

**Mémoire présenté devant le CNAM pour l'obtention du Master
Droit Economie Gestion, mention Actuariat
et l'admission à l'Institut des Actuares**

le 10 Juillet 2025

Par : Jean-Michel VAUTRIN

Titre: Modèle de provisionnement des sinistres sur données détaillées en
assurance de responsabilité civile des risques industriels dans
l'environnement de Solvabilité II

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

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

Président du Jury :
M. Stéphane LOISEL

signatures

Entreprise :
Nom : HDI Global SE

*Membres présents du jury de
l'Institut des Actuares :*

Mme Edith BOCQUAIRE
M. Michel GERMAIN
M. Ravin DUMANOIR

Directeur de mémoire :
Nom : Philippe KUTEK
Signature :

*Membres présents du jury du
Cnam :*

M. Olivier DESMETTRE

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

Secrétariat :

Bibliothèque :

Signature du responsable entreprise

Signature du candidat

Table des matières

Table des matières.....	2
Remerciements.....	4
Résumé.....	5
Abstract.....	6
Glossaire.....	7
1 Introduction.....	8
2 Cadre de l'étude.....	11
2.1 L'assurance de responsabilité civile des risques industriels.....	11
2.2 Enjeux du provisionnement des sinistres.....	17
2.3 Solvabilité II et provisionnement des sinistres.....	21
3 Présentation des données et outils utilisés.....	32
3.1 Structure des sinistres du portefeuille.....	32
3.2 Indexation des variables quantitatives.....	35
3.3 Outils utilisés.....	36
4 Méthodes de provisionnement.....	38
4.1 Modèles sur données agrégées.....	39
4.2 Modèle sur données détaillées.....	44
5 Mise en œuvre des modèles de provisionnement.....	49
5.1 Modèle sur données détaillées – module 1 – Génération des sinistres non ouverts (IBNYR).....	49
5.2 Modèle sur données détaillées – module 2 – Développement des sinistres déclarés.....	64
5.3 Chain-Ladder.....	86
5.4 Chain-Ladder stochastique I (Mack).....	89
5.5 Chain-Ladder stochastique II (GLM/Bootstrap).....	90
6 Résultats et impact dans l'environnement de Solvabilité II.....	92
6.1 Résultats à l'ultime.....	92
6.2 Résultats à horizon 1 an.....	94
6.3 Potentiel et limite d'exploitation du modèle détaillé.....	97
7 Complément sur la modélisation individuelle des sinistres depuis 2013 et comparaison avec le modèle étudié.....	99

7.1	Revue et discussion sur une sélection d'articles.....	99
7.2	Pistes originales du modèle détaillé de ce mémoire et autres idées de prolongement de l'étude	107
8	Conclusion	110
9	Bibliographie.....	113
10	Annexes	115
10.1	Principales théories de modélisation utilisées	115
10.2	Tableaux de résultats des régressions statistiques et des ajustements paramétriques du modèle sur données détaillées des sinistres.....	134
10.3	Triangles de liquidation des sinistres	140
10.4	La société HDI Global SE au 31/12/2010.....	142
10.5	Informations sur les sinistres.....	142
10.6	Informations sur les primes.....	143
10.7	Informations principales sur les contrats.....	144

Remerciements

Je tiens à remercier en premier lieu mes enfants, Diane, Flore et Appoline, pour leur soutien et leur patience durant toutes ces années de cours du soir et de travail sur ce mémoire.

Je remercie vivement M. Philippe Kutek, qui a aimablement accepté d'être mon tuteur pour ce mémoire, pour ses conseils éclairés et sa disponibilité qui m'ont permis de mener à bien cette étude.

Je remercie MM. Jean-Marie Haquette et Dominique Guérit, respectivement actuel et ancien directeurs généraux d'HDI Global SE à Paris, qui m'ont permis de réaliser cette étude au sein de leur entreprise, ainsi que M. Jens Wohlthat, ancien membre du comité de direction d'HDI Global SE. J'y associe l'ensemble de l'équipe d'HDI Global SE à Paris pour leur disponibilité, leur obligeance et leur expertise qui m'ont aidé tout au long de cette étude. En particulier M. Pascal Deliry, ancien directeur indemnisation d'HDI Global SE à Paris, et Mme Catherine Renodon Delubria, ancienne directrice produits et relations grands comptes d'HDI Global SE à Paris qui m'ont tous deux énormément appris sur les spécificités des contrats RC des risques industriels et leurs conséquences indemnitaires.

Je souhaite remercier aussi l'ensemble de ceux que j'ai eu le privilège de côtoyer durant ces deux décennies de travaux actuarielles et assurantielles sans lesquels cette étude n'aurait pas été possible. A ce titre je tiens à remercier toute l'équipe pédagogique du CNAM et en particulier M. Stéphane Loisel, ainsi que M. Guillaume Gorge qui m'a ouvert les portes de l'actuariat, de l'assurance, et de la réassurance il y a 25 ans, et Mme Anne-Cécile Kerdranvat sans qui mon chemin de vie aurait sans doute été très différent.

Enfin, je tiens également à rendre hommage à la mémoire de M. Michel Fromenteau qui a su de si nombreuses fois transformer des journées de travail stressantes ou maussades en soirées passionnantes et radieuses par ses grandes qualités humaines et pédagogiques pendant ses cours du soir au CNAM.

Résumé

Le provisionnement des sinistres est un enjeu majeur pour l'assureur, étant un poste essentiel de son bilan en quantifiant les prestations futures auxquelles il s'est engagé. Pour le régulateur également, car il exige de l'assureur une pratique rigoureuse dans son estimation et dans sa résilience aux conditions extrêmes pour en déduire sa solvabilité. Ceci est d'autant plus complexe pour l'assureur dans le cadre d'un portefeuille aux risques hétérogènes et volatiles telle la couverture en responsabilité civile générale de risques industriels. En parallèle, des études ont développé des méthodologies d'estimations individuelles détaillée des provisions pour sinistre, en alternative des méthodes usuelles basées sur des données agrégées de liquidation des sinistres, afin d'en perfectionner l'estimation par l'emploi d'informations supplémentaires disponibles dans les données de l'assureur.

Dans ce contexte, ce mémoire présente la conception et le développement d'un modèle détaillée appliqué à un tel portefeuille d'assurance, et permet d'identifier plusieurs éléments clés de la liquidation des sinistres impactant le besoin de provisionnement, tout en fournissant une granularité de résultats rendant possible des calculs ultérieurs, comme sur la réassurance, et des reporting répondant à des besoins analytiques et réglementaires exigeants.

Même si sa réalisation fût faite il y a une quinzaine d'années, la littérature actuarielle a depuis confirmé l'intérêt général de cette approche ainsi que d'options de modélisation choisies à l'époque, tout en ouvrant des voies sur de nouvelles techniques, tels les modèles à apprentissage automatique, et qui peut être encore étendu aujourd'hui à l'utilisation d'agents IA.

Mots clés : modélisation individuelle des sinistres, GLM, modèle de survie, loi normale contaminée, garantie en responsabilité civile générale, risques industriels.

Abstract

Claims reserving is a major issue for insurers, as it is a major item on their balance sheet, quantifying the future benefits to which they have committed. And for the regulator too, as he requires to the insurer to be rigorous in its estimates and in its ability to withstand extreme conditions to deduce his solvency. Especially for the insurer when dealing with a heterogeneous and volatile risks portfolio, such as general liability cover for industrial risks. At the same time, studies have developed methodologies for micro-level claims reserving, as an alternative to the usual methods based on aggregated claims settlement, to improve estimation by using additional information available in the insurer's data.

In this context, this thesis presents the design and development of a micro-level reserving model applied to such as insurance portfolio and makes it possible to identify several key elements of the settlement of claims impacting the reserving request, while providing a results granularity allowing subsequent calculations, such as reinsurance application, and reporting meeting demanding analytical and regulatory needs.

Although it was developed about fifteen years ago, the actuarial literature has since confirmed the general interest of this approach and of some keys modelling options chosen at the time, while also opening to new techniques, such as machine-learning models, which can still be extended today to the use of AI agents.

Keywords: Micro-level reserving, GLM, survival model, contaminated normal distribution, general third-party liability coverage, industrial risks.

Glossaire

BE	<i>Best estimate</i>
BSCR	<i>Basic solvency capital requirement</i>
CART	<i>Classification and regression tree</i>
CL	<i>Chain-Ladder</i>
EIOPA	<i>European Insurance and Occupational Pensions Authority</i>
ERM	<i>Entreprise risk management</i>
FFSA	<i>Fédération française des sociétés d'assurances</i>
FGU	<i>From ground-up</i>
GEMA	<i>Groupement des entreprises mutuelles d'assurance</i>
GLM	<i>Modèle linéaire généralisé (general linear model)</i>
GTPL	<i>General third party liability</i>
IA	<i>Intelligence artificielle</i>
IARD	<i>Incendie, Accidents et Risques Divers</i>
IBNER	<i>Incurred but not enough reported</i>
IBNYR	<i>Incurred but not yet reported</i>
ID	<i>Identifiant</i>
iid	<i>Indépendants et identiquement distribués</i>
IPC	<i>Indice de prix à la consommation</i>
MCR	<i>Capital minimum requis (Minimum capital requirement)</i>
ML	<i>Machine learning</i>
ORSA	<i>Own Risk and Solvency Assessment</i>
QRT	<i>Quantitative reporting template</i>
RBNS	<i>Report to the insurer But Not Settled</i>
RC	<i>Responsabilité civile</i>
RCG	<i>Responsabilité civile générale</i>
RM	<i>Risk margin</i>
SCR	<i>Capital de solvabilité requis (Solvency capital requirement)</i>
U.E.	<i>Union Européenne</i>
VaR	<i>Value at risk</i>

1 Introduction

L'activité d'assurance repose sur un engagement de prestations, principalement financières, en cas de réalisation des conditions prévues au contrat d'assurance. A cela se confronte deux visions adverses : l'assureur qui souhaite immobiliser le moins possible de capital pour faire face à ses engagements afin de maximiser sa profitabilité, et le législateur qui souhaite protéger les intérêts de ces concitoyens, et qui veille au respect des obligations des assureurs par l'intermédiaire du régulateur. Le principe directeur de ce dernier s'énonce simplement : l'assureur doit faire face à l'ensemble de ses engagements à tout instant. Pour cela le législateur européen a adopté en 2009 la directive Solvabilité II qui est appliquée dans l'U.E. depuis le 1^{er} janvier 2016. Son objectif est triple : s'assurer de l'adéquation des fonds propres de l'assureur au regard de ses engagements pris, de la mise en place d'une gouvernance des risques et de leurs bonnes évaluations, et enfin d'en établir une communication transparente auprès du public et du régulateur. L'assureur doit ainsi souscrire des contrats en contrôlant continuellement l'impact des risques sous-jacents sur son niveau de solvabilité qui doit au moins atteindre le niveau requis par le législateur. Ce principe motive l'assureur à opter pour une stratégie qui peut être résumée par cette simple phrase : on ne souscrit que ce que l'on peut diversifier et comme rappeler dans l'ouvrage « Manuel d'Assurance » (Bénéplanc et al., 2022). Pour cela l'assureur dispose de deux grands leviers : par accumulation grâce aux conséquences de la loi des grands nombres sur ses risques, et en cession, par la réassurance, qui lui permet de retirer de son portefeuille les risques adverses à sa stratégie de solvabilité et de capital.

Toute réflexion dans ce domaine commence par la mise en place, par l'assureur, de méthodes d'estimations des engagements et des risques fiables et adéquates. Ceci est particulièrement important pour l'assurance de risques industriels par la nature même de leur secteur d'activité : affaires souscrites par réseaux non-propriétaires, nombre restreint et peu homogène de clients, sinistralité plus volatile que celle observée sur les risques de masse, montants d'engagements des polices conséquents. Ainsi cette analyse est fondamentale pour ce type de portefeuille pour valider et ajuster la stratégie de souscription au regard de la structure du portefeuille. Dans ce cadre, les provisions pour sinistres à payer, qui matérialisent les montants d'engagements de l'assureur pour faire face à l'ensemble des prestations futures issues des affaires souscrites, sont le principal poste au bilan en volume et réclament donc une attention toute particulière. D'autant plus lorsqu'elles concernent des garanties à développement long comme la responsabilité civile générale, qui en plus de l'incertitude sur les montants ultimes d'indemnités des polices d'assurance des risques industriels ajoute aussi une dimension temporelle obligeant l'assureur à immobiliser des ressources sur de longues périodes tout en étant incertain de l'échéancier réel des futurs flux de trésorerie. Ce qui est le cas de cette étude qui se base sur les données de la branche responsabilité civile générale des risques industriels de la succursale française de l'assureur HDI Global SE.

Même si les méthodes usuelles basées sur des triangles agrégés de liquidation des sinistres sont généralement robustes et assez simples à mettre en œuvre dans le cadre de portefeuille de risques dit de masses, comme sur l'assurance non-vie des particuliers, elles se révèlent beaucoup plus fragiles sur un portefeuille de contrats RCG de risques industriels, d'autant plus de taille limitée comme celui qui sert de base à cette étude. Au-delà des faiblesses intrinsèques des méthodes basées sur les triangles, comme sur l'estimation des valeurs périphériques de ces triangles, ici un seul sinistre, par une forte et brusque variation de son montant indemnitaire, est susceptible de dévier à lui seul significativement les estimateurs du triangle, aliénant la nature même de la méthode qui cherche à caractériser les macro-tendances de liquidation des sinistres de ce portefeuille. Toutefois lorsque cette étude a été initiée en 2009, l'avancement des techniques actuarielles et des capacités informatiques a rendu envisageable de développer une approche basée directement sur les données détaillées des polices et des sinistres, comme présenté dans l'article « Loss reserving using claim-

level data » (Guszcza et Lommele, 2006). Dès 2004, le mémoire « Modèle de provisionnement sur données détaillées en assurance non-vie » (Beneteau, 2004) propose déjà une approche appliquée de provisionnement sur données détaillées d'un portefeuille de RC professionnelle sur la partie IBNER uniquement. Et nombres d'études se sont déroulés ensuite, comme nous le verrons dans le chapitre 7, jusqu'à des mémoires réalisés récemment comme « Assurance construction : Provisionnement Dommages-ouvrage exploitant les typologies des sinistres et les risques assurés » (Skoubani, 2022) avec un enjeu important sur la garantie décennale en construction ou encore « Approches de provisionnement individuel en réassurance non-vie » (Giuly, 2018) qui cherche à caractériser les provisions de chaque traité d'un portefeuille de réassurance. Ici, l'objectif de cette étude sera la conception et l'application d'un prototype de modèle de provisionnement des sinistres dit détaillé par dossier (micro-level reserving en Anglais) sur un portefeuille d'assurance de risques industriels de couverture en responsabilité civile générale. Le principe, développé en détail dans ce document, est assez simple finalement : l'observation de la base des contrats et de la base comptable des transactions a permis d'identifier de manière empirique les principaux mécanismes qui animent le déroulement de la vie des sinistres étudiés dans les comptes de l'assureur, de les traduire en modèles mathématiques pour en reproduire le comportement stochastique, et de comparer quelques résultats clés avec les méthodes actuarielles éprouvées reposant sur des données agrégées, et ce aussi bien pour les sinistres dits IBNER que pour les sinistres dits IBNYR. Ainsi seront analysées les informations quantitatives et qualitatives qui influent sur le processus de liquidation de ce type de sinistres et de voir en quoi cela permet d'explicitier des mécanismes clés sous-jacents, ainsi que la plus-value de résultats détaillés afin de permettre des calculs dédiés plus précis ultérieurement, en particulier dans le cas de montages d'assurance et de réassurance complexes mal reflétés dans les méthodes agrégées usuelles. Le développement de ce prototype était achevé, ainsi que 90% de la rédaction de ce document, lorsqu'il fut mis de côté pour des raisons professionnelles et personnelles en 2013. Il put être repris en 2024, donnant une opportunité, rare dans le cadre d'un mémoire d'actuariat, de comparer cette étude avec le développement de l'état de l'art de ce thème dans la décennie qui a suivi. Cette situation a permis d'enrichir l'analyse originelle en présentant une revue d'articles remarquables sur ce thème parus au cours de la dernière décennie et en les commentant par rapport au travail de cette étude réalisée antérieurement.

Il est d'abord présenté au chapitre 2 le cadre de l'étude avec une première partie rappelant la définition de l'assurance de responsabilité civile générale ainsi que ses principaux mécanismes légaux et contractuels, base fondamentale et préliminaire à toute analyse actuarielle qu'elle soit tarifaire ou d'estimation des provisions. Puis, après quelques rappels sur le marché d'assurance français et la typologie de la liquidation d'un sinistre, la directive de solvabilité II est rappelée en se concentrant sur les éléments liés au provisionnement des sinistres d'un portefeuille de responsabilité civile générale, et en particulier sur la formule standard permettant le calcul de solvabilité, SCR et MCR. Ce chapitre permet ainsi d'obtenir une vision d'ensemble de l'environnement réglementaire et du contexte de marché du portefeuille étudié par la suite.

Le chapitre 3 présente des informations clés sur le portefeuille analysé et, en particulier, met en évidence la très grande dispersion du montant des sinistres historiques motivant l'investigation dans une méthode de provisionnement alternatives.

Ensuite sont présentées les méthodes déterministes et stochastiques appliquées dans cette étude au chapitre 4, à commencer par les méthodes usuelles, puis le concept du modèle détaillé proposée de ce mémoire.

La mise en œuvre de ce modèle détaillé fait l'objet du chapitre 5 en décrivant en détail sa structure et son articulation modulaire, où chaque module reflète un comportement clé observé de la liquidation d'un sinistre. Ensuite pour chaque module est décrit le modèle mathématique qui a été

utilisé, si besoin, en explicitant la méthode de simulation correspondante, et en développant les commentaires pour les résultats les plus importants.

Le chapitre 6 présente les résultats des différents modèles sur base des simulations et utilise ses résultats pour le calcul réglementaire du SCR et du MCR, puis explique les avantages et limites des modèles au regard des résultats obtenus.

Le dernier chapitre, 7, explore une sélection d'articles parus dans la dernière décennie et postérieurs aux travaux décrits précédemment en commentant les principales différences, puis en montrant les aspects du modèle détaillé de ce mémoire qui restent pertinents et originaux aujourd'hui, avant de terminer ce chapitre sur les possibles voies nouvelles à explorer au regard de l'évolutions des connaissances en cours.

Enfin la conclusion reprend l'ensemble des éléments de ce mémoire, appuie sur ses aspects importants, la construction du modèle, ses plus-values et ses limites, avant d'éclairer sur les possibles voies de développement ultérieurs de ce thème riche et passionnant.

Il convient de préciser un dernier aspect pratique et limitatif : les données utilisées ne sont plus accessibles, même si les résultats clés ont été conservés ainsi que l'ensemble des algorithmes développés et utilisés.

2 Cadre de l'étude

2.1 L'assurance de responsabilité civile des risques industriels

2.1.1 Définitions

Tout d'abord, il convient de définir ce qu'est une assurance de responsabilité civile des risques industriels afin de bien en appréhender les risques. Nous proposons de reprendre les définitions fournies par François Chaumet dans son ouvrage de 2011 « Les assurances de responsabilité de l'entreprise – 5^e édition » :

La responsabilité civile

Les responsabilités pesant sur l'entreprise sont essentiellement de deux ordres, la responsabilité pénale, et la responsabilité civile. Nous laissons de côté le volet pénal, qui a pour but de punir les auteurs d'infractions troublants l'ordre public, pour nous concentrer sur la partie civile.

La responsabilité civile implique nécessairement l'existence d'un dommage à réparer, quelle que soit la gravité de la faute commise à l'origine de celui-ci, et même, de plus en plus souvent sans faute de la part de l'auteur du dommage. Elle peut être retenue à l'encontre de la personne qui a personnellement causé le dommage, mais celle-ci peut être également amenée à répondre du fait d'autrui ou du fait des choses. C'est la victime du dommage qui va en demander réparation, en exerçant une action civile obéissant aux règles de la procédure civile et se voyant appliquer les délais de prescription du droit civil.

La responsabilité civile, une fois établie, donnera lieu à la réparation du dommage causé à la victime. Cette réparation est indépendante de la gravité de la faute mais proportionnée à l'ampleur du dommage. C'est à l'auteur du dommage de réparer celui-ci, mais s'agissant d'une responsabilité civile, la réparation sera le plus souvent prise en charge par l'assurance.

Le dommage

Le dommage peut être causé à des biens (dommage matériel), à une personne physique (dommage corporel). Il peut résulter d'une atteinte à l'honneur ou à la vie privée ou encore d'une atteinte aux sentiments (préjudice d'affection).

Le dommage pour être réparable, doit :

- Constituer la violation d'un intérêt légitime juridiquement protégé, c'est-à-dire ne pas être contraire à l'ordre public ou aux bonnes mœurs.
- Être certain, ce qui signifie qu'un dommage éventuel ou hypothétique n'est pas réparable, en revanche un dommage futur l'est.
- Être personnel, c'est-à-dire ne peut donner lieu à réparation qu'en faveur de la victime qui l'a subi personnellement

- Être direct, ce qui implique qu'il ne soit pas la conséquence trop lointaine du fait générateur.

Une catégorie de dommage nous intéresse également dans le cadre de l'assurance d'activités industrielles : le dommage immatériel. Le dommage immatériel concerne tout préjudice pécuniaire résultant de la perte de jouissance d'un droit, de l'interruption d'un service rendu par une personne ou par un bien meuble (déplaçable) ou immeuble (bâtiment) ou de la perte d'un bénéfice attendu ou espéré. Il s'agit donc d'un dommage financier. Dans cette catégorie de dommage, nous pouvons distinguer :

- Le dommage immatériel consécutif : il peut être la conséquence d'un dommage corporel ou matériel garanti lorsqu'un contrat de responsabilité civil prévoit une extension au dommage immatériel.
- Le dommage immatériel non consécutif : il doit être la conséquence d'un dommage corporel ou matériel non garanti engageant la responsabilité d'un tiers assuré pour qu'une perte d'exploitation soit mise à la charge de l'assureur du tiers ayant contracté une garantie responsabilité civile.

Le fait générateur

La responsabilité civile suppose également un fait générateur qui, en matière délictuelle ou quasi-délictuelle (c'est-à-dire en dehors d'un dommage indépendant de tout lien juridique entre l'auteur et la victime) peut être :

- Soit le fait personnel de l'auteur du dommage, sans intervention d'une chose ou d'une personne dont celui-ci devrait répondre, et qui engage sa responsabilité pour faute sur le fondement, selon le cas, d'une faute délibérée ou d'une obligation générale de prudence et de diligence.
- Soit la responsabilité du fait des choses ou du fait des animaux.
- Soit la responsabilité du fait d'autrui.

En matière contractuelle, c'est-à-dire en présence d'un dommage lié à l'existence d'un contrat entre l'auteur du dommage et la victime, cela peut être :

- Soit le manquement personnel de l'un des contractants dans l'exécution de ses obligations à l'égard de l'autre.
- Soit la non-exécution ou l'exécution défectueuse de ses obligations par l'un des contractants à l'égard de l'autre mais du fait d'autrui concourant ou faisant obstacle à l'exécution du contrat, ou du fait des choses utilisées pour l'exécution du contrat.

Le lien de causalité

La responsabilité civile suppose enfin un lien de causalité entre le fait générateur et le dommage apprécié plus ou moins rigoureusement en fonction des circonstances de la cause et de la nature (délictuelle, quasi-délictuelle, ou contractuelle) de la responsabilité.

En présence de dommages, d'un coût souvent élevé, et de l'obligation qu'a l'auteur des dommages de les réparer, la pratique de l'assurance de responsabilité civile contre les conséquences pécuniaires d'une telle obligation n'a pu que se développer avec le temps. Sans elle, le responsable devrait y faire face personnellement en y répondant sur son patrimoine. En conséquence, l'assurance de responsabilité civile pour les entreprises industrielles ou commerciales est de toute première importance afin de garantir leurs activités et leurs conséquences en cas de mise en responsabilité.

Principales couvertures de responsabilité civile prises par les entreprises

Voici une synthèse des principales catégories d'assurance de responsabilité civile prises par les entreprises :

- La responsabilité civile d'exploitation : l'objet de cette couverture est de garantir les conséquences financières des dommages causés par toute personne attachée à l'entreprise vis-à-vis d'un tiers dans le cadre de l'activité habituelle ou occasionnelle de l'entreprise, pourvu que cette activité soit préalablement déclarée à l'assureur.
- La responsabilité civile après livraison ou travaux : cette garantie couvre les préjudices subis par des tiers, y compris les acquéreurs, et imputables aux produits livrés ou aux travaux exécutés par l'assuré.
- La responsabilité civile professionnelle : Cette couverture est directement liée à l'activité exercée par l'assuré afin d'en couvrir les conséquences financières sur des tiers ayant pour origine des fautes, des erreurs, des omissions, des inexactitudes ou des négligences commises au détriment des clients et des tiers. Pour un certain nombre de professions réglementées, cette assurance est obligatoire (architecte, agent d'assurance...)
- La défense-recours : c'est une garantie accessoire à la garantie principale de responsabilité civile. L'assureur finance la défense de l'assuré lorsque sa responsabilité civile est mise en cause par un tiers pour un préjudice causé dans le cadre de son activité professionnelle sans que sa responsabilité soit démontrée au démarrage de la procédure.

L'application dans le temps de l'assurance de responsabilité civile

Contrairement à l'assurance dommage, la réalisation d'un sinistre de responsabilité civile est composée de plusieurs faits successifs, dont certains peuvent survenir avant ou après la période de validité du contrat d'assurance. Ces événements sont au nombre de trois :

- Le fait dommageable : constitue la cause génératrice du dommage. Pour protéger les victimes, la jurisprudence a défini le fait dommageable comme la date de livraison d'un bien défectueux en cas de vente, ou la date d'exécution des travaux en cas de prestations de services.
- Le dommage : c'est le préjudice (corporel, matériel, ou immatériel) subi par la victime.

- La réclamation est la demande en réparation amiable ou contentieuse formulée par la victime d'un dommage ou ses ayants droit et adressée à l'assuré ou à son assureur.

L'application dans le temps des garanties est une particularité importante de l'assurance de responsabilité civile. Depuis la loi de sécurité financière du 01/08/2003, il existe deux modes d'applications¹ :

- Base fait dommageable : La garantie RC en fait dommageable est déclenchée dès lors que le fait dommageable survient entre la prise d'effet initiale et la résiliation/expiration de la garantie, quelle que soit la date de la réclamation. Le délai pour transmettre une réclamation à l'assureur est donc illimité en durée, sous réserve de l'application des règles de prescription : 30 ans en contractuel, 10 ans en extracontractuel. La datation du fait dommageable détermine la période d'assurance à laquelle le sinistre est affecté comptablement.
- Base réclamation : La garantie RC en base réclamation est déclenchée dès lors que le fait dommageable est antérieur à la résiliation/expiration de la garantie et que la première réclamation est adressée à l'assuré ou à son assureur entre la prise d'effet initiale de la garantie et le terme d'un délai subséquent. Une reprise du passé inconnu est automatiquement délivrée par l'assureur RC qui émet une garantie en base réclamation. La subséquente est déclenchée dès lors que la garantie RC en base réclamation n'est plus renouvelée par l'assureur. La base réclamation combine trois notions essentielles :
 - La période de validité : L'assureur RC qui émet la garantie prend en charge les réclamations qui lui sont transmises entre la date d'effet et la date de résiliation/expiration. La date de la première réclamation est donc le critère déterminant pour affecter comptablement les sinistres à la période de validité concernée.
 - La reprise du passé inconnu : Le terme « passé inconnu » recouvre les faits dommageables susceptibles de causer un dommage garanti à autrui, dont l'assuré n'a pas déjà eu connaissance à la date de souscription de la garantie en base réclamation. Les réclamations rattachées à ces faits dommageables antérieurs, inconnus à la date d'effet, seront prises en charge par l'assureur de la période de validité. Le contrat en base réclamation doit couvrir tout le passé inconnu, afin de maintenir dans le temps une continuité de couverture RC sur les activités passées de l'assuré. Cette garantie de reprise du passé inconnu est illimitée en durée. A contrario, le passé connu, c'est-à-dire les faits dommageables déjà connus de l'assuré à la date de souscription et susceptibles de causer un dommage garanti à autrui, sont exclus de la garantie RC souscrite en base réclamation.
 - La subséquente : La loi prévoit, pour une garantie souscrite en base réclamation, une subséquente minimale de 5 ans en cas de résiliation (changement d'assureur), d'expiration ou de suppression (cessation d'une des activités assurées). Ce délai supplémentaire permet la prise en charge

¹ Source : http://scor-front1.heb.fr.colt.net/www/fileadmin/uploads/publics/NTNV2004_07_fr_NTSV07fr.pdf

d'une réclamation adressée à l'assuré et/ou son assureur après la résiliation/expiration de la garantie. L'objet de la subséquente est donc limité aux réclamations rattachées à des faits dommageables connus à la date de résiliation/expiration de la garantie. En continuité de base réclamation et à chaque changement d'assureur, la subséquente sera déclenchée pour des réclamations rattachées à des faits dommageables relevant du passé connu. Lorsque le fait dommageable n'est connu qu'après expiration ou résiliation de la garantie, la mise en jeu de la subséquente implique que l'assuré n'ait pas resouscrit de police RC, ou bien ait resouscrit une police RC en fait dommageable qui, par définition, ne reprend pas le passé. Pour ce fait dommageable qui devient connu après expiration ou résiliation de la garantie, la subséquente relevant de la dernière garantie émise en base réclamation sera appelée à jouer en priorité. On clarifie ainsi la problématique éventuelle de cumuls de subséquentes, en cas de plafonds de garantie insuffisants ou de fréquents changements d'assureur. Un sinistre issu d'une réclamation transmise pendant la subséquente sera affecté à un plafond spécifique de garantie, qui ne pourra pas être inférieur à celui de la dernière période de validité.

Les deux modes d'application des garanties ne sont pas équivalents, et comportent des différences notables :

La base fait dommageable permet une délimitation des garanties dans le temps bien définie que ce soit pour l'assuré ou l'assureur. Mais du fait du délai, parfois long, nécessaire à la manifestation du préjudice, l'assureur rencontrera des difficultés à gérer ses risques du fait de la durée pendant laquelle il est susceptible de recevoir des réclamations, même si des prescriptions existent entre les tiers ayant subis un dommage et l'assuré, ainsi qu'entre l'assuré et l'assureur. Pour rappel la prescription en responsabilité civile est de 30 ans.

La base réclamation présente l'avantage pour l'assuré d'une couverture simple d'application puisqu'il ne s'adresse qu'à son assureur actuel, dans la majorité des cas, lors de la manifestation d'un préjudice. De même pour l'assureur, son exposition ne concerne que les contrats en cours et ceux pour lesquelles la subséquente entre en jeu. Il doit tenir compte aussi bien du profil de risque de l'assuré au moment de la souscription, mais aussi du profil de risque antérieur à la souscription impliqué par la reprise du passé. Il doit également s'assurer de la continuité de la garantie au-delà de son propre engagement pour anticiper son exposition à la suite de l'activation de la garantie subséquente.

Spécificité des contrats de responsabilité civile des risques industriels

Une particularité de cette garantie pour les risques industriels est le sur-mesure qui s'applique aux contrats de responsabilité civile. Que ce soit pour les particuliers ou les risques dits commerciaux, les assureurs proposent des contrats aux clauses et montants de garantie prédéfinis. A l'opposé, les risques industriels nécessitent l'adaptation des différents paramètres du contrat d'assurance aux risques proprement liés à l'exercice de l'activité de l'assuré, à sa politique de gestion des risques, à son environnement réglementaire, ou encore à l'exigence de garantie demandée par l'actionnariat de l'assuré. Ainsi, les montants de garanties nécessaires sont évalués contrat par contrat, les clauses

adaptées individuellement, le tout en respect des conditions d'acceptation du traité de réassurance de l'assureur.

2.1.2 Quelques chiffres du marché français

Afin de mettre en perspective le portefeuille étudié au sein du marché français, voici quelques chiffres clés extraits du document « Les Assurances de Biens et de Responsabilité – Données clés 2009 » réalisé conjointement par la FFSA et le GEMA, et « L'Assurance Française - Les Données clés 2023 » édité par France Assureurs afin de couvrir la période depuis la date de ce portefeuille, 2010, jusqu'à la rédaction de ce mémoire.

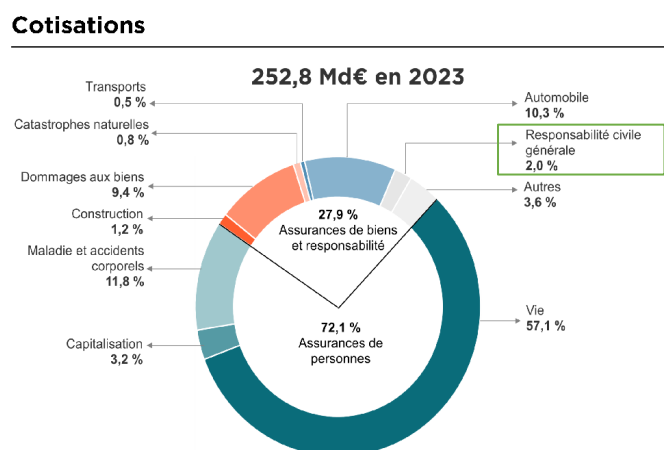


Figure 1: Répartition des cotisations d'assurance en France en 2023 (source France Assureurs)

Le marché de l'assurance en France est passé de 199.9 milliards d'euros de cotisations toute branche confondue en 2009 à 252.8 milliards d'euros en 2023, soit une progression de 26.5% en 14 ans très légèrement supérieure à la progression de l'indice de prix à la consommation (IPC) sur la même période qui a été de 25.9%².

Sur la seule partie de responsabilité civile générale, les cotisations des affaires directes ont évolué de 3.5 milliards d'euros en 2009 à 3.8 milliards d'euros en 2020, soit 9.1% sur la période, inférieur à l'IPC qui progressait alors de 12.3%, pour ensuite atteindre 5 milliards d'euros en 2023, soit une progression de 42.9% depuis 2009 très nettement supérieure à l'IPC cette fois-ci. Cette augmentation tarifaire a donc eu lieu ces dernières années et ce pour des raisons multiples. L'inflation est en grande partie responsable de cette situation, car combinée aux caractéristiques de développement long de la garantie RC générale, cela implique un décalage entre le rendement des actifs adossés aux provisions pour sinistre à payer investis sur une anticipation d'inflation modeste par rapport à la réelle inflation des coûts des sinistres survenues depuis 2021. Ce n'est donc pas uniquement la tarification des futurs engagements qui sont à adapter, mais il faut aussi intégrer l'impact négatif sur les paiements des engagements déjà pris par les assureurs. Et il ne faut pas oublier le marché de la réassurance qui est international, qui fournit les assureurs, et qui a durci ces conditions depuis 2021, dû à cette inflation, mais aussi à des développements défavorables en sinistralité, et pas uniquement RC. La combinaison de ces facteurs a donc obligé les assureurs à revoir leurs politiques tarifaires pour répercuter cette situation, mais aussi à limiter leurs expositions

² Source INSEE

par des restrictions de garanties ou de capacités, accentuant la dégradation tarifaire pour les assurés, jusqu'à être bien supérieure à l'inflation exprimée à travers l'IPC.

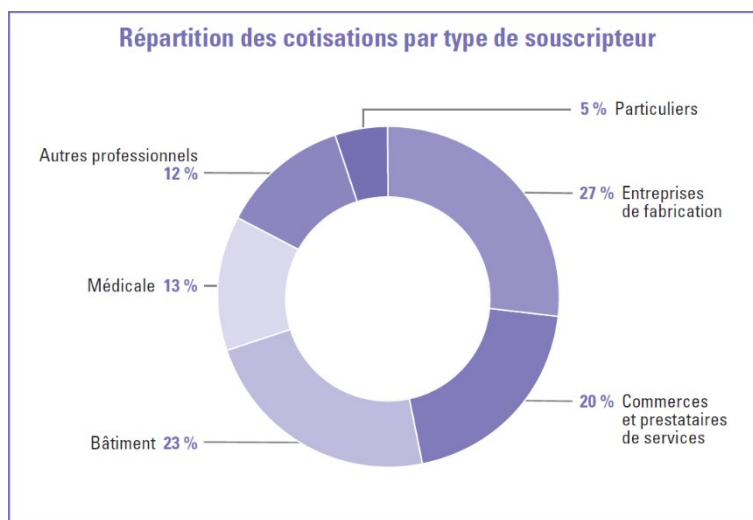


Figure 2: Répartition des cotisations en RC générale par type de souscripteur en 2009 (source France Assureurs)

Un dernier point à considérer est que l'assurance RCG est une couverture essentiellement de professionnels avec 5% seulement de cotisations de particuliers, proportion restée stable depuis 2009.

Le portefeuille sur lequel nous avons réalisé cette étude a comme cœur de cible la couverture RCG des entreprises de fabrication et de transformation industriels et pharmaceutiques, et n'inclut pas de couvertures obligatoires RC professionnelles. Les entreprises des autres secteurs d'activités sont présentes dans ce portefeuille dans de plus faibles proportions en termes de volume de primes. Il n'y a pas de particuliers dans le portefeuille considéré.

2.2 Enjeux du provisionnement des sinistres

Nécessité du provisionnement des sinistres

L'activité d'assurance est en cycle inversé par rapport aux autres activités économiques. D'un point de vue schématique, une entreprise vend un produit ou une prestation en connaissant son coût net au moment de la transaction, et par conséquent son résultat financier. Dans le cas de l'assureur, celui-ci vend une promesse de prestation dont il ne connaît pas le montant final au moment de la transaction. Le montant de la transaction, c'est-à-dire la prime, est estimé par l'assureur sur la base des probabilités des prestations qu'il devra assumer, en y incluant l'ensemble des frais accessoires pour réaliser ces prestations. De son côté le client attend le versement des prestations en cas de survenance d'événements tels que définis au contrat. Il se dessine donc une relation de confiance entre les deux parties qui impose à l'assureur d'être capable de faire face à ses engagements à tout instant. Ainsi il doit les estimer correctement, et les mettre de côté, ce sont les provisions techniques des sinistres, en plaçant la contrepartie financière sur des supports suffisamment sûrs et liquides pour répondre en temps et en volume à ses obligations.

Les provisions techniques des sinistres peuvent avoir deux modes de gestion : par répartition ou par capitalisation. En voici les définitions disponibles sur le site de la FFSA³ :

- Les assurances gérées en répartition : La gestion par répartition consiste, pour l'assureur, à redistribuer aux victimes des sinistres la masse des cotisations payées par l'ensemble des assurés. Cette répartition s'opère par année : les cotisations de l'année paient les sinistres de l'année. Les assurances de biens et de responsabilité ainsi que certaines assurances de personnes comme les assurances complémentaires santé et dommages corporels sont gérées en répartition.
- Les assurances gérées en capitalisation : La capitalisation est une technique de gestion financière : l'assureur place une partie des primes collectées pour les faire fructifier et réinvestit les revenus financiers ainsi obtenus de manière à accroître la somme initiale. L'assureur constitue de cette manière un capital qui lui servira à payer, au moment voulu, la prestation prévue dans le contrat. La capitalisation concerne les assurances qui présentent les particularités suivantes :
 - Elles sont souscrites pour une longue période et comportent un aspect d'épargne ;
 - Elles portent sur des risques non constants dont la fréquence augmente ou diminue au cours du contrat. C'est le cas de la probabilité de décès ou de survie de la personne humaine.

Les principales branches gérées en capitalisation sont l'assurance vie, les contrats de capitalisation ou encore la prévoyance collective.

³ http://www.ffsa.fr/ffsa/jcms/c_24059/les-activites-dassurance?cc=fn_7345#corps3

Chronologie schématique des sinistres non-vie dans les comptes de l'assureur

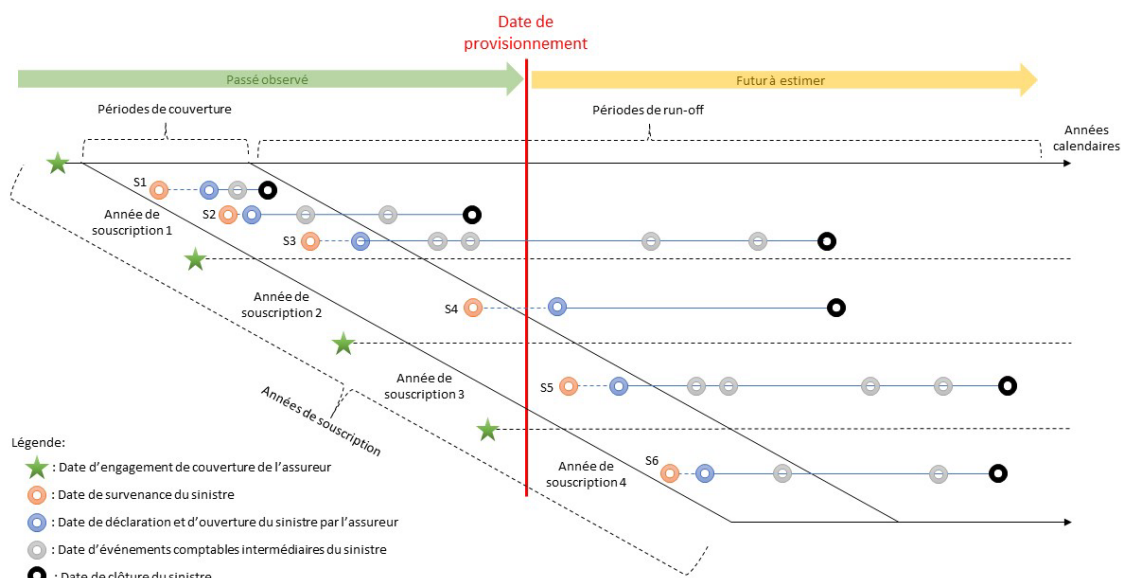


Figure 3: Chronologie schématique des sinistres d'un contrat d'assurance non-vie

Afin de clarifier l'ensemble des engagements de l'assureur, la Figure 3 propose une représentation schématique simplifiée de la chronologie des sinistres d'un contrat d'assurance non-vie sur 4 années de souscriptions consécutives reprenant les différents cas de figures tels qu'ils peuvent être observés. Un sinistre S_i initie son parcours chronologique par un événement, ou fait générateur, considéré comme la survenance du sinistre dû à un tiers, personne physique ou morale. Puis après un certain délai, les faits sont révélés à l'assuré et le préjudice potentiel est évalué comme susceptible de déclencher la couverture du contrat d'assurance qui motive l'information de cette situation à l'assureur et qui correspond à la date de déclaration. Dans les données utilisées, les dates de déclarations des sinistres et d'ouvertures des sinistres sont bien souvent confondues ou très proches, ce qui justifie de les considérer comme identique pour cette étude. Ceci est lié à la politique de provisionnement des sinistres de la compagnie qui procède à l'ouverture comptable des sinistres dès leurs connaissances, et ce au moins avec une provision forfaitaire si aucun montant de préjudice ne peut être encore estimé. Nous verrons plus loin que cet aspect d'ouverture forfaitaire devra être pris en compte dans le modèle détaillé. Puis, une fois les responsabilités clairement établies, le préjudice de l'assuré déterminé, ainsi que son montant financier évalué correctement, l'assureur a alors la capacité de clôturer le dossier. Comptablement cela signifie qu'à cette date la provision attachée à ce dossier est close, c'est-à-dire supprimée. Entre les dates d'ouvertures et de clôtures des sinistres, différents événements comptables intermédiaires sont susceptibles de survenir au fil des informations collectées par l'assureur. Ainsi à chacune de ces dates d'événements comptables, qu'elles soient pour l'ouverture du sinistre, la clôture du sinistre ou bien intermédiaires, plusieurs transactions comptables peuvent avoir lieu simultanément : des paiements ou bien des mouvements de provisions qui peuvent concernés aussi bien l'indemnisation des sinistres que les frais liés à la gestion des sinistres, tels des honoraires d'avocats, des frais d'experts externes ou encore des frais de déplacement de collaborateurs de l'assureur.

Concernant les cas de réouvertures de sinistres après une première ou plusieurs clôtures : même si elles sont possibles en principe, en pratique très peu de cas de réouverture de dossiers sinistre ont été observés dans la données analysées, là aussi dû à la politique de provisionnement de l'assureur considéré. C'est pourquoi nous n'en tiendrons pas compte dans cette étude, tout en étant conscient

que cet aspect peut avoir une importance matérielle dans d'autres configurations de risques et de politiques de provisionnement.

Ainsi, à la date de provisionnement, les sinistres peuvent être déjà clos (S1 et S2), ou bien ouverts mais non clos (S3), ou biens survenus mais non déclarés (S4), ou même non encore survenus mais concernant une période de couverture en cours (S5) ou une période de couverture non débutée mais sur laquelle l'assureur est déjà légalement engagé (S6), soit par la signature du contrat d'assurance antérieur à la date de provisionnement, soit par la non activation de la clause de résiliation à la date prévue au contrat dans le cadre de contrats à tacite reconduction. Dans cet exemple, l'engagement de l'assureur est la somme des flux comptables futurs, c'est à dire postérieurs à la date de provisionnement sur la Figure 3.

Ainsi l'estimation des engagements de l'assureur consiste à faire ce recensement sur l'ensemble des contrats de son portefeuille en y appliquant des méthodes adaptées aux caractéristiques de ses risques afin d'estimer les futurs flux comptables des contrats auxquels il est engagé. Il est alors possible d'y appliquer les outils de mathématiques financières y afférant (Portait et Poncet, 2008)

Place du provisionnement des sinistres dans le bilan d'un assureur non-vie

Le graphique suivant présente un exemple issu d'un cas réel d'équilibre de bilan d'une compagnie d'assurance non-vie :

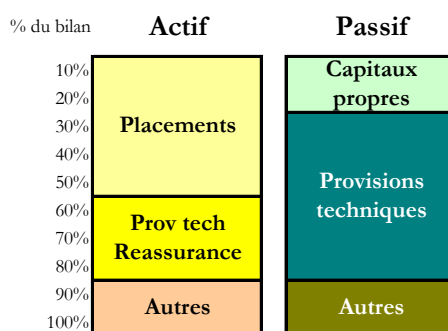


Figure 4: Structure schématique du bilan d'une société d'assurance

Même si ce les ratios ne sont qu'illustratifs bien entendu et dépendent des situations particulières de chaque assureur, il est à noter que les provisions techniques, suivant la norme Solvabilité II, représentent entre 60% et 80% du montant total du bilan, d'après les rapports financiers des principaux assureurs IARD du marché français⁴.

Quantification des provisions pour sinistre à payer

Le provisionnement des sinistres à payer est un exercice pointu pour l'assureur : il s'agit pour lui d'estimer à tout instant, les coûts ultimes des sinistres, y compris sur les sinistres survenus et déclarés mais insuffisamment provisionnés (IBNER) ou sur les sinistres survenus mais non encore déclarés (IBNYR).

⁴ Rapports financiers d'AXA France IARD 2023, de Groupama 2023, de MMA IARD 2023 et Allianz IARD 2022.

De nombreux facteurs vont rendre l'évaluation difficile :

- Le temps de manifestation du sinistre
- La chaîne d'information entre les différents acteurs (assuré, courtier, assureur, expert...)
- L'interprétation des garanties tout au long de la chaîne précédemment citée
- La survenance d'un éventuel contentieux sur le point précédent, accompagné d'un règlement amiable ou judiciaire d'un certain montant et au bout d'un processus d'une certaine durée
- L'impact de la réassurance en vigueur

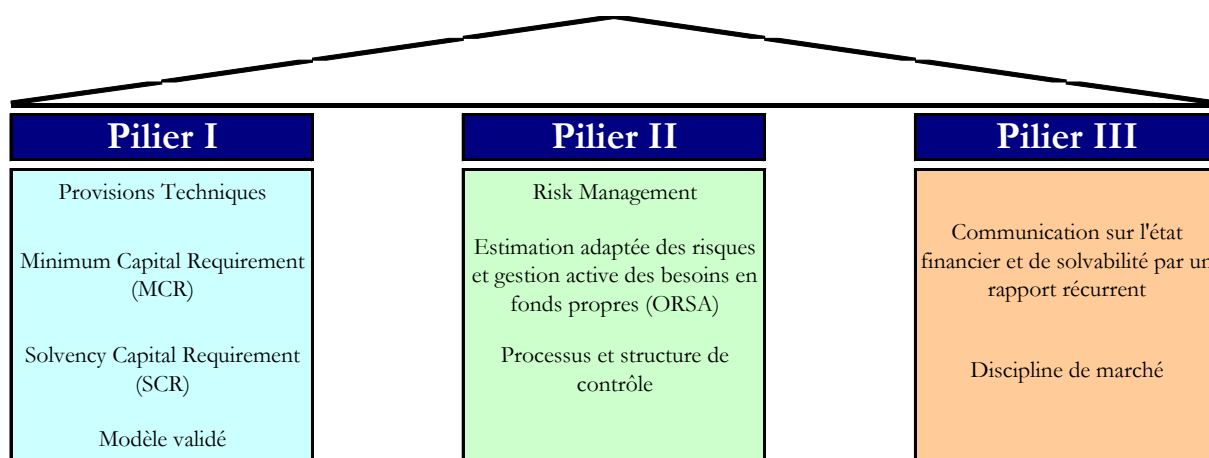
Contrairement à l'assurance des risques de masses, l'assurance des risques industriels s'expose à des sinistres souvent complexes du fait des caractéristiques de cessions utilisés et des clauses contractuels compliquées qui rendent les évaluations délicates. Ceci est donc à cerner correctement pour apparaître le plus précisément possible au bilan de l'assureur, mais aussi dans le cadre du processus d'estimation des capitaux requis de Solvabilité II.

2.3 Solvabilité II et provisionnement des sinistres

2.3.1 Introduction à la directive Solvabilité II⁵

Solvabilité II est une directive de l'union européenne à destination des assureurs et réassureurs de l'union européenne. Elle a été votée le 22 avril 2009 au parlement européen, transposée dans les droits des états adhérents de l'union européenne, puis est entrée en application le 01/01/2016.

