



Mémoire présenté le :

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

Par : LE TUAN Anh

Titre : Les méthodes de provisionnement en assurance non-vie

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

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

Membre présents du jury de l'IA

Signatures

Entreprise

Nom : Petro VietNam Insurance-PVI

Directeur de mémoire en entreprise

Nom : M. PLANCHET Frédéric

Signature :

Invité

Nom :

Signature :

Membres présents du jury de l'ISFA

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

Signature du responsable entreprise

Secrétariat :

Mme Christine DRIGUZZI

Bibliothèque :

Mme Patricia BARTOLO

Signature du candidat

Résumé

Mots clés

Assurance Non-Vie, provisionnement, méthodes déterministes, méthodes stochastiques, bootstrap, Merz-Wüthrich, modélisation des valeurs extrêmes, risque de provisionnement à un an.

Le sujet de ce mémoire est de proposer un modèle de provisionnement non-vie. Une des tâches les plus importantes pour une société d'assurances est l'évaluation de ses provisions techniques et plus précisément les réserves pour sinistres. L'objectif de ce mémoire est d'implémenter, d'analyser et d'établir une étude comparative des méthodes de calcul des provisions techniques en assurances non-vie. Enfin, nous nous sommes concentrés sur l'évaluation du risque à un an et l'évaluation des risques de l'assurance non-vie.

Une vue d'ensemble de la nature de l'assurance Non-Vie est donnée dans la Partie I pour que l'on puisse mieux comprendre le concept de provisionnement. Il illustre les théories et les méthodes de base de calcul des provisions applicables au marché Vietnamien.

La partie 2 présente les méthodes déterministes et stochastiques, en particulier la méthode Chain Ladder, la méthode de London-Chain, la méthode de Cape Cod, la méthode de Bornhuetter-Fergusson, le modèle de Mack, le modèle Log-normal, la modélisation GLM. Les méthodes déterministes sont intuitives et simples à utiliser, mais les données doivent être stables tout au long de la période étudiée afin de produire des résultats précises. Une telle stabilité n'est pas toujours valable. Donc, l'application des méthodes stochastiques afin de quantifier l'incertitude liée à l'estimation des montants de réserves est effectuée.

La partie 3 établie la distribution des réserves, nous appliquons une technique d'échantillonnage en provisionnement qui est la méthode Bootstrap. Cette méthode nous permet d'estimer la distribution des réserves à partir d'un échantillon d'observations indépendantes et identiquement distribuées.

Finalement, nous nous sommes focalisés sur l'évaluation du risque à un an. Dans ce but, nous avons adapté les méthodes retenues précédemment et comparé nos résultats le modèle de Merz-Wüthrich. Nous avons utilisé la théorie des valeurs extrêmes pour éliminer les risques graves. L'ensemble des résultats présentés dans ce mémoire ont été obtenus à l'aide d'outils développés sous le logiciel R.

Abstract

Key Words

Non-life insurance, provisioning, deterministic methods, stochastic methods, bootstrap, Merz-Wüthrich, modeling extreme values, risk provisioning in one year.

The subject of this paper is to propose a non-life provisioning model. One of the most important tasks for an insurance company is assessing its technical reserves and specifically claims reserves. The objective of this paper is to implement, analyze and establish a comparative study of methods for calculating technical provisions in non-life insurance. Finally, we focused on risk assessment at one year and the risk assessment of non-life insurance.

An overview of the nature of the non-life insurance is given in Part I for one can better understand the concept of provisioning. It illustrates the basic theories and methods of calculation of the applicable provisions in the Vietnamese market.

Part 2 presents the deterministic and stochastic methods, especially the Chain Ladder method, the London-Chain, the method of Cape Cod, the method of Bornhuetter-Fergusson, model Mack, the log-normal model, modeling GLM. Deterministic methods are intuitive and simple to use, but the data must be stable throughout the study period in order to produce accurate results. Such stability is not always valid. So the application of stochastic methods to quantify the uncertainty of reserve estimates of amounts is performed.

Part 3 established the distribution of reserves. We apply a sampling technique that is Bootstrap. This method allows us to estimate the distribution of reserves from a sample of independent and identically distributed observations.

Finally, We focused on risk assessment at one year. To this end, we have adapted the methods previously used and compared our results model Merz-Wüthrich. We used the theory of extreme values to eliminate serious risks. All the results presented in this report were obtained using tools developed under the R software.

Remerciements

Je tiens à remercier mes collègues au département de stratégique à PVI pour leur aide et soutien pendant tout le déroulement de mon mémoire.

J'adresse mes remerciements à l'Institut de Science Financière et d'Assurances, notamment à mon tuteur monsieur Frédéric Planchet, professeur à l'ISFA, pour sa patience et son dévouement dans les lectures de mon mémoire, ainsi que pour ses conseils éclairés.

Je remercie monsieur Brias Samy, Gestionnaire des relations internationales, pour son aide très enthousiaste tout au long l'époque que je vivais en France.

Table des matières

Résumé	2
Abstract	3
Remerciements	4
Introduction	6
1 Provisionnement non-vie	8
1.1 Présentation de PVI	8
1.2 Provision de l'assurance non-vie	9
1.2.1 Les provisions pour primes non-acquises au Viet Nam	10
1.2.2 Les provisions pour sinistres au Viet Nam	14
2 Rappel les méthodes déterministes et stochastiques	16
2.1 Méthodes déterministes	16
2.1.1 Méthode de Chain Ladder	16
2.1.2 Méthode de London-Chain	24
2.1.3 Méthode de Bornhuetter-Fergusson	27
2.1.4 Méthode de Cape Cod	31
2.2 Méthodes stochastiques	34
2.2.1 Le modèle de Mack	34
2.2.2 Le modèle Log-normal	38
2.2.3 La modélisation GLM	42
3 Le risque de provisionnement et bootstrap	49
3.1 Etude de la base de données	49
3.2 Le bootstrap	52
4 Mesurer le risque de provisionnement à horizon un an	56
4.1 La méthode de Merz-Wüthrich	56
4.2 Calcul du SCR	62

4.3	La méthode GLM Bootstrap	65
4.4	Modélisation des valeurs extrêmes	67
4.5	L'incertitude à un an	76
Conclusion		77
Bibliographie		79
Annexes		82
	Annexe 1 : Triangle des paiements cumulés(Dom.Biens)	82
	Annexe 2 : Triangle des paiements cumulés(Dom.Energie)	82
	Annexe 3 : Estimations des paramètres des modèles linéaires généralisés et leur risque associé	82
	Annexe 4 : Loi de probabilité de la variable de réponse	84
	Annexe 5 : Théorie des valeurs extrêmes	86
Table des figures		90
Liste des tableaux		91

Introduction

L'objectif de la réglementation assurantienne est de préserver le système financier et de protéger le consommateur de produits d'assurance. Les autorités prudentielles en charge du contrôle du secteur assurantiel, telles que le ministère des Finances au Vietnam, l'Autorité de Contrôle Prudentiel et de Résolution (ACPR) en France, s'assurent du respect des règles de commercialisation des produits d'assurance et que les compagnies d'assurance soient bien en mesure de tenir leurs engagements vis-à-vis de leurs assurés.

Le monde assurantiel est régi par des règles différentes, les compagnies d'assurance européennes obéissent à la réglementation Solvabilité I établie dans les années 1970 et Solvabilité II sera mise en place dès 2016, les compagnies d'assurance des États-Unis sont réglementées au niveau de l'Etat. Actuellement les organismes de gestion et de supervision de l'assurance dans le cadre du ministère des Finances au Vietnam est en train d'élaborer les règles d'assurance conformément au droit international pour renforcer la protection des assurés et améliorer la gestion des risques en définissant de nouvelles exigences qui reflèteront au mieux le profil de risque de l'assureur. L'objectif de ce projet est de fournir des méthodes de calcul des réserves d'assurance non-vie, les règles du Vietnam, à comprendre pourquoi certaines méthodes sont inadaptées et à proposer des solutions pour modéliser au mieux les réserves.

Le mémoire s'articule selon le plan suivant :

- dans la première partie, nous rappellerons l'environnement réglementaire au Viet Nam. Nous détaillons aussi les différentes étapes de la vie d'un sinistre, de sa survenance à sa clôture, afin de mieux se familiariser avec le type d'information qui sera disponible dans les données individuelles de sinistres.

- dans la seconde partie, nous exposerons les principales méthodes agrégées de provisionnement, en premier lieu un ensemble de méthodes déterministes qui sont Chain-Ladder, London-Chain, Bornhuetter-Ferguson, Cape Cod et nous appliquerons des méthodes à fondement stochastique, à savoir le modèle de Mack, le modèle Log-normal et la modélisation GLM.

Nous résumons les résultats de toutes les méthodes utilisées dans cette partie, et nous établissons une analyses de ces résultats afin de déterminer la méthode la plus appropriée au portefeuille risque de la société.

- dans la troisième partie, nous ajoutons aussi une nouvelle technique d'échantillonnage en provisionnement qui est la méthode Bootstrap. Les modèles Bootstrap développés par England et Verrall se basent sur des simulations. Le modèle de Mack a été étudié, et adapté à la présence de la contrainte sur le taux de recouvrement. L'utilisation des modèles GLM a ensuite permis de donner une dimension stochastique au modèle de Chain Ladder, en utilisant respectivement une distribution de Poisson et une distribution normale. Nous avons également confronté ces modèles à la loi Gamma. Avec un grand nombre de simulations, la méthode bootstrap permet de déterminer la distribution suivie par les pertes ultimes.

- dans la quatrième partie, nous nous sommes ensuite focalisés sur l'évaluation du risque à un an. Dans ce but, nous avons adapté les méthodes retenues précédemment et implémenté une procédure utilisant des simulations Monte Carlo. Nous avons ainsi pu comparer nos résultats avec ceux donnés par la formule standard et le modèle de Merz-Wüthrich.

L'application de la méthode du bootstrap dans le cadre du provisionnement à 1 an permet d'obtenir la distribution des *Claims Development Results*, dont l'écart-type permet d'en déduire le SCR correspondant au risque de réserve. Nous avons utilisé la méthode GLM bootstrap, les distributions permettent alors d'en déduire les écarts-type et SCR du risque de réserve correspondant au quantile à 99,5% du risque de provisionnement à 1 an.

Nous appliquerons également la théorie des valeurs extrêmes pour modéliser les risques de gravité. Nous allons faire des tests graphiques pour regarder si l'espérance et la variance de nos risques sont finies ou infinies : cela nous donnera une première idée de la valeur de l'indice de queue. On définira d'abord des seuils pour pouvoir segmenter le portefeuille. Ensuite, on étudiera la fréquence au sein de chaque classe. Après cela, nous modéliserons ensuite le coût de ces sinistres.

Enfin, nous pouvons appliquer ces modélisations au risque de prime et en déduire les quantiles à 99,5% des charges de sinistres attritionnels et graves. Nous allons analyser et de comparer les différentes méthodes pour faire des commentaires et conclusions.

Chapitre 1

Provisionnement non-vie

1.1 Présentation de PVI

Insurance Corporation PetroVietnam (PVI) a été créé en 1996 en tant que compagnie d'assurance captive pour l'Oil & Gas Corporation Vietnam. Actuellement, la compagnie d'assurance de PVI est le leader sur le marché de l'assurance non-vie au Vietnam, ce qui représente la plus grande part de marché dans le domaine de l'assurance de l'énergie, l'assurance aviation, l'assurance de biens, assurance automobile, l'assurance d'ingénierie. PVI Insurance a obtenu le certificat international de la note A.M.Best à la cote de capacité financière à B ++ en 2014. Le marché de l'assurance non-vie au Vietnam est en croissance et aussi beaucoup de potentiel à exploiter. Par conséquent, l'assurance PVI est dans le besoin d'amélioration afin de construire le meilleur environnement de travail, l'optimisation des ressources.

PVI est une société avec les actionnaires des plus grandes sociétés financières dans la région et dans le monde :

- PetroVietNam-Le plus grand groupe économique au Vietnam(35.5%).
- Talanx-Finance Société d'assurance leader en Allemagne(31.82%).
- OIF-Sociétés de gestion de Fonds royale d'Oman(9.77%).

Système de gestion de la société PVI est développé selon les normes internationales, les normes IFRS et la méthode de calcul des réserves d'assurance internationales sont appliquées à la société PVI. Mais cette connaissance est encore limitée, car ils doivent se conformer à la réglementation sur le marché du Vietnam.

1.2 Provision de l'assurance non-vie

Les provisions techniques que les compagnies d'assurance doivent constituer pour faire face à leurs engagements. Les provisions techniques sont inscrites au passif du bilan des compagnies d'assurance ; elles représentent une part très importante de celui-ci environ 60 % à 80 %. Leur évaluation annuelle, voire trimestrielle a donc un impact important sur le résultat de la compagnie, ce qui en fait l'une des priorités majeures des compagnies d'assurance. Les provisions techniques correspondent au montant que l'assureur doit mettre de côté afin de pouvoir faire face au règlement des engagements envers ses assurés. En effet, en échange du versement de primes, l'assureur s'engage à indemniser l'assuré en cas de survenance de sinistres définis contractuellement. Les provisions techniques sont composées à la fois des provisions pour primes et des provisions pour sinistres.

En assurance non-vie, il prévaut le régime de survenance : chaque assureur paye aux assurés les sinistres qu'il couvre au moment de la survenance de celui-ci. En revanche, il est fréquent que s'écoulent plusieurs années entre la date de survenance et la date du versement des indemnités. Les assureurs constituent ainsi des provisions afin de procéder à ses règlements de sinistres. En effet, selon l'article 76 de la directive européenne (Parlement Européen, 2009), Il en existe deux types :

- PSAP (Provisions pour Sinistres À Payer) : "Valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés y compris les capitaux constitutifs des rentes non encore mises à la charge de l'entreprise".

- IBNR (Incured But Not Reported) : Il s'agit des sinistres survenus mais non encore déclarés à l'assureur.

Les sinistres peuvent être représentés par plusieurs dates : la date de survenance, la date de déclaration, les dates de paiements et enfin la date de clôture.

Nous symbole "*Best Estimate*" utilisé pour calcul des provisions techniques. Le *Best Estimate* se définit comme : "La moyenne pondérée en fonction de leur probabilité des flux futurs de trésorerie compte tenu de la valeur temporelle de l'argent, laquelle est estimée sur la base d'une courbe de taux sans risque pertinente". Il représente l'estimation la plus juste des provisions de l'assureur, qui lui permettront d'honorer ses engagements. Mathématiquement, il s'agit de l'espérance de la différence des flux futurs sortants et des flux futurs entrants.

En assurance non-vie, le *Best Estimate* se décompose en deux *Best Estimates* :

celui des provisions pour primes et celui des provisions pour sinistres. Les provisions pour primes visent à couvrir le risque que les primes non acquises ou futures soient insuffisantes pour faire face aux sinistres qui ne sont pas survenus à la date d'inventaire mais néanmoins couverts par le contrat. La provision pour sinistres concerne les règlements des sinistres déjà survenus et est principalement constituée de la PSAP (provision pour sinistres à payer).

Le *Best Estimate* des sinistres vise à passer d'une évaluation prudente à une évaluation économique. Il n'existe pas de méthode systématique pour le calcul de cette grandeur mais la nouvelle directive définit un ensemble de principes à respecter pour l'évaluer tels que l'usage de méthodes reconnues notamment celles reposant sur l'utilisation de triangles de liquidation et la justification auprès des autorités de contrôle prudentiel des hypothèses retenues pour le calcul du *Best Estimate*.

1.2.1 Les provisions pour primes non-acquises au Viet Nam

a) Méthodes conformes au pourcentage du total des primes :

- Pour les opérations de marchandises transportées par voies routières, maritimes, fluviaux, ferroviaires et aériens assurance : 25% du total des primes a conservé de l'exercice de l'activité d'assurance.

- Pour les autres activités d'assurance : 50% du total des primes a conservé de l'exercice de l'activité d'assurance.

b) Méthodes conformes à terme en cours de validité du contrat d'assurance :

- Méthode 1/8 : Cette méthode suppose la prime du contrat d'assurance émis dans un quartier est répartie entre les mois du trimestre, ou en d'autres termes, toutes la police d'assurance d'un trimestre donné sont supposés être assurance en vigueur entre ce trimestre. La provision pour primes non acquises est calculée par la formule suivante :

$$\text{Provision pour primes non acquises} = \text{Primes retenues} \times \text{Coefficients primes non acquises}$$

Exemple : La provision pour primes non acquises au moment de l'31/12/n est calculé comme suit :

Pour les contrats d'assurance d'une durée de 1 an et est toujours en vigueur sur 31/12/n :

moment où le contrat d'assurance expire		Coefficients primes non acquises
Année	Trimestre	Coefficient
n+1	I	1/8
	II	3/8
	III	5/8
	IV	7/8

TABLE 1.1 – Les coefficients primes non acquises par trimestre (1 an)

La provision pour primes non acquises est calculée comme suit :

$$\text{Provision pour primes non acquises} = \text{Primes retenues(Trimestre I)} * 1/8 + \text{Primes retenues(Trimestre II)} * 3/8 + \text{Primes retenues(Trimestre III)} * 5/8 + \text{Primes retenues(Trimestre IV)} * 7/8$$

Pour les contrats d'assurance avec dépassement d'une durée une année : dénominateur des coefficients primes non acquises sera égale à la période de validité d'assurance (en années) multiplié par 8.

Provision primes non acquises au moment du contrat d'assurance 31/12/n avec 2 ans la période de validité et reste en vigueur le 31/12/n est calculé comme suit :

moment où le contrat d'assurance expire		Coefficients primes non acquises
Année	Trimestre	Coefficient
n+1	I	1/16
	II	3/16
	III	5/16
	IV	7/16
n+2	I	9/16
	II	11/16
	III	13/16
	IV	15/16

TABLE 1.2 – Les coefficients primes non acquises par trimestre (2 an)

La provision pour primes non acquises est calculée comme ci-dessus.

- Méthode 1/24 : Cette méthode suppose la prime de la police d'assurance émise en un mois de l'activité de l'assurance est répartie sur les mois, ou en d'autres termes, tous les contrats d'assurance d'un mois donné est supposé entrer en vigueur à la mi-mois. Les coefficients primes non acquises sont exprimé de la manière suivante :

Pour les contrats d'assurance d'une durée de 1 an et m ans est toujours en vigueur sur 31/12/n :

moment où le contrat d'assurance expire		Coefficients primes non acquises
Année	mois	Coefficient
n+1	1	1/24
	2	3/24
	3	5/24
	.	.
	.	.
	.	.
	10	19/24
	11	21/24
	12	23/24

TABLE 1.3 – Les coefficients primes non acquises par mois (1 an)

moment où le contrat d'assurance expire		Coefficients primes non acquises
Année	mois	Coefficient
n+1	1	1/(24*m)
	2	3/(24*m)
	3	5/(24*m)
	.	.
	.	.
	.	.
	10	19/(24*m)
	11	21/(24*m)
	12	23/(24*m)
n+2	1	25/(24*m)
	2	27/(24*m)
	3	29/(24*m)
	.	.
	.	.
	.	.
	10	43/(24*m)
	11	45/(24*m)
	12	47/(24*m)
...	.	.

TABLE 1.4 – Les coefficients primes non acquises par mois (m an)

- La méthode de calcul de la provision pour primes non acquises par jour : Cette méthode peut être appliquée pour calculer provision pour primes non acquises tous les temps du contrat selon la formule générale suivante :

$$PPNA = \frac{PR \times NR}{TR}$$

Avec :

- PPNA : La provision pour primes non acquises
- PR : Les primes retenues
- NR : Le nombre de jours restant couvert par le contrat d'assurance
- TR : La couverture de jour total dans le cadre du contrat d'assurance

La provision pour primes non acquises s'articule autour de deux natures de primes :

- Les primes émises : A la souscription ou au renouvellement du contrat liant l'assureur à l'assuré, les primes sont émises pour toute la période de couverture du contrat (souvent un an). Il s'agit de l'engagement de l'assuré envers l'assureur. La prime émise peut coïncider avec l'encaissement de celle-ci elle est intégralement payée dès la signature (dans le cas contraire, la partie non réglée constitue une créance de l'assureur sur l'assuré).
- Les primes acquise : Elles correspondent à la part de prime acquise à l'assureur au fur et à mesure de l'exécution du contrat. Elle constitue le principal produit du compte de résultat d'un assureur non-vie.

Par conséquent, La provision pour primes non acquises (PPNA) est la part des primes que l'assureur doit conserver pour faire face aux risque à venir. Les PPNA se calculent par différent entre les primes émises et les primes acquises. Elles constituent un poste du bilan très important pour les assureurs émettant des contrats pluriannuels. A contrario, elle peut aussi être négligeable (au 31/12) pour des assureurs émettant des contrat annuels avec des dates de renouvellement au 1^{er} janvier.

La provision pour primes non acquises dépend de la durée du contrat d'assurance et selon la nature des différentes branches d'assurance. Elle est choisie en fonction des données et calculé automatiquement dans le logiciel de comptabilité.

1.2.2 Les provisions pour sinistres au Viet Nam

Les provisions pour sinistres correspondent aux montants que l'assureur doit mettre de côté pour régler l'ensemble des sinistres. Celles-ci sont principalement composées des provisions pour sinistres à payer (PSAP) qui correspondent au montant intégral des dépenses nécessaires au règlement des sinistres survenus et non payés. La PSAP est calculée exercice par exercice et se décompose de la façon suivante :

$$\text{PSAP} = \text{Provision Dossier} / \text{Dossier} + \text{IBNER} + \text{IBNYR}.$$

Avec :

— Provision Dossier/Dossier : montant que l'assureur s'attend à payer pour un sinistre déclaré (évaluée par les gestionnaires de sinistres de la compagnie).

— IBNER (Incurred But Not Enough Reported) : montant que l'assureur provisionne pour couvrir le risque d'une sous-évaluation du sinistre lors de la constitution de la provision Dossier/ Dossier.

— IBNYR (Incurred But Not Yet Reported) : montant destiné à couvrir le coût lié aux sinistres déjà survenus mais dont l'assureur n'a pas encore connaissance, appelés tardifs.

Alors que les provisions Dossier/ Dossier sont évaluées au cas par cas dès la déclaration d'un sinistre, les IBNR (IBNER + IBNYR) sont évaluées au global à l'aide de méthodes statistiques de provisionnement.

Le principe de base de l'évaluation des provisions de sinistres est l'évaluation par les gestionnaires de sinistres des sommes restant à payer dossier par dossier. Au montant ainsi obtenu, les sociétés doivent ajouter :

- une estimation pour "sinistres tardifs" (sinistres survenus dans l'année mais non encore déclarés – on utilise souvent l'abréviation anglaise IBNR : "incurred but not reported").

- une provision pour frais de gestion de sinistres internes (que les gestionnaires ne peuvent pas appréhender).

Deux types de données permettent d'évaluer la sinistralité : les règlements qui correspondent au montant déjà versé à l'assuré afin d'indemniser tout ou partie du dommage subi et les charges qui correspondent à l'estimation du coût total du

sinistre pour l'assureur. La charge peut se décomposer de la façon suivante :

$$\text{Charge} = \text{Règlements} + \text{Provision Dossier/Dossier.}$$

Notons que la charge peut évoluer à la hausse ou à la baisse au fur et à mesure de la liquidation de la provision Dossier/Dossier grâce à de nouvelles informations portées à la connaissance de l'assureur. De plus, les frais de gestion des sinistres sont inclus dans les règlements et la provision Dossier/Dossier.

• Notations

- Soit n le nombre total d'années de développement ;
- Soit m le nombre total d'années de survenance ;
- Soit i tel que $1 \leq i \leq m$, l'année de survenance considérée ;
- Soit j tel que $1 \leq j \leq n$, l'année de développement considérée ;
- Soit $C_{i,j}$, le montant cumulé de règlements ou de charges de l'année de survenance i pour l'année de développement j .

• **Remarque** : Si nous considérerons que $n = m$, c'est-à-dire qu'il y a autant d'années de survenance que d'années de développement.

Considérons un triangle dont le nombre d'années de développement et de survenance est 4. Il sera de la forme :

ij	1	2	3	4
1	$C_{1,1}$	$C_{1,2}$	$C_{1,3}$	$C_{1,4}$
2	$C_{2,1}$	$C_{2,2}$	$C_{2,3}$	
3	$C_{3,1}$	$C_{3,2}$		
4	$C_{4,1}$			

TABLE 1.5 – Exemple de triangle de liquidation

Il existe deux types de méthodes de provisionnement attritionnelles : les méthodes déterministes et les méthodes stochastiques. Les méthodes déterministes permettent de déterminer un estimateur du niveau de réserves nécessaire mais elles présentent l'inconvénient de ne fournir aucune information sur la distribution sous-jacente des provisions. Contrairement aux méthodes déterministes, celles stochastiques apportent une estimation de la variabilité des réserves et permettent de définir l'erreur d'estimation ainsi que des intervalles de confiance. Nous présenterons dans la partie suivante les principes de déterministes et stochastiques.

Chapitre 2

Rappel les méthodes déterministes et stochastiques

Les méthodes déterministes permettent d'avoir une estimation de cette provision. L'objectif est alors de savoir à quel point la provision connue a posteriori diffère de l'estimation qui avait été faite auparavant ; les méthodes stochastiques permettent de répondre à cette problématique en évaluant la variabilité de la charge sinistre prévue par le modèle. La propriété fondamentale de ces méthodes est la modélisation des règlements incrémentaux par des variables aléatoires. Des estimations et des intervalles de confiance peuvent alors être obtenus et la sinistralité d'exercices futurs simulée.

2.1 Méthodes déterministes

2.1.1 Méthode de Chain Ladder

La méthode Chain Ladder est une méthode déterministe fréquemment utilisée car facile à mettre en œuvre. Elle s'applique à des triangles de paiements cumulés ou des triangles de charges.

L'objectif de cette méthode est l'estimation des montants futurs de la charge totale des sinistres par année de survenance.

On considère une branche dont les sinistres se déroulent sur $(N + 1)$ années. Nous notons :

- i : l'année d'origine, $i = 0, \dots, N$
- j : le délai de développement, $j = 0, \dots, N$
- x_{ij} : la mesure de sinistralité pour l'année d'origine i et le délai de règlement j .

Une ligne de ces triangles représente les montants versés au cours du temps pour une année de survenance ; ainsi, la première ligne du triangle 2.1 représente les sommes successivement versées par l'assurance au titre des sinistres survenus en 2003. Chaque colonne représente ce qu'on appelle une année de déroulement, le nombre d'année entre la survenance du sinistre et le paiement de l'assuré.

Application au triangle Dommages Auto :

	0	1	2	3	4	5	6	7	8	9	10	11	12
2003	32,23	4,87	0,286	0,144	0,096	0,114	0,027	0,016	0,015	0,012	0,011	0,006	0,002
2004	32,85	5,05	0,365	0,277	0,263	0,215	0,189	0,166	0,153	0,142	0,128	0,008	
2005	34,37	5,54	0,327	0,281	0,228	0,213	0,147	0,144	0,128	0,061	0,013		
2006	31,18	5,02	0,348	0,311	0,249	0,124	0,135	0,124	0,122	0,119			
2007	33,54	6,45	0,259	0,215	0,185	0,125	0,123	0,121	0,12				
2008	34,93	6,02	0,227	0,213	0,165	0,151	0,133	0,129					
2009	36,26	6,25	0,288	0,237	0,226	0,182	0,131						
2010	35,49	6,82	0,351	0,228	0,145	0,134							
2011	34,79	5,19	0,221	0,168	0,148								
2012	32,68	5,85	0,321	0,258									
2013	32,99	5,59	0,348										
2014	35,43	5,71											
2015	36,54												

FIGURE 2.1 – Triangle de paiements décumulés (Dommages Auto)

En date $N + 1$, les paiements des sinistres antérieurs à cette date sont classiquement mis sous la forme d'un triangle de liquidation. Le plus couramment utilisé est le triangle des paiements cumulés, c'est-à-dire le triangle $(C_{i,j})_{i=0\dots N, j=0\dots N}$ avec :

$$C_{ij} = \sum_{h=0}^j x_{ih} \quad (2.1)$$

Ce montant correspond à la somme des paiements cumulés entre l'année de survenance i et l'année de développement j .

