Ecart d'estimation des provisions pour les exercices 2007 à 2011	52
3.8	Bilan de la méthode	53
II	Provisionnement par une méthode ligne à ligne déterministe	54
4	Présentation de la sinistralité	55
4.1	Vie d'un sinistre	55
4.2	Traitement d'un sinistre	55
5	Projection individuelle des sinistres atypiques	57
5.1	Projection de la charge des sinistres atypiques	57
5.1.1	Développement de la matrice des états	57
5.1.2	Calcul des facteurs de développement	60
5.1.3	Projection des charges à l'ultime	62
5.1.4	Application de la méthode	62
5.2	Evolution des provisions pour les sinistres atypiques	64
5.3	Calcul de l'écart d'estimation sur le périmètre 2007-2011	65
5.4	Calcul d'un capital réglementaire théorique	66
5.5	Application du traité de réassurance	68
5.5.1	Calcul des provisions à constituer	68
5.5.2	Evolution de l'estimation des provisions	69
5.5.3	Ecart d'estimation des provisions pour les exercices 2007 à 2011	69
5.6	Bilan de la méthode	70

III	Estimation des provisions avec un modèle d'apprentissage automatique	72
6	Les forêts aléatoires	73
6.1	Présentation générale	73
6.2	Les arbres de décisions	73
6.2.1	Construction d'un arbre	73
6.2.2	Le critère de coupure : CART	75
6.3	Les forêts d'arbres aléatoires	75
6.3.1	Ré-échantillonnage aléatoire : Bagging	76
6.3.2	Construction de la forêt aléatoire	76
6.4	Choix des variables	77
6.4.1	Erreur Out-of-bag	77
6.4.2	Importance des variables	78
6.5	Avantages et inconvénients	78
7	Implémentation du modèle	79
7.1	Variables de sinistralités	79
7.2	Sélection des variables et paramétrage du Random Forest	80
7.2.1	Périmètre de calibration	80
7.2.2	Choix du nombre d'arbres	81
7.2.3	Choix des variables	81
7.3	Projection des charges à l'ultime	83
7.4	Evolution de l'estimation des provisions pour les exercices 2007 à 2014	90
7.5	Calcul de l'écart d'estimation sur le périmètre 2007-2011	90
7.6	Calcul d'un capital réglementaire théorique	92
7.7	Application du traité de réassurance 130k€ XS 230k€	93
7.7.1	Calcul des provisions à constituer	93
7.7.2	Evolution de l'estimation des provisions	96
7.7.3	Écart d'estimation des provisions pour les exercices 2007-2011	97
7.8	Bilan de la méthode	98
	Conclusion	99
	Bibliographie	100
	Lexique	102
	Annexes	105

Introduction

La récolte de données massives et le développement d'algorithmes permettant leur traitement fournissent aux assureurs de nouvelles perspectives pour la gestion de leur risque. L'évolution de l'environnement réglementaire imposée par Solvabilité 2 oblige désormais d'avoir un suivi plus juste des risques.

Soucieux de vouloir réduire la volatilité de l'estimation des provisions et leurs variations dans le temps, les assureurs sont à l'écoute des nouvelles méthodes de projection de celle-ci. Ce mémoire propose de développer deux méthodes alternatives au provisionnement classique agrégé, afin d'améliorer l'estimation et la projection des provisions au cours du temps en estimant la charge sinistre par sinistre.

La donnée constitue une réelle richesse de l'assureur, cependant son utilisation dans les modèles de provisionnement est limitée. Ce mémoire se focalise à développer une application d'un algorithme d'apprentissage statistique dans le cadre du provisionnement non-vie permettant d'intégrer un volume et une variété de données jusqu'à présent inexploitées pour quantifier le risque de réserve.

En exploitant la base de sinistre d'un assureur français, ce mémoire vise à rappeler les méthodes de provisionnement agrégées (méthode de Chain-Ladder et modèle de Mack) afin d'en construire une mesure de référence et apprécier les contributions des deux autres méthodes.

L'explosion des moyens en termes de puissance de calcul permet de développer un premier modèle de projection individuelle exploitant d'avantage le profil de risque de chaque sinistre. Ce modèle étaye les méthodes agrégées et vise à les améliorer.

La troisième approche exploite un algorithme d'apprentissage automatique reposant sur la construction d'arbres de décisions dont l'objectif est de sélectionner et intégrer toute l'information pertinente à la disposition de l'assureur afin de prédire la provision à constituer. L'objectif de ce mémoire est d'intégrer ce nouvel outil dans des problématiques de provisionnement et analyser l'impact qu'il peut amener sur la qualité des estimations par son application opérationnelle.

Première partie

Contexte et enjeux

Chapitre 1

Contexte réglementaire

La directive réglementaire Solvabilité 2, entrée en vigueur au 1^{er} janvier 2016, est une réforme européenne dont l'objectif pour les compagnies d'assurances est d'identifier, d'analyser, de mesurer, de gérer et de communiquer sur leurs risques. Dans ce contexte, les assureurs sont amenés à estimer leurs provisions avec une plus grande précision et doivent pouvoir contrôler leur évolution d'un exercice à l'autre. Dans le cadre de ce mémoire, deux nouvelles méthodes sont développées afin d'aider l'assureur à appréhender le risque de réserve en souscription non-vie. Dans un premier temps, les grands principes de la réforme Solvabilité 2 sont rappelés, puis les méthodes de provisionnement classiques sont présentées afin de pouvoir les comparer aux deux méthodes développées par la suite.

1.1 Introduction à Solvabilité 2

1.1.1 Approche sous la directive Solvabilité 2

Le dispositif Solvabilité 2 est une directive dont l'objectif est d'harmoniser les réglementations au niveau européen afin de mieux couvrir les assurés contre les risques de faillites des assureurs.

Solvabilité 2 est construit autour de trois piliers :

- **pilier 1** : Ce premier pilier est un pilier d'exigence quantitative, il permet d'évaluer le montant de fonds propres dont les compagnies d'assurance doivent disposer afin de faire face aux risques qu'elles encourent dans leur activité.

Il est caractérisé par le calcul :

SCR (Solvency Capital Requirement) : représente le niveau de fonds propres dont doit disposer un assureur afin de ne pas se retrouver en ruine économique dans 99,5% des cas.

MCR (Minimum Capital Requirement) : correspond à l'exigence minimale de fonds propres dont doit disposer un assureur pour exercer son activité.

- **pilier 2** : Ce pilier complète le premier pilier avec des exigences quantitatives et qualitatives. L'objectif est de mesurer, gérer et maîtriser ses risques.
- **pilier 3** : Exigence de transparence avec les publications d'informations, l'exigence de reporting auprès du régulateur. Il harmonise la communication faite auprès des assureurs et de l'ACPR.

1.1.2 Le calcul de SCR

Le **Capital de Solvabilité Requis** traite tous les risques suivants :

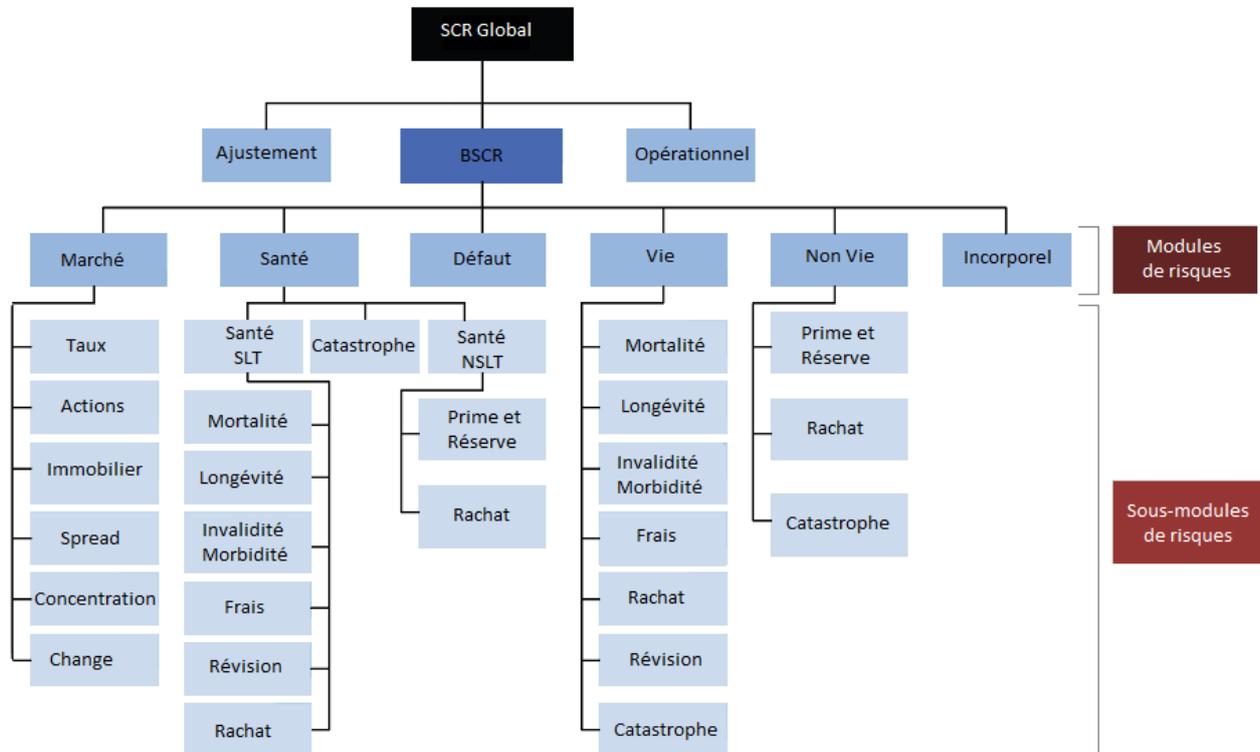


FIGURE 1.1 – Décomposition de la formule standard par module de risques

La formule standard est l'approche générale qu'adoptent les assureurs pour calculer le SCR Global. Le calcul du SCR sous la formule standard s'effectue selon une approche modulaire, il se réalise en quatre étapes :

1. Évaluation des **SCR sous-modulaires**
2. Déduction des **SCR modulaires** par agrégation des SCR de sous-modules
3. Calcul du **BSCR** en agrégeant les SCR modulaires
4. Calcul du **SCR global** correspond au SCR de base et au SCR relatif au risque opérationnel auquel les ajustements pour capacité d'absorption des pertes sont ajoutés.

1.1.3 Le risque de réserve en souscription non-vie

Le risque traité dans ce mémoire est le risque de réserve en souscription non-vie.

Le **risque de réserve** mesure l'incertitude de la compagnie à évaluer convenablement la charge ultime de sa sinistralité passée. Le risque est de faire une mauvaise évaluation de la charge de sinistres restante. L'évaluation du SCR au titre du risque de réserve s'effectue au travers d'une formule fermée estimant l'écart entre le quantile à 99.5% et la moyenne des estimations des provisions.

$$NL_{pr} = 3 \cdot \sigma \cdot V$$

Avec :

- V : mesure du volume sous risque.
- σ : coefficient de variation du risque de réserve. Ce coefficient correspond au ratio écart-type sur moyenne.

1.2 Le bilan prudentiel sous Solvabilité 2

Solvabilité 2 introduit la notion de bilan économique, cela signifie que les actifs sont désormais comptabilisés en valeur de marché et les provisions sont remplacées par un *Best Estimate* et une Marge pour risque.

Un bilan simplifié sous Solvabilité 2 est présenté ci-dessous :

Actifs	Passifs
Investissements et autres actifs	Fonds propres
	Marge pour risque
Provisions techniques à la charge des réassureurs	Best Estimate
	Autres dettes

FIGURE 1.2 – Bilan prudentiel sous Solvabilité 2

- Le *Best Estimate* est défini comme l'espérance de la valeur actuelle de la somme des flux futurs actualisés
- La **Marge pour risque** permet d'intégrer une marge de prudence dans le calcul des provisions techniques. Elle représente le coût de l'immobilisation des fonds propres.

D'après le code des assurances (article R 331-6), la provision pour sinistres à payer (PSAP) est la valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés. Comme toutes les provisions techniques, elle doit être suffisante (article R 331-1) pour le règlement intégral des engagements vis à vis des assurés ou bénéficiaires de contrats. La provision réglementaire pour sinistres à payer calculée exercice par exercice, est déterminée brute de recours (articles R 331-15 et R 331-1). Les recours à recevoir sont évalués séparément sous la forme de prévisions de recours. Par différence, on obtient la PSAP nette de recours figurant au passif du bilan. Celle-ci est évaluée en brute de réassurance. En contre-partie la provision pour sinistres réassurés apparaît à l'actif du bilan.

Une mauvaise estimation des provisions techniques impacte le bilan mais également le résultat et les fonds propres. La figure suivante présente les fonds propres de façon schématique :

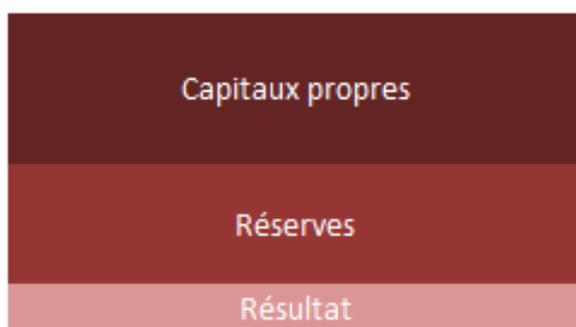


FIGURE 1.3 – Fonds propres simplifiés

Le résultat est obtenu par différence entre les produits et les charges. La variation des provisions s'inscrivant dans le compte de résultat, les erreurs d'estimations entraînent une variation d'une année sur l'autre et impactent le résultat. Un calcul au plus juste permet de réduire la variation des provisions inscrites dans le compte de résultat et contribue à la stabilité du résultat d'une année à l'autre.

1.3 Les méthodes de provisionnement en souscription non-vie

Dans cette section, deux méthodes classiques de provisionnement sont présentées : la méthode de Chain-Ladder et le modèle de Mack. Ces deux méthodes agrégées permettent la projection des triangles de paiements ou de charges.

1.3.1 La méthode de Chain-Ladder

La méthode de Chain-Ladder est une méthode déterministe basée sur le calcul de facteurs de développement permettant de projeter le triangle de provisionnement et d'en déduire les provisions à constituer.

1.3.1.1 Présentation de la méthode

Le modèle de Chain-Ladder s'appuie sur l'hypothèse que les années d'origines sont indépendantes les unes des autres :

- $\forall j \in [0, n - 1]$, les facteurs de développement individuels $\frac{C_{i,j+1}}{C_{i,j}}$ sont indépendants.

L'hypothèse forte d'indépendance entre les années de survenance des sinistres n'est pas vérifiable dans un cadre opérationnel. Cependant, la méthode de Chain-Ladder du fait de sa simplicité de mise en oeuvre et de sa traçabilité est très répandue sur la place.

		Années de développement				
		0	1	2	...	n
Années de survenance	0	$C_{0,0}$	$C_{0,1}$	$C_{0,2}$...	$C_{0,n}$
	1	$C_{1,0}$	$C_{1,1}$			
	2	$C_{2,0}$		\ddots		
	\vdots	\vdots				
	\vdots					
	n	$C_{n,0}$				

FIGURE 1.4 – Triangle supérieur des paiements cumulés

Sous l'hypothèse d'indépendance des années de survenance, la relation suivante doit être vérifiée :

- $\forall j \in [0, n - 1]$,

$$\frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = \dots = \frac{C_{n-j-1,j+1}}{C_{n-j-1,j}}$$

En pratique ces égalités ne sont qu'approximativement vérifiées. Il est naturel de choisir comme facteurs de développement communs :

- $\forall j \in [0, n - 1]$,

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

Les facteurs de développement permettent de calculer les cadences de développement. La cadence de développement représente la proportion de sinistres payés après k années.

— $\forall i \in [0, n - 1]$,

$$p_i = \prod_{k=i}^{n-1} \frac{1}{f_k}$$

Les charges ultimes peuvent être déduites par la formule suivante :

— $\forall i \in [0, n]$,

$$S_i = C_{i,n-i} \cdot \prod_{k=n-i}^{n-1} f_k$$

Les charges ultimes ainsi obtenues permettent de calculer les provisions à constituer :

— $\forall i \in [0, n]$,

$$R_i = S_i - C_{i,n-i}$$

Puis d'obtenir la provision globale :

$$R = \sum_{i=0}^n R_i$$

Le rectangle de liquidation peut être complété avec la formule suivante :

— $\forall i \in [0, n]$,

$$C_{i,j} = C_{i,n-i} \cdot \prod_{k=n-i}^{j-1} f_k$$

1.3.1.2 Validation des hypothèses

L'hypothèse sur laquelle la méthode de Chain-Ladder repose doit être préalablement validée. Il est à noter que toute la modélisation est empirique.

— Si $\forall j$ fixé, $\exists f_j$ tel que :

$$\forall i \in [0, n - j - 1], C_{i,j+1} = f_j \cdot C_{i,j}$$

Les $n - j$ couples $(C_{i,j}, C_{i,j+1})_{i \in [0, n-j-1]}$ doivent être relativement alignés sur une droite passant par l'origine.

La validation essentielle repose sur le triangle de développement formé des facteurs, pour $i + j \leq n - 1$:

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

L'hypothèse sous-jacente à la méthode de Chain-Ladder n'est acceptable que si, pour $j \in [0, n - 2]$, les éléments de la $j^{\text{ème}}$ colonne du triangle de développement sont "constants". Si ce n'est pas le cas, un examen permet d'effectuer des choix alternatifs de facteurs f_j , en particulier si les colonnes du triangle de développement montrent la présence de valeurs extrêmes ou de structures non aléatoires.

1.3.2 Le modèle récursif de Mack

1.3.2.1 Présentation du modèle

À l'inverse du modèle de Chain-Ladder, le modèle de Mack n'est pas paramétrique dans le sens où aucune hypothèse de distribution n'est faite sur les composantes du triangle. Il est conditionnel dans le sens où les espérances sont prises connaissant les réalisations du triangle supérieur. Ce modèle s'applique aux paiements cumulés (ou aux charges) sous les deux hypothèses suivantes :

Hypothèse 1 : Les années d'origine sont indépendantes.

Les variables aléatoires $(C_{i,j})_{j=1..n}$ et $(C_{k,j})_{j=1..n}$ sont pour $i \neq k$ indépendantes.

Hypothèse 2 : Pour $j \in [0, n]$, il existe un paramètre f_j tel que conditionnellement :

$\forall i \in [0, n - j - 1]$,

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

Il est à noter que pour tous les rapports du triangle de développement $f_j = \frac{C_{i,j+1}}{C_{i,j}}$ cette hypothèse peut s'écrire :

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

Théorème 1 :

En notant $D = \{C_{i,j} ; i + j \leq n + 1\}$ l'information apportée par le triangle supérieur, la relation suivante est vérifiée :

$$E[C_{i,n}|D] = f_{n-i+1} \times \dots \times f_{n-1} \times C_{i,n-i+1}$$

Théorème 2 :

Sous les hypothèses précédentes, les facteurs de développement de Chain-Ladder $\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}$ sont des estimateurs sans biais de f_j et sont non corrélés.

Théorème 3 :

- (i) L'estimateur $\hat{C}_{i,n} = f_{n-i+1} \times \dots \times f_{n-1} \times C_{i,n-i+1}$ est un estimateur sans biais de $E[C_{i,n}|D] = f_{n-i+1} \times \dots \times f_{n-1} \times C_{i,n-i+1}$.
- (ii) L'estimateur du montant de provisions de l'exercice i , $\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i+1}$ est un estimateur sans biais de $E[R_i|D]$ avec $R_i = C_{i,n} - C_{i,n-i+1}$.

Hypothèse 3 : Pour $j \in [0, n]$, il existe un paramètre σ_j^2 tel que conditionnellement :

$$V[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = \sigma_j^2 \cdot C_{i,j}, \text{ pour } i = 0, \dots, n$$

Le paramètre $\hat{\sigma}_j^2$ peut être estimé comme suit, pour $j \in [0, n - 2]$:

$$\hat{\sigma}_j^2 = \frac{1}{n - j - 1} \cdot \sum_{i=0}^{n-j-1} C_{i,j} \cdot \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2$$

Pour obtenir l'estimateur $\hat{\sigma}_{n-1}^2$ deux cas sont possibles :

- Si $f_{n-1} = 1$ alors $\hat{\sigma}_{n-1} = 0$
- Sinon $f_{n-1} = 1$ alors $\hat{\sigma}_{n-1}^2 = \min\left(\frac{\hat{\sigma}_{n-2}^2}{\hat{\sigma}_{n-3}^2}; \min(\hat{\sigma}_{n-2}^2; \hat{\sigma}_{n-3}^2)\right)$

L'écart quadratique moyen mesure l'incertitude présente dans l'estimation de $C_{i,j}$ par $\hat{C}_{i,j}$, il s'écrit comme :

$$\begin{aligned} MSEP(\hat{C}_{i,n}) &= E[(C_{i,n} - \hat{C}_{i,n})^2] \\ &= V(C_{i,n}|D) + (E[C_{i,n}|D] - \hat{C}_{i,n})^2 \end{aligned}$$

Cet écart mesure également l'incertitude présente dans les provisions :

$$\begin{aligned} MSEP(\hat{R}_i) &= E[(R_i - \hat{R}_i)^2] \\ &= E[(C_{i,n} - \hat{C}_{i,n})^2] \end{aligned}$$

Corolaire : Sous les hypothèses 1,2 et 3, l'erreur de prédiction moyenne pour les provisions de chaque année se calcule comme suit :

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

L'erreur de prédiction pour la totalité des provisions peut être estimée sous les hypothèses de Mack avec la formule suivante :

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

Les démonstrations de ces résultats sont présentées en Annexe.

Le calcul de l'erreur de prédiction (MSEP), permet d'en déduire la volatilité du risque de réserve, comme suggéré dans la Formule Standard :

$$\sigma = \frac{\sqrt{MSEP(\hat{R})}}{RES}$$

Le capital réglementaire à constituer est égal au quantile à 99,5% tel que :

$$X \sim \mathcal{LN}(RES; (\sigma RES)^2)$$

La Formule Standard retient cette modélisation comme prudente et permet l'approximation du capital réglementaire par $3 \cdot \sigma \cdot RES$.

1.3.2.2 Validation du modèle

Des contrôles graphiques des hypothèses 2 et 3 doivent être mis en place :

Vérification hypothèse 2 : Comme pour Chain-Ladder, pour j fixé, les $n-j$ couples $(C_{i,j}, C_{i,j+1})_{i \in [0, n-j-1]}$ doivent être relativement alignés sur une droite passant par l'origine.

Vérification hypothèse 3 : Pour tout j fixé et $\forall i = 1, \dots, n-j$ le nuage de points des paiements cumulés $(C_{i,j})_{i,j=0..n}$ et des résidus $(\frac{C_{i,j+1} - \hat{f}_j \cdot C_{i,j}}{\sqrt{C_{i,j}}})_{i,j=0..n}$, ne doit faire apparaître aucune tendance.

1.3.3 Mise en oeuvre dans le cadre du provisionnement

Les sinistres sont classés en deux catégories en fonction de l'estimation de leur charge à la fin de la première année de développement :

- Sinistres attritionnels : il s'agit des sinistres classiques et récurrents. Ces sinistres ont une fréquence importante et une sévérité peu élevée
- Sinistres atypiques : il s'agit de sinistres graves plus rares, mais ayant un coût élevé

La frontière entre sinistres attritionnels et atypiques appelée **Seuil de gravité**, sera déterminée par la suite.

Dans le cadre de la modélisation, les deux types de sinistre sont projetés de manière différente :

- * Pour les sinistres attritionnels, le triangle des paiements est projeté. Les montants des sinistres attritionnels étant faible, leur développement est court. Il est donc préférable d'évaluer le montant des règlements.

- * Les sinistres atypiques ayant un développement long, les charges donnent une meilleure représentation de ces sinistres. Pour cette raison, la projection s'effectue sur le triangle de charge.

Parallèlement aux calculs de provisions et de MSEP, un autre enjeu des deux méthodes de provisionnement développées par la suite est la simplification de l'application de traité de réassurance non proportionnel sinistre par sinistre. Le traité appliqué par la suite est présenté dans la section suivante.

1.4 Traité de réassurance non proportionnel : l'excédent de sinistre

La réassurance est un contrat par lequel l'assureur (le cédant) va se décharger d'une partie de son risque auprès d'une institution financière, le réassureur, en échange du versement d'une prime.

Dans ce mémoire, le traité de réassurance par excédents de sinistre par tête, du fait de la projection des sinistres atypiques de manière individuelle, a un intérêt tout particulier. En effet, la réassurance par excédents de sinistre est un traité de réassurance non proportionnel. Ce type de réassurance trouve toute son utilité lorsqu'un sinistre a un montant élevé et ainsi permet de lisser les pics de sinistralité importants. Ce traité de réassurance est recommandé lorsque les capitaux assurés sont difficiles à estimer et peuvent entraîner une forte volatilité des sinistres.

L'excédent de sinistre par tête noté également XS tête est un contrat dans lequel le réassureur s'engage à payer, dans une période donnée, un montant inférieur ou égal à la portée à chaque fois qu'une police est sinistrée pour un montant supérieur à la priorité. La limite est la somme de la priorité et de la portée.

L'assureur s'engage en contrepartie à payer une prime annuelle globale.

La figure suivante présente schématiquement le principe d'une réassurance par excédent de sinistre.

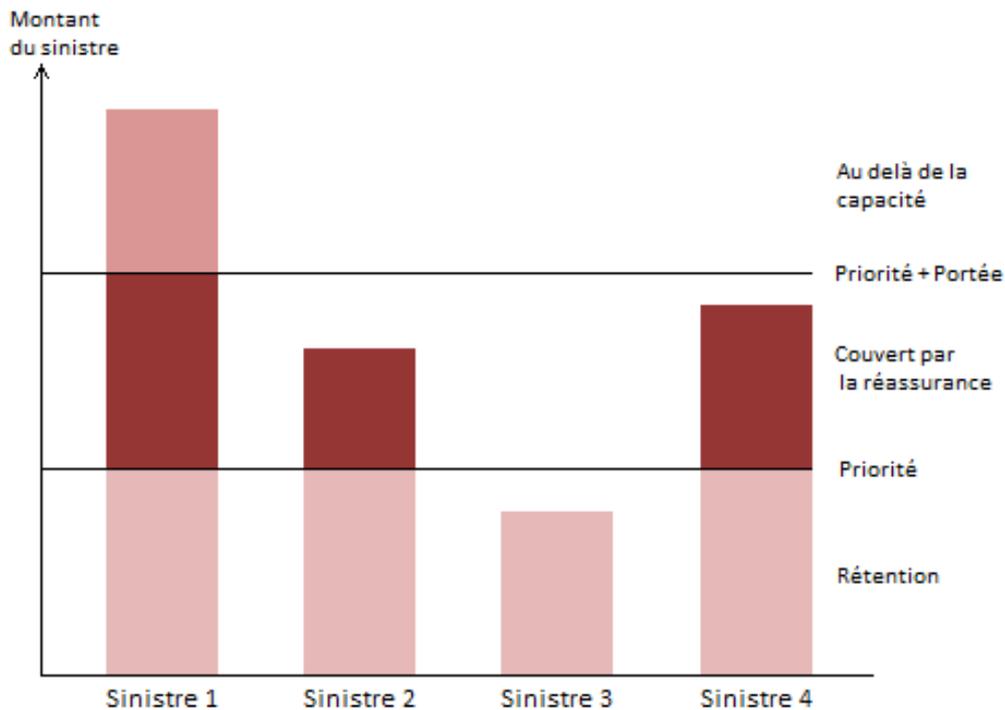


FIGURE 1.5 – XS par risque

Notation : *portée XS priorité*

Ce type de traité s'applique sinistre par sinistre. Dans la projection des sinistres avec la méthode de Chain-Ladder, les sinistres sont agrégés, l'application du traité sinistre par sinistre n'est pas applicable en l'état.

Dans le cadre du provisionnement des sinistres atypiques, entrant dans le périmètre de la réassurance, les assureurs procèdent à une application agrégée du traité. Le triangle supérieur est reconstruit : la charge des sinistres couverte par la réassurance est retranchée. Ce nouveau triangle est ainsi projeté avec la méthode de Chain-Ladder.

Le nouveau montant de règlement versé par l'assureur est déterminé par la même méthode. Le paiement est pris en charge par l'assureur lorsqu'il est inférieur à la *priorité* ou supérieur à la *portée + priorité*.