Cette directive repose sur trois piliers conçus pour assurer un processus d'Enterprise Risk Management (ERM) étendu qui combine des objectifs quantitatifs, qualitatifs, de contrôles, de règles de gouvernance des risques, ainsi que d'exigences de communications :



⁵ https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii_en

Figure 5: les trois piliers de Solvabilité II

- Pilier I : C'est la partie quantitative de la directive qui définit les seuils nécessaires à l'évaluation des provisions techniques et des fonds propres nécessaires. Schématiquement, le bilan d'une société d'assurance se transforme ainsi :

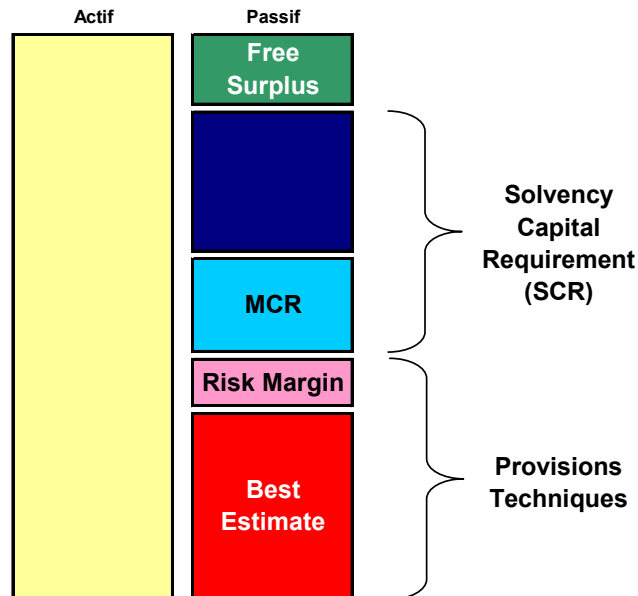


Figure 6: ventilation structurelle des principales composantes du bilan d'un assureur dans Solvabilité II

Le Best Estimate, principale composante des provisions techniques, est défini comme la moyenne pondérée des flux futurs actualisés découlant des engagements de l'assureur à la date d'analyse. Ce calcul doit être réalisé à partir d'hypothèses réalistes, sans biais, et sur des bases prospectives. Il inclut l'escompte de flux futurs de trésorerie à un taux sans risque.

Le « Risk Margin » est la deuxième composante de la provision technique. Il s'agit de risques non cessibles qui concernent le coût du transfert du portefeuille à un assureur tiers. L'idée est d'ajouter au Best Estimate le coût d'immobilisation en capital nécessaire à un tiers reprenant les provisions pour supporter le risque jusqu'à l'extinction complète de ces provisions, et donc des engagements sous-jacents.

Le MCR

(Minimum Capital Requirement) est la marge de solvabilité minimum dont doit disposer la société d'assurance pour exercer son activité. Il s'agit donc d'un seuil d'intervention du régulateur qui alors peut décider d'une mise sous tutelle ou même supprimer l'agrément de l'assureur provoquant sa cessation d'activité. Sa règle d'estimation est définie de manière unique par le régulateur et sur des bases de calculs simples.

Le SCR (Solvency Capital Requirement) est le capital destiné à absorber une quantité significative de sinistres imprévus. Une diminution sensible pourra faire l'objet d'un avertissement du régulateur. Son calcul est basé sur une approche stochastique avec un seuil de probabilité de ruine de 0.5% à horizon 1 an. La méthode d'estimation peut être basée sur un modèle dit « interne » propre à chaque assureur et qui doit être validé au préalable par le régulateur, ou bien sur un modèle dit « standard » défini par le régulateur dont les éléments clés liés au provisionnement seront décrits ci-dessous page 26. Ce modèle standard repose sur un système modulaire et hiérarchique de différents types de risques qu'il convient

de modéliser. Les calculs se font via des scénarios stress et des facteurs d'approximation. La structure du SCR peut être vue ainsi, en se limitant à l'assurance non-vie :

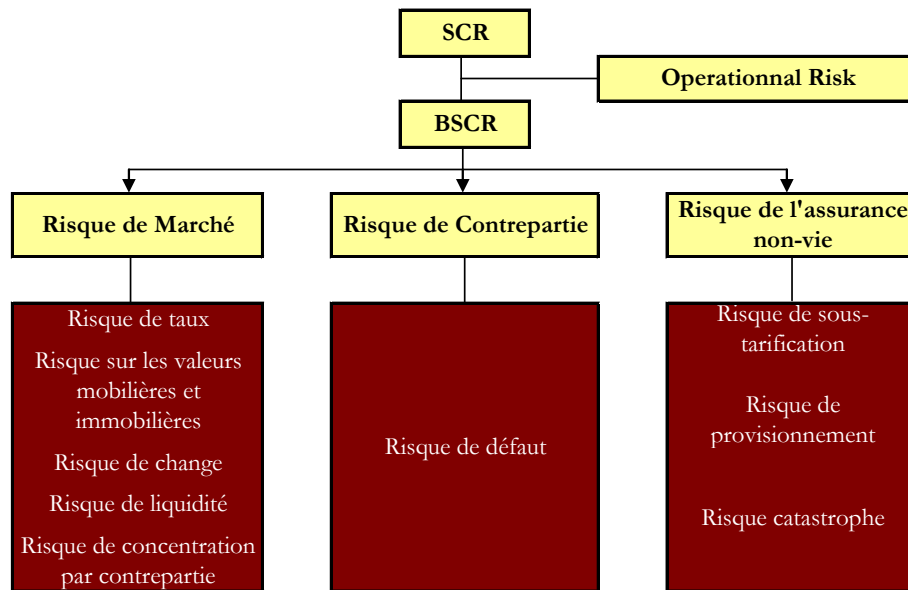


Figure 7: principaux modules du SCR non-vie

La validation du modèle interne, réalisée par le régulateur local, repose sur une série de critères incluant des tests statistiques, des tests de calibrage, des tests d'utilisation, une documentation... Même si le modèle interne est validé, la formule standard doit être utilisée en parallèle pendant une période de deux ans.

- Pilier II : Cette partie définit l'ensemble des processus et des critères de contrôles internes afin d'encadrer et de fiabiliser les calculs réalisés sur le Pilier I. Cela peut se décomposer en cinq parties :
 - Encadrement critique sur la gouvernance, les éléments du bilan, le risk management, l'audit interne, et les modèles utilisés.
 - Estimation d'un capital supplémentaire nécessaire en cas de faiblesse des calculs du modèle ou de l'inadéquation de la gouvernance ou du risk management.
 - Système de gouvernance robuste et documenté.
 - Politique de risk management afin d'identifier, de mesurer, de réduire le risque et d'en fournir des comptes-rendus.
 - Obligation d'avoir des fonctions d'audit interne.

Tous ces éléments ont vocations à s'assurer de l'existence d'une forte culture du risque au sein des sociétés d'assurance et d'introduire ces notions au centre de leur gouvernance.

- Pilier III : Ce pilier concerne les informations à communiquer au régulateur et celles à rendre publiques. Le rapport financier de solvabilité à rendre public doit inclure les informations suivantes :
 - Une vue d'ensemble de l'activité de l'entreprise

- La structure de gouvernance, et particulièrement celle destinée aux risques.
- Hypothèses et méthodes d'estimation des actifs et des provisions techniques.
- Le processus de risk management
- Les capitaux de solvabilité, MCR et SCR, incluant si nécessaire les capitaux requis supplémentaires, et les capitaux disponibles.
- En cas d'utilisation d'un modèle interne, les informations méthodologiques, d'hypothèses, et de validation de celui-ci.

Ceci permet de rendre transparents les processus d'ERM aux yeux du marché et d'en faciliter la compréhension. Ce qui a pour objectif ultime de générer la confiance sur la gouvernance de l'entreprise et de fiabiliser l'interprétation de ses résultats financiers.

2.3.2 Principaux montants de Solvabilité II liés au provisionnement des sinistres

L'intégralité des instructions et documents sont disponibles sur le site de l'« European Insurance and Occupational Pensions Authority » (EIOPA)⁶. Voici les principaux éléments du pilier I de la directive en lien avec le provisionnement des sinistres. Nous restreindrons le périmètre de cette partie aux informations concernant la branche de responsabilité civile générale à laquelle se rattache notre étude.

De plus nous négligerons la réassurance afin de nous concentrer sur la zone d'influence directe des modèles d'estimations de provisionnement du portefeuille sur l'acceptation des risques. Ainsi, nous considérons la réassurance comme étant exogène aux caractéristiques des risques en acceptation. Ceci n'est pas rigoureusement exact puisque les traités de réassurance obligatoire sont pris en compte dans le processus de souscription soit pour savoir si les risques peuvent être inclus dans le traité de réassurance en cours, avec un complément possible par une réassurance facultative, soit en déclinant la souscription du risque.

L'objectif sera de comparer le capital de solvabilité requis calculé en appliquant la formule standard avec la VaR (« value at risk ») au percentile 99,5% et à horizon 1 an issues des modèles stochastiques présentés dans l'étude en se focalisant sur l'impact des scénarios extrêmes des provisions pour sinistres impactant les calculs du SCR et du MCR. En premier lieu ces provisions sont utilisées pour calculer les provisions techniques de Solvabilité II constituées de deux composants : Le Best Estimate et le Risk Margin.

Le Best Estimate

L'estimation du Best Estimate, première composante des provisions techniques, consiste en l'évaluation de la moyenne pondérée des flux futurs de trésorerie des composants suivants (voir schéma Figure 8) : Sinistres à payer, frais de gestion des sinistres (direct et indirect), autres frais, et déduit des primes futurs recevables au titre des engagements en cours. Ces valeurs sont regardées

⁶ https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii_en

sous deux dimensions complémentaires : expositions passées (Provisions pour sinistres ou « Claims Provision » en anglais) et exposition futures (Provisions pour primes « Premium Provision » en anglais). Enfin les différents montants de flux futurs sont actualisés avec les courbes de taux sans risques pour en estimer leurs valeurs actuelles.

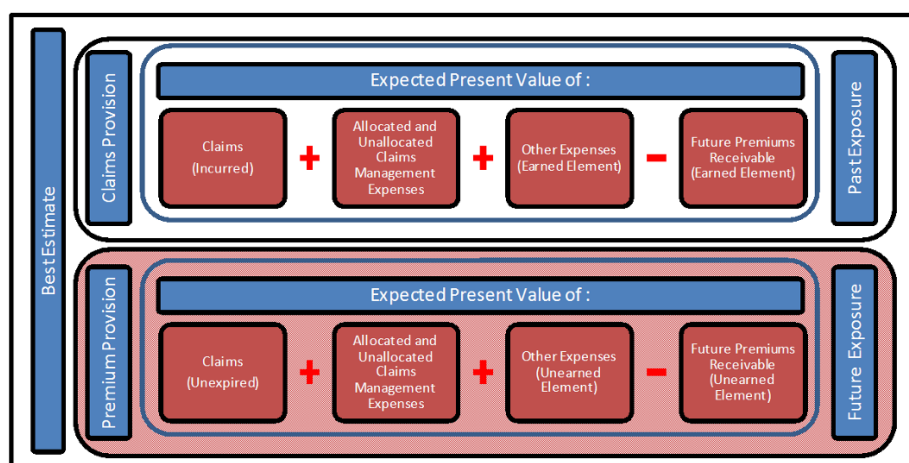


Figure 8: Composants du Best Estimate⁷

Les instructions données préconisent d'utiliser une méthode actuarielle et statistique qui reflète le plus fidèlement possible les flux de trésorerie futurs du portefeuille analysé. Ainsi pour la partie sinistre, indemnitaire et frais direct de gestion de sinistre, nous utiliserons successivement les méthodes traditionnelles basées sur les données agrégées : Chain-Ladder, Mack et GLM avec bootstrap. Puis un modèle basé sur les données détaillées sera analysé. Toutes ces méthodes sont explicitées dans la suite de cette étude. Le Best Estimate ne nécessite pas pour autant l'utilisation d'un modèle stochastique puisque seule l'espérance pondérée des flux futurs est nécessaire, mais seront utiles dans la partie SCR dans le cadre d'un modèle interne.

Les autres frais, qui concernent les frais généraux et les frais de gestion des actifs adossés aux provisions pour sinistres, ainsi que les primes futures recevables, qui concernent les primes non perçues nettes de risques de défauts concernant les engagements du portefeuille, ne seront pas abordés dans cette étude.

Risk Margin⁸ :

Le Risk Margin, deuxième composante des provisions techniques, représente l'estimation du coût de transfert à un autre assureur de l'ensemble des engagements d'assurance du portefeuille considéré en tenant compte de la diversification de l'ensemble du portefeuille. En voici la formule générale :

$$RM = CoC. \sum_{t \geq 0} \frac{SCR(t)}{(1 + r(t + 1))^{t+1}}$$

⁷ Source « Solvency II Technical Provisions for General Insurers » publié par l'Institut and Faculty of Actuaries.

⁸ Journal officiel de l'Union européenne du 17 janvier 2015 « RÈGLEMENT DÉLÉGUÉ (UE) 2015/35 DE LA COMMISSION du 10 octobre 2014 complétant la directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II) » - Chapitre III

Où :

RM est le risk margin, CoC le taux de coût de capital, $SCR(t)$ est le capital de solvabilité requis après t années, et $r(t + 1)$ le taux sans risque de maturité $t + 1$ années.

Il est possible de faire un calcul simplifié, sous réserve de validation par le régulateur, avec l'approximation de la valeur $SCR(t)$ si elle n'est pas aisément disponible, par exemple, en la considérant comme proportionnelle au Best Estimate à partir de la formule suivante :

$$SCR(t) = \frac{SCR(0)}{BE_{net}(0)} BE_{net}(t) , \text{ avec } BE_{net}(t) \text{ le Best Estimate net de réassurance après } t \text{ années.}$$

Cette approche a le mérite de conserver la structure de liquidation des sinistres dans le temps (« claims run-off pattern » en anglais), mais implique aussi que la structure de risque du portefeuille reste inchangée dans le temps, ce qui ne peut être valide lorsque le portefeuille est composé de risques à maturités différentes.

Le Risk Margin est mentionné dans cette étude, uniquement ici, pour donner une vision exhaustive des provisions techniques sous Solvabilité II, et aussi montré son potentiel lien direct avec les provisions pour sinistres à travers l'utilisation du Best Estimate en cas de simplification pour son estimation.

SCR – formule standard :

Le SCR (voir Figure 7) a pour formule générale dans la formule standard :

$$SCR = BSCR + Adj + SCR_{op}$$

Où :

- $BSCR$ (Basic Solvency Requirement) représente le capital requis de base.
- Adj représente l'ajustement lié à la capacité d'absorption des pertes par les provisions techniques et les impôts différés.
- SCR_{op} représente le risque opérationnel qui désigne le risque de perte résultant de procédures internes, de membres du personnel ou de systèmes inadéquats ou défaillants, ou d'événements extérieurs.

Le $BSCR$ est calculé à partir de la formule suivante, des éléments i et j de la liste d'agrégation des risques et de leurs corrélations $Corr_{i,j}$:

$$BSCR = \sqrt{\sum_{i,j} Corr_{i,j} * SCR_i * SCR_j} + SCR_{incorporels}$$

Ainsi l'estimation du $BSCR$ est la combinaison des différents modules de risque de marché, de contrepartie, de souscription, ainsi que de leurs sous-modules de risque, plus le risque lié aux immobilisations incorporelles.

- Le risque de marché reflète le risque lié à la volatilité de la valeur de marché des instruments financiers ayant un impact sur la valeur des actifs et des passifs de

l'entreprise concernée. Il reflète de manière adéquate toute inadéquation structurelle entre les actifs et les passifs, en particulier au regard de leur duration.

- Le risque de contrepartie : mesure le risque de défaut des tiers ayant un engagement envers la société. Les principales entreprises concernées ici sont les réassurances participant aux différentes cessions de l'assureur, ainsi que les émetteurs d'obligations auxquelles a souscrit l'assureur.
- Le risque de souscription désigne le risque de perte ou de changement défavorable de la valeur des engagements d'assurance, en raison d'hypothèses inadéquates en matière de tarification et de provisionnement. Pour le risque de souscription non-vie, il se ventile en trois sous-modules :
 - a. le sous-module «risque de primes et de réserve en non-vie» qui mesure :
 - le risque de sous-tarification des contrats qui se traduit par une insuffisance des provisions pour primes constituées pour couvrir les sinistres à venir, en dehors de la réalisation d'événement de type catastrophique.
 - Le risque de sous-provisionnement qui se traduit par une insuffisance des provisions pour sinistres constituées pour couvrir les sinistres déjà survenus.
 - b. Le sous-module « risque catastrophe » est le risque de perte, ou de changement défavorable de la valeur des engagements d'assurance, résultant de l'incertitude importante, liée aux événements extrêmes ou exceptionnels, qui pèse sur les hypothèses retenues en matière de prix et de provisionnement. 4 sous-modules le compose :
 - Le sous-module « risque de catastrophe naturelle
 - Le sous-module « risque de catastrophe en réassurance dommages non proportionnelle »
 - Le sous-module « risque de catastrophe d'origine humaine »
 - Le sous-module « autres risques de catastrophe en non-vie »
 - c. Le sous-module « risque de cessation en non-vie » est le risque de perte, ou de changement défavorable de la valeur des engagements d'assurance, résultant de fluctuations affectant le niveau ou la volatilité des taux de cessation, d'échéance, de renouvellement et de rachat des polices, et résultant de la conjugaison des événements soudains suivants :
 - La cessation de 40% des contrats d'assurance dans le cas desquels cette cessation a pour effet d'entraîner une augmentation des provisions techniques sans la marge de risque
 - Lorsque des contrats de réassurance couvrent des contrats d'assurance ou de réassurance qui seront émis à l'avenir, la baisse de

40% du nombre de ces futurs contrats d'assurance ou de réassurance utilisé dans le calcul des provisions technique.

Ainsi pour un assureur non-vie, l'exigence de capital $SCR_{non-life}$ se calcul ainsi :

$$SCR_{non-life} = \sqrt{\sum_{i,j} CorrNL_{i,j} * SCR_i * SCR_j}$$

La somme couvre toutes les combinaisons possibles (i, j) des sous-modules risque de primes et de réserve en non-vie, risque de catastrophe en non-vie, et risque de cessation en non-vie, ainsi que leurs corrélations $CorrNL_{i,j}$.

Dans le cadre de cette étude, seul le sous-module « risque de primes et de réserve en non-vie » : $SCR_{nl prem res}$ sera calculé. Le risque catastrophe fera l'objet d'un commentaire au regard des données disponibles et des résultats.

Tout d'abord, voici la formule générale d'estimation du $SCR_{nl prem res}$:

$$SCR_{nl prem res} = 3 * \sigma_{nl} * V_{nl}$$

Où :

- σ_{nl} représente l'écart-type du risque de primes et de réserve en non-vie
- V_{nl} représente la mesure de volume pour le risque de primes et de réserve en non-vie qui est égale à la somme des mesures du volume pour le risque de primes et de réserve des différents segments non-vie⁹.

La formule standard d'estimation de σ_{nl} est :

$$\sigma_{nl} = \frac{1}{V_{nl}} \sqrt{\sum_{s,t} CorrS_{(s,t)} * \sigma_s * V_s * \sigma_t * V_t}$$

Où :

- $CorrS_{(s,t)}$ représente le coefficient de corrélation pour le risque de primes et de réserve en non-vie du segment s et du segment t.
- σ_s et σ_t représentent les écarts types du risque de primes et de réserve en non-vie des segments s et t.
- V_s et V_t représentent les mesures du volume pour le risque de primes et de réserve des segments s et t.

Et où la valeur σ_s de chaque segment s est calculé par la formule :

$$\sigma_s = \frac{\sqrt{\sigma_{(prem,s)}^2 * V_{(prem,s)}^2 + \sigma_{(prem,s)} * V_{(prem,s)} * \sigma_{(res,s)} * V_{(res,s)} + \sigma_{(res,s)}^2 * V_{(res,s)}^2}}{V_{(prem,s)} + V_{(res,s)}}$$

Les valeurs données pour le risque de responsabilité générale (GTPL pour « General Third Party Liability ») $\sigma_{(prem,s)}$ $\sigma_{(res,s)}$ et sont⁹ :

- $\sigma_{(prem,GTPL)} = 14\%$
- $\sigma_{(res,GTPL)} = 11\%$

Et

- $V_{(prem,s)}$ représente la mesure de volume pour le risque de primes du risque du segment.
- $V_{(res,s)}$ représente la mesure de volume pour le risque de réserve du segment.

Ensuite l'estimation de V_s se fait par la formule suivante :

$$V_s = (V_{(prem,s)} + V_{(res,s)}) * (0,75 + 0,25 * DIV_s)$$

Où DIV_s représente le facteur de diversification géographique du segment s . Il est calculé ainsi, en décomposant les valeurs par région géographique notée r :

$$DIV_s = \frac{\sum_r (V_{(prem,r,s)} + V_{(res,r,s)})^2}{(\sum_r (V_{(prem,r,s)} + V_{(res,r,s)}))^2}$$

Le portefeuille sera considéré comme mono-géographique par prudence dans le calcul, même si des expositions internationales sont présentes dans le portefeuille.

Ainsi : $DIV_{GTPL} = 1$

Puis $V_{(prem,s)}$ est estimé par :

$$V_{(prem,s)} = \max[P_s; P_{(last,s)}] + FP_{(existing,s)} + FP_{(future,s)}$$

Où :

- P_s représente une estimation des primes à acquérir dans le segment s au cours des 12 mois à venir.
- $P_{(last,s)}$ représente les primes acquises dans le segment s au cours des 12 derniers mois.
- $FP_{(existing,s)}$ représente la valeur actuelle attendue des primes à acquérir dans le segment s après les 12 mois à venir pour les contrats existants.
- $FP_{(future,s)}$ représente le montant à comptabiliser pour les contrats dont la date de comptabilisation initiale survient dans les 12 mois à venir, ventilée en 2 parties :

⁹ Journal officiel de l'Union européenne du 17 janvier 2015 « RÈGLEMENT DÉLÉGUÉ (UE) 2015/35 DE LA COMMISSION du 10 octobre 2014 complétant la directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II) » - Annexe II

- Contrats à durée égale ou inférieure à 1an : valeur actuelle attendue des primes à acquérir, mais à l'exclusion des primes à acquérir au cours des 12 mois qui suivent la date de comptabilisation initiale
- Contrats à durée supérieure à 1 an : montant égal à 30% de la valeur actuelle attendue des primes à acquérir après les 12 mois à venir

Il sera considéré que le portefeuille est composé de contrats à durée d'un an sans tacite reconduction.

Le dernier composant à estimer, $V_{(res,s)}$, le volume de risque de réserve pour un segment s , est donné par la meilleure estimation des provisions pour sinistres à payer, qui pourra être obtenu par les différentes méthodes présentées dans ce document.

MCR :

Le « Minimum Capital Requirement » est le résultat d'un calcul basé sur les émissions de primes et les provisions techniques puis encadrer par le SCR et par un seuil plancher absolu dépendant uniquement des types de branches d'assurance concernées et indépendant des niveaux d'activités de celles-ci, appelé *AMCR*, et définie par le régulateur.

Ainsi, nous avons la succession de formules suivante :

$$MCR = \max(MCR_{combined}; AMCR)$$

Où :

- $MCR_{combined} = \min(\max(MCR_{linear}; 0.25 * SCR); 0.45 * SCR)$
- *AMCR* représente le seuil plancher absolu du minimum de capital requis.
 - Comme le risque du portefeuille est de la responsabilité générale, le *AMCR* est égal à 4,0 millions d'euros¹⁰.

Puis le MCR_{linear} se calcule ainsi :

$$MCR_{linear} = MCR_{(linear,nl)} + MCR_{(linear,l)}$$

Où chaque terme représente la formule linéaire pour les engagements vie et non-vie.

Et

$$MCR_{(linear,nl)} = \sum_s \sigma_s * TP_{(nl,s)} + \beta_s * P_s$$

Où :

¹⁰ Arrêté du 21 octobre 2022 fixant dans le code des assurances les montants libellés en euros du seuil absolu de minimum de capital requis (JORF n°0251 du 28 octobre 2022), et Code des Assurances (Articles R352-29 et R321-1).

- $TP_{(nl,s)}$ représente les provisions techniques sans marge de risque dans le segment s après déduction des montants recouvrables au titre des contrats de réassurance et des véhicules de titrisation, avec un plancher de zéro.
- P_s désigne les primes émises pour le segment s au cours des 12 derniers mois, après déduction des primes des contrats de réassurance, avec un plancher de zéro.
- σ_s est le facteur afférent aux provisions techniques pour le segment s .
 - Pour la responsabilité générale¹¹ : $\sigma_{GPL} = 10,3\%$
- β_s est facteur afférent aux primes émises pour le segment s .
 - Pour la responsabilité générale¹¹ : $\beta_{GPL} = 13,1\%$

¹¹ Journal officiel de l'Union européenne du 17 janvier 2015 « RÈGLEMENT DÉLÉGUÉ (UE) 2015/35 DE LA COMMISSION du 10 octobre 2014 complétant la directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II) » - Annexe XIX

3 Présentation des données et outils utilisés

3.1 Structure des sinistres du portefeuille

Les données utilisées pour mener cette étude sont celles d'un portefeuille de contrats de responsabilité civile générale d'une succursale d'assurance spécialisée dans les risques industriels et souscrivant des polices pour des assurés ayant leur siège social en France. Les expositions sont donc très majoritairement françaises, mais il y a également une petite partie des expositions à l'étranger.

Les données utilisées sont directement issues du système d'information de production de la succursale permettant ainsi de travailler sur l'ensemble des caractéristiques disponibles des polices, des sinistres, et sur chacun des mouvements comptables des sinistres. La situation comptable au 31/12/2010 est utilisée pour l'ensemble des sinistres déclarés depuis le 01/01/1992, et l'ensemble des polices ayant eu au moins un jour de couverture entre le 01/01/1992 et le 31/12/2010.

Caractéristiques du portefeuille

Les contrats d'assurance en responsabilité civile des entreprises industrielles ont des caractéristiques particulières que l'on retrouve dans le portefeuille étudié :

- Les grandes entreprises cèdent leurs risques à travers des montages découpant par tranche d'engagement les règlements des sinistres susceptibles de survenir.
- L'ensemble des contrats possèdent des limites d'engagements y compris ceux concernant les plus petites entreprises.
- Les contrats nécessitant des engagements souvent importants, la coassurance est présente dans une grande majorité des polices. Cependant, l'assureur considéré est en position d'apériteur sur les trois-quarts des contrats composant le portefeuille étudié.
- Les contrats sont généralement annuels avec ou sans reconduction tacite. Il y a également quelques contrats particuliers qui portent sur une activité occasionnelle et limitée dans le temps, comme dans le cadre d'essais cliniques.
- Le portefeuille étudié contient historiquement aussi bien des contrats fonctionnant sur la base du fait dommageable que sur la base de la réclamation.
- L'appellation « type de couverture » évoquée dans le reste du document concerne les différents types d'assurance de responsabilité civile vue précédemment (RC exploitation, RC des employeurs...). Celles-ci sont regroupées en cinq catégories pour les besoins de l'étude.

Caractéristiques des sinistres

Nous proposons ici de montrer les principales caractéristiques de la sinistralité du portefeuille étudié qui ont pilotées nos choix dans les méthodes de provisionnements.

Montants en €	Nombre de sinistres clos	% Charge Ultime Totale
Sans paiements	12,791	0%
0 - 10k	12,372	7%
10k - 100k	2,380	22%
100k - 500k	374	24%
500k - 1m	50	11%
1m - 5m	25	15%
5m - 10m	3	6%
Plus de 10m	1	14%
Total	27,996	100.0%

Tableau 1: Profil de la charge sinistre de la branche RC Générale - années de survenance 1992-2009

Le tableau ci-dessus présente le profil des paiements des sinistres clos pour les années de survenances comprises entre 1992 et 2009. Les sinistres sont ventilés par tranche de coût final à la clôture des sinistres incluant les paiements indemnitaires et les paiements pour frais.

Un point important qui ressort de ces chiffres est le poids des sinistres atypiques dans la sinistralité totale. A partir des quatre plus hautes tranches, nous voyons que les 79 sinistres les plus importants, soit 0.3% des dossiers sinistres clos, représentent plus de 45% de la charge à l'ultime sur la période considérée. En ajoutant la tranche précédente, les proportions passent à 1,6% des dossiers sinistres clos générant 70% de la charge à l'ultime.

Remarque : la base sinistre synthétisée dans le tableau précédent représente 383,000 lignes de transactions dans la base de production, où chaque enregistrement représente un paiement ou une modification de provisions sur sa partie indemnitaire ou sa partie pour frais.

A partir des informations sinistres disponibles, une segmentation du portefeuille à partir de différentes variables qualitatives a été possible. Celle-ci se révélera adapter dans la suite de l'étude.

La première d'entre elles est le type de garantie générant la prestation de l'assureur. Elle est ventilée en 5 classes :

- RC exploitation
- RC produits
- RC des employés : cette couverture couvre spécifiquement les employés d'une entreprise contre les conséquences de l'activité de l'entreprise. En France, elle fait souvent référence à la faute inexcusable. Dans la plupart des pays anglo-saxons, cette couverture fait l'objet d'une police spécifique souscrite par l'employeur.
- RC activité médicale : cette classe regroupe principalement deux activités industrielles du domaine médical voisines que sont, d'une part, les contrats principaux de responsabilité civile générale des entreprises de l'industrie pharmaceutique, et d'autres parts, les essais cliniques. Ceci n'est pas à confondre avec la couverture de responsabilité professionnelle obligatoire des professionnels de santé exerçant à titre libéral ou des établissements de santé, dite communément RC Médicale, qui n'existe pas dans le portefeuille analysé.

- Générale : Cette valeur sert de valeur par défaut en cas d'information indisponible dans la base de gestion. Elle est également systématiquement utilisée pour les sinistres antérieurs à 2000, la variable type de couverture étant seulement utilisée de manière consistante à partir de cette année-là.

Le type de dommage consécutif est également une variable qualitative significative. Elle est ventilée en 4 classes différentes :

- Dommages matériels
- Pertes financières
- Dommages corporels
- Générale : comme pour le type de couverture, cette valeur sert de valeur par défaut en cas d'information indisponible dans la base de gestion. Elle est également systématiquement utilisée pour les sinistres antérieurs à 2000, la variable type de dommage étant seulement utilisée de manière consistante à partir de cette année-là.

Le positionnement de l'assureur sur le contrat d'assurance s'est révélé être une variable significativement influente sur plusieurs régressions importantes de l'étude. Celle-ci est une variable qualitative à 2 classes qui sépare les sinistres issus des contrats où l'assureur est en position d'apériteur de ceux où il se trouve en position de simple coassureur.

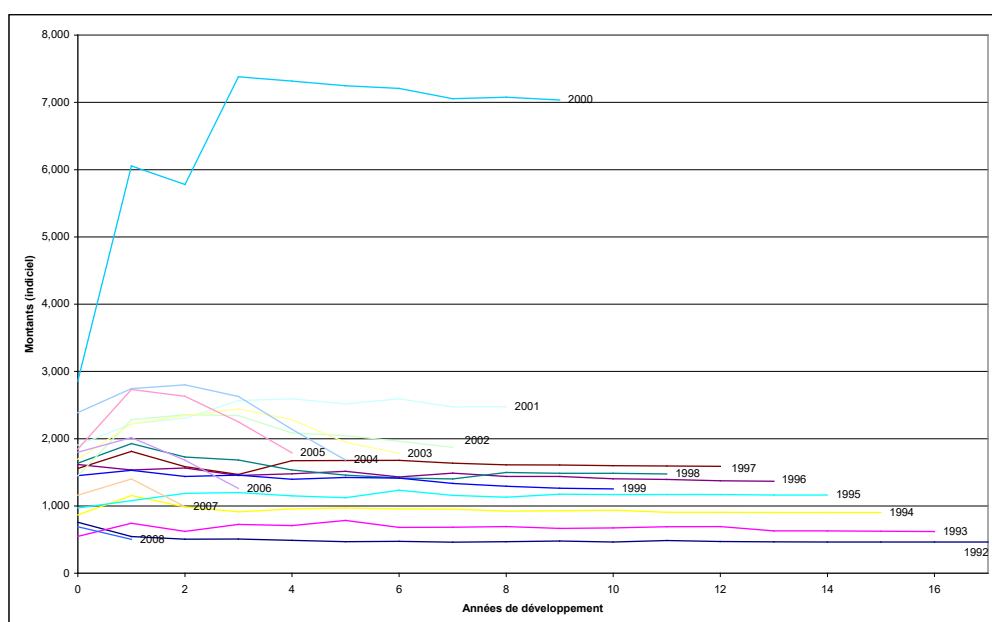


Figure 9: Développement des sinistres clos en apériton au 31/12/2009 par année de survenance

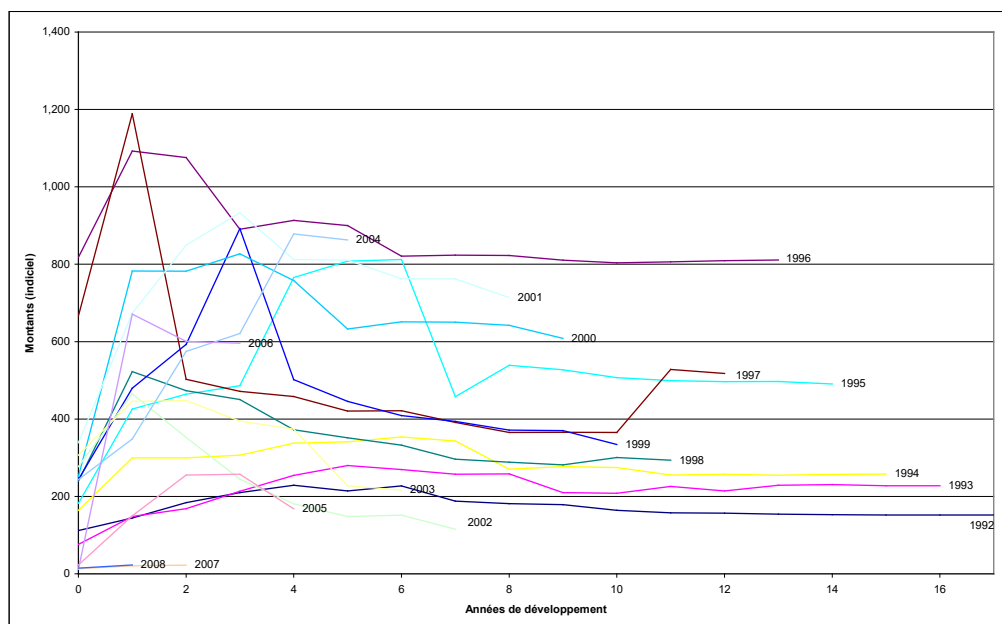


Figure 10: Développement des sinistres clos en coassurance au 31/12/2009 par année de survenance

Les deux graphiques précédents montrent les développements de la charge totale des sinistres clos au 31/12/2009 par année de survenance. La différence de stabilisation dans le temps des charges totales respectives montre une nette différence entre ces deux catégories. Ceci s'explique principalement par le fait que dans le cas d'une apériton, l'assureur a une implication directe avec les différents intervenants de la gestion du sinistre : assuré, expert, avocat... De plus il est également systématiquement au courant de la moindre évolution du dossier. Dans le cadre de la coassurance, la vision de l'assureur peut se révéler plus diffuse : l'assureur en position de coassurance est tributaire des informations fournies et de la gestion réalisée par l'apériteur. Et d'autres critères peuvent venir parasiter cette vision en coassurance : la politique de souscription sur certaines classes d'activité privilégiées en coassurance, et le positionnement en XS coassurance pour sur le haut des programmes d'assurance peuvent être une également explication.

La variable qualitative concernant l'activité principale de l'assuré utilisée est issue de la base de données de l'assureur reposant sur une nomenclature très détaillée avec 5 niveaux hiérarchiques de précision. Pour rendre cette variable plus significative, un traitement par regroupement a été effectué de façon à utiliser l'information la plus précise possible en fonction du nombre de sinistres observés par catégories et sous-catégories. Cependant il convient d'avoir en tête sa faiblesse quant à son utilisation, car beaucoup d'assurés industriels exercent plusieurs activités. Cela pose deux problèmes :

- La ventilation des diverses activités des assurés, qui sont rarement mono-activités pour les entreprises les plus importantes, ne peut être prise en compte
- De plus cette information est fortement statique par assuré dans la base de données de l'assureur, car rarement revue après sa création pour être mise à jour.

3.2 Indexation des variables quantitatives

L'analyse des observations sur une longue période implique de prendre en compte l'impact de l'évolution des conditions économiques dans lesquelles elles ont eu lieu dans le but de pouvoir les

comparer dans des conditions homogènes. Dans le domaine de la responsabilité civile des risques industriels, choisir un index peut être délicat à la vue de l'hétérogénéité des expositions des différents assurés. Par simplification, l'indice IPC a été utilisé pour indexer les flux passés en équivalent euro au 31/12/2010.

Il convient de noter que le régleur sinistre a pour mission d'anticiper les paiements futurs, indemnitaires ou pour frais, par la création d'une provision dédiée à chaque sinistre, dite provision « dossier », en fonction des informations qui lui sont connus et de son expérience. À charge de l'actuaire ensuite, dans le cadre de son calcul de provisions pour sinistres, de vérifier et d'intégrer le cas échéant l'effet de l'inflation sur les provisions techniques. Mais par simplification ici les provisions pour sinistres à payer au 31/12/2010 seront utilisées sans retraitement.

A l'opposé, les flux de paiements futurs modélisés seront actualisés en valeur au 31/12/2010 par l'utilisation d'une courbe de taux sans risque. Nous utiliserons celle disponible mensuellement sur le site internet de l'Institut des Actuaire.

Deux autres effets méritent d'être mentionnés. Tout d'abord l'évolution de la matière assurable. L'évolution de l'activité industrielle sur le long terme n'est pas négligeable. Nous pouvons prendre comme exemple le secteur informatique : au début des années 90, ce domaine concernait en France encore la production de matériel ainsi que le développement de logiciels. Aujourd'hui, les principales entreprises de ce secteur ont comme cœur de métier l'intégration de progiciels, c'est-à-dire réaliser la mise en production de progiciels dans les systèmes et processus des entreprises. Le risque lié à l'assurance de responsabilité civile générale de ce domaine a donc fortement évolué dans sa nature même. Cet effet ne sera pas pris en compte dans les modèles : c'est-à-dire que nous supposons que la politique de souscription cherchera à conserver le même niveau de risque en adaptant la matière souscrite.

L'autre effet notable concerne les cycles de souscription. Il est observé sur le marché de l'assurance des risques industriels des périodes où les primes sont réputées comme étant basses, et d'autres où elles sont réputées comme étant élevées. L'alternance de ces périodes a fait l'objet d'étude visant à caractériser des effets de cycles : c'est-à-dire à définir la fréquence de ces cycles et les événements déclencheurs de retournement de cycles, voire d'essayer de déterminer des modèles prédictifs à partir de différents éléments quantitatifs et qualitatifs connus sur le marché. Cet aspect influe d'abord sur le couple (prime, exposition) des contrats, mais comme les modèles présentés dans cette étude n'utilisent pas les Primes mais seulement la base de transactions sinistres et des informations d'exposition pour le modèle détaillé, son impact sera considéré ici comme minime.

3.3 Outils utilisés

Les données de gestion qui servent de base d'analyse à l'étude sont stockées sur une base ORACLE et convenablement préparées pour l'ensemble des besoins de « reporting » de la société dans une base spécifique.

Le logiciel SAS Enterprise Guide 4.2 est utilisé pour préparer les données et réaliser l'ensemble des travaux statistiques de cette étude du fait de sa capacité à manipuler de grandes quantités de données et sa technologie statistique disponible en 2009. Les modules SAS/Base, SAS/Stats, et SAS/IML seront particulièrement utilisés.

L'ensemble du code SAS utilisé pour les régressions et les simulations du modèle sur données détaillées est disponible dans le document « Matériel supplémentaire – Code SAS du modèle détaillé » et téléchargeable au lien suivant :

https://drive.google.com/file/d/1qCCuSvjhUbNEP82TAmy7oBJ61vv_dRf3/view?usp=drive_link

Ou sur simple requête auprès de l'auteur sur LinkedIn :

<https://www.linkedin.com/in/jean-michel-vautrin-96b3423/>

Le logiciel Excel est également utilisé pour sa simplicité de mise en œuvre dans la présentation de tableaux et de graphiques.

4 Méthodes de provisionnement

Avant d'évoquer les méthodes de provisionnement elles-mêmes, il convient de revenir sur le but recherché de ces calculs. Les compagnies d'assurance optent pour une méthode à la lumière de deux objectifs qui s'imposent à elles :

- Le législateur impose à toute compagnie d'assurance d'être capable de faire face à l'ensemble de ses engagements, et ce à tout instant. En assurance dommage, et en l'absence de la connaissance de la charge sinistre ultime de chacun des contrats vendus, il doit donc réaliser une estimation statistique transparente sur l'évolution de sa charge sinistre afin d'en estimer le montant ultime, puis incorporer une réserve supplémentaire, si nécessaire, afin de compléter les réserves déjà ouvertes sur chacun des sinistres connus.
- La réforme Solvabilité II impose de son côté aux compagnies d'assurances d'avoir une réserve en capital suffisante pour faire face aux imprévus susceptibles de survenir dans l'ensemble des risques liés à l'exercice de leurs activités. Cette approche utilise des critères statistiques et financiers pour la mesure du capital nécessaire, plus précisément un horizon temporel et un niveau de confiance.

En se focalisant sur le provisionnement des sinistres, ces deux mesures s'appuient sur la même base de risque sous-jacent.

Au travers la littérature actuarielle, un observateur peut voir que les méthodes de provisionnement sont nombreuses : il n'existe pas de méthode universelle. Néanmoins, il est possible de retenir certaines caractéristiques majeures de ces différentes méthodes afin de les classer. C'est l'approche proposée par Taylor et al. (2003) dont voici une synthèse :

- La stochasticité : Ceci dépend de la présence d'un terme d'erreur stochastique explicite au sein du modèle. On parlera de modèle stochastique si oui, et de modèle déterministe en son absence.
- Le dynamisme : Un modèle est soit statique, soit dynamique. Le critère de différence est la présence de paramètres dépendant du temps pour un modèle dynamique.
- La structure du modèle : Ce point discerne les modèles s'intéressant aux larges phénomènes de sinistralité, donc à travers une approche de données agrégée, des modèles ayant une structure fine pour capturer les microphénomènes des sinistres qui auront donc une approche d'analyse sur des données détaillées.

Ces trois critères peuvent être mélangés pour définir 2^3 typologies de modèles. Par exemple le modèle traditionnel Chain-Ladder peut être défini comme Agrégé / Déterministe / Statique, alors que le modèle de Mack qui complète le modèle Chain-Ladder en estimant la volatilité des paramètres du triangle pourra être défini comme Agrégé / Stochastique / Statique. Ainsi l'ensemble des modèles destinés au calcul des provisions pour sinistres à payer peuvent être vu tel un arbre généalogique.

L'ensemble de ces approches théoriques ont été développées non pas par simple défi intellectuel, mais bien pour fournir des solutions pragmatiques face aux différents cas de figures qui se présentent à l'assureur. Ainsi, le choix du modèle s'appuiera sur la structure et la qualité des données disponibles, sur les caractéristiques du risque sous-jacent, et sur les résultats que l'on souhaite obtenir. Chacune de ces contraintes peut réduire significativement les choix de méthodes :

- Si seule une agrégation des données est disponible, il ne sera pas possible d'utiliser une méthode détaillée
- Si l'objectif est d'avoir un résultat stochastique, cela disqualifie l'ensemble des méthodes déterministes.

Un des critères majeurs inclus dans Solvabilité II est d'estimer l'incertitude, et vient donc ajouter une nouvelle contrainte dans le choix de la méthode. Celle-ci sera alors celle validant les contraintes opérationnelles et estimant l'incertitude le risque sous-jacent de la manière la plus robuste.

Nous pouvons voir en arrière-plan se dessiner l'impact de ces réflexions sur le fonctionnement opérationnel d'une compagnie d'assurance. Imaginons que dans la situation du portefeuille étudié dans cette étude, l'assureur ne dispose que de données agrégées. Du fait de la faible robustesse des outils de mesures disponibles sur ses données, l'assureur sera tenté de modifier ses systèmes d'informations en profondeur afin de pouvoir exploiter les informations nécessaires à l'application d'une méthode plus finement segmentée, voir complètement détaillée.

4.1 Modèles sur données agrégées

4.1.1 Chain Ladder standard

La méthode Chain-Ladder est une des méthodes d'estimation des provisions les plus utilisées par sa simplicité de mise en œuvre et sa relative consistance sur des risques bien mutualisés et stables dans le temps (Denuit et Charpentier, 2005). Voici ci-dessous la forme couramment rencontrée de cette méthode :

Soit $C_{i,k}$, le montant cumulé des sinistres de l'année de survenance i jusqu'à l'année de développement k . Si n années sont connues alors $k \leq n$. Ce montant peut être basé sur les paiements ou le coût total des sinistres, c'est-à-dire les paiements effectués plus les réserves en cours à l'année de développement k . Les valeurs des $C_{i,k}$ sont connues pour

$i + k \leq n + 1$, et l'on cherche donc les valeurs des $C_{i,k}$ tel que $i + k > n + 1$.

La matrice suivante représente l'ensemble des valeurs connues :

$$C = \begin{pmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,n-1} & C_{1,n} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,n-1} & \\ \vdots & \vdots & \ddots & & \\ C_{n-1,1} & C_{n-1,2} & & & \\ C_{n,1} & & & & \end{pmatrix}$$

La méthode consiste à estimer les facteurs de développement suivants :

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}} \text{ avec } 1 \leq k \leq n-1$$

Les déroulements inconnus sont déduits par :

$$\hat{C}_{i,k} = C_{i,n+1-i} \cdot \prod_{j=n+1-i}^{n-k+1} \hat{f}_j \text{ pour } i + k > n + 1.$$

Le montant ultime des sinistres pour l'année i est alors donné par $\hat{C}_{i,n}$ avec $k = n$ dans la formule précédente.

La réserve $\hat{R}$ pour l'ensemble des années peut alors s'écrire :

$$\hat{R} = \sum_{i=2}^n C_{i,n+1-i} \cdot (\prod_{j=n+1-i}^{n-1} \hat{f}_j - 1).$$

Les réserves nécessaires estimées par cette méthode sont donc proportionnelles aux dernières positions connues pour chaque année. Le facteur proportionnel est calculé en fonction des développements constatés sur les années de survenance ayant déjà atteint le niveau de développement requis pour les années plus récentes. Cette logique peut s'appliquer également pour d'autres unités de temps que l'année comme le trimestre ou le mois. Ceci est à ajuster en fonction de la typologie du risque sous-jacent, la logique de cette approche est de globaliser les micro-événements pour faire ressortir les impacts macros de l'évolution des sinistres sur les réserves.

Néanmoins la méthode Chain-Ladder pose plusieurs problèmes :

- Elle ne permet pas de calculer une volatilité des provisions pour sinistre dans le cadre de Solvabilité II.
- Dans le cas d'une sinistralité particulièrement volatile, comme sur le portefeuille considéré, un sinistre atypique peu influencer sensiblement les facteurs de développement du triangle. Nous avons alors un comportement d'un sinistre élémentaire, que l'on peut qualifier de micro-événement, qui modifiera sensiblement les paramètres sensés refléter les comportements macro des sinistres du portefeuille. Ceci va alors à l'encontre du but recherché de la méthode. S'il y a un nombre limité de sinistres atypiques, nous pouvons essayer de les traiter à part, avec l'aide de l'expertise des régleurs sinistres, mais dans le cadre du portefeuille présent (voir le Tableau 1: Profil de la charge sinistre de la branche RC Générale - années de survenance 1992-2009), cette approche viderait alors le triangle de la grande majorité de sa charge à l'ultime.

Ces particularités viennent s'ajouter aux problèmes de robustesse de cette méthode dus à la forte influence des valeurs périphériques du triangle (Verdonck et al., 2009) :

- Le dernier facteur de développement du triangle, $C_{1,n}$, ne dépend que d'une seule valeur et influence l'ensemble des projections du triangle.
- De même, en cas d'absence de valeur pour la dernière année de survenance, $C_{n,1}$, l'utilisation de la méthode n'est plus immédiate, imposant alors des hypothèses sur la valeur la plus travaillée au travers des facteurs du triangle.

4.1.2 Chain-Ladder stochastique - I (Mack 1993)

La recherche sur la volatilité des triangles de liquidation a abouti sur plusieurs modèles différents. Pour ne pas dévier de l'objectif d'analyse de cette étude, seuls deux des modèles les plus marquants

seront détaillés ici. Tout d'abord voici le modèle stochastique proposé par T. Mack qui est basé sur les triangles cumulés des paiements (Partrat et al., 2008).

Ce modèle repose sur trois hypothèses :

1. $E(C_{i,k+1}|C_{i1}, \dots, C_{ik}) = C_{ik}f_k$ pour $1 \leq i \leq n$ et $1 \leq k \leq n-1$
2. $\{C_{i1}, \dots, C_{in}\}, \{C_{j1}, \dots, C_{jn}\}$ sont indépendants $\forall i, j$.
3. $Var(C_{i,k+1}|C_{i1}, \dots, C_{ik}) = C_{ik} \cdot \sigma_k^2$ où $1 \leq i \leq n, 1 \leq k \leq n-1$

Ainsi, sous les hypothèses (1) et (2), les estimateurs $\hat{f}_k$ de f_k sont sans biais. La méthode de Mack fournit en espérance les mêmes résultats que la méthode originale de Chain-Ladder.

En ajoutant l'hypothèse (3), les estimateurs $\hat{\sigma}_k^2$ de σ_k^2 sont également sans biais, avec

$$\hat{\sigma}_k^2 = \frac{1}{n-k-1} \sum_{i=1}^{n-k} C_{ik} \left(\frac{C_{i,k+1}}{C_{ik}} - \hat{f}_k \right)^2$$

Néanmoins cette estimation pose une difficulté pour $k = n-1$. T. Mack a proposé la solution suivante :

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

L'estimation de l'erreur carré moyenne prédictive de $\hat{R}_i$ fournit par la méthode de T. Mack est :

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

Et pour l'ensemble $\hat{R}$, cette estimation est donnée par :

$$MSEP(\hat{R}) = \sum_{i=2}^n \left\{ MSEP(\hat{R}_i) + \hat{C}_{in} \left(\sum_{l=i+1}^n \hat{C}_{ln} \right) * \sum_{k=n-i+1}^{n-1} \frac{2\hat{\sigma}_k^2}{\hat{f}_k^2 \sum_{h=1}^{n-k} C_{hk}} \right\}$$

Une fois ces estimations faites, la recherche d'un percentile α de la distribution de $\hat{R}$ est réalisée par son ajustement à une loi paramétrique par la méthode des moments, généralement la loi Normale ou loi Log-Normale en admettant a priori l'hypothèse de Normalité ou de Log Normalité de $\hat{R}$.

Les remarques faites sur la méthode Chain-Ladder s'appliquent d'office à cette méthode sur l'estimation de $E(\hat{R})$, mais peuvent également se prolonger sur l'estimation de la volatilité autour de cette valeur, c'est-à-dire de $MSEP(\hat{R})$.