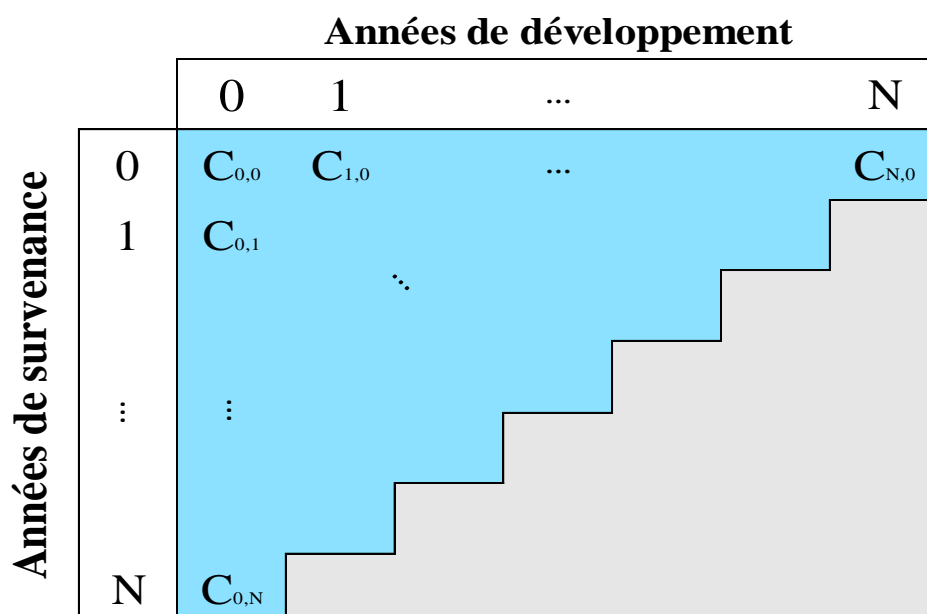


FIGURE 2.2 – Triangle de paiements cumulés

Ensuite, nous appliquons la formule 2.1 :

	0	1	2	3	4	5	6	7	8	9	10	11	12
2003	32,23	37,1	37,386	37,53	37,626	37,74	37,767	37,783	37,798	37,81	37,821	37,827	37,829
2004	32,85	37,9	38,265	38,542	38,805	39,02	39,209	39,375	39,528	39,67	39,798	39,806	
2005	34,37	39,91	40,237	40,518	40,746	40,959	41,106	41,25	41,378	41,439	41,452		
2006	31,18	36,2	36,548	36,859	37,108	37,232	37,367	37,491	37,613	37,732			
2007	33,54	39,99	40,249	40,464	40,649	40,774	40,897	41,018	41,138				
2008	34,93	40,95	41,177	41,39	41,555	41,706	41,839	41,968					
2009	36,26	42,51	42,798	43,035	43,261	43,443	43,574						
2010	35,49	42,31	42,661	42,889	43,034	43,168							
2011	34,79	39,98	40,201	40,369	40,517								
2012	32,68	38,53	38,851	39,109									
2013	32,99	38,58	38,928										
2014	35,43	41,14											
2015	36,54												

FIGURE 2.3 – Triangle de paiements cumulés (Dommages Auto)

Elle peut s'appliquer sur différents types de données (paiements cumulés, charge de sinistres ou encore le nombre de sinistres déclarés ou survenus). La méthode prend en compte les montants de la charge totale des sinistres par année de survenance et suppose une relation explicative entre les années de développement et la sinistralité future.

Cette somme représente la somme versée par l'assureur au titre de sinistres ayant eu lieu l'année i au cours des j années qui ont suivi. La méthode de Chain-Ladder s'applique à des triangles cumulés, elle est basée sur l'utilisation des facteurs de développement, supposés constants pour chaque année et s'appuie sur les hypothèses suivantes.

• Hypothèses

- $\forall j = 0, \dots, N - 1$, $f_j = \frac{C_{i,j+1}}{C_{i,j}}$ sont indépendants de l'année d'origine i .
- $\forall j = 0, \dots, N - 1$, $\frac{C_{0,j+1}}{C_{0,j}} = \frac{C_{1,j+1}}{C_{1,j}} = \dots = \frac{C_{N-j-1,j+1}}{C_{N-j-1,j}}$

Dans la pratique, ces égalités ne sont, au mieux, qu'approximativement vérifiées. Il est donc naturel de choisir l'estimateur du facteur de développement suivant :

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

On en déduit les Coefficients de passage :

j	\widehat{f}_j	j	\widehat{f}_j
0	1,1679	6	1,0029
1	1,0077	7	1,0027
2	1,0058	8	1,0021
3	1,0047	9	1,0012
4	1,0038	10	1,0001
5	1,0031	11	1,0001

TABLE 2.1 – Coefficients de passage (Dommages Auto)

Les facteurs de développement permettent également d'introduire la notion de cadence de règlement : la cadence de règlement p_i signifie qu'une proportion p_i du sinistre a été payée après i années de développement.

$$p_i = \frac{1}{f_i \times \dots \times f_{N-1}} \quad (2.2)$$

L'objectif de cette méthode est l'estimation des montants futurs de la charge totale des sinistres par année de survenance. À partir des facteurs de développement, nous pouvons compléter la partie inférieure du triangle en se servant de l'expression suivante :

$$\forall j \geq N - i + 1, \widehat{C}_{i,j} = C_{i,N-i} \cdot \prod_{h=N-i}^{j-1} \widehat{f}_h \quad (2.3)$$

Ainsi, la méthode Chain-Ladder calcule un facteur de développement moyen à partir des données observées permettant ainsi d'estimer le triangle inférieur de développement. Nous allons utiliser un triangle de paiements cumulés ($C_{i,j}$) et évaluer les facteurs de développement sur les données historiques. Ces derniers représentent le taux de croissance des montants cumulés d'une année à la suivante, elles donnent donc une prévision des montants totaux payés dans le futur.

Ils permettent d'obtenir les provisions pour chaque exercice de survenance, et la provision globale :

-les provisions par exercice de survenance :

$$\widehat{R}_i = \widehat{C}_{i,N} - C_{i,N-i} \quad (2.4)$$

-les provisions totales :

$$\widehat{R} = \sum_{i=1}^N \widehat{R}_i \quad (2.5)$$

Enfin, on obtient la complétion du triangle de paiement :

	0	1	2	3	4	5	6	7	8	9	10	11	12	Provision
2003	32,23	37,1	37,386	37,53	37,626	37,74	37,767	37,783	37,798	37,81	37,821	37,827	37,829	0
2004	32,85	37,9	38,265	38,542	38,805	39,02	39,209	39,375	39,528	39,67	39,798	39,806	39,808	0,002105
2005	34,37	39,91	40,237	40,518	40,746	40,959	41,106	41,25	41,378	41,439	41,452	41,459	41,462	0,009646
2006	31,18	36,2	36,548	36,859	37,108	37,232	37,367	37,491	37,613	37,732	37,78	37,787	37,789	0,056979
2007	33,54	39,99	40,249	40,464	40,649	40,774	40,897	41,018	41,138	41,226	41,279	41,286	41,288	0,150166
2008	34,93	40,95	41,177	41,39	41,555	41,706	41,839	41,968	42,082	42,172	42,226	42,233	42,235	0,267474
2009	36,26	42,51	42,798	43,035	43,261	43,443	43,574	43,701	43,82	43,913	43,97	43,977	43,98	0,405762
2010	35,49	42,31	42,661	42,889	43,034	43,168	43,304	43,43	43,548	43,641	43,697	43,705	43,707	0,538885
2011	34,79	39,98	40,201	40,369	40,517	40,675	40,803	40,922	41,033	41,121	41,173	41,18	41,183	0,665622
2012	32,68	38,53	38,851	39,109	39,294	39,447	39,571	39,687	39,794	39,88	39,93	39,938	39,94	0,830742
2013	32,99	38,58	38,928	39,157	39,342	39,496	39,62	39,735	39,843	39,928	39,979	39,987	39,989	1,060656
2014	35,43	41,14	41,458	41,702	41,9	42,063	42,195	42,318	42,433	42,524	42,578	42,586	42,588	1,447934
2015	36,54	42,676	43,006	43,259	43,464	43,633	43,77	43,898	44,017	44,111	44,167	44,175	44,178	7,637661
														13,07363

FIGURE 2.4 – Complétion du triangle de paiement (Dommages Auto)

Une fois que le triangle de paiement a été complété, il est possible de connaître une estimation des réserves nécessaires. Pour chaque année, le paiement total relatif à une année de survenance se trouve dans la dernière colonne, sur la ligne correspondante.

Néanmoins, elle présente certaines limites qu'il est nécessaire de souligner :

- L'hypothèse forte du modèle selon laquelle la cadence de développement future sera à l'image de celle du passé est difficilement respectée en pratique. En effet, certains phénomènes influent sur le portefeuille à un moment t mais ne sont pas amenés à se reproduire ultérieurement. Parmi eux, nous pouvons citer, à titre d'exemple : des modifications dans la gestion des sinistres, des changements du profil de risque des assurés, l'évolution de la composition du portefeuille. . .

- L'estimation de la charge ultime sur les années plus récentes doit donc être considérée avec beaucoup de prudence. La charge ultime n'est calculée qu'avec un faible recul sur des assiettes de primes parfois très grandes. En effet, l'estimation de la charge ultime est fonction des montants de sinistres déjà déclarés dans le triangle. Or pour une branche ayant des déclarations tardives, par exemple la construction, ces montants sont faibles dans les années les moins développées. L'erreur d'estimation peut donc être fortement amplifiée dans le cas de branches longues.

- Cette méthode paraît plus fiable sur des branches où les risques sont homogènes car elle utilise l'expérience passée des années d'origine les plus anciennes.

Le but de ce mémoire est de créer des modèles pouvant donner une représentation fidèle des distributions des réserves nécessaires. Il s'agit donc d'estimer à partir de triangles de paiements de différentes branches la distribution de la réserve agrégée, le tout en prenant en compte le plus d'informations possible.

Les réserves de Chain-Ladder pour chaque année de survenance sont obtenus par la différence entre l'estimation du montant de la charge totale $i+j$ et le dernier montant connu. En prenant les données de l'entreprise PVI, on calcule les provisions pour l'ensemble des branches comme le montre la table 2.2.

Le modèle déterministe est couramment utilisé par les assureurs et nous donne une estimation des réserves à détenir. Bien que très intéressant à connaître, cette valeur ne donne aucune information sur la distribution des paiements futurs.

Ainsi, nous constatons que pour le triangle de la branche Dommages Auto, 85% des sinistres sont réglés dès la première année et 99% au bout de la troisième. La branche Dommages aux Biens a également une liquidation courte, 90% des paiements étant effectués dès la deuxième année.

En revanche, comme nous pouvions nous y attendre, les branches des Energies ont une liquidation beaucoup plus longue : il faut attendre la cinquième année pour atteindre 95%.

De plus, l'objectif même de l'assureur étant de déterminer le montant des prestations qu'il devra verser dans les années à venir, nous en déduisons le tableau des provisions suivant :

i	Dom.Auto	Dom.Biens	Dom.Energie
1	0	0	0
2	0,201	0,214	0,644
3	0,267	0,249	0,896
4	0,405	0,386	1,125
5	0,538	0,509	1,236
6	0,665	0,633	2,636
7	0,830	0,726	5,859
8	1,060	0,892	7,566
9	1,447	1,218	9,096
10	7,637	6,158	12,054
Total	13,073	10,985	41,112

TABLE 2.2 – Tableau des provisions pour l'ensemble des branches

Le modèle Chain Ladder est simple à appliquer et à comprendre, ce qui explique son succès dans la plupart des compagnies d'assurance. Ces atouts lui permettent d'être fréquemment utilisée en amont par les actuaires afin d'avoir une idée générale du montant. Cependant, Il possède cependant plusieurs limites. La provision dépend fortement du dernier montant de charge connu ; si le triangle est peu stable et irrégulier, ce modèle n'est pas adapté des réserves.

• Validation de modèle

Malgré sa simplicité d'utilisation, il est important d'effectuer certaines vérifications pour valider cette méthode.

Les deux premières hypothèses se justifient de la même manière que pour la méthode de Chain-Ladder : pour tout j fixé, les couples $(C_{i,j}, C_{i,j+1})_{i=0\dots N-j-1}$ doivent être sensiblement alignés sur une droite passant par l'origine.

Afin de valider la dernière hypothèse, il convient d'étudier les résidus standardisés :

$$r_{i,j} = \frac{C_{i,j+1} - \hat{f}_j \cdot C_{i,j}}{\sqrt{C_{i,j}}} \quad (2.6)$$

La représentation graphique des couples $(C_{i,j}, r_{i,j})_{i=0\dots N-j-1}$ ne doit présenter aucune structure non aléatoire.

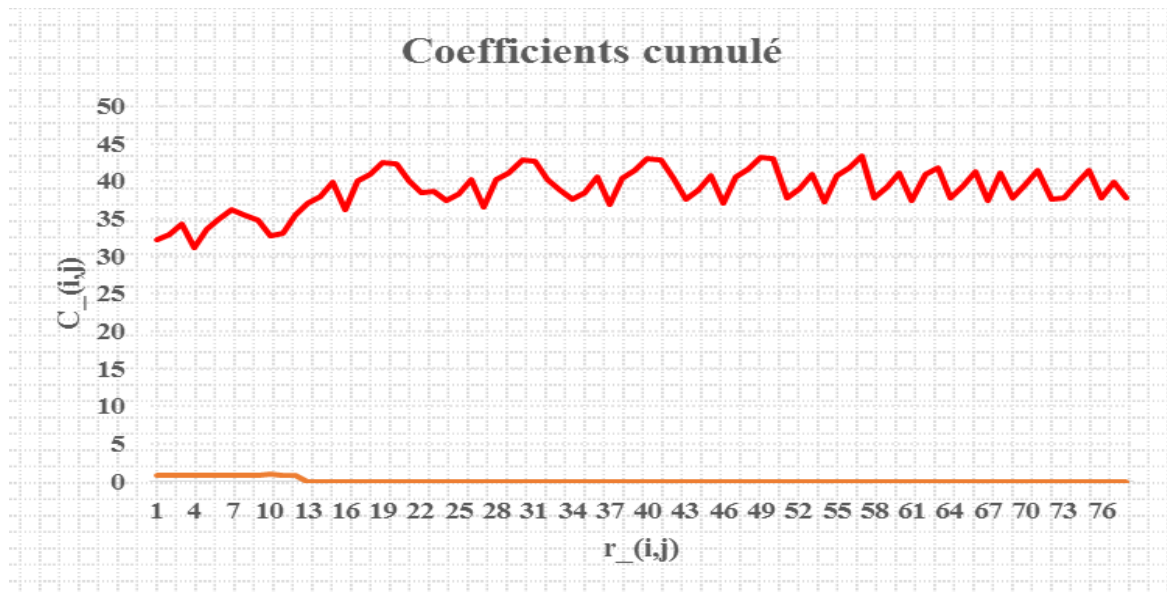


FIGURE 2.5 – Graphiques des couples d'origine

La représentation graphique des couples $(C_{i,j}, r_{i,j})_{i=0\dots N-j-1}$ est sensiblement alignés par une droite passant par l'origine.

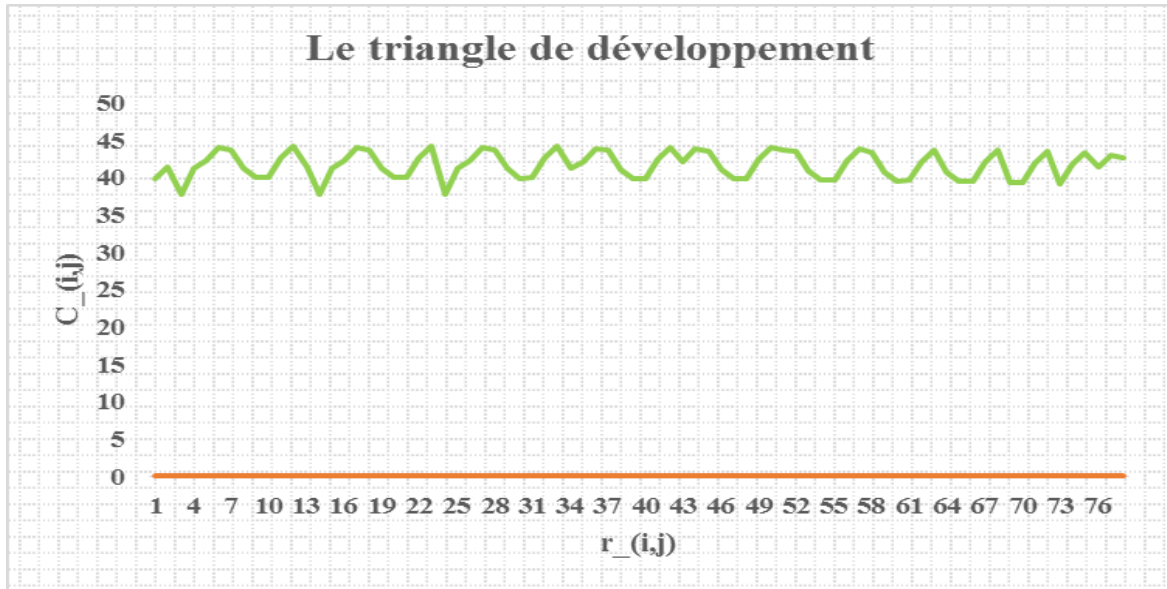


FIGURE 2.6 – Graphiques des couples du triangle de développement

La représentation graphique des couples $(C_{i,j}, r_{i,j})_{i=0 \dots N-j-1}$ est également sensiblement alignés par une droite passant par triangle de développement. Donc, on accepte la validation des conditions de Chain-Ladder.

2.1.2 Méthode de London-Chain

Cette méthode suppose une relation affine entre les montants de la charge totale des sinistres d'une année de développement à une autre.

Cette méthode fait partie des méthodes dites autorégressives qui s'appuient sur une hypothèse moins contraignante que le modèle de CL selon laquelle $C_{i,j+1}$ est une fonction affine de $C_{i,j}$.

Dans la méthode de London Chain, on suppose que la dynamique des $(C_{i,j})$ est de la forme $C_{i,j+1} = \lambda_j \cdot C_{i,j} + \alpha_j$.

Cela signifie que, comme précédemment, cette méthode repose sur l'hypothèse selon laquelle les points $(C_{i,j}, C_{i,j+1})$ sont sensiblement alignés, à la différence que la droite n'est plus censée passer par l'origine, l'hypothèse d'alignement des couples $(C_{i,j}, C_{i,j+1})$ suivant une droite qui passe par l'origine n'est pas toujours validée.

Il s'agit donc d'estimer $2n$ paramètres, à savoir λ_j et α_j pour $j=0 \dots n-1$. Le couple (λ_j, α_j) est estimé par la méthode des Moindres Carrés Ordinaires, en minimisant la fonction suivante :

$$(\lambda_j, \alpha_j) = \underset{i=0}{\operatorname{Argmin}} \sum_{i=0}^{n-j-1} (C_{i,j+1} - \alpha_j - \lambda_j C_{i,j})^2 \quad (2.7)$$

La solution au problème de minimisation est donnée par :

$$\bar{C}_j = \frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j} \quad (2.8)$$

$$\bar{C}_{j+1} = \frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j+1} \quad (2.9)$$

Les estimateurs de λ_j et α_j sont respectivement :

$$\hat{\lambda}_j^{LC} = \frac{\frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j} \cdot C_{i,j+1} - \bar{C}_j \cdot \bar{C}_{j+1}}{\frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j}^2 - \bar{C}_j^2} \quad (2.10)$$

$$\hat{\alpha}_j^{LC} = \bar{C}_{j+1} - \hat{\lambda}_j^{LC} \cdot \bar{C}_j \quad (2.11)$$

On rappelle que l'expression de λ_j^{LC} est aussi égale à : $\frac{Cov(colonne_j, colonne_{j+1})}{Var(colonne_j)}$

Nous estimons les paramètres de notre modèle avec la méthode de moindres carrés ordinaire qui nous donne les résultats suivants :

$\hat{\lambda}$	Coef.L-C	$\hat{\alpha}$	Constante
$\hat{\lambda}_1$	1,41	$\hat{\alpha}_1$	8988,1
$\hat{\lambda}_2$	1,03	$\hat{\alpha}_2$	342,69
$\hat{\lambda}_3$	1,08	$\hat{\alpha}_3$	-1328,4
$\hat{\lambda}_4$	1,17	$\hat{\alpha}_4$	-9107,5
$\hat{\lambda}_5$	0,62	$\hat{\alpha}_5$	12376,4
$\hat{\lambda}_6$	1,00	$\hat{\alpha}_6$	0

TABLE 2.3 – Coefficients de la méthode London-Chain

Nous présentons ci-dessus les résultats obtenus par la méthode de London-Chain. De même que pour les techniques présentées précédemment, une fois avoir estimé les paramètres, nous complétons la partie inférieure du triangle des paiements cumulés puis nous en déduisons la provision totale à constituer. Le montant total des réserves Dom.Auto estimé par la méthode London-Chain est 11,718 M\$.

• Validation des hypothèses

Nous utilisons le test de Student pour tester l'hypothèse suivante :

$$\begin{cases} H_0 : \alpha_j = 0 & (\text{le facteur n'est pas significatif}) \\ H_1 : \alpha_j \neq 0 & (\text{le facteur est significatif}) \end{cases} \quad (2.12)$$

Nous allons introduire une variable aléatoire dans le modèle, qui permet de décrire l'erreur de spécification :

$$\varepsilon_j = C_{i,j+1} - \hat{\lambda}_j C_{i,j} - \hat{\alpha}_j \quad (2.13)$$

Cette variable aléatoire est supposée suivre une loi normale : $\varepsilon_j \sim N(0, \sigma_\varepsilon^2)$
 Cette condition peut se vérifier à l'aide d'un test de normalité des erreurs.

L'étape suivante va donc consister en l'estimation de cette variance. Il vient :

$$\widehat{Var}(\hat{\alpha}_j) = \widehat{Var}(\varepsilon_j) \left(\frac{1}{n-j} + \frac{\bar{C}_j^2}{\sum_{i=0}^{n-j-1} (C_{i,j} - \bar{C}_j^2)} \right) \quad (2.14)$$

Avec :

$$\bar{C}_j = \frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j} \quad (2.15)$$

$$\widehat{Var}(\varepsilon_j) = \frac{1}{n-j-p} \sum_{i=0}^{n-j-1} \varepsilon_j^2 \quad (2.16)$$

Où p est le nombre de variables explicatives.

La statistique du test est donnée par :

$$T = \frac{\hat{\alpha}_j - \alpha_j}{\sqrt{\widehat{Var}(\hat{\alpha}_j)}} \quad (2.17)$$

Du fait de la normalité des erreurs, la statistique de test T suit une loi de Student à $(n - j - p)$ degrés de liberté .(i.e. $T \sim S_{n-j-p}$).
sous l'hypothèse H_0 :

$$T = \frac{\hat{\alpha}_j}{\sqrt{\widehat{Var}(\hat{\alpha}_j)}} \sim S_{n-j-p} \quad (2.18)$$

- Si $T > t_{\alpha;n-j-p}$ on rejette l'hypothèse H_0 avec une marge d'erreur α . Autrement dit, le facteur incrémental est significativement différent de zéro et par conséquent il joue un rôle dans le modèle.

- Inversement, si $T < t_{\alpha;n-j-p}$ on accepte l'hypothèse H_0 .

Le test de significativité de Student nous a indiqué que l'utilisation de la méthode London-Chain n'est pas le choix le plus approprié à nos données. Après avoir utilisé les estimateurs de cette méthode, les résultats de la provision finale à constituer montre un écart très important avec la valeur estimée par la méthode Chain-ladder. Ce qui confirme l'invalidité des estimations de London-Chain et confirme aussi les résultats du test de Student.

Nous devons mentionner que cette méthode a montré des résultats plus cohérents quand elle est appliquée sur un autre portefeuille de données dans la société. De la même manière que pour la méthode de Chain-Ladder, cette approche requiert une grande stabilité au sein du portefeuille étudié.

Elle lui sera néanmoins privilégiée si l'ajout du facteur incrémental α_j permet un meilleur ajustement du modèle à nos données, une information qui sera nécessaire de valider à l'aide de tests statistiques.

2.1.3 Méthode de Bornhuetter-Fergusson

Une méthode qui a été présentée par Bornhuetter et Ferguson (BF) en 1972 dans leur article "The Actuary and IBNR", où ils ont introduit une nouvelle approche pour l'estimation des sinistres tardifs, mais son principe peut être appliqué plus généralement. Là où pour Chain-Ladder le montant de la charge ultime dépend de la dernière position connue (si elle est nulle, la charge ultime est nulle), la méthode de BF introduit une donnée exogène pour palier à ce problème. Le modèle de BF est une alternative très répandue du modèle de Chain-Ladder et est souvent utilisé pour les triangles dits « instables ».

La méthode de Bornhuetter-Fergusson repose sur des hypothèses :

1. Les montants cumulés $C_{i,j}$ sont indépendants suivant les années de surveillance i .

2. Il existe des paramètres $\alpha_0, \dots, \alpha_I > 0$ et des cadences de paiements cumulés

$\beta_i^{BF}, \dots, \beta_J^{BF} > 0$ avec $\beta_J^{BF} = 1$ tels que $\forall 0 \leq i \leq I, 0 \leq j \leq J - 1$, et

$\forall 0 < k \leq J - j$ nous ayons :

$$E(C_{i,0}) = \beta_0^{BF} \alpha_i \quad (2.19)$$

$$E(C_{i,J}) = \alpha_i \quad (2.20)$$

Sous ces hypothèses, pour tout i, j , $E(C_{i,j}) = \beta_j^{BF} \alpha_i$. Ce qui peut rappeler les modèles à facteurs évoqués auparavant.

L'estimateur de la méthode "Bornhuetter-Ferguson" est donné par :

$$\hat{C}_{i,j+1} = C_{i,I-i} + \hat{\alpha}_i(1 - \hat{\beta}_{I-i}^{BF}), \quad \forall 0 \leq i \leq I \quad (2.21)$$

Où :

- $\hat{\beta}_{I-i}^{BF}$ est un estimateur a posteriori de β_{I-i}^{BF}
- $\hat{\alpha}_i$ est un estimateur a priori de la charge ultime espérée $E(C_{i,j+1})$

L'estimateur $\hat{\beta}_{I-i}^{BF}$ peut être déterminé en recourant à l'estimation des facteurs de développement issue du modèle de Chain-Ladder :

$$\hat{\beta}_j^{BF} = \prod_{k=j}^{J-1} \frac{1}{\hat{f}_k} \quad (2.22)$$

• Application au Dom.Auto Chain-ladder

Année	loss ratios ultimes(Estimé en J)	loss ratios ultimes(Estimé en J-1)
2003	78,3%	78,3 %
2004	76,4%	76,4%
2005	74,8%	74,8%
2006	77,3%	77,3%
2007	79,2%	78,7%
2008	83,5%	81,2%
2009	72,5%	71,2%
2010	73,5%	74,2%
2011	79,9%	78,4%
2012	82,6%	81,7%
2013	83,7%	84,5%
2014	83,3%	81,4%
2015	80,6%	

TABLE 2.4 – Loss Ratio ultimes Bornhuetter-Ferguson du triangle Dom.Auto

L'estimateur $\hat{\alpha}_i$ est une information exogène qui sera fournie par un expert. Elle est déterminée à partir d'un certain "Loss-ratio" qui donnera la valeur ultime de la sinistralité de chaque année de survenance.

La réserve de cette méthode sera obtenue par la différence entre le coût total estimé et la dernière valeur observée :

$$\hat{R}_i = \hat{C}_{i,J} - C_{i,I-i} = \hat{\alpha}_i(1 - \hat{\beta}_{I-i}^{BF}) \quad (2.23)$$

Le montant total des réserves sera la somme des réserves calculés pour chaque année de survenance :

$$\hat{R} = \sum_{i=0}^I \hat{R}_i \quad (2.24)$$

Cette approche présente l'avantage d'introduire un facteur exogène au moyen de l'avis d'un professionnel et ainsi d'être moins dépendante des données observables.