Chapitre 2

Problématique du mémoire

2.1 Objectif du mémoire

La vocation de ce mémoire est de présenter deux nouvelles méthodes alternatives aux méthodes de provisionnement actuelles permettant de mieux prendre en compte les différentes informations disponibles.

Outre l'intérêt d'estimer les provisions au plus juste, les méthodes développées permettent de simplifier la gestion des traités de réassurance appliqués sinistre par sinistre et donc en particulier les traités en excédents de sinistres.

L'approche développée pour estimer la charge des sinistres dans les deux modèles est une approche "*hybride*", les sinistres attritionnels étant toujours projetés avec la méthode de Chain-Ladder.

La première méthode est inspirée de l'article de Murphy et McLennan[18]. Les charges des sinistres atypiques sont projetées à l'aide de facteurs de développement individuels conditionnés à l'état du sinistre.

La deuxième méthode développée est une approche par apprentissage non supervisé. La charge ultime de chaque sinistre atypique est projetée en utilisant les caractéristiques du sinistre et les charges connues de l'assureur.

Ce mémoire attachera une grande importance à la robustesse des modèles développés. Ces méthodes seront comparées entre elles et les avantages et inconvénients de chacune seront détaillés.

2.2 Étude préalable de la base de données

La base de données utilisée provient d'un assureur français. Elle recense tous les sinistres de la branche automobile pour la garantie responsabilité civile corporelle de 2000 à 2015. Pour chacun des sinistres, nous disposons d'une image de chaque événement survenu au cours de son développement. La base contient également de nombreuses variables liées à la vie du sinistre, comme les caractéristiques du conducteur ou de l'évènement survenu par exemple.

La garantie Responsabilité civile corporelle, notée RCC, couvre les dommages corporels que le conducteur ou son véhicule ont causé à une tierce personne.

Le tableau suivant présente les statistiques de sinistralité vues à la date de déclaration et à la fin de l'année 2015.

Vision	Nombre de sinistres	Minimum	Premier quartile	Median	Moyenne	Troisième quartile	Maximum
Charge en fin de première année	71 407	- 75 000	600	3 049	5 142	3 200	4 081 000
Fin de l'année 2015	71 407	- 84 070	-	370	6 584	2 234	6 837 000

TABLE 2.1 – Statistiques sur les charges des sinistres

Le nombre de sinistres survenus entre 2000 et 2015 est de 71 407 pour une charge estimée moyenne de 5 142€ à la fin de l'année de déclaration et de 6 584€ à la fin de l'année 2015.

Le montant de la charge totale estimé sur la totalité du périmètre s'élève à 464 136 927€, vu à la fin de l'année 2015.

L'assureur, dans la pratique, segmente en deux sa base de sinistres par rapport à la charge estimée à la déclaration. Il construit deux triangles puis projette les sinistres avec deux méthodes de Chain-Ladder. Dans un premier temps, ce seuil est fixé arbitrairement à 20 000€.

Le tableau suivant présente les statistiques sur les charges estimatives supérieures à 20 000€, pour les sinistres vues à la date de déclaration et leur évolution à la fin de l'année 2015.

Vision	Nombre de sinistres	Minimum	Premier quartile	Median	Moyenne	Troisième quartile	Maximum
Charge en fin de première année	2 219	20 000	30 950	70 000	113 600	77 830	4 081 000
Fin de l'année 2015	2 219	- 7 950	16 610	42 120	124 400	96 380	6 837 000

TABLE 2.2 – Statistiques descriptives sur le montant des sinistres estimés à plus de 20 000€ à la date de déclaration

Le coût moyen des sinistres estimés à plus de 20 000€ augmentent entre leur date de déclaration et l'année 2015.

Les deux graphiques suivants représentent la répartition des charges à la date de déclaration et à la date de clôture.

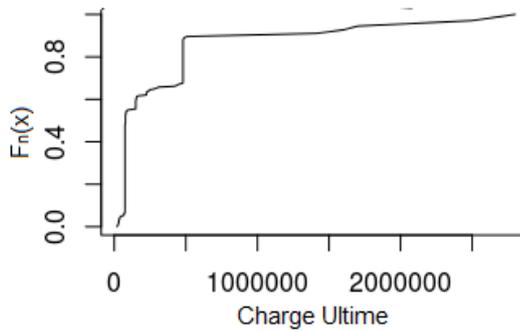


FIGURE 2.1 – Répartition de la charge des sinistres à la date de déclaration

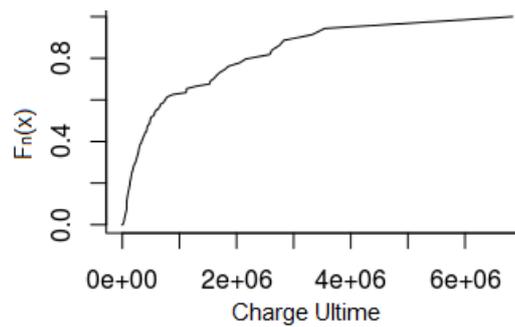


FIGURE 2.2 – Répartition de la charge des sinistres au cours de l'année 2015

La charge des sinistres augmentent entre la date de déclaration et l'année 2015. Les sinistres de la garantie responsabilité civile sont adaptés et bénéficient de suffisamment d'historique afin de réaliser un développement ligne à ligne.

Le choix du périmètre est important afin que le nombre de sinistres par année soit suffisant et qu'il n'y ait pas de disparité d'une année sur l'autre.

Le tableau suivant présente la charge de sinistralité vue à la fin de l'année de déclaration.

Année de survenance	Nombre de sinistres	Charge en fin de première année	Charge moyenne	Ecart-type
2000	3 152	12 228 404	3 880	2 252
2001	3 166	13 578 181	4 289	4 441
2002	3 387	13 180 824	3 892	5 328
2003	3 678	14 027 909	3 814	4 065
2004	3 745	18 038 718	4 817	4 261
2005	4 219	21 699 104	5 143	4 027
2006	4 620	28 749 896	6 223	13 742
2007	4 760	23 217 195	4 878	27 084
2008	4 553	25 392 154	5 577	21 444
2009	4 737	31 339 263	6 616	49 769
2010	5 443	23 087 143	4 242	12 885
2011	5 583	30 264 923	5 421	28 278
2012	4 852	26 211 413	5 402	18 614
2013	5 189	30 904 381	5 956	14 727
2014	6 164	33 873 753	5 495	14 912
2015	4 159	21 361 624	5 136	13 360
Base RCC	71 407	367 154 885	5 142	104 306

TABLE 2.3 – Statistiques descriptives de la charge des sinistres RCC à l'ouverture

La charge moyenne à la fin de l'année de déclaration est estimée à 5 142€.

L'histogramme suivant présente le nombre de sinistres par année de survenance, il permet de suivre son évolution.

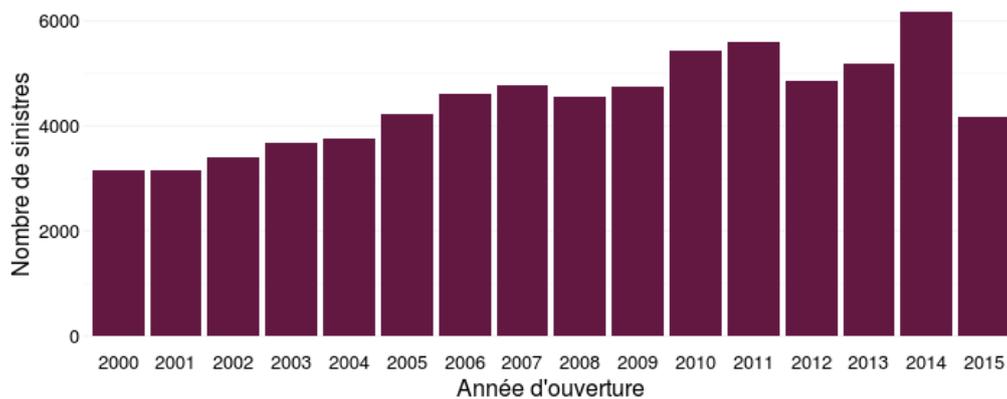


FIGURE 2.3 – Nombre de sinistres par année d'ouverture

Le nombre de sinistres augmente au cours du temps. L'année 2015 n'étant pas clôturée, celle-ci est retirée du périmètre de modélisation.

Les écoulements des sinistres sont représentés dans le graphique suivant. Ce graphique permet de visualiser la charge par année au cours du temps.

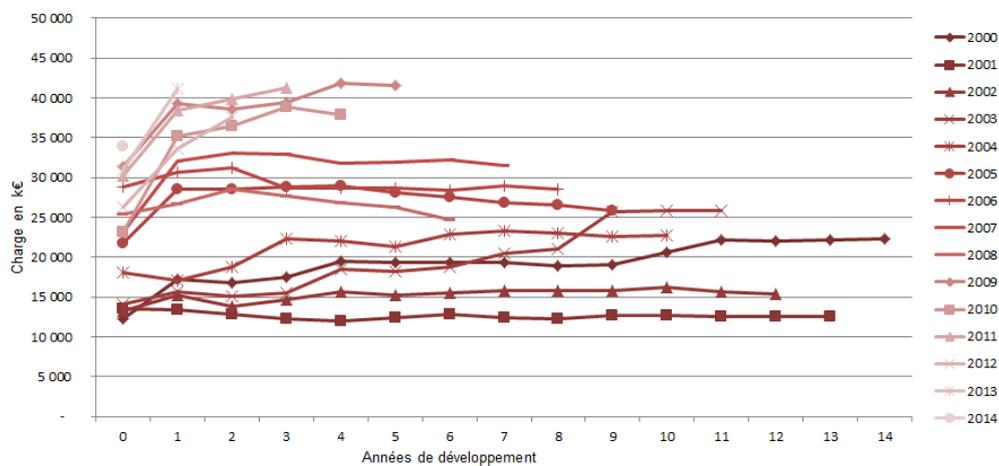


TABLE 2.4 – Ecoulements des sinistres

Le tableau suivant représente le triangle des règlements de la garantie RCC.

K€		Année de développement														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Année de survenance	2000	785	4 448	7 914	9 982	11 991	13 694	15 253	16 045	16 350	18 026	18 191	19 586	22 000	22 390	22 554
	2001	616	3 834	6 334	8 162	9 601	10 060	10 329	10 762	10 939	11 224	11 682	11 983	12 207	12 237	
	2002	868	4 613	7 776	9 652	11 948	12 521	13 206	13 875	14 224	14 326	14 550	14 595	14 691		
	2003	773	5 317	8 997	10 562	12 024	13 433	13 970	14 209	14 658	14 722	14 863	17 433			
	2004	740	5 212	9 190	11 877	14 584	15 055	15 595	17 062	17 596	17 778	18 803				
	2005	1 250	8 174	12 445	16 030	18 464	20 215	22 538	23 611	24 031	24 441					
	2006	1 546	7 713	12 710	16 013	18 873	21 849	23 432	24 039	24 753						
	2007	925	7 136	15 000	19 066	23 134	24 468	26 175	27 691							
	2008	1 605	8 048	13 755	17 961	20 095	21 787	22 456								
	2009	2 289	9 653	17 922	21 522	29 057	30 717									
	2010	2 222	11 084	17 962	22 841	26 370										
	2011	2 214	11 372	18 877	24 006											
	2012	1 621	9 532	17 036												
	2013	2 114	10 234													
2014	2 158															

TABLE 2.5 – Triangle des paiements

Le montant des règlements a été multiplié par trois entre 2000 et 2014. Cela signifie que l'activité est en forte croissance sur cette période.

Le triangle des charges présenté dans le tableau suivant confirme la croissance de l'assureur.

K€		Année de développement														
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Année de survenance	2000	12 228	17 230	16 851	17 547	19 544	19 375	19 326	19 319	18 953	19 037	20 576	22 132	22 059	22 212	22 329
	2001	13 578	13 436	12 777	12 249	11 994	12 431	12 800	12 338	12 313	12 662	12 610	12 541	12 479	12 494	
	2002	13 181	15 247	13 874	14 603	15 659	15 265	15 511	15 783	15 845	15 793	16 264	15 595	15 326		
	2003	14 028	15 587	15 092	15 551	18 434	18 252	18 829	20 484	20 984	25 721	25 912	25 861			
	2004	18 039	17 134	18 775	22 287	21 954	21 350	22 919	23 246	22 947	22 657	22 676				
	2005	21 699	28 527	28 541	28 857	28 937	28 145	27 595	26 892	26 487	25 895					
	2006	28 750	30 585	31 229	28 671	28 604	28 647	28 330	28 898	28 569						
	2007	23 217	32 072	32 995	32 982	31 746	31 970	32 155	31 558							
	2008	25 392	26 722	28 589	27 674	26 884	26 308	24 779								
	2009	31 339	39 284	38 597	39 463	41 774	41 559									
	2010	23 087	35 195	36 513	38 806	37 898										
	2011	30 265	38 457	39 910	41 221											
	2012	26 211	33 689	37 630												
	2013	30 904	41 107													
2014	33 874															

TABLE 2.6 – Triangle des charges

La base de données fait référence à une unique ligne d'activité, qui sous un référentiel Solvabilité 2, appartient à ligne d'activité 4 « Assurance de responsabilité civile automobile ».

La fiabilité des données joue un rôle très important dans le calcul des provisions techniques. Dans le cadre de ce mémoire, la seule base à notre disposition est la base de sinistre. Les données comptables ne sont pas disponibles afin d'assurer la vérification des retraitements effectués sur des données.

Les données seront considérées suffisamment fiables, cependant aucun contrôle ne le vérifiera. Cette hypothèse restera admise dans la suite de ce mémoire.

Chapitre 3

Estimation des provisions par méthode agrégée

3.1 Choix du seuil de gravité

Le périmètre d'étude étant fixé, le seuil de gravité doit être déterminé afin de segmenter la base en deux et différencier les sinistres atypiques des sinistres attritionnels.

L'objectif est de trouver le seuil qui minimise l'erreur sur la projection de l'ultime pour les sinistres attritionnels, par la méthode de Chain-Ladder.

1. Le périmètre temporel est à définir avec les cadences de règlements. Dans cette étude, les années varient entre 2000-2009.
2. Le seuil de gravité est défini. Ce seuil permet de diviser la base de sinistres en deux. Les sinistres dont la charge à la fin de la première année est supérieure au seuil sont considérés comme atypiques.
3. Le triangle des paiements des sinistres attritionnels est construit.
4. Le triangle des paiements est projeté avec la méthode de Chain-Ladder. Les charges ultimes obtenues sont comparées aux charges ultimes réelles connues.

La figure suivante représente les taux d'erreurs entre les paiements projetés avec Chain-Ladder en date de calcul et les charges ultimes réellement observées à fin 2015.

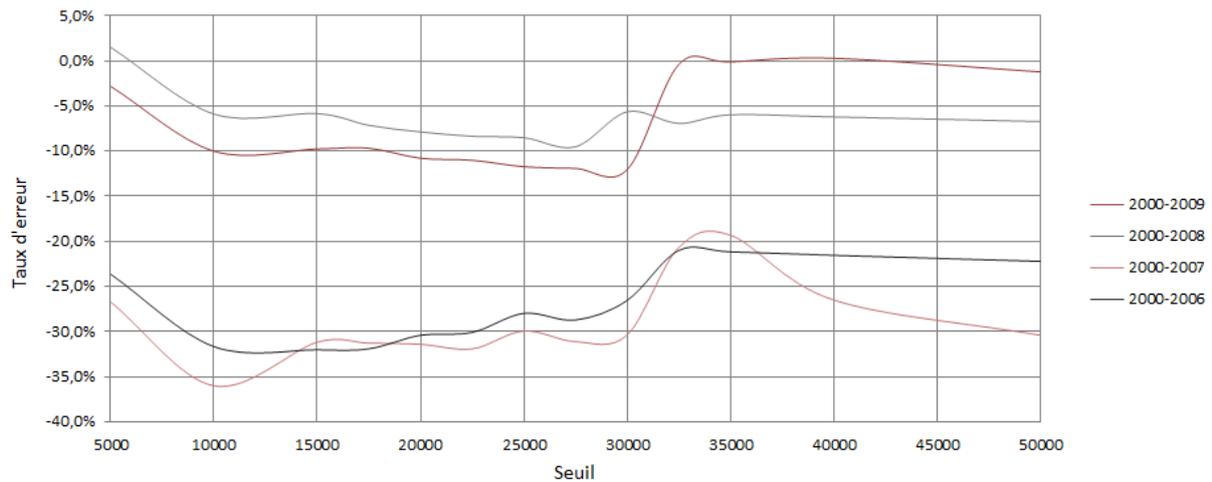


FIGURE 3.1 – Erreur de la charge ultime totale en fonction du seuil de coupure

Le taux d'erreur est minimisé sur chaque segment lorsque le seuil est compris entre 30 000 et 35 000. Afin d'affiner la précision et de confirmer ces observations, le graphique suivant représente l'erreur d'estimation de l'ultime pour les années 2006 à 2009, sur une vision fin de première année.

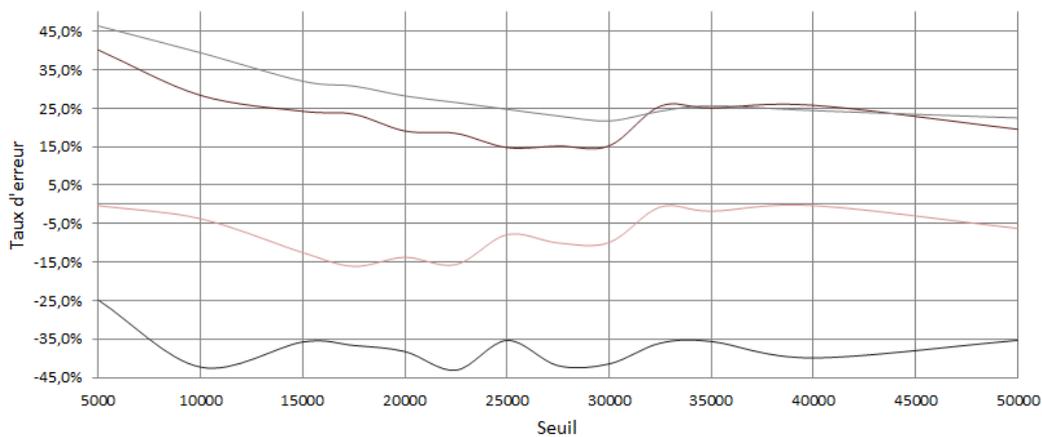


FIGURE 3.2 – Erreur de la charge ultime de chaque année en fonction du seuil de coupure

L'erreur sur la projection de la charge ultime est minimisée pour les années 2006 et 2007 pour un seuil compris entre 30 000 et 35 000. Pour les années 2008 et 2009 cette erreur est minimisée pour un seuil compris entre 25 000 et 30 000.

La modélisation des provisions pour les sinistres attritionnels étant faite de manière agrégée, il convient de choisir un seuil qui minimise l'erreur au global. Au vu des graphiques, ce seuil se situe entre 30 000 et 35 000. Dans la suite du mémoire ce seuil sera fixé à **32 500**.

Triangles de liquidation

Le seuil de gravité étant fixé à 32 500€, la base de sinistres est segmenté en deux : 1 530 sinistres sont évalués à plus de 32500€ à la fin de la première année contre 65 718 pour la base des sinistres attritionnels. Le tableau suivant présente la répartition des charges estimées à fin 2015 pour les sinistres survenus entre 2000 et 2014.

Sinistres	Nombre de sinistres	Charge estimée
Attritionnels	65 718	225 476 442
	97,7%	50,9%
Atypiques	1 530	217 298 858
	2,3%	49,1%
Total	67 248	442 775 300

TABLE 3.1 – Répartition des montants de charges et de nombres

La charge des sinistres atypiques représente 49,1% de la charge totale pour un montant de 442,78M€, alors qu'elle ne représente en nombre que 2,3% des sinistres, d'où l'utilité d'affiner leurs prédictions.

Le tableau suivant représente le triangle de paiements des sinistres attritionnels.

K€	Année de développement														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2000	313	2 513	4 491	5 999	7 577	8 079	8 625	8 966	9 096	10 726	10 863	10 942	13 204	13 206	13 207
2001	342	2 197	3 911	5 292	6 023	6 328	6 502	6 745	6 831	7 113	7 414	7 446	7 477	7 477	
2002	338	2 842	4 840	5 969	6 995	7 622	7 975	8 515	8 681	8 782	8 966	8 983	9 074		
2003	468	3 904	6 338	7 527	8 259	8 982	9 182	9 325	9 693	9 744	9 801	12 361			
2004	455	3 278	6 372	7 809	8 597	8 853	9 232	9 855	10 274	10 445	11 471				
2005	441	4 767	8 014	9 790	11 080	12 503	13 047	13 214	13 401	13 794					
2006	440	3 813	7 258	9 205	10 204	10 592	11 006	11 212	11 362						
2007	522	4 986	9 987	12 910	14 849	15 780	16 729	17 050							
2008	848	4 870	8 808	10 949	11 650	12 238	12 515								
2009	869	6 038	10 458	12 574	13 696	14 399									
2010	1 001	6 527	11 327	14 435	16 279										
2011	1 077	6 423	11 411	13 633											
2012	838	5 847	10 123												
2013	982	6 209													
2014	1 000														

TABLE 3.2 – Triangle des paiements des sinistres attritionnels

Le montant des paiements totaux réalisés sur cette période est de 188M€.

Concernant les sinistres graves, les charges sont projetées :

K€	Année de développement														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2000	4 255	7 153	7 029	7 545	7 649	7 738	7 750	7 626	7 267	7 309	7 468	9 019	8 947	9 116	9 288
2001	6 391	6 031	5 589	5 112	5 098	5 195	5 069	5 059	4 990	5 284	5 003	5 074	5 099	5 099	
2002	6 576	7 850	6 236	6 393	6 709	6 225	6 419	6 398	6 544	6 545	6 470	6 070	5 758		
2003	5 904	5 159	5 017	5 101	6 335	6 402	5 901	5 856	6 409	6 916	6 895	6 837			
2004	8 247	7 726	7 109	7 788	7 457	7 625	7 961	8 462	8 173	7 878	7 817				
2005	9 297	13 035	13 768	13 393	13 345	12 352	12 204	11 801	11 974	11 451					
2006	17 570	18 346	18 176	16 805	16 948	16 993	16 432	16 289	15 964						
2007	12 297	13 867	13 608	14 888	13 432	13 768	13 943	13 275							
2008	14 511	14 201	15 669	14 005	13 707	13 229	11 793								
2009	19 782	24 107	22 901	23 250	25 364	25 229									
2010	10 549	17 877	17 285	18 576	19 211										
2011	17 190	22 071	22 452	23 274											
2012	14 595	18 089	21 205												
2013	17 763	22 990													
2014	18 108														

TABLE 3.3 – Triangle des charges des sinistres atypiques

Les paiements réalisés à la fin de l'année 2014 pour chaque année de survenance sont présentés dans le tableau suivant.

Paiements	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
	9 347	4 760	5 617	5 072	7 332	10 646	13 391	10 641	9 941	16 317	10 090	10 373	6 913	4 025	1 158

TABLE 3.4 – Paiements réalisés pour les sinistres atypiques à la fin de l'année 2014

Au total, le montant des paiements effectués à la fin de l'année 2014 pour les sinistres graves s'élève à 125,6M€.

3.2 Application des méthodes agrégées

La base des sinistres est divisée en deux parties par rapport au seuil défini dans la section précédente. Les sinistres attritionnels correspondent aux sinistres dont la charge à la fin de la première année est inférieure au seuil. Le triangle des sinistres attritionnels est un triangle des paiements cumulés alors que le triangle des sinistres atypiques est un triangle de charge cumulé.

3.2.1 Application de la Méthode de Chain-Ladder

3.2.1.1 Vérification des hypothèses

Sinistres attritionnels

Sous l'hypothèse d'indépendance des années de survenance, la relation suivante doit être vérifiée :

$$- \forall j \in [0, n - 1],$$

$$\frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = \dots = \frac{C_{n-j-1,j+1}}{C_{n-j-1,j}}$$

Pour $j = 1, 2, 3, 4$ et $\forall i = 0, \dots, n - j - 1$ le nuage de point $(C_{i,j}, C_{i,j+1})$ est tracé :

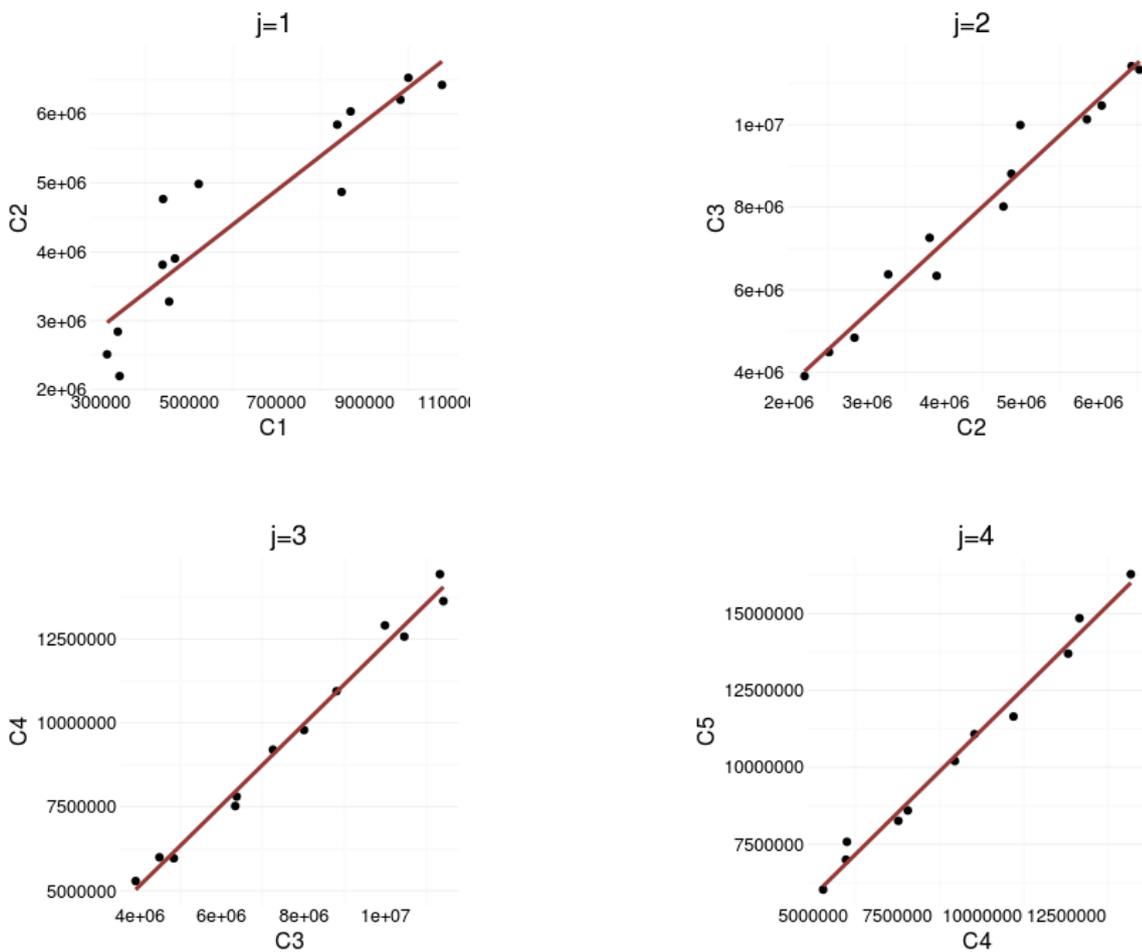


FIGURE 3.3 – Vérification de l'existence de facteurs de développement

Une tendance linéaire se dessine, cela vérifie l'hypothèse de l'existence de facteurs de développement.

La deuxième hypothèse à vérifier est l'indépendance entre les facteurs de développement. Pour $j = 1, \dots, 4$, les facteurs de développement individuels doivent être constants. La figure représente les facteurs de développement notés $(f_{i,j})_{i=1, \dots, n-j-1}$ en fonction des années de survenance.

