



INSTITUT DE SCIENCE FINANCIERE ET D'ASSURANCES (ISFA)
UNIVERSITE DE LYON 1



INSTITUT INTERNATIONAL DES ASSURANCES (IIA)
YAOUNDE

PROVISIONNEMENT NON VIE ET RISQUE DE MODELE

Mémoire présenté pour l'obtention des diplômes de
Master Professionnelle Actuariat de l'IIA et Master Actuariat de l'ISFA

Par

MENDZANA Benoit

Sous

L'encadrement :

Professionnel de :

DONGMO KENFACK Landry
Actuaire, AXA Cameroun

Académique du :

Dr TCHAPDA DJAMEN Idriss
Actuaire, BNP Paribas

Août 2021

DEDICACE

A
MES PARENTS

REMERCIEMENTS

Je commence par rendre grâce à Dieu pour avoir rendu ce travail possible.

Je dis merci à l'administration et au personnel enseignant de l'Institut International des Assurances (IIA) et de l'Institut des Sciences Financières et d'Assurances (ISFA) pour leur suivi académique.

Je suis reconnaissant au Docteur TCHAPDA DJAMEN Idriss, mon encadreur académique, qui a bien voulu m'apporter son expertise scientifique.

J'exprime ma profonde gratitude à Monsieur KEPEDEN Thierry, Directeur Général d'AXA Cameroun pour m'avoir donné l'opportunité d'effectuer mon stage professionnel dans la structure qu'il dirige ; à Monsieur DONGMO Landry, Actuaire à AXA Cameroun pour sa disponibilité, ses conseils et son encadrement professionnel ; à tout le personnel de la Direction Technique d'AXA Cameroun pour leurs apports multiformes.

Enfin, je remercie du fond du cœur ma tante MBALLA Stéphanie, mes sœurs MBIA Bibiane Francine et NGONO Marie Cornellie, mes frères NGAH Jean-Marie et NYASSA OHANDJA Hervé, Monsieur SANAMA Celin d'Aloys et Monsieur KAMGUIA Eric pour leurs encouragements, leurs soutiens moral et financier.

Sommaire

Dédicace	i
REMERCIEMENTS	ii
Liste des abréviations	iv
Liste des figures	v
Liste des tableaux	vi
Résumé	viii
Abstract	ix
Introduction générale.....	1
Partie I : Évaluation des provisions pour sinistres à payer	3
Chapitre 1 : Cadre de l'Étude et Notations	4
I- Le code CIMA et le provisionnement IARD	4
II- Généralités sur la gestion des sinistres	5
III- Structure d'Accueil.....	8
Chapitre 2 : Méthodes classiques de provisionnement	11
I- Méthodes déterministes	11
II- Méthodes stochastiques	26
Chapitre 3 : Analyse des Résultats	55
I- Choix de la méthode.....	55
II- Comparaison des méthodes, analyse des résultats et perspectives	55
Partie II : LE RISQUE DE MODELE	60
Chapitre 4 : Description du risque de modèle.....	62
I- Sources de risque de modèle	62
II- Quantification de l'erreur de modèle.....	68
Chapitre 5 : Evaluation de l'erreur de modèle en provisionnement non vie.....	78
I- Evaluation de l'erreur de modèle	78
II- Quantification de l'erreur de modèle.....	78
Conclusion et perspectives générales	82
Bibliographie.....	84
ANNEXES.....	I

LISTE DES ABREVIATIONS

CIMA : Conférence Interafricaine des Marchés d'Assurance

IBNeR : Incurred But Not enough Reported

IBNyR : Incurred But Not yet Reported

IBNR : Incurred But Not Reported

PSAP : Provisions Pour Sinistres à Payer

LISTE DES FIGURES

Figure 1 : Répartition des primes émises en 2018 par branche d'assurance en zone CIMA.....	4
Figure 2 : Cycle de vie d'un sinistre.....	5
Figure 3 : Décomposition de la charge ultime d'un sinistre.....	7
Figure 4 : Triangle de liquidation.....	8
Figure 5 : Répartition des primes émises par branche au deuxième trimestre 2021 chez AXA Cameroun.....	9
Figure 6: CC-Plots auto corporelle.....	17
Figure 7: Test d'indépendance des facteurs (j=0 à 3) auto corporelle.....	18
Figure 8: Distribution des résidus de Pearson en fonction des règlements estimés, des années de survéance, des années calendaires et des années de développement de la branche auto corporelle.....	31
Figure 9 : Evolution des $\ln(f_j - 1)$ et prédiction par un modèle linéaire.....	34
Figure 10 : Les résidus de Pearson normalisés du modèle log-normale.....	38
Figure 11 : Q-Q plot des Résidus du modèle Log-Normale vs Distribution normale.....	39
Figure 12 : Représentation des résidus de Pearson GLM Poisson.....	45
Figure 13 : Q-Q plot des Résidus GLM Poisson vs Distribution normale.....	45
Figure 14 : Histogramme et densité des résidus avec l'ajustement Gaussien en pointillés (auto corporelle).....	52
Figure 15 : Fonctions de répartition avec l'ajustement Gaussien en pointillés (auto corporelle).....	52
Figure 16 : Représentation de la distribution des provisions (auto corporelle).....	54
Figure 17 : Cadences de règlements Auto matérielle vs Auto corporelle.....	56
Figure 18 : Analyse comparée des cadences des règlements en Santé, Incendie et Auto corporelle.....	58
Figure 19 : Cartographie du risque de modèle.....	67
Figure 20_AI : CC-Plots auto matérielle.....	I
Figure 21_AI : Test d'indépendance des facteurs (j=0 à 3) auto matérielle.....	I
Figure 22_AI : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder.....	III
Figure 23_AI : Evolution des $\ln(f_j - 1)$ et prédiction par un modèle linéaire : branche auto matérielle.....	IV
Figure 24_AII : CC-Plots incendie.....	V
Figure 25_AII : Test d'indépendance des facteurs (j=0 à 3) incendie.....	V
Figure 26_AII : Validation des hypothèses de Mack en incendie : cadence de règlements presque régulière.....	VII
Figure 27_AII : Evolution des $\ln(f_j - 1)$ et prédiction par un modèle linéaire en incendie.....	VIII
Figure 28_AII : QQ-plot GLM quasi-poisson incendie.....	IX
Figure 29_AIII : CC-Plots santé.....	X
Figure 30_AIII : Test d'indépendance des facteurs (j=0 à 3) santé.....	X
Figure 31_AIII : Validation des hypothèses de Mack en santé : cadence de règlements presque régulière.....	XII
Figure 32_AIII : QQ-plot GLM quasi-poisson santé.....	XIII

LISTE DES TABLEAUX

Tableau 1 : Règlements cumulés de la branche auto corporelle	15
Tableau 2 : D-triangle de la branche auto corporelle et statistique descriptives.....	15
Tableau 3 : Triangle des rangs du D-triangle en auto corporelle	15
Tableau 4: Statistiques issues du triangle des rangs en auto corporelle.....	16
Tableau 5: Résultat du test d'absence d'effet calendaire en auto corporelle.....	16
Tableau 6: Développement du triangle inférieur auto corporelle par la méthode Chain Ladder	19
Tableau 7: Tableau récapitulatif des provisions auto corporelle par la méthode Chain Ladder	19
Tableau 8 : Illustration de la méthode de pondération « Min/Max »	21
Tableau 9 : Triangle complété par Chain Ladder pondérée « Min/Max »	22
Tableau 10 : Provision auto corporelle par la méthode Chain Ladder pondérée « Min/Max » ..	22
Tableau 11 : Test de significativité des paramètres additifs modèle London Chain.....	25
Tableau 12 : Triangle complété par la méthode London Chain	25
Tableau 13 : Provision auto corporelle par la méthode London Chain	25
Tableau 14 : Intervalles de confiance normal et log normal du modèle Mack branche auto corporelle.....	32
Tableau 15 : résultats du F-test sur les paramètres du modèle linéaire ajusté sur le logarithme des facteurs de transition	34
Tableau 16 : Provision auto corporelle après application du tail factor	35
Tableau 17 : Triangle des incréments auto corporelle.....	37
Tableau 18 : Résultat du test de Fisher	37
Tableau 19 : Les résidus de Pearson normalisés du modèle log-normale	37
Tableau 20 : Résultats du test de Shapiro-Wild	38
Tableau 21 : Paramètres estimés du modèle Log-Normale.....	39
Tableau 22 : Provisions par la méthode log-Normale.....	40
Tableau 23 : Résultats de l'estimation des paramètres du modèle GLM Poisson.....	43
Tableau 24 : Résidus du modèle GLM Poisson sur-dispersé.....	44
Tableau 25 : Résultat du test de Shapiro-Wild des résidus GLM Poisson.....	45
Tableau 26 : Provisions par la méthode GLM Poisson	47
Tableau 27 : Provision obtenue par distribution prédictive par convolution	49
Tableau 28 : Résidus de Pearson GLM Poisson	51
Tableau 29 : Provision obtenue par Bootstrap GLM Poisson sur-dispersé (auto corporelle).....	53
Tableau 30 : Réserves obtenues pour les sous branches automobile par différents modèles.....	56
Tableau 31 : Réserves obtenues pour les branches auto corporelle, incendie et santé par les différents modèles.....	57
Tableau 32 : Résultats des Best Estimate des provisions par modèle pour la $Var_{95\%}$	78
Tableau 33 : Résultats des mesures absolue et relative du risque de modèle pour la $Var_{95\%}$..	79
Tableau 34_AI : Développement du triangle inférieur auto matérielle par la méthode Chain Ladder	II
Tableau 35_AI : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder	II
Tableau 36_AI : Test de significativité des paramètres additifs modèle London Chain en auto matérielle.....	II
Tableau 37_AI : Développement du triangle inférieur auto matérielle par la méthode London-Chain	III
Tableau 38_AI : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder	III

Tableau 39_AI : Résultats du modèle de Mack en auto matérielle.....	IV
Tableau 40_AI : Provision auto matérielle après prise en compte du tail factor	IV
Tableau 41_AII : Développement du triangle inférieur incendie par la méthode Chain Ladder.	VI
Tableau 42_AII : Tableau récapitulatif des provisions en incendie par la méthode Chain Ladder	VI
Tableau 43_AII : Test de significativité des paramètres additifs modèle London Chain en incendie.....	VI
Tableau 44_AII : Développement du triangle inférieur en incendie par la méthode London-Chain	VII
Tableau 45_AII : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder	VII
Tableau 46_AII : Résultats du modèle de Mack en incendie	VIII
Tableau 47_AII : Provision incendie après prise en compte du tail factor	VIII
Tableau 48_AIII : Développement du triangle inférieur santé par la méthode Chain Ladder	XI
Tableau 49_AIII : Tableau récapitulatif des provisions santé par la méthode Chain Ladder	XI
Tableau 50_AIII : Test de significativité des paramètres additifs du modèle London Chain en santé.....	XI
Tableau 51_AIII : Développement du triangle inférieur santé par la méthode London Chain....	XI
Tableau 52_AIII : Tableau récapitulatif des provisions santé par la méthode London Chain....	XII
Tableau 53_AIII : Résultats du modèle de Mack en santé	XII
Tableau 54_AIII : Provision santé après prise en compte du tail factor.....	XIII
Tableau 55_AIII : Résultat du test de Shapiro-Wild des résidus GLM Poisson en santé	XIII
Tableau 56_AIII : Résultats du modèle GLM en santé	XIII

RESUME

Le risque est présent dans la quasi-totalité des activités humaines, dès qu'il y a un aléa, il y a nécessairement un risque. En effet, un certain nombre d'aléas peuvent venir grever significativement le patrimoine des agents économiques qui cherchent donc à diminuer leur risque. Le principe de l'assurance est de couvrir une partie des risques des agents dont le montant est fonction de la sinistralité future, donc aléatoire, en échange de prime dont le montant est fixé par avance. Le cycle de production d'un assureur est donc inversé. Ainsi, pour être à mesure de couvrir les sinistres futurs, non seulement l'assureur doit tarifier la prime au plus juste, il doit surtout constituer des réserves afin d'honorer tous les engagements pris envers les assurés. Ces réserves appelées provisions techniques occupent environ 80% du passif du bilan d'une compagnie d'assurance. Par conséquent, une bonne estimation des provisions est primordiale pour assurer la solvabilité et la pérennité d'une compagnie d'assurance. Notre étude porte premièrement sur l'estimation des Provisions pour Sinistres à Payer en assurance dommage à partir de différents modèles et de différentes branches d'assurance. L'estimation de ces provisions est une tâche plus ou moins compliquée, et repose sur des modèles souvent assez complexes dont l'application se fait parfois en supposant la validité des hypothèses des modèles considérés alors que les tests statistiques ne les confirment pas systématiquement. Le choix d'un modèle peut donc induire des erreurs dans l'évaluation de la provision et être à l'origine de risque de modèle. Ainsi, le deuxième objectif de notre étude est de quantifier l'erreur de modèle par une approche dite de « modèle de référence » en utilisant la mesure absolue du risque de modèle qui sera basée sur la *VaR*.

Mots-clés : Provisions techniques, Provisions pour Sinistres à Payer, Risque de modèle, Erreur de modèle, *VaR*.

ABSTRACT

Risk is present in almost all human activities, as soon as there is a hazard, there is necessarily a risk. Indeed, a certain number of contingencies can significantly burden the assets of economic agents who therefore seek to reduce their risk. The principle of insurance is to cover part of the agents' risks, the amount of which depends on future claims, and therefore random, in exchange for a premium, the amount of which is fixed in advance. The production cycle of an insurer is therefore reversed. Thus, to be able to cover future claims, not only must the insurer price the premium as accurately as possible, it must above all build up reserves in order to honor all the commitments made to policyholders. These reserves called technical provisions occupy about 80% of the liabilities of an insurance company's balance sheet. Therefore, a good estimate of provisions is essential to ensure the solvency and sustainability of an insurance company. Our study focuses first on estimating the Provisions for Claims Payable in property and casualty insurance based on different models and different branches of insurance. The estimation of these provisions is a more or less complicated task, and is based on models which are often quite complex, the application of which is sometimes done by assuming the validity of the assumptions of the models considered whereas the statistical tests do not systematically confirm them. The choice of a model can therefore lead to errors in the valuation of the provision and be the source of model risk. Thus, the second objective of our study is to quantify the model error by a so-called "reference model" approach using the absolute measure of the model risk which will be based on the VaR.

Keywords: Technical provisions, Provisions for Claims Payable, Model risk, Error of model, Var.

INTRODUCTION GENERALE

L'assureur non vie ne connaît pas à priori le coût futur des garanties offertes aux assurés, compte-tenu de l'inversion du cycle de production en assurance. Il faut par conséquent qu'il estime le montant des sinistres dont il aura la charge et effectue les provisions nécessaires afin de respecter ses engagements. Les estimations de l'assureur devraient être les plus justes possibles puisqu'elles permettent à la compagnie d'assurance de garantir sa solvabilité. En effet, une surévaluation des provisions augmente le niveau de capital requis et sa durée de détention dans la compagnie tandis qu'une sous-évaluation des provisions peut mettre l'assureur dans l'incapacité d'honorer ses engagements auprès des assurés. Les provisions techniques sont donc définies comme étant la meilleure estimation possible des flux futurs de l'assureur (Best Estimate) et d'une marge de risque. Cette optique de détermination des provisions techniques est une partie intégrante de la réglementation des compagnies d'assurance en Europe (Solvabilité II).

Que ces évaluations soient faites à partir de modèles statistiques purs ou en y intégrant des jugements d'expert, ces estimations seront toujours basées sur l'historique de données que l'assureur possède ou sur leur projection. La précision des modèles et les mesures de risque utilisées sont donc des éléments essentiels pour le calcul de la solvabilité de toute compagnie d'assurance. Les décisions prises sur la base de ces modèles assurantiels pour le calcul des réserves dépendent essentiellement des résultats statistiques des modèles calibrés et des jugements d'experts. Souvent les données disponibles ne permettent pas de valider les hypothèses des modèles et d'estimer de manière précise leurs paramètres. Les jugements alors basés sur ces modèles sont donc biaisés et peuvent ainsi induire des erreurs dans le choix du modèle approprié pour la détermination précise des réserves.

Le cadre de ce mémoire se situe en assurance non-vie et utilise des données de branches d'activité d'AXA Cameroun. Le provisionnement des sinistres chez AXA Cameroun est fait par le service de l'actuariat. C'est un processus qui se veut complexe car il se fait en gérant d'une façon particulière les sinistres graves et en émettant certaines hypothèses de travail propre à chaque branche et à sa gestion. Il se fait trimestriellement, impliquant une quantité importante de données. Même si elle est parfois ajustée aux branches d'activité et à certaines spécificités, la méthode utilisée est un modèle déterministe classique, le Chain-Ladder. La présente étude

visé à évaluer la pertinence de cette méthode, ainsi que son efficacité. L'objet de ce mémoire est donc d'évaluer les provisions pour sinistres à payer et de proposer une estimation de l'erreur due à la mauvaise spécification du modèle, erreur connue dans la littérature sous le nom de risque de modèle. L'objectif principal est de trouver une approche permettant d'évaluer le risque de modèle dans une classe de modèles en adéquations avec les données d'étude issu d'une compagnie d'assurance non vie. Sans pour autant tester une liste exhaustive de méthodes développées par la communauté actuarielle, le but est également de s'intéresser à des méthodes qui pourraient être particulièrement adaptées au cadre des branches d'activité d'AXA Cameroun.

Dans nos travaux, nous allons tenter de relever le double défi que constitue l'évaluation des provisions pour sinistres à payer et la quantification de l'erreur lié au choix du modèle utilisé pour cette évaluation.

En partant du postulat que le modèle de Chain Ladder et ses extensions sont les plus utilisés par les assureurs (pour leur facilité de compréhension, de calcul et d'adaptabilité aux données), le choix est fait de présenter des méthodes d'évaluation des provisions basées sur ces modèles dans le cadre d'une approche de triangles de liquidation. Ainsi, après avoir rappelé leur cadre théorique, nous implémentons à l'aide des logiciels Excel et R les méthodes déterministes du Chain-Ladder, Chain-Ladder pondérées, London-Chain ; le modèle stochastique récursif de Mack ; les modèles stochastiques factoriels log-normal et GLM Poisson et enfin les méthodes stochastiques de distributions prédictives par convolution et par Bootstrap.

Nous portons par ailleurs tout au long de la première partie de ce mémoire une attention particulière à étudier les hypothèses, intérêts et limites théoriques de chacune des méthodes présentées et à les confronter à nos données lors d'applications numériques.

Et dans la seconde partie nous présentons et définissons le risque de modèle ainsi que la cartographie des risques sous-jacents, nous exposons les approches d'évaluation de l'erreur de modèle et les mesures de risques associées et enfin nous appliquons la méthodologie mise en place pour l'estimation de la charge ultime d'un modèle de provisionnement non-vie.

PARTIE I : ÉVALUATION DES PROVISIONS POUR SINISTRES A PAYER

Chapitre 1 : Cadre de l'Etude et Notations

Introduction

Ce chapitre introductif permet de définir les notions utiles qui seront utilisées tout au long de ce mémoire. Nous définissons la provision pour sinistre à payer (PSAP) et ce qu'elle couvre ainsi que sa méthode générale d'estimation consistant à projeter la charge totale à l'ultime de l'assureur. Nous rappelons la représentation usuelle sous forme d'un triangle des données chez les assureurs non-vie pour le calcul statistique des provisions (appelé triangle de liquidation) et enfin nous présentons brièvement notre structure d'accueil et l'organisation de ses périmètres.

I- Le code CIMA et le provisionnement IARD

La Conférence Interafricaine des marchés d'assurance encadre les activités d'assurance de quatorze pays : le Bénin, le Burkina-Faso, le Cameroun, la Centrafrique, les Iles Comores, le Congo, la Cote d'Ivoire, le Gabon, la Guinée Equatoriale, le Mali, le Niger, le Sénégal, le Tchad et le Togo. Les activités de la zone sont marquées par une forte prédominance de l'assurance dommage qui occupe la majeure partie des primes émises. (66%) contre 34% pour l'assurance vie en 2018. (KOUADIO E, 2020).

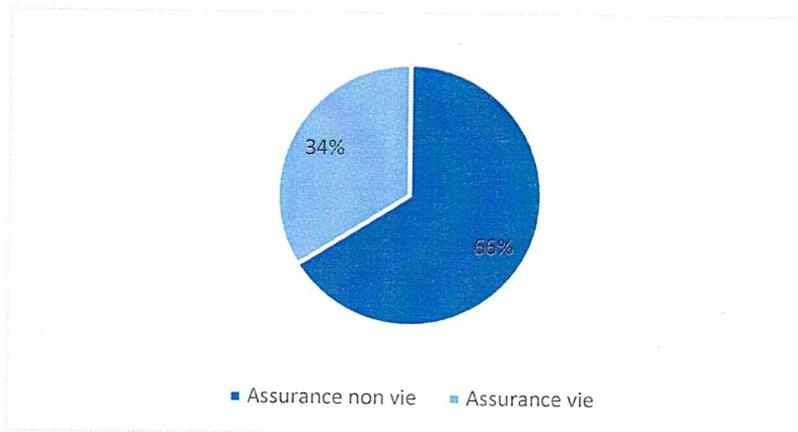


Figure 1 : Répartition des primes émises en 2018 par branche d'assurance en zone CIMA

L'Article 334 du code CIMA préconise quatre engagements réglementés que les compagnies d'assurance doivent respecter. Parmi ces engagements figurent les provisions techniques. Celles-ci doivent être suffisantes pour le règlement de tous les sinistres. Les provisions pour sinistres à payer (PSAP) occupent la majorité des provisions techniques en assurance non vie. En référence à l'article 334-8 (3) du code CIMA, cette provision correspond à la valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au

règlement de tous les sinistres survenus et non payés, y compris les capitaux consécutifs des rentes non encore mises à la charge de l'entreprise.

L'article 334-12 du code CIMA renseigne sur les modalités de calcul de la PSAP.

Celle-ci doit être calculée exercice par exercice de la manière suivante :

Sans préjudice de l'application des règles spécifiques à certaines branches prévues à la présente section, l'évaluation des sinistres connus est effectuée dossier par dossier, le coût d'un dossier comprenant toutes les charges externes individualisables ; elle est augmentée d'une estimation du coût des sinistres survenus mais non déclarés. Les modalités d'estimation du coût des sinistres survenus mais non déclarés ou sinistres déclarés tardifs sont fixées par circulaire de la Commission de Contrôle des Assurances.

La provision pour sinistres à payer doit toujours être calculée pour son montant brut, sans tenir compte des recours à exercer ; les recours à recevoir font l'objet d'une évaluation distincte.

Par dérogation aux dispositions du deuxième alinéa du présent article, l'entreprise peut, avec l'accord de la Commission de Contrôle des Assurances, utiliser des méthodes statistiques pour l'estimation des sinistres survenus au cours des deux derniers exercices.

II- Généralités sur la gestion des sinistres

1- Cycle de vie d'un sinistre

Le cycle de vie d'un sinistre est très important pour l'estimation de ses provisions. En effet, des événements comme la date de survenance ou la date de déclaration vont totalement conditionner les paramètres pour la modélisation des lois utilisées pour le calcul des réserves.

On peut résumer le cycle de vie d'un sinistre de la manière suivante.

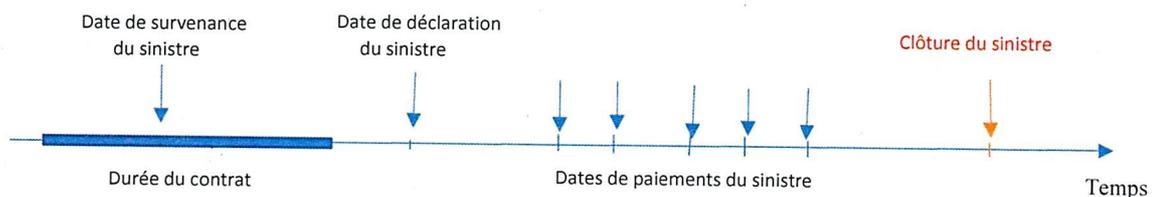


Figure 2 : Cycle de vie d'un sinistre

Il est indispensable de prendre en compte le délai entre la date de survenance du sinistre et sa date de déclaration. Il peut s'écouler plusieurs années entre les deux.

Par ailleurs, une fois qu'un sinistre est déclaré, il peut se passer également plusieurs années avant que le sinistre soit entièrement réglé ; comme dans les situations de dommage corporel et

de responsabilité civile, lorsqu'un ou plusieurs procès sont nécessaires pour statuer précisément sur la responsabilité des acteurs impliqués. Sans oublier le fait qu'un sinistre fermé peut aussi être réouvert par la suite, dans le cas, par exemple, où de nouveaux éléments apparaîtraient, de nouvelles pièces seraient ajoutées au dossier lors de la procédure judiciaire pour déterminer les différentes responsabilités (pour les garanties liées aux responsabilités civiles par exemple). Cet aspect spécifique, les réouvertures, ainsi que les provisions pour recours que l'assureur peut également constituer dans certains cas, ne sont pas traités dans ce mémoire. L'assureur doit donc faire face à plusieurs inconnues vis-à-vis des sinistres qu'il doit gérer, notamment leur délai de règlement et leurs montants finaux. Il va alors devoir utiliser ou concevoir des méthodes pour provisionner suffisamment en prévision de ces sinistres en adaptant ces méthodes spécifiquement à ses branches d'activité.

2- Charge ultime

L'assureur doit, du fait du cycle inversé de production de son activité, constituer des provisions techniques afin de pouvoir payer, entre autres, plusieurs mois ou années après, les sinistres dont il n'a pas complètement, ou pas du tout connaissance, au moment de l'arrêté comptable qu'il effectue régulièrement.

Au moment d'un arrêté, la charge ultime d'un sinistre correspond à la somme du montant payé pour ce sinistre et des provisions techniques constituées.