Un désavantage ou difficulté de cette méthode est qu'elle utilise un loss ratio attendu qui n'est pas toujours connu. Les résultats dépendent fortement de ce ratio ainsi que de l'aliment choisi. En revanche, et contrairement à la méthode du « loss ratio », cette méthode tient compte des sinistres déjà connus dans le calcul de la charge ultime de sinistres à l'aide d'une cadence de développement.

• **Méthode de la moyenne arithmétique des "Bornhuetter Ferguson"**

Nous nous sommes proposés de considérer le "Loss-ratio" ultime comme la moyenne des $\frac{S}{P}$ estimés des dix années de survénance, qui sont les données disponibles dans notre étude.

Étant donné un ensemble $X = x_1, \dots, x_n$ ainsi que les poids non-négatifs correspondants $W = w_1, \dots, w_n$.

Nous définissons la moyenne arithmétique \bar{x} par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (2.25)$$

La moyenne arithmétique pondérée \bar{x} est calculée suivant la formule :

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \quad (2.26)$$

Nous devons par la suite calculer la charge totale ultime pour chaque année de survénance i en appliquant la formule suivante :

$$S_i^{ultimate} = \left(\frac{S}{P}\right)^{ultimate} P_i \quad (2.27)$$

Avec P_i le montant de la prime acquise relatif à l'année de survénance i .

Après avoir calculé les $S_i^{ultimate}$ nous ajustons ces montants par les facteurs de développements \hat{f}_k obtenus par la méthode Chain-Ladder . Ces facteurs vont nous permettre de déterminer la valeur réelle des réserves qui sera la différence entre ce que nous estimons par une méthode ordinaire comme Chain-Ladder et ce que l'avis d'expert nous donne comme charge totale ultime. Les estimations (qui sont exogènes) nous informent sur ce que la sinistralité de la branche peut atteindre à la fin de J années de développement.

Afin d'obtenir les valeurs des S_j^{ajust} , nous avons eu recours à l'expression suivante :

$$S_i^{ajust} = \frac{S_i^{ultimate}}{\prod_{k=I-i+1}^{I-1} \hat{f}_k} \iff S_i^{ajust} = \frac{\left(\frac{S}{P}\right)^{ultimate} P_i}{\prod_{k=I-i+1}^{I-1} \hat{f}_k} \quad (2.28)$$

Le montant de la réserve totale \hat{R} est estimé à partir des valeurs des S_i^{ajust} et des $S_i^{ultimate}$:

$$\hat{R} = \left(\sum_{k=2}^I S_i^{ultimate}\right) - \left(\sum_{k=1}^I S_i^{ajust}\right) \quad (2.29)$$

Nous obtenons les valeurs du montant de la provision totale suivantes :

Méthode de Bornhuetter-Furgeson	S/P	Provision totale en M\$
loss ratios ultimes	68,2%	14,457
La moyenne arithmétique sur 10 ans	76,4%	16,073

TABLE 2.5 – Provisions Dom.Auto estimées par la méthode de Bornhuetter-Furgeson

Les estimations données par les différents choix de "Loss-ratio" s'approchent des résultats de notre méthode de référence. L'estimation de la provision totale avec une moyenne arithmétique des "Loss-ratio" sur les dix années récentes donne une provisions totale supérieure à celle donnée par la méthode Chain-Ladder.

Le tableau ci-dessus montre que le choix du "Loss ratio" a une influence directe sur le montant de la provision totale estimée par la méthode Bornhuetter-Ferguson. Ainsi, l'avis d'expert doit être bien évalué et doit refléter une connaissance solide de la structure du portefeuille risques de la société.

L'actuaire apporte son expérience dans l'estimation de la charge ultime a priori, ce qui rend l'application de cette méthode spécifique à chaque actuaire. La provision estimée paraît plus adéquate que la méthode Chain Ladder car elle est propre à la branche étudiée et indépendante de la dernière charge connue. Ainsi, elle est souvent supérieure à la charge sinistre estimée par Chain Ladder mais permet d'en réduire l'instabilité.

2.1.4 Méthode de Cape Cod

Cette méthode est en quelque sorte une extension de la méthode de Bornhuetter-Ferguson. Elle tente de s'affranchir du caractère subjectif de la détermination a priori de la charge ultime de sinistres.

Cette méthode estime et applique un Loss Ratio $\beta_{\omega_s}^{CC}$ sur un groupe ω_s d'années d'origine a priori semblables.

Avec ω_s le s^{me} groupe d'années d'origine comparables, $s = 1, 2, \dots, S$ et S le nombre de groupe d'années d'origine comparables.

Reprenons les hypothèses du modèle de Bornhuetter-Ferguson pour poser celles du modèle Cape Cod :

1. Les montants cumulés $C_{i,j}$ sont indépendants suivant les années de surveillance i .
2. Il existe des paramètres $\alpha_0, \dots, \alpha_I > 0$ et des cadences de paiements cumulés

$\beta_i^{CC}, \dots, \beta_J^{CC} > 0$ avec $\beta_J^{CC} = 1$ tels que $\forall 0 \leq i \leq I, 0 \leq j \leq J - 1$, nous ayons :

$$E(C_{i,0}) = \beta_0^{CC} \alpha_i \quad (2.30)$$

$$E(C_{i,J}) = \alpha_i \quad (2.31)$$

Les facteurs de développement $\beta_{i,j}^{CC}$:

$$\beta_{i,j}^{CC} = \frac{E[C_{i,J}]}{E[C_{i,j}]} \quad (2.32)$$

L'estimation de la charge ultime devient donc :

$$\begin{aligned} \hat{C}_{i,J}^{CC} &= C_{i,j} + \left(1 - \frac{1}{\beta_{i,j}^{CC}}\right) \times E(\hat{C}_{i,J}^{\omega_s}) \\ &= C_{i,j} + \left(1 - \frac{1}{\beta_{i,j}^{CC}}\right) \times L_{\omega_s}^{CC} \times \alpha_i \end{aligned} \quad \text{pour tout } i \in \omega_s$$

De plus, sous hypothèse d'indépendance entre les années d'origine au sein d'un même groupe ω_s on pose :

$$\beta_{i,j} = \beta_j$$

La somme des provisions pour le s^{me} groupe peut alors s'écrire :

$$\sum_{i \in \omega_s} R_i^{CC} = \beta_{\omega_s}^{CC} \times \sum_{i \in \omega_s} \left(1 - \frac{1}{\beta_{j-i+1}}\right) \times \alpha_i \quad (2.33)$$

et la somme totale des provisions :

$$R^{CC} = \sum_{s=1}^S \sum_{i \in \omega_s} R_i^{CC} \quad (2.34)$$

De même que pour les méthodes Loss Ratio et Bornhuetter Ferguson, les provisions correspondent à une part non absorbée de l'aliment multipliée par un niveau de risque.

L'estimation du Loss ratio dans la méthode Cape Cod découle de l'estimation de l'ultime de cette méthode :

$$\sum_{i=1}^I C_{i,j-i+1} + \sum_{i=1}^I R_i = \sum_{i=1}^I (\beta_i \times \alpha_i) \text{ avec } \beta_i = \beta_{\omega_s}^{CC} \text{ pour } i \in \omega_s \quad (2.35)$$

Dès lors pour le s^{me} groupe :

$$\sum_{i \in \omega_s} (\beta_i \times \alpha_i) = \beta_{\omega_s}^{CC} \sum_{i \in \omega_s} \alpha_i \quad (2.36)$$

Ainsi :

$$\sum_{i \in \omega_s} C_{i,j-i+1} + \beta_{\omega_s}^{CC} \times \sum_{i \in \omega_s} \left(1 - \frac{1}{\beta_{j-i+1}}\right) \times \alpha_i = \beta_{\omega_s}^{CC} \sum_{i \in \omega_s} \alpha_i \quad (2.37)$$

$$\sum_{i \in \omega_s} C_{i,j-i+1} = \beta_{\omega_s}^{CC} \times \sum_{i \in \omega_s} \left(\frac{\alpha_i}{\beta_{j-i+1}}\right) \quad (2.38)$$

$$\beta_{\omega_s}^{CC} = \frac{\sum_{i \in \omega_s} C_{i,j-i+1}}{\sum_{i \in \omega_s} \left(\frac{\alpha_i}{\beta_{j-i+1}}\right)} \quad (2.39)$$

Cette méthode nécessite elle une bonne connaissance de son portefeuille afin de pouvoir effectuer les segmentations adéquates des années de souscription semblables. Elle est efficace sur des branches ayant des politiques de souscription homogènes.

Les ultimes ainsi obtenus sont ci-dessous :

Année	loss ratios ultimes(Estimé en N)	loss ratios ultimes(Estimé en N-1)
2003	78,3%	78,3 %
2004	76,4%	76,4%
2005	74,8%	74,8%
2006	77,3%	77,3%
2007	79,2%	78,7%
2008	83,5%	81,2%
2009	72,5%	71,2%
2010	73,5%	74,2%
2011	79,9%	78,4%
2012	82,4%	81,3%
2013	83,9%	85,1%
2014	83,8%	81,7%
2015	81,2%	

TABLE 2.6 – Loss Ratio ultimes Cape Cod du triangle Dom.Auto

On obtient ici un montant de provision total supérieur à celui obtenu par la méthode Chain Ladder et la méthode Bornhuetter-Ferguson puisque $\text{sum}(R)$ vaut ici 16,217 M\$.

La méthode traditionnelle de Cape Cod utilise le loss ratio comme une donnée endogène, laquelle est déterminée à partir du triangle des charges de sinistres. La spécificité de cette méthode tient à l'application d'un unique loss ratio à un groupe d'années homogènes.

En comparant les écarts types relatifs (en pourcentage du sinistre moyen obtenu) entre les résultats des méthodes Cape Cod et Bornhuetter-Ferguson, la variabilité est similaire sur les années ayant plus d'une année de développement. Sur l'année n'ayant qu'une année de développement, la méthode de Chain-Ladder comme la méthode de London-Chain obtient des résultats beaucoup plus stables.

2.2 Méthodes stochastiques

2.2.1 Le modèle de Mack

Le modèle de Mack (1993) est un des premiers modèles stochastique qui a reproduit les estimations de Chain ladder. En effet, le montant des provisions estimé est identique. Il a cependant l'avantage d'estimer une erreur de prédiction des provisions.

Cette méthode est fondée sur les trois hypothèses suivantes :

H1- L'indépendance des années d'origine :

$C_{i,0}, \dots, C_{i,n}$ et $C_{k,0}, \dots, C_{k,n}$ sont indépendants pour $i \neq k$

H2- Il existe des facteurs f_j tel que :

$$E(C_{i,j+1}/C_{i,1}, \dots, C_{i,n}) = f_j C_{i,j} \text{ pour } 0 \leq i \leq n, 0 \leq j \leq n$$

H3- Il existe σ_j tel que :

$$Var(C_{i,j+1}/C_{i,1}, \dots, C_{i,n}) = \sigma_j^2 C_{i,j} \text{ pour } 0 \leq i \leq n, 0 \leq j \leq n$$

Soit $D = \{C_{i,j}/i + j \leq n\}$ l'ensemble des variables observables.

Sous les hypothèses ci-dessus, $E(C_{i,n}/D) = C_{i,n-i} f_{n-i} \dots f_{n-1}$

Dans ce modèle, les facteurs (f_j) sont estimés par les facteurs de développement de Chain Ladder (\hat{f}_j) qui sont sans biais et non corrélés.

Nous calculons les coefficients de développement estimés par le modèle de Mack, qui sont les mêmes coefficients estimés par la méthode Chain- Ladder. Thomas Mack dans son modèle, a calculé la variance $\hat{\sigma}_j^2$ associée aux estimations des facteurs de développements. Elle est donnée par l'expression suivante :

$$\begin{cases} \hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, 0 \leq j \leq n-2 \\ \hat{\sigma}_{n-1}^2 = \min(\hat{\sigma}_{n-2}^4 / \hat{\sigma}_{n-3}^2, \min(\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2)) \end{cases} \quad (2.40)$$

Ces estimateur est également sans biais .

Les estimateurs des facteurs de développement de Chain-Ladder étant sans biais et non corrélés, le modèle de Mack permet de satisfaire l'égalité :

$$\begin{aligned} E[\widehat{C}_{i,N}] &= E[E[\widehat{f}_{N-i} \times \dots \times \widehat{f}_{N-1} \times C_{i,N-i} \mid D_N]] \\ &= E[E[C_{i,N} \mid D_N]] \\ &= E[C_{i,N}] \end{aligned}$$

Et comme $R_i = C_{i,N} - C_{i,N-i}$ alors $E[R_i \mid D] = E[\widehat{C}_{i,N} \mid D] - C_{i,N-i+1}$

Nous pouvons alors calculer l'écart quadratique moyen (mean squared error of prediction) qui mesure l'incertitude de prédiction. Pour les provisions par années d'exercice, il est défini par $MSEP(\widehat{R}_i) = E[(\widehat{R}_i - R_i)^2/D]$.

Le modèle de Mack est très utilisé car il fournit les mêmes estimations de réserves que la méthode de Chain-Ladder et la troisième hypothèse introduite (vision stochastique) permet également au modèle d'évaluer l'erreur de prédiction.

Sous les hypothèses énoncées et si $\widehat{C}_{i,n-i} = C_{i,n-i}$, on peut l'estimer par :

$$\widehat{MSEP}(\widehat{R}_i) = \widehat{C}_{i,n}^2 \sum_{j=n-i}^{n-1} \frac{\widehat{\sigma}_j^2}{\widehat{f}_j} \left(\frac{1}{\widehat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{n-j} C_{i,k}} \right), i = 1, \dots, n \quad (2.41)$$

L'estimation du MSEP de la provision totale est alors donnée par :

$$\widehat{MSEP}(\widehat{R}) = \sum_{i=1}^n \left\{ \widehat{MSEP}(\widehat{R}_i) + \widehat{C}_{i,n} \left(\sum_{k=i+1}^n \widehat{C}_{k,n} \right) \sum_{j=n-i}^{n-1} \frac{2\widehat{\sigma}_j^2}{\widehat{f}_j^2 \sum_{u=0}^{n-j} C_{u,j}} \right\} \quad (2.42)$$

On peut aussi donner l'erreur standard relative i.e. l'écart-type rapporté au montant de charge ultime $\frac{\widehat{MSEP}(\widehat{R})}{\widehat{R}}$

Afin de tester la validité de l'hypothèse, nous devons estimer les résidus de notre modèle. Par la suite, nous vérifions graphiquement l'existence d'une tendance spécifique de ces résidus. L'expression des résidus estimés par la méthode des moindres carrés est :

$$r_{i,j} = \frac{C_{i,j+1} - \widehat{f}_j \cdot C_{i,j}}{\sqrt{C_{i,j}}} \quad (2.43)$$

Nous utilisons les facteurs de développement \widehat{f}_j estimés dans la partie précédente .

• **Application aux Dommages Auto :**

Le nuage de points $\left\{ C_{i,j}, \frac{C_{i,j+1} - \widehat{f}_j C_{i,j}}{\sqrt{C_{i,j}}} \right\}$ des données de l'assureur se présente ainsi :

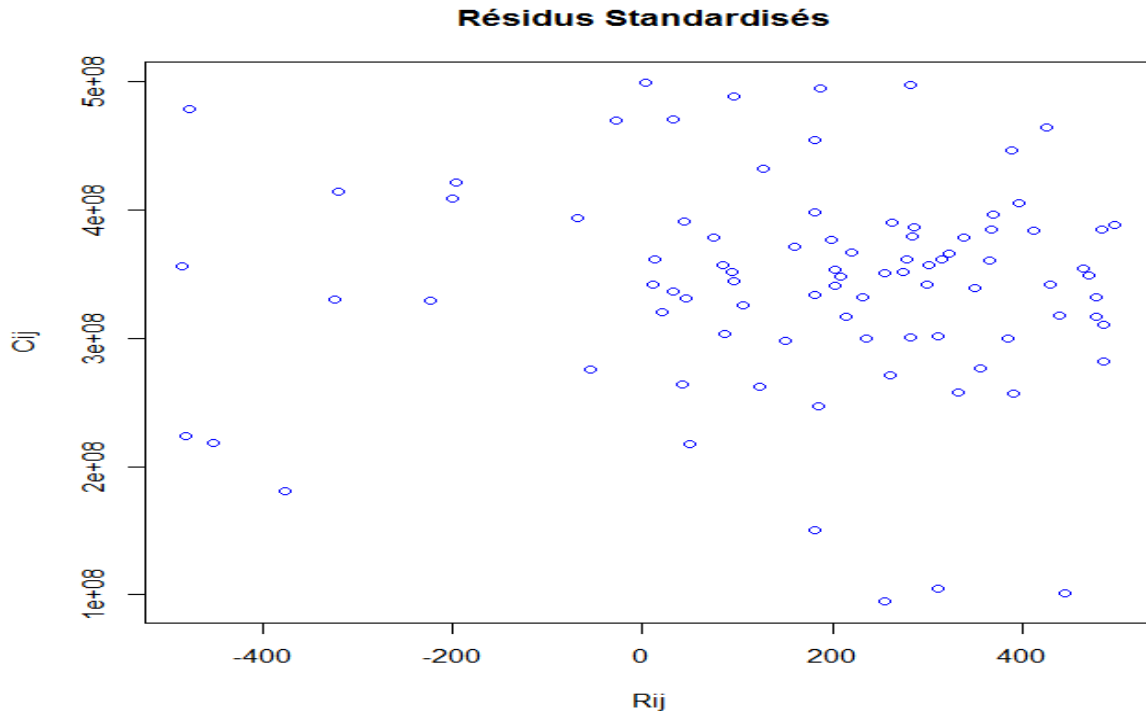


FIGURE 2.7 – Représentation graphique des résidus standardisés

Graphiquement, l'hypothèse est validée parce que les points $r_{i,j} = \frac{C_{i,j+1} - \hat{f}_j \cdot C_{i,j}}{\sqrt{C_{i,j}}}$ ne présentent pas une tendance spécifique, il convient donc de montrer que les résidus standardisés présentent une structure aléatoire.

Le nuages de points ne peuvent exprimer toute la tendance des sinistres payés. La concentration du nuage de points comme concentrée près de 0, les hypothèses du modèle Mack de plus en plus grande fiabilité.

Ensuite, nous calculons la variance $\hat{\sigma}_j^2$:

j	$\hat{\sigma}_j^2$	j	$\hat{\sigma}_j^2$
0	0,049	6	$1,091^{-07}$
1	$3,215^{-04}$	7	$1,387^{-07}$
2	$7,494^{-05}$	8	$1,757^{-08}$
3	$7,458^{-05}$	9	$2,345^{-09}$
4	$7,965^{-05}$	10	$3,221^{-09}$
5	$4,344^{-06}$	11	$2,354^{-10}$

FIGURE 2.8 – L'estimation de l'écart-type proposée par Mack

Le modèle de Mack permet donc de compléter le modèle de provisionnement de Chain-Ladder par le calcul des volatilités des paiements cumulés. Après avoir calculer la charge ultime, Les réserves estimées pour chaque année de survenance et MSEP sont :

i	\hat{R}_i	$\widehat{MSEP}(\hat{R}_i)$	$\sqrt{\widehat{MSEP}(\hat{R}_i)/\hat{R}_i}$
1	0,002104634	1,0206 ⁻⁰³	52,82%
2	0,00964631	3,1175 ⁻⁰³	18,21%
3	0,056978901	0,0211	13,41%
4	0,150165714	2,6161	56,23%
5	0,267473695	25,1149	42,55%
6	0,405762354	36,9681	39,83%
7	0,538884918	40,9611	32,76%
8	0,66562193	58,5419	27,58%
9	0,830741943	556,6855	22,35%
10	1,060656293	1991,4683	21,59%
11	1,447933558	11810,2243	19,87%
12	7,637660829	620948,0325	18,75%
total	13,07363108	635470,6418	18,16%

TABLE 2.7 – Mean squared error of prediction

Le modèle a une approche récursive semblable à celle du modèle binomial négatif et celle du modèle Normal. Le modèle de Mack a la même structure que l'approximation normale du modèle binomial négatif.

Sous le modèle de Mack, des hypothèses limitées sont faites comme par exemple sur la distribution des données sous-jacentes. Seuls les premiers deux moments sont spécifiés. La moyenne et la variance de C_{ij} sont :

Moyenne = $\lambda_j C_{i,j-1}$ et Variance = $\sigma_j^2 C_{i,j-1}$ respectivement.

Nous notons que la moyenne et la variance de C_{ij} sous le modèle du Mack sont semblables à les celles de C_{ij} dans l'approximation Normale du modèle binomial négatif. Il s'agit ici de l'incertitude à l'ultime, nous adoptons donc une vision à long terme. En outre, la volatilité à l'ultime peut être estimée à l'aide de méthodes simulatoires de type Bootstrap et sera présenté dans la section suivante.

2.2.2 Le modèle Log-normal

Modèle Log-Normal a été présenté par Kremer(1982) et utilisé par Renshaw(1989), Verrall(1989), Zehnwirth(1989) et Christofides(1990) entre autre. L'usage de ce modèle produit normalement des valeurs prévues proches de celles de la technique de Chain Ladder simple. Mais ce n'est pas garanti, et il peut y avoir des différences.

Nous allons utiliser une loi log-normale pour modéliser nos données, qui nous permet de résoudre les deux derniers problèmes soulevés : la loi log-normale est positive, asymétrique, avec une skewness strictement positive et sa queue est épaisse. Sa fonction d'excès moyen est croissante. Elle possède une très grande flexibilité pour la modélisation.

Nous supposons donc que les incréments $(X_{i,j})$ suivent une loi log-normale de paramètres μ, ω^2 (cela revient à dire que le logarithme de cette loi suit une loi normale d'espérance μ et de variance ω^2) où

$$\mu_{i,j} = \alpha_i + \beta_j + h \quad (2.44)$$

Soit la variable $Y_{i,j} = \ln(X_{i,j})$ (appelée log-incrément).

Ainsi : $E(X_{i,j}) = e^{\mu_{i,j} + \frac{\omega^2}{2}}$ et $var(X_{i,j}) = e^{2\mu_{i,j} + \omega^2}(e^{\omega^2} - 1)$. La régression Log Normale est donc un modèle linéaire Normal appliqué aux log-incréments du triangle : $Y = M\zeta + \epsilon$ où :

-Y est le vecteur des log-incréments du triangle pris ligne à ligne (de taille N)

-M est la matrice de régression. Elle correspond à la matrice Jacobienne de la transformation $\mu : \zeta \rightarrow \mu(\zeta) = \mu_{ij}$ définie par

$$\begin{cases} \frac{\partial \mu_{ij}}{\partial h} &= 1 \\ \frac{\partial \mu_{ij}}{\partial \alpha_k} &= 1 \text{ si } k = i, 0 \text{ sinon} \\ \frac{\partial \mu_{i,j}}{\partial \beta_l} &= 1 \text{ si } l = j, 0 \text{ sinon} \end{cases} \quad (2.45)$$

$\zeta = (h, \alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n)$ le vecteur des paramètres de régression (de taille p)

ϵ le vecteur des erreurs, distribué selon une loi $N(0, \omega^2)$.

L'estimateur sans biais du maximum de vraisemblance de ζ est $\hat{\zeta} = (M' M)^{-1} Y$.

Notons (e_i) le résidu, estimation de l'erreur (ε_i) ; $e_i = Y_i - \hat{Y}_i$ où \hat{Y}_i est la i^{me} valeur prévue par le modèle.

L'estimateur sans biais de ω^2 est $S^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2$.

L'estimation des incréments est alors donnée par $\hat{\mu}_{ij} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + S^2/2}$ et donc celle des provisions est $\hat{R} = \sum_{i+j > n} \hat{\mu}_{ij}$.

• Risque d'estimation et de prédiction

La matrice de variance-covariance des provisions $\hat{\Sigma}_{\hat{R}}$ est estimée à l'aide de la méthode Delta qui sera présenté dans la section suivante.

L'erreur d'estimation des provisions, erreur commise dans l'estimation des paramètres du modèle, est donnée par :

$$\hat{\omega}_{\hat{R}} = \sqrt{\sum_{i,j} (\hat{\Sigma}_{\hat{R}_{ij}})}$$

Enfin, nous obtenons l'erreur de prédiction :

$$\widehat{sep}_{\hat{R}} = \sqrt{V(\hat{R}) + \hat{\omega}_{\hat{R}}^2}$$

Nous avons obtenu les résultats des branches d'assurance caractérisées :

Branche	Dom.Auto	Dom.Biens	Dom.Habitation
Provisions	14,25641	11,69584	18,22461
sep	97,2%	37,1%	61,4%
S^2	0,538	0,339	0,213

TABLE 2.8 – Tableau des provisions, erreurs relatives et paramètres de dispersion

• Validation du modèle par examen des résidus

Nous utiliserons les résidus studentisés pour test d'indépendance des résidus définis comme suit :

$$e_i^* = \frac{e_i}{s\sqrt{1-m_i}} \quad (2.46)$$

où :

-s est l'estimation de ω l'erreur standard du modèle

- m_i est le i^{me} élément de la diagonale de la matrice de Hat, matrice de projection des Y_i sur les \hat{Y}_i : $H = M(M'M)^{-1}M'$ Leur étude permet en particulier de détecter les valeurs «atypique», observations pour lesquelles $|e_i^*| > 2$.

Nous utiliserons un test de Kolmogorov-Smirnov pour vérifier l'hypothèse de normalité des résidus. Application aux Dommages Auto :

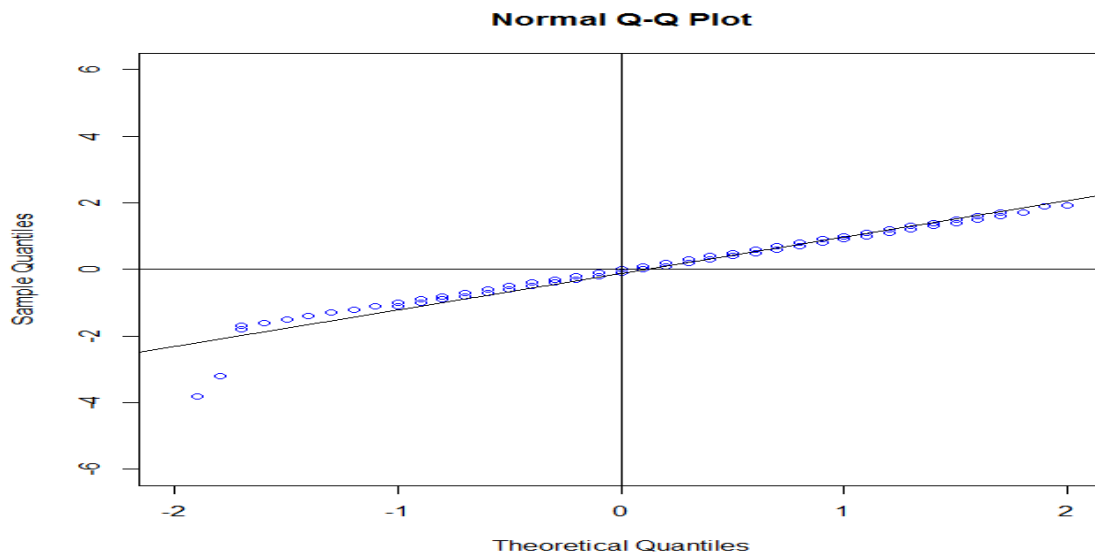


FIGURE 2.9 – le graphique QQ-plot sur les résidus

Le graphique QQ-plot sur les résidus est également sensiblement alignés des points sur une droite de pente positive passant par 0.