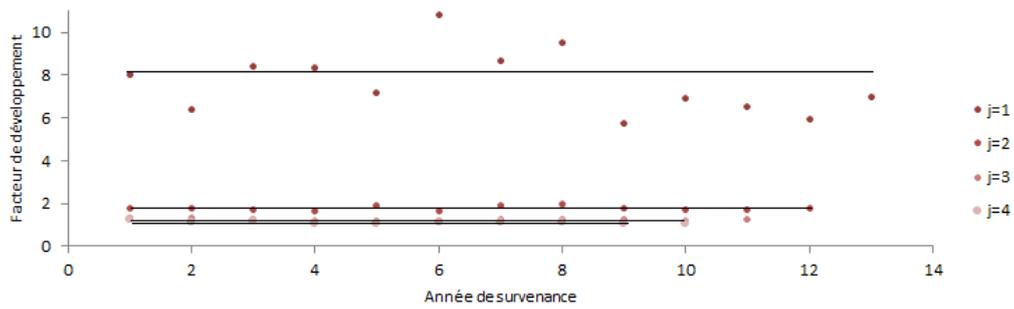


FIGURE 3.4 – Indépendance des facteurs de développement

La tendance est constante, l'hypothèse d'indépendance entre les facteurs de développement est vérifiée.

Sinistres atypiques

Les deux hypothèses de Chain-Ladder sont vérifiées comme le montrent les graphiques suivants :

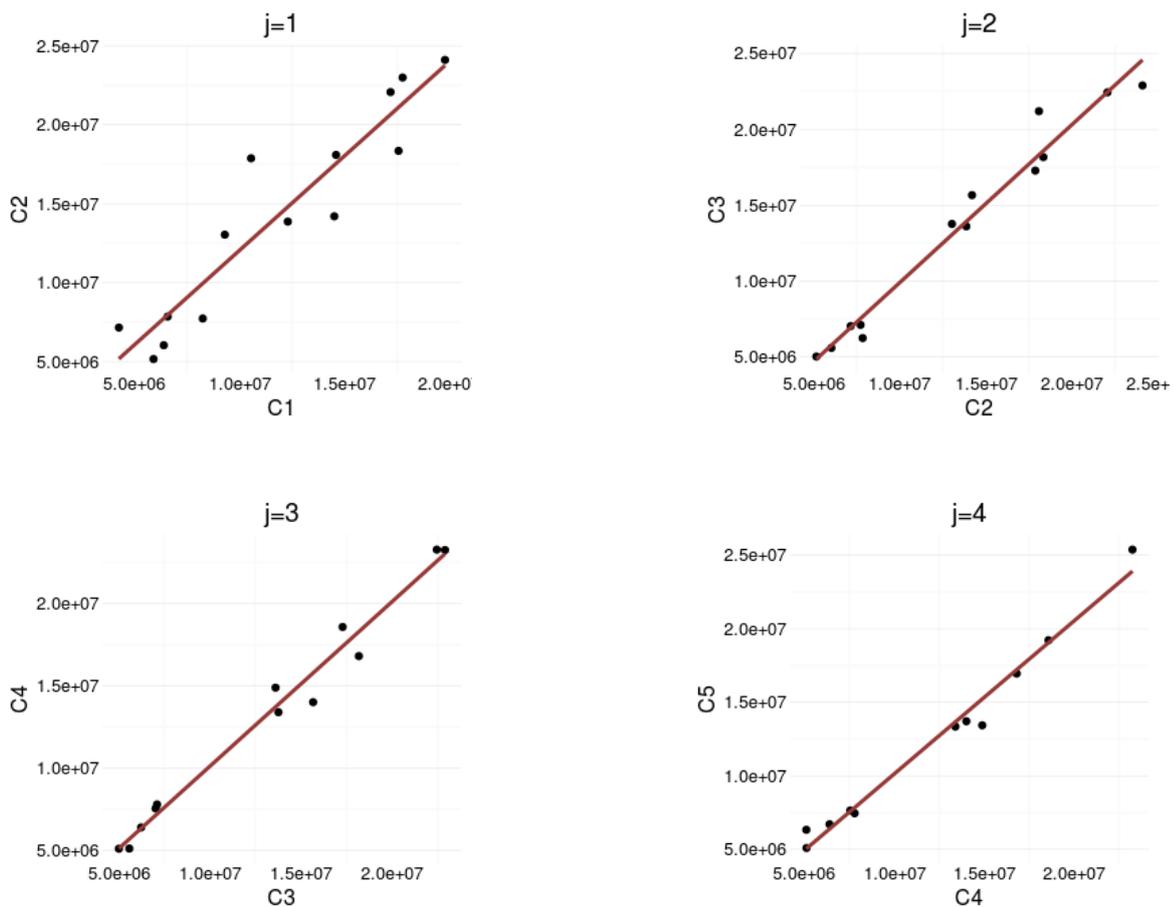


FIGURE 3.5 – Vérification de l'existence de facteurs de développement

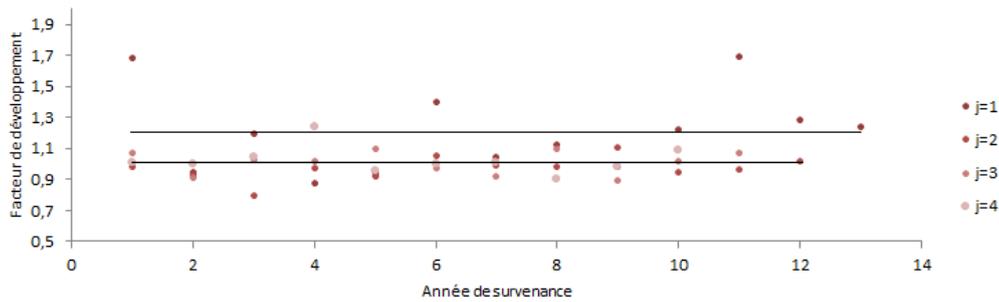


FIGURE 3.6 – Indépendance des facteurs de développement

Le triangle des sinistres graves étant projeté en charges, les facteurs de développement sont moins élevés que pour le triangle des paiements car l'assureur estime la charge de sinistres au plus proche dès la première année.

3.2.1.2 Application aux triangles de liquidation

Sinistres attritionnels

Après application de la méthode de Chain-Ladder, les facteurs de développement suivant sont obtenus. Les cadences de règlements sont ensuite déduites. Ces résultats sont présentés dans le tableau suivant :

Année de développement	Facteur de développement	Cadences cumulées	Cadences décumulées
0	7,188	4%	4%
1	1,781	27%	23%
2	1,245	48%	21%
3	1,124	60%	12%
4	1,065	68%	7%
5	1,042	72%	4%
6	1,031	75%	3%
7	1,022	77%	2%
8	1,045	79%	2%
9	1,036	83%	4%
10	1,073	86%	3%
11	1,087	92%	6%
12	1,000	100%	8%
13	1,000	100%	0%
14	1,000	100%	0%

TABLE 3.5 – Facteurs de développement et cadences des règlements

Le triangle est développé à l'aide des facteurs de développement présentés précédemment. Les provisions sont ensuite calculées :

$$provision = Charges Ultimes - Paiements$$

K€	Année de développement														Paiement	Réserve	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13			14
2000	313	2 513	4 491	5 999	7 577	8 079	8 625	8 966	9 096	10 726	10 863	10 942	13 204	13 206	13 207	13 207	-
2001	342	2 197	3 911	5 292	6 023	6 328	6 502	6 745	6 831	7 113	7 414	7 446	7 477	7 477	7 478	7 477	1
2002	338	2 842	4 840	5 969	6 995	7 622	7 975	8 515	8 681	8 782	8 966	8 983	9 074	9 075	9 076	9 074	2
2003	468	3 904	6 338	7 527	8 259	8 982	9 182	9 325	9 693	9 744	9 801	12 361	13 438	13 439	13 441	12 361	1 080
2004	455	3 278	6 372	7 809	8 597	8 853	9 232	9 855	10 274	10 445	11 471	12 303	13 375	13 376	13 377	11 471	1 907
2005	441	4 767	8 014	9 790	11 080	12 503	13 047	13 214	13 401	13 794	14 296	15 333	16 669	16 671	16 673	13 794	2 878
2006	440	3 813	7 258	9 205	10 204	10 592	11 006	11 212	11 362	11 878	12 310	13 203	14 353	14 354	14 356	11 362	2 994
2007	522	4 986	9 987	12 910	14 849	15 780	16 729	17 050	17 429	18 219	18 882	20 252	22 016	22 018	22 021	17 050	4 971
2008	848	4 870	8 808	10 949	11 650	12 238	12 515	12 908	13 195	13 793	14 295	15 332	16 668	16 669	16 672	12 515	4 156
2009	869	6 038	10 458	12 574	13 696	14 399	15 006	15 477	15 821	16 539	17 141	18 384	19 985	19 987	19 990	14 399	5 591
2010	1 001	6 527	11 327	14 435	16 279	17 340	18 071	18 639	19 053	19 917	20 642	22 139	24 067	24 070	24 073	16 279	7 793
2011	1 077	6 423	11 411	13 633	15 330	16 328	17 017	17 551	17 941	18 755	19 437	20 847	22 663	22 665	22 668	13 633	9 035
2012	838	5 847	10 123	12 608	14 177	15 101	15 738	16 232	16 592	17 345	17 976	19 280	20 960	20 962	20 964	10 123	10 841
2013	982	6 209	11 061	13 776	15 491	16 500	17 196	17 736	18 129	18 952	19 641	21 066	22 901	22 904	22 906	6 209	16 698
2014	1 000	7 191	12 811	15 956	17 941	19 110	19 916	20 541	20 997	21 950	22 748	24 399	26 524	26 526	26 530	1 000	25 529
Total																169 956	93 476

FIGURE 3.7 – Triangle des paiements développés et provisions

La charge ultime des sinistres attritionnels s'élève à 263,43M€ et la valeur des provisions, non actualisée est de 93,48M€.

Sinistres atypiques

Le tableau suivant présente les facteurs de développement calculés avec le triangle de charges des sinistres graves.

Année de développement	Facteur de développement	Cadences cumulées	Cadences décumulées
0	1,204	81%	81%
1	1,003	97%	16%
2	1,008	97%	0%
3	1,018	98%	1%
4	0,989	100%	2%
5	0,977	99%	-1%
6	0,988	97%	-2%
7	0,997	95%	-1%
8	1,001	95%	0%
9	0,992	96%	0%
10	1,045	96%	1%
11	0,982	98%	2%
12	1,012	99%	1%
13	1,019	100%	1%
14	1,000	100%	0%

TABLE 3.6 – Facteurs de développement des sinistres graves

Le triangle des sinistres atypiques est présenté dans le tableau suivant.

K€	Année de développement														Paiement	Réserve		
	0	1	2	3	4	5	6	7	8	9	10	11	12	13			14	
2000	4 255	7 153	7 029	7 545	7 649	7 738	7 750	7 626	7 267	7 309	7 468	9 019	8 947	9 116	9 288	9 347	-	59
2001	6 391	6 031	5 589	5 112	5 098	5 195	5 069	5 059	4 990	5 284	5 003	5 074	5 099	5 099	5 194	4 760		434
2002	6 576	7 850	6 236	6 393	6 709	6 225	6 419	6 398	6 544	6 545	6 470	6 070	5 758	5 827	5 936	5 617		319
2003	5 904	5 159	5 017	5 101	6 335	6 402	5 901	5 856	6 409	6 916	6 895	6 837	6 715	6 796	6 924	5 072		1 852
2004	8 247	7 726	7 109	7 788	7 457	7 625	7 961	8 462	8 173	7 878	7 817	8 170	8 024	8 120	8 273	7 332		940
2005	9 297	13 035	13 768	13 393	13 345	12 352	12 204	11 801	11 974	11 451	11 358	11 869	11 657	11 798	12 019	10 646		1 373
2006	17 570	18 346	18 176	16 805	16 948	16 993	16 432	16 289	15 964	15 973	15 842	16 556	16 260	16 456	16 765	13 391		3 374
2007	12 297	13 867	13 608	14 888	13 432	13 768	13 943	13 275	13 239	13 246	13 138	13 730	13 485	13 647	13 903	10 641		3 262
2008	14 511	14 201	15 669	14 005	13 707	13 229	11 793	11 651	11 619	11 625	11 530	12 050	11 835	11 977	12 202	9 941		2 261
2009	19 782	24 107	22 901	23 250	25 364	25 229	24 650	24 353	24 286	24 300	24 100	25 186	24 737	25 034	25 504	16 317		9 187
2010	10 549	17 877	17 285	18 576	19 211	18 998	18 562	18 338	18 288	18 298	18 148	18 966	18 627	18 851	19 205	10 090		9 115
2011	17 190	22 071	22 452	23 274	23 694	23 431	22 893	22 617	22 555	22 568	22 383	23 391	22 974	23 250	23 687	10 373		13 314
2012	14 595	18 089	21 205	21 382	21 769	21 527	21 033	20 779	20 722	20 734	20 564	21 490	21 107	21 361	21 762	6 913		14 849
2013	17 763	22 990	23 060	23 252	23 672	23 409	22 872	22 596	22 534	22 547	22 362	23 369	22 952	23 228	23 664	4 025		19 640
2014	18 108	21 794	21 860	22 042	22 441	22 191	21 682	21 421	21 362	21 374	21 199	22 154	21 759	22 020	22 434	1 158		21 276
Total																125 622		101 137

FIGURE 3.8 – Triangle des charges développées et provisions

La charge ultime totale des sinistres atypiques est de 226,76M€, les paiements réalisés à la fin de l'année 2014 s'élèvent à 125,62M€. Le montant des provisions à constituer est donc de 101,14M€.

Calcul des provisions

Le montant des provisions pour les :

- Sinistres attritionnels est de 93,48M€
- Sinistres atypiques est de 101,14M€

Le montant des provisions totales est de 194,61M€

3.3 Evolution de la sinistralité

Avant d'étudier l'évolution des provisions projetées avec la méthode de Chain-Ladder, une présentation de la sinistralité atypique est proposée. Dans un premier temps, le périmètre d'étude est fixé puis l'évolution des charges ultimes par exercice est présentée.

Les cadences cumulées présentées dans le tableau 3.6 montre qu'après cinq années de développement la charge ultime est connue. La charge ultime des sinistres survenus entre 2000 et 2011 est connue à la fin de l'année 2015. La base de test sera constituée de ces sinistres pour étudier la robustesse de la prédiction. Afin de visualiser la tendance de la sinistralité pour chacun des exercices, le graphe suivant représente l'évolution de la charge ultime pour les exercices 2000 à 2011.

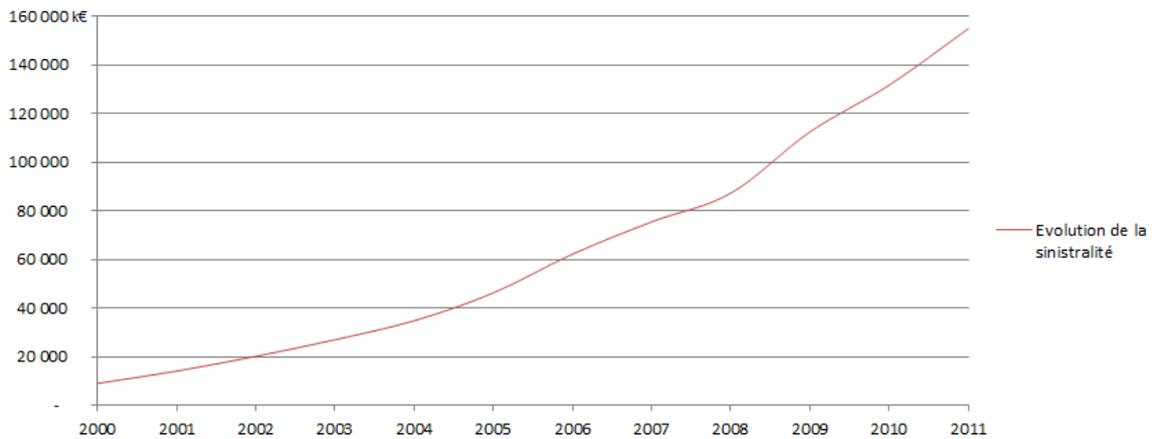


FIGURE 3.9 – Evolution de la sinistralité entre 2000 et 2011

La sinistralité a une tendance haussière et lisse sur le périmètre représenté. Le graphique 3.10 présente l'évolution du montant de provisions sur ce même périmètre observé en 2015.

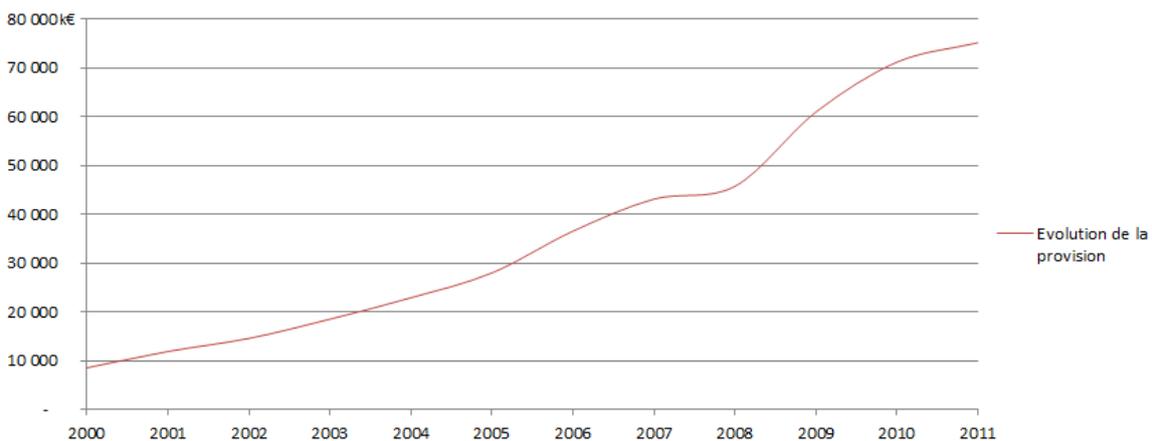


FIGURE 3.10 – Evolution de la provision entre 2000 et 2011

Comme les charges ultimes, les provisions ont une tendance haussière et monotone. L'estimation des provisions au cours du temps devra reproduire cette tendance.

3.4 Evolution de l'estimation des provisions

Afin de suivre l'évolution du montant des provisions au cours du temps, la méthode de Chain-Ladder a été appliquée pour les exercices de 2007 à 2014. Huit triangles de charges des sinistres atypiques ont été projetés au 31/12/N, $N = 2007, \dots, 2014$. Pour chacun des triangles, la somme des paiements au 31/12/N a été retranchée afin d'obtenir le montant des provisions. Le graphique suivant permet de visualiser l'évolution des provisions que l'assureur doit constituer au cours de chaque exercice N.

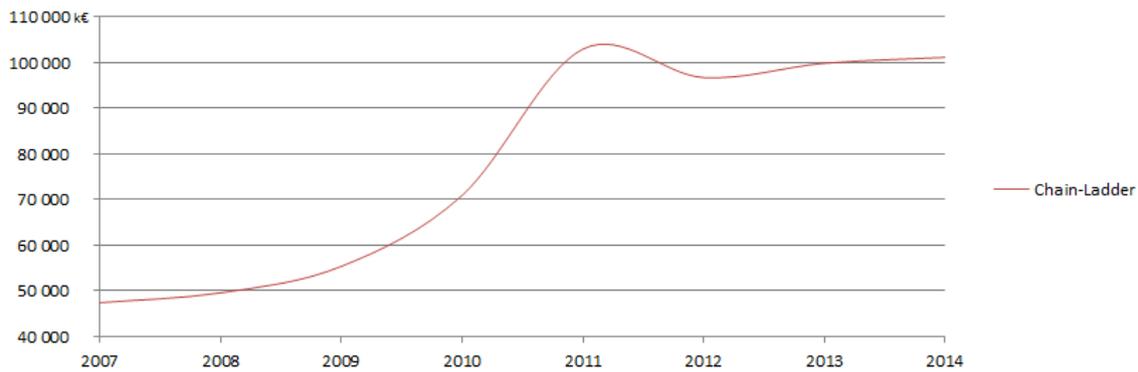


FIGURE 3.11 – Evolution des provisions estimées avec la méthode de Chain-Ladder entre 2007 et 2014

La courbe précédente montre que l'estimation des provisions avec la méthode de Chain-Ladder n'est pas stable dans le temps. Entre 2009 et 2011, l'assureur a estimé que ses provisions augmentaient de 86%. Entre 2011 et 2012, l'estimation des provisions a diminué de 8%. Ce graphique montre que l'estimation des provisions avec la méthode de Chain-Ladder n'est pas stable d'un exercice à l'autre alors que la tendance observée sur le graphique 3.10 laisse apparaître une tendance haussière et monotone.

3.5 Calcul de l'écart d'estimation sur le périmètre 2007-2011

Afin de valider le modèle de Chain-Ladder, les montants des provisions estimées pour les exercices 2007 à 2011 sont comparés aux montants de provisions réels, c'est-à-dire le montant de provisions que l'assureur devait avoir en sa possession pour faire face à ses engagements lors de ces cinq exercices. Ce montant est constaté lors de l'année 2015. Le graphique suivant permet de visualiser l'erreur faite sur ces exercices.

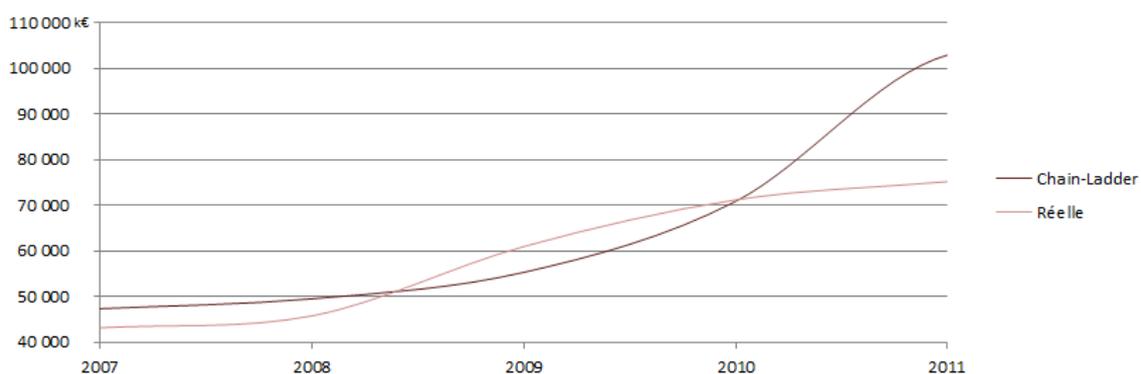


FIGURE 3.12 – Comparaison des provisions réelles et des provisions estimées avec la méthode de Chain-Ladder

Les provisions estimées avec la méthode de Chain-Ladder ne permettent pas de déterminer un montant juste de provisions à constituer. Cette projection surestime le montant de provisions à constituer

lors des exercices 2007, 2008 et 2011 et sous estime pour 2009 et 2010. Le tableau suivant permet de visualiser les écarts d'estimation de la méthode entre Chain-Ladder et l'estimation réelle.

	2007	2008	2009	2010	2011
Réserves réelles	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%

TABLE 3.7 – Ecarts de l'estimation des provisions avec Chain-Ladder sur les exercices 2007 à 2011

Ce tableau complète le graphique précédent et permet de voir que l'estimation des provisions avec la méthode de Chain-Ladder est très volatile au cours du temps. L'erreur moyenne d'estimation sur cette période est de 12,91%. Ces tests fournissent une base de calcul afin de pouvoir estimer la qualité des modèles traités dans les chapitres suivants.

3.6 Application de la Méthode de Mack

3.6.1 Vérification des hypothèses

Les hypothèses de la méthode de Chain-Ladder étant vérifiées, il reste à montrer que le nuage de points $\left(C_{i,j}, \frac{C_{i,j+1} - \hat{f}_j \cdot C_{i,j}}{\sqrt{C_{i,j}}}\right)_{i,j=0..n}$ ne présente aucune structure non aléatoire.

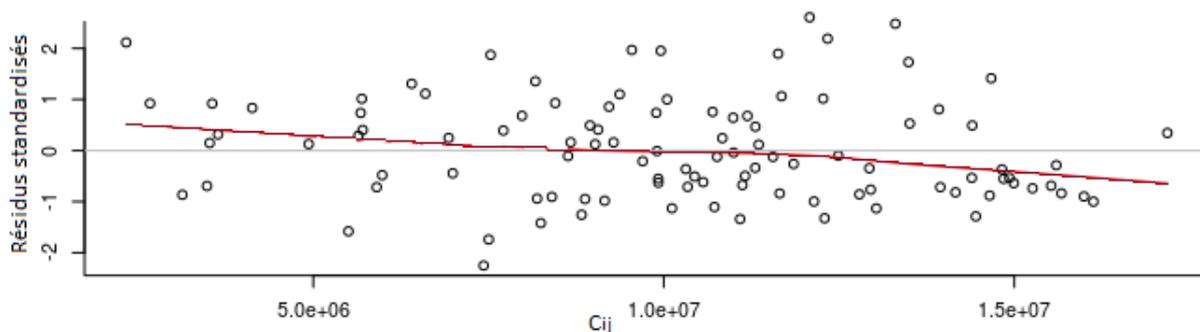


FIGURE 3.13 – Graphique des résidus des sinistres attritionnels

Les nuages de points des deux figures ne font apparaître aucune tendance, les hypothèses du modèle de Mack sont vérifiées.

3.6.2 Calcul des volatilités

La méthode de Chain-Ladder peut être complétée par un calcul de volatilités des paiements et charges cumulés, comme le stipule l'hypothèse 3 de Mack :

Pour $j \in [0, n]$, $\exists \sigma_j^2$ tq. :

$$V[C_{i,j+1} | C_{i,1}, \dots, C_{i,j}] = \sigma_j^2 \cdot C_{i,j}, \text{ pour } i = 0, \dots, n$$

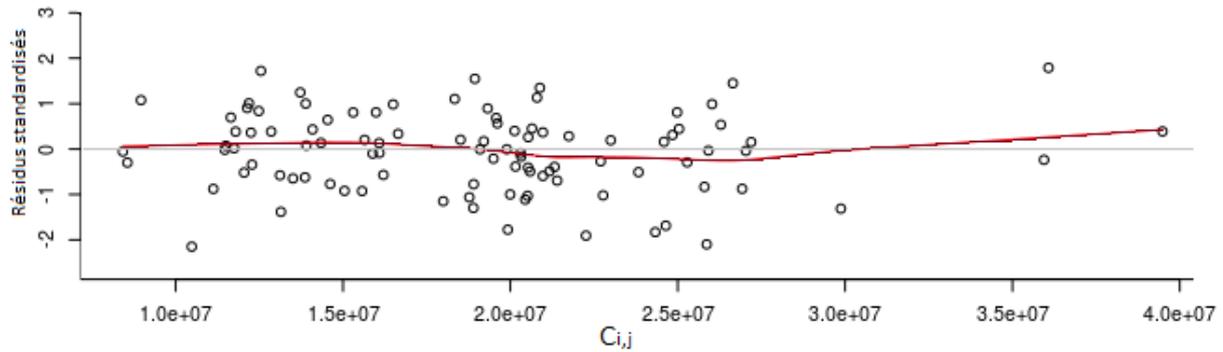


FIGURE 3.14 – Graphique des résidus sinistres atypiques

Année de développement	Sinistres attritionnels	Sinistres atypiques
0	1 081,595	551,689
1	44,837	77,341
2	13,841	55,043
3	18,774	60,744
4	7,490	14,282
5	1,719	17,437
6	3,991	6,765
7	1,273	18,909
8	32,844	21,497
9	3,500	3,926
10	31,414	77,982
11	69,249	0,530
12	0,000	0,859
13	0,000	0,530
14	-	-

TABLE 3.8 – Paramètres de volatilités

La volatilité des sinistres attritionnels est calculée sur les estimations des paiements qui sont d'avantage volatiles que les estimations de charges. Ce phénomène a par ailleurs déjà été observé sur les facteurs de développement de Chain-Ladder.

Les facteurs de développement pour les deux types de sinistres lors de la dernière année de développement sont égaux à 1. Le paramètre de volatilité de la dernière année est nul pour la quatorzième année de développement.

3.6.3 Calcul du Capital Réglementaire

Le modèle de Mack permet de calculer la volatilité sur les provisions par année de développement pour chaque base de sinistres. Il est alors possible de calculer les erreurs moyennes de prédiction sur les provisions par année de survenance ($M\hat{S}EP(\hat{R}_i)$). Le tableau suivant présente ces calculs.