4.1.3 Chain-Ladder stochastique II (GLM/Bootstrap)

Une autre approche utilisée, est l'application d'un modèle linéaire généralisé (GLM) aux triangles agrégés de liquidation des sinistres. En combinant ce type de modèle avec la technique de rééchantillonnage appelée Bootstrap, et la méthode d'estimation numérique par simulation dite de Monte-Carlo, cette approche permet de générer une multitude de triangles de liquidation probable qui permet ainsi de définir une distribution des coûts des sinistres à différents horizons, et ainsi des provisions nécessaires. (England et Verall, 2002)

Modélisation des triangles de liquidation des sinistres par un GLM

Les triangles de liquidations des sinistres peuvent être analysés sous plusieurs angles. Il est possible d'analyser l'évolution des paiements ou des charges totales par année de survenance. Ceci peut être vu à partir des incréments par année de développements ou bien des cumuls par année de développement. Ces approches, quoique généralement utilisées, ne s'adaptent pas au portefeuille considéré : les années de survenance ont des montants trop hétérogènes pour obtenir une robustesse suffisante dans l'estimation des modèles. Nous proposons alors de nous intéresser plutôt aux facteurs de développements des valeurs cumulées de paiements ou de charges totales. Dans ces cas-là, les triangles des facteurs ont une colonne et une ligne de moins que les triangles des valeurs des incréments ou des cumuls. (Planchet et al. 2005)

En reprenant la notation développée dans la section dédiée à la méthode Chain-Ladder, nous avons :

$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}}$ où $C_{i,j}$ représentent la valeur utilisée dans le triangle à l'année de survenance i et l'année de développement j . Cette valeur peut-être la valeur cumulée des paiements ou des charges totales.

Le triangle des facteurs de développement se présente ainsi :

$$\hat{F} = \begin{pmatrix} \hat{f}_{1,1} & \hat{f}_{1,2} & \cdots & \hat{f}_{1,n-1} & \hat{f}_{1,n-1} \\ \hat{f}_{2,1} & \hat{f}_{2,2} & \cdots & \hat{f}_{2,n-1} & \\ \vdots & \vdots & \ddots & & \\ \hat{f}_{n-1,1} & \hat{f}_{n-1,2} & & & \\ \hat{f}_{n-1,1} & & & & \end{pmatrix}$$

Il est possible ainsi d'interpréter $\hat{f}_k$ comme étant la moyenne des facteurs $\hat{f}_{i,k}$ du vecteur $\hat{F}_k$ pondérés par les $C_{i,k}$. En effet :

$$\begin{aligned} \hat{f}_k &= \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{i,k}} = \frac{C_{1,k+1} + \cdots + C_{i,k+1} + \cdots + C_{n-k,k+1}}{C_{1,k} + \cdots + C_{i,k} + \cdots + C_{n-k,k}} \\ &= \frac{C_{1,k} \frac{C_{1,k+1}}{C_{1,k}} + \cdots + C_{i,k} \frac{C_{i,k+1}}{C_{i,k}} + \cdots + C_{n-k,k} \frac{C_{n-k,k+1}}{C_{n-k,k}}}{C_{1,k} + \cdots + C_{i,k} + \cdots + C_{n-k,k}} \end{aligned}$$

$$\hat{f}_k = \frac{C_{1,k}\hat{f}_{1,k} + \dots + C_{i,k}\hat{f}_{i,k} + \dots + C_{n-k,k}\hat{f}_{n-k,k}}{C_{1,k} + \dots + C_{i,k} + \dots + C_{n-k,k}}$$

Et donc :

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k}\hat{f}_{i,k}}{\sum_{i=1}^{n-k} C_{i,k}}$$

L'application d'un modèle GLM à ces triangles peut alors se faire avec les caractéristiques suivantes :

- Variable réponse basée sur les incréments avec comme loi utilisée la loi de Poisson, la loi de Poisson sur-dispersée, ou bien encore la loi Gamma
- Fonction de lien canonique
- Années de survenance et années de développement utilisées comme des variables qualitatives.

Cette approche fonctionne difficilement sur le portefeuille analysé par la dispersion des montants entre les années de survenance. Les expositions peuvent varier nettement d'une année de souscription à l'autre rendant instable les montants des incréments des sinistres d'une année de survenance à l'autre. C'est pourquoi nous proposons d'appliquer le GLM sur le triangle $\hat{F}$ des facteurs de développement avec les caractéristiques suivantes :

- Variable réponse basée sur les facteurs de développements de valeurs cumulées avec comme loi utilisée la loi de Poisson.
- Fonction de lien canonique : fonction logarithme
- Variable de pondération : $C_{i,k}$
- Années de survenance et années de développement utilisées comme des variables qualitatives.

Bootstrap et application sur un triangle de liquidation modélisé par un GLM

La technique du Bootstrap est une méthode de rééchantillonnage dont l'objectif est d'obtenir des informations sur une loi statistique uniquement à travers un échantillon de réalisations de cette loi et sans poser d'hypothèses externes. Ceci implique que l'on suppose les éléments de l'échantillon utilisé comme étant indépendants et identiquement distribués (iid). (Planchet et al., 2005)

Cette méthode consiste à réaliser m tirages avec remise d'un n échantillon afin d'observer la variabilité d'un paramètre, avec m suffisamment grand. Ainsi il sera possible d'estimer le paramètre à partir des rééchantillons plutôt qu'en utilisant le n échantillon de départ.

Dans le cadre d'un triangle de sinistres, les éléments du triangle sont les réalisations à rééchantillonner. Mais ceux-ci ne respectent pas parfaitement l'hypothèse iid. Bien que supposé indépendants, les éléments du triangle ne suivent pas tous la même loi : le GLM fournit des

paramètres différents par année de survenance et par année de développement. C'est pourquoi la méthode du Bootstrap est appliquée aux résidus qui respectent l'hypothèse iid. Par la simplicité d'implémentation, les résidus de Pearson sont les plus souvent utilisés.

En pratique, on utilise les n résidus de Pearson du triangle originel sur lesquels on réalise un n tirage avec remise qui constitue ainsi un nouveau triangle de résidus. A partir de ceux-ci on réestime les nouvelles valeurs du triangle, puis on obtient l'estimation des provisions issues de ce nouveau triangle par la méthode du Chain-Ladder.

Cette opération est réalisée m fois fournissant m estimations des provisions et décrivant ainsi sa loi statistique.

4.2 Modèle sur données détaillées

D'un point de vue pratique, un modèle de provisionnement peut être vu comme un arbitrage entre l'utilisation de méthodes réduisant la volatilité des prévisions, et le coût d'utilisation de ces méthodes en termes de développement et de temps de traitement par rapport au bénéfice final escompté. La granularité d'un modèle de provisionnement dépend à la fois des données disponibles, mais aussi de l'homogénéité des risques sous-jacents. En effet, en partant du principe que l'on a accès à l'intégralité des informations des sinistres, l'homogénéité des risques va être un facteur décisif dans l'ajustement de la granularité du modèle : des risques très homogènes pourront ne nécessiter qu'un unique triangle, un portefeuille séparé en un nombre limité de classe de risques nécessitera alors autant de triangles dédiés à chacune des classes. Et ainsi, un portefeuille aux risques particulièrement hétérogènes nécessitera une caractérisation très fine du modèle de provisionnement.

Ce qui est proposé ici, du fait des caractéristiques des risques du portefeuille analysé, c'est d'appliquer un modèle ayant une granularité fine cherchant à caractériser l'ensemble des sinistres individuellement.

4.2.1 Préambule : rappels académiques de méthodes utilisées

Afin d'aide à la compréhension des éléments qui vont suivre, des rappels académiques des principales méthodes appliquées ont été ajoutées en annexes :

- Les modèles GLM : 10.1.1 page 115
- Simulation de Monte-Carlo : 10.1.2 page 121
- Régression Logistique : 10.1.3 page 125
- Régression Bêta : 10.1.4 page 127
- Loi Normale Contaminée : 10.1.5 page 129
- Modèle de survie : 10.1.6 page 129

4.2.2 Concept

La vie des provisions techniques pour sinistre d'une société d'assurance peut être vue comme une suite d'événements, de l'ouverture des sinistres jusqu'à leurs clôtures, tel qu'introduit par la Figure 3 page 19. En réalisant des modèles simples pour chacun de ces événements élémentaires, il devient possible de caractériser l'ensemble de la vie d'un sinistre en chaînant l'ensemble de ces sous-modèles.

A partir des données disponibles, nous allons distinguer la modélisation des sinistres déclarés de ceux restant à déclarer. Nous allons donc définir un premier module qui nous permettra de générer des sinistres non déclarés, avec leurs caractéristiques propres, tels ceux déjà déclarés, mais avec la particularité qu'ils entreront dans les bases comptables dans le futur. Ainsi nous obtenons l'ensemble des sinistres rattaché aux expositions du portefeuille à la date d'analyse. Puis nous prendrons l'ensemble des sinistres pour modéliser le développement individuel de chacun d'eux, par mouvement indemnitare et par mouvement pour frais, et ce jusqu'à leurs clôtures. Nous pourrions ensuite réaliser l'ensemble des travaux d'analyses autour des résultats du modèle (voir Figure 11: Schéma d'ensemble du processus de modélisation ci-dessous).

Principe général

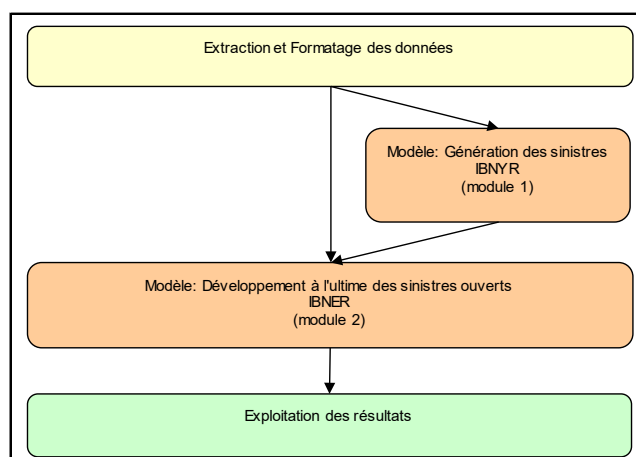


Figure 11: Schéma d'ensemble du processus de modélisation

Si l'on regarde la chronologie d'un sinistre S_i , les deux modules principaux cherchent à caractériser la vie entière d'un sinistre, où le module 1, dédié aux sinistres non ouverts, va de la survenance à l'ouverture du dossier dans les comptes de l'assureur, y compris toute transaction initiale attachée à cette ouverture, et où le module 2, dédié aux sinistres ouverts, poursuit la chronologie des sinistres jusqu'à leurs clôtures (voir Figure 12) :

Modules utilisés pour modéliser la vie d'un sinistre S_i

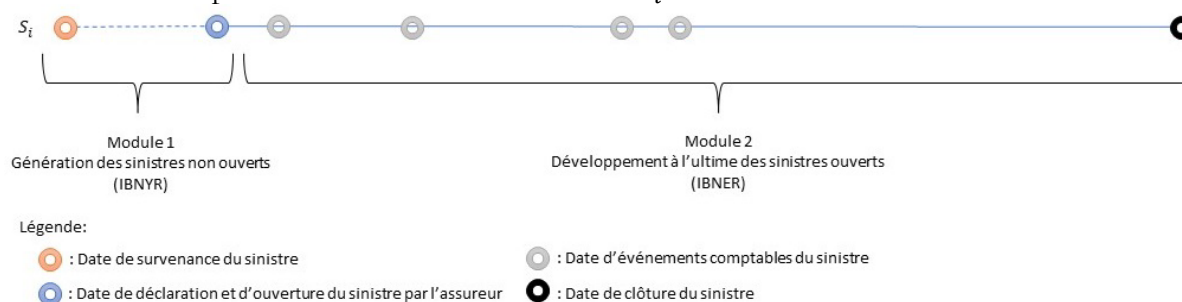


Figure 12: Modules utilisés pour modéliser la vie d'un sinistre S_i

Il convient de noter que seuls les engagements présents dans la base de données peuvent être utilisés, et donc toute police dont l'engagement ferme a été légalement validé par l'assureur à la date d'analyse, par signature du contrat ou par courriel auprès de l'assuré ou de son mandataire, mais sans être encore renseignée dans le système de gestion des polices ne peut être prise en compte.

Les montants de sinistres sont vus « FGU » (« from the ground-up » en anglais, ou au 1^{er} euro), c'est-à-dire bruts de toutes franchises et limites. Ce n'est que lors de l'exploitation des résultats que l'ensemble des franchises et des limites est à nouveau appliqué pour observer ce qui est réellement à la charge de l'assureur (voir Figure 55: Exemple de densité tronquée).

D'autres parts, nous allons caractériser les différentes étapes de la vie d'un sinistre au niveau indemnitaire que ce soit sur l'évolution de ses provisions ou de ses paiements. Historiquement, les provisions ne distinguaient pas, jusqu'à très récemment, la partie indemnitaire de la partie pour frais. Cela a deux conséquences dans notre analyse :

- Nous considérerons les provisions globales déduites des frais réellement payés comme étant des provisions indemnitaires.
- Pour les frais, nous nous intéresseront qu'à caractériser les flux de paiements, et donc à ne pas prendre en compte les provisions dédiées aux frais.

Module de génération de sinistres non ouverts (IBNYR)

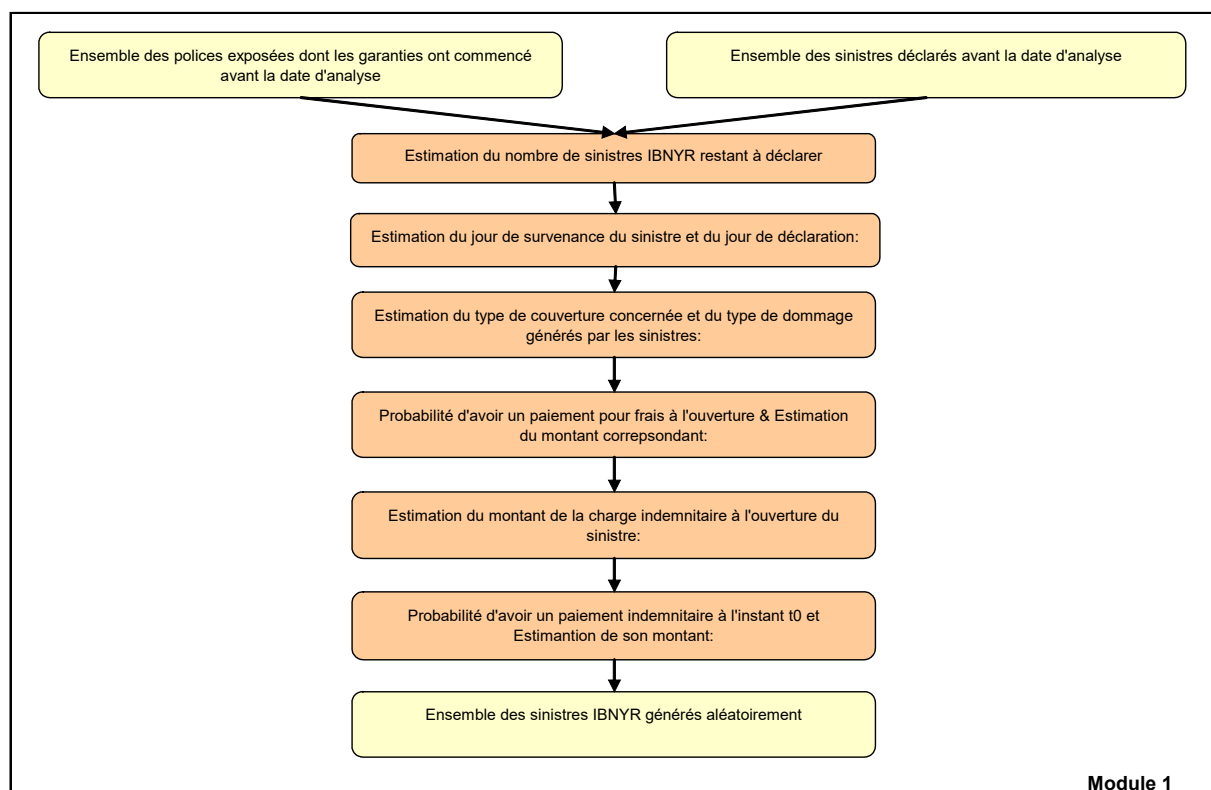


Figure 13: Schéma de la modélisation des sinistres non ouverts (IBNYR)

Ce premier module (voir Figure 13: Schéma de la modélisation des sinistres non ouverts (IBNYR)) va croiser l'historique des contrats et leurs conséquences en termes de sinistralité :

A partir de l'ancienneté des contrats, de leurs durées de couvertures, et du nombre de sinistres déjà déclarés, nous allons estimer le nombre de sinistres restant à déclarer.

Puis nous allons estimer le jour de survenance et le jour de déclaration de ces sinistres. Nous allons ensuite définir quelles sont les caractéristiques de ces sinistres, c'est-à-dire quels sont les types de couvertures générant ces sinistres et quels sont les types de dommages donnant droit aux réclamations.

Nous estimons ensuite s'il y a ou non un paiement pour frais dès l'ouverture et de quel montant le cas échéant.

Puis, nous estimons la charge indemnitaire totale à l'ouverture. Ce modèle se décompose en deux parties, tout d'abord nous estimons si le sinistre est ouvert à un montant récurrent, c'est-à-dire un montant fréquemment utilisé à l'ouverture lorsque le régleur sinistre n'a pas d'informations concrètes pour affiner le montant de provision nécessaire. Ce montant peut être vu comme forfaitaire. Dans le cas contraire, nous estimons ce montant suivant l'historique observé des ouvertures de sinistres que l'on pourra qualifier de non forfaitaires.

Nous terminons cette partie en caractérisant la probabilité d'avoir un paiement indemnitaire dès l'ouverture et de quel montant le cas échéant. Ce montant est vu comme un ratio de la charge totale indemnitaire estimée à l'étape précédente.

Ainsi à l'issu des calculs de ce module, nous nous retrouvons avec une base de sinistres qui seront déclarés dans le futur et sur lesquels nous pouvons appliquer une méthodologie de développement commune avec les sinistres déjà déclarés à la date d'analyse dans le module 2.

Module de développement à l'ultime des sinistres ouverts (IBNER)

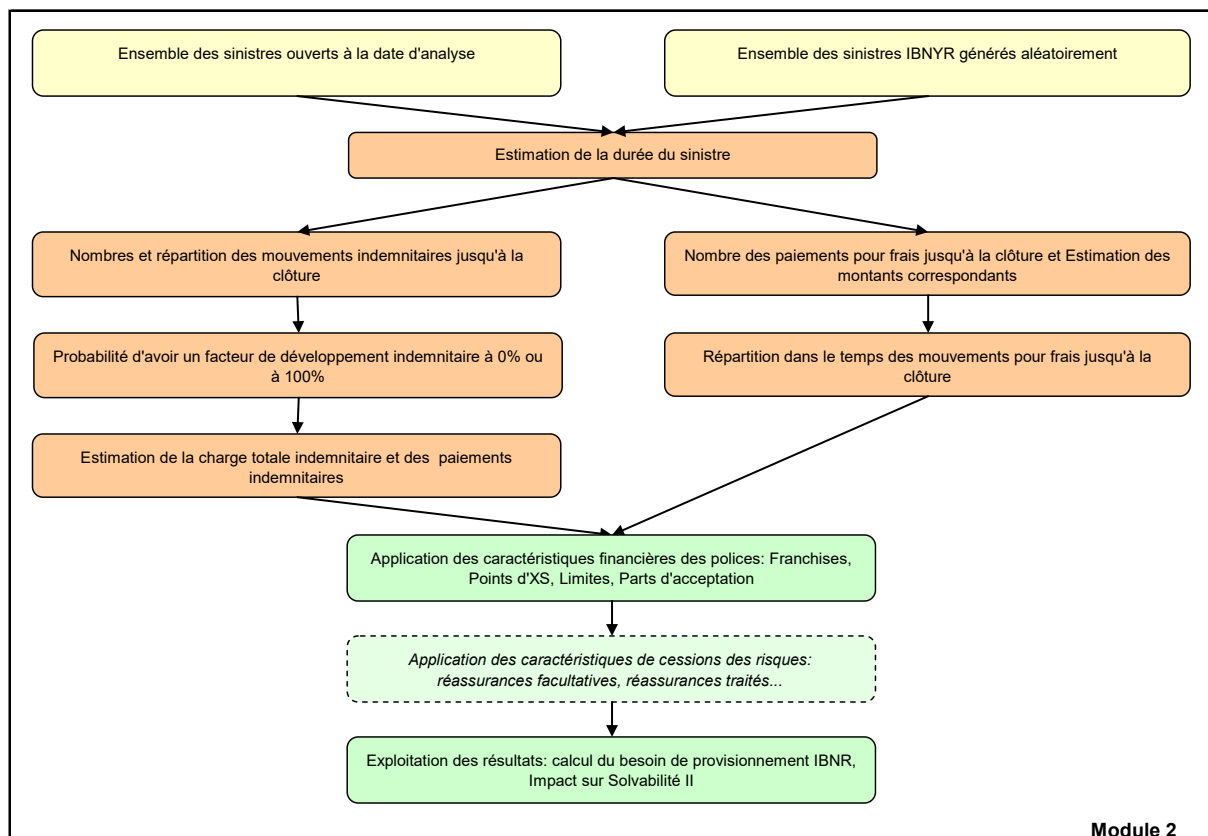


Figure 14: Schéma de la modélisation du développement à l'ultime des sinistres ouverts (IBNER)

Le deuxième module reprend l'ensemble des sinistres ouverts au jour de l'analyse et l'ensemble des sinistres survenus et non déclarés générés par le premier module. La première étape de ce module consiste à estimer la date de clôture des sinistres. Ceci est réalisé à partir d'un modèle de survie.

Puis le processus se sépare en deux parties distinctes : l'indemnisation et les frais.

Concernant l'indemnisation, on caractérisera le nombre de mouvements indemnitaires restant jusqu'à la clôture et leurs répartitions sur cette période. Ensuite, nous estimerons leurs impacts sur la charge totale indemnitaire des sinistres. Ces impacts peuvent être vus en trois catégories principales : ou bien le mouvement a un impact neutre (facteur de développement de la charge indemnitaire égale à 1), ou bien il annule la provision indemnitaire (facteur de développement sur la charge indemnitaire égale à 0), ou toute autre évolution non comprise dans les deux situations précédentes. La provision indemnitaire restante à la clôture est alors vue comme le montant de paiement indemnitaire final du sinistre.

Concernant les frais, nous commençons par estimer le nombre de paiements pour frais restant jusqu'à la clôture des sinistres, ainsi que les montants correspondants. Puis ces mouvements sont répartis dans le temps jusqu'à la clôture.

5 Mise en œuvre des modèles de provisionnement

Même si les principes des différents modèles ont été présentés précédemment, il est nécessaire d'approfondir leurs mises en œuvre pour en comprendre les différentes caractéristiques. Ainsi nous pourrions conclure sur leurs capacités et limites propres dans l'objectif de cette étude.

Les tableaux de résultats des différentes régressions se trouvent en annexes 10.2 page 134.

5.1 Modèle sur données détaillées – module 1 – Génération des sinistres non ouverts (IBNYR)

5.1.1 Estimation du nombre de sinistres IBNYR restant à déclarer

Estimation du nombre de sinistres IBNYR restant à déclarer:
4 sous-modèles

L'objectif est de connaître le nombre de sinistres restant à déclarer pour chaque contrat observé en portefeuille, aussi bien concernant les contrats en cours que les contrats antérieurs afin de capturer l'ensemble des engagements de l'assureur sur le portefeuille.

Pour cela, nous comptons le nombre de dossiers sinistres déclarés de l'ensemble du portefeuille. Ce comptage est structuré par génération pour chaque contrat et par année de développement. La notion de contrat par génération s'interprète comme étant un contrat d'une durée maximale d'un an à la date d'échéance principale. Les contrats renouvelés plusieurs années sont donc vu comme autant de génération de contrats. Nous pouvons parler aussi d'une vision des contrats par année de souscription. Pour cela, nous prendrons comme référentiel d'observation d'un contrat la date de début de couverture du contrat.

Comme le but est d'estimer le nombre de sinistres restant à déclarer pour chaque génération de chaque contrat du portefeuille sachant une situation de sinistralité à l'instant de l'analyse, nous allons segmenter en trois cas principaux la situation possible combinant à quel moment d'observation des contrats nous nous trouvons et quelle est la situation de sinistralité, en termes de nombre de sinistres. :

- Cas 1 : Les générations de contrats au début de la période de couverture.
- Cas 2 : Les générations de contrats au-delà de la première année d'observation et n'ayant subi aucun sinistre
- Cas 3 : Les générations de contrats au-delà de la première année d'observation et ayant subi des sinistres

A cela un 4^{ème} cas vient s'ajouter afin d'homogénéiser les Cas 1 et 3 en traitant à part des contrats ayant une fréquence de sinistres atypiques.

- Cas 3 bis « contrats à fréquence atypique » : Les générations de contrats au-delà de la première année d'observation et ayant subi un nombre significatif de sinistres par rapport aux autres contrats du portefeuille.

Dans ce dernier cas, les contrats sont analysés à part du reste du portefeuille et font l'objet de modèles individuels dédiés par contrat afin de refléter plus fidèlement leurs comportements propres.

Remarque : Nous pouvons anticiper l'enrichissement de ces modèles en prévoyant également l'intégration des informations de sinistralité sur les périodes précédant l'entrée en portefeuille des contrats ; ces informations étant reçues dans le cadre des études préalables de souscription. Ainsi le cas 3 ci-dessus serait amené à disparaître pour requalifier les contrats concernés sur les autres cas possibles.

• Cas 1 : Les générations de contrats au début de la période de couverture

L'objectif est d'estimer le nombre de sinistre qui seront déclarés pendant la première année de couverture d'une génération d'un contrat. Nous allons donc estimer un modèle de comptage en fonction des caractéristiques des contrats.

Remarque : Nous pouvons anticiper l'enrichissement de ce modèle en prévoyant également l'intégration des informations de sinistralité sur les périodes précédant l'entrée en portefeuille des contrats ; ces informations étant reçues dans le cadre des études préalables de souscription. Ainsi ce cas sera amené à disparaître pour être assimilé au Cas 2 ou 3.

Dans ce cas le nombre de sinistres à déclarer est estimé par un modèle GLM (Pedan, 2001) ayant les caractéristiques suivantes :

- Loi : Binomiale Négative
- Fonction lien : Log
- Variables explicatives :
 - Variable de classe - Contrats en Apéritif ou en Coassurance : TYPE_C_USED
 - Variable de classe - Type d'activité assuré : OCCUPANCY_USED
- Pas de constante

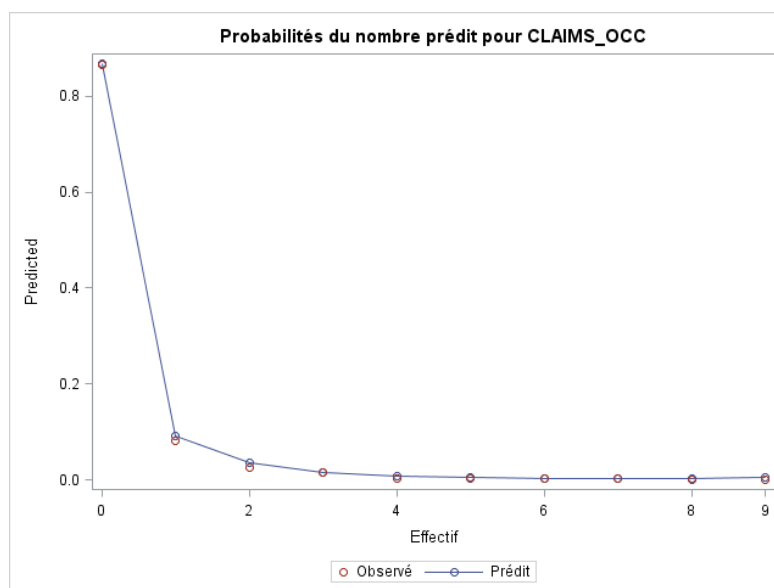


Figure 15: Nombre de sinistres déclarés pendant la première année de couverture – Loi Binomiale Négative

Le modèle retenu a une capacité prédictive acceptable d'après le graphe précédent.

Pour information, la loi de Poisson testée également est apparue moins pertinente.

• Cas 2 : Les générations de contrats au-delà de la première année d'observation et n'ayant subi aucun sinistre

L'objectif est d'estimer le nombre de sinistre qui seront déclarés pendant l'année de développement suivant la date de l'analyse pour les générations de contrats n'ayant subi aucun sinistre jusqu'à la date d'analyse. L'approche est donc similaire au cas précédent en y ajoutant l'information sur le nombre d'année de développement observée à la date d'analyse.

Dans ce cas le nombre de sinistres à déclarer est estimé par un modèle GLM ayant les caractéristiques suivantes :

- Loi : Binomiale Négative
- Fonction lien : Log
- Variables explicatives :
 - Variable de classe - Contrats en Apéritif ou en Coassurance : TYPE_C_USED
 - Variable de classe - Type d'activité assuré : OCCUPANCY_USED
 - Variable de classe - Nombre d'années de développement : IBNYR_DEV_YEAR_CNT
- Pas de constante

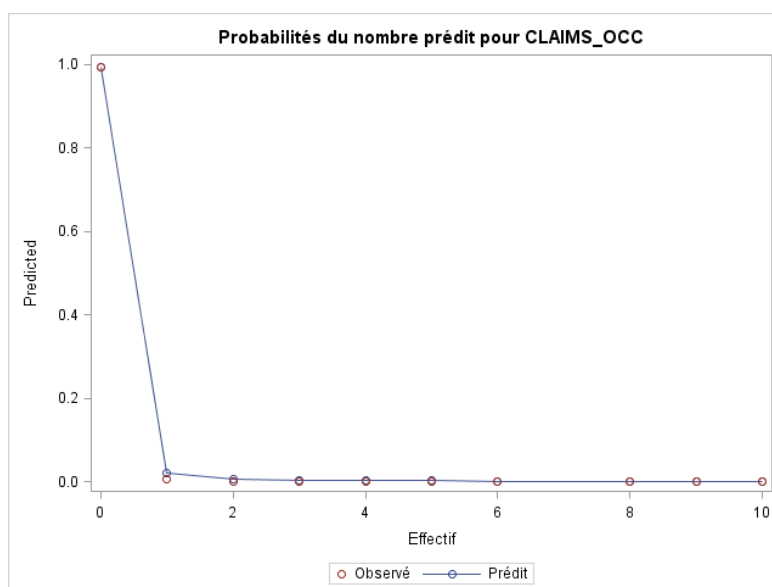


Figure 16: Nombre de sinistres déclarés pendant la prochaine année de développement – Loi Binomiale Négative

Le modèle retenu a une capacité prédictive acceptable d'après le graphe précédent.

Globalement, pour cette catégorie de générations de contrats, nous pouvons observer que la probabilité d'avoir un premier sinistre est assez faible. En analysant les paramètres liés aux années

de développements observés (voir tableaux en annexes page 134), la probabilité est négligeable, mais non-nulle, au-delà de la 5^{ème} année de développement.

Là aussi, et pour information, la loi de Poisson testée également est apparue moins pertinente.

• Cas 3 : Les générations de contrats au-delà de la première année d'observation et ayant subi des sinistres

L'objectif est d'estimer le nombre de sinistre qui seront déclarés pendant l'année de développement suivant la date de l'analyse pour les générations de contrats ayant subi des sinistres à la date d'analyse. L'approche est différente cette fois-ci par rapport aux deux premiers cas.

Nous allons estimer le facteur de développement du nombre de sinistres cumulés pour l'année de développement à venir plutôt que le nombre explicite de sinistre.

Dans ce cas, nous estimerons le logarithme du facteur de développement du nombre de sinistre cumulé par un modèle GLM ayant les caractéristiques suivantes :

- Loi : Normal
- Fonction lien : Identité
- Variables explicatives :
 - Variable de classe – Nombre d'années de développement : IBNYR_DEV_YEAR_CNT
 - Pas de constante

Une seule variable explicative est donc utilisée ici : dans cette configuration car le nombre d'années de développement écoulées à la date de l'analyse est apparu comme suffisamment significatif par rapport aux autres variables disponibles pour obtenir un résultat acceptable. En particulier, la variable type d'activité n'apporte pas d'information suffisante dans la régression.

• Cas 3 bis : Les générations de contrats au-delà de la première année d'observation et ayant une fréquence atypique de sinistres

Comme dans le Cas 3, l'objectif est d'estimer le nombre de sinistre qui seront déclarés pendant l'année de développement suivant la date de l'analyse pour les générations de contrats ayant subi des sinistres à la date d'analyse. Les contrats concernés sont uniquement ceux ayant un profil atypique de fréquence de sinistre par rapport au reste du portefeuille analysé précédemment. Le critère de sélection des contrats concernés est 20 sinistres survenus au bout de 5 années de développement pour l'ensemble des générations des contrats concernés.

	Nombre de Contrats	Nombre de génération de contrats	Nombre de Sinistres
Fréquence Standard	6,562	22,224	2,759
Fréquence Atypique	184	1,036	4,805

Tableau 2: Répartition des contrats entre fréquence standard et fréquence atypique

La même régression qu'au Cas 3 est réalisée, mais par contrat cette fois-ci, afin d'avoir des paramètres de développement adaptés au profil propre de chacun des contrats concernés.

Le logarithme du facteur de développement du nombre de sinistre cumulé par contrat est donc estimé par un modèle GLM ayant les caractéristiques suivantes :

- Groupement par Contrat
- Loi : Normal
- Fonction lien : Identité
- Variables explicatives :
 - Variable de classe – Nombre d'années de développement : IBNYR_DEV_YEAR_CNT
 - Pas de constante

5.1.2 Estimation du jour de survenance du sinistre et de son jour de déclaration

Estimation du jour de survenance du sinistre et du jour de déclaration:
 Loi Uniforme sur la période de garantie de la police et Modèle
 Linéaire Généralisée (Loi Normale)

• Estimation du jour de survenance des sinistres :

L'objectif est d'estimer les jours de survenances et les jours de déclarations correspondant des sinistres restant à déclarer estimés précédemment.

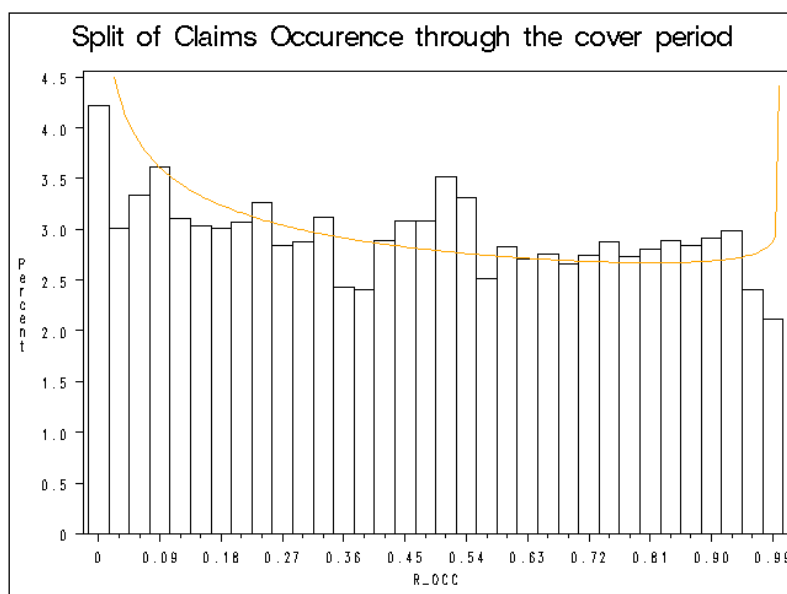


Figure 17: Répartition de la survenance des sinistres au cours de la période de couverture des contrats et ajustement d'une loi Béta aux données historiques

Au regard du graphique précédent, il ne semble pas y avoir de tendance marquée dans la survenance des sinistres au cours de la vie des contrats. Ainsi, nous prenons comme hypothèse qu'il n'y a pas de saisonnalité dans la survenance des sinistres. Nous allons estimer la date de survenance de ces sinistres suivant une loi uniforme entre les dates de début et de fin de couverture des contrats.

• Estimation du jour de déclaration des sinistres :

La date de déclaration est considérée comme un événement se produisant à un certain pourcentage de la durée de vie d'un sinistre, c'est-à-dire entre sa date de survenance et sa date de clôture. Nous nommerons ce pourcentage : ratio de déclaration du sinistre. La loi qui est naturellement privilégiée pour modéliser un ratio compris entre 0 et 1 est la loi Bêta. Mais elle ne convient pas aux données observées, c'est pourquoi nous avons opté de transformer le ratio de déclaration par la fonction LOGIT :

$$\text{Logit}(r^{\text{notif}}) = \log\left(\frac{r^{\text{notif}}}{1 - r^{\text{notif}}}\right)$$

Ainsi, nous passons de la variable r^{notif} définie sur $[0; 1]$ à la variable $\text{Logit}(r^{\text{notif}})$ définie sur $]-\infty; +\infty[$. Sur cette nouvelle variable, nous appliquons le modèle GLM suivant :

- Loi : Normale
- Fonction lien : Identité
- Variables explicatives :
 - Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
 - Variable de classe de la base de garantie du contrat (E en base survenance, C en base déclaration) : KZ_VERSAFLL (E / C)
 - Constante utilisée

Nous pouvons noter que pour les contrats où l'assureur est en position de coassureur, la déclaration du sinistre survient plus tard que pour les contrats où l'assureur est en apéritition. De même, la base de garantie en « mode survenance » a un effet retard sur la déclaration des sinistres. Ces deux effets intuitifs sont donc confirmés par les résultats de la régression. (Voir Tableau 23: Modèle d'estimation du délai de déclaration des sinistres - Résultats de la régression)

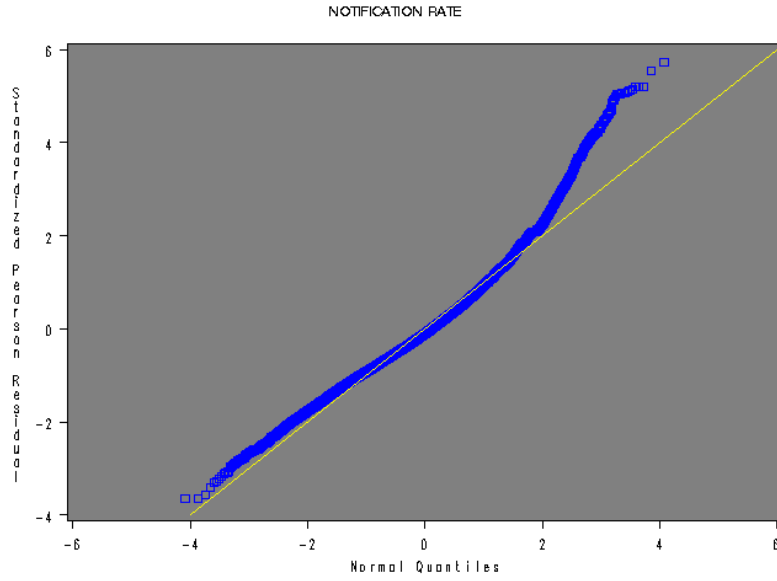


Figure 18: Modélisation du ratio de déclaration des sinistres - Résidus de Pearson standardisés

En pratique, nous réaliserons cette estimation dans le module 2, lorsque les durées totales des sinistres, c'est-à-dire de la survenance du sinistre jusqu'à sa clôture dans les comptes de l'assureur, seront évaluées. A partir de ces estimations, nous évaluons la date de déclaration sachant qu'un certain temps s'est écoulé à la date d'analyse depuis leurs survenances. Ce qui revient à la simulation suivante de la variable R^{notif} avec $U \in [0; 1]$ et X le vecteur des variables explicatives du modèle :

$$\begin{aligned}
 U &= F(r_0^{notif} + r^{notif} | R^{notif} \geq r_0^{notif} | X) \\
 &= P(r_0^{notif} < R^{notif} < r_0^{notif} + r^{notif} | R^{notif} \geq r_0^{notif} | X) \\
 &= \frac{P(r_0^{notif} < R^{notif} < r_0^{notif} + r^{notif} | X)}{P(R^{notif} \geq r_0^{notif} | X)} \\
 &= \frac{F(r_0^{notif} + r^{notif} | X) - F(r_0^{notif} | X)}{1 - F(r_0^{notif} | X)}
 \end{aligned}$$

Ainsi :

$$F(r_0^{notif} + r^{notif} | X) = U * (1 - F(r_0^{notif} | X)) - F(r_0^{notif} | X)$$

5.1.3 Estimation du type de couverture et du type de dommage des sinistres IBNYR

Estimation du type de couverture concernée et du type de dommage
générés par les sinistres:
Modèle Logistique

L'objectif est de déterminer le type de garantie mis en jeu par la réclamation et le type de dommage subi par l'assuré.

A partir des données disponibles depuis 2000, nous créons une variable *GD* qui est la combinaison de la garantie mise en jeu et du dommage subi. Voici les différentes classes de garanties et types de dommage utilisés :

Garanties	Types de dommage
RC Exploitation	Matériel
RC Produit	Perte financière
RC des employeurs	Corporel
RC du secteur médical	autre ou non précisé
autre ou non précisé	

Tableau 3: Liste des garanties et des types de dommages utilisés dans les modèles

Nous regroupons l'ensemble des contrats en 4 classes différentes suivant une classification interne propre aux données disponibles :

- Les contrats dont la base de garantie principale combine les garanties RC exploitation et RC Produit (classe de contrats A).
- Les contrats dont la base de garantie principale combine les garanties RC exploitation, RC Produit, et RC des employeurs (classe de contrats B).
- Les contrats sur les essais cliniques (classe de contrats C).
- Les contrats autres ou non précisés (classe de contrats D).

Le modèle estime donc pour chacune des quatre classes de contrats la probabilité de survenance d'un couple garantie/type de dommage à travers la variable *GD*.

Pour illustration, nous présentons ci-dessous les probabilités observées pour la classe de contrat A :

GD (type de dommage et garantie en jeu)	Fréquence	Pourcentage	Fréquence Cumulée	Pctage. Cumulé
Corporel / RC employeurs	188	4.7	188	4.7
Corporel / RC activités Médical	35	0.9	223	5.6
Corporel / RC Produit	304	7.7	527	13.3
Corporel / RC Exploitation	249	6.3	776	19.5
Corporel / Autre	4	0.1	780	19.6
Perte Financière / RC activités Médical	6	0.2	786	19.8
Perte Financière / RC Produit	374	9.4	1,160	29.2
Perte Financière / RC Exploitation	69	1.7	1,229	30.9
Perte Financière / Autre	169	4.3	1,398	35.2
Matériel / RC employeurs	2	0.1	1,400	35.2
Matériel / RC activités Médical	3	0.1	1,403	35.3
Matériel / RC Produit	1,164	29.3	2,567	64.6
Matériel / RC Exploitation	646	16.3	3,213	80.9
Matériel / Autre	41	1.0	3,254	81.9
Autre / RC employeurs	2	0.1	3,256	81.9
Autre / RC activités Médical	3	0.1	3,259	82.0
Autre / RC Produit	285	7.2	3,544	89.2
Autre / RC Exploitation	156	3.9	3,700	93.1
Autre / Autre	274	6.9	3,974	100.0

Tableau 4: Probabilités observés du couple garantie/type de dommages pour la classe de contrats A

Le modèle logistique (Tenenhaus, 2012) employé a les caractéristiques suivantes :

- Loi : Binomiale
- Fonction lien : Logit
- Variables explicatives :
 - Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
 - Variable de classe de la base de garantie du contrat (E en base survenance, C en base déclaration) : KZ_VERSAFLL (E / C)
 - Constante par valeur de *GD*

La même méthode est répétée pour les trois autres classes de contrats.

L'estimation par simulation se fait de manière simple : le modèle fournit les valeurs de probabilités cumulées de chaque couple garantie/type de dommage. Ainsi, en tirant une valeur d'une loi uniforme, on regarde à quel intervalle de probabilités cumulées cette valeur appartient, et à quelle catégorie cet intervalle se rapporte.

5.1.4 Modèle de paiements pour frais à l'ouverture des sinistres

Probabilité d'avoir un paiement pour frais à l'ouverture & Estimation du montant correspondant:
Loi de Bernouilli et Loi Log-Normale

L'objectif de cette partie est de savoir s'il y a ou non un paiement pour frais dès l'ouverture du sinistre et si oui, d'en estimer le montant.

En premier lieu nous utilisons une loi de Bernoulli basée sur les fréquences observées des paiements pour frais à l'ouverture des sinistres :

Payements pour frais	Fréquence	Pourcentage	Fréquence	Pctage.
non	12,250	96.7	12,250	96.7
oui	422	3.3	12,672	100.0

Tableau 5: Fréquence observée des paiements pour frais à l'ouverture des sinistres

Puis nous ajustons une loi normale sur le logarithme des montants de paiements pour frais réalisés à l'ouverture. Ceci revient à un modèle GLM avec les caractéristiques suivantes :

- Loi : Normale
- Fonction lien : Identité
- Variables explicatives :
 - Constante utilisée

Voici les principaux résultats de l'ajustement :

Paramètres	DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95%		Khi-2 de Wald	Pr > Khi-2
Intercept	1	6.7714	0.0571	6.7	6.8833	14,081	<.0001
Scale	1	1.6947	0.0404	1.6	1.7757		

Tableau 6: Modèle des montants de paiements pour frais à l'ouverture des sinistres - Paramètres Loi Log-Normale

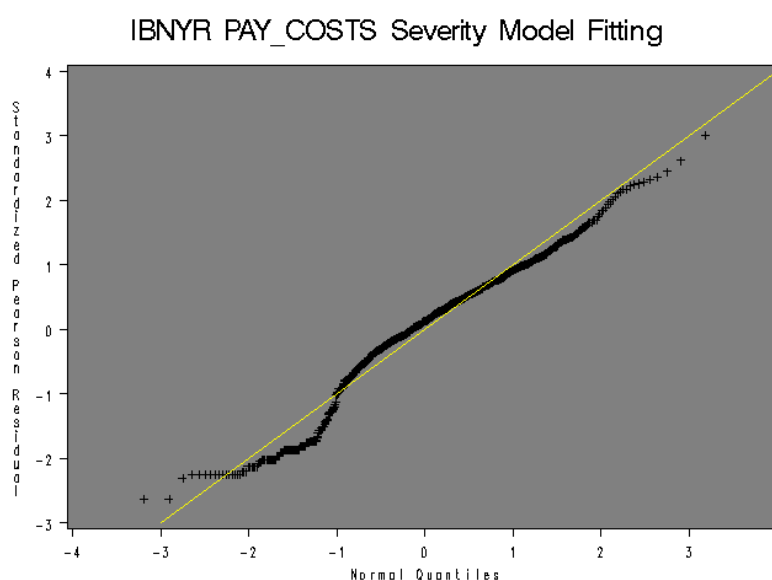


Figure 19: Modèle des montants de paiements pour frais à l'ouverture des sinistres - Résidus de Pearson standardisés

Ainsi en cas de paiement pour frais à l'ouverture, le montant est en moyenne de €3,700.

5.1.5 Estimation du montant de la charge indemnitaire à l'ouverture des sinistres

Estimation du montant de la charge indemnitaire à l'ouverture du sinistre:
Modèle Logistique et Loi Log-Normale via Modèle Linéaire Généralisé

L'objectif est d'estimer le montant de la charge indemnitaire « FGU », c'est-à-dire des paiements et des provisions brutes de franchises, à l'ouverture des sinistres.

A l'ouverture d'un sinistre, le régleur sinistre utilise les éléments à sa connaissance pour enregistrer le montant de charge indemnitaire. De manière très synthétique, il se trouve dans l'un des trois cas suivants :

- Les éléments de responsabilité et de dommage sont suffisamment clairs pour qu'il puisse enregistrer une estimation de charge indemnitaire adaptée.
- La mise en cause de l'assuré est faite sans qu'il soit possible d'estimer ni sa réelle responsabilité ni le montant des dommages consécutifs.
- Les éléments en sa possession sont intermédiaires entre les deux cas précédents.

Ainsi le panel des possibilités du régleur sinistre va aller de l'ouverture du sinistre à un montant conservatoire fixe, que nous qualifieront de forfaitaire dans cette étude, dans le simple but d'avoir une ouverture administrative du dossier sinistre tout en sachant que le montant indemnitaire enregistré alors n'a de justification technique qu'une moyenne à l'ultime de ce type de dossier. En connaissant quelques éléments supplémentaires, il peut aussi choisir un autre montant forfaitaire dans une grille préalablement établie ou en fonction de son expérience. Enfin à partir d'éléments tangibles qui peuvent être combinés à sa propre connaissance sur le type de sinistre concerné, il peut réaliser un enregistrement spécifique du montant de la charge indemnitaire.

D'un point de vue statistique, il apparaît donc que le montant de charge indemnitaire à l'ouverture des sinistres est une distribution continue, certes, mais où l'on observe des masses de Dirac non négligeables sur certaines valeurs correspondant aux montants forfaitaires utilisés.

Le modèle va prendre en compte ces aspects par la manière suivante : un modèle logistique va estimer la probabilité d'application des principales valeurs forfaitaires d'ouverture de sinistres observées, en y ajoutant une classe de valeurs repérant les cas où les valeurs ne sont pas forfaitaires. Les montants d'ouverture sont considérés comme forfaitaires lorsqu'ils ont une fréquence d'utilisation observée significative.

Montant indemnitaire d'ouverture (en €)	Fréquence observée	% Observé	Montant indemnitaire d'ouverture (en €)	Fréquence observée	% Observé
-1	4,309	14%	20,000	2,492	8%
1	865	3%	30,000	878	3%
500	2,291	7%	50,000	1,718	5%
1,000	2,647	8%	75,000	431	1%
2,500	4,410	14%	100,000	812	3%
5,000	4,623	15%	250,000	365	1%
7,500	982	3%	500,000	84	0.3%
10,000	3,334	11%	1,000,000	44	0.1%
15,000	1,026	3%	2,500,000	5	0.02%

Tableau 7: Fréquences observées des montants indemnitaires d'ouverture des sinistres

Le Tableau 7: Fréquences observées des montants indemnitaires d'ouverture des sinistres ci-dessus présente les montants forfaitaires retenus et leurs fréquences d'utilisation à l'ouverture. La valeur -1 permet de repérer les montants non-forfaitaires. Nous observons que seulement 14% des ouvertures de sinistres se font avec des montants indemnitaires non forfaitaires.

Estimation du montant de charge indemnitaire forfaitaire à l'ouverture

Le modèle logistique prend donc les caractéristiques suivantes :

- Loi : Binomiale
- Fonction lien : Logit
- Variables explicatives :
 - Variable de classe Type de garantie (voir le détail dans le Tableau 3: Liste des garanties et des types de dommages utilisés dans les modèles) : MAIN_TYP_BEZ_40
 - Logarithme du nombre de jours écoulés depuis la survenance du sinistre : LN_DAY_OCC
 - Logarithme du montant de paiements pour frais : LN_CUMUL_PAY_COSTS
 - Constante par valeurs présentes dans le Tableau 7: Fréquences observées des montants indemnitaires d'ouverture des sinistres

Le type de garantie mis en jeu par la réclamation est significatif alors que le type de dommage n'a pu être retenu dans ce sens. Le délai de déclaration est également vu comme significatif ainsi que le montant de paiement pour frais à l'ouverture. Dans ces deux cas, plus ces valeurs sont grandes, plus le montant forfaitaire de charge indemnitaire à l'ouverture sera important.