Test	p-value	acceptation/rejet (seuil à 5%)
Le test d'indépendance des résidus	0,7215	on accepte
Le test de normalité des résidus	0,0358	on rejette

FIGURE 2.10 – Le test d'indépendance et de normalité des résidus

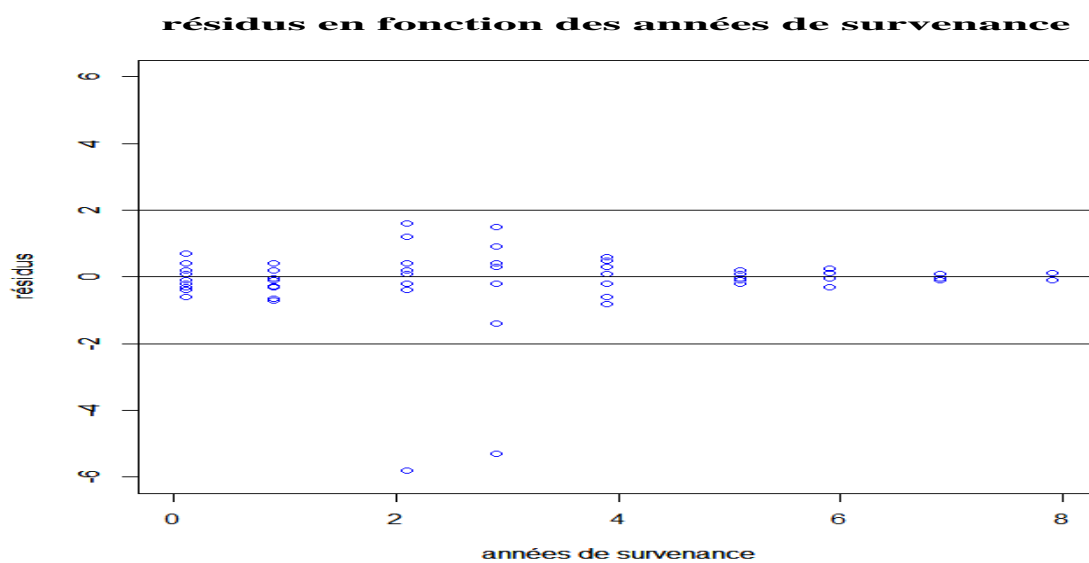


FIGURE 2.11 – le graphique résidu en fonction des années de survenance

Valeurs atypiques : (0,3), (0,4), (1,4), (3,5)

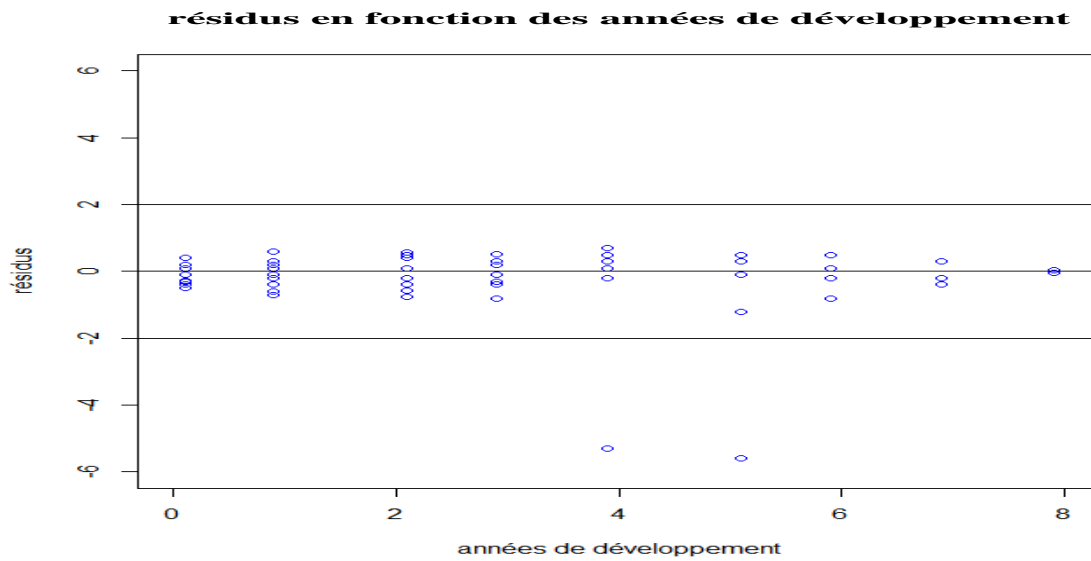


FIGURE 2.12 – le graphique résidu en fonction des années de développement

Nous pouvons noter que l'indépendance des résidus est toujours acceptée. L'hypothèse de normalité n'est quant à elle qu'approximativement respectée, en particulier pour la branche Auto.

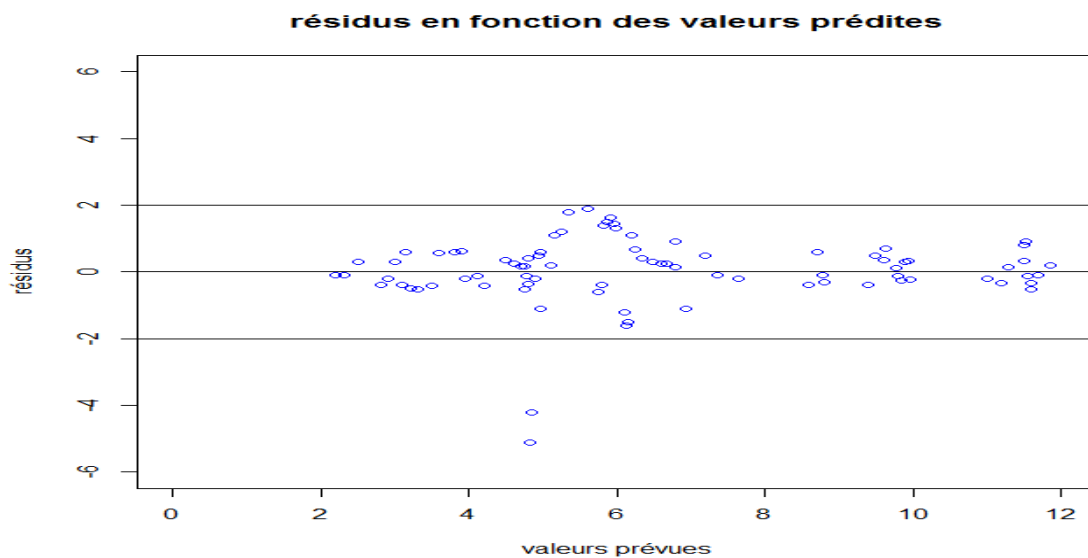


FIGURE 2.13 – le graphique résidu en fonction des valeurs prédites

Enfin, l'examen des graphiques des résidus en fonction des valeurs prévues et des différentes variables explicatives ne doit faire apparaître aucune structure non aléatoire.

2.2.3 La modélisation GLM

Le modèle linéaire généralisé a été développé initialement en 1972 par Nelder et Wedderburn. Ils sont une généralisation du modèle linéaire normal et sont formés de trois composantes : la composante aléatoire, la composante systématique et la fonction de lien.

• La composante aléatoire

On suppose que les variables aléatoires indépendantes $(X_{i,j})$, Nous supposons maintenant qu'elles suivent une loi de probabilité de type exponentielle. Leur densité est définie par la formule suivante :

$$f(x_{ij}, \theta_{ij}, \phi) = \exp \left\{ \frac{x_{ij}\theta_{ij} - b(\theta_{ij})}{\phi} + c(x_{ij}, \phi) \right\} \quad (2.47)$$

Où :

- θ_{ij} est appelé paramètre naturel de la famille exponentielle.
- ϕ est appelé paramètre de dispersion. Il est indépendant de i et j .
- b et c sont des fonctions régulières spécifiques à la distribution.

• La composante systématique

La composante systématique est une composante déterministe du modèle. La plupart des modèles de provisionnement sont basés sur un prédicteur linéaire η de la forme :

$$\eta_{ij} = \mu + \alpha_i + \beta_j \quad i, j = 0, \dots, n \quad (2.48)$$

Avec $\alpha_0 = \beta_0 = 0$

Où :

- μ : représente l'inflation, supposée constante. Dans certains modèles ce paramètre est nul.
- α_i : paramètre dépendant du semestre de survenance.
- β_j : paramètre relatif au délai de règlement.

• La fonction de lien

Il existe une relation fonctionnelle entre les composantes aléatoires et la composante systématique :

$$\eta_{ij} = g(\mu_{ij}) \quad i, j = 0 \dots n \quad (2.49)$$

Où g est appelé fonction de lien et est supposée monotone et différentiable.

Les fonctions de lien standards sont :

- lien identité : $\eta_{ij} = \mu_{ij}$

- lien logarithmique : $\eta_{ij} = \ln(\mu_{ij})$ soit $\mu_{ij} = \exp(\eta_{ij})$

On considère que les montants de recouvrement incrémentaux $X_{i,j}$ sont indépendants et suivent une loi de poisson de paramètre $\alpha_i \beta_j$.

On aura donc :

$$\begin{cases} P(X = x) = \exp[(x \ln \lambda - \lambda) + d(x)] & x \in N \\ \text{soit } \theta = \ln \lambda; \phi = 1; E(X) = \lambda; V(\mu) = \mu; W(\mu) = \mu \end{cases} \quad (2.50)$$

La contrainte $\beta_0 + \dots + \beta_n = 1$ permet en outre l'interprétation des β_j comme la fraction des recouvrements ayant eu lieu au cours du j^{me} semestre et des α_i comme le montant des recouvrements perçu le semestre i .

L'estimation des paramètres est obtenue en maximisant la log-vraisemblance du modèle linéaire généralisé.

$$L = \prod_{i,j=0}^n \exp(-\alpha_i \beta_j) \frac{(\alpha_i \beta_j)^{x_{i,j}}}{X_{i,j}!} \quad i + j \leq n \quad (2.51)$$

$$\log L = \sum_{i,j=0}^n -\alpha_i \beta_j + X_{i,j} \log(\alpha_i \beta_j) - \log(X_{i,j}!) \quad i + j \leq n \quad (2.52)$$

Le modèle de Poisson sur-dispersé $P_{sd}(\lambda, \phi)$ peut s'écrire comme :

$$\begin{cases} X \approx P^{sd}(\lambda, \phi) \text{ ssi } \frac{X}{\phi} \approx P\left(\frac{\lambda}{\phi}\right) \\ E(X) = \lambda; V(X) = \phi E(X); V(\mu) = \mu \end{cases} \quad (2.53)$$

Ensuite, nous testerons également une modélisation par la loi Gamma.

La spécification de la densité est la suivante :

Loi Gamma $G(r, \frac{r}{\mu})$:

$$\begin{cases} f(x) = \exp \left[\left(-\frac{x}{\mu} - \ln(\mu) \right) r + d(x, r) \right] & x > 0 \\ \text{soit } \theta = -\frac{1}{\mu}; \phi = \frac{1}{r}; E(X) = \mu; V(\mu) = \mu^2; W(\mu) = 2\mu^3 \end{cases} \quad (2.54)$$

La fonction de vraisemblance associée s'écrit :

$$L = \prod_{i,j=0}^n \frac{1}{\Gamma(1/\phi) \cdot X_{i,j}} \left(\frac{X_{i,j}}{\phi \alpha_i \beta_j} \right)^{1/\phi} \exp \left(-\frac{X_{i,j}}{\phi \alpha_i \beta_j} \right) \quad i + j \leq n \quad (2.55)$$

$$\log L = \sum_{i,j=0}^n \frac{1}{\phi} \log \left(\frac{X_{i,j}}{\phi \alpha_i \beta_j} \right) - \frac{X_{i,j}}{\phi \alpha_i \beta_j} - \log(X_{i,j}) - \log(\Gamma(1/\phi)) \quad i + j \leq n \quad (2.56)$$

L'estimation des incréments est alors donnée par $\hat{\mu}_{i,j} = e^{\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j}$ et donc celle des provisions est $\hat{R} = \sum_{i+j > n} \hat{\mu}_{ij}$

• Risque d'estimation et de prédiction

L'application de la méthode Delta est détaillée ci-dessous. En notant :

$$\sigma_R^2(\hat{\mu}), \sigma_R^2(\hat{\alpha}_i) \text{ et } \sigma_R^2(\hat{\beta}_j) \text{ les variances et } Cov_R(\hat{\mu}), Cov_R(\hat{\alpha}_i) \text{ et } Cov_R(\hat{\beta}_j)$$

les covariances asymptotiques des paramètres, on peut définir la variance asymptotique de l'estimateur du prédicteur linéaire $\hat{\eta}$ par :

$$\sigma_R^2(\hat{\eta}) = \sigma_R^2(\hat{\mu}) + \sigma_R^2(\hat{\alpha}_i) + \sigma_R^2(\hat{\beta}_j) + 2Cov_R(\hat{\mu}, \hat{\alpha}_i) + 2Cov_R(\hat{\mu}, \hat{\beta}_j) + 2Cov_R(\hat{\alpha}_i, \hat{\beta}_j)$$

Soit de façon équivalente, avec l'écriture matricielle :

$\sum_R(\hat{\eta}) = J_\eta \sum_R(\hat{\zeta}) J_\eta'$ où $\sum_R(\hat{\zeta})$ désigne la matrice de variance covariance des estimateurs.

$\hat{\zeta} = \hat{\mu}, (\hat{\alpha}_i)_{i=0\dots n}, (\hat{\beta}_j)_{j=0\dots n}$ de dimension $[2(n+1)+1, 2(n+1)+1]$, et J_η la matrice jacobienne relative à la transformation $\eta : \zeta \rightarrow \eta(\zeta) = \eta_{i,j} = \mu + \alpha_i + \beta_j$ de taille $[(n+1)^2, 2(n+1)+1]$ définie par :

$$\frac{\partial \eta_{i,j}}{\partial \mu} = 1 \quad (2.57)$$

$$\frac{\partial \eta_{i,j}}{\partial \alpha_k} = \begin{cases} 1 & \text{si } k = i \\ 0 & \text{si } k \neq i \end{cases} \quad (2.58)$$

$$\frac{\partial \eta_{i,j}}{\partial \beta_k} = \begin{cases} 1 & \text{si } k = j \\ 0 & \text{si } k \neq j \end{cases} \quad (2.59)$$

A cette étape, il est nécessaire de mettre à zéro les lignes pour lesquelles la contrainte a été appliquée c'est-à-dire celles où le $\hat{\eta}_{i,j} = 0$.

On définit ensuite la matrice de variance covariance des $\hat{\mu}_{ij}$, notée $\sum_R(\hat{\mu})$, à partir de la matrice jacobienne D issue de la relation entre les $\eta_{i,j}$ et les $\mu_{i,j}$. La matrice D, de taille $[(n+1)^2, (n+1)^2]$ est une matrice diagonale dont les valeurs sont $(g^{-1})'(\eta_{i,j})$ soit dans le cas du lien logarithmique directement les valeurs estimées $\hat{\mu}_{i,j}$. Les valeurs de $\hat{\mu}_{i,j}$ qui ont été utilisées sont en réalité celles qui sont obtenues après la mise en place de la contrainte, soit les $\hat{\mu}_{i,j}$. On obtient alors :

$$\sum_R(\hat{\mu}) = D \sum_R(\hat{\eta}) D. \quad (2.60)$$

Enfin, la matrice de variance covariance du montant espéré de recouvrement est donnée par :

$$\sum_R([E(R_i)]) = J_\mu \sum_R(\hat{\mu}) J'_\mu \quad (2.61)$$

avec J_μ la matrice Jacobienne $[n+1, (n+1)^2]$ de la fonction qui relie les $\hat{\mu}_{i,j}$ à $E(R_i)$, définie par :

$$\frac{\partial E(R_i)}{\partial \mu_{k,l}} = \begin{cases} 0 & \text{si } k \neq i \\ 0 & \text{si } k = i, l \leq n-i \\ l & \text{si } k = i, l > n-i \end{cases} \quad (2.62)$$

Il n'est pas nécessaire de modifier la matrice J_μ puisque on a bien :

$$E(R_i) = \sum_{j=n-i+1}^n \hat{\mu}_{i,j}$$

La variance du montant total de recouvrement espéré $E(R)$ est donnée par la relation

$$\hat{\sigma}_{\hat{R}}^2([E(R)]) = J_R \sum_R([E(R_i)]) J'_R \quad (2.63)$$

où J_R est la matrice ligne $[1, n+1]$ contenant que des 1, correspondant à la matrice jacobienne de l'application reliant les $E(R_i)$ à $E(R)$.

$$V(\hat{R}) = \hat{\phi} \sum_{i+j > n} V(\hat{\mu}_{ij}), \quad \text{avec} \quad \hat{\phi} = \frac{1}{N-p} \sum_{i+j \leq n} \frac{(x_{ij} - \hat{\mu}_{ij})^2}{V(\hat{\mu}_{ij})} \quad (2.64)$$

L'erreur de prédiction est alors donnée par

$$\widehat{sep}_{\hat{R}} = \sqrt{V(\hat{R}) + \hat{\sigma}_{\hat{R}}^2} \quad (2.65)$$

Les résultats obtenus sont les suivants :

Modèle	modèle Log Poisson surdispersé		modèle Log Gamma	
Branches	Dom.Auto	Dom.Biens	Dom.Auto	Dom.Biens
Provisions	13,073	10,982	13,367	11,078
sep	11,2 %	14,3 %	48,2 %	49,7 %
$\widehat{\phi}$	112	87	0,07	0,14

TABLE 2.9 – Tableau des provisions, erreurs relatives et paramètres de dispersion

Le modèle Log Poisson nous donne bien les mêmes montants de provisions que la méthode Chain-Ladder ; en revanche, l'erreur de prédiction est relativement inférieure dans le cadre de ce modèle. En ce qui concerne le modèle Log Gamma, les montants sont sensiblement proches des résultats précédents. Nous pouvons néanmoins remarquer que l'erreur de prédiction est plus importante pour les branches à liquidation courte, comme pour la régression Log Normale.

• Validation du modèle par examen de résidus

Le graphe des résidus peut permettre de détecter la présence de valeurs aberrantes. Pour chaque cellule du triangle supérieur (tel que $i+j < n$), 3 types de résidus peuvent être définis :

- les résidus bruts : $r_{i,j} = X_{i,j} - \widehat{\mu}_{i,j}$; En pratique ces résidus ne sont pas adaptés.

- Les résidus non standardisés de Pearson : $r_{ij}^{(p)} = \frac{X_{ij} - \widehat{\mu}_{ij}}{\sigma(\widehat{\mu}_{ij})}$

- Les résidus de la déviance :

$$\begin{cases} r_{ij}^{(d)} = \text{sgn}(X_{i,j} - \widehat{\mu}_{i,j})\sqrt{d_{ij}} \\ d_{ij} = 2X_{i,j}(\tilde{\theta}_{ij} - \widehat{\theta}_{ij}) - [b(\tilde{\theta}_{ij}) - b(\widehat{\theta}_{ij})] \end{cases} \quad (2.66)$$

où $\tilde{\theta}_{ij} = b^{-1}(X_{i,j})$ et $\widehat{\theta}_{ij} = b^{-1}(\widehat{\mu}_{i,j})$

A partir des résidus de déviance, il est possible de déterminer un estimateur convergent de ϕ par $\widehat{\phi} = \frac{D}{t-p}$ avec t le nombre de données du modèle et p le nombre de paramètres de la régression.

Des indicateurs de la qualité d'ajustement du modèle aux données et d'analyse des résidus sont donnés par :

- La statistique du χ^2 de Pearson :

$$\chi^2 = \sum_{i,j=0}^n \frac{(X_{i,j} - \hat{\mu}_{ij})^2}{V(\hat{\mu}_{ij})} \quad i + j \leq n \quad (2.67)$$

- La D eviance :

$$D = 2 \sum_{i,j=0}^n X_{i,j}(\theta_{ij} - \hat{\theta}_{ij}) - [b(\theta_{ij}) - [b(\theta_{ij}) - b(\hat{\theta}_{ij})]] \quad i + j \leq n \quad (2.68)$$

Plus la d eviance D est faible, plus l'ajustement par le mod ele est bon. Cependant, cet indicateur ne permet pas de comparer plusieurs mod elisations entre-elles, sauf dans le cas o   la composante al eatoire est identique.

Pour pouvoir comparer les mod eles   structure (V, ϕ) diff erentes, Nelder et Al. (1987) ont introduit la quasi-vraisemblance  tendue q^+ . Pour choisir un mod ele, il suffit de minimiser la fonction :

$$-2q^+(X, \mu) = \frac{1}{\phi} \sum_{i+j \leq n} d_{i,j} + \sum_{i+j \leq n} \ln[2\pi\phi V(X_{i,j})] \quad (2.69)$$

Les graphiques suivants repr esentent la distribution des r esidus de Pearson dans chacune des mod elisations.

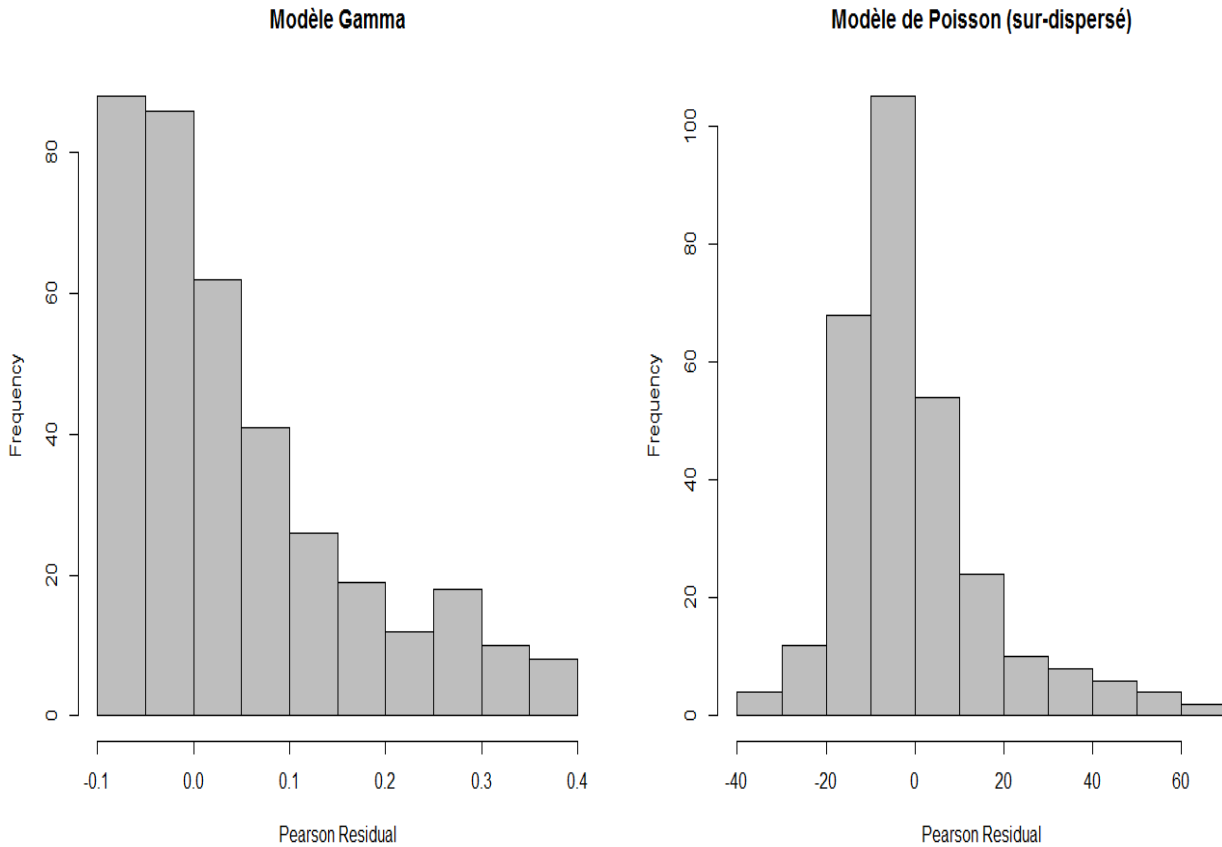


TABLE 2.10 – Distribution des r esidus de Pearson

Le tableau suivant donne la statistique $-2q^+$ obtenue pour chacune d'entre-elles.

statistique $-2q^+$		
Poisson sur dispersé	Poisson simple	Gamma
1078	858 373	2066

TABLE 2.11 – Valeurs de la statistique $-2q^+$

Les résultats nous indiquent que le modèle de poisson simple ne semble pas adapté à nos données. Nous allons donc l'exclure de nos résultats et nous focaliser sur les distributions gamma et poisson sur-dispersé.

Méthode	C-Ladder		L.Normale		L.Poisson Surd.		Gamma	
Coefficients	Provi	sep	Provi	sep	Provi	sep	Provi	sep
Dom.Auto	13,073	9,4%	16,945	91,9%	13,073	8,2%	13,365	39,8%
Dom.Biens	10,981	9,8%	12,601	58,3%	10,981	8,9%	11,077	41,2%
Dom.Energie	41,095	6,1%	44,521	26,7%	41,095	8,3%	41,837	22,6%

TABLE 2.12 – Tableau récapitulatif des provisions et des erreurs de prédiction associées

Le modèle Log Poisson, qui donne les mêmes montants de provisions que la méthode Chain-Ladder, avec toutefois une erreur de prédiction plus faible, semble être le modèle qui s'adapte le mieux aux données. Le modèle Log Gamma estime des montants du même ordre, et confirme donc les résultats précédents. En revanche, la régression Log Normale surestime l'ensemble des montants. L'erreur de prédiction associée est d'ailleurs supérieure à celle des autres méthodes, quelque soit la branche considérée, plus particulièrement pour les branches à liquidation courte.

Par la suite, nous allons donc principalement retenir le modèle Log Poisson Surdispersé et le modèle Log Gamma à titre de comparaison. Nous avons choisi le modèle Log Poisson Surdispersé parce que sa erreur de prédiction est faible.

Comme nous l'avons dit précédemment, ces méthodes ne donnent qu'une estimation des provisions. Après avoir pris en compte l'inflation future et actualisé les paiements, nous obtenons les Best Estimates suivants :

Méthode	Chain-Ladder	Log Normale	Log Poisson Surd.	Gamma
Dom.Auto	13,098	16,986	13,098	13,395
Dom.Biens	10,992	12,621	10,992	11,089
Dom.Energie	40,831	44,219	40,831	41,552

TABLE 2.13 – Tableau récapitulatif des Best Estimates

Chapitre 3

Le risque de provisionnement et bootstrap

3.1 Etude de la base de données

Le but le plus important de toute la compagnie d'assurance non-vie qui est efficace sur toutes les opérations d'assurance. la différence entre l'assurance vie et non-vie est la période de validité de la police d'assurance, contrats d'assurance non-vie est généralement de courte durée de 1 à 3 ans. L'année récurrente se déroule du 1/1/n au 31/12/n. Le 31/12/n est la date d'inventaire(ou de fin d'exercice).

Les sinistres sont rattachés à l'année d'origine(origin year) qui peut être :

- l'année de survenance(accident year)
- l'année de souscription(underwriting year)
- l'année de déclaration(report year)

Le rattachement le plus fréquent est celui de l'année de survenance.

L'effet réel des activités d'assurance sont enregistrés à l'année de souscription.

Le *Best Estimate* est constitué d'une composante « Prime » qui correspond à la valeur actuelle des engagements d'assurance couvrant des sinistres non-survenus à la date d'évaluation pour lesquels la compagnie est engagée et d'une composante « Sinistres » qui correspond à la valeur actuelle des coûts futurs des sinistres survenus, qu'ils soient connus ou non de l'assureur, à la charge de l'assureur.

Lors du calcul en *Best Estimate*, la frontière entre la provision de prime et la provision de sinistres est assez difficile à définir pour les activités gérées en souscription. En effet, le *Best Estimate* de sinistres, dès lors qu'il est appréhendé par

génération souscription, par mécanique calculatoire, couvre à la fois les sinistres survenus et sinistres à venir.

Le *Best Estimate* sinistres devrait être déterminé en dissociant les éléments bruts de réassurance, les effets de l'inflation, les recours, la provision pour frais de gestion des sinistres (calculée brute de réassurance et à la charge exclusive de l'assureur) et les impacts de la réassurance.

Il paraît donc raisonnable de calculer, dans un premier temps, un best estimate au brut de réassurance, puis passer au net de réassurance. Le calcul du best estimate devrait ainsi comporter les différentes phases ci-dessous :

- Projection des triangles de charges et de paiements au brut de recours encaissés par catégorie avec prise en compte de l'inflation,
- Projection des triangles de recours encaissés par catégorie,
- Estimation de l'impact de la réassurance en fonction de la nature des traités (différentes clauses particulières) et de l'évolution de ces derniers,
- Actualisation des cash-flows futurs au taux sans risque.

L'assurance Non-Vie est assujettie à des fluctuations liées à l'environnement économique. Ces modifications impacteront le niveau des primes, les conditions dans lesquelles se feront les affaires nouvelles mais également leur rentabilité, avec un décalage temporel. Ces fluctuations caractérisent le cycle de souscription de l'assurance Non-Vie, ce dernier est directement lié au calcul du *Best Estimate* de primes.