Ces provisions se répartissent de la manière suivante :

a) Provisions dossier/dossier (D/D). Ces provisions sont estimées au cas par cas par l'équipe de gestionnaires des sinistres qui est spécialisée par branche d'activité et qui est experte en ce qui concerne les différents cas de figure qui peuvent survenir au cours du cycle de vie du sinistre.

b) Provisions IBNR (Incurred But Not Reported) : elles-mêmes subdivisées en :

i) Provisions IBNeR (Incurred But Not enough Reported) : provisions visant à compléter (positivement ou négativement) les provisions dossier/dossier ;

ii) Provisions IBNyR (Incurred But Not yet Reported) : provisions servant à couvrir les sinistres survenus mais non encore déclarés à l'assureur.

On appelle généralement PSAP, pour Provisions pour Sinistres À Payer, la somme de toutes ces provisions techniques.

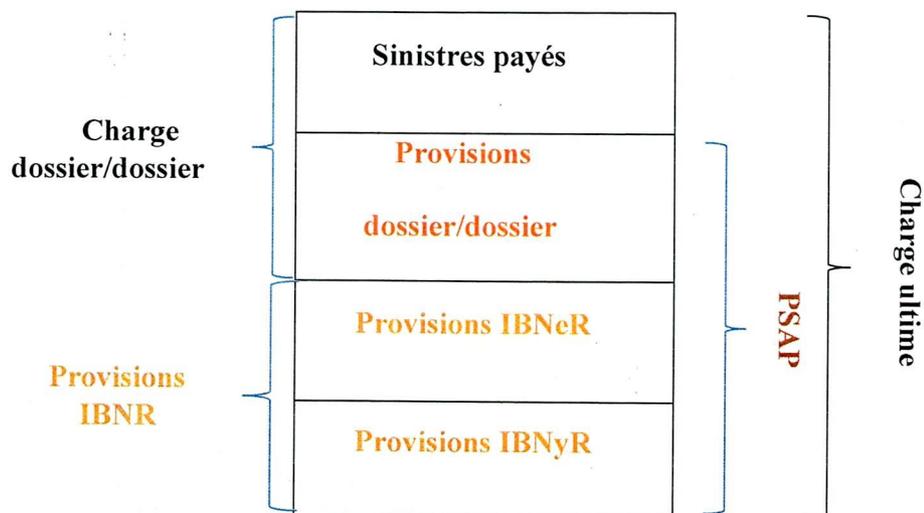


Figure 3 : Décomposition de la charge ultime d'un sinistre

3- Triangles de liquidation

Un grand nombre de méthodes actuarielles d'estimation des provisions de sinistres se basent sur des triangles agrégés de montants cumulés de sinistres. Ils sont bien souvent le point de départ, la première étape indispensable, pour pouvoir appliquer ces méthodes. On comprend ainsi aisément qu'il est primordial de pouvoir constituer ces triangles de données sensibles en s'assurant d'une certaine qualité, d'un certain contrôle, en maîtrisant entre autres, leur origine et la manière dont ils sont constitués.

Il existe plusieurs manières de présenter ces triangles. Dans ce mémoire, nous utilisons la représentation et les notations ci-après :

- Les lignes i correspondent aux années de survenance du sinistre ;
- Les colonnes j représentent les années de développement, c'est-à-dire le délai entre la survenance du sinistre et son règlement ;
- Les diagonales correspondent aux années calendaires et contiennent les informations relatives aux années comptables $i + j$;
- $X_{i,j}$ représente l'information relative aux règlements effectués pour l'année de survenance i et après j années.

Le triangle de liquidation se présente donc sous la forme suivante :

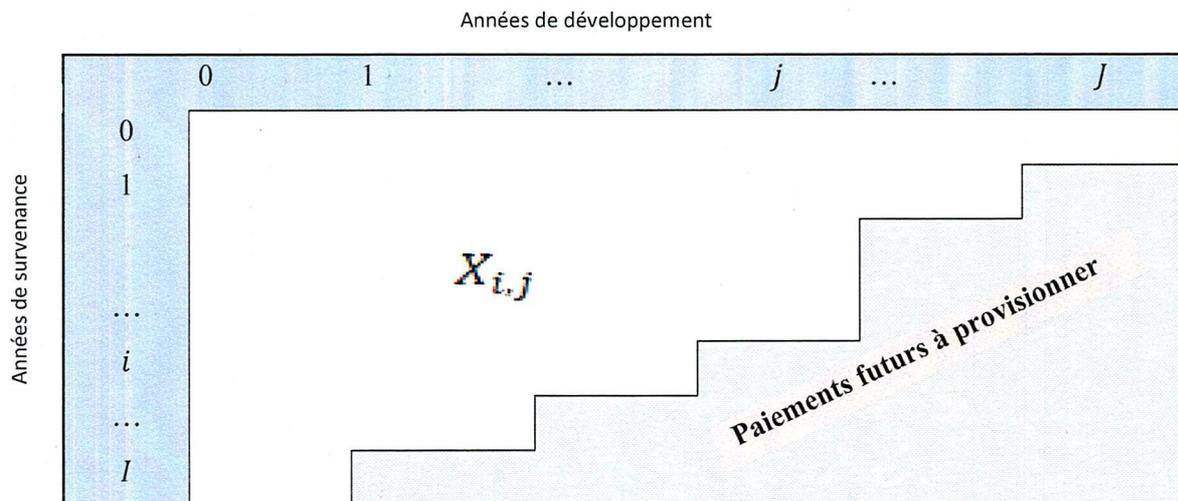


Figure 4 : Triangle de liquidation

4- Best Estimate des provisions

Le Best Estimate des provisions à l'instant t est défini comme la somme des règlements futurs actualisés et pondérés par leurs probabilités d'occurrence. Il correspond donc à l'espérance conditionnelle des règlements futurs actualisés compte-tenu de l'information disponible F_t .

Afin de ne pas interférer avec le risque de taux et pour ne pas conditionner nos résultats de modélisations futures à une hypothèse quelconque de taux d'intérêt, nous utiliserons tout au long de ces travaux une approche du Best Estimate pour laquelle les règlements futurs ne seront pas escomptés. Un tel Best Estimate est alors défini par :

$$R^t = \sum_{i=1}^I \sum_{j=I-i+1}^J E[C_{i,j} | F_t].$$

III- Structure d'Accueil

1- Présentation d'AXA Cameroun

La filiale du groupe AXA au Cameroun, commercialise une gamme variée de produits, communs à toutes les compagnies d'assurance et repartis par branche.

Ainsi, nous pouvons avoir : Individuelle Accident, Responsabilité Civile Générale, Santé, Transport, Incendie, Aviation, Dommages aux Biens, Assistance, Automobile, Crédit et Acceptation.

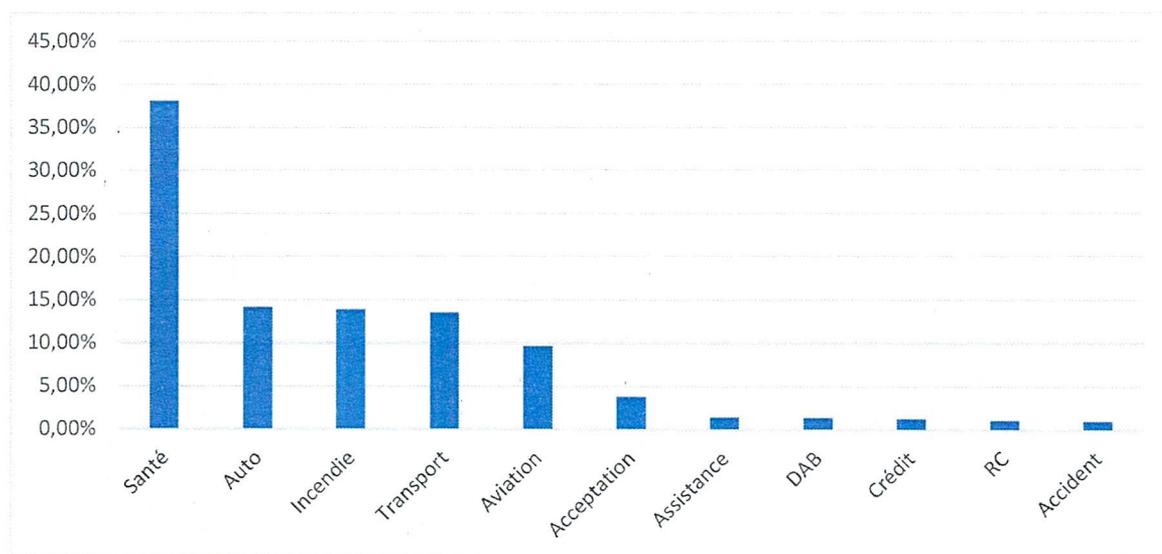


Figure 5 : Répartition des primes émises par branche au deuxième trimestre 2021 chez AXA Cameroun

2- Organisation des périmètres chez AXA Cameroun

En plus du découpage naturel par branche d'activité (Santé, Aviation, Responsabilité Civile, ...) et par entité gérant le contrat (Département Transport, ...), les informations sur les sinistres sont réparties également par :

a- Catégorie de sinistres

La distinction est faite en deux catégories : attritionnel ou grave. Le seuil des sinistres graves étant fixé à 10 millions de francs CFA, sachant que le montant d'un sinistre peut évoluer au cours du temps, il est tout à fait possible qu'entre sa date d'ouverture et sa date de fermeture, un sinistre change de catégorie. Ce découpage est nécessaire car la fréquence des sinistres graves est complètement éloignée de celle des sinistres attritionnels avec des montants en rien comparables. Les mélanger aurait pour conséquence de "polluer" les triangles de données avec des valeurs extrêmes qui ne permettraient pas de vérifier les hypothèses qu'il est nécessaire de respecter pour appliquer les méthodes traditionnelles de calcul de provisionnement de sinistres. Dans nos travaux, nous nous sommes intéressés uniquement aux méthodes de provisionnement pour calculer les réserves des sinistres attritionnels. Les sinistres graves sont traités d'une manière totalement différente.

b- Sous branches

Une branche d'activité peut être subdivisée en plusieurs sous-branches avec des modes de gestion et donc des délais de déclaration, de paiements très différents, d'où la nécessité de les traiter séparément pour conserver le côté homogène des données.

Par ailleurs, bien que les arrêtés comptables effectués chez AXA Cameroun soient trimestriels, les montants cumulés agrégés des sinistres dans les triangles sont construits avec des données annuelles. Tout comme pour le découpage en sous-branche, ce choix a un effet important sur les résultats et considérer des données trimestrielles, semestrielles ou annuelles par exemple, a une conséquence directe sur les résultats des méthodes ne serait-ce que parce que cela a un impact direct sur la volatilité des données.

Parmi toutes les branches d'activité d'AXA Cameroun, nous ne nous pencherons que sur certaines (Automobile, Incendie et Santé) dans le cadre de nos travaux pour tester les différentes méthodes de provisionnement abordées. Elles ont chacune leurs spécificités propres, aussi bien au niveau de la durée de leur déroulement, de leur système de gestion, que de leur législation.

Conclusion

Dans Ce chapitre la présentation des éléments utiles à l'estimation de la PSAP a été faite. Le prochain chapitre présente différentes méthodes proposant une estimation de la charge totale à l'ultime. Cette charge à l'ultime est directement à relier au montant à provisionner au titre de la PSAP.

Chapitre 2 : Méthodes classiques de provisionnement

Introduction

Nous allons présenter quelques méthodes de provisionnement, que l'on classe ici en deux catégories : les méthodes déterministes, et les méthodes stochastiques. Pour cela, nous travaillons sur un triangle de développement dont les lignes représentent les années d'origine notées de 0 à I , et les colonnes celles des années de développement notées de 0 à J . La valeur J correspond donc à la durée de développement, et I l'année civile actuelle (la construction se base sur l'hypothèse importante qui limite à J le nombre d'années de développement).

Définitions et notations

- $X_{i,j}$: le paiement effectué l'année j pour des sinistres survenus l'année i .
- $C_{i,j}$: le paiement cumulé, soit :

$$C_{i,j} = \sum_{k=0}^j X_{i,k}$$

- R_i : montant de réserves nécessaires pour financer les encours futurs des sinistres d'année d'origine i , et se définit comme suit :

$$R_i = C_{i,J} - C_{i,I-i}$$

- R : réserve globale, définie par :

$$\sum_{i=0}^I R_i$$

I- Méthodes déterministes

1- Chain-Ladder

a- Cadre théorique

Cette méthode de calcul de provision (que l'on appelle aussi "cadence de règlement") est relativement ancienne puisqu'on peut trouver la trace de ses premières apparitions dans des livres de droit des assurances des années 1930, comme par exemple (Astesan, 1938). C'est une méthode générique qui peut s'appliquer sur tous les types de triangles (de développement, de charges etc...).

L'idée ici est de considérer que l'évolution des encours cumulés d'une année de développement à une autre reste identique pour toutes les années de survenance. Cette évolution est caractérisée

par un facteur multiplicatif propre à chaque année de développement. Sa simplicité et son intuitivité font de ce modèle un modèle de référence en termes de provisionnement, et incitent à l'utiliser pour se faire une première idée des montants de réserve. Les autres modèles tendent d'ailleurs à s'aligner sur les résultats obtenus par Chain-Ladder, comme nous le verrons par la suite.

Pour appliquer cette méthode, il est nécessaire de disposer d'un portefeuille homogène et grand, sans événements extrêmes. Ces conditions sont indispensables pour obtenir des résultats satisfaisants. En effet, la présence d'un grand sinistre, dont le déroulement au cours du temps peut être différent des sinistres classiques (étalement, en général plus long), perturbera l'hypothèse d'égalité des facteurs de développement pour toutes les années d'origine i . Toutefois, un moyen de pallier à ce problème est d'effectuer un pré-traitement visant à ne retenir que les risques attritionnels (répétitifs, nombreux, de petite taille), et donc retirer les risques graves et extrêmes.

Hypothèses du modèle

Cette modélisation repose sur les hypothèses :

- les montants cumulés sont indépendants par année de survenance (test d'absence d'effet calendaire), soit : $\forall i \neq k, (C_{i,1}, \dots, C_{i,n})$ et $(C_{k,1}, \dots, C_{k,n})$ sont indépendants ;
- qu'il existe une relation de proportionnalité entre les montants cumulés, d'une année de développement à la suivante. De plus, il est supposé que le coefficient de proportionnalité d'une année de développement soit identique pour toutes les années de survenance.

C'est-à-dire :

$$\forall 0 \leq i \leq I-1, 0 \leq j \leq J-1, C_{i,j+1} = f_j \times C_{i,j}$$

Où f_j , est estimé avec l'historique de données par :

$$\hat{F}_j := \frac{\sum_{k=0}^{I-j} C_{k,j+1}}{\sum_{k=0}^{I-j} C_{k,j}}, \quad \forall j \geq I-i+1.$$

Il est alors possible d'estimer les montants des sinistres survenus en année i , après j années de développement, par la relation suivante :

$$\forall j \geq I-i, \hat{C}_{i,j} := C_{i,I-i} \prod_{k=I-i}^{j-1} \hat{F}_k$$

Ce qui permet de calculer les réserves pour chaque année de survenance et la réserve globale :

$$\forall i \in \{0, \dots, I\}, \hat{R}_i := \hat{C}_{i,J} - \hat{C}_{i,I-i} \quad \text{et} \quad \hat{R} := \sum_{i=0}^I \hat{R}_i$$

Test des hypothèses

Nous allons présenter les méthodes utilisées pour la vérification des hypothèses sous-jacentes du modèle de Chain Ladder. La qualité des résultats de ce modèle dépend du fait que les données du triangle respectent ces hypothèses.

Indépendance des années de survenance

Ce test proposé par Thomas Mack permet de tester l'indépendance des années de survenance qui peut être remise en cause dû à des effets dit « calendaires ». Ces effets peuvent intervenir par exemple à cause de changements dans la gestion des sinistres ou de l'inflation.

Nous calculons les facteurs de développement individuel $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ et nous définissons les éléments d'une diagonale D_j par :

$$D_j = \{C_{j,1}, C_{j-1,2}, \dots, C_{2,j-1}, C_{1,j}\}, \quad \forall 1 \leq j \leq I = J$$

Pour chaque colonne du triangle des facteurs de développement individuels, nous comparons ses éléments à la médiane de cette même colonne. Nous calculons ensuite pour chaque diagonale (excepté la première qui ne contient qu'un élément) le nombre de facteurs L_k appartenant au groupe des rangs élevés et ceux S_k appartenant au groupe des rangs bas. Dans le cas théorique d'absence d'effet calendaire, la probabilité d'être supérieure ou inférieure à la médiane est de 50 % et donc les valeurs de L_k et S_k doivent être très rapprochées.

Nous définissons la variable $Z_k = \min(S_k; L_k)$ qui théoriquement doit être proche de $\frac{S_k + L_k}{2}$.

Dans l'hypothèse d'absence d'effet calendaire, la variable Z_k suit une loi binomiale $\mathcal{B}(n_k, p)$ avec $n_k = S_k + L_k$ et $p = 0,05$. Thomas Mack a démontré que :

$$E(Z_k) = \frac{n_k}{2} - C_{n_k-1}^{m_k} \times \frac{n_k}{2^{n_k}} \quad \text{et} \quad V(Z_k) = \frac{n_k(n_k-1)}{4} - C_{n_k-1}^{m_k} \times \frac{n_k(n_k-1)}{2^k} + E(Z_k) - (E(Z_k))^2$$

où m_k est la partie entière de $\frac{n_k-1}{2}$

Nous définissons la variable aléatoire $Z = \sum_{k=2}^{n-1} Z_k$ tel que

$$E(Z) = \sum_{k=2}^{n-1} E(Z_k) \quad \text{et} \quad V(Z) = \sum_{k=2}^{n-1} V(Z_k)$$

Comme la somme de variables aléatoires binomiales converge vers une variable aléatoire normale, nous supposons que la variable Z suit une loi normale. Nous pouvons donc construire un intervalle de confiance à 95% permettant de tester notre hypothèse principale. L'intervalle de confiance est :

$$E(Z) - N^{-1}(97,5\%) \times V(Z) \leq Z \leq E(Z) + N^{-1}(97,5\%) \times V(Z)$$

Existence des facteurs de développement

Pour vérifier cette hypothèse, nous allons vérifier la linéarité de $C_{i,j+1}$ par rapport à $C_{i,j}$ pour chaque année de développement. L'hypothèse sera acceptée si ces points sont approximativement alignés sur une droite passant par l'origine. Nous effectuons alors une régression linéaire des montants de l'année de développement $j+1$ en fonction de l'année j . Si une droite (courbe de tendance) passant par l'origine du repère est obtenue, et que le coefficient de détermination est proche de 1 alors il y a une bonne adéquation au modèle de Chain Ladder.

Stabilité des règlements

Le d-triangle est constitué des facteurs de développement individuels. Pour tester la stabilité des règlements pour chaque année de développement, il suffit de vérifier que les coefficients de passage partiels ont une variabilité limitée et donc les montants de règlement sont stables. Nous allons tester cette hypothèse graphiquement en observant le comportement des facteurs de développement partiels autour de leur moyenne.

b- Cadre pratique

Dans cette application numérique, nous déterminons le montant des Best Estimate des provisions.

Nous confrontons cependant nos données aux hypothèses du modèle. Nous voulons donc dès à présent tester la bonne adéquation de nos données à ce modèle.

Présentation des données

Les données sont issues des bases d'AXA Cameroun. Elles correspondent aux règlements cumulés de la branche auto corporelle aux premiers semestres 2013 à 2021, puisque notre étude à lieu juste après le premier semestre 2021. Dans un souci de confidentialité, ces données ont subi certaines modifications (multiplication des valeurs par un facteur commun) qui n'impactent en aucune façon les conclusions tirées dans ces travaux.

Tableau 1 : Règlements cumulés de la branche auto corporelle

	0	1	2	3	4	5	6	7	8
2013	7 294	21 105	62 031	76 194	88 583	94 916	103 522	104 178	111 766
2014	5 552	29 118	57 697	104 696	123 713	136 049	150 906	160 270	
2015	2 234	24 011	39 433	55 122	69 065	70 027	80 136		
2016	5 425	23 930	41 152	72 870	84 763	98 889			
2017	3 271	22 771	111 495	121 248	152 664				
2018	1 260	24 698	68 142	90 501					
2019	562	10 316	47 397						
2020	4 734	88 028							
2021	7 879								

Validation du modèle

Indépendance des années de survenance

Test d'absence d'effet calendaire.

Pour notre triangle des paiements, les résultats du test sont les suivants :

Tableau 2 : D-triangle de la branche auto corporelle et statistique descriptives

	0	1	2	3	4	5	6	7
2013	2,8935	2,9392	1,2283	1,1626	1,0715	1,0907	1,0063	1,0728
2014	5,2444	1,9814	1,8146	1,1816	1,0997	1,1092	1,0621	
2015	10,7486	1,6423	1,3979	1,2529	1,0139	1,1444		
2016	4,4114	1,7197	1,7707	1,1632	1,1667			
2017	6,9613	4,8963	1,0875	1,2591				
2018	19,5991	2,7590	1,3281					
2019	18,3561	4,5946						
2020	18,5939							

Moyenne	10,8511	2,9332	1,4378	1,2039	1,0880	1,1147	1,0342
Ecart-type	7,0120	1,3347	0,2943	0,0482	0,0635	0,0273	0,0394
Coef-Var	64,62%	45,50%	20,47%	4,01%	5,83%	2,45%	3,81%

Tableau 3 : Triangle des rangs du D-triangle en auto corporelle

	0	1	2	3	4	5	6	7
2013	8	3	5	5	3	3	2	1
2014	6	5	1	3	2	2	1	
2015	4	7	3	2	4	1		
2016	7	6	2	4	1			
2017	5	1	6	1				
2018	1	4	4					
2019	3	2						
2020	2							

Tableau 4: Statistiques issues du triangle des rangs en auto corporelle

k	L_k	S_k	Z_k	n_k	m_k	$E(Z_k)$	$V(Z_k)$
2	1	1	1	2	0	0,5	0,25
3	1	2	1	3	1	0,75	0,1875
4	1	3	1	4	1	1,25	0,4375
5	1	3	1	4	1	1,25	0,4375
6	5	1	1	6	2	2,0625	0,62109375
7	1	4	1	5	2	1,5625	0,37109375
8	6	1	1	7	3	2,40625	0,55371094
Total			7			9,78125	2,85839844

L'hypothèse d'indépendance des années de survenance est testée comme décrit ci-dessus, en calculant les valeurs de Z , son espérance et sa variance :

$$Z = 7; E(Z) = 9,78125 \text{ et } V(Z) = 2,85840$$

Au seuil de 5%, $N^{-1}(97,5\%) = 1,96$

Donc : $E(Z) - N^{-1}(97,5\%) \times V(Z) = 4,178789$

et $E(Z) + N^{-1}(97,5\%) \times V(Z) = 15,383711$

La valeur obtenue pour la variable est : $z = 7$. Elle appartient bien à l'intervalle de confiance à 95% que nous avons déterminé. L'hypothèse est donc valide.

Tableau 5: Résultat du test d'absence d'effet calendaire en auto corporelle

Niveau de confiance	95%
Intervalle de confiance	[4,178789; 15,383711]
Résultat du test	Hypothèse non rejetée

Existence des facteurs de développement

Si l'hypothèse selon laquelle il existe une relation linéaire entre les paiements cumulés d'une année de déroulement à l'autre est vérifiée, les points

$$(C_{i,j}; C_{i,j+1}) \text{ pour } i = 0, 1, \dots, i - j$$

doivent être alignés sur une droite passant par l'origine.

Afin de vérifier cette hypothèse, nous avons tracé le CC-plot représentant l'ensemble de ces points pour chaque année de développement.

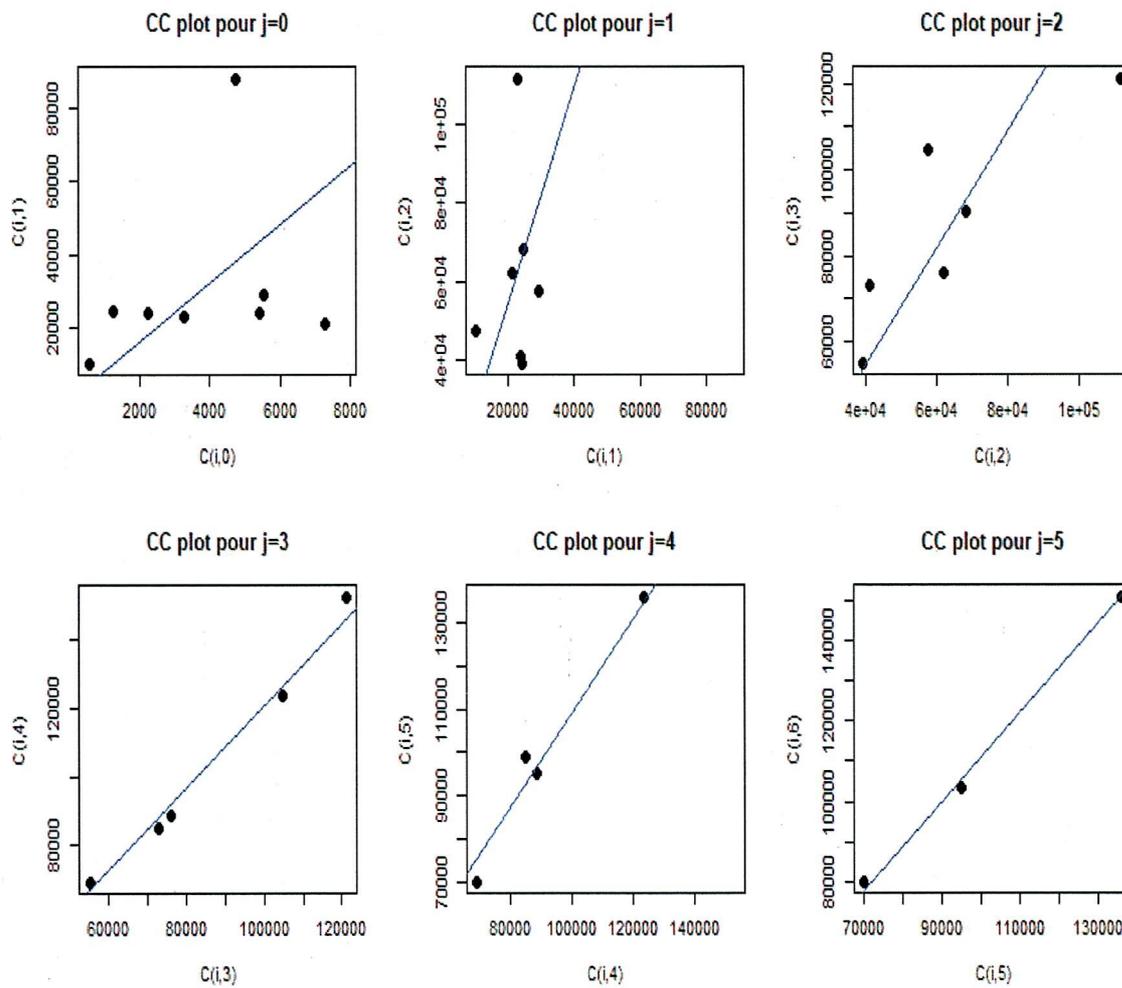


Figure 6: CC-Plots auto corporelle

Au regard de ces figures, l'hypothèse paraît raisonnable à partir de $j=2$ car les points semblent alignés sur des droites issues de l'origine pour chaque année de développement à partir de la troisième année. Pour les deux premières années les points sont un peu plus dispersés que pour les autres années.

