Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuaires

Par : Laëtitia Laloum

Titre du mémoire : Approche hybride au calcul de la provision IBNR : méthode de fréquence / coût pour l'estimation des sinistres tardifs

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

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

Membres présents du jury de l'Institut des Actuaires signature

Membres présents du jury de la filière

Entreprise :

Nom : OPTIMIND

Signature :

Directeur de mémoire en entreprise :

Nom : CANON THOMAS

Signature :

Invité :

Nom : PESNEAUD ANTOINE

Signature :

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

Signature du responsable entreprise

Signature du candidat

ISUP

Mémoire d'actuariat

Approche hybride au calcul de la provision IBNR : méthode de fréquence / coût pour l'estimation des sinistres tardifs

Auteure :
Laetitia LALOUM

11 avril 2021

Résumé

Mots clés : *provisionnement, provisions techniques, méthode fréquence coût, charges, modèle de fréquence, modèle de nombres, loi de fréquence,loi des nombres, délais de déclaration, Incurred But Not Reported (IBNR), Incurred But Not yet Reported (IBNyR), Incurred But Not enough Reported (IBNeR), Provision pour Sinistres A Payer, Chain Ladder, ajustement de lois statistiques, sinistres tardifs, maximum de vraisemblance, loi Négative Binomiale*

L'inversion du cycle de production engendre un réel besoin pour les assureurs et les réassureurs de constituer des provisions techniques. Les méthodes traditionnelles telles que Chain Ladder sont des approches qui agrègent les données en triangles, ne permettant pas du tout d'avoir une vision individuelle des sinistres dans le portefeuille. Malgré leur facilité d'implémentation, le fait d'agréger les sinistres écarte des informations précieuses qui pourraient s'avérer très enrichissantes dans les prédictions des sinistres tardifs, comme les délais de déclaration. Ce mémoire étudie une alternative aux méthodes traditionnelles d'estimation de la réserve *IBNR* et propose une piste d'amélioration en distinguant la modélisation des deux composantes de la réserve *IBNR* : les *IBNyR* et les *IBNeR*. La première est liée aux sinistres tardifs (survenus et pas encore déclarés) alors que la seconde est liée aux sinistres qui sont déclarés mais dont la charge placée est potentiellement insuffisante. D'abord les sinistres tardifs sont calculés par une approche fréquence-coût grâce à l'ajustement de lois statistiques aux délais de déclaration connus et aux charges placées sur la période d'observation. Ensuite, les sinistres connus et potentiellement sous-provisionnés, les *IBNeR*, sont estimés avec un Chain Ladder effectué sur les sinistres en cours sur la période d'évaluation.

Abstract

Key words : *reserving, technical provisions, frequency/cost method, incurred claims, frequency model, frequency law, reporting delays, Incurred But Not Reported (IBNR), Incurred But Not yet Reported (IBNyR), Incurred But Not enough Reported (IBNeR), Chain Ladder, statistic law adjustment, late claims, maximum likelihood, Negative Binomial law*

The reversal of the production cycle creates a real need for insurers and reinsurers to set up technical provisions. Traditional methods such as Chain Ladder are approaches that aggregate data into triangles, not allowing for an individual view of claims in the portfolio at all. Despite the ease of implementation and justification of these methods, aggregating claims rules out valuable information such as reporting delays that could turn out to be very useful in predicting late claims. This paper explores an alternative to the traditional methods of estimating the *Incurred But Not Reported* reserve and proposes an improvement by distinguishing between the two components of this reserve : the *Incurred But Not yet Reported* and the *Incurred But Not enough Reported*. The first is related to late claims (occurred and not yet reported) while the second is related to claims that are reported but whose claim charge is potentially insufficient. Thus, we model *IBNyR* and *IBNeR* claims separately. First, late claims are modeled using an approach that estimates their frequency and associated charge separately by adjusting statistic laws to the reporting delays for the frequency portion and to the charge placed over the evaluation period for the expense portion. Known and potentially underprovisioned claims, the IBNeR, are estimated with a Chain Ladder performed on claims that are in progress over the evaluation period and still in progress at the evaluation date.

Note de synthèse

L'inversion du cycle de production caractérise le secteur de l'assurance. Les produits sont vendus avant d'en connaître leur coût définitif, ce qui engendre un réel besoin pour les assureurs de constituer des provisions techniques. Ainsi, l'estimation des provisions requiert une précision importante et les méthodes de calcul utilisées doivent être adaptées aux caractéristiques des sinistres.
Les méthodes de provisionnement traditionnelles telles que Chain Ladder sont des approches agrégées : elles compressent les données en triangles, ne permettant pas du tout d'avoir une vision individuelle des sinistres dans le portefeuille. Malgré leur facilité d'implémentation et de communication des résultats, le fait d'agréger les sinistres écarte des informations précieuses qui pourraient améliorer la précision des estimations, notamment celles effectuées au titre des sinistres tardifs.

La réserve IBNR [1] est composée de deux éléments. La partie IBNyR [2] représente la provision qui doit être placée au titre des sinistres qui sont déjà survenus mais qui ne sont pas encore déclarés à l'assureur alors que la seconde partie IBNeR [3] couvre l'insuffisance potentielle des provisions déjà établies pour les sinistres survenus et déjà déclarés. Ce mémoire étudie une approche alternative aux méthodes traditionnelles utilisées dans le calcul de la réserve IBNR car elle introduit l'utilisation d'une modélisation "fréquence / sévérité" pour les sinistres tardifs. L'approche proposée dans les travaux est inspirée des travaux de *Schnieper (1991)* et de *Pietro Parodi (2013)* où l'on retrouve cette séparation des sinistres IBNyR et IBNeR. En effet, par la nature et la quantité d'information disponible pour chacun de ces éléments, il semble intéressant de proposer des méthodes différentes et adaptées.

L'objectif initial de ce mémoire était de faire une modélisation séparée des IBNyR (sinistres tardifs survenus et pas encore déclarés) et des IBNeR (sinistres déjà connus de l'assureur en cours d'indemnisation). Cependant nous avons rencontré deux obstacles qui n'ont pas permis aux travaux d'aboutir de manière satisfaisante :

- Tout d'abord, les sinistres à disposition sont issus de la garantie Responsabilité Civile Automobile sur une période entre 2009 et 2018, garantie pour laquelle nous nous sommes aperçus assez tardivement que les sinistres tardifs sont très rares. En effet les sinistres corporels et / ou matériels sont déclarés très rapidement dans la pratique, ce qui laisse peu de place au besoin de modéliser les IBNyR.
- Par ailleurs, nous ne disposions pas des données relatives à l'exposition au risque sur le portefeuille à disposition et de surcroît il n'est pas issu du marché français, ce qui a empêché l'analyse complète des données et des résultats. Du fait du manque d'expertise et de connaissances sur l'environnement des données ainsi que sur leur contexte économique et juridique, nous avons été

1. Incurred But Not Reported
2. Incurred But Not Yet Reported, en français survenus mais non déclarés
3. Incurred But Not Enough Reported, en français survenus mais "pas assez" provisionnés

amenés à faire des simplifications, hypothèses et retraitements tout au long des travaux qui ont fait que la mise en pratique de la méthode fut compliquée.

De manière générale en automobile, les sinistres déclarés au-delà de deux ans après leur survenance ne peuvent pas être des tardifs au sens traditionnel, il s'agit bien souvent de réouvertures. Nous avons pris la décision de modéliser en fréquence/sévérité les sinistres tardifs (déclarés dans les deux ans après la survenance). La modélisation des réouvertures peut être effectuée de la même manière en calibrant une loi adaptée, mais cela n'a pas été traité dans le sujet faute de temps. Les fondements théoriques de la méthode présentée restent valables malgré tout et pourront s'appliquer dans le cas de garanties où l'on retrouve bien des sinistres tardifs par exemple en Responsabilité Médicale où le délai entre la survenance et la réclamation peut être relativement long, mais également pour les réouvertures.

Dans une première partie de la modélisation, l'objectif est de prédire le nombre de sinistres tardifs. Pour ce faire, nous avons analysé les délais de déclaration pour estimer au plus juste le nombre de sinistres qui seront déclarés dans le futur. Dans un second temps, nous avons étudié la distribution des charges des sinistres du portefeuille afin d'être en mesure d'allouer une charge aux nombres de tardifs estimés.

Une fois la modélisation des sinistres tardifs effectuée, nous nous penchons sur le sujet des IBNeR, sinistres déjà connus mais dont la variation future doit être estimée. Cette dernière est calculée par une méthode Chain Ladder avec un triangle de charges de sinistres qui sont encore en cours à la date d'évaluation et qui ont été déclarés dès leur survenance afin de ne pas laisser place à de nouveaux sinistres dans le portefeuille. La somme des IBNyR et des IBNeR permettra ensuite d'établir une estimation de la réserve *IBNR*.

Calcul de la réserve IBNR au 31/12/2018 avec la méthode Chain Ladder

Tout d'abord nous présentons les résultats obtenus avec Chain Ladder pour la période 2009-2018, une des méthodes les plus utilisées en pratique car cela nous permettra de comparer la performance de notre modèle face à une méthode traditionnelle. Dans cette partie, les charges des sinistres connues ont été agrégées en triangle. L'année de rattachement (en ligne) est l'année de survenance et les colonnes représentent l'année de développement. Après validation des hypothèses de Chain Ladder, la réserve IBNR est estimée à 56 207 974€. Elle est obtenue en déduisant les montants connus au 31/12/2018 des ultimes projetés et offre une estimation de la provision à placer pour tous les sinistres survenus sur la période d'évaluation.

Survenance	0	1	2	3	4	5	6	7	8	9	Ultime	Réserve
2009	14 485 401	27 810 864	34 679 862	38 534 466	41 810 003	43 990 855	45 087 721	44 859 167	45 332 597	45 195 672	45 195 672	-
2010	13 135 693	26 737 946	31 364 706	35 285 311	37 845 715	39 372 862	40 486 855	40 016 346	40 109 641	39 988 492	39 988 492	- 121 149
2011	14 146 964	27 141 883	31 681 244	35 114 914	37 157 607	38 619 399	38 879 888	39 334 870	39 597 513	39 477 911	39 477 911	143 042
2012	16 011 252	28 414 095	33 816 625	37 657 678	40 311 220	41 871 818	42 881 945	42 797 844	43 083 611	42 953 479	42 953 479	71 534
2013	12 888 928	24 768 498	30 165 924	34 341 997	36 759 707	38 144 421	38 954 887	38 878 488	39 138 085	39 019 870	39 019 870	875 449
2014	12 041 707	22 220 071	28 385 083	32 093 163	34 221 616	35 653 976	36 411 526	36 340 115	36 582 763	36 472 267	36 472 267	2 250 651
2015	12 110 485	24 864 658	30 985 738	34 967 113	37 442 127	39 009 282	39 838 124	39 759 993	40 025 475	39 904 581	39 904 581	4 937 467
2016	11 351 040	24 971 808	31 283 840	35 092 516	37 576 406	39 149 181	39 980 995	39 902 584	40 169 019	40 047 690	40 047 690	8 763 850
2017	12 101 769	27 326 301	33 326 023	37 383 326	40 029 362	41 704 807	42 590 921	42 507 392	42 791 219	42 661 970	42 661 970	15 335 669
2018	11 448 098	22 674 504	27 652 885	31 019 507	33 215 104	34 605 336	35 340 606	35 271 295	35 506 806	35 399 560	35 399 560	23 951 461
	1,98	1,22	1,12	1,07	1,04	1,02	1,00	1,01	0,997		401 121 492	56 207 974

FIGURE 1 – Résultat du calcul de la réserve IBNR sur le triangle de charges des sinistres survenus entre 2009 et 2018

Calcul de la composante IBNyR

L'approche utilisée pour l'estimation de la partie *IBNyR* sépare la modélisation des nombres des sinistres tardifs et la modélisation de la charge qui leur est associée. La loi de la fréquence des sinistres tardifs est calculée suite à l'ajustement d'une loi statistique aux délais de déclaration. La charge est modélisée suite à l'analyse des charges placées pour les sinistres connus.
Dans les deux cas, nous cherchons à ajuster la meilleure loi statistique théorique aux données observées dans la base afin de trouver la distribution qui sera la plus adaptée à chaque variable.
En combinant la loi des nombres de sinistres tardifs et la loi ajustée pour les charges et en simulant de nombreux scénarios, nous serons capables d'en estimer son montant grâce à la méthode de Monte Carlo.

Ajustement d'une loi statistique aux délais de déclaration

L'objectif ici est d'ajuster la meilleure loi statistique aux délais de déclaration observés pour en déduire le nombre de sinistres survenus sur la période 2009-2018 qui ne sont pas encore déclarés. Nous analysons l'ajustement de plusieurs lois afin de choisir celle qui paraît la mieux adaptée à nos données.

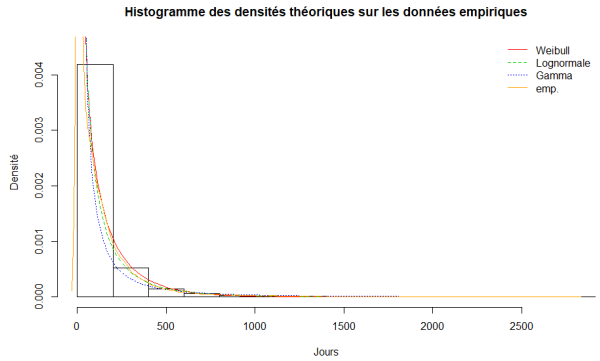

FIGURE 2 – Comparaison des lois candidates aux délais de déclaration empiriques

Grâce aux différentes sorties graphiques, statistiques descriptives et aux tests d'adéquation de lois (Kolmogorov Smirnoff, Anderson Darling, Cramer Von Mises), la loi Weibull est jugée la plus pertinente pour représenter les délais de déclaration parmi les lois candidates.

Ajustement d'une loi pour les nombres de sinistres tardifs

Après avoir ajusté une loi aux délais de déclaration, nous l'exploitons dans le but d'obtenir le nombre ultime de sinistres tardifs pour chaque année de survenance. La formule d'obtention des sinistres tardifs ultimes, démontrée dans le mémoire s'écrit :
$\hat{N}_t - n_t = \frac{t}{\sum_{u=0}^{t} F(t-u)} * n_t - n_t$ où F est la fonction de répartition de la loi choisie pour représenter les délais de déclaration et n_t est le nombre de sinistres déclarés à la date d'évaluation.
Le but ultime de cette partie est d'obtenir une loi pour le nombre de *IBNyR*. Celle-ci est représentée par une loi Binomiale Négative de paramètres (r, p).
Nous avons pu comparer ce nombre de tardifs projetés par l'approche de P.Parodi avec le nombre de sinistres tardifs projetés par la méthode de Schnieper. Avec Parodi, les sinistres tardifs modélisés apparaissent à partir de l'année 2015 ce qui est cohérent avec les délais de déclaration que nous observons dans la base. Pour les années de survenance 2009 - 2014, aucun sinistre tardif n'est modélisé, ce qui est

Année de survenance	Ultime	Nombre projetés de sinistres IBNyR (λ)	r	p
2009	5750	0	154,94	0,000
2010	4662	0	154,94	0,000
2011	4034	0	154,94	0,000
2012	4513	0	154,94	0,000
2013	4190	0	154,94	0,000
2014	3881	0	154,94	0,000
2015	4147	5	154,94	0,032
2016	4147	35	154,94	0,186
2017	4025	146	154,94	0,486
2018	2470	609	154,94	0,797
Agregé	**41 820**	**796**	**154,94**	**0,836**
Moyenne	**4 182**			
Variance	**652 139**			
Ratio Variance Moyenne	**155,94**			

FIGURE 3 – Obtention des paramètres de la distribution de la loi des nombres

différent de l'approche de Schnieper qui recense des sinistres tardifs (réouvertures) jusqu'à 7 ans après la survenance. Le modèle développé à partir de l'approche de Pietro Parodi a du sens étant donné que la quasi-majorité des sinistres sont déclarés dans les deux ans mais ne permet pas d'aller chercher les observations plus rares. Les évènements où les sinistres sont déclarés très tardivement sont plutôt rares et il est possible qu'il ne soit pas capté par la queue de la loi Weibull. Cela nous a permis de voir qu'une piste d'amélioration serait d'utiliser la théorie des valeurs extrêmes afin d'attribuer une distribution spécifique aux délais très élevés.
Bien qu'elle soit très sensible au choix de l'exposition pour chaque année de survenance, la méthode de Schnieper permet de différencier les nouveaux sinistres et les sinistres sans-suite ce qui présente un réel avantage. Dans les calculs, les nouveaux sinistres sont mesurés par rapport à l'exposition alors que les sans suite sont mesurés par rapport aux nombres de sinistres précédemment observés ou estimés (incluant donc les nouveaux sinistres estimés). La méthode de Schnieper est ainsi très dépendante des informations sur l'exposition. On voit ci-dessous la sensibilité de l'approche de Schnieper vis-à-vis de l'exposition. Les résultats obtenus par la méthode de P.Parodi est assez proche des résultats de Schnieper quand on considère une baisse de l'exposition similaire à celle observée sur les déclarations de nouveaux sinistres en première année d'ouverture.

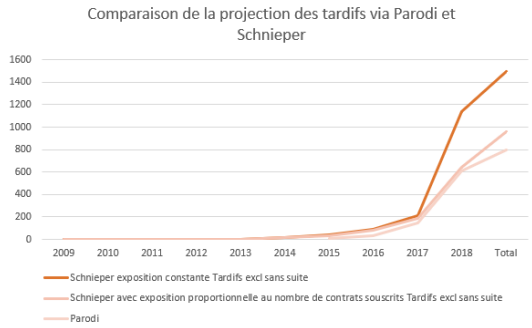

FIGURE 4 – Comparaison des sinistres tardifs obtenus avec Schnieper et avec Parodi

Dans notre application de Parodi, nous n'avons pas tenu compte des sinistres qui seront potentiellement déclarés sans suite. Cela présente un axe d'amélioration à apporter au calcul car si nous sommes dans

une optique de projection des sinistres tardifs, certains d'entre eux seront classés sans suite et donc ne donneront pas lieu à un sinistre pour l'assureur. Ainsi, il aurait été préférable de préparer les données en tenant compte de la proportion des sinistres qui pourraient être classés sans suite afin d'apporter à l'approche de Parodi cette même vision que l'on peut retrouver dans la méthode de Schnieper. Les tardifs seraient ainsi projetés en prenant la base de données complète et les sans suite en seraient déduits par la suite.

Ajustement d'une loi statistique aux charges

L'objectif ici est d'associer un montant de charge aux sinistres tardifs qui ont été estimés dans la partie précédente.
Pour les charges, le procédé d'ajustement d'une loi statistique bien qu'il soit similaire à celui des délais de déclaration requiert un travail préalable plus important. Les trois points d'attention sont les suivants :

- Pour les sinistres qui sont clos depuis un certain temps, il est important d'étudier la problématique de l'inflation.
- Les sinistres dont le statut est "en cours" dans le portefeuille ont la particularité de ne pas être entièrement réglés et leurs montants de charges sont sujets à des variations et sont également soumis à l'inflation.
- Nous devons faire un choix important : la distribution de la charge est-elle la même pour toutes les années de survenance ou faut-il prendre en compte certains phénomènes propres à chaque année de survenance ?

Après avoir traité les 3 problématiques, nous appliquons un procédé similaire à celui de l'ajustement d'une loi aux délais de déclaration.

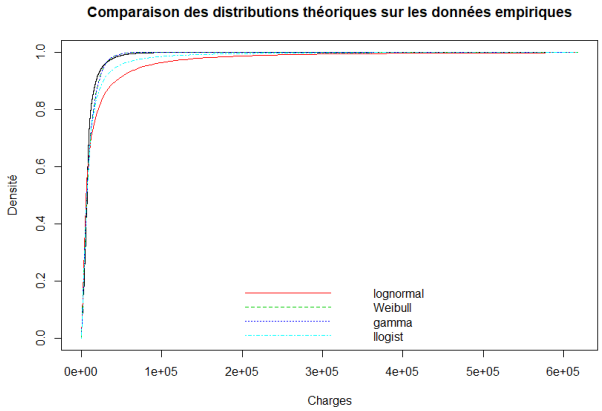

FIGURE 5 – Comparaison de la distribution des données empiriques et de la distribution des lois testées

Notre analyse montre que la loi Weibull avec les paramètres $\lambda = 0.9862$ et $\kappa = 9208.992$ est celle qui coïncide le mieux possible avec les données empiriques parmi celles que nous avons testées. Cet ajustement mériterait d'être amélioré avec une meilleure prise en compte des phénomènes d'inflation et d'analyse des sinistres sans suite. Cependant, les données du portefeuille et du marché étant très limitée lors de nos travaux, nous n'avons pu optimiser cette modélisation.

Mise en commun de la loi des nombres et de la loi du montant de charges

Par l'application de la méthode de Monte Carlo et grâce à ses propriétés, nous sommes capables d'estimer le montant de la réserve *IBNR*.
En effet, en simulant le nombre de sinistres tardifs tiré de la loi Binomiale Négative un certain N_{sim} nombre de fois et en allouant à chaque fois un montant de charges tiré d'une loi Weibull au nombre généré, nous sommes capables d'obtenir la distribution de la réserve *IBNyR*. Plus précisément, l'algorithme combine la loi des nombres de sinistres tardifs et la loi ajustée aux charges et se déroule comme suit :

- Pour chaque scénario $j = 1$ jusqu'à N_{sim}, on simule n_j sinistres tardifs tirés d'une loi Négative Binomiale de paramètres $p = 0.816$ et $r = 154.94$.
  Pour chaque $j \in [1, N_{sim}]$, $n_j \sim F_X(x)$ où $F_X(x)$ est la fonction de répartition de la loi $\mathrm{NB}(r, p)$
- Pour chaque $i \in [1, n_j]$, on simule une charge sinistre x_i^j tiré d'une loi de Weibull de paramètres $\lambda = 0.9862$ et $\kappa = 9208.992$.
- Calculer le montant agrégé de la charge sinistre pour chaque scénario j, tel que $S_j = \sum_{i=1}^{n_j} x_i^{(j)}$ $\forall j = 1, .., N_{sim}$ que nous rangeons dans un vecteur de taille N_{sim}.

Le résultat final est un vecteur de taille N_{sim} où nous avons toutes les valeurs possibles de la charge ultime des sinistres tardifs simulés.
Par la méthode de Monte Carlo et grâce à la loi des grands nombres, le montant ultime de la charge au titre des sinistres tardifs est la moyenne empirique de toutes les valeurs possibles calculées :

$$\boxed{\delta = \frac{1}{N_{sim}} * \sum_{j=1}^{N_{sim}} S_j}$$

Calcul de la composante IBNeR

Le caractère connu des sinistres pris en compte dans la réserve *IBNeR* nous amène à conserver la méthode Chain Ladder pour son estimation sans faire de retraitement vis-à-vis de l'inflation. Le triangle est construit par l'agrégation des charges placées au titre des sinistres en cours à la date d'évaluation. L'année de rattachement est l'année de déclaration, ce qui permet de ne pas introduire de nouveaux sinistres d'une année de développement à l'autre.
La réserve devant être placée au titre des sinistres en cours est de 17 579 583€. Nous observons un rattrapage sur les années plus anciennes (2010-2014) dû à un excédent de provision dossier/dossier.

Survenance	0	1	2	3	4	5	6	7	8	9	Ultime	Réserve
2009	4 315 074	5 662 927	6 952 301	7 595 574	8 253 566	8 576 683	8 634 905	8 284 186	8 325 257	8 108 664	8 108 664	-
2010	6 668 937	8 763 137	9 660 717	10 438 966	10 872 858	11 068 845	11 283 530	10 752 821	10 575 784	10 300 640	10 300 640	- 275 144
2011	6 750 257	9 277 136	10 277 364	10 738 330	11 185 603	11 537 550	11 195 483	11 185 076	11 105 190	10 816 273	10 816 273	- 368 803
2012	13 255 614	17 086 754	18 991 629	20 062 158	20 727 623	21 028 195	21 333 046	20 721 565	20 573 568	20 038 318	20 038 318	- 1 294 728
2013	11 591 989	15 788 351	17 717 754	18 991 679	19 867 506	20 452 436	20 544 762	19 955 876	19 813 348	19 297 876	19 297 876	- 1 154 561
2014	11 280 646	14 622 280	16 626 697	18 150 390	18 646 168	19 108 082	19 194 339	18 644 161	18 511 001	18 029 411	18 029 411	- 616 757
2015	11 071 950	15 742 735	18 372 708	19 708 984	20 528 781	21 037 332	21 132 298	20 526 571	20 379 967	19 849 753	19 849 753	140 769
2016	13 010 248	19 127 614	22 393 143	24 002 673	25 001 067	25 620 408	25 736 062	24 998 375	24 819 832	24 174 109	24 174 109	1 780 966
2017	17 224 926	25 532 778	29 124 566	31 217 923	32 516 436	33 321 952	33 472 373	32 512 936	32 280 722	31 440 894	31 440 894	5 908 116
2018	19 151 264	26 482 999	30 208 457	32 379 720	33 726 558	34 562 052	34 718 071	33 722 927	33 482 072	32 610 989	32 610 989	13 459 725
	1,38	1,14	1,07	1,04	1,02	1,00	0,97	0,99	0,97		194 666 926	17 579 583

FIGURE 6 – Projection de la charge des sinistres en cours à l'ultime

Validation des résultats : comparaison avec Chain Ladder et l'observé

Une limite importante dans notre validation réside dans le fait que la méthode mise en place dans ces travaux ne comporte pas de cadence de liquidation pour les tardifs et les réouvertures. En effet, nous ne

savons pas répartir les charges par diagonale pour la partie des tardifs. Ainsi, les chiffres de notre méthode ci-dessous sur-estiment les diagonales en partie à cause des projections des tardifs qui sont vues à l'ultime ; cela n'est pas le cas sur la partie IBNeR qui, elle, est en Chain Ladder.
De manière générale, la méthode Chain Ladder donne des estimations inférieures aux montants réellement observés alors que la méthode prenant en compte les tardifs via une approche fréquence / sévérité donne au contraire des montants supérieurs aux données observées (cohérent avec la remarque préliminaire ci-dessus).

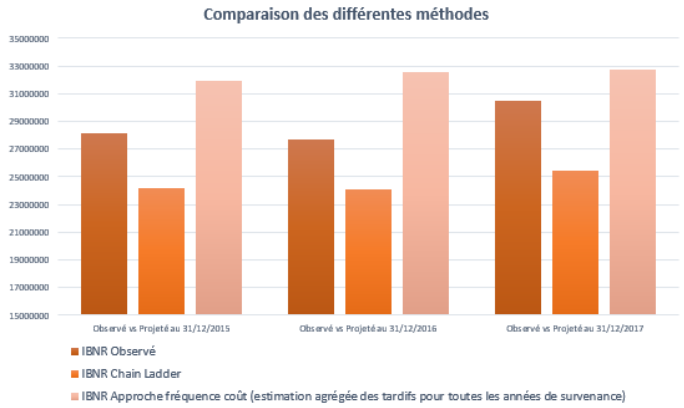

FIGURE 7 – Comparaison des projections avec la charge placée observée aux dates d'arrêtés 31/12/2015, 2016 et 2017

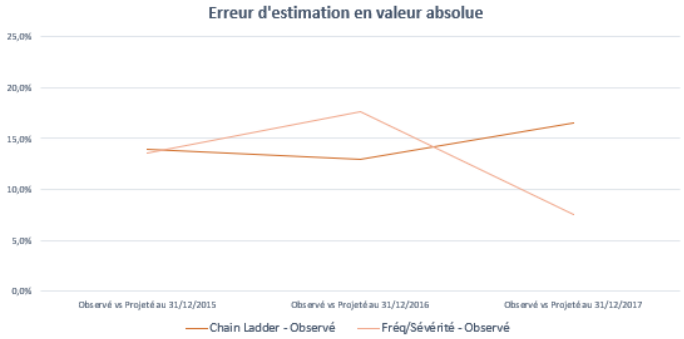

FIGURE 8 – Erreur d'estimation en % associée à chaque méthode

Les performances de la méthode de Chain Ladder et de l'approche fréquence / sévérité sont toutes deux peu satisfaisantes, l'une sous-estimant d'en moyenne 14% le montant d'IBNR à placer et l'autre le surestimant du même pourcentage.Nous avons vu durant nos analyses que considérer l'inflation permettrait d'isoler la composante inflation de la composante aggravation / amélioration et ainsi mieux estimer la charge allouée par Chain Ladder.
En ce qui concerne la performance de notre application de la méthode de P.Parodi, il est difficile d'en tirer des conclusions pertinentes car elle repose sur beaucoup d'hypothèses (prise en compte du taux d'inflation dans les projections, prise en compte des sinistres sans suite, calibration d'une seule et même loi pour les sinistres tardifs et les réouvertures, absence de cadence de liquidation et de paiement...). Notre méthodologie ne considère pas non plus que les sinistres tardifs puissent être sans suite, ce qui constitue

un axe d'amélioration majeur.

Pour conclure, la méthode présentée à de nombreux avantages et pourrait être intéressante en provisionnement, surtout dans des domaines comme la Responsabilité Civile Médicale où il y a un délai potentiellement important entre la survenance du dommage et sa réclamation. L'approche proposée est plus complète que la méthode Chain Ladder, mais les projections semblent surestimer la charge ultime et nécessite de réelles améliorations sur plusieurs aspects :

- Il serait préférable de mieux prendre en compte l'effet de l'inflation sur les projections en s'appuyant sur des hypothèses et des analyses plus robustes. Cela permettrait de n'observer que l'aggravation ou l'amélioration des sinistres sans le biais lié à l'inflation.
- Du fait de la branche considérée (la RC automobile corporelle), nous étions dès le début limité à un nombre de sinistres tardifs très réduit dans le temps (les deux premières années suivant la survenance). Cependant cela nous a donné l'occasion de nous pencher sur le cas des réouvertures de sinistres qui est un réel problème pour le provisionnement des sinistres graves en assurance directe mais également un sujet de tarification au niveau de la réassurance par exemple.
- Afin de modéliser les tardifs et les réouvertures de manière optimale, il aurait fallu disposer de l'exposition et calibrer deux lois distinctes aux délais des sinistres tardifs d'une part et aux réouvertures d'autre part afin de les utiliser séparément lors de la projection des nombres de sinistres.
- Pour la modélisation des sinistres tardifs, il faudrait approfondir le fait que des sinistres tardifs peuvent donner lieu à des sans suites et procéder à une approche plus poussée pour la modélisation de ces derniers. Il faudra aussi améliorer la modélisation en ce qui concerne les cadences de liquidation et de paiement.

Malgré tout, ce type de modélisation offre un cadre de calcul qui se rapproche de ce que l'on peut retrouver en tarification et propose une approche cohérente pour la tarification et le provisionnement. Elle serait également intéressante à étudier sur d'autres garanties et en réassurance.

Summary note

The inversion of the production cycle characterizes the insurance sector. Products are sold before their final cost is known, creating a real need for insurers to build up technical reserves. Thus, the estimation of reserves requires a high degree of precision and the calculation methods used must be adapted to the characteristics of the claims.
Traditional provisioning methods such as Chain Ladder are aggregated approaches : they compress data into triangles, not allowing for an individual view of the claims in the portfolio at all. Despite their ease of implementation and reporting, aggregating claims rules out valuable information that could improve the accuracy of estimates, especially those made for late claims.

The *Incurred But Not Reported* reserve is made of two components. The IBNyR [1] part represents the reserve that must be placed for claims that have already occurred but not yet reported to the insurer, while the second part, IBNeR [2], covers the potential shortfall in the reserves already established for claims that have occurred and have already been reported.
This paper studies an alternative approach to the traditional methods used in the calculation of the IBNR reserve as it introduces the use of a "frequency/severity" modeling for late claims. The approach proposed in the work is inspired by the work of *Schnieper (1991)* and *Pietro Parodi (2013)* where we find this separation of IBNyR and IBNeR claims. Indeed, by the nature and quantity of information available for each of these elements, it seems interesting to propose different and adapted methods.

The initial objective of this paper was to model separately IBNyRs (late claims that have occurred and are not yet reported) and IBNeRs (claims that are already known to the insurer and are being paid). However, we encountered two obstacles that did not allow the work to be completed satisfactorily :

- First of all, the claims at our disposal are from the motor third party liability coverage over a period between 2009 and 2018, a coverage for which we realized rather late that late claims are very rare. Indeed, bodily injury and/or property claims are declared very quickly in practice, which leaves little room for the need to model IBNyR.
- Moreover, we did not have the data relating to the exposure to risk on the portfolio available and, moreover, it was not from the French market, which prevented a complete analysis of the data and results. Due to the lack of expertise and knowledge of the data environment as well as its economic and legal context, we had to make simplifications, assumptions and restatements throughout the work, which made the implementation of the method complicated.

In general, in the automobile industry, claims that are reported more than two years after their occurrence cannot be considered as late claims in the traditional sense, as they are often reopenings. We have decided

1. Incurred But Not Yet Reported
2. Incurred But Not Enough Reported

to model late claims (declared within two years of occurrence) in frequency/severity. The modeling of reopenings can be done in the same way by calibrating an adapted law, but this has not been treated in the subject due to lack of time. The theoretical foundations of the method presented remain valid in spite of everything and can be applied in the case of guarantees where late claims are found, for example in Medical Liability where the delay between the occurrence and the claim can be relatively long, but also for reopenings.
In the first part of the modeling, the objective is to predict the number of late claims. To do so, we analyzed the reporting delays to estimate as accurately as possible the number of claims that will be reported in the future. Secondly, we studied the distribution of claims expenses in the portfolio in order to be able to allocate an expense to the estimated number of late claims. Once the modeling of late claims is done, we turn to the subject of IBNeR, claims already known but whose future variation must be estimated. The latter is calculated using a Chain Ladder method with a triangle of claims expenses that are still in progress at the valuation date and that were reported as soon as they occurred in order not to leave room for new claims in the portfolio. The sum of the IBNyRs and IBNeRs will then be used to estimate the IBNR reserve.

Calculation of the IBNR reserve at 31/12/2018 with the Chain Ladder method

First of all we present the results obtained with Chain Ladder for the period 2009-2018, one of the most widely used methods in practice because it will allow us to compare the performance of our model against a traditional method. In this section, the known claims expenses have been aggregated into a triangle. The year of attachment (on the line) is the year of occurrence and the columns represent the year of development. After validation of the assumptions of Chain Ladder, the IBNR reserve is estimated at €56 207 974. It is obtained by deducting the amounts known as of 12/31/2018 from the projected ultimate amounts and provides an estimate of the provision to be placed for all claims occurring over the assessment period.