Toute société doit mettre en place les outils de gestion de données lui permettant de disposer d'un historique de données fiables suffisant pour les branches qu'elle pratique (la profondeur des historiques requis est indiquée dans les fiches par nature de garantie). Ce point est particulièrement crucial pour les sociétés dont toute ou partie de la gestion est déléguée, pour les entités nouvellement créées, les branches nouvellement lancées ou pour les fusions/acquisitions de portefeuille conduisant souvent à la coexistence de données hétérogènes dont le traitement peut poser des difficultés.

Dans la mesure du possible, la compagnie doit reconstituer les historiques de données afin d'appréhender le déroulement de la catégorie et d'intégrer l'ensemble

des éléments de rupture. Les données nécessaires sont les suivantes (l'historique de données nécessaire étant fonction de la durée de la catégorie) :

- Primes : historique des primes acquises et émises par exercice de rattachement
- Sinistres : règlements, recours encaissés, charges de sinistres, nombre de sinistres, coût moyen en distinguant les sinistres graves des sinistres attritionnels (triangles de liquidation),
- Frais : frais de gestion des sinistres, commissions fixes ou variables, frais d'administration et autres frais.

De façon générale, on retient comme hypothèse que les historiques passés intègrent des éléments d'inflation dans les données courantes et que les évolutions futures conservent ce type d'évolutions. On considère alors que l'inflation future correspond à l'inflation passée.

Néanmoins, dans certaines conditions, ces hypothèses doivent être remises en cause. C'est régulièrement le cas pour des sinistres à développement long dont la gestion peut s'étaler sur des périodes relativement différentes. La difficulté consiste alors à déterminer l'inflation passée et prévoir celle du futur sur une période de plusieurs décennies.

Par conséquent, comprendre le cycle de la couverture des produits d'assurance non-vie, les facteurs qui influent sur l'activité d'assurance, la construction de modèles internes pour calculer les réserves d'assurance et des simulations de tendance activités commerciales sont des facteurs importants à faire par un ou une actuair.

3.2 Le bootstrap

La méthode de Bootstrap a été introduite par Efron en 1979 dans le but d'estimer le biais et la variabilité d'un estimateur dans un contexte non paramétrique. Le principe consiste à simuler un grand nombre d'échantillons de taille N , en tirant aléatoirement avec remise N observations à partir d'un échantillon initial de N variables aléatoires indépendantes et identiquement distribuées (i.i.d.) (X_1, X_2, \dots, X_N) de fonction de répartition commune F . Cette méthode est donc aussi appelée méthode de « rééchantillonnage ».

L'utilisation du bootstrap suppose que les éléments de l'échantillon de départ soient indépendants et identiquement distribués(iid). Les variables $(X_{i,j})$ ne sont en général pas identiquement distribuées. Il est donc préférable d'avoir recours aux résidus du modèle, en particulier les résidus de Pearson car plus simples à calculer.

Nous somme présenter d'abord le principe de rééchantillonnage des résidus dans deux modèle de Poisson et Chain Ladder. Le modèle Log Poisson donnant les mêmes résultats que la méthode Chain-Ladder, l'utilisation de cette dernière est une solution pour diminuer le temps de calcul.

Pour corriger le biais dans la comparaison des estimations analytiques et bootstrap d'erreur de prédiction, il est conseillé d'ajuster les résidus en intégrant le nombre de paramètres de régression dans l'erreur bootstrap de prédiction. On définit les résidus ajustés par $r_{ij}^{aj} = \sqrt{\frac{N}{N-p}} r_{ij}^p$ où N est le nombre d'éléments de l'échantillon et p le nombre de paramètres de régression.

La méthode du bootstrap :

- Ajustement un modèle linéaire et estimation des valeurs prévues $(\hat{\mu}_{ij})$.
- Calcul des résidus de Pearson du modèle :

$$r_{ij}^p = \frac{x_{ij} - \hat{\mu}_{ij}}{\sqrt{v(\hat{\mu}_{ij})}} \quad (*) \quad (3.1)$$

- Les résidus de Pearson ajustés sont donnés par :

$$r_P^{ajt} = \sqrt{\frac{n}{1/2n(n+1) - 2n + 1}} \times r_{ij}^p \quad (3.2)$$

- Pour $b=1 \dots B$ ($B=5000$ ou 10000) :

-Rééchantillonnage avec remise des résidus.

Ces résidus bootstrapés sont notés $(r_{ij}^p)^b$ -En inversant la relation(*), obtention d'un nouveau triangle, appelé «pseudo-triangle» , composé des valeurs

$$(x_{ij})^b = \hat{\mu}_{ij} + (r_{ij}^p)^b \sqrt{V(\hat{\mu}_{ij})} \quad (3.3)$$

Le modèle réalisé avec les données sources est utilisé avec chaque jeu de données. Autant de projections que de jeux de données sont obtenues. A partir de ces projections, une distribution des règlements futurs est estimée. Il suffira alors de calculer les moments de cette loi et de choisir un quantile en fonction du risque choisi pour obtenir le montant de la provision.

Le bootstrap est une méthode d'inférence statistique basée sur l'utilisation de l'ordinateur qui peut répondre sans formule à beaucoup de questions statistiques réelles. Il est incontestable que l'utilisation des techniques de rééchantillonnage a été rendue possible grâce à la généralisation des moyens de calculs performants. Ces techniques reposent, au départ, sur des idées simples. Toutefois, il faut bien admettre que les développements apportés aux méthodes de base leur ont fait perdre une partie de cette simplicité.

-À l'aide d'une nouvelle modélisation GLM appliquée au pseudo-triangle, détermination d'une nouvelle estimation des provisions notée \hat{R}^b .

-À partir du B-échantillon des provisions (\hat{R}^b), nous pouvons alors estimer la moyenne, la variance et l'erreur de prédiction du montant de provisions.

$$E^B(\hat{R}) = \frac{1}{B} \sum_{b=1}^B \hat{R}^b \quad (3.4)$$

Risque d'estimation :

$$Var^B(\hat{R}) = \frac{1}{B-1} \sum_{b=1}^B [\hat{R}^b - E^B(\hat{R})]^2 \quad (3.5)$$

Risque de processus :

$$\hat{V}(R) = \hat{\phi} \sum_{i+j>n} V(\hat{\mu}_{ij}) \quad (3.6)$$

Erreur de prédiction :

$$sep = \sqrt{\hat{V}(R) + Var^B(\hat{R})} \quad (3.7)$$

Nous avons utilisé cette méthode pour 10000 simulations avec les modèles Log Poisson surdispersé et Log Gamma.

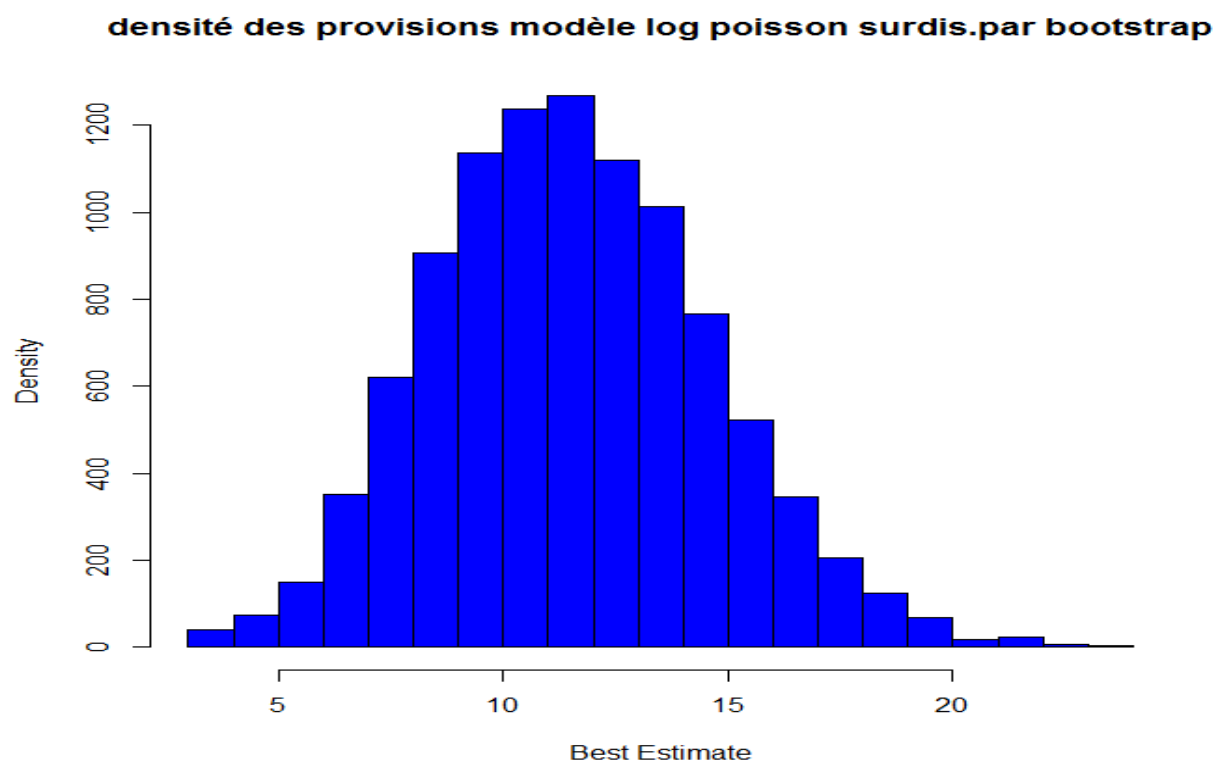


FIGURE 3.1 – Graphiques des provisions par modèle log poisson pour la branche Auto

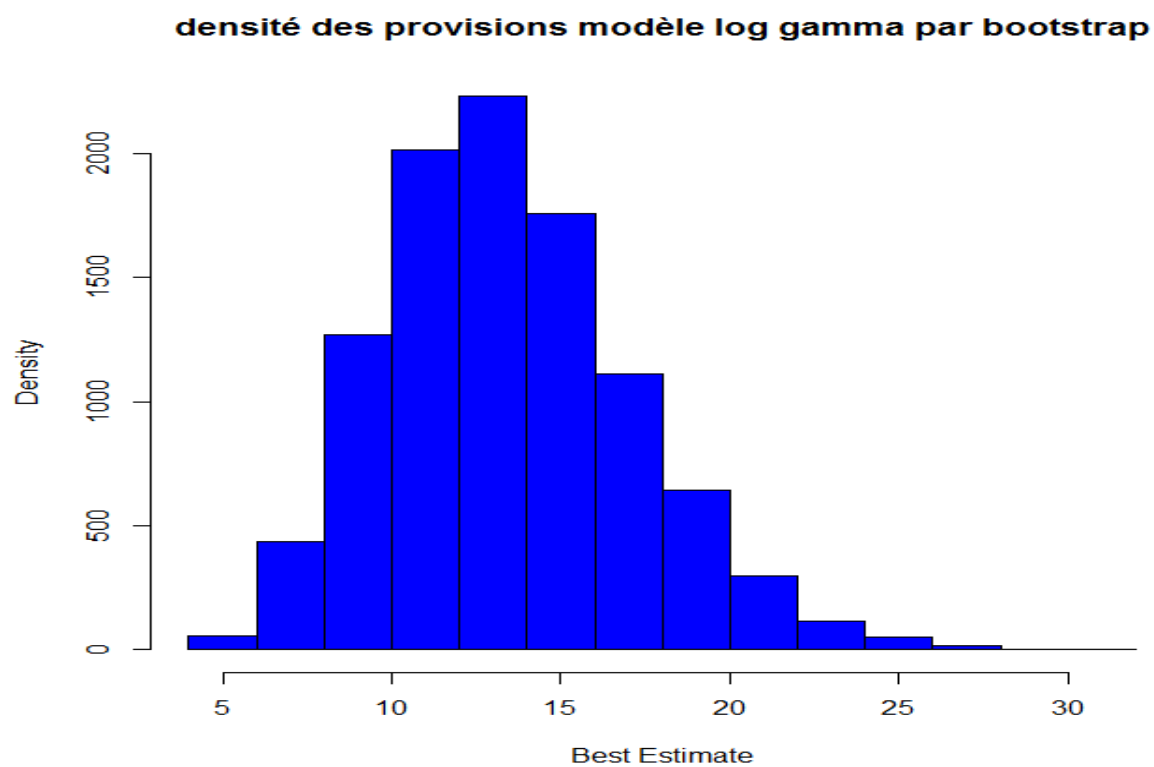


FIGURE 3.2 – Graphiques des provisions par modèle log gamma pour la branche Auto

Branches	Modèles	\hat{R}	$E^B(\hat{R})$	$Var^B(\hat{R})$	$\hat{V}(R)$	sep/R
Dom.Auto	Poisson	13,153	13,323	637,823	2169,142	8,92%
	Gamma	13,448	13,737	31286,61	26350,83	41,64%
Dom.Biens	Poisson	12,669	11,211	765,637	1296,724	7,47%
	Gamma	11,141	11,372	22878,797	18710,976	41,64%
Dom.Energie	Poisson	43,591	43,439	765,637	23602,339	9,67%
	Gamma	44,612	45,092	116550,106	44924,071	26,34%

TABLE 3.1 – Tableau récapitulatif des résultats donnés par le bootstrap

Le calcul de la variance des provisions simulées par Bootstrap fournit alors un estimateur de l'erreur de prédiction. Dans l'approche simulatoire, et en supposant les incréments indépendants, on estime la variance des réserves définie comme la variance de la somme des incréments du triangle inférieur :

Branches	Modèles	VaR à 95%	VaR à 99,5%
Dom.Auto	Poisson	14,367	15,213
	Gamma	20,942	26,619
Dom.Biens	Poisson	12,369	13,127
	Gamma	17,292	25,845
Dom.Energie	Poisson	50,068	54,194
	Gamma	60,585	74,113

TABLE 3.2 – VaR à 95% et VaR à 99,5% des provisions

Ces résultats tiennent compte de l'inflation future et de l'actualisation. Ils sont donc comparables à des *Best Estimates*. Le bootstrap ne s'applique que sur des résidus indépendants et identiquement distribués. Le bootstrap n'est pas robuste pour l'estimation des quantiles élevés dans la mesure où souvent les données que l'on possède ne reflètent pas les observations concernant ces quantiles. Il faut ajuster la queue de distribution.

De manière générale, le modèle Log Gamma donne des distributions à queue plus épaisse que le modèle Log Poisson Surdispersé.

Chapitre 4

Mesurer le risque de provisionnement à horizon un an

4.1 La méthode de Merz-Wüthrich

La méthode de Merz et Wüthrich proposent une méthode pour obtenir une estimation à 1 an de la variance de la variation calendaire de la charge de sinistres ultime estimée. Ceux-ci se sont appuyés sur le modèle de Mack qui mesure l'incertitude à l'ultime pour proposer une formule approchée fermée de l'erreur quadratique moyenne de prédiction conditionnelle du CDR observable, par rapport à 0 et par rapport au CDR réel. Pour estimer le risque à un an, nous allons chercher à analyser l'incertitude liée aux Claims Development Results (CDR) i.e. liée à la différence entre deux estimations successives du montant des *Best Estimates*.

Les hypothèses de Merz et Wüthrich sont similaires à celles de Mack, en rajoutant toutefois la condition de Markov :

-Les paiements cumulés $C_{i,j}$ des différentes années de survenance sont indépendants

-($C_{i,j}$) $_{j \geq 0}$ est une chaîne de Markov et il existe $f_j > 0$ et $\sigma_j > 0$ telles que, $\forall 1 \leq j \leq I, \forall 1 \leq i \leq I$,

$$\begin{cases} E[C_{i,j} | C_{i,j-1}] = f_{j-1} \cdot C_{i,j-1} \\ Var[C_{i,j} | C_{i,j-1}] = \sigma_{j-1}^2 \cdot C_{i,j-1} \end{cases} \quad (4.1)$$

D_I et D_{I+1} désignent l'information disponible aux dates $t = I$ et $t = I + 1$ respectivement :

$$D_I = \{C_{i,j}; i + j \leq I, i \leq I\} \quad (4.2)$$

$$D_{I+1} = \{C_{i,j}; i + j \leq I + 1, i \leq I + 1\} \quad (4.3)$$

On observe une diagonale supplémentaire des paiements dans l'ensemble D_{I+1} . En termes mathématiques, la tribu D_I est étendue à la tribu D_{I+1} .

Les estimateurs de Chain-Ladder dépendent de la date de considération des données :

En $t = N$,

$$\widehat{f}_j^I = \frac{\sum_{i=0}^{I-j-1} C_{i,j+1}}{\sum_{i=0}^{I-j-1} C_{i,j}} \quad (4.4)$$

En $t = N + 1$,

$$\widehat{f}_j^{I+1} = \frac{\sum_{i=0}^{I-j} C_{i,j+1}}{\sum_{i=0}^{I-j} C_{i,j}} \quad (4.5)$$

Grâce aux estimateurs de Chain-Ladder, il ressort les estimateurs sans biais des charges futures :

En $t = N$, $\forall j \geq I - i + 1$,

$$\widehat{C}_{i,j}^I = C_{i,I-i} \cdot \prod_{h=I-i}^{j-1} \widehat{f}_h^I \quad (4.6)$$

$$R_i^I = C_{i,I} - C_{i,I-i} \quad (4.7)$$

En $t = N + 1$, $\forall j \geq I - i + 2$

$$\widehat{C}_{i,j}^{I+1} = C_{i,I-i+1} \cdot \prod_{h=I-i+1}^{j-1} \widehat{f}_h^{I+1} \quad (4.8)$$

$$R_i^{I+1} = C_{i,I} - C_{i,I-i+1} \quad (4.9)$$

• Claims Development result (CDR)

Le modèle sous-jacent que les auteurs utilisent est celui proposé par Mack pour l'approche Chain Ladder. Les hypothèses de ce dernier modèle doivent donc être vérifiées. Les premiers travaux ayant pour but de mesurer le risque de provisionnement à un an. Ils considèrent la différence entre l'estimation de la charge ultime calculée en $t = I$ et la prédiction de la charge ultime calculée en $t = I + 1$. Cette différence est appelée Claims Development Result, notée CDR_i pour l'année de survenance i :

$$CDR_i(I + 1) = E[C_{i,I} \mid D_I] - E[C_{i,I} \mid D_{I+1}] \quad (4.10)$$

$$CDR(I + 1) = \sum_{i=1}^I CDR_i(I + 1) \quad (4.11)$$

L'erreur de prédiction des Claims Development Result peut être vue de deux manières différentes :

- Le point de vue prospectif, une prédiction du CDR sur l'année $I + 1$, à partir des observations des sinistres en date $t = I$

- Le point de vue rétrospectif, en date $t = I + 1$, et on obtient une réalisation du CDR observable :

$$\widehat{CDR}_i^{(I+1)} = \widehat{C}_{i,I}^I - \widehat{C}_{i,I}^{I+1} \quad (4.12)$$

Les $(C_{i,j})_{j \geq 0}$ possèdent les propriétés d'une martingale, :

$$E[CDR_i(I + 1) \mid D_I] = 0 \quad (4.13)$$

Cette dernière propriété signifie que le CDR réel serait égal à 0 si les facteurs de développement étaient parfaitement connus. Cependant, Le CDR réel n'étant pas observable, nous utilisons un estimateur de cette quantité : le CDR observable. Le CDR observable correspond donc à la différence entre le montant des règlements cumulés ultimes estimé à partir de l'information disponible à la date $t = I$ et celui estimé à partir de l'information disponible à la date $t = I + 1$. Les éléments inconnus du CDR réel sont remplacés par les estimateurs associés à la méthode de Chain-Ladder : les sinistres à l'ultime $E[C_{i,I} \mid D_I]$ et $E[C_{i,I+1} \mid D_{I+1}]$ sont remplacés respectivement par les estimateurs $\widehat{C}_{i,I}^I$ et $\widehat{C}_{i,I}^{I+1}$:

$$\widehat{CDR}_i(I + 1) = \widehat{C}_{i,I}^I - \widehat{C}_{i,I}^{I+1} \quad (4.14)$$

• Erreur de prédiction à 1 an

L'objectif de la méthode est donc de calculer l'erreur de prédiction à 1 an. Il s'agit de quantifier deux termes : une vision prospective avec l'erreur de prédiction des CDR observable par rapport à 0 et une vision rétrospective en quantifiant la distance entre le CDR réel et le CDR observable :

$$MSEP_{\widehat{CDR}_{(I+1)|D_I}}(0) = E[(\widehat{CDR}_i(I + 1) - 0)^2 \mid D_I] \quad (4.15)$$

$$MSEP_{\widehat{CDR}_{(I+1)|D_I}}(\widehat{CDR}_i(I + 1)) = E[(\widehat{CDR}_i(I + 1) - CDR_i(I + 1))^2 \mid D_I] \quad (4.16)$$

Les travaux de Merz et Wüthrich s'appuient sur l'hypothèse $\frac{\widehat{\sigma}_j^2}{\widehat{f}_j^2} \ll C_{I-j+1,j}$ per-

mettant de recourir à la simplification $\prod_{j=1}^I (1 + a_j) \approx 1 + \sum_{j=1}^I a_j$

Merz et Wüthrich proposent alors un premier estimateur de la variance des CDR :

$$\widehat{Var}(CDR_i(I+1) \mid D_I) = (\widehat{C}_{i,I}^I)^2 \widehat{\Psi}_i^I \quad (4.17)$$

Avec :

$$\widehat{\Psi}_{i,I}^I = \frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot C_{i,I-i}} = \frac{1}{(\widehat{f}_{I-i}^I)^2 \cdot C_{i,I-i}} \cdot \frac{1}{I-j+1} \sum_{j=0}^{I-i-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \widehat{f}_j \right)^2 \quad (4.18)$$

Sous les conditions de Chain-Ladder, en date $t = I + 1$, les estimateurs des MSEP conditionnelles sont définies pour chaque année de survénance $i = 1 \dots I$. Les deux erreurs définies précédemment peuvent alors se calculer par des formules fermées.

• MSEP du CDR observable par rapport au CDR réel

Nous avons alors comme estimateur de cette erreur de prédiction :

$$\widehat{MSEP}_{\widehat{CDR}(I+1)|D_I}(\widehat{CDR}_i(I+1)) = (\widehat{C}_{i,I}^I)^2 (\widehat{\Phi}_{i,I}^I + \widehat{\Delta}_{i,I}^I) \quad (4.19)$$

Avec :

$$\widehat{\Phi}_{i,I}^I = \sum_{j=I-i+1}^{I-1} \left(\frac{C_{I-j,j}}{S_j^{I+1}} \right)^2 \cdot \frac{\widehat{\sigma}_j^2}{(\widehat{f}_j^I)^2 \cdot C_{I-j,j}} \quad (4.20)$$

$$\widehat{\Delta}_{i,I}^I = \frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot S_{I-i}^I} + \sum_{j=I-i+1}^{I-1} \left(\frac{C_{I-j,j}}{S_j^{I+1}} \right)^2 \cdot \frac{\widehat{\sigma}_j^2}{(\widehat{f}_j^I)^2 \cdot S_j^I} \quad (4.21)$$

Où :

$$S_j^I = \sum_{i=0}^{I-j-1} C_{i,j} \quad (4.22)$$

$$S_j^{I+1} = \sum_{i=0}^{I-j} C_{i,j} \quad (4.23)$$

Le premier terme $\widehat{\Phi}_{i,I}^I$ correspond à l'incertitude intrinsèque du processus stochastique tandis que le second terme $\widehat{\Delta}_{i,I}^I$ correspond au biais de ce processus.

• MSEP du CDR observable par rapport à 0

$$\widehat{MSEP}_{\widehat{CDR}(I+1)|D_I}(0) = (\widehat{C}_{i,I}^I)^2 (\widehat{\tau}_{i,I}^I + \widehat{\Delta}_{i,I}^I) \quad (4.24)$$

Avec :

$$\widehat{\tau}_{i,I}^I = \widehat{\Phi}_{i,I}^I + \widehat{\Psi}_{i,I}^I \geq \widehat{\Phi}_{i,I}^I \quad (4.25)$$

La MSEP se décompose en deux termes : l'erreur d'estimation faite sur la première diagonale et l'erreur d'estimation faite sur les diagonales suivantes.

• **CDR agrégé**

L'erreur de prédiction du CDR réel agrégé considère l'erreur d'estimation pour la diagonale suivante ($j = I - i$) mais écrase toutes les erreurs d'estimation des diagonales d'après ($j \geq I - i$) par le facteur $\frac{C_{I-j,j}}{S_j^{I+1}} \leq 1$.

$$MSEP_{\sum_{i=1}^I \widehat{CDR}(I+1)|D_I} \left(\sum_{i=1}^I \widehat{CDR}_i(I+1) \right) = \sum_{i=1}^I \widehat{MSEP}_{\widehat{CDR}(I+1)|D_I} (\widehat{CDR}_i(I+1)) + 2 \cdot \sum_{k>i>0} \widehat{C}_{i,I}^I \cdot \widehat{C}_{k,I}^I (\widehat{\Phi}_{i,I}^I + \widehat{\Delta}_{i,I}^I) \quad (4.26)$$

Avec :

$$\widehat{\Delta}_{i,I}^I = \frac{C_{i,I-i}}{S_{I-i}^{I+1}} \frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot S_{I-i}^I} + \sum_{j=I-i+1}^{I-1} \left(\frac{C_{I-j,j}}{S_j^{I+1}} \right)^2 \cdot \frac{\widehat{\sigma}_j^2}{(\widehat{f}_j^I)^2 \cdot S_j^I} \quad (4.27)$$

Pour la MSEP conditionnelle autour du 0, nous devons alors définir :

$$\widehat{\xi}_{i,I}^I = \widehat{\Phi}_{i,I}^I + \frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot S_{I-i}^{I+1}} \quad (4.28)$$

Et ainsi : $MSEP_{\sum_{i=1}^I \widehat{CDR}(I+1)|D_I}(0) =$

$$\sum_{i=1}^I \widehat{MSEP}_{\widehat{CDR}(I+1)|D_I}(0) + 2 \cdot \sum_{k>i>0} \widehat{C}_{i,I}^I \cdot \widehat{C}_{k,I}^I (\widehat{\xi}_{i,I}^I + \widehat{\Delta}_{i,I}^I) \quad (4.29)$$

En développant les formules, nous obtenons :

$$\begin{aligned} MSEP_{\sum_{i=1}^I \widehat{CDR}(I+1)|D_I}(0) = & \sum_{i=1}^I \left[(\widehat{C}_{i,I}^I)^2 \left(\frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot C_{i,I-i}^I} + \frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot S_{I-i}^I} + \sum_{j=I-i+1}^{I-1} \frac{C_{I-j,j}}{S_j^{I+1}} \frac{\widehat{\sigma}_j^2}{(\widehat{f}_j^I)^2 \cdot S_j^I} \right) \right] \\ & + 2 \cdot \sum_{k>i>0} \widehat{C}_{i,I}^I \cdot \widehat{C}_{k,I}^I \left[\frac{\widehat{\sigma}_{I-i}^2}{(\widehat{f}_{I-i}^I)^2 \cdot S_{I-i}^{I+1}} + \sum_{j=I-i+1}^{I-1} \frac{C_{I-j,j}}{S_j^{I+1}} \frac{\widehat{\sigma}_j^2}{(\widehat{f}_j^I)^2 \cdot S_j^I} \right] \end{aligned} \quad (4.30)$$

Cette dernière formule fermée nous permet donc de calculer le risque de provisionnement à un an pour l'ensemble du portefeuille considéré.

Le modèle de Merz et Wüthrich peut soulever quelques critiques :

-Il ne permet pas d'apprécier correctement la volatilité au niveau de la queue de distribution.

-Il est fondé sur le modèle de Mack dont les hypothèses peuvent être non pertinentes au niveau de la queue de distribution.

-Il est fait l'hypothèse que le *Best Estimate* à horizon un an est sans biais. Dans la réalité, ce n'est pas forcément le cas.

-L'hypothèse $\frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \ll C_{I-j+1,j}$ peut également être mise en doute.