Stabilité des règlements

Pour vérifier l'hypothèse, nous comparons graphiquement les facteurs de développement individuels à leur moyenne, les figures suivantes en présente l'exemple de cette comparaison pour les années de développement de 1 à 4.

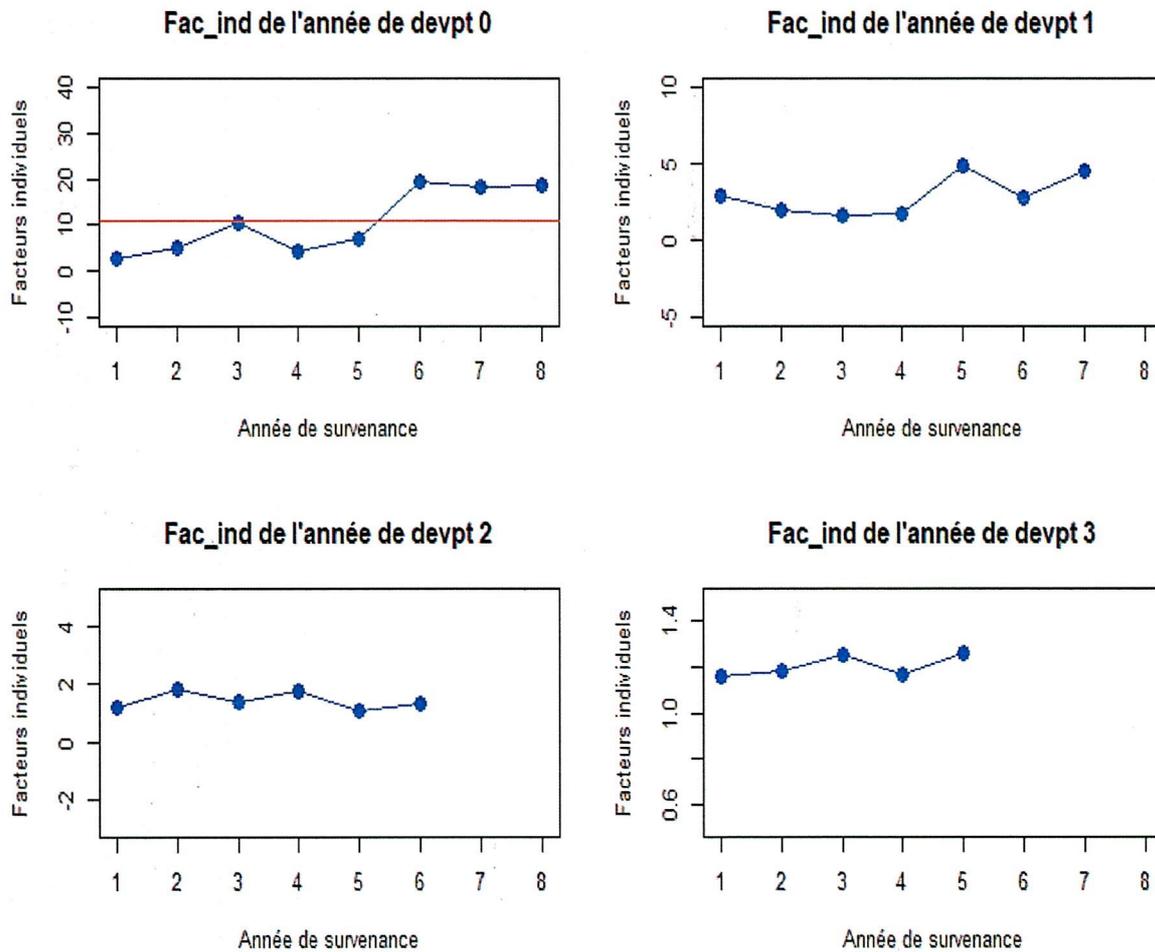


Figure 7: Test d'indépendance des facteurs ($j=0$ à 3) auto corporelle

Pour la première année de développement, l'hypothèse d'égalité des coefficients sur les années de survénance n'est pas vérifiée, les coefficients par année sont plutôt volatiles autour de la moyenne. En revanche, elle paraît plus vraisemblable à partir de la 2^{ème} années de développement. Il en est de même pour les autres années. Les facteurs individuels de Chain-Ladder sont proches des moyennes des facteurs évalués.

Nous considérons donc que les hypothèses du modèle de Chain-Ladder sont ici vérifiées, et nous pouvons appliquer le modèle.

Les facteurs de développements ainsi que le triangle projeté en application de la méthode Chain-Ladder sont donnés ci-dessous.

Tableau 6: Développement du triangle inférieur auto corporelle par la méthode Chain Ladder

	0	1	2	3	4	5	6	7	8
2013	7 294	21 105	62 031	76 194	88 583	94 916	103 522	104 178	111 766
2014	5 552	29 118	57 697	104 696	123 713	136 049	150 906	160 270	171 945
2015	2 234	24 011	39 433	55 122	69 065	70 027	80 136	83 292	89 359
2016	5 425	23 930	41 152	72 870	84 763	98 889	109 919	114 248	122 570
2017	3 271	22 771	111 495	121 248	152 664	166 740	185 337	192 636	206 668
2018	1 260	24 698	68 142	90 501	109 155	119 219	132 516	137 735	147 768
2019	562	10 316	47 397	64 946	78 333	85 555	95 098	98 843	106 043
2020	4 734	88 028	241 223	330 538	398 668	435 426	483 991	503 053	539 697
2021	7 879	63 372	173 660	237 959	287 007	313 470	348 432	362 155	388 535
fac_estimé	8,043589	2,740312	1,370257	1,206118	1,092204	1,111534	1,039385	1,072842	1

Tableau 7: Tableau récapitulatif des provisions auto corporelle par la méthode Chain Ladder

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	111 766	111 766	0	0,00%
2014	171 945	160 270	11 674	1,12%
2015	89 359	80 136	9 223	0,88%
2016	122 570	98 889	23 681	2,26%
2017	206 668	152 664	54 005	5,16%
2018	147 768	90 501	57 267	5,47%
2019	106 043	47 397	58 646	5,60%
2020	539 697	88 028	451 669	43,15%
2021	388 535	7 879	380 657	36,36%
Total	1 884 352	837 529	1 046 823	100%

Avec la méthode Chain-Ladder et en supposant que le déroulement des sinistres est terminé ($J \leq I$), à la fin du deuxième trimestre 2021, l'assureur doit provisionner à hauteur de

1 046 823 FCFA sa branche auto corporelle.

Notons que cette modélisation est faite sous l'hypothèse forte que les déroulements sont clos au bout de J années. Pour pallier à cette restriction, il est possible d'introduire un tail factor ou facteur de queue comme nous le verrons dans la suite.

Nous continuons de travailler sous cette hypothèse forte et pour simplifier nous notons

$$I = J = n.$$

c- Critiques de la méthode Chain-Ladder

Nous avons vu que cette méthode a l'avantage d'être aisée à appliquer. Cependant, elle repose sur des hypothèses qui peuvent ne pas s'avérer réalistes dans la pratique.

En effet, le schéma de développement, selon lequel le cumul d'une année de développement est proportionnel à celui de l'année précédente, doit être similaire pour toutes les années d'origine. Ainsi, ce schéma ne tient pas compte d'une évolution probable dans le déroulement des règlements des sinistres au cours du temps. Par exemple, en cas de changement de jurisprudence, il peut y avoir une variation notable au niveau des montants de paiement. Ou encore en cas de changement de management au niveau de la gestion des sinistres visant à accélérer ou ralentir les rythmes de remboursements. Pour résumer, cette méthode repose sur l'hypothèse d'une très grande stabilité au niveau des facteurs de développement de la sinistralité.

Variantes de Chain-Ladder

2- Chain-Ladder pondérée

a- Cadre théorique

La méthode Chain-Ladder pondérée est une généralisation de la méthode classique permettant un choix de facteurs beaucoup plus large. Dans ces travaux, inspirés de (Busson, 2012), nous nous intéressons à la version pondérée en reprenant le formalisme utilisé par (Mack, 1999). Les facteurs de développement sont définis de la manière suivante :

$$\forall j \in \{0, 1, \dots, n\}, \quad \hat{f}_j = \frac{\sum_{i=0}^{n-j-1} f_{i,j} \times \omega_{i,j} \times C_{i,j}^\alpha}{\sum_{i=0}^{n-j-1} \omega_{i,j} \times C_{i,j}^\alpha}, \quad \text{avec: } f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$$

La pondération $\omega_{i,j}$ est « judicieusement » choisie par l'expertise de l'actuaire en fonction de l'allure des **facteurs individuels**, de la fiabilité que l'on peut leur accorder, de leur volatilité et de considérations externes notamment prudentielle.

En ce sens, le poids $\omega_{i,j}$ représente la confiance en ce facteur $f_{i,j}$ que lui accorde l'actuaire souvent pris pour 0 ou 1.

Le paramètre α représente le type de pondération souhaité :

- $\alpha = 0$: Chain Ladder simple, c'est-à-dire les \hat{f}_j sont des moyennes arithmétiques des $f_{i,j}$;
- $\alpha = 1$: Chain Ladder classique, c'est-à-dire les $f_{i,j}$ sont pondérés par les volumes (autrement dit par paiements cumulés pour un triangle de liquidation de paiements cumulés) ;

- $\alpha = 2$: Chain Ladder vue comme une régression, c'est-à-dire les \hat{f}_j sont des estimateurs issus d'une minimisation de moindres carrés.

Les choix standards de pondération

Le choix des pondérations peut être guidé par plusieurs facteurs selon la nature des données. Lorsque l'on cherche à calculer la PSAP par les méthodes statistiques, on souhaite projeter la charge à l'ultime espérée (ou moyenne). Lors de cet exercice, l'actuaire souhaite donc écarter les sinistres qu'il reconnaît comme « anormaux ».

En ce sens, la pondération sert à réduire l'importance de certaines données jugées non représentatives :

- événement exceptionnel ;
- changement dans la gestion des sinistres ;
- etc...

Dans ces travaux, nous illustrons deux méthodes de pondération définies ci-dessous.

Pondération « expert » :

L'estimation des facteurs de développement est réalisée en prenant en compte seulement les facteurs de développement individuels choisis par l'expert.

Formellement le facteur estimé pour une année de développement j fixé s'écrit :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} f_{i,j} \times \omega_{i,j} \times C_{i,j}^\alpha}{\sum_{i=0}^{n-j-1} \omega_{i,j} \times C_{i,j}^\alpha}, \quad \text{avec: } \omega_{i,j} = 1_{\{\text{choix expert}\}}$$

Pondération « Min/Max » :

Les facteurs de développement individuels maximum et minimum par année de développement sont pondérés par 0. Cette méthode permet ainsi de censurer les facteurs de développement « extrêmes ».

Tableau 8 : Illustration de la méthode de pondération « Min/Max »

		Facteurs de développement individuels $f_{i,j}$								Poids $\omega_{i,j}$ associés aux facteurs $f_{i,j}$							
		0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7
2013		2,8935	2,9392	1,2283	1,1626	1,0715	1,0907	1,0063	1,0728	0	1	1	0	1	0	1	1
2014		5,2444	1,9814	1,8146	1,1816	1,0997	1,1092	1,0621		1	1	0	1	1	1	1	
2015		10,7486	1,6423	1,3979	1,2529	1,0139	1,1444			1	0	1	1	0	0		
2016		4,4114	1,7197	1,7707	1,1632	1,1667				1	1	1	1	0			
2017		6,9613	4,8963	1,0875	1,2591					1	0	0	0				
2018		19,5991	2,7590	1,3281						0	1	1					
2019		18,5991	4,5946								1						
2020		18,5991															

Le facteur de développement estimé pour une année de développement j fixée s'écrit :

$$\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} f_{i,j} \times \omega_{i,j} \times C_{i,j}^\alpha}{\sum_{i=0}^{n-j-1} \omega_{i,j} \times C_{i,j}^\alpha}, \quad \text{avec } \omega_{i,j} = 1_{\{f_{i,j} \neq \max_j(f_{i,j}) \text{ et } f_{i,j} \neq \min_j(f_{i,j})\}}$$

b- Cadre pratique branche auto corporelle

Pondération « expert » :

L'hypothèse de la méthode Chain-Ladder classique s'appliquant aux données directement, l'expert juge qu'il n'a pas à pondérer les données. Pour $\omega_{i,j} = 1$ et $\alpha = 1$ et l'on retrouve les mêmes résultats que Chain-Ladder classique.

Pondération « Min/Max » :

En développant le triangle grâce à la méthode de Chain-Ladder, nous obtenons l'estimation de la charge totale à l'ultime de l'assureur. Nous en déduisons le montant à provisionner. Les tableaux ci-dessous résument les informations obtenues.

Tableau 9 : Triangle complété par Chain Ladder pondérée « Min/Max »

	0	1	2	3	4	5	6	7	8
2013	7 294	21 105	62 031	76 194	88 583	94 916	103 522	104 178	111 766
2014	5 552	29 118	57 697	104 696	123 713	136 049	150 906	160 270	171 945
2015	2 234	24 011	39 433	55 122	69 065	70 027	80 136	82 876	88 913
2016	5 425	23 930	41 152	72 870	84 763	98 889	109 688	113 439	121 702
2017	3 271	22 771	111 495	121 248	152 664	165 732	183 830	190 116	203 965
2018	1 260	24 698	68 142	90 501	108 534	117 825	130 691	135 160	145 006
2019	562	10 316	47 397	67 837	81 355	88 319	97 963	101 313	108 693
2020	4 734	88 028	246 371	352 620	422 884	459 084	509 216	526 629	564 990
2021	7 879	84 453	236 367	338 302	405 713	440 442	488 539	505 245	542 048
fac_dev_pon	10,7193	2,7988	1,4313	1,1993	1,0856	1,1092	1,0342	1,0728	1

Tableau 10 : Provision auto corporelle par la méthode Chain Ladder pondérée « Min/Max »

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	111 766	111 766	0	0,00%
2014	171 945	160 270	11 674	0,96%
2015	88 913	80 136	8 777	0,72%
2016	121 702	98 889	22 813	1,87%
2017	203 965	152 664	51 301	4,20%
2018	145 006	90 501	54 505	4,46%
2019	108 693	47 397	61 296	5,02%
2020	564 990	88 028	476 962	39,05%
2021	542 048	7 879	534 170	43,73%
Total	2 059 028	837 529	1 221 498	100%

Avec cette pondération l'assureur provisionne à hauteur de **1 221 498 FCFA** pour la branche auto corporelle, soit 16,69% de plus que pour la version non pondérée. En utilisant donc la version pondérée « Min/Max » du Chain Ladder pour estimer la provision auto corporelle, l'assureur est beaucoup plus prudent.

3- London-Chain

Cette méthode a été introduite par (Benjamin et Eagles, 1986) afin de calculer des provisions aux Lloyd's.

a- Cadre théorique

Hypothèses du modèle

Ce modèle reprend les hypothèses précédentes, mais repose sur une relation moins contraignante entre les états $C_{i,j}$ et $C_{i,j+1}$ selon laquelle l'un est fonction affine de l'autre :

$\forall j \in \{0,1, \dots, n-1\}, \exists f_j$ et a_j tels que :

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

Estimateurs du modèle

On a dans ce modèle deux fois plus de paramètres à estimer que dans la méthode de Chain-Ladder. Les f_j et les a_j pour $j = 0,1, \dots, n$. Ce modèle se présente sous la forme de régressions linéaires et ainsi, la manière qui semble la plus naturelle pour estimer ses paramètres consiste en la méthode des moindres carrés. Les estimateurs vérifient donc, pour tout

$j = 0,1, \dots, n$:

$$(\hat{F}_j, \hat{a}_j) = \operatorname{argmin}\{\sum_{i=0}^{n-j-1} (C_{i,j+1} - a_j - f_j \times C_{i,j})^2\},$$

ce qui donne les estimations suivantes :

$$\hat{F}_j = \frac{\frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j} \times C_{i,j+1} - \bar{C}_j \times \bar{C}_{j+1}}{\frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j}^2 - \bar{C}_j^2}, \quad \hat{a}_j = \bar{C}_{j+1} - \hat{F}_j \times \bar{C}_j$$

$$\text{Où } \bar{C}_j = \frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j}, \quad \bar{C}_{j+1} = \frac{1}{n-j} \sum_{i=0}^{n-j-1} C_{i,j+1}$$

Notons que : $\hat{F}_j = \frac{\operatorname{cov}(\text{colonne}_j, \text{colonne}_{j+1})}{\operatorname{var}(\text{colonne}_j)}$

b- Avantage du London-Chain

La méthode de London-Chain est plus élaborée et souple que la méthode de Chain-Ladder dans la mesure où elle tient compte d'une tendance multiplicative, mais aussi d'une tendance additive (ou incrémentale). En posant cette dernière à zéro, on retrouve le modèle de Chain-Ladder. Ainsi, contrairement à la méthode de Chain-Ladder qui ne pouvait être appliquée que lorsque les points $(C_{i,j}, C_{i,j+1})$ étaient sensiblement alignés (à j fixé) sur une droite passant par l'origine, la méthode de London-Chain suppose elle aussi un alignement des points sur une même droite, mais relâche l'hypothèse d'alignement avec l'origine. Le relâchement de cette contrainte peut être validé à l'aide d'un test statistique concernant la significativité du facteur incrémental.

De manière formelle, le test d'hypothèse est le suivant :

$$\begin{cases} \mathcal{H}_0: a_j = 0 \\ \mathcal{H}_1: a_j \neq 0 \end{cases}$$

Dans le cas où l'on suppose la normalité des résidus ϵ_j , il est montré que l'estimateur \hat{a}_j suit une loi normale $\mathcal{N}(a_j, \frac{\sigma_j^2}{n-j} + \frac{\sigma_j^2 \times \bar{c}_j^2}{(n-j-1)S_{C_j}^2})$ où σ_j^2 est la variance de ϵ_j et $S_{C_j}^2$ est un estimateur sans biais de la variance de C_j . On définit ensuite la statistique de test suivante

$$T_j = \frac{\hat{a}_j - a_j}{S_{\epsilon_j} \sqrt{\frac{1}{l-j} + \frac{\bar{c}_j^2}{(l-j-1)S_{C_j}^2}}}$$

Où $S_{C_j}^2$ est un estimateur sans biais de σ_j^2 .

Ainsi, sous les hypothèses précédentes, $T_j \sim t_{n-j-2}$, ce qui nous permet d'obtenir une zone de rejet de l'hypothèse nulle \mathcal{H}_0 pour un niveau de confiance arbitraire. On pourra alors par exemple valider le modèle à 95 % dans le cas où la p-value associée à ce test est inférieure 5%.

c- Cadre pratique : branche auto corporelle

Les hypothèses du modèle ont toutes été vérifiées dans la méthode Chain-Ladder, hormis la significativité des paramètres additifs $(a_j)_{j=0, \dots, n-1}$.

Principe : Régression linéaire sans contrainte des $(C_{i,j+1})_{i=0, \dots, n-j}$ sur $(C_{i,j})_{i=0, \dots, n-j}$

$\forall j \in \{0, \dots, n-1\}$. Obtention des paramètres multiplicatifs et additifs, et des p-values pour chacun de ces paramètres.

Pour notre triangle, les résultats des estimations sont :

Tableau 11 : Test de significativité des paramètres additifs modèle London Chain

Paramètre testé	Valeur	p-valeur	Significativité
a ₀	27231,92	0,289	Non Pertinent
a ₁	65200	0,309	Non Pertinent
a ₂	50500	0,079	Non Pertinent
a ₃	-7030	0,548	Non Pertinent
a ₄	-8726,994	0,673	Non Pertinent
a ₅	4230	0,685	Non Pertinent

Ainsi, aucun des $(a_j)_{j=0,\dots,n-1}$ n'est significatif à un niveau de 95 %. Le modèle de London-Chain n'est donc pas nécessairement le plus ajusté à notre branche. Néanmoins, malgré la pertinence limitée de l'application de ce modèle à notre triangle, nous garderons les résultats pour la suite, à titre de comparaison.

Les tableaux ci-dessous résument les résultats de la méthode de London Chain appliquée à la branche auto corporelle.

Tableau 12 : Triangle complété par la méthode London Chain

	0	1	2	3	4	5	6	7	8
2013	7 294	21 105	62 031	76 194	88 583	94 916	103 522	104 178	111 766
2014	5 552	29 118	57 697	104 696	123 713	136 049	150 906	160 270	171 945
2015	2 234	24 011	39 433	55 122	69 065	70 027	80 136	76 494	82 066
2016	5 425	23 930	41 152	72 870	84 763	98 889	109 964	111 804	119 949
2017	3 271	22 771	111 495	121 248	152 664	171 111	187 959	204 134	219 004
2018	1 260	24 698	68 142	90 501	109 429	120 798	133 625	139 813	149 998
2019	562	10 316	47 397	74 472	89 114	97 158	108 095	109 591	117 574
2020	4 734	88 028	96 971	112 753	137 631	153 617	169 067	181 770	195 010
2021	7 879	41 356	71 472	93 063	112 676	124 577	137 706	144 644	155 181
Pente, f _j	2,6568	0,54633	0,7722	1,2674	1,1637	1,0799	1,1838	1,0728	
Constante, a _j	20424	48878	37872	-5271,2	-6545,2	3170,2	-18370	0	

Tableau 13 : Provision auto corporelle par la méthode London Chain

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	111 766	111 766	0	0,00%
2014	171 945	160 270	11 675	2,41%
2015	82 066	80 136	1 930	0,40%
2016	119 949	98 889	21 060	4,34%
2017	219 004	152 664	66 340	13,68%
2018	149 998	90 501	59 497	12,27%
2019	117 574	47 397	70 177	14,47%
2020	195 010	88 028	106 982	22,06%
2021	155 181	7 879	147 302	30,37%
Total	1 322 493	837 529	484 963	100%

Le montant des réserves obtenu par la méthode London-Chain est de **484 963 FCFA**, soit un montant de réserves nettement plus petit que celui obtenu par la méthode Chain-Ladder. L'écart relatif s'élève à 53,67 %.

Ce résultat n'est tout de même pas surprenant car nous avons constaté dans la validation des hypothèses de ce modèle qu'il n'était pas ajusté à nos données.

II- Méthodes stochastiques

Intérêt

L'idée dans cette section est de pouvoir quantifier la variabilité des réserves estimées, notamment par la construction d'intervalles de confiance, et ainsi d'obtenir une marge d'erreur sur le montant des provisions. Pour cela, les méthodes stochastiques sont adaptées, puisqu'elles considèrent les réserves sous un angle probabiliste en considérant leur distribution.

1- Modèle de Mack (1993)

a- Cadre théorique

Hypothèses du modèle

Le modèle proposé par (Mack, 1993) est un modèle non paramétrique, conditionnel à la réalisation du triangle, et qui permet d'estimer les erreurs commises lors du calcul des réserves. Il s'agit en fait d'un modèle stochastique pour la méthode de Chain-Ladder. Les hypothèses sous-jacentes à ce modèle sont les suivantes :

(H1) Les paiements cumulés $C_{i,j}$ sont indépendants par année de survenance c'est-à-dire :

$$\forall i, k \{ C_{i,1}, \dots, C_{i,n} \} \text{ et } \{ C_{k,1}, \dots, C_{k,n} \} \text{ sont indépendants.}$$

(H2) Il existe des paramètres f_1, \dots, f_n tel que :

$$\forall i \in \{ 0, \dots, n \}, \quad j \in \{ 0, \dots, n - 1 \}, \quad \mathbb{E}(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = C_{i,j} f_j$$

Remarque : Précisons que l'hypothèse (H1) n'est plus vérifiée dès lors qu'intervient un changement de mode de gestion, ou dans les taux d'inflation de la branche considérée. Des mises en *as if* permettent cependant de limiter ce point. Sous ces deux hypothèses, Mack (Mack, 1993) a montré que ce modèle stochastique permet d'obtenir exactement les mêmes montants de provision que la méthode standard de Chain Ladder, qui est la référence en matière de provisionnement.