Année de survenance	Sinistres attritionnels	Sinistres atypiques
2000	-	-
2001	0,00	832
2002	0,03	2 117
2003	100 545,04	2 962
2004	145 132,99	108 860
2005	160 691,53	126 028
2006	200 779,07	172 789
2007	231 397,44	186 946
2008	217 273,40	188 123
2009	232 527,28	265 035
2010	260 215,78	254 957
2011	286 212,76	347 411
2012	304 826,68	401 813
2013	415 714,48	499 197
2014	4 869 570,00	1 090 875

TABLE 3.9 – MSEP par année de survenance

Une fois les *MSEP* obtenues, Mack explicite une formule fermée permettant d'obtenir la MSEP des provisions globales. La volatilité des provisions et le capital réglementaire sont calculés dans le tableau suivant :

	Sinistres attritionnels	Sinistres atypiques
$MSEP(\hat{R})$	40 571	14 396
σ	0,03748	0,03020
Réserve	93 476	101 137
Capital Réglementaire	10 510	9 164

TABLE 3.10 – Capital réglementaire pour la méthode agrégée

Le capital réglementaire que l'assureur doit détenir est de 19,67M€, conformément au calcul de la formule standard.

3.7 Application d'un traité de réassurance non proportionnel : Excédent de sinistre

Afin de pouvoir comparer les trois modèles de projection des sinistres sur plusieurs points opérationnels, le traité de réassurance mis en place chez l'assureur est appliqué avec la méthode agrégée dans la section suivante.

3.7.1 Calcul des provisions à constituer

Le traité de réassurance présenté est celui que l'assureur a mis en place avec un réassureur. Il s'agit d'un traité par excédents de sinistre.

- La priorité est fixée à 130 000€.
- La portée est fixée à 230 000€.

Le réassureur effectue les règlements pour un sinistre lorsque ceux-ci sont compris entre 130 000€ et 360 000€. Afin de simplifier l'étude, le traité de réassurance est appliqué uniquement à la sinistralité atypique.

Le tableau suivant présente le nouveau montant de paiement par année de survenance que l'assureur a versé.

Paielements	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
	7 088	4 184	4 179	4 057	5 222	8 523	9 881	8 130	6 990	13 607	7 722	8 590	5 371	3 564	1 084

TABLE 3.11 – Règlements par année de survenance

Le montant de paiements réglés par l'assureur à la fin de l'année 2014 après application de la réassurance est de 98,19M€ contre 125,62M€ sans la réassurance. Le réassureur a réglé 27,43M€ à l'assureur pour couvrir les sinistres.

Le tableau suivant présente le triangle des charges pour les sinistres atypiques, restant dans le passif de l'assureur.

k€	Année de développement														
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2000	3 310	4 724	4 673	5 345	5 421	5 506	5 501	5 439	5 307	5 317	5 313	6 899	6 814	6 953	7 101
2001	5 876	5 051	4 522	4 149	4 133	4 268	4 137	4 126	4 144	4 268	4 138	4 246	4 275	4 275	
2002	5 719	6 478	4 624	4 824	5 168	4 916	5 015	4 991	5 101	5 103	4 997	4 586	4 264		
2003	4 440	4 004	3 763	3 873	4 853	4 859	4 484	4 430	5 186	5 698	5 682	5 632			
2004	6 792	5 585	5 479	5 665	5 328	5 307	5 562	6 039	5 966	5 672	5 648				
2005	7 126	11 285	11 879	11 430	11 421	10 351	10 204	9 801	9 815	9 292					
2006	13 736	14 604	14 803	13 233	13 251	13 251	12 479	12 306	11 981						
2007	9 546	11 138	10 433	11 694	10 486	10 890	10 889	10 373							
2008	11 288	10 748	11 416	10 006	9 866	8 868	7 732								
2009	16 240	20 654	19 299	19 723	21 596	21 625									
2010	8 796	14 205	13 737	14 832	15 080										
2011	14 554	17 794	17 822	18 169											
2012	11 960	14 383	16 415												
2013	14 354	18 147													
2014	14 722														

TABLE 3.12 – Triangle de charges après application d'un traité de réassurance

Afin de calculer les provisions à la charge de l'assureur, ce triangle est projeté avec la méthode de Chain-Ladder pour en déduire les charges ultimes. Le tableau suivant présente les facteurs de développement permettant de développer le triangle.

Année de développement	Facteur de développement	Cadences cumulées	Cadences décumulées
0	1,187	84%	84%
1	0,987	100%	16%
2	1,004	98%	-1%
3	1,017	99%	0%
4	0,982	100%	2%
5	0,968	99%	-2%
6	0,987	95%	-3%
7	1,008	94%	-1%
8	0,995	95%	1%
9	0,989	95%	0%
10	1,061	96%	1%
11	0,976	96%	0%
12	1,013	97%	0%
13	1,021	99%	0%
14	1,000	100%	0%

TABLE 3.13 – Facteurs de développement après application du traité de réassurance

Le tableau suivant présente le triangle des sinistres atypiques projetés à la charge de l'assureur.

K€	Année de développement														Paiement	Réserve	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13			14
2000	3 310	4 724	4 673	5 345	5 421	5 506	5 501	5 439	5 307	5 317	5 313	6 899	6 814	6 953	7 101	7 088	14
2001	5 876	5 051	4 522	4 149	4 133	4 268	4 137	4 126	4 144	4 268	4 138	4 246	4 275	4 275	4 366	4 184	183
2002	5 719	6 478	4 624	4 824	5 168	4 916	5 015	4 991	5 101	5 103	4 997	4 586	4 264	4 318	4 410	4 179	230
2003	4 440	4 004	3 763	3 873	4 853	4 859	4 484	4 430	5 186	5 698	5 682	5 632	5 497	5 566	5 684	4 057	1 627
2004	6 792	5 585	5 479	5 665	5 328	5 307	5 562	6 039	5 966	5 672	5 648	5 994	5 850	5 924	6 050	5 222	828
2005	7 126	11 285	11 879	11 430	11 421	10 351	10 204	9 801	9 815	9 292	9 192	9 755	9 521	9 641	9 846	8 523	1 323
2006	13 736	14 604	14 803	13 233	13 251	13 251	12 479	12 306	11 981	11 924	11 796	12 519	12 217	12 372	12 635	9 881	2 754
2007	9 546	11 138	10 433	11 694	10 486	10 890	10 889	10 373	10 454	10 404	10 292	10 922	10 660	10 794	11 024	8 130	2 894
2008	11 288	10 748	11 416	10 006	9 866	8 868	7 732	7 631	7 690	7 653	7 571	8 035	7 842	7 941	8 110	6 990	1 120
2009	16 240	20 654	19 299	19 723	21 596	21 625	20 923	20 648	20 810	20 710	20 487	21 742	21 219	21 487	21 944	13 607	8 337
2010	8 796	14 205	13 737	14 832	15 080	14 803	14 322	14 134	14 245	14 176	14 024	14 883	14 525	14 709	15 022	7 722	7 300
2011	14 554	17 794	17 822	18 169	18 486	18 146	17 557	17 327	17 462	17 378	17 192	18 245	17 806	18 031	18 415	8 590	9 824
2012	11 960	14 383	16 415	16 481	16 768	16 461	15 926	15 717	15 840	15 764	15 594	16 550	16 152	16 356	16 704	5 371	11 333
2013	14 354	18 147	17 916	17 988	18 302	17 966	17 383	17 154	17 289	17 206	17 021	18 064	17 629	17 852	18 231	3 564	14 667
2014	14 722	17 481	17 259	17 328	17 630	17 307	16 745	16 525	16 654	16 574	16 396	17 401	16 982	17 196	17 562	1 084	16 478
Total																98 192	78 912

TABLE 3.14 – Triangle de charges après application d'un traité de réassurance

Le traité de réassurance est également appliqué aux paiements réalisés. Cela permet de calculer le nouveau montant de paiement que l'assureur a versé sur la période pour les sinistres atypiques.

La charge ultime projetée avec la méthode Chain-Ladder est de 177,1M€ (contre 226,76M€ sans prise en compte de la réassurance). La charge cédée au réassureur s'élève donc à 49,66M€. La nouvelle provision de l'assureur pour les sinistres atypiques s'élève ainsi à 78,91M€. Le traité de réassurance entraîne une diminution de 22,23M€ (-22%) de la charge globale des sinistres atypiques.

3.7.2 Evolution de l'estimations des provisions

Comme précédemment, une étude graphique de l'évolution des provisions par exercice est menée. Le graphique suivant présente les montants des provisions estimées avec la méthode de Chain-Ladder au cours des exercices 2007 à 2014 après application du traité de réassurance.

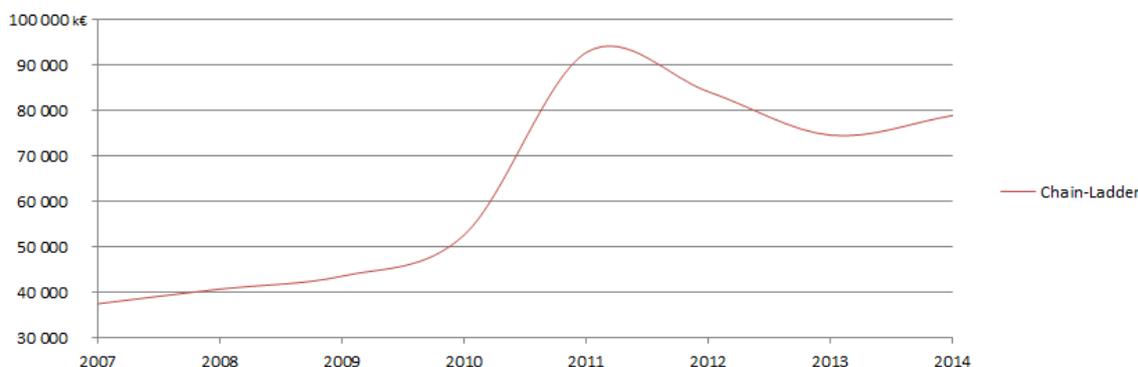


FIGURE 3.15 – Evolution de l'estimation des provisions pour les exercices 2007 à 2014

La méthode de Chain-Ladder ne permet pas d'assurer la stabilité d'estimation des projections. Entre 2009 et 2011, l'estimation des provisions augmente de 114%. La prise en compte de la réassurance amplifie les variations de provisions déjà observées sans réassurance.

3.7.3 Ecart d'estimation des provisions pour les exercices 2007 à 2011

Afin de visualiser l'erreur de projection de la méthode de Chain-Ladder après application du traité de réassurance, le graphique suivant reprend l'estimation des provisions sur les exercices 2007 à 2011 et les compare à la provision réellement constatée en 2015 pour ces exercices.

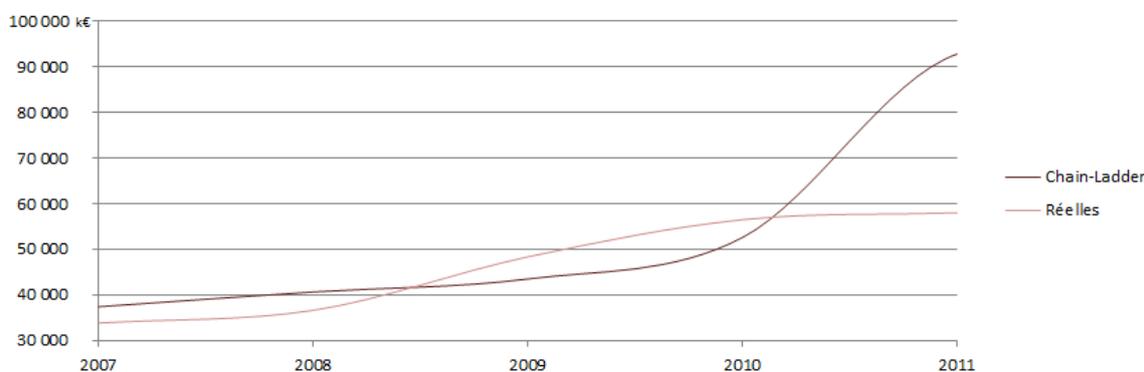


FIGURE 3.16 – Comparaison de l'estimation des provisions avec la provision constatée en 2015 pour ces exercices

La méthode de Chain-Ladder ne permet pas d'estimer les provisions au plus juste, cela entraîne des variations de provisions au cours de chaque exercice. L'année 2011 montre même que le montant de provisions estimées avec la méthode de Chain-Ladder est nettement supérieur à celui réellement observé en 2015. Le tableau suivant présente les écarts d'estimation :

	2007	2008	2009	2010	2011
Réserves réelles	33 792	36 596	48 331	56 476	57 987
Chain-Ladder	37 379	40 623	43 474	52 604	92 841
	10,6%	11,0%	-10,0%	-6,9%	60,1%

TABLE 3.15 – Ecart d'estimation des provisions par exercice

Le tableau précise les écarts visibles sur le graphique précédent. L'application du traité de réassurance entraîne une hausse des écarts d'estimation sur les provisions avec la méthode de Chain-Ladder. En moyenne l'erreur d'estimation sur ces exercices est de 19,73%, elle culmine à 60,1% pour l'exercice 2011.

3.8 Bilan de la méthode

La méthode de Chain-Ladder est simple d'application et permet d'obtenir un montant de provisions proche du montant réel pour les sinistres de coûts faibles. Le triangle des sinistres est facilement développé et les cadences des écoulements sont calculés rapidement.

Cependant pour les sinistres atypiques, comme vu précédemment, l'estimation peut-être améliorée. De plus, pour chaque exercice les provisions fluctuent ce qui entraînent l'assureur à enregistrer des variations de provisions impactant son résultat annuel. La méthode de Chain-Ladder servira de base de référence dans la suite de l'étude pour être comparée avec les nouvelles approches.

L'objectif des deux autres méthodes développées dans la suite est, d'une part, d'améliorer les estimations des provisions par rapport à la méthode de Chain-Ladder, et d'autre part, de réduire l'évolution des provisions d'un exercice à l'autre. Les deux modèles construits sont appelés modèles hybrides car l'estimation des provisions des sinistres attritionnels est obtenue à l'aide de la méthode de Chain-Ladder. Ces deux méthodes simplifieront également l'application d'un traité de réassurance par excédents de sinistres. Enfin un calcul de capital réglementaire théorique sera réalisé pour évaluer l'impact de ces méthodes dans un cadre opérationnel.

Deuxième partie

Provisionnement par une méthode ligne à ligne déterministe

Chapitre 4

Présentation de la sinistralité

Alors que les modèles de place présentés précédemment s'attachent à agréger les sinistres, d'autres méthodes exploitent les caractéristiques de la vie d'un sinistre pour estimer son coût ultime. La méthode présentée par la suite s'attache quant à elle à simuler une partie du développement du sinistre, afin d'améliorer la prédiction de son coût.

4.1 Vie d'un sinistre

Le risque peut être défini comme la probabilité qu'un fait ou qu'un événement considéré comme un mal ou un dommage survienne. Afin de se prémunir contre un risque, un assuré au travers d'un contrat d'assurance le transfère à un assureur. L'assureur va porter le risque à la place de l'assuré en échange d'une prime. Lorsqu'un risque survient, on parle de sinistre. Le schéma suivant présente la vie d'un sinistre dès que le risque est transféré à un assureur :



FIGURE 4.1 – Vie d'un sinistre

Afin de faire face à un sinistre, l'assureur va constituer des provisions. La partie précédente a présenté la méthode de place la plus utilisée sur le marché, basée sur la méthode de Chain-Ladder. Comme détaillée dans la partie précédente, cette méthode agrégée n'exploite qu'une très faible partie de l'information disponible.

L'objectif de cette partie est de présenter les différents éléments de la vie d'un sinistre susceptibles d'influer sur la nature du risque et donc sur sa provision.

4.2 Traitement d'un sinistre

La méthode de Chain-Ladder utilise un très petit nombre de variables pour projeter les sinistres et en déduire les provisions :

- La date de survenance
- La date de déclaration
- La date et le montant des paiements (pour les sinistres attritionnels)
- L'estimation des charges (pour les sinistres atypiques)

Dans la méthode développée par la suite, l'idée est de projeter la trajectoire des sinistres atypiques, c'est-à-dire simuler la partie du traitement des sinistres. Le schéma suivant présente de façon détaillée le traitement d'un sinistre.

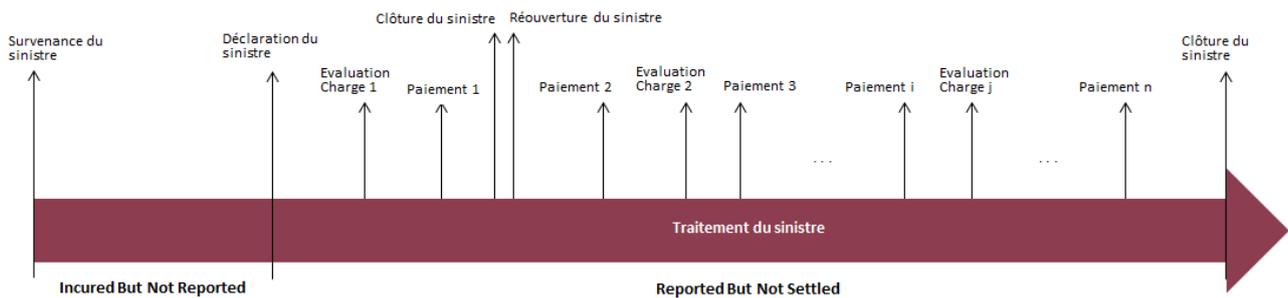


FIGURE 4.2 – Traitement d'un sinistre

Le principe d'agrégation de la méthode de Chain-Ladder ne permet de différencier les sinistres déclarés ou non, clôturés ou non, annulés ou non...

L'idée est donc d'intégrer cette dimension aux calculs des provisions.

Chapitre 5

Projection individuelle des sinistres atypiques

Afin d'améliorer la qualité d'estimation des provisions, les sinistres atypiques peuvent être projetés de manière plus détaillée. La méthode hybride développée utilise la méthode classique de Chain-Ladder pour projeter les sinistres atypiques et une méthode empirique pour projeter les sinistres atypiques. Cette dernière est inspirée par K. Murphy et A. McLennan [18] et projette l'état d'un sinistre d'une part, puis développe sa charge en fonction de son état d'autre part.

Dans la première partie un seuil de gravité a été fixé à 32 500€. Ce seuil permet d'optimiser l'utilisation de la méthode de Chain-Ladder pour les sinistres attritionnels. Il est conservé pour faciliter la comparaison entre les différentes méthodes.

5.1 Projection de la charge des sinistres atypiques

Afin d'estimer la charge ultime des sinistres graves, la projection s'opère en deux temps. Dans un premier temps, son état est simulé pour chacune des années de développement, puis une fois son état connu, son coût est projeté à l'aide de facteurs de développement.

5.1.1 Développement de la matrice des états

Dans notre méthodologie, quatre états seront considérés comme possible :

- Ouvert : le sinistre est en cours de traitement.
- Clos : le traitement du sinistre est terminé.
- Réouvert : le sinistre est retraité, après avoir été fermé.
- Annulé : le traitement du sinistre est interrompu après connaissance d'un nouvel élément.

La matrice d'états se présente comme dessous :

		Année de développement						
		0	1	2	...	j	...	n
Sinistre	1	Ouvert	Ouvert	Ouvert	...			
	2	Ouvert	Ouvert	Annulé				
	3	Ouvert	Ouvert	Clos	...	Réouvert	...	Clos
			
	i	Ouvert	Ouvert	Ouvert				
			
l	Ouvert	Clos						

TABLE 5.1 – Matrice des états

Afin de projeter la matrice d'états à l'ultime, quinze matrices de transitions sont construites pour chaque année de développement.

La construction des matrices de transition est réalisée en calculant la fréquence observée sur les changements d'état.

$$P_{Ouvert \rightarrow Ouvert} = \frac{Nb \text{ Sinistres}_{Ouvert}(t)}{Nb}$$

$$P_{Ouvert \rightarrow Clos} = \frac{Nb \text{ Sinistres}_{Clos}(t)}{Nb}$$

$$P_{Ouvert \rightarrow Réouvert} = \frac{Nb \text{ Sinistres}_{Réouvert}(t)}{Nb}$$

$$P_{Ouvert \rightarrow Annulé} = \frac{Nb \text{ Sinistres}_{Annulé}(t)}{Nb}$$

avec $Nb = Nb \text{ Sinistres}_{Ouvert}(t) + Nb \text{ Sinistres}_{Clos}(t) + Nb \text{ Sinistres}_{Réouvert}(t) + Nb \text{ Sinistres}_{Annulé}(t)$

Une matrice de transition se présente comme suit :

	Ouvert	Clos	Réouvert	Clos sans suite
Ouvert	$p_{1,1}$	$p_{1,2}$	$p_{1,3}$	$p_{1,4}$
Clos	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	$p_{2,4}$
Réouvert	$p_{3,1}$	$p_{3,2}$	$p_{3,3}$	$p_{3,4}$
Clos sans suite	$p_{4,1}$	$p_{4,2}$	$p_{4,3}$	$p_{4,4}$

TABLE 5.2 – Matrice de transition

Cette matrice doit respecter quelques propriétés :

- $\forall i, j, p_{i,j} \geq 0$
- $\forall i, \sum_{j>0} p_{i,j} = 1$

Dans la pratique, pour projeter l'état d'un sinistre i de $t - 1$ à t , on effectue un tirage à l'aide d'une loi uniforme $[0, 1]$. Notons q ce tirage. Dans la matrice de transition de l'instant $t - 1$ à t si :

- $q < p_{1,1}$ alors l'état en t pour le sinistre i sera 1
- $p_{1,1} \leq q < p_{1,2} + p_{1,1}$ alors l'état en t pour le sinistre i sera 2
- $p_{1,2} + p_{1,1} \leq q < p_{1,3} + p_{1,2} + p_{1,1}$ alors l'état en t pour le sinistre i sera 3
- $q \geq p_{1,4} + p_{1,3} + p_{1,2} + p_{1,1}$ alors l'état en t pour le sinistre i sera 4

Le schéma suivant présente les passages possibles entre les différents états.

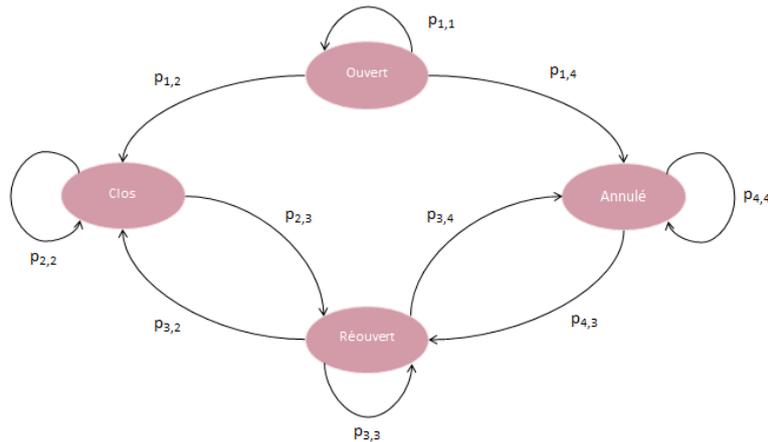


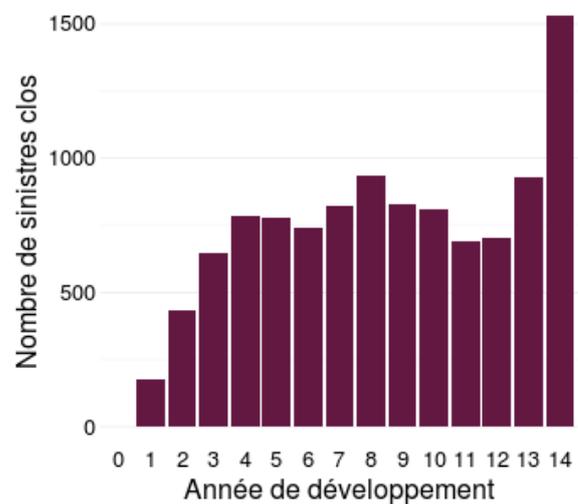
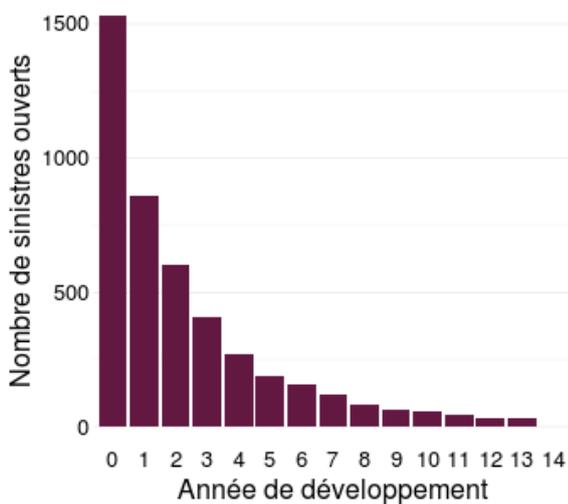
FIGURE 5.1 – Schéma de passage entre les états

Il est à noter que des changements d'états ne peuvent pas se produire, cela entraîne :

$$p_{1,3} = p_{2,1} = p_{2,4} = p_{3,1} = p_{4,1} = p_{4,2} = 0$$

Hypothèse Le dernier état de chaque sinistre sera **clos**. Cette hypothèse impose que le nombre d'années de développement de chaque sinistre soit le même.

La matrice d'état est ainsi complétée à l'aide des matrices de transition. Les graphiques suivants reprennent l'évolution de chaque état en fonction de l'année de développement.



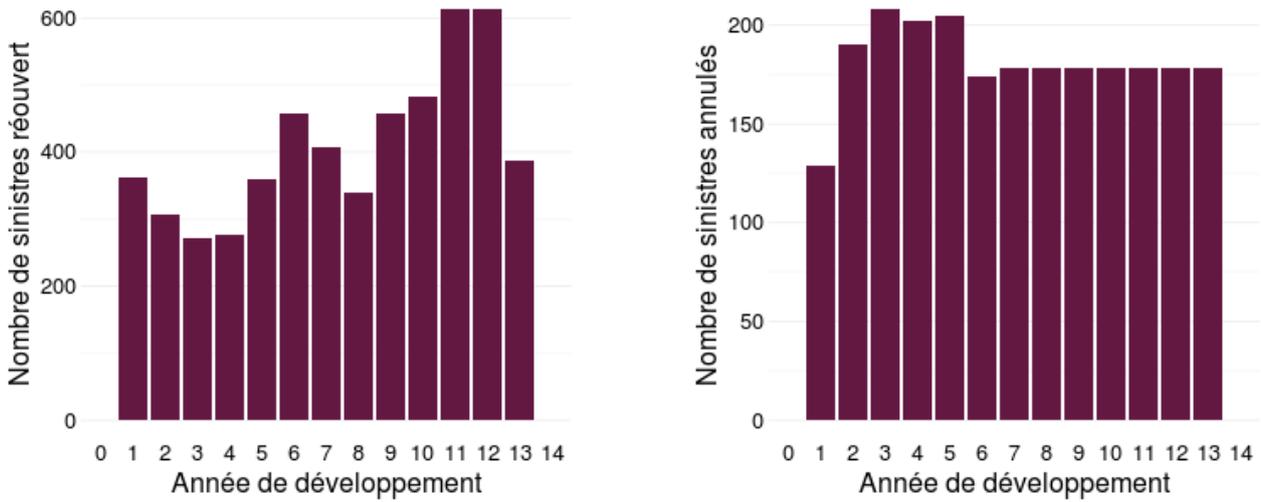


FIGURE 5.2 – Evolution des états

Le nombre de sinistres ouverts décroît rapidement lors des trois premières années puis à partir de la cinquième année, il tend lentement vers 0. L'ensemble des sinistres sont clôturés après 14 années de développement.

Le nombre de sinistres réouverts ne laisse pas apparaître de tendance particulière, même si un pic de sinistres réouverts est atteint lors des douzième et treizième années de développement. Le nombre de sinistres annulés reste stable dans le temps.

5.1.2 Calcul des facteurs de développement

Une fois les états des sinistres connus jusqu'à l'ultime, il convient de calculer les facteurs de développement pour chacun des quatre états disponibles.

La matrice de liquidation se présente comme suit :

		Année de développement						
		0	1	2	...	j	...	n
Sinistre	1	$C_{1,0}$	$C_{1,1}$	$C_{1,2}$...			
	2	$C_{2,0}$	$C_{2,1}$	$C_{2,2}$				
	3	$C_{3,0}$	$C_{3,1}$	$C_{3,2}$...	$C_{3,j}$...	$C_{3,n}$
			
	i	$C_{i,0}$	$C_{i,1}$	$C_{i,2}$				
			
	l	$C_{l,0}$	$C_{l,1}$	$C_{l,2}$...			