Les travaux de Merz et Wüthrich supposent, comme ceux de Mack, que le déroulement des sinistres suit un modèle de Chain Ladder. Elle fournit en effet une expression fermée de la volatilité à un an de la variation calendaire de la charge de sinistres ultime estimée. L'importante utilisation de la méthode Chain Ladder induit une préférence pour ce modèle Merz Wüthrich qui est relativement simple à implémenter et non paramétrique. Il est possible dans certains triangles que l'incertitude à un an soit supérieure à l'incertitude à long terme. Cette méthode ne permet pas a priori d'appliquer un facteur de queue de développement, ce qui peut véritablement s'éloigner de la réalité de la sinistralité de certaines branches. C'est pourquoi on s'intéresse désormais à une autre approche, à savoir la technique du bootstrap, afin de trouver une alternative aux problèmes évoqués.

4.2 Calcul du SCR

Le SCR est le Capital de Solvabilité Requis (Solvency Capital Requirement). Il est égal à la Value at Risk à 99,5% à un an des fonds propres de base, soit la perte maximale possible à horizon un an avec une probabilité de 99,5%. Le calcul du SCR peut se faire par une formule standard ou par un modèle interne (calculs des FP propres aux sociétés) complet ou partiel.

• Formule standard

- Calcul SCR par la matrice de corrélation entre les branches :

$$SCR = \rho(\sigma) \times V \quad (4.31)$$

Où :

$$-V = \sum_* V_*$$

- V_* est *Best Estimate* des provisions pour sinistres à payer pour une branche.

$$-\sigma = \sqrt{\frac{1}{V^2} \sum_{i,j} Corr_{*i,j} \sigma_i \sigma_j V_i V_j}$$

- $Corr_{*i,j}$ est la matrice de corrélation entre les branches.

$Corr_{*}$	Dom. Auto	Dom. Biens	Dom. Energie	Dom. Habitation
Dom. Auto	1	•	•	•
Dom. Biens	0.5	1	•	•
Dom. Energie	0.25	0.5	1	•
Dom. Habitation	0.5	0.5	0.25	1

TABLE 4.1 – Matrice de corrélation entre les branches

- Calcul SCR par la méthode de Merz et Wüthrich :

À partir du calcul de l'erreur de prédiction (MSEP), nous pouvons en déduire la volatilité des réserves à 1 an, comme il est suggéré dans la Formule Standard :

$$\sigma^* = \frac{\sqrt{MSEP}}{BE_{CL}} \quad (4.32)$$

Avec :

MSEP : L'erreur de prédiction à 1 an, calculée précédemment

BE_{CL} : Le *Best Estimate* du montant des provisions estimé par la méthode de Chain-Ladder.

La première méthode pour déterminer le SCR repose sur la mesure de l'incertitude à horizon un an par une formule fermée, en l'occurrence, dans ce mémoire, celle proposée par Merz et Wüthrich.

L'application du modèle de Merz et Wüthrich s'avère ensuite assez simple puisque qu'une formule fermée permet de calculer directement la MSEP et le quantile du risque de réserve en découle directement selon l'approche sélectionnée.

Nous avons appliqué pour calculer le SCR avec différents paramètres :

Paramètres	P.donnés dans la matrice			P.donnés par modèle de Merz-W		
Coefficients	σ	SCR		σ^*	SCR	
		poisson	gamma		poisson	gamma
Dom.Auto	9 %	3,77	3,85	13,5 %	6,41	6,55
Dom.Biens	10 %	3,51	3,55	14,5 %	5,57	5,62
Dom.Energie	11,78 %	12,08	3,54	5,5 %	7,18	7,34

TABLE 4.2 – Résultats des SCR donnés par la formule standard avec différents paramètres

• Simulations de lois Log Normales

Nous supposons que les provisions suivent une loi Log Normale de paramètres (μ, σ^2) avec :

$$\mu = \ln(BE_{CL}) - \frac{1}{2}\ln\left(1 + \frac{MSEP}{BE_{CL}^2}\right); \quad \sigma^2 = \ln\left(\frac{MSEP}{BE_{CL}^2 + 1}\right) \quad (4.33)$$

En effectuant 10 000 simulations de cette loi, nous obtenons une distribution provision.

Toutefois, il est important de remarquer que le modèle de Merz-Wüthrich évalue le risque à un an des provisions et non du *Best Estimate*. Nous allons donc faire l'hypothèse suivante :

Quelque soit le triangle à l'ultime, $BE = \alpha\%$ Provisions.

À partir des résultats de Merz-Wuthrich, nous estimons la VaR à 99,5 % à un an des provisions et en déduisons la VaR à 99,5 % du BE à un an :

$$VaR_{99,5\%}^1 \left(\frac{BE_1}{1+r} + \frac{P_1}{\sqrt{1+r}} - BE_0 \right) = \alpha\% VaR_{99,5\%}^1 (provisions_1 + P_1 - provisions_0) \quad (4.34)$$

Application pour 10 000 simulations :

Branches	SCR provisions	alpha	SCR BE
Dom.Auto	6,435	1,00313	6,442
Dom.Biens	5,328	1,00213	5,351
Dom.Energie	6,938	0,92724	6,862

TABLE 4.3 – Résultats des SCR donnés par la loi log-normale

Pour les branches Dommages Auto et Dommages aux Biens le paramètre alpha est supérieur à 1. Ceci est dû au fait que ce sont des branches à liquidation courtes : l'actualisation n'a que peu d'impact sur les flux futurs (à faibles montants).

Par ailleurs, cette méthode semble sous-estimer le montant du SCR pour les branches des Energie.

4.3 La méthode GLM Bootstrap

La méthode Bootstrap GLM (Generalized Linear Model), également appelée GLM bootstrapé introduite par England et Verrall. Cette dernière méthode permet de contourner les inconvénients du modèle de Merz et Wüthrich, elle permet d'analyser plus finement les sources d'incertitudes au niveau de la queue de distribution et permet de distinguer à la fois l'erreur de prédiction et l'erreur d'estimation du modèle. De plus, la méthode du bootstrap permet d'obtenir la distribution complète des provisions, au contraire de la formule fermée de Merz et Wüthrich qui ne propose que les deux premiers moments.

Nous allons utiliser les méthodes traditionnelles et rééchantillonnage des résidus pour simuler un grand nombre de nouvelles « diagonales » d'incrémentes les paiements survenus entre n et $n+1$:

-modélisation GLM, tirage de la première diagonale notée(P_1^b)

-intégration au triangle initial et obtention d'un « trapèze ».

-application d'une nouvelle modélisation GLM et estimation des provisions à un an, (BE_1^b).

-on réitère ce procédé B fois.

Nous partons d'un bilan à $t=0$ de la forme suivante :

ACTIF	PASSIF
A_0	$FP_0 = A_0 - BE_0$
	BE_0

Nous supposons que l'actif est investi au taux sans risque r et que les paiements surviennent en milieu d'année.

À $t=1$, le bilan sera de la forme :

ACTIF	PASSIF
$A_1 = A_0(1+r) - P_1(1+r)^{1/2}$	$FP_1 = A_1 - BE_1$
	BE_1

Les pertes d'une année sur l'autre sont représentées par $FP_0 - \frac{FP_1}{1+r}$, soit :

$$Pertes = (A_0 - BE_0) - \frac{(A_1 - BE_1)}{1 + r} \Leftrightarrow Pertes = \frac{BE_1}{1 + r} + \frac{P_1}{\sqrt{1 + r}} - BE_0$$

Le calcul du $(SCR_{Non - Vie})_{provisions}$ revient donc à évaluer :

$$VaR_{99,5\%}^1 \left(\frac{BE_1}{1 + r} + \frac{P_1}{\sqrt{1 + r}} - BE_0 \right) \quad (4.35)$$

Ensuite, nous pouvons calculer le SCR : $VaR_{99,5\%}^1 \left(\frac{BE^1}{1 + r_1} + \frac{P_1}{\sqrt{1 + r_1}} - BE_0 \right)$ où r_1 est le taux d'actualisation à un an et BE_0 le *Best Estimate* du montant de provisions à provisionner calculer en $t=0$ par l'une des méthodes exposées dans la partie précédente.

Application du bootstrap à un an :

Branches	Modèles	moyenne	SCR
Dom.Auto	Poisson	-1,01	2,731
	Gamma	0,06	11,482
Dom.Biens	Poisson	-0,084	2,165
	Gamma	0,002	12,829
Dom.Energie	Poisson	-0,632	9,177
	Gamma	-0,373	22,031

TABLE 4.4 – Résultats du bootstrap GLM à un an

Il semblerait que cette méthode sous-estime le SCR des branches à développement court dans le cas Log Poisson. En revanche les résultats pour les branches longues sont relativement semblables à ceux trouvés par la formule standard ou à l'aide du modèle de Merz-Wüthrich, mis à part le SCR obtenu par modélisation Gamma sur la Dom.Energie (nous avons remarqué précédemment que ce modèle n'était pas adapté au triangle Dom.Energie).

4.4 Modélisation des valeurs extrêmes

L'évaluation des sinistres graves fait notamment appel à la Théorie des Valeurs Extrêmes (en annexe). En pratique la tarification IARD est en général effectuée dans le cadre très général des modèles fréquence-coût :

$$S = \sum_{i=1}^N X_i + I_G \times G$$

avec N le nombre de sinistres (souvent supposé suivre une loi de Poisson), X le coût unitaire d'un sinistre (en général gamma ou log-normal), I_G l'indicatrice de survenance d'un sinistre grave (souvent supposé suivre une loi log-normale) et G le coût d'un sinistre grave (par exemple de type Pareto). La base des sinistres individuels permet de calibrer ces différentes méthodes. Pour chaque méthode illustrée, nous utiliserons les notations suivantes :

- u est le seuil d'acceptation des graves

- p_u la probabilité (empirique) de dépasser le seuil : $p_u = \frac{1}{n} \cdot \sum_{i=1}^n \Omega_{X_i > u}$

Après avoir appliqué chaque méthode de choix du seuil, nous pourrions discuter du choix retenu dans le cadre de notre modélisation.

La recherche du seuil de modélisation des graves ne concerne que le périmètre Dommages aux biens puisque la branche Dommages automobiles n'est constituée que de sinistres attritionnels, modélisés par une approche globale. En effet, les sinistres automobiles sont le plus souvent limités au coût du véhicule, limitant le montant maximal d'indemnisation.

Nous étudions ensuite la distribution de la charge de sinistres à travers la représentation de la fonction de répartition.

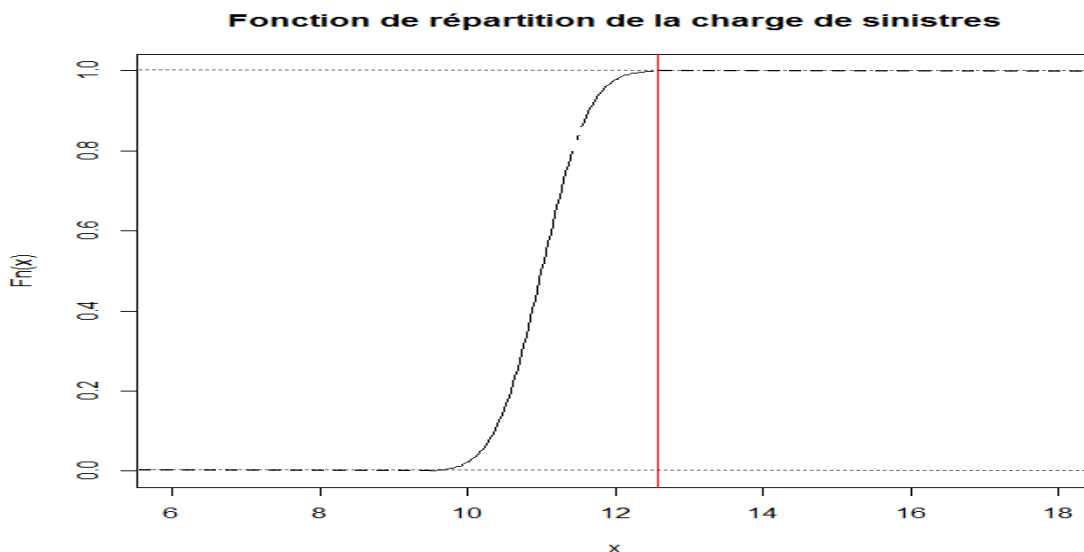


FIGURE 4.1 – Etudes de la distribution de la charge de sinistres

Sur la fonction de répartition de la charge de sinistres, nous observons que la charge de sinistres à 12,366 M\$ est proche d'un quantile à 99%.

Estimation de Hill

Le seuil u correspond au seuil à partir duquel l'estimateur de Hill devient constant.

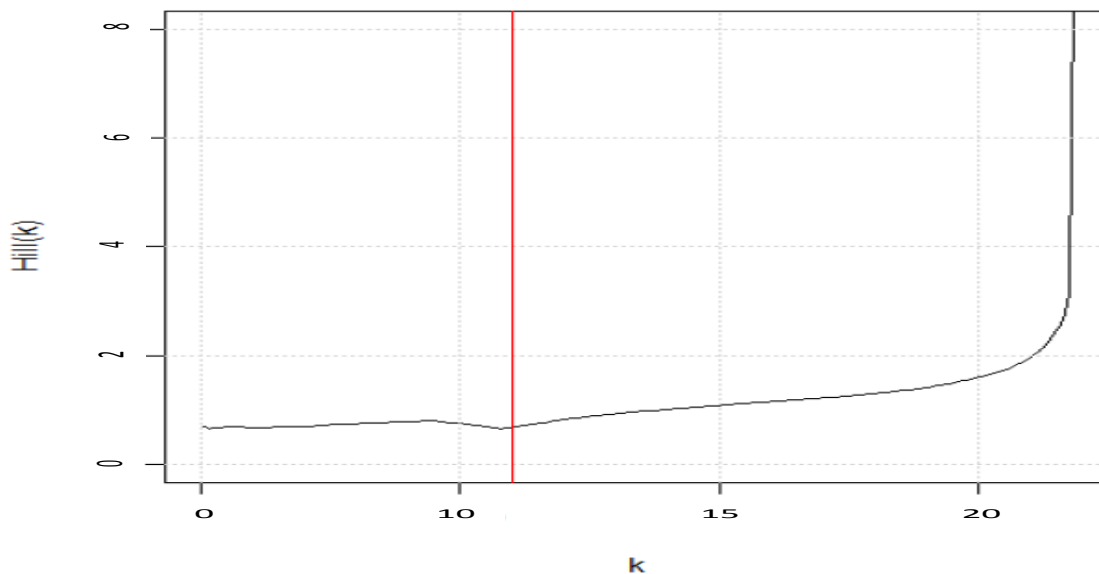


FIGURE 4.2 – Graphique de l'estimateur de Hill en fonction du seuil

Nous observons que l'estimation est relativement stable jusqu'à un seuil de 10,5 M\$. Nous en déduisons d'après cette méthode que l'on peut prendre un seuil de graves jusqu'à 10,5 M\$.

Le Pareto Quantile Plot

C'est une représentation très utile pour visualiser graphiquement si les observations sont distribuées selon une loi du domaine de Fréchet. Dans ce domaine, le graphe $\left(\log \frac{n+1}{j}, \log X_{n-j+1,n}\right)$ serait approximativement linéaire avec une pente de γ pour les petites valeurs de k , c'est-à-dire les points extrêmes. Nous construisons ce graphe sur la base des sinistres.

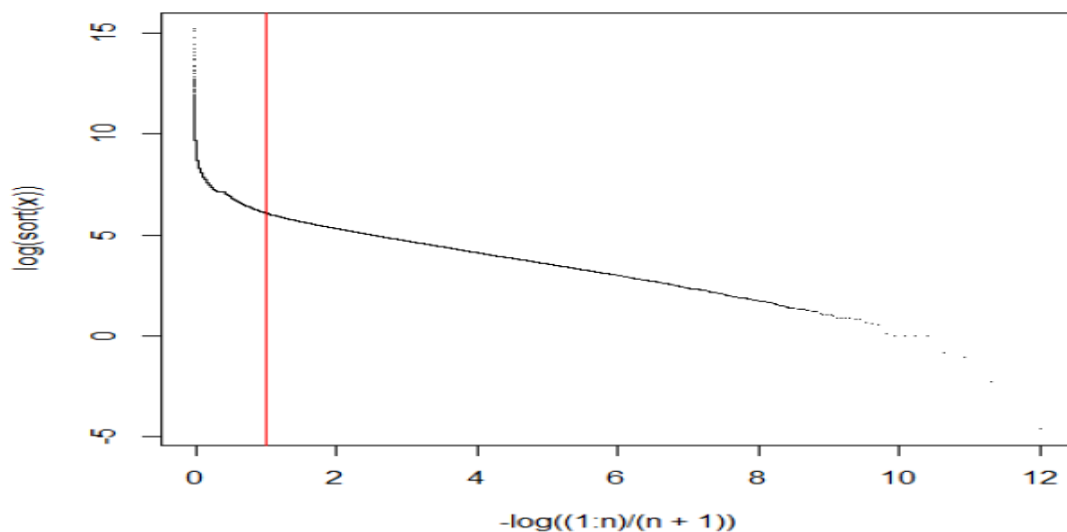


FIGURE 4.3 – Le graphique du Pareto quantile plot

D'après la représentation ci-dessus, nous déduisons que le seuil se situe au tour de 1 en échelle $\log(\cdot)$, ce qui correspond à un montant proche de 10,38 M\$.

Fonction Generalized Pareto Distributions

La recherche du quantile par la fonction GPD permet de calibrer un nouveau choix de seuil, toujours graphiquement. Il correspond à l'abscisse à partir duquel l'estimateur du quantile devient approximativement linéaire.

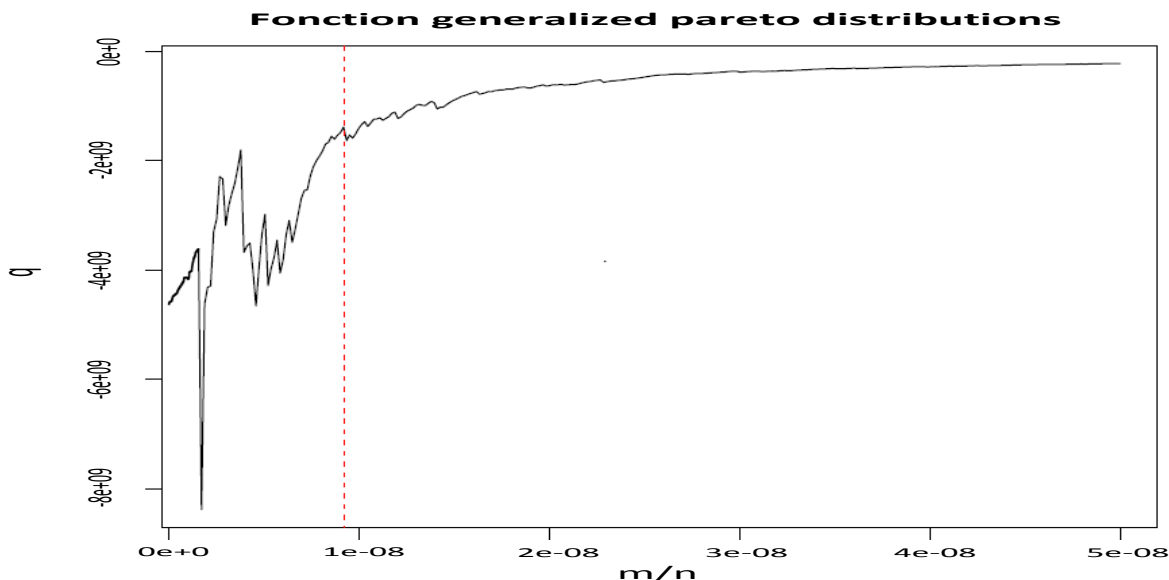


FIGURE 4.4 – Le graphique du Pareto quantile plot

L'estimateur du quantile devient relativement constant à partir du rapport $\frac{m}{n} \approx 8e-8$. Cette estimation équivaut à sélectionner le seuil $u = 11,5$.

Comme attendu, chaque méthode d'estimation fournit un seuil différent pour la modélisation des sinistres graves. Mais en mettant de côté la méthode des valeurs records, qui est significativement moins adaptée par rapport aux autres méthodes, les seuils estimés sont relativement proches : $u \in [10,5;12,5]$.

La dernière méthode est de tester si le seuil choisi est acceptable compte tenu des modélisations effectuées dans le modèle. Le seuil u doit être choisi tel que :

- Les S/P des sinistres attritionnels suivent une loi log-normale
- Le nombre de sinistres graves suivent une loi de Poisson
- Les coûts des sinistres graves suivent une loi de Pareto

Ainsi, nous pouvons calculer pour chaque seuil choisi, la p-value du test de Kolmogorov-Smirnov pour chacune de ces hypothèses.

La théorie des valeurs extrêmes permet de limiter l'étude des seuils de déclenchement des sinistres graves. Pour chaque valeur de seuil de l'intervalle, le seuil finalement retenu sera celui permettant le meilleur ajustement des paramètres de calibrage.

Pour chaque seuil, les valeurs des p-value des tests de Kolmogorov-Smirnov sont calculées :

Seuil	10,5	11,5	12,5
N : Poisson	0,8423	0,8411	0,8369
X : log-normal	0,8221	0,8236	0,8025
G : Pareto	0,1258	0,1126	0,1513

FIGURE 4.5 – Tests de Kolmogorov-Smirnov, selon le choix du seuil

Le seuil $u = 12,5$ semble être pertinent puisqu'il permet d'accepter des branches différents et est compris dans l'intervalle des seuils proposés par le théorème des valeurs extrêmes.

Nous déterminerons par la suite la sensibilité du calcul du SCR par rapport au choix de ce seuil.

Sinistralité attritionnelle

La sinistralité attritionnelle est abordée d'un point de vue global en modélisant la loi du ratio S/P , ratio entre la charge à l'ultime des sinistres attritionnels (c'est-à-dire ne dépassant pas le seuil de déclenchement des sinistres graves) et les primes perçues pour l'année considérée. Les ratios S/P des deux branches sont modélisés par des lois log-normales

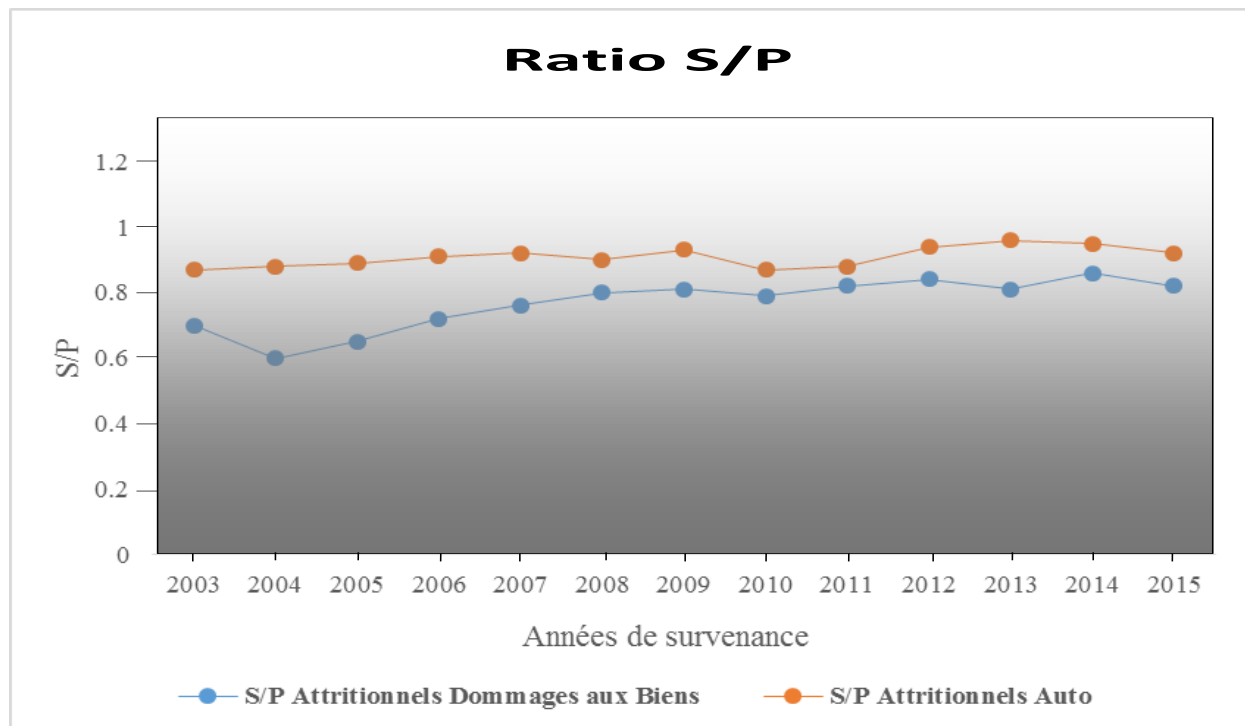


FIGURE 4.6 – Ratios S/P en fonction des années de survenance

Une loi log-normale est calibrée par la méthode du maximum de vraisemblance sur chacun des échantillons de ratios S/P. Pour valider le graphiquement, nous avons utilisé le test d'adéquation de Kolmogorov-Smirnov.

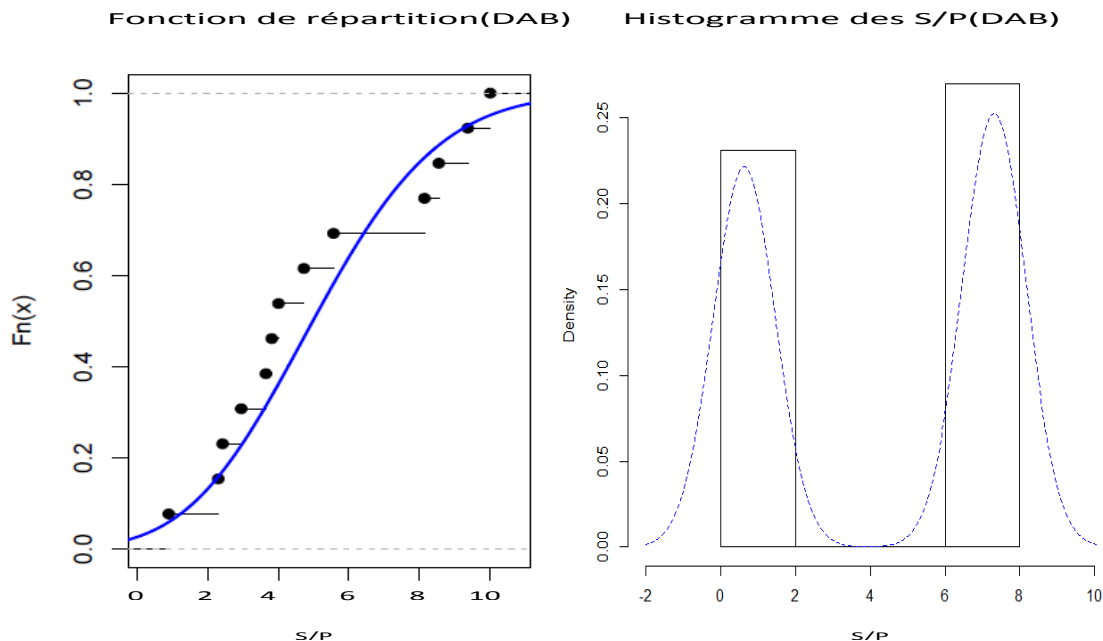


FIGURE 4.7 – Analyses de l'adéquation de la loi log-normale

Le test d'adéquation de lois de Kolmogorov-Smirnov valide statistiquement l'ajustement de loi : la p-value correspondante est de 0,71564.

Nous avons accepté le modèle log-normal comme décrit dans la section précédente.

Sinistralité atypique

L'analyse des sinistres graves s'opère uniquement sur le périmètre Dommages aux Biens. Les sinistres sont étudiés individuellement, par une approche par fréquence et sévérité. La fréquence est calibrée sur le nombre de sinistres par année de survenance tandis que la sévérité est étudiée indépendamment de l'année de survenance.