Occurence	0	1	2	3	4	5	6	7	8	9	Ultime	Reserve
2009	14 485 401	27 810 864	34 679 862	38 534 466	41 810 003	43 990 855	45 087 721	44 859 167	45 332 597	45 195 672	45 195 672	-
2010	13 135 693	26 737 946	31 364 706	35 285 311	37 845 715	39 372 862	40 486 855	40 016 346	40 109 641	39 988 492	39 988 492	- 121 149
2011	14 146 964	27 141 883	31 681 244	35 114 914	37 157 607	38 619 399	38 879 888	39 334 870	39 597 513	39 477 911	39 477 911	143 042
2012	16 011 252	28 414 095	33 816 625	37 657 678	40 311 220	41 871 818	42 881 945	42 797 844	43 083 611	42 953 479	42 953 479	71 534
2013	12 888 928	24 768 498	30 165 924	34 341 997	36 759 707	38 144 421	38 954 887	38 878 488	39 138 085	39 019 870	39 019 870	875 449
2014	12 041 707	22 220 071	28 385 083	32 093 163	34 221 616	35 653 976	36 411 526	36 340 115	36 582 763	36 472 267	36 472 267	2 250 651
2015	12 110 485	24 864 658	30 985 738	34 967 113	37 442 127	39 009 282	39 838 124	39 759 993	40 025 475	39 904 581	39 904 581	4 937 467
2016	11 351 040	24 971 808	31 283 840	35 092 516	37 576 406	39 149 181	39 980 995	39 902 584	40 169 019	40 047 690	40 047 690	8 763 850
2017	12 101 769	27 326 301	33 326 023	37 383 326	40 029 362	41 704 807	42 590 921	42 507 392	42 791 219	42 661 970	42 661 970	15 335 669
2018	11 448 098	22 674 504	27 652 885	31 019 507	33 215 104	34 605 336	35 340 606	35 271 295	35 506 806	35 399 560	35 399 560	23 951 461
	1,98	1,22	1,12	1,07	1,04	1,02	1,00	1,01	0,997		401 121 492	56 207 974

FIGURE 9 – Result of the calculation of the IBNR reserve on the expense triangle for claims incurred between 2009 and 2018

Calculation of the IBNyR component

The approach used for the estimation of the *IBNyR* part separates the modeling of the frequency of late claims and the modeling of the associated load. The late loss frequency distribution is calculated by fitting a statistical distribution to the reporting delays. The expense is modeled following the analysis of expenses placed for known claims.
In both cases, we seek to fit the best theoretical statistical law to the observed data in the database in

order to find the distribution that will best fit each variable.
By combining the late loss frequency distribution and the fitted distribution for expenses and by simulating many scenarios, we will be able to obtain a distribution for the reserve $IBNyR$ and to estimate its amount using the Monte Carlo method.

Adjusting a statistical law to reporting delays

The objective here is to adjust the best statistical law to the reporting delays observed in the database in order to deduce the number of claims incurred over the period 2009-2018 that are not yet reported. We analyze the adjustment of several laws in order to choose the one that seems best suited to our data.

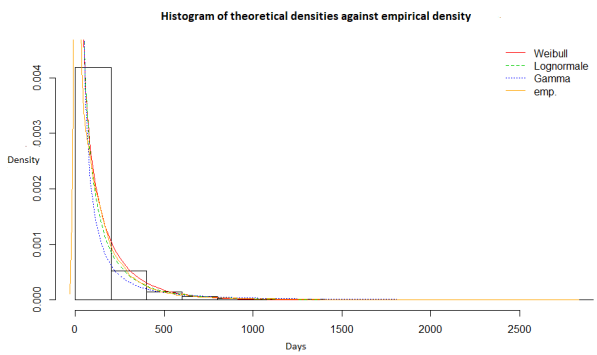

FIGURE 10 – Comparison of candidate laws to empirical reporting delays

Thanks to the various graphical outputs, descriptive statistics and law suitability tests (Kolmogorov Smirnoff, Anderson Darling, Cramer Von Mises), the Weibull law is considered to be the most relevant to represent the reporting deadlines among the candidate laws.

Adjustment of a frequency law for late claims

After adjusting a law for reporting delays, we operate it with the goal of obtaining the ultimate number of late claims for each year of occurrence. The formula for obtaining the ultimate late claims, demonstrated in the brief, is written as follows :
$\hat{N}_t - n_t = \frac{t}{\sum_{u=0}^{t} F(t-u)} * n_t - n_t$ where F is the distribution function of the law chosen to represent reporting delays. The ultimate goal of this part is to obtain a law for the frequency of the $IBNyR$. This is represented by a Binomial Negative Frequency distribution of parameters (r, p).

Occurrence year	Ultimate	Projected IBNyR number	r	p
2009	5750	0	154,94	0,000
2010	4662	0	154,94	0,000
2011	4034	0	154,94	0,000
2012	4513	0	154,94	0,000
2013	4190	0	154,94	0,000
2014	3881	0	154,94	0,000
2015	4147	5	154,94	0,032
2016	4147	35	154,94	0,186
2017	4025	146	154,94	0,486
2018	2470	609	154,94	0,797
Agregated	**41 820**	**796**	**154,94**	**0,836**
Mean	**4 182**			
Variance	**652 139**			
Ratio Variance Mean	**155,94**			

FIGURE 11 – Obtaining the parameters of the distribution of the frequency law

We were able to compare this number of late claims projected by the P.Parodi approach with the number of late claims projected by the Schnieper method. With Parodi, the modeled late claims appear from the year 2015 onwards, which is consistent with the reporting delays we observe in the database. For the years of occurrence 2009 - 2014, no late claims are modeled, which is different from the Schnieper approach that identifies late claims (reopenings) up to 7 years after occurrence. The model developed from Pietro Parodi's approach makes sense given that almost all claims are reported within two years, but does not allow for the rarer observations. The events where claims are reported very late are extremely rare and may not be captured by the tail of the Weibull distribution. This allowed us to see that an avenue for improvement would be to use extreme value theory to assign a specific distribution to very late events. Although it is very sensitive to the choice of exposure for each year of occurrence, the Schnieper method allows us to differentiate between new claims and non-sequential claims, which is a real advantage. In the calculations, new claims are measured against the exposure, while no claims are measured against the number of previously observed or estimated claims. The Schnieper method is very dependent on exposure information. The sensitivity of the Schnieper approach to exposure is shown below. The results obtained by P. Parodi's method are quite close to Schnieper's results when we consider a decrease in exposure similar to the one observed on new claims in the first year of opening. In our application of Parodi, we

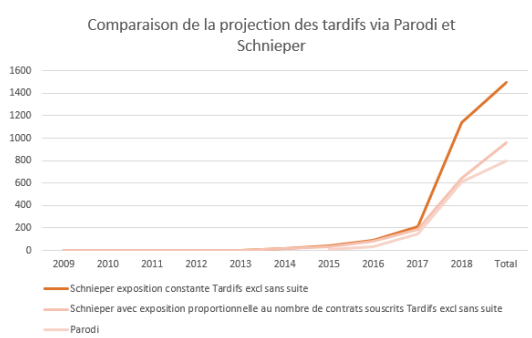

FIGURE 12 – Comparaison des sinistres tardifs obtenus avec Schnieper et avec Parodi

have not taken into account the claims that will potentially be declared as no claims. This presents an area for improvement in the calculation because if we are projecting late claims, some of these will be closed and therefore not result in a claim for the insurer. Thus, it would have been preferable to prepare

the data taking into account the proportion of claims that might be closed out in order to bring the same vision to Parodi's approach that can be found in Schnieper's method. The late claims would be projected by taking the entire database and the no claims would be deducted from it afterwards.

Adjustment of a statistical law to expenses

The objective here is to associate a burden amount with the late claims that were estimated in the previous section.
The process of adjusting a statistical law for expenses, while similar to that for reporting delays, requires more work beforehand. The three points of attention are as follows :

- For claims that have been closed for some time, it is important to study the inflation issue.
- Claims with an "open" status in the portfolio have the additional characteristic of not being fully settled and their expense amounts are subject to variations.
- We also have to make an important choice : is the distribution of the expense the same for all the years of occurrence or do we have to take into account certain phenomena specific to each year of occurrence ?

After addressing the 3 issues, we apply a process similar to that of adjusting a law to reporting timeframes.

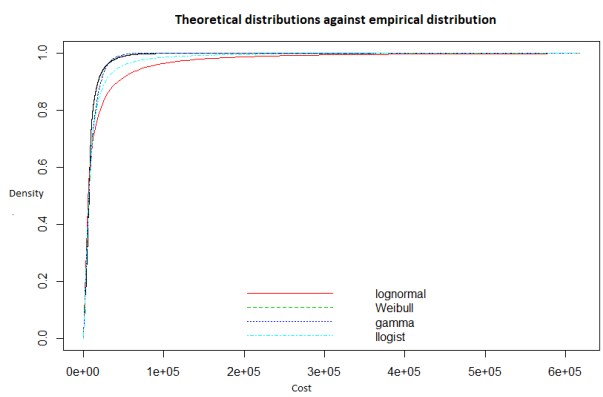

FIGURE 13 – Comparison of the distribution of the empirical data and the distribution of the tested laws

Our analysis shows that the Weibull law with the parameters $\lambda = 0.9862$ and $\kappa = 9208.992$ is the one that coincides best with the empirical data among those we tested. This fit could be improved with a better consideration of inflation and no claims analysis. However, since the portfolio and market data were very limited during our work, we were unable to optimize this model.

Pooling of the law of frequency and the law of the amount of charges

By applying the Monte Carlo method and thanks to its properties, we are able to estimate the amount of the reserve *IBNR*.
Indeed, by simulating the number of late claims drawn from the Binomial Negative law a certain N_{sim} number of times and by allocating each time an amount of charges drawn from a Weibull law to the generated number, we are able to obtain the distribution of the reserve *IBNyR*. More precisely, the algorithm combines the late loss frequency law and the expense-adjusted law and proceeds as follows :

— For each scenario $j = 1$ up to N_{sim}, we simulate n_j late claims from a Binomial Negative Law of parameters $p = 0.816$ and $r = 154.94$.
For each $j \in [1, N_{sim}]$, $n_j \sim F_X(x)$ where $F_X(x)$ is the distribution function of the NB(r, p) distribution function. each $i \in [1.n_j]$, we simulate a sinister charge x_i^j from a Weibull law of parameters $\lambda = 0.9862$ and $\kappa = 9208.992$.

— Calculate the aggregate amount of the sinister load for each j scenario, such that $S_j = \sum_{i=1}^{n_j} x_i^{(j)}$ $\forall j = 1, ..., N_{sim}$ that we put into a vector of size N_{sim}.

The final result is a vector of size N_{sim} where we have all possible values of the ultimate load of the simulated late claims.
Using the Monte Carlo method and the law of large numbers, the ultimate amount of the late claims expense is the empirical average of all possible values calculated :

$$\boxed{\delta = \frac{1}{N_{sim}} * \sum_{j=1}^{N_{sim}} S_j}$$

Calculation of the IBNeR component

The known nature of the claims included in the IBNeR reserve leads us to retain the Chain Ladder method for its estimation without adjusting for inflation, although other individual methods could be explored to achieve the result. The triangle is constructed by aggregating the expenses placed on outstanding claims at the valuation date. The year of attachment is the reporting year, which ensures that no new claims are introduced from one development year to the next.

Occurrence	0	1	2	3	4	5	6	7	8	9	Ultime	Reserve
2009	4 315 074	5 662 927	6 952 301	7 595 574	8 253 566	8 576 683	8 634 905	8 284 186	8 325 257	8 108 664	8 108 664	-
2010	6 668 937	8 763 137	9 660 717	10 438 966	10 872 858	11 068 845	11 283 530	10 752 821	10 575 784	10 300 640	10 300 640	- 275 144
2011	6 750 257	9 277 136	10 277 364	10 738 330	11 185 603	11 537 550	11 195 483	11 185 076	11 105 190	10 816 273	10 816 273	- 368 803
2012	13 255 614	17 086 754	18 991 629	20 062 158	20 727 623	21 028 195	21 333 046	20 721 565	20 573 568	20 038 318	20 038 318	- 1 294 728
2013	11 591 989	15 788 351	17 717 754	18 991 679	19 867 506	20 452 436	20 544 762	19 955 876	19 813 348	19 297 876	19 297 876	- 1 154 561
2014	11 280 646	14 622 280	16 626 697	18 150 390	18 646 168	19 108 082	19 194 339	18 644 161	18 511 001	18 029 411	18 029 411	- 616 757
2015	11 071 950	15 742 735	18 372 708	19 708 984	20 528 781	21 037 332	21 132 298	20 526 571	20 379 967	19 849 753	19 849 753	140 769
2016	13 010 248	19 127 614	22 393 143	24 002 673	25 001 067	25 620 408	25 736 062	24 998 375	24 819 832	24 174 109	24 174 109	1 780 966
2017	17 224 926	25 532 778	29 124 566	31 217 923	32 516 436	33 321 952	33 472 373	32 512 936	32 280 722	31 440 894	31 440 894	5 908 116
2018	19 151 264	26 482 999	30 208 457	32 379 720	33 726 558	34 562 052	34 718 071	33 722 927	33 482 072	32 610 989	32 610 989	13 459 725
	1,38	1,14	1,07	1,04	1,02	1,00	0,97	0,99	0,97		194 666 926	17 579 583

FIGURE 14 – Projection of the cost of claims in progress to the ultimate

The reserve to be invested in respect of current claims is €17 579 583. We observe a catch-up on the older years (2010-2014) due to an excess of file/file provision.

Validation of the results : comparison with Chain Ladder and the observed

An important limitation in our validation is that the method implemented in this work does not include a liquidation rate for latecomers and reopeners. Indeed, we do not know how to split the loads diagonally for the late part. Thus, the figures in our method below overestimate the diagonals in part because of the projections of the latecomers (not on the IBNeR part which is in Chain Ladder) which are seen at the ultimate.
Generally speaking, the Chain Ladder method gives lower estimates than the amounts actually observed, whereas the method that takes into account late arrivals via a frequency/severity approach gives amounts that are higher than the observed data (consistent with the preliminary remark above).
The performance of the Chain Ladder method and the frequency/severity approach are both unsatisfactory,

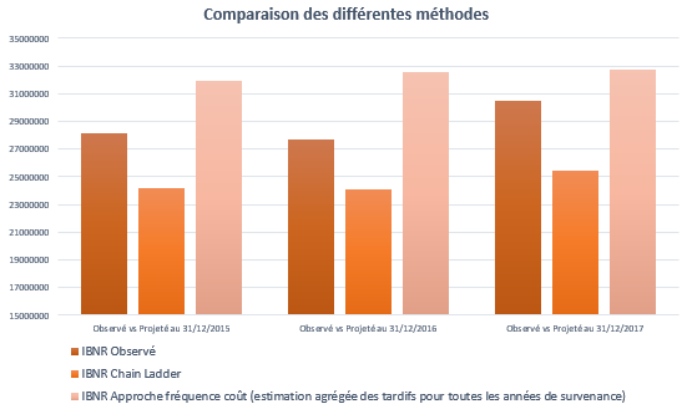

FIGURE 15 – Comparison of projections with observed at cut-off dates 31/12/2015, 2016 and 2017

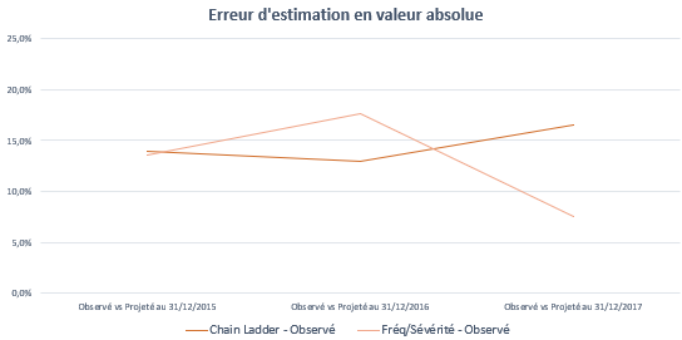

FIGURE 16 – Estimation error in % associated with each method

one underestimating the amount of IBNR to be placed by an average of 14% and the other overestimating it by the same percentage.
We saw during our analyses that considering inflation would allow us to isolate the inflation component from the aggravation/improvement component and thus better estimate the load allocated by Chain Ladder. Regarding the performance of our application of P. Parodi's method, it is difficult to draw relevant conclusions because it is based on many assumptions (taking into account the inflation rate in the projections, taking into account no claims, calibration of a single law for late claims and reopenings, absence of liquidation and payment rate...). Our methodology also does not consider that late claims can be no claims, which is a major area for improvement.

To conclude, the method presented has many advantages and could be interesting in provisioning, especially in areas such as Medical Liability where there is a potentially significant delay between the occurrence of the damage and its claim. The proposed approach is more complete than the Chain Ladder method, but the projections seem to overestimate the ultimate burden and requires real improvements on several aspects :

- It would be preferable to better account for the effect of inflation on the projections by relying on more robust assumptions and analyses. This would allow only the worsening or improvement of claims to be observed without the inflation bias.
- Because of the line of business under consideration (bodily injury motor liability), we were limited

from the outset to a very small number of late claims (the first two years following occurrence). However, this gave us the opportunity to examine the case of claim reopenings, which is a real problem for the provisioning of serious claims in direct insurance, but also a subject of pricing in reinsurance, for example.

— In order to model late claims and reopeners in an optimal way, it would have been necessary to have the exposure and to calibrate two separate laws for late claims on the one hand and for reopeners on the other hand in order to use them separately when projecting the number of claims.

— For the modeling of late claims, the fact that late claims may result in no claims should be further explored and a more sophisticated approach to modeling late claims should be taken. The modeling also needs to be improved with respect to liquidation and payment rates.

Nevertheless, this type of modeling offers a computational framework that is close to what can be found in pricing and offers a consistent approach to pricing and provisioning. It would also be interesting to study on other guarantees and in reinsurance.

Remerciements

J'aimerais remercier chaleureusement l'équipe OPTIMIND pour son accueil, sa bonne humeur et cet agréable environnement de travail dans lequel j'ai pu grandir pendant une année, et en particulier toutes les personnes qui m'ont aidée ou m'ont proposé leur aide pour mener à bien mon alternance.

Je remercie tout particulièrement M. Amine Zeffane et M. Emmanuel Guiffart, mes managers pour leurs précieux conseils au cours de cette année. J'aimerais également remercier M. Antoine Pesneaud et M.Thomas Canon, mes tuteurs en entreprise pour leur accompagnement, leur disponibilité et leur guidance très précieuse tout au long de l'établissement de ce mémoire.

Je tiens à remercier l'ensemble du corps enseignant de l'ISUP pour la formation de qualité dispensée et particulièrement à M. Oliver Wintenberger mon tuteur académique pour son encadrement.

Je tiens également à remercier Madame Anne Bontoux, Madame Sylvie Chanh, Monsieur François Blanchet, Monsieur François Gelan et Monsieur Mohamed Amine Bouayad de m'avoir accordé leur temps précieux et pour toutes leurs connaissances du métier qu'ils ont partagé afin de m'aider dans l'élaboration de ce mémoire.

Et pour finir, je souhaite remercier mes parents, mes frères et mon compagnon pour leur soutien quotidien et irremplaçable.

Table des matières

Note de synthèse **2**

Summary note **10**

Remerciements **17**

Table des matières **21**

Introduction **22**

I Généralités de l'assurance non-vie **24**

1 Contexte **25**
1.1 Généralités . . . 25
1.1.1 Définition du contrat d'assurance . . . 25
1.1.2 L'inversion du cycle de production . . . 25
1.2 Description de l'assurance automobile . . . 25
1.2.1 Les principales garanties . . . 25
1.2.2 Exposition au risque automobile . . . 26
1.2.3 La nature des sinistres en assurance automobile . . . 26
1.3 Le marché de l'assurance automobile au Maroc . . . 27
1.3.1 La garantie Responsabilité Civile . . . 27
1.3.2 Deux types d'indemnisation différents . . . 28
1.3.3 Quelques chiffres clés . . . 28
1.4 Environnement juridique . . . 29
1.4.1 En Europe . . . 29
1.4.2 En France . . . 29
1.4.3 En Espagne . . . 31
1.5 Spécificités de la vie d'un sinistre . . . 31
1.5.1 Le cycle de vie du sinistre . . . 31
1.5.2 La décomposition des sinistres . . . 33

2 Méthodes de provisionnement classiques **37**
2.1 Méthode de triangulation/agrégation des données . . . 37
2.2 La méthode la plus classique : Chain-Ladder . . . 38
2.2.1 Triangles de liquidation . . . 38
2.2.2 Hypothèses de calcul . . . 39
2.2.3 Résultats de la méthode . . . 39

2.2.4 Avantages et limites de la méthode . . . 39
2.3 La méthode Bornhuetter Fergusson . . . 40
2.3.1 Principe . . . 40
2.3.2 Hypothèses . . . 40
2.3.3 Calcul . . . 41
2.4 La méthode stochastique de Mack . . . 41
2.4.1 Hypothèses de calcul . . . 41
2.4.2 Validation du modèle . . . 42
2.4.3 Avantages et inconvénients de la méthode . . . 43
2.5 La méthode de Schnieper . . . 43
2.5.1 Notations . . . 43
2.5.2 Hypothèses de la méthode . . . 43
2.5.3 Estimations des paramètres . . . 44
2.5.4 Avantages et inconvénients . . . 45
2.6 Approche proposée par Pietro Parodi . . . 45
2.7 Inconvénients des méthodes agrégées, intérêt des méthodes individuelles et présentation de la problématique . . . 46

3 Analyse exploratoire des données 48
3.1 Présentation de la base . . . 48
3.2 Retraitements effectués . . . 48
3.3 Analyses graphiques . . . 49

4 Résultats des méthodes de provisionnement classiques 54
4.1 Modélisation par Chain Ladder de la réserve IBNR . . . 54
4.1.1 Vérification des hypothèses . . . 54
4.1.2 Résultats de la méthode . . . 56
4.2 Application de Chain Ladder sur méthode inflatée . . . 57
4.2.1 Choix de l'inflation . . . 57
4.2.2 Mise en as-if des sinistres et calcul . . . 58
4.2.3 Calculs . . . 58
4.3 Méthode de Mack . . . 61
4.4 Application de Schnieper sur le triangle de nombre . . . 62

II Approche hybride d'estimation de la réserve IBNR : modélisation séparée du nombre et du coût pour les sinistres tardifs 65

5 Estimation des sinistres tardifs par une approche fréquence / coût 66
5.1 Notions théoriques . . . 67
5.1.1 Ajustement d'une loi de distribution à un échantillon de données . . . 67
5.1.2 Statistiques descriptives et mesures de forme . . . 67
5.1.3 Méthode du maximum de vraisemblance . . . 68
5.1.4 Ajustement des lois candidates aux données empiriques de la base . . . 68
5.1.5 Test d'adéquation aux lois / Goodness-of-fit . . . 69
5.1.6 Méthode de Monte Carlo . . . 70
5.2 Application (1) : modélisation du nombre des sinistres tardifs . . . 72
5.2.1 Ajustement d'une distribution des délais de déclaration . . . 72
5.2.2 Estimation du nombre de sinistres tardifs à l'ultime par la modélisation des délais de déclaration . . . 76
5.2.3 Obtention des paramètres de la loi pour le nombre des sinistres . . . 79
5.3 Application (2) : Modélisation de la charge des sinistres tardifs . . . 80
5.3.1 Détermination des ultimes pour les sinistres clos . . . 82
5.3.2 Prise en compte des sinistres en cours / IBNER . . . 82
5.3.3 La loi des distributions de la charge est-elle commune à toutes les années de survenance ? . . . 83
5.3.4 Ajustement d'une loi statistique aux montants des charges . . . 84

5.4 Application (3) : mise en commun des nombres et de la charge des sinistres tardifs 85
5.4.1 Déroulement de la simulation pour l'estimation de la charge des sinistres tardifs . . 85
5.4.2 Résultats de la simulation 86

6 Obtention du montant de la réserve au titre des sinistres déclarés encore en cours 87
6.1 Contexte 87
6.2 Implémentation de Chain Ladder sur les sinistres en cours 88
6.3 Montant de la réserve IBNR 89

7 Validation de la méthode 90

Bibliographie 96

Liste des abréviations 98

Table des figures 98

Annexes 101

A Chiffres clés pour l'Assurance Automobile en France 102

B La convention d'Indemnisation et de Recours Corporel Automobile (IRCA) en France 104

C Liste des pays adhérant au système carte verte 107

D La Nomenclature Dintilhac 109
D.1 Préjudices de la victime directe 109
D.1.1 Préjudices patrimoniaux 109
D.1.2 Préjudices extra-patrimoniaux 109
D.2 Préjudices des victimes indirectes 110
D.2.1 Préjudices des victimes indirectes en cas de décès de la victime directe 110
D.2.2 Préjudices des victimes indirectes en cas de survie de la victime directe 110

E Estimateur du maximum de vraisemblance 111

F Tests d'adéquation statistiques 112
F.1 Le test de Kolmogorov Smirnoff 112
F.2 Le test Anderson-Darling et le test de Cramer Von Mises 113

Introduction

L'inversion du cycle de production est un élément qui caractérise le secteur de l'assurance. Les produits sont vendus avant d'en connaître leur coût définitif, ce qui engendre un réel besoin pour les assureurs et les réassureurs de constituer des provisions techniques. Ces provisions, aussi connues comme les réserves de l'assureur, sont l'un des éléments le plus important indiqués au bilan ; elles permettent d'assurer que les engagements pris envers la clientèle (promesses d'indemnisation en cas de survenance des risques garantis au contrat) soient respectés, et par conséquent participent fortement à la viabilité de l'entreprise. La modélisation des provisions techniques est strictement règlementée, les hypothèses utilisées dans les calculs doivent être clairement justifiées et l'objectif ultime du provisionnement est d'obtenir des estimations futures les plus précises qui soient.

L'objectif initial de ce mémoire était de trouver une alternative aux méthodes traditionnelles d'estimation des sinistres tardifs en faisant une modélisation séparée des IBNyR (sinistres tardifs survenus et pas encore déclarés) et des IBNeR (sinistres déjà connus de l'assureur en cours d'indemnisation). Cependant nous avons rencontré deux obstacles qui n'ont pas permis aux travaux d'aboutir de manière satisfaisante :

- Tout d'abord, les sinistres à disposition sont issus de la garantie Responsabilité Civile Automobile, garantie pour laquelle nous nous sommes aperçus assez tardivement que les sinistres tardifs sont très rares. En effet les sinistres corporels et / ou matériels sont déclarés relativement rapidement dans la pratique, ce qui laisse peu de place au besoin de modéliser les IBNyR.
- Par ailleurs, nous ne disposions pas de données relatives à l'exposition au risque sur le portefeuille à disposition et il n'est pas issu du marché français, ce qui a empêché l'analyse complète des données et des résultats. Du fait du manque d'expertise et de connaissances sur l'environnement des données ainsi que sur leur contexte économique et juridique, nous avons été amenés à faire des simplifications, hypothèses et retraitements tout au cours des travaux qui ont fait que la mise en pratique de la méthode fut compliquée.

En automobile, les sinistres déclarés au-delà de deux ans après leur survenance ne peuvent pas être des tardifs au sens traditionnel, il s'agit très probablement de réouvertures. Les réouvertures de sinistres corporels en automobile posent un vrai problème dans le marché français, notamment pour les graves. Ainsi, nous avons pris la décision de modéliser en fréquence/sévérité, d'une part les sinistres tardifs déclarés dans les deux premières années et d'autre part les réouvertures (les sinistres déclarés après cinq ans). Les fondements théoriques de la méthode présentée restent valables malgré tout et pourront s'appliquer dans le cas de garanties où l'on retrouve bien des sinistres tardifs comme par exemple en Responsabilité Médicale.

Les méthodes de provisionnement traditionnelles utilisées dans la modélisation de la réserve IBNR telles que Chain Ladder, Mack... sont des approches dites "agrégées" : elles compressent les données en triangles, ne permettant pas du tout d'avoir une vision individuelle des sinistres dans le portefeuille. Malgré la facilité

d'implémentation et de communication des résultats des méthodes classiques, le fait d'agréger les sinistres écarte des informations précieuses, en particulier des informations qui sont propres à la fréquence et au coût (tardifs, sans suite, délais de déclaration, statut des sinistres dans le portefeuille) et qui pourraient s'avérer très utiles dans l'amélioration des prédictions concernant les sinistres non connus et dès lors dans l'estimation des provisions techniques.

Les travaux présentés cherchent à appliquer autant que possible les travaux de Pietro Parodi, et distingue les deux composantes de la réserve IBNR (les IBNyR et les IBNeR) comme le propose la méthode de Schnieper.
Ce mémoire nous a permis de comprendre différents aspects très importants pour la modélisation tels que le traitement de l'inflation, les problématiques de réouvertures, les sinistres déclarés sans suite et nous avons pu tester en provisionnement quelques approches utilisées classiquement en tarification.
La méthode présentée offre de nombreux avantages : elle permet d'inclure dans la modélisation des caractéristiques individuelles des sinistres connus comme les délais de déclaration, elle ne mélange pas les charges des sinistres en cours et des sinistres clos et elle ne se base pas sur la cadence passée pour en déduire la cadence future des charges. Cette méthode limite également le jugement d'experts en proposant une manière plus sophistiquée de calculer les facteurs de queues dans le cas où le déroulement des sinistres est plus long que l'historique des données à disposition.

Ce mémoire présentera tout d'abord les généralités de l'assurance non-vie et surtout l'environnement juridique de nos données, puis nous détaillerons la théorie des différentes méthodes de provisionnement usuelles. Nous étudierons ensuite les première analyses exploratoires des données et l'application des méthodes de provisionnement usuelles. Enfin notre modélisation se déroulera en deux parties : pour la modélisation des IBNyR, nous chercherons d'abord à ajuster une loi statistique aux délais de déclaration qui permettra de projeter le nombre de sinistres tardifs puis nous calibrons une loi aux charges connues pour associer un montant de charge aux sinistres tardifs projetés ; puis pour les IBNeR , nous conserverons la méthode Chain Ladder pour les sinistres hors tardifs. Les résultats totaux seront par la suite comparés à ceux obtenu avec un Chain Ladder classique afin d'évaluer la performance de l'approche proposée.
Pour conclure, nous présenterons les avantages et les inconvénients de cette méthode ainsi que les aspects qui pourraient faire l'objet d'améliorations et d'approfondissement pour que cette dernière puisse être considérée dans la pratique.

Première partie

Généralités de l'assurance non-vie

Chapitre 1

Contexte

1.1 Généralités

1.1.1 Définition du contrat d'assurance

Tout individu ou entreprise est tenu d'assurer ses biens mobiliers, immobiliers ainsi que sa responsabilité. L'assurance non-vie rassemble tous les contrats de type IARD (Incendies, Accidents et Risques Divers) dont le types de risques garantis ne sont ni liés, ni conditionnés à la vie de l'assuré.
Un contrat d'assurance possède plusieurs caractéristiques : il est consensuel (il naît d'un commun accord entre deux parties, l'assureur et le souscripteur), aléatoire (sa réalisation est subordonnée à la survenance d'un événement incertain) et synallagmatique (il fait naître des obligations réciproques entre l'assureur et l'assuré).
L'assureur rédige un contrat d'adhésion où il s'engage à indemniser l'assuré (c'est la personne sur qui repose le risque - cela peut être le souscripteur lui-même ou une tierce personne), dans le cas de la survenance d'un évènement couvert dans les garanties du contrat, en contrepartie d'une prime versée par le souscripteur. Le risque couvert est défini par les parties, généralement par des conditions générales et des conditions particulières du contrat. Sa survenance doit obligatoirement être indépendante de la volonté de chacune des parties. Dans le cas d'événements certains, impossibles ou dépendants de la volonté de l'assuré, ces derniers sont exclus du contrat d'assurance.

1.1.2 L'inversion du cycle de production

L'inversion du cycle de production est un élément important qui distingue le secteur de l'assurance des autres secteurs du marché. Le produit d'assurance est vendu avant que l'on en connaisse son coût définitif, ce qui engendre le besoin pour l'assureur de constituer des provisions techniques. Ces provisions représentent une large partie du passif de l'assureur et doivent être correctement estimées afin que ce dernier soit en mesure de répondre à ces engagements à tout moment.

1.2 Description de l'assurance automobile

1.2.1 Les principales garanties

— La garantie dommage automobile : elle couvre les dommages subis par le véhicule assuré (bris de glace, vol). Le montant de l'indemnisation est déterminé par l'examen du véhicule par un expert

indépendant.
- La Responsabilité Civile : cette garantie couvre les dommages subis par les tiers. En assurance automobile, il s'agit de réparer les dommages causés par le véhicule assuré à autrui : dégâts matériels, blessures, décès.
- La garantie conducteur : Le conducteur n'est pas un tiers et les dommages qu'il subit ne rentrent ni dans la catégorie Responsabilité Civile ni dans la garantie dommages automobiles. Ainsi, il existe une garantie spécifique, que les conducteurs peuvent souscrire s'ils le souhaitent. Cette garantie est indépendante de la Responsabilité Civile (encore une fois le conducteur n'est pas un tiers) et prend en charge les frais liés aux blessures ou au décès du conducteur lorsqu'il est responsable de l'accident.

1.2.2 Exposition au risque automobile

La nature du véhicule responsable est en général considérée comme un facteur explicatif de la fréquence et la sévérité des sinistres.
Il existe plusieurs catégories de véhicules pouvant être assurés dans le cadre de l'assurance automobile :
- 1^{e}catégorie : les véhicules qui sont inférieurs ou égal à 3.5 tonnes
- 2^{e}catégorie : supérieurs à 3.5 tonnes
- 3^{e}catégorie : les deux roues/voiturettes électriques
- 4^{e}catégorie : les engins de types agricoles
- 5^{e}catégorie : ceux de plus de 9 places affectés au transport de personnes et les transports publics de marchandises

Les deux dernières catégories peuvent notamment être générateurs d'accidents graves avec des dommages importants (matériels, immatériels ou corporels).

1.2.3 La nature des sinistres en assurance automobile

Les dommages causés par des accidents automobiles peuvent être de trois natures différentes :
- Matériels : ce sont les dommages qui sont causés à des véhicules ou à des biens tiers.
- Corporels : ce sont les dommages physiques qui sont causés aux tiers ou au conducteur(s) responsable(s) touchés lors de l'accident de voiture (le conducteur doit avoir souscrit au préalable une couverture spécifique pour se protéger)
- Immatériels : ce sont les pertes financières qui sont subies par les assurés - il s'agit le plus souvent de la privation de jouissance d'un bien, l'interruption d'un service ou alors tout dommage autre que corporel ou matériel.

Un bon exemple pour illustrer ce point est l'accident du tunnel du Mont Blanc. Le sinistre s'est produit du 24 au 26 mars 1999 dans le tunnel du Mont-Blanc entre la France et l'Italie, suite à l'embrasement d'un camion frigorifique semi-remorque belge transportant de la margarine et de la farine. Le violent incendie cause :
- La mort de 39 personnes (dommages corporels).
- La fermeture du tunnel pendant une durée d'environ 3 ans, ce qui a entraîné des pertes d'exploitation / pertes financières importantes (dommage immatériels).
- Des dommages au tunnel (dommages matériels)

Pour la garantie dommage automobile, il ne peut y avoir que des dommages matériels. Pour la garantie Responsabilité Civile Automobile, on peut rencontrer les trois types de dommages. En revanche, pour la garantie conducteur, il ne peut s'agir que de dommages corporels.

1.3 Le marché de l'assurance automobile au Maroc

Les données traitées dans le mémoire n'étant pas des données issues du marché français mais du marché marocain, il est important de tâcher de bien comprendre le contexte de l'assurance automobile du pays afin de mieux appréhender les données.
Le marché de l'assurance non-vie au Maroc est dominé par l'assurance automobile qui représentait, d'après la FSMAR [1], environ 48,9% des primes brutes en 2018.

Le système légal au Maroc est une combinaison des systèmes islamique, espagnol et français. La plupart des procédures légales sont basées sur les pratiques françaises. Le processus d'indemnisation est régi par un cadre précis dénommé le *Dahir du 02 Octobre 1984*.

La fréquence des sinistres corporels au Maroc est en baisse grâce à l'augmentation des mesures liées à la sécurité routière telle que la mise en place des radars qui deviennent de plus en plus fréquents.
En effet, les sinistres au Maroc peuvent être très graves, en partie à cause de l'état des routes et à la surcharge de camions et de cars, véhicules lourds qui peuvent être impliqués dans des accidents et dont les conséquences peuvent être très sévères dû au nombre de personnes transportées.

Le tarif de la garantie Responsabilité Civile au Maroc est un tarif unique, qui dépend de certains facteurs comme les bonus-malus de l'assuré ou le nombre de chevaux fiscaux du véhicule. Cette notion de tarif unique pour la Responsabilité Civile a pour but de limiter la compétition sur le marché. Cependant, la garantie dommages automobiles n'est pas tenue à ce tarif unique qui est libre pour tous les assureurs marocains.