Le tableau de résultat de la régression est disponible en annexes : Tableau 29 page 137.

Estimation du montant de charge indemnitaire non forfaitaire à l'ouverture

Pour les cas d'ouverture sur la base de montants indemnitaires non forfaitaires, un modèle GLM (Duchesne, 2010) est utilisé avec les caractéristiques suivantes sur le logarithme de ces montants :

- Loi : Normale

- Fonction lien : Identité
- Variables explicatives :
 - Variable de classe Type de dommage (voir le détail dans le Tableau 3: Liste des garanties et des types de dommages utilisés dans les modèles) : MAIN_CAUSE_BEZ_40
 - Logarithme du montant de paiements pour frais : LN_CUMUL_PAY_COSTS
 - Constante utilisée

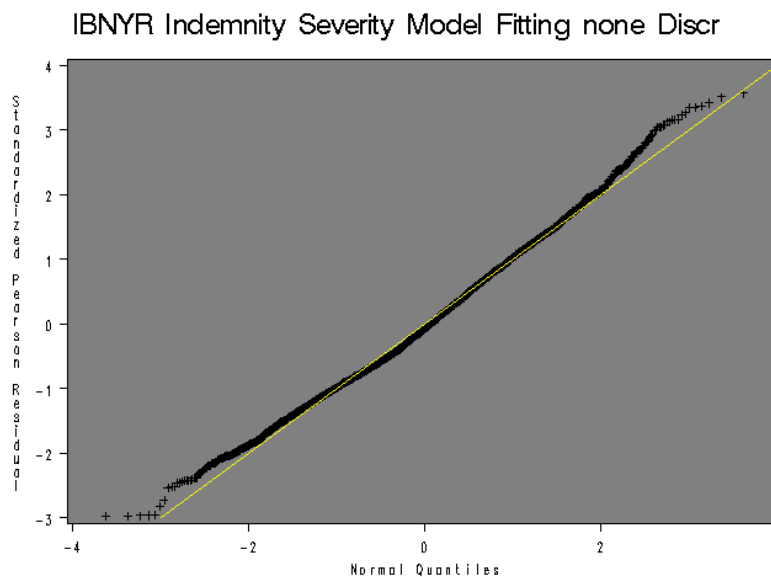


Figure 20: Modèle d'estimation du montant d'ouverture non-forfaitaire des sinistres - Résidus de Pearson standardisés

La valeur de référence utilisée dans la régression pour le type de dommage est « matériel ». Les types de dommages « corporel » et « pertes financières » apparaissent avec un montant d'ouverture plus important par rapport à la valeur de référence. A l'opposé, le montant de paiement pour frais à l'ouverture a tendance à faire baisser le montant d'ouverture indemnitaire.

Nous remarquons pourquoi le processus d'estimation du montant de paiement pour frais se trouve avant la partie indemnitaire : les frais apportent de l'information sur la partie indemnitaire que ce soit sur les montants forfaitaires ou non-forfaitaires.

Nous relevons que sur la partie forfaitaire, c'est le type de garantie mis en jeu qui apporte de l'information sur le montant indemnitaire d'ouverture, alors que sur la partie non-forfaitaire c'est le type de dommage qui apporte cette fois-ci de l'information.

5.1.6 Estimation du montant de paiement indemnitaire à l'ouverture des sinistres

Probabilité d'avoir un paiement indemnitaire à l'instant t0 et Estimation de son montant:
Modèle Logistique & Loi Béta

L'objectif est d'estimer s'il y a ou non un paiement indemnitaire dès l'ouverture d'un sinistre, et si oui, de quel montant.

Le modèle retenu ici est le même que celui utilisé ensuite dans le module 2 d'estimation des sinistres IBNER, c'est-à-dire concernant les sinistres déjà ouverts à la date d'analyse.

L'estimation se réalise en deux temps : tout d'abord, un modèle logistique va estimer s'il y a ou non un paiement indemnitaire en fonction des caractéristiques des sinistres. Puis un modèle va estimer le montant de paiement indemnitaire le cas échéant.

Estimation de la probabilité de paiement indemnitaire à l'ouverture

Un modèle logistique est utilisé pour cette partie avec les caractéristiques :

- Loi : Binomiale
- Fonction lien : Logit
- Variables explicatives :
 - Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
 - Logarithme du nombre de jours restant avant la clôture du sinistre : LN_DAY_BEF_CLOSED
 - Logarithme de la charge totale indemnitaire : LN_CUMUL_FGU_INDEMN
 - Variable de classe Au moins un paiement indemnitaire a déjà été réalisé sur le sinistre : FLAG_LAG_PAY_INDEMN
 - Variable de classe Au moins un paiement pour frais réalisé sur le sinistre : FLAG_COSTS
 - Variable de classe Facteur de développement de la charge indemnitaire du mouvement considéré égale à 1 (neutre) : FLAG_FD_1
 - Pas de constante

Nous remarquons que la probabilité de paiements indemnitaires augmente avec l'approche de la date de clôture. De même cette probabilité est plus importante lorsqu'il y a déjà eu un paiement indemnitaire auparavant ou un paiement pour frais. Le montant de charge indemnitaire exerce également une influence sur la probabilité de paiement, cette probabilité augmente avec le montant de charge indemnitaire, mais cette influence reste limitée. Enfin si le mouvement considéré a pour

conséquence un facteur de développement de la charge indemnitaire égale à 1, ceci augmente la probabilité d'avoir un paiement indemnitaire sur le mouvement en question.

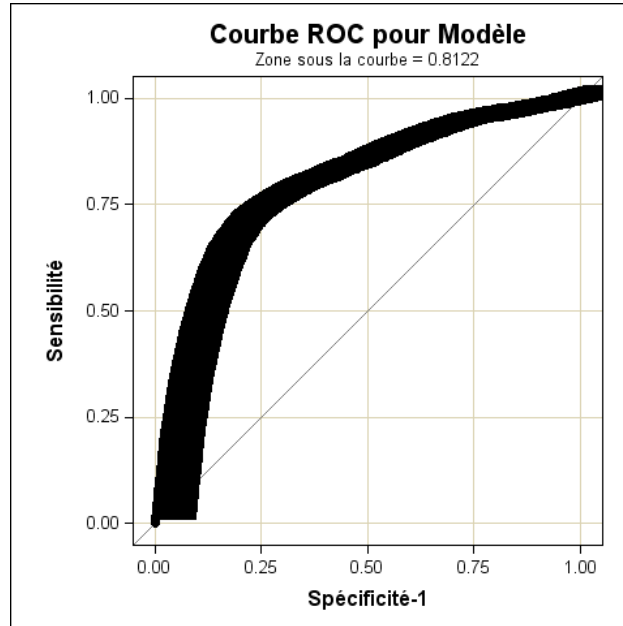


Figure 21: Modèle de survenance d'un paiement indemnitaire - Courbe ROC

La zone de la courbe ROC est supérieure à 0.8, ce que nous donne une prédictibilité du modèle acceptable.

Estimation du montant de paiement indemnitaire à l'ouverture

Concernant le montant du paiement en lui-même, nous posons que le cumul des paiements indemnitaires est un ratio de la charge totale indemnitaire. Si $r_0^{pay_indemn}$ est ce ratio de départ, nous cherchons alors la valeur $r_0^{pay_indemn} + r^{pay_indemn}$ tel que $r_0^{pay_indemn} \in [0; 1]$ et $r_0^{pay_indemn} + r^{pay_indemn} \in [r_0; 1]$. Par analogie avec l'estimation de $r_0^{notif} + r^{notif}$ au paragraphe 5.1.2, nous avons :

$$F(r_0^{pay_indemn} + r^{pay_indemn} | X) = U * (1 - F(r_0^{pay_indemn} | X)) - F(r_0^{pay_indemn} | X)$$

Dans le cas de l'analyse d'un sinistre à l'ouverture : $r_0^{pay_indemn} = 0$

Dans le cas de l'analyse d'un sinistre précédemment ouvert : $r_0^{pay_indemn} \in [0; 1]$

Pour modéliser cet aspect, nous avons utilisé un ajustement de loi Bêta par estimation du maximum de vraisemblance sur les ratios observés avec les caractéristiques suivantes (Smithson et Verkuilen, 2005) :

Paramètre	Symbole	Valeur estimée
Threshold	Theta	0
Scale	Sigma	1
Shape	Alpha	0.6145
Shape	Beta	0.3263
Mean		0.3469
Std Dev		0.3417

Tableau 8: Modèle d'estimation du ratio de paiement indemnitaire - Paramètres Loi Bêta

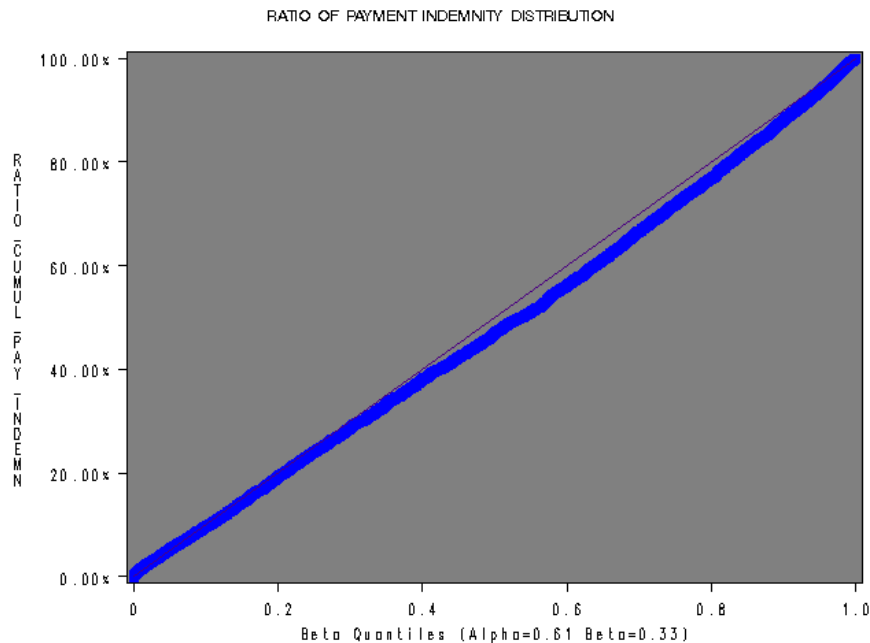


Figure 22: Modèle d'estimation du ratio de paiement indemnitaire - QQ-Plots

Un modèle introduisant des variables explicatives à la loi réponse Bêta a été testée, mais aucune des variables utilisées n'a apporté d'informations supplémentaires justifiant la pertinence de cette approche sur cette partie.

5.2 Modèle sur données détaillées – module 2 – Développement des sinistres déclarés

5.2.1 Estimation de la durée d'un sinistre - Modèle de survie de Cox

Estimation de la durée du sinistre:
Loi de Weibull via Modèle à hasard proportionnel de Cox

L'objectif est d'estimer le temps, en nombre de jours, s'écoulant entre la date de survenance d'un sinistre et sa date de clôture.

Pour cela nous avons utilisé le modèle à hasards proportionnels de Cox (Smith et Smith, 2002) avec les caractéristiques suivantes :

- Variable de stratification :

- Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
- Variables explicatives :
 - Variable de classe Type de dommage (voir le détail dans le Tableau 3: Liste des garanties et des types de dommages utilisés dans les modèles) : MAIN_CAUSE_BEZ_40
 - Variable de classe de la base de garantie du contrat (E en base survenance, C en base déclaration) : KZ_VERSAFL (E / C)

Dans un modèle de Cox, la stratification consiste à estimer une fonction de risque de base par valeur de la variable de classe utilisée pour la stratification ; les paramètres des autres variables explicatives restant identiques et utilisés en commun pour les différentes fonctions de base. Ici, nous avons stratifié la variable TYPE_C_USED car d'une part elle apporte une quantité d'information significative sur la durée de vie d'un sinistre, mais elle ne remplit pas suffisamment les critères de proportionnalité nécessaires à son utilisation comme simple variable explicative.

Les valeurs de références pour l'estimation du risque de base sont la base survenance (E) pour la variable base de garantie (KZ_VERSFALL), et « Autres/non précisé » pour le type de dommage généré par le sinistre (MAIN_CAUSE_BEZ_40).

Avant d'aller plus en avant, nous devons valider l'hypothèse de proportionnalité au cours du temps sur les deux variables explicatives afin de considérer le modèle comme acceptable. Nous allons valider graphiquement l'hypothèse de proportionnalité au cours du temps. Nous proposons d'utiliser deux méthodes : la première est graphique en regardant si les différentes courbes

$t = \ln(-\ln\hat{S}(t))$ pour chaque valeur d'une classe de variables sont parallèles, puis en vérifiant si les co-variables dépendant du temps de chacune des deux variables explicatives peuvent être considérées comme non significatives. Sur ce dernier point nous utilisons le logarithme du temps.

Pour la variable Base de garantie (KZ_VERSFALL) :

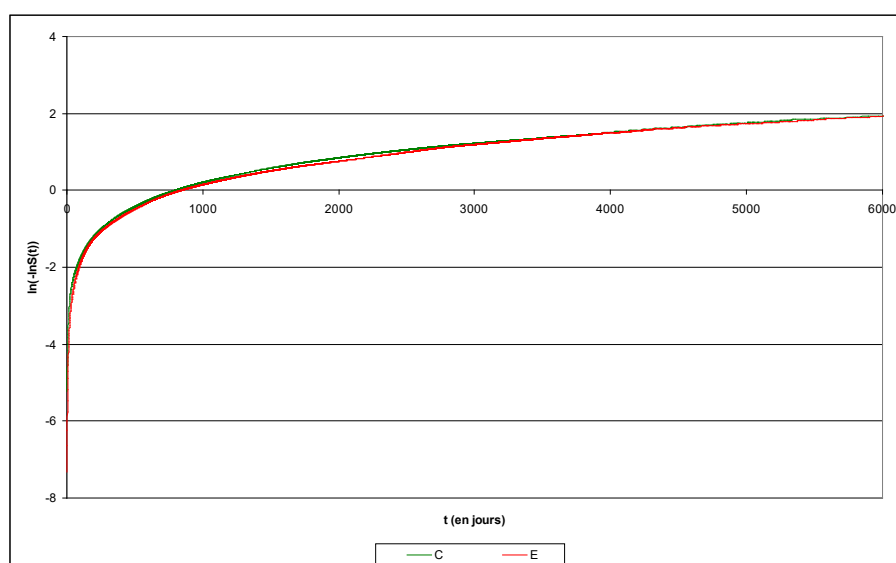


Figure 23: Estimation de la durée de vie des sinistres dans les comptes de l'assureur - Validation graphique de l'hypothèse de proportionnalité sur la variable Base de garantie

Graphiquement, les deux courbes semblent être parallèles, l'hypothèse de hasards proportionnels peut donc être considérée comme valide.

Paramètre		DDL	Valeur estimée	Erreur	Khi-2	Pr > Khi-2
KZ_VERSFALL	C	1	0.29477	0.02551	133.4909	<.0001
LOG_TIME*KZ_VERSFALL	C	1	-0.03548	0.0041	74.7646	<.0001

Tableau 9: Estimation de la durée de vie des sinistres dans les comptes de l'assureur – Validation de l'hypothèse de proportionnalité sur la variable Base de garantie par test de significativité de la covariable temps

Même si la valeur estimée du paramètre de la covariable temps n'est pas négligeable, nous la considérerons comme suffisamment faible pour rendre acceptable ce modèle.

Pour la variable Type de dommage (MAIN_CAUSE_BEZ_40). :

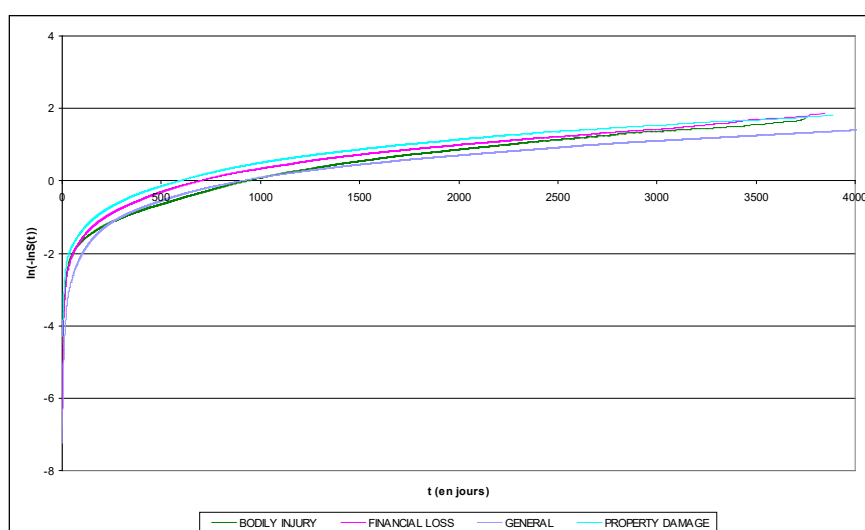


Figure 24: Estimation de la durée de vie des sinistres dans les comptes de l'assureur - Validation graphique de l'hypothèse de proportionnalité sur la variable Type de dommage

Même si graphiquement, il n'y a pas de grandes divergences, il n'est nécessaire de se reporter au deuxième test avant de conclure :

Paramètre		DDL	Valeur estimée	Erreur	Khi-2	Pr > Khi-2
MAIN_CAUSE_BEZ_40	Corporel	1	0.21073	0.04707	20.043	<.0001
MAIN_CAUSE_BEZ_40	Perte Financière	1	0.42023	0.03821	120.9252	<.0001
MAIN_CAUSE_BEZ_40	Matériel	1	0.93228	0.02931	1011.49	<.0001
LOG_TIME*MAIN_CAUSE	Corporel	1	-0.02077	0.00772	7.2356	0.0071
LOG_TIME*MAIN_CAUSE	Perte Financière	1	-0.02299	0.00637	13.0188	0.0003
LOG_TIME*MAIN_CAUSE	Matériel	1	-0.0872	0.00504	298.9637	<.0001

Tableau 10: Estimation de la durée de vie des sinistres dans les comptes de l'assureur – Validation de l'hypothèse de proportionnalité sur la variable Type de dommage par test de significativité de la covariable temps

Toutes les covariables dépendantes du temps ont des valeurs de paramètres inférieurs à 10% des variables correspondantes. Nous considérerons là aussi ces covariables comme suffisamment faible pour rendre acceptable le modèle.

Ainsi, nous considérons l'hypothèse de hasards proportionnels comme suffisamment respecter pour valider le modèle et poursuivre notre étude.

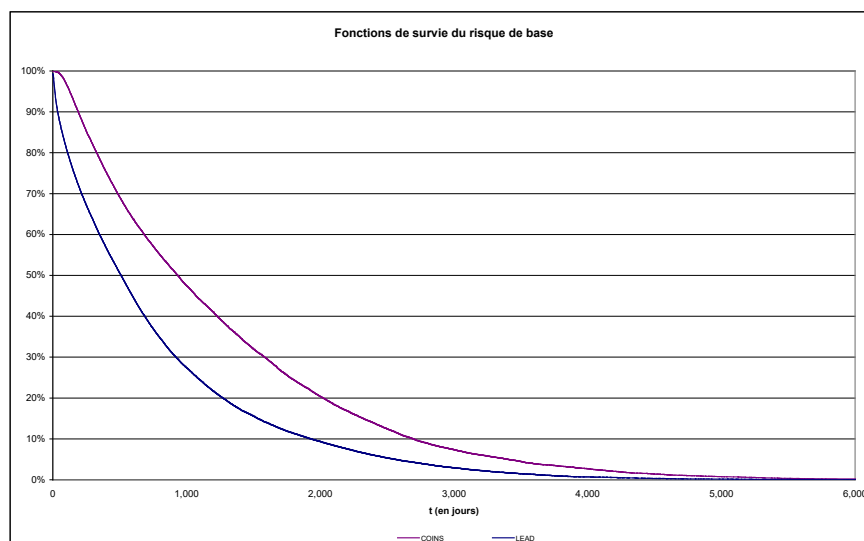


Figure 25: Estimation de la durée de vie des sinistres dans les comptes de l'assureur – Comparaison entre les risques de bases pour les sinistres en apériture et les sinistres en coassurance.

Il est intéressant de voir l'impact de la position de l'assureur, c'est-à-dire lorsqu'il se trouve apériteur ou coassureur sur le contrat concerné, sur la durée de survie du sinistre dans ses comptes. La position de coassureur apparaît comme très nettement défavorable en allongeant sensiblement la durée de survie des sinistres par rapport aux sinistres où il se trouve apériteur. Dans le portefeuille considéré, l'assureur semble capable de clôturer plus rapidement les dossiers que l'ensemble des autres apériteurs du marché. C'est un des résultats marquant de l'étude.

Nous pouvons voir aussi que les courbes convergent en queue de distribution signifiant que les dossiers sinistres les plus complexes semblent être traités avec les mêmes difficultés dans les deux cas.

Nous avons opté pour un ajustement paramétrique de la loi de survie du risque de base afin de pouvoir réaliser l'ensemble des simulations du modèle. Pour cela nous avons testé trois lois éligibles dans le cadre du modèle de Cox (Bender et al., 2005) :

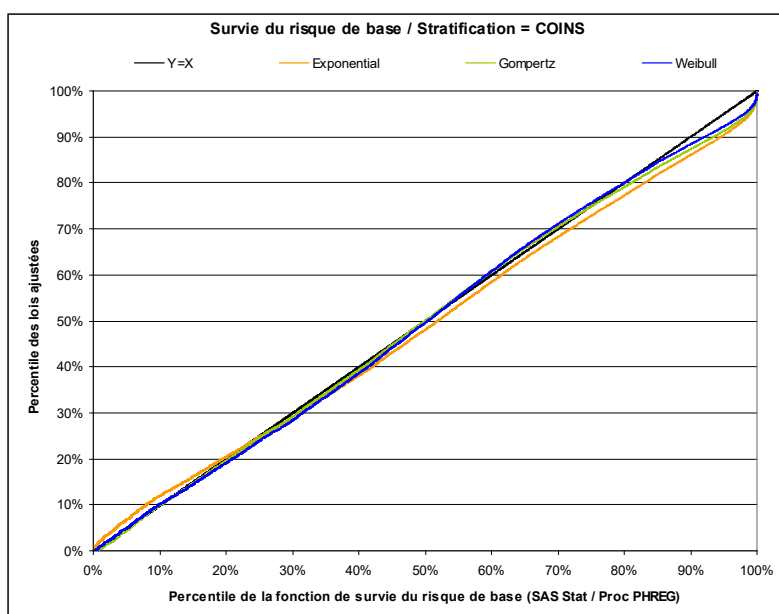


Figure 26: Estimation de la durée de vie des sinistres dans les comptes de l'assureur – Ajustement paramétrique Coassurance - PP-Plots

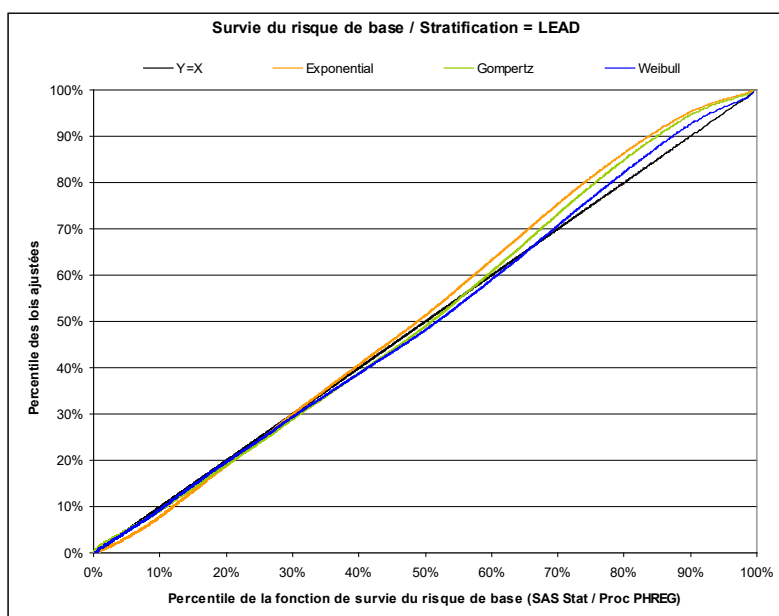


Figure 27: Estimation de la durée de vie des sinistres dans les comptes de l'assureur – Ajustement paramétrique Apérition - PP-Plots

D'après le graphe en Figure 27, nous avons retenu la loi de Weibull pour les deux risques de base avec les paramètres suivants après ajustement numérique des paramètres par la méthode des moindres carrées :

Paramètres Loi de Weibull sur modèle de Cox		
	Cas Apériteur	Cas Coassureur
Alpha	0.874	1.073
Bêta	722.216	1306.177

Tableau 11: Ajustement Loi de Weibull sur modèle de Cox – Paramètres estimées

5.2.2 Estimation du nombre de paiements pour frais et de leurs montants

Nombre des paiements pour frais jusqu'à la clôture et Estimation des montants correspondants:
Modèle Logistique & Modèle Linéaire Généralisé (Loi Log-Normale)

L'objectif est de savoir combien de paiements pour frais vont survenir depuis la date d'analyse jusqu'à la clôture et de quels montants.

Estimation du nombre de paiements pour frais restant jusqu'à la clôture

Pour l'estimation du nombre de paiements, nous avons observé le nombre total de paiements pour frais par sinistre, et nous avons utilisé un modèle GLM avec les caractéristiques suivantes :

- Loi : Binomiale négative
- Fonction lien : Log
- Variables explicatives :
 - Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
 - Logarithme du nombre de jours d'ouverture effective du sinistre dans les comptes de l'assureur : LN_DAY_NOTIF
 - Constante utilisée

Il apparaît logiquement que la durée de présence d'un sinistre dans les comptes de l'assureur a un effet majeur sur le nombre de paiements pour frais. Le positionnement de l'assureur a un effet moindre mais qui a été conservé dans le modèle retenu. Le montant des paiements indemnitaires réalisés jusqu'à cet instant sur le sinistre considéré apporte également de l'information dans une moindre mesure sur le nombre de paiements pour frais à venir.

Le tableau de résultat de la régression est disponible en annexes : Tableau 33 page 138.

Remarque : comme au point 5.1.1, voici la comparaison graphique entre les observations et les différents modèles testés. Comme précédemment, nous choisissons de retenir le modèle basé sur une loi binomiale négative.

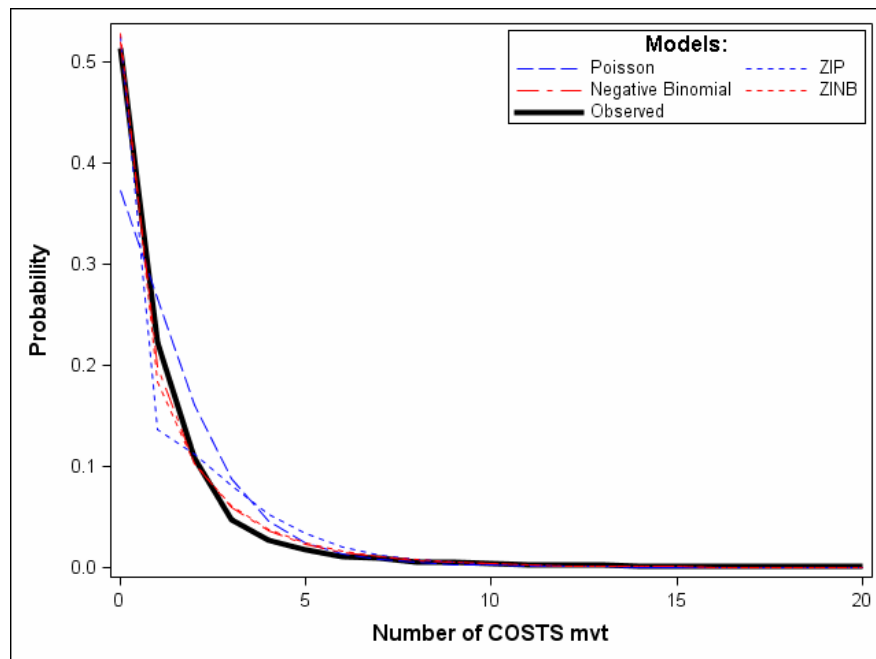


Figure 28: Estimation du nombre de paiements pour frais - Comparaison entre les modèles testés

Estimation des montants de paiements pour frais

Nous avons ensuite analysé le montant de ces paiements pour frais. Pour cela nous avons actualisé ces montants par l'indice de prix à la consommation (IPC) de l'INSEE pour se positionner au 31/12/2010. Puis nous avons utilisé un modèle GLM sur le logarithme de ces paiements pour frais indexés avec les caractéristiques suivantes :

- Loi : Normale
- Fonction lien : Identité
- Variables explicatives :
 - Variable de classe Type de dommage (voir le détail dans le Tableau 3: Liste des garanties et des types de dommages utilisés dans les modèles) : MAIN_CAUSE_BEZ_40
 - Logarithme du nombre de jours restant avant la clôture du sinistre dans les comptes de l'assureur : LN_DAY_BEF_CLOSED
 - Constante utilisée

Les résultats de la régression nous montrent que les montants de paiements pour frais diminuent au fur et à mesure que l'on se rapproche de la date de clôture. De plus, les pertes financières génèrent les montants de frais les plus importants, suivi des dommages matériels. Les dommages corporels génèrent des montants de frais un peu plus faible que les autres types de dommages.

Le tableau de résultat de la régression est disponible en annexes : Tableau 34 page 138.

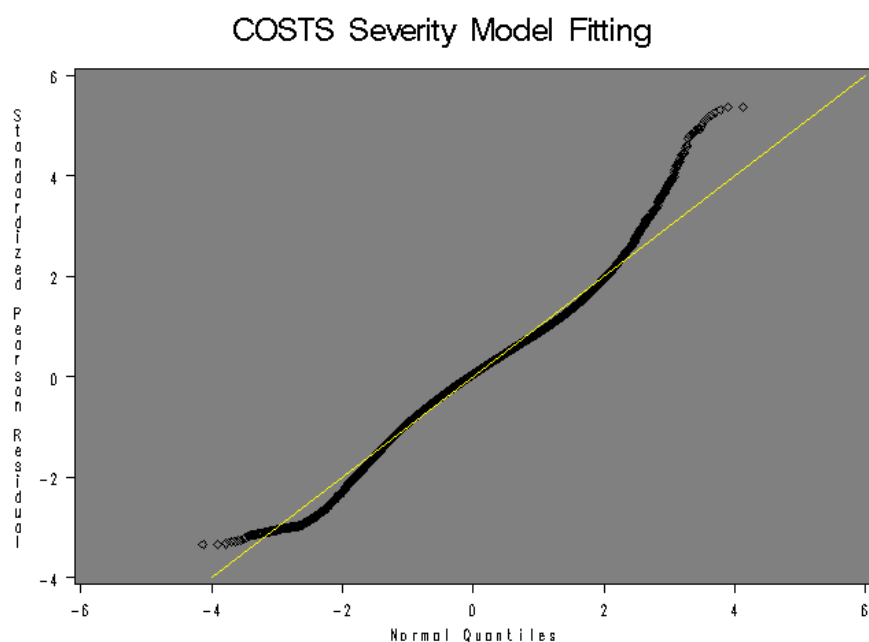


Figure 29: Estimation des montants de paiements pour frais - Résidus de Pearson

5.2.3 Répartition des paiements pour frais jusqu'à la clôture

Répartition dans le temps des mouvements pour frais jusqu'à la clôture:
Loi Bêta

L'objectif est d'estimer la répartition des paiements pour frais au cours de la présence des sinistres dans les comptes de l'assureur.

Pour cela nous avons observé la position de survenance des paiements pour frais durant la période de présence comptable du sinistre. Nous avons regardé cette position de manière relative, ainsi, nous avons considéré à quels pourcentages de la durée de présence comptable des dossiers sinistres, les paiements pour frais se positionnent. Ces survenances relatives sont donc comprises sur $[0; 1]$. Nous avons alors utilisé un ajustement d'une loi Bêta sur les observations dont voici les principaux résultats :

Paramètre	Symbole	Valeur estimée
Threshold	Theta	0
Scale	Sigma	1
Shape	Alpha	0.8561
Shape	Beta	1.3764
Mean		0.3835
Std Dev		0.2704

Tableau 12: Modèle d'estimation de la répartition des paiements pour frais - Paramètres Loi Bêta

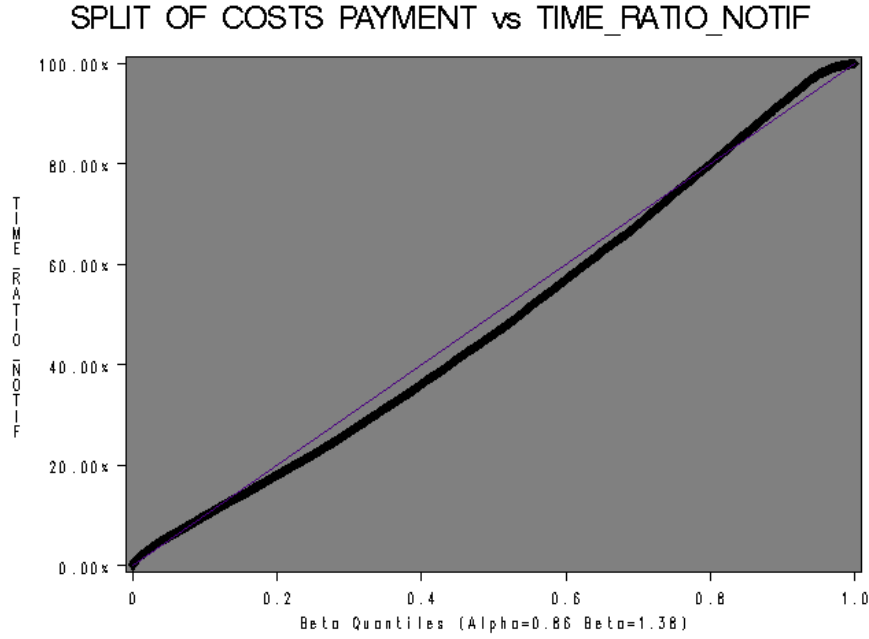


Figure 30: Modèle d'estimation de la répartition des paiements pour frais - QQ-Plots

Pour réaliser la simulation, nous regardons la durée de présence d'un sinistre dans les comptes de l'assureur comme un ratio compris entre 0 et 1. A la date d'analyse nous avons ce ratio qui prend la valeur $r_0^{time_open} \in [0; 1]$, et nous cherchons donc $r_0^{time_open} + r^{time_open} \in [r_0; 1]$. Par analogie avec l'estimation de $r_0^{notif} + r^{notif}$ au paragraphe 5.1.2, nous avons :

$$F(r_0^{time_open} + r^{time_open} | X) = U * (1 - F(r_0^{time_open} | X)) - F(r_0^{time_open} | X)$$

Dans le cas de l'analyse d'un sinistre à l'ouverture : $r_0^{time_open} = 0$

Dans le cas de l'analyse d'un sinistre déjà ouvert à la date d'analyse : $r_0^{time_open} \in [0; 1]$.

5.2.4 Estimation du nombre de mouvements indemnitaires et de leurs répartitions jusqu'à la clôture des sinistres

Nombres et répartition des mouvements indemnitaires jusqu'à la clôture:
Modèle Linéaire Généralisé et Regression Bêta

L'objectif est d'estimer le nombre de mouvements indemnitaires restant depuis la date d'analyse jusqu'à la clôture d'un sinistre, et d'en déterminer leurs répartitions. Nous ne cherchons ici qu'à connaître uniquement les mouvements indemnitaires aléatoires, c'est-à-dire, tous sauf ceux d'ouvertures et de clôtures qui sont considérés comme certains. Nous allons réaliser cette estimation en deux étapes :

Estimation du nombre de mouvements indemnitaires

D'abord nous allons estimer le nombre de mouvements indemnitaires aléatoires restant. Pour cela nous avons réalisé une modélisation GLM avec les paramètres suivants sur les observations :

- Loi : Binomiale négative
- Fonction lien : Log
- Variables explicatives :
 - Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
 - Logarithme du nombre de jours couverts par le contrat : LN_DAY_COVER
 - Constante utilisée

Comme pour le nombre de paiements pour frais, le nombre de mouvements indemnitaires restant, hors mouvement d'ouverture ou de clôture, a pour variable explicative la durée de présence du sinistre dans les comptes de l'assureur qui apporte une information significative, ainsi que le positionnement de l'assureur sur le contrat, mais dans une moindre mesure.

Le tableau de résultat de la régression est disponible en annexes : Tableau 35 page 139.

Remarque : comme au point 5.1.1, voici la comparaison graphique entre les observations et les différents modèles testés. Comme précédemment, nous avons choisi de retenir le modèle basé sur une loi binomiale négative.

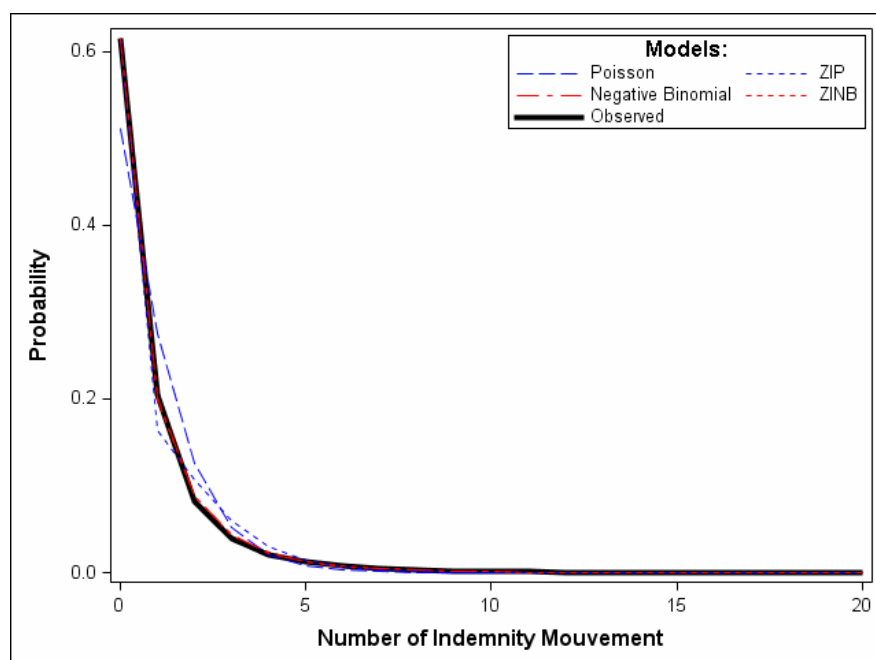


Figure 31: Estimation du nombre de mouvements indemnitaires - Comparaison entre les modèles testés

Répartition des mouvements indemnitaires restant jusqu'à la clôture

La durée d'ouverture d'un sinistre est relativisée entre 0 et 1, c'est-à-dire, lorsqu'elle est comprise entre sa déclaration et sa clôture. La répartition des mouvements indemnitaires est estimée par l'intermédiaire d'un model GLM étendu à la loi Bêta dont voici les principales caractéristiques :

- Loi : Bêta
- Fonction lien : voir explications du principe au point en annexes 10.1.4. page 127.
- Variables explicatives :
 - Variable de classe Positionnement de l'assureur sur le contrat (Apériteur ou Coassureur) : TYPE_C_USED (COINS / LEAD)
 - Constante utilisée

Comme la seule variable explicative est une variable de classe à deux valeurs, cela donne directement les deux lois Bêta s'appliquant à ces deux catégories issues du modèle.

Le tableau de résultat de la régression est disponible en annexes : Tableau 36 page 139.

Pour les sinistres pour lesquels l'assureur est l'apériteur, nous avons les paramètres suivants :

Paramètre	Symbole	Valeur estimée
Threshold	Theta	0
Scale	Sigma	1
Shape	Alpha	0.7561
Shape	Beta	0.8271
Mean		0.4776
Std Dev		0.3108

Tableau 13: Répartition des mouvements indemnitaires au cours de la durée d'ouverture des sinistres - Paramètres Loi Bêta - Cas Apériton

SPLIT INDEMNITY MOVEMENT vs TIME_RATIO_NOTIF — LEAD

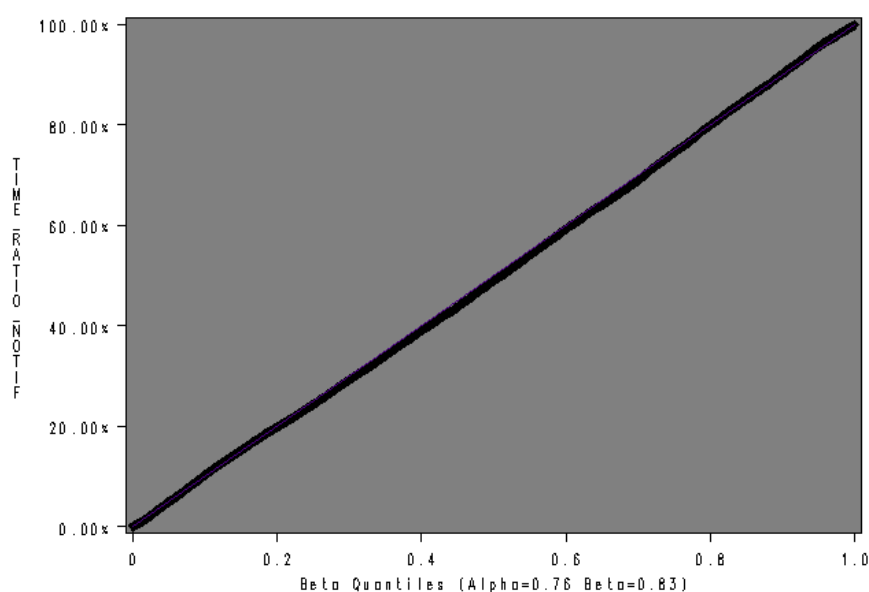


Figure 32: Répartition des mouvements indemnitaires au cours de la durée d'ouverture des sinistres — QQ-Plots - Cas Apériton

Pour les sinistres pour lesquels l'assureur est coassureur, nous avons les paramètres suivants

Paramètre	Symbole	Valeur estimée
Threshold	Theta	0
Scale	Sigma	1
Shape	Alpha	0.6435
Shape	Beta	0.8961
Mean		0.4180
Std Dev		0.3095

Tableau 14: Répartition des mouvements indemnitaires au cours de la durée d'ouverture des sinistres - Paramètres Loi Bêta - Cas Coassurance

SPLIT INDEMNITY MOVEMENT vs TIME_RATIO_NOTIF — COINS

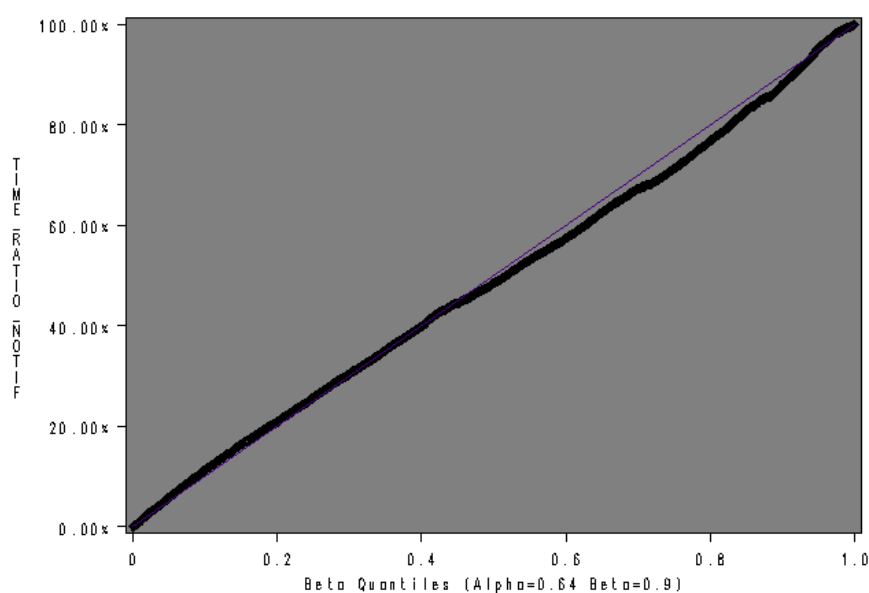


Figure 33: des mouvements indemnitaires au cours de la durée d'ouverture des sinistres — *QQ-Plots* - Cas Coassurance

La simulation se réalise exactement de la même manière que pour la répartition des paiements pour frais (voir la dernière partie du paragraphe 5.2.3).

5.2.5 Estimation de la probabilité des facteurs de développements à 0% ou 100%

Probabilité d'avoir un facteur de développement indemnitaire à 0% ou à 100%
Modèle Logistique

Nous souhaitons caractériser la probabilité des facteurs de développement de la charge indemnitaire à chaque mouvement indemnitaire. Les données disponibles indiquent que ces facteurs peuvent être divisés en trois classes distinctes :

- Lorsque le facteur de développement est égal à 0 ; c'est-à-dire lorsque la charge indemnitaire est simplement annulée.
- Lorsque le facteur de développement est égal à 1 ; c'est-à-dire lorsque la charge indemnitaire reste inchangée lors du mouvement considéré.
- Tous les autres cas qui nécessiteront une approche quantitative spécifique.

Ainsi nous allons caractériser ces 3 classes de valeurs par une modélisation logistique. Cette approche a dû être réalisée avec deux modèles distincts dans le cas où il y a eu un paiement indemnitaire avant le mouvement considéré, puis dans le cas contraire.

Estimation de la probabilité des classes de facteurs de développements – cas sans paiement indemnitaire au préalable

Un modèle logistique est utilisé pour cette partie avec les caractéristiques suivantes :

- Loi : Binomiale
- Fonction lien : Logit
- Variables explicatives :
 - Variable de classe de la base de garantie du contrat (E en base survenance, C en base déclaration) : KZ_VERSAFLL (E / C)
 - Logarithme du nombre de jours restant avant la clôture du sinistre : LN_DAY_BEF_CLOSED

Nous observons que lorsque le temps restant avant la clôture du sinistre diminue, la probabilité d'avoir un facteur de développement à 0 augmente. A la clôture, cette probabilité atteint environ 80%.

Le tableau de résultat de la régression est disponible en annexes : Tableau 37 page 139.

Estimation de la probabilité des classes de facteurs de développements – cas avec paiement indemnitaire au préalable

Un autre modèle logistique est utilisé pour cette deuxième partie avec les caractéristiques suivantes :

- Loi : Binomiale
- Fonction lien : Logit
- Variables explicatives :
 - Logarithme du nombre de jours restant avant la clôture du sinistre : LN_DAY_BEF_CLOSED

Seule la variable explicative représentant le temps restant avant la clôture s'est révélée significative dans cette régression. Elle a également pour influence de tirer vers les classes inférieures, mais avec une incidence assez faible puisqu'il y a très peu de sinistres ayant eu des paiements indemnitaires dont la charge indemnitaire s'annule par la suite.

Le tableau de résultat de la régression est disponible en annexes : Tableau 38 page 139.

5.2.6 Estimation de la charge totale indemnitaire

Estimation de la charge totale indemnitaire et des paiements indemnitaires:
Modèle Linéaire Généralisé & Loi Normale Contaminée

L'objectif est d'estimer la charge indemnitaire à l'issu du mouvement concerné et le montant cumulé des paiements indemnitaires. Nous avons vu précédemment que nous nous basons sur la caractérisation des facteurs de développement de la charge indemnitaire propre à chacun des mouvements. Ici nous regardons les cas où le facteur de développement n'est ni égale à 0 ou égale à 1, cela étant déjà caractérisé au préalable dans le modèle précédent.

Pour cela nous avons classé l'évolution de la charge indemnitaire en trois catégories :

- La charge indemnitaire précédant le mouvement considéré est inférieure à 10 000 euros.
- La charge indemnitaire précédant le mouvement considéré est supérieure ou égale à 10 000 euros, mais sans qu'il y ait encore eu de paiements indemnitaires.
- La charge indemnitaire précédant le mouvement considéré est supérieure ou égale à 10 000 euros, avec un cumul de paiement indemnitaire strictement positif.

Cas où la charge indemnitaire précédant le mouvement est inférieure à 10000 euros

Dans ce cas la valeur du facteur de développement est extrêmement volatile. Ceci s'explique simplement par le fait que sur des montants indemnitaires faibles, le facteur de développement n'a pas réellement de signification, car lors d'un mouvement indemnitaire sur un sinistre de faible importance, le régleur sinistre réalise plutôt une réévaluation pure et simple de la charge indemnitaire qu'une translation de la valeur précédente.

Ainsi, nous avons opté pour un ajustement d'une loi Log Normale sur la nouvelle charge indemnitaire générée par le mouvement considéré par estimation du maximum de vraisemblance.

Les résultats de l'ajustement sont les suivants :

Paramètre	Symbole	Valeur estimée
Mean	Mu	8.1149
Std Dev	Sigma	1.7589

Tableau 15: Estimation du montant de charges indemnitaires - Charges indemnitaires précédentes inférieures à 10000€ - Paramètres estimés par EMV

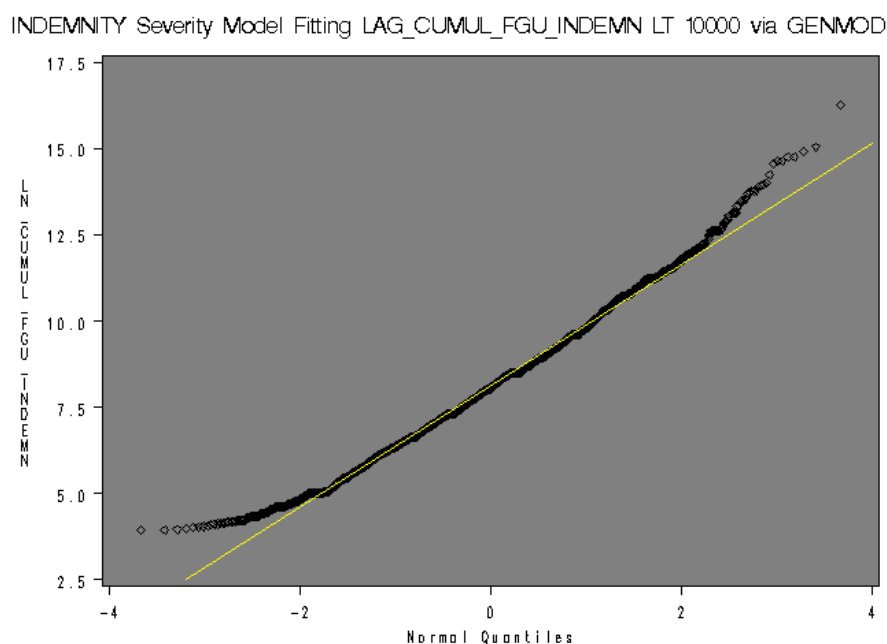


Figure 34: Estimation du montant de charges indemnitaires - Charges indemnitaires précédentes inférieures à 10000€ - *QQ-Plots*

Cas où la charge indemnitaire précédant le mouvement est supérieure ou égale à 10000 euros

Concernant les deux autres cas, l'origine de cette dichotomie s'explique sur l'observation des graphes suivants qui présente la densité observée du logarithme du facteur de développement en fonction du logarithme du nombre de jours restant avant la clôture. Un facteur de développement à 1 prend donc la valeur 0 ici ; un facteur de développement à 0 est représenté arbitrairement par la valeur -30, ainsi $LN_FD_FGU_INDEMN(0) = -30$. Le jour de la clôture d'un sinistre, le nombre de jours restant est donc égale à 0, est représenté par la valeur $LN_DAY_BEFORE_CLOSED(0) = 0$ par simplification également.