Sous (H1) et (H2) et conditionnellement à $D_n = \{ C_{i,j} : i + j \leq n \}$, on obtient les résultats suivants.

$$(P1) : \forall i \geq 1, \quad \mathbb{E}(C_{i,n} \setminus D_n) = f_{n-1} \times \dots \times f_{n-i} C_{i,n-i}$$

En effet :

$$\mathbb{E}(C_{i,n} \setminus D_n) = \mathbb{E}(C_{i,n} \setminus C_{i,0}, \dots, C_{i,n-i}) = \mathbb{E}(\mathbb{E}(C_{i,n} \setminus C_{i,0}, \dots, C_{i,n-i}) \setminus C_{i,0}, \dots, C_{i,n-i})$$

$$\text{Or} \quad \mathbb{E}(C_{i,j} \setminus C_{i,0}, \dots, C_{i,j-i}) = f_{j-1} C_{i,j-1}$$

$$\text{D'où} \quad \mathbb{E}(C_{i,n} \setminus D_n) = f_{n-1} \mathbb{E}(C_{i,n-1} \setminus C_{i,0}, \dots, C_{i,n-i})$$

Par le même raisonnement en double conditionnement, l'Hypothèse (H2) donne plus généralement :

$$\mathbb{E}(C_{i,n+1} \setminus D_n) = f_j C_{i,j}$$

Théorème de Mack (T1)

Les facteurs Chain-Ladder $\hat{f}_j = \frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}}$ sont sans biais et non corrélés sous l'hypothèses

(H1).

En effet :

Soit

$$j \in \{ 0, \dots, n-1 \}, \quad D_j = \{ C_{i,k} : k \leq j, i + j \leq n \}$$

la partie du triangle supérieur limitée à l'année de règlement j .

$$\text{On a : } \mathbb{E}(C_{i,j+1} \setminus D_j) = f_j C_{i,j}$$

$$\mathbb{E}(\hat{f}_j \setminus D_j) = \mathbb{E}\left(\frac{\sum_{i=0}^{n-j-1} C_{i,j+1}}{\sum_{i=0}^{n-j-1} C_{i,j}} \setminus D_j\right) = \frac{\sum_{i=0}^{n-j-1} \mathbb{E}(C_{i,j+1} \setminus D_j)}{\sum_{i=0}^{n-j-1} C_{i,j}} = f_j$$

En conditionnant

$$\mathbb{E}(\hat{f}_j) = \mathbb{E}\left(\mathbb{E}(\hat{f}_j \setminus D_j)\right) = f_j,$$

\hat{f}_n est un estimateur non biaisé de f_j .

Pour $j < k$:

$$\mathbb{E}(\hat{f}_j \hat{f}_k) = \mathbb{E}\left(\mathbb{E}(\hat{f}_j \hat{f}_k \setminus D_k)\right) = \mathbb{E}(\hat{f}_j \hat{f}_k) = f_j f_k$$

Par l'égalité

$$\mathbb{E}(\hat{f}_k \setminus D_k) = f_k \quad \text{on en déduit : } cov(\hat{f}_j \hat{f}_k) = \mathbb{E}(\hat{f}_j \hat{f}_k) - \mathbb{E}(\hat{f}_j) \mathbb{E}(\hat{f}_k) = 0$$

Les \hat{f}_j ne sont pas corrélés. \square

L'absence de biais des facteurs de développement se « transmettent » aux valeurs calculées à l'aide des facteurs de développement Chain-Ladder.

Ainsi nous avons que les $\hat{C}_{i,j}$ et $\hat{C}_{i,n}$ ainsi que \hat{R}_i et \hat{R} sont sans biais.

L'incertitude à l'ultime de Mack

Par la méthode Chain-Ladder, nous avons obtenu un estimateur du montant de provision \hat{R} . On cherche à quantifier l'incertitude de l'estimateur \hat{R} associée à une variable aléatoire R .

L'incertitude présente dans la prédiction de \hat{R}_i est mesurée par l'erreur quadratique moyenne conditionnellement à D_n que l'on note MSEP.

Nous avons :

$$MSEP_{\hat{R}_i}(R_i) = \mathbb{E}((\hat{R}_i - R_i)^2 | D_n) = \mathbb{E}(\hat{C}_{i,n} - C_{i,n} | D_n) = MSEP_{\hat{C}_{i,n}}(C_{i,n})$$

L'erreur standard associée est notée : $sep(\hat{R}_i) = \sqrt{MSEP_{\hat{R}_i}(R_i)}$

Afin d'estimer ces quantités, Thomas Mack a introduit une 3ème hypothèse sur la variabilité au sein du rectangle.

Hypothèse de Mack (H3) :

$\forall j \in \{0, \dots, n-1\}$, il existe un paramètre σ_j^2 tel que :

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

(P2) : L'égalité de l'hypothèse (H3) s'écrit en termes de facteur comme

$$Var(f_{i,j} | C_{i,0}, \dots, C_{i,n-i}) = \frac{\sigma_j^2}{C_{i,j}}$$

Sous (H1), (H2) et (H3) de manière analogue aux démonstrations de \hat{f}_n on montre que pour

$$j \leq n-2:$$

$$\sigma_j^2 = \frac{1}{n-j-1} \sum_{i=0}^{n-j-1} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2, \quad \hat{\sigma}_{n-1}^2 = \min \left(\frac{\hat{\sigma}_{n-2}^2}{\hat{\sigma}_{n-3}^2}, \min(\hat{\sigma}_{n-3}^2, \hat{\sigma}_{n-2}^2) \right)$$

est un estimateur sans biais de σ_j^2 .

Cet estimateur compare les facteurs $f_{i,j} = \frac{C_{i,j+1}}{C_{i,j}}$ à leurs estimateurs donnés par Chain-Ladder.

Sous l'hypothèse (H3), Thomas Mack démontre que l'estimateur de Chain-Ladder est à variance minimale dans la classe des estimateurs sans biais.

Théorème de Mack (T2)

Si l'on pose : $\hat{C}_{i,n-i} = C_{i,n-i}$ sous les hypothèses (H1), (H2), et (H3), on peut estimer la $MSEP(R_i)$ pour $i = 0, \dots, n$ par :

$$\widehat{MSEP}(R_i) = \hat{C}_{i,n}^2 \sum_{j=n-i}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left[\frac{1}{\hat{C}_{i,j}^2} + \frac{1}{\sum_{k=0}^{n-j} C_{k,j}} \right]$$

Remarque :

On note que $\hat{C}_{i,n}$ est un prédicteur D-mesurable de $C_{i,n}$ d'où on obtient la décomposition en Erreur de processus + Erreur d'estimation

suivante :

$$MSEP_{\hat{R}_i}(R_i) = \underbrace{Var(C_{i,n} \setminus D)}_{\text{Erreur de processus}} + \underbrace{(\mathbb{E}(C_{i,n} \setminus D) - \hat{C}_{i,n})^2}_{\text{Erreur d'estimation}} = \hat{C}_{i,n}^2 \sum_{j=n-i}^{n-1} \frac{\hat{\sigma}_j^2}{\hat{f}_j^2} \left[\frac{1}{\hat{C}_{i,j}^2} + \frac{1}{\sum_{k=0}^{n-j} C_{k,j}} \right]$$

Le théorème (T2) nous donne la MSEP pour une année de survénance i donnée. Le corollaire ci-dessous donne une estimation de la MSEP pour toutes les années de survénances agrégées.

Corollaire :

$$\widehat{MSEP}_{\hat{R}_i}(R_i) = \sum_{i=1}^n \{ MSEP(\hat{R}_i) + \hat{C}_{i,n} (\sum_{k=i+1}^n \hat{C}_{k,n}) \sum_{j=n-i}^{n-1} \frac{2\hat{\sigma}_j^2}{\hat{f}_j^2 \sum_{h=0}^{n-j} C_{h,j}} \} \text{ estime } MSEP_{\hat{R}_i}(R_i).$$

En résumé, les hypothèses sous-jacentes à la méthode de Mack sont :

(H1) Les paiements cumulés $C_{i,j}$ sont indépendants par année de survénance c'est-à-dire :

$$\forall i \neq k, \{ C_{i,1}, \dots, C_{i,n} \} \text{ et } \{ C_{k,1}, \dots, C_{k,n} \} \text{ sont indépendants.}$$

(H2) Il existe des paramètres f_1, \dots, f_n tel que :

$$\forall i \in \{ 0, \dots, n \}, \quad j \in \{ 0, \dots, n-1 \}, \quad \mathbb{E}(C_{i,j+1} \setminus C_{i,1}, \dots, C_{i,j}) = C_{i,j} f_j$$

(H3) $\forall j \in \{ 0, \dots, n-1 \}$, il existe un paramètre σ_j^2 tel que :

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

Remarque :

Ces hypothèses peuvent s'écrire sous la forme suivante :

$$C_{i,j+1} = f_j C_{i,j} + \varepsilon_{i,j} \sqrt{C_{i,j}}, \quad 0 \leq i \leq n$$

Où

$$(\varepsilon_{i,j})_{i \geq 0} \text{ sont i. i. d et de variance } \sigma_j^2.$$

Cette formule permet d'estimer $(f_j)_{j \geq 0}$ par l'estimateur des moindres carrés.

Sous les hypothèses (H1) et (H2) les estimateurs des moindres carrés sont obtenus par :

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

Soit le même que celui de Chain-Ladder.

De même les estimateurs des paramètres σ_j^2 sont estimés par :

$$\widehat{\sigma}_j^2 = \widehat{\mathbb{E}}[\varepsilon_{i,j}^2] = \frac{1}{n-j} \sum_{i=1}^{n-j} \left(\frac{C_{i,j+1} - \hat{f}_j C_{i,j}}{\sqrt{C_{i,j}}} \right)^2$$

Afin de vérifier l'hypothèse (H3) on peut étudier le graphique des résidus en fonction des $C_{i,j}$ et on vérifie si l'hypothèse de variance utilisée conduit bien à des résidus aléatoires, ne manifestant pas de tendance, $\varepsilon_{i,j} = \frac{C_{i,j+1} - \hat{f}_j C_{i,j}}{\sqrt{C_{i,j}}}$.

Construction d'un intervalle de confiance

Dans le but de déterminer un intervalle de confiance pour nos provisions calculées par année de survenance, il est nécessaire de définir l'écart type des R_i . L'écart type est déduit du MSEP calculé précédemment. En effet il est défini par :

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

Le modèle de Mack ne fournit que la distribution pour les provisions calculées à ultime. Cependant il est nécessaire de supposer l'hypothèse paramétrique sur la forme de la distribution des \hat{R}_i afin de pouvoir définir des intervalles de confiance autour de ceux-ci.

Afin d'estimer les intervalles de confiance, on suppose que les montants des provisions suivent la loi log-normale de paramètres μ_i et σ_i .

Classiquement l'hypothèse sur la distribution du montant des provisions est de supposer que les provisions suivent une loi normale de moyenne les montants estimés \hat{R}_i et d'écart type $\widehat{sep}(\hat{R}_i)$. L'intervalle de confiance à 95% est donné par :

$$[\hat{R}_i \pm 1,96 \times \widehat{sep}(\hat{R}_i)].$$

Cette hypothèse peut être remise en cause du fait de la distribution symétrique de la loi normale qui ne semble pas se vérifier dans la pratique.

C'est d'ailleurs pour cette raison que nous avons choisi d'utiliser la loi log-normale qui présente l'avantage d'être une distribution asymétrique, avec des queues de distribution plus épaisses.

Rappel : Soit $X \sim \log - normale$ de paramètres μ_i et σ_i . La fonction de densité est donnée par :

$$f(x, \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$$

$$\mathbb{E}[X] = e^{\left(\mu + \frac{\sigma^2}{2}\right)} \quad \text{et} \quad V[X] = (e^{\sigma^2} - 1)e^{(2\mu + \sigma^2)}$$

Ainsi nous aurons :

$$\mathbb{E}[X_i] = e^{\left(\mu_i + \frac{\sigma_i^2}{2}\right)} = \hat{R}_i \Rightarrow \mu_i = \ln(\hat{R}_i) - \frac{\sigma_i^2}{2}$$

$$V[X_i] = (e^{\sigma_i^2} - 1)e^{(2\mu_i + \sigma_i^2)} = (\widehat{sep}(\hat{R}_i))^2 \Rightarrow \sigma_i^2 = \ln\left[1 + \left(\frac{\widehat{sep}(\hat{R}_i)}{\hat{R}_i}\right)^2\right]$$

L'intervalle de confiance à 95% est donné par :

$$IC = [\exp(\mu_i - 1,96 \times \sigma_i^2); \exp(\mu_i + 1,96 \times \sigma_i^2)]$$

$$IC = [\hat{R}_i \times \exp\left(-\frac{\sigma_i^2}{2} - 1,96 \times \sigma_i^2\right); \hat{R}_i \times \exp\left(-\frac{\sigma_i^2}{2} + 1,96 \times \sigma_i^2\right)]$$

b- Cadre pratique

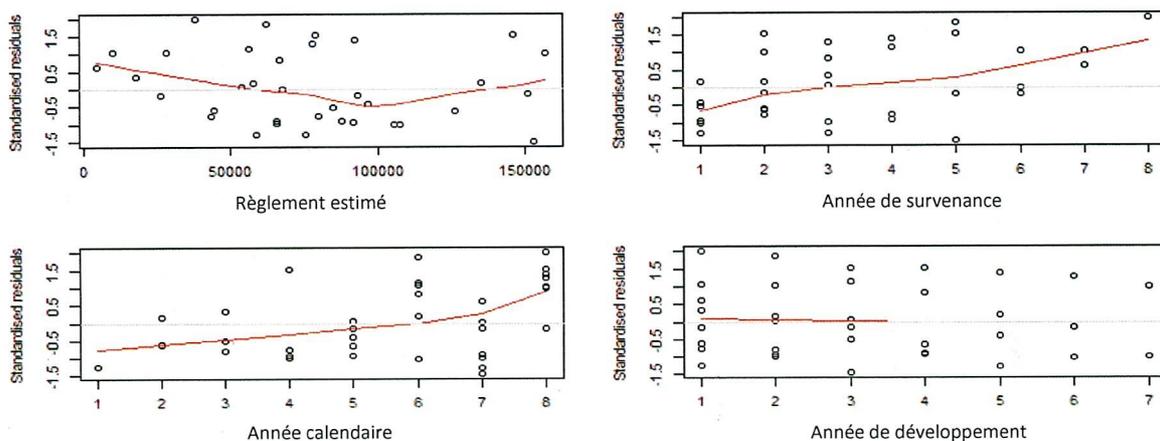
Nous appliquons le modèle de Mack au triangle de paiements cumulés auto corporelle afin d'estimer le montant espéré des paiements futurs ainsi que l'erreur commise.

Nous allons dans un premier temps regarder si la branche considéré vérifie les hypothèses sous-jacentes du modèle de Mack.

Les hypothèses (H1) et (H2) ont déjà été vérifiées lors des tests réalisés pour Chain-Ladder classique.

L'hypothèse (H3) peut se tester graphiquement. Nous construisons donc les graphiques des résidus.

Figure 8: Distribution des résidus de Pearson en fonction des règlements estimés, des années de survenance, des années calendaires et des années de développement de la branche auto corporelle.



Nous ne remarquons aucune tendance au sein des graphiques de résidus, par conséquent, nous pouvons dire que leur distribution est aléatoire.

Nous **validons** ainsi l'hypothèse faite sur la variance conditionnelle (**H3**).

Les trois hypothèses vérifiées, nous appliquons le modèle de Mack aux données afin d'estimer l'erreur de prédiction des provisions réalisées avec la méthode Chain-Ladder.

En reprenant les formules fournies auparavant, nous pouvons en déduire les provisions \hat{R}_i , l'estimation de l'erreur commise sur ces réserves (sa racine carrée notée $\widehat{sep}(\hat{R}_i)$) dans le tableau ci-après), ainsi que les intervalles de confiance à 95% autour des provisions basés sur la distribution normale d'une part et log normale d'autre part.

Tableau 14 : Intervalles de confiance normal et log normal du modèle Mack branche auto corporelle

Année de survenance	Provision	$\widehat{sep}(\hat{R})$	Loi log-normale		Loi normale	
			Borne_inf	Borne_sup	Borne_inf	Borne_sup
2013	0	0				
2014	11 674	1 949	10 911	12 152	7 854	15 494
2015	9 223	4 952	4 946	13 350	483	18 929
2016	23 681	6 809	19 478	26 593	10 335	37 027
2017	54 005	14 080	45 938	59 447	26 408	81 602
2018	57 267	12 950	50 655	61 592	31 885	82 649
2019	58 646	30 270	32 802	82 793	683	117 975
2020	451 669	177 286	317 492	556 772	104 188	799 150
2021	380 657	279 327	131 916	713 974	- 166 824	928 138
Total	1 046 823	362 749	791 994	1 235 310	335 835	1 757 810

Le coefficient de variation ($\frac{\widehat{sep}(\hat{R})}{\hat{R}}$) est de 35% c'est-à-dire que la variance de l'estimation $\mathbb{E}[R]$ représente 35% du montant total des provisions. Cela s'explique par le fait qu'en 2021 la valeur des règlements de la première année de développement est faible (nous rappelons que les règlements sont ceux jusqu'en juin 2021) ce qui augmente l'erreur d'estimation.

Perspective :

Ainsi, en se basant sur une distribution log-normale des provisions, non seulement l'assureur connaît le montant moyen des réserves (1 046 823) pour la branche auto corporelle, il a également l'information sur le montant de réserves nécessaire (1 235 310) en cas de dégradation exceptionnelle de la sinistralité dans ladite branche. Par conséquent, avec 1 235 310 FCFA, l'assureur est solvable avec une probabilité de 95%.

Et pour une distribution normale des provisions, l'assureur avec un montant de 1 757 810 FCFA sera capable de faire face à une dégradation exceptionnelle de la sinistralité dans 95% des cas.

Avec cette extension stochastique du Chain Ladder, l'assureur dispose donc des informations qui lui permettent d'être assez prudent.

c- Limites de la méthode de Mack

En reprenant les hypothèses du modèle de Chain Ladder, le modèle de Mack en reprend également les limites présentées précédemment.

2- Notion de Tail factor : Gestion de la liquidation incomplète des triangles

Pour les risques longs (par exemple ceux pouvant être entièrement réglés après plusieurs dizaines d'années), le triangle n'est parfois pas complet, c'est-à-dire que nous ne disposons pas du développement complet de la sinistralité. Il s'agit de troncature droite. Cela conduit à une sous-estimation des charges ultimes, sauf en cas de recours durant les dernières années. Il est donc nécessaire d'estimer une queue de développement du triangle.

Nous présentons dans la suite une méthode d'estimation de cette queue de développement. Cette méthode peut être utilisée dans les modèles reposant sur les hypothèses de Chain-Ladder et est basée sur les coefficients de déroulement. (Habib et Riban, 2012)

a- Extrapolation des facteurs de développement

Dans les modèles de type Chain-Ladder, nous estimons des facteurs de développement

$\hat{f}_j, j = 0, \dots, n$ en considérant que la $n^{\text{ème}}$ année est l'année de clôture. Comme nous l'avons évoqué précédemment, cela n'est pas toujours le cas. Pour pallier à ce problème, on peut avoir recours à une estimation des facteurs de développement pour les années suivantes (pour $k \in \{n, \dots, ult\}$). Pour cela, nous cherchons à ajuster une fonction $y = f(k)$, régulière et supérieure à 1, afin de les extrapoler. Il semble judicieux, à la vue de l'allure de la courbe des facteurs \hat{f}_k en fonction des années de développement, d'extrapoler par une fonction de type "exponentielle inverse". On va en fait chercher les paramètres a et b tels que la fonction $f(k) = e^{a-bk} + 1$ soit la plus proche de l'évolution des \hat{f}_k en fonction des années de développement. La fonction f va ensuite nous permettre d'extrapoler et d'estimer les valeurs des coefficients de Chain-Ladder pour des années de développement supérieures à la dernière année connue. Le passage au logarithme permet de résoudre ce problème par les moindres carrés. On réalise ainsi une extrapolation linéaire de $\ln(\hat{f}_k - 1)$. Le *tail factor* f_{ult} est alors obtenu grâce à la relation : $f_{ult} = \prod_{k \geq n} \hat{f}_k$

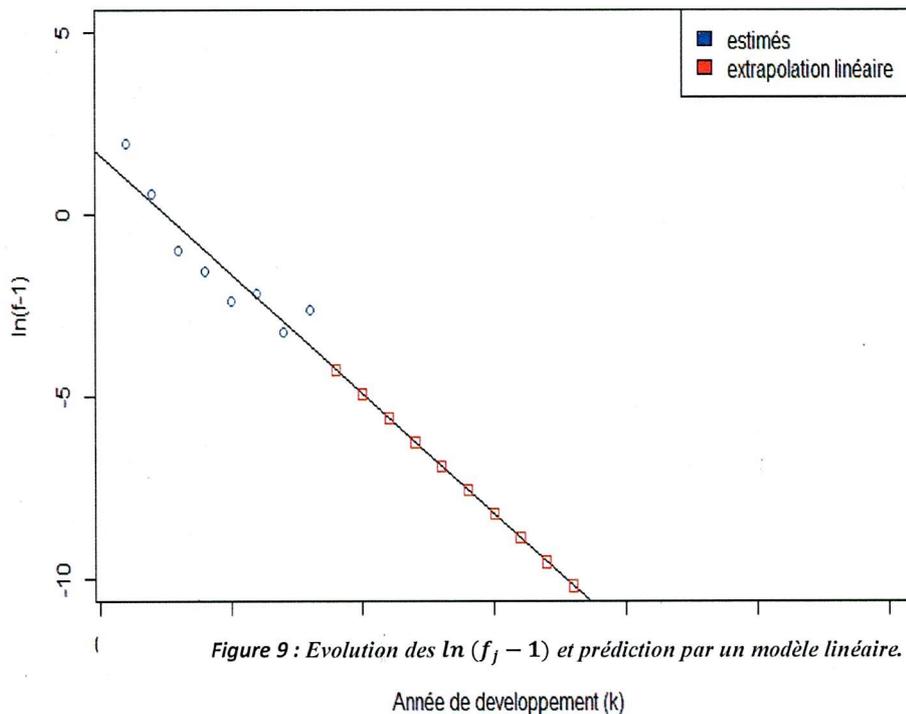
Et ainsi on a : $C_{i,ult} = f_{ult} \times \hat{C}_{i,n}$

Mais il faut toutefois que ce tail factor soit plausible et en concordance avec l'expérience acquise sur le développement des sinistres.

b- Application

Nous appliquons aux données de notre triangle de développement de la branche auto corporelle l'extrapolation exponentielle.

Voici la représentation qui nous permet de visualiser le modèle linéaire ajusté sur le logarithme des facteurs de transition.



En faisant un F-test sur les paramètres du modèle linéaire ajusté sur le logarithme des facteurs de transition, nous obtenons les résultats résumés dans le tableau ci-après.

Tableau 15 : résultats du F-test sur les paramètres du modèle linéaire ajusté sur le logarithme des facteurs de transition

Paramètre	Estimation	p-value
Constance (intercept)	1,653	0,0305
Coefficient	-0,659	0,0013
R ² – ajusté :		0,816

Les paramètres du modèle sont donc significatifs à 95%, car les p-value sont inférieures à 5% et le $R_{ajusté}^2 = 0,816$.

On peut alors appliquer au triangle initial l'ensemble des coefficients obtenus, et ainsi calculer un montant de réserves dans un cadre où l'on considère que la durée de liquidation totale du triangle a été atteinte.

En se basant à l'extrapolation, nous obtenons les résultats résumés dans le tableau suivants.

Tableau 16 : Provision auto corporelle après application du tail factor

tail factor	1,029047				
Année de survenance	Charge ultime Chain Ladder	Charge ultime après tail factor	Dernier paiement	Provision	% de la provision totale
2013	111 766	115 013	111 766	3 247	0,29%
2014	171 945	176 939	160 270	16 669	1,51%
2015	89 359	91 955	80 136	11 819	1,07%
2016	122 570	126 130	98 889	27 241	2,47%
2017	206 668	212 672	152 664	60 008	5,45%
2018	147 768	152 060	90 501	61 559	5,59%
2019	106 043	109 123	47 397	61 726	5,60%
2020	539 697	555 374	88 028	467 346	42,43%
2021	388 535	399 821	7 879	391 942	35,58%
Total	1 884 351	1 939 087	837 529	1 101 558	100%

Le montant estimé des réserves est de **1 101 558 FCFA**, montant logiquement supérieur au cas d'absence de tail factor, avec une différence relative de 5,23%.

Ceci nous amène à penser que les sinistres ne sont pas entièrement clos en branche auto corporelle après 9 ans. Il y a donc nécessité de revoir la profondeur du triangle pour cette branche si l'on se limite à utiliser la méthode du Chain-Ladder.

Il est à noter que l'extension stochastique du Chain-Ladder nous donne de pallier à ce problème, puisque nous avons constaté ci-dessus qu'en estimant la provision à 95%, nous sommes largement au-dessus du montant trouvé après application du tail factor.

3- Méthode de régression log-normale

a- Présentation du modèle

Dans le cadre du modèle Log-normale on suppose que les règlements $Y_{i,j}$ ($Y_{i,j} > 0$) sont distribués selon une loi Log-normale.

$Y_{i,j} \sim \text{LogN}(m_{i,j}, \sigma^2)$, où σ^2 est un paramètre de dispersion.

On pose :

$$X_{i,j} = \ln(Y_{i,j}) = \mu + \alpha_i + \beta_j + \varepsilon_{i,j}, \quad \mathbb{E}(X_{i,j}) = \mu + \alpha_i + \beta_j \text{ avec } \mathbb{E}(\varepsilon_{i,j}) = 0$$

Les erreurs $\varepsilon_{i,j}$ étant i.i.d, distribuées selon la loi Normale $\mathcal{N}(0, \sigma^2)$.