Aussi, il existe le Fonds de Garantie des Accidents de la circulation marocain, dont les ressources sont financées en à environ 71% par la contribution des assurés, 18% par les produits des placements et environ 9% par la participation des entreprises d'assurance et de réassurance. Le reste des contributions vient des amendes. Ce fonds permet d'indemniser les victimes de dommages corporels causés par des véhicules terrestres à moteur (non liés à une voie ferrée) dans le cas de sinistres particuliers où l'auteur du fait est inconnu, ne dispose pas d'assurance ou bien alors quand l'assureur lui-même est insolvable.

1.3.1 La garantie Responsabilité Civile

La garantie Responsabilité Civile au Maroc est une garantie obligatoire pour toute personne physique ou morale. Ainsi, tout véhicule terrestre à moteur non lié à une voie ferrée doit être couvert pour une assurance Responsabilité Civile. Cette obligation d'assurance s'applique également aux conjoints ou descendants de l'assuré ou du conducteur.

La notion de partage de responsabilité est effectuée suite au rapport de la police. Les deux assureurs sont informés, et se mettent ensuite d'accord sur la responsabilité mais il n'existe pas de convention officielle qui évite à une victime de se retrouver devant plusieurs interlocuteurs. Le droit marocain fonde ainsi la responsabilité pour la réparation du préjudice sur la notion de faute dont la victime doit donner une preuve.
L'indemnisation est ensuite versée à la victime de l'accident, à l'exclusion du conducteur, du souscripteur du contrat (le propriétaire du véhicule assuré), les représentants légaux de la personne morale propriétaire

1. *Fédération marocaine des sociétés d'assurance et de réassurance*

du véhicule s'ils sont transportés dans le véhicule au moment de l'accident et les salariés du conducteur pendant leur service.

1.3.2 Deux types d'indemnisation différents

Indemnisation sur le territoire local

Lorsqu'un sinistre automobile survient sur le territoire marocain, alors il existe un dispositif d'indemnisation précis pour les victimes d'accidents causés par des véhicules terrestres à moteur, régit par le *Dahir du 2 Octobre 1984 au Maroc* qui se présente sous la forme d'un barème (on retrouve également ce principe d'indemnisation dans le droit espagnol). Ainsi, selon les dommages causés à la victime, le degré de gravité des blessures ou le niveau de perte d'autonomie, l'indemnisation est fixée à l'avance.
Il est possible que l'indemnisation soit jugée insuffisante par les victimes. Dans ce cas, si le barème est rejeté, alors il est possible que l'affaire soit portée au tribunal et des facteurs plus subjectifs rentrent en compte dans le processus d'indemnisation.
Cependant, il est important de noter que ce barème fonctionne bien et que les cas où les montants diffèrent de manière significative du barème sont rares.

Le système des cartes vertes

Au Maroc, comme dans de nombreux pays [1], il existe une extension de garantie pour l'assurance automobile, appelée la carte verte. Elle permet aux assurés de circuler librement à l'entrée d'un pays adhérent au système des cartes vertes sans souscription d'assurance frontière, son assurance locale suffisant à le garantir dans tous les pays adhérents. Le *Bureau Central Marocain des Sociétés d'Assurance - BCMA* est l'organisme officiel au Maroc chargé de la gestion de ce système.
Ainsi un assuré marocain victime ou responsable d'un sinistre automobile dans un pays adhérent au système carte verte sera indemnisé (dans le cas où il est victime) et c'est la loi du pays dans lequel est survenu le sinistre qui régira le processus d'indemnisation.
Dans ce cadre, il se peut que les délais de déclaration dans ces cas soient légèrement plus longs car le sinistre n'est pas forcément déclaré immédiatement à l'assureur marocain, mais le dispositif fonctionne plutôt bien et les délais de déclaration ne sont pas significatifs dans ces cas-là.

1.3.3 Quelques chiffres clés

L'assurance automobile au Maroc est en hausse de 7,2% en 2019 [2] en termes de primes émises. Nous avons pu avoir accès au bilan statistique provisoire des accidents et victimes au titre des premiers mois de l'année 2020 en comparaison avec les données provisoires de 2019. Ainsi, les résultats sont les suivants :

- Les accidents corporels au Maroc sont en baisse d'environ 25%
- Les accidents mortels en baisse d'environ 26%.
- Les blessés graves sont en baisse d'environ 29%
- Les blessés légers sont en baisse d'environ 28%

Cette baisse peut s'expliquer en partie par l'augmentation des mesures de la sécurité routière et par la pandémie du COVID qui a entrainé un nombre plus faible de véhicules sur les routes.
Aussi, le document explique que les premières victimes des accidents de la route sont des usagers vulnérables tels que les piétions et les conducteurs de deux ou trois roues, suivi par les usagers des voitures de tourisme.

1. voir annexe
2. Rapport du secteur des assurances 2019 - Autorité de Contrôle des Assurances et de la Prévoyance Sociale

1.4 Environnement juridique

Au cours des travaux, nous n'avons pas été pas en mesure de nous procurer des connaissances exhaustives juridiques ni d'expertise locale précise sur le Maroc qui auraient permis l'interprétation complète des résultats. Le droit marocain étant en partie inspiré du droit français et afin d'éclairer la compréhension des données, certaines de nos interprétations et analyses qui suivent ont été inspirées de l'environnement français. Ainsi le paragraphe suivant offre un tour d'horizon de cette garantie en Europe et notamment en France et en Espagne, dont la législation du Maroc est inspirée.

1.4.1 En Europe

Il existe six types de directives régissant l'assurance automobile en Europe.

- La 1e directive n°2/166/CEE du 24 avril 1972 rend la Responsabilité Civile Automobile obligatoire pour tous les membres de la Communauté Economique Européenne.
- La 2e directive n° 90/323/CEE du 30 décembre 1983 étend l'assurance Responsabilité Civile Automobile obligatoire aux dommages matériels pour tous les membres de la Communauté Economique Européenne.
- La 3e directive n° 90/232/CE du 14 mai 1990 offre une couverture à tous les passagers transportés (à l'exception des complices impliqués dans un vol de véhicule)
- La 4e directive n° 2000/26/CE du 16 mai 2000 permet aux victimes transfrontières de poursuivre l'assureur du conducteur responsable en justice.
- La 5e directive n°2005/14/CE du 11 mai 2005 fixe des garanties minimales révisées tous les 5 ans
- Et enfin, les directives codifiées 2009/103/CE du 16 septembre 2009 régissent qu'aucune transposition n'est nécessaire dans le droit national pour être applicable.

1.4.2 En France

Le Code des assurances

L'article L. 211-1 du Code des assurances indique qu'il est obligatoire de souscrire une assurance automobile pour « tout véhicule automoteur destiné à circuler sur le sol et qui peut être actionné par une force mécanique sans être lié à une voie ferrée, ainsi que toute remorque, même non attelée ».

La loi du 5 juillet 1985 - Badinter

La loi no 85-677 du 5 juillet 1985, dite loi Badinter, crée un régime spécial d'indemnisation des victimes d'accident de circulation. Cette loi française vise à l'amélioration de la situation des victimes d'accidents de la circulation en harmonisant et en accélérant les procédures d'indemnisation. Concernant le délai d'indemnisation, l'assureur qui garantit la responsabilité civile du conducteur du véhicule terrestre à moteur doit indemniser la victime dans les huit mois qui suivent l'accident de la route ou tout au moins proposer une offre prévisionnelle si l'état de la victime n'est pas consolidé. L'offre définitive d'indemnisation doit alors être faite dans un délai de cinq mois suivant la date à laquelle l'assureur a été informé de cette consolidation. L'offre, qu'elle soit provisionnelle ou définitive, doit indiquer l'évaluation de chaque chef de préjudice. Le montant des indemnités dues aux victimes est affiché dans les tribunaux et fixé par les juges ou par une transaction entre assureurs et assurés. Par conséquent, les tardifs en assurance automobile sont rares voire inexistants en France.

Le Fonds de Garanties Des Assurances Obligatoires de dommages

Le Fonds de Garanties des Assurances Obligatoires de dommages (FGA0) intervient au titre de la solidarité nationale dans l'indemnisation des victimes qui ne sont pas prises en charge par les assurances lorsque le responsable du préjudice n'est pas identifié ou assuré. Par un communiqué du 18 juin 2019, le FGAO rappelle que les trottinettes électriques (entre autres) sont considérées comme des véhicules terrestres à moteur, et par conséquent soumis à l'obligation d'assurance RC automobile.

La Nomenclature Dintilhac

En France, il n'existe pas de barème d'indemnisation. Pour l'indemnisation des victimes corporels, on retrouve la notion de réparation intégrale du préjudice - la victime doit être replacée dans l'état où elle se trouvait avant la survenance du sinistre, l'indemnité ayant pour mesure le préjudice subi. Ce principe impose une appréciation concrète du préjudice subi et sa mise en oeuvre relève ainsi du pouvoir souverain d'appréciation des juges du fonds. En effet, la détermination des postes de préjudices indemnisables est effectuée selon des règles d'évaluation prévues en droit commun qui sont appliquées par les tribunaux lorsqu'ils fixent la dette du responsable.
Le rapport de M. Jean-Pierre Dintilhac a pour but d'élaborer une nomenclature commune des préjudices corporels [1] en séparant les préjudices patrimoniaux (économiques) des préjudices extra-patrimoniaux (non-économiques). Le préjudice corporel comprend :

- les préjudices physiques
- les préjudices psychiques
- les préjudices professionnels
- les préjudices familiaux

La loi du 21 décembre 2006 impose aux professionnels (juges, experts, avocats, sécurité sociale) le recours à une méthodologie précise pour déterminer les préjudices subis. La nomenclature Dintilhac fixe des principes pour l'évaluation de la réparation résultant d'infractions ayant causé des dommages corporels à une victime. Elle fixe vingt postes pour les victimes directes et sept postes pour les victimes indirectes. La nomenclature n'a pas de force obligatoire, elle est simplement indicative et un instrument pour les praticiens. Elle n'est pas non plus exhaustive, c'est-à-dire que le juge pourra décider d'indemniser un poste qui ne figure pas dans la nomenclature.

La garantie du conducteur

Aujourd'hui en France, la loi Badinter du 5 Juillet de 1985 s'applique en cas d'accident de la circulation et établit que les piétons, cyclistes et les passagers d'une voiture accidentée, à l'exception du conducteur, sont considérés comme victimes et sont donc indemnisées par l'assurance, même s'ils sont en tort.
Un projet de loi de réforme de la responsabilité civile présenté en mars 2017 par l'ex-ministre de la Justice Jean-Jacques Urvoas prévoit d'étendre le droit à l'indemnisation aux conducteurs victimes d'accident de la route impliquant un autre véhicule terrestre à moteur, sauf en cas de « faute inexcusable » ayant causé l'accident. Cette réforme pourrait engendrer un coût supplémentaire d'environ 48% [2] pour les assureurs.

1. liste des préjudices de Dintilhac en annexe
2. Source : Argus de l'assurance - RC auto : la réforme qui fait peur aux assureurs 02/07/2018

1.4.3 En Espagne

En Espagne, la loi en matière d'accident de la circulation impose depuis 1995 le recours à un barème nommé le "*Baremo de indemnizaciones en accidentes de circulación*" dit « Le Baremo ».
Une grande refonte de ce barème a été opérée par la loi 35/2015 du 22 septembre 2015, entrée en vigueur le 1er janvier 2016. Ce barème a été divisé en trois grands tableaux relatifs à l'indemnisation, en cas de :

- Décès
- Séquelles permanentes
- Blessures temporaires

Une distinction est également opérée au sein de ces trois grands tableaux entre les préjudices :

- Personnels de base
- Personnels particuliers
- Patrimoniaux

Le principe en droit espagnol est donc la barèmisation des préjudices extra-patrimoniaux (préjudices personnels) comme des préjudices patrimoniaux (préjudices économiques).

1.5 Spécificités de la vie d'un sinistre

Afin de comprendre le contexte de modélisation des sinistres qui ne sont pas encore totalement connus de l'assureur, nous définissons différents éléments.

1.5.1 Le cycle de vie du sinistre

L'année de rattachement

Elle représente l'année à laquelle est rattaché le sinistre. Cela peut être l'année de survenance, de souscription ou de réclamation.
Au Maroc, l'année de rattachement en RC Automobile est définie différemment dans la police originale selon que l'on veut définir les obligations de déclaration d'un sinistre par l'assuré ou alors la prise en charge de l'assureur.
En France, la loi exige que l'année de rattachement en assurance automobile soit l'année de survenance. En revanche, par exemple, en RC médicale, l'année de rattachement est l'année de réclamation.

L'année de développement et l'année comptable

Un sinistre n'est pas toujours payé en totalité dès la déclaration et les règlements peuvent s'écouler sur plusieurs années. Dans le cadre de la Responsabilité Civile Automobile, selon la gravité du sinistre et des conséquences corporelles de celui-ci sur les victimes, les indemnités peuvent s'écouler sur plusieurs années appelées années de développement, pour diverses raisons (attente des décisions de justice, complications médicales par exemple). L'année comptable correspond à la diagonale du triangle de liquidation et permet d'identifier les montants réglés sur une année en particulier. L'année de développement quant à elle permet d'identifier le délai avec l'année de survenance.

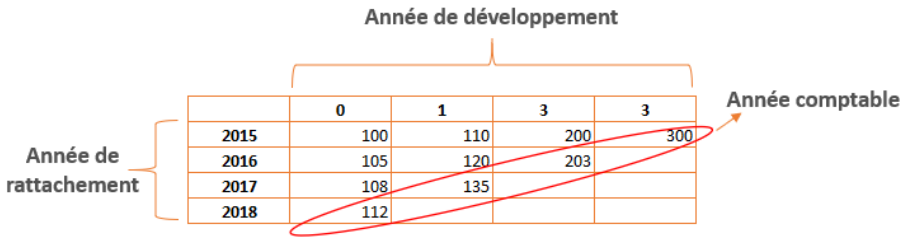

FIGURE 1.1 – Schéma récapitulatif des années de référence

Illustration du cycle de vie du sinistre

Dans le cadre du provisionnement, il est important de bien comprendre comment se dessine la vie d'un sinistre, de la date de survenance à la clôture du dossier. Le cycle de vie d'un sinistre est présenté dans la figure ci-dessous :

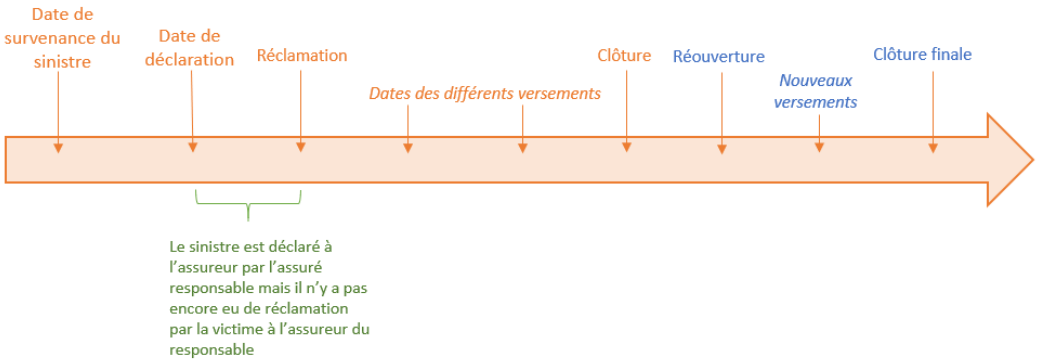

FIGURE 1.2 – Le cycle de vie d'un sinistre

Il est important de noter la différence entre le délai de déclaration et le délai de réclamation bien que ces deux dates soient souvent les mêmes. En effet, l'assuré peut avoir connaissance d'un sinistre et le déclarer bien que le tiers ne fasse pas de réclamation. Par exemple, dans le cas de la responsabilité civile médicale, le médecin peut penser être responsable d'un sinistre causé lors d'une opération (une erreur), en aviser son assureur bien que finalement, la victime (le patient) n'ait rien et donc ne déclare pas de sinistre. Ainsi, en responsabilité civile, la déclaration est faite par l'assuré alors que la réclamation est faite par le tiers.

Date et délai de réclamation

La date de réclamation correspond au moment où la victime ou son assureur font une demande en réparation - en effet, c'est le tiers impacté qui est indemnisé en Responsabilité Civile et c'est à lui que seront versées les indemnités.
Le délai de réclamation est un élément important de la modélisation des sinistres qui ne sont pas encore déclarés. Dans ces cas, il peut s'écouler plusieurs années entre la date de survenance et la date de déclaration ou de réclamation et il s'avère donc nécessaire d'étudier les sinistres de manière individuelle afin de déterminer les éléments qui caractérisent au mieux leur délai de réclamation. L'utilisation de cet élément

dans le calcul de la réserve *IBNyR* constitue la base du modèle de nombre construit dans la suite du mémoire.

Délais d'évaluation et de règlement

Une fois le sinistre déclaré, plusieurs années peuvent s'écouler avant que la totalité du montant dû par l'assureur à l'assuré soit versé. Parmi les raisons à cela, on retrouve notamment le cas où la charge varie sur plusieurs années de par le délai de consolidation (le temps que l'assuré atteigne un état de santé qui ne changera plus) et des réouvertures des sinistres. Un autre cas correspond aux liquidations longues du fait par exemple des règlements en rente.

En ce qui concerne les réouvertures de sinistres, nous n'avons pas pu trouver d'informations spécifiques sur le fonctionnement des réouvertures des dossiers au Maroc. Ayant eu accès à plus de documentation française, les commentaires qui suivent sont spécifiques au cas français. La réouverture d'un dossier en aggravation, qu'elle soit médicale ou situationnelle, est rendue possible en vertu du principe de réparation intégrale.
Le principe de la réparation intégrale sous-tend que la victime est en droit d'obtenir l'indemnisation de tous ses dommages. En conséquence, si le préjudice subi vient à être modifié dans le temps, ce principe induit que l'indemnisation soit ajustée à la nouvelle situation.
L'indemnisation initiale ou subséquente n'ayant pas un caractère libératoire, il est possible qu'elle soit suivie d'une nouvelle demande (voire même de plusieurs autres demandes étalées dans le temps) visant à réparer une aggravation de l'état de santé ou un changement dans la situation de la victime ce qui peut ainsi éventuellement conduire à une réouverture de dossier pour aggravation.
Deux catégories d'aggravations post-indemnitaires sont à distinguer :

- L'aggravation médico-légale : il s'agit de l'évolution péjorative de l'état de santé d'une victime, caractérisée par un préjudice corporel aggravé par rapport aux conclusions médicales ayant servies de base à l'évaluation définitive initiale du dommage.
  Pour donner lieu à indemnisation, ce préjudice nouveau doit être imputable à l'accident à l'origine du dommage initial et n'avoir fait l'objet d'aucune indemnisation. Ainsi sont écartées les aggravations en lien avec un état antérieur ou le vieillissement de la victime, et il appartient à l'expert médical d'être particulièrement diligent sur l'appréciation du lien de causalité. Cependant il faut noter qu'une prédisposition pathologique révélée par l'accident n'est pas de nature à écarter l'indemnisation.
- L'aggravation situationnelle : de conception récente, on vise sous cette appellation les cas où le préjudice de la victime est aggravé non en raison d'une détérioration de son état de santé mais par des modifications environnementales ou contextuelles.
  Il en est ainsi de la naissance d'un ou plusieurs enfants après transaction ou décision, de la modification de la cellule familiale (divorce, décès du conjoint ou parent faisant office de tierce personne...) ou même de l'augmentation des tarifs de l'institution où se trouve placée la victime.

1.5.2 La décomposition des sinistres

Les payés

Les payés correspondent aux sommes réglées par l'assureur dans le cadre de l'indemnisation du sinistre. Lorsque le sinistre est réglé en totalité, alors il est clos.

Les recours

Lorsqu'un assureur indemnise son client alors que l'assuré n'est pas responsable du sinistre, il a un droit de subrogation. Dans ce cas, l'assureur exerce un recours contre le responsable du sinistre à la place de son client, afin d'être lui-même indemnisé par l'assureur du responsable. *Si l'assuré A n'est pas responsable de l'accident survenu avec l'assuré B, alors l'assureur A indemnise son client A pour les dommages matériels ou corporels subis, puis exerce un recours contre l'assureur B afin d'être remboursé de l'indemnisation qu'il a effectuée en premier lieu.*

En France, d'un point de vue juridique, la convention IRCA (Convention d'Indemnisation et de Recours Corporel Automobile) instaurée le 1er avril 2002, a pour objet d'organiser les relations entre assureurs dans l'indemnisation des victimes de blessures légères consécutives à un accident de la circulation. En instaurant la désignation d'un unique interlocuteur pour instruire la procédure d'offre et en fixant les modalités de recours entre assureurs, cette convention permet de faciliter le règlement de la plus grande partie des sinistres corporels.
Il existe également la convention IRSA qui est la Convention d'Indemnisation directe de l'assuré et de Recours entre Sociétés d'assurance Automobile. Elle est signée par la plupart des sociétés d'assurance en France et est destinée à faciliter l'indemnisation des dommages matériels en cas d'accident de la circulation.
Au Maroc, il existe la convention d'indemnisation directe (CID) qui permet aux compagnies d'assurances d'indemniser directement leurs assurés sans qu'ils aient à s'adresser à l'assurance du responsable de l'accident. Cependant, cette convention ne s'applique qu'aux accidents matériels automobiles quels que soit les montants des dommages subis par le véhicule. Les responsabilités sont déterminées à partir du constat amiable ou du procès-verbal de la police et du barème de responsabilité détaillé dans la convention.

La Provision pour Sinistres à Payer

La **Provision pour Sinistres A Payer**, dite par la suite **"PSAP"** couvre les sinistres qui sont déjà survenus qu'ils soient déjà déclarés ou non. Un des enjeux majeurs des assureurs non-vie est de prédire de manière la plus juste les évolutions futures des sinistres déclarés.

La **PSAP** regroupe deux provisions :

- La provision dossier-dossier couvre les sinistres déjà déclarés mais dont le règlement n'est pas terminé en date d'inventaire. Elle est estimée au cas par cas par les gestionnaires sinistres en fonction des informations disponibles suite à la déclaration du sinistre. Elle est réévaluée et mise à jour dès lors qu'une nouvelle information est ajoutée au dossier.
- La provision *Incurred But Not Reported* **(IBNR)** couvre les variations de sinistres qui ne sont pas encore connus par l'assureur et qui peut être également décomposée en deux parties :
  - La provision *Incurred but Not **Enough** Reported (IBNeR)*
  - La provision *Incurred But Not **Yet** Reported* (IBNyR)

La somme des IBNeR et des IBNyR donne le montant de la réserve IBNR. La plupart du temps, lors de l'estimation de la réserve IBNR, les méthodes de calculs utilisées par les actuaires ne distinguent pas les deux composantes.

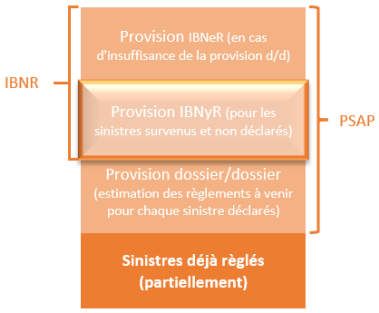

FIGURE 1.3 – Schéma récapitulatif de la PSAP en non-vie

Les sinistres IBNeR : La partie IBNeR d'un sinistre représente l'insuffisance potentielle de la provision placée initialement par le gestionnaire sinistre, c'est à dire à la déclaration puis à chaque arrêté comptable. La difficulté liée à ces *IBNER* réside dans l'estimation plus ou moins juste du montant de l'indemnisation finale attendue. En effet, nous sommes obligés d'attendre la consolidation pour connaître le coût définitif du sinistre en RC Corporelle.

Les sinistres IBNyR : Les IBNyR sont les sinistres réclamés tardivement, c'est à dire des sinistres qui n'ont pas engendré immédiatement des dommages mais qui au bout de "x" années donnent lieu à des demandes indemnisations. Ce genre de cas est assez rare en Responsabilité Civile Automobile mais n'est pas inexistant. Prenons l'exemple simple d'un piéton bousculé par une voiture. Sur le coup, ce dernier se sent bien mais commence à souffrir du dos quelques temps plus tard. Après expertises, il est établi que ces douleurs sont dues au choc de l'accident. Le fait générateur étant établi comme l'accident survenu auparavant, alors l'assurance en responsabilité civile est valide et les assureurs se doivent d'indemniser le piéton.
Une grande difficulté dans l'estimation des IBNR est la détermination de la date à laquelle les séquelles dues à l'accident seront déclarées.

Dans la pratique, il y a très peu de sinistres IBNyR en RC Automobile même si l'on admet que les conventions peuvent retarder un peu le processus. Dans le cas de dommages corporels, les sinistres sont déclarés très rapidement, surtout dans le cas où les coûts médicaux liés à la prise en charge des victimes sont importants. En revanche, ce n'est pas le cas en RC médicale car dans le cas où un chirurgien fait une erreur lors d'une opération, il peut s'écouler plusieurs années avant que le patient s'en aperçoive (un patient peut commencer à avoir des douleurs dans l'estomac quelques mois ou années après son opération suite à une erreur du médecin) et déclare un sinistre dont l'évènement déclencheur est quelques années auparavant. Par conséquent, on peut retrouver des IBNyR en RC médicale, ce qui est matière à étude mais ce n'est pas le cas en RC automobile.
Au début des travaux, nous souhaitions nous consacrer aux sinistres tardifs, mais nous nous sommes tardivement rendu compte que le sujet ne se prêtait pas bien à nos données. Ainsi comme nous le verrons par la suite, nous soupçonnons fortement que les délais de déclaration les plus élevés proviennent de réouvertures plutôt que de nouveaux sinistres qui seraient déclarés tardivement. Les travaux présentés par la suite pourront cependant s'adapter aux deux cas de figures.

La charge du sinistre

C'est la somme des règlements et de la provision pour sinistres à payer [1]. La charge sinistre comptable correspond aux règlements et aux provisions de l'exercice comptable. Comme expliqué précédemment, le cycle inversé de production de l'activité d'un assureur l'oblige à constituer des réserves afin de pouvoir payer les sinistres dont il n'a pas connaissance (ou partiellement connaissance) au moment de l'arrêté comptable. La charge d'un sinistre correspond à la somme du montant payé pour ces sinistres et des provisions techniques constituées.

1. La PSAP -Provision Pour Sinistres A Payer est définie dans la partie suivante

Chapitre 2

Méthodes de provisionnement classiques

Pour rappel, les provisions dossier-dossier sont estimées par le gestionnaire sinistre. Ainsi, les méthodologies de valorisation présentées dans cette partie concernent seulement la réserve IBNR sur la base des données avec les provisions dossier/dossier. Elles doivent permettre d'obtenir des estimations précises à l'ultime pour les sinistres déjà survenus et le degré d'incertitude qu'elles comportent doit être connu de l'assureur.
Il est important de noter que le choix de la méthode a un impact sur le résultat de l'assureur. Les hypothèses utilisées doivent être explicitées et justifiées. Ici nous présenterons d'abord trois méthodes traditionnelles, deux déterministes (Chain Ladder et Bornhuetter Fergusson) et une stochastique (Mack). Ces méthodes sont toutes trois des méthodes dites agrégées et sont très utilisées en assurance non-vie. Ensuite, nous présenterons la méthode de Schnieper, méthode qui distingue la modélisation des IBNyR et les IBNeR.

2.1 Méthode de triangulation/agrégation des données

Les méthodes de provisionnement traditionnelles se font à partir de triangles. Les triangles de nombres ou de liquidation montrent l'évolution des sinistres **cumulés** (ou décumulés) en fonction du temps. Les sinistres sont regroupés/cumulés dans un premier temps par année de rattachement puis par année de développement et cette agrégation prend la forme de triangles.

Exemple d'un triangle

Le triangle de charges prend la forme suivante :

	0	1	2	3
2017	1050	2300	2900	3110
2018	4100	6400	7710	
2019	1000	1300		
2020	500			

FIGURE 2.1 – Triangle de charge cumulé

Ainsi, par exemple, pour l'année de survenance 2017, on voit que la charge payée en premier année de développement (année 0) était à 1 050 pour passer l'année suivante à 2300 soit 1 250€ en plus.

Il nous est impossible de différencier les variations provenant de l'évolution à la hausse ou à la baisse des sinistres des variations provenant de la réclamation de nouveaux sinistres (tardifs). Cela représente bien toute la perte d'information à laquelle la création d'un triangle de liquidation donne lieu. Cette approche s'applique aussi bien pour la charge que pour les nombres. Lorsqu'un triangle de charge est agrégé, cela entraîne un niveau d'informations plus faible encore car les variations de nombres sont incluses dans les variations de la charge (on mélange les variations de nombres et de la sévérité)

2.2 La méthode la plus classique : Chain-Ladder

C'est la méthode de calcul de la provision IBNR la plus répandue sur le marché de l'assurance non-vie du fait de sa facilité d'implémentation. Elle permet de compléter le triangle (et donc de projeter les variations de sinistres à l'ultime) en utilisant des facteurs multiplicatifs, appelés facteurs de développement, qui proviennent des variations observées en j+1 versus j. La méthode est applicable à différents types de données (règlements cumulés, charges de sinistres, nombre de sinistres déclarés...).

2.2.1 Triangles de liquidation

Le principe est d'utiliser des facteurs de proportionnalité appelés facteurs de développement calculé en prenant le ratio entre le niveau de sinistralité vu au bout d'un certain temps et ce même niveau vu un an auparavant.

Année de survenance	Année de développement									
	1	2	3	...	j	...	...	...	n-1	n
1	C_{11}	C_{12}	C_{13}	...	C_{1j}	...	...	...	$C_{1,n-1}$	$C_{1,n}$
2	C_{21}	C_{22}	C_{23}	...	C_{2j}	...	...	...	$C_{2,n-1}$	$C_{2,n}$
3	C_{31}	C_{32}	C_{33}	...	C_{3j}	...	...	...	$C_{3,n-1}$	
...	...	...	...	...	...	...	...	...	...	
i	C_{i1}	C_{i2}	C_{i3}	...	C_{ij}	...	...			
...	...	...	...	...	...	...				
...	...	...	...	...	...					
...	...	...	...							
n-1	$C_{n-1,1}$	$C_{n-1,2}$								
n	C_n									

FIGURE 2.2 – Structure du triangle utilisé dans la méthode Chain Ladder

Il conviendra de noter que :
— $i \in 1, ...n$ sont les années de survenance en lignes
— $j \in 1, ...n$ sont les années de développement en colonnes
— $C_{i,j}$ sont les paiements/charges/nombres connus et cumulés pour $1 \leq i \leq n$ et $j \leq n - i$

Les facteurs de développement sont définis par :

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

On peut également calculer les facteurs de développements individuels, très importants dans la vérification des hypothèses de la méthode : $f_{ij} = \frac{C_{i,j+1}}{C_{i,j}}$

Ainsi, le triangle cumulé est complété en un rectangle complet par la formule suivante :

$$\hat{C_{i,j+1}} = C_{i,j} \prod_{j=n-i}^{n-1} f_j$$

2.2.2 Hypothèses de calcul

La méthode Chain Ladder repose sur l'hypothèse forte suivante :

$\forall j \in (1, ..., n)$, les facteurs de développement individuels f_{ij} sont indépendants de l'année de survenance i

Cette hypothèse est simplement vérifiable graphiquement : les couples $(C_{i,j}, C_{i,j+1})$ doivent être sensiblement alignés sur une droite qui passe par l'origine.

2.2.3 Résultats de la méthode

Grâce à ses facteurs, nous pouvons estimer plusieurs choses :

— Les charges / règlements / nombres ultimes (cela dépend du triangle avec lequel on travaille et sur lequel on applique la méthode) par année de survenance :

$$\hat{C_{ij}} = C_{i,n-i} \prod_{j=n-i}^{n-1} f_j$$

— Si on travaille sur des triangles de charges / règlements cumulés, alors on peut calculer les provisions par exercice de survenance :

$$\hat{R_i} = \hat{C_{i,n}} - C_{i,n-i}$$

.

— Les provisions totales sont donc estimées par

$$\hat{R} = \sum_{i=1}^{n} \hat{R_i}$$

2.2.4 Avantages et limites de la méthode

Avantages

La méthode est simple d'implémentation et offre une réelle facilité d'interprétation et de communication des résultats.
Il y aussi un avantage au niveau des données d'entrée requises. Les *inputs* sont des triangles cumulés ; ils sont faciles à obtenir et ne requièrent pas de retraitements importants. Pour finir, les résultats de Chain Ladder présentent une base solide pour l'implémentation des méthodes stochastiques (Mack (1993) et Merz et Wutrich (2008)).

Inconvénients

En pratique, les hypothèses ne sont pas toujours validées car certains coefficients ne dégagent pas de tendance linéaire. Ainsi, le principe selon lequel les sinistres survenus lors d'années différentes se développent de façon similaire n'est pas vérifié.
Aussi, la méthode ne tient pas compte de l'évolution probable de certains facteurs influençant le déroulement des règlements des sinistres. Les ajustements doivent être faits à la main par l'actuaire expert chargé des calculs et dûment justifiés par la suite. En effet, les valeurs de la dernière année de survenance du triangle découlent du produit de n-1 estimations de facteurs ; cela montre une forte sensibilité à la première valeur estimée. Il suffit que l'année la plus récente n'ait plus du tout les mêmes caractéristiques que les années d'historique pour que les prédictions soient erronées et pas du tout représentatives de l'année considérée. Aussi, les derniers coefficients de développement peuvent être très instables car ils ne sont basés que sur deux valeurs ce qui rend l'estimation très peu robuste. Cela introduit un biais qui n'existerait pas dans le cas d'une modélisation ligne à ligne. Enfin, lorsque le triangle ne permet pas de dérouler les sinistres sur une période assez longue (le triangle se déroule sur 10 ans par exemple mais les sinistres sur 15), alors ils sont parfois complétés de *tail factor* qui sont choisis par les experts.

2.3 La méthode Bornhuetter Fergusson

2.3.1 Principe

Il est en pratique possible d'avoir des informations complémentaires au triangle de liquidation comme le montant des primes par exemple, qui permettrait d'affiner l'estimation de la provision. La méthode Bornhuetter Fergusson permet d'utiliser de telles données dites "exogènes" au triangle de liquidation qui permettent de réduire l'instabilité de Chain Ladder.
Nous ne disposons malheureusement pas de l'information sur les primes dans les données étudiées, ce qui ne nous permet pas de déduire un loss ratio pour le portefeuille considéré. De la même manière, nous ne disposons pas d'une information marché qui nous permettrait d'en estimer un. Ainsi, cette méthode ne sera pas appliquée dans le cadre du mémoire mais nous présentons néanmoins son fonctionnement.

2.3.2 Hypothèses

Une propriété fondamentale de cette méthode est l'indépendance entre la provision à constituer et le dernier montant connu des règlements et permet donc de s'exonérer du lien entre la charge ultime et le dernier règlement cumulé.
Deux hypothèses doivent être vérifiées avant d'appliquer la méthode :

- Les sinistres cumulés $C_{i,j}$ sont indépendants suivant l'année de survenance i
- Il existe des paramètres $\mu_0, \mu_1, ...\mu_i$ et des cadences $\beta_0, ...\beta_i$ avec $\beta_j = 1$ tels que :
  $\forall 0 \leq i \leq I$ et $0 \leq j \leq J-1$ et $0 \leq k \leq J-j$ on a :

$$\mathrm{E}(C_{i,0}) = \beta_0 \mu_i$$

et

$$\mathrm{E}(C_{i,j+k} | C_{i,0}, ...C_{i,j}) = C_{i,j} + (\beta_{j+k} - \beta_j)\mu_i$$

2.3.3 Calcul

Soit P_i la prime de l'année i, alors est un estimateur du ratio de perte dit *loss ratio* ultime pour l'année i.
Soit U_i la charge ultime de l'année i, nous avons donc :

$$\hat{U}_i = LR \times P_i$$

L'estimateur $\hat{C_{i,J}}^{BF}$ de la charge ultime par cette méthode pour l'année de survenance i est :

$$\hat{C_{i,J}}^{BF} = C_{i,I-i} + (1 - \hat{z}_{I+1-i})U_i$$

où $\hat{z_j}$ la cadence de règlement cumulée après j années de développement, souvent issue de la méthode de Chain Ladder.
Ainsi, l'estimation de la charge ultime correspond à une pondération entre une estimation Chain Ladder et la charge de sinistres ultime supposée connue.
On peut ainsi déduire $R_i^{\hat{B}F}$ le montant de la provision à constituer pour l'année de survenance i tel que

$$(1 - \hat{z}_{I+1-i})U_i$$

L'estimateur de la provision globale est

$$R^{BF} = \hat{\sum_{i=1}^{I}} R_i^{\hat{B}F}$$

Il faudra cependant noter que l'actuaire apporte son expérience dans l'estimation de la charge ultime a priori ce qui rend l'application de cette méthode spécifique à chacun. La provision estimée paraît plus adéquate que Chain Ladder car elle peut être propre à la branche étudiée et indépendante de la dernière charge connue (contrairement à Chain Ladder qui est sensible à ce facteur).