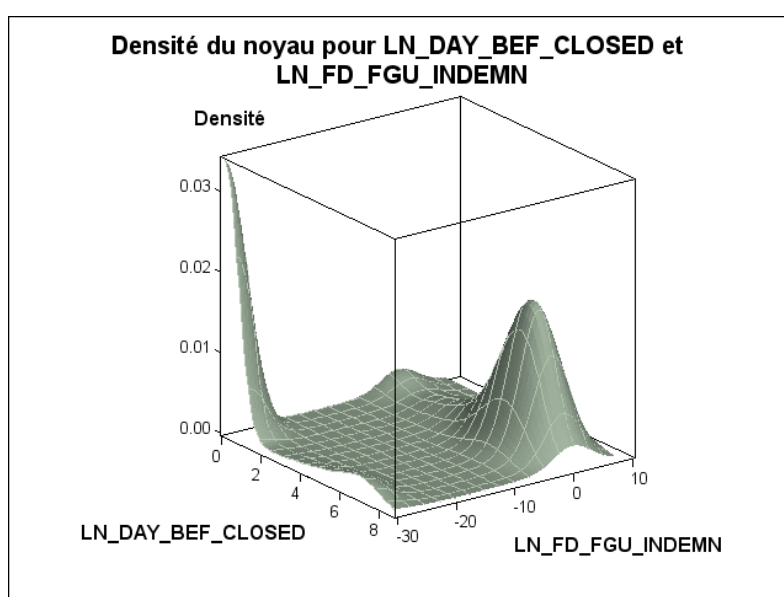


Figure 35: Densité observée du facteur de développement de la charge indemnitaire en fonction du nombre de jours restant avant la clôture – Cas sans paiements indemnitaires précédemment

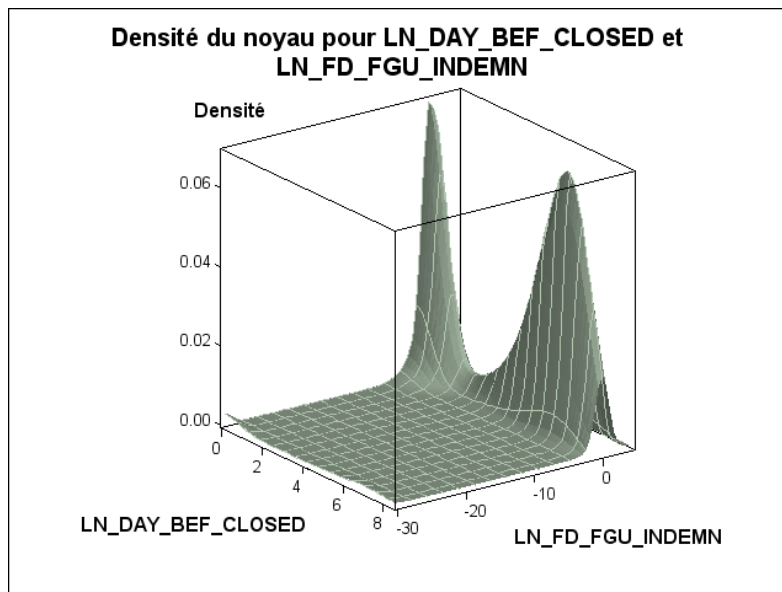


Figure 36: Densité observée du facteur de développement de la charge indemnitaire en fonction du nombre de jours restant avant la clôture – Cas avec paiements indemnitaires précédemment

La comparaison des deux graphes nous amène à deux constatations importantes :

- Au cours de la vie du sinistre la volatilité des facteurs de développements est beaucoup plus forte en l'absence de paiements indemnitaires que lorsqu'il y en a eu auparavant. L'explication est que le régleur sinistre ne verse un paiement indemnitaire que lorsqu'il a des éléments tangibles sur le coût attendu du sinistre, les modifications de charge indemnitaire pouvant survenir par la suite sont alors limitées en amplitude.
- A la clôture, en l'absence de paiements indemnitaires auparavant, il y a une forte proportion des sinistres qui se terminent avec un facteur de développement à 0, ce qui représente environ 80% des dossiers sinistres. L'observation au point précédent concernant les sinistres ayant eu des paiements indemnitaires reste valable : le facteur de développement à la clôture reste concentré autour de 1.

Nous avons donc utilisé deux modèles distincts pour caractériser ces deux catégories de comportements.

Cas où la charge indemnitaire précédant le mouvement est supérieure ou égale à 10000 euros et sans paiement indemnitaire préalable

Tout d'abord, nous avons utilisé un premier modèle GLM pour représenter le logarithme des facteurs de développement entre deux mouvements successifs pour les cas où il n'y a pas eu de paiements indemnitaires auparavant. En voici ses caractéristiques :

- Loi : Normale
- Fonction lien : Identité
- Variables explicatives :
 - Logarithme du nombre de jours restant avant la clôture : LN_DAY_BEFORE_CLOSED
 - Variable de classe prenant la valeur 0.5 avant la clôture et 1 à la clôture : TIME_FLAG_BIN

- Pas de constante

Ici la variable TIME_FLAG_BIN remplace la constante classique d'un modèle GLM afin d'utiliser deux constantes différentes pour chacun des cas représentés par les valeurs de la variable.

Nous pouvons remarquer que les révisions de charge indemnitaire au cours de la vie d'un sinistre sont tendanciellément à la baisse : l'espérance du facteur de développement indemnitaire d'un sinistre à 1000 jours de la clôture est égale à 0.74, et 0.46 à 100 jours de la clôture.

Le tableau de résultat de la régression est disponible en annexes : Tableau 39 page 139.

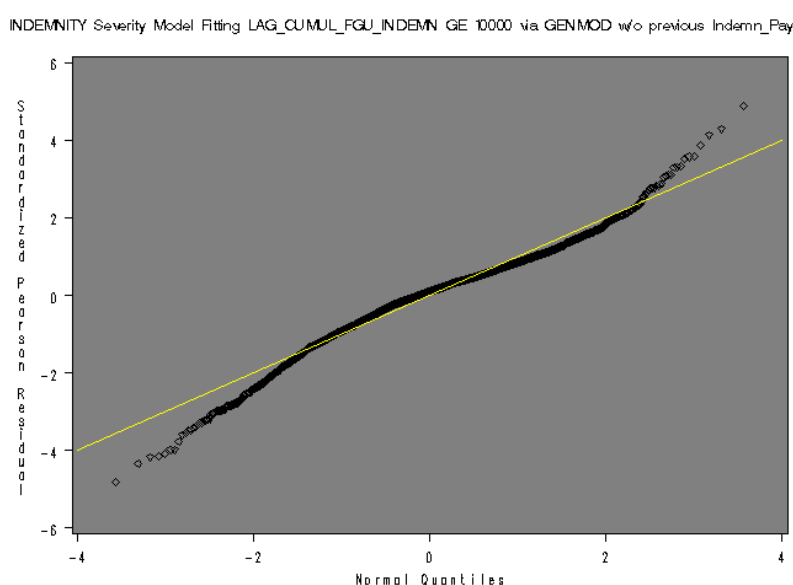


Figure 37: Estimation du facteur de développement de la charge indemnitaire - Cas sans paiements indemnitaires préalables – Résidus de Pearson

Cas où la charge indemnitaire précédant le mouvement est supérieure ou égale à 10000 euros et avec paiement indemnitaire préalable

Dans l'autre cas, lorsqu'il y a eu précédemment un paiement indemnitaire sur le sinistre considéré ; nous avons dans un premier temps appliqué un modèle GLM également sur le logarithme des facteurs de développement entre deux mouvements successifs avec les caractéristiques suivantes :

- Loi : Normale
- Fonction lien : Identité
- Variables explicatives :
 - Logarithme du nombre de jours restant avant la clôture : LN_DAY_BEFORE_CLOSED
 - Ratio du montant de paiements indemnitaires rapporté à la charge indemnitaire avant le mouvement considéré : LAG_RATIO_CUMUL_PAY
 - Constante utilisée

Le tableau de résultat de la régression est disponible en annexes : Tableau 40 page 140.

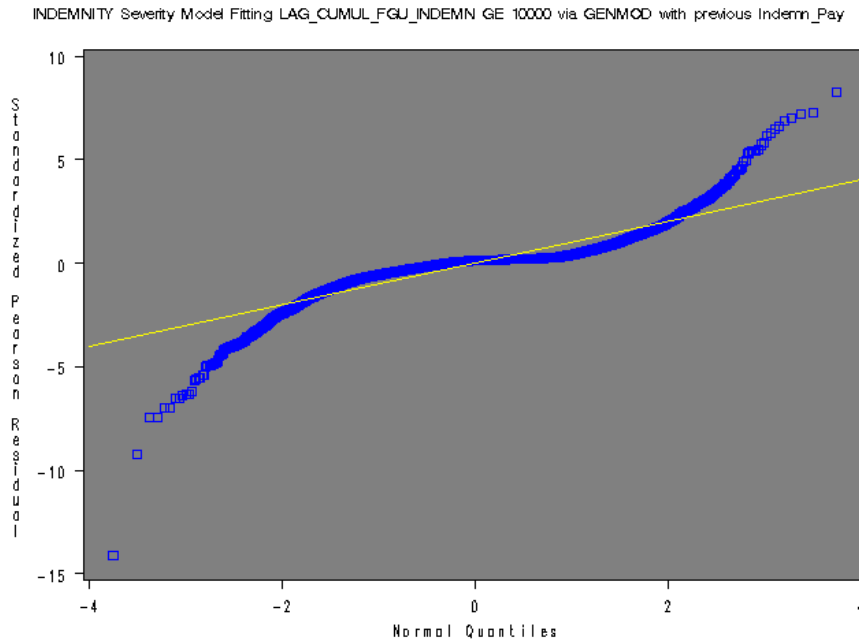


Figure 38: Estimation du facteur de développement de la charge indemnitaire - Cas avec paiements indemnitaires préalables – Résidus de Pearson

Le modèle obtenu ne peut être retenu en l'état. Les résidus observés sont concentrés autour de 0, mais il y a également un certain nombre de valeurs sensiblement dispersées.

Nous proposons d'utiliser une loi mélange connue sous le nom de Loi Normale Contaminée (voir l'annexe 10.1.5 page 129 pour son descriptif théorique). Cette loi est un mélange de deux lois normales ayant la même moyenne, mais aux écart-types différents. La régression de cette loi consiste alors à estimer ces deux écart-types ainsi que la probabilité d'être dans un cas ou dans un autre. Ici nous allons repartir des valeurs prédites par le GLM initial décrit ci-dessus et utiliser la PROC NLMIXED afin d'estimer les trois paramètres recherchés (Gilbert et Chen 2007).

Nous introduisons toutefois une variable explicative pour décrire la probabilité de se situer dans un cas ou dans l'autre de cette manière :

$$p = \frac{\exp(zp)}{1 + \exp(zp)} \text{ et } zp = \alpha X$$

Avec X qui est la variable explicative représentant le taux des paiements cumulés par rapport à la charge totale effectuée juste avant le mouvement considéré.

Des résidus du modèle basé sur la Loi Normale Contaminée (CND pour « Contaminated Normal Distribution en anglais) ont été estimés afin de les comparer avec les résidus du modèle GLM initial. Leur calcul a été fait ainsi :

$$r_i^{CND} = \hat{p}_{small} \frac{r_i}{\sqrt{\hat{v}_{small}}} + (1 - \hat{p}_{small}) \frac{r_i}{\sqrt{\hat{v}_{large}}}$$

Où r_i sont les résidus bruts du modèle GLM de base, $\hat{p}_{small}$ l'estimation de la probabilité d'être dans la loi Normale avec une faible volatilité, $\hat{V}_{small}$ et $\hat{V}_{large}$, les variances des deux lois normales avec l'une ayant une variance petite et l'autre ayant une variance grande.

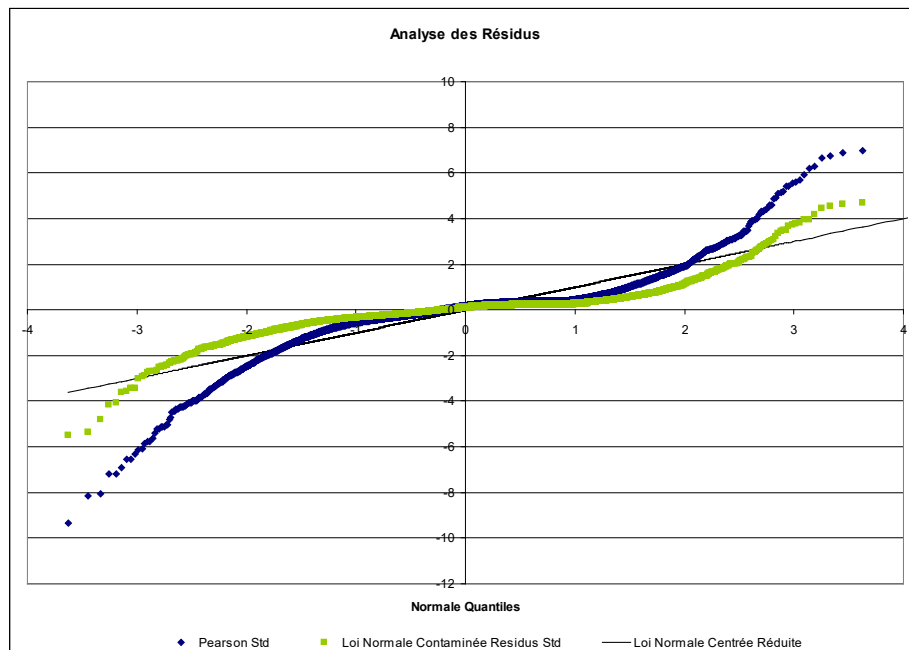


Figure 39: Estimation du facteur de développement le la charge indemnitaire - Cas avec paiements indemnitaires préalables – Comparaison des Résidus de Pearson et des Résidus avec la Loi Normale Contaminée

Les résidus extrêmes, à droite et à gauche, de la régression sur la CND dévient moins que pour la régression de base. Sur ces points-là, le modèle CND améliore les défauts du GLM de base. Sur le reste des résidus, les deux approches présentent des lacunes, mais il ne faut pas oublier que nous cherchons à caractériser un phénomène très concentré en 0 ($\ln(1)$), représentant un facteur de développement à 1, avec quelques cas, rares, mais très volatiles. Ainsi, nous optons pour conserver le modèle GLM modifié par une CND.

La probabilité moyenne estimée d'être dans le cas d'une volatilité faible est de 88%. Ainsi, dans le cas où un paiement indemnitaire a déjà été effectué précédemment, le facteur de développement entre deux mouvements successifs a 9 chances sur 10 d'être contenu dans une valeur très proche autour de 1.

5.2.7 Simulation du modèle détaillé

La simulation du modèle détaillé se déroule suivant la méthodologie présentée ci-dessous, une fois l'ensemble des paramètres des différents modèles estimés. Concrètement, 10,000 itérations ont été calculées par la méthode de Monte-Carlo pour estimer le développement futur des flux de trésorerie sur les 7 années suivant la date de reporting, et ce pour des raisons de capacités de calculs : celles-ci générant déjà plus de 90 millions de lignes dans la table de simulation. Cette table de résultats de simulation suit la logique d'une table de transaction comptable classique où chaque ligne représente une transaction pour un sinistre donné (paiement ou modification de provision), avec son propre ID le reliant à une police donnée, ayant sa propre caractéristique de type de préjudice (dommage matériel, financier ou corporel), et l'estimation de sa date de clôture. L'ensemble des transactions d'un sinistre donné permet de voir son déroulement de liquidation jusqu'à sa clôture estimée pour

une simulation. L'ensemble des simulations pour un sinistre donné permet d'avoir une estimation probabiliste de la liquidation d'un sinistre, où la date clôture, le nombre d'évènements comptables (ou transactions) entre la date de reporting et la date de clôture, et enfin les montants eux-mêmes à chaque transaction vont varier pour chaque simulation.

- Récupération de la dernière position des dossiers sinistres ouverts à la date d'analyse souhaitée.
- Récupération de l'historique d'exposition du portefeuille à la date d'analyse souhaitée.
- Simulation de l'ensemble des sous-modèles du module 1 en suivant le séquençement du schéma correspondant (Figure 13: Schéma de la modélisation des sinistres non ouverts (IBNYR). Ceci pour obtenir les sinistres survenus mais non déclarés à la date d'analyse.
- Simulation de l'ensemble des sous-modèles du module 2 en suivant le séquençement du schéma correspondant (Figure 14: Schéma de la modélisation du développement à l'ultime des sinistres). Ceci pour obtenir la chronologie comptable des sinistres non déclarés et déclarés jusqu'à leurs clôtures.
- Création d'une base des sinistres à l'ultime et estimation des différentes statistiques à l'ultime (distribution, espérance, écart-type)
- Création d'une base des sinistres à horizon 1 an et estimation des différentes statistiques à cet horizon (distribution, espérance, écart type). En remarque, nous précisons que seuls les sinistres qui ont été ouverts durant cet horizon d'1 an sont inclus dans cette partie ; les sinistres non déclarés simulés qui s'ouvrent au-delà de cet horizon ne sont donc pas pris en compte.

Afin de calibrer le nombre d'itérations nécessaires pour obtenir des résultats suffisamment robustes, nous avons observé l'évolution de la moyenne, de l'écart type, et de la Value at Risk à 99.5% de la charge totale sinistre de l'assureur en fonction du nombre d'itérations. Cette opération a été réalisée à l'ultime et à horizon 1 an.

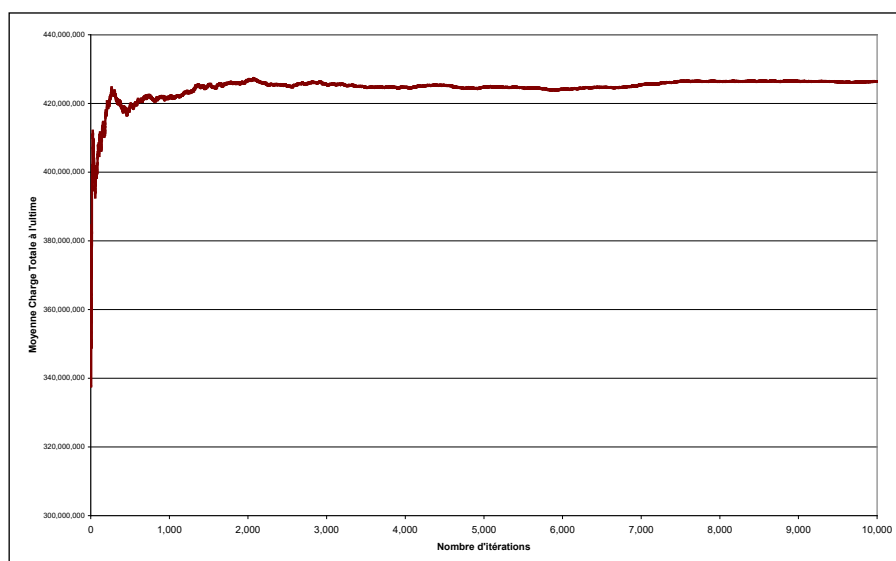


Figure 40: Modèle détaillé - Vitesse de convergence des simulations – Charge à l'ultime - Moyenne

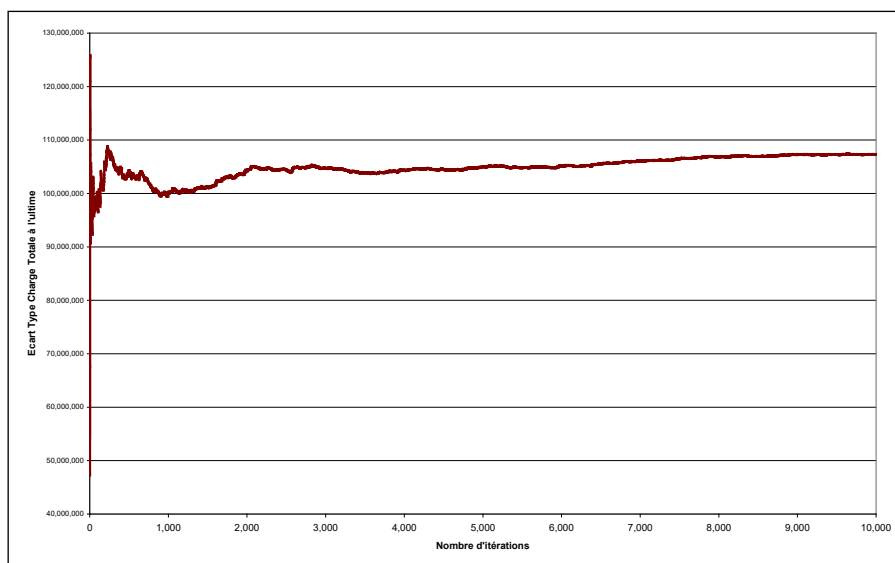


Figure 41: Modèle détaillé - Vitesse de convergence des simulations – Charge à l'ultime – Ecart Type

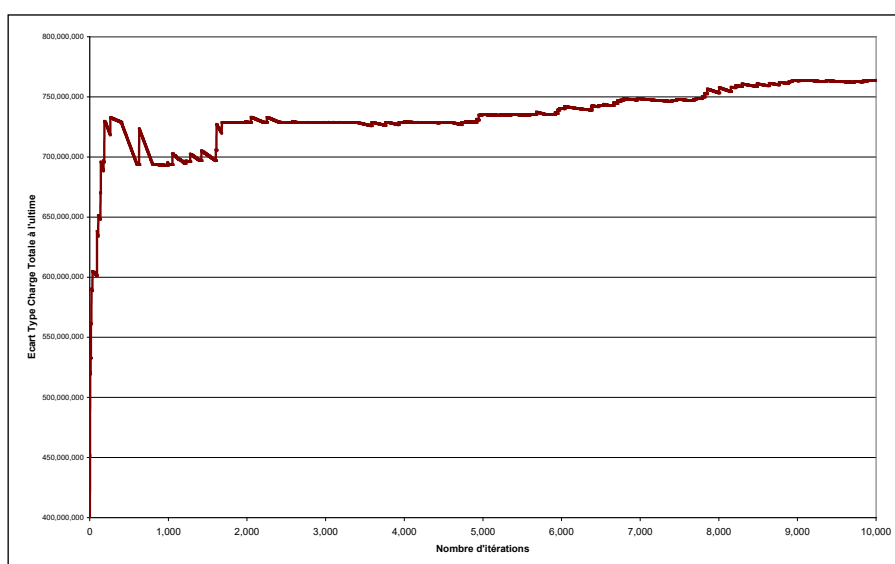


Figure 42: Modèle détaillé - Vitesse de convergence des simulations – Charge à l'ultime – VaR 99.5%

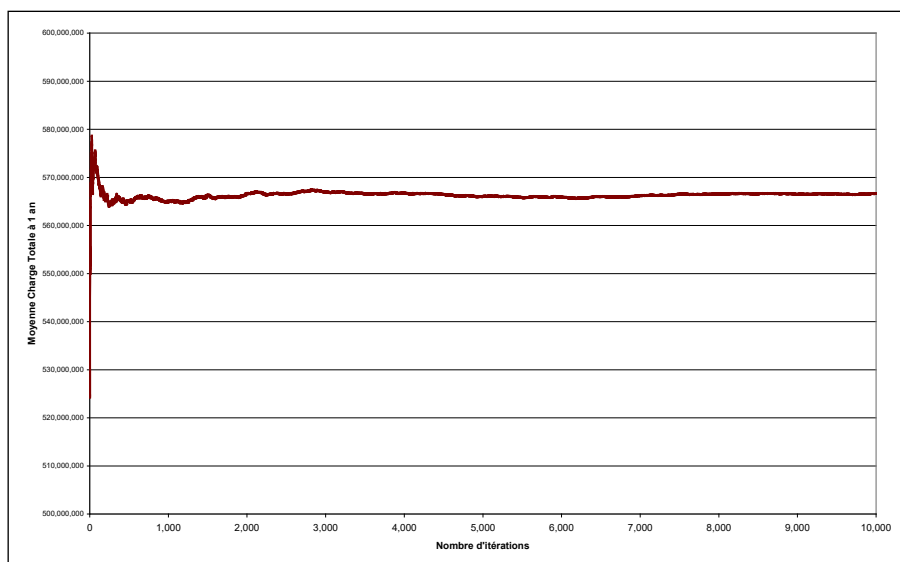


Figure 43: Modèle détaillé - Vitesse de convergence des simulations – Charge totale à horizon 1 an – Moyenne

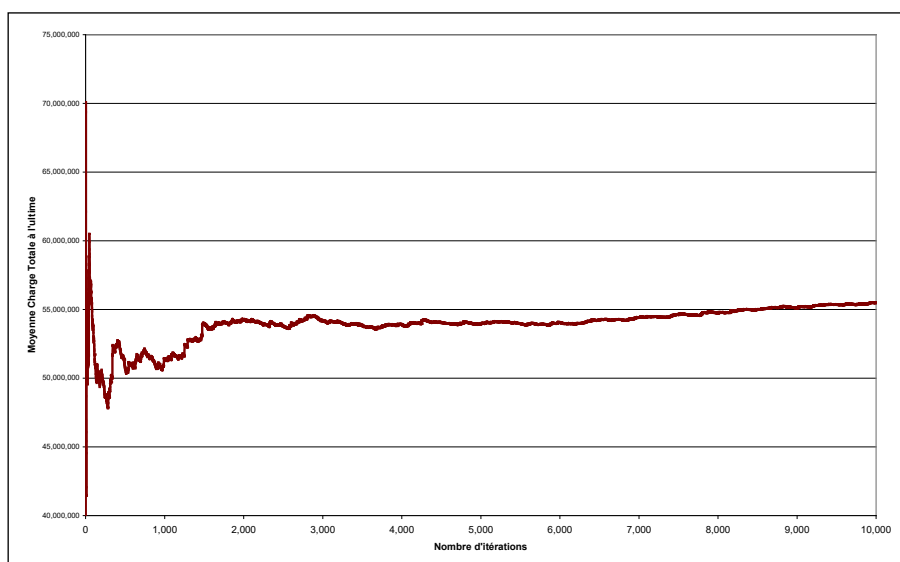


Figure 44: Modèle détaillé - Vitesse de convergence des simulations – Charge totale à horizon 1 an – Ecart Type

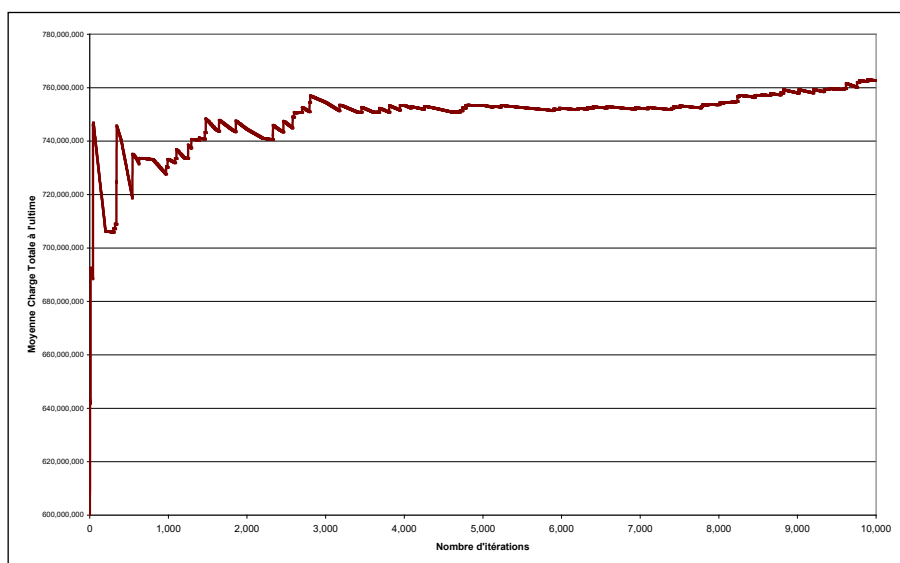


Figure 45: Modèle détaillé - Vitesse de convergence des simulations – Charge totale à horizon 1 an – VaR 99.5%

D'après les graphiques ci-dessus, les convergences de la moyenne et de l'écart type de la charge totale à l'ultime et à horizon 1 an peuvent être considérées comme satisfaisantes à partir d'une simulation sur 10 000 itérations. Concernant la Value at Risk à 99.5%, la stabilisation de cette statistique est plus difficile, même au bout de 10 000 itérations. Néanmoins, pour des raisons de contraintes techniques, nous considérerons les 10 000 itérations comme suffisantes pour réaliser la simulation du modèle détaillé, même si nous retiendrons que la précision de la VaR à 99.5% n'est pas encore optimale après 10 000 itérations.

5.3 Chain-Ladder

Les triangles utilisés pour la méthode Chain-Ladder et la méthode stochastique dérivée proposée par Mack sont :

- Le triangle des paiements cumulés par année de survenance et année de développement entre 1992 et 2010, soient 19 années.
- Le triangle des charges totales par année de survenance et année de développement sur la même période que le triangle des paiements cumulés.

La méthode Chain-Ladder ne pose pas de difficultés de mise en œuvre. L'ensemble des sinistres ont été utilisés, sans écarter les mouvements atypiques. Nous avons conservé les facteurs de développement bruts, sans chercher à leur appliquer de méthodes de lissage à travers les années de développement.

Voici les facteurs de développements calculés, et quelques statistiques complémentaires autour de ces facteurs :

Années de dvpt	Paiements cumulés			Charges Totales		
	Facteur de dvpts	Ecart-type Pondéré	Coefficient de variation	Facteur de dvpts	Ecart-type Pondéré	Coefficient de variation
1-2	4.078	1.864	45.71%	1.499	0.655	43.66%
2-3	1.666	0.405	24.30%	1.019	0.087	8.49%
3-4	1.530	1.219	79.67%	1.050	0.081	7.75%
4-5	1.116	0.092	8.24%	0.999	0.058	5.79%
5-6	1.095	0.075	6.86%	0.978	0.069	7.04%
6-7	1.084	0.059	5.41%	1.044	0.083	7.93%
7-8	1.056	0.062	5.91%	0.989	0.045	4.54%
8-9	1.046	0.038	3.59%	0.997	0.017	1.75%
9-10	1.034	0.041	3.97%	0.965	0.066	6.86%
10-11	1.015	0.006	0.61%	1.029	0.048	4.70%
11-12	1.064	0.112	10.52%	0.997	0.020	2.03%
12-13	1.011	0.016	1.54%	0.993	0.015	1.52%
13-14	1.018	0.020	1.98%	0.984	0.041	4.14%
14-15	0.988	0.023	2.35%	0.995	0.019	1.95%
15-16	0.999	0.006	0.56%	1.003	0.010	0.96%
16-17	0.997	0.004	0.45%	0.999	0.008	0.82%
17-18	1.001	0.000	0.01%	0.994	0.007	0.70%
18-19	1.004			1.001		

Tableau 16: Facteurs de développement des triangles de liquidation des sinistres – Méthode Chain-Ladder

Nous remarquons que les facteurs des trois premières années de développements sont moins volatiles sur le triangle des charges totales que sur celui des paiements cumulés. Comme du fait de la méthodologie Chain-Ladder, cette volatilité se propage à travers l'ensemble des estimations du triangle par multiplication successive des valeurs par les différents facteurs, nous pouvons donc conclure que les résultats issus des triangles de charges totales seront plus consistants que ceux issus des triangles de paiements cumulés.

Ceci est confirmé graphiquement par la représentation des nuages de points des trois premières années de développement comparées aux facteurs de développement correspondant sur les deux triangles :

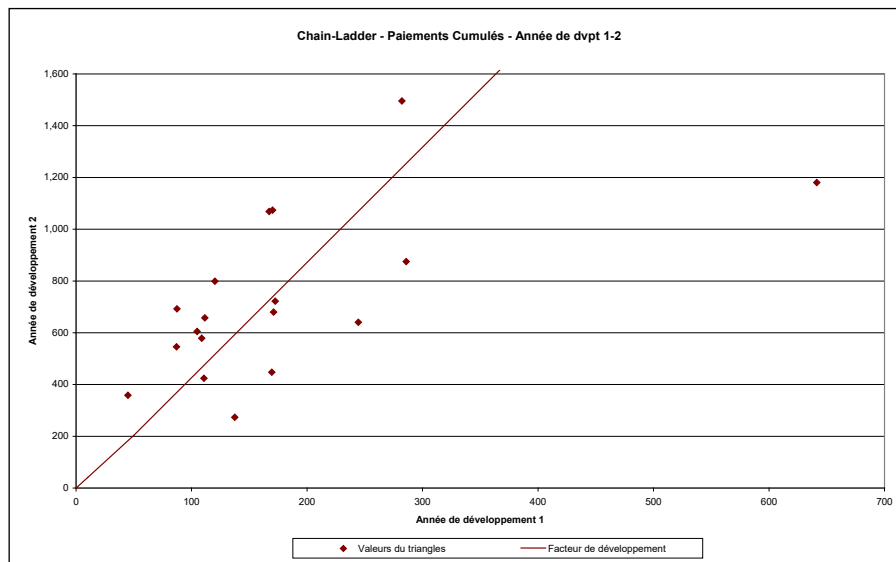


Figure 46: Chain-Ladder sur triangle des paiements cumulés – Nuage de points des années de développement 1 et 2

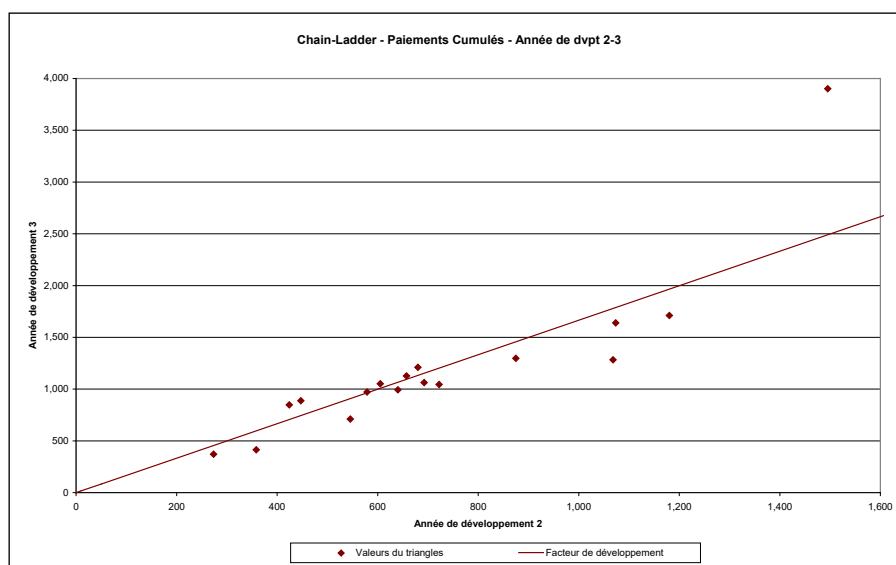


Figure 47: Chain-Ladder sur triangle des paiements cumulés – Nuage de points des années de développement 2 et 3

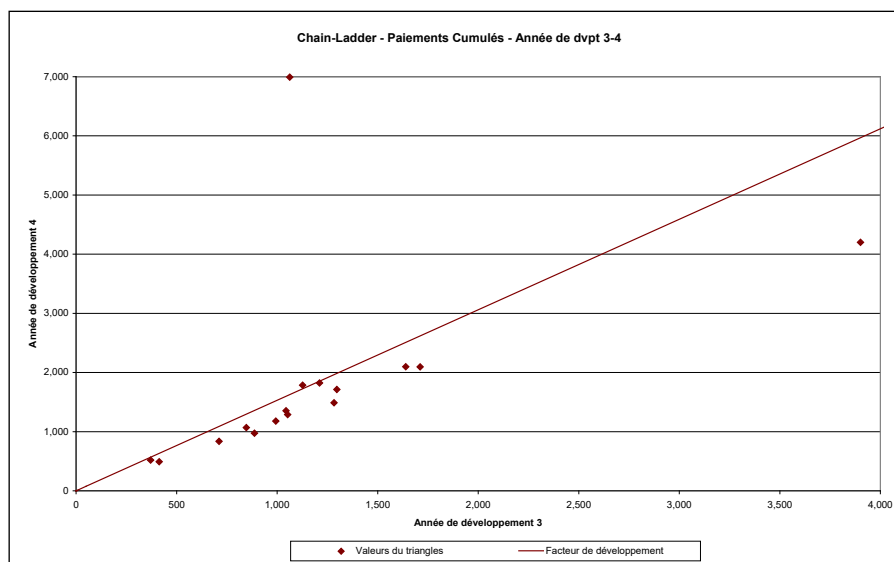


Figure 48: Chain-Ladder sur triangle des paiements cumulés – Nuage de points des années de développement 3 et 4

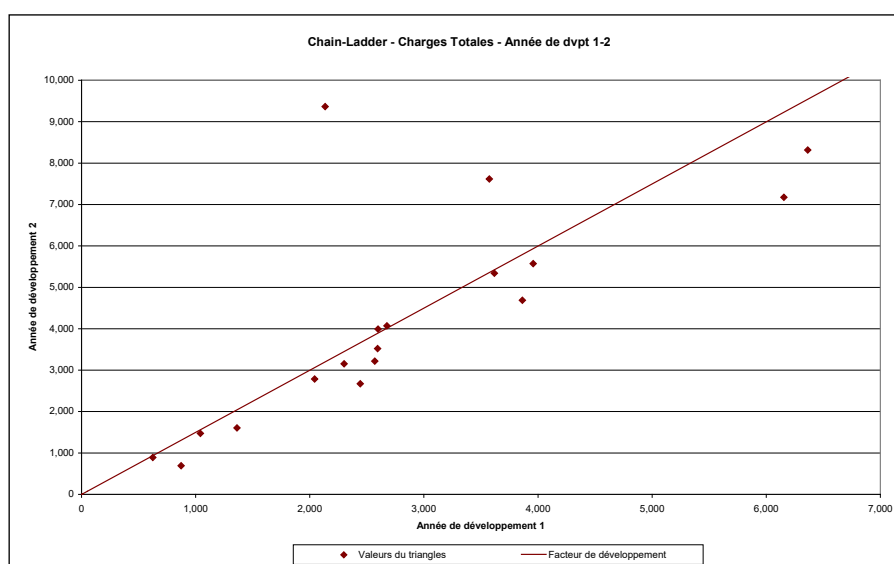


Figure 49: Chain-Ladder sur triangle des charges totales – Nuage de points des années de développement 1 et 2

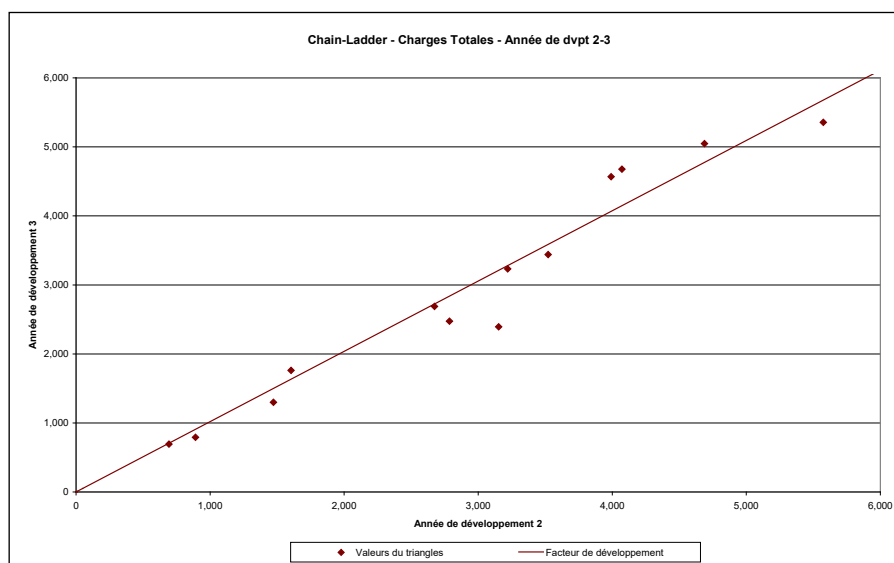


Figure 50: Chain-Ladder sur triangle des charges totales – Nuage de points des années de développement 2 et 3

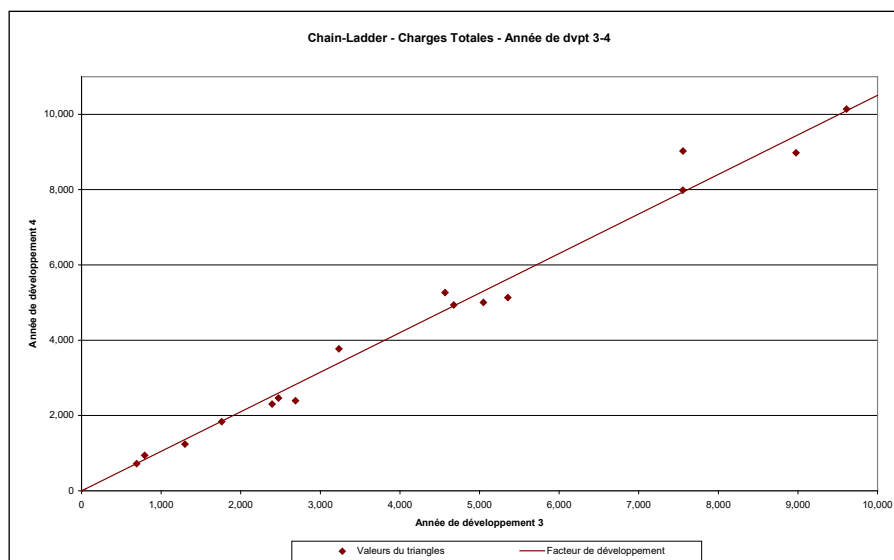


Figure 51: Chain-Ladder sur triangle des charges totales – Nuage de points des années de développement 3 et 4

Ainsi nous observons que les points des années 2 à 3 et 3 à 4 sont répartis de manière moins volatile autour du facteur de développement pour le triangle des charges totales que sur le triangle des paiements cumulés. Ceci nous permet de confirmer que le triangle des charges totales est plus stable que le triangle des paiements cumulés et indique que la fiabilité des résultats des triangles de paiements sera moindre au regard de ceux des triangles de charges totales.

5.4 Chain-Ladder stochastique I (Mack)

La méthode stochastique de Mack a été mise en œuvre avec quelques ajustements. La principale hypothèse posée pour son utilisation est le caractère Log-Normale des densités des années de survenance. Et comme certaines années sont apparues avec des valeurs R_i négatives, les ajustements suivants ont été appliqués avec comme objectif de conserver constant la somme $\sum_{i=1993}^{2010} R_i$ et donc la cohérence avec la méthode Chain-Ladder:

- Pour les paiements cumulés, l'année 1997 a été ajusté par le montant moyen observé durant les années précédentes, et la différence engendrée a été compensée proportionnellement sur les deux années suivantes
- Pour la charge totale, le même principe a été appliqué en annulant la période 1993-2003, puis compensée sur l'ensemble des années suivantes.

Année de survenance	Paiements cumulés					Charges Totales				
	RI	RI Adj	SEP(RI)	Paramètres Log Normale		RI	RI Adj	SEP(RI)	Paramètres Log Normale	
				alpha	beta				alpha	beta
1993	30,805	30,805	77	10.335	0.000	8,005	0			
1994	55,211	55,211	2,416	10.919	0.002	-48,348	0			
1995	34,852	34,852	87,651	8.476	1.991	-88,702	0			
1996	21,608	21,608	164,333	1.679	4.075	-42,889	0			
1997	-253,029	35,619	549,750	-9.747	6.296	-166,131	0			
1998	115,643	69,317	632,458	1.583	4.380	-471,822	0			
1999	604,914	362,591	1,053,871	10.415	2.197	-1,276,325	0			
2000	6,512,393	6,512,393	5,741,744	15.524	0.575	-2,103,463	0			
2001	3,854,830	3,854,830	3,641,448	14.961	0.638	828,351	0			
2002	3,619,765	3,619,765	3,191,749	14.936	0.575	-1,647,680	0			
2003	5,587,706	5,587,706	3,650,127	15.473	0.355	-845,414	0			
2004	8,424,332	8,424,332	4,574,951	15.913	0.258	2,640,584	2,464,586	8,638,963	11.372	2.587
2005	8,937,224	8,937,224	4,425,825	15.982	0.219	7,392,373	6,899,664	11,587,706	14.849	1.340
2006	22,371,330	22,371,330	7,504,472	16.918	0.107	13,806,175	12,885,980	18,221,168	15.768	1.098
2007	8,690,390	8,690,390	4,467,754	15.950	0.235	7,528,334	7,026,563	13,095,995	14.643	1.498
2008	14,665,136	14,665,136	19,196,935	16.003	0.998	6,859,812	6,402,598	11,020,375	14.724	1.377
2009	25,210,508	25,210,508	23,517,361	16.847	0.626	14,582,614	13,610,668	18,689,886	15.865	1.060
2010	36,189,212	36,189,212	26,642,234	17.310	0.433	35,026,998	32,692,413	23,649,102	17.214	0.421

Tableau 17: Paramètres Log Normale par année de survenance – Méthode de Mack

Ainsi chaque loi Log-Normale correspondant à une année de survenance est simulée par la méthode de Monte-Carlo avec 10 000 itérations.

5.5 Chain-Ladder stochastique II (GLM/Bootstrap)

La mise en œuvre du modèle Chain-Ladder stochastique combinant une modélisation GLM et la méthode du Bootstrap a été réalisée, non pas sur les montants des deux triangles utilisés par la méthode Chain-Ladder et la méthode de Mack, mais sur les facteurs de développement de ces deux triangles.

Les deux régressions ont été réalisées avec les caractéristiques suivantes :

- Loi : Poisson
- Fonction lien : Log
- Variables explicatives :
 - Variable de classe sur l'année de survenance : USED_I
 - Variable de classe sur l'année de développement : DEV_J
 - Pas de constante utilisée

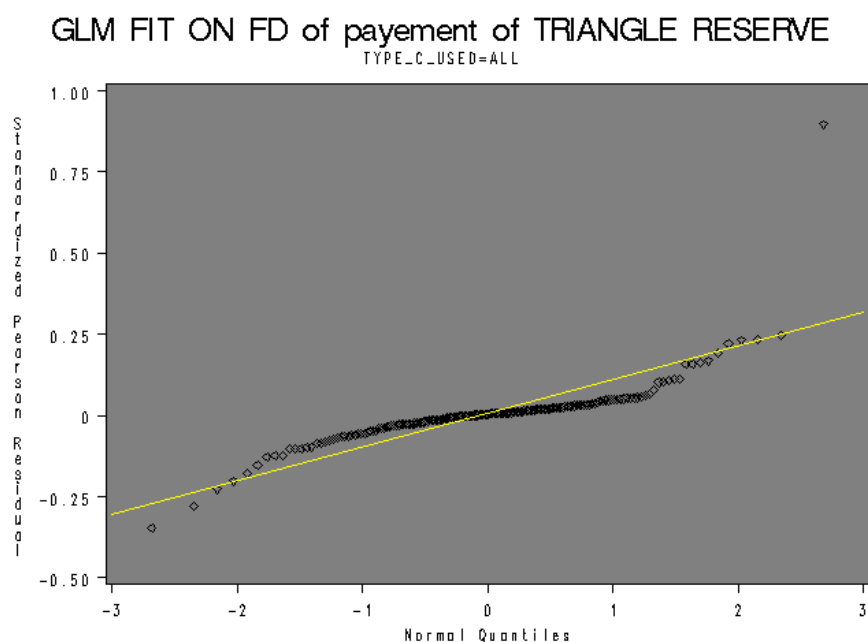


Figure 52: Régression GLM sur triangle des paiements cumulés – Résidus standardisés de Pearson

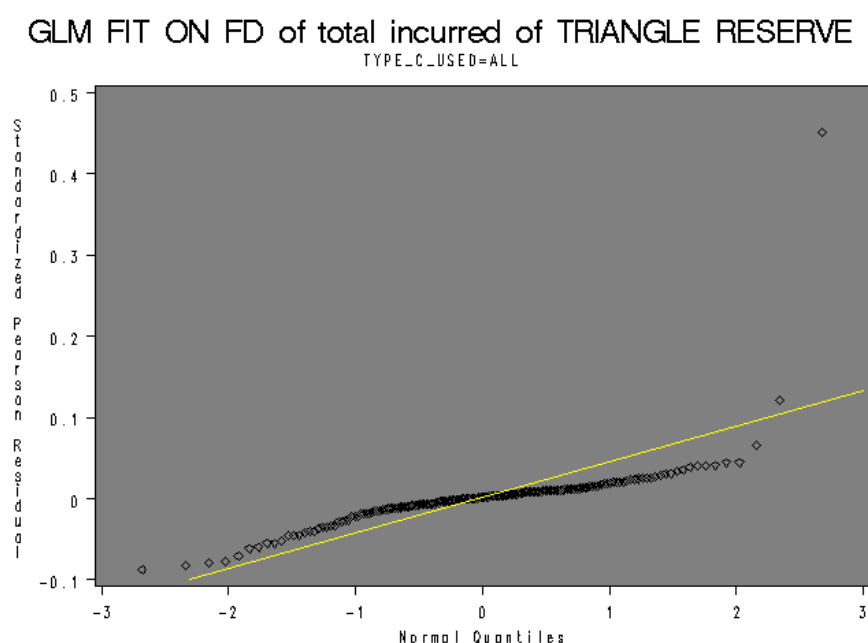


Figure 53: Régression GLM sur triangle des charges totales – Résidus standardisés de Pearson

Dans les deux cas, un ou plusieurs points se retrouvent très nettement décalés par rapport à l'alignement attendu de ces résidus. Ceci pourrait rendre difficile de retenir ces modèles pour une utilisation classique des paramètres issus de la régression, mais ici nous utilisons la méthode du Bootstrap et non les paramètres estimés directement. Ces points atypiques vont alors transmettre leurs décalages sous forme de volatilité dans les résultats de la simulation du Bootstrap.