Les variables α_i et β_j sont binaires et permettent de capter respectivement les effets des années de survenance et les cadences de règlements.

$$\alpha_i = \begin{cases} 1 & \text{si le sinistre est survenue l'année} \\ 0 & \text{sinon} \end{cases}$$

$$\beta_i = \begin{cases} 1 & \text{si le sinistre est réglé l'année } i \\ 0 & \text{sinon} \end{cases}$$

Estimation des paramètres

L'estimation des paramètres se fait à partir du modèle de régression multiple (Sabam et Ngoufo, 2015)

$$X = M\beta + \varepsilon$$

Avec : X : Le vecteur des variables à expliquer ;

M : la matrice de regression associée ;

β : le vecteur des variables explicatives

$$\varepsilon_{i,j} = X_{i,j} - \hat{X}_{i,j}$$

En introduisant les contraintes $\alpha_0 = \beta_0 = 0$, nous estimons les paramètres par maximum de vraisemblance :

$$\hat{\beta} = (M'M)^{-1}M'X ; \quad \widehat{MSE} = \frac{\varepsilon^2}{n - p - 1}$$

La partie inférieure du triangle est complétée par la formule suivante :

$$\mu_{i,j} = \mathbb{E}[Y_{i,j}] = \exp\left(\mu + \alpha_i + \beta_j + \frac{\sigma^2}{2}\right),$$

$$V[Y_{i,j}] = \exp\left(2(\mu + \alpha_i + \beta_j) + \frac{\sigma^2}{2}\right) \times (\exp(\sigma^2) - 1)$$

b- Application

Nous appliquons la méthode log-normale à la branche auto corporelle. Nous commençons par la validation du modèle et l'analyse des résidus de Pearson, puis nous terminons par la production des résultats.

Tableau 17 : Triangle des incréments auto corporelle

	0	1	2	3	4	5	6	7	8
2013	7 294	13 811	40 927	14 162	12 389	6 332	8 606	656	7 589
2014	5 552	23 566	28 578	46 999	19 017	12 336	14 857	9 365	
2015	2 234	21 777	15 423	15 689	13 942	963	10 108		
2016	5 425	18 505	17 222	31 718	11 893	14 127			
2017	3 271	19 500	88 724	9 753	31 416				
2018	1 260	23 438	43 444	22 358					
2019	562	9 754	37 081						
2020	4 734	83 293							
2021	7 879								

Analyse des Résidus et Validations du Modèle

Pour tester la significativité des paramètres du modèle nous utilisons le test de Fisher. Les estimations sont faites à partir du logiciel R. les résultats obtenus sont présentés dans le tableau ci-dessous.

Test de Fisher

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

Tableau 18 : Résultat du test de Fisher

Test	Statistique du test	Degré de liberté	p-value
Fisher	5,83	(16,28)	2,63e ⁻⁰⁵

Nous constatons que la p-value de la statistique de Fisher est strictement inférieure à 5% donc nous rejetons l'hypothèse de non significativité des paramètres.

Par ailleurs le coefficient de détermination ajusté du modèle est égal à 0,64. Nous pouvons donc considérer que le modèle est significatif.

Analyse des Résidus

Nous allons vérifier graphiquement si les résidus sont distribués de façon aléatoire et tester leurs normalités à l'aide du test de Shapiro-Wild.

Tableau 19 : Les résidus de Pearson normalisés du modèle log-normale

	1	2	3	4	5	6	7	8	9
1	1,4	-0,328926	0,34242	-0,23988	-0,10993	0,34055	-0,14708	-1,2572	-1,06E-16
2	0,19818	-0,507533	-0,9689	0,42479	-0,42274	0,32993	-0,31092	1,2572	
3	0,30101	0,649449	-0,49374	0,29096	0,44086	-1,64654	0,458		
4	0,64302	-0,339659	-1,13656	0,40059	-0,54344	0,97606			
5	-0,05093	-0,326013	0,8889	-1,14721	0,63526				
6	-0,89401	0,273368	0,34988	0,27076					
7	-1,0488	0,030845	1,018						
8	-0,54847	0,548469							
9	-1,76E-16								

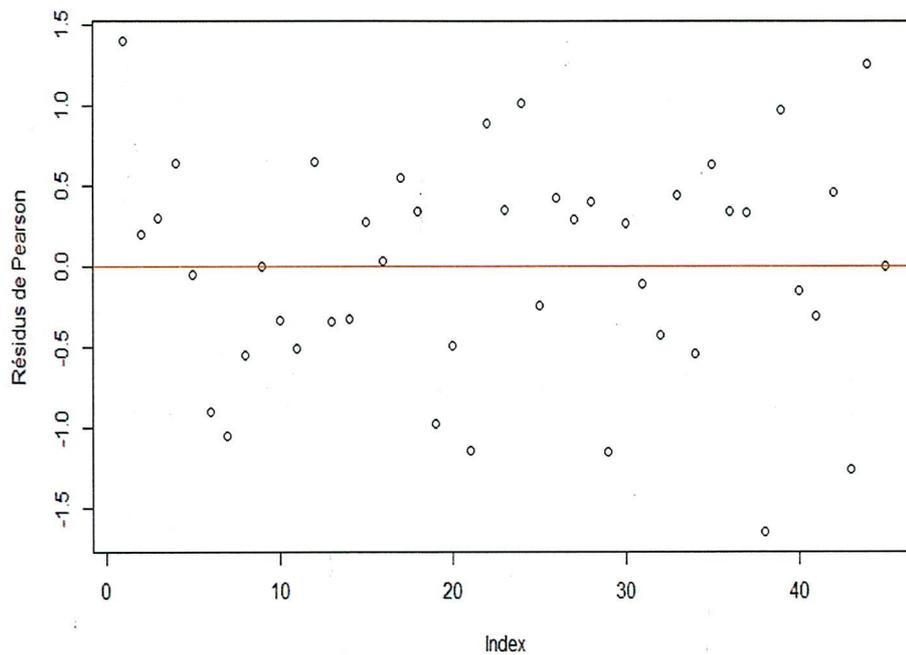


Figure 10 : Les résidus de Pearson normalisés du modèle log-normale

D'après le graphique précédent, les résidus sont distribués aléatoirement, il y a absence de structure.

Nous vérifions l'hypothèse de normalité des résidus par le test de Shapiro-Wild et le Q-Q plot des résidus.

Test de Shapiro-Wild

Pour tester la normalité des résidus nous utilisons le test de Shapiro-Wild.

Tableau 20 : Résultats du test de Shapiro-Wild

Test	Statistique du test	p-value
Shapiro-Wild	0,98	0,63

La p-value de la Statistique de Shapiro-Wild est supérieure à 5%, l'hypothèse de normalité des résidus est acceptée. Nous confirmons cela par le Q-Q plot. Nous avons fait la comparaison graphique de la distribution des résidus à la distribution d'une loi normale. Si les points sont alignés le long d'une droite, la distribution des résidus peut être considérée comme normale.

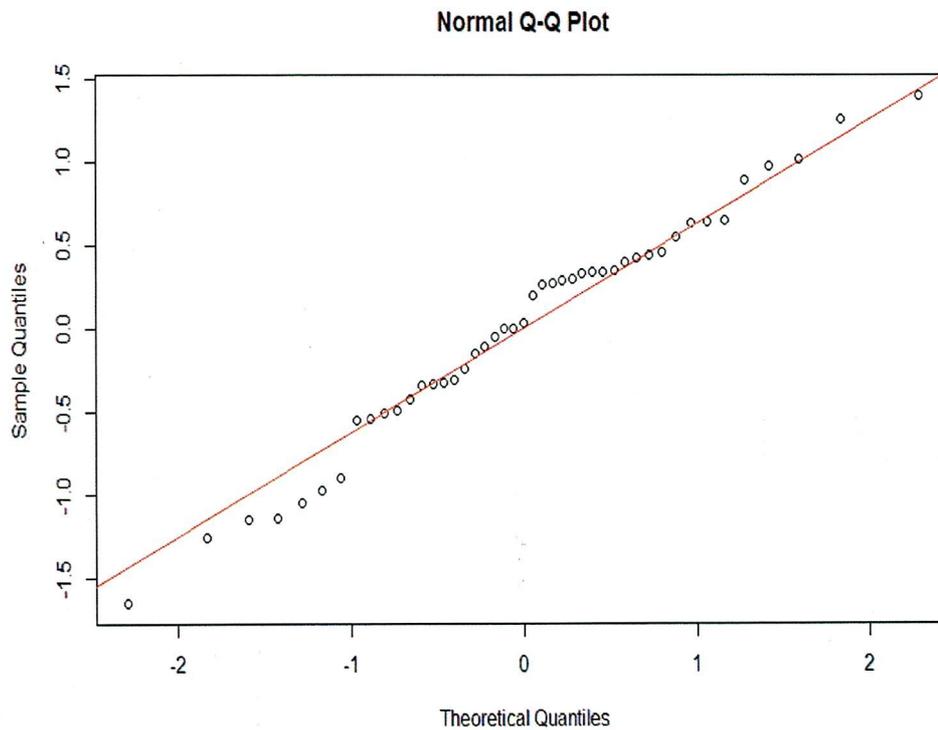


Figure 11 : Q-Q plot des Résidus du modèle Log-Normale vs Distribution normale

L'analyse du Q-Q Plot des résidus et le résultat du test de Shapiro-Wild nous permettent de conclure à la normalité des résidus.

Résultats obtenus

Les estimations sont faites à partir du logiciel R. les résultats obtenus sont présentés dans le tableau ci-dessous.

Tableau 21 : Paramètre estimés du modèle Log-Normale

Coefficient	μ	α_0	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
Estimation	7,79	0	2	2,56	1,96	1,72	0,69	1,39	-0,31	1,14

Coefficient	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Estimation	0	0,67	-0,32	0,30	0,34	0,05	-0,63	1,10	1,18

A partir des paramètres estimés nous complétons la partie inférieure du triangle des règlements et déduisons les provisions à l'ultime.

Tableau 22 : Provisions par la méthode log-Normale

Année de survenance	Provision	% de la provision totale
2013	0	0,00%
2014	18 819	2,00%
2015	8 608	0,91%
2016	32 446	3,45%
2017	42 433	4,51%
2018	49 786	5,29%
2019	36 574	3,88%
2020	326 664	34,69%
2021	426 359	45,28%
Total	941 689	100,00%

La provision à l'ultime est de **941 689 FCFA**, elle est légèrement inférieure à celle de Chain Ladder car l'estimation des provisions avec la loi log normale intègre le paramètre de volatilité.

c- Les limites du modèle Log-normal

Le modèle de régression log-normal présente l'avantage de donner de bon résultat et simple à mettre en œuvre. Cependant l'application de cette méthode impose une condition à priori sur les incréments ($Y_{i,j} > 0$) qui n'est pas toujours vérifiée. La valeur attribuée aux incréments négatifs ou nul influence les paramètres estimés.

4- Modèle Linéaire Généralisé (GLM)

Nous allons présenter plus généralement en quoi consiste les méthodes de provisionnement basées sur les modèles GLM. Ces derniers ont été introduits par (Nelder et Wedderburn, 1972). Il s'agit en fait d'extensions du modèle linéaire Normal. Ils ont commencé à être utilisés pour la détermination stochastique des provisions pour sinistres à partir des années 1990.

L'idée principale d'un modèle GLM est de capter la tendance des règlements incrémentaux au fil des années de survenance et des années de développement en supposant que les règlements suivent une loi paramétrique. Parmi les familles de distribution possibles (Poisson sur-dispersée, log-normal, Gamma, etc...) la théorie prouve qu'avec la famille Poisson sur-dispersée, le GLM reproduit exactement les mêmes prédictions que la méthode de Chain Ladder (Renshaw et al, 1998).

a- Présentation du modèle

Le modèle utilise trois composantes : les composantes aléatoires, systématiques et la fonction de lien.

Composante aléatoire

On suppose que les règlements non cumulés $(Y_{i,j})_{i,j=0,\dots,n-1}$ sont des variables aléatoires indépendantes de fonction de densité :

$$f(Y_{i,j}; \theta_{i,j}; \varphi) = \exp \left\{ \frac{[\theta_{i,j} Y_{i,j} - b(\theta_{i,j})] \omega_{i,j}}{\varphi} + c(Y_{i,j}, \varphi) \right\}$$

Où :

- $\theta_{i,j}$ est un paramètre naturel de la famille exponentiel ;
- φ est un paramètre de dispersion, il est indépendant de i et de j , avec $\varphi = 1$ pour une loi de poisson et $\varphi \neq 0$ pour la loi de poisson sur-dispersée ;
- $\omega_{i,j}$ est un pondéré donné ;
- b et c sont de fonctions spécifiques de la distribution, b est une fonction régulière.

On en déduit les relations sur l'espérance $\mu_{i,j} = E(Y_{i,j})$ des règlements non cumulés et la variance des règlements par :

$$\mu_{i,j} = E(Y_{i,j}) = b'(\theta_{i,j})$$

$$V(Y_{i,j}) = \varphi b''(\theta_{i,j}) = \varphi b''(b'^{-1}(\mu_{i,j})) = \varphi V(\mu_{i,j})$$

$$\mu_3(Y_{i,j}) = \varphi^2 V(\mu_{i,j}) V'(\mu_{i,j})$$

$$\gamma_1(Y_{i,j}) = \frac{\mu_3(Y_{i,j})}{[V(Y_{i,j})]^{\frac{3}{2}}} = \sqrt{\varphi} \frac{V'(\mu_{i,j})}{[V(\mu_{i,j})]^{\frac{1}{2}}}$$

La fonction V est une fonction variance de la distribution. Elle prend différente forme selon la distribution étudiée.

La composante systématique

La composante systématique est une composante déterministe basée sur un prédicteur linéaire de la forme : $\eta_{i,j} = \mu + \alpha_i + \beta_j \quad i, j = 0, \dots, n-1$

Pour raison d'identification du modèle, on pose $\alpha_0 = 0 = \beta_0 = 0$.

Dans le cadre du provisionnement :

- μ : est un paramètre du contrôle global, qui représente l'inflation supposé constance ;
- α_i : est un paramètre qui capte la tendance des règlements par rapport aux années de survenance ;
- β_j : est un paramètre qui capte la tendance des règlements par rapport aux années de développement.

La fonction lien

La fonction lien établit un pont entre les composantes aléatoires et systématiques sous forme de la donnée d'une fonction réelle g strictement monotone et dérivable telle que :

$$\eta_{i,j} = g(\mu_{i,j}) \quad \text{ou} \quad \mu_{i,j} = g^{-1}(\eta_{i,j}) \quad i, j = 0, \dots, n-1$$

Les fonctions standards sont les fonctions :

- La fonction identité : $\eta_{i,j} = \mu_{i,j}$, pour laquelle les effets des facteurs sont additifs ;
- La fonction logarithmique : $\eta_{i,j} = \ln(\mu_{i,j})$ ou $\mu_{i,j} = e^{\eta_{i,j}}$, elle représente un avantage parce qu'elle introduit un effet multiplicatif.

Modélisation par la loi poisson sur-dispersée

Les règlements $Y_{i,j}$ suivent la loi de poisson sur-dispersée de paramètres $(\mu_{i,j}, \varphi)$, la densité de probabilité de $Y_{i,j}$ en $y_{i,j}$ est :

$$\begin{cases} \mathbb{P}(Y_{i,j} = y_{i,j}) = \frac{(\mu_{i,j}/\varphi)^{y_{i,j}/\varphi} \exp(-\mu_{i,j}/\varphi)}{(y_{i,j}/\varphi)!} \\ E[Y] = \mu, & \text{Var}[X] = \varphi E[X] \end{cases}$$

L'estimation des paramètres se fait par maximum de vraisemblance, la fonction de vraisemblance s'écrit :

$$\begin{aligned} L(Y_{i,j}, \mu_{i,j}) &= \log \prod_{i+j \leq n} \frac{(\mu_{i,j}/\varphi)^{y_{i,j}/\varphi} \exp(-\mu_{i,j}/\varphi)}{(y_{i,j}/\varphi)!} \\ &= \frac{1}{\varphi} \sum_{i+j \leq n} (y_{i,j}(\mu + \alpha_i + \beta_j) - \exp(\mu + \alpha_i + \beta_j)) \end{aligned}$$

Avec $f(y_{i,j}, \varphi)$ qui regroupe tous les termes qui ne dépendent que de $y_{i,j}$ et de φ .

Remarque : $f(y_{i,j}, \varphi)$ et φ ne jouent aucun rôle dans la maximisation de la vraisemblance du paramètre $\theta = \mu_{i,j}$.

Par conséquent l'estimateur du maximum de vraisemblance du paramètre θ se fait à partir du programme suivant :

$$\hat{\theta} = \underset{\mu, \alpha_i, \beta_j}{\operatorname{argmax}} \hat{l}(Y_{i,j}, \mu_{i,j}) = \sum_{i+j \leq n} (y_{i,j}(\mu + \alpha_i + \beta_j) - \exp(\mu + \alpha_i + \beta_j))$$

Ainsi chaque paramètre est obtenu par dérivation par rapport à chaque composante. On obtient le système d'équation suivant :

$$\frac{\partial \hat{\theta}}{\partial \mu} = \sum_{i+j \leq n} y_{i,j} - \sum_{i+j \leq n} \exp(\mu + \alpha_i + \beta_j) = 0$$

$$\frac{\partial \hat{\theta}}{\partial \alpha_i} = \sum_{j \leq n-i} y_{i,j} - \sum_{j \leq n-i} \exp(\mu + \alpha_i + \beta_j) = 0$$

$$\frac{\partial \hat{\theta}}{\partial \beta_j} = \sum_{i \leq n-j} y_{i,j} - \sum_{i \leq n-j} \exp(\mu + \alpha_i + \beta_j) = 0$$

Le système d'équation étant complexe pour être résolu analytiquement, le recours est fait à la méthode de Newton Raphson.

b- Cadre pratique

Avant de faire une estimation de la réserve totale en branche Auto corporelle et de l'erreur commise par le modèle GLM Poisson, nous allons procéder à la validation de ce dernier et à l'analyse des résidus de Pearson.

Validation du modèle

Nous allons appliquer le modèle GLM Poisson à notre triangle de développement provenant de la branche Auto corporelle. Nous commençons d'abord par la validation de ce dernier et l'analyse des résidus.

Nous avons obtenu la valeur estimée pour chaque paramètre du modèle GLM et leur p-value grâce à la fonction *glm* du logiciel R afin de tester l'estimation de ces paramètres avec un seuil de significativité de 5%.

Tableau 23 : Résultats de l'estimation des paramètres du modèle GLM Poisson

Paramètre	μ	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8
Estimation	7,73	1,95	2,64	2,10	1,83	1,21	1,49	0,55	1,21
p-value	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16	<2e-16

Paramètre	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Estimation	0,43	-0,22	0,09	0,61	0,27	-0,05	1,57	1,25
p-value	<2e-16							

Nous pouvons constater que tous les p-value du modèle GLM Poisson de nos paramètres sont inférieure au seuil de significativité. Par conséquent, nos paramètres sont bien pertinents.

Analyse des résidus

L'analyse des résidus permet de mesurer la qualité d'ajustement d'un modèle. Dans le cadre d'un GLM, on utilise les résidus de Pearson parce qu'ils sont plus pertinents. L'analyse des résidus est indispensable pour tester l'indépendance des règlements par année de survenance, par année de développement et par rapport aux valeurs estimées.

Définition des résidus de Pearson

Le résidu de Pearson de la cellule (i, j) est défini par :

$$r_{i,j}^P = \frac{y_{i,j} - \mu_{i,j}}{\sqrt{V(\mu_{i,j})}}, \quad \forall i + j \leq n$$

Toutefois, en ajustant ces résidus afin d'avoir une variance unitaire, on considère classiquement :

$$r_{i,j}^P = \sqrt{\frac{n}{n-k}} \times \frac{X_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\text{Var}(\hat{\mu}_{i,j})}}, \quad \text{où } k \text{ est le nombre de paramètres estimés dans le modèle.}$$

Tableau 24 : Résidus du modèle GLM Poisson sur-dispersé

	1	2	3	4	5	6	7	8	9
1	134	-21,6	65,49	-40,4	-18,4	-18,6	-18,35	-66,4	-3,18E-13
2	44,3	-8,03	-116,08	139,36	-23,1	7,32	-6,28	53,5	
3	12,6	101,15	-79,14	9,39	31,8	-83,26	29,23		
4	74,7	9,57	-119,41	101,77	-36,5	80,17			
5	-18	-73,92	157,34	-167,6	41,8				
6	-40,2	20,35	9,28	-16,99					
7	-43,4	-55,54	51,01						
8	-75,3	28,35							
9	-7,14E-13								

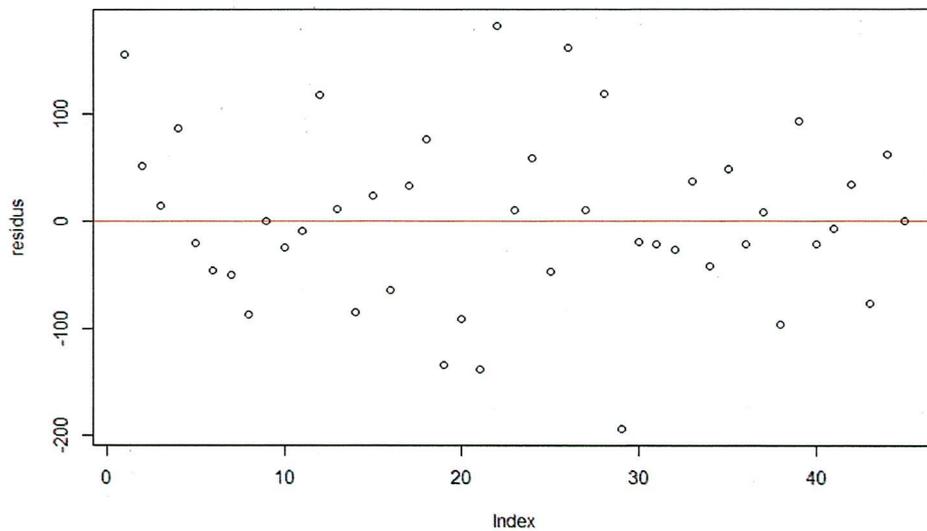


Figure 12 : Représentation des résidus de Pearson GLM Poisson

Le graphique des résidus ci-dessus présente une structure aléatoire. Nous pouvons donc conclure à une indépendance des résidus de Pearson.

Normalité des résidus

Nous vérifions cette hypothèse par le test de Shapiro-Wild et le QQ-plot des résidus.

Tableau 25 : Résultat du test de Shapiro-Wild des résidus GLM Poisson

Test	Statistique	p-value
Shapiro-Wild	0,9872	0,895

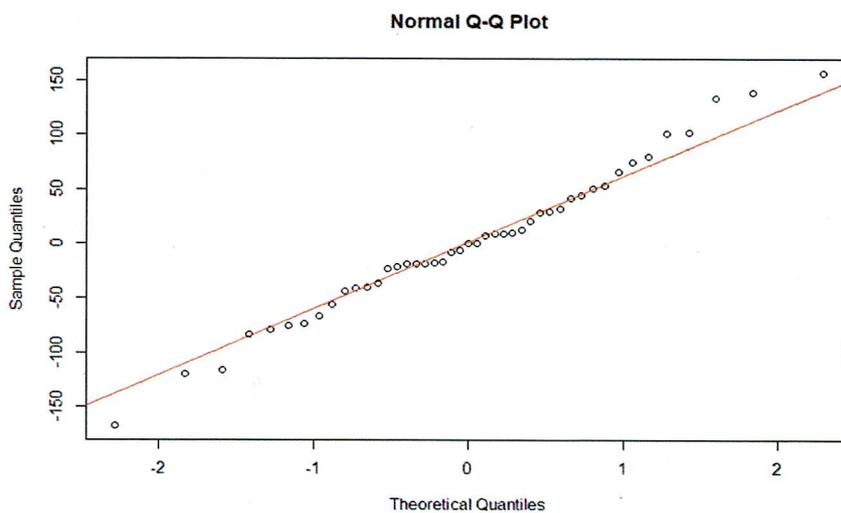


Figure 13 : Q-Q plot des Résidus GLM Poisson vs Distribution normale

De ce qui précède, Nous pouvons conclure que les résidus de Pearson sont distribués suivant une loi normale.

Définition des provisions et estimation du risque d'erreur

Les provisions pour l'année de survenance $i = 0, \dots, n-1$ sont :

$$E(R_i) = \sum_{j=n-i}^{n-1} E(y_{i,j}) = \sum_{j=n-i}^{n-1} \mu_{i,j}$$

$$V(R_i) = \sum_{j=n-i}^{n-1} V(y_{i,j}) = \varphi \sum_{j=n-i}^{n-1} V(\mu_{i,j})$$

$$\mu_3(R_i) = \sum_{j=n-i}^{n-1} \mu_3(y_{i,j}) = \varphi^2 \sum_{j=n-i}^{n-1} V(\mu_{i,j})V'(\mu_{i,j})$$

Les provisions sont estimées par maximum de vraisemblance

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

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

$$\hat{\mu}_3(R_i) = \hat{\varphi}^2 \sum_{j=n-i}^{n-1} V(\hat{\mu}_{i,j})V'(\hat{\mu}_{i,j})$$

Ainsi on déduit l'estimateur de la provision totale par :

$$\hat{E}(R) = \sum_{i=0}^{n-1} \hat{E}(R_i); \quad \hat{V}(R) = \sum_{i=0}^{n-1} \hat{V}(R_i); \quad \hat{\mu}_3(R) = \sum_{i=0}^{n-1} \hat{\mu}_3(R_i)$$

Estimation du risque d'erreur

Dans une régression GLM avec un lien logarithmique par exemple, nous avons (Charpentier et Dutang, 2012) :

$$E(Y_{i,j} | \mathcal{F}_n) = \mu_{i,j} = \exp(\eta_{i,j})$$

ou encore

$$\hat{Y}_{i,j} = \hat{\mu}_{i,j} = \exp(\hat{\eta}_{i,j}).$$

La *delta method* nous permet d'écrire que :

$$V(\hat{Y}_{i,j}) \approx \left| \frac{\partial \mu_{i,j}}{\partial \eta_{i,j}} \right|^2 V(\hat{\eta}_{i,j}),$$

ce qui se simplifie dans le cas où le lien est logarithmique, c'est-à-dire :

$$\frac{\partial \mu_{i,j}}{\partial \eta_{i,j}} = \mu_{i,j}$$

Aussi, pour une loi de Poisson sur-dispersée comme dans (Renshaw et Verrall, 1998),

$$\mathbb{E} \left([Y_{i,j} - \hat{Y}_{i,j}]^2 \right) \approx \hat{\phi} \hat{\mu}_{i,j} + \hat{\mu}_{i,j}^2 \hat{V}(\hat{\eta}_{i,j})$$

pour la partie inférieure du triangle.