2.4 La méthode stochastique de Mack

La méthode de Mack, introduite en 1993, vise à estimer la volatilité totale à l'ultime sur les montants de provision. Elle fait intervenir la notion d'incertitude dans la méthode de Chain Ladder. Elle permet de donner une marge d'erreur sur le montant des provisions et est également connue sous le nom de *distribution free Chain Ladder* car aucune hypothèse n'est faite concernant la distribution des triangles.

2.4.1 Hypothèses de calcul

La méthode s'appuie sur trois hypothèses :

— H1 : Les années de survenance des sinistres sont indépendantes les unes des autres, c'est à dire C_{ij} et C_{kj} sont indépendants si $i \neq k$

— H2 : Le passage d'une année de développement à l'autre est décrit en termes d'espérance.
$\forall i = 1...., n, \forall j = 1, ..., n-1$ on a $\mathrm{E}(C_{i,j+1} \,|C_{i1}, ..., C_{ij}) = \lambda_j C_{ij}$ où $\lambda_j = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}}$

— H3 : $\forall i = 1....,n, \forall j = 1,...,n-1$, il existe un paramètre σ_j tel que :
$\text{Var}(C_{ij+1}|C_{i1},....,C_{ij}) = \sigma_j^2 C_{ij}$.

Ces hypothèses peuvent s'écrire sous la forme suivante :

$$C_{i,j+1} = \lambda_j C_{i,j} + \epsilon_{i,j}\sqrt{C_{i,j}}$$

Cette formule permet d'estimer les λ_j pour tous les j par l'estimateur des moindres carrés. Sous les hypothèses H1 et H2 ; les estimateurs des moindres carrés (sans biais et non corrélés) sont obtenus par :

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

De la manière, les estimateurs des paramètres $\hat{\sigma_j}$ sont estimés par :

$$\hat{\sigma_j} = \frac{1}{n-j}\sum_{i=1}^{n-j}\frac{C_{i,j+1}\lambda_j\hat{C}_{i,j}}{\sqrt{C_{i,j}}}$$

Les coefficients de développement estimés par la méthode de Mack coïncident avec ceux estimés par la méthode de Chain Ladder, par conséquent les provisions estimées par les deux méthodes sont égales.

La méthode permet alors de quantifier l'erreur de prédiction. L'erreur quadratique moyenne de prévision (MSEP) des provisions représente l'écart moyen entre les valeurs estimées et les valeurs observées réellement. Pour chaque année de survenance i, l'estimateur MSEP($\hat{R}_i$) est donné par :

$$\hat{msep}(\hat{R}_i) = \hat{C_{i,n}^2}\sum_{j=n-i}^{n-1}\frac{\sigma_j^2}{\hat{\lambda_j^2}}(\frac{1}{\hat{C_{i,j}}} + \frac{1}{\sum_{i=1}^{n+1} C_{i,j}})$$

La MSEP du montant total des provisions pour l'ensemble du portefeuille est donné par :

$$\hat{msep}(\hat{R}) = \sum_{i=1}^{n}(\hat{msep}(\hat{R}_i) + 2\hat{C_{i,n}^2}\sum_{i=i+1}^{n}\hat{C_{i,j}}\sum_{j=n+1}^{n-1}\frac{\hat{\sigma_j^2}}{\hat{\lambda_j^2}\sum_{i=1}^{n+1-j} C_{i,j}})$$

Pour définir des intervalles de confiance autour des $\hat{R}_i$, ous prenons l'hypothèse que les provisions suivent une loi normale de moyenne $\hat{R}_i$ et d'écart-type $\hat{sep}(\hat{R}_i)$ où

$$\hat{sep}(\hat{R}_i) = \sqrt{\hat{msep}(\hat{R}_i)}$$

. L'intervalle de confiance à 95% est alors donné par

$$[\hat{R}_i \pm 1.96\hat{sep}(\hat{R}_i)]$$

2.4.2 Validation du modèle

Il est nécessaire de valider les trois hypothèses pour pouvoir appliquer les formules.

— H1 : elle peut être validée de manière graphique, de la même manière que l'hypothèse de Chain Ladder. Pour toute année de développement j fixée, les couples $(C_{ij}, C_{ij+1})_{i=1,...,n-j}$ doivent

sensiblement faire apparaître une droite passant par l'origine.
— H2 : cette deuxième hypothèse ne peut être validée par de la modélisation, et la méthode est considérée acceptable si les facteurs de développement individuels sont relativement constants.
— H3 : cette hypothèse peut également être validée graphiquement, en s'assurant que les résidus $\frac{C_{ij+1}-\hat{f}_j C_{ij}}{\sqrt{C_{ij}}}$ ne présentent pas de tendance.

2.4.3 Avantages et inconvénients de la méthode

Avantages

La méthode permet de fournir une estimation de l'erreur de prédiction. Il donne une formule fermée qui calcule de manière explicite la variance du montant des provisions qui permet de mesurer l'incertitude de la charge ultime.

Inconvénients

La méthode est excessivement sensible aux facteurs de développement. Dans le cas d'une année atypique, une seule irrégularité dans les facteurs peut perturber l'estimation de la volatilité. De plus, elle donne l'erreur de prédiction mais ne permet pas d'obtenir la distribution de la provision. L'assureur peut alors estimer le niveau de prudence de la provision à l'aide de Mack uniquement en fixant la loi de distribution par lui-même.

2.5 La méthode de Schnieper

2.5.1 Notations

— Soit $C_{i,j}$ est le montant cumulé des sinistres pour chaque année de rattachement i et de développement j.
— Soit $N_{i,j}$ le montant des nouveaux sinistres, qu'on peut assimiler aux IBNyR.
— Soit $D_{i,j}$ le montant cumulé agrégé des variations des sinistres observées entres les années de développement $j-1$ et j pour chaque année de rattachement i. Cette composante correspond aux IBNeR.
— Soit E_i l'exposition pour chaque année de rattachement i

Les relations suivantes sont supposées vraies dans la méthode $\forall i = (0, ..., n) \forall j = (0, ...n)$:

$$N_{i,0} = X_{i,0}$$

$$X_{i,j} = X_{i,j-1} - D_{i,j} + N_{i,j}$$

Les tableaux de chacun des trois éléments $X_{i,j}$, $N_{i,j}$, $D_{i,j}$ sont triangulaires, pour $i + j \leq n$
Ainsi, à partir des données individuelles, il est possible de reconstituer les trois triangles.

2.5.2 Hypothèses de la méthode

Soit H_k l'ensemble des variables $N_{i,j}$ et $D_{i,j}$ observées jusqu'à l'année comptable k.
$H_k = \{N_{i,j}; D_{i,j} | i + j \leq k + 1\}$ et
H_0 est l'ensemble des variables connues pour l'instant présent et

H_{i-j-2} est l'ensemble des variables existant jusqu'à l'émergence des variables $N_{i,j}$ et $D_{i,j}$.
Les hypothèses sont les suivantes :

- (H1) : $\mathrm{E}(N_{i,j}|H_{i+j-2}) = E_i * \lambda_j$: les sinistres non déclarés ne dépendent pas du passé, ils sont le produit de l'exposition pour chaque année de rattachement avec un facteur de développement.
- (H2) : $\mathrm{E}(D_{i,j}|H_{i+j-2}) = X_{i,j-1} * \delta_j$
  Les variations d'IBNeR sont le produit des sinistres reportés des années de développement passés avec un facteur de développement. Si l'information sur le statut du sinistre (ouvert ou clos) est disponible, alors il est préférable de remplacer les $X_{i,j}$ dans H2 par les montants de sinistres ouverts seulement.
- (H3) : Les ensembles de variables $\{N_{1,j}; D_{1,j}|j = 1, 2..., n\}$,
  $\{N_{2,j}; D_{2,j}|j = 1, 2..., n\}$,...,$\{N_{n,j}; D_{n,j}|j = 1, 2..., n\}$ sont indépendants (les années de survenance sont indépendantes).

(H1) modélise l'hypothèse que les montants des nouveaux sinistres dépendant uniquement de l'indicateur mesurant le volume d'exposition E_i et d'un facteur λ_j dépendant de l'année d'observation j.
(H2) modélise le fait que les variations des montants des sinistres dépendent uniquement des montants précédents des sinistres et d'un facteur δ_j. En effet, la part de variation d'un siniste dépend pour une grande partie de la qualité de ses estimations antérieures et notamment de la dernière. Ainsi plus on dispose d'années de développement pour un sinistre, plus l'estimation du coût s'affine.
(H3) suppose enfin que la variation des montants des sinistres et la survenance de nouveaux sinistres pour une année de rattachement donnée sont indépendantes des autres années.

2.5.3 Estimations des paramètres

A partir des hypothèses (A1), (A2), et (A3), il est possible de déterminer des estimateurs sans biais de paramètres λ et δ pour $j = 1, ..., n$.
Soit $\hat{\lambda}$ et $\hat{\delta}$ pour $j = 1, .., n$, les estimateurs de λ_j et δ_j dont les formules sont les suivantes :

$$\hat{\lambda_j} = \frac{\sum_{i=1}^{n+1-j} N_{i,j}}{\sum_{i=1}^{n+1-j} E_i}$$

$$\hat{\delta_j} = \frac{\sum_{i=1}^{n+1-j} D_{i,j}}{\sum_{i=1}^{n+1-j} X_{i,j-1}}$$

Grâce à ses estimateurs et aux trois hypothèses H1, H2 et H3, il est possible de déterminer des estimateurs pour $X_{i,j}$, Di, j et $N_{i,j}$ quand $i + j \geq n + 1$. Ainsi, on obtient :

- $\hat{N_{i,j}} = E_i * \hat{\lambda_j}$ $\forall$i+j $\geq n + 1 - i$
- $\hat{D_{i,j}} = X_{i,j-1} * \hat{\delta_j}$ $\forall$i+j $\geq n + 1 - i$
- $\hat{X_{i,j}} = \hat{X_{i,j-1}} * (1 - \hat{\delta_j}) + Ei * \hat{\lambda_j}$

Ainsi en procédant parallèlement à la diagonale de chaque triangle, il est possible de compléter les triangles des $N_{i,j}$,Di, j et Xi, j

2.5.4 Avantages et inconvénients

La méthode de Schnieper est simple d'implémentation, logique et peut être appliqué au niveau agrégé ou individuel. Elle permet de modéliser séparément les sinistres tardifs de l'évolution des sinistres connus (IBNyR des IBNeR). Cependant, elle est sensible à l'insuffisance des données comme on peut retrouver par ailleurs dans la méthode de Chain Ladder.

2.6 Approche proposée par Pietro Parodi

Dans la continuité de Schnieper, l'approche proposée par Pietro Parodi soutient que les méthodes de provisionnement basées sur des triangles sont inadéquates pour modéliser avec précision la distribution des réserves, bien qu'elles puissent être suffisamment bonnes pour produire une estimation ponctuelle de ces dernières. P.Parodi explique que la représentation en triangle implique une perte d'information cruciale sur les sinistres individuels, ce qui rend difficile l'obtention d'une distribution pour les IBNR.
Dans son étude, l'auteur met en oeuvre une approche pour éviter une telle perte d'informations. Il explique qu'il est nécessaire d'adopter une approche similaire à celle utilisée pour la tarification, où un modèle de fréquence et de gravité distinct sont élaborés puis combinés par simulation de Monte Carlo (ou d'autres techniques numériques) qui permettent ensuite de produire la distribution des pertes / IBNR.
Dans un premier temps, un modèle de fréquence pour le nombre de sinistres déjà survenus (mais non déclarés) est produit grâce à la distribution des délais de déclaration puis dans un second temps, c'est un modèle de sévérité qui est produit pour les sinistres individuels (en calibrant une loi sur les charges déjà connues). Les deux lois (fréquence et sévérité) calibrées sont ensuite combinées avec la méthode de Monte Carlo, ce qui entraîne l'obtention d'une distribution pour les IBNR.
Le résumé de l'approche proposée est la suivante :

- Calibrage d'une loi pour les délais de déclaration basé sur les délais empiriques.
- Utilisation de cette loi pour estimer le nombre de sinistres IBNR (à partir des sinistres déclarés à la date d'observation par année de survenance).
- Calibrage de la meilleure loi pour les nombres de sinistres IBNR (Poisson, Négative Binomiale).
- Modélisation de la loi de sévérité pour les sinistres IBNR
- Combinaison de la distribution des nombres et de la distribution des montants des IBNR via une simulation Monte Carlo pour produire une distribution pour les sinistres IBNR.

Par conséquent, l'approche proposée tente d'illustrer comment une méthode non agrégée ou individuelle peut être utilisée en pratique et montre qu'exploiter les données individuelles d'un portefeuille peut potentiellement entraîner une meilleure estimation des réserves. En effet, dans une première étape de validation de la méthode de calcul présentée dans l'étude et dans le cadre d'une comparaison avec les méthodes existantes, une série d'expériences avec des ensembles de données artificielles ont été entreprises, et dont les résultats montrent une réduction drastique de l'erreur de prévision du nombre de sinistres IBNR et du montant total IBNR par rapport à la méthode standard Chain Ladder. Un élément prometteur également provient du fait que les expériences montrent que la distribution des réserves IBNR est beaucoup plus proche (en termes de distance Kolmogorov-Smirnov) de la "vraie" par rapport à celle basée sur la méthode de Mack. La méthode promet donc une évaluation plus précise de l'incertitude autour des réserves et c'est ce que nous tentons d'appliquer dans la suite de ce mémoire.

2.7 Inconvénients des méthodes agrégées, intérêt des méthodes individuelles et présentation de la problématique

En agrégeant les données, des informations précieuses sur les caractéristiques individuelles des sinistres sont écartées qui pourraient pourtant améliorer la prédiction. En effet, les techniques traditionnelles ont été développées avant l'essor de l'informatique et par conséquent à l'époque où les ordinateurs n'avaient pas la grande puissance de calcul que nous avons aujourd'hui.
A cause du faible nombre d'observations dans un triangle agrégé, ces méthodes ne prennent pas en compte l'information disponible sur le contrat d'assurance, s'avèrent très sensibles aux valeurs extrêmes, ne permettent pas de détecter les potentielles différences individuelles et enfin sont susceptibles de créer une sur-paramétrisation.

Ces différents inconvénients ont motivé le développement du provisionnement individuel. Les travaux présentés dans la suite sont basés sur l'argument que les méthodes agrégées de provisionnement, aussi sophistiquées qu'elles soient, ne permettent pas d'obtenir de manière précise une estimation de la réserve au titre des sinistres non encore déclarés. La triangulation des données mène à une perte énorme d'informations propres aux sinistres individuels qui ensuite empêche d'extraire des informations détaillées sur la distribution des *IBNyR*.

Dans le cadre de ce mémoire, nous proposons donc une approche hybride à la constitution de la réserve IBNR. Nous allons distinguer les deux composantes de la réserve dans les calculs afin d'améliorer les projections effectuées pour les sinistres qui sont encore inconnus de l'assureur :

Pour la réserve IBNyR, nous proposons d'implémenter une approche fréquence / coût qui permettra de prendre en compte un phénomène clé influençant la modélisation des nombres de sinistres tardifs : les délais de déclaration. En effet, le nombres de sinistres qui sont portés à l'attention de l'assureur dépendent du temps qui s'est écoulé depuis leur survenance.
Par conséquent, avoir une vision du temps écoulé entre la survenance et la déclaration pour les sinistres du portefeuille peut éventuellement amener à améliorer les projections des sinistres tardifs et ainsi d'améliorer la pertinence des résultats par rapport à des approches agrégées qui ne prennent pas en compte cette variable temps qui semble pourtant importante. Pour cette partie, nous allons ajuster des lois statistiques connues aux délais de déclaration afin d'estimer le nombre ultime de sinistres tardifs puis faire de même pour les charges pour obtenir une loi représentative de la charge placée pour les sinistres du portefeuille. En effectuant de nombreuses simulations tirées de ces deux lois, il nous sera ensuite possible d'obtenir la distribution de la réserve *IBNyR* et son montant.

Pour la partie IBNeR, nous conserverons une approche déterministe, Chain Ladder car les caractéristiques des sinistres sont déjà connues, ce qui introduit beaucoup moins de biais lors des projections. Nous jugeons pertinent d'utiliser la vision du déroulement des charges placées pour les sinistres déjà déclarés et de continuer à développer la même cadence sur le futur.
Les variables utilisées lors de la modélisation de chaque partie ne sont pas les mêmes car dans le cas des sinistres tardifs, nous n'avons aucune information sur les sinistres car ils ne sont pas déclarés. Par conséquent la volatilité inhérente au nombre et au coût de ces sinistres est bien plus importante que celle de la composante IBNeR. Le risque d'introduire un biais dans les projections est beaucoup plus élevé que lors des estimations liées aux sinistres en cours où toutes les informations récoltées sur les charges depuis

la déclaration sont prises en compte. C'est d'ailleurs pour cela qu'une approche stochastique est préférée pour la partie *IBNyR* alors que nous conservons une approche déterministe pour les *IBNeR*.

Il est important de noter que la méthode présentée dans ce mémoire est plus coûteuse en temps de calcul. Il sera alors intéressant de bien comprendre les avantages et les inconvénients qu'elle apporte afin de pouvoir conclure objectivement si elle présente une réelle piste d'amélioration au provisionnement, le but n'étant pas de remplacer les méthodes traditionnelles mais plutôt de les améliorer en y apportant de l'information supplémentaire.

Chapitre 3

Analyse exploratoire des données

3.1 Présentation de la base

Les données présentées ci-dessous ont été modifiées pour l'aboutissement du mémoire. Elles ne représentent pas les données réelles.
La base de données utilisée pour la modélisation est une base sinistres pour la garantie Responsabilité Civile Corporelle Automobile au Maroc. Elle présente 41 024 sinistres survenus entre 2009 et 2018 vus au 31/12/2018. Chaque ligne représente une date de vue du sinistre depuis sa survenance jusqu'au 31/12/2018. Nous ne connaissons pas le traitement des recours estimés dans les provisions (on ne sait pas si les recours sont provisionnés ou non), élément important en théorie car cela permet de savoir si les bases de données de différents assureurs sont comparables ou non. Les variables de la base sont les suivantes :

— La date de survenance du sinistre
— La date de déclaration du sinistre
— Les dates de début et fin d'exposition au risque
— Le sexe de l'assuré
— L'âge de l'assuré
— L'ancienneté du permis du conduire
— L'ancienneté du véhicule
— Le règlement effectué (en euros) au titre du sinistre à chaque année de vue
— Les recours encaissés (en euros)
— Le montant de la réserve à chaque date de vue (en euros)

3.2 Retraitements effectués

Afin d'avoir une vision plus claire des données, les retraitements suivants ont été effectués :

Création des différents triangles

Les triangles suivants ont été créés afin d'analyser les sinistres du portefeuille :
— Le triangle de payés
— Le triangle de provisions
— Le triangle de recours
— Le triangle de charge totale (règlements sur l'année considérée + réserve - recours)

— Le triangle des nombres (important dans notre modélisation de tardifs)

Récupération du statut des sinistres

Ensuite, le statut du sinistre à chaque date de vue a été estimé selon les hypothèses suivantes :
— Un sinistre est **sans suite** si sa charge vue au 31/12/2018 est à 0.
— Un sinistre est **clos** si sa provision vue au 31/12/2018 est à 0.
— Un sinistre est **en cours** sinon

Délai entre la survenance et la déclaration / réclamation

Les délais de déclaration en années et en jours ont été récupérés en prenant la différence entre les dates de déclaration et les dates de survenance.

Délai entre la déclaration/réclamation et la clôture

Une fois que le statut des sinistres est connu pour chaque date de vue, il est intéressant de calculer le temps écoulé entre la déclaration à l'assureur et la clôture du dossier. Pour cela, la première année de vue lors de laquelle le sinistre est indiqué comme clos est assumé comme étant l'année de clôture. Suite à ça, le nombre d'années entre la déclaration et la date de clôture défini bien le délai de règlement du sinistre. Si le sinistre n'est pas clos, alors le délai de règlement est marqué comme "N/A".

Dates manquantes en termes de survenance et déclaration/réclamation

Pour les dates manquantes identifiées dans la base de données, nous avons pris l'hypothèse de les remplacer par le 30 juin de l'année considérée car en théorie en automobile, les sinistres arrivent de manière homogène sur l'année.

Récupération des données individuelles

Dans la base, un sinistre apparaît sur plusieurs lignes car il y a une vue chaque année, entre la déclaration et le 31/12/2018. Ils sont observés plusieurs fois, à la fin de chaque année qui s'écoule entre la déclaration et le 31/12/2018. Bien que ces observations soient très pertinentes dans le cadre du provisionnement, il est aussi nécessaire d'avoir une vision des sinistres de manière individuelle, c'est à dire d'avoir une ligne pour chaque sinistre où l'on y retrouve toutes ses caractéristiques. Cette base permet également de prédire les nombres. Si l'on se base sur les vues pour prédire les nombres, alors cela biaise les estimations car on prendra en compte plusieurs fois le même sinistre.

3.3 Analyses graphiques

Pour information, les assurés du portefeuille ont en moyenne 49 ans avec une ancienneté de permis d'environ 21 ans. Les véhicules assurés dans le portefeuille ne sont pas très récents, avec une ancienneté de 8,63 ans en moyenne (ce ne sont pas des informations que nous pourrons utiliser par la suite dans nos modèles).

	Age de l'assuré	Ancienneté du permis de condure	Ancienneté du véhicule
Moyenne	49,10	20,86	8,63

FIGURE 3.1 – Caractéristiques des assurés et des véhicules

Les sinistres ont été regroupés par année de survenance afin d'avoir une vision plus précise de la façon dont ils sont répartis sur l'historique des données.
Le nombre de sinistres par année de survenance à l'exception de 2018 est relativement stable. Nous ne disposons malheureusement pas des mesures d'exposition tels que le volume de primes ou le nombre de véhicules. Il est donc difficile d'interpréter la baisse en nombre entre 2009 et la période 2010-2017. En plus d'avoir peu de recul sur l'année 2018, nous n'avons pas d'autres mesures pour expliquer la forte baisse des nombres de sinistres (de 4000 à 2000). Le mix du portefeuille n'étant pas disponible, il a été difficile de savoir si les variations en nombres proviennent d'un changement de profil spécifique des contrats souscrits.

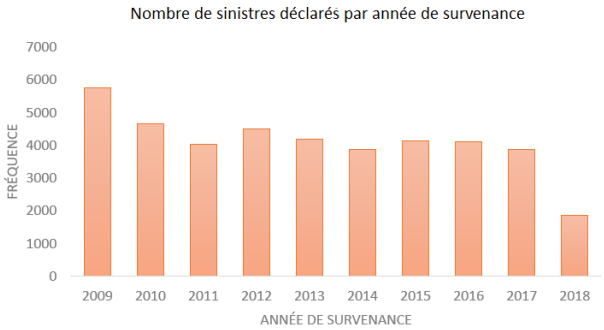

FIGURE 3.2 – Nombre de sinistres par année de survenance

La majorité des sinistres est déclarée dans les deux ans après la survenance. 75% des sinistres sont déclarés lors de leur année de survenance et 21% en n+1, ce qui est normal compte tenu des survenances potentielles en fin d'année (décalage de saisie / gestion). Enfin, 2% en n+2 et la proportion de sinistres déclarés au-delà de trois années après la survenance est extrêmement faible. D'ailleurs, en assurance automobile en France, le délai de prescription est de 2 ans à partir de l'événement pour l'assuré et de 10 ans pour les tiers. Nous ne disposons pas de ces informations pour le Maroc.

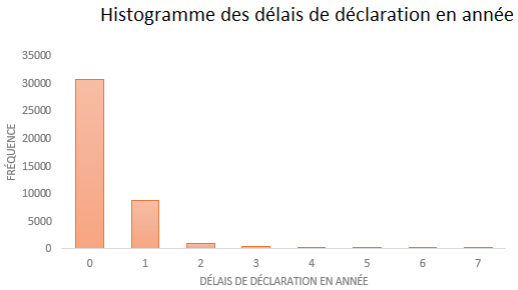

FIGURE 3.3 – Histogramme des délais de déclaration en années

On observe une médiane à 39 jours soit 1,3 mois et un maximum à 2802 jours soit 7 ans. Une telle valeur pour le maximum par rapport à la médiane confirme le fait qu'au-delà de deux ans il ne peut s'agir que de réouvertures de sinistres.

Minimum	Maximum	Mediane
0 jours / 0 mois	2802 jours / 7 ans	39 jours / 1,3 mois

FIGURE 3.4 – Statistiques descriptives des délais de déclaration en jours

Etant donné que l'information sur les délais de déclaration forme un élément important du mémoire, le graphique ci-dessous permet de les visualiser de manière plus détaillée, en particulier car nous avons à disposition l'information en jours pour chaque année de survenance.
Le délai moyen observé est de 112 jours et nous observés trois périodes bien distinctes :

- Entre 2009 et 2011, les délais de déclaration est autour de 130 jours.
- Entre 2012 et 2017, le délai reste stable autour de 100 jours. Ici, il est possible qu'il y ai eu un changement dans les process de gestion permettant un enregistrement plus rapide des sinistres.
- En 2018, le délai diminue fortement autour de 60 jours, ce qui s'explique par les données arrêtées au 31/12/2018, la période d'observation ne permettant pas de voir les délais de déclarations les plus longs pour les sinistres survenus durant le second semestre.

Il est difficile d'expliquer les raisons à ces variations de délais selon les périodes car nous n'avons pas accès à la politique de gestion appliquée par la compagnie. Nous tirons cependant la conclusion que le délai de 112 jours soit 4 mois sur la période d'historique complète 2009-2017 paraît relativement long pour des sinistres de nature corporels.

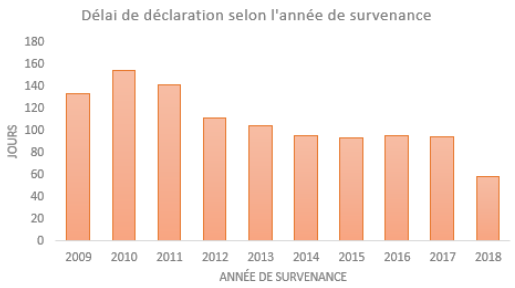

FIGURE 3.5 – Délais de déclaration moyens en jours par année de survenance

Lors du prétraitement de la base, la variable statut du sinistre a pu être reconstituée pour chaque date de vue. Ainsi, il est possible de voir le changement de statut des sinistres réparti selon leur année de survenance entre le moment où ils ont été déclarés et à la fin de la période d'observation présentée par la base, le 31/12/2018.
Ici nous pouvons voir que plus l'on recule dans le temps (exercices de survenance les plus anciens), plus la proportion de sinistres clos augmente, les exercices de survenance les plus récents (de 2016 à 2018) ont quant à eux un pourcentage de dossiers clos très faible, ce qui paraît plutôt logique. De plus, en 2009, nous avons encore 0.5% de sinistres qui ne sont pas encore clos après 10 années de développement, ce qui démontre le besoin de considérer une *tail* (un facteur de queue) permettant d'aller au-delà des 10 années de développement observées.

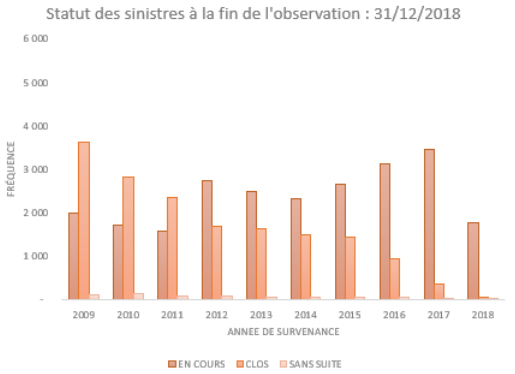

FIGURE 3.6 – Statut des sinistres à la survenance puis à la fin de la période d'observation

Concernant les charges moyennes par sinistres, nous pouvons voir dans le graphique suivant que :

- Les montants moyens à l'ouverture de 2009 à 2010 sont relativement atypiques, tout comme 2011 qui paraît exceptionnellement élevé
- De 2012 à 2014, les montants moyens à l'ouverture sont assez cohérents
- On observe ensuite une charge moyenne assez constante pour 2015-2018

Les montants à l'ouverture relèvent en général de la politique de gestion des sinistres des sociétés et sont souvent des montants forfaitaires revus par la suite dès que les éléments du dossier ont été récupérés.

Par ailleurs, on peut également observer que les charges moyennes vues sur la dernière diagonale sont systématiquement inférieures aux montants retenus à l'ouverture. On observerait donc plutôt une ten-

dance au sur provisionnement. Cependant, il ne faut pas oublier que les sinistres graves peuvent mettre un certain nombre d'années à se clôturer : ainsi, en France notamment, on doit en attendre que les victimes mineures aient atteint leur majorité pour évaluer la totalité des dommages. Le delta entre charge moyenne à l'ouverture et sur la dernière diagonale se réduit au fur et à mesure des années de développements, ce qui est logique car moins les années sont développées, plus les sinistres sont évalués à leur montant à l'ouverture.

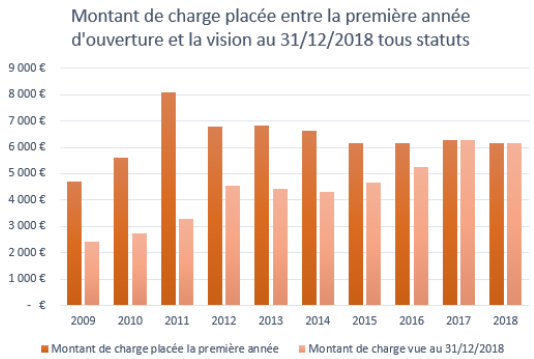

FIGURE 3.7 – Charge moyenne placée la première année d'ouverture et au 31/12/2018 par année de survenance pour tous les sinistres

Dans le graphique suivant qui se concentre sur les sinistres clos, on voit que les montants moyens à l'ouverture (histogramme orange foncé) sont faibles comparés aux montants observés en dernière diagonale (histogramme orange clair), ce qui est normal car les sinistres qui sont clos rapidement sont les sinistres "peu graves" ou "peu coûteux". Concernant les sinistres moyens au 31/12/2018, les survenances 2015, 2016 et 2018 sont en décroissance par rapport aux autres, ce qui est logique car ce sont encore des années peu développées et donc ce sont les sinistres les moins graves qui sont clôturés L'analyse des sinistres clos au 31.12.2018 sur les exercices 2009-2014 ne montre pas d'inflation significative. Cela est peu intuitif en RC Automobile Corporelle car au moins en France les sinistres corporels ont eu des inflations importantes au moins pour les graves.

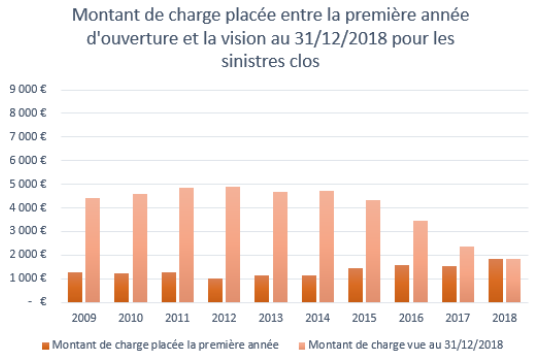

FIGURE 3.8 – Charge moyenne placée la première année d'ouverture et au 31/12/2018 par année de survenance pour les sinistres clos

Pour résumer rapidement, les données laissent penser qu'il pourrait y avoir une *tail* et l'inflation des sinistres semble marginale.

Chapitre 4

Résultats des méthodes de provisionnement classiques

4.1 Modélisation par Chain Ladder de la réserve IBNR

Dans cette partie, la réserve IBNR est calculée par la méthode de Chain Ladder. La méthode Chain Ladder ne distingue pas les *IBNyR* et les *IBNeR*. Tous les montants connus à la date d'évaluation sont projetés sans distinguer ceux qui sont liés à des sinistres qui sont connus mais qui ne sont pas encore réglés en totalité de ceux qui sont liés à des sinistres non encore connus. Ainsi, c'est en étudiant 'évolution des sinistres sur tout l'historique que nous serons capables de calculer les sinistres à l'ultime et ainsi de déterminer le montant de la réserve IBNR à placer (la réserve IBNR est calculée en prenant la différence entre l'ultime projeté et les sinistres observés sur la dernière diagonale).

Le triangle de charges est le suivant :

	Année de développment									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	14 485 401	27 810 864	34 679 862	38 534 466	41 810 003	43 990 855	45 087 721	44 859 167	45 332 597	45 195 672
2010	13 135 693	26 737 946	31 364 706	35 285 311	37 845 715	39 372 862	40 486 855	40 016 346	40 109 641	
2011	14 146 964	27 141 883	31 681 244	35 114 914	37 157 607	38 619 399	38 879 888	39 334 870		
2012	16 011 252	28 414 095	33 816 625	37 657 678	40 311 220	41 871 818	42 881 945			
2013	12 888 928	24 768 498	30 165 924	34 341 997	36 759 707	38 144 421				
2014	12 041 707	22 220 071	28 385 083	32 093 163	34 221 616					
2015	12 110 485	24 864 658	30 985 738	34 967 113						
2016	11 351 040	24 971 808	31 283 840							
2017	12 101 769	27 326 301								
2018	11 448 098									

FIGURE 4.1 – Triangle de charges cumulé

4.1.1 Vérification des hypothèses

Après avoir vérifié que les hypothèses de Chain Ladder sont bien respectées, nous serons en mesure d'appliquer la méthode Chain Ladder sur les données du portefeuille considéré. Pour j fixé, nous avons supposé l'existence d'un coefficient f_j $C_{ij+1} = f_j C_{ij}$ où les f_{ij} sont indépendants de l'année de survenance i. Pour vérifier cela, les couples (C_{ij}, C_{ij+1}) doivent être sensiblement alignés sur une droite passant par l'origine, ce qui reflète la stabilité des rapports entre les montants.

Les graphiques montrent bien que les hypothèses de Chain Ladder sont plutôt respectées sur les années de développement observées car on observe une tendance linéaire quasi partout, excepté pour la première année où il y a une instabilité assez importante des coefficients individuels. Le premier facteur de développement est généralement ainsi du fait des provisionnement forfaitaire le temps de réunir des informations sur les sinistres et les réserver au mieux.
Avec Chain Ladder, l'hypothèse d'indépendance est assez restrictive et difficile à valider dans de nombreux cas. Le passé doit être suffisamment régulier et la branche doit être peu volatile. Il est compliqué de traiter des sinistres graves et ponctuels qui casseraient rapidement la régularité de l'historique.