Ces deux modèles vont nous servir à regarder les charges à l'ultime par année de survenance et également les montants à 1 an afin de regarder les résultats suivant les critères de Solvabilité 2.

10 000 simulations ont été réalisées pour obtenir les résultats globaux.

6 Résultats et impact dans l'environnement de Solvabilité II

Cette section présente les résultats des différents modèles sous plusieurs angles. Tout d'abord nous regarderons les résultats à l'ultime, puis les résultats à horizon 1 an dans l'optique de Solvabilité II.

Les résultats présentés vont se concentrer sur le montant de capital requis pour couvrir les flux futurs des sinistres des engagements du portefeuille à la date de reporting, ici le 31.12.2010, à travers les trois modèles stochastiques utilisés pour cette étude :

- Modèle de Mack sur triangles agrégés des paiements et des charges totales
- Modèle GLM à Bootstrap sur triangles agrégés des paiements et des charges totales
- Modèle sur données détaillées

Quelques commentaires sur les résultats présentés :

- Comme indiqué page 82, le modèle détaillé estime concrètement le développement des 7 années suivantes la date de reporting des flux de paiements et de modifications de provisions des sinistres.
- En toute transparence et comme indiqué en introduction, les données utilisées ne sont plus disponibles, et les résultats qui ont pu être sauvegardés et présentés ci-dessous sont limités aux principaux résultats agrégés des montants de provisionnements.

6.1 Résultats à l'ultime

Voici les résultats des simulations des différents modèles utilisés de l'estimation des provisions techniques nécessaires à l'ultime, en excédent des paiements déjà effectués.

	C-L/Mack Paiements	GLM/Bootstrap Paiements	C-L/Mack Charge Totale	GLM/Bootstrap Charge Totale	Modèle détaillé
Espérance	146	263	461	434	426
Ecart Type	162	188	43	105	107
Probabilité					
5.0%	105	33	417	305	268
10.0%	111	70	423	326	298
20.0%	120	116	431	355	333
50.0%	139	229	451	415	417
80.0%	166	381	482	499	511
90.0%	184	494	506	564	567
95.0%	201	611	532	631	620
99.0%	248	905	617	783	711
99.5%	278	1,010	659	851	763
99.8%	314	1,301	736	935	810

Tableau 18: Distributions des provisions nécessaires à l'ultime simulées

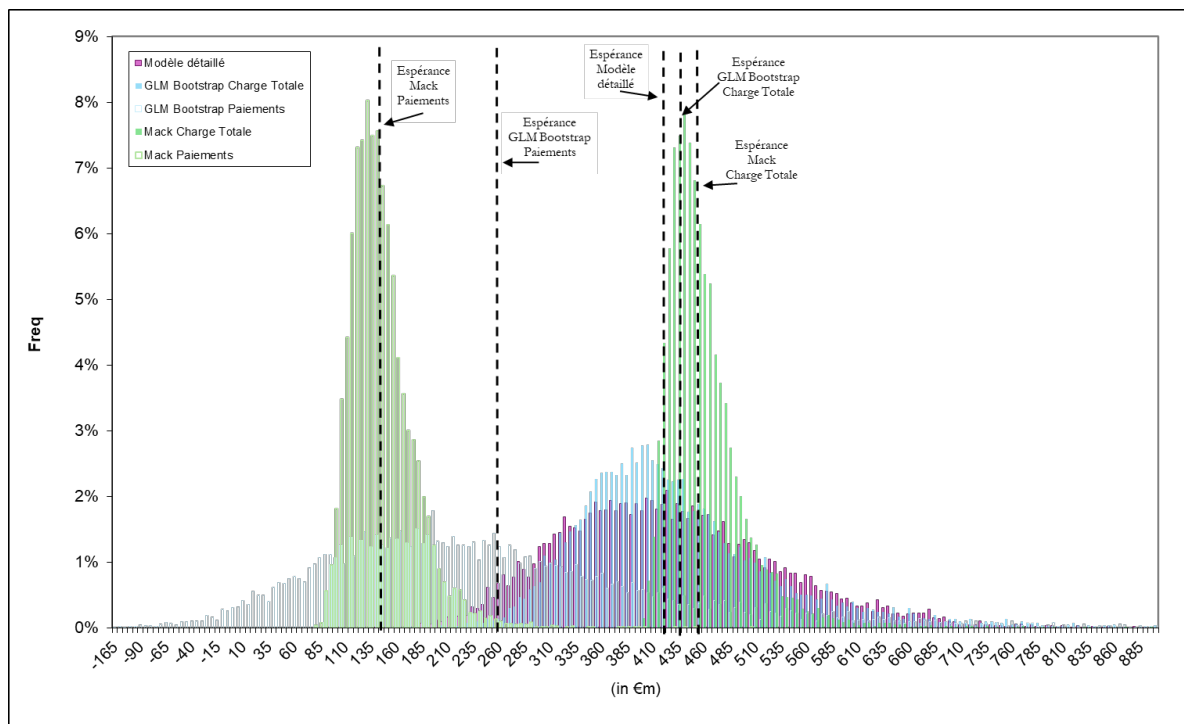


Figure 54: Densités des provisions nécessaires à l'ultime simulées

Il convient tout d'abord d'appréhender ces résultats au regard de la provision techniques des dossiers ouverts à la date de reporting au 31/12/2010, soit €380m.

Une observation immédiate, qui sera commentée plus en détail dans la section suivante, concerne la proximité des espérances des modèles basés sur la charge totale par opposition aux résultats des modèles basés sur les paiements qui sont plus dispersés et même inférieures aux provisions techniques à la date de reporting.

6.1.1 Résultats à l'ultime – pertinence des modèles analysés

On observe ainsi que les modèles agrégés basés sur les triangles de paiements donnent des espérances de provisions substantiellement inférieures aux approches sur les triangles de charges totales et du modèle détaillé, et aussi de la provision initiale des dossiers, laissant suggérer des bonis de liquidations substantielles. Mais il faut garder à l'esprit que les triangles des paiements ne considèrent pas, par construction, l'information des provisions existantes, et donc ne prennent en compte que la cadence des paiements. Cette caractéristique peut être recevable dans le cadre d'un portefeuille pour lequel les paiements de sinistres génèrent des flux réguliers au niveau de granularité choisi pour construire les triangles. Mais ici cela n'est pas le cas car les polices sous-jacentes ne sont pas homogènes du fait de la diversité des garanties, des activités des assurés, et même de la variété des juridictions des expositions. Ceci s'exprime aussi au travers d'une volatilité plus conséquente de ces modèles.

Un autre point notable est la proximité des résultats, en espérance, des modèles agrégés de charges totales et du modèle détaillé. On peut même constater cette proximité sur l'ensemble de la distribution entre le modèle GLM de charge totale et le modèle détaillé. Tandis que le modèle de Mack de charge totale donne une volatilité bien moindre que tous les autres modèles. L'objectif est de caractériser le comportement probable des sinistres le plus fidèlement possible et pas de

minimiser les différents résultats d'un modèle pour le privilégier plutôt qu'un autre. Pour rappel nous avons dû ajuster plus de la moitié de ce triangle dû à des estimations négatives de liquidations pour pouvoir appliquer cette méthode, on peut donc comprendre que ce sont autant d'informations perdues dans cette simulation affaiblissant la pertinence de cette approche.

Il reste donc cette similitude entre le modèle GLM de charge totale et du modèle détaillé qui permet d'avoir une indication positive sur la cohérence des résultats de ce modèle par rapport à une méthode éprouvée tout en donnant l'important avantage de fournir une grande richesse de détails permettant alors, par exemple, d'appliquer directement et sans approximations les structures de cessions en réassurance, même complexes comme les cessions non proportionnelles ou à multiples mécanismes.

6.1.2 Résultats à l'ultime – conséquences dans Solvabilité II

Dans l'environnement de Solvabilité II, l'estimation à l'ultime va permettre d'obtenir la valeur *BE* (Best Estimate), soit le montant des provisions techniques requises pour couvrir les engagements à venir du portefeuille, ou encore les futurs flux de paiements des sinistres. Cette valeur *BE* va ensuite être regardée sous deux angles et faire l'objet de projection de flux de trésorerie distincte dans les reporting de Solvency II (QRTs) entre « Claims Provision » pour les flux de trésorerie pour sinistre correspondant à des expositions passées, et « Premium Provision » ceux correspondant à des expositions futures (voir Figure 8). En pratique cette distinction peut se faire au regard de la cadence d'acquisition des primes au sens comptable : la prime acquise caractérisant les expositions passées, la prime non-acquise caractérisant les expositions futures.

La valeur *BE* peut être choisie comme le résultat de l'estimation des provisions techniques dans le cadre d'un modèle déterministe, comme la méthode Chain-Ladder ici, ou bien comme l'espérance des provisions techniques dans le cadre d'un modèle stochastique, ici Mack, GLM à Bootstrap, ou modèle détaillé. Ce dernier fournissant des informations permettant la possibilité d'affiner l'allocation du *BE* entre « Claims Provision » et « Premium Provision » puisque la date de survenance des sinistres IBNYR simulés permet de déterminer si la survenance est antérieure ou postérieure à la date de reporting.

6.2 Résultats à horizon 1 an

A présent le portefeuille est regardé au travers des règles prudentielles de Solvabilité II, en estimant la valeur du SCR tout d'abord par la formule standard, puis par les modèles stochastiques.

6.2.1 Estimations du SCR par la formule standard, et du MCR

Voici l'estimation du capital de solvabilité requis en suivant la formule standard, comme indiqué dans la partie 2.3.2 page 24, y compris les calculs intermédiaires clés.

$V_{(prem,GTPL)}$ est basé sur les primes acquises des 12 derniers mois, le portefeuille étant considéré composé de contrats de durée 1 an uniquement et sans tacites reconduction.

$V_{(res,GTPL)}$ est estimé à partir de l'espérance des différents modèles présentés dans la partie 6.1.

Enfin pour le calcul du MCR, $TP_{(nl,GTPL)}$ est estimée avec la valeur des provisions « dossier » pour sinistre à payer au 30/12/2010.

(en million d'euros)	Modèles Best Estimate				
	C-L/Mack Paielements	C-L/Mack Charge Totale	GLM/Bootstrap Paielements	GLM/Bootstrap Charge Totale	Modèle détaillé
$V_{(prem,GTPL)}$			75		
$V_{(res,GTPL)}$	146	461	263	434	426
V_{GTPL}	222	536	338	509	502
σ_{GTPL}	10.4%	10.6%	10.5%	10.6%	10.6%
$SCR_{GTPL,prem,res}$	69	171	107	162	159
$TP_{(nl,GTPL)}$			379		
MCR	31	49	48	49	49

Tableau 19: Estimations du SCR et du MCR

Sous les conditions de calculs indiquées pour ce portefeuille, les SCR estimés par la formule standard suivent logiquement une répartition de résultats analogues aux Best Estimate des différents modèles utilisés, avec les SCR basés sur les triangles de paiements en retrait par rapport aux autres modèles qui donnent des estimations très proches.

Les MCR calculés, eux, sont très similaires car le MCR_{linear} est situé entre 25% et 45% des différentes estimations de SCR, à l'exception de la valeur basée sur le Chain-ladder du triangle de paiements qui donne une MCR en-deçà des autres estimations.

6.2.2 Modèles Stochastiques

Les résultats ci-dessous présentent la distribution des réserves nécessaires à horizon 1 an dans les scénarios extrêmes, c'est-à-dire autour de la probabilité de 99,5%, valeur de référence pour le calcul du SCR.

	GLM/Bootstrap Paielements	GLM/Bootstrap Charge Totale	Modèle détaillé
Probabilité			
99.0%	459	443	344
99.5%	465	447	375
99.8%	472	452	415
99.9%	479	456	441

Tableau 20: Distributions des provisions pour sinistres nécessaires à horizon 1 an

Même si les régulateurs appliquant Solvabilité II exigent le calcul d'estimation de 99,5% des scénarios à horizon 1 an des risques d'une compagnie d'assurance comme référence d'estimation du SCR, incluant l'ensemble des contrats de tous les type de garanties, le risque de marché, le risque de contrepartie, les effets de corrélation et de diversification entre l'ensemble des risques, il est

intéressant d'observer comment se comportent les différents modèles stochastiques dans cette zone probabiliste à horizon 1 an.

Opérationnellement, l'estimation ne pose pas de difficulté :

- Dans le cadre de la simulation des modèles GLM, il suffit de regarder le développement de la première diagonale représentant la position probabiliste à la clôture de l'année suivante à la date de reporting, donc 1 an après.
- Pour le modèle détaillé, c'est encore plus simple, puisqu'il suffit d'utiliser uniquement les transactions de la table de simulation des 12 mois qui suivent la date de reporting.

Les résultats des deux modèles GLM sont comparables et peu volatiles. Avec toutes les limitations mentionnées sur l'observation d'un SCR sur ce seul risque, la probabilité de 99,5% est d'environ 20% supérieur aux provisions à la date de reporting. Alors que les résultats du modèle détaillé sont plus volatiles, et donnent une valeur de probabilité à 99,5% très proche des provisions à la date de reporting. Comme les valeurs à l'ultime était proche entre les modèles basés sur la charge totale, la question autour de la cadence de règlements des sinistres se pose entre les approches basées sur les triangles et le modèle détaillé. Plusieurs pistes peuvent générer cette situation :

Tout d'abord, les méthodes basées sur les triangles ne regardent les montants que de manières homogènes comme un tout. Or dans ce portefeuille, un nombre limité de sinistres peuvent représenter des pourcentages conséquents des provisions à la date de reporting. Si l'un d'eux est revu durant l'horizon d'1 an, ce sera sur son montant global et pas juste d'un facteur linéaire limité, à l'image du facteur de développement de la méthode Chain-Ladder. Surtout que si le modèle détaillé simule sa clôture dans cette période, il peut très bien être clos sans paiement indemnitaire, générant un fort boni de liquidation, et ceci sans que pour autant un sinistre IBNYR important ne soit ouvert dans la simulation sur la même période d'1 an pour compenser ce boni.

Également, cette différence du modèle détaillé avec les autres modèles sur la vision à 1 an est bien moindre à l'ultime et dénote potentiellement un cadencement du règlement des sinistres plus lent identifié par le modèle détaillé. Ceci peut être l'indication que l'horizon 1 an est trop limité comme critère de prudence par rapport au potentiel développement adverse des sinistres jusqu'à l'ultime des engagements du portefeuille. Sur le Tableau 18 page 92, la charge restante à l'ultime à 1 chance sur 10 (percentile à 90%) d'être supérieure de 50% à la provision des sinistres à la date de reporting, et 1 chance sur 100 (percentile à 99%) d'en être supérieure de 88%, soit quasiment un doublement. La structure du portefeuille très soumise à un nombre limité de sinistres combiné au développement long de la responsabilité civile semble être une explication à cette incertitude sur 1 an. Une idée d'amélioration consisterait à inclure les dates d'échéances d'expertises, de réunions de conciliations ou de décisions judiciaires pour mieux refléter le cadencement des sinistres, au moins sur les dossiers sinistres identifiés avec un potentiel de charges conséquentes.

Ces caractéristiques sont une invitation à approfondir les résultats pour comprendre ces différences, dans le cadre d'une nouvelle étude afin de perfectionner le travail de ce mémoire.

6.3 Potentiel et limite d'exploitation du modèle détaillé

Le modèle détaillé est ici présenté sous une forme simplifiée avec comme axe d'utilisation le provisionnement ainsi que son impact dans le pilier I de Solvabilité II. Mais les résultats offerts par cette approche permettent d'envisager beaucoup d'autres utilisations possibles.

En premier lieu, les résultats simulés par sinistres individuels offrent la possibilité d'appliquer l'exacte configuration de la réassurance quel qu'elle soit, ce qui ne peut être réalisé avec les approches agrégées. Ainsi les résultats post-réassurance peuvent être obtenus ouvrant ainsi la voie à un travail d'analyse de son impact et à son optimisation le cas échéant.

Comme le modèle fournit une simulation du développement des sinistres sur les années comptables futures, il peut être utilisé pour estimer le pilier II de Solvabilité II et plus particulièrement le « business plan » des prochaines années comptables à trois ans ou même cinq ans.

Toujours dans l'univers comptable, les détails fournis par le modèle permettent de construire un bilan stochastique, et ainsi d'obtenir une distribution stochastique des résultats comptables après impôt et des autres critères financiers intermédiaires en y intégrant l'ensemble des paramètres comptables, réglementaires, et fiscaux. Dans ces conditions, le moindre changement de configuration réglementaire ou fiscal peut être mesuré à travers l'ensemble du bilan de la compagnie d'assurance.

Dans la continuité du bilan stochastique, il est bien sûr possible d'intégrer les résultats de ce modèle dans les calculs financiers d'actifs/passifs en termes de structuration des investissements adossés aux provisions techniques. Il est facile d'imaginer que la problématique de volatilité dans l'estimation des provisions se répercute également dans le choix des investissements et la liquidité requise.

Un autre axe d'utilisation peut être dans le domaine de la gestion du portefeuille de contrats, et en premier lieu sur le domaine de la tarification. Sous couvert d'offrir une dichotomie suffisamment fine pour décrire la nature et la qualité des risques souscrits, le modèle permet une liaison dynamique entre la distribution de la charge sinistre attendue et l'allocation du capital requis pour cette charge sinistre. Il est donc possible de déterminer la marge technique nécessaire à la rémunération du capital ainsi immobilisé sous les objectifs de rentabilité attendus.

Toujours dans le domaine de la gestion du portefeuille de contrats, le modèle permet de donner l'impact marginal de différentes caractéristiques du portefeuille utilisées en paramétrage du modèle et d'offrir ainsi un support aux arbitrages de marketing, comme sur le fait de privilégier ou non un positionnement en Apéritif ou en Coassurance etc... Et bien sûr de tester différents scénarios de positionnement marché possible.

Tous ces éléments d'exploitation interagissent les uns sur les autres, et le modèle permet d'avoir une image assez large des conséquences sur un grand nombre de domaines de la compagnie d'assurance. Et ces éléments cités ne sont évidemment que des exemples non exhaustifs.

Même si le modèle détaillé est conçu de manière modulaire, permettant une maintenance et des modifications ponctuelles sans altération fonctionnelle de l'ensemble, il est évidemment très sensible aux deux domaines suivants :

- Les données : que ce soit sur l'aspect de l'accès à la donnée, à sa quantité nécessaire pour estimer les paramètres du modèle, et aussi sur l'aspect qualitatif de cette donnée.
- La modélisation elle-même : chaque sous-modèle estime des paramètres à partir de données, voir ci-dessus, mais aussi de méthode et d'hypothèses qui doivent être validées pour obtenir des résultats cohérents module par module. Même si la construction des différentes composantes du modèle détaillé a fait l'objet d'analyses pour les valider dans le cadre de ce mémoire, un approfondissement serait nécessaire pour en améliorer pour bien cerner les avantages et limites, aussi bien sur chaque composant individuel que combinés.

Ceci implique une veille constante sur l'exploitation du modèle, sur la source des données et la qualité des sous-modèles afin de minimiser l'ensemble des risques de modèle de cette méthode.

Enfin dans le cadre de Solvabilité II, la mise en œuvre du modèle détaillé comme composant d'un modèle interne nécessitera la mise en place de procédures multiples et conséquentes sur le domaine de la validation du modèle, des contrôles, et de sa documentation. Sans oublier, le défi que peut représenter les communications avec le régulateur autour d'une approche ad-hoc qui lui sera nécessairement inconnue et donc complètement nouvelle. Une estimation d'un rapport coût/bénéfice est donc indispensable, aussi bien quantitatif que qualitatif, comme préalable à la validation d'un tel projet de mis en œuvre. Mais si les gains sont avérés et substantiels pour ce modèle, la question qui se pose est de trouver le moyen de minimiser ses coûts de mise en place, surtout par rapport aux communications avec le régulateur, voir. Évidemment cette problématique concerne toutes formes d'innovations, et pas uniquement ce modèle. Mais pour autant, il reste tout à fait pertinent d'avoir ce modèle comme opinion alternative afin de détecter des incohérences avec un modèle principal et ouvrir des pistes d'analyses et d'amélioration le cas échéant.

7 Complément sur la modélisation individuelle des sinistres depuis 2013 et comparaison avec le modèle étudié

7.1 Revue et discussion sur une sélection d'articles

Dans la quinzaine d'années écoulée, la littérature actuarielle s'est étoffée de nombre d'études sur le thème de la modélisation détaillée / individuelle des sinistres. L'objectif principal est l'amélioration de la prédictibilité des engagements de l'assureur en mettant à profit des améliorations dans trois différents axes :

- En termes de sciences appliquées, en appliquant les derniers développements théoriques, comme le « machine learning » par exemple
- En termes d'accès à la donnée
- En termes de capacité de traitement

Nous allons principalement nous concentrer sur le premier point bien sûr, même si les deux autres sont cruciaux pour la viabilité opérationnelle des calculs de l'actuaire. Ainsi à travers des articles représentatifs, mais non exhaustifs, sur le thème du provisionnement des sinistres sur données détaillées, je vous propose un panorama du développement des connaissances dans ce domaine depuis 2013.

En premier lieu, un article est fréquemment mentionné, et de par son ancienneté nous ramène un peu plus loin dans le temps, et peut être vu comme initiateur dans ce domaine, même s'il reste sur un plan théorique et formel : il s'agit de l'article « The claims reserving problem in non-life insurance : Some structural ideas » (Arja, 1989) qui présente l'intérêt de modèles stochastiques pour le provisionnement des sinistres, et en particulier en considérant celui-ci comme un processus ponctuel, c'est à dire un processus stochastique dont la réalisation est un ensemble de points isolés dans le temps. Ceci permet d'introduire l'information connue par l'assureur à la date du provisionnement dans la prédiction du montant des sinistres comme conditionnelle à la probabilité de celui-ci. Il montre également que la déclaration de sinistres peut être considérée comme un processus de Poisson. Même si l'article oriente son champ d'application sur des données agrégées, ces résultats sont tout aussi importants pour une granularité fine par dossier sinistre individuel.

L'article « Individual Claim Reserving Model – Add Value » (Boumezoued et Courchene, 2017) offre une intéressante introduction à ce domaine en en présentant des aspects clés. Tout d'abord en rappelant les limites des méthodes de triangulation sur données agrégées :

- La perte d'information lors de l'agrégation des données
- La rigidité des méthodes sur les données et l'intervalle de temps appliquée
- Le poids de l'âge moyen alors que des variables alternatives peuvent être porteuse d'information significatives également.
- La faiblesse de la prise en compte des changements sous-jacent des données à travers les triangles.
- La marge d'erreur significative sur les données les moins matures.

De l'autre côté, la modélisation individuelle des sinistres est présentée comme complémentaire aux triangulations, car elle permet :

- D'améliorer la fiabilité des résultats en intégrant des informations inutilisées dans les triangles.
- De vérifier tout résultat, en particulier ceux suspect, par la granularité des calculs appliqués.
- D'améliorer la compréhension des facteurs qui déterminent la sinistralité, et donc lié l'univers du provisionnement avec celui de la souscription et de la tarification.
- D'analyser en détail la sources des différences entre les différents modèles.

Il convient d'ajouter que cette notion de complémentarité est importante car elle permet de comparer les estimations d'une valeur recherchée faites par des méthodes différentes. C'est d'autant plus important dans le cadre de calculs complexes, et par expérience, cela est soit porteur d'information précieuse dans le cas de divergence des résultats, soit un facteur de validation majeur lors de résultats convergents, aussi bien pour l'actuaire qui fait les calculs que pour tout organe de contrôle de ces calculs.

L'article introduit ensuite les grandes étapes clés de l'implémentation d'un tel modèle en passant par le choix de la structure dynamique, c'est à dire être en temps discret ou continu tout en précisant que le temps continu apporte le plus de précision à la compréhension du portefeuille, à la classification des données et par cet aspect à l'importance de l'exploration des données, à l'estimation des paramètres et à leurs capacités explicatives du modèle, et enfin une fois les étapes de calibration des paramètres et de simulation, à la capacité prédictive du modèle qui va au-delà de l'objet premier du modèle, c'est-à-dire l'estimation des provisions, mais inclut bien d'autres détails sur l'estimation des futurs flux de trésorerie de paiements des sinistres, comme une meilleure précision sur l'incertitude des montants estimés.

Mais un des articles les plus marquants de la période, et aussi régulièrement cité dans ce domaine est « Micro-level stochastic loss reserving for general insurance » (Antonio et Plat, 2014). L'intérêt de l'article est de présenter un modèle de sinistres individuels basé sur une approche, stochastique bien entendu, sur chacun des événements clés des sinistres, mais y compris une étude de cas sur des données réelles de sinistres RC issues d'une compagnie d'assurance européenne. Ce modèle est ensuite comparé à l'approche Chain-Ladder et aussi double Chain-Ladder pour conclure sur la meilleure performance du modèle individuel par rapport aux autres approches sur données agrégées. Au regard de son intérêt, nous allons développer les éléments de cet article plus en détail et discuter des similarités et différences avec le modèle détaillé de ce mémoire.

« Micro-level stochastic loss reserving for general insurance » (Antonio et Plat 2014)

En effet, dans son introduction l'article pose le cadre de l'analyse en définissant d'abord pourquoi l'assureur doit établir une provision, et décrit brièvement la différence bien connue entre sinistre préalablement déclaré à la date de provisionnement, les IBNER dans ce mémoire (« RBNS » dans l'article pour « Report to the insurer But Not Settled ») et les sinistres non encore déclarés à la date de provisionnement, IBNYR dans ce mémoire (« IBNR » dans l'article).

Un schéma de principe général de développement d'un sinistre, voir Figure 17 page 53, présente les différentes étapes de la vie du sinistre pour l'assureur, même si seuls des paiements sont indiqués comme événements comptables intermédiaires. Ceci permet de définir ce qui est considéré comme une donnée sinistre détaillée (« Micro-Level »). L'environnement réglementaire est aussi introduit concernant la norme prudentielle Solvabilité II et aussi la norme comptable internationale IFRS4.

Sa remplaçante aujourd'hui, la norme IFRS17, centrée sur la projection des futurs flux de trésorerie par classe de sinistres homogènes, renforce encore l'attractivité d'une approche du provisionnement au niveau détaillé. L'article constate également la perte d'information générée par l'agrégation des données dans le cadre d'application des méthodes traditionnelles de triangulation où chaque valeur d'une cellule d'un triangle recèle un nombre d'information sous-jacentes très importantes occultées. Alors que les capacités modernes de calculs et en continues progressions permettent de dépasser et d'envisager d'autres approches, et ce qui est aussi un des points de départ de ce mémoire.

Puis l'article décrit les données utilisées issues d'un portefeuille de couverture en responsabilité civile de particuliers d'une compagnie d'assurance européenne. Ceci impliquera évidemment de comprendre ces résultats au regard de cette différence majeure avec ce mémoire qui porte sur des assurés professionnels y compris classés comme « grand risque » dans le langage des assureurs du secteur. Ceci se confirme pour les données retenues pour l'exposition du portefeuille qui se limitent à la période de couverture de la police sans prendre en compte de caractéristiques propres de l'assuré, ce qui peut se comprendre dans le cadre d'assuré particulier, mais qui s'est révélée être une source d'information pour le portefeuille de ce mémoire comme variables explicatives. Les sinistres sont classés en deux catégories de préjudice : dommage matériel et dommage corporel, mais pas de pertes financières. Comptablement les sinistres se matérialisent par 1,5 millions d'enregistrements pour environ 500,000 sinistres au total. Ce qui indique 3 enregistrements en moyenne par sinistre : survenance, ouverture, et paiements / provisions. C'est une différence majeure sur la structure des données, car dans ce mémoire, c'est la base de données des transactions comptables qui a été exploitée, ce qui peut anticiper une meilleure capacité à estimer les positions intermédiaires des sinistres entre leurs ouvertures et leurs clôtures pour le modèle détaillé de ce mémoire. Le type de dommage est également retenue comme variable explicative sur la durée de développement des sinistres, mais sans la perte financière bien entendu, et indique aussi un lien entre sinistre dit « important » et durée de développement du sinistre, ce qui n'a pas été caractérisé dans les données de ce mémoire. Cette différence peut être dû à la combinaison de différents facteurs comme la complexité des contrats et des sinistres RC industriels, ainsi qu'à la nécessaire différence de politique de gestion des sinistres entre ces deux portefeuilles.

La modélisation est basée sur la compréhension des auteurs de la structure d'un sinistre qui est vu comme la combinaison suivante : la date de survenance (T_i), le délai de déclaration (U_i), et le processus de développement (X_i) qui est un processus à saut. Les sinistres déclarés (IBNER) et les sinistres non encore déclarés (IBNYR) suivent deux processus disjoints et considérés comme indépendant. Même si le séquençage d'un sinistre est similaire au modèle de ce mémoire (survenance, déclaration, développement jusqu'à la clôture), la modélisation a été structurée sur la base des données contractuelles et de la table des transactions comptables en découpant les principales parties identifiées en modèles individuels pour former un système de modèle composite construit par empirisme mais qui peut être compris comme un seul et unique processus de Poisson au niveau agrégé. La déclaration d'un sinistre n'y est vu que comme un événement particulier, compris entre sa survenance et sa clôture, c'est à dire sur la durée d'engagement légale et contractuelle de la prestation de l'assureur sur ce sinistre. Comme dans cet article, ne sont considérés que les sinistres portés à la connaissance de l'assureur et donc ayant eu une substance comptable explicite, laissant de côté logiquement les potentiels sinistres jamais déclarés en l'absence d'information, même si l'on peut y percevoir une forme de troncature des données sinistres sur cet aspect d'un point de vue contractuel et légal qui pourrait faire l'objet d'une étude à part entière.

L'approche par modèle de survie a été aussi choisi dans cet article, mais deux modèles ont été utilisés pour caractériser séparément le délai de déclaration puis la durée de développement jusqu'à la clôture, au lieu d'un seul modèle de survie dans ce mémoire qui caractérise le délai de la survenance à la clôture, avec un modèle GLM qui caractérise le ratio de délai de déclaration dans

la durée de vie du sinistre. Les deux approches donnent la capacité d'utiliser des variables explicatives et, évidemment, il conviendrait de tester celle de cet article sur les données de ce mémoire où la question principale serait liée l'intensité de l'interaction entre le délai de déclaration et le délai de clôture qui s'en suit, et où la classification des contrats et des sinistres en groupes homogènes, comme l'activité de l'assuré ou la nature du dommage, pourrait apporter de l'information aux modèles, même si seules la position de l'assureur (apériteur ou coassureur) et le type de garanties (base fait dommageable ou base réclamation) ont pu apporter de l'information en tant que variables explicatives du ratio qui caractérise ici le délai de déclaration. Cependant et pour rappel, dans ce mémoire le type de dommage (matériel, corporel ou financier) a pu être utilisé comme variable explicative dans le modèle de survie caractérisant le délai entre la survenance et la clôture d'un sinistre. Ce qui implique une forme de stabilité dans la proportionnalité entre ces deux durées.

L'une des différences majeures entre l'article et le modèle de ce mémoire concerne le processus de développement : un seul événement comptable de 3 différents types (paiement sans clôture du sinistre, paiement et clôture, et enfin clôture sans paiement) peut survenir dans la vie du sinistre. Dans le mémoire, le modèle de survie caractérise la durée du sinistre jusqu'à sa clôture. Le modèle va aussi déterminer s'il y a un paiement final ou non à cette date de clôture, mais également le nombre d'événements comptables intermédiaires pouvant chacun donner lieu à un paiement indemnitaire ou pour frais, ou encore à une réévaluation de la charge totale du sinistre, et distribuer ces événements intermédiaires entre la date de déclaration du sinistre et la date de clôture. Enfin et pour rappel, dans le modèle de ce mémoire, le dernier événement comptable, la clôture du sinistre, se caractérise par la suppression de la provision à cette date, qu'il y ait ou non-paiement, ce qui permet de considérer que le modèle fournit une vision à l'ultime du développement de l'ensemble des sinistres et donc des engagements de l'assureur. Par conséquence, le nombre d'événements comptables combinés à leur type rend l'approche dans l'article plus simple, mais peu adaptée à un portefeuille où des sinistres majeurs subissent plusieurs réévaluations ou paiements indemnitaires ou pour frais au gré des expertises et des décisions judiciaires. Ceci rendant la projection des flux de trésorerie peut réaliste pour le portefeuille de ce mémoire.

Comme dans ce mémoire, les données utilisées dans cet article ont quelques valeurs qui nécessitent une adaptation ad-hoc du modèle pour coller à ces données, comme le délai de déclaration qui a une forte concentration dans les tout premiers jours qui suivent la survenance du sinistre. Ceci a motivé les auteurs a utilisé une loi composite entre une loi standard et des lois dites dégénérées (« degenerate distributions » en anglais), à l'image des masses de Dirac utilisées dans ce mémoire, comme pour l'estimation de la charge sinistre à l'ouverture qui a un poids statistique spécifique sur la valeur utilisée comme montant forfaitaire en l'absence d'information suffisante.

Le reste de l'article se focalise ensuite sur l'adéquation du modèle avec les données, les méthodes de simulations, sur ses résultats et sa capacité prédictives. L'ensemble de ces éléments est circonstanciel aux données utilisées et présentent moins d'intérêt à comparer avec ce mémoire. Mais la conclusion de l'article insiste sur plusieurs points importants, à commencer pour la faisabilité en condition réelle d'un tel modèle, puis sur la qualité prédictive du modèle qui ressort comme meilleur que les méthodes traditionnelles qui, elles, tendent à surestimer les paiements. Et l'article se termine en ouvrant la voie vers une application pour la réassurance, sur l'extension de la comparaison entre modèle détaillé et modèle agrégé, ainsi que sur l'impact de l'inflation.

En conclusion, le modèle présenté dans cet article présente des similitudes logiques avec le modèle de ce mémoire, comme sur le principe du modèle de survie pour caractériser la durée de vie d'un sinistre, mais aussi, bien sûr, des différences qui s'expliquent par la nature même des données et du type de portefeuille analysé, ce qui mériterait une analyse comparative plus approfondie sur les données du mémoire. Cela confirme la faisabilité d'un tel modèle, qui, une fois implémenté, permet

d'analyser les résultats dans une approche « bottom-up » qui évite l'application d'hypothèses pour compenser les pertes d'information liées à l'agrégation des données des modèles traditionnels.

L'article suivant présente une alternative intéressante avec une approche non-paramétrique :

« Non Parametric individual claim reserving in insurance » (Baudry et Robert, 2017)

Le constat de départ des auteurs porte sur le caractère rigide de l'approche paramétrique des modèles présentés dans les articles de références sur le provisionnement des sinistres individuels qu'ils considèrent comme pouvant manquer de flexibilité au regard de la diversité des données. Ainsi leur attention s'est portée sur les techniques d'apprentissage automatique (« machine learning » en Anglais) en partant des travaux de Wüthrich (2017) qui propose un premier arbre de régression appliquée au provisionnement de sinistres individuels pour estimer uniquement le nombre de paiements, et non leurs montants. Ils appliquent ainsi la méthode « ExtraTrees » pour estimer séparément les provisionnements IBNYR et IBNER (respectivement IBNR et RBNS dans l'article) à partir des informations quantitatives et qualitatives clés pour caractériser l'hétérogénéité des sinistres, les cadences inhérentes (survenance, déclaration, et paiement), ainsi que les potentiels changements de comportements qui peuvent être liés à différentes sources d'évolutions autour des produits, de l'environnement légal, ou du processus de gestion des sinistres lui-même.

A ce stade, nous pouvons constater que les deux premiers objectifs de l'étude recouvrent ceux du modèle de ce mémoire. La question ouverte est sur le dernier point et la capacité à caractériser des changements tendanciels. On peut facilement anticiper que le modèle du mémoire aura un effet de persistance défavorable à reproduire rapidement un changement légal par exemple. Et l'on peut aussi penser à des évolutions brusques dans la nature des risques assurés comme dans le domaine automobile avec la production massive et rapide de véhicules hybrides et électriques ou, chez les fournisseurs de ce secteur, avec un développement important de la digitalisation et connectivité des produits fabriqués.

Dernier aspect annoncé par les auteurs est l'attention particulière qui sera donnée à la volatilité de l'erreur de prédiction qui est un critère majeur de la pertinence d'un modèle prédictif.

Après cette introduction, l'article définit le provisionnement en deux catégories (IBNYR et IBNER), et les différentes valeurs à estimer : le processus des paiements IBNER, et le couple fréquence / processus des paiements pour les IBNYR. Ces estimations sont faites suivant une grille temporelle définie, à l'image des méthodes de triangulation, qui peut être définies sur base annuelle, trimestrielle, mensuelle etc. C'est une différence importante avec le modèle de ce mémoire qui alloue distribue les transactions dans un espace temporel continu indépendamment de toute cadence d'observation liée aux processus de reporting financiers propres à la compagnie d'assurance, à la manière d'un flux standard d'enregistrements comptables de sinistres. La segmentation entre IBNYR et IBNER correspond partiellement à celle appliquée dans notre modèle, car le processus dédié aux IBNYR se concentre uniquement à générer les caractéristiques et les montants initiaux des sinistres à leurs ouvertures. Les flux de paiements suivants jusqu'aux clôtures des sinistres suivent le même processus des paiements que pour les sinistres IBNER conditionnellement aux informations connues entre la date d'ouverture et de reporting des sinistres considérés.

Les auteurs décrivent ensuite la construction des jeux de données tests et d'apprentissage utilisés pour les 3 classes de valeurs de leur modèle (paiements IBNER, fréquence IBNYR et paiements

IBNYR), ainsi que l'estimation prédictive finale, c'est-à-dire le montant global des provisions requises. Puis ils présentent la méthode de triangulation Chain-Ladder et l'estimation du montant global des provisions par cette méthode, en translatant les dates de sinistres de l'axe temporel des périodes de souscription sur les périodes de survenance, ceci afin de pouvoir comparer leur modèle suggéré précédemment avec cette approche usuelle.

Un cas pratique est utilisé en illustration concernant un portefeuille d'assurance de téléphone portable de 2 années d'existence, environ 500,000 polices, et offrant 3 niveaux de couverture (Bris, Bris + Oxydation, et Bris + Oxydation + Vol), 4 marques d'appareils, et 4 types d'appareils. Les informations exploitées concernant les polices sont : Marque, Prix, Type d'appareil et couverture. Et concernant les informations sinistres les informations suivantes sont exploitées : Type de dommage (ou de couverture engagée), délai de déclaration, nombre de jours écoulé depuis la déclaration. Dans ce portefeuille, il apparaît que le type de couverture et le type de dommage sont identiques, à la différence du portefeuille de ce mémoire qui nécessite cette dichotomie. Si l'on peut voir évidemment des similarités d'approche concernant les variables qualitatives, le montant d'indemnisation en revanche ne peut être vu comme un ratio d'une valeur connue à l'avance dans un portefeuille RC de risques industriels, comme le prix du téléphone ici, car bien évidemment il est beaucoup plus compliqué d'avoir des informations qualitatives et quantitatives des tiers pouvant engager la responsabilité de l'assuré. Enfin le paiement se fait en un flux unique versé à une date estimée par le modèle et borné. Evidemment cet aspect est très réducteur par rapport aux caractéristiques des flux de trésorerie du portefeuille considéré dans ce mémoire.

Les montants estimés des réserves sont réalisés mensuellement sur 2 années avec la méthode de Chain-Ladder (CL) et celle du Machine-Learning (ML) suggérée. Même si les deux approches nécessitent un délai d'apprentissage, le ML est un peu plus rapide que le CL. Mais là où le ML est significativement plus performant c'est sur la volatilité des estimations où les auteurs montrent bien que l'intervalle de confiance est bien moins important sur le ML que sur le CL. Ceci rend attractif ce type de modèle pour le tester également dans le cadre du portefeuille de ce mémoire pour voir s'il est susceptible d'avoir également une volatilité réduite. Cela pourrait être en remplacement complet des sous-modèles suggérés, ou simplement partiellement sur les aspects les plus importants comme le modèle de durée, la survenance des événements comptables, l'estimation du facteur de développement de la charge indemnitaire totale, ou le déclenchement d'un paiement indemnitaire.

Les auteurs font ensuite variés différentes hypothèses du modèle afin de voir comment s'adaptent les deux méthodes, mais le plus notable de ces tests concerne la variation de 10% des délais de paiements, que ce soit négativement en allongeant le délai de 1 jour ou bien positivement en raccourcissant ce délai de 1 jour. Les résultats montrent sans ambiguïté l'effet de persistance très nettement supérieur de la méthode CL qui, sans surprise, nécessite des délais beaucoup plus importants pour intégrer ce changement tout en générant des déviations d'estimations très importantes, alors que la méthode ML s'adapte en très peu de temps et sans que les prédictions diffèrent trop significativement des valeurs attendues. Ce résultat là aussi rend très attractif l'analyse de la méthode ML dans le cadre de ce portefeuille pour voir comment tout changement de caractéristique, comme le délai de traitement des sinistres influe sur les estimations, que ce changement soit endogène, comme une amélioration dans le processus interne de gestion des sinistres, ou exogène comme une complexification dans l'environnement réglementaire générant l'allongement des procédures de recours entre les différents tiers potentiellement impliqués dans un sinistre, par exemple.

Une meilleure performance du ML est aussi remarquée sur un test de choc de sinistralité temporaire, aussi bien en termes de capacité d'adaptation que de volatilité des estimations qui reste bien moindre que celle du CL.

L'étude se concentre ensuite sur la comparaison entre la méthode d'apprentissage automatique à partir de l'algorithme ExtraTrees et la méthode bien connue de triangulation du Chain-Ladder. Les résultats sont très encourageants et confirment définitivement l'intérêt de cette nouvelle approche. Et la question immédiate est : quelle serait le résultat en l'appliquant sur ce portefeuille RC de risques industriels ? Il serait vraiment intéressant d'analyser les capacités d'adaptation et la volatilité des estimations en comparaison des sous-modèles proposés dans ce mémoire, comme le modèle de Cox sur la survie des sinistres ou le facteur de développement qui à nécessiter un sous-modèle hybride plus complexe qu'un modèle GLM standard.

Il est intéressant de noter qu'en 2018 un autre article propose une approche également non-paramétrique basée sur l'algorithme CART (« Classification And Regression Tree ») :

« Micro-Level vs Macro-Level reserving in non-life insurance: Why and When? » (Lopez et al. 2018)

Les auteurs présentent un modèle basé sur des données explicatives, une durée jusqu'à la clôture des sinistres et leurs montants de charges finales. Ils introduisent la variable de pondération de Kaplan-Meier pour pouvoir réaliser l'estimation de fonction de répartition du couple durée / montant conditionnel aux variables explicatives en présence de données tronquées. Dans ce mémoire pour rappel, le modèle de Cox a été privilégié par intégrer les variables explicatives.

Point intéressant, les auteurs de l'article indiquent qu'il y a un lien direct entre la durée d'un sinistre et son montant final. Cette affirmation est challengée par les caractéristiques du portefeuille analysé dans ce mémoire : dans un contexte de couverture RC de risques industriels n'ayant pas uniquement des expositions françaises mais aussi internationales y compris en terme de domiciliation juridique des contrats, le jeu des expertises, contre-expertises, conciliations amiables, recours juridiques incluant les différentes procédures d'appels rendent le développement d'un sinistre, certes plus long, mais surtout incertain pour son montant final et pas nécessairement plus coûteux, ce qui est une nuance dans ce contexte par rapport à l'affirmation initial de l'article.

L'article, qui se concentre sur l'estimation des IBNER, conclut là aussi que l'approche détaillée est plus réactive aux changements de caractéristiques de développements de sinistres, mais nécessite une base de données étendues aussi bien en termes d'historique des données qu'en termes de données explicatives. Il est aussi mentionné que dans un contexte de risques connus et de court-terme, les méthodes traditionnelles, type Chain-Ladder, restent un meilleur compromis pour estimer les montants de provisions.

Dans son introduction, les auteurs résument les faiblesses des méthodes agrégées traditionnelles comme la potentielle fragilité des estimations des valeurs périphériques des triangles, la difficulté de capturer les caractéristiques de risques hétérogènes, ou encore tout changement dans la chaîne opérationnelle de traitement des sinistres, comme la gouvernance de règlement des sinistres, le contexte réglementaire et juridiques, ou encore le changement du contrat d'assurance sous-jacent ou de la structure de réassurance. L'ensemble de ces constatations sont également le point de départ de ce mémoire.

Encore plus récemment, dans l'article :

« Micro-Level Parametric Duration-Frequency-Severity modelling for outstanding claim payments » (Yanez et Pigeon, 2021).

Les auteurs proposent une méthode paramétrique structurellement similaire à l'approche de ce mémoire. Elle repose en effet sur les 3 principales étapes que sont l'estimation de la durée des sinistres depuis la survenance jusqu'à la clôture, l'estimation des sinistres déjà ouverts IBNER, et enfin des sinistres non encore déclarés IBNYR. Les estimations sont vues sous 3 grands composants : la durée des sinistres, la fréquence des paiements et le montant de ces paiements. La similitude avec l'approche de ce mémoire est très marquante car, même si ce mémoire cherche également à caractériser certaines autres caractéristiques des paiements, pas uniquement indemnitaires, ces importants mécanismes sont similaires.

Sur la durée d'un sinistre, les auteurs séparent celle-ci en trois périodes, avec en premier lieu le délai de survenance qui est défini entre le début de l'année de survenance et la date exacte de survenance. Ce qui dans le cadre de ce mémoire correspond au délai compris entre la date de début de l'année de souscription et la date exacte de survenance du sinistre considéré durant cette année de souscription. La nature du risque de responsabilité civile ne montre pas de saisonnalité au cours d'une année de souscription ici, mais sur le principe c'est une possibilité à prendre en compte, comme dans le cadre de couvertures directement ou indirectement liées aux événements climatiques par exemple. La période suivante est naturellement le délai de déclaration du sinistre où les auteurs, comme dans ce mémoire, ne considèrent pas de décalage entre la déclaration à l'assureur au sens juridique et l'ouverture comptable du sinistre par l'assureur. Puis vient le délai de clôture du sinistre pour conclure le volet durée. Si l'ensemble de ces 3 périodes recouvrent les mêmes éléments à estimer que dans ce mémoire, l'approche diffère néanmoins car, pour rappel, un seul réel modèle de durée est utilisé dans ce mémoire pour estimer la durée totale de l'engagement juridique de l'assureur, c'est-à-dire depuis la survenue d'un sinistre jusqu'à sa clôture, et où le délai de déclaration n'est vu que comme un ratio de cette durée totale. Il serait intéressant de creuser l'approche de cet article appliquée aux données de ce mémoire, même si l'on peut anticiper que d'estimer ces 3 périodes séparément permet d'avoir plus de flexibilité et d'intégrer plus directement tout changement dans le processus de gestion des sinistres provenant de l'assureur ou de l'évolution de pratiques du marché de l'assurance. Evidemment ces deux premiers délais nécessitent des approches dédiées alors que pour la dernière période, la durée jusqu'à la clôture, les auteurs suggèrent également d'utiliser un modèle de survie comme dans ce mémoire.

L'article présente la fréquence des paiements à partir d'une partition d'intervalle de temps, approche courante, qui consiste à regarder le développement des sinistres à chaque fin de périodes prédéterminées, en pratique à des périodes de reporting financiers comme lors d'une clôture comptable trimestrielle ou annuelle. C'est une différence importante avec ce mémoire qui fait cette analyse sur une base de temps continu en plus de différencier la distribution de ces événements comptables suivant la position de l'assurance comme apériteur du contrat ou coassureur du contrat.

Concernant le dernier composant, l'estimation de la valeur des paiements, les auteurs explicitent l'aspect hétérogène et diffus de ce composant. Cette position des auteurs, de nature académique dans leur article, confirme la nécessité d'avoir une approche ad hoc sur cette partie, tels les choix qui furent nécessaires dans ce mémoire pour cette catégorie, même si c'est le développement du montant de la charge totale des sinistres qui a été ici directement analysé.

Un autre aspect intéressant de cet article est, comme de ce mémoire, de proposer une estimation des sinistres IBNYR, ce qui est peu fréquent au travers de l'ensemble des articles recensés dans ce domaine. Pour cela les auteurs estiment le nombre de sinistres restant à être déclarer à partir des données existantes en partant de l'estimation du nombre total de sinistres qui seront déclarer dans la période i en fonction de l'exposition du portefeuille sur cette même période i et qui suit une

distribution de Poisson. Puis les autres valeurs quantitatives, délais et montants, sont estimés à partir des mêmes données que pour la partie IBNER. Le principe est identique à l'approche de ce mémoire, même si dans le mémoire l'estimation des différentes valeurs est chaînée car les sinistres IBNYR sont vus comme un cas particulier de sinistres ayant une position comptable d'ouverture dans le futur, et dont les estimations de liquidations suivantes suivent alors la même approche que pour les sinistres IBNER déjà ouverts.

L'approche des auteurs est illustrée par l'analyse de données réelles d'un portefeuille de couverture « Accident Benefits » d'un assureur canadien appliquées aux 3 principales étapes de leur modèle, puis comparer à des méthodes agrégées (Mack et GLM), puis à un autre modèle sur données détaillées, celui de Antonio et Plat de 2014 dont l'article a été commenté précédemment. Cette couverture concerne l'ensemble des dommages corporels survenant lors d'un accident auto, aussi bien ceux du conducteur que ceux des tiers, passagers ou non. Les principales variables explicatives sont le genre, la région, le type de sinistre (multi-véhicule, choc avec piéton etc...), âge du véhicule, âge du blessé, délai de déclaration, et la provision initiale. Les lois statistiques sont sélectionnées pour les 3 composants du modèle : la loi de Weibull pour le délai de clôture, la loi binomiale négative pour la fréquence des sinistres, et la loi Gamma pour la sévérité des sinistres. Il est à noter que les deux premières lois sont similaires à celles employées dans ce mémoire, ce qui ne pouvait être de même pour la sévérité au regard des approches différentes employées pour cette dernière partie. L'approche traditionnelle est vue au travers de quatre modèles stochastiques : Mack avec bootstrap suivant d'abord une loi de Poisson, puis une loi de Gamma pour le deuxième modèle. Les deux autres modèles sont basés sur l'approche GLM avec comme variables explicatives l'année de survenance et l'année de développement avec les deux mêmes distributions que précédemment, Poisson et Gamma, pour ces 3^e et 4^e modèles. Enfin le modèle détaillé d'Antonio et Plat est appliqué en suivant les mêmes intervalles de temps que pour le modèle des auteurs, mais où la partie sévérité, basé sur le montant des paiements individuels, est utilisé sans variables explicatives.