De plus, car il sera nécessaire de sommer tous les termes de la partie inférieure du triangle pour déterminer le montant total de provisions,

$$\text{Cov}(\hat{Y}_{i,j}, \hat{Y}_{k,l}) \approx \hat{\mu}_{i,j} \hat{\mu}_{k,l} \text{Cov}(\hat{\eta}_{i,j}, \hat{\eta}_{k,l}).$$

Le montant de provision que l'on cherche à estimer étant la somme des prédictions de paiements à venir,

$$\hat{R} = \sum_{i+j>n} \hat{Y}_{i,j},$$

alors

$$\mathbb{E} \left([R - \hat{R}]^2 \right) \approx \left(\sum_{i+j>n} \hat{\phi} \hat{\mu}_{i,j} \right) + \hat{\mu}'_F \hat{V}(\hat{\eta}_F) \hat{\mu}_F,$$

où les vecteurs $\hat{\mu}_F$ et $\hat{\eta}_F$ sont des restrictions des vecteurs $\hat{\mu}$ et $\hat{\eta}$ aux indices $i + j > n$ (c'est-à-dire à la partie inférieure du triangle à prédire).

En appliquant donc tout ceci à notre triangle de règlements, nous obtenons les résultats résumés dans le tableau suivant :

Tableau 26 : Provisions par la méthode GLM Poisson

	0	1	2	3	4	5	6	7	8	provision	% prov totale
2013	2266	15963	31725	18496	14109	7612	10057	3948	7589	0	0,00%
2014	3487	24559	48807	28455	21706	11711	15473	6073	11674	11 674	1,12%
2015	1812	12763	25365	14788	11280	6086	8041	3156	6067	9 223	0,88%
2016	2485	17506	34792	20284	15473	8348	11030	4329	8322	23 681	2,26%
2017	4191	29518	58664	34202	26089	14076	18597	7299	14032	54 004	5,16%
2018	2996	21105	41945	24454	18654	10064	13297	5219	10033	57 267	5,47%
2019	2150	15146	30101	17549	13387	7223	9542	3745	7200	58 646	5,60%
2020	10944	77084	153196	89315	68130	36758	48565	19062	36644	451 670	43,15%
2021	7879	55494	110288	64299	49048	26463	34963	13723	26380	380 658	36,36%
Total										1 046 823	100%

Nous retrouvons bien le fait que le modèle GLM Poisson donne les mêmes montants des réserves que la méthode de Chain-Ladder, avec un écart-type de l'erreur de prédiction de la provision

$$\widehat{sep}(\hat{R}) = 69\,898 \text{ et un coefficient de variation } CV = \frac{\widehat{sep}(\hat{R})}{\hat{R}} \approx 7\%.$$

Et pour le quasi-poisson, l'écart-type de l'erreur de prédiction de la provision est

$$\widehat{sep}(\hat{R}) = 401\,100 \text{ et un coefficient de variation } CV = \frac{\widehat{sep}(\hat{R})}{\hat{R}} = 38\%.$$

Intervalle de confiance

Après estimation des provisions et de l'erreur standard de prédiction nous déterminons un intervalle de confiance des provisions à 95 % :

$$IC = [\hat{R} - 1,96 \times \widehat{sep}(\hat{R}); \hat{R} + 1,96 \times \widehat{sep}(\hat{R})]$$

Nous obtenons ainsi, pour Poisson : $IC_p = [909\ 823; 1\ 183\ 823]$

Pour quasi-poisson : $IC_q = [260\ 667; 1\ 832\ 979]$

Distribution prédictive

Pour l'estimation de la loi de la provision globale R (dite loi prédictive), nous aborderons les approches :

- par convolution ;
- par des techniques de Bootstrap.

5- Distributions prédictives par convolution

a- Présentation

Lorsque les données (paiements non cumulés) du triangle supérieur $(x_{i,j})_{i+j \leq n}$, sont des réalisations de variables aléatoires $(X_{i,j})_{i+j \leq n}$ considérées indépendantes, si la loi des $X_{i,j}$ est additive (par exemple : Poisson, poisson sur-dispersée, Normale), la distribution prédictive de R peut s'obtenir par convolution directe. (Djagana, 2020)

Cas de la loi de Poisson : $X_{i,j} \sim P(\mu_{i,j})$

Dans ce cas :

$$R = \sum_{i=1}^n \sum_{j=n-i+1}^n X_{i,j} \sim P(\mu_R)$$

Avec

$$\mu_R = \sum_i \sum_j \mu_{i,j}$$

Si

$$\mu_R \geq 50,$$

Alors

$$R \sim \text{approx}N(\mu_R, \mu_R)$$

Et donc

$$P(R \leq \mu_R) \approx \Phi\left(\frac{1-\mu_R}{\sqrt{\mu_R}}\right)$$

Et le quantile d'ordre $1 - \eta$ de R est :

$$\hat{q}_{1-\eta}(R) \approx \hat{q}_{1-\eta}^{(P)}(R) = \hat{\mu}_R + \sqrt{\hat{\mu}_R} q_{1-\eta}$$

Avec $q_{1-\eta}$ le quantile d'ordre $1 - \eta$ d'une loi normale centrée réduite.

Il suffit d'estimer $\mu_{i,j}$ et μ_R par maximum de vraisemblance pour obtenir un estimateur de maximum de vraisemblance de $q_{1-\eta}^{(P)}(R)$.

b- Cadre pratique

Nous appliquons la méthode à notre triangle des règlements de la branche auto corporelle. La validité du modèle a été déjà fait au niveau de la méthode GLM Poisson.

Les résultats obtenus aux quantiles d'ordres respectifs : 50%, 60%, 75%, 95%, 99% et 99,5% sont résumés dans le tableau suivant.

Tableau 27 : Provision obtenue par distribution prédictive par convolution

Quantile (VaR)	50%	60%	75%	95%	99%	99,5%
Provision	1 046 824	1 047 111	1 047 600	1 048 404	1 049 202	1 049 461

c- Limite de la distribution prédictive par convolution

Comme toute approximation normale, cette approximation sous-estime la queue de distribution et donc les quantiles.

6- Bootstrap GLM

a- Présentation

La méthode du Bootstrap adaptée au provisionnement nous permet d'estimer la distribution complète du montant des provisions. Il est ainsi possible d'estimer les moments d'ordre quelconque de la distribution de la provision. On l'utilise généralement avec le modèle de Poisson, ce qui permet de « bootstrapper » avec la méthode Chain Ladder donnant des résultats identiques. (England et Verrall, 2002)

Cette méthode suit le principe de rééchantillonnage. Soit (X_1, \dots, X_n) un échantillon initial *iid* de variables aléatoires réelles et de fonction de densité f inconnue. On note θ le paramètre d'intérêt avec $\hat{\theta} = f(X_1, \dots, X_n)$ l'estimation de θ à partir de l'échantillon initial.

Un échantillon « Bootstrap » est le tirage avec remise de n éléments dans l'échantillon initial avec la même probabilité. Notons l'échantillon « Bootstrap » (X_1^*, \dots, X_n^*) et $\hat{\theta}^* =$

$f(X_1^*, \dots, X_n^*)$. Le Bootstrap correspond à la répétition de ce principe N fois afin d'obtenir N échantillons.

La distribution de l'estimateur $\hat{\theta}$ est approchée par la distribution empirique calculée sur les N réalisations bootstrappées.

Dans les modèles de régression en provisionnement, les variables de réponses ne sont généralement pas *iid*. C'est la raison pour laquelle on utilise les résidus des modèles. Dans la pratique on utilise fréquemment les résidus de Pearson :

$$r_{i,j}^p = \frac{X_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\text{Var}(\hat{\mu}_{i,j})}}$$

En ajustant ces résidus afin d'avoir une variance unitaire, on a :

$$r_{i,j}^p = \sqrt{\frac{n}{n-k}} \times \frac{X_{i,j} - \hat{\mu}_{i,j}}{\sqrt{\text{Var}(\hat{\mu}_{i,j})}}, \text{ où } k \text{ est le nombre de paramètres estimés dans le modèle.}$$

L'implémentation de la méthode du Bootstrap sur un triangle est décrit par les étapes suivantes :

1. Obtention des $\hat{\mu}_{i,j}$ via l'utilisation d'un modèle appliqué au triangle ;
2. Calculs des résidus $r_{i,j}^p$;
3. Rééchantillonnage des $r_{i,j}^p$ afin d'avoir un « nouveau » triangle supérieur de résidus ;
4. Inversion de la formule des résidus afin d'obtenir le « nouveau » triangle supérieur ;
5. Développement du triangle inférieur via le modèle sélectionné ;
6. Tirage des incréments d'espérance $\hat{\mu}_{i,j}$ et de variance $\varphi \times \text{Var}(\hat{\mu}_{i,j})$;
7. Calcul de la valeur de R dans cette simulation ;
8. On recommence à l'étape 3.

Ainsi, l'on obtient une distribution de la valeur de la provision. A l'aide de cette distribution on connaît ainsi les différents moments. Nous sommes capables d'estimer différents quantiles et en particulier la $\text{MSEP}(R)$.

Remarque :

La version standard du Bootstrap GLM ne comprend pas l'étape 6 décrite ci-dessus. En version standard, la méthode du Bootstrap permet d'estimer seulement l'erreur d'estimation. En matière de solvabilité, on s'intéresse à la

$$\begin{aligned} \text{MSEP}(R) &= \text{Var}(R) + E \left[\left(\hat{R} - E(R) \right)^2 \right] \\ &= \text{Erreur de processus} + \text{Erreur d'estimation} \end{aligned}$$

La technique usuelle de Bootstrap GLM permet d'estimer le terme « d'erreur d'estimation », il faut ainsi lui rajouter l'erreur de processus. L'étape 6 consiste à estimer l'erreur de processus.

b- Cadre pratique

Nous allons appliquer la méthode Bootstrap GLM à notre triangle des règlements de la branche Auto corporelle.

Analyse des hypothèses du Bootstrap

Les hypothèses relatives aux modèles stochastiques de provisionnement de Poisson sur-dispersé ont été testées au cours de l'application numérique du GLM Poisson. Nous allons à présent nous concentrer sur les hypothèses du Bootstrap.

Tableau 28 : Résidus de Pearson GLM Poisson

	1	2	3	4	5	6	7	8	9
1	134	-21,6	65,49	-40,4	-18,4	-18,6	-18,35	-66,4	-3,18E-13
2	44,3	-8,03	-116,08	139,36	-23,1	7,32	-6,28	53,5	
3	12,6	101,15	-79,14	9,39	31,8	-83,26	29,23		
4	74,7	9,57	-119,41	101,77	-36,5	80,17			
5	-18	-73,92	157,34	-167,6	41,8				
6	-40,2	20,35	9,28	-16,99					
7	-43,4	-55,54	51,01						
8	-75,3	28,35							
9	-7,14E-13								

Simulation des résidus

Nous allons bootstrapper les résidus parmi les pseudo-résidus obtenus. Puis utiliser un modèle paramétrique. Nous supposons que la distribution des résidus de Pearson suit une loi normale, puis nous représentons l'histogramme, la fonction densité et de répartition des résidus avec un ajustement Gaussien. (Charpentier et Dutang, 2005)

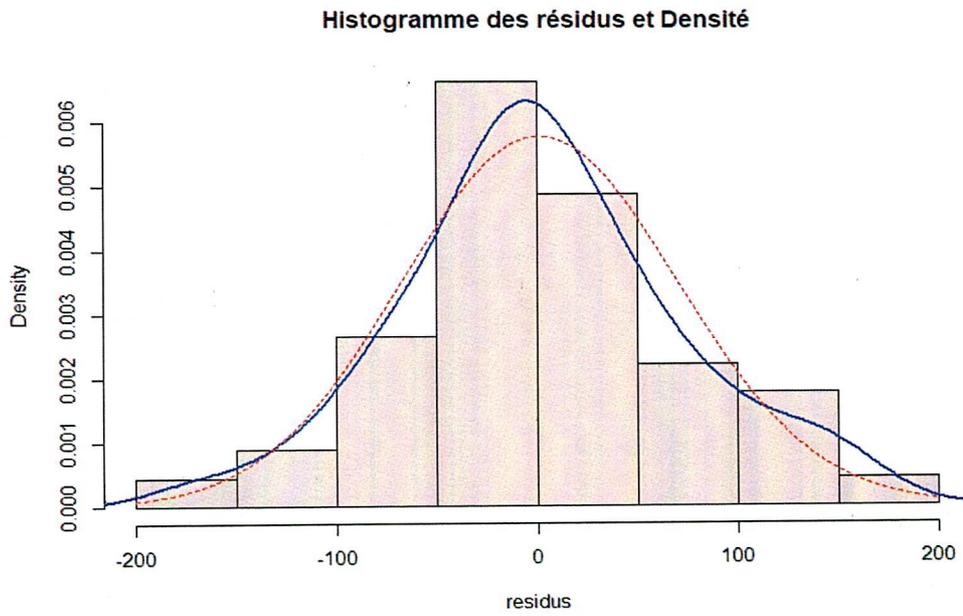


Figure 14 : Histogramme et densité des résidus avec l'ajustement Gaussien en pointillés (auto corporelle)

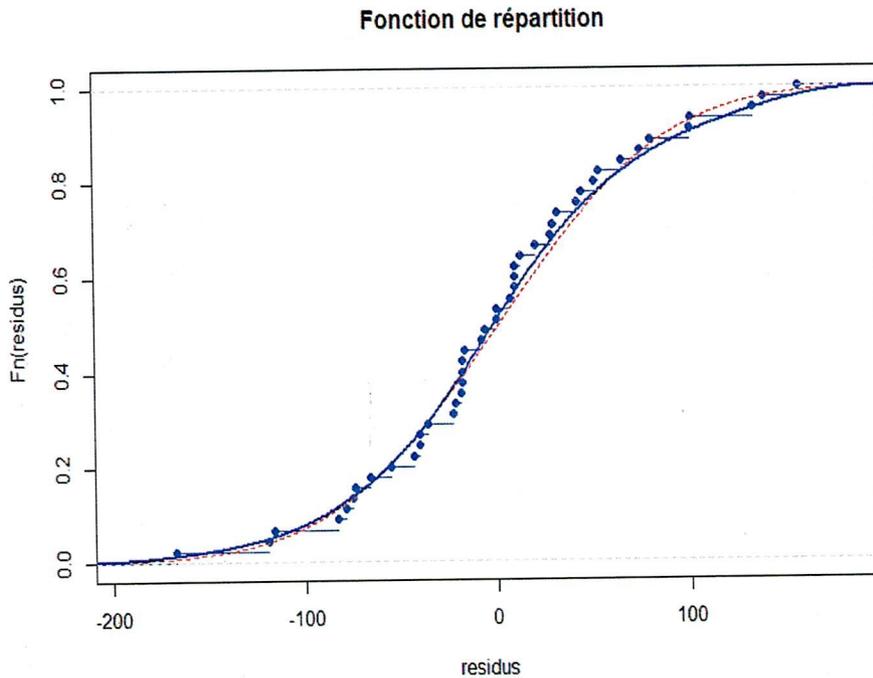


Figure 15 : Fonctions de répartition avec l'ajustement Gaussien en pointillés (auto corporelle)

Une fois simulé un pseudo-triangle d'incrément de paiements, on prédit un montant de provision \hat{R}^B (par exemple via une méthode Chain Ladder). La variance des \hat{R}^B correspond à l'erreur d'estimation.

Afin de prendre en compte l'erreur de modèle, plusieurs méthodes peuvent être utilisées. L'une des plus simples, consiste à noter qu'à partir du pseudo triangle $Y_{i,j}^B$, nous pouvons obtenir des prédictions pour la partie inférieure, $\hat{Y}_{i,j}^B$. Compte tenu du modèle Poissonnier, on peut alors simuler une trajectoire possible d'incrément de paiements en simulant les $Y_{i,j}^B$ à l'aide de loi de Poisson de paramètre $\hat{Y}_{i,j}^B$.

Toutefois, simuler des lois de Poisson risque d'être trop conservateur. En effet, comme nous l'avons vu sur la régression quasi Poisson, le paramètre de sur-dispersion φ est significatif,

$$\varphi = 6\ 180.$$

Il peut alors être pertinent de générer des lois avec davantage de variance (abusivement, on parlera de simulation d'une loi quasi Poisson). La première idée pourra être d'utiliser une loi Gamma. En notant que :

$$E(Y) = \lambda \text{ et } V(Y) = \varphi\lambda,$$

la loi de Gamma $\mathcal{G}(\alpha, \beta)$ vérifiera :

$$\alpha\beta = \lambda \text{ et } \alpha\beta^2 = \varphi\lambda.$$

Une autre idée peut être d'utiliser le lien qui existe entre la loi de Poisson et la loi binomiale négative (qui est un mélange de lois de Poisson, c'est-à-dire dont l'hétérogénéité résiduelle n'a pas pu être modélisée par nos facteurs lignes et colonnes). Pour une loi binomiale négative de moyenne μ et de variance $\mu + \frac{\mu^2}{k}$, on pose $\mu = \lambda$ et $k = \lambda(\varphi\lambda - 1)^{-1}$.

On a donc généré des scénarii de paiements avec le logiciel R.

Nous avons simulé 20 000 triangles et obtenu la distribution des montants de provisions estimées, mais aussi des scénarii de paiements (et donc de provisions nécessaires).

Nous résumons les résultats obtenus dans le tableau ci-après :

Tableau 29 : Provision obtenue par Bootstrap GLM Poisson sur-dispersé (auto corporelle)

Quantile	50%	60%	75%	95%	99%	99,5%
Provision	938 931	1 076 448	1 202 882	1 627 453	2 064 352	2 164 579

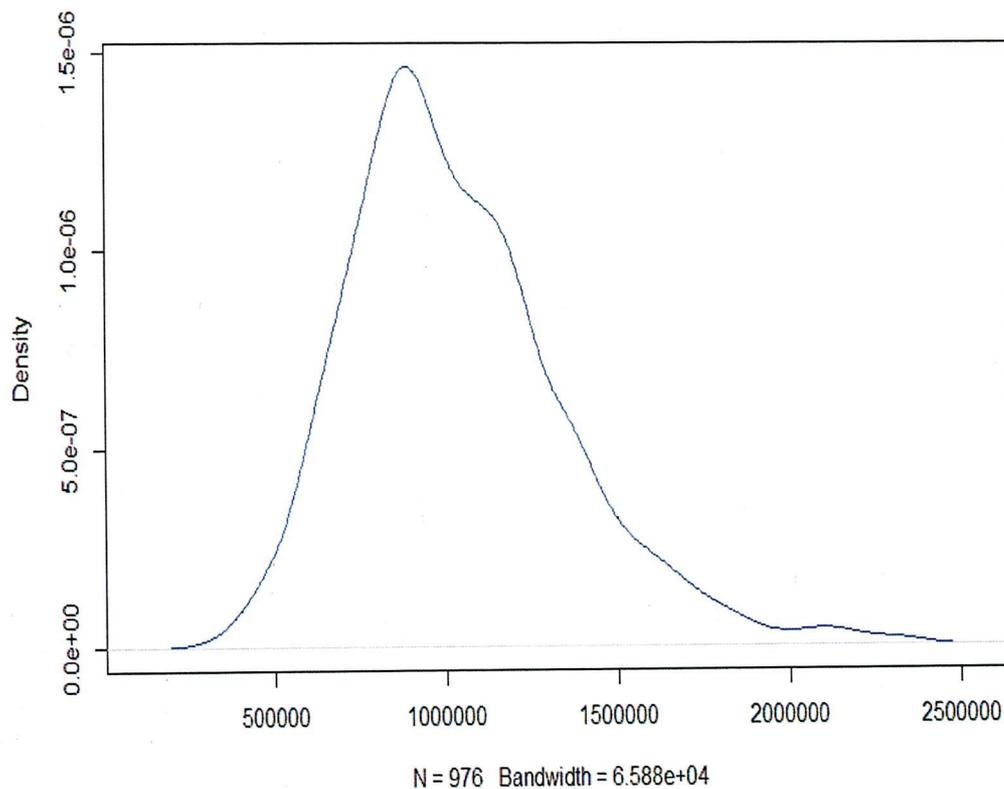


Figure 16 : Représentation de la distribution des provisions (auto corporelle)

c- Limites de l'approche Bootstrap

- Le Bootstrap ne s'applique que sur les résidus *i. i. d* et cette hypothèse doit être vérifiée.
- Le Bootstrap n'est pas robuste pour l'estimation de quantiles élevés dans la mesure où souvent les données que l'on possède ne reflètent pas les observations concernant ces quantiles.

Conclusion

Les méthodes déterministes permettent d'estimer le montant moyen que la compagnie d'assurance devra payer dans le futur pour les sinistres survenus. Cependant, l'assureur n'a aucune information quant à la volatilité de cette estimation. Pour pallier à cette limite, l'assureur peut faire recours à des méthodes stochastiques et ainsi avoir l'erreur liée à la volatilité de ses réserves à l'ultime.

Face à cette pléthore de méthodes, il faut bien que l'assureur face un choix qui se veut objectif. Le prochain chapitre traite de la comparaison et du choix de méthodes.

Chapitre 3 : Analyse des Résultats

Introduction

L'implémentation des différentes méthodes a été faite sur quatre branches d'activité : Auto corporelle, Auto matérielle, Incendie et Santé. Il est question de faire deux analyses comparées, une sur les sous branches Auto corporelle et Auto matérielle de la branche Automobile question de justifier de leur subdivision. Et une deuxième analyse sur trois branches à déroulé différent pour comprendre l'impact de la profondeur d'un triangle.

I- Choix de la méthode

Il s'agit pour l'actuaire de choisir parmi les diverses méthodes à disposition, sachant qu'il n'existe pas de méthode standard. En effet, le choix doit tenir compte de nombreux paramètres, notamment du risque en question, des contraintes réglementaires, de la politique de la compagnie, de la branche d'assurance étudiée, etc.

Nous avons jusque-là évoqué et appliqué plusieurs méthodes de provisionnement stochastiques et déterministes, mais sur un seul et même triangle. Nous avons remarqué qu'elles n'aboutissent pas aux mêmes montants de réserves finales. Nous avons choisi d'orienter cette comparaison sur la particularité de la nature du triangle. Dans la suite, nous avons donc voulu étudier si ces divergences persistaient d'une branche d'assurance à l'autre ; ce qui nous a menés à appliquer ces méthodes sur des triangles provenant de secteurs d'assurance différents.

II- Comparaison des méthodes, analyse des résultats et perspectives

Nous mettons en exergue les résultats de trois branches dont une subdivisée en deux sous branches. Il s'agit de la branche Automobile (subdivisée en Auto corporelle et Auto matérielle), des branches Santé et Incendie.

1- Analyse des résultats des sous branches d'une branche

Nous comparons dans cette sous-section les résultats obtenus des branches (sous branches) Auto corporelle et Auto matérielle. L'objectif étant d'apprécier la nécessité de leur séparation. Le tableau ci-dessous résume les différents résultats obtenus par cinq méthodes en Auto corporelle et en Auto matérielle, entre parenthèses l'écart relatif par rapport à la méthode de Chain-Ladder.

Tableau 30 : Réserves obtenues pour les sous branches automobile par différents modèles

Modèle \ Branche	Auto corporelle	Auto matérielle
Chain-Ladder / Mack	1 046 823	533 295
Chain-Ladder pondéré (Min/Max)	1 221 498 (+16,69%)	543 100 (+1,84%)
London-Chain	484 963 (-53,67%)	883 211 (+65,61%)
Après Prise en compte du tail factor	1 101 558 (+5,23%)	534 686 (+0,26)
Log normale	941 689 (-10,04%)	608 338 (+14,07%)

Les impressions que nous avons au regard de ce tableau sont :

a- Les deux sous branches de la branche Automobile ont des comportements totalement différents.

Car, dans la classe de modèles choisis, celui de Chain-Ladder en Auto matérielle donne le montant de réserves le moins élevé. Ceci n'est pas le cas pour la sous branche Auto corporelle puisque Chain-Ladder ne donne le montant de réserves le moins élevé que par rapport au Chain-Ladder pondéré et après la prise en compte du tail factor, mais par rapport aux modèles London-Chain et log-normale, Chain-Ladder est bien plus prudent. Cette nette différence de comportement nous permet de comprendre pourquoi la branche Automobile a été subdivisée en deux sous branches. Ceci est conforté par la figure 17 ci-après de cadences de règlements qui sont totalement différents d'une branche à l'autre.