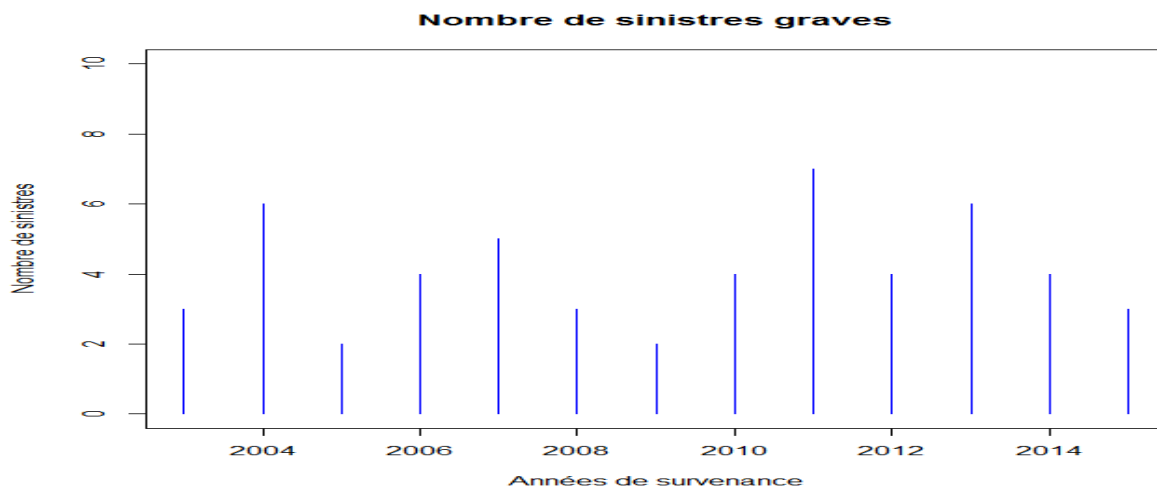


FIGURE 4.8 – Nombre de sinistres graves par année de survenance

Ce nombre de sinistres peut être modélisé par la loi de Poisson. Cette loi est indépendamment des primes acquises pour les années de survenance correspondantes. En effet, il est possible de montrer graphiquement que le nombre de sinistres graves survenant une année est indépendant des primes acquises cette année-là, puisque nous ne pouvons pas déceler de tendances dans le nombre de sinistres en fonction des années de survenance et des primes acquises.

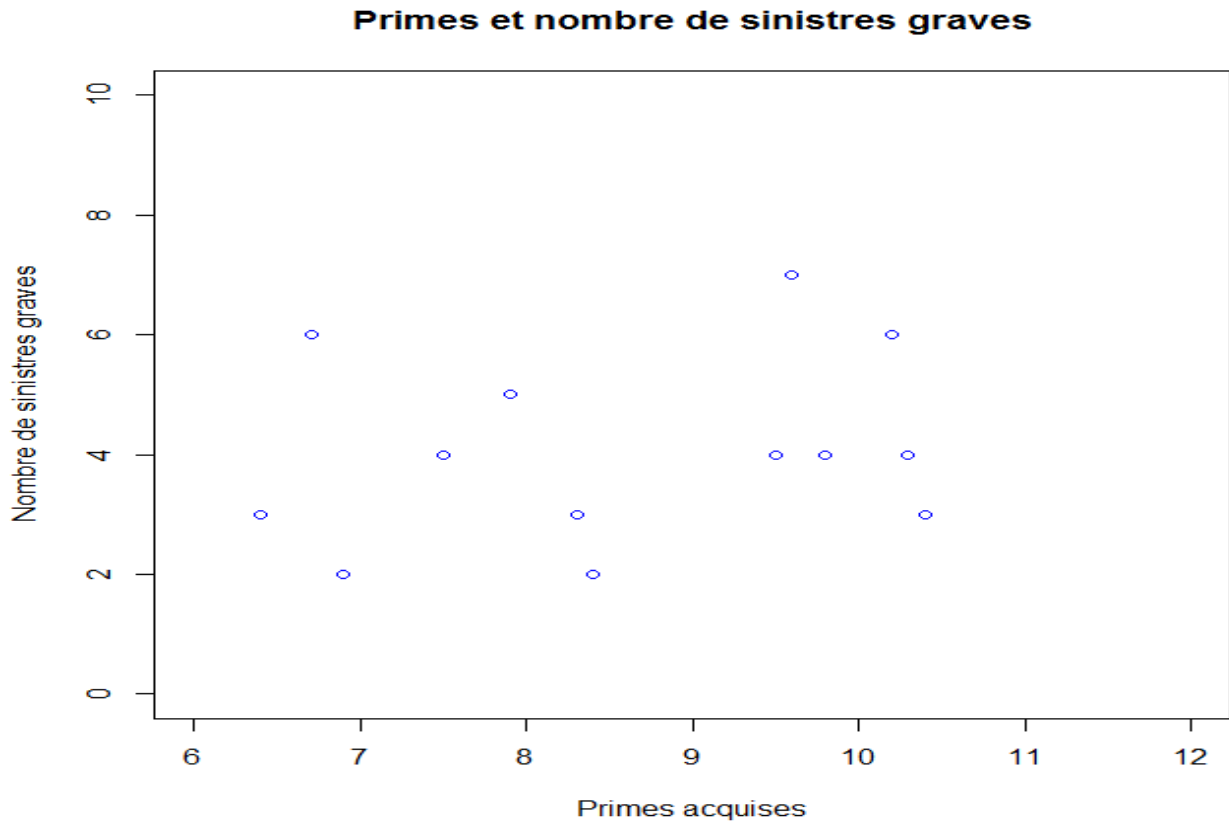


FIGURE 4.9 – Représentation du nombre de sinistres graves en fonction des primes acquises

Le graphique ci-dessus ne permet pas de repérer de dépendance entre le nombre de sinistres graves et les primes acquises. Même si les primes et le nombre total de sinistres sont corrélées positivement, les sinistres graves sont suffisamment rares pour que l'hypothèse d'indépendance paraisse correcte. Le calibrage de la distribution par la méthode du maximum de vraisemblance permet alors d'en déduire les estimations des paramètres de la loi Poisson avec $\lambda = 2,1146$.

Coût des sinistres graves

Les coûts des sinistres sont calibrés sur l'ensemble des sinistres graves de l'historique. De même que la fréquence, le coût d'un sinistre est indépendant de son année de survenance et par conséquent des primes acquises cette année-là. Une représentation de l'étendue des sinistres graves pour chaque année de survenance ne permet pas de déceler une éventuelle tendance.

La loi de Pareto simple ajustée modélise le coût d'un sinistre grave et peut être comparée aux observations empiriques :

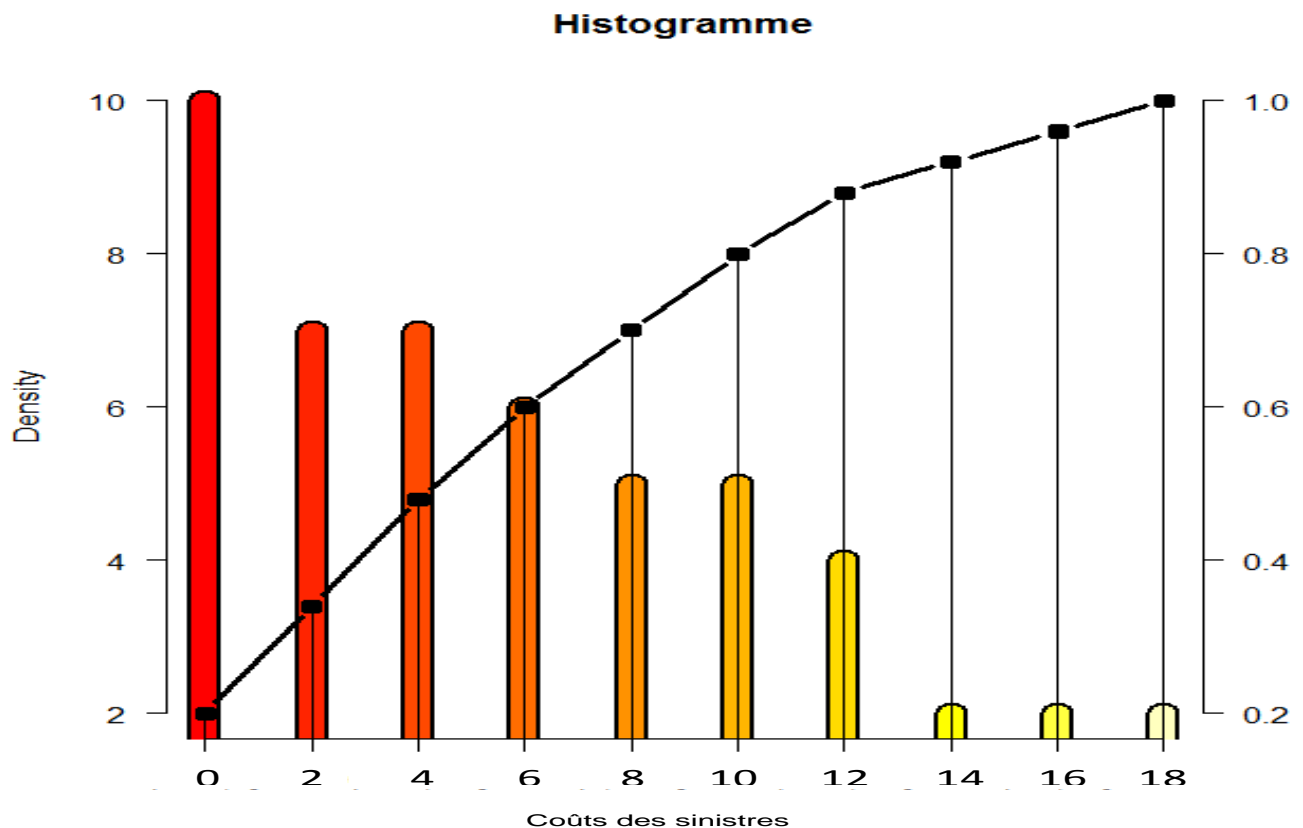


FIGURE 4.10 – Résultats du bootstrap GLM à un an

Le modèle de Pareto simple semble assez adapté à la modélisation des coûts de sinistres atypiques. De plus, le test d'adéquation de Kolmogorov-Smirnov indique une p-value de 0,166 : Le test d'adéquation est accepté.

Application aux Dommages automobiles et Dommages aux biens

La corrélation implicite, notée α^* s'obtient en appliquant la modélisation au triangle agrégé des deux branches, permettant d'en déduire le $SCR_{DAB+Auto}$. Le coefficient de corrélation implicite est ensuite défini tel que :

$$SCR_{DAB+Auto} = \sqrt{SCR_{DAB}^2 + 2.\alpha^*.SCR_{DAB}.SCR_{Auto} + SCR_{Auto}^2} \quad (4.36)$$

Et ainsi :

$$\alpha^* = \frac{SCR_{DAB+Auto}^2 - SCR_{DAB}^2 - SCR_{Auto}^2}{2.SCR_{DAB}.SCR_{Auto}} \quad (4.37)$$

Cette méthode permet d'en déduire la corrélation implicite $\alpha \approx 50\%$, permettant de confirmer que les corrélations de la Formule Standard sont cohérentes et prudentes avec les diversifications réellement observées chez l'assureur.

Nous supposons donc que les incréments $(X_{i,j})$ (Dom.Auto) suivent une loi log-normale de paramètres μ, ω^2 (cela revient à dire que le logarithme de cette loi suit une loi normale d'espérance μ et de variance ω^2) et quantile $N_{0,995}$ d'une loi normale centrée réduite :

$$VaR(C_{2016}^{Auto}; 99, 5\%) = Primes_{2016}^{Auto} \cdot \exp(\widehat{\mu} + \widehat{\omega} \cdot N_{0,995}) \cdot \sum_{i=0}^4 \frac{p_i}{(1 + r_i)^{i+0,5}}$$

Avec :

- C_{2016}^{Auto} : les cadences de paiements

- r_i le taux d'actualisation entre les années 2016 et 2016 + i. Le taux peut être calibré par un modèle de taux CIR_{++} .

Développement	Cadence	Taux d'actualisation
0	74%	0,02345
1	18%	0,02689
2	4%	0,03125
3	2%	0,03345
4	1%	0,03698
5	0%	0,03845
6	0%	0,04125
7	0%	0,04412
8	0%	0,04568
9	0%	0,04728
10	0%	0,04812
11	0%	0,04912
12	0%	0,04956

TABLE 4.5 – Cadences de paiements et actualisation

Donc :

$$SCR_{Auto} = VaR(C_{2016}^{Auto}; 99, 5\%) - Primes_{2016}^{Auto} = 48,72 - 46,7 = 2,02$$

Concernant les "Dommages aux biens". On a :

$$VaR(C_{2016}^{DAB}; 99, 5\%) = VaR \left(\widehat{S/P} \cdot Primes_{2016}^{DAB} \cdot \sum_{i=0}^N \frac{p_i^A}{(1 + r_i)^{i+0,5}} + \sum_{k=1}^{\widehat{N}_s} \widehat{X}_k \cdot \sum_{i=0}^N \frac{p_i^G}{(1 + r_i)^{i+0,5}}; 99, 5\% \right)$$

Avec :

\widehat{N}_s (variable aléatoire de loi de Poisson) le nombre de sinistres graves simulé et \widehat{X}_k (variable aléatoire de loi de Pareto) le coût du k^{ime} sinistre grave.

Les calculs des cadences de paiements pour chaque type de sinistres permettent d'en déduire les charges actualisées.

Développement	Cadence Attri	Cadence Graves	Taux d'actualisation
0	36%	24%	0,02345
1	32%	19%	0,02689
2	11%	22%	0,03125
3	9%	16%	0,03345
4	7%	12%	0,03698
5	4%	6%	0,03845
6	1%	1%	0,04125
7	0%	0%	0,04412
8	0%	0%	0,04568
9	0%	0%	0,04728
10	0%	0%	0,04812
11	0%	0%	0,04912
12	0%	0%	0,04956

TABLE 4.6 – Cadences de paiements et actualisation (DAB)

Le capital réglementaire en découle :

$$SCR_{DAB} = VaR(C_{2016}^{DAB}; 99,5\%) - Primes_{2016}^{DAB} = 42,19 - 40,43 = 1,76$$

Donc :

$$SCR_{DAB+Auto} = \sqrt{1,76^2 + 2 * 0,5 * 1,76 * 2,02 + 2,02^2} = 3,276$$

Les résultats de la théorie des Valeurs Extrêmes mis en place fournissent finalement les résultats suivants :

Branche	Auto	DAB	Agrégation
Méthode	SCR	SCR	SCR
Formule Standard	3,77	3,51	5,306
Théorie des Valeurs Extrêmes	2,02	1,76	3,276

TABLE 4.7 – Comparaison des SCR (en M\$)

La Théorie des Valeurs Extrêmes mis en place pour le risque de prime apporte une réduction de SCR de 38,26% par rapport à l'utilisation de la Formule Standard.

4.5 L'incertitude à un an

La compagnie d'assurance cherche donc à détenir un montant de fonds propres qui permettrait de couvrir le CDR dans 99,5 % des cas. La compagnie d'assurance est exposée à un risque de provisionnement lors de valeurs négatives du CDR, c'est-à-dire, lors de l'observation de malice. Elle va donc calculer le niveau en dessous duquel le CDR ne peut passer que dans 0,5 % des cas. Ce montant correspond à l'opposé de la Value-at-Risk (VaR) du CDR à 0,5 %. Ainsi, le SCR lié au risque de réserve pour l'année de survenance i est défini par :

$$SCR_{rserve}^i = -VaR_{0,5\%}(\widehat{CDR_i}^{n+1}) \quad (4.38)$$

La formule agrégée est :

$$SCR_{rserve} = -VaR_{0,5\%}(\widehat{CDR_{Total}}) = -VaR_{0,5\%}(\sum_{i=1}^n \widehat{CDR_i}^{n+1}) \quad (4.39)$$

Il semble également que, globalement, les incertitudes à un an obtenues par les méthodes bootstrap sont légèrement supérieures à celles obtenues par une formule fermée, à savoir la méthode Merz et Wüthrich. En effet, dans ce type de méthodes simulatoires, on observe une plus grande variabilité des règlements. De plus, concernant les hypothèses du modèle, s'il existe des résidus atypiques dans les triangles, alors cela aura une forte influence sur la détermination de l'erreur de prédiction. C'est d'ailleurs le cas dans nos triangles dans la mesure où on a pu constater des facteurs de développement atypiques. En revanche, le bootstrap apparaît comme une méthode qui simule bien la réalité de la sinistralité, sauf si on observe des tendances et que l'on aimerait en tenir compte. Elle est à privilégier surtout lors de l'étude d'un triangle stable.

On observe une très grande différence entre les SCR obtenus pour les méthodes. Certains SCR sont ainsi bien inférieurs au SCR standard comme la modélisation des valeurs extrêmes. Une telle baisse permet à l'assureur de réduire sa marge de solvabilité mais permet surtout de mieux apprécier son propre profil de risque. La méthode de Merz-Wüthrich, qui apparaît déjà comme la méthode de référence, permet d'obtenir un SCR proche de la formule standard.

Finalement, la méthode de provisionnement à un an que l'organisme assureur doit utiliser en priorité dépend nécessairement du montant du SCR qu'il juge le plus approprié. Il s'agit à la fois de respecter le principe de prudence à l'origine même de la définition du SCR, et d'effectuer un arbitrage entre les méthodes selon les critères à privilégier, comme un SCR proche de la formule standard, l'insertion d'un facteur de queue ou encore l'influence de données exogènes au triangle de liquidation.

Conclusion

Construire un modèle optimal calcule la réserve de primes dans une compagnie d'assurance non-vie est une combinaison de l'art de nombreux facteurs, nous ne pouvons pas dire que l'utilisation d'une méthode est préférable à toute autre dans tous les cas. Le choix d'une méthode dépend avant tout de son adéquation avec les données. Elle tend à surestimer les provisions et semble très sensible aux faibles valeurs qui sont présentes dans les triangles à développement court. Une des caractéristiques les plus importantes des assureurs non-vie à la suite de la décision d'apporter dans le modèle de calcul, le modèle est construit affectent les résultats des entreprises, le modèle est la combinaison harmonieuse entre le solide et les résultats de l'activité d'assurance. Par conséquent, dans cet article nous avons abordé un grand nombre de modèles et de modèles d'utilisation pour être flexible, de temps en temps, chacune des politiques mises en avant dans le court terme de chaque entreprise, en utilisant des résultats cohérents apportera succès pour les entreprises. La méthode Chain-Ladder reste toujours la méthode la plus simple à mettre en œuvre et son utilisation dans la méthode bootstrap permet de diminuer le temps de calcul. Les méthodes Bornhuetter-Fergusson et Cape Code permet d'augmenter la garantie de l'instabilité des données. Ensuite, nous nous sommes intéressés aux modèles stochastiques qui sont : le modèle de Mack, le modèle Log-normal et la modélisation GLM nous donnent une perspective plus claire sur la distribution des réserves. Nous avons cependant préféré l'utilisation des modèles GLM, en particulier le modèle Log Poisson Surdispersé. Il semble s'adapter à tout type de branche et donne des erreurs de prédiction relativement faibles.

Ces dernières années, le sujet de provisionnement a été à l'origine de plusieurs controverses, principalement au sujet du choix du modèle le plus performant à utiliser pour obtenir la meilleure estimation des réserves. Mais, il n'y a pas de réponse particulière à une telle question. Certains modèles s'adapteront à certains problèmes et type de données. Lorsqu'un calcul de provision est réalisé, les données devraient être examinées en détail pour trouver un modèle approprié, plutôt que d'utiliser la même approche de modélisation dans toutes les circonstances, que le modèle soit déterministe ou stochastique. Calculer des réserves est un exercice pratique d'analyse des données pour lequel il est vital de s'efforcer de comprendre et apprendre des données plutôt que d'imposer la même approche dans toutes les situations.

Pour l'évaluation du risque à un an, nous avons cherché à analyser l'incertitude liée aux Claims Development Results (CDR) liée à la différence entre deux esti-

mations successives du montant des provisions. Plusieurs méthodes ont ainsi été développées dans ce mémoire pour déterminer une telle incertitude. Celle de Merz et Wüthrich est désormais une référence en la matière, mais diverses méthodes bootstrap ont également montré leurs atouts. La méthode de Merz et Wüthrich est utilisée également dans le calcul du SCR non-vie et comparée aux méthodes GLM bootstrap, la formule standard. Nous avons effectué une analyse plus approfondie sur les types de sinistres (attritionnels/graves). Nous avons choisi la Theorie des Valeurs Extrêmes. Un certain nombre de risques peuvent être qualifiés de “risques extrêmes”, car ils présentent une probabilité d’occurrence très faible (ou un temps de retour très élevé). Ils correspondent à des événements rares. Ce modèle présente aussi l’avantage d’être simple et plus flexible que la Formule Standard. Ainsi, l’évolution du profil de risque de la compagnie peut entraîner une adaptation du modèle en conséquence, comme nous l’avons par exemple envisagé au sujet des tendances des ratios de sinistralité. Elle permet d’estimer la distribution des sinistres graves, et par conséquent d’obtenir un montant du SCR. Elle donne des résultats plus faibles que la formule standard.

L’évaluation de la force financière d’une compagnie d’assurance non-vie inclut une analyse des règlements attendus et une estimation de la variabilité possible du provisionnement. Toute erreur de calcul dans le provisionnement pourra causer l’insolvabilité de certains assureurs. il est important de choisir la meilleure méthode pour l’estimation des réserves afin d’éviter tous effets négatifs dans la position financière futur de l’entreprise.

Le travail de l’actuaire est d’assurer le suivi de la situation de la volatilité du marché à fournir des solutions à des situations dans le temps, les modèles d’assurance différente applique aux entreprises et aux différentes opérations d’assurance. Calculer et prédire la performance de l’assurance est importante et nous avons discuté principalement dans la thèse. La prime, réserve de prime calculée, les prévisions de la volatilité, la maîtrise des risques, la réassurance, d’optimiser le travail qui est lié au modèle que l’actuaire devrait se développer.

Bibliographie

- [1] Frédéric PLANCHET : Le modèle collectif, Support de cours 2003-2004.
- [2] PLANCHET, F. (2011). bootstrap et méthodes de triangles, [actu.dactuaires.typepad.com/ laboratoire/2011/07/bootstrap-et-méthodes-detriangles. html](http://actu.dactuaires.typepad.com/laboratoire/2011/07/bootstrap-et-methodes-detriangles.html).
- [3] Frédéric PLANCHET et Pierre-E.THÉRON : Allocation d'actifs selon le critère de maximisation des fonds propres économiques en assurance non-vie,2005.
- [4] Frédéric PLANCHET : Mesure de l'incertitude sur le taux de couverture des engagements dans un cadre orsa ; le cas de l'assurance non-vie,2010.
- [5] Peter D. ENGLAND et Richard D. VERRALL : Stochastic Claims Reserving in General Insurance (with discussion). British Actuarial Journal 8, III, 2002.
- [6] PARTRAT Christian, BESSON Jean-Luc , Assurance non-vie, Modélisation, Simulation, Economica.
- [7] Bradley EFRON : Bootstrap Methods : Another Look at the Jackknife. The Annals of Statistics, 1979.
- [8] Frédéric PLANCHET et Pierre-E.THÉRON : Simulation de trajectoires de processus continus . Les cahiers de recherche de l'ISFA, WP 2004.
- [9] E. Ohlsson and B. Johansson. : Non-life insurance pricing with Generalized Linear Models. Springer Verlag, 2010.
- [10] M. V. Wüthrich and M. Merz. Stochastic Claims Reserving Methods in Insurance. Wiley Interscience, 2008.
- [11] M. ; Dhaene J. Kaas, R. ; Goovaerts and M. Denuit. Modern Actuarial Risk Theory. Springer Verlag, 2009.
- [12] M. Denuit and A. Charpentier. Mathématiques de l'assurance non-vie :Tarification et provisionnement. Tome 2. Economica, 2005.
- [13] BORNHUETTER, R.L., FERGUSON, R.E. (1972). The actuary and IBNR, Proc Casualty Actuarial Society, Vol. LIX, 181-195.

- [14] JAZIRI, S. (2011). Méthodes de provisionnement non-vie et risque de réserve à un an, Mémoire d'actuariat, ISFA.
- [15] LACOUME, A. (2008). Mesure du risque de réserve sur un horizon de un an, Mémoire d'actuariat, ISFA.
- [16] MERZ, M., WÜTHRICH, M. V. (2007). Prediction Error of the Expected Claims Development Result in the Chain Ladder Method, North American Actuarial J.
- [17] Hélène COMPAIN, Analyse du risque de provisionnement non-vie dans le cadre de la réforme Solvabilité II (2009-2010), DAUPHINE.
- [18] OHLSSON, E., LAUZENINGKS, J. (2008). The one-year non-life insurance risk, ASTIN Colloquium in Manchester.
- [19] VOGT, A. (2010). Modélisation stochastique en assurance non vie : calcul de l'exigence de capital réglementaire sous Solvency II, Mémoire d'actuariat, UFR-FSEG.
- [20] MACK T., Distribution-free calculation of the standard error of Chain Ladder reserve estimates, Astin bulletin, Vol 23, 1993.
- [21] GILLET A., SERRA B., Effets de la dépendance entre différentes branches sur le calcul des provisions, Mémoire d'actuariat ENSAE, 2002.
- [22] KRAUTH G., Provisionnement et corrélation entre branches, Mémoire d'actuariat CNAM, 2007.
- [23] Frédéric PLANCHET : Utilisation de la théorie des valeurs extrêmes dans le contexte solvabilité 2, Support de cours 2015-2016, Juin 2016.
- [24] F. PLANCHET, «Assurances des biens et de responsabilité - Enjeux actuariels,» Formation Groupama, 2012.
- [25] Paul EMBRECHTS, Sidney I. RESNICK et Gennady SAMORODNITSKY : Extreme Value Theory as a Risk Management Tool. North American Actuarial Journal, Volume 3, Number 2, 1999.
- [26] PLANCHET F., THÉRON P.E., JUILLARD M. [2011] Modèles finan-

ciers en assurance, seconde édition, Paris : Economica.

[27] J. BEIRLANT, Y. GOEGEBEUR, J. SEGERS, J. TEUGELS, D. D. WAAL et C. FERRO, *Statistics of Extremes : Theory and Applications*, WILEY, 2004.

[28] A. GUILLOU et A. YOU, «Introduction à la théorie des valeurs extrêmes : Applications en Actuariat,» chez Université de Strasbourg & Société Générale Insurance, Strasbourg, 2011.

[29] Romain LAÏLY, «Construction d'un Modèle Interne Partiel en Assurance non-vie,» Diplôme d'Actuaire EURIA, 2014.

[30] PARTRAT C., BESSON J.L., [2004] *Assurance non-vie – modélisation, simulation*, Paris : Economica.

[31] VASECHKO O.A.; GRUN-REHOMME M.; BENLAGHA N. [2009] «Modélisation de la fréquence des sinistres en assurance automobile », *Bulletin Français d'Actuariat*, vol. 9, n18.

[32] Calcul stochastique de provisions techniques d'assurance non-vie - Christophe Bonnefoy et Bruno Devictor, ENSAE 2007.

[33] Provisionnement technique en assurance non-vie – C. PARTRAT, J.M. NESSI, E. LECOEUR, Economica, 2007.

[34] *Methods and Models of Loss Reserving Based on Run-Off Triangles : A Unifying Survey* - Klaus D. Schmidt, Casualty Actuarial Society Forum, Fall 2006.

[35] The concept of reserving and reserving methodologies in general insurance – Elena Pantzopoulou, City University of London, 2003.

[36] S. PITREBOIS, P. D. LONGUEVILLE, M. DENUIT et J.-M. WALHIN, «Etude de techniques IBNR modernes,» *actu-L* 2, pp. 29-62, 2002.

[37] J. NELDER et R. WEDDERBURN, «Generalized linear models,» *Journal of the Royal Statistical Society, Series A* 135, pp. 370-384, 1972.

[38] R. W. KEENER, «Statistical Theory : Notes for a Course in Theoretical Statistics,» Springer, p. 27–28 ; 32–33, 2006.

Annexes

Annexe 1 : Triangle des paiements cumulés(Dom.Biens).