TABLE 5.3 – Matrice de liquidation

Cette matrice permet de calculer les facteurs de développement individuels. La formule de calcul pour les facteurs de développement individuels est la même que dans la méthode de Chain-Ladder :

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

La matrice des facteurs de développement individuels est ainsi obtenue :

		facteurs de développement						
		0 à 1	1 à 2	2 à 3	...	j-1 à j	...	n-1 à n
Sinistre	1	$f_{1,0}$	$f_{1,1}$	$f_{1,2}$...			
	2	$f_{2,0}$	$f_{2,1}$	$f_{2,2}$				
	3	$f_{3,0}$	$f_{3,1}$	$f_{3,2}$...	$f_{3,j-1}$...	$f_{3,n-1}$
			
	i	$f_{i,0}$	$f_{i,1}$	$f_{i,2}$				
			
	l	$f_{l,0}$	$f_{l,1}$	$f_{l,2}$...			

TABLE 5.4 – Matrice des facteurs de développement individuels

On notera $\forall j \in [0, n-1]$, $F_{j-1} = (f_{1,j-1}, \dots, f_{n,j-1})$ le vecteur des facteurs de développement de la $(j-1)^{\text{ème}}$ à la $j^{\text{ème}}$ année.

Le tableau suivant présente quelques statistiques sur les facteurs de développement individuels par année de développement.

	Minimum	Quantile 5%	Quantile 10%	Quantile 25%	Médiane	Moyenne	Quantile 75%	Quantile 90%	Quantile 95%	Maximum	
0 → 1	-	0,101	-	-	0,523	1,000	1,123	1,093	1,901	2,739	57,940
1 → 2	-	22,910	-	-	0,771	1,000	1,060	1,010	1,455	2,024	94,750
2 → 3	-	0,165	-	0,027	0,876	1,000	1,526	1,007	1,239	1,796	517,900
3 → 4	-	3,312	-	0,287	0,915	1,000	0,951	1,004	1,159	1,444	10,220
4 → 5	-	2,676	0,003	0,591	0,975	1,000	0,995	1,003	1,142	1,506	8,712
5 → 6	-	-	0,470	0,717	0,994	1,000	1,136	1,003	1,167	1,281	28,980
6 → 7	-	-	0,750	0,872	1,000	1,000	0,980	1,005	1,075	1,164	1,801
7 → 8	-	-	0,401	0,650	0,941	1,000	0,942	1,001	1,063	1,254	2,713
8 → 9	-	-	0,698	0,857	1,000	1,000	0,985	1,005	1,045	1,227	1,705
9 → 10	-	-	0,595	0,950	1,000	1,000	0,993	1,001	1,042	1,373	2,062
10 → 11	0,780	0,871	0,958	1,000	1,000	1,170	1,005	1,355	1,843	4,217	
11 → 12	0,800	0,909	0,968	1,000	1,000	0,989	1,001	1,026	1,031	1,041	
12 → 13	0,999	1,000	1,000	1,000	1,000	1,015	1,011	1,048	1,069	1,090	
13 → 14	1,000	1,000	1,000	1,000	1,011	1,057	1,036	1,160	1,220	1,280	

TABLE 5.5 – Statistiques descriptives des facteurs de développement individuels

5.1.3 Projection des charges à l'ultime

Une fois la matrice des états complètement développée et les facteurs de développement calculés, les charges des sinistres peuvent être projetées jusqu'à l'ultime.

Pour développer le sinistre i de l'année $j - 1$ à j , le vecteur de développement F_{j-1} est sélectionné. Nous regardons ensuite l'état du sinistre à l'année $j - 1$ et j , puis nous sélectionnons parmi les facteurs de développement du vecteur de développement F_{j-1} ceux dont les sinistres ont le même état en $j - 1$ et j .

Il reste à tirer aléatoirement un facteur de développement appartenant à ce vecteur extrait. La charge est ensuite développée en multipliant la facteur de développement avec la charge de l'année précédente.

$$C_{i,j} = f_{i,j-1} \cdot C_{i,j-1}$$

Ce processus est répété pour chaque année de développement jusqu'à la dernière année et pour chaque sinistre afin de compléter la matrice de liquidation. On additionne ensuite la dernière colonne pour connaître la charge ultime des sinistres connus.

5.1.4 Application de la méthode

Le processus global est alors simulé un grand nombre de fois, afin d'en déduire la charge ultime globale estimée :

$$ChargeUltime = \frac{1}{M} \sum_{k=1}^M ChargeUltime_k$$

Le graphique suivant présente la convergence de la charge Ultime en fonction du nombre de simulations.

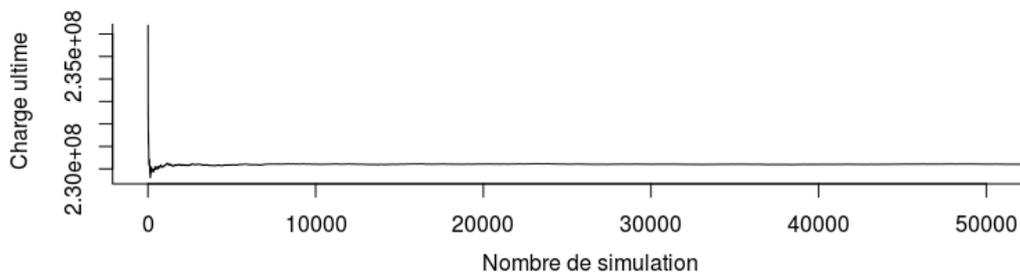


FIGURE 5.3 – Convergence de la charge ultime en fonction du nombre de simulations

À noter que la convergence vers une charge ultime moyenne se stabilise à partir de 5000 simulations. Ces simulations permettent d'en déduire la distribution de la charge ultime des sinistres déclarés

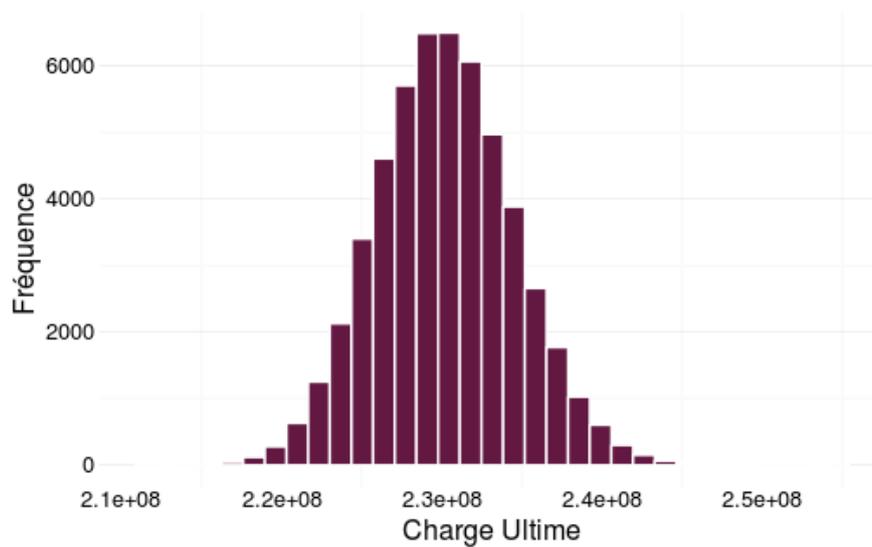


FIGURE 5.4 – Distribution des charges ultimes

Minimum	Quantile 25%	Médiane	Moyenne	Quantile 75%	Maximum
213 000 000	227 300 000	230 100 000	230 200 000	233 100 000	253 500 000

TABLE 5.6 – Statistiques descriptives des charges ultimes

$$\text{Coût Sinistres}_{\text{atypiques}} = 230,2\text{M€}$$

Le tableau suivant présente le triangle développé avec le montant de provision provisionné chaque année.

K€	Année de développement															Païement	Réserve	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14			
2000	4 255	7 153	7 029	7 545	7 649	7 738	7 750	7 626	7 267	7 309	7 468	9 019	8 947	9 116	9 288	9 347	-	59
2001	6 391	6 031	5 589	5 112	5 098	5 195	5 069	5 059	4 990	5 284	5 003	5 074	5 099	5 099	5 170	4 760	-	410
2002	6 576	7 850	6 236	6 393	6 709	6 225	6 419	6 398	6 544	6 545	6 470	6 070	5 758	6 296	6 146	5 617	-	528
2003	5 904	5 159	5 017	5 101	6 335	6 402	5 901	5 856	6 409	6 916	6 895	6 837	7 229	7 221	7 319	5 072	-	2 247
2004	8 247	7 726	7 109	7 788	7 457	7 625	7 961	8 462	8 173	7 878	7 817	7 912	7 778	7 673	7 744	7 332	-	412
2005	9 297	13 035	13 768	13 393	13 345	12 352	12 204	11 801	11 974	11 451	12 695	12 238	12 187	13 269	13 500	10 646	-	2 854
2006	17 570	18 346	18 176	16 805	16 948	16 993	16 432	16 289	15 964	16 858	16 503	16 471	16 454	16 747	17 156	13 391	-	3 766
2007	12 297	13 867	13 608	14 888	13 432	13 768	13 943	13 275	14 224	13 573	13 623	13 597	13 609	13 609	13 936	10 641	-	3 295
2008	14 511	14 201	15 669	14 005	13 707	13 229	11 793	12 907	11 844	11 895	11 522	11 340	11 315	11 601	11 768	9 941	-	1 827
2009	19 782	24 107	22 901	23 250	25 364	25 229	25 420	25 301	25 734	24 819	26 271	25 948	25 886	29 069	29 683	16 317	-	13 365
2010	10 549	17 877	17 285	18 576	19 211	19 683	21 852	20 604	20 080	20 015	20 139	19 736	19 509	19 195	19 357	10 090	-	9 267
2011	17 190	22 071	22 452	23 274	23 279	23 629	23 138	23 334	22 952	22 363	23 850	23 792	23 779	23 657	24 292	10 373	-	13 919
2012	14 595	18 089	21 205	19 075	21 592	21 429	21 044	21 074	21 652	21 646	21 912	21 583	21 542	21 440	22 079	6 913	-	15 167
2013	17 763	22 990	22 878	24 774	23 042	22 188	21 680	21 585	22 122	22 097	21 847	21 770	21 657	21 323	21 697	4 025	-	17 672
2014	18 108	21 420	21 459	25 680	23 927	22 161	21 062	21 429	21 708	21 670	21 001	20 740	20 747	20 711	21 065	1 158	-	19 908
Total																125 622	104 577	

TABLE 5.7 – Triangle développé des sinistres atypiques déclarés

Avec un coût ultime projeté de 230,2M€ et des paiements effectués à 125,6M€, le montant des provisions s'élève alors à 104,58M€ pour les sinistres atypiques.

À cela il convient de rajouter les sinistres attritionnels afin d'obtenir le montant total des provisions de l'assureur. Pour rappel, les sinistres attritionnels sont projetés avec la méthode classique de Chain-Ladder, développée dans la première partie.

Calcul des provisions

Le montant des provisions pour les :

- Sinistres attritionnels est de 93,48M€
- Sinistres atypiques est de 104,58M€

Le montant des provisions totales est de **198,06M€**

Le tableau suivant récapitule la charge ultime et les provisions obtenues avec chaque modèle de projection.

k€		Chain-Ladder	Projection individuelle
Charge Ultime	Sinistres atypiques	226 759	230 200
	Sinistres attritionnels	263 432	263 432
	Total	490 192	493 632
Paielements effectués	Sinistres atypiques	125 622	125 622
	Sinistres attritionnels	169 956	169 956
	Total	295 579	295 579
Réserves	Sinistres atypiques	101 137	104 577
	Sinistres attritionnels	93 476	93 476
	Total	194 613	198 054

TABLE 5.8 – Présentation des résultats de projection avec chaque méthode

La méthode de projection individuelle estime une augmentation de 1,77% des provisions par rapport à la méthode de Chain-Ladder.

5.2 Evolution des provisions pour les sinistres atypiques

Afin de comparer l'évolution des provisions avec la méthode de Chain-Ladder, les provisions globales de l'assureur sont recalculées pour les exercices 2007 à 2014 vue à chaque fin d'exercice. Le graphique suivant présente la comparaison entre les deux projections.

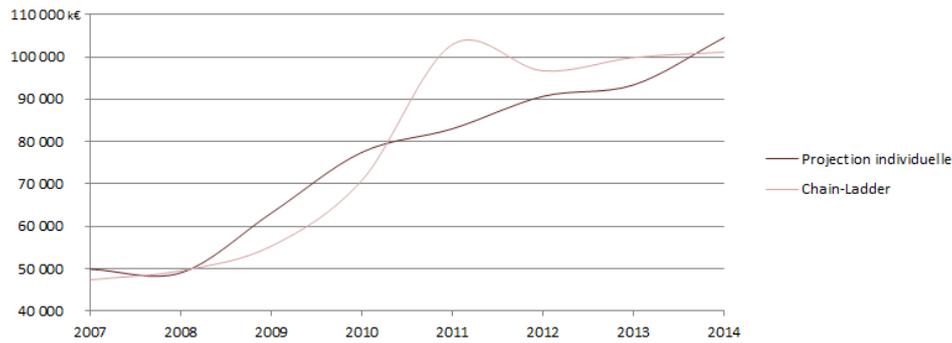


FIGURE 5.5 – Comparaison de l'évolution des provisions avec la méthode de Chain-Ladder

La méthode de projection individuelle réduit les fortes variations présente dans le modèle de Chain-Ladder et se rapproche d'avantage de la tendance observée sur la figure 3.10. La tendance est à la hausse, du fait de l'accroissement de l'activité mais la progression est moins brutale et d'avantage lissée.

Entre 2009 et 2011, l'estimation des provisions avec le deuxième modèle augmente de 31% là où le modèle de Chain-Ladder montre une augmentation de 86%. L'augmentation réelle des provisions entre 2009 et 2011 observée en 2015 donne une hausse de 23%. La méthode de projection individuelle a pour conséquence de réduire la volatilité de l'estimation des provisions d'un exercice à un autre et de mieux répliquer la tendance monotone de la sinistralité.

Pour vérifier la robustesse du modèle, les projections des provisions obtenues avec le nouveau modèle sont comparées avec les provisions réelles dans la section suivante.

5.3 Calcul de l'écart d'estimation sur le périmètre 2007-2011

Comme pour le modèle de Chain-Ladder, le graphique suivant se concentre sur les exercices 2007 à 2011, l'évolution des provisions réellement constatées en 2015 y est ajoutée afin de pouvoir comparer la projection des deux méthodes.

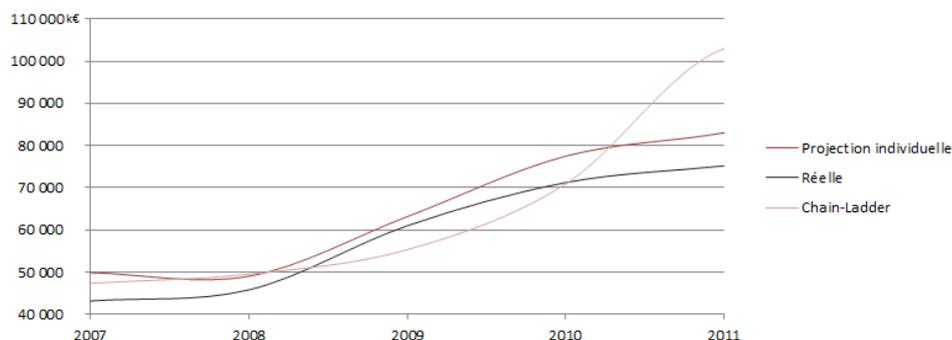


FIGURE 5.6 – Validation de la projection individuelle avec la base test

En comparaison avec la méthode de Chain-Ladder, le modèle de projection individuelle estime mieux les provisions au cours du temps. La projection individuelle des sinistres surestime le montant

de provisions à constituer lors de ces cinq exercices. Le tableau suivant permet de comparer les écarts sur les provisions pour ces cinq années d'exercices.

	2007	2008	2009	2010	2011
Réserves réelles	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%
Projection individuelle	49 841	48 990	63 127	77 479	83 030
	15,5%	7,1%	3,5%	8,8%	10,4%

TABLE 5.9 – Ecart entre la provision réelle et l'estimation faite avec les deux modèles

Le tableau appuie les observations faites sur le graphique précédent : la projection individuelle des sinistres atypiques permet de réduire l'erreur d'estimation. L'écart moyen d'estimation avec ce modèle est de 9,08% contre 12,91% avec la méthode de Chain-Ladder.

Ce nouveau modèle de projection a pour avantage de réduire la volatilité des provisions d'un exercice à l'autre et également de les estimer au plus juste. Il surperforme la méthode de Chain-Ladder sur ces points.

5.4 Calcul d'un capital réglementaire théorique

Dans le but d'avoir un comparatif opérationnel entre ce modèle et le modèle de place, un calcul de capital réglementaire théorique est réalisé avec les provisions estimées avec la méthode de projection individuelle.

La méthode développée précédemment permet d'obtenir la distribution des provisions. Afin d'obtenir le capital réglementaire avec l'estimation fournie par la formule standard, une loi lognormale est ajustée sur cette distribution. Vérifions que cette loi permet de modéliser correctement la distribution des provisions.

La densité des provisions ainsi que la loi lognormale ajustée sont représentées dans la première figure. Le deuxième graphique présente les diagramme QQ-plot.

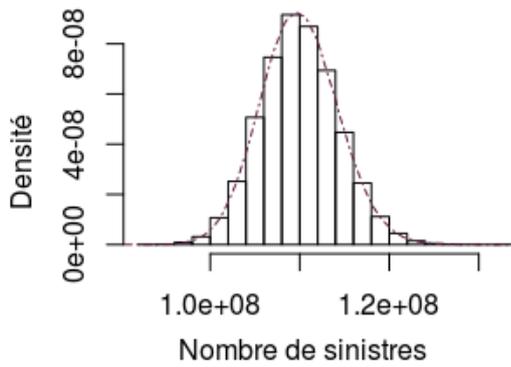


FIGURE 5.7 – Histogramme de la distribution des provisions

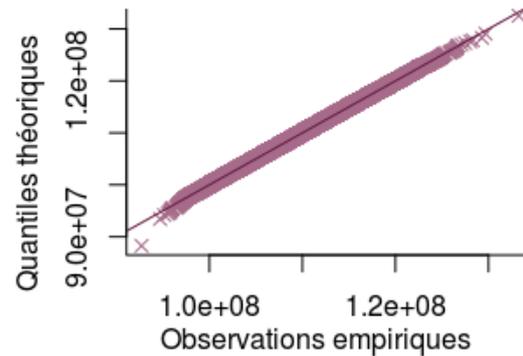


FIGURE 5.8 – Diagramme Quantiles-Quantiles

Les deux graphiques précédents permettent de valider visuellement que la loi lognormale ajustée modélise correctement la distribution des provisions.

Pour confirmer le bon ajustement, un test de Kolmogorov-Smirnov est réalisé. Ce test nous donne une $p\text{-value} = 0,9014$. La modélisation de la distribution des provisions par une loi lognormale est retenue. La volatilité de la loi ajustée permet de calculer le capital réglementaire en utilisant l'approximation de la formule standard : $3 * \sigma * RES$.

Le tableau suivant présente les paramètres de volatilité est le montant des provisions à constituer avec cette méthode.

	Sinistres attritionnels	Sinistres atypiques
σ	0,03716	0,03933
Réserve	93 476	104 586
Capital Réglementaire	10 420	12 341

TABLE 5.10 – Charge ultime des sinistres atypiques par année après application du traité de réassurance

Le montant total du capital réglementaire à constituer avec cette méthode est de 22,76M€. Cela représente une augmentation de 16% du capital réglementaire par rapport à l'estimation avec la méthode précédente.

Une partie de cette augmentation s'explique par le fait que les provisions à constituer ont augmenté de 1,06% entre les deux méthodes. D'autre part, la volatilité des provisions avec la méthode de projection individuelle augmente de 0,9% pour les sinistres atypiques. Elle reste identique pour les sinistres attritionnels. Ces deux facteurs expliquent l'augmentation de capital réglementaire à constituer.

5.5 Application du traité de réassurance

5.5.1 Calcul des provisions à constituer

Afin de comparer les deux méthodes, le traité de réassurance est de nouveau appliqué avec la méthode de projection individuelle pour sinistres atypiques.

La priorité est fixée à 130 000€ et la portée à 230 000€. Le montant de sinistres compris entre 130 000€ et 360 000€ est cédé à un réassureur.

Pour rappel le tableau suivant présente les paiements que l'assureur a réglé, ces paiements ont été calculés dans la partie précédente.

Paiements	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
	7 088	4 184	4 179	4 057	5 222	8 523	9 881	8 130	6 990	13 607	7 722	8 590	5 371	3 564	1 084

TABLE 5.11 – Paiements réalisés par l'assureur après mise en place du traité de réassurance non proportionnel

L'application du traité de réassurance se fait sinistre par sinistre. La charge ultime de chaque sinistre est estimée, il suffit de verser au réassureur la partie de la charge comprise entre 130k€ et 360k€.

Le tableau suivant présente le coût ultime des sinistres restant à la charge de l'assureur pour la méthode de Chain-Ladder et la projection individuelle, ainsi que la provision à constituer avec les deux méthodes.

k€	Chain-Ladder		Projection individuelle	
	Charge Ultime	Réserves	Charge Ultime	Réserves
2000	7 101	14	7 101	14
2001	4 366	183	4 455	271
2002	4 410	230	4 232	53
2003	5 684	1 627	5 795	1 738
2004	6 050	828	5 207	- 15
2005	9 846	1 323	9 005	482
2006	12 635	2 754	14 358	4 478
2007	11 024	2 894	15 090	6 960
2008	8 110	1 120	7 560	569
2009	21 944	8 337	22 770	9 163
2010	15 022	7 300	15 806	8 084
2011	18 415	9 824	18 660	10 070
2012	16 704	11 333	17 914	12 543
2013	18 231	14 667	15 104	11 540
2014	17 562	16 478	15 858	14 774
Total	177 104	78 912	178 915	80 723

TABLE 5.12 – Charge ultime et provisions des sinistres atypiques par année après application du traité de réassurance

La somme des charges ultimes des sinistres déclarés après application de la réassurance est estimée à 178,9M€ : l'assureur a cédé pour 51,29M€ de sa charge au réassureur.

La méthode de Chain-Ladder estime la charge ultime des sinistres atypiques après application du traité de réassurance à 177,10M€. Pour l'exercice 2014, les provisions estimées sont proches, il est intéressant de visualiser leur évolution sur plusieurs exercices afin de comparer l'impact du traité de réassurance pour ces deux méthodes.

Afin de valider la robustesse de ce modèle et la comparer à la méthode de Chain-Ladder, l'évolution des provisions entre les années 2007 et 2014 est présentée dans la section suivante.

5.5.2 Evolution de l'estimation des provisions

Comme pour la section précédente, l'évolution des provisions estimées pour les exercices 2007 à 2014 après application du traité de réassurance est comparée avec l'évolution des provisions estimées avec la méthode de Chain-Ladder vue dans la partie précédente. Le graphique suivant présente cette comparaison.

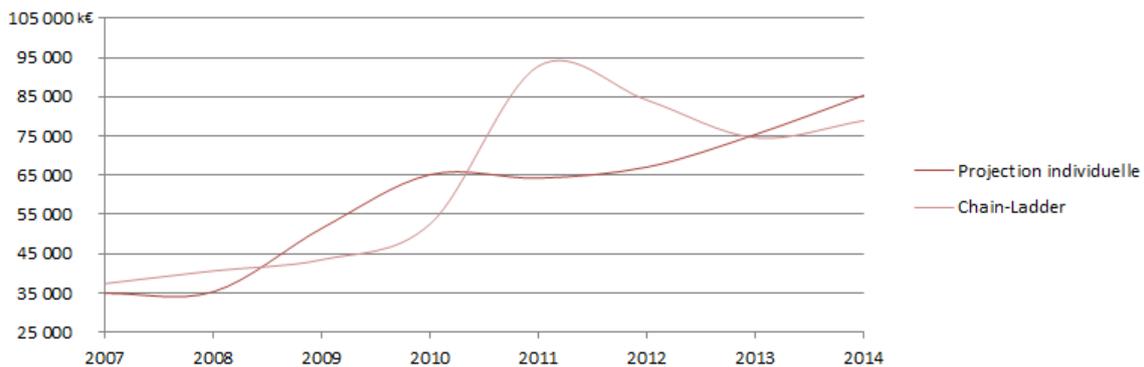


FIGURE 5.9 – Evolution des provisions estimées avec la projection individuelle après application de la réassurance

Comme dans la section précédente, le modèle de projection individuelle permet de diminuer la volatilité des provisions d'un exercice à l'autre par rapport à la méthode de Chain-Ladder. De plus, l'application du traité de réassurance semble encore amplifier la volatilité des provisions.

Afin de valider la robustesse du modèle après application du traité de réassurance, l'estimation des provisions est comparée aux provisions réellement observées en 2015.

5.5.3 Ecart d'estimation des provisions pour les exercices 2007 à 2011

Le graphique suivant permet de visualiser l'évolution des estimations des provisions avec le modèle de projection individuelle et de Chain-Ladder et de les comparer aux provisions réelles.

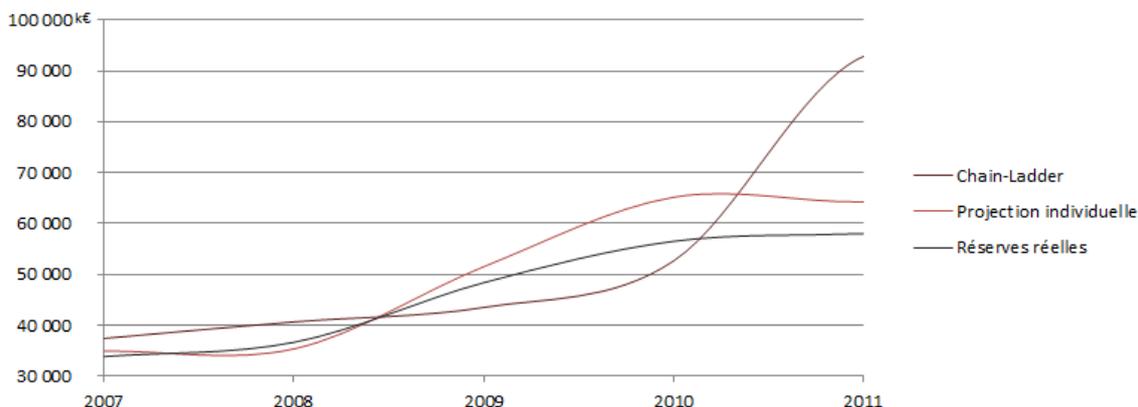


FIGURE 5.10 – Comparaison de l'évolution des estimations des provisions avec la provision réelle

En comparaison avec la méthode de Chain-Ladder, le modèle de projection individuelle estime le montant des provisions à constituer avec une meilleure précision. Le tableau suivant présente les écarts pour chaque exercice.

	2007	2008	2009	2010	2011
Réserves réelles	33 792	36 596	48 331	56 476	57 987
Chain-Ladder	37 379	40 623	43 474	52 604	92 841
	10,6%	11,0%	-10,0%	-6,9%	60,1%
Projection individuelle	34 885	35 308	51 470	65 148	64 252
	3,2%	-3,5%	6,5%	15,4%	10,8%

TABLE 5.13 – Comparaison des écarts de projection entre les deux modèles

Le tableau reprend les observations faites sur le graphique précédent. La projection individuelle des sinistres atypiques permet de réduire l'erreur d'estimation. L'écart moyen d'estimation avec ce modèle est de 7,88% contre 19,73% avec la méthode de Chain-Ladder. L'erreur d'estimation est réduite après application du traité, la méthode de projection individuelle s'adapte très bien à ce type de réassurance.

Ce nouveau modèle de projection a pour avantage de réduire la volatilité des provisions d'un exercice à l'autre par rapport à la méthode de Chain-Ladder. Il permet d'estimer les provisions avec une bonne précision.

5.6 Bilan de la méthode

Après avoir projeté les charges à l'ultime pour chaque sinistre et déduit le montant global de provisions à constituer lors des exercices 2007 à 2014, l'évolution des provisions a été représentée. La méthode de projection individuelle réduit la volatilité des provisions d'un exercice à l'autre. La provision d'un exercice à un autre est strictement croissante là où la méthode de Chain-Ladder surestime

le montant à constituer certaines années et le sous-évalue d'autres années. Comme pour la projection avec la méthode de Chain-Ladder, la tendance globale des provisions sur la période 2007-2014 reste haussière.