Les résultats sont vus à deux dates d'horizon différentes, 31.12.2017 et 31.12.2019, où les conclusions sont les mêmes concernant les différents modèles testés. Le modèle détaillé d'Antonio et Plat sous-estime les valeurs observées y compris sur des percentiles à 99%. Les autres modèles sont assez proches, mais le modèle à 3 composants des auteurs diverge moins dans les percentiles à 95% et 99% et semble ainsi le plus intéressant, avec l'avantage supplémentaire de pouvoir fournir une probabilité individuelle par sinistre. Il est à noter que les auteurs concluent également que le choix des variables explicatives est important et permet d'améliorer l'estimation des provisions, et qu'il peut être attractif également dans le cadre d'un modèle tarifaire ou dans l'analyse et le choix de la réassurance. Ceci fait écho bien entendu aux résultats issus de ce mémoire également.

7.2 Pistes originales du modèle détaillé de ce mémoire et autres idées de prolongement de l'étude

Au travers des articles vus précédemment, les travaux sur le « Micro-level reserving » consacrés au type de portefeuille analysé dans ce mémoire ne sont pas les plus courants, et ce malgré une plus-value opérationnelle quant à l'application et au calcul de l'impact de structures contractuelles, parfois complexes, entre la police de risques industrielles, une potentielle réassurance facultative, et un traité de réassurance, pour arriver à une vision nette pour l'assureur. Les caractéristiques de ce portefeuille posent des problématiques absentes des risques de masses et homogènes, et nécessitent de les comprendre pour réaliser un modèle qui incorpore ces comportements propres qui influent significativement sur le processus de liquidation des sinistres. Ainsi il a été identifié l'importance de la position contractuelle de l'assureur : Apériteur ou coassureur (voir 5.2.1 page 64)

Aucun des articles ne mentionne cette variable explicative, alors que dans les conditions de ce portefeuille, elle influe très nettement sur la cadence de règlement des sinistres, avec des conséquences qui dépassent le seul sujet de provisionnement en débordant sur la stratégie même de portefeuille et sa tarification.

Autre caractéristique originale de ce portefeuille, et absente des articles précédents, est le déroulement long et parfois complexe d'un sinistre, aussi bien en termes du délai de manifestation du préjudice, de la compréhension du fait générateur, de la caractérisation des responsabilités, et de la quantification du préjudice. L'enchaînement de ces étapes ne se limite pas à étendre la durée de règlement du sinistre au travers des expertises, des réunions de coassureurs, des médiations, ou encore des possibles recours juridiques. Mais il a été identifié qu'un régleur sinistre n'enclenche de paiement indemnitaire que lorsque l'issue de ce processus est devenu quasi certain, ce qui a permis d'utiliser la caractéristique de statut du paiement indemnitaire comme clé dans les développements suivant son enclenchement.

Une autre approche originale a été de considérer le développement des sinistres en temps continu et indépendant des périodes de reporting comptables. Ce choix a été fait dès le départ de l'étude pour limiter le nombre d'opérations à réaliser par itération simulée. Ainsi cette approche est insensible aux périodes et fréquences de reporting et permet de simuler des événements comptables plus fidèlement à la réalité des transactions observées.

Un dernier aspect original a noté dans le modèle de ce mémoire a été de structurer les parties IBNYR et IBNER comme conditionnelle à l'ensemble de la vie d'un sinistre depuis sa survenance. L'originalité est d'abord liée à la rareté même de l'estimation des IBNYR dans les articles sur le « Micro Level reporting ». De plus ici, les mêmes modèles de développement et de durée du sinistre sont utilisés, conditionnellement au statut de chaque sinistre. Un sinistre IBNYR est alors vu comme un sinistre ayant une date d'ouverture postérieure à la date de reporting et ayant des caractéristiques et montants propres qui permettent de l'utiliser ensuite comme n'importe quel autre sinistre préalablement ouvert. C'est une structure qui pourrait être intéressante d'approfondir pour en identifier tous les avantages et potentielles limites.

Même si les articles précédents et ce mémoire couvrent de nombreux aspects, il s'agit d'une matière dynamique qui a vu dans les 10 dernières années à la fois de nouvelles possibilités techniques s'ouvrir, mais aussi l'émergence de nouveaux risques. D'ailleurs l'un peut même répondre à l'autre : l'IA offre des possibilités d'exploration très intéressantes, comme effectuer une veille sur l'émergence de nouveaux risques. En effet, un agent IA pourrait être entraîné sur la base des jurisprudences internationales, croiser les similitudes et divergences entre les différentes juridictions, et estimer l'impact potentiel d'un risque émergent, lui-même étant issu d'une veille IA sur la base des milliards d'informations publiées chaque jour, permettant ainsi très en amont pour l'actuaire de prendre en compte ce potentiel développement adverse dans les provisions, et pour le responsable du portefeuille de l'intégrer dans sa politique de souscription.

Parmi les nouveaux risques deux sont bien illustratifs de cette évolution : tout d'abord les enjeux autour des conséquences et responsabilités sur l'emploi de polluants éternels¹². Ce risque est même considéré comme systémique car ces produits issus de l'industrie chimique se retrouvent dans nombre de productions de biens usuels exposant un très grand nombre de personnes à de potentiels impacts sur leur santé. Cela ouvre la voie à des litiges pour préjudices aussi bien matériels

¹² PFAS ET (RÉ)ASSURANCE - Institut des Actuaire et Guy Carpenter – 20.06.2024

que corporels et motivent de nombreuses discussions du législateur qui a déjà adapté la directive sur la prescriptibilité du dommage. Et logiquement un large panel de couvertures RC (entreprises, employeurs, produits, environnement, et mandataires sociaux) se retrouve exposé à des recherches en responsabilités et d'indemnisations. Cela oblige l'actuaire à revoir l'exposition du portefeuille en cours et passé au regard des règles de prescriptions pour réajuster les provisions sur des bases nouvelles potentiellement mal anticipées lors de la tarification des polices concernées plusieurs années auparavant.

Il n'est pas exclu, étant donné son potentiel caractère systémique, que les polluants éternels déclenchent des class actions en France. Cette disposition est le deuxième exemple d'émergence de risques, mais cette fois-ci sur l'aspect réglementaire et législatif. Établie par le législateur avec la loi « Hamon » du 17.03.2014, elle permet, sous certaines conditions, à un groupe de personnes ayant subi un préjudice similaire de se regrouper pour porter une action en justice contre une même entreprise ou organisation. Cela permet, d'une part, d'éviter la multiplication de procédures individuelles, et d'autre part, de renforcer le poids de la démarche, surtout contre les grandes entreprises. Les domaines éligibles à une classe action sont la santé, la consommation, l'environnement, la protection des données personnelles, le travail, et la location immobilière. Ainsi, une des toutes premières classes intentées est contre la société pharmaceutique Sanofi (déclarée recevable le 5 janvier 2022 par le tribunal judiciaire de Paris), dont la procédure est en cours, et concerne un contentieux en responsabilité pour produits défectueux dans le domaine du droit de la santé. Cette disposition, là aussi, motive l'actuaire à réévaluer la nature des expositions des contrats passés et en cours au regard de ce changement réglementaire.

Ces deux exemples sont bien représentatifs de l'évolution des risques : en supplément de l'émergence de risques matériels, le risque réglementaire vient s'ajouter à la matière première de l'assureur, et peut se révéler être un réel défi par ses potentielles conséquences rétroactives, plusieurs années après la souscription de polices faites dans un marché concurrentiel et un contexte réglementaire différent. Face à cela, l'actuaire peut établir des analyses allant des plus conceptuelles, avec l'introduction de variables latentes dans ces modèles, au plus concrètes, comme l'application de scénarios déterministes, tels des « stress tests », sur ces expositions pour en vérifier la résilience et identifier les potentiels faiblesses. En cela un modèle détaillé peut permettre l'application de scénarios bien définies, déterministes ou stochastiques et avec potentiellement l'aide de l'IA, sur le portefeuille et peut aider à en observer les impacts au niveau de chaque contrat, ce qui dans le cadre des risques industriels serait un atout majeur pour éclairer les décisions stratégiques du portefeuille.

8 Conclusion

Cette étude avait pour objectif la conception et l'application d'un modèle stochastique de provisionnement des sinistres sur base de données détaillées (« Micro-Level loss stochastic reserving model ») pour un portefeuille de couverture en responsabilité civile générale de risques industriels dans l'environnement de la directive Solvabilité II. Cela a été réalisé à la fois par l'investigation dans les données disponibles, et par les nombreux échanges auprès des experts de la gestion de ce type de sinistres, aux caractéristiques parfois complexes, qui ont fourni un précieux soutien dans l'interprétation des données observées.

Après des rappels sur les aspects légaux et contractuels de la couverture en responsabilité civile générale, quelques informations sur le marché français correspondant, des éléments clés de la directive Solvabilité II liés à l'objet de cette étude, et sur quelques méthodes usuelles de provisionnement des sinistres sur données agrégées, la problématique liée aux caractéristiques du portefeuille ont été présentées en identifiant la concentration de la charge sinistre historique sur un nombre très limité de sinistres et dans un contexte contractuel parfois complexe. Ceci a amené à la recherche d'une méthode plus adaptée que les approches usuelles pour estimer le provisionnement des sinistres. La piste d'un modèle basé sur les données détaillées a été suivie permettant d'enrichir la modélisation de variables explicatives et de garder un niveau de granularité adapté à la possible application ultérieure de conditions contractuelles ciblées. Une approche ad-hoc a été employée afin de comprendre les données de liquidation des sinistres pour décomposer l'ensemble du processus en différentes étapes. Les principales étapes ont été regroupées dans un ensemble dédié à la liquidation des sinistres déjà ouverts à la date de reporting : durée du sinistre dans les comptes de l'assureur jusqu'à sa clôture, nombre d'événements comptables jusqu'à cette date ainsi que leurs répartitions continues dans le temps, et caractérisation de chacun de ces événements comptables comme pouvant être soit une réévaluation de la charge soit un paiement. Un deuxième ensemble d'étapes a concerné les sinistres restant à déclarer dont l'enjeu a porté sur l'estimation de leurs nombres, mais également d'en déterminer les différentes caractéristiques et les différents montants à leurs ouvertures de façon à pouvoir être employé comme entrée dans l'ensemble d'étapes principales de liquidation des sinistres déjà ouverts.

Chaque étape identifiée a fait l'objet d'un modèle dédié pour en simuler le comportement par l'ajustement de modèles de la famille des GLM incluant l'utilisation de variables explicatives qualitatives et quantitatives provenant aussi bien des informations des polices que des sinistres. Certaines étapes ont dû être scindées en plusieurs sous-étapes comme le modèle estimant le montant de charge indemnitaire à l'ouverture d'un sinistre qui a dû distinguer ceux ayant un simple montant forfaitaire d'ouverture par manque d'information de ceux ayant une première estimation spécifique. L'ensemble de ces modèles recèlent nombre d'enseignements sur le comportement du processus de liquidation dont deux ont été particulièrement marquants. En premier lieu sur le modèle de survie des sinistres qui a explicité l'impact du choix de positionnement de l'assureur sur une police, entre apériteur et coassueur, sur la durée de liquidation des sinistres s'y appliquant et donc sur leurs coûts de gestion. Puis l'autre étape remarquable concerne le développement du montant de sinistre lors d'un événement comptable. Deux comportements de liquidation distincts ont été identifiés lorsqu'un sinistre a déjà eu un paiement indemnitaire précédemment ou non. Les facteurs de développement des sinistres ont des profils très différents dans chacune de ces deux situations définissant ainsi deux processus stochastiques distinctes où un sinistre peut passer d'un régime aux caractéristiques volatiles avant tout paiement indemnitaire à un régime beaucoup plus certain ayant des facteurs de développement proche de 1 lorsque des paiements indemnitaires ont déjà eu lieu. Ce qui explicite le fait qu'un régleur sinistre initie des paiements indemnitaires lorsqu'il a suffisamment d'éléments sur le coût final d'un sinistre. Et réciproquement lors de la clôture d'un sinistre, il a été vu que sans paiement indemnitaire préalable, un sinistre avait alors 80% de chance

de se clôturer sans aucune charge indemnitaire. L'intérêt a été de pouvoir intégrer ces caractéristiques propres par sinistre, susceptibles de créer des variations individuelles et potentiellement intenses, mais peu appréhendables avec l'approche linéaire des triangles de liquidation.

Les résultats à l'ultime de la moyenne des provisions nécessaires des triangles de liquidation des sinistres basés sur les charges totales et du modèle détaillé ont montré des convergences encourageantes d'une part, et ont aussi révélé les faiblesses des triangles de liquidation basés sur les simples paiements qui semblent sous-estimer ce besoin de provision dans le contexte de ce portefeuille, montrant que l'absence d'information sur l'évaluation des provisions à la date de reporting pouvait donner des résultats particulièrement optimistes sur les bonis de liquidation et mal à propos avec une pratique de provisionnement prudente au regard des résultats issus des triangles de charges totales. Le modèle détaillé et l'approche GLM à bootstrap sont similaires sur les résultats à l'ultime, aussi bien en moyenne qu'en terme de volatilité, ce qui est là aussi prometteur pour le modèle détaillé qui, lui, permettra de refléter beaucoup plus précisément les impacts des différentes cessions en réassurance, facultatives ou obligatoires, en particulier pour celles non-proportionnelles, ce qui est toujours délicat à partir des données agrégées d'un triangle. Il en va de même grâce à cette granularité de calcul pour les résultats à horizon 1 an bien entendu.

Evidemment, arriver à un modèle détaillé, et basé sur une large palette de variables explicatives, apte à fournir des résultats proches des méthodes éprouvées et à un niveau de granularité très utile pour nombre de calcul ou de reporting ultérieurs ne suffit pas à le rendre admissible pour autant, et c'est une limite de cette étude. Une phase plus approfondie de validation est requise pour cela. D'ailleurs cette étape est examinée avec attention par le régulateur pour agréer l'utilisation d'un modèle interne dans le cadre de Solvabilité II. Ici, les résultats à horizon 1 ans du modèle détaillé et des autres modèles indiquent un besoin de validation plus approfondi, comme sur les cadences de règlements par exemple. De même que des tests rétroactifs systématiques seraient nécessaires pour améliorer la fiabilité des modèles utilisés par module du modèle détaillé. Enfin, il serait intéressant de comparer les charges brutes simulées par le modèle détaillé avec des coûts observés ou potentiels sur le marché, français ou international suivant la nature des risques des polices couvertes.

D'autres limitations, d'ordre méthodologique, auraient pu être identifiées au regard de la quinzaine d'années écoulées depuis la conception de ce modèle détaillé. Mais il n'a pas émergé de la revue de la sélection d'articles parus depuis d'éléments remettant en cause lourdement l'approche de cette étude. Bien au contraire, plusieurs choix de modélisation, optées à l'époque, sont également présents dans ces articles, comme sur l'emploi d'un modèle de survie pour estimer la durée de vie des sinistres, ou encore de réaliser les calculs en temps continu plutôt que discret pour favoriser la précision du modèle.

Les perspectives ouvertes par cette étude sont nombreuses comme son potentiel champ d'application. De la revue d'articles émerge la possibilité d'utiliser des méthodes non-paramétriques pour améliorer l'adaptabilité du modèle à de nouvelles conjonctures, comme un changement d'ordre réglementaire, contractuel ou de processus de gestion des sinistres. L'IA offre également un nouveau champ de développement permettant, par exemple, tout aussi bien d'offrir une veille sur les risques émergents, sur un marché précis ou au niveau international, et ce bien en amont de toute demande de réclamation dans le portefeuille de l'assureur, ou même d'avoir une idée de la transposition juridique potentiel dans son marché d'un risque apparu dans un autre environnement légal par entraînement d'un agent IA sur différentes jurisprudences et transpositions observées dans le passé.

Ainsi, cette étude confirme le potentiel des modèles détaillés, tout d'abord dans l'estimation des provisions pour sinistre à payer sur un portefeuille de risques hétérogènes, comme ce portefeuille de risques RCG de risques industriels, et également comme outil attractif d'aide à la compréhension et à la décision dans bien d'autres domaines clés de l'assureur, techniques, financiers ou opérationnels, et ce d'autant plus dans un environnement exigeant qui impose à l'assureur une connaissance de plus en plus fine de son portefeuille. Ainsi cela lui permet d'avoir à disposition des éléments factuels enrichis pour prendre des décisions tout autant ciblées que stratégiques.

9 Bibliographie

- Antonio et Plat (2014). Micro-level stochastic loss reserving for general insurance. *University of Amsterdam and KU Leuven (Belgium)*.
- Arjas (1989). The claims reserving problem in Non-Life Insurance: Some structural ideas. *ASTIN Bulletin: The Journal of the IAA*.
- Baudry et Robert (2017). Non-Parametric individual claim reserving in insurance. *Preprint, ressources-actuarielles.net*.
- Bender et al. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in medicine, Wiley Online Library*.
- Bénéplanc et al. (2022). Manuel d'Assurance. *Presses Universitaires de France*.
- Beneteau (2004). Modèle de provisionnement sur données détaillées en assurance non-vie. *ENSAE Paristech, actuaries.org*.
- Boumezoued et Courchene (2017). Individual Claim Reserving Model – Adding Value. *Milliman white paper*.
- Chaumet (2011). Les assurances de responsabilité de l'entreprise (5^e édition). *Editions L'Argus de l'Assurance*.
- Denuit et Charpentier (2005). Mathématiques de l'assurance non-vie, tome 2. *Economica*.
- Duchesne (2010). Théorie et applications des méthodes de régression - Instructions informatiques. *Département de mathématiques et de statistique Université Laval*.
- England et Verall (2002). Stochastic claims reserving in general insurance. *British Actuarial Journal, cambridge.org*.
- Gilbert et Chen (2007). Using SAS Proc NLMIXED for Robust Regression. *SAS Global Forum 2007*.
- Guszcza et Lommele (2006). Loss reserving using claim-level data. *Casualty Actuarial Forum, casact.org*.
- Giuly (2018). Approches de provisionnement individuel en réassurance non-vie. *Mémoire actuariat - ressources-actuarielles.net*
- Lopez et al. (2018). Micro-Level vs Macro-Level reserving in non-life insurance: Why and When? *hal-emse.ccsd.cnrs.fr*.
- Partrat et al. (2008). Provisionnement technique en assurance non-vie. *Economica*.
- Pedan (2001). Analysis of Count Data Using the SAS System. *support.sas.com/resources/papers/*
- Planchet et al. (2005). Modèles financiers en assurance, analyses de risque dynamiques. *Economica*.
- Portait et Poncet (2008). Finance de marché. *Dalloz*.

Saporta (2008). Probabilités, analyses des données et statistique. *Editions Technip*.

Skoubani (2022). Assurance construction : Provisionnement Dommages-ouvrage exploitant les typologies des sinistres et les risques assurés. *Mémoire actuariat - ressources-actuarielles.net*

Smith et Smith (2002). Kaplan Meier and Cox proportional hazards modeling : Hands on survival analysis. *DoD Center for Deployment Health research*.

Smithson et Verkuilen (2005). A better lemon squeezer? Maximum-likelihood regression with Beta-distributed dependant variables. *Psychological methods, psycnet.apa.org*.

Taylor G et al. (2003). Loss reserving: past, present and future. *University of Melbourne Centre of Actuarial Studies Research Paper, papers.ssrn.com*.

Tenenhaus (2012). La Régression Logistique. *Groupe HEC, et Chambre et d'industrie de Paris*

Valleron et Kutek (2003). Cotation des traités non proportionnels en développement long. La Réassurance, approche technique. *Economica*.

Yanez et Pigeon (2021). Micro-Level Parametric Duration-Frequency-Severity modelling for outstanding claim payments. *Insurance: Mathematics and Economics, Elsevier*.

10 Annexes

10.1 Principales théories de modélisation utilisées

10.1.1 Les modèles GLM

Généralités sur les GLM

Les modèles linéaires généralisés ont été proposés par Nelder et Wedderburn en 1972, et regroupent l'ensemble des modèles qui cherchent à exprimer l'espérance d'une variable réponse Y comme combinaison linéaire de variables explicatives $X_1 \dots X_n$.

Trois composantes définissent l'appartenance d'un modèle à la catégorie des modèles linéaires généralisés :

La première composante est la forme de la distribution de la variable réponse Y . Elle doit appartenir à la famille exponentielle. Ainsi, la densité pour une variable continue, ou la fonction de masse pour une variable discrète, doit pouvoir s'écrire sous la forme suivante :

$$f(y_i; \theta_i; \varphi) = \exp \left(\frac{y_i \theta_i - b(\theta_i)}{a(\varphi)} + c(y_i; \varphi) \right)$$

Où :

- θ_i est un paramètre réel, dit paramètre naturel
- φ est un paramètre réel strictement positif, appelé paramètre de dispersion ou de nuisance
- a est une fonction non-nulle définie sur les réels de la forme $a(\varphi) = \frac{\varphi}{w_i}$ où w_i est un facteur de pondération entre les observations i .
- b est une fonction définie sur les réels et deux fois différentiable.
- c est une fonction définie sur R^2

Remarque : dans la suite du document les formules seront indiquées avec le seul paramètre de dispersion φ . En cas d'utilisation d'un facteur de pondération entre les observations i , φ est à remplacer par $\frac{\varphi}{w_i}$.

La deuxième composante est le prédicteur linéaire η : Celui-ci doit s'écrire sous forme de combinaison linéaire des variables explicatives $X_1 \dots X_k$:

$$\eta = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

La troisième composante est la fonction de lien g entre la variable réponse Y et le prédicteur linéaire η . En posant $E(Y_i) = \mu_i$, g relie ainsi l'espérance de Y_i à η_i par : $g(\mu_i) = \eta_i$.

Deux résultats importants de cette méthode sont les relations :

$$\mu_i = E(Y_i) = \int_R y_i \cdot f(y_i; \theta_i; \varphi) \cdot dy_i = b'(\theta_i) \text{ et}$$

$$V(Y_i) = \int_R y_i^2 \cdot f(y_i; \theta_i; \varphi) \cdot dy_i - \left(\int_R y_i \cdot f(y_i; \theta_i; \varphi) \cdot dy_i \right)^2 = b''(\theta_i) \varphi = V(\mu) \varphi$$

En inversant la relation de l'espérance nous obtenons également : $\theta_i = (b')^{-1}(E(Y_i))$,

$$\text{Et donc : } \theta_i = (b')^{-1} \circ g^{-1}(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$$

Ainsi il devient possible d'estimer le paramètre θ_i à partir de la combinaison linéaire

$$\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k.$$

Estimation des paramètres d'un GLM

Le vecteur de paramètres $\beta = (\beta_1, \dots, \beta_k)$ ainsi que le paramètre de dispersion ϕ sont estimés par la méthode du maximum de vraisemblance. Ceux-ci sont donc asymptotiquement distribués selon une loi normale. En supposant les variables réponses $y_1, \dots, y_n$ indépendantes, nous avons par définition la vraisemblance du modèle :

$$L(y, \theta, \varphi) = \prod_{i=1}^n f(y_i; \theta_i; \varphi)$$

$$\text{et la log-vraisemblance : } l(y; \theta; \varphi) = \ln(\prod_{i=1}^n f(y_i; \theta_i; \varphi)) = \sum_{i=1}^n l_i(y_i; \theta_i; \varphi)$$

$$\text{avec } l_i(y_i; \theta_i; \varphi) = \ln(f(y_i; \theta_i; \varphi)) = \frac{y_i \theta_i - b(\theta_i)}{a(\varphi)} + c(y_i; \varphi).$$

L'estimateur $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_k)$ de $\beta = (\beta_1, \dots, \beta_k)$ est alors solution de :

$$\sum_{i=1}^n \frac{\partial l_i}{\partial \beta_j}(\hat{\beta}_j) = 0 \text{ pour } j = 1, \dots, k$$

Or il n'y a pas de calculs formels directs pour résoudre ce système d'équations, il faut passer par une méthode numérique. Deux méthodes numériques sont largement utilisées dans ce cadre : la méthode du score de Fisher et la méthode de Newton-Raphson utilisée par défaut par le logiciel SAS dans la procédure dédiée aux modèles GLM (GENMOD). Cette dernière repose sur la matrice Hessienne H des dérivées croisées des paramètres et le vecteur gradient s des dérivées des paramètres :

$$s = [s_j] = \left[\frac{\partial L}{\partial \beta_j} \right] \text{ et } H = [h_{ij}] = \left[\frac{\partial^2 L}{\partial \beta_i \partial \beta_j} \right]$$

$$\text{L'estimation de } s \text{ est donnée par : } s = \sum_{i=1}^n \frac{(y_i - \mu_i) y_i}{V(\mu_i) g'(\mu_i) \varphi}$$

$$\text{et } H \text{ est donnée par le produit matriciel : } H = -X' W_0 X$$

où X est la matrice des variables explicatives et W_0 est une matrice diagonale dont le i ème élément de la diagonale est donnée par :

$$w_{0i} = w_{ei} + (y_i - \mu_i) \frac{V(\mu_i)g''(\mu_i) + V'(\mu_i)g'(\mu_i)}{(V(\mu_i))^2 (g'(\mu_i))^3 \varphi}$$

avec $w_{ei} = \left(\varphi V(\mu_i) (g'(\mu_i))^2 \right)^{-1}$

Après avoir initialisé les valeurs du vecteur β à $\hat{\beta}_0$, la $r + 1$ ème itération est estimée par :

$$\hat{\beta}_{r+1} = \hat{\beta}_r - H^{-1}s$$

L'algorithme se poursuit jusqu'à ce que $\hat{\beta}_{r+1} - \hat{\beta}_r \leq \alpha$ où α est la valeur de précision pour laquelle l'estimation $\hat{\beta}$ du vecteur β sera jugée acceptable. $\hat{\beta}_0$ et α sont préétablis dans le logiciel SAS.

Valeur prédite et intervalle de confiance

Puisque nous avons $\hat{\beta}$, nous obtenons immédiatement $\hat{\mu}_i = g^{-1}(X_i' \hat{\beta})$

La variance du prédicteur linéaire η_i est donnée par : $\sigma_{\eta_i}^2 = X_i' \Sigma X_i$

où Σ est la matrice de covariance de $\hat{\beta}$.

Ainsi, il est possible de donner un intervalle de confiance autour de μ_i avec une précision de $(1 - \alpha)$ par :

$$g^{-1} \left(X_i' \hat{\beta} \pm z_{1-\alpha/2} \sigma_{\eta_i} \right)$$

où $z_{1-\alpha/2}$ représente le quantile d'ordre $1 - \alpha/2$ de la loi normale standard.

Quelques lois et relations disponibles avec les GLM

Voici les principales lois et relations valables directement sous la théorie des GLM.

➤ Loi Normale :

Densité : $f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right) y \in]-\infty; +\infty[$

Fonction lien canonique : $g(\mu) = \mu$

Dispersion : $\varphi = \sigma^2$

Variance : $V(Y) = \sigma^2$

Remarque : en transformant la variable réponse par la fonction logarithme, il est possible d'utiliser alors la loi log-normale pour modéliser la variable réponse.

➤ Loi Binomiale :

Densité : $f(y) = \binom{n}{r} \mu^r (1 - \mu)^{n-r}$ pour $y = \frac{r}{n}, r = 0, 1, \dots, n$

Fonction lien canonique : $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$

Dispersion : $\varphi = 1$

Variance : $V(Y) = \frac{\mu(1-\mu)}{n}$

Remarque : Ce type de GLM est aussi appelé modèle logistique.

➤ Loi de Poisson :

Densité : $f(y) = \frac{\mu^y e^{-\mu}}{y!}$ pour $y \in \mathbb{N}$

Fonction lien canonique : $g(\mu) = \ln(\mu)$

Dispersion : $\varphi = 1$

Variance : $V(Y) = \mu$

Remarque : il est possible d'étendre cette approche en modifiant la valeur du paramètre de dispersion φ . La variance deviendra $V(Y) = \varphi\mu$. Il s'agira alors d'un modèle dit Poisson sur-dispersé ou sous-dispersé suivant la valeur de φ . Cette modification est aussi valable pour la loi binomiale.

➤ Loi Binomiale Négative :

Densité : $f(y) = \frac{\Gamma((y+1)k^{-1})}{\Gamma(y+1)\Gamma(k^{-1})} \frac{(k\mu)^y}{(1+k\mu)^{\frac{y+1}{k}}}$ pour $y \in \mathbb{N}$

Fonction lien canonique : $g(\mu) = \ln(\mu)$

Variance : $V(Y) = \mu(1 + k\mu)$

Concernant l'objet de cette partie, nous avons les éléments suffisants pour modéliser les triangles de liquidation. Mais il existe plusieurs extensions à la théorie des GLM dont certaines seront développées et utilisées dans le modèle détaillé que nous proposons.

Tests d'adéquation

Plusieurs outils sont utilisés pour permettre de valider la pertinence du modèle. En premier lieu voici un rappel des trois tests classiquement utilisés qui permettent d'estimer l'apport des variables explicatives au modèle en testant $H_0: \beta_j = 0$ contre $H_1: \beta_j \neq 0$

- Test de Wald : la statistique de Wald s'écrit $W = \frac{\hat{\beta}^2}{v(\hat{\beta})}$ où W suit une loi du $\chi^2(1)$ sous H_0 . Cette méthode utilise les paramètres estimés par le modèle et l'estimation de leurs variances d'estimations.
- Test du score : Ce test utilise la dérivée de la vraisemblance sous H_0 . Sa statistique s'écrit : $score = \left[\frac{\partial L}{\partial \beta} \right]'_{\beta=\hat{\beta}_{H_0}} \left[\frac{\partial^2 L}{\partial \beta^2} \right]^{-1}_{\beta=\hat{\beta}_{H_0}} \left[\frac{\partial L}{\partial \beta} \right]_{\beta=\hat{\beta}_{H_0}}$

Celle-ci suit asymptotiquement une loi du $\chi^2(1)$ sous H_0 .

- Test de rapport de vraisemblance : Dans un cadre général, ce test consiste à comparer deux hypothèses de paramétrage d'un modèle entre $H_0: \beta = \beta_0$ et $H_1: \beta = \beta_1$. Pour un modèle GLM, l'hypothèse $H_1: \beta = \hat{\beta}$ sans restriction est comparée à l'hypothèse $H_0: \beta = \tilde{\beta}$ où m paramètres sont neutralisés en prenant la valeur 0.

Soit $l_0(y, \theta, \varphi)$ la log-vraisemblance du modèle sous H_0 et $l_1(y, \theta, \varphi)$ la log-vraisemblance du modèle sous H_1 . La statistique Λ du test s'écrit :

$$\Lambda(y) = 2(l_1(y, \theta, \varphi) - l_0(y, \theta, \varphi))$$

Λ suit asymptotiquement une loi du χ^2 à m degrés de liberté sous l'hypothèse H_0 . Si la valeur s est le résultat au calcul de $\Lambda(y)$, alors H_0 est rejeté si, avec un seuil de certitude à $1 - \alpha$: $P_{\chi_m^2}(\Lambda(y) \leq s | H_0) < \alpha$

A partir de ce test, il devient possible de mesurer la contribution d'une variable explicative X_i au modèle.

Le critère d'information d'Akaike :

Le critère d'information d'Akaike (AIC) va être utile car il permet de fournir une mesure de performance du modèle sur la base uniquement de sa log-vraisemblance et du nombre de

paramètres estimés, alors que le test du rapport de vraisemblance ne permet de comparer des modèles que deux à deux et avec la contrainte que l'un des modèles soit identique à l'autre avec pour seule différence des paramètres en moins. Avec le AIC, il est possible de comparer des modèles cherchant à expliquer le comportement d'une même variable réponse y mais qui peuvent être de structures et d'hypothèses relativement différentes. Le AIC s'écrit ainsi :

$$AIC = -2 \ln(L) + 2k$$

où L est la vraisemblance du modèle et k le nombre de paramètres estimés dans le modèle.

Le modèle préféré selon ce critère sera celui qui aura le AIC le plus faible.

Statistiques de déviance

Ces valeurs sont également utiles pour évaluer la bonne adéquation d'un modèle GLM. Il en existe deux :

La statistique du Chi-deux de Pearson est défini par $\chi^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}$. Celle-ci suit une loi du Chi-deux à $n - p$ degrés de liberté où n est le nombre d'observations et p le nombre de paramètres estimés. La statistique du Chi-deux de Pearson standardisée est $\chi^{2*} = \frac{\chi^2}{\phi}$

La statistique de la Déviance D est issu d'un test de rapport de vraisemblance entre le modèle estimé et le modèle dit saturé contenant autant de paramètres qu'il y a d'observations.

$D = 2(l(y, y) - l(y, \mu))$. Lorsque ϕ est connu, nous pouvons définir la déviance standardisée par $D^* = \frac{D}{\phi}$. Lorsque ϕ est inconnu, D n'est pas considéré comme une statistique.

En pratique, si D^* et χ^{2*} sont tops éloignés de 1, cela révélera un phénomène de sur-dispersion ou de sous-dispersion. Dans ce cas, et lorsque la loi utilisée est loi binomiale ou la loi de Poisson, il est possible d'étendre la théorie des GLM pour intégrer ce phénomène en modifiant la valeur de ϕ pour ces deux lois par :

$\hat{\phi} = \frac{D}{n-p}$ en utilisant la statistique de la déviance ou

$\hat{\phi} = \frac{\chi^2}{n-p}$ en utilisant la statistique du Chi-deux de Pearson.

D'autre part, des valeurs éloignées entre D^* et χ^{2*} est un indice de mauvaise adéquation du modèle.

Les résidus :

L'analyse des résidus est également un élément important de la validation d'un modèle GLM. En effet ceux-ci vont permettre de repérer les valeurs aberrantes dans la population observée, mais aussi de vérifier la compatibilité de la variable réponse avec les différentes caractéristiques du modèle comme les variables explicatives utilisées ou la loi statistique appliquée.

Le résidu de base, dit résidu brut, est défini comme : $r_i = y_i - \hat{\mu}_i$. C'est la différence entre la valeur observée et la valeur moyenne prévue par le modèle pour une observation i .

Le résidu de Pearson est défini comme la racine-carrée de la contribution de l'observation i à la statistique du Chi-deux de Pearson :

$$r_i^p = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}$$

La version standardisée du résidu de Pearson se définit comme :

$$r_i^{p*} = \frac{r_i^p}{\sqrt{\varphi(1 - h_i)}} = \frac{y_i - \hat{\mu}_i}{\sqrt{\varphi V(\hat{\mu}_i)(1 - h_i)}}$$

où $\varphi V(\hat{\mu}_i)$ est la variance de l'observation i et h_i est le levier associé à l'observation i .

Ainsi r_i^{p*} et r_i^p ont une espérance égale à 0 et r_i^{p*} a une variance égale à 1.

Le résidu de la déviance est défini de manière quasiment analogue que le résidu de Pearson comme étant la racine carrée de la contribution de l'observation i sur la statistique de la déviance, mais en prenant en compte également le signe du résidu brut :

$$r_i^D = \text{signe}(y_i - \hat{\mu}_i) \sqrt{d_i}$$

La version standardisée du résidu de la déviance s'écrit :

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

10.1.2 Simulation et méthode de Monte-Carlo

Tout d'abord, voici une présentation des principes de la méthode de Monte-Carlo qui va nous permettre de combiner l'ensemble des sous-modèles. (Saporta, 2008)

La méthode de Monte-Carlo est une méthode couramment utilisée pour réaliser des estimations numériques d'objets mathématiques complexes. Ce que nous entendons par complexe ici réside dans le fait qu'il n'existe pas de calcul formel possible pour ces objets mathématiques, ou bien lorsque, dans un cadre opérationnel, il est plus rapide d'en estimer une valeur approchée acceptable que d'en estimer la valeur exacte. Les différents secteurs utilisant des outils de mathématiques appliqués sont amenés à employer cette méthode comme en physique, chimie, biologie, sociologie, économie, finance...

Le principe de base est d'estimer la fonction f recherchée en réalisant un échantillonnage aléatoire de f . Le résultat est donc la réalisation d'une variable aléatoire. Comme la méthode s'appuie sur un processus aléatoire, deux estimations successives de la même fonction ne donneront pas

exactement le même résultat. C'est là où réside l'intérêt de la loi faible des grands nombres qui permet d'atténuer ce problème : lorsque l'on augmente le nombre de tirages aléatoires, on augmente également la précision du résultat.

Remarque : voici pour rappel le théorème de la loi faible des grands nombres

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires iid. Alors :

$$\forall \varepsilon > 0, \lim_{n \rightarrow +\infty} P \left(\left| \frac{X_1 + \dots + X_n}{n} - E(X) \right| \geq \varepsilon \right) = 0$$

Cette méthode peut servir à beaucoup de domaines différents des mathématiques appliquées comme le calcul d'intégration. En assurance, son utilisation est très utile également comme dans le cadre d'estimation de lois tronquées dont voici un exemple : un assureur a estimé une loi paramétrique s'ajustant convenablement aux montants des sinistres d'un de ces produits avant application des franchises ou des limites de garanties. Pour estimer son espérance de montants de sinistres, il lui faudra alors réaliser l'estimation de cette loi paramétrique tronquée sur les valeurs hautes par la limite de garantie, et sur les petites valeurs par la franchise. De plus, s'il souhaite connaître l'impact de la réassurance, il devra réaliser une nouvelle série de troncatures correspondant aux cessions de réassurance de ce produit. Le graphique suivant illustre cet exemple. Même s'il existe des calculs formels applicables dans des conditions particulières (Valleron et Kutek, 2003), dans un cadre général il est nécessaire d'utiliser une méthode d'estimation numérique pour résoudre ce type de problème. Ainsi nous allons tirer un n-échantillon de la loi paramétrique de base par une méthode de simulation adaptée, estimer pour chacun des tirages les différentes lois recherchées (montants de sinistres nets des conditions de la police de base, puis nets de réassurance), puis définir les principales caractéristiques d'espérance et de dispersion de ces lois tronquées.

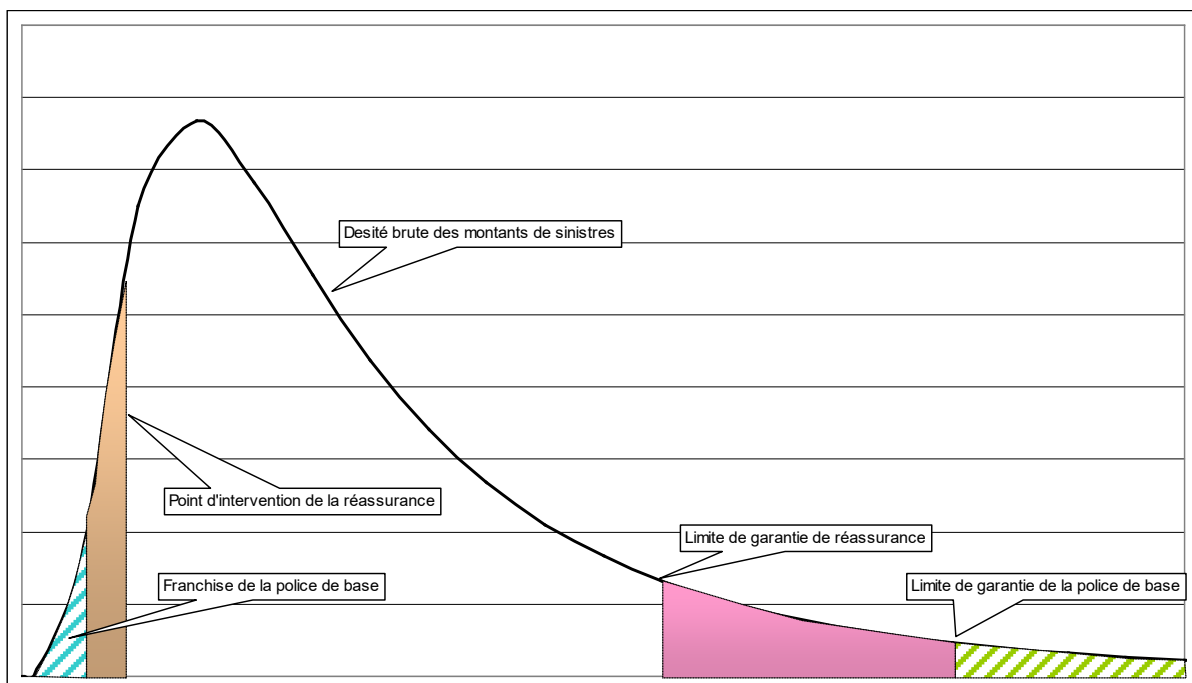


Figure 55: Exemple de densité tronquée

Il apparaît donc que l'une des principales applications de la méthode de Monte-Carlo en assurance est de pouvoir générer des échantillons aléatoires de lois adaptées au comportement de risques

sous-jacents afin de pouvoir réaliser des estimations de différentes transformations et combinaisons entre les risques.

D'autre part, les lois statistiques ne sont pas aisément échantillonnables à partir de leur forme de base, mais il existe des méthodes de simulation qui permettent de simplifier ce problème dont voici un rappel synthétique :

La méthode d'inversion

Soit $F(X)$ la fonction de répartition de la loi de X pour laquelle il est possible d'obtenir F^{-1} sous une forme analytique. Soit $U = F(X)$ qui est alors uniformément distribuée sur $[0,1]$. Il devient possible d'estimer X comme : $X = F^{-1}(U)$.

En tirant aléatoirement n valeurs u dans $[0,1]$, nous obtenons alors un n -échantillon de la loi de X , où $U \rightarrow Unif$.

La méthode du rejet

Soit f la densité de X , et m un majorant de f . On réalise un tirage aléatoire sur $[0,1]$ noté u , puis un tirage aléatoire sur $[0, m]$ noté v . Si $v < f(u)$, on conserve u qui est alors un échantillon de X .

La démonstration vient du fait que la probabilité conditionnelle que u soit gardé sachant que

$u = x$ est égale à $P(V < f(U)) = \frac{f(x)}{m}$, et inconditionnellement est égale à $\frac{1}{m}$. D'où :

$$P(x < U < x + dx / U \text{ est gardée}) = \frac{\frac{f(x)}{m} dx}{\frac{1}{m}} = f(x) dx$$

Afin d'optimiser cette méthode, on cherche à utiliser une densité de probabilité d'une variable aléatoire Y que l'on sait simuler et qui a pour propriété : $\frac{f}{g} \leq c$ où c est une constante.

On génère alors le couple $(y; u)$ avec u uniforme. Si $u < \frac{f(y)}{cg(y)}$, y est considéré comme un échantillon de la loi de X , sinon on rejette y et l'on recommence l'opération.

Quelques lois classiques nécessitent l'utilisation de méthodes de simulation particulières, nous rappellerons ici uniquement celles nécessaires aux lois utilisées dans notre étude.

Simulation d'une loi de Poisson

Le principe est de simuler un processus de Poisson de cadence 1 sur une période égale à λ . Les temps d'attentes T_i entre deux événements sont indépendants et suivent une loi exponentielle de paramètre λ . Ainsi on génère des tirages dans cette loi exponentielle jusqu'à ce que la somme des valeurs tirées dépasse λ . La réalisation n de la variable de Poisson de paramètre λ est le plus grand entier n tel que $\sum_{i=1}^n -\ln r_i < \lambda$ ou de manière équivalente tel que $\prod_{i=1}^n r_i > e^{-\lambda}$.

Lorsque le paramètre λ de la loi de poisson à simuler est grand, il est préférable d'approximer cette loi par une loi normale.

Simulation d'une loi binomiale négative

En partant de la structure de cette loi comme étant une loi de Poisson mélange lorsque son paramètre λ suit une loi gamma, il suffit alors de simuler Y suivant une loi de gamma $Gam(n, 1)$ puis X qui suit une loi de Poisson de paramètre $\frac{Y}{1-p}$, alors X suit une loi binomiale négative de paramètres n et p .

Simulation d'une loi normale

La méthode la plus couramment utilisée pour simuler la loi normale est la méthode de Box et Müller. Elle part du théorème suivant :

Si U_1 et U_2 sont deux variables uniformes sur $[0,1]$ indépendantes, alors X_1 et X_2 définies par :

$$\begin{cases} X_1 = (-2 \ln U_1)^{1/2} \cos 2\pi U_2 \\ X_2 = (-2 \ln U_1)^{1/2} \sin 2\pi U_2 \end{cases} \text{ sont deux variables normales centrées réduites indépendantes.}$$

Donc en tirant un couple $(u_1; u_2)$ nous obtenons un 2-échantillon x_1 et x_2 de X qui suit une loi normale centrée réduite. Par transformation ensuite de la loi normale centrée réduite avec les paramètres μ et σ , il est possible d'en déduire des échantillons pour l'ensemble des lois normales.

Il existe une autre méthode dite méthode polaire de Marsaglia qui est une variante de la méthode de Box et Müller et qui permet d'éviter le calcul des fonctions cosinus et sinus, consommant moins de temps de calcul en principe.

En considérant U_1 comme une coordonnée d'abscisse et U_2 comme une coordonnée d'ordonnée nous avons le point d'origine du référentiel $(0; 0)$, le point $(U_1; 0)$, et le point $(U_1; U_2)$ qui forment un triangle rectangle. En posant $\rho^2 = U_1^2 + U_2^2 < 1$, et θ l'angle aléatoire en $(0; 0)$, nous avons $\cos \theta = \frac{U_1}{\rho}$ et $\sin \theta = \frac{U_2}{\rho}$. D'où le résultat suivant :

$$\begin{cases} x_1 = u_1 \sqrt{\frac{-2 \ln \rho^2}{\rho^2}} \\ x_2 = u_2 \sqrt{\frac{-2 \ln \rho^2}{\rho^2}} \end{cases}$$

Générateur de nombres pseudo-aléatoires

A travers les méthodes développées ci-dessus nous voyons clairement qu'un élément fondamental des simulations réside en la nécessité d'avoir un générateur de réalisation de la loi uniforme performant. Au-delà du débat sur la définition même de la notion d'aléa, il n'existe pas d'algorithmes informatiques capables de réaliser cet exercice dans des conditions nécessaires : les valeurs générées ne sont pas strictement aléatoires, mais seront qualifiées de pseudo-aléatoires.

Ainsi l'un des critères fondamentaux d'un générateur est d'avoir une périodicité la plus longue possible afin d'avoir un minimum de chance de voir se répéter des valeurs déjà tirées et donc de biaiser les résultats des simulations.

Les générateurs congruents linéaires sont très largement répandus et reposent sur la formule de récurrence suivante :

$$X_{n+1} = (aX_n + c) \bmod m$$

Des valeurs utilisables par exemple sont celles de la version dite « standard minimal » où

$a = 16807$, $c = 0$, et $m = 2^{31} - 1$.

La valeur X_0 est la graine de la suite, appelée « seed » en anglais, et permet de générer l'ensemble de la suite récursive. Il est intéressant de la fixer manuellement lorsque l'on souhaite répéter des simulations à l'identiques. Par défaut, les logiciels de statistiques génèrent la graine à partir de l'horloge interne de l'ordinateur utilisé.

Comme la quantité de réalisation de loi uniforme générée va être importante dans notre étude, nous allons utiliser les fonctions de simulations de SAS s'appuyant sur le générateur Mersenne Twister réputé très consistant et rapide : fonctions RAND().

Nous laissons le lecteur approfondir cette partie présentée synthétiquement ici, mais d'une problématique très riche, en se référant aux nombreux ouvrages traitant de la génération de nombres aléatoires.

10.1.3 Régression Logistique

Le modèle logistique fait parti des modèles GLM comme évoqué précédemment dans la synthèse théorique des modèles GLM. Ce type de modèle permet d'estimer la probabilité d'une population de prendre un certain nombre de valeur dénombrable et limitée. Une application courante est de déterminer la loi d'une variable aléatoire susceptible de prendre deux états comme « 0 » ou « 1 », ou encore « vrai » ou « faux ». Dans ces exemples la variable réponse suivra une loi de Bernoulli

qui est un cas élémentaire de la loi Binomiale. Si les états possibles de la variable réponse sont supérieurs à deux, la loi de cette variable réponse suivra alors une loi Binomiale classique. Dans ce cas, nous pouvons illustrer en exemple la loi de l'espèce de poisson pêché en rivière avec comme valeurs possibles « brochet », « sandre », « perche », « gardon », et « autres ».

Une autre application possible est d'estimer la masse d'observations ayant une surreprésentation non-négligeable dans une loi empirique continue et qui forment alors des masses de Dirac. C'est le cas dans notre étude sur le montant de la charge totale à l'ouverture d'un sinistre qui peut prendre soit une valeur forfaitaire (masse de Dirac), soit un montant non forfaitaire plus adapté à la réalité du dossier sinistre.

Ce type de régression utilise certaine caractéristique particulière que nous rappelons ici :

- La variable réponse suit une loi Binomiale
- La fonction de lien canonique est la fonction logit $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$.
- Dispersion : $\varphi = 1$
- Variance : $V(Y) = \frac{\mu(1-\mu)}{n}$

Remarque : D'autres fonctions liens peuvent être utilisées dans le cadre d'une régression logistique que nous citons ci-dessous uniquement pour information :

- La fonction probit : $g(\mu) = \Phi^{-1}(\mu)$ où Φ est la densité de la loi normale centrée réduite.
- La fonction log-log complémentaire : $g(\mu) = \ln(-\ln(1 - \mu))$

Comme type de modèle GLM, toute la théorie sur les estimations et les tests s'appliquent donc pour le modèle logistique. Il existe plusieurs outils complémentaires propres au modèle logistique pour en valider l'adéquation comme le test Hosmer et Lemeshow. Un autre outil intéressant est le graphique dit de la courbe ROC pour « receiver operating characteristic » en anglais qui montre le rapport entre la sensibilité et la spécificité d'une variable explicative X dans le modèle.

- La sensibilité représente la capacité du modèle à prédire un événement
- La spécificité représente la capacité du modèle à prédire un non-événement

La courbe ROC prend en abscisse la valeur $(1 - \text{la spécificité de } X)$ et en ordonnée la sensibilité de X . L'aire sous la courbe ROC est alors une mesure du pouvoir prédictif de la variable explicative X .

Le logiciel SAS offre de multiples procédures pour estimer un modèle logistique. Les procédures PROC LOGISTIC et PROC GENMOD sont les plus classiques.