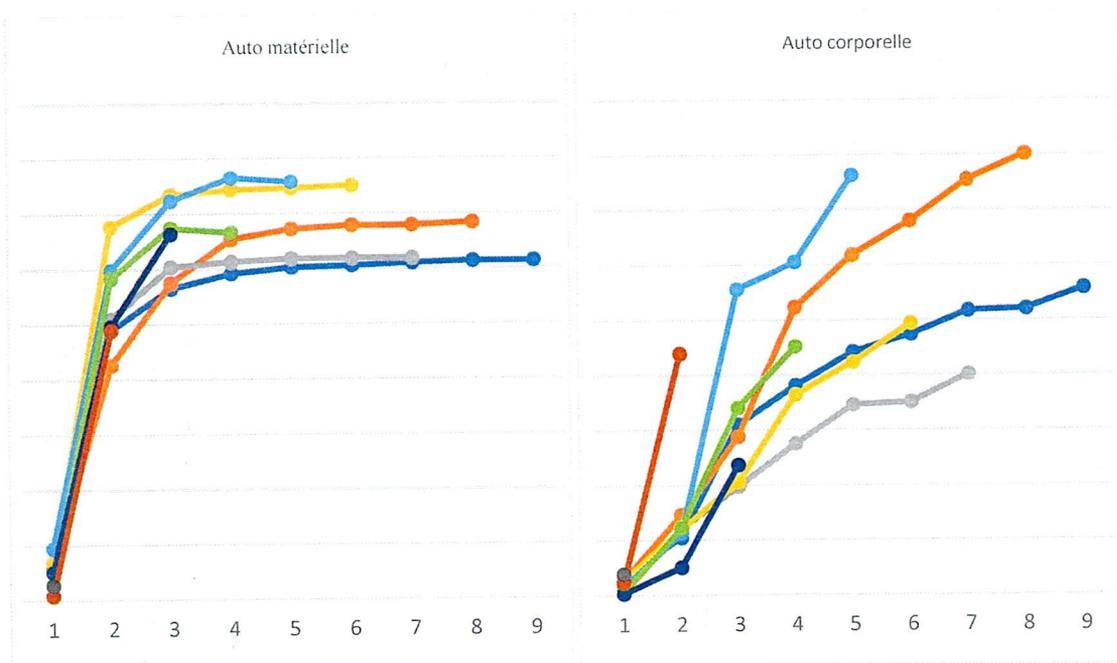


Figure 17 : Cadences de règlements Auto matérielle vs Auto corporelle

b- L'adéquation des données à un modèle avant l'application de ce dernier à une branche est primordiale.

L'illustration est faite par l'écart relatif assez considérable qui existe dans les deux cas entre les modèles Chain-Ladder et London-Chain, tout simplement parce que les paramètres additifs du modèle London-Chain dans les deux cas ne sont pas significatifs (cf. tableaux 11 et 36_AI), par conséquent London-Chain n'est pas adéquat pour l'estimation des réserves en Auto corporelle et en Auto matérielle.

c- La branche Auto corporelle est à déroulé long et Auto matérielle à déroulé moyen.

En fait, en se basant sur notre modèle de référence Chain-Ladder, la provision après prise en compte d'un tail factor donne un écart relatif de (+5,23%) pour la branche (sous branche) Auto corporelle. Ceci montre que l'information n'est pas totalement captée en Auto corporelle. De plus, le triangle des règlements des incréments Auto corporelle ne possède aucune valeur nulle (cf. tableau 17). Par conséquent, l'assureur gagnerait à revoir la profondeur de son triangle pour cette branche d'activité.

Par contre, en Auto matérielle, l'écart relatif de la provision après prise en compte d'un tail factor par rapport à Chain-Ladder est de (+0,26%). Ce qui supposerait que l'information est totalement captée après 9 ans. Ceci peut d'ailleurs être confirmé par la figure 17 (A) où il est constaté une stabilité des règlements vers la fin du déroulé.

2- Analyse comparée des résultats des branches : Auto corporelle, Incendie et Santé

Le tableau ci-après confronte les résultats obtenus en Auto corporelle, Incendie et Santé.

Nous voulons dans cette analyse comparée évaluer les différents modèles étudiés par rapport à la durée de développement (déroulé court, moyen et long) des branches d'assurance dommage.

Tableau 31 : Réserves obtenues pour les branches auto corporelle, incendie et santé par les différents modèles

Modèle ↓ \ branche →	Auto corporelle	Incendie	Santé
Chain-Ladder / Mack	1 046 823	414 280	4 529 640
London-Chain	484 963 (-53,67%)	629 261 (+51,89%)	3 381 353 (-25,35%)
Tail factor	1 101 558 (+5,23%)	422 645 (+2,02%)	4 535 107 (+0,12%)
Bootstrap (moyenne)	1173 820 (+12,13%)	439 799 (+6,16%)	4 614 784 (+1,88%)
Convolution ($Var_{99,5\%}$)	1 049 461 (+0,25%)	415 772 (0,36%)	4 534 211 (0,10%)

Le tableau 31 ci-dessus donne de noter que :

a- La branche Santé est à déroulé court, Incendie à déroulé moyen et Auto corporelle à déroulé long.

Ceci est vu à partir de la ligne 4 des résultats après prise en compte du tail factor. La confirmation est faite par la figure 18 d'évolution des cadences des règlements pour chacune de ces branches. En Santé, les règlements n'évoluent plus à partir de la 5^e année de développement ; en Incendie, les règlements évoluent jusqu'à la fin du déroulé mais très peu vers les dernières années de développement et enfin en Auto corporelle, aucune stabilité en vue durant tout le déroulé.

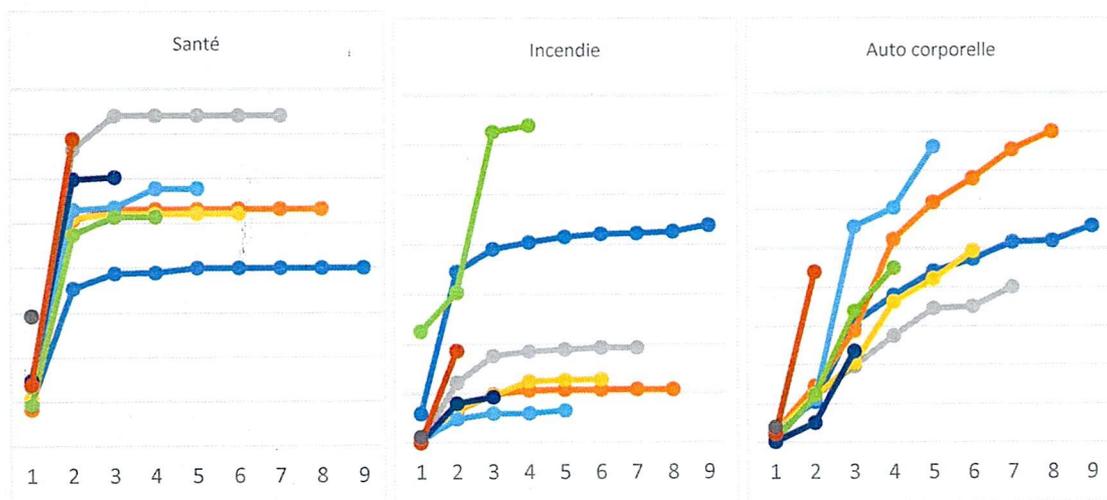


Figure 18 : Analyse comparée des cadences des règlements en Santé, Incendie et Auto corporelle

b- Les branches à déroulé court ne sont presque pas exposées au risque de modèle.

Car dans notre classe de modèles, la variation des réserves est négligeable en Santé. Les résultats donnés par différents modèles dans cette branche sont presque identiques. A part London-Chain qui n'est pas en adéquation avec les données d'étude.

c- Le choix du modèle impacte moyennement la provision pour les branches à déroulé moyen.

Dans notre branche Incendie par exemple, les écarts relatifs des résultats des différentes méthodes par rapport à Chain-Ladder varient entre 0,36% et 6,16% en ne tenant pas compte du modèle de London-Chain qui n'est pas en adéquation avec nos données. Ainsi, l'assureur bien qu'ayant une bonne profondeur de triangle pour cette branche, devrait tout de même challenger de temps à autre son modèle interne (Chain-Ladder) à d'autres modèles, notamment stochastiques, en vue de prendre en compte l'erreur d'estimation, pour qu'il ne soit pas surpris en cas d'aggravation de la sinistralité dans une telle branche.

d- Les branches à déroulé long sont très sensibles au choix du modèle.

Car nous constatons au regard des tableaux 30 et 31 que les variations des réserves sont assez poussées d'un modèle à l'autre. Une telle branche donne ainsi un sens à l'étude du risque de modèle et à la nécessité de bien maîtriser ce dernier.

Conclusion et perspectives

Il était question dans ce chapitre de faire une synthèse des résultats d'une classe de modèles choisie et adopter la méthode la plus adéquate selon les branches pour le calcul de la provision.

Il en ressort que, le choix du modèle pour l'évaluation de la provision des branches à déroulé court ne se pose pas. Il suffit que les données de la branche soient en adéquation avec le modèle choisi. Par conséquent un simple modèle déterministe suffit pour avoir la provision d'une telle branche. La provision de l'assureur en branche Santé est donc optimale.

Pour les branches à déroulé moyen, l'assureur bien qu'ayant une bonne profondeur de triangle pour de telles branches, est privé d'une information cruciale sur la volatilité (variabilité) de ses provisions en utilisant un modèle déterministe. Par conséquent, il gagnerait à utiliser des modèles stochastiques pour de telles branches afin, non seulement d'avoir une estimation moyenne de sa provision, mais de bénéficier de l'information que des modèles stochastiques fournissent sur l'erreur d'estimation. Ainsi, pour des branches Incendie et Auto matérielle, l'assureur a certes des provisions acceptables, mais est légèrement exposé en cas de sinistralité exceptionnelle.

Enfin, pour les branches à déroulé long, nous avons vu qu'elles étaient assez sensibles au choix du modèle. Par conséquent, l'assureur ne devrait pas se contenter d'une simple évaluation moyenne de la provision et de l'information sur l'erreur d'estimation, mais devrait également prendre en compte l'erreur de modèle. Ceci nous conduit donc à étudier de fond en comble dans la deuxième partie de ce mémoire le risque de modèle dont nous appliquerons la méthodologie pour la quantification de l'erreur de modèle en Auto corporelle.

PARTIE II :

LE RISQUE DE MODELE

Un modèle est une méthode quantitative, un système ou une approche qui applique les théories, techniques et hypothèses mathématiques, statistiques, économiques et financières pour transformer des données en entrée en estimations quantitatives. Un modèle mathématique est un outil d'aide à la décision qui est construit dans un but précis (description de données, prédiction d'un cours...). Un modèle n'est jamais parfaitement réaliste. Le choix du modèle et des paramètres afférents peut amener à des résultats différents bien que modélisant la même quantité. D'où la notion de risque de modèle.

Il y a plusieurs sens à l'expression "risque de modèle". Communément, on définit le risque de modèle comme la différence de prix que procure plusieurs modèles de valorisation d'un instrument ayant les mêmes données en entrée (hypothèses sur la quantité d'intérêt, données historiques...). En effet, nous supposons qu'il existe un unique "*vrai modèle*" et l'erreur de modèle vient du risque auquel est exposé, par exemple, un décideur en utilisant un modèle erroné ou mal spécifié. Nous retiendrons la définition très générale de Kherkof, le risque de modèle est l'aléa engendré par le fait de "travailler avec un modèle potentiellement incorrect".

Dans cette partie, nous traitons en premier lieu des sources des risques liés à un modèle, ce qui nous permet d'élaborer une cartographie de risque de modèle ; puis nous rappelons quelques mesures qui nous permettent de quantifier l'erreur de modèle enfin nous appliquons cela pour l'estimation de la charge ultime d'un modèle de provisionnement non vie.

Chapitre 4 : Description du risque de modèle

Introduction

Le développement des modèles dans le domaine de la finance dans les années 70 avec notamment le modèle de Black-Scholes-Merton ainsi que leur complexification a engendré des questionnements quant aux risques que pouvaient générer ces modèles. Pour faire face à ce nouveau risque, les entreprises et particulièrement les banques ont mis en place des systèmes de gestion du risque de modèle.

Il est important de noter que ce risque peut, au même titre que les autres risques financiers, entraîner des pertes financières conséquentes. Dans le contexte actuel où des modèles sont créés et utilisés pour la gestion et la prise de décisions stratégiques à chaque niveau des entreprises du secteur bancaire et assurantiel, il est indispensable de ne pas négliger ce risque, car Le risque est défini comme étant inhérent au modèle.

Dans ce chapitre nous identifions les sources du risque de modèle, ce qui nous conduira à l'élaboration d'une cartographie de risque de modèle et après nous définissons la principale mesure de risque que nous utiliserons et les mesures de risque de modèle liées à différentes approches de quantification de l'erreur de modèle.

I- Sources de risque de modèle

Crouhy et al (*Crouhy et al, 1998.*) définissent le risque de modèle comme étant "tous risques induits par l'application d'un modèle statistique". Cette définition concerne aussi bien les modèles stochastiques que les modèles économétriques et comprend toutes les sources du risque. Le risque de modèle est aussi défini de manière plus restreinte, en excluant le risque opérationnel, comme "tous risques induits par le choix, la spécification et l'estimation d'un modèle statistique". Tenant donc compte de tout ceci, nous pouvons classer des différentes sources de risque de modèle en 6 types (Wang et al, 2012) :

- Le modèle est hors sujet
- Le modèle est incorrect (mal spécifié)
- L'implémentation du modèle est erronée
- Le modèle est mal calibré
- Les données sont mal traitées
- La mise en application est inadaptée

Nous allons, par la suite, développer chacune de ces sources de risques et tenter d'élaborer une cartographie des sources du risque de modèle pour avoir une vision plus claire et synthétique du risque de modèle.

1- Modèle sorti du contexte

Dans ce cas le modèle ne répond tout simplement pas à la problématique de départ. Cela peut se produire si le modèle est basé sur des données n'étant pas appropriées ou sur des théories économiques sans fondement. Nous considérerons que ce cas est relativement marginal puisqu'en pratique, dans les grandes institutions, les modèles font l'objet de revues, de back-testing et d'audit. Par conséquent, cette source de risque sera négligée.

2- Modèle incorrect

a- Solutions analytiques fausses

La solution analytique résultant du modèle est fautive. Un exemple clé est la recherche de maximum global qui peut dans certains cas ne résulter qu'en un maximum local. Ce risque appartient au risque de spécification du modèle.

b- Hypothèses sur le processus stochastique fausses

Dans le cas où le modèle fait appel à des processus stochastiques, certaines hypothèses peuvent être faites à tort, par exemple, en faisant des approximations par un processus déterministe. En pratique, il est difficile de modéliser un processus stochastique car nous ne pouvons nous fier qu'aux données observées pour émettre des hypothèses. Ce risque sera considéré comme le risque stochastique par la suite.

c- Mauvais choix de variables

Ce risque peut survenir dans des modèles à multiples variables lorsque les variables sont malencontreusement échangées, mal identifiées ou manquantes. Ce risque appartient au risque de spécification du modèle puisqu'il est lié aux variables afférentes au modèle.

d- Mauvaises hypothèses sur la distribution des paramètres

Si la distribution des paramètres est mal choisie, cela peut conduire à de grandes déviations des résultats. Par exemple, si on fait l'hypothèse qu'un paramètre est gaussien alors que ce dernier possède une distribution à queue épaisse. Ce risque appartient au risque de spécification.

e- Mauvaises hypothèses sur la nature des marchés

Les marchés sont très influencés par l'environnement économique, géographique et politique. Tous ces facteurs doivent être pris en compte de la manière la plus plausible possible afin de se rapprocher de la réalité. Cependant, certains marchés sont très complexes à modéliser et cela peut entraîner une nouvelle fois une erreur de spécification.

f- La non prise en compte des coûts de trading

Les coûts de trading sont bien souvent ignorés des modèles. Cependant, ces coûts fluctuent et peuvent être modélisés par un processus stochastique. Par conséquent, dans des cas d'extrême fluctuation, la non prise en compte de ces coûts entraînent une erreur majeure du modèle. Cette erreur est liée à des hypothèses sur le modèle et est donc rattachée aux erreurs de spécification.

g- Mauvaises hypothèses sur les sous-jacents des actifs

La plus grande erreur à cet égard repose dans la mauvaise prise en compte des dépendances entre les actifs. Ceci est aussi une erreur de spécification.

3- L'implémentation du modèle erronée

Le développement croissant de modèles de plus en plus complexes a rendu nécessaire l'amélioration de l'implémentation de ces derniers. Plusieurs critères sont indispensables dans l'implémentation des modèles comme le temps de déroulement de l'algorithme, sa stabilité et sa clarté afin qu'elle soit compréhensible et vérifiable par d'autres collaborateurs.

a- Méthodes numériques inappropriées

Le choix de la méthode numérique peut être inadapté et entraîner des erreurs d'approximation exacerbées en période de crise. Cette possible erreur doit être prise en compte en amont de l'implémentation car les algorithmes étant généralement longs, il est souvent difficile de trouver et traiter l'erreur *a posteriori*. Cette erreur est une erreur d'implémentation.

b- Nombre de simulations insuffisant

Le nombre de simulations doit être correctement choisi. En effet, si nous effectuons beaucoup de simulations, l'utilisation du modèle peut être fortement ralentie. Cependant, beaucoup d'algorithmes, en particulier ceux utilisant les méthodes de Monte-Carlo ou les méthodes de Bootstrap, requièrent la réalisation d'un nombre important de simulations pour assurer leur convergence. Cette erreur est une erreur d'implémentation.

c- Choix des données

Le choix des données peut s'avérer assez complexe dans la mesure où le nombre de données disponibles croit énormément et que le choix du type de données, de leurs sources et de leur quantité n'est pas évident. Par exemple, les données utilisées proviennent parfois de plusieurs sources et ne sont pas par conséquent identiques. De plus, la prolifération des données a encouragé l'utilisation d'historiques beaucoup plus longs pour la calibration des modèles bien que cela entraîne des erreurs (données trop anciennes et incohérentes, période de crise mal gérée...). Cette erreur est une erreur de données.

4- Le modèle mal calibré

Cette source d'erreur provient de mauvais choix de méthodes statistiques pour les estimations.

a- Choix des méthodes statistiques utilisées

Il existe une quantité non négligeable de méthodes statistiques pouvant donner des estimations différentes et le choix de l'une par rapport à l'autre peut être assez subjectif.

b- Incertitude des estimations

Les méthodes statistiques engendrent des erreurs d'estimation dépendant des données d'entrée qui sont parfois elles même estimées. La source d'erreur est à la fois dans les données et l'estimation.

c- Gestion des données aberrantes

Le retraitement des données est indispensable afin d'avoir les données les plus "propres" possibles. Il faut s'assurer que les données aberrantes soient supprimées et qu'il n'y ait pas de données manquantes ou tronquées pouvant engendrer un biais d'estimation. Ces erreurs sont des erreurs de données qui peuvent être liées à une erreur purement "humaine" ou opérationnelle lorsqu'il s'agit d'erreurs de frappe par exemple.

d- Calibration des modèles

Chaque modèle doit être régulièrement vérifié, calibré, réévalué et ajusté. Il faut alors définir à quelle fréquence ces vérifications doivent avoir lieu et s'assurer que les hypothèses de départ en particulier concernant les dynamiques des paramètres sont toujours valables. Cette erreur sera qualifiée d'erreur de maintenance des modèles.

e- Jugements subjectifs

Les modèles sont souvent basés à la fois sur un historique de données et sur des considérations plus subjectives issues d'expériences passées et d'avis d'experts. Les deux entrées présentent des avantages bien qu'il faut prendre beaucoup de précautions dans leur utilisation.

5- Mauvais traitement des données

La qualité des modèles est intimement liée à la qualité de ces données d'entrée.

a- Horizon de temps et fréquence des données

Comme précisé précédemment, la taille de l'historique de données doit être mûrement réfléchi afin qu'en sortie, les résultats soient significatifs sans perdre en qualité de prédiction.

b- Erreur dans la base de données

Les données étant issues de sources diverses, il est souvent difficile de vérifier la rectitude des bases fournies.

6- Mauvaise mise en application

a- Hypothèses passées plus valides

Le modèle ayant fait ses preuves par le passé peut ne plus être justifié dans certains contextes en particulier si les hypothèses ne sont plus valides. Cette erreur est due à un problème d'utilisation du modèle comme pour le cas suivant.

b- Réutilisation des modèles

Un modèle ayant correctement fonctionné pour un certain actif peut être utilisé pour un actif similaire mais présentant quelques caractéristiques distinctes.

Nous présentons ci-dessous la cartographie du risque de modèle au vu de la description des différents risques que nous venons d'établir (Davesne, 2015).

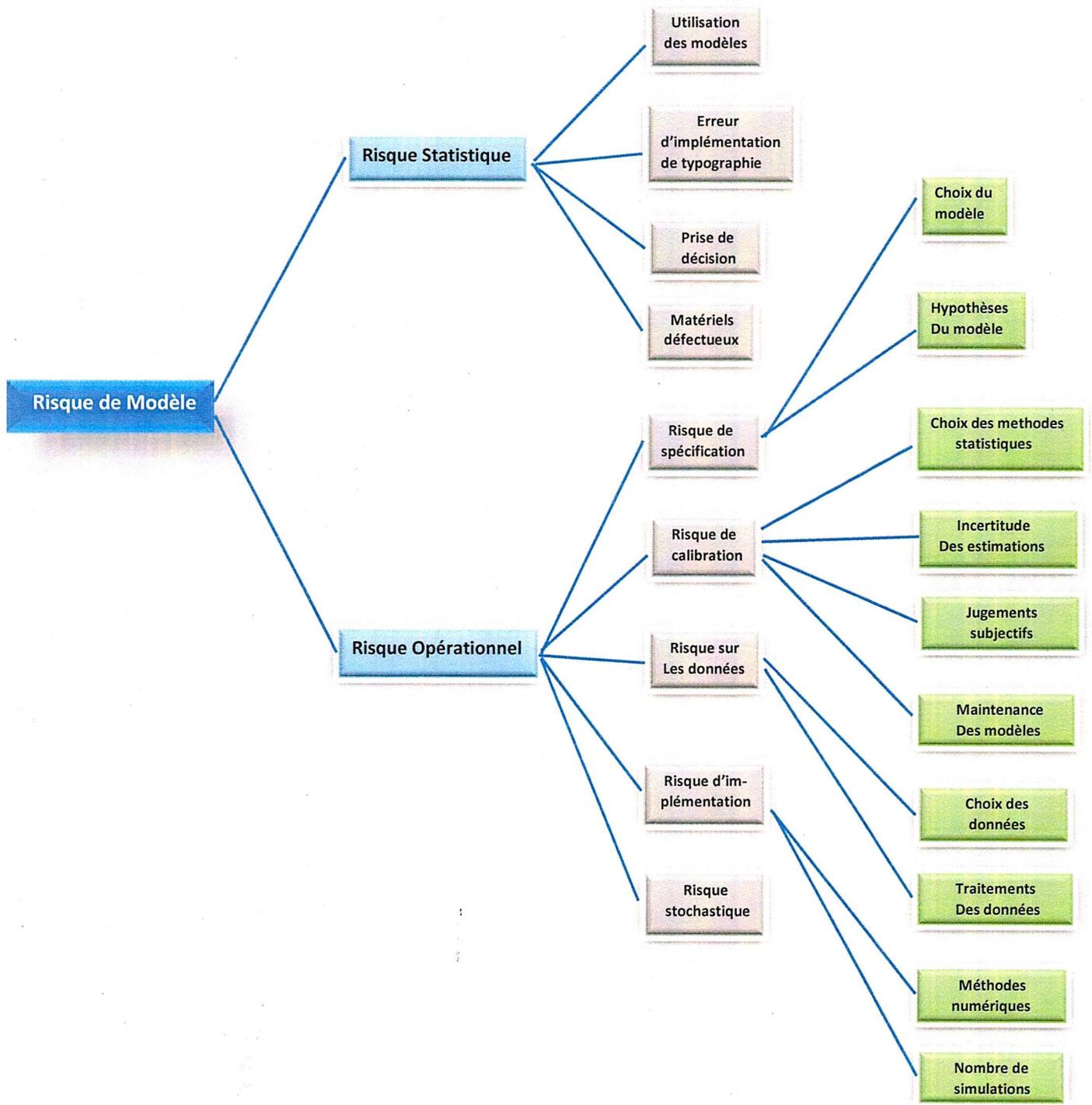


Figure 19 : Cartographie du risque de modèle

Tous les risques présentés interviennent dans le processus de création du modèle interne d'une entreprise d'assurance. Dans notre étude, nous nous concentrerons uniquement sur le risque de spécification du modèle et en particulier sur le choix du modèle par soucis de simplification. Cependant, parmi les autres risques définis, beaucoup sont déjà plus ou moins bien pris en

compte par les politiques de gestion du risque, notamment le risque opérationnel. Nous considérons donc que ces risques sont déjà traités par ailleurs.

II- Quantification de l'erreur de modèle

Dans cette section inspirée des travaux de Davesne (Davesne, 2015), nous présentons le modèle théorique de notre étude ainsi que la théorie mathématique afférente aux mesures de risque qui nous permettra par la suite de quantifier l'erreur de modèle.

1- Le modèle

Dans les problèmes statistiques, nous cherchons à inférer ou prédire une quantité d'intérêt inconnue, notée y , sachant une quantité x connue qui représente les données historiques, le contexte et des avis d'experts plus subjectifs. L'objectif est donc de modéliser la probabilité suivante $P(y/x)$, en d'autres termes, de déterminer, au vu de la variable x , l'incertitude sur y . Spécifier cette probabilité peut s'avérer complexe, c'est pourquoi nous nous référons à des modèles afin d'approcher cette distribution de probabilité. Un modèle M permet de définir une relation hypothétique entre les variables x et y . Un modèle comprend donc toutes les hypothèses émises sur la variable y et est calibré à l'aide de l'historique de données. Le modèle M_θ est représenté par deux éléments, (M, θ) où M correspond aux hypothèses sur la forme structurelle du modèle et θ correspond aux paramètres du modèle spécifiques à la structure du modèle. Les paramètres sont estimés après spécification du modèle à partir de l'historique de données x . La variable d'intérêt, obtenue à partir du modèle $(M, \hat{\theta})$, est alors notée $\hat{y}_{M, \hat{\theta}}$.