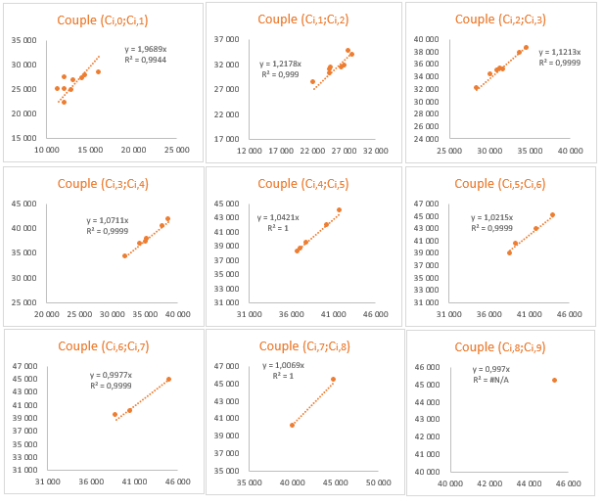

FIGURE 4.2 – Vérification de l'hypothèse de linéarité des couples

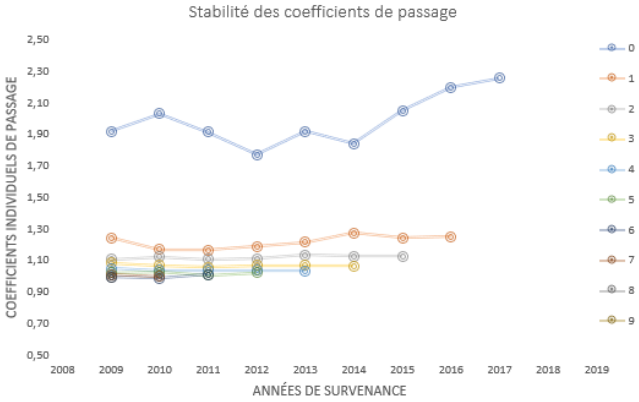

FIGURE 4.3 – Etude de la linéarité des coefficients de passage individuels par année de développement

Les facteurs de développement individuels sont obtenus de la manière suivante : $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$.
On obtient ainsi le triangle de facteurs de développement ci - dessous.

— Pour les facteurs de développement individuels entre les années 0 et 1, nous observons deux périodes en termes d'évolution des coefficients individuels, d'une part 2009-2014 qui montrent

une volatilité autour d'une moyenne et d'autre part 2015-2018 qui semble montrer une tendance à la hausse. Le coefficient Chain Ladder retenu ne permet pas de voir cette information. Selon les informations fournies par les experts, il serait probable que l'on retienne plutôt une projection de la tendance observée sur 2015-2018 plutôt qu'une valeur moyenne. Nous avons conservé la valeur de Chain Ladder n'ayant pas accès à des experts.

— Pour les facteurs de développement individuels entre les années 1 et 2, nous observons également une rupture potentielle entre 2013 et 2014 mais bien moins importante que pour les années de développement précédentes. Nous conserverons également l'estimateur donné par Chain Ladder.
— Le reste des données observées est relativement stables et permettent de conserver les valeurs calculées par Chain Ladder.
— Compte tenu des valeurs observées très proches de 1, voire légèrement inférieures pour les années de développement les plus récentes, nous décidons dans ce jeu d'hypothèses de considérer un facteur de queue (*tail factor*) à 1.

	Années de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009		1,92	1,25	1,11	1,09	1,05	1,02	0,99	1,01	1,00
2010		2,04	1,17	1,13	1,07	1,04	1,03	0,99	1,00	
2011		1,92	1,17	1,11	1,06	1,04	1,01	1,01		
2012		1,77	1,19	1,11	1,07	1,04	1,02			
2013		1,92	1,22	1,14	1,07	1,04				
2014		1,85	1,28	1,13	1,07					
2015		2,05	1,25	1,13						
2016		2,20	1,25							
2017		2,26								
2018										
Min		1,77	1,17	1,11	1,06	1,04	1,01	0,99	1,00	1,00
Max		2,26	1,28	1,14	1,09	1,05	1,03	1,01	1,01	1,00
Moyenne		1,99	1,22	1,12	1,07	1,04	1,02	1,00	1,01	1,00
Mediane		1,92	1,23	1,13	1,07	1,04	1,02	0,99	1,01	1,00
Chain Ladder		1,98	1,22	1,12	1,07	1,04	1,02	1,00	1,01	1,00
Selected		**1,98**	**1,22**	**1,12**	**1,07**	**1,04**	**1,02**	**1,00**	**1,01**	**1,00**

FIGURE 4.4 – Triangle des facteurs de développement individuels

On voit avec les différentes statistiques descriptives de base qu'il n'y a pas d'outliers dans les observations et que la liquidation des sinistres semble stable, ainsi les facteurs de développement obtenus par Chain ladder sont assez fiables.

La réserve IBNR est ensuite calculée sur les triangles de charges totales, même s'il faut garder en tête que cette méthode est vraiment une approche très simplifiée.

4.1.2 Résultats de la méthode

La réserve IBNR est estimée à 56 207 974€ pour ce portefeuille de sinistres. Elle est obtenue en déduisant les montants connus au 31/12/2018 des ultimes projetés, ce qui permet de connaître le montant qui sera dû au titre des sinistres survenus sur la période 2009-2018.

	Années de développement											
Survenance	0	1	2	3	4	5	6	7	8	9	Ultime	Réserve
2009	14 485 401	27 810 864	34 679 862	38 534 466	41 810 003	43 990 855	45 087 721	44 859 167	45 332 597	45 195 672	45 195 672	-
2010	13 135 693	26 737 946	31 364 706	35 285 311	37 845 715	39 372 862	40 486 855	40 016 346	40 109 641	39 988 492	39 988 492	- 121 149
2011	14 146 964	27 141 883	31 681 244	35 114 914	37 157 607	38 619 399	38 879 888	39 334 870	39 597 513	39 477 911	39 477 911	143 042
2012	16 011 252	28 414 095	33 816 625	37 657 678	40 311 220	41 871 818	42 881 945	42 797 844	43 083 611	42 953 479	42 953 479	71 534
2013	12 888 928	24 768 498	30 165 924	34 341 997	36 759 707	38 144 421	38 954 887	38 878 488	39 138 085	39 019 870	39 019 870	875 449
2014	12 041 707	22 220 071	28 385 083	32 093 163	34 221 616	35 653 976	36 411 526	36 340 115	36 582 763	36 472 267	36 472 267	2 250 651
2015	12 110 485	24 864 658	30 985 738	34 967 113	37 442 127	39 009 282	39 838 124	39 759 993	40 025 475	39 904 581	39 904 581	4 937 467
2016	11 351 040	24 971 808	31 283 840	35 092 516	37 576 406	39 149 181	39 980 995	39 902 584	40 169 019	40 047 690	40 047 690	8 763 850
2017	12 101 769	27 326 301	33 326 023	37 383 326	40 029 362	41 704 807	42 590 921	42 507 392	42 791 219	42 661 970	42 661 970	15 335 669
2018	11 448 098	22 674 504	27 652 885	31 019 507	33 215 104	34 605 336	35 340 606	35 271 295	35 506 806	35 399 560	35 399 560	23 951 461
	1,98	1,22	1,12	1,07	1,04	1,02	1,00	1,01	1,00	1,00	Réserve IBNR	56 207 974

FIGURE 4.5 – Résultat du calcul de la réserve IBNR sur le triangle de charges par la méthode Chain Ladder

4.2 Application de Chain Ladder sur méthode inflatée

4.2.1 Choix de l'inflation

Le fait de travailler sur des données historiques pourrait être considéré si et seulement si nous admettons que l'inflation passée va se reproduire de manière identique dans le futur, ce qui est en effet possible dans certains cas mais limite tout de même les possibilités.
Ainsi, il serait intéressant de considérer les triangles de charges totales à inflation constante, par exemple au 31/12/2018 car les facteurs de développement ne seraient ainsi plus biaisés par les inflations contenues dans les variations observées.
Nous ne disposons pas de données nous permettant de dégager clairement une inflation, ni de données marché qui permettraient de l'approcher sur le segment RC automobile corporelle. Après quelques recherches sur le Maroc, nous avons trouvé finalement les taux d'inflation suivants sur 2014-2020. Cependant, nous n'avons aucun moyen d'en évaluer la pertinence. Il faudrait pour cela évaluer l'inflation corporels automobiles au Maroc. Néanmoins, par soucis d'exhaustivité, nous avons fait l'exercice en considérant une inflation de 1,29% passée et future.

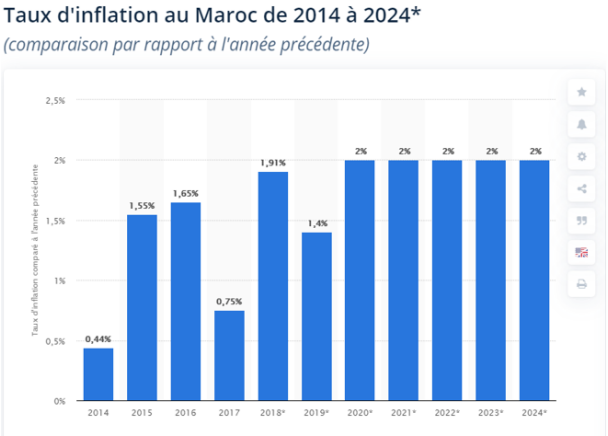

FIGURE 4.6 – Taux d'inflation au Maroc entre 2014 et 2024 (comparaison par rapport à l'année précédente)

4.2.2 Mise en as-if des sinistres et calcul

Nous différencions donc les règlements des provisions pour *as-ifer* les sinistres, c'est à dire, pour mettre tous les sinistres en valeur monnaie 2018 selon les hypothèses d'inflation définies ci-dessus.

Formules mathématiques pour la mise en as-if en 2018

Soit I_i l'indice pour l'année de survenance i. Alors :

$$I_{2009} = 100$$

et

$$I_{i+1} = I_i \times (1 + r)$$

où r est le taux d'inflation pris constant dans notre cas à 1,29%.

Il est d'usage de considérer les règlements non-cumulés car le montant réglé en année de développement j (où $0 \leq j \leq n-1$) années après la survenance i correspond à l'année comptable $i + j$. Soit $P_{i,j}$ le montant payé pour l'année de survenance i en année de développement j :
Pour $j = 0$ alors

$$P_{i,0}^{as-if} = P_{i,0} \times \frac{indice_{2018}}{indice_i}$$

.
Pour $1 \leq j \leq 9$ et $i + j \leq 2018$ alors

$$P_{i,j}^{as-if} = P_{i,j-1}^{as-if} + (P_{i,j} - P_{i,j-1}) \times \frac{indice_{2018}}{indice_{i+j}}$$

En revanche, pour les provisions, l'hypothèse communément admise est que les provisions sont revues à chaque date d'arrêté comptable.
Soit $R_{i,j}$ le montant de provisions pour l'année de survenance i en année de développement j :
Pour $i + j \leq 2018$ alors

$$R_{i,j}^{as-if} = R_{i,j} \times \frac{indice_{2018}}{indice_{i+j}}$$

Cette dernière hypothèse est assez forte notamment quand on constate que les règlements et les provisions n'ont pas évolué entre j et $j + 1$. Dans ce cas, travailler sur des données *as-ifées* aura un effet qu'il faudra corriger car, une fois inflatés, les sinistres inchangés entre j et $j + 1$ seront revalorisés et montreront une baisse.

4.2.3 Calculs

Nous pouvons ainsi déduire le triangle de charges

$$C_{i,j}^{as-if} = R_{i,j}^{as-if} + P_{i,j}^{as-if}$$

	Années de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	16 256 598	30 818 337	37 964 871	41 714 147	44 793 096	46 684 248	47 439 538	46 833 152	46 961 945	46 482 976
2010	14 554 107	29 251 459	33 894 346	37 692 821	40 005 379	41 228 060	42 021 222	41 203 841	40 978 607	
2011	15 474 950	29 314 748	33 798 875	37 036 322	38 797 514	39 952 737	39 879 914	40 013 602		
2012	17 291 183	30 297 246	35 617 232	39 224 566	41 563 717	42 769 645	43 411 866			
2013	13 741 991	26 073 125	31 370 179	35 317 160	37 433 549	38 487 859				
2014	12 675 186	23 093 119	29 144 734	32 601 479	34 426 765					
2015	12 585 233	25 513 336	31 420 236	35 082 997						
2016	11 645 786	25 298 089	31 314 658							
2017	12 257 881	27 331 148								
2018	11 448 098									

FIGURE 4.7 – Triangle de charges après la mise en as-if

Nous obtenons ainsi le triangle de facteurs de développement individuels suivant et l'on remarque que :

- Pour les facteurs de développement individuels entre les années 0 et 1, nous observons une tendance à la hausse depuis 2014. Le coefficient Chain Ladder retenu ne permet pas de voir cette information. Selon les informations fournies par les experts, il serait probable que l'on retienne plutôt une projection de la tendance observée sur 2014-2018 qu'une valeur moyenne bien inférieure aux valeurs observées.
- Pour les facteurs de développement individuels entre les années 1 et 2, nous observons également une rupture potentielle entre 2013 et 2014.
- Le reste des données observées est relativement stables et permettent de conserver les valeurs calculées par Chain Ladder.
- Compte tenu des valeurs observées très proches de 1, voire légèrement inférieures pour les années de développement les plus récentes, nous décidons dans ce jeu d'hypothèses de considérer un facteur de queue (*tail factor*) à 1.

	Années de développement										
Survenance	0	1	2	3	4	5	6	7	8	9	
2009		1,90	1,23	1,10	1,07	1,04	1,02	0,99	1,00	0,99	
2010		2,01	1,16	1,11	1,06	1,03	1,02	0,98	0,99		
2011		1,89	1,15	1,10	1,05	1,03	1,00	1,00			
2012		1,75	1,18	1,10	1,06	1,03	1,02				
2013		1,90	1,20	1,13	1,06	1,03					
2014		1,82	1,26	1,12	1,06						
2015		2,03	1,23	1,12							
2016		2,17	1,24								
2017		2,23									
2018											
Min		1,75	1,15	1,10	1,05	1,03	1,00	0,98	0,99	0,99	
Max		2,23	1,26	1,13	1,07	1,04	1,02	1,00	1,00	0,99	
Moyenne		1,97	1,21	1,11	1,06	1,03	1,01	0,99	1,00	0,99	
Mediane		1,90	1,22	1,11	1,06	1,03	1,02	0,99	1,00	0,99	
Chain Ladder		1,95	1,20	1,11	1,06	1,03	1,01	0,99	1,00	0,99	Tail
Selected		1,95	1,20	1,11	1,06	1,03	1,01	0,99	1,00	0,99	1,00

FIGURE 4.8 – Triangle de facteurs de développement individuel du triangle de charge *as-ifé*

Nous pouvons ensuite compléter le triangle de charges en as-if 2018 présenté ci-dessous :
Afin de pouvoir déterminer le triangle des règlements et des provisions, il faut obtenir une cadence de règlements sur les données as-if et pour cela il faut considérer les $\frac{P_{i,j}^{as-if}}{C_{i,j}^{ultime}}$.
Nous décidons également de prendre l'hypothèse très forte d'un paiement total en année de développement

	Années de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	16 256 598	30 818 337	37 964 871	41 714 147	44 793 096	46 684 248	47 439 538	46 833 152	46 961 945	46 482 976
2010	14 554 107	29 251 459	33 894 346	37 692 821	40 005 379	41 228 060	42 021 222	41 203 841	40 978 607	40 560 663
2011	15 474 950	29 314 748	33 798 875	37 036 322	38 797 514	39 952 737	39 879 914	40 013 602	39 969 769	39 562 114
2012	17 291 183	30 297 246	35 617 232	39 224 566	41 563 717	42 769 645	43 411 866	42 978 865	42 931 783	42 493 918
2013	13 741 991	26 073 125	31 370 179	35 317 160	37 433 549	38 487 859	38 965 555	38 576 903	38 534 643	38 141 625
2014	12 675 186	23 093 119	29 144 734	32 601 479	34 426 765	35 536 291	35 977 354	35 618 506	35 579 487	35 216 609
2015	12 585 233	25 513 336	31 420 236	35 082 997	37 190 854	38 389 462	38 865 937	38 478 278	38 436 126	38 044 113
2016	11 645 786	25 298 089	31 314 658	34 733 203	36 820 044	38 006 701	38 478 425	38 094 631	38 052 900	37 664 796
2017	12 257 881	27 331 148	32 913 563	36 506 656	38 700 050	39 947 297	40 443 107	40 039 717	39 995 855	39 587 934
2018	11 448 098	22 355 373	26 921 481	29 860 433	31 654 508	32 674 688	33 080 233	32 750 282	32 714 405	32 380 749

FIGURE 4.9 – Triangle de charges as-if 2018 complété

10. Dans la réalité, cette cadence aurait été lissée sur plusieurs années de développement et les montants auraient été plus exposés à l'effet d'inflation.

	Années de développement										
Survenance	0	1	2	3	4	5	6	7	8	9	
2009	1%	4%	12%	20%	28%	35%	39%	42%	43%	44%	
2010	1%	4%	10%	19%	28%	34%	38%	40%	41%		
2011	1%	4%	11%	22%	30%	34%	37%	39%			
2012	0%	4%	13%	21%	29%	33%	35%				
2013	0%	4%	13%	24%	30%	34%					
2014	0%	5%	16%	25%	33%						
2015	1%	7%	17%	27%							
2016	1%	6%	16%								
2017	1%	6%									
2018	1%										
Min	0%	4%	10%	19%	28%	33%	35%	39%	41%	44%	
Max	1%	7%	17%	27%	33%	35%	39%	42%	43%	44%	
Moyenne	1%	5%	13%	23%	30%	34%	37%	40%	42%	44%	
Mediane	1%	4%	13%	22%	29%	34%	37%	40%	42%	44%	Tail
Sélection	1%	5%	13%	23%	30%	34%	37%	40%	42%	44%	100%

FIGURE 4.10 – Cadence de règlements

Enfin, nous réinjectons l'inflation sur les diagonales post-2018 où nous avons gardé par simplicité la valeur de 1,29% par an pour projeter les années futures.

	Années de développement											
Survenance	0	1	2	3	4	5	6	7	8	9	Ultime	IBNR
2009	14 485 401	27 810 864	34 679 862	38 534 466	41 810 003	43 990 855	45 087 721	44 859 167	45 332 597	45 195 672	45 530 776	335 104
2010	13 135 693	26 737 946	31 364 706	35 285 311	37 845 715	39 372 862	40 486 855	40 016 346	40 109 641	39 998 104	39 998 104	- 111 537
2011	14 146 964	27 141 883	31 681 244	35 114 914	37 157 607	38 619 399	38 879 888	39 334 870	39 606 931	39 492 052	39 203 162	- 131 708
2012	16 011 252	28 414 095	33 816 625	37 657 678	40 311 220	41 871 818	42 881 945	42 809 709	43 099 847	42 974 863	42 350 263	- 531 682
2013	12 888 928	24 768 498	30 165 924	34 341 997	36 759 707	38 144 421	38 955 635	38 880 101	39 143 883	39 030 252	38 183 876	39 455
2014	12 041 707	22 220 071	28 385 083	32 093 163	34 221 616	35 639 452	36 399 277	36 328 636	36 575 330	36 469 061	35 420 353	1 198 737
2015	12 110 485	24 864 658	30 985 738	34 967 113	37 421 441	38 990 014	39 821 432	39 744 136	40 014 075	39 897 792	38 472 465	3 505 351
2016	11 351 040	24 971 808	31 283 840	35 071 525	37 555 023	39 127 990	39 961 737	39 884 224	40 154 919	40 038 310	40 335 332	9 051 492
2017	12 101 769	27 326 301	33 303 428	37 350 792	39 994 768	41 669 376	42 556 998	42 474 477	42 762 663	42 638 520	42 954 734	15 628 433
2018	11 448 098	22 639 317	27 596 511	30 949 737	33 140 262	34 527 669	35 263 060	35 194 691	35 433 453	35 330 601	35 592 583	24 144 485
											Total	53 128 131

FIGURE 4.11 – Projection des IBNR avec une méthode Chain Ladder inflatée

Pour rappel, le montant d'IBNR obtenu avec la méthode non-inflatée était de 56 207 974€ alors que le montant d'IBNR obtenu avec la méthode inflatée est de 53 138 131€ soit une différence de 3 079 844€. Nous pouvons donc en déduire que le fait de considérer l'inflation sur les données historiques et futures a un impact direct sur le montant des réserves (dans notre cas, l'impact est à la baisse). Par cette méthode, le fait de retirer l'inflation inhérente aux montants payés a entraîné une baisse dans les coefficients de

développement individuels par rapport à la méthode Chain Ladder classique. Par conséquent, le triangle est développé de manière moins sévère qu'en gardant l'inflation ce qui explique cette baisse de plus de 3 millions d'IBNR. Afin d'illustrer cette interprétation, nous avons augmenté le taux d'inflation passé de 1,29% à 3% et le taux d'inflation futur à 2%,, ce qui conduit à une baisse du montant d'IBNR de plus de 10 millions d'euros par rapport au calcul Chain Ladder sans tenir compte de l'inflation.
Les pratiques actuelles du marché calculent les provisions sans prendre en compte l'inflation. Compte tenu que notre mesure d'inflation reste assez simpliste, nous préférons continuer sans cette étape de correction de l'inflation, le sujet de l'estimation de l'inflation des sinistres constituant quasiment un sujet de mémoire à lui seul.

4.3 Méthode de Mack

Comme nous l'avons expliqué dans la partie théorique, la méthode de Mack fait intervenir la notion d'incertitude dans la méthode déterministe de Chain Ladder. Ainsi, l'apport de T.Mack par rapport à Chain Ladder déterministe est la possibilité de quantifier l'erreur de prédiction. L'erreur quadratique moyenne de prévision (MSEP) des provisions représente l'écart moyen entre les valeurs estimées et les valeurs observées réellement.

Ainsi, si l'on reprend les étapes indiquées dans la partie théorique et les triangles, coefficients de développement et résultats obtenus lors de l'application de la méthode Chain Ladder nous obtenons les estimations suivantes :

j	λj	σ²j
0		
1	1,98	329 173
2	1,22	42 480
3	1,12	4 032
4	1,07	2 827
5	1,04	1 451
6	1,02	3 674
7	1,00	5 736
8	1,01	1 430
9	1,00	356

FIGURE 4.12 – Estimateurs des moindres carrés de λ_j et σ_j

i	Ĉi,n	Ri	MSEP(Ri)	Sep(Ri)	IC 95% LB	IC 95% UB
2009	45 195 672					
2010	39 988 492	- 121 149	0,00097	0,03122	- 121 149	- 121 149
2011	39 477 911	143 042	0,00027	0,01629	143 042	143 042
2012	42 953 479	71 534	0,00013	0,01128	71 534	71 534
2013	39 019 870	875 449	0,00011	0,01062	875 449	875 449
2014	36 472 267	2 250 651	0,00010	0,00996	2 250 651	2 250 651
2015	39 904 581	4 937 467	0,00009	0,00961	4 937 467	4 937 467
2016	40 047 690	8 763 850	0,00007	0,00841	8 763 850	8 763 850
2017	42 661 970	15 335 669	0,00002	0,00495	15 335 669	15 335 669
2018	35 399 560	23 951 461	0,00001	0,00281	23 951 461	23 951 461

FIGURE 4.13 – Calculs des erreurs quadratiques moyennes

Total Réserve	56 207 974
MSEP R	10 201 457 294 043
Erreur de prédiction	3 193 972

FIGURE 4.14 – Erreur de prédiction finale

Ainsi, l'erreur de prédiction de la réserve IBNR calculée par Chain Ladder est de 3 193 972€.

4.4 Application de Schnieper sur le triangle de nombre

La méthode de Schnieper est basée sur la séparation des *IBNyR* et des *IBNeR*. Ainsi, en partant d'un triangle comme utilisé pour la méthode Chain Ladder, nous allons diviser ce triangle en deux nouveaux triangles D et N représentant les *IBNyR* et les *IBNeR*.

- Soit $C_{i,j}$ est le montant cumulé des sinistres pour chaque année de rattachement i et de développement j.
- Soit $N_{i,j}$ le montant des nouveaux sinistres, qu'on peut assimiler aux IBNyR.
- Soit $D_{i,j}$ le montant cumulé agrégé des variations des sinistres observées entres les années de développement $j-1$ et j pour chaque année de rattachement i. Cette composante correspond aux IBNeR.
- Soit E_i l'exposition pour chaque année de rattachement i

Les relations suivantes sont supposées vraies dans la méthode :

$$N_{i,0} = X_{i,0} \forall i = 1, ..., n \& \forall j = 1, ..., n$$

$$\hat{X_{i,j}} = X_{i,j-1} - D_{i,j} + N_{i,j} \forall i = 1, ..., n \& \forall j = 1, ..., n$$

Ainsi, à partir des données individuelles, nous avons pu reconstituer les trois triangles de nombres suivants :

Xij	Année de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	3 110	4 753	5 164	5 398	5 531	5 612	5 652	4 764	4 447	4 447
2010	2 353	3 750	4 180	4 377	4 496	4 571	4 630	3 769	3 778	
2011	1 754	3 200	3 564	3 740	3 857	3 946	3 423	3 463		
2012	2 383	3 607	3 987	4 228	4 382	4 472	4 557			
2013	1 903	3 230	3 647	3 916	4 094	4 219				
2014	1 835	3 080	3 516	3 765	3 939					
2015	1 982	3 479	3 916	4 184						
2016	1 854	3 582	4 154							
2017	1 945	3 938								
2018	1 877									

FIGURE 4.15 – Triangle de nombre cumulé agrégé $X_{i,j}$

Nij	Année de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	3 110	1 225	160	-	-	-	3	2	-	
2010	2 353	975	-	-	-	11	4	3		
2011	1 754	904	147	-	21	12	3			
2012	2 383	811	1	3	-	5				
2013	1 903	805	96	-	7					
2014	1 835	709	-	22						
2015	1 982	817	-							
2016	1 854	962	-							
2017	1 945	1 033								
2018	1 877									

FIGURE 4.16 – Nombre agrégé des nouveaux sinistres $N_{i,j}$

Dij	Année de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009		- 418	- 251	- 234	- 133	- 81	- 37	890	317	-
2010		- 422	- 430	- 197	- 119	- 64	- 55	864	- 9	
2011		- 542	- 217	- 176	- 96	- 77	526	- 40		
2012		- 413	- 379	- 238	- 154	- 85	- 85			
2013		- 522	- 321	- 269	- 171	- 125				
2014		- 536	- 436	- 227	- 174					
2015		- 680	- 437	- 268						
2016		- 766	- 572							
2017		- 960								
2018										

FIGURE 4.17 – Variations du nombre de sinistres $D_{i,j}$

Afin de prédire le nombre de sinistres qui ne sont pas déclarés, nous devons estimer les paramètres $\hat{\lambda_j} = \frac{\sum_{i=1}^{n+1-j} N_{i,j}}{\sum_{i=1}^{n+1-j} E_i} \forall j = 1, 2..n$ et $\hat{\delta_j} = \frac{\sum_{i=1}^{n+1-j} D_{i,j}}{\sum_{i=1}^{n+1-j} X_{i,j-1}}$
Compte tenu du fait que nous n'avions pas les données sur les expositions dans le portefeuille, nous avons pris une exposition constante pour chaque année de survenance égale à 10 000. Ainsi, nous avons pu complété les trois triangles $N_{i,j}$, $D_{i,j}$ et $X_{i,j}$ à l'aide des formules énoncées dans la partie théoriques, présentés ci-dessous.
Nous voyons ici que les déclarations de tardifs se font massivement dans les deux premières années de développement puis par la suite, le nombre de tardifs s'affaiblit. Dans le cas de notre portefeuille, les déclarations tardives ne dépassent pas la 7^{e}année de développement.

Nij	Année de développement										
Survenance	0	1	2	3	4	5	6	7	8	9	Eij
2009	4,157	1,298	160	69	34	27	3	2			10,000
2010	3,295	1,025	199	87	38	11	4	3	-	-	10,000
2011	2,821	975	157	45	21	12	3	-	-	-	10,000
2012	3,425	887	124	52	19	5	1	2	-	-	10,000
2013	3,157	883	98	40	7	5	3	2	-	-	10,000
2014	2,995	762	93	22	9	12	3	2	-	-	10,000
2015	3,178	855	81	28	21	12	3	2	-	-	10,000
2016	3,005	1,004	103	49	21	12	3	2	-	-	10,000
2017	2,857	1,022	127	49	21	12	3	2	-	-	10,000
2018	1,861	968	127	49	21	12	3	2	-	-	10,000
λj	0.3075	0.0968	0.0127	0.0049	0.0021	0.0012	0.0003	0.0002	-	-	

FIGURE 4.18 – Triangles des nouveaux sinistres $N_{i,j}$ complété

Nos données montraient des sinistres sans suite dès leur déclaration. Pour remédier à cela, nous avons décidé de les regrouper avec les sinistres qui sont sans suite en première année de développement.
Les données sont assez atypiques puisqu'il n'y a quasiment aucun sinistre sans suite après la première année de développement, ce qui est vraiment inhabituel en Responsabilité Civile Corporelle automobile.

Dij	Année de développement										
Survenance	0	1	2	3	4	5	6	7	8	9	Total ss suite
2009		62									62
2010	79	96	4							-	100
2011	79	97							-	-	97
2012	97	100						-	-	-	100
2013	50	60	1	1			-	-	-	-	62
2014	68	80				-	-	-	-	-	80
2015	53	58			-	-	-	-	-	-	58
2016	41	46		0	-	-	-	-	-	-	46
2017	32	36	1	0	-	-	-	-	-	-	37
2018	28	41	0	0	-	-	-	-	-	-	41
fj		0.02197992	0.00015095	3.3259E-05	0	0	0	0	0	0	

FIGURE 4.19 – Triangle du nombre de sinistres sans suite $D_{i,j}$ complété

Xij	Année de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	4,157	5,393	5,553	5,622	5,656	5,683	5,686	5,688	5,688	5,688
2010	3,295	4,224	4,419	4,506	4,544	4,555	4,559	4,562	4,562	4,562
2011	2,821	3,699	3,856	3,901	3,922	3,934	3,937	3,937	3,937	3,937
2012	3,425	4,212	4,336	4,388	4,407	4,412	4,413	4,415	4,415	4,415
2013	3,157	3,980	4,077	4,116	4,123	4,128	4,131	4,132	4,132	4,132
2014	2,995	3,677	3,770	3,792	3,801	3,813	3,816	3,817	3,817	3,817
2015	3,178	3,975	4,056	4,084	4,105	4,117	4,120	4,122	4,122	4,122
2016	3,005	3,963	4,066	4,115	4,136	4,148	4,151	4,153	4,153	4,153
2017	2,857	3,843	3,969	4,018	4,039	4,051	4,054	4,056	4,056	4,056
2018	1,861	2,788	2,914	2,963	2,985	2,997	2,999	3,001	3,001	3,001
										Total

FIGURE 4.20 – Triangle des $X_{i,j}$ complété

Le tableau ci-dessous montre que la méthode de Schnieper est extrêmement sensible à l'évolution de l'exposition E_i. Dans notre cas, nous prenons une exposition constante E_i de 10 000 pour chaque année de survenance alors que le nombre de nouveaux sinistres en première année de développement semble décroître au fil des années. Nous avons aussi effectué les calculs en considérant que l'exposition varie de la même façon que les nouveaux sinistres observés en première année de développement. Ainsi, l'exemple montre bien que si le choix de l'exposition n'est pas correctement appuyé par des informations sur la compagnie et le marché, cela peut introduire un biais significatif.
Nous voyons ici que selon les variations d'exposition retenues, nous pouvons avoir des variations d'estimations de sinistres tardifs de 30% environ. Une mesure d'exposition pertinente pour analyser les nombres est clé. Celle-ci doit d'ailleurs être complétée par des informations sur la qualité du portefeuille : ainsi par exemple un changement de mix dans la nature des risques souscrits a une forte influence sur la fréquence de sinistralité.

	Synthèse		Synthèse	
	Exposition constante		*Exposition proportionnelle au nombre de contrats souscrits*	
	Exposition	Tardifs excl sans suite	Exposition variable	Tardifs excl sans suite
2009	10 000	-	10 000	-
2010	10 000	-	7 926	-
2011	10 000	-	6 786	-
2012	10 000	2	8 239	-
2013	10 000	4	7 594	4
2014	10 000	16	7 205	15
2015	10 000	38	7 645	36
2016	10 000	87	7 229	79
2017	10 000	213	6 873	186
2018	10 000	1 140	4 477	641
Total		1 500		961

FIGURE 4.21 – Résultats

Deuxième partie

Approche hybride d'estimation de la réserve IBNR : modélisation séparée du nombre et du coût pour les sinistres tardifs

Chapitre 5

Estimation des sinistres tardifs par une approche fréquence / coût

Tout d'abord, l'intérêt d'implémenter une approche séparée fréquence x coût en provisionnement (approche classiquement utilisée pour l'analyse de la prime pure en tarification) va permettre de prendre en compte les délais de déclaration ainsi que le statut des sinistres, élément de poids qui n'est pas pris en compte dans les méthodes traditionnelles. Ainsi, avec cette étude séparée des nombres et des montants, nous avons la possibilité de capter les phénomènes influant sur le nombre et le coût de manière bien plus précise que les méthodes agrégées.
Le déroulement de notre approche, inspirée de la recherche de Pietro Parodi dans sa publication *Triangle free reserving : a non-traditionnal framework for estimating reserves and reserve uncertainty (2013)* est le suivant :

Dans un premier temps, nous souhaitons prédire le nombre de sinistres survenus mais non encore déclarés à l'assureur, que nous appelons aussi les sinistres tardifs. L'objectif alors est de savoir combien de temps s'écoule entre la date de survenance et la date de déclaration afin de projeter au mieux le nombre de sinistres qui sont déjà survenus mais dont l'assureur n'a pour le moment pas connaissance. Pour cela, il faut modéliser les délais de déclaration (en jours) en ajustant une loi statistique aux délais qui sont observés dans l'historique de la base.
Cette approche permet de placer beaucoup de poids sur la variable de temps, élément qui n'est pas présent dans les méthodes traditionnelles et qui pourrait apporter plus de précision dans l'estimation des IBNR. De plus, les délais de déclaration sont renseignés en jours et permettent d'utiliser une maille très fine de données dans les estimations.

Ensuite, grâce à la distribution des délais de déclaration, nous serons capables d'obtenir une estimation du nombre de sinistres tardifs au titre de chaque année de survenance et d'en déduire la loi des nombres et ses paramètres. Le choix de cette loi pour les nombres des sinistres tardifs se fait entre la loi de Poisson et la loi Négative Binomiale.

Dans un second temps, nous analysons les montants de charges connus pour les sinistres en cours et pour les sinistres clos. L'objectif est d'ajuster une loi statistique aux montants des charges qui nous permettra d'associer un coût aux sinistres tardifs que nous projetons dans la première partie. L'ajustement d'une loi pour le montant des charges requiert d'amener les sinistres à l'ultime par une projection des *IBNeR*.

Ainsi, nous serons en mesure d'obtenir la distribution de la réserve IBNyR en combinant la distribution du nombre des sinistres tardifs avec la distribution des charges sinistres à l'ultime. En simulant un grand nombre de scénarios, nous serons à même d'avoir assez d'observations pour estimer la charge due aux sinistres tardifs. Grâce à la propriété très forte de la loi des grands nombres, plus nous simulons de variables aléatoires, plus l'espérance de la distribution obtenue de notre réserve IBNyR convergera vers sa moyenne empirique. C'est ainsi que nous pourrons obtenir un résultat unique pour le montant de la réserve au titre des sinistres tardifs.