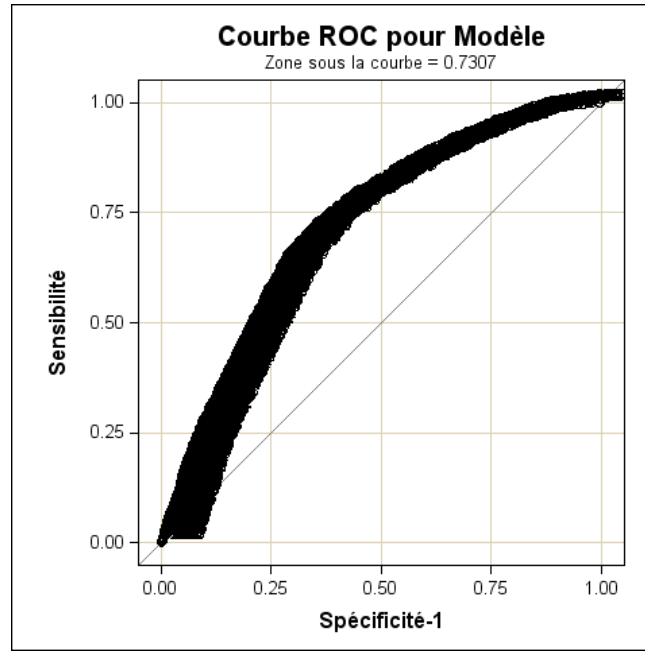


Figure 56: Exemple de courbe ROC

10.1.4 Régression Bêta

La régression Bêta consiste à ajuster une loi Bêta à un comportement observé en tenant compte de variables dépendantes (Smithson et Verkuilen, 2005). C'est une extension de la théorie des modèles GLM en appliquant une loi bêta à la variable réponse. Cette loi a des caractéristiques intéressantes pour modéliser certain phénomène puisqu'elle est définie sur $[0,1]$ et qu'elle est capable de prendre une grande variété de forme dans cet intervalle. Ainsi elle sera très utile pour modéliser des ratios.

La densité d'une variable aléatoire Y suivant une loi Bêta peut s'écrire ainsi :

$$f(y; \alpha; \beta) = \frac{1}{B(\alpha; \beta)} y^{\alpha-1} (1-y)^{\beta-1}$$

où $y \in [0,1]$, $\alpha, \beta > 0$ et $\frac{1}{B(\alpha; \beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ avec $\Gamma(\cdot)$ la fonction gamma.

L'espérance de Y est donnée par : $E(Y) = \frac{\alpha}{\alpha+\beta}$

La variance de Y est donnée par : $V(Y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

La méthode proposée par Smithson M. et Verkuilen J. (2005) permet d'étendre la méthode de régression des modèles linéaires généralisés à la loi bêta. En voici les principaux résultats, nous laissons le lecteur se référer à l'article original pour un approfondissement.

Les paramètres α et β sont tous les deux des paramètres de forme (« shape » en anglais). Or ceux-ci sont difficilement interprétables dans le cadre d'espérance conditionnelle. C'est pourquoi les paramètres α et β vont être transformés en paramètres de position et de dispersion.

Posons : $\mu = E(Y)$ et $\sigma^2 = V(Y)$

Et notons : $\varphi = \alpha + \beta$.

Alors : $\alpha = \mu\varphi$ et $\beta = \varphi - \mu\varphi$

Ainsi $\sigma^2 = \frac{\mu(1-\mu)}{1+\varphi}$.

φ est interprétable comme une variable de précision, car lorsque φ diminue, σ^2 augmente.

La log-vraisemblance de l'observation y_i est donnée par :

$$\ln L(\alpha; \beta; y_i) = \ln \Gamma(\alpha + \beta) - \ln \Gamma(\alpha) - \ln \Gamma(\beta) + (\alpha - 1) \ln(y_i) + (\beta - 1) \ln(1 - y_i).$$

Deux fonctions liens vont être utilisées, l'une pour le paramètre de position μ , l'autre pour le paramètre de précision φ .

Tout d'abord la fonction f définie vectoriellement par $f(\mu_i) = X_i\beta$ doit être monotone et différentiable. La variable réponse étant définie sur $[0,1]$, la fonction logit sera utilisée car c'est une fonction courante qui permet de respecter cette caractéristique. Les autres fonctions liens utilisées pour une régression logistique sont aussi applicables. Nous avons donc :

$$\ln\left(\frac{\mu_i}{1-\mu_i}\right) = X_i\beta.$$

Ainsi la régression Bêta utilise la même fonction de lien que la régression logistique sur son paramètre de position.

Par analogie avec $f(\mu_i)$, nous allons poser la fonction lien $h(\varphi_i)$ comme étant vectoriellement :

$$h(\varphi_i) = W_i\delta$$

En utilisant la fonction lien log, nous avons la relation suivante :

$$\ln(\varphi_i) = -W_i\delta$$

La forme négative est utilisée par soucis de clarté dans l'interprétation du paramètre δ_j : comme φ_i est un paramètre de précision, δ_j peut alors être vu comme un paramètre de dispersion dans cette configuration.

Nous obtenons ainsi deux sous-modèles :

$$\mu_i = \frac{\exp(X_i\beta)}{1+\exp(X_i\beta)} \text{ et } \varphi_i = \exp(-W_i\delta).$$

L'estimation de la log-vraisemblance, des différents paramètres et des erreurs d'estimations asymptotiques sont réalisés par approximation numérique avec la procédure PROC NLMIXED de SAS par la méthode de Newton-Raphson évoquée précédemment dans la synthèse théorique des modèles GLM.

10.1.5 Loi Normale Contaminée

En Anglais connue sous le nom de CND (contaminated normal distribution), cette loi est un cas particulier de mélange de lois normales : une CND est un mélange de deux lois normales ayant la même espérance. Cette loi est couramment appliquée pour améliorer la robustesse des régressions, « robust regression » en anglais, afin de filtrer les valeurs aberrantes d'une population de valeurs gaussiennes et d'en améliorer l'homogénéité.

Son utilisation dans le modèle est faite pour améliorer la qualité d'ajustement d'une loi normale sur une variable observée à caractère symétrique à la fois concentrées autour de sa moyenne, mais avec également un certain nombre d'éléments dispersés.

Sa densité peut s'écrire comme :

$$f(y_i) = p\Phi\left(\frac{y-\mu_i}{\sigma_1}\right) + (1-p)\Phi\left(\frac{y-\mu_i}{\sigma_2}\right) \text{ où } \Phi \text{ est la densité de la loi normale centrée réduite.}$$

Si Y suit une loi CND($\mu, \sigma_1, \sigma_2, p$), cela peut s'interpréter comme : Y suit une loi normale d'espérance μ avec la probabilité p que l'écart-type soit égal à σ_1 et la probabilité $(1-p)$ que l'écart-type soit égal à σ_2 .

La densité peut aussi être écrite sous une forme légèrement différente avec un seul paramètre d'écart-type en posant : $\sigma_2 = c\sigma_1$ où c est un réel.

L'estimation des paramètres est réalisée classiquement par l'estimation du maximum de vraisemblance. Toutefois, il n'y a pas de simplification de la formule de vraisemblance ou de log-vraisemblance, c'est-à-dire que l'on conserve la forme de base de la définition de la vraisemblance à partir de la densité, et que l'on réalise une estimation numérique des paramètres par une méthode récursive comme la méthode de Newton Raphson. Cette opération est réalisée à l'aide de la procédure PROC NLMIXED de SAS.

10.1.6 Modèle de survie des sinistres – Modèle à hasard proportionnel de Cox

Une des problématiques de base de la modélisation détaillée consiste à pouvoir caractériser la survie probable des sinistres, c'est-à-dire la persistance d'un sinistre dans les comptes de l'assureur. En cela nous souhaitons mesurer la durée entre la réalisation d'un événement pouvant générer contractuellement une prestation de l'assureur et la libération de tout engagement de l'assureur. Cette fin d'engagement peut intervenir à la suite du versement des prestations et des frais inhérents ou bien à l'invalidation complète ou partielle de l'engagement de l'assureur. Ainsi durant toute la durée du sinistre le potentiel de la charge totale du sinistre, c'est-à-dire la somme des prestations et des frais restants à verser doit être suffisamment provisionnée dans les comptes de l'assureur, et ce à n'importe quel instant de la vie du sinistre. Ainsi le modèle de survie doit caractériser la durée durant laquelle le sinistre ne lui est pas encore connu.

Les modèles de survies les plus connus sont le modèle de Kaplan-Meier et le modèle de Cox. Ils sont souvent utilisés dans le domaine médical afin de comparer des effets thérapeutiques sur des échantillons de populations. Nous proposons d'utiliser ici le modèle de Cox qui permet d'intégrer

différentes informations sur les caractéristiques des sinistres en tant que variables explicatives sans qu'elles soient contraintes comme constantes dans le temps. (Smith et Smith, 2002)

Généralités sur les modèles de survie

Dans un modèle de survie, la fonction de loi de survie $S(t)$ et la fonction de la loi de durée de vie $F(t)$ sont définies comme :

$$S(t) = P(T > t) = \int_t^{\infty} f(u) du = 1 - F(t)$$

où $S(t) = 1$ si $t = 0$ et $S(t) \rightarrow 0$ lorsque $t \rightarrow +\infty$.

La fonction de risque instantané de décès $h(t)$ est définie comme :

$$h(t) = P(t < T < t + \Delta / T > t)$$

$$\text{lorsque } \Delta \rightarrow 0, h(t) = \frac{P(t < T \leq t + \Delta)}{P(T > t)} = \frac{f(t)}{S(t)}.$$

La relation entre $h(t)$ et $S(t)$ peut être établie également ainsi :

$$h(t) = \frac{\frac{dF(t)}{dt}}{S(t)} = \frac{-\frac{dS(t)}{dt}}{S(t)} = -\frac{d}{dt} \ln S(t)$$

La fonction cumulative de risque instantané $H(t)$ est vue comme :

$$H(t) = \int_0^t h(u) du$$

Ainsi :

$$H(t) = -\ln S(t)$$

Le Modèle de hasard proportionnel de Cox

Dans un modèle linéaire, la variable résultat ou sa transformée peut s'écrire sous forme d'une équation linéaire de type $Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ où les X_i sont des variables explicatives continues ou indicelles afin de représenter des caractéristiques discrètes, et les β_i sont les coefficients de régressions à estimer. Le modèle de Cox reprend cette forme pour définir $h(t)$ comme une combinaison linéaire à valeur positive quelles que soient les valeurs prises par les X_i et les β_i . L'équation linéaire est alors transformée en relation log-linéaire afin de garantir la positivité de la variable résultat :

$$\log h(t/X) = \log h_0(t) + \beta_1 X_1 + \dots + \beta_k X_k \text{ ou sous forme matricielle : } h(t/X) = h_0(t) e^{\beta'X}.$$

La variable t représente le délai entre la date d'origine et la survenue de l'événement, ici la clôture du sinistre. Nous remarquons que seul $h_0(t)$ dépend de ce délai qui est le taux instantané lorsque le vecteur X est nul.

Nous remarquons ainsi la propriété de proportionnalité du modèle :

$$\frac{h(t/X)}{h_0(t)} = e^{\beta'X}$$

Ceci indique que le risque d'un individu est proportionnel au risque de base $h_0(t)$ et ne dépend pas du temps dans le modèle de Cox.

La fonction de survie $S(t/X)$ du modèle est définie par :

$$S(t/X) = S_0(t)e^{\beta'X}$$

Avec $S_0(t)$, la fonction de survie de base lorsque le vecteur X est nul, et égale à $e^{-H_0(t)}$.

Vraisemblance partielle de Cox

L'estimation du vecteur de paramètres β peut être réalisée par l'approche suivante :

Les n différentes durées observées sont triées par ordre croissant. $r(t_i)$ représente l'ensemble des sinistres encore présents au bout de $t_i - \varepsilon$ jours, et utilisée pour se placer juste avant la clôture du sinistre i en t_i .

Les données censurées représentent les sinistres non clôturés à la date de fin d'observation de l'étude : il s'agit d'une censure à droite. Comme la date de survenance du sinistre est connue par définition à l'ouverture du sinistre, il n'y a pas de censure à gauche. Les éléments censurés sont repérés par la variable booléenne δ_i qui prend la valeur 0 si le sinistre i n'est pas clos à la fin de la période d'observation, sinon elle prend la valeur 1.

$$\text{Pour un sinistre } j \in r(t_i) : P(T_j \leq t_i / j \in r(t_i)) = \frac{h_0(t_i)e^{\beta'X_j}}{\sum_{k \in r(t_i)} h_0(t_i)e^{\beta'X_k}} = \frac{e^{\beta'X_j}}{\sum_{k \in r(t_i)} e^{\beta'X_k}}$$

La vraisemblance partielle de Cox est le produit de ces probabilités pour l'ensemble des clôtures :

$$L(\beta) = \prod_{i=1}^n \left(\frac{e^{\beta'X_i}}{\sum_{k \in r(t_i)} e^{\beta'X_k}} \right)^{\delta_i}$$

On en déduit la log-vraisemblance de Cox :

$$l_n(\beta) = \ln(L(\beta)) = \sum_{i=1}^n \delta_i (\beta'X_i - \ln(\sum_{k \in r(t_i)} e^{\beta'X_k})).$$

Mais cette approche repose sur le fait que chaque t_i ne correspond qu'à une observation unique. Or la granularité du modèle, le nombre de jours en l'occurrence, n'est pas continu et ne permet donc pas de valider cette hypothèse. C'est pourquoi l'approximation $l_n^*(\beta)$ de Breslow (1974) sera utilisée pour réaliser ce calcul :

$$l_n^*(\beta) = \sum_{i=1}^n \left(\beta \left(\sum_{j=1}^{d_i} X_j - d_i \ln \left(\sum_{j \in r(i)} e^{\beta' X_j} \right) \right) \right)$$

$$\hat{\beta} = \operatorname{argmax} l_n^*(\beta) \Leftrightarrow \frac{dl_n^*(\beta)}{d\beta} = 0$$

Cette estimation sera réalisée par calcul numérique.

Nous noterons que la vraisemblance de Cox est qualifiée de partielle car cette méthode fournit les estimations du vecteur de paramètres β sans chercher à caractériser le risque de base $h_0(t)$. De plus le modèle de Cox est un modèle semi-paramétrique puisqu'il ne pose aucune hypothèse sur la forme de $h_0(t)$.

Tests d'adéquations dans le cadre d'un modèle à hasard proportionnel

La trilogie des tests classiques sont applicables pour ce type de modèle : test de Wald, test de rapport de vraisemblance, et test du score. Ceux-ci sont donc très utiles pour vérifier la pertinence des variables explicatives utilisées. L'importante différence d'un modèle de Cox par rapport à un modèle GLM traditionnel est l'hypothèse de hasard proportionnel du risque de base. Celle-ci est donc à valider par un des tests suivants :

Un premier test possible consiste à réaliser une régression avec chacune des variables explicatives associées à une covariable composée de cette même variable et de la variable temps. Ceci donne le modèle suivant :

$$h(t/X_i) = h_0(t) \exp(\beta_1 X_i + \beta_2 X_i t)$$

Ainsi si l'hypothèse de hasard proportionnel est respectée, le paramètre β_2 ne doit pas être significatif. En d'autres termes l'information apportée par $X_i t$ est négligeable face l'information déjà fournie par X_i seule.

Un deuxième test possible est une approche graphique à partir du logarithme de la fonction cumulative du risque de base. Les courbes $t = \ln(\hat{H}(t))$ ou de manière équivalente

$t = \ln(-\widehat{\ln S}(t))$ des différentes variables quantitatives ou de chaque classe des variables de classes doivent être parallèles.

Les différents tests présentés ici sont finalement empiriques et permettent d'avoir un résultat uniquement heuristique sur la propriété de hasard proportionnel des données utilisées.

Simulation du temps de survie conditionnelle

Comme nous allons réaliser une approche d'estimation numérique par simulation du modèle de provisionnement détaillé, nous allons donc devoir réaliser la simulation du modèle de Cox. La méthode d'inversion s'applique directement à ce modèle. L'objectif est de connaître la durée t restant à vivre sachant que le sinistre est survenu depuis t_0 , avec $t_0 \geq 0$. T étant une variable aléatoire ayant comme fonction de répartition $F(t/X)$, nous pouvons définir une variable aléatoire U ainsi :

$$U = F(t_0 + t/T \geq t_0/X) = P(T \leq t_0 + t/T \geq t_0/X) \text{ où } U \mapsto \text{Unif}[0,1]$$

Si nous posons $F(t/X) = 1 - S(t/X) = 1 - \exp(-H_0(t)e^{\beta'X})$, alors U peut s'écrire comme :

$$\begin{aligned} U &= \frac{P(t_0 < T < t_0 + t/X)}{P(T \geq t_0/X)} = \frac{F(t_0 + t/X) - F(t_0/X)}{S(t_0/X)} \text{ d'où} \\ U &= \frac{(1 - \exp(-H_0(t_0 + t)e^{\beta'X})) - (1 - \exp(-H_0(t_0)e^{\beta'X}))}{\exp(-H_0(t_0)e^{\beta'X})} \\ &= 1 - \frac{\exp(-H_0(t_0 + t)e^{\beta'X})}{\exp(-H_0(t_0)e^{\beta'X})} \end{aligned}$$

U étant une variable uniforme et donc symétrique, alors il est équivalent d'écrire

$$U = \frac{\exp(-H_0(t_0 + t)e^{\beta'X})}{\exp(-H_0(t_0)e^{\beta'X})}$$

En passant au logarithme nous obtenons :

$$\ln(U) = -H_0(t_0 + t)e^{\beta'X} + H_0(t_0)e^{\beta'X}$$

et donc :

$$t + t_0 = H_0^{-1} \left(H_0(t_0) - \frac{\ln(U)}{e^{\beta'X}} \right)$$

Ainsi nous pouvons donc simuler $\forall t_0 \geq 0$ la distribution du temps restant avant la clôture d'un sinistre sachant qu'il est survenu depuis t_0 jours, mais à condition de caractériser le risque de base.

Estimation du risque de base

Nous utiliserons la méthode avancée par Breslow pour estimer le risque de base. $\hat{S}_0(t)$ est estimée par la formule suivante :

$$\hat{S}_0(t) = \prod_{t_i < t} 1 - \frac{\delta_i}{\sum_{k \in r(t_i)} e^{X'_k \beta}}$$

Ajustement paramétrique du risque de base

Afin d'effectuer efficacement ces simulations, il paraît préférable de basculer dans un univers paramétrique.

D'après Lee ET, Go OT (1997), les lois Exponentiel, de Weibull, et de Gompertz respectent l'hypothèse de hasard proportionnel du modèle de Cox. Après ajustement de ces lois à la survie du risque de base, la loi de Weibull se révèle la plus appropriée à nos données (voir annexes). Ses paramètres sont estimés numériquement par la méthode des moindres carrés.

Communément, la densité d'une variable aléatoire T qui suit une loi de Weibull s'écrit :

$$f_0(t) = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha}$$

Si nous posons $v = \alpha$, $\lambda = \left(\frac{1}{\beta}\right)^\alpha$, et $\lambda, v > 0$ alors nous obtenons les relations suivantes sur le risque de base :

Densité : $f_0(t) = \lambda v t^{v-1} e^{-\lambda t^v}$

Fonction de survie : $S_0(t) = e^{-\lambda t^v}$

Fonction de hasard : $h_0(t) = \lambda v t^{v-1}$

Fonction cumulative de risque instantané : $H_0(t) = \lambda t^v$ et $H_0^{-1}(t) = \left(\frac{t}{\lambda}\right)^{\frac{1}{v}}$

Espérance : $E(T) = \frac{1}{\sqrt[v]{\lambda}} \Gamma\left(\frac{1}{v} + 1\right)$ et Variance : $V(T) = \frac{1}{\sqrt[v]{\lambda^2}} \left(\Gamma\left(\frac{2}{v} + 1\right) - \Gamma^2\left(\frac{1}{v} + 1\right) \right)$ où Γ est la fonction gamma.

Ainsi la simulation de $t + t_0$ par ajustement à la loi de Weibull devient :

$$t + t_0 = \left(t_0^v - \frac{\ln(U)}{\lambda e^{\beta'X}} \right)^{\frac{1}{v}}$$

10.2 Tableaux de résultats des régressions statistiques et des ajustements paramétriques du modèle sur données détaillées des sinistres

Estimation du nombre de sinistres IBNYR restant à déclarer

Paramètres estimés par l'analyse du maximum de vraisemblance		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		0	0	0	0	0	.	.
LN_DAY_COVER_OUT		1	0.4574	0.0096	-0.4762	-0.4385	2254.55	<.0001
TYPE_C_USED	COINS	1	-0.4866	0.0498	-0.5842	-0.389	95.48	<.0001
TYPE_C_USED	LEAD	1	-1.5248	0.052	-1.6268	-1.4227	858.27	<.0001
LN_CLAIMS_CNT_OCC		1	0.7094	0.0105	0.6888	0.73	4557.5	<.0001
Dispersion		1	0.0385	0.0009	0.0367	0.0403		

Tableau 21: Modèle du nombre de sinistres restant à déclarer – Contrats en cours de couverture - Résultats de la régression

Paramètres estimés par l'analyse du maximum de vraisemblance		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		0	0	0	0	0	.	.
LN_DAY_COVER		1	0.648	0.0597	0.531	0.765	117.78	<.0001
TYPE_C_USED	COINS	1	5.1212	0.3407	4.4535	5.789	225.96	<.0001
TYPE_C_USED	LEAD	1	3.4795	0.3382	2.8166	4.1424	105.84	<.0001
LN_DAY_CLAIMS_WATCH_		1	-0.595	0.0219	-0.6379	-0.5521	739.43	<.0001
LN_RATIO_CLAIMS		1	0.7718	0.0117	0.7489	0.7948	4342.47	<.0001
Dispersion		1	0.046	0.0012	0.0436	0.0483		

Tableau 22: Modèle du nombre de sinistres restant à déclarer – Contrats expirés - Résultats de la régression

Estimation du délai de déclaration des sinistres

Paramètres estimés par l'analyse du maximum de vraisemblance		DDL	Valeur estimée	Erreur type	Intervalle de de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		1	-2.1017	0.0188	-2.1386	-2.0649	12507.8	<.0001
TYPE_C_USED	COINS	1	1.2761	0.0207	1.2355	1.3167	3797.64	<.0001
TYPE_C_USED	LEAD	0	0	0	0	0	.	.
KZ_VERSFALL	C	1	-0.3803	0.0213	-0.4221	-0.3385	318.44	<.0001
KZ_VERSFALL	E	0	0	0	0	0	.	.
Scale		1	1.6729	0.0071	1.6591	1.6868		

Tableau 23: Modèle d'estimation du délai de déclaration des sinistres - Résultats de la régression

Probabilités observés du couple garantie / type de dommage

GD (type de dommage et garantie en jeu)	Fréquence	Pourcentage	Fréquence Cumulée	Pctage. Cumulé
Corporel / RC employeurs	74	5.5	74	5.5
Corporel / RC activités Médical	18	1.3	92	6.9
Corporel / RC Produit	39	2.9	131	9.8
Corporel / RC Exploitation	90	6.7	221	16.5
Corporel / Autre	2	0.2	223	16.7
Perte Financière / RC activités Médical	1	0.1	224	16.7
Perte Financière / RC Produit	49	3.7	273	20.4
Perte Financière / RC Exploitation	15	1.1	288	21.5
Perte Financière / Autre	67	5.0	355	26.5
Matériel / RC Produit	197	14.7	552	41.2
Matériel / RC Exploitation	237	17.7	789	58.9
Matériel / Autre	47	3.5	836	62.4
Autre / RC employeurs	1	0.1	837	62.5
Autre / RC Produit	284	21.2	1,121	83.7
Autre / RC Exploitation	88	6.6	1,209	90.3
Autre / Autre	130	9.7	1,339	100.0

Tableau 24: Probabilités observés du couple garantie/type de dommages pour la classe de contrats B

GD (type de dommage et garantie en jeu)	Fréquence	Pourcentage	Fréquence Cumulée	Pctage. Cumulé
Corporel / RC activités Médical	13	86.7	13	86.7
Autre / RC activités Médical	2	13.3	15	100.0

Tableau 25: Probabilités observés du couple garantie/type de dommages pour la classe de contrats C

GD (type de dommage et garantie en jeu)	Fréquence	Pourcentage	Fréquence Cumulée	Pctage. Cumulé
Corporel / RC employeurs	8	0.5	8	0.5
Corporel / RC activités Médical	7	0.4	15	0.8
Corporel / RC Produit	162	9.1	177	10.0
Corporel / RC Exploitation	161	9.1	338	19.0
Perte Financière / RC Produit	91	5.1	429	24.1
Perte Financière / RC Exploitation	25	1.4	454	25.5
Perte Financière / Autre	23	1.3	477	26.8
Matériel / RC Produit	468	26.3	945	53.2
Matériel / RC Exploitation	519	29.2	1,464	82.3
Matériel / Autre	66	3.7	1,530	86.1
Autre / RC employeurs	20	1.1	1,550	87.2
Autre / RC activités Médical	1	0.1	1,551	87.2
Autre / RC Produit	111	6.2	1,662	93.5
Autre / RC Exploitation	10	0.6	1,672	94.0
Autre / Autre	106	6.0	1,778	100.0

Tableau 26: Probabilités observés du couple garantie/type de dommages pour la classe de contrats D

Estimation du type de sinistres survenus et non déclarés

Paramètres		DDL	Valeur estimée	Erreur	Khi-2 de Wald	
Corporel / RC employeurs		1	-3.7077	0.1623	521.9286	<.0001
Corporel / RC activités Médical		1	-3.4591	0.1463	558.9091	<.0001
Corporel / RC Produit		1	-2.8652	0.1165	604.9799	<.0001
Corporel / RC Exploitation		1	-2.3966	0.0997	578.2841	<.0001
Corporel / Autre		1	-2.3881	0.0994	577.198	<.0001
Perte Financière / RC Produit		1	-2.0876	0.0911	525.0188	<.0001
Perte Financière / RC Exploitation		1	-2.0526	0.0902	517.289	<.0001
Perte Financière / Autre		1	-1.1226	0.0739	230.8428	<.0001
Matériel / RC Produit		1	-0.2098	0.0673	9.7155	0.0018
Matériel / RC Exploitation		1	0.3672	0.0678	29.3231	<.0001
Matériel / Autre		1	0.3773	0.0679	30.9235	<.0001
Autre / RC employeurs		1	0.3874	0.0679	32.5623	<.0001
Autre / RC Produit		1	1.2066	0.0754	256.1321	<.0001
Autre / RC Exploitation		1	1.7824	0.0856	433.8357	<.0001
TYPE_C_USED	COINS	1	-0.9671	0.0583	275.2416	<.0001
KZ_VERSFALL	C	1	-0.0441	0.0554	0.6354	0.4254

Tableau 27: Modèle de type de sinistres IBNYR - Classe de contrats A - Résultats de la régression

Paramètres		DDL	Valeur estimée	Erreur	Khi-2 de Wald	
Corporel / RC employeurs		1	-3.795	0.1442	692.6161	<.0001
Corporel / RC activités Médical		1	-3.5537	0.1347	695.9846	<.0001
Corporel / RC Produit		1	-3.1503	0.1221	665.1944	<.0001
Corporel / RC Exploitation		1	-2.5124	0.1088	533.045	<.0001
Corporel / Autre		1	-2.5007	0.1086	529.9204	<.0001
Perte Financière / RC activités Médical			-2.4949	0.1085	528.3599	<.0001
Perte Financière / RC Produit		1	-2.2297	0.1048	452.7928	<.0001
Perte Financière / RC Exploitation		1	-2.1542	0.1039	430.25	<.0001
Perte Financière / Autre		1	-1.8368	0.1005	333.9242	<.0001
Matériel / RC Produit		1	-1.0029	0.0948	111.8781	<.0001
Matériel / RC Exploitation		1	0.1509	0.092	2.6904	0.101
Matériel / Autre		1	0.4715	0.0927	25.8732	<.0001
Autre / RC employeurs		1	0.479	0.0927	26.6855	<.0001
Autre / RC Produit		1	2.2244	0.1095	412.7967	<.0001
Autre / RC Exploitation		1	2.8742	0.1239	538.3386	<.0001
TYPE_C_USED	COINS	1	-1.5042	0.0649	536.335	<.0001
KZ_VERSFALL	C	1	-0.0612	0.0784	0.6092	0.4351

Tableau 28: Modèle de type de sinistres IBNYR - Classe de contrats B - Résultats de la régression

Estimation de la charge indemnitaire forfaitaires à l'ouverture des sinistres

Paramètre		DDL	Valeur estimée	Erreur	Khi-2 de Wald	
Intercept	-1	1	-3.0436	0.0421	5232.499	<.0001
Intercept	1	1	-2.7994	0.0414	4566.776	<.0001
Intercept	500	1	-2.2857	0.0404	3208.066	<.0001
Intercept	1,000	1	-1.8285	0.0396	2127.676	<.0001
Intercept	2,500	1	-1.1953	0.0389	944.0942	<.0001
Intercept	5,000	1	-0.5701	0.0385	219.827	<.0001
Intercept	7,500	1	-0.4318	0.0384	126.4161	<.0001
Intercept	10,000	1	0.0898	0.0384	5.4663	0.0194
Intercept	15,000	1	0.2765	0.0385	51.5569	<.0001
Intercept	20,000	1	0.8385	0.0392	457.188	<.0001
Intercept	30,000	1	1.1013	0.0398	764.6165	<.0001
Intercept	50,000	1	1.8551	0.043	1860.248	<.0001
Intercept	75,000	1	2.1557	0.0451	2285.728	<.0001
Intercept	100,000	1	3.1527	0.0571	3045.652	<.0001
Intercept	250,000	1	4.4858	0.0935	2299.668	<.0001
Intercept	500,000	1	5.4869	0.1469	1394.705	<.0001
Intercept	1,000,000	1	7.7705	0.4477	301.2656	<.0001
MAIN_TYP_BEZ_40	RC des employeurs	1	-0.7093	0.0576	151.8059	<.0001
MAIN_TYP_BEZ_40	activités médicales	1	0.1136	0.0879	1.6709	0.1961
MAIN_TYP_BEZ_40	RC Produit	1	-0.1224	0.0318	14.7783	0.0001
MAIN_TYP_BEZ_40	RC Exploitation	1	0.507	0.0353	206.4298	<.0001
LN_DAY_OCC		1	0.1955	0.00658	881.5846	<.0001
LN_CUMUL_PAY_COSTS		1	0.4002	0.0123	1060.55	<.0001

Tableau 29: Modèle d'estimation du montant d'ouverture forfaitaire des sinistres - Résultats de la régression

Estimation de la charge indemnitaire non forfaitaires à l'ouverture des sinistres

Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance à 95%		Khi-2 de Wald	Pr > Khi-2
Intercept		1.0000	9.2389	0.1162	9.0111	9.4667	6,320.2	<.0001
MAIN_CAUSE_BEZ_40	Corporel	1.0000	1.2306	0.2564	0.7280	1.7332	23.0	<.0001
MAIN_CAUSE_BEZ_40	Pertes Financiers	1.0000	1.2741	0.2305	0.8223	1.7259	30.6	<.0001
MAIN_CAUSE_BEZ_40	Autre/Non précisé	1.0000	-0.3807	0.1213	-0.6185	-0.1429	9.8	0.0017
MAIN_CAUSE_BEZ_40	Matériel	0.0000	0.0000	0.0000	0.0000	0.0000	.	.
LN_CUMUL_PAY_COSTS		1.0000	-0.1431	0.0141	-0.1708	-0.1154	102.6	<.0001
Scale		1.0000	2.2657	0.0244	2.2184	2.3141		

Tableau 30: Modèle d'estimation du montant d'ouverture non-forfaitaire des sinistres - Résultats de la régression

Estimation de la probabilité de paiement indemnitaire à l'ouverture

Paramètre		DDL	Valeur estimée	Erreur	Khi-2 de Wald	
TYPE_C_USED	COINS	1	0.4031	0.0135	889.225	<.0001
LN_DAY_BEF_CLOSED		1	-0.3508	0.00633	3068.44	<.0001
LN_CUMUL_FGU_INDEMN		1	0.1315	0.00391	1131.87	<.0001
FLAG_LAG_PAY_INDEMN	0	1	-0.6081	0.0173	1230.205	<.0001
FLAG_COSTS	0	1	-0.9203	0.0156	3495.128	<.0001
FLAG_FD_1	0	1	-0.3621	0.0161	506.6811	<.0001

Tableau 31: Modèle de survenance d'un paiement indemnitaire - Résultats de la régression

Estimation de la durée de vie des sinistres (régression du modèle de Cox)

Paramètre		DDL	Valeur estimée	Erreur type	Khi-2	Pr > Khi-2	Rapport de risque	Intervalle de confiance à 99% du rapport de risque	
MAIN_CAUSE_BEZ_40	Corporel	1	-0.16987	0.01086	244.7602	<.0001	0.844	0.821	0.868
MAIN_CAUSE_BEZ_40	Perte Financière	1	0.02054	0.00934	4.8364	0.0279	1.021	0.996	1.046
MAIN_CAUSE_BEZ_40	Matériel	1	0.19845	0.00771	663.0369	<.0001	1.22	1.196	1.244
KZ_VERSFALL	C	1	0.09509	0.00594	256.0765	<.0001	1.1	1.083	1.117

Tableau 32: Estimation de la durée de vie des sinistres dans les comptes de l'assureur - Résultats de la régression

Estimation du nombre de paiements pour frais restant avant la clôture

Paramètre		DDL	Valeur estimée	Erreur type	Valeur du test t	Approx. De Pr > t
Intercept		1	-3.97051	0.06552	-60.6	<.0001
LN_DAY_NOTIF		1	0.53826	0.00909	59.2	<.0001
TYPE_C_USED	LEAD	1	0.78964	0.02066	38.23	<.0001
LN_PAY_INDEMN		1	0.07303	0.00220	33.19	<.0001
Alpha		1	1.45429	0.02442	59.56	<.0001

Tableau 33: Estimation du nombre de paiements pour frais - Résultats de la régression

Estimation des montants de paiements pour frais après ouverture

Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		1	6.7798	0.0227	6.7353	6.8243	89297.1	<.0001
MAIN_CAUSE_BEZ_40	Corporel	1	-0.2998	0.0291	-0.3568	-0.2428	106.17	<.0001
MAIN_CAUSE_BEZ_40	Perte financière	1	0.2522	0.0222	0.2088	0.2957	129.36	<.0001
MAIN_CAUSE_BEZ_40	Autre / Non précisé	1	-0.1557	0.0164	-0.1878	-0.1235	90.09	<.0001
MAIN_CAUSE_BEZ_40	Matériel	0	0	0	0	0	.	.
LN_DAY_BEF_CLOSED		1	0.1013	0.0031	0.0953	0.1074	1069.79	<.0001
Scale		1	1.1176	0.0043	1.1092	1.126		

Tableau 34: Estimation des montants de paiements pour frais - Résultats de la régression

Estimation du nombre de mouvements indemnitaires restant jusqu'à la clôture

Paramètre		DDL	Valeur estimée	Erreur type	Valeur du test t	Approx. de Pr > t
Intercept		1	-5.58464	0.085143	-65.59	<.0001
LN_DAY_NOTIF	LEAD	1	0.770885	0.011747	65.62	<.0001
TYPE_C_USED		1	0.348636	0.021398	16.29	<.0001
Alpha		1	1.21309	0.027773	43.68	<.0001

Tableau 35: Estimation du nombre de mouvements indemnitaires - Résultats de la régression

Estimation de la répartition des mouvements indemnitaires sur la durée relative de présence du sinistre dans les comptes de l'assureur

Paramètre		Valeur estimée	Erreur type	Valeur du test t	Pr > t	Alpha	Inférieur	Supérieur
Xb Intercept		-0.090	0.009	-9.72	<.0001	0.05	-0.108	-0.072
Xb TYPE_C_USED	COINS	-0.241	0.016	-15.32	<.0001	0.05	-0.272	-0.210
Wd Intercept		-0.460	0.009	-50.25	<.0001	0.05	-0.477	-0.442
Wd TYPE_C_USED	COINS	-0.028	0.016	-1.78	0.0748	0.05	-0.059	0.003

Tableau 36: Répartition des mouvements indemnitaires au cours de la durée d'ouverture des sinistres - Résultats de la régression

Estimation de la classe de facteurs de développement

Paramètre		DDL	Valeur estimée	Erreur type	Khi-2 de Wald	Pr > Khi-2
Intercept	0	1	1.37310	0.01690	6580.04	<.0001
Intercept	0.5	1	3.26100	0.02370	18971.40	<.0001
KZ_VERSFALL	C	1	0.11950	0.01140	108.98	<.0001
LN_DAY BEF CLOSED		1	-0.45600	0.00381	14300.17	<.0001

Tableau 37: Estimation de la classe de facteurs de développement – cas sans paiement indemnitaire préalable - Résultats de la régression

Paramètre		DDL	Valeur estimée	Erreur type	Khi-2 de Wald	Pr > Khi-2
Intercept	0	1	-3.20110	0.05990	2860.47	<.0001
Intercept	0.5	1	1.20710	0.03080	1537.64	<.0001
LN_DAY BEF CLOSED		1	-0.15750	0.00622	640.55	<.0001

Tableau 38: Estimation de la classe de facteurs de développement – cas avec paiement indemnitaire préalable - Résultats de la régression

Estimation du facteur de développement (cas où la charge indemnitaire est supérieure à 10000 euros juste avant le mouvement considéré)

Paramètre		DDL	Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept		0	0	0	0	0	.	.
LN_DAY_BEFCLOSED		1	0.2107	0.0209	0.1698	0.2516	102.01	<.0001
TIME_FLAG_BIN	0.5	1	-1.7534	0.1337	-2.0155	-1.4914	171.99	<.0001
TIME_FLAG_BIN	1	1	-0.667	0.0934	-0.8501	-0.484	51	<.0001
Scale		1	1.5713	0.0188	1.5348	1.6086		

Tableau 39: Estimation du facteur de développement le la charge indemnitaire - Cas sans paiements indemnitaires préalables - Résultats de la régression

Paramètre		Valeur estimée	Erreur type	Intervalle de confiance de Wald à 95 %		Khi-2 de Wald	Pr > Khi-2
Intercept	1	-1.2836	0.0204	-1.3235	-1.2437	3976.09	<.0001
LAG_RATIO_CUMUL_PAY_	1	0.9497	0.0214	0.9077	0.9917	1964.6	<.0001
LN_DAY_BEFCLOSED	1	0.1039	0.0034	0.0973	0.1105	953.38	<.0001
Scale	1	0.8924	0.0075	0.8778	0.9073		

Tableau 40: Estimation du facteur de développement de la charge indemnitaire - Cas avec paiements indemnitaires préalables - Résultats de la régression

Paramètre		Valeur estimée	Erreur type	DDL	Valeur du test t	Pr > t	Alpha	Inférieur	Supérieur
Variance Small	1	1.7344	0.03319	6990	52.26	<.0001	0.05	1.6693	1.7995
Variance Large	1	102.37	8.2785	6990	12.37	<.0001	0.05	86.1455	118.6
P_LAG_R_CUMUL_PAY	1	4.2329	0.1276	6990	33.18	<.0001	0.05	3.9828	4.483

Tableau 41: Estimation du facteur de développement de la charge indemnitaire - Cas avec paiements indemnitaires préalables - Résultats de la régression de la loi Normale Contaminée

10.3 Triangles de liquidation des sinistres

Occurrence Year	Dvp. Year									
	1	2	3	4	5	6	7	8	9	10
1992	44.94	358.40	413.31	492.47	509.75	524.91	528.76	555.45	587.53	605.10
1993	137.39	273.59	370.76	519.98	581.41	611.11	693.58	772.47	795.11	815.64
1994	87.05	545.76	710.82	835.37	911.64	985.35	1,057.55	1,085.93	1,112.62	1,127.29
1995	169.54	447.26	887.11	973.99	1,149.27	1,288.34	1,399.43	1,459.71	1,641.65	1,699.62
1996	167.15	1,068.30	1,282.70	1,487.63	1,653.33	1,707.93	1,791.25	1,892.12	2,102.44	2,175.46
1997	110.72	424.09	846.97	1,068.29	1,209.17	1,602.46	1,779.29	1,923.63	2,030.51	2,063.94
1998	244.45	640.18	993.29	1,179.15	1,319.97	1,482.09	1,642.86	1,789.93	1,816.82	1,842.73
1999	172.43	722.08	1,043.92	1,354.20	1,448.67	1,700.85	1,889.07	2,384.33	2,491.62	2,708.31
2000	87.42	692.40	1,062.63	6,991.42	7,218.58	7,422.51	7,604.81	7,692.87	7,824.70	7,800.32
2001	170.15	1,073.54	1,639.37	2,096.52	2,686.83	2,907.72	3,299.74	3,417.65	3,493.56	3,881.61
2002	111.45	657.45	1,126.82	1,785.80	2,007.05	2,169.41	2,308.73	2,392.49	2,638.22	
2003	285.87	874.98	1,296.78	1,712.05	1,999.77	2,227.90	2,748.18	2,875.96		
2004	171.00	680.05	1,210.69	1,822.58	2,477.13	3,003.18	3,212.83			
2005	641.37	1,180.36	1,711.08	2,095.94	2,277.25	2,424.13				
2006	282.07	1,495.49	3,901.12	4,200.27	4,477.25					
2007	104.76	605.12	1,052.20	1,288.63						
2008	108.76	578.90	971.17							
2009	120.23	799.08								
2010	208.84									

Tableau 42: Triangle de liquidation des sinistres – Paiements cumulés – partie 1

Occurrence Year	Dvp. Year								
	11	12	13	14	15	16	17	18	19
1992	621.56	625.61	625.44	626.50	635.72	635.86	636.60	637.25	639.51
1993	821.13	821.92	854.02	881.52	883.68	873.59	865.66	866.77	
1994	1,156.59	1,151.74	1,158.31	1,158.61	1,162.31	1,168.14	1,168.12		
1995	1,709.45	1,710.85	1,715.26	1,718.25	1,723.27	1,722.73			
1996	2,197.88	2,228.56	2,234.93	2,344.40	2,242.26				
1997	2,120.83	2,164.48	2,237.79	2,252.19					
1998	1,873.07	1,870.14	1,862.90						
1999	2,744.82	3,519.66							
2000	7,913.64								
2001									
2002									
2003									
2004									
2005									
2006									
2007									
2008									
2009									
2010									

Tableau 43: Triangle de liquidation des sinistres – Paiements cumulés – partie 2

Occurrence Year	Dvp. Year									
	1	2	3	4	5	6	7	8	9	10
1992	872.79	693.39	693.39	721.10	720.22	685.78	705.08	653.95	654.56	661.70
1993	624.13	890.85	792.17	941.63	964.11	1,065.66	958.57	950.74	963.42	888.68
1994	1,041.69	1,471.84	1,300.12	1,241.06	1,317.29	1,334.79	1,332.73	1,322.30	1,210.77	1,224.25
1995	1,361.53	1,604.41	1,761.83	1,836.88	2,068.42	2,083.54	2,199.58	1,759.44	1,793.24	1,828.92
1996	2,442.03	2,673.59	2,688.75	2,395.64	2,443.72	2,472.81	2,312.53	2,371.35	2,341.43	2,345.27
1997	2,300.82	3,152.45	2,393.88	2,304.88	2,768.11	2,746.28	2,735.13	2,682.75	2,668.11	2,666.48
1998	2,042.11	2,785.27	2,473.80	2,463.27	2,260.13	2,160.37	2,140.63	2,082.60	2,048.92	2,098.58
1999	2,570.05	3,219.72	3,233.06	3,769.70	3,318.03	3,355.01	3,870.37	3,848.95	3,863.75	3,881.08
2000	3,573.51	7,616.10	7,556.62	9,024.55	9,018.00	8,997.75	9,153.09	9,158.58	9,226.31	9,144.92
2001	2,597.83	3,991.81	4,567.02	5,264.63	5,197.18	5,177.89	6,490.75	6,649.97	6,690.82	5,633.57
2002	2,135.09	9,364.72	9,610.10	10,140.51	9,900.29	9,918.35	10,205.95	10,105.76	10,006.28	
2003	3,863.51	4,687.46	5,048.31	5,004.63	4,931.15	4,949.87	5,072.82	5,050.47		
2004	6,154.15	7,172.75	7,554.54	7,987.33	8,119.89	6,623.18	6,679.40			
2005	3,957.93	5,573.95	5,356.31	5,131.57	4,902.67	5,099.86				
2006	6,364.13	8,317.19	8,977.08	8,980.31	9,196.33					
2007	2,675.50	4,071.70	4,676.17	4,935.32						
2008	2,594.74	3,521.73	3,439.62							
2009	3,617.53	5,341.24								
2010	3,763.05									

Tableau 44: Triangle de liquidation des sinistres – Charges totales – partie 1

Occurrence Year	Dvp. Year								
	11	12	13	14	15	16	17	18	19
1992	696.20	710.28	695.57	688.57	644.95	647.32	654.12	655.75	656.35
1993	894.67	936.96	952.76	906.14	908.63	894.59	884.50	874.23	
1994	1,231.67	1,179.93	1,190.48	1,187.00	1,189.60	1,200.55	1,201.11		
1995	1,800.74	1,755.93	1,757.05	1,754.51	1,754.76	1,770.13			
1996	2,305.90	2,343.41	2,328.33	2,416.88	2,421.83				
1997	2,713.37	2,725.47	2,722.26	2,535.31					
1998	2,147.02	2,129.27	2,057.59						
1999	4,410.60	4,369.41							
2000	9,252.04								
2001									
2002									
2003									
2004									
2005									
2006									
2007									
2008									
2009									
2010									

Tableau 45: Triangle de liquidation des sinistres – Charges totales – partie 2

10.4 La société HDI Global SE au 31/12/2010

HDI Global SE en France est une entité opérationnelle de la société d'assurance HDI Global SE basée à Hanovre, elle-même filiale du groupe Talanx, troisième groupe d'assurance en Allemagne. Le groupe Talanx comporte comme branche d'activité principale l'assurance vie, l'assurance non-vie, la réassurance vie et non-vie avec comme compagnie Hannover Re, troisième groupe de réassurance au monde, et a également une branche importante d'activité en bancassurance et une autre en activité financière. Les activités de holding de Talanx ont démarré en 1996 sur la base des sociétés et filiales HDI.

L'opération française est née en 2006 de la fusion de la succursale HDI, mutuelle industrielle allemande créée en 1903 et qui opérait en France depuis 2000, et de l'opération française du groupe Gerling, fondé lui en 1904 également en Allemagne. L'activité d'HDI Global SE a comme cible l'assurance des grands risques industriels et des « grandes » PME dite « middle market » en assurance non-vie.

Les deux centres de profits antécédents en France, HDI et Gerling, avaient des systèmes d'information différents. Depuis le 01/01/2008, toutes les informations sont gérées dans le même système, GERAS issu de Gerling, l'historique des informations de HDI ayant été entièrement migré sans perte d'informations majeures.

10.5 Informations sur les sinistres

Les sinistres utilisés sont extraits de la base de données de l'outil de gestion technique de la succursale. Ils concernent les années de souscription allant de 1986 jusqu'à 2010 et approchent le nombre de 60 000 dossiers ouverts sur cette période pour les deux branches observées. La notion d'année de souscription permet d'avoir une approche générationnelle de l'activité d'assurance, et donc de mettre en relation direct les émissions de primes avec les prestations qu'elles ont générées.

La base de données sinistre contient l'ensemble des mouvements de chaque sinistre accompagné d'informations concernant la définition du mouvement, la définition du sinistre, et un lien avec l'avenant et la police concernés.

Voici les principales variables disponibles dans cette base de données qui nous serviront dans le cadre de cette étude :

- Numéro de branche
- Part de la compagnie
- Sinistre Actif/Clos
- Année de souscription
- Date de survenance du sinistre
- Date de déclaration du sinistre
- Date de clôture du sinistre
- Pays du sinistre
- Type de dommage
- Cause du dommage
- Numéro d'identifiant du sinistre
- Numéro d'avenant auquel est attaché le sinistre
- Mois comptable du mouvement

- Type de mouvement (Réserve / Paiement / Recours)
- Impact du mouvement sur les paiements (valeurs prenant $\{0, 1, -1\}$)
- Impact du mouvement sur la réserve (valeurs prenant $\{0, 1, -1\}$)
- Impact du mouvement sur le recours (valeurs prenant $\{0, 1, -1\}$)
- Monnaie
- Date du mouvement
- Montant du mouvement à la part de la succursale
- Type de bénéficiaire
- Montant en réassurance du mouvement à la part de la compagnie
- Identifiant unique du mouvement

Les données de la base sont mises à jour quotidiennement et permettent de suivre en quasi-temps réel les principaux indicateurs de l'activité.

Les IBNR sont également présents dans la base et sont enregistrés par mouvement comme tout autres sinistres avec la particularité de faire référence à des polices fictives de régulation de bilan. Ainsi il est possible de suivre l'historique des IBNR sur l'ensemble de la période.

Le détail des mouvements permet de suivre la dynamique du déroulement des sinistres. Chaque mouvement n'impact pas uniquement soit la réserve, soit le paiement ou encore les recours : il fournit un impact combiné à ses trois facteurs à la fois. C'est la variable d'impact du mouvement par catégorie qui indique l'influence du mouvement en le multipliant soit par 0 (aucun impact), soit par 1 (impact additif positif), soit par -1 (impact additif négatif).

Donc pour connaître le montant des réserves R_i du sinistre i , par exemple, voici l'opération à réaliser :

$R_i = \sum_j fR_{i,j} \cdot M_{i,j}$, où $fR_{i,j}$ est l'impact sur la réserve du mouvement j du sinistre i , et $M_{i,j}$ est le montant du mouvement j du sinistre i .

L'analogie pour les paiements et les recours d'un sinistre i est immédiat. La charge totale T_i d'un sinistre i est obtenu par $T_i = \sum_j (fR_{i,j} + fP_{i,j}) \cdot M_{i,j}$, où $fP_{i,j}$ est l'impact sur les paiements du mouvement j du sinistre i .

A noter que les paiements sont calculés nets de recours.

10.6 Informations sur les primes

Dans la logique générationnelle présentée ci-dessus, les primes émises sur les années de souscription allant de 1986 à 2010 sont prises en comptes

La base de données des primes émises contient l'ensemble des mouvements de primes accompagné d'informations concernant la définition du mouvement, la définition de l'objet de la prime, et un lien avec l'avenant et la police concernés.

Voici les principales variables disponibles dans cette base de données qui nous serviront dans le cadre de cette étude :

- Numéro de branche
- Part de la compagnie
- Mois et année comptable d'enregistrement
- Année de souscription
- Cadence de la prime (annuel, semestriel, forfaitaire...)
- Date de la quittance
- Numéro de la quittance
- Date de début du comptant
- Date de fin du comptant
- Date d'échéance
- Part de la compagnie
- Date d'encaissement total du mouvement de prime
- Montant de la prime émise à la part de la compagnie
- Code monnaie
- Type de prime (Normal, CatNat, GAREAT...)
- Identifiant unique du mouvement de prime
- Numéro d'avenant auquel est attaché le sinistre

Sans entrer dans le détail, chaque mouvement de prime est relié à une base contenant les mouvements annexes des primes que sont les frais, les taxes, et les commissions.

10.7 Informations principales sur les contrats

Le point central du système d'information est la base des avenants des polices comprenant l'ensemble des informations définissant la police.

Voici les principales variables disponibles dans cette base de données qui nous serviront dans le cadre de cette étude :

- Numéro de branche
- Identifiant unique de l'avenant
- Numéro de la police
- Date de l'avenant
- Échéance principale
- Prochaine échéance
- Date de fin d'avenant
- Code type d'activité principale
- Code compagnie d'origine