Dans un deuxième temps, les estimations des provisions sur les années 2007 à 2011 ont été comparées à la base de test. Le modèle de projection individuelle surestime les provisions au cours de chaque exercice mais permet de reproduire la tendance réelle de l'évolution du montant provisions sur cette période. L'erreur d'estimation moyenne du montant de provision sur le périmètre de test est de 9,1% contre 12,9% pour la méthode de Chain-Ladder. La méthode de projection individuelle en plus de diminuer la volatilité des provisions permet de réduire l'estimation des provisions par exercice.

Dans un troisième temps, un calcul de capital réglementaire pour l'exercice 2014 a été présenté lors de l'application de cette méthode. Le capital réglementaire théorique estimé avec cette méthode de projection augmente de 16% par rapport à celui déterminé avec la méthode de Chain-Ladder et le modèle de Mack. Deux raisons expliquent en partie cette augmentation, d'une part le montant des provisions plus important pour l'exercice 2014 que pour la méthode de Chain-Ladder, et d'autre part la hausse de l'incertitude liées à la volatilité de la distribution des provisions.

Enfin, le traité de réassurance non proportionnel de l'assureur a été appliqué. La projection individuelle permet de simplifier son utilisation en l'appliquant à la charge ultime projetée sinistre par sinistre. Après application du traité de réassurance, l'évolution de l'estimation des provisions avec le modèle de projection individuelle est amplifié, cependant, l'erreur d'estimation moyenne sur la période est réduite. Cette méthode de provisionnement s'adapte bien à la mise en place d'un traité de réassurance par tête.

La méthode de projection individuelle permet d'obtenir la distribution de charge ultime pour chaque sinistre et également d'intégrer les états de ces sinistres. Cela a pour conséquence de réduire l'erreur d'estimation des provisions et la volatilité d'un exercice à l'autre. Elle permet également d'obtenir un triangle des charges complet.

Cependant ce modèle est très calculatoire et repose sur un grand nombre de simulations pour obtenir des résultats qui convergent, ce qui implique un temps de calcul assez long en comparaison de la méthode de Chain-Ladder qui permet d'obtenir un montant de provisions presque instantanément.

Les modèles précédents sont des modèles déterministes pour Chain-Ladder et empiriques pour la projection individuelle. Dans la partie suivante, l'utilisation d'un algorithme d'apprentissage statistique va permettre de poursuivre l'estimation de la charge ultime par une projection individualisée de chaque sinistre en incluant des variables qualitatives dans le modèle et ainsi utiliser toute l'information que l'assureur a en sa possession.

Troisième partie

Estimation des provisions avec un modèle d'apprentissage automatique

Chapitre 6

Les forêts aléatoires

L'algorithme d'apprentissage automatique des forêts aléatoires est une méthode ensembliste agrégeant plusieurs arbres de décisions décorrélés désignant une famille d'algorithmes de *machine-learning*. Chaque arbre découpe l'espace des variables explicatives en groupes homogènes. Cet algorithme a l'avantage de pouvoir construire un modèle performant en utilisant un grand nombre de variables et un petit nombre d'observations comme c'est le cas pour la sinistralité atypique. Dans le cadre du provisionnement non-vie, les forêts aléatoires permettent de traiter de façon individuelle chaque sinistre, à travers une modélisation ligne à ligne.

6.1 Présentation générale

Les forêts aléatoires introduites par L. Breiman[6] en 2001 font partie de la famille des algorithmes supervisés construites à l'aide d'une base d'apprentissage :

$$D_n = \{(X_1, Y_1); \dots; (X_n, Y_n)\}$$

L'estimation par apprentissage automatique consiste à obtenir une prédiction Y à partir d'une entrée X avec :

- $X = (X^{(1)}, \dots, X^{(p)})$ le vecteur des variables explicatives, à valeurs dans \mathbf{R}^p .
- $Y = (Y_1, \dots, Y_p)$ la variable à expliquer.

L'objectif de la modélisation est d'estimer une fonction de régression :

$$m(x) = E[Y|X = x]$$

6.2 Les arbres de décisions

6.2.1 Construction d'un arbre

Un arbre aléatoire segmente l'espace des variables en un ensemble de sous-ensembles, pour appliquer un modèle trivial sur chacun. Les observations de chaque variable qualitative sont elles-mêmes transformées en variables quantitatives pouvant prendre comme valeur 0 ou 1, ce procédé porte le nom de "*Encodage one hot*". La méthode d'arbre la plus classique, appelée CART, se concentre sur des partitions binaires récursives. L'espace est d'abord séparé en deux régions, en choisissant une variable d'entrée et en définissant un seuil sur cette variable. Ces deux régions sont appelés noeuds. Les régions

sont ensuite définies sont séparées en deux nouvelles sous-régions, et ainsi de suite jusqu'à ce qu'un critère d'arrêt mette fin au développement d'une branche de l'arbre.

Pour décider de la variable de séparation et du seuil à appliquer, la solution privilégiée est de procéder de manière exhaustive. En testant pour toutes les variables et en parcourant les observations dans l'ordre de cette variable, la séparation optimale est celle qui minimise une erreur définie à l'avance.

La figure 6.1 présente un exemple de construction d'arbre de décision avec deux variables explicatives $X = (X_1, X_2)$.

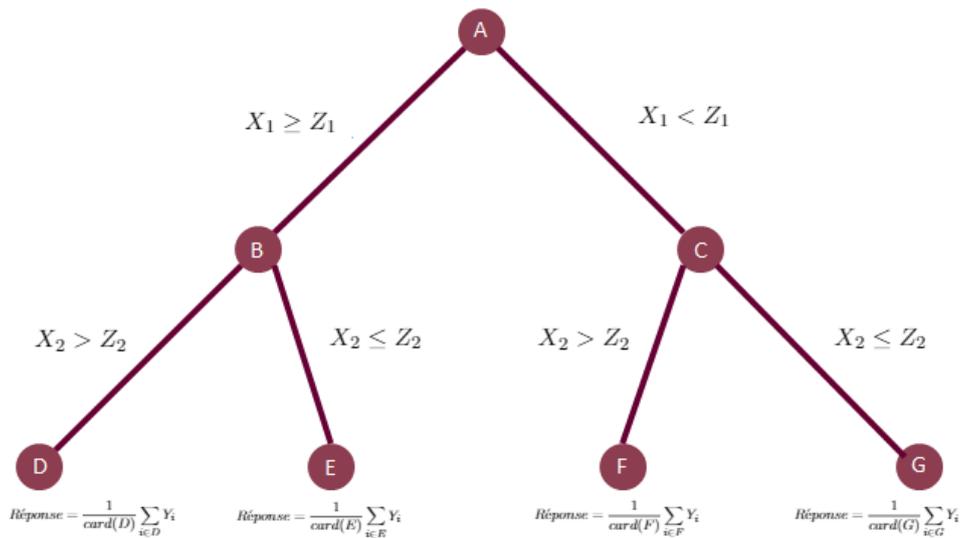


FIGURE 6.1 – Arbre de décision construit avec deux variables explicatives

Dans cet exemple, A est la racine de l'arbre. Dans un espace de dimension 2, A est représenté par le plan cartésien \mathbf{R}^2 . B et C sont des noeuds et D, E, F et G sont des feuilles de l'arbre.

Les deux figures suivantes représentent respectivement les noeuds B et C et les feuilles D, E, F et G dans le plan cartésien.

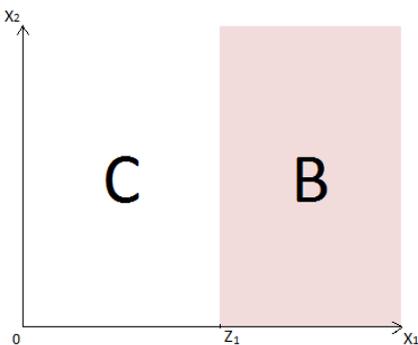


FIGURE 6.2 – Représentation des noeuds dans le plan cartésien

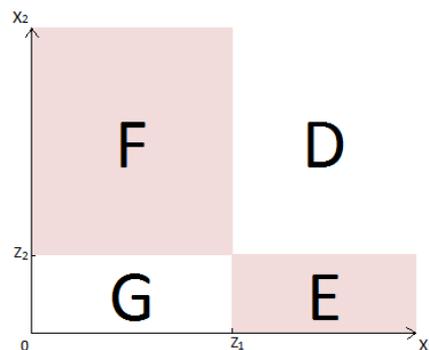


FIGURE 6.3 – Représentation des feuilles dans le plan cartésien

Chaque feuille est caractérisée par la moyenne des Y_i associées aux observations X_i contenues dans cette même feuille.

6.2.2 Le critère de coupure : CART

Le critère exposé dans cette section intervient dans la construction de l'arbre.

Soit A un noeud, on note N_A le nombre d'observations appartenant au noeud A .

Une **coupure** dans A est un couple (j, z) tel que :

- $j \in \{1, \dots, p\}$ désigne la variable sur laquelle se fait la segmentation
- z est la valeur du critère de segmentation sur la variable

On note C_A l'ensemble des coupures possibles, c'est-à-dire séparant A en deux noeuds non vides.

Le schéma suivant représente une coupure quelconque sur un noeud.

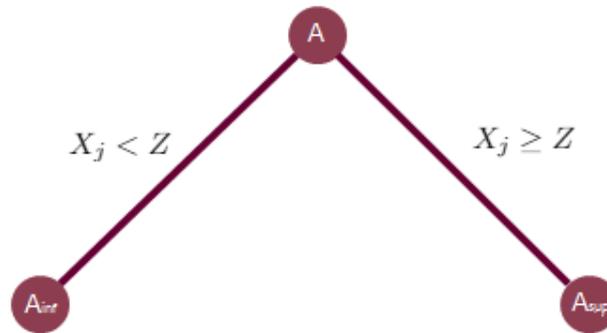


FIGURE 6.4 – Coupure d'un noeud à la $j^{\text{ième}}$ coordonnée par rapport à z

Avec :

- $A_{inf} = \{x \in A; x_j < z\}$
- $A_{sup} = \{x \in A; x_j \geq z\}$

On note :

- \bar{Y}_A : la moyenne des $Y_i \in A$
- $\bar{Y}_{A_{inf}}$: la moyenne des $Y_i \in A_{inf}$
- $\bar{Y}_{A_{sup}}$: la moyenne des $Y_i \in A_{sup}$

Le critère de coupure s'écrit comme suit :

$$L_n(j, z) = \frac{1}{N_A} \sum_{i=1}^n (Y_i - \bar{Y}_A)^2 \mathbf{1}_{X_i \in A} - \frac{1}{N_A} \sum_{i=1}^n (Y_i - \bar{Y}_{A_{inf}} \mathbf{1}_{A_{inf}} - \bar{Y}_{A_{sup}} \mathbf{1}_{A_{sup}})^2 \mathbf{1}_{X_i \in A}$$

Pour chaque noeud A , la meilleur coupure (\hat{j}_n, \hat{z}_n) est celle maximisant $L_n(j, z)$.

6.3 Les forêts d'arbres aléatoires

Les arbres de type CART ne sont pas très performants en terme de prédiction. Leur intérêt réside dans la possibilité de les agréger ensemble pour améliorer leur capacité prédictive. C'est le principe

des techniques d'agrégation d'arbres, telles que les forêts aléatoires. Une forêt aléatoire est, comme son nom l'indique, un ensemble d'arbres aléatoires. Son principe consiste à construire une large collection d'arbres non corrélés et de prendre la moyenne de leurs prédictions. L'idée est donc de construire des arbres très complexes, de biais faibles, et de les agréger afin de réduire la variance du modèle. Ces derniers sont construits indépendamment les uns des autres en utilisant à chaque fois un ré-échantillonnage de la base d'apprentissage.

6.3.1 Ré-échantillonnage aléatoire : Bagging

Lors de la construction d'une forêt aléatoire, une technique de ré-échantillonnage est utilisée avant la croissance de l'arbre. Dans cette optique, la méthode bootstrap (B. Efron, 1979)[8] permet d'ajouter de l'information supplémentaire. Les deux principales hypothèses sur lesquelles repose la méthode sont l'indépendance des observations et la distribution identique des lois.

Le rééchantillonnage

Soit un échantillon de variables aléatoires réelles (x_1, \dots, x_q) indépendantes et identiquement distribuées. On souhaite estimer le paramètre θ avec $\hat{\theta} = f(x_1, \dots, x_q)$. Pour cela un échantillon du bootstrap est construit :

- Un tirage de q éléments avec remise est réalisé. La probabilité de réalisation à chaque tirage est $\frac{1}{q}$
- Notons (x_1^*, \dots, x_q^*) ce tirage et $\hat{\theta}^* = f(x_1^*, \dots, x_q^*)$

Cette procédure est répétée pour chaque arbre (M fois). On obtient : $\hat{\theta}_k^* = f(x_1^{(k)*}, \dots, x_q^{(k)*})_{k=1, \dots, M}$

Ainsi :

$$\text{Moyenne } \hat{\theta}^* = \frac{1}{M} \sum_{k=1}^M \hat{\theta}_k^*$$

$$\text{Ecart-type } \sigma_{\hat{\theta}^*} = \sqrt{\frac{1}{M-1} \sum_{k=1}^M (\hat{\theta}_k^* - \hat{\theta}^*)^2}$$

Dans la forêt aléatoire, chaque arbre est développé avec un de ces tirages.

Lors de la construction d'un nouveau noeud, un sous-ensemble de variables explicatives est utilisé pour la construction de l'arbre. La recherche de la division binaire optimale est donc modifiée car elle ne s'applique que sur un sous ensemble aléatoire de variables. Cette sélection est censée contrer le risque de corrélation des arbres produits par un simple bagging. Ce risque est particulièrement élevé dans la situation dans laquelle certaines variables détiennent un pouvoir explicatif important et apparaissent donc dans la plupart des arbres construits, les rendant ainsi fortement corrélés entre eux.

Les arbres sont ensuite agrégés. Cette méthode est appelée *bagging*.

6.3.2 Construction de la forêt aléatoire

L'algorithme va construire M arbres aléatoires de la manière suivante :

- Pour chaque arbre, q observations sont sélectionnées avec remise parmi les n observations initiales (*Bagging*).

- À chaque noeud, seul un sous-ensemble de variables explicatives est utilisé.
- Chaque noeud est ensuite coupée de manière à maximiser le critère de coupure (*CART*).
- Le procédé est réitéré tant que le nombre d'observations dans chaque feuille reste supérieur au paramètre $nodesize \in \{1, \dots, n\}$.

Le schéma suivant présente l'agrégation des estimateurs des M arbres.

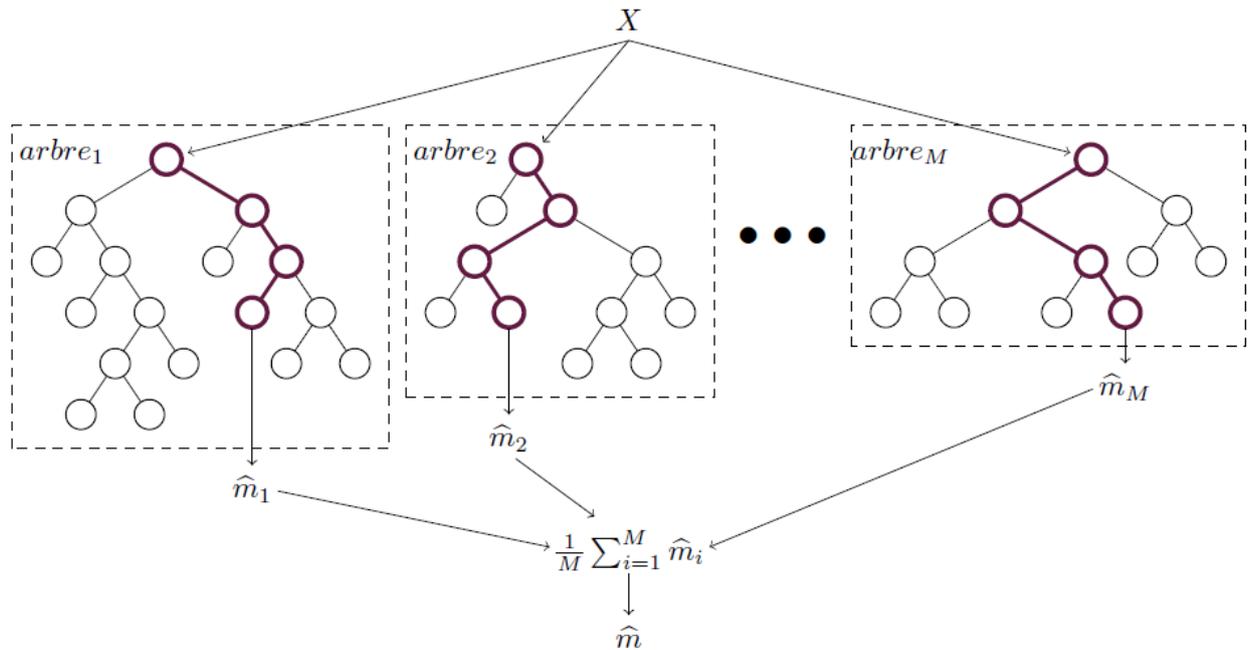


FIGURE 6.5 – Agrégation des estimateurs de M arbres aléatoires

L'estimateur final $\hat{m}(x)$ de la fonction de régression s'obtient alors en agréant les estimateurs de chaque arbre :

$$\hat{m}(x) = \frac{1}{M} \sum_{j=1}^M \hat{m}_j$$

6.4 Choix des variables

Les forêts aléatoires permettent de faire de la sélection de variables et de mesurer l'importance des variables.

6.4.1 Erreur Out-of-bag

Les arbres de la forêt aléatoire sont construits de manière indépendante. Il est donc possible de réaliser une mesure d'erreur par validation croisée tirant parti du bootstrap déjà réalisé pour éliminer des observations.

Soit une observation (x_i, y_i) issue de la base d'apprentissage. Soit F_i^* la sous forêt construite avec un échantillon ne contenant pas cette observation.

L'erreur *Out-of-bag* se définit comme l'erreur de prédiction moyenne des arbres composant la sous-forêt F_i^* sur l'observation (x_i, y_i) .

6.4.2 Importance des variables

Avec l'erreur *Out-of-bag*, il est maintenant possible de donner un ordre d'importance aux variables explicatives X dans la prédiction de la variable Y .

L'objectif est de déterminer l'importance de l'information marginale qu'apporte une variable lors de la construction d'une forêt.

La mesure d'importance d'un régresseur correspond à l'augmentation marginale de l'erreur *out-of-bag* due à la permutation des observations de cette variable. Si une variable est peu importante, le réarrangement de ses valeurs n'impacte pas la précision de la prédiction.

6.5 Avantages et inconvénients

Les forêts aléatoires sont utilisées pour résoudre des problèmes d'apprentissage statistique. Leurs avantages sont :

- Elles dépendent d'un petit nombre de paramètres
- Elles permettent de traiter avec une bonne précision des jeux de données composés de peu d'observations et d'un grand nombre de variables

Deux mécanismes rendent difficile la compréhension de l'algorithme :

- **Le bootstrap-aggregating (bagging)** : c'est un schéma d'agrégation qui permet de générer des échantillons bootstrap à partir de l'échantillon initial, puis de construire un estimateur de chacun de ces échantillons. Il reste à agréger ces estimateurs pour faire de la prédiction.
- **Le critère de coupure, Classification And Regression Tree (CART)** : permet de sélectionner la meilleure coupure lors de la construction de l'arbre.

Chapitre 7

Implémentation du modèle

L'algorithme des forêts aléatoires présenté dans le chapitre précédent, va permettre de projeter la charge ultime de chacun des sinistres atypiques présents dans la base. Dans un premier temps les variables caractérisants les sinistres sont sélectionnées puis classifiées. Puis pour chaque année de survenance, une forêt d'arbre est construite par apprentissage sur les années antérieures afin de projeter chacun des sinistres. Enfin l'évolution de l'estimation des provisions des sinistres atypiques est comparée aux méthodes précédentes.

7.1 Variables de sinistralités

Les méthodes précédentes pour la projection utilisent principalement des variables quantitatives : dates, montants des charges et des paiements. La méthode de Murphy[18] permet d'aller plus loin en tenant compte de l'état des sinistres.

Cependant, aucune des méthodes précédentes n'intègre des informations liées au sinistre. La figure suivante présente une partie des variables à la disposition de l'assureur concernant la survenance du sinistre.

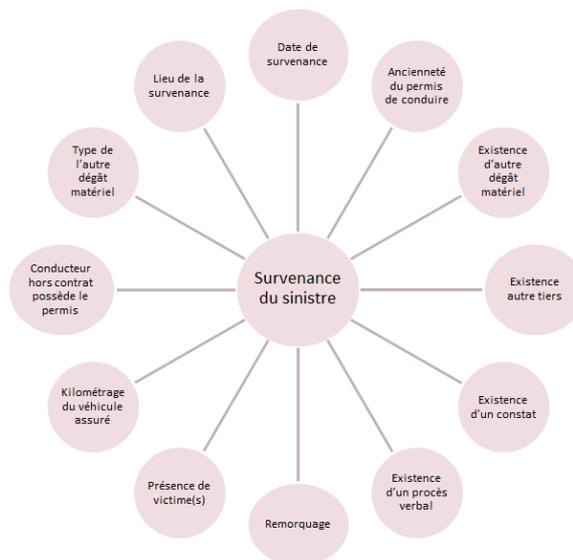


FIGURE 7.1 – Variables liées à la survenance d'un sinistre

Les variables présentées ci-dessus, ainsi que d'autres liées à la gestion du sinistre sont intégrées au modèle afin de déterminer si les informations qu'elles contiennent permettent de déterminer la charge ultime plus finement.

7.2 Sélection des variables et paramétrage du Random Forest

L'objectif du modèle est de projeter la charge estimée par le gestionnaire pour déterminer la charge ultime et en déduire les provisions de l'assureur pour cette année.

7.2.1 Périmètre de calibration

Tout d'abord pour calibrer le modèle, le périmètre doit être fixé. La charge ultime doit être connue afin de vérifier que la projection avec l'algorithme de *Random Forest* soit correcte.

Les cadences d'estimation des charges sont présentées dans le tableau suivant.

Année de développement	Facteur de développement	Cadences cumulées	Cadences décumulées
0	1,204	81%	81%
1	1,003	97%	16%
2	1,008	97%	0%
3	1,018	98%	1%
4	0,989	100%	2%
5	0,977	99%	-1%
6	0,988	97%	-2%
7	0,997	95%	-1%
8	1,001	95%	0%
9	0,992	96%	0%
10	1,045	96%	1%
11	0,982	98%	2%
12	1,012	99%	1%
13	1,019	100%	1%
14	1,000	100%	0%

TABLE 7.1 – Facteurs de développement et cadence des charges

Après cinq années de développement, la charge se stabilise à 100%. Ainsi, afin d'avoir le modèle le plus robuste, la calibration du modèle est réalisée sur les années 2000 à 2010. Les tests seront quant à eux réalisés à partir de l'année 2007. Les sinistres de ces années seront considérés comme clos.

Pour calibrer le modèle, deux bases sont nécessaires :

- La base d'apprentissage contenant les sinistres dont les années de survenances sont antérieures à l'année considérée pour le test ;
- La base de test contenant les sinistres de l'année de survenance actuelle.

Une fois le modèle calibré, les charges des années 2011 à 2014 seront projetées.

7.2.2 Choix du nombre d'arbres

Avant de commencer à sélectionner les variables qui permettront la projection des charges à l'ultime, il faut déterminer le nombre d'arbres nécessaires pour que l'algorithme des forêts aléatoires converge. Le graphique suivant représente l'erreur *out of bag* en fonction du nombre d'arbre.

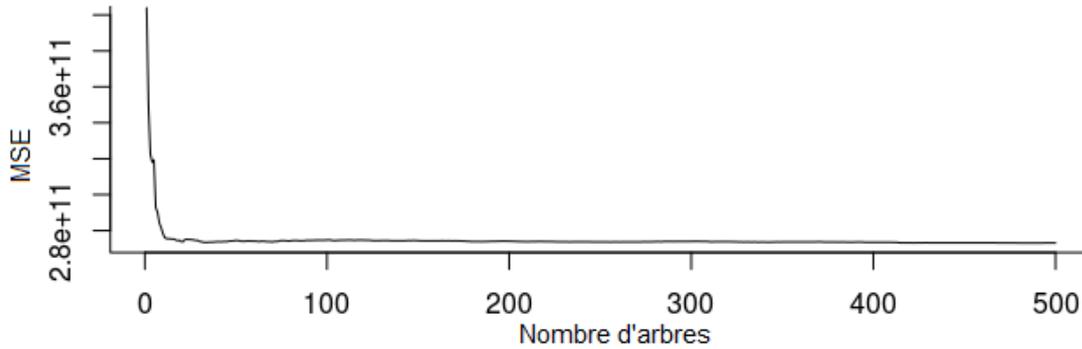


FIGURE 7.2 – Erreur du modèle en fonction du nombre d'arbre

Le graphique permet de visualiser la stabilisation de l'erreur *out of bag* après quelques dizaines d'arbres. Le choix de 500 arbres est justifié et permet de produire un modèle robuste en réduisant la volatilité de la réponse.

7.2.3 Choix des variables

Le nombre d'arbres est fixé à 500 comme présenté précédemment.

Pour la sélection de variable les paramètres (feuilles et noeuds) sont optimisés¹ et fixés à :

- $m_{try} = \sqrt{p}$ avec p le nombre de variables. Ici $m_{try} = \sqrt{24} = \sqrt{18 + 6} \approx 5$
- $nodesize = 1$

La figure suivante présente l'importance des variables en fonction de l'erreur *Out of bag*. Ce graphique permet de sélectionner les variables à retenir pour dans le modèle de projection.

1. E. Scornet, **Apprentissage et forêts aléatoires** : "La littérature est peu proluxe sur la façon dont les performances des forêts sont influencées par les paramètres de l'algorithme (M , m_{try} , an , $nodesize$). Cependant les paramètres par défaut semblent en général de bons choix" p.16

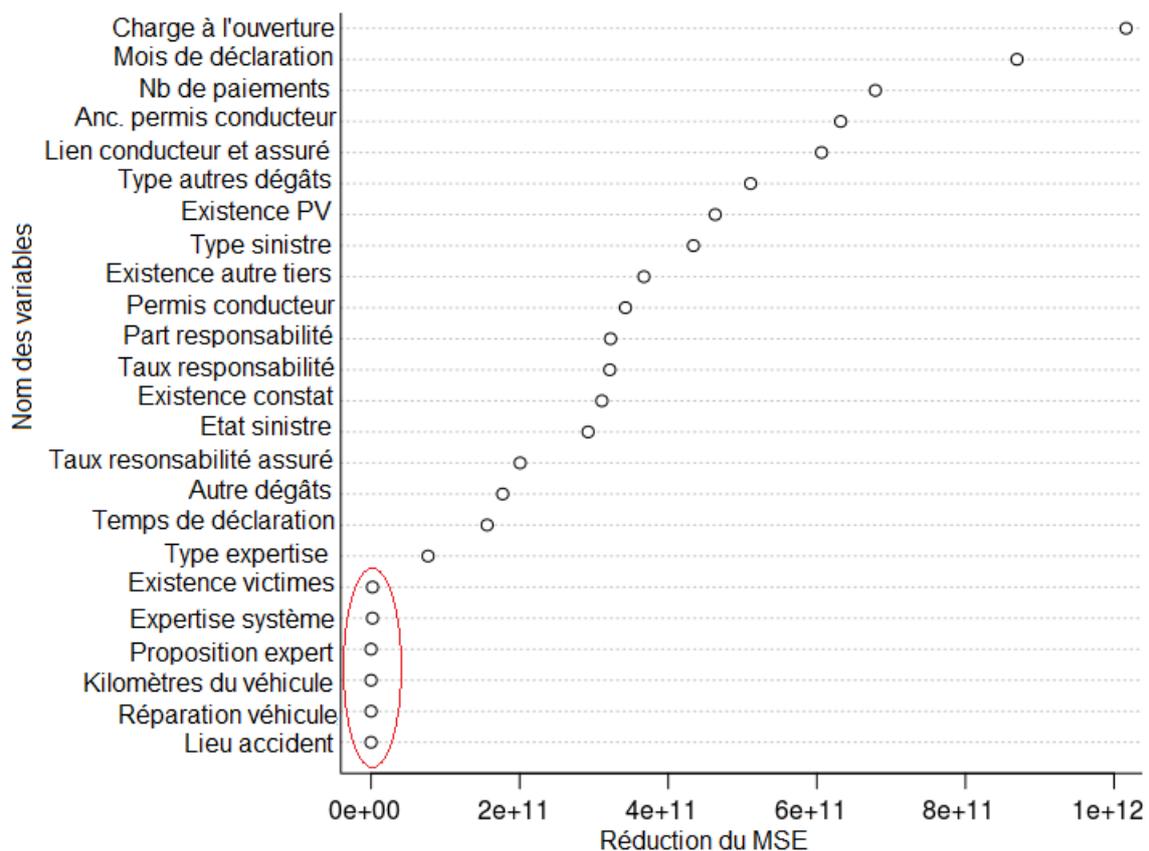


FIGURE 7.3 – Importance des variables dans le modèle des forêts aléatoires

Ce graphique montre l'importance des variables à retenir dans la suite pour réaliser le modèle de projection. Certaines variables se démarquent des autres, les informations qu'elles apportent sont pertinentes pour la projection des charges à l'ultime, comme le mois déclaration ou le nombre de paiements. D'autres variables, au contraire, n'apportent aucune information supplémentaire, comme le lieu de l'accident ou le kilométrage du véhicule. Les six variables les moins impactantes sont en outre exclues du modèle.