5.1 Notions théoriques

5.1.1 Ajustement d'une loi de distribution à un échantillon de données

L'ajustement d'une loi de distribution à un échantillon de données (dans notre cas le nombre de sinistres tardifs et les charges) consiste, une fois le type de loi choisie, à estimer les paramètres de la loi de telle sorte que l'échantillon soit représenté le plus vraisemblablement possible (au sens du maximum de vraisemblance) par la loi ou qu'au moins certaines statistiques de l'échantillon (moyenne, variance par exemple) correspondent le mieux possible à celles de la loi. Ce travail demande une bonne analyse statistique ainsi qu'un jugement d'expert correctement justifié. C'est un processus itératif où il faut évaluer plusieurs distributions avant de trouver la plus adéquate. Sur R, les fonctions *plotdist* et *descdist* permettent de choisir le type de lois les mieux adaptées aux données puis la fonction *fitdist* permet d'estimer les paramètres de la loi avec l'estimateur du maximum de vraisemblance [1]

Avant d'ajuster une ou plusieurs lois statistiques à nos variables d'intérêt, il est nécessaire d'étudier la forme de la densité des données empiriques car cela permet de savoir vers quelles lois diriger notre étude. Le choix des lois "éligibles" est la première étape et est effectuée graphiquement grâce à la fonction *plotdist* préinstallée dans R qui sort l'histogramme de densité et la fonction de distribution empirique. Aussi, le graphique *Cullen and Frey* est un bon moyen d'éliminer les lois qui ne correspondent pas à notre recherche.

5.1.2 Statistiques descriptives et mesures de forme

En plus des graphiques empiriques, les statistiques descriptives peuvent également aider à choisir quelles sont les distributions paramétriques les plus adaptées pour décrire les données. Les coefficients d'asymétrie et d'aplatissement sont des caractéristiques de forme liées aux troisième et quatrième moments et sont utiles dans le choix de la distribution. Les formules des coefficients d'asymétrie et d'aplatissement pour un échantillon $(X_i)_i \overset{i.i.d}{\sim} X$ sont les suivantes :

$$sk(X) = \frac{\mathrm{E}[(X-\mathrm{E}(X))^3]}{\mathrm{V}(X)^{\frac{3}{2}}} \text{ et } \hat{sk} = \frac{\sqrt{n(n-1)}}{n-2}\frac{m_3}{{m_2}^{\frac{3}{2}}}$$
$$kr(X) = \frac{\mathrm{E}[(X-\mathrm{E}(X))^4]}{\mathrm{V}(X)2} \text{ et } \hat{kr} = \frac{n-1}{(n-2)(n-3)}((n+1)\frac{m_4}{{m_2}^2} - 3(n-1)) + 3$$
où m_2,m_3,m_4 sont les moments empiriques définies par $m_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k$

Ces deux statistiques de forme sont très sensibles aux valeurs extrêmes surtout dans les petits échantillons, il faut donc les utiliser avec prudence.

1. voir annexe

La fonction *descdist* préinstallée dans R fournit les statistiques descriptives classiques (minimum, maximum, moyenne, écart type, coefficient d'aplatissement et coefficient d'asymétrie). La fonction présente également un graphique des couples de coefficients (asymétrie ; aplatissement), appelé *Cullen and Frey Graph* des distributions connues et de la distribution empirique qui aide à déterminer les meilleures lois candidates.
Pour certaines lois statistiques (loi normale, uniforme, logistique, exponentielle), les coefficients $sk(X)$ et $kr(X)$ ont une valeur unique et sont représentées par des points sur le graphique. En ce qui concerne les autres distributions, les coefficients peuvent prendre plusieurs valeurs, et les couples sont par conséquent représentés par des lignes ou par des zones entières grisées. Le couple (asymétrie, aplatissement) des données observées est représenté par le point bleu.
Il est néanmoins important de noter que les coefficients d'aplatissement et d'asymétrie ne sont pas robustes et il faut les utiliser de manière prudente. Afin de prendre en compte l'incertitude des valeurs estimées pour ces coefficients, une procédure de bootstrap est effectuée et permet d'avoir une meilleure vision de la distribution empirique et d'empêcher de rejeter certaines distributions qui semblent éloignées du point bleu. Cette procédure est illustrée par le nuage de point jaune. Le graphique *Cullen and Frey* est surtout un moyen d'éliminer dans un premier temps les distributions qui ne sont pas du tout appropriées aux données.

5.1.3 Méthode du maximum de vraisemblance

Après avoir sélectionné les lois candidates qui semblent les plus adaptées pour chaque variable, nous les ajustons à nos données et en estimons les paramètres pour chacune grâce à la fonction R *fitdist*.
Cette fonction estime les paramètres des lois selon plusieurs méthodes (maximum de vraisemblance, méthodes des moments...).
Nous choisissons de retenir les paramètres issus de la méthode du maximum de vraisemblance. La vraisemblance mesure la probabilité que les observations proviennent d'un échantillon de loi paramétrée par θ. Trouver le maximum de vraisemblance consiste alors à trouver le paramètre le plus vraisemblable pour l'échantillon.
Pour rappel, la fonction de vraisemblance $L(\theta)$ prend la forme suivante :

$$L(\theta) = \prod_{i-1}^{n} f(x_i|\theta)$$

avec θ le vecteur de paramètres de la distribution, x_i les n observations(montant de la charge placée au titre du ième sinistre ou le délai de déclaration du ième sinistre) et $f(.|\theta)$ la fonction de densité de la distribution paramétrique statistique
Usuellement, c'est la log-vraisemblance qui est considérée : $l(\theta) = ln(L(\theta))$
L'intérêt de cette méthode est qu'au-delà de trouver un estimateur pour les paramètres de la loi, elle permet aussi de trouver un estimateur muni d'excellentes qualités : il est consistant, asymptotiquement sans biais et de variance minimale. La plupart du temps, il est unique et se détermine explicitement.
Pour trouver l'estimateur, il s'agit maximiser la log-vraisemblance $\log(L(\theta)$ en annulant la dérivée première afin d'obtenir les estimateurs $\hat{\theta}$. Nous détaillons la méthode en annexe.

5.1.4 Ajustement des lois candidates aux données empiriques de la base

La fonction *fitdist* offre une sortie R avec des résultats numériques et des résultats graphiques.
Les résultats numériques obtenus sont :

— Les estimateurs des paramètres de la distribution $\underline{\hat{\theta}}$ par la méthode *mle*[1]
— Les écarts types des estimateurs
— La valeur du log de la fonction de vraisemblance.
— Les critères d'informations AIC et BIC : ce sont des critères qui permettent de mesurer la qualité d'un modèle statistique. Si l'on considère un ensemble de modèles candidats, le modèle choisi est celui qui aura la plus faible valeur d'AIC.
— La matrice de corrélation entre les estimateurs (si la distribution a plusieurs paramètres)

Les résultats graphiques obtenus sont :
— Un graphique de densité représentant la densité de la distribution ajustée et l'histogramme de la densité empirique. Plus les deux coïncident, plus l'ajustement est correct.
— Un graphique représentant la fonction de distribution ajustée et la fonction de distribution empirique. De la même manière, plus les deux coïncident, plus l'ajustement est correct.
— Un Q-Q plot, qui permet de confronter les quantiles observés avec les quantiles de la distribution ajustée.
— Un P-P plot qui trace les deux fonctions de distribution cumulées l'une contre l'autre.

Les graphiques de densité et de distribution sont considérés comme les plus révélateurs en termes de tests d'ajustements. Le QQ plot souligne les défauts d'ajustement dans les queues des distributions alors que le PP plot les souligne au niveau du centre de la distribution. Les fonctions R *denscomp, cdfcomp, qqcomp et ppcomp* permettent de tracer simultanément les quatre graphiques afin de pouvoir comparer les données empiriques avec les différentes distributions ajustées. C'est ainsi qu'il va être possible de choisir les distributions les plus adaptées aux données.

5.1.5 Test d'adéquation aux lois / Goodness-of-fit

La fonction *gofstat* sur R donne plusieurs éléments qui permettent d'évaluer la qualité d'ajustement des distributions.

Le critère AIC[2]est une mesure de la qualité d'un modèle statistique. Si l'on considère un ensemble de lois candidates, la loi choisie est celle qui offrira la plus faible valeur d'AIC.

En ce qui concerne les tests statistiques, le test d'adéquation à la loi connue se présente comme suit ($F0$ est la fonction de répartition de la loi candidate et F est la fonction de répartition empirique) :

$$\boxed{\begin{array}{l} H0 : F{=}F0 \\ H1 : F{\neq}\mathrm{F0} \end{array}}$$

La fonction R *gofstat* offre trois valeurs de tests statistiques calculés par les 3 méthodes suivantes :
— La statistique Kolmogorov Smirnoff (KS) permet d'estimer si le caractère d'un échantillon peut être approximé par une loi de probabilité connue. La statistique mesure l'écart maximum qui existe entre la fonction de répartition empirique et la fonction de répartition théorique. Sous l'hypothèse H0 où les données empiriques peuvent en effet être assimilés à la loi théorique, alors la statistique tend vers 0. La meilleure distribution candidate sera donc celle qui offre la plus faible valeur pour la statistique.

1. Maximum Likelihood Estimation, en anglais maximum de vraisemblance
2. Critère d'Information Akaike

— La statistique Cramer Von Mises (CVM) vérifie aussi la compatibilité entre l'échantillon de nos données et une distribution choisie. Plus la loi est ajustée aux données, plus la statistique sera faible.
— La statistique Anderson- Darling (AD) qui détermine si les données suivent une loi de distribution spécifique. Pareillement, plus la loi est ajustée aux données, plus la statistique sera faible.

Les trois tests débutent avec la même logique à savoir examiner la distance entre la fonction de répartition théorique supposée sous H0 et la fonction de répartition empirique considérée. Ils diffèrent en fait lors du calcul de la statistique. Les tests AD et CVM prennent en compte toutes les observations dans le calcul du test statistique alors que KS ne calcule que la distance maximale entre les deux fonctions de répartition. Dans le cas de CVM, le même poids est donné à toutes les observations, ce qui n'est pas le cas pour AD qui pondère de manière plus forte les observations situées dans les queues de la distribution. La statistique AD es intéressante si nous souhaitons étudier l'écart des deux lois dans les queues des distributions, mais comme nous n'avons pas beaucoup de valeurs extrêmes, cela sera sans doute peu nécessaire. [1]

5.1.6 Méthode de Monte Carlo

Principe de la méthode

Cette méthode permet d'estimer une quantité en simulant des variables aléatoires tirées d'une distribution choisie.
Dans notre cas, la quantité que nous souhaitons estimer est le montant de la réserve *IBNyR*, c'est à dire la réserve placée au titre des sinistres tardifs.
Notons δ la quantité que nous souhaitons estimer.
Le problème écrit sous la forme d'une espérance est le suivant :

$$\boxed{\delta = \mathrm{E}_v[h(X)] = \int_{\Re^d} h(x)f(dx)}$$

Le montant de la charge dûe au titre des sinistres tardifs est traité comme une variable aléatoire $\mathbf{X}$=$(X_1, X_2..., X_d)$ suivant la loi v sur $\Re^d$

La solution du problème lorsqu'on le traite avec les méthodes de Monte Carlo est de simuler une suite $(\mathbf{X}_n)_{n\geq 1}$ indépendantes et identiquement distribuées sur la loi v.

Grâce à la loi des grands nombres (si n grand), l'espérance $\mathrm{E}_v[h(X)]$ converge vers la moyenne empirique :

$$\boxed{\bar{h_n} = \frac{1}{n} * \sum_{i=1}^{n} h(X_i)}$$

La méthode permet de modéliser les probabilités de différents résultats qui sont difficilement prédictibles du à l'intervention des variables aléatoires. C'est une technique qui prend en compte l'incertitude des prédictions et des projections. Lorsqu'il existe une incertitude significative dans l'élaboration de projections, la méthode de Monte Carlo propose une solution efficace alternative au remplacement de la variable inconnue par la moyenne.

1. Théorie en annexe

Validité de la méthode

La convergence de la méthode est assurée par la loi des grands nombres sous l'hypothèse que $\mathrm{E}_v[h(X)]$ existe (ie la loi h est intégrable par rapport à la mesure v).
Dans notre cas, le choix de h sera guidé par les parties de la modélisation précédentes où nous aurons choisi les distributions les plus adaptées pour définir le nombre et le coût.

Théorème : Loi faible et loi forte des grands nombres

Soit $(\mathbf{X}_n)_{n\geq 1}$, la suite de variables aléatoires réelles $i.i.d^1$ de loi v telle que $h(\mathbf{X}_1)$ soit v-intégrable.
Alors pour tout $\epsilon \geq 0$
$\Pr[|\bar{h_n} - \mathrm{E}[h(\mathbf{X}_1)]| \leq \epsilon] \underset{n\to+\infty}{\longrightarrow} 1$ (Loi faible)
$\Pr[|\bar{h_n} - \mathrm{E}[h(\mathbf{X}_1)]| \underset{n\to+\infty}{\longrightarrow} 0] = 1$ (Loi forte)

Selon la loi faible, la probabilité de commettre une erreur plus grande que ϵ est nulle, tandis qu'avec la loi forte, une fois que l'erreur d'estimation est inférieure à ϵ alors elle ne peut que décroître.

Erreur Monte Carlo & vitesse de convergence

L'erreur quadratique est donnée par :

$$\mathrm{E}[(\bar{\mathrm{h}}_\mathrm{n} - [\mathrm{h}(\mathbf{X})]^2] = \frac{\sigma^2}{n} \text{ avec } \sigma^2 = \mathrm{Var}[\mathrm{h}(\mathbf{X})]$$

Par ailleurs, $\frac{\sigma^2}{n} \underset{n\to\infty}{\longrightarrow} 0$
Il est donc possible de réduire l'erreur en augmentant le nombre n de simulations au prix du temps de calcul qui sera de plus en plus long.
Aussi, la vitesse de convergence de la méthode est notée $\mathcal{O}(n^{-\frac{1}{2}})$. Cette vitesse est très lente car elle signifie qu'il faut multiplier par 4 le nombre d'itérations pour diviser l'erreur par 2, ce qui engendre un coût de temps de calcul très lourd, mais elle présente l'avantage de ne pas dépendre de la dimension.

5.2 Application (1) : modélisation du nombre des sinistres tardifs

Dans cette partie, le but est de trouver une distribution pour le nombre des sinistres tardifs.

5.2.1 Ajustement d'une distribution des délais de déclaration

Choix des lois candidates : analyse statistique et graphique

Dans la base de données, les délais sont exprimés en jours. La moyenne est de 112 jours (notons que dans cette partie, nous avons ajuster la variable jours en ajoutant 1 afin de pouvoir ajuster le maximum de lois, la loi log normale par exemple n'étant pas définie sur zéro). Le maximum est de 2803 jours (environ 7 ans après la survenance). Le fait que l'information sur les délais de déclaration soit si fine est un avantage. Comme nous le verrons dans le graphique suivant, la majorité des sinistres sont déclarés dans les 2 ans après leur survenance, il est donc important d'avoir une vision plus fine qu'une maille annuelle pour pouvoir correctement effectuer les prévisions. De De plus, les $IBNyR$ sont par définition déjà survenus mais ne sont pas encore connus de l'assureur. Avoir une vision du temps écoulé entre la survenance et la déclaration permet de prédire au mieux les sinistres tardifs.

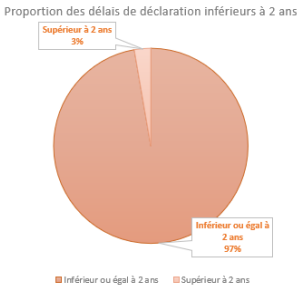

FIGURE 5.1 – Proportion des sinistres dont le délai de déclaration est inférieur à 2 ans

Pour rappel, ci-dessous l'histogramme des délais de déclaration des 41 024 sinistres observés dans la base (les pointillés bleus représente la moyenne) :

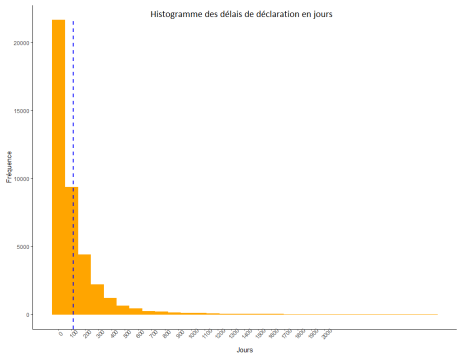

FIGURE 5.2 – Histogramme des délais en jours

Suite à nos recherches, nous avons vu que les tardifs en automobile sont déclarés au maximum dans les deux ans suivant la survenance. Ainsi, nos données nous indiquent d'une part les sinistres tardifs dont les délais de déclaration en jours sont inférieurs à 730 jours et d'autre part les réouvertures dont les délais sont supérieurs à 730 jours. La fonction *fitdist* appliquée sur les délais de déclaration permet d'avoir une idée de la forme de la loi candidate recherchée. Il en ressort ici que pour les sinistres tardifs, les distributions de type Gamma, Weibull et Beta sont les meilleures candidates car la distribution empirique des délais est une distribution à valeurs positives avec une queue légère à droite.

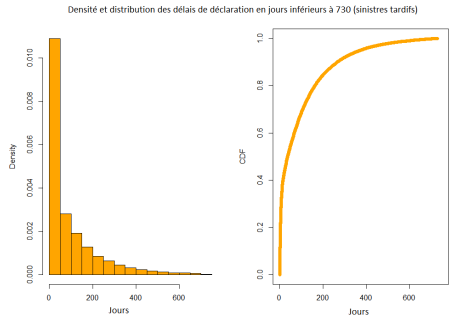

FIGURE 5.3 – Densité et distribution empirique des délais de déclaration inférieurs à 730 jours (2 ans)

Les réouvertures, contrairement aux sinistres tardifs n'épousent pas la forme des distributions de type Weibull, Gamma ou Beta.
L'approche la plus adéquate aurait été de calibrer deux lois distinctes sur les sinistres tardifs et les réouvertures afin de pouvoir prédire de manière pertinente les sinistres futurs qui découleront d'une part des déclarations tardives et d'autre part des réouvertures.
Cependant, le contexte sanitaire ne m'a pas permis de pouvoir analyser les deux concepts séparément (manque de temps, chômage technique, prise de décisions difficiles) et c'est pour cela que dans la suite du mémoire, nous calibrons qu'une seule loi pour les délais de déclaration. Cette approche est d'ailleurs celle retenue dans le papier de recherche de Pietro Parodi.

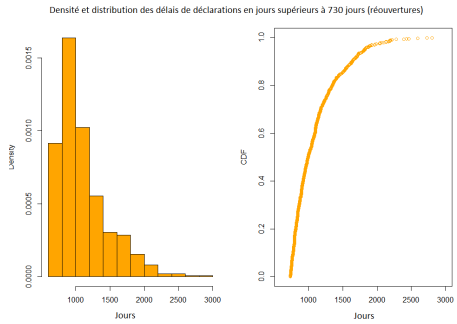

FIGURE 5.4 – Densité et distribution empirique des délais de déclaration supérieurs à 730 jours (2 ans)

Ainsi, si l'on se concentre sur les sinistres tardifs, le graphique *Cullen & Frey* permet directement de voir que les distributions normale, uniforme, exponentielle et logistique ne sont pas du tout adaptées et que les observations ont des propriétés plus semblables à celles des loi Weibull et Gamma.

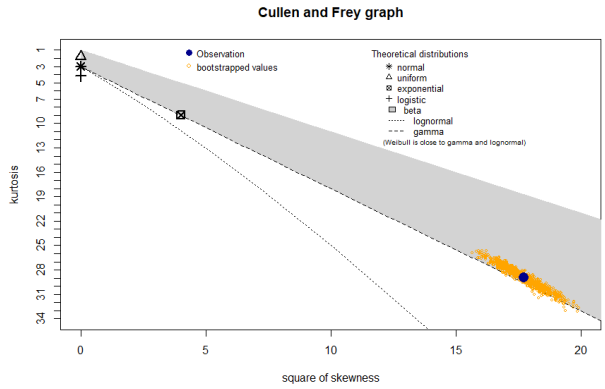

FIGURE 5.5 – Graphique Cullen & Frey sur les délais de déclaration en jours

Aussi, les statistiques descriptives révèlent un estimateur de skewness aux alentours de 4.11 et un coefficient d'aplatissement autour de 28. Le coefficient d'asymétrie étant positif et le coefficient d'aplatissement étant supérieur à 3 permettent de conforter notre première analyse.

Statistiques descriptives	Valeurs
Minimum	1
Maximum	2802
Médiane	40
Moyenne	112,56
Ecart type	196,77
Coefficient d'asymétrie	4,11
Coefficient d'aplatissement	28,18

FIGURE 5.6 – Statistiques descriptives des délais de déclaration en jours

Les lois Weibull, Gamma et Beta qui sont des distributions positives et non bornées à droite semblent être les meilleures candidates.

Choix de la meilleure loi candidate

Les fonctions *denscomp* et *cdfcomp* préinstallées dans R permettent de superposer les densités et les distributions des différentes lois candidates simultanément sur celles des données empiriques.
Parmi les trois distributions testées, la densité et la distribution de la loi gamma semble être celle qui s'éloigne le plus des données empiriques surtout à l'extérieur de la queue de distribution. Les lois beta et weibull coïncident bien avec la densité / distribution empirique, mais pour avoir plus de visibilité sur l'adéquation des données à ce lois, nous pouvons utiliser la fonction *gofstat* dans R qui sort tous les résultats des tests statistiques.

En termes de tests d'adéquation, la distribution la plus adéquate aux délais de déclaration observés est celle qui aura les valeurs de statistiques les plus faibles. Aussi, il faudrait que la valeur du critère AIC soit la plus petite parmi les trois possibilités. Ici, les tests d'adéquation à la loi de Weibull offrent les valeurs les plus faibles pour les trois tests (K-S, CVM, AD). L'ajustement à la loi Weibull donne aussi l'AIC le plus faible.

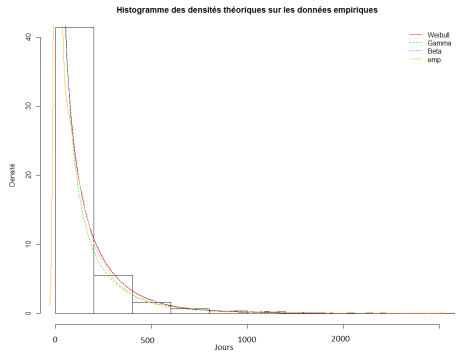

FIGURE 5.7 – Histogramme des données et densités empiriques

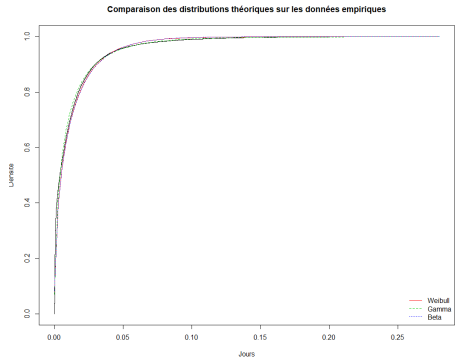

FIGURE 5.8 – Distribution des données observées et des lois testées

Tests statistiques	Gamma	Weibull	Beta
Kolmogorov Smirnoff	0,10	0,09	0,10
Cramer Von Mises	93,22	64,57	94,99
Anderson Darling	612,03	446,21	624,01
AIC	- 305 366,50	- 307 634,50	- 305 131,50

FIGURE 5.9 – Résultats des tests d'adéquation pour les 3 lois candidates

La loi de Weibull est donc la loi choisie pour représenter les délais de déclaration.
Après avoir conclu que la distribution la plus adaptée aux données parmi toutes les options testées est une loi de Weilbull de paramètres $\lambda = 0.6065721$ et $\kappa = 75.1723929$, nous allons mesurer l'incertitude de cette adéquation. L'écart type de l'estimateur correspond à la racine carrée de la variance des estimateurs et permet de créer des intervalles de confiance autour de la valeur des paramètres.
Par ailleurs, en simulant 50 000 observations tirées d'une loi de Weibull suivant ces paramètres et en comparant leur densité avec celle des données de la base, on voit que la loi n'est pas très bien ajustée au début (entre 0 et 365jours), ce qui n'est pas très grave car comme nous allons rediscrétiser par la suite pour faire une estimation annuelle, nous ne cherchons pas à avoir un ajustement exact au jours près.

Paramètres	Forme	Echelle
Weibull	0,6070	75,7120
Ecart-type	0,0023	0,6500

FIGURE 5.10 – Paramètres de la loi choisie

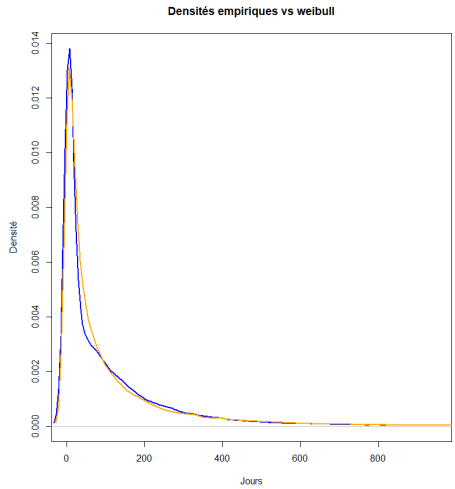

FIGURE 5.11 – Superposition de la densité empirique et d'une densité de Weibull avec les paramètres calculés

5.2.2 Estimation du nombre de sinistres tardifs à l'ultime par la modélisation des délais de déclaration

Après avoir ajusté la loi la plus adéquate et cohérente possible à nos délais de déclaration, nous pouvons l'utiliser afin de prédire le nombre de sinistres tardifs/$IBNyR$.

A l'aide de la distribution des délais de déclaration, il est possible de prédire le nombre de sinistres tardifs pour chaque année de survenance. La période d'observation dans la base est 2009-2018 soit 10 ans de déroulement (pour la validation nous choisirons d'enlever les trois dernières années d'observation au fur et à mesure pour pouvoir comparer les estimations au réel).

Estimation du nombre de sinistres tardifs sur la période 2009-2018

Soit $v(t)$ la densité du nombre des sinistres (inconnue pour le moment)
Soit n_t le nombre de sinistres déclarés dans l'intervalle $[0, t]$.
Soit N_t le nombre total de sinistres déclarés dans l'intervalle $[0, t]$
Soit N_t - n_t le nombre de sinistres tardifs dans l'intervalle $[0, t]$
Soit $F(t)$ la distribution des délais de déclaration des sinistres survenus et déclarés dans l'intervalle $[0, t]$

Pour $v(t)$ connu, le nombre prévu total de sinistres survenant dans l'intervalle $[0, t]$ est :

$$\mathrm{E}(N_t) = \int_0^t v(u)du$$

Années de survenance	Nombre de sinistres déclarés au 31/12/2018
2009	5750
2010	4662
2011	4034
2012	4513
2013	4190
2014	3881
2015	4142
2016	4112
2017	3879
2018	1861
Total	41024

FIGURE 5.12 – Nombre de sinistres observés au 31/12/2018 répartis par année de survenance

Le nombre de sinistres déclarés attendu dans l'intervalle $[0, t]$ est donné par :
:

$$\mathrm{E}(n_t) = \int_0^t v(u)F(t-u)du$$

En pratique, le nombre n_t de sinistres déclarés est connu, alors que le nombre total de sinistres survenus N_t doit être estimé.
On peut donc réécrire $\mathrm{E}(n_t) = n_t$ et par conséquent :

$$\hat{N}_t = \mathrm{E}(N_t) = \frac{\int_0^t v(u)du}{n_t} * n_t \longrightarrow \mathrm{E(N_t)} = \frac{\int_0^t v(u)du}{\int_0^t v(u)F(t-u)du} * n_t$$

Si l'on prend l'hypothèse que la probabilité de survenance d'un sinistre est uniforme sur toute la période alors $\int_0^t v(u)du = \frac{1}{t}$ soit :

$$\hat{N}_t = \mathrm{E}(N_t) = \frac{t}{\int_0^t F(t-u)du} * n_t$$

Le nombre de sinistres tardifs correspond au nombre total de sinistres survenus sur la période moins ceux qui ont été déclarés soit $N_t - n_t$. Il faut donc soustraire n_t de chaque côté de l'équation et :

$$\hat{N}_t - n_t = \frac{t}{\int_0^t F(t-u)du} * n_t - n_t$$

En pratique, le nombre de sinistres déclarés sur la période $[0, t]$ est connu pour des valeurs discrètes de t, dans notre cas t est exprimé en jours.
Alors, la forme discrète de l'équation permettant de calculer le nombre de sinistres non encore déclarés est la suivante :

$$\hat{N}_t - n_t = \frac{t}{\sum_{u=0}^t F(t-u)} * n_t - n_t$$

La formule mathématique a été programmée sur R de manière à avoir une boucle qui calcule le nombre de sinistres tardifs attendu pour chaque année de survenance.

Année de survenance	t	$\sum_{u=0}^{t} F(t-u)$	Nombre de sinistres déclarés	Ultime	Sinistres tardifs
2009	365	365	5750	5 750	-
2010	365	365	4662	4 662	-
2011	365	365	4034	4 034	-
2012	365	365	4513	4 513	-
2013	365	365	4190	4 190	-
2014	365	365	3881	3 881	-
2015	365	364,5547	4142	4 147	5
2016	365	361,8932	4112	4 147	35
2017	365	351,7312	3879	4 025	146
2018	365	275,0119	1861	2 470	609
			41024	41820	796

FIGURE 5.13 – Sinistres tardifs survenus sur la période 2009-2018

Soit les délais de déclaration qui suivent une loi de Weilbull avec $F(t) = 1 - \exp -(\frac{x}{\lambda})^{\kappa}$ la fonction de distribution avec $\lambda = 0.6065721$ et $\kappa = 75.1723929$

Les sinistres tardifs modélisés apparaissent à partir de l'année 2015 ce qui est cohérent avec les délais de déclaration que nous observons dans la base. Pour les années de survenance 2009 - 2014, aucun sinistre tardif n'est modélisé, ce qui est différent de l'approche de Schineper qui recense des sinistres tardifs (réouvertures) jusqu'à 7 ans après la survenance. Le modèle développé à partir de l'approche de Pietro Parodi a du sens étant donné que la quasi-majorité des sinistres sont déclarés dans les deux ans mais ne permet pas d'aller chercher les observations plus rares. Les évènements où les sinistres sont déclarés très tardivement sont extrêmement rares et il est possible qu'il ne soit pas capté par la queue de la loi Weibull. Une piste d'amélioration ici serait d'utiliser la théorie des valeurs extrêmes afin d'attribuer une distribution spécifique aux délais très élevés.

Pour chaque année de survenance, le nombre ultime de sinistres déclarés est déterminé en sommant le nombre de sinistres qui sont déjà déclarés et connus de l'assureur avec le nombre de sinistres projetés.

Nous pouvons comparer les sinistres projetés par l'approche de Parodi avec ceux obtenus avec la méthode de Schnieper. Bien qu'elle soit très sensible au choix de l'exposition pour chaque année de survenance, la méthode de Schnieper permet de différencier les nouveaux sinistres et les sinistres sans-suite ce qui présente un réel avantage. Dans les calculs, les nouveaux sinistres sont mesurés par rapport à l'exposition alors que les sans suite sont mesurés par rapport aux nombres de sinistres précédemment observés ou estimés (incluant donc les nouveaux sinistres estimés). La méthode de Schnieper est donc très dépendante des informations sur l'exposition. On voit ci-dessous la sensibilité de l'approche de Schnieper vis-à-vis de l'exposition. Les résultats obtenus par la méthode de P.Parodi est assez proche des résultats de Schnieper quand on considère une baisse de l'exposition similaire à celle observée sur les déclarations de nouveaux sinistres en première année d'ouverture.

Dans notre application de Parodi, nous n'avons pas tenu compte des sinistres qui vont être potentiellement déclarés sans suite. Cela présente un axe d'amélioration à apporter au calcul car si nous sommes dans une optique de projection des sinistres tardifs, certains d'entre eux seront classés sans suite et donc ne donneront pas lieu à un sinistre pour l'assureur. Ainsi, il serait préférable de préparer les données en tenant compte de la proportion des sinistres qui pourraient être classés sans suite afin d'apporter à l'approche de Parodi cette même vision que l'on peut retrouver dans la méthode de Schnieper. Les tardifs seraient ainsi projetés en prenant la base de données complète et les sans suite en seraient déduits par la suite.

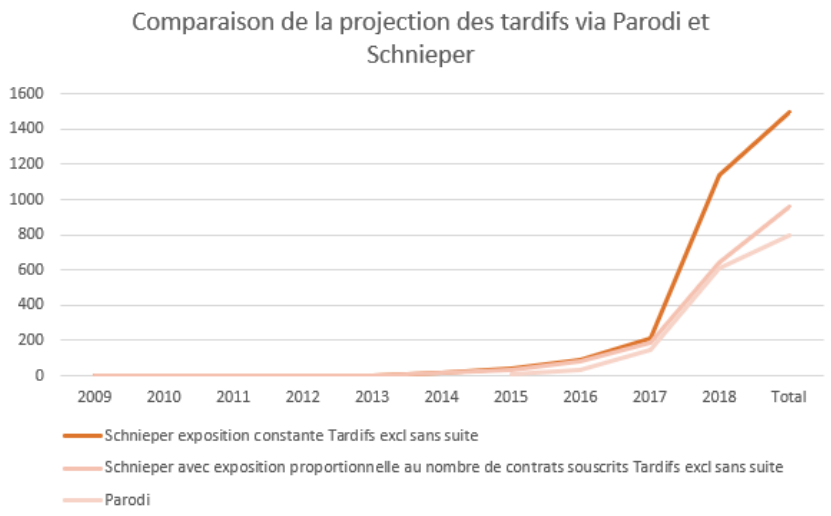

FIGURE 5.14 – Comparaison des sinistres tardifs obtenus avec Schnieper et avec Parodi

5.2.3 Obtention des paramètres de la loi pour le nombre des sinistres

Nous avons utilisé la distribution des délais de déclaration pour projeter le nombre de sinistres tardifs pour chaque année de survenance. Nous allons pouvoir maintenant dériver la loi des nombres pour les tardifs.

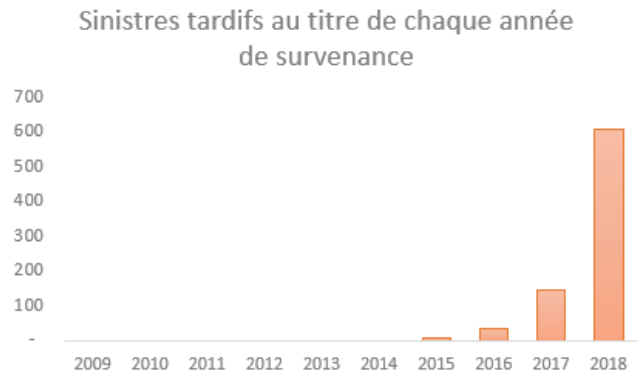

FIGURE 5.15 – Sinistres tardifs par année de survenance

Le choix de la distribution des nombres se fait entre la loi de Poisson et la loi Négative Binomiale. C'est en analysant la moyenne et la variance qu'il est possible de choisir la plus adéquate.

La distribution de Poisson

La loi de Poisson est généralement utilisée en assurance non-vie pour modéliser le nombre d'évènements dans une fenêtre de temps prédéterminée. Le paramètre λ de la loi de Poisson définie à la fois l'espérance et la variance de la distribution. En assurance non-vie, cette distribution est utilisée si la moyenne et la variance des données observée sont proches.

Distribution Négative Binomiale

La loi Négative Binomiale possède deux paramètres, $r \in [0, \infty)$ et $p \in [0, 1]$, qui la rendent plus flexible que la loi de Poisson. Elle représente un choix approprié lorsque la variance est plus élevée que la moyenne. Elle est utilisée pour modéliser les fréquences qui sont sujettes à un haut niveau de variabilité, car elle permet de prendre en compte la dispersion des données. Dans notre cas, la variance est bien plus élevée que la moyenne, justifiant donc l'utilisation de cette distribution.
Nous supposons donc que les nombres des *IBNyR* suit une loi Négative Binomiale.

Détermination des paramètres de la loi Négative Binomiale

Les deux paramètres de la distribution Négative Binomiale sont $r \in [0, \infty)$ et $p \in [0, 1]$ où :

— Le ratio variance/moyenne, calculé sur le nombre de sinistres ultimes (nombre de sinistres observés + nombre de sinistres tardifs projetés) correspond à $1 + r$. Par conséquent,

$$r = \frac{Variance_{NombresUltimes}}{Moyenne_{NombresUltimes}} - 1$$

.