	0	1	2	3	4	5	6	7	8	9	10	11	12
2003	26,45	29,08	29,334	29,566	29,919	30,256	30,329	30,382	30,382	30,382	30,382	30,382	30,382
2004	27,62	32,6	32,765	33,042	33,305	33,52	33,709	33,709	33,709	33,709	33,709	33,709	33,709
2005	25,71	30,13	30,227	30,508	30,736	30,949	31,096	31,129	31,129	31,129	31,129	31,129	31,129
2006	22,34	27,88	28,158	28,169	28,253	28,296	28,352	28,352	28,352	28,352	28,352	28,352	28,352
2007	18,94	26,59	26,957	27,111	27,325	27,354	27,497	27,57	27,57	27,57	27,57	27,57	27,57
2008	19,36	25,23	25,567	25,621	25,837	25,91	26,115	26,192	26,192	26,192	26,192	26,192	26,192
2009	21,32	28,67	28,958	29,551	29,626	29,789	29,822	29,822	29,822	29,822	29,822	29,822	29,822
2010	24,83	29,31	29,671	29,747	30,179	31,022	31,022	31,022	31,022	31,022	31,022	31,022	31,022
2011	25,73	33,62	33,887	34,45	34,682	34,682	34,682	34,682	34,682	34,682	34,682	34,682	34,682
2012	24,68	30,61	31,001	31,533	31,533	31,533	31,533	31,533	31,533	31,533	31,533	31,533	31,533
2013	25,73	34,57	34,938	34,938	34,938	34,938	34,938	34,938	34,938	34,938	34,938	34,938	34,938
2014	26,89	33,24	33,24	33,24	33,24	33,24	33,24	33,24	33,24	33,24	33,24	33,24	33,24
2015	23,48	23,48	23,48	23,48	23,48	23,48	23,48	23,48	23,48	23,48	23,48	23,48	23,48

Annexe 2 : Triangle des paiements cumulés(Dom.Energie).

	0	1	2	3	4	5	6	7	8	9	10	11	12
2003	39,56	43,79	47,02	49,160	51,003	52,047	52,48	53,213	53,537	54,069	54,384	54,81	55,574
2004	36,29	40,27	43,03	45,260	46,523	47,738	47,927	48,364	48,896	49,531	49,894	50,236	50,236
2005	29,71	35,13	38,42	40,701	42,929	44,142	44,289	44,433	45,087	45,412	45,778	45,778	45,778
2006	28,92	34,16	38,5	40,811	42,060	42,496	42,609	43,241	43,974	44,3	44,3	44,3	44,3
2007	27,82	34,27	38,52	41,735	43,367	44,90	44,926	44,927	45,37	45,37	45,37	45,37	45,37
2008	25,49	28,36	30,58	31,793	32,821	33,871	33,963	34,807	34,807	34,807	34,807	34,807	34,807
2009	21,35	24,7	26,98	28,217	29,430	31,174	31,607	31,607	31,607	31,607	31,607	31,607	31,607
2010	20,14	24,62	26,98	28,208	29,240	30,242	30,242	30,242	30,242	30,242	30,242	30,242	30,242
2011	18,28	22,17	23,54	24,708	25,763	25,763	25,763	25,763	25,763	25,763	25,763	25,763	25,763
2012	23,82	26,75	29,14	30,366	30,366	30,366	30,366	30,366	30,366	30,366	30,366	30,366	30,366
2013	19,32	23,16	26,52	26,52	26,52	26,52	26,52	26,52	26,52	26,52	26,52	26,52	26,52
2014	22,27	29,62	29,62	29,62	29,62	29,62	29,62	29,62	29,62	29,62	29,62	29,62	29,62
2015	16,29	16,29	16,29	16,29	16,29	16,29	16,29	16,29	16,29	16,29	16,29	16,29	16,29

Annexe 3 : Estimations des paramètres des modèles linéaires généralisés et leur risque associé.

Modèle de Poisson			Modèle Gamma		
paramètres de régression	Estimateur du max de vrais.	Ecart type (poisson simple)	Ecart type (poisson surdispersé)	Estimateur du max de vrais.	Ecart type
Intercept	0,14471	0,00024	0,0175	0,06254	
aref =a0	0	0	0	0	
a1	-0,1283	0,0012	0,1245	-0,1246	0,1589
a2	0,0452	0,0036	0,1545	-0,1545	0,1687
a3	0,4479	0,0052	0,1455	0,1466	0,1354
a4	-0,3188	0,0005	0,2644	-0,2178	0,1547
a5	-0,3188	0,0044	0,1546	0,1234	0,1268
a6	0,2135	0,0065	0,1453	-0,1145	0,2154
a7	-0,1214	0,0055	0,2465	0,3244	0,2564
a8	0,0321	0,0026	0,1547	-0,1349	0,1267
a9	-0,1554	0,0004	0,2546	0,1647	0,1239
a10	0,3658	0,0001	0,1544	0,1338	0,2168
a11	0,1334	0,0086	0,1536	-0,1435	0,1298
a12	-0,1235	0,0034	0,1564	-0,2164	0,1248
a13	0,1247	0,0032	0,2714	0,2485	0,1472
a14	-0,2365	0,0054	0,1445	0,1326	0,1589
a15	-0,3657	0,0064	0,1565	-0,2165	0,1266
a16	0,2368	0,0001	0,1456	-0,1556	0,1478
a17	0,1478	0,0064	0,1547	-0,2354	0,1354
a18	0,3365	0,0048	0,1268	0,1254	0,1238
a19	0,6871	0,0015	0,2548	0,2418	0,1257
a20	0,1547	0,0065	0,1256	0,1259	0,1397
bref=b0	0,2364	0	0	0	0
b1	0,3654	0,0057	0,1264	0,2887	0,1278
b2	-0,3668	0,0014	0,1475	-0,1579	0,2159
b3	0,1487	0,0055	0,1264	-0,1356	0,2256
b4	-0,6412	0,0057	0,2324	0,1255	0,2579
b5	0,1257	0,0038	0,1655	0,1268	0,2365
b6	0,2668	0,0015	0,2155	-0,2448	0,2148
b7	-0,1478	0,0014	0,1325	0,2541	0,2117
b8	0,3685	0,0035	0,2484	0,1237	0,1268
b9	0,3356	0,0036	0,1213	0,3657	0,2148
b10	-0,1685	0,0017	0,1354	-0,4464	0,1257
b11	0,2368	0,0001	0,3264	0,2664	0,1258
b12	0,1478	0,0006	0,1355	-0,1346	0,2015
b13	-0,3564	0,0007	0,1534	0,2014	0,2684
b14	-0,1264	0,0006	0,2856	0,1026	0,1218
b15	-0,3646	0,0004	0,1244	-0,2018	0,1169
b16	-0,2157	0,0003	0,2156	0,2684	0,1258
b17	-0,3564	0,0001	0,1538	0,2368	0,2364
b18	0,1238	0,0047	0,2456	-0,2446	0,2154
b19	0,3354	0,0035	0,1643	-0,1264	0,1297
b20	0,3689	0,0025	0,2154	0,1287	0,2015

Annexe 4 :Loi de probabilité de la variable de réponse.

Loi Normale

$$f(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \text{ pour } -\infty < y < \infty$$
$$\phi = \sigma^2$$
$$\text{Var}(Y) = \sigma^2$$

Loi Gamma :

$$f(y) = \frac{1}{\tau(v)y} \left(\frac{yv}{\mu}\right)^v \exp\left(-\frac{yv}{\mu}\right), \text{ pour } 0 < y < \infty$$
$$\phi = v^{-1}$$
$$\text{Var}(Y) = \frac{\mu^2}{v}$$

Loi de Poisson

$$f(y) = \frac{\mu^y e^{-\mu}}{y!}, \text{ pour } y = 0, 1, 2, \dots$$
$$\text{Var}(Y) = \mu$$

Loi Binomiale

$$f(y) = \binom{n}{r} \mu^r (1-\mu)^{n-r}, \text{ pour } y = r/n, r = 0, 1, 2, \dots, n$$
$$\text{Var}(Y) = \mu(1-\mu)/n$$

Loi Binomiale Négative

$$f(y) = \frac{\theta(y+1/k)}{\theta(y+1)\theta(1/k)} \frac{(k\mu)^k}{(1+k\mu)^{y+1/k}} \text{ pour } y = 0, 1, 2, \dots$$
$$\text{Var}(Y) = \mu(1-\mu)/n$$

-Le modèle de la loi de Poisson surdispersée

Soient μ, ϕ deux réels strictement positifs. Une variable aléatoire X suit la loi de Poisson surdispersée de paramètre (μ, ϕ) si et seulement si X/ϕ suit une loi de Poisson de paramètre μ / ϕ . Dans ce cas, nous notons $X \Xi \text{ODP}(\mu, \phi)$.

Avec le paramètre ϕ en plus, une loi de Poisson surdispersée généralise une loi de Poisson habituelle. Elle permet ainsi une relation plus flexible entre la variance et l'espérance de la variable. Concrètement, $\text{Var}[X] = \phi E[X]$. Par ailleurs, la famille de loi de Poisson surdispersée possède une propriété intéressante : elle est invariante par l'additivité montrée dans le théorème connu suivant :

- Soient X_1, X_2 deux variables aléatoires telles que $X_1 \Xi \text{ODP}(\mu_1, \phi)$ et $X_2 \Xi \text{ODP}(\mu_2, \phi)$. Alors, nous avons $X_1 + X_2 \Xi \text{ODP}(\mu_1 + \mu_2, \phi)$.

Ces propriétés ont un rôle important dans le modèle linéaire généralisé basé sur l'hypothèse suivante :

- (GLM, loi de Poisson surdispersée) Soient $\Delta = (Y_{i,j})_{i \geq 1, j \geq 1}$ les règlements non cumulés, alors pour tout $i, j = 1, 2, \dots, n$:

+ Les $(Y_{i,j})_{i,j}$ sont indépendants.

+ $Y_{i,j}$ suit la loi de Poisson surdispersée de paramètre $(\mu_{i,j}, \phi)$ avec $\phi > 0$.

+ Il existe des paramètres réels $\mu, \alpha_1, \alpha_2, \dots, \alpha_n$ et $\beta_1, \beta_2, \dots, \beta_n$ tels que

$$\mu_{i,j} = \exp(\mu + \alpha_i + \beta_j)$$

+ $\alpha_1 = 0$ et $\beta_1 = 0$

Les paramètres $(\alpha_i)_i$ ont pour rôle de capter la tendance des règlements par rapport aux années d'origine. De même façon, les $(\beta_j)_j$ décrivent la tendance des règlements par rapport aux années de développement alors que μ permet un contrôle global.

Nous avons à notre disposition la réalisation $y_{i,j}$ des variables aléatoires $Y_{i,j}$ avec $i + j \leq n$. Cet échantillon servira à estimer le vecteur des paramètres $\theta = (\mu, \alpha, \beta)$ où $\alpha = (\alpha_2, \dots, \alpha_n)$ et $\beta = (\beta_2, \dots, \beta_n)$ par la méthode de maximum de vraisemblance. Une fois avoir obtenu l'estimateur $\hat{\theta}$ de θ , nous remplirons les cases manquantes, c'est-à-dire les cases (i,j) où $n + 1 < i + j \leq 2n$ par leurs valeurs espérées estimées $\hat{E}[Y_{i,j}] = \hat{\mu}_{i,j} = \exp(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j)$.

-Estimation des paramètres

Etant donné que la variable $Y_{i,j}$ suit la loi de Poisson surdispersée de paramètres $(\mu_{i,j}, \phi)$, la densité de probabilité de $Y_{i,j}$ en $y_{i,j}$ est :

$$P(Y_{i,j} = y_{i,j}) = \frac{(\mu_{i,j}/\phi)^{y_{i,j}/\phi} \exp(-\mu_{i,j}/\phi)}{(y_{i,j}/\phi)!} \quad (4.40)$$

Par ailleurs, l'indépendance des variables $Y_{i,j}$ et nous permet d'écrire le logarithme de la vraisemblance du modèle comme suit :

$$\begin{aligned} l(\Delta, \theta) &= \log \prod_{i+j \leq n} \frac{(\mu_{i,j}/\phi)^{y_{i,j}/\phi}}{(y_{i,j}/\phi)!} \\ &= \frac{1}{\phi} \sum_{i+j \leq n} (y_{i,j}(\mu + \alpha_i + \beta_j) - \exp(\mu + \alpha_i + \beta_j)) - f(y_{i,j}, \phi) \end{aligned} \quad (4.41)$$

avec $f(y_{i,j}, \phi)$ et ϕ ne jouent aucun rôle dans la maximisation de la vraisemblance par rapport au paramètre θ . De ce fait, l'estimateur de maximum de vraisemblance du paramètre θ est le résultat du programme suivant :

$$\hat{\theta} = \operatorname{argmax}_{\theta} l^*(\Delta, \theta) = \sum_{i+j \leq n} (y_{i,j}(\mu + \alpha_i + \beta_j) - \exp(\mu + \alpha_i + \beta_j)) \quad (4.42)$$

Pour résoudre ce programme, nous dérivons $l^*(\Delta, \theta)$ par rapport à chaque composante pour obtenir les conditions de premier ordre. La solution de ces conditions sera l'estimateur de maximum de vraisemblance $\hat{\theta} = (\hat{\mu}, \hat{\alpha}, \hat{\beta})$. Les conditions de premier ordre sont :

$$\begin{aligned} \text{par rapport à } \mu : & \sum_{i+j \leq n} x_{i,j} - \sum_{i+j \leq n} \exp(\mu + \alpha_i + \beta_j) = 0 \\ \text{par rapport à } \alpha_i : & \sum_{j \leq n-i} x_{i,j} - \sum_{j \leq n-i} \exp(\mu + \alpha_i + \beta_j) = 0 \\ \text{par rapport à } \beta_j : & \sum_{i \leq n-j} x_{i,j} - \sum_{i \leq n-j} \exp(\mu + \alpha_i + \beta_j) = 0 \end{aligned}$$

Une manière classique d'étudier la qualité de l'ajustement d'un modèle linéaire généralisé est d'utiliser les résidus de Pearson, qui, dans le cas présent, se définissent comme :

$$r_{i,j}^P = \frac{y_{i,j} - \hat{m}u_{i,j}}{\sqrt{\hat{\phi}\hat{\mu}_{i,j}}} \quad i,j = 1,2,\dots,n \text{ tel que } i + j \leq n + 1$$

Si le modèle s'ajuste bien aux données, les $r_{i,j}^P$ ne devraient pas contenir encore beaucoup d'informations et devraient être approximativement normalement distribués. Nous pouvons le vérifier en utilisant un QQ-plot entre l'échantillon $r_{i,j}^P$ et la loi normale centrée réduite.

Après avoir estimé les paramètres (μ, α, β) par la méthode du maximum de vraisemblance, nous pouvons à ce stade obtenir le tableau de sinistralité complet, c'est-à-dire les cellules (i,j) , où $n + 1 \leq i + j \leq 2n$, avec leur montant espéré estimé $\hat{\mu}_{i,j} = \exp(\hat{\mu} + \hat{\alpha}_i \hat{\beta}_j)$. Ainsi, la provision espérée à horizon 1 an et ultime par année d'origine et au niveau global sont estimées par :

$$E[\hat{X}_i^1] = \hat{\mu}_{i,n+1-i} \quad (4.43)$$

$$E[\hat{X}^1] = \sum_{i=0}^n E[\hat{X}_i^1] \quad (4.44)$$

$$E[\hat{X}_i^u] = \sum_{j=n+1-i}^n \hat{\mu}_{i,j} \quad (4.45)$$

$$E[\hat{X}^u] = \sum_{i=0}^n E[\hat{X}_i^u] \quad (4.46)$$

Annexe 5 : Théorie des valeurs extrêmes

On définit les variables aléatoires M_n et m_n qui traduisent respectivement le maximum et le minimum d'une réalisation d'une variable aléatoire X , par :

$$M_n = \max(X_i)_{1 \leq i \leq n} \text{ et } m_n = \min(X_i)_{1 \leq i \leq n}$$

En théorie des valeurs extrêmes, le but visé est de déterminer la loi que suit le maximum ou le minimum en fonction de celle de la variable aléatoire X .

Pour ce faire, on calcule la fonction de répartition :

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

$F_{M_n}(x)$ converge donc vers 0 ou 1 selon la valeur de $F(x) \in [0,1]$. La loi limite de M_n est donc dégénérée, il convient de renormaliser linéairement M_n :

$$M_n^* = \frac{M_n - b_n}{a_n}$$

Avec a_n et b_n deux suites déterministes convenablement choisies afin d'éviter la dégénérescence de la loi limite, le théorème de Fisher-Tippett montre que la loi limite M_n^* de M_n suit une loi GEV définie ci-dessous : $X \sim \text{GEV}(\mu, \sigma, k)$ avec $\sigma > 0$ avec :

- Si $k \neq 0$: $F_{\text{GEV}}(x) = \exp \left(- \left[1 + k \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/k} \right)$ pour $1 + k \left(\frac{x - \mu}{\sigma} \right) > 0$

- Si $k = 0$: $F_{\text{GEV}}(x) = \exp \left(- \exp \left(- \left(\frac{x - \mu}{\sigma} \right) \right) \right)$ pour $x \in \mathbb{R}$

Trois paramètres sont pris en compte :

- μ est un paramètre de position
- σ est un paramètre d'échelle
- k est un paramètre de forme ($k = 0$: loi de Gumbel, $k > 0$: loi de Fréchet, $k < 0$: loi de Weibull)

Estimateur de Hill

La méthode de Hill est une méthode graphique, elle repose sur le graphe de Hill (Hill-plot), développé dans l'article *An application of Extreme Value Theory for Measuring Financial Risk*.

L'estimateur de Hill est défini de la façon suivante :

$$\hat{\gamma}_{X,k,n}^{(H)} = \frac{1}{k} \sum_{j=1}^k \log(X_{(n-j+1)}) - \log(X_{(n-k)}) \quad (4.47)$$

Avec $X_{(1)}, \dots, X_{(n)}$ les statistiques d'ordre associées à l'échantillon X_1, \dots, X_n .

L'estimateur de Hill correspond au coefficient linéaire de la pente du nuage de points

$$\left(\log\left(\frac{n+1}{j}\right), \log(X_{n-j+1}) \right)_j \text{ (Pareto quantile plot)}$$

Cet estimateur n'est valable que dans le cas d'un indice k positif (distribution de Fréchet).

Le seuil e correspond à l'indice k à partir duquel l'indice de Hill devient approximativement constant, en suivant le même raisonnement que précédemment.

Fonction GPD et estimation de la queue de distribution

La distribution GPD utilisée dans cette partie se présente sous la forme suivante :

$$G(x) = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{\frac{-1}{\xi}} & \text{si } \xi \neq 0 \\ 1 - e^{-\frac{x}{\beta}} & \text{si } \xi = 0 \end{cases} \quad (4.48)$$

Il s'agit d'une méthode qui consiste à choisir le seuil comme étant le quantile d'ordre $\frac{m_n}{n}$ de la loi des données, avec m_n le nombre des excès fixé tel que $1 < m_n < n$, tendant vers l'infini avec la taille n de l'échantillon, mais restant petit devant n .

Plus précisément $u_n = F^{-1}(1 - m_n/n)$ où F^{-1} est la fonction inverse de F et m_n est tel que :

$$\lim_{n \rightarrow \infty} m_n = +\infty \text{ et } \lim_{n \rightarrow \infty} \frac{m_n}{n} = 0$$

Les travaux de Reiss et Thomas (1997) ont mis en avant un test d'adéquation de la queue de distribution des données. Le modèle doit s'adapter aussi bien à l'ensemble de la distribution des données qu'à la queue de la distribution :

$$\forall x \geq u, F(x) = P(X \leq x) = (1 - F(u)) \cdot F_u(x - u) + F(u)$$

Et d'après la théorie des valeurs extrêmes : $F_u(x - u) \approx H_{k,\sigma}(x - u)$ pour u élevé. De plus $F(u)$ peut être approchée par l'estimateur empirique $F_n(u) = 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{X_i > u}$.

Cela permet de déduire l'estimation de $F(x)$:

$$\forall x \geq u, \hat{F}(x) = P(X \leq x) = (1 - F_n(u)) \cdot H_{k,\sigma}(x - u) + F_n(u)$$

$\hat{F}(x)$ est une distribution de Pareto généralisée, de même paramètre de forme k mais dont l'échelle et la position sont différentes.

L'estimateur du quantile extrême $q_{GPD,n}$ issu de l'approximation par la loi de Pareto généralisée, sera utilisé pour la détermination du seuil. Il est développé dans l'article de Kaarik et Zegulova et est égal à :

$$\hat{q}_{GPD,n}(p) = \hat{\mu}_n + \frac{\hat{\sigma}_n}{\hat{k}_n} \left[\left(\frac{n \cdot p}{m_n} \right)^{-\hat{k}} - 1 \right] \quad (4.49)$$

Le choix de u_n permet de calibrer les paramètres σ_n et k : le tout permet de calculer l'estimation du quantile de la loi GPD pour le niveau de probabilité p .

Le modèle CIR++

Le modèle CIR++ étant une extension du modèle de CIR, nous allons tout d'abord présenter ce dernier.

Le modèle de CIR, établi en 1985, porte le nom des trois personnes à son initiative : Cox, Ingersoll et Ross. Sous ce modèle et sous la probabilité risque neutre Q , le taux d'intérêt court est modélisé par l'équation différentielle stochastique suivante :

$$\begin{cases} dx_t = k[\theta - x_t]dt + \sigma\sqrt{x_t}dW_t \\ x(0) = x_0 \end{cases} \quad (4.50)$$

avec k, θ , et σ des constantes positives et où,

- k est la vitesse de retour à la moyenne du taux,
- θ représente la moyenne de long terme du taux court,
- σ représente la volatilité du taux de court terme,
- W_t est un mouvement brownien.

Comme le modèle de Vasicek, le modèle de CIR permet un retour à la moyenne. Cependant, on remarque que ce qui le différencie du modèle de Vasicek est le terme $\sqrt{x_t}$. Ce terme permet ainsi d'interdire au taux d'intérêt court de prendre des valeurs négatives.

Notons que la simulation des incréments (la discrétisation) peut être faite de manière exacte ou bien, il est possible de recourir à des approximations.

Dans le cas exact, les incréments du taux d'intérêt court suivent une loi du chi-deux décentrée à d de degré de liberté, c'est à dire que pour tout s tel que $0 < s < t$, $r(t)$ est défini par :

$$x(t) = \frac{\sigma^2(1 - e^{-k(t-s)})}{4k} \chi_d^2 \left(\frac{4ke^{-k(t-s)}}{\sigma^2(1 - e^{-k(t-s)})} x(s) \right) \quad (4.51)$$

où, $d = \frac{4\theta k}{\sigma^2}$ est le nombre de degré de liberté de la loi du chi-deux.

De cette formule, nous pouvons en déduire l'équation donnée ci-dessous, utilisée pour déterminer par récursion les valeurs du taux d'intérêt court à chaque date :

$$x_{t+h} = \frac{\sigma^2(1 - e^{-kh})}{4k} \chi_d^2 \left(\frac{4ke^{-kh}}{\sigma^2(1 - e^{-kh})} x_t \right) \quad (4.52)$$

Par la suite, Brigo et Mercurio donnent une extension du modèle de CIR et il définissent ainsi, sous la probabilité risque neutre Q , le taux d'intérêt court instantané comme la somme d'une fonction déterministe $\phi^{CIR}(t, \alpha)$ et d'un processus

de Cox-Ingersoll-Ross (CIR) de paramètres θ, σ, k et x_0 :

$$\begin{cases} r(t) = x(t) + \phi^{CIR}(t, \alpha) \\ dx(t) = k[\theta - x(t)]dt + \sigma\sqrt{x(t)}dW(t) \\ x(0) = x_0 \end{cases} \quad (4.53)$$

avec,

$$\phi^{CIR}(t, \alpha) = f^M(0, t) - f^{CIR}(0, t, \alpha),$$

$$f^{CIR}(0, t, \alpha) = \frac{2k\theta(e^{th} - 1)}{2h + (k + h)(e^{th} - 1)} + x_0 \frac{4h^2 e^{th}}{[2h + (k + h)(e^{th} - 1)]^2},$$

$$\text{et, } h = \sqrt{k^2 + 2\sigma^2}$$

Notons $f^M(0, t)$ le taux forward instantané et $P^M(0, T)$, le prix de l'obligation zérocoupon de maturité T, observé sur le marché à la date 0.

Le prix du zéro-coupon de maturité T à la date t (noté P(t, T)) est donné par :

$$P(t, T) = \bar{A}(t, T)e^{-B(t, T)r(t)} \quad (4.54)$$

où,

$$\begin{cases} \bar{A}(t, T) = \frac{P^M(0, T)A(0, t)e^{-B(0, t)x_0}}{P^M(0, t)A(0, T)e^{-B(0, T)x_0}}A(t, T)e^{B(t, T)}\phi^{CIR}(t, \alpha), \\ A(t, T) = \left(\frac{2he^{(k+h)(T-t)/2}}{2h + (k + h)(e^{(T-t)/h} - 1)} \right)^{(2k\theta)/\sigma^2}, \\ B(t, T) = \left(\frac{2(e^{(T-t)/h} - 1)}{2h + (k + h)(e^{(T-t)/h} - 1)} \right) \end{cases} \quad (4.55)$$

Table des figures

2.1	Triangle de paiements décumulés (Dommages Auto)	17
2.2	Triangle de paiements cumulés	17
2.3	Triangle de paiements cumulés (Dommages Auto)	18
2.4	Complétion du triangle de paiement (Dommages Auto)	20
2.5	Graphiques des couples d'origine	22
2.6	Graphiques des couples du triangle de développement	23
2.7	Représentation graphique des résidus standardisés	36
2.8	L'estimation de l'écart-type proposée par Mack	36
2.9	le graphique QQ-plot sur les résidus	40
2.10	Le test d'indépendance et de normalité des résidus	40
2.11	le graphique résidu en fonction des années de survenance	40
2.12	le graphique résidu en fonction des années de développement	41
2.13	le graphique résidu en fonction des valeurs prédites	41
3.1	Graphiques des provisions par modèle log poisson pour la branche Auto	54
3.2	Graphiques des provisions par modèle log gamma pour la branche Auto	54
4.1	Etudes de la distribution de la charge de sinistres	67
4.2	Graphique de l'estimateur de Hill en fonction du seuil	68
4.3	Le graphique du Pareto quantile plot	68
4.4	Le graphique du Pareto quantile plot	69
4.5	Tests de Kolmogorov-Smirnov, selon le choix du seuil	70
4.6	Ratios S/P en fonction des années de survenance	70
4.7	Analyses de l'adéquation de la loi log-normale	71
4.8	Nombre de sinistres graves par année de survenance	71
4.9	Représentation du nombre de sinistres graves en fonction des primes acquises	72
4.10	Résultats du bootstrap GLM à un an	73

Liste des tableaux

1.1	Les coefficients primes non acquises par trimestre (1 an)	11
1.2	Les coefficients primes non acquises par trimestre (2 an)	11
1.3	Les coefficients primes non acquises par mois (1 an)	12
1.4	Les coefficients primes non acquises par mois (m an)	12
1.5	Exemple de triangle de liquidation	15
2.1	Coefficients de passage (Dommages Auto)	19
2.2	Tableau des provisions pour l'ensemble des branches	21
2.3	Coefficients de la méthode London-Chain	25
2.4	Loss Ratio ultimes Bornhuetter-Ferguson du triangle Dom.Auto . .	28
2.5	Provisions Dom.Auto estimées par la méthode de Bornhuetter-Furgeson	30
2.6	Loss Ratio ultimes Cape Cod du triangle Dom.Auto	33
2.7	Mean squared error of prediction	37
2.8	Tableau des provisions, erreurs relatives et paramètres de dispersion	39
2.9	Tableau des provisions, erreurs relatives et paramètres de dispersion	46
2.10	Distribution des résidus de Pearson	47
2.11	Valeurs de la statistique $-2q^+$	48
2.12	Tableau récapitulatif des provisions et des erreurs de prédiction associées	48
2.13	Tableau récapitulatif des Best Estimates	48
3.1	Tableau récapitulatif des résultats donnés par le bootstrap	55
3.2	VaR à 95% et VaR à 99,5% des provisions	55
4.1	Matrice de corrélation entre les branches	62
4.2	Résultats des SCR donnés par la formule standard avec différents paramètres	63
4.3	Résultats des SCR donnés par la loi log-normale	64
4.4	Résultats du bootstrap GLM à un an	66
4.5	Cadences de paiements et actualisation	74
4.6	Cadences de paiements et actualisation (DAB)	75
4.7	Comparaison des SCR (en M\$)	75