À noter que le nombre de victimes n'apporte aucune information pertinente. Cela s'explique par le fait que tous les sinistres contenus dans la base de graves en comportent. Afin d'améliorer d'avantage le modèle, l'assureur pourrait par exemple affiner cette variable en la sous-catégorisant.

Le tableau suivant récapitule les variables utilisées pour la projection des charges à l'ultime en fonction de la méthode utilisée.

Variables	Chain-Ladder	Projection individuelle	Random Forest
Date de déclaration	x	x	x
Date de survenance	x	x	x
Dates des paiements	x	x	x
Dates des évaluation des charges	x	x	x
Type du sinistre			x
Etat du sinistre		x	x
Charge sinistre	x	x	x
Règlements	x	x	x
Proposition de l'expert après le sinistre			
Ancienneté du permis			x
Autres dégâts matériels			x
Autres tiers			x
Existence d'un constat			x
Existence d'un PV			x
Présence de victimes			
Kilomètres du véhicule assuré			
Lien entre le conducteur et l'assuré			x
Possession du permis de conduire pour le conducteur			x
Proposition du système après le sinistre			
Type d'expertise			x
Solution retenu après expertise			
Taux d'indemnisation			x
Taux de gestion			x
Taux de responsabilité			x
Type des autres dégâts matériels			x
Lieu accident			

TABLE 7.2 – Variables utilisées dans chaque méthode

Le modèle *Random Forest* permet de sélectionner 20 variables pertinente pour réaliser la régression.

7.3 Projection des charges à l'ultime

Comme présenté dans le chapitre précédent, l'algorithme *Random Forest* ne prend que trois paramètres en entrée. Le nombre d'arbres a été fixé dans la section précédente à 500. Les autres *inputs* de l'algorithme *Random Forest* sont les suivants :

- *noeuds* = 5
- *feuille* = 1

Ces paramètres sont optimisés pour réaliser la régression.

Deux méthodes projetant la charge à l'ultime sont développées dans la suite, leurs résultats seront comparés avec la base de test pour choisir le plus performant.

Dans un premier temps les charges des sinistres sont projetées à l'ultime avec les vingt variables retenues dans la section précédente, c'est-à-dire avec seulement la charge connue à la fin de l'année de déclaration, c'est-à-dire que le modèle de projection reste le même pour chaque année de survenance, et seul la base d'apprentissage évolue.

Dans la seconde méthode, toutes les charges connues de l'assureur sont utilisées pour la projection, un modèle de projection et une base d'apprentissage sont développés par année de survenance.

Modèle de projection à une charge

La projection des charges à l'ultime avec l'algorithme des forêts aléatoires est réalisée dans un premier temps en **vision année 1**, cela signifie que seule la charge connue en fin de première année rentre en compte dans le modèle. L'apprentissage est réalisé sur les années antérieures.

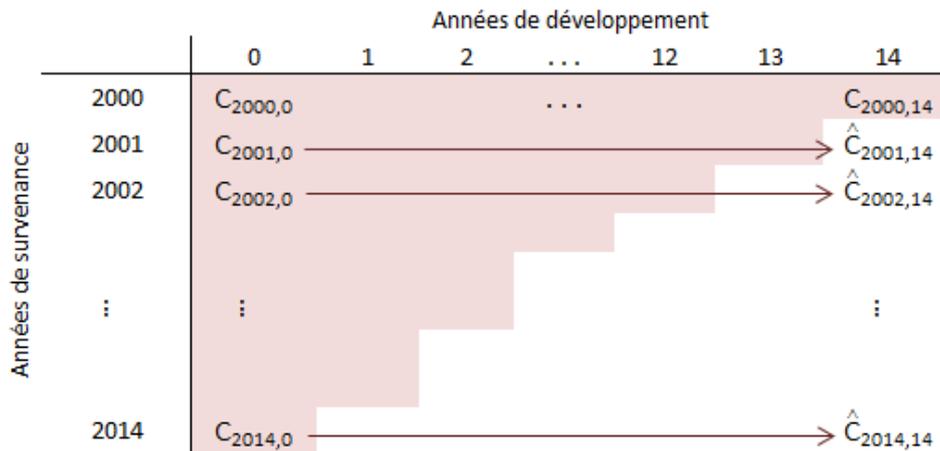


FIGURE 7.4 – Schéma présentant la projection utilisant uniquement les charges à la fin de l'année de déclaration

Pour projeter les charges à l'ultime des sinistres survenus en 2002, la base d'apprentissage du modèle est construite avec les sinistres survenus en 2000 et 2001. Les variables explicatives ont été explicitées précédemment. La variable à expliquer est la charge à l'ultime, c'est-à-dire la charge de chaque observation vue en année 2015.

La cadence cumulée atteint 100% après cinq années de développement, les sinistres postérieurs à 2011 ne sont pas suffisamment développés, ils ne peuvent pas être utilisés dans la base d'apprentissage.

Le tableau suivant présente l'estimation des provisions vues en fin d'année 2014, projetées avec les trois méthodes. Pour rappel le montant de paiements effectués à ce jour pour les sinistres atypiques est de 125,62M€.

k€	Random Forest	Chain-Ladder	Projection individuelle
2000	- 59	- 59	- 59
2001	4 026	434	410
2002	1 875	319	528
2003	835	1 852	2 247
2004	- 656	940	412
2005	- 465	1 373	2 854
2006	- 995	3 374	3 766
2007	3 145	3 262	3 295
2008	3 055	2 261	1 827
2009	2 502	9 187	13 365
2010	7 262	9 115	9 267
2011	10 463	13 314	13 919
2012	10 937	14 849	15 167
2013	14 927	19 640	17 672
2014	23 642	21 276	19 908
Total	80 495	101 137	104 577

TABLE 7.3 – Comparaison des provisions projetées avec le premier modèle *Random Forest*

Le montant des provisions obtenues avec l'algorithme d'apprentissage non supervisé est de 80,49M€, cela représente une baisse de 20% par rapport à la provision estimée avec la méthode de Chain-Ladder et une baisse de 23% avec la projection individuelle des sinistres. L'estimation de la provision est de 4,03M€ et 1,88M€ pour les années 2001 et 2002 alors que le modèle de projection individuelle en prévoit respectivement 410k€ et 528k€ et de la méthode de Chain-Ladder 434k€ et 319k€. Les provisions sont surestimées pour ces années alors que les provisions prédites pour les années 2004, 2005 et 2006 sont négatives cela à pour effet de compenser les provisions surestimées.

C'est après ces constatations qu'une méthode plus complexe utilisant un modèle de projection pour chaque année de survenance est développé.

Modèle de projection à plusieurs charges

Afin d'améliorer la projection en ajoutant de l'information dans le modèle, les charges de chaque année de développement disponible sont ajoutées.

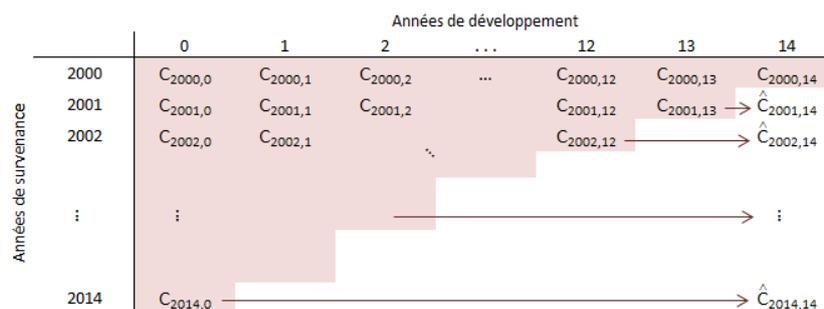


FIGURE 7.5 – Schéma présentant la projection prenant en compte toutes les charges

Pour tester la pertinence de toutes ces nouvelles variables, la classification des variables est à nouveau réalisée sur la base d'apprentissage construite. Le graphique d'importance suivant est construit avec une base contenant les sinistres survenues entre 2000 et 2007.

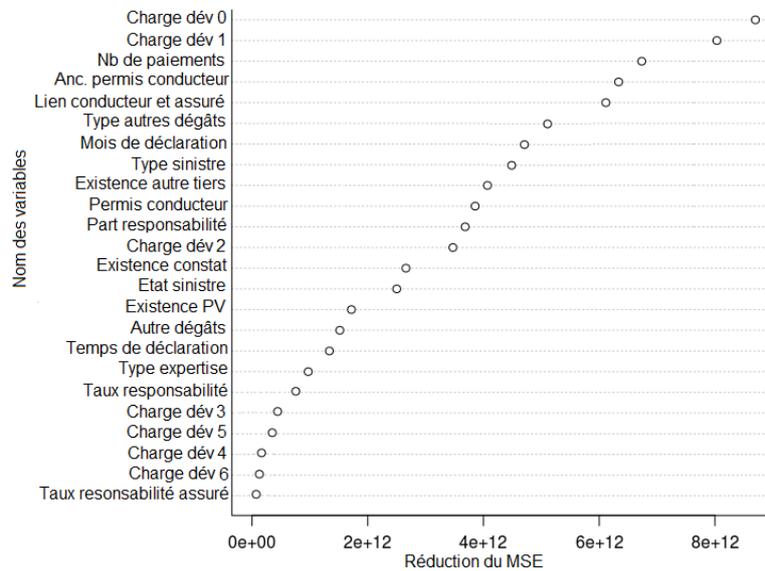


FIGURE 7.6 – Importance des variables dans le modèle des forêts aléatoires

Les charges des 2 premières années de développement sont très pertinentes pour la projection des charges à l'ultime. En comparaison avec le graphique d'importance figure 7.3, seule la variable "Mois de déclaration" perd en importance. La charge après 2 années de développement est également pertinente même si elle apporte moins d'informations. Au-delà, les charges continuent à apporter de l'information mais cela reste moindre, cependant afin de ne perdre aucune information disponible, elles seront conservées pour la projection.

L'ajout de ces nouvelles variables complexifie le modèle car la projection pour chaque année de survenance possède une base d'apprentissage avec un nombre de variables différent. Comme précédemment, les sinistres survenus après 2011 n'entrent pas en compte dans l'apprentissage. Pour l'année 2014, la charge en fin d'année est l'unique information disponible. Le modèle reste le même que précédemment.

Afin d'illustrer la méthode de projection, un exemple est présenté sous forme schématique. Dans cet exemple, l'objectif est de projeter les charges des sinistres atypiques de l'année 2007 à l'ultime. Le schéma suivant présente la projection :

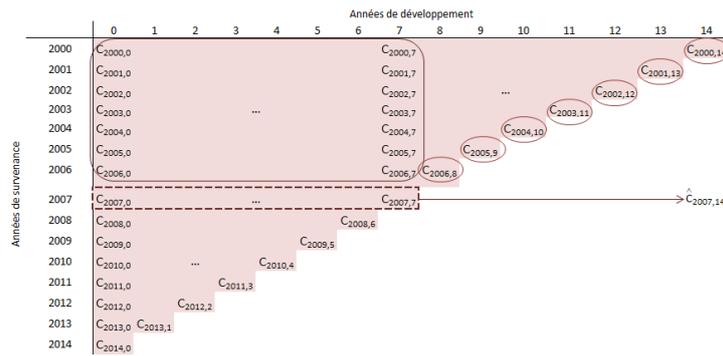


FIGURE 7.7 – Schéma de projection des charges des sinistres de l'année 2007

Comme le présente le schéma, seules les charges des sept premières années sont connues pour les sinistres survenus en 2007. La base d'apprentissage pour la projection des sinistres de 2007 est présentée dans le tableau suivant :

	Variables explicatives											Variable à expliquer
	Charge année dev. 0	Charge année dev. 1	Charge année dev. 2	Charge année dev. 3	Charge année dev. 4	Charge année dev. 5	Charge année dev. 6	Charge année dev. 7	Ancienneté permis	...	Existence PV	Charge Ultime
	Base d'apprentissage	$C_{2000,0}$	$C_{2000,1}$	$C_{2000,2}$	$C_{2000,3}$	$C_{2000,4}$	$C_{2000,5}$	$C_{2000,6}$	$C_{2000,7}$...	
	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots		...		\vdots
	$C_{2006,0}$	$C_{2006,1}$	$C_{2006,2}$	$C_{2006,3}$	$C_{2006,4}$	$C_{2006,5}$	$C_{2006,6}$	$C_{2006,7}$				$C_{2006,8}$

TABLE 7.4 – Base d'apprentissage pour la projection des charges des sinistres de 2007

Cette base d'apprentissage permet de développer la forêt aléatoire qui servira à projeter les charges des sinistres de l'année 2007 à l'ultime. Le tableau suivant présente les sinistres de 2007 à projeter.

	Variables explicatives											Variable à expliquer
	Charge année dev. 0	Charge année dev. 1	Charge année dev. 2	Charge année dev. 3	Charge année dev. 4	Charge année dev. 5	Charge année dev. 6	Charge année dev. 7	Ancienneté permis	...	Existence PV	Charge Ultime
Base de projection	$C_{2007,0}$	$C_{2007,1}$	$C_{2007,2}$	$C_{2007,3}$	$C_{2007,4}$	$C_{2007,5}$	$C_{2007,6}$	$C_{2007,7}$...		$C_{2007,14}$

TABLE 7.5 – Base des sinistres survenus en 2007

La variable à expliquer est la charge ultime des sinistres de 2007, observée en 2014, dans cet exemple. $\hat{C}_{2007,14}$ contient toutes les estimations des charges ultimes des sinistres survenues en 2007.

Le tableau suivant présente l'estimation des provisions pour chaque année de survenance, vues en fin d'année 2014 et projetées avec les trois méthodes. Pour rappel le montant de paiements effectués à ce jour pour les sinistres atypiques est de 125,62M€.

k€	Chain-Ladder	Projection individuelle	Random Forest 1	Random Forest 2
2000	-	59	-	59
2001	434	410	4 026	718
2002	319	528	1 875	199
2003	1 852	2 247	835	1 661
2004	940	412	-	656
2005	1 373	2 854	-	465
2006	3 374	3 766	-	995
2007	3 262	3 295	3 145	2 881
2008	2 261	1 827	3 055	2 106
2009	9 187	13 365	2 502	5 937
2010	9 115	9 267	7 262	8 385
2011	13 314	13 919	10 463	12 846
2012	14 849	15 167	10 937	14 098
2013	19 640	17 672	14 927	19 028
2014	21 276	19 908	23 642	23 642
Total	101 137	185 072	80 495	92 975

TABLE 7.6 – Comparaison des provisions projetées avec le deuxième modèle *Random Forest*

La provision est égale à 92,96M€, cela représente une augmentation de 15,5% entre le premier modèle d'apprentissage statistique et ce dernier prenant en compte toute la charge connue. Ce dernier modèle possède d'avantage d'informations que le premier modèle, toutes les provisions estimées sont positives contrairement au modèle précédent.

Afin de comparer les deux modèles et d'en sélectionner un pour la suite des modélisations, les provisions estimées sur les exercices 2007 à 2011 sont représentées sur le graphique suivant.

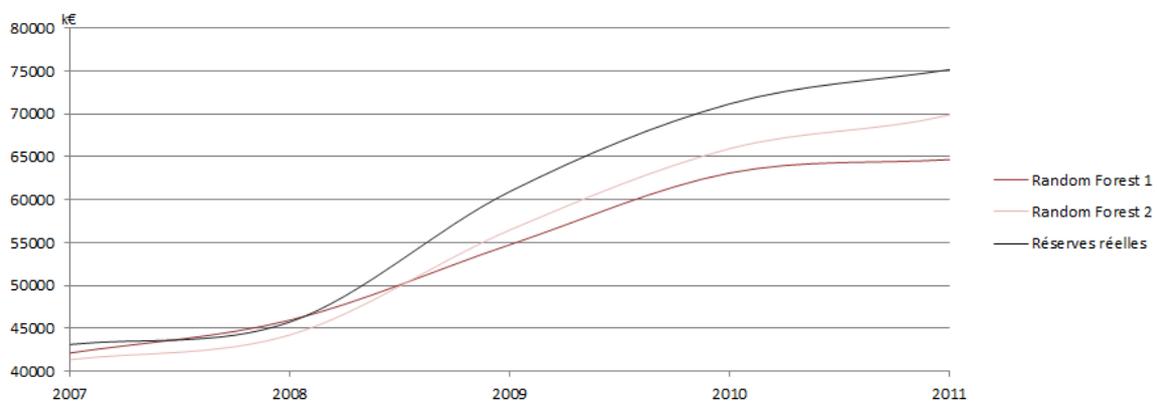


FIGURE 7.8 – Comparaison des deux modèles *Random Forest*

Graphiquement, comme attendu, le modèle de projection à une charge est moins précis que le modèle de projection à plusieurs charges. Ces deux modèles permettent d'avoir une projection sans variations

singulières d'un exercice à l'autre. L'erreur moyenne du premier modèle est de 7,7% contre 5,8% avec le second modèle.

Dans toute la suite de l'étude, la projection des provisions est réalisée avec le deuxième modèle d'apprentissage non supervisé.

Calcul des provisions

Le montant des provisions pour les :

- Sinistres attritionnels est de 93,48M€
- Sinistres atypiques est de 92,98M€

Le montant des provisions totales est de **186,45M€**

Le tableau suivant récapitule la charge ultime et les provisions obtenues avec chaque modèle de projection.

k€		Chain-Ladder	Projection individuelle	Random Forest
Charge Ultime	Sinistres atypiques	226 759	230 200	218 597
	Sinistres attritionnels	263 432	263 432	263 432
	Total	490 192	493 632	469 549
Paielements effectués	Sinistres atypiques	125 622	125 622	125 622
	Sinistres attritionnels	169 956	169 956	169 956
	Total	295 579	295 579	295 579
Réserves	Sinistres atypiques	101 137	104 577	92 975
	Sinistres attritionnels	93 476	93 476	93 476
	Total	194 613	198 054	186 451

TABLE 7.7 – Présentation des résultats de projection avec chaque méthode

Le montant de provision estimé avec la méthode d'apprentissage statistique pour l'année 2014 est réduit de 4,19% par rapport à la méthode de Chain-Ladder et de 5,86% avec la projection individuelle.

7.4 Evolution de l'estimation des provisions pour les exercices 2007 à 2014

Afin de comparer l'évolution de l'estimation des provisions des sinistres atypiques obtenues avec les trois modèles, celles-ci sont présentées dans le graphique suivant pour les exercices 2007 à 2014.

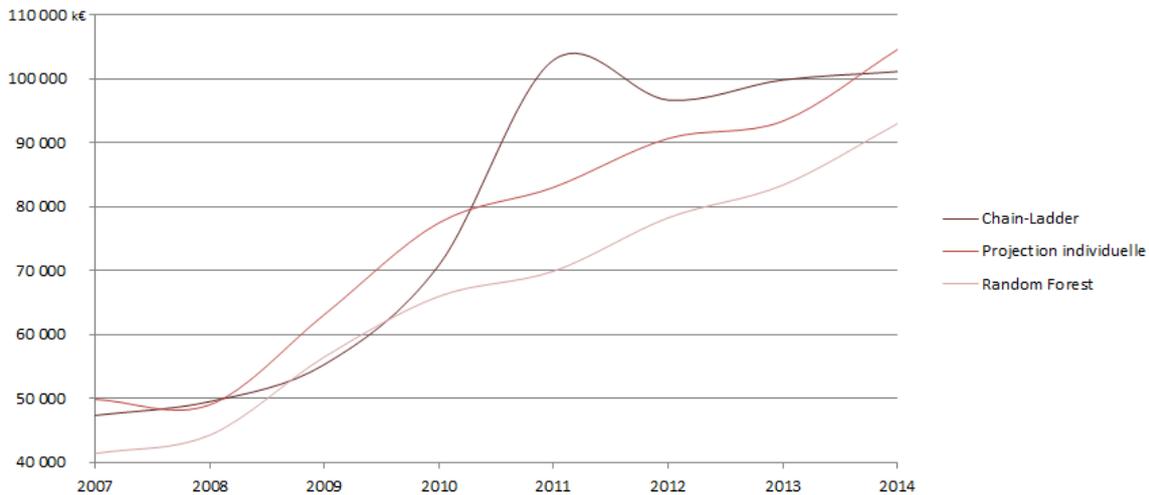


FIGURE 7.9 – Evolution des provisions estimées avec l'algorithme *Random Forest*

La projection réalisée avec le modèle d'apprentissage statistique ne laisse apparaître aucune variation brutale, notamment entre 2009 et 2011 où ce modèle prévoit 24% d'augmentation alors que la projection avec Chain-Ladder estime une augmentation de 86%, la méthode de projection individuelle de 32% des provisions et l'observation réelle de 23%. La projection avec le modèle d'apprentissage automatique donne une tendance haussière et monotone qui se rapproche de l'évolution de la provision observée figure 3.10.

Le but de la projection des sinistres ligne à ligne est de réduire d'une part la volatilité de l'estimation des provisions d'un exercice à l'autre et d'autre part de réduire l'erreur d'estimation. Comme vu dans la seconde partie, la projection individuelle permet de diminuer la volatilité des provisions entre deux exercices et d'améliorer la méthode de Chain-Ladder. Avec la projection par apprentissage automatique, la variation des provisions est encore réduite. La tendance reste à la hausse, comme les deux modèles précédents, les provisions augmentent d'un exercice à l'autre. La projection par apprentissage automatique conduit à une meilleure perennité dans l'évolution des provisions.

7.5 Calcul de l'écart d'estimation sur le périmètre 2007-2011

Afin de valider la robustesse de la projection avec le modèle d'apprentissage non supervisé, les projections des provisions réalisées avec les trois modèles sont représentées graphiquement sur les exercices 2007 à 2011 et sont comparées avec le montant de provisions dont l'assureur devait disposer pour ces exercices.

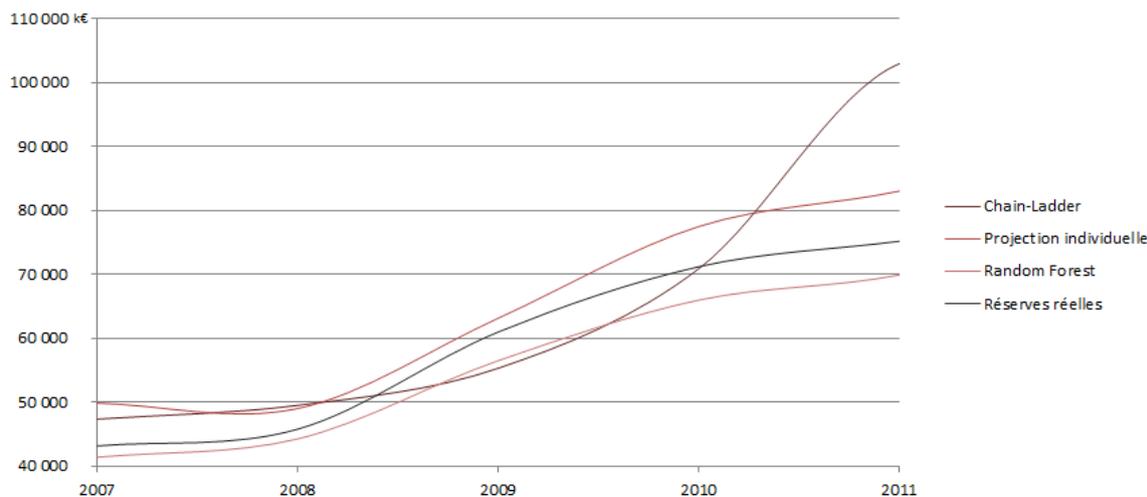


FIGURE 7.10 – Comparaison de l'estimation des provisions entre les trois méthodes

Graphiquement, la projection réalisée avec l'algorithme des forêts aléatoires sous-estime sur ces cinq exercices les provisions à constituer. Le tableau suivant permet de comparer les écarts des trois projections sur chacun des exercices.

	2007	2008	2009	2010	2011
Réserves réelles	43 136	45 762	60 970	71 193	75 184
Chain-Ladder	47 329	49 491	55 300	70 877	102 967
	9,7%	8,1%	-9,3%	-0,4%	37,0%
Projection individuelle	49 841	48 990	63 127	77 479	83 030
	15,5%	7,1%	3,5%	8,8%	10,4%
Random Forest	41 366	44 246	56 479	65 966	69 902
	-4,1%	-3,3%	-7,4%	-7,3%	-7,0%

TABLE 7.8 – Écart entre les trois projections et la provision réelle

Comme vu sur le graphique, la provision estimée avec le modèle *Random Forest* sous-estime la provision à constituer. Cependant, l'erreur moyenne d'estimation sur ces trois exercices est de 5,8% contre une erreur à 12,9% en moyenne pour le modèle de Chain-Ladder et de 9,1% pour le modèle de projection individuelle.

Le modèle d'apprentissage statistique basé sur les caractéristiques de chaque sinistre amoindrit le coût des sinistres, améliore les estimations des provisions et diminue les variations d'un exercice à l'autre par rapport aux deux autres modèles.

Afin d'avoir d'autres points de comparaison entre ces trois méthodes, le calcul du capital réglementaire théorique est réalisée avec cette nouvelle méthode de projection et le traité de réassurance non proportionnel de l'assureur est appliqué avec ce modèle.

7.6 Calcul d'un capital réglementaire théorique

Afin de déterminer le capital réglementaire théorique avec ce modèle de projection, la distribution des provisions est simulée. L'algorithme *Random Forest* construit son modèle de régression à l'aide d'une base d'apprentissage sur laquelle il effectue un bootstrap. Pour amplifier le bootstrap, 5000 simulations sont réalisées. À chaque simulation, seule 90% de la base d'apprentissage est utilisée pour calibrer le modèle.

La distribution des provisions est obtenue, comme le suggère la formule standard, celle-ci peut être simulée avec une loi lognormale bien calibrée.

La première figure présente la densité des provisions sur laquelle une loi lognormale a été calibrée. Le deuxième graphique expose le diagramme Quantiles-Quantiles.

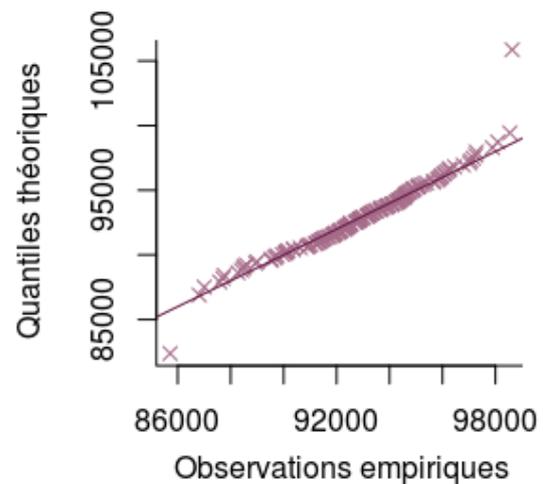
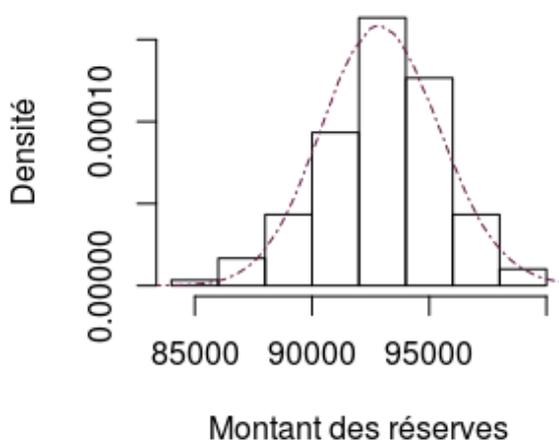


FIGURE 7.11 – Histogramme de la distribution des provisions

FIGURE 7.12 – Diagramme Quantiles-Quantiles

Visuellement, la loi lognormale semble être adaptée pour la modélisation des provisions projetées avec l'algorithme *Random Forest*. Pour le confirmer, un test de Kolmogorov-Smirnov est réalisé. La *p-value* est égale à 0,8578. La loi lognormale est retenue pour modéliser les provisions.