— Le ratio effectif Variance/Moyenne permet de prendre en compte l'incertitude liée à l'estimation des paramètres. En règle générale, il suffit d'ajouter le terme $\frac{1}{n}$ où n est le nombre d'années sur laquelle la variance est calculée, dans notre cas 10 années.

— L'espérance de la loi Négative Binomiale est dénotée par $\lambda = \frac{rp}{(1-p)}$.
Par conséquent

$$p = \frac{\lambda}{r + \lambda}$$

λ correspond au nombre de sinistres tardifs projetés sur la période considérée au titre de chaque année de survenance.

En théorie, si X suit une loi Binomiale Négative de paramètres r et p, alors X correspond au nombre d'échecs précédant le r^{ieme} succès et p représente la probabilité d'obtenir un "succès". Nous pouvons assimiler le succès à l'obtention d'un sinistre tardif et p correspond à la probabilité de voir un sinistre tardif déclaré sur la période.

Le nombre de sinistres tardifs suit donc une loi Négative Binomiale de paramètres $p = 0.836$ et $r = 154.94$. Nous utilisons les paramètres issus de l'agrégation des sinistres tardifs pour toutes les années de survenance. Les prédictions par année individuelle introduisent un risque d'erreur assez élevé sur les années les plus récentes dû au manque d'historique les concernant. Ce problème revient aussi lors de l'implémentation de la méthode Chain Ladder. Cette incertitude liée au manque d'historique peut être rectifiée par des méthodes de crédibilité qui ne sont pas traitées dans ce mémoire. La théorie de la crédibilité consiste à prendre en compte des sources d'informations externes (le développement d'autres lignes d'activité par exemple) lors du provisionnement.

5.3 Application (2) : Modélisation de la charge des sinistres tardifs

L'objectif ici est d'associer un montant de charge aux nombres de sinistres tardifs qui ont été estimés dans la partie précédente afin de pouvoir ensuite déterminer la réserve IBNyR en combinant les deux.

Année de survenance	Ultime	Nombre projetés de sinistres IBNyR (λ)	r	p
2009	5750	0	154,94	0,000
2010	4662	0	154,94	0,000
2011	4034	0	154,94	0,000
2012	4513	0	154,94	0,000
2013	4190	0	154,94	0,000
2014	3881	0	154,94	0,000
2015	4147	5	154,94	0,032
2016	4147	35	154,94	0,186
2017	4025	146	154,94	0,486
2018	2470	609	154,94	0,797
Agregé	**41 820**	**796**	**154,94**	**0,836**
Moyenne	**4 182**			
Variance	**652 139**			
Ratio Variance Moyenne	**155,94**			

FIGURE 5.16 – Dérivation des paramètres de la loi Négative Binomiale

Il est important de différencier les sinistres clos et des sinistres en cours. En effet, les sinistres en cours sont sujets à des IBNeR car ils ne sont pas encore définitifs, alors que les sinistres clos, quant à eux, ne sont plus soumis à variation hormis les cas de réouvertures.
Nous rappelons qu'en cas de réouverture, du fait de la manière dont la base est constituée, un sinistre réouvert est considéré comme un nouveau sinistre.

La charge moyenne est de 7 436,77€ pour les sinistres en cours et de 10 130,21€ pour les sinistres clos. On observe que la charge moyenne est supérieure pour les sinistres clos sur l'ensemble des années sauf pour les deux années les plus récentes. Cela s'explique par le fait que les deux dernières années sont très peu développées et que les sinistres qui sont clos pour les années très récentes ont été réglés ou classés sans suite très vite et sont très probablement peu coûteux, ce qui explique les faibles niveaux de charges.

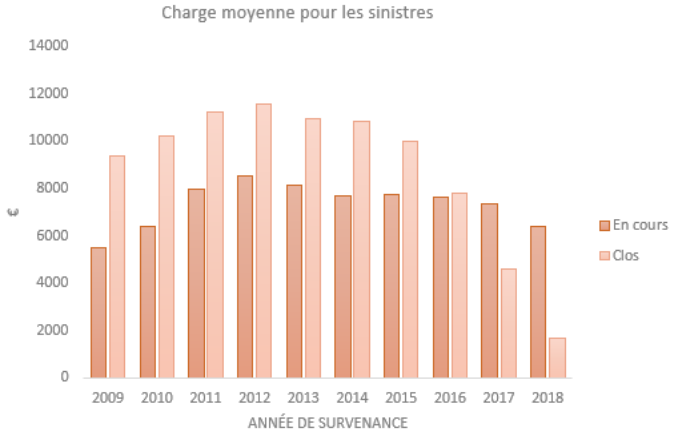

FIGURE 5.17 – Charge moyenne au titre de chaque année de survenance selon le statut du sinistre au 31/12/2018

5.3.1 Détermination des ultimes pour les sinistres clos

Nous connaissons le montant ultime des sinistres clos dans la base et nous voulons nous servir de cette information pour estimer le montant des sinistres clos pour les sinistres tardifs qui surviendront à partir du 1er janvier 2019.
Les sinistres clos observés ont été payés entre leur année de déclaration et le 31/12/2018 ; leur charge est donc une somme des montants payés avec différentes dates de valeurs (par exemple, pour une déclaration en 2009, la charge est composée des montants payés en 2009,2010, 2011 jusqu'à 2018). Ainsi, selon leur date de déclaration et leur cadence de paiement, les montants ultimes clos ne sont pas homogènes et par conséquent pas comparables entre eux.
Pour les *as-ifés*, afin de servir à notre modélisation des tardifs, nous aurions besoin de connaître deux informations : la date de déclaration des tardifs et leur cadence de liquidation à utiliser dans la modélisation. Ces aspects n'ont pas été traités dans le papier de P.Parodi.
Nos tardifs se décomposent en deux catégories :

- D'une part, les "vrais tardifs", au sens de la RC automobile corporelle, c'est à dire les déclarations dans les deux ans suivant la survenance, ce qui dans notre cas ne concerne que les survenances 2017 et 2018.
- D'autre part les réouvertures au sens de la RC automobile corporelle, c'est à dire les réouvertures après deux ans. Cela concerne toutes les années de survenance de notre échantillon.

Dans le premier cas, les dates de déclaration sont très concentrées mais dans le cas des réouvertures, la plage des délais de déclaration est bien plus large. Nous sommes également limités par notre faible connaissance de l'inflation des sinistres passée et prospective à considérer afin de déterminer la meilleure estimation possible de l'ultime des sinistres clos tardifs. Il nous faudrait enfin également réaliser une étude sur les cadences de règlements des sinistres clos pour les deux catégories de tardifs susmentionnés.
Compte tenu du contexte très particulier de la réalisation de ce mémoire, nous n'avons pas pu nous pencher très en détail sur ce point et avons juste retenu une approche extrêmement simplifiée pour considérer l'inflation. Nous avons donc pris la moyenne des taux d'inflation de 2009 à 2018 et l'avons appliqué sur les sinistres clos via la formule suivante : $Charges_{post/inflation} = Charges_{sinistreclos} * (1+i)^{(2018-datedecloture)}$ avec i l'inflation choisie.

5.3.2 Prise en compte des sinistres en cours / IBNER

L'idée pour parer au problème de l'incertitude des charges des sinistres en cours est d'identifier les éventuels surestimations ou sous-estimations des charges et de les rectifier. Pour cela, nous comparons les charges moyennes pour les sinistres en cours et pour les sinistres clos au titre de chaque année de survenance :

AJUSTEMENT POUR LES SINISTRES INCURRED BUT NOT ENOUGH REPORTED

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
EN COURS	5488,22529	6406,13636	7999,27922	8512,94464	8124,54183	7699,9208	7750,44683	7644,41601	7371,39135	6427,00411
CLOS	9371,46202	10236,9506	11254,8833	11595,377	10933,3666	10819,3136	10001,3128	7820,56374	4627,16271	1675,57143
Ratio	1,70755782	1,59799137	1,40698718	1,36208768	1,34572101	1,40512012	1,29041758	1,02304267	0,62771904	0,26070801

FIGURE 5.18 – Ratios pour le réajustement des sinistres en cours

Ce tableau montre que pour les sinistres survenus en 2009, les charges des sinistres clos sont 70% plus élevées que les charges des sinistres qui sont encore en cours. La même interprétation peut être faite pour chaque année de survenance et il nous est possible de réajuster les montants des sinistres en cours grâce

à ses résultats. Le fait de procéder ainsi comporte une limite pour les exercices récents car cela amène à considérer un montant moyen relativement faible, qui est expliqué par la duration très faible des sinistres observés qui sont a priori des sinistres de faibles sévérités.

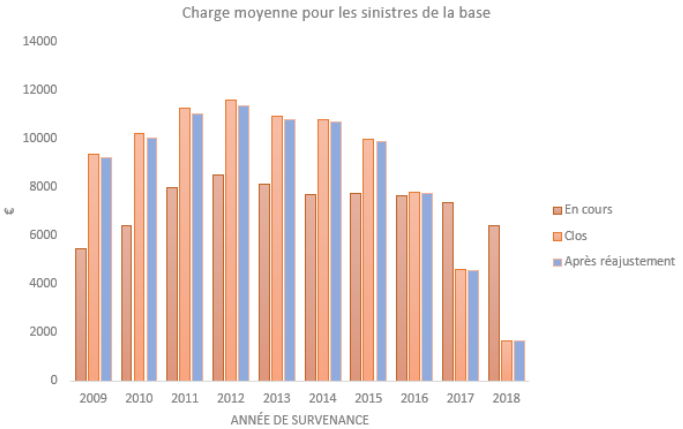

FIGURE 5.19 – Charge moyenne des sinistres après réajustement pour les sinistres en cours (en bleu)

Nous avons également considéré deux options alternatives pour amener les sinistres en suspens à l'ultime : nous pourrions reprendre les coefficients de développement de la méthode Chain Ladder sur l'intégralité du triangle et les appliquer au triangle des sinistres en cours, ou alors procéder à un Chain Ladder sur les charges des sinistres en cours du 31/12/2018. Nous n'avons pas pu réaliser ces études.

5.3.3 La loi des distributions de la charge est-elle commune à toutes les années de survenance ?

L'une des difficultés majeures dans l'obtention d'une distribution statistique pour les charges découle du fait que les sinistres soient tous rattachés à des années de survenance différentes.
Il est effectivement possible que chaque année de survenance présente une distribution bien spécifique pour les charges des sinistres survenus à cette période.
Il existe trois éléments qui expliquent le lien qu'il pourrait y avoir entre l'année de survenance et la distribution des charges :

- Le mix du portefeuille peut changer d'une année à l'autre. Dans le cadre de nos travaux, n'ayant pas d'information sur le portefeuille, nous supposons que la proportion au niveau des souscriptions pour ce type de garantie est stable sur l'historique.
- La nature, la fréquence et la gravité des risques peuvent évoluer suite à des changements technologiques et environnementaux.
- L'inflation - le coût des sinistres les plus récents ne sont pas comparables à ceux des sinistres plus anciens dû à l'inflation, phénomène économique et persistant qui fait monter l'ensemble des charges.

Ici, nous supposons que le mix des contrats souscrits est stable sur l'historique et qu'il n'y a pas eu de changements environnementaux majeurs au cours de la période d'observation. La problématique sur l'inflation a été abordée dans la partie précédente.
Néanmoins, il est possible d'étudier l'évolution des charges pour chaque année de survenance graphiquement. Après avoir ajusté le montant des charges pour les sinistres en cours et pour l'inflation, nous avons tracé la densité empirique de la variable pour chaque année de survenance.

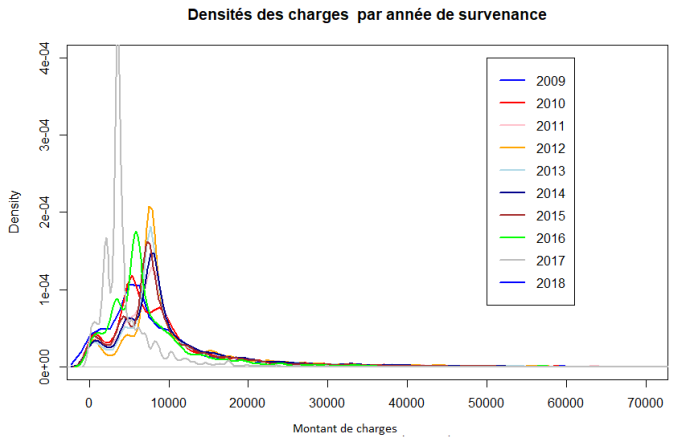

FIGURE 5.20 – Densité empirique des charges pour chaque année de survenance

Seule l'année 2017 est vraiment atypique, mais là encore nous n'avons pas d'information pour savoir si cela est lié à nos différentes hypothèses ou si cela est dû au portefeuille. La densité pour les années les plus récentes est un peu plus instable car nous manquons de données et d'historique. Nous pensons qu'il est raisonnable d'ajuster une loi statistique commune à toutes les années de survenance également afin d'éviter le sur-paramétrage de nos modèles.

5.3.4 Ajustement d'une loi statistique aux montants des charges

Après avoir ajuster les données des charges pour prendre en compte les *IBNER*, l'inflation et après avoir valider visuellement notre hypothèse selon laquelle la loi utilisée pour représenter les charges est commune à toutes les années de survenance, nous pouvons implémenter la même méthode d'ajustement de lois statistiques que celle présentée pour les délais de déclaration. Dans un premier temps, nous ré-évaluons le graphique *Cullen Frey* qui a changé après les différents ajustements effectués.

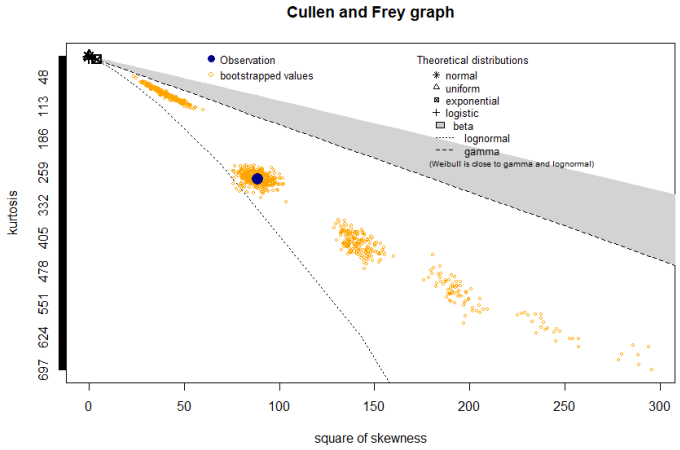

FIGURE 5.21 – Graphique Cullen Frey des charges après réajustements

Les lois qui nous semblent intéressantes de tester sont : la loi weibull, gamma et lognormale.

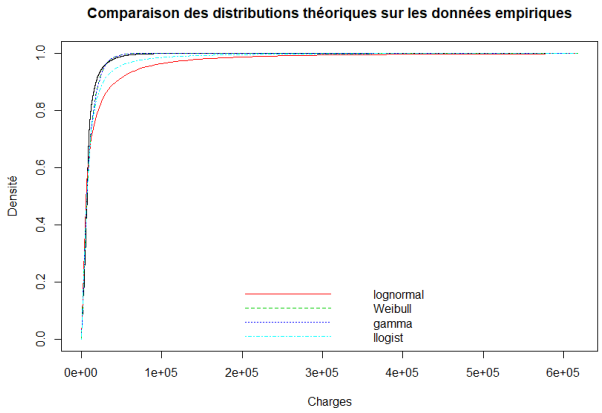

FIGURE 5.22 – Comparaison de la distribution des données empiriques et de la distribution des lois testées

Après analyse des statistiques descriptives et du résultat du *Goodness Of Fit*, il en ressort que la loi Weibull avec les paramètres $\lambda = 0.9862$ et $\kappa = 9208.992$ est celle qui coïncide le mieux possible avec les données empiriques. Il faut dorénavant combiner la distribution des nombres Négative Binomiale et la distribution des charges pour obtenir une distribution agrégée des sinistres tardifs.

Test statistique	Lognormale	Weibull	Gamma	Log-logistique
Kolmogorv-Smirnoff	0,2154	0,1221	0,1252	0,098
Cramer Von Mises	598,0236	185,3205	193,0419	154,246
Anderson Darling	3368,1257			1103,639
AIC	858278,9	831476,6	831451,1	837245,6

FIGURE 5.23 – Tests d'adéquation des lois statistiques

5.4 Application (3) : mise en commun des nombres et de la charge des sinistres tardifs

En simulant le nombre de sinistres tardifs un certain N_{sim} nombre de fois et en allouant à chaque fois un montant de charges au titre de ces sinistres tardifs, nous sommes capables d'obtenir la distribution de la réserve *IBNyR*.
En effet, grâce aux propriétés de la méthode de Monte Carlo qui consiste à simuler une suite d'observations identiquement distribuées toutes issues de la distribution totale de la charge des sinistres tardifs puis d'en estimer son espérance par la moyenne empirique, nous sommes capables d'obtenir une estimation du montant de cette réserve.

5.4.1 Déroulement de la simulation pour l'estimation de la charge des sinistres tardifs

Sans distinction des années de survenance, le protocole est le suivant :

— **Etape 1** : Créer un vecteur composé de 100 000 éléments, tous initialisés à zéro qui correspond aux nombres de scénarios qui forment la simulation Monte Carlo. Plus le nombre de scénarios est élevé, plus le temps de calcul est long mais plus l'erreur est faible.

— **Etape 2** : Pour chaque scénario allant de $j = 1$ à 100000, simuler un nombre n_j correspondant au nombre de sinistres tardifs tirés d'une distribution Négative Binomiale de paramètres $r = 154.94$ et $p = 0.836$
— **Etape 3** : Pour chaque sinistre allant de 1 à n_j, on simule le montant de la charge x_i^j tirée d'une loi Weibull de paramètre $\lambda = 0.9862$ et $\kappa = 9208.992$
— **Etape 4** : Pour chaque simulation j, la charge totale au titre des n_j sinistres tardifs est calculée par la somme des charges individuelles des n_j sinistres tel que $S_j = \sum_{i=1}^{n_j} x_i^j$
— **Etape 5** : La distribution **totale** des *IBNyR* correspond à la distribution empirique des montants S_j obtenus pour les 100 000 scénarios.
— **Etape 6** : La moyenne empirique de la distribution des S_j correspond au montant total dû au titre des sinistres IBNyR au 31/12/2018.

5.4.2 Résultats de la simulation

Nous effectuons la simulation de deux manières différentes :

Cas 1 : les nombres des sinistres tardifs agrégée sur toute la période

En appliquant ce protocole sur R, nous obtenons que la réserve devant être placée au titre des sinistres tardifs sur la période 2009-2018 est de 7 374 157€. Ce montant est obtenu en prenant la moyenne empirique de toutes les simulations. Les quantiles de la distribution découlant des 100 000 simulations sont affichés ci-dessous :

Pourcentiles	Valeur
0,995	9 288 548 €
0,990	9 092 859 €
0,900	8 281 126 €
0,750	7 832 241 €
0,500	7 355 662 €
0,250	6 894 182 €

FIGURE 5.24 – Quantiles de la distribution

Cas 2 : les nombres des sinistres tardifs prise en compte individuellement pour chaque année de survenance

En appliquant le protocole sur R en changeant les paramètres de la loi des nombres propres à chaque année de survenance, nous obtenons que la réserve devant être placée au titre des sinistres tardifs sur la période 2009-2018 est de 7 363 773€. Bien que l'écart soit très faible dans notre cas, nous jugeons l'utilisation de lois spécifiques à chaque année de survenance très incertaine et à utiliser avec prudence, les années les plus récentes étant très instables dû au manque d'information.

Chapitre 6

Obtention du montant de la réserve au titre des sinistres déclarés encore en cours

6.1 Contexte

Cette partie a pour but de venir compléter le calcul de la réserve IBNR afin de pouvoir valider les résultats.
Pour rappel, la réserve *Incurred But Not Enough Reported*, composante de la Provision Pour Sinistres A Payer est composée de deux choses :

- La partie due au titre des sinistres *IBNyR*[2] qui sont les sinistres survenus mais inconnus de l'assureur.
- La partie due au titre des sinistres *IBNeR*[3] qui sont déjà connus de l'assureur mais qui vise à compléter l'insuffisance potentielle des provisions dossier/dossier placée par le gestionnaire sinistre.

Dans ce mémoire, nous proposons une méthode alternative aux méthodes usuelles en distinguant le calcul des deux composantes.
Dans le chapitre 5, nous avons proposé une approche fréquence/coût pour projeter la charge à placer au titre des sinistres tardifs sur la période 2009-2018. Comme nous l'avons expliqué, cette approche permet de prendre en compte les caractéristiques individuelles des sinistres du portefeuille et d'offrir une estimation du nombre et de la charge des sinistres inconnus en étudiant l'impact des délais de déclaration et du statut des sinistres.

Pour l'estimation de la partie *IBNeR*, les sinistres sont déjà déclarés à l'assureur et nous avons déjà l'information quant à la charge placée jusqu'au 31/12/2018. La projection de ces sinistres à l'ultime se fera donc avec la méthode Chain Ladder. Il existe d'autres méthodes de calcul pour cette composante, tels les modèles linéaires généralisés où les facteurs de développement des charges des sinistres *IBNeR* sont modélisés par une combinaison linéaire de plusieurs variables explicatives. Une fois modélisés, ces derniers sont utilisés pour projeter les charges des sinistres en cours jusqu'à l'ultime. Le détail de cette méthode n'est pas traité dans ce mémoire car nous n'avons pas assez de variables explicatives dans la base.

2. *Incurred But Not Yet Reported*
3. *Incurred But Not Enough Reported*

6.2 Implémentation de Chain Ladder sur les sinistres en cours

Pour le calcul de la réserve au titre des sinistres *IBNeR*, nous avons reconstitué le triangle des charges sur les sinistres qui sont en cours à la date d'évaluation en veillant bien à ne pas prendre en compte de nouveaux sinistres d'une année de développement à l'autre.
L'année de rattachement ici est l'année de déclaration ce qui assure que la première colonne est la première année de vue de chaque sinistre et que des nouveaux sinistres ne sont pas introduits au cours de la construction du triangle. Nous avons déjà une vision du développement de la charge placée pour les sinistres en cours selon leur année de déclaration et continuer de développer cette même cadence ne nous paraît pas aberrant. Cette logique, en plus du fait que les hypothèses soient respectées, justifient ainsi l'utilisation de Chain Ladder pour cette partie.

	Années de développement									
Survenance	0	1	2	3	4	5	6	7	8	9
2009	4 315 074	5 662 927	6 952 301	7 595 574	8 253 566	8 576 683	8 634 905	8 284 186	8 325 257	8 108 664
2010	6 668 937	8 763 137	9 660 717	10 438 966	10 872 858	11 068 845	11 283 530	10 752 821	10 575 784	10 300 640
2011	6 750 257	9 277 136	10 277 364	10 738 330	11 185 603	11 537 550	11 195 483	11 185 076	11 105 190	10 816 273
2012	13 255 614	17 086 754	18 991 629	20 062 158	20 727 623	21 028 195	21 333 046	20 721 565	20 573 568	20 038 318
2013	11 591 989	15 788 351	17 717 754	18 991 679	19 867 506	20 452 436	20 544 762	19 955 876	19 813 348	19 297 876
2014	11 280 646	14 622 280	16 626 697	18 150 390	18 646 168	19 108 082	19 194 339	18 644 161	18 511 001	18 029 411
2015	11 071 950	15 742 735	18 372 708	19 708 984	20 528 781	21 037 332	21 132 298	20 526 571	20 379 967	19 849 753
2016	13 010 248	19 127 614	22 393 143	24 002 673	25 001 067	25 620 408	25 736 062	24 998 375	24 819 832	24 174 109
2017	17 224 926	25 532 778	29 124 566	31 217 923	32 516 436	33 321 952	33 472 373	32 512 936	32 280 722	31 440 894
2018	19 151 264	26 482 999	30 208 457	32 379 720	33 726 558	34 562 052	34 718 071	33 722 927	33 482 072	32 610 989
	1,38	1,14	1,07	1,04	1,02	1,00	0,97	0,99	0,97	

FIGURE 6.1 – Triangle de charges cumulées pour les sinistres en cours dans le portefeuille

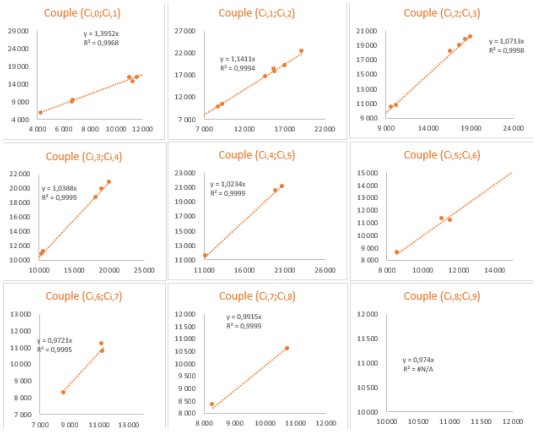

FIGURE 6.2 – Vérification des hypothèses de Chain Ladder

Année de survenance	Ultime	Diagonale (montants connus)	Réserve
2009	8 108 664	8 108 664	-
2010	10 300 640	10 575 784	- 275 144
2011	10 816 273	11 185 076	- 368 803
2012	20 038 318	21 333 046	- 1 294 728
2013	19 297 876	20 452 436	- 1 154 561
2014	18 029 411	18 646 168	- 616 757
2015	19 849 753	19 708 984	140 769
2016	24 174 109	22 393 143	1 780 966
2017	31 440 894	25 532 778	5 908 116
2018	32 610 989	19 151 264	13 459 725
	194 666 926	177 087 344	17 579 583

FIGURE 6.3 – Synthèse des résultats pour la période 2009-2018

La réserve devant être placée au titre des sinistres en cours est de 17 579 583€. Nous observons un rattrapage sur les années plus anciennes (2010-2014) dû à un excédent de provision dossier/dossier. Cette provision est plus élevée que celle au titre des sinistres tardifs car le nombre de sinistres en cours est bien plus élevé que les sinistres tardifs attendus pour la période d'évaluation.

6.3 Montant de la réserve IBNR

Par ailleurs, la réserve IBNR est la somme de la réserve pour les sinistres *IBNyR* et de la réserve pour les sinistres *IBNeR* soit 7 374 157€ + 17 579 583€ qui équivaut à 24 953 740€. Nous décidons de ne pas utiliser le montant obtenu avec les calculs effectués sur les années de survenance individuelles car nous risquerions d'introduire de l'erreur à cause de l'instabilité numérique des années les plus récentes (dû au manque d'historique de données). Nous effectuons les mêmes calculs pour les 3 années les plus récentes de l'historique car dans le chapitre suivant, nous confronterons ces résultats aux résultats obtenus en appliquant Chain Ladder et aux données réellement observées dans la base.

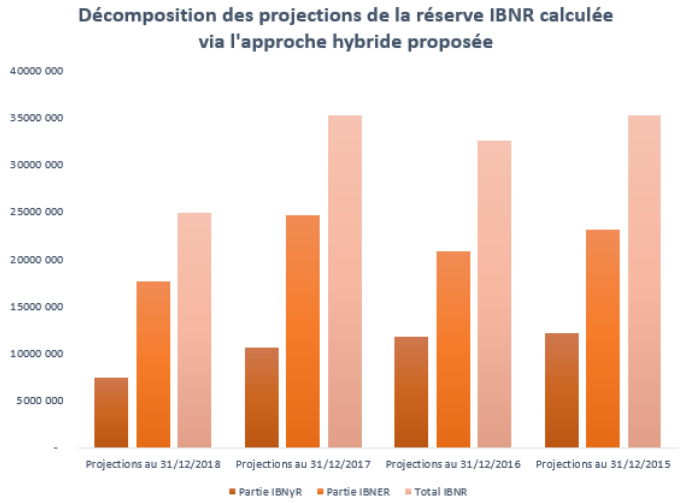

FIGURE 6.4 – Résultats pour les projections effectuées au 31/12 pour les 4 dernières années d'historique

Chapitre 7

Validation de la méthode

Pour valider l'approche fréquence/coût au calcul de la réserve IBNR, nous allons comparons les diagonales observées avec les diagonales projetées par la méthode Chain Ladder et la méthode proposée inspirée de travaux de P.Parodi.

En pratique, valider une méthode est une étape fastidieuse car cela nécessite d'analyser sa performance sur plusieurs portefeuilles de données et ce sur plusieurs années afin de vraiment pouvoir la comparer aux données observées. Dans cette partie, nous avons fait au mieux en retirant une par une les trois dernières années d'observation de la base. Ainsi, nous avons pu comparer ce qui a été observé avec les projections de la méthode Chain Ladder et de la méthode présentée. Cela nous permet alors de voir si une méthode qui distingue les *IBNyR* des *IBNeR* améliore les estimations.
Les projections au 31/12/2018 ne sont pas présentées dans cette partie car nous n'avons pas à disposition ce qui a été observé au 31/12/2019, ce qui ne permet pas d'effectuer une comparaison réel/projeté.

Une limite importante dans notre validation réside dans le fait que la méthode mise en place dans ces travaux ne comporte pas de cadence de liquidation pour les tardifs et les réouvertures. En effet, nous ne savons pas répartir les charges par diagonale pour la partie des tardifs. Ainsi, les chiffres ci-dessous surestiment les diagonales en partie à cause des projections des tardifs (pas sur la partie IBNeR qui, elle, est en Chain Ladder). Ainsi, nous validons cette approche en considérant que les tardifs seraient tous déclarés sur la diagonale projetée et réglés en totalité dès leur survenance.
Nous voyons bien ici l'importance de faire une analyse exhaustive des données et notamment de travailler avec des cadences de liquidation et de règlement.

Concernant les résultats chiffrés, de manière générale, on voit que la méthode Chain Ladder donne des estimations inférieures aux montants réellement observés alors que la méthode prenant en compte les tardifs via une approche fréquence / sévérité donne au contraire des montants supérieurs aux données observées (cohérent avec la remarque préliminaire ci-dessus).

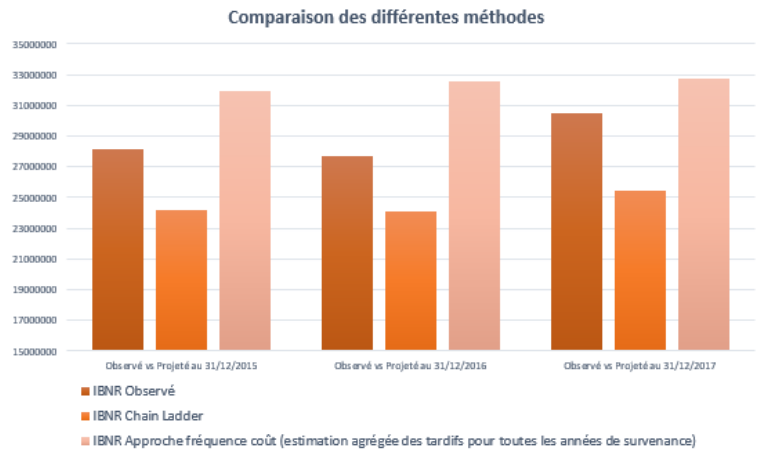

FIGURE 7.1 – Comparaison des projections avec la charge placée observée aux dates d'arrêtés 31/12/2015, 2016 et 2017

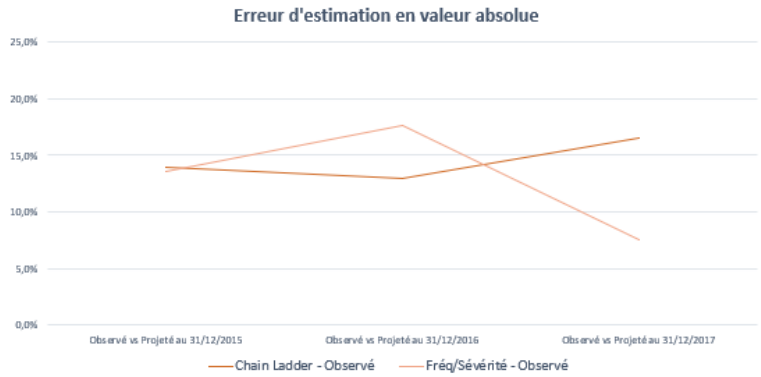

FIGURE 7.2 – Erreur d'estimation en % associée à chaque méthode

— Le montant observé au 31/12/2016 est plus élevé que les projections de Chain Ladder effectuée au 31/12/2015 (+13%]. Cette erreur d'estimation de Chain Ladder peut s'expliquer par un manque d'historique et par conséquent des coefficients de développement instables qui ne répliquent pas bien la vraie cadence. La méthode fréquence/sévérité donne une estimation plus élevée que Chain Ladder et est justifié par la remarque ci-dessus.
— Lorsqu'on compare les projections effectuées au 31/12/2016 avec ce qui a été observé entre 2016 et 2017, la méthode Chain Ladder est plus proche mais continue de sous-estimer les résultats par rapport aux charges observées. Ce problème récurrent semble être dû aux coefficients de développement des années les plus récentes qui sont inférieurs à 1. Compte tenu du fait que ces facteurs ne sont basés que sur deux valeurs, leur pertinence est fortement questionnable et est potentiellement la raison de cette sous-estimation.
— Pour les projections du 31/12/2017 sur l'année 2018, l'approche fréquence/sévérité est la plus proche des données observées.

Les résultats observés avec les chiffres obtenus par l'adaptation de P.Parodi sont cohérents avec nos hypothèses qui mériteraient bien sûr un approfondissement mais qui malgré tout semblent prometteurs pour améliorer les modèles utilisés en provisionnement.

Les performances de la méthode de Chain Ladder et de l'approche fréquence / sévérité sont toutes deux peu satisfaisantes, l'une sous-estimant d'en moyenne 14% le montant d'IBNR à placer et l'autre le sur-estimant du même pourcentage. Nous pourrions nous attendre à ce que l'erreur de la méthode Chain Ladder diminue avec l'allongement de l'historique des données, ce qui n'est pas le cas avec trois années d'observation supplémentaires.

Nous avons vu durant nos analyses que considérer l'inflation et travailler avec les facteurs de développement sur des données inflatées permettrait d'isoler la composante inflation de la composante aggravation / amélioration et ainsi mieux estimer la charge allouée par Chain Ladder. Cela génère une estimation plus élevée que la réalité à deux niveaux : lors de l'estimation des ultimes d'une part et également au niveau de l'estimation des *IBNeR* d'autre part.
En ce qui concerne la performance de la méthode de P.Parodi, il est difficile d'en tirer des conclusions pertinentes car dans le cadre de nos travaux, cette dernière repose sur beaucoup d'hypothèses (prise en compte du taux d'inflation dans les projections, prise en compte des sinistres sans suite, calibration d'une seule et même loi pour les sinistres tardifs et les réouvertures, absence de cadence de liquidation et de paiement...). Notre méthodologie ne considère pas non plus que les sinistres tardifs puissent être sans suite, ce qui constitue un axe d'amélioration majeur. Cependant, nous pouvons voir que malgré les améliorations nécessaires quant aux hypothèses utilisées, elle s'avère très intéressante car elle se base sur une meilleure connaissance et compréhension des données traitées et permet une granularité plus fine dans le modèle.

Si l'on se concentre sur la projection du nombre de sinistres tardifs, ils sont relativement proches du nombre de sinistres réellement déclarés.

FIGURE 7.3 – Erreur dans la prévision du nombre de sinistres tardifs

Nous avons également voulu savoir si utiliser la distribution empirique des délais de déclaration entraînerait un changement dans les prédictions. Pour rappel, nous avions trouvé qu'une loi de Weibull était la plus adaptée pour représenter les délais de déclaration. En utilisant la distribution empirique des délais, nous avons refait tous les calculs jusqu'à obtention de la réserve IBNR finale et les résultats sont les suivants :

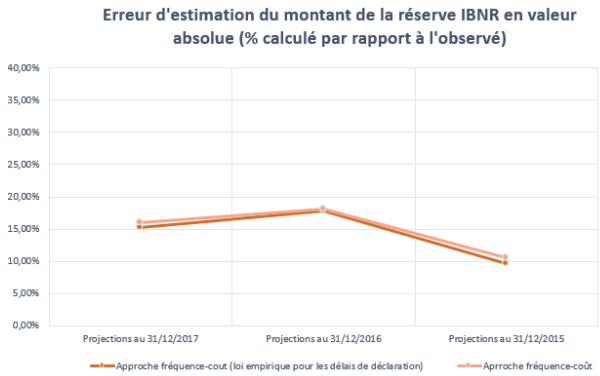

FIGURE 7.4 – Ecart avec l'observé selon l'utilisation ou non de la distribution empirique des délais de déclaration

L'écart entre les deux résultats est extrêmement faible. L'utilisation de la distribution empirique semble améliorer légèrement les prévisions mais il faut être très prudent dans ce genre de pratique car la distribution empirique sera forcément biaisée par les délais courts qui sont majoritaires dans la base (ce qui explique que les résultats sont très similaires à la loi de Weibull qui capture très mal les délais les plus élevés) et ne pourra jamais prendre en compte des délais de déclaration plus longs que l'historique de données utilisé. Aussi, cela confirme le fait que l'erreur réside très probablement dans la loi choisie pour représenter les charges qui demandent une analyse plus approfondie.

Conclusion

Nous pouvons conclure en résumant les avantages et les inconvénients de la méthode et en présentant les axes d'améliorations que nous avons noté au cours de ce mémoire.
La méthode présente de nombreux avantages : tout d'abord, il semblerait qu'une approche coût/fréquence pour estimer les IBNyR soit une approche intéressante en provisionnement, surtout dans des domaines comme la Responsabilité Civile Médicale où il y a un délai potentiellement important entre la survenance du dommage et sa réclamation. Ce type de modélisation offre un cadre de calcul qui se rapproche de ce que l'on peut retrouver en tarification et propose ainsi une approche cohérente pour la tarification et le provisionnement.
L'approche proposée est plus complète que la méthode Chain Ladder, mais les projections semblent surestimer la charge ultime et nécessite de réelles améliorations sur plusieurs aspects :

- Il serait préférable de mieux prendre en compte l'effet de l'inflation sur les projections en s'appuyant sur des hypothèses et analyses plus robustes. Cela permettrait de n'observer que l'aggravation ou l'amélioration des sinistres sans le biais lié à l'inflation.
- Du fait de la branche considérée (la RC automobile corporelle), nous étions dès le début limité à un nombre de sinistres tardifs très réduit dans le temps (les deux premières années suivant la survenance). Cependant cela nous a donné l'occasion de nous pencher sur le cas des réouvertures de sinistres qui est un réel problème pour le provisionnement des sinistres graves en assurance directe mais également un sujet de tarification au niveau de la réassurance par exemple.
- Afin de modéliser les tardifs et les réouvertures de manière optimale, il aurait fallu disposer des expositions et calibrer deux lois distinctes aux délais des sinistres tardifs d'une part et aux réouvertures d'autre part afin de les utiliser séparément lors de la projection des nombres de sinistres.
- Pour la modélisation des sinistres tardifs, il faudrait approfondir le fait que des sinistres tardifs peuvent aussi donner lieu à des sans suites et procéder à une approche plus poussée pour la modélisation de ces derniers, notamment en ce qui concerne les cadences de liquidation et de paiement.

Malheureusement, la méthode est coûteuse en temps de calcul. Comme nous l'avons vu dans la théorie de la méthode Monte Carlo, l'erreur d'estimation décroît lorsque le nombre de simulations augmente, ce qui entraîne directement une augmentation en temps de calcul. Elle est également plus complexe que Chain Ladder à implémenter car elle requiert d'avoir les informations sur les sinistres individuels. Elle pourrait néanmoins s'inscrire dans un environnement de validation des calculs afin d'appuyer ou non les résultats des méthodes agrégées.
Avec la puissance des outils de calcul qui existent aujourd'hui, le développement de méthodes de provisionnement individuel devient plus facile. Bien que les travaux demandent de nombreux axes d'amélioration, l'approche semble ouvrir une piste intéressante et pourrait plus facilement s'appliquer en réassurance, où l'on peut en effet retrouver des délais de déclaration assez longs (et donc des sinistres tardifs) liés au fait

que les assureurs attendent par exemple d'atteindre un certain montant de sinistres avant de les déclarer aux réassureurs.

Bibliographie

[1] ACAPS. Rapport du secteur des assurances. 2019.

[2] Apref. Réouvertures des dossiers en (ré)assurance. 2012.

[3] ATEL. Pays de la zone carte verte.

[4] Mr Hamid BESRI. Indemnisation des victimes de circulation au maroc. 2008.

[5] Anne Bontoux. Méthode de cotation en excedent de sinistre appliquées à la réassurance rc automobile. 1997.

[6] Clément Bourry. Evaluation des provisions techniques et du capital économique associé au risque de réserve en assurance construction. *Institut des Actuaires*.

[7] MARION CAMIER. Argus de l'assurance - des règles communes de gestion des dommages corporels. 2012.

[8] Michel Denuit Julien Truffin Carole BettonVille, Robin Von Oirbeek. Semi-markov multistate individual loss reserving model in general insurance. 2020.

[9] Gilbert Colletaz. Statistique non paramétrique. 2020.

[10] Agence Nationale de la Sécurité Routière. Note sur les statistiques provisoires des accidents corporels de la circulation routière au titre du mois d'août 2020. 2020.

[11] Marie Laure Delignette-Muller. fitdistrplus : An r package for fitting distributions. 2020.

[12] Jean Pierre Dintilhac. Nomenclature des postes de préjudices : rapport de m. dintilhac. 2005.

[13] Aman-Yah Andréa Ehui Mayoro FALL. Application de méthodes de machine learning au provisionnement non-vie. 2017-2018.

[14] FFA. Données clés assurances de biens de de resposabilité. 2019.

[15] Lise He. Méthode de provisionnement et analyse de la solvabilité d'un entreprise d'assurance non-vie. 2004.

[16] https ://fr.statista.com/statistiques/875501/taux-d-inflation maroc/. Taux d'inflation au maroc de 2014 à 2024.

[17] https ://www.benezra-victimesdelaroute.fr/accident-corporel-de-la-route-dun-francais-en espagne/. Victime d'un accident de la circulation en espagne : quelle indemnisation dans le pays du tout barème.

[18] INSEE. Taux d'inflation - données annuelles de 1991 à 2019. 2020.

[19] Shin Woong KANG. Quantification of operational risk within insurance company. *Institut des Actuaires*, 2015.

[20] Magali Kelle. Provisionenment pour sinistres à payer : Analyses et modélisation sur données détaillées.

[21] Uri Korn. *A Frequency-Severity Stochastic Approach to Loss Development.*

[22] Chistophe Dutang Marie Laure Delignette-Muller. Documentation r. *fitdistrplus : An R package for Fitting Distributions*, 2020.

[23] Zoltan Milotain. Frequency and severity models in reserving. 2016.

[24] Pietro Parodi. Triangle-free reserving : a non-traditional framework for estimating reserves and reserve uncertainty. 2013.

[25] Glenn G.Meyers Philip E.Heckman. The calculation of aggregate loss distributions from claim severity and claim count distributions.

[26] Frédéric Planchet. Modèles fréquence coût : Quelles perspectives d'évolution ? 2014.

[27] Julien STOEHR. Méthodes de monte carlo. 2019.

[28] Jun Yan. Loss cost modeling vs frequency and severity modeling. *2010 CAS RateMaking and Product Management Seminar*, 2011.

Liste des abréviations

AD : Anderson Darling

CVM : Cramer Von Mises

FFA : Fédération Française des Assurances

PSAP : Provision Pour Sinistre A Payer

IA : Institut des Actuaires

IBNER : *Incurred But Not Enough Reported*

IBNR : *Incurred But Not Reported*

IBNYR :*Incurred But Not Yet Reported*

PPNA : Provision Pour Primes Non Acquises

PRC Provision Pour Risques En Cours

Provision D/D : Provision Dossier/Dossier

Table des figures

1 Résultat du calcul de la réserve IBNR sur le triangle de charges des sinistres survenus entre 2009 et 2018 . . . 3
2 Comparaison des lois candidates aux délais de déclaration empiriques . . . 4
3 Obtention des paramètres de la distribution de la loi des nombres . . . 5
4 Comparaison des sinistres tardifs obtenus avec Schnieper et avec Parodi . . . 5
5 Comparaison de la distribution des données empiriques et de la distribution des lois testées . . . 6
6 Projection de la charge des sinistres en cours à l'ultime . . . 7
7 Comparaison des projections avec la charge placée observée aux dates d'arrêtés 31/12/2015, 2016 et 2017 . . . 8
8 Erreur d'estimation en % associée à chaque méthode . . . 8
9 Result of the calculation of the IBNR reserve on the expense triangle for claims incurred between 2009 and 2018 . . . 11
10 Comparison of candidate laws to empirical reporting delays . . . 12
11 Obtaining the parameters of the distribution of the frequency law . . . 13
12 Comparaison des sinistres tardifs obtenus avec Schnieper et avec Parodi . . . 13
13 Comparison of the distribution of the empirical data and the distribution of the tested laws . . . 14
14 Projection of the cost of claims in progress to the ultimate . . . 15
15 Comparison of projections with observed at cut-off dates 31/12/2015, 2016 and 2017 . . . 16
16 Estimation error in % associated with each method . . . 16

1.1 Schéma récapitulatif des années de référence . . . 32
1.2 Le cycle de vie d'un sinistre . . . 32
1.3 Schéma récapitulatif de la PSAP en non-vie . . . 35

2.1 Triangle de charge cumulé . . . 37
2.2 Structure du triangle utilisé dans la méthode Chain Ladder . . . 38

3.1 Caractéristiques des assurés et des véhicules . . . 50
3.2 Nombre de sinistres par année de survenance . . . 50
3.3 Histogramme des délais de déclaration en années . . . 51

3.4 Statistiques descriptives des délais de déclaration en jours 51
3.5 Délais de déclaration moyens en jours par année de survenance 52
3.6 Statut des sinistres à la survenance puis à la fin de la période d'observation . . . 52
3.7 Charge moyenne placée la première année d'ouverture et au 31/12/2018 par année de survenance pour tous les sinistres 53
3.8 Charge moyenne placée la première année d'ouverture et au 31/12/2018 par année de survenance pour les sinistres clos 53

4.1 Triangle de charges cumulé 54
4.2 Vérification de l'hypothèse de linéarité des couples 55
4.3 Etude de la linéarité des coefficients de passage individuels par année de développement 55
4.4 Triangle des facteurs de développement individuels 56
4.5 Résultat du calcul de la réserve IBNR sur le triangle de charges par la méthode Chain Ladder 57
4.6 Taux d'inflation au Maroc entre 2014 et 2024 (comparaison par rapport à l'année précédente) 57
4.7 Triangle de charges après la mise en as-if 59
4.8 Triangle de facteurs de développement individuel du triangle de charge *as-ifé* . . 59
4.9 Triangle de charges as-if 2018 complété 60
4.10 Cadence de règlements 60
4.11 Projection des IBNR avec une méthode Chain Ladder inflatée 60
4.12 Estimateurs des moindres carrés de λ_j et σ_j 61
4.13 Calculs des erreurs quadratiques moyennes 61
4.14 Erreur de prédiction finale 62
4.15 Triangle de nombre cumulé agrégé $X_{i,j}$ 62
4.16 Nombre agrégé des nouveaux sinistres $N_{i,j}$ 63
4.17 Variations du nombre de sinistres $D_{i,j}$ 63
4.18 Triangles des nouveaux sinistres $N_{i,j}$ complété 63
4.19 Triangle du nombre de sinistres sans suite $D_{i,j}$ complété 64
4.20 Triangle des $X_{i,j}$ complété 64
4.21 Résultats 64

5.1 Proportion des sinistres dont le délai de déclaration est inférieur à 2 ans 72
5.2 Histogramme des délais en jours 72
5.3 Densité et distribution empirique des délais de déclaration inférieurs à 730 jours (2 ans) 73
5.4 Densité et distribution empirique des délais de déclaration supérieurs à 730 jours (2 ans) 73
5.5 Graphique Cullen & Frey sur les délais de déclaration en jours 74
5.6 Statistiques descriptives des délais de déclaration en jours 74
5.7 Histogramme des données et densités empiriques 75

5.8 Distribution des données observées et des lois testées 75
5.9 Résultats des tests d'adéquation pour les 3 lois candidates 75
5.10 Paramètres de la loi choisie 76
5.11 Superposition de la densité empirique et d'une densité de Weibull avec les paramètres calculés 76
5.12 Nombre de sinistres observés au 31/12/2018 répartis par année de survenance . . 77
5.13 Sinistres tardifs survenus sur la période 2009-2018 78
5.14 Comparaison des sinistres tardifs obtenus avec Schnieper et avec Parodi 79
5.15 Sinistres tardifs par année de survenance 79
5.16 Dérivation des paramètres de la loi Négative Binomiale 81
5.17 Charge moyenne au titre de chaque année de survenance selon le statut du sinistre au 31/12/2018 81
5.18 Ratios pour le réajustement des sinistres en cours 82
5.19 Charge moyenne des sinistres après réajustement pour les sinistres en cours (en bleu) 83
5.20 Densité empirique des charges pour chaque année de survenance 84
5.21 Graphique Cullen Frey des charges après réajustements 84
5.22 Comparaison de la distribution des données empiriques et de la distribution des lois testées 85
5.23 Tests d'adéquation des lois statistiques 85
5.24 Quantiles de la distribution 86

6.1 Triangle de charges cumulées pour les sinistres en cours dans le portefeuille . . . 88
6.2 Vérification des hypothèses de Chain Ladder 88
6.3 Synthèse des résultats pour la période 2009-2018 89
6.4 Résultats pour les projections effectuées au 31/12 pour les 4 dernières années d'historique 89

7.1 Comparaison des projections avec la charge placée observée aux dates d'arrêtés 31/12/2015, 2016 et 2017 91
7.2 Erreur d'estimation en % associée à chaque méthode 91
7.3 Erreur dans la prévision du nombre de sinistres tardifs 92
7.4 Ecart avec l'observé selon l'utilisation ou non de la distribution empirique des délais de déclaration 93

A.1 Répartition du nombre de sinistres dommages total en 2019 en France - Données Clés 2019 *Fédération Française de l'assurance* 102
A.2 Fréquence des accidents corporels de la circulation entre 2015 et 2019 en France -Données Clés 2019 *Fédération Française de l'assurance* 103

Annexe A

Chiffres clés pour l'Assurance Automobile en France

Parmi les grands pays européens, les primes moyennes en France restent très compétitives. Les garanties évoluent peu et s'étendent au rythme des évolutions technologiques dans les véhicules ou des nouvelles mobilités utilisées par les employeurs pour leurs collaborateurs. Le chiffre d'affaires de l'assurance automobile s'établit à 22,1 Md€ en 2018 et représente 39% de l'ensemble des cotisations des assurances de dommages aux biens et de responsabilité civile.

En 2019, sur 13,2 millions de dommages de biens et responsabilités déclarés, 66% sont issus de contrats automobiles. Les sinistres découlant de la garantie Responsabilité Civile Corporelle représente **x pourcent** des sinistres automobiles en 2019.

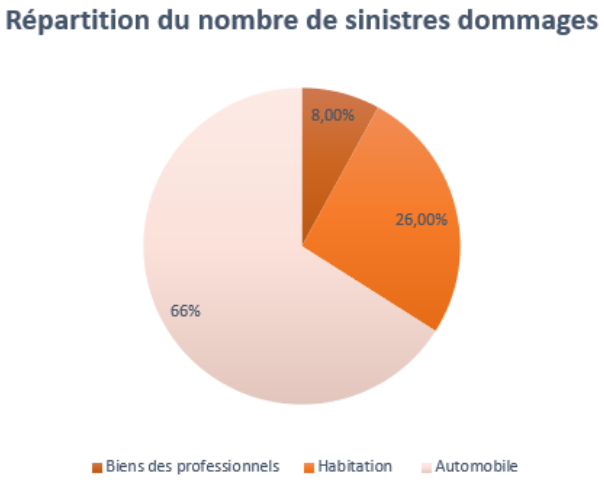

FIGURE A.1 – Répartition du nombre de sinistres dommages total en 2019 en France - Données Clés 2019 *Fédération Française de l'assurance*

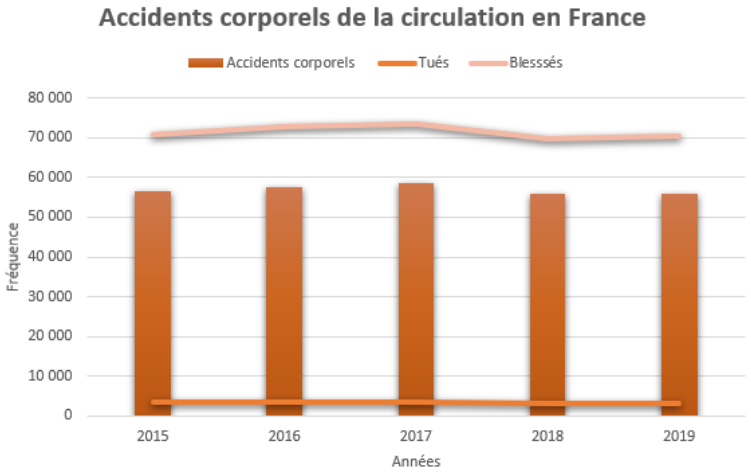

FIGURE A.2 – Fréquence des accidents corporels de la circulation entre 2015 et 2019 en France -Données Clés 2019 *Fédération Française de l'assurance*

Annexe B

La convention d'Indemnisation et de Recours Corporel Automobile (IRCA) en France

Il est possible que les règles établies par la convention IRCA retardent la connaissance du sinistre par l'assuré indemnisateur :
Cette convention permet tout d'abord d'éviter qu'une victime se trouve face à une multiplicité d'interlocuteurs, ou à une absence d'interlocuteurs en cas d'accident mettant en cause plusieurs véhicules. Ell est applicable aux accidents survenus depuis le 1er avril 2002 et concerne les accidents de la circulation ayant entraîné des atteintes à la personne, survenus en France métropolitaine, dans les départements d'Outre Mer et dans la principauté de Monaco. Cette convention concerne les dossiers des victimes ayant subi une atteinte à leur intégrité physique et/ou psychique (AIPP) inférieure ou égale à 5%. Le taux d'AIPP correspond à l'invalidité que va garder la victime postérieurement à la consolidation et est évaluée en degré sur une échelle de 0 à 100% par le médecin-expert. Ainsi, bien que ce taux d'AIPP retenu puisse sembler faible, les accidents qui leur sont liés représentent tout de même plus de 90% des personnes indemnisées.
Les litiges entre assureurs concernant la mise en oeuvre de la convention Irca ne sont pas soumis aux tribunaux. Ils se traitent dans le cadre d'une procédure de conciliation et d'arbitrage. Dans la convention Irca, le recours à l'arbitrage est obligatoire pour les litiges relatifs à l'attribution du mandat, à l'application du Protocole d'accord entre les assureurs et les organismes sociaux (PAOS) et aux recours en contribution d'un montant fixé à 100 000€. Pour les autres litiges, le recours à l'arbitrage est subordonné à l'accord des parties.

Il appartient à l'assureur du véhicule impliqué de gérer et, dans la majorité des cas, de présenter une offre d'indemnisation aux victimes de dommages corporels. Si un même accident résulte de l'implication de plusieurs véhicules, il convient alors de déterminer quel doit être l'assureur tenu de présenter cette offre.
Ainsi, l'un des principaux objectifs de la convention Irca est de désigner, " dès la survenance de l'accident, l'assureur chargé, dans le respect de la loi du 5 juillet 1985 et de son décret d'application du 6 janvier 1986, d'instruire le dossier de la victime et de satisfaire à la procédure d'offre ».
La désignation de l'assureur chargé d'instruire le dossier et de présenter une offre d'indemnisation est donc faite, dès la survenance de l'accident, victime par victime, parmi les assureurs

des véhicules impliqués dans l'accident. Il est à noter qu'en application de l'article L. 211-10 du code des assurances, l'assureur mandaté est tenu d'instruire le dossier, même s'il invoque une exception légale ou contractuelle.

S'agissant des règles d'attribution du mandant à proprement parler, deux grandes catégories de victimes se dégagent de la convention Irca :

— Les occupants d'un véhicule assuré auprès d'une société adhérente. Si la victime, passager ou conducteur, se trouvait avant l'accident dans ou sur un véhicule assuré auprès d'une société adhérente, le mandat est attribué à l'assureur de ce véhicule (assureur direct)

— Pour les autres victimes (cyclistes, piétons, occupants d'un véhicule non assuré ou assuré auprès d'une société non adhérente à la convention ou appartenant à l'État), l'assureur mandaté est celui du véhicule qui est entré en contact avec elles (règle du véhicule percuteur). En présence de plusieurs véhicules garantis auprès d'une société adhérente et impliqués dans un même accident, l'assureur mandaté est celui du véhicule qui porte le plus faible numéro minéralogique [1]. Son mandat s'étend à toutes les victimes autres que les occupants des véhicules.

À l'égard de la victime, l'assureur direct est chargé, principalement, de prendre toutes les dispositions qui permettent de respecter les règles de la procédure d'offre d'indemnisation à la victime en application des articles L. 211-9 et suivants du code des assurances. Pour ce faire, il lui adresse la lettre d'information sur ses droits, ainsi que le questionnaire corporel qui permettra de prendre connaissance de l'importance des dommages et de l'identité des tiers payeurs susceptibles de lui verser des prestations. L'assureur direct prend également les mesures nécessaires à l'évaluation du dommage corporel de la victime. Selon l'importance du préjudice, cette dernière pourra rendre l'expertise médicale obligatoire, autoriser un simple avis sur pièces médicales ou permettre le règlement sur la base des seules pièces produites par la victime.

En cas d'expertise médicale, l'assureur mandaté choisit l'expert sur une liste établie, mise à jour et diffusée par la commission médicale, instance faisant partie des organes de la convention Irca. Le médecin expert est missionné sur la base de la mission « droit commun 2006 », mise à jour en 2009, intégrant la nomenclature « Dintilhac » des postes de préjudice.

Il est à noter que, en cas d'AIPP inférieure ou égale à 5%, les conclusions médicales sont incontestables, sauf en cas d'erreur médico-légale manifeste (par exemple, si le médecin méconnaît un état antérieur ou une aggravation, ou s'il retient à tort une imputabilité). En cas d'AIPP comprise entre 6 et 10%, les conclusions médicales sont incontestables (sauf en cas d'erreur médico-légale manifeste), dès lors que l'assureur mandaté a invité les autres assureurs à participer à l'expertise et que ceux-ci n'ont pas utilisé cette possibilité. En cas d'AIPP supérieure à 10%, les conclusions de l'expert médical peuvent être contestées.

Par ailleurs, l'assureur mandaté doit donner suite aux recours des tiers payeurs. En outre, il a la qualité d'assureur gestionnaire dans le cadre du PAOS. Enfin, à l'égard des autres assureurs de véhicules impliqués, l'assureur mandaté est tenu de leur communiquer les informations qu'il détient tout au long de la procédure. Ainsi, dès l'ouverture du dossier, il doit transmettre les premiers renseignements qu'il possède sur la nature et l'importance des dommages subis par la victime. De même, en cas d'expertise médicale, il doit informer de la saisine et du nom de l'expert médical missionné. À ce titre, si, dans un délai de huit mois, l'assureur mandaté n'a pas informé les autres assureurs du nom de l'expert missionné par lui et n'a pas communiqué une copie de son rapport d'expertise médicale, l'assiette de son recours est limitée au montant du forfait. À

1. plaque d'immatriculation

noter que ces échanges d'informations sont aujourd'hui automatisés.

En cas d'AIPP supérieure à 5% ou de décès de la victime, l'assureur qui reconnaît à son assuré, selon les règles du droit commun, une responsabilité prépondérante doit revendiquer le mandat à l'assureur initialement désigné.
L'assureur qui, à tort, ne se serait pas reconnu mandaté, ne peut pas contester l'indemnité payée pour son compte par un autre assureur et doit supporter en outre une majoration de 10% du montant des sommes versées, avec un minimum de 1 000€.
L'assureur substitué dans le mandat s'engage, tant à l'égard de la victime que des tiers payeurs, à ne pas remettre en cause les accords déjà passés ni, plus généralement, les mesures prises par le précédent assureur mandaté. Il rembourse à ce dernier les sommes qu'il avait engagées au titre de provisions versées à la victime ou d'indemnités aux tiers payeurs.
La convention Irca fixe les règles concernant les recours en contribution entre assureurs qui interviennent lorsque l'assureur gestionnaire a indemnisé la victime et les organismes sociaux. L'assiette de recours intègre les indemnités versées à la victime au titre de ses dommages corporels et à ses proches, les sommes payées aux organismes sociaux dans le cadre du PAOS et, éventuellement, les montants perçus par l'employeur de la victime au titre des compléments de salaires et des charges patronales. Cette assiette peut être réduite lorsque la victime ne conserve pas d'incapacité, ou en cas d'AIPP inférieure à 5% Dans le premier cas, pour les accidents survenus après le 1er avril 2007, les assureurs ont mis en place un recours forfaitaire. Ce forfait est fixé chaque année par la commission d'application (1 490€ au 1er janvier 2011). En cas d'AIPP inférieure ou égale à 5%, l'assiette de recours est encadrée, c'est-à-dire que pour les préjudices autres qu'économiques (incapacité permanente, souffrances endurées, préjudice esthétique), l'assureur gestionnaire n'obtiendra le remboursement de ce qu'il a payé à ses clients que dans la mesure où les montants de chacun de ces postes de préjudice se situent dans une fourchette prédéterminée. Par ailleurs, dans cette même hypothèse, les sommes réclamées par l'assureur mandaté ne peuvent être contestées ni en ce qui concerne le droit à indemnisation ni au sujet du montant des indemnités versées.

Annexe C

Liste des pays adhérant au système carte verte

— Autriche
— Allemagne
— Belgique
— Bulgarie
— Chypre
— Croatie
— Espagne
— Estonie
— France
— Finlande
— Grèce
— Guadeloupe
— Guyanne Française
— Hongrie
— Italie
— Irlande
— Islande
— Luxembourg
— Lituanie
— Lettonie
— Malte
— Martinique
— Mayotte
— Maroc
— Monaco
— Norvège
— Pays Bas
— Portugal
— Pologne
— République Tchèque

— Roumanie
— Réunion
— Roumanie
— Royaume Uni
— Saint Pierre et Miquelon
— Serbie
— Slovaquie
— Slovénie
— Suède
— Suisse

Annexe D

La Nomenclature Dintilhac

D.1 Préjudices de la victime directe

D.1.1 Préjudices patrimoniaux

Préjudices patrimoniaux temporaires (avant consolidation)

— Dépenses de santé actuelles (D.S.A)
— Fais divers (F.D)
— Pertes de gains professionnels actuels

Préjudices patrimoniaux permanents (après consolidation)

— Dépenses de santé futures (D.S.F)
— Frais de logement adapté (F.L.A)
— Frais de véhicule adapté (F.V.A)
— Assistance par tierce personne (A.T.P)
— Pertes de gains professionnels futurs (P.G.P.F)
— Incidence professionnelle (I.P)
— Préjudice scolaire, universitaire ou de formation (P.S.U)

D.1.2 Préjudices extra-patrimoniaux

Préjudices extra-patrimoniaux temporaires (avant consolidation)

— Déficit fonctionnel temporaire (D.F.T)
— Souffrances endurées (S.E)
— Préjudice esthétique temporaire (P.E.T)

Préjudices extra-patrimoniaux permanents (après consolidation)

— Déficit fonctionnel permanent (D.F.P)
— Préjudice d'agrément (P.A)
— Préjudice esthétique permanent (P.E.P)
— Préjudice sexuel (P.S)

— Préjudice d'établissement (P.E)
— Préjudices permanents exceptionnels (P.P.E)

Préjudices extra-patrimoniaux évolutifs (hors consolidation)

— Préjudices liés à des pathologies évolutives (P.EV.)

D.2 Préjudices des victimes indirectes

D.2.1 Préjudices des victimes indirectes en cas de décès de la victime directe

Préjudices patrimoniaux

— Frais d'obsèques (F.O)
— Pertes de revenus des proches (P.R)
— Frais divers des proches (F.D)

Préjudices extra-patrimoniaux

— Préjudice d'accompagnement (P.AC)
— Préjudice d'affection (P.AF)

D.2.2 Préjudices des victimes indirectes en cas de survie de la victime directe

Préjudices patrimoniaux

— Pertes de revenus des proches (P.R)
— Frais divers des proches (F.D)

Préjudices extra-patrimoniaux

— Préjudice d'affection (P.AF)
— Préjudices extra-patrimoniaux exceptionnels (P.EX)

Annexe E

Estimateur du maximum de vraisemblance

Soit une suite de variables aléatoires $\mathbf{Y} = Y_1, ..., Y_n$ indépendantes et identiquement distribuées, suivant une loi de densité $f(\mathbf{Y}_i|\theta)$ avec θ inconnu, correspondant au paramètre de la distribution (θ peut également représenter un vecteur de paramètres inconnus).
Pour $\mathbf{Y}$ une série de variables continues, $f(\mathbf{Y}_i|\theta)$ est la fonction de densité pour de chaque variable Y_i. La fonction de vraisemblance, dénotée $L(\theta)$ est la densité jointe des $Y_i = y_i, ...Y_n = y_n$ telle que :

$$L_n(\theta) = \prod_{i=1}^{n} f(y_i|\theta)$$

La valeur de θ maximisant cette fonction de vraisemblance est définie comme étant $\hat{\theta_n}$ **l'estimateur du maximum de vraisemblance** (en anglais, $m.l.e$) [2] où :

$$L(\hat{\theta_n}) \geq L(\theta)$$

Il est important de noter que la fonction de vraisemblance et par conséquent l'estimateur du maximum de vraisemblance dépendent de la taille de l'échantillon n. La fonction de vraisemblance est composé d'au moins un maximum mais $\hat{\theta_n}$ n'a pas forcément une valeur unique. Dans les distributions utilisées au cours de ce mémoire, nous considérons que $\hat{\theta_n}$ est une valeur unique.
Etant donné qu'il est plus facile de travailler avec des sommes que des produits, nous procédons à une transformation logarithimique de la fonction de vraisemblance, soit :

$$l_n(\theta) = \log L_n(\theta) = \log(\prod_{i=1}^{n} f(y_i|\theta)) = \sum_{i=1}^{n} \log(f(y_i|\theta))$$

où $l_n(\theta)$ est définie comme étant la **log-vraisemblance**. Si $\hat{\theta_n}$ maximise $L_n(\theta)$ alors il maximise également $l_n(\theta)$. Pour trouver le maximum d'une fonction, il faut trouver la solution de l'équation

$$\frac{d}{d\theta} l_n(\theta) = 0$$

2. maximum likelihood estimator

Annexe F

Tests d'adéquation statistiques

Le principe de ces tests est de comparer une fonction de répartition théorique F_X spécifiée sous $H0$ à une fonction de répartition empirique G_X.
Ces tests reposent sur le théorème suivant :

Théorème 1 : Si X est une variable aléatoire de fonction de répartition $F(x)$ inversible, alors $Y = F(X)$ est une variable uniforme sur $[0, 1]$

Démonstration
$\forall y \in [0, 1]$,
$\Pr[F_(X) \leq y] = \Pr[F_(X) \leq y]$ par construction de Y
$\Pr[F_(X) \leq y] = \Pr[X \leq F^{-1}(y)]$ car F monotone non décroissante
$\Pr[F_(X) \leq y] = F(F^{-1}(y))$ '
$\Pr[F_(X) \leq y] = y$

Ce théorème donne une solution si on veut générer des nombres x comme réalisations d'une fonction de répartition F presque quelconque.

F.1 Le test de Kolmogorov Smirnoff

Afin de statuer le caractère approprié de la loi de probabilité ayant une fonction de répartition F, il faut regarder la distance qui sépare la fonction théorique a priori de la fonction de répartition empirique G_X.
Une mesure possible de cette distance est :

$$\boxed{D = \sup_x |G(x) - F(x)|}$$

Sous $H0$, $G_X = F_X$, c'est à dire la fonction de répartition empirique correspond à la fonction de répartition théorique. Ainsi, une distance proche de 0 est favorable à l'hypothèse nulle et plus D est élevé, plus on est tenté de rejeter $H0$.
En pratique, nous devons espérer que D possède une certaine distribution qui permettra de calculer des valeurs critiques et par conséquent de juger si D est raisonnablement assez petit pour ne pas rejeter H0.

Il est possible de montrer que la distribution statistique de D ne dépend pas de la fonction supposée F_X.

$|G(x) - F(x)| = |\frac{1}{n}\sum_{i=1}^{n} I[x_i \leq x] - F(x)|$
$|G(x) - F(x)| = |\frac{1}{n}\sum_{i=1}^{n} I[F(x_i) \leq F(x)] - F(x)|$
$|G(x) - F(x)| = |\frac{1}{n}\sum_{i=1}^{n} I[y_i) \leq y] - y|$
et $y_i = F(x_i)$ est la réalisation d'une uniforme sur [0,1] alors pour un x donné, l'écart absolu entre $F(x)$ et $G(x)$ ne dépend pas de $F(x)$.
En conséquence, $D = \sup_x |G(x) - F(x)| = \sup_x |F_{UNIF}(y) - y|$

Théorème 2 (Théorème de Kolmogorov Smirnoff) :
Pour un ensemble de n variables aléatoires i.i.d de fonction de répartition continue F, on a :
$\Pr[\sqrt{n}D \leq x] \underset{n\to+\infty}{\longrightarrow} K(X)$ où $K(X)$ est la fonction de répartition de Kolmogorov définie par :
$K(X) = 1 - 2\sum_{i=1}^{\infty}(-1)^{i-1}\mathrm{e}^{-2i^2x^2}$

Pour les faibles valeurs de n, il existe des tables qui donnent les valeurs critique aux seuils de risques usuels. Pour n grand, il est possible de calculer $K(X)$

F.2 Le test Anderson-Darling et le test de Cramer Von Mises

Ce test débute avec la même logique que le test de Kolmogorov Smirnoff, c'est à dire qu'il étudie la distance entre la fonction de répartition théorique supposée sous $H0$ et la fonction de répartition empirique.
Ils diffèrent du test K-S pour deux raisons :

— La distance prend en compte l'écart quadratique $[F(x) - G(x)]^2$ et non l'écart absolu $|F(x) - G(x)|$
— Le test de K-S fait un focus sur la distance maximale entre les deux fonctions (utilisation du *supremum*), alors que les deux autres considèrent toutes les observations lors du calcul de la statistique de test.

Par contre, la statistique de tests dans ces cas va dépendre de la fonction $F(x)$ retenue dans $H0$.
L'expression générale des statistiques de tests pour ces deux tests est :

$Q = n\int_{-\infty}^{+\infty}(F(x) - G(x))^2 p(x)dF(x)$ où $p(x)$ est une fonction de pondération caractérisant l'un des deux tests.
Pour le test d'Anderson Darling : $p(x) = [F(x)(1 - F(x)]^{-1}$ et
$D = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\log(F(x_i)) + (2n+1-2i)\log(1 - F(x_i))]$

Pour le test Cramer Von Mises : $p(x) = 1$ et
$D = \sum_{i=1}^{n}(\frac{2i-1}{2n} - F(x_i))^2 + \frac{1}{12n}$. x_i est la ième plus petite valeur de l'échantillon.

Comme pour le test CVM $p(x) = 1$, un même poids est donné à toutes les observations quel que soit la position dans l'échantillon. Pour AD, la pondération donne plus de poids aux observations

situées dans les queues de distributions. Ainsi la statistique pour AD est intéressante lorsque nous sommes intéréssés par l'acart entre les deux fonctions pour les observations situées dans les queues.