Comme pour les précédentes méthodes, le capital réglementaire des sinistres attritionnels est obtenu avec le modèle de Mack. Le tableau suivant présente le capital que l'assureur doit détenir pour être considéré comme solvable.

	Sinistres attritionnels	Sinistres atypiques
σ	0,0374777	0,02724
Réserve	93 476	92 975
Capital Réglementaire	10 510	7 599

TABLE 7.9 – Capital réglementaire et volatilité des provisions

La volatilité des provisions sur les sinistres atypiques diminue d'environ 0,3% par rapport au modèle de Mack et de 1,2% par rapport au deuxième modèle. L'estimation des provisions avec ce modèle est moins volatile qu'avec les deux précédents.

La conséquence de cette diminution de volatilité est que le capital réglementaire que doit détenir l'assureur s'élève à 18,11M€. Cela représente une diminution de 8% par rapport à la projection réalisée avec Chain-Ladder et 20% par rapport à la projection avec la deuxième méthode.

Le tableau suivant récapitule le montant de capital réglementaire nécessaire avec chacun des trois modèles :

€	Capital Réglementaire
Chain-Ladder	19 673 701 100,0%
Projection individuelle	22 761 148 115,7%
Random Forest	18 108 994 92,0%

TABLE 7.10 – Comparaison entre les trois modèles du capital réglementaire

Après avoir présenté une méthode permettant de calculer le capital réglementaire, le traité de réassurance non proportionnel de l'assureur est appliqué afin de comparer les résultats entre les trois méthodes.

7.7 Application du traité de réassurance 130k€ XS 230k€

Comme pour la méthode de projection individuelle, le modèle permet de projeter la charge de chaque sinistre à l'ultime, le traité de réassurance peut être appliqué sinistre par sinistre en cédant la charge de chaque sinistre comprise entre 130 000€ et 360 000€.

7.7.1 Calcul des provisions à constituer

Le tableau suivant présente les projections des charges à l'ultime par année de survenance obtenues avec les trois méthodes développées tout au long de cette étude.

k€	Random Forest	Chain-Ladder	Projection individuelle
2000	7 101	7 101	7 101
2001	4 608	4 366	4 455
2002	4 150	4 410	4 232
2003	5 761	5 684	5 795
2004	5 690	6 050	5 207
2005	9 215	9 846	9 005
2006	11 742	12 635	14 358
2007	10 243	11 024	15 090
2008	7 646	8 110	7 560
2009	21 298	21 944	22 770
2010	15 074	15 022	15 806
2011	18 357	18 415	18 660
2012	17 656	16 704	17 914
2013	15 357	18 231	15 104
2014	16 434	17 562	15 858
Total	170 331	177 104	178 915

TABLE 7.11 – Présentation des charges ultimes après application du traité de réassurance vues à fin 2014

Après application du traité de réassurance la charge ultime restante à l'assureur est de 170,33M€. Ci-dessous la répartition des charges ultimes estimées avec les trois méthodes après l'application du traité de réassurance.

k€		Sans Réassurance	Avec Réassurance	Cédées au réassureur
Charge Ultime	Chain-Ladder	226 759	177 104 78,1%	49 655 21,9%
	Projection individuelle	230 200	178 915 77,7%	51 285 22,3%
	Random Forest	218 597	170 331 77,9%	48 266 22,1%

TABLE 7.12 – Comparaison des charges ultimes après application du traité de réassurance vues à fin 2014

Le montant de charges cédées est de 48 266€, cela représente 22,1% de la charge. Les proportions des charges cédées avec les trois méthodes sont équivalentes. Les projections des sinistres avec une méthode ligne à ligne permettent d'appliquer le traité de réassurance tête par tête comme cela se fait opérationnellement, le risque est ainsi mieux géré.

Le tableau ci-dessous rappelle le montant des règlements et la répartition entre l'assureur et le réassureur.

k€	Sans Réassurance	Avec Réassurance	Réglés par le réassureur
Paiements effectués	125 622	98 192 78,2%	27 430 21,8%

TABLE 7.13 – Répartition paiement après mise en place du traité de réassurance

Le montant des paiements restants à la charge du réassureur est de 98,19M€, soit 21,8% des règlements payés par le réassureur. Le tableau suivant présente le nouveau montant de provisions de l'assureur.

k€	Random Forest	Chain-Ladder	Projection individuelle
2000	14	14	14
2001	425	183	271
2002	- 29	230	53
2003	1 704	1 627	1 738
2004	467	828	15
2005	692	1 323	482
2006	1 861	2 754	4 478
2007	2 113	2 894	6 960
2008	656	1 120	569
2009	7 691	8 337	9 163
2010	7 352	7 300	8 084
2011	9 767	9 824	10 070
2012	12 285	11 333	12 543
2013	11 793	14 667	11 540
2014	15 350	16 478	14 774
Total	72 139	78 912	80 723

TABLE 7.14 – Présentation des estimations des provisions après application du traité de réassurance

Les provisions que l'assureur doit constituer avec le modèle des forêts aléatoire s'élèvent à 72,14M€. Le tableau ci-dessous permet de comparer les écarts et les montants cédés avec chaque méthode de projection.

k€		Sans Réassurance	Avec Réassurance	Cédées au réassureur
Réserves	Chain-Ladder	101 137	78 912 78,0%	22 225 22,0%
	Projection individuelle	104 577	80 723 77,2%	23 855 22,8%
	Random Forest	92 975	72 139 77,6%	20 836 22,4%

TABLE 7.15 – Présentation des estimations des provisions après application du traité de réassurance

Avec l'estimation des provisions par la méthode d'apprentissage statistique, l'assureur cède 22,4% de sa provision au réassureur. Comme pour les charges, les proportions des charges cédées avec les trois méthodes sont équivalentes.

Afin de suivre la variation des provisions après mise en place du traité de réassurance, l'évolution des provisions pour les exercices 2007 à 2014 est présentée dans la suite.

7.7.2 Evolution de l'estimation des provisions

Comme précédemment, le graphique ci-après présente l'évolution de l'estimation des provisions pour chacune des méthodes lors des exercices 2007 à 2014.

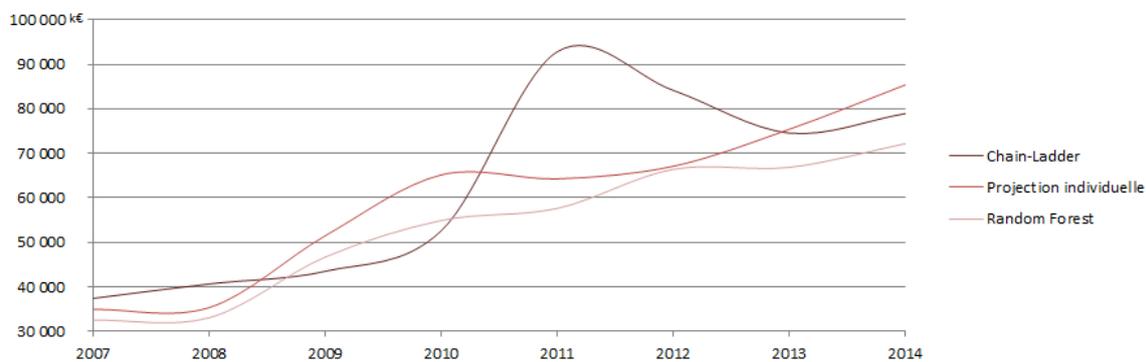


FIGURE 7.13 – Évolution des provisions estimées avec l'algorithme des forêts aléatoire après application du traité de réassurance

La projection avec l'algorithme *Random Forest* ne laisse apparaître aucune variation brutale des provisions entre chaque exercice. La tendance est à la hausse constante. En comparaison la méthode de projection individuelle ou de Chain-Ladder accentuent les variations des estimations des provisions après application du traité de réassurance.

Pour valider l'estimation des provisions avec le modèle d'apprentissage statistique, celles-ci sont comparées aux provisions connues des exercices 2007 à 2011.

7.7.3 Écart d'estimation des provisions pour les exercices 2007-2011

Le graphique suivant présente les provisions estimées et réelles pour les exercices des années 2007 à 2011 :

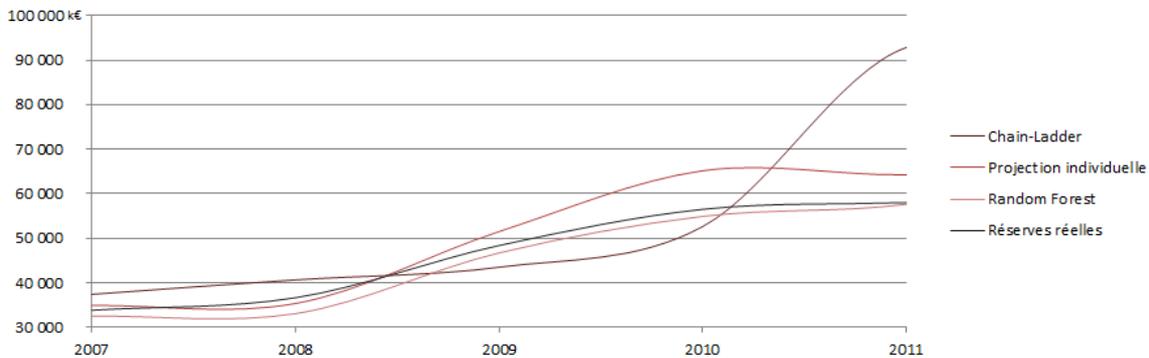


FIGURE 7.14 – Comparaison des provisions estimées par les trois méthodes

Graphiquement, la projection des provisions à l'ultime pour chaque exercice avec l'algorithme *Random Forest* est plus précise que les deux méthodes précédentes. De plus, la nouvelle méthode de projection permet de continuer à diminuer les variations des provisions d'un exercice à l'autre. Le tableau suivant présente les écarts d'estimation faits avec chaque méthode. Comme précédemment, cela permet tester la qualité de chaque modèle.

	2007	2008	2009	2010	2011
Réserves réelles	33 792	36 596	48 331	56 476	57 987
Chain-Ladder	37 379	40 623	43 474	52 604	92 841
	10,6%	11,0%	-10,0%	-6,9%	60,1%
Projection individuelle	34 885	35 308	51 470	65 148	64 252
	3,2%	-3,5%	6,5%	15,4%	10,8%
Random Forest	32 457	33 026	46 674	54 894	57 608
	-4,0%	-9,8%	-3,4%	-2,8%	-0,7%

TABLE 7.16 – Écart entre les trois projections et la provision réelle

Les écarts confirment les observations faites sur le graphique. L'écart moyen après application du traité de réassurance pour la projection avec le modèle *Random Forest* est de 4,1%, cet écart d'estimation diminue de 1,7% par rapport à la projection sans la mise en place du traité de réassurance. Cette méthode de projection comme la méthode de projection individuelle s'adapte bien à la réassurance gérée tête par tête. La méthode de projection individuelle présente un écart moyen de 6,5% et la méthode de Chain-Ladder un écart de 19,7%. Au vu de ces tests, la méthode de projection par apprentissage automatique est de meilleure qualité.

7.8 Bilan de la méthode

La projection des sinistres avec l'algorithme d'apprentissage automatique permet d'inclure des variables qualitatives jusqu'à présent inexploitées. La base de sinistres dont nous disposons possède de nombreux renseignements sur le sinistre et son traitement.

Après avoir classifiées toutes les variables disponibles, l'ajout des informations liées à la survenance et au traitement du sinistre permettent d'obtenir un modèle de projection avec l'algorithme des forêts aléatoires. Dans un premier temps, deux modèles d'apprentissage ont été développés. Le deuxième étant l'amélioration du premier, il a été décidé de l'utiliser pour poursuivre la modélisation. Notons que le premier modèle améliore les estimations en comparaison de la projection avec la méthode de Chain-Ladder et de la projection individuelle.

Après avoir projeté les provisions pour les exercices 2007 à 2014 avec l'algorithme d'apprentissage statistique, la comparaison avec le modèle de projection individuelle et la méthode de Chain-Ladder ont montré que, graphiquement, les estimations d'un exercice à l'autre étaient moins volatiles. L'évolution du montant de provisions d'un exercice à l'autre avec cette nouvelle méthode de modélisation des sinistres est plus stable dans le temps.

Afin de valider la robustesse de ce nouveau modèle, les estimations ont été comparées sur la même base de test 2007 à 2014 que précédemment. L'erreur moyenne sur cette période est de 5,8% contre 9,1% pour l'estimation avec la projection individuelle et 12,9% avec la méthode de Chain-Ladder. En plus de réduire la volatilité, le modèle améliore l'estimation des provisions. Notons cependant que l'estimation obtenue sous-évalue lors de chacun des exercices le montant de provisions à constituer alors que le modèle de projection individuelle le surestime.

Enfin, un calcul de capital réglementaire théorique et le traité de réassurance ont été appliqués. La projection des provisions avec le modèle d'apprentissage automatique réduit a priori la volatilité des provisions.

Le traité de réassurance appliqué est le même que précédemment, comme pour la projection individuelle, l'estimation du montant des provisions est améliorée sur le périmètre de test, de plus l'évolution des provisions est moins volatile avec ce nouveau modèle. La modélisation des provisions après la mise en place du traité de réassurance par excédent de sinistre s'adapte bien avec le modèle d'apprentissage supervisé.

L'estimation des provisions avec l'algorithme *Random Forest* a l'avantage de réduire la volatilité des provisions au cours du temps et d'améliorer les estimations. D'un point de vue opérationnel, les temps de calculs pour la projection sont négligeables par rapport à la méthode de projection individuelle. De plus, les forêts aléatoires donne des résultats pertinents malgré le petit nombre d'observations et le grand nombre de variables. Les nouvelles méthodes d'apprentissage statistique vont permettre aux assureurs d'exploiter le grand nombre de données collectées et d'optimiser leur collecte de données. La limite de cette projection est que la justification des estimations des provisions par l'algorithme *Random Forest* est difficile à expliquer notamment auprès du régulateur. La traçabilité dans les arbres de décisions et leur construction est très difficile à présenter et à appréhender.

Conclusion

L'estimation des provisions spécifiquement pour les sinistres atypiques par des méthodes agrégées comme Chain-Ladder ou le modèle de Mack conduit à une mauvaise maîtrise du risque au cours du temps.

Dans le cadre de cette étude, l'estimation des provisions par une projection individualisée des sinistres permet de réduire l'erreur d'estimation et la volatilité sur les provisions d'un exercice à l'autre. Cette méthode très calculatoire s'avère surestimer le montant de provisions sur chacun des exercices de la base de test.

Afin d'exploiter les nombreuses caractéristiques récoltées par l'assureur sur la vie de chaque sinistre, une méthode de projection par apprentissage automatique permet d'une part de classifier les données à la disposition de l'assureur, et d'autre part de prédire la charge ultime à l'aide de ces informations. Il en ressort que la projection réalisée réduit l'erreur d'estimation et atténue d'avantage la volatilité des provisions d'un exercice à l'autre.

Ce nouvel outil issue de l'apprentissage statistique permet donc d'améliorer les performances prédictives et calculatoires en appréhendant les variables liées au développement d'un sinistre. Cependant le manque de traçabilité ne permet pas à l'assureur de l'utiliser comme unique outil de provisionnement mais plus comme un atout permettant un suivi de son risque de réserve. La seconde méthode de projection individuelle des sinistres trouve un véritable intérêt lors de l'utilisation d'un traité de réassurance non-proportionnelle.

Bibliographie

- [1] ACPR . (2015). Analyses et synthèse : Situation d'un échantillon de groupes d'assurance actifs en France à fin 2014.
- [2] A. LIAW and M. WIENER. (2015). Breiman and cutler's random forests for classification and regression. R package version 4.6-12.
- [3] ANGOUA Y. (2015-2016). Cours de provisionnement non-vie, EURIA.
- [4] BELIN V.H. (2015-2016). Cours de Contexte règlementaire et prudentiel, euria.
- [5] BREIMAN L. (2001). Bagging predictors, machine learning.
- [6] BREIMAN L. (2001). Random forest, machine learning.
- [7] CHALNOT J-B., GREMILLET M., MIEHE P., and J. TRUFIN. (2015). Implementing the individual claims reserving method, a new approach in non-life reserving.
- [8] EFRON B. (1979). Bootstrap methods : Another look at the jackknife. the annals of statistics.
- [9] EIOPA . (2014). Eiopa : Delegated acts : Mesures de niveau 2.
- [10] ENGLAND P.D. and VERRALL R.J. (2002). Stochastic claims reserving in General Insurance.
- [11] GUILLOT A. (2015). Apprentissage statistique en tarification non-vie : quel avantage opérationnel?, Institut des Actuaire.
- [12] JIN X. and FREES E.W. (2013). Comparing micro- and macro-level loss reserving models.
- [13] JIN X. (2014). Micro-Level Loss Reserving models with applications in workers compensation insurance.
- [14] LAILY R. (2014). Construction d'un modèle interne partiel en assurance non-vie, Institut des Actuaire.
- [15] MACK T. (1993). Distribution-free calculation of standard error of Chain-Ladder reserve estimates.
- [16] MACK T. (1999). The standard error of Chain-Ladder reserve estimate : recursive calculation and inclusion of a tail factor.
- [17] M. GESMANN, D. MURPHY, Y. ZHANG, A. CARRATO, G. CRUPI, C. DUTANG, A. CHARPENTIER, M. WUTRICH, and F. CONCINA. (2015). Statistical methods and models for claims reserving in general insurance. R package version 0.2.2.
- [18] MURPHY K. and McLENNAN A. (2006). A method for projecting individual large claims.
- [19] R Core Team. (2015). R : A language and environment for statistical computing.
- [20] SAC J., PETIT M., and DONIO M. (2016). Formule standard usp.

- [21] SCORNET E. (2015). Apprentissage et forêts aléatoires.
- [22] VERRALL R.J. (2000). An investigation into stochastic claims reserving models and the Chain-Ladder technique.

Lexique

Année de développement : L'année de développement fait référence à la vie d'un sinistre en assurance non-vie.

Année de survenance : L'année de survenance correspond à l'année d'origine du sinistre en assurance non-vie.

Apprentissage automatique : L'apprentissage non supervisé est une méthode d'apprentissage automatique où l'algorithme traite les données comme des variables aléatoires et construit un modèle de « densités jointes » pour cet ensemble de données.

Arbre de décision C'est un outil d'aide à la décision représentant un ensemble de choix sous la forme graphique d'un arbre.

Assurance : Contrat par lequel l'assureur s'engage à indemniser l'assuré, moyennant une prime ou une cotisation, de certains risques ou sinistres éventuels.

Assurance Non-vie : L'assurance non-vie regroupe les opérations qui n'ont pas pour objet la vie de l'assuré.

Bootstrap : Technique de rééchantillonnage avec remise utilisant des simulations Monte-Carlo.

MCR : Le capital minimum requis représente le niveau minimum de fonds propres en dessous duquel l'intervention de l'autorité de contrôle est automatique.

MSEP : Mean square error prediction, correspond à l'erreur moyenne de prédiction.

PSAP : La provision pour sinistre à payer vise à couvrir les paiements des sinistres survenus avant une date d'évaluation.

Réassurance : La réassurance est l'assurance des sociétés d'assurance. Elle permet à l'assureur de se décharger d'une partie de son risque.

Résponsabilité civile : C'est la responsabilité considérée du point de vue des dommages qu'une action volontaire ou non, ainsi qu'une absence d'action prévue dans un contrat, ont pu causer.

Résponsabilité civile corporelle : C'est un contrat qui garantit les conséquences pécuniaires encourues par l'assuré lorsque celui-ci cause un dommage corporel à un tiers que ce soit par sa négligence.

SCR : Capital de solvabilité requis, il s'agit du capital requis sous la Directive Solvabilité II pour qu'une compagnie d'assurance soit solvable dans un an dans 99,5% des cas.

Sinistralité attritionnelle : Il s'agit de la sinistralité classique, récurrente. Elle est composée des sinistres ayant une fréquence importante et une sévérité peu élevée.

Sinistralité atypique : Il s'agit des sinistres graves qui sont plus rares mais de forte sévérité.

Annexes

Démonstration des théorèmes du modèle de Mack

Théorème 1 :

En notant $D = \{C_{i,j} ; i + j \leq n + 1\}$ l'information apportée par le triangle supérieur, la relation suivante est vérifiée :

$$E[C_{i,n}|D] = f_{n-i+1} \times \cdots \times f_{n-1} \times C_{i,n-i+1}$$

Preuve :

$$\begin{aligned} E[C_{i,n}|D] &= E[C_{i,n}|C_{i,1} \cdots C_{i,n+1-i}] \\ &= E[E[C_{i,n}|C_{i,1} \cdots C_{i,n-1}]|C_{i,1} \cdots C_{i,n+1-i}] \\ &= E[C_{i,n-1} \times f_{n-1}|C_{i,1} \cdots C_{i,n+1-i}] \\ &= E[E[C_{i,n-1}|C_{i,1} \cdots C_{i,n-2}]|C_{i,1} \cdots C_{i,n+1-i}] \times f_{n-1} \\ &= E[C_{i,n-2} \times f_{n-2}|C_{i,1} \cdots C_{i,n+1-i}] \times f_{n-1} \\ &= \cdots \\ &= E[C_{i,n+1-i}|C_{i,1} \cdots C_{i,n+1-i}] \times f_{n-1} \times f_{n-2} \times \cdots \times f_{n+1-i} \\ &= C_{i,n-i+1} \times f_{n-1} \times \cdots \times f_{n-i+1} \end{aligned}$$

Théorème 2 :

Sous les hypothèses précédentes, les facteurs de développement de Chain-Ladder $\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}$ sont des estimateurs sans biais de f_j et sont non corrélés.

Preuve :

Soit $B_k = \{C_{i,j}|j \leq k, i + j \leq n + 1\}, 1 \leq k \leq n$.

Démontrons que les estimateurs des facteurs de développement sont des estimateurs sans biais :

L'hypothèse 1 du modèle de Mack donne :

$$E[C_{i,k+1}|B_k] = E[C_{i,k+1}|C_{i,1} \cdots C_{i,k}] = f_k \cdot C_{i,k}$$

Nous en déduisons que :

$$\begin{aligned} E[\hat{f}_k|B_k] &= E\left[\sum_{j=1}^{n-k} \frac{C_{j,k+1}}{C_{j,k}}|B_k\right] \\ &= \frac{\sum_{j=1}^{n-k} E[C_{j,k+1}|B_k]}{\sum_{j=1}^{n-k} C_{j,k}} \\ &= f_k \end{aligned}$$

Il en découle alors :

$$E[f_k] = E[E[\hat{f}_k|B_k]] = E[\hat{f}_k]$$

Les estimateurs des facteurs de développement sont des estimateurs sans biais.

Démontrons la non corrélation des estimateurs des facteurs de développement :

$$\begin{aligned} E[\hat{f}_k \hat{f}_j] &= E[E[\hat{f}_k \hat{f}_j|B_k]] \\ &= E[\hat{f}_j \cdot E[\hat{f}_k|B_k]] \\ &= E[\hat{f}_j] \cdot f_k \\ &= E[\hat{f}_k] \cdot E[\hat{f}_j] \end{aligned}$$

Théorème 3 :

- (i) L'estimateur $\hat{C}_{i,n} = f_{n-i+1} \times \cdots \times f_{n-1} \times C_{i,n-i+1}$ est un estimateur sans biais de $E[C_{i,n}|D] = f_{n-i+1} \times \cdots \times f_{n-1} \times C_{i,n-i+1}$.
- (ii) L'estimateur du montant de provision de l'exercice i , $\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i+1}$ est un estimateur sans biais de $E[R_i|D]$ avec $R_i = C_{i,n} - C_{i,n-i+1}$.

Preuve :

Prouvons que $\hat{C}_{i,j}$ est un estimateur sans biais de $C_{i,j}$:

$$\begin{aligned} E[\hat{C}_{i,n}] &= E[E[\hat{C}_{i,n}|D]] \\ &= E[f_{n-i+1} \times \cdots \times f_{n-1} \times C_{i,n-i+1}] \\ &= E[E[C_{i,n}|D]] \\ &= E[C_{i,n}] \end{aligned}$$

Nous pouvons ainsi en déduire que \hat{R}_i est un estimateur sans biais de R_i :

$$E[\hat{R}_i] = E[\hat{C}_{i,n} - C_{i,n-i+1}] = E[\hat{C}_{i,n}] - E[C_{i,n-i+1}] = E[C_{i,n}] - E[C_{i,n-i+1}] = E[R_i]$$

Présentation des tests statistiques

Le diagramme Quantile-Quantile

Le diagramme Quantile-Quantile, noté QQ-plot est un outil graphique qui permet d'estimer la pertinence d'une loi de probabilité théorique ajustée sur la distribution des observations. Ce diagramme permet également de comparer deux distributions.

Principe À partir d'une série d'observations statistiques (x_1, \dots, x_n) , les quantiles de cette distribution (q_1, \dots, q_n) sont calculés.

Les quantiles (q_1^*, \dots, q_n^*) de la distribution théorique ajustée sur les observations sont calculés. Pour rappel, notons F la fonction de répartition de la loi ajustée. Les quantiles théoriques sont calculés comme :

$$q_i^* = F^{-1}\left(\frac{i}{n}\right)$$

Le nuage de points (q_i, q_i^*) est tracé. Si ces points sont alignés sur la première bissectrice, le modèle ajusté est validé.

Le test de Kolmogorov-Smirnov

Le test de Kolmogorov-Smirnov, noté ks-test, est un test statistique qui permet de valider qu'un échantillon donné suit une loi théorique ou que deux échantillons suivent la même loi.

Ce test s'appuie sur les fonctions de répartitions empiriques. Soit F_n , la fonction de répartition d'un échantillon de loi inconnue \mathcal{L} :

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i \leq x} \text{ avec } \delta_{x_i \leq x} = \begin{cases} 1 & \text{si } x_i \leq x, \\ 0 & \text{sinon.} \end{cases} \text{ dans la pratique } n > 50.$$

Soit F une fonction de répartition d'une loi continue connue. L'hypothèse H_0 : "La loi \mathcal{L} a pour fonction de répartition F ", est vérifiée dans $1 - \alpha$ cas si :

$$D_n = \sup_x |F_n(x) - F(x)| \leq D_{n,\alpha}$$

La distance de Kolmogorov-Smirnov $D_{n,\alpha}$ est à retrouver dans le tableau suivant :

$n \backslash \alpha$	0,1	0,05	0,01
1	0,9500	0,9750	0,9950
2	0,7764	0,8419	0,9293
3	0,6360	0,7076	0,8290
4	0,5652	0,6239	0,7342
5	0,5095	0,5633	0,6685
6	0,4680	0,5193	0,6166
7	0,4361	0,4834	0,5758
8	0,4096	0,4543	0,5418
9	0,3875	0,4300	0,5133
10	0,3697	0,4092	0,4889
11	0,3524	0,3912	0,4677
12	0,3381	0,3754	0,4491
13	0,3255	0,3614	0,4325
14	0,3142	0,3489	0,4176
15	0,3040	0,3376	0,4042
16	0,2947	0,3273	0,3920
17	0,2863	0,3180	0,3809
18	0,2785	0,3094	0,3706
19	0,2714	0,3014	0,3612
20	0,2647	0,2941	0,3524
21	0,2586	0,2872	0,3443
22	0,2528	0,2809	0,3367
23	0,2475	0,2749	0,3295
24	0,2424	0,2693	0,3229
25	0,2377	0,2640	0,3166
30	0,2176	0,2417	0,2899
35	0,2019	0,2242	0,2690
40	0,1891	0,2101	0,2521
45	0,1786	0,1984	0,2380
50	0,1696	0,1884	0,2260
60	0,1551	0,1723	0,2067
70	0,1438	0,1598	0,1917
80	0,1347	0,1496	0,1795
90	0,1271	0,1412	0,1694
100	0,1207	0,1340	0,1608
$n > 100$	$\frac{1,223}{\sqrt{n}}$	$\frac{1,358}{\sqrt{n}}$	$\frac{1,629}{\sqrt{n}}$

TABLE 17 – Table de Kolmogorov-Smirnov