Dans cette étude, nous prédisons des variables d'intérêt, au préalable, nous avons défini plusieurs modèles qui nous paraissaient cohérents pour la prédiction et l'inférence de cette variable d'intérêt au vu de l'historique de données, de fondements théoriques et des pratiques actuarielles. Comme nous l'avons expliqué précédemment, la définition de l'erreur de modèle est vaste et ce risque est issu de multiples sources. Nous nous concentrons dans cette étude uniquement sur l'erreur de spécification du modèle, c'est à dire sur l'erreur commise lors du choix de modèle.

Plus précisément, pour chaque quantité d'intérêt considérée, il existe un modèle dit "de référence", noté M^* qui correspond au modèle utilisé en pratique pour modéliser ce risque sans prendre en compte une éventuelle erreur de modèle. Le "vrai modèle" M_{θ_0} , avec θ_0 le vrai paramètre, correspond au modèle qui permet d'obtenir la vraie distribution inconnue de cette variable.

L'erreur de modèle sera alors évaluée à partir d'une classe de modèles $M = (M_1, \dots, M_n)$, appelée aussi ensemble de tolérance, où chaque modèle M_i est un "bon" candidat pour l'estimation de la variable d'intérêt.

Dans les sous-sections suivantes, nous présentons quelques rappels sur les mesures de risques.

2- Mesures de risque

Les définitions rappelées ci-dessous sont celles prises dans le livre d'Arthur Charpentier « Mathématique de l'assurance non vie » (*Charpentier, 2005*).

a- Définition :

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Une mesure de risque \mathcal{R} est une fonction définie par :

$$\begin{aligned} \mathcal{R} &: \Omega \rightarrow \mathbb{R}_+ \\ X &\mapsto \mathcal{R}(X) \end{aligned}$$

C'est la fonction qui à chaque variable aléatoire X associe un nombre $\mathcal{R}(X)$ quantifiant le niveau de danger inhérent au risque.

b- Quelques propriétés des mesures de risque

- i) - **Invariance en loi** : $X \approx_L Y \implies \mathcal{R}(X) = \mathcal{R}(Y)$
- ii) - **Croissance** : $X \geq Y \implies \mathcal{R}(X) \geq \mathcal{R}(Y)$
- iii) - **Invariance par translation** : $\forall k \in \mathbb{R}, \mathcal{R}(X + k) = \mathcal{R}(X) + k$
- iv) - **Homogénéité positive** : $\forall \lambda \in \mathbb{R}_+, \mathcal{R}(\lambda X) = \lambda \mathcal{R}(X)$
- v) - **Sous-additivité** : $\mathcal{R}(X + Y) \geq \mathcal{R}(X) + \mathcal{R}(Y)$
- vi) - **Convexité** : $\forall \beta \in [0, 1], \mathcal{R}(\beta X + (1 - \beta)Y) \geq \beta \mathcal{R}(X) + (1 - \beta) \mathcal{R}(Y)$

Dans le cadre des réglementations des systèmes bancaires et assurantiels, l'utilisation de mesures de risque s'est développée pour le calcul du capital réglementaire. Ces différentes propriétés sont souhaitables et naturelles pour les mesures de risque permettant de calculer ce capital. En effet, la propriété d'invariance en loi garantit que deux risques identiques sont couverts par le même niveau de capital. La propriété de croissance assure qu'un risque plus "dangereux" nécessitera pour le couvrir, une réserve de capital plus importante que pour un risque moins dangereux. La propriété d'invariance par translation implique que si on enlève un montant k d'une position pour l'investir sans risque alors le capital réglementaire sera réduit du même montant. L'homogénéité de la mesure de risque impose que le niveau de risque soit proportionnel à la taille du portefeuille. Pour finir, les propriétés de sous-additivité et de

convexité garantissent que la diversification des risques implique une réduction du risque global et donc du capital réglementaire.

Pour notre étude, nous nous focaliserons sur la mesure de risque « Value-at-Risk » qui vérifie les quatre premières propriétés mais pas la propriété de sous-additivité.

c- La « Value-at-Risk » (*VaR*)

Nous adoptons la définition suivante pour la *VaR* aussi appelée “mesure de perte potentielle” ou encore “perte potentielle maximale”.

Définition : La *VaR* correspond au niveau maximal de perte pour un niveau de confiance donné et un horizon de temps fixé.

Soit $X \in \Omega$ une variable aléatoire, un risque. La *VaR* au niveau α de X est la fonction définie par :

$$VaR_{\alpha}(X) = \inf \{x \in \mathbb{R}, \mathbb{P}(X \leq x) \geq \alpha\}$$

Si X est une variable aléatoire continue alors : $VaR_{\alpha}(X) = q_{\alpha}(X)$.

La *VaR* est une mesure développée dans les années 90 à la suite de plusieurs crises ayant fortement impacté les marchés. Elle a d’abord été popularisée dans le secteur bancaire avant d’être utilisée dans les milieux assurantiels. C’est la mesure retenue dans le cadre réglementaire Solvabilité II (appliqué dans le secteur des assurances de l’Union Européen). Nous utilisons cette mesure pour définir le montant des réserves requises.

Dans la sous-section précédente, nous avons introduit plusieurs notations. En gardant ces notations, nous notons ρ la mesure du risque lié au choix du modèle :

$$\begin{aligned} \rho : \mathbb{R} &\rightarrow \mathbb{R} \\ \mathcal{R}(\hat{y}_{M,\hat{\theta}}) = \mathcal{R}(M) &\mapsto \rho(M) = \rho(\mathcal{R}(\hat{y}_{M,\hat{\theta}})) \end{aligned}$$

La réserve totale à constituer est alors :

$$\rho_{total} = \mathcal{R}(\hat{y}_{M^*,\hat{\theta}}) + \rho(M^*)$$

Où :

- ρ_{total} : réserve totale pour faire face à tous les risques considérés ;
- $\mathcal{R}(\hat{y}_{M^*,\hat{\theta}})$: réserve pour le modèle de référence M^* ;
- $\rho(M^*) = \rho_{erreur}$: réserve pour prendre en compte l’erreur de modèle.

Dans la suite, nous allons étudier différentes approches pour quantifier l'erreur de modèle $\rho(M)$.

En particulier, nous détaillons les deux approches les plus développées dans la littérature ; à savoir l'approche dite "*worst case*" et l'approche dite Bayésienne.

3- Approches de quantification de l'erreur de modèle

a- L'approche « *worst case* »

Description

Cette approche consiste à mesurer l'erreur maximale commise en choisissant M^* comme modèle de référence par rapport à un ensemble de modèle M (contenant le modèle de référence et hypothétiquement le vrai modèle). Plusieurs mesures ont été proposées afin d'évaluer cette erreur. Dans cette sous-section, les notations précédentes sont conservées. Kerkhof et al (*Kerkhof et al, 2002*) propose une mesure du risque de modèle "*worst case*". Le principe est de comparer la mesure de risque obtenue avec le modèle de référence avec celle obtenue avec le pire modèle, celui qui nécessite le plus de réserve.

Définition : La mesure "*worst case*" est définie comme suit :

$$\rho(M^*) = \max_{M_i \in M} \mathcal{R}(M_i) - \mathcal{R}(M^*)$$

Cette approche est très prudente dans la mesure où l'erreur est évaluée par rapport au pire des cas. Cette méthode ne s'intéresse en aucun cas au vrai modèle qui pourrait être bien moins conservateur et dont la mesure de la réserve est *a priori* moins importante que celle obtenue avec le "pire modèle".

Remarque :

R1. Nous pouvons remarquer que si le modèle de référence est celui qui requiert le plus de réserve au vu de la mesure sélectionnée (généralement la *VaR*) alors la mesure de l'erreur de modèle devient nulle. En effet, la réserve requise par le modèle de référence est déjà maximale dans ce cas et l'erreur de modèle n'est donc pas sous-estimée.

R2. Une hypothèse importante dans cette approche est que le vrai modèle appartienne à l'ensemble de modèles. Le choix de l'ensemble de modèle n'est pas anodin. Si l'ensemble est trop petit, la probabilité que le vrai modèle soit dans cette ensemble sera petite. A l'inverse un ensemble grand pourra contenir des modèles très éloignés du vrai modèle et la mesure "*worst case*" pourrait devenir très grande.

Quelques mesures de risque liées à l'approche *worst case*

La mesure de Cont

Cont (Cont, 2006) a proposé une autre mesure du risque de modèle basée sur le pire scénario.

Définition : La mesure de Cont est définie par :

$$\rho(M^*) = \max_{M_i \in M} \mathcal{R}(M_i) - \min_{M_i \in M} \mathcal{R}(M_i)$$

Cette mesure est indépendante du modèle de référence. C'est une mesure très conservatrice qui est très coûteuse en pratique car la réserve est maximale. Ce n'est donc pas une approche permettant d'optimiser l'allocation de capital.

La mesure absolue

Barrieu et al (Barrieu et al, 2013) propose deux mesures de risques construites à partir des précédentes.

Définition : La mesure absolue du risque de modèle

$$AM = \rho(M^*) = \frac{\max_{M_i \in M} \mathcal{R}(M_i)}{\mathcal{R}(M^*)} - 1$$

Cette mesure quantifie à quel point le pire scénario est mauvais par rapport au modèle de référence. Si le pire scénario est donné par le modèle de référence alors la mesure absolue est nulle. Donc le risque de modèle est d'autant plus faible que le modèle de référence est proche du pire modèle.

La mesure relative

Définition : La mesure relative du risque de modèle :

$$RM = \rho(M^*) = \frac{\max_{M_i \in M} \mathcal{R}(M_i) - \mathcal{R}(M^*)}{\max_{M_i \in M} \mathcal{R}(M_i) - \min_{M_i \in M} \mathcal{R}(M_i)}$$

Cette mesure donne la position relative du modèle de référence sur l'intervalle $[\min_{M_i \in M} \mathcal{R}(M_i); \max_{M_i \in M} \mathcal{R}(M_i)]$, en d'autres termes elle permet de définir si le modèle de référence est plus proche du meilleur scénario ou du pire. De même, cette mesure est nulle si le modèle de référence correspond au pire scénario et tend vers 1 lorsque ce dernier se rapproche du meilleur scénario.

Il est nécessaire avant d'évaluer le risque de modèle avec une de ces mesures de bien réfléchir au type de risque que l'on cherche à étudier. Comme nous utiliserons la value-at-risk (*VaR*) comme mesure de risque \mathcal{R} . Dans ce cas, nous nous intéressons à la partie droite de la

distribution (risque de hausse) et non à la partie gauche (risque de baisse). Nous traitons notamment de modèles de provisionnement en assurance IARD où nous nous intéresserons au risque de hausse substantielle des engagements de l'assureur. Cette hausse sera évaluée à l'aide du quantile à $\alpha\%$ de la distribution à l'ultime et non le quantile à $(1 - \alpha)\%$ (qui correspondrait à un risque de baisse).

Critiques de l'approche

Cette approche présente un avantage majeur, sa simplicité. En effet, la mise en place est relativement aisée ; dès lors que les modèles ont été choisis, il faut calculer la réserve pour chacun des modèles et la mesure du risque de modèle est alors immédiate par l'une des formules ci-dessus.

Cependant, cette approche comporte plusieurs inconvénients notamment un inconvénient économique. D'une part, cette approche requiert un ensemble de modèle qui soient en adéquation avec les hypothèses faites sur notre quantité d'intérêt et qui s'adaptent bien aux données. Comme nous l'avons expliqué auparavant, ce choix présente une première difficulté, le nombre de modèles sélectionnés. Par ailleurs, pour que la théorie ci-dessus soit valable, il est nécessaire que le vrai modèle appartienne à cet ensemble, or ne connaissant pas ce modèle, cette hypothèse est difficilement vérifiable. Le choix du modèle de référence est aussi capital puisque que ce dernier influence énormément le calcul de l'erreur de modèle.

D'autre part, économiquement, nous remarquons que ces différentes mesures de risque peuvent être, comme nous le verrons ultérieurement, très prudentes et donc augmentent significativement la réserve totale ρ_{totale} telle que définit précédemment. En effet, ces mesures s'intéressent uniquement aux cas extrêmes et non au cas le plus vraisemblable. Dans un souci de gestion du risque et d'optimisation des stratégies d'allocation du capital, cette approche ne paraît pas efficace.

b- L'approche bayésienne

Principe

L'approche Bayésienne a été proposée en 1994 par Draper (*Draper, 1994*). Comme son nom l'indique, cette approche est fondée sur la théorie de Bayes qui s'est développée à partir des années 60 dans différents domaines d'étude en particulier les statistiques, l'économétrie et l'intelligence artificielle. Draper réutilise cette théorie basée sur les probabilités conditionnelles pour introduire des poids afin de pondérer les différents modèles.

L'objectif de l'approche bayésienne est d'estimer la densité de probabilité $p(y \setminus x)$ en prenant en compte à la fois l'erreur de spécification du modèle et l'erreur liée aux paramètres. Ainsi, la prise en compte, dès l'estimation de la densité, des possibles erreurs de modélisation permet, contrairement à l'approche *worst case*, de diminuer l'erreur de modèle en adoptant une stratégie plus optimale.

Le fondement théorique de cette approche expliquée ci-dessous s'inspire de l'article de Cairns (*Cairns, 2000*). Le but est de trouver la distribution de $p(y \setminus x)$

$$p(y \setminus x) = \sum_i p(y \setminus x, M_i) p(M_i \setminus x)$$

Où :

- $p(y \setminus x)$: la densité conditionnelle de y sachant les données historiques x ;
- $p(y \setminus x, M_i)$: la densité conditionnelle de y sachant les données historiques x et le modèle M_i ;
- $p(M_i \setminus x)$: la probabilité *a posteriori* du modèle M_i , C'est la probabilité que le modèle M_i soit le vrai modèle sachant l'historique.

Par ailleurs, nous pouvons réexprimer les quantités intervenant dans la somme :

$$p(y \setminus x, M_i) = \int p(y \setminus x, M_i, \theta_i) p(\theta_i \setminus x, M_i) d\theta_i$$

Où :

- $p(y \setminus x, M_i, \theta_i)$: la densité conditionnelle de y sachant les données historiques x , le modèle M et le paramètre du modèle θ_i ;
- $p(\theta_i \setminus x, M_i)$: la distribution à *posteriori* du paramètre θ_i .

Nous cherchons les distributions à *posteriori* $p(M_i \setminus x)$ et $p(\theta_i \setminus x, M_i)$. En utilisant la loi de Bayes. Nous obtenons :

$$p(M_i \setminus x) = \frac{p(x \setminus M_i) p(M_i)}{\sum_j p(x \setminus M_j) p(M_j)} = c \times p(x \setminus M_i) p(M_i)$$

Et

$$p(\theta_i \setminus x, M_i) = \frac{p(x \setminus M_i, \theta_i) p(\theta_i \setminus M_i) p(M_i)}{p(M_i, x)} = c_i \times p(x \setminus M_i, \theta_i) p(\theta_i \setminus M_i)$$

où :

- $p(x \setminus M_i)$: la fonction de vraisemblance du modèle M_i ;
- $p(M_i)$: la probabilité à *priori* du modèle M_i . Cette probabilité doit être choisie consciencieusement ;

- $p(x \setminus M_i, \theta_i)$: la fonction de vraisemblance du modèle (M_i, θ_i) ;
- $p(\theta_i \setminus M_i)$: la distribution *à priori* du paramètre θ_i associé au modèle M_i . Cette probabilité est aussi choisie par nos soins de manière plus ou moins subjective.
- c et c_i : des constantes de normalisation qui assurent que les probabilités se somment à 1.

Plusieurs auteurs ont proposé des formules permettant d'approximer la vraisemblance des modèles $p(x \setminus M_i)$ et la distribution $p(\theta_i \setminus x, M_i)$ qui peuvent s'avérer difficile à calculer en pratique.

Remarque :

— Draper a montré que $p(\hat{\theta}_i \setminus M_i)$ est négligeable en particulier lorsque la loi de probabilité de cet *a priori* est continue. En d'autres termes, les résultats de l'approche bayésienne ne sont pas très sensibles au choix de l'*a priori* pour les paramètres. Cependant, ce n'est pas le cas pour l'*a priori* relatif au modèle. Ce dernier point pourrait nous laisser penser que l'approche bayésienne n'est peut-être pas la bonne approche pour mesurer l'incertitude du modèle à cause de sa sensibilité. En effet, 2 modèles peuvent avoir la même adéquation aux données mais donner des prédictions différentes.

— Il existe une autre approche appelée l'approche bayésienne partielle. Cette approche consiste à trouver le modèle M^* qui a la plus grande probabilité *a posteriori*. Une fois ce modèle trouvé (on considère que c'est le vrai modèle), on trouve la distribution de y sachant x de la même manière en ne conditionnant que par rapport à ce modèle M^* . Cette approche ne permet donc pas de prendre en compte l'erreur de spécification mais uniquement l'erreur de paramétrisation.

— Afin de simplifier la modélisation, l'approximation suivante peut être suivie :

$p(y \setminus x, M_i) \approx p(y \setminus x, M_i, \hat{\theta}_i)$ où $\hat{\theta}_i$ est l'EMV (Estimateur du Maximum de Vraisemblance). Cette approximation n'est valide que dans le cas où les données sont en nombre suffisant et en considérant que l'EMV est assez fiable. Dans ce cas, on néglige l'erreur liée à la paramétrisation en considérant qu'avec un grand nombre de données et avec les méthodes actuelles d'estimation des paramètres, cette source est négligeable.

Étapes de l'approche bayésienne :

- I- Choix de la classe de modèles M : en se basant sur des avis d'experts et sur la littérature académique. Ce choix est subjectif mais comme expliqué précédemment, cet ensemble de modèles doit contenir le vrai modèle pour que les résultats soient vérifiés. Cette décision n'est donc pas anodine dans notre étude et doit mériter une grande attention.
- II- Choix de l'*a priori* sur le modèle $p(M_i)$: (i.e. la probabilité que M_i soit le vrai modèle) d'après des jugements subjectifs (intuitions, expérience, avis d'experts). Rappel : l'*a priori* sur les paramètres est négligé.
- III- Estimation des paramètres θ_i : pour chaque modèle M_i puis projection de la valeur de y à l'aide des (M_i, θ_i) c'est-à-dire calcul de $p(y \setminus x, M)$. Cette étape consiste à définir pour chaque modèle la réserve qui lui est associé en fonction de la mesure de risque choisie (généralement la *Var*).
- IV- Détermination de l'*a posteriori* de chaque modèle $p(M_i \setminus x)$: Cette probabilité va permettre d'introduire des poids à chaque modèle qui seront utilisés pour évaluer la projection de la valeur de y , $p(y \setminus x)$.
- V- Calcul de l'erreur de modèle : L'erreur de modèle est définie comme suit,

$$\rho(M) = \frac{\mathcal{R}_{Bootstrap}(M)}{\mathcal{R}(M^*)} - 1$$

où $\mathcal{R}_{Bootstrap}(M)$ correspond à la mesure de risque obtenue en faisant un Bootstrap parmi nos différents modèles pondérés à l'aide des probabilités *a posteriori*.

Critiques de l'approche

L'approche bayésienne présente l'avantage de prendre en compte immédiatement et simultanément l'erreur de spécification et de paramétrisation. Cela permet d'avoir une mesure du risque plus précise. Ainsi les réserves pour couvrir le risque de modèle sont moins importantes en comparaison avec l'approche *worst case*. Cette approche est donc conseillée pour une entreprise qui cherche à optimiser sa stratégie d'allocation du capital tout en ayant une bonne gestion de ses risques.

Cependant, elle présente tout de même plusieurs inconvénients. Le premier, similaire à l'approche *worst case*, réside dans le choix de l'ensemble de modèles qui doit potentiellement contenir le vrai modèle.

Le second concerne le choix des *a priori* des modèles de la classe qui peut s'avérer très complexe et loin d'être une évidence. Ce choix peut fortement impacter la probabilité *a posteriori* et ne peut donc pas être fait de manière aléatoire. Généralement, ce choix est subjectif et basé plus ou moins sur des connaissances théoriques ou liées à l'expérience. Il est nécessaire d'être très prudent lors de cette étape.

En outre, dans l'approche théorique plusieurs approximations sont faites sur les distributions afin de simplifier la modélisation de l'approche. Cependant ces approximations ne sont vraies que si le nombre de données est suffisant. De même, l'erreur liée aux paramètres ne peut être négligée que si nous possédons un grand nombre de données historiques permettant de supposer que l'EMV est fiable. La stabilité de l'EMV peut être démontrée à l'aide de mesures d'incertitude.

Pour finir, cette approche peut s'avérer difficile à mettre en application. En effet, la vraisemblance peut ne pas être calculable pour certains modèles. Cette dernière peut aussi être tellement grande (d'un ordre de grandeur bien supérieur à celui des autres modèles) que quasiment l'intégralité du poids (plus de 99%) est donné à ce modèle. Dans ce cas, l'approche par pondération n'apporte pas d'information sensiblement différente de l'approche *worst case*.

Conclusion

Dans ce chapitre il était question de décrypter les notions de risque, de modèle, afin de décrire le risque de modèle. Nous avons ainsi inventorié quelques sources de risque de modèle et rappelé quelques mesures de risques nécessaires pour quantifier l'erreur de modèle.

Dans le chapitre suivant nous appliquons toute cette méthodologie afin de déterminer la charge ultime d'un modèle de provisionnement IARD.

Chapitre 5 : Evaluation de l'erreur de modèle en provisionnement non vie

Introduction

Comme nous l'avons précisé précédemment, ce cadre pratique sera consacré à l'application de l'étude du risque de modèle à une classe de modèles de provisionnement IARD de la branche Auto corporelle issue de la base de données d'AXA Cameroun

I- Evaluation de l'erreur de modèle

Parmi tous les modèles présentés dans la première partie, nous n'avons décidé de ne garder que ceux qui nous semblaient pertinents. Nous rappelons que l'ensemble des modèles doit contenir un nombre suffisant de modèles pour comprendre le « *vrai modèle* » mais ne doit pas être alourdi par des modèles non pertinents.

Nous avons donc défini une classe de modèles constituée des modèles Chain-Ladder, Mack avec distribution prédictive log-normale, Mack avec distribution prédictive normale, GLM Poisson sur-dispersé, distribution prédictive par convolution avec loi Poisson et Bootstrap Poisson.

Nous utilisons l'approche '*worst case*' dans laquelle l'erreur de modèle est évaluée par rapport au pire cas.

Le choix du modèle de référence s'est porté sur le modèle de Chain-Ladder pour la quantification du risque de modèle par rapport à la VaR à 95% parce que, non seulement la validation des hypothèses du modèle de Chain-Ladder a donné des résultats presque corrects, mais également ce dernier est celui utilisé par notre structure d'accueil.

La VaR à 95% de chacun de ces modèles est résumé dans le tableau suivant.

Tableau 32 : Résultats des Best Estimate des provisions par modèle pour la $VaR_{95\%}$

Modèle	Mack Log-normal	Mack Normal	GLM Quasi-poisson	Convolution	Bootstrap Poisson
$VaR_{95\%}$	1 235 310	1 757 810	1 832 979	1 048 404	1 627 453

II- Quantification de l'erreur de modèle

Nous quantifions l'erreur de modèle sur la $VaR_{95\%}$ en utilisant l'approche dite de « modèle de référence » en calculant la valeur absolue de ce risque.

Les résultats sont résumés dans le tableau suivant.

Tableau 33 : Résultats des mesures absolue et relative du risque de modèle pour la $VaR_{95\%}$

$VaR_{95\%}$						
Modèle	Chain-Ladder	Mack Log-normale	Mack Normale	GLM Quasi-poisson	Convolution	Bootstrap Poisson
i	0	1	2	3	4	5
AM_i	0	0,1800	0,6791	0,7510	0,0015	0,5547

D'où

AM	0,7510
----	--------

Nous pouvons constater que la valeur de AM est assez élevée pour la $VaR_{95\%}$. La valeur $(1+AM)$ correspondrait à la proportion par laquelle il faudrait multiplier le montant du modèle de référence pour bien prendre en compte le risque de modèle sur le segment étudié. Dans ce cas, le montant obtenu serait relativement conséquent.

Cette étude nous montre parfaitement la limite de la méthode de calcul pour la quantification du risque de modèle par l'approche dite de « modèle de référence » surtout lorsque l'on utilise des modèles dont les hypothèses ne sont pas totalement vérifiées, ce qui arrive parfois dans la réalité en l'absence de données suffisantes.

Nous présentons ci-dessous une piste d'amélioration pour notre approche « *worst case* » de quantification de l'erreur de modèle.

Dans cette sous-section, nous allons proposer une amélioration de la méthode ci-dessus. Pour ce type d'approche, il est impératif que la classe de modèle sélectionnée soit représentative du risque que l'on souhaite modéliser.

Une alternative intéressante serait de prendre en compte l'ensemble des valeurs individualisées pour quantifier le risque de modèle et non la borne supérieure de ces valeurs. Le risque de modèle pourrait alors être calculé à partir de ces valeurs individualisées et de facteur de crédibilité associé à ces différentes valeurs (Lallement, 2014).

L'enjeu principal concerne toutefois la détermination de ces facteurs de crédibilité. La valeur à affecter à ces facteurs devra dépendre de deux critères essentiels : l'avis d'un expert et les résultats statistiques des modèles. Les facteurs de crédibilité seront donc alors pondérés par ces deux avis.

L'avis d'un expert est essentiel car il permet de juger d'une part quelle classe de modèle est la plus adaptée par rapport aux différentes typologies de risques étudiés et des informations disponibles, et d'autre part quels sont les degrés de comparabilité de ces modèles.

L'avis statistique, quant à lui, résulte des différents tests réalisés par les actuaires que ce soit en termes de vérification des hypothèses des modèles, avec un poids plus conséquent affecté aux hypothèses importantes, et de calibrage de leurs paramètres, avec la nécessité de prendre en compte les tests de significativité des paramètres s'ils sont disponibles.

En prenant en compte ces facteurs de crédibilité, la valeur obtenue pour la mesure absolue et relative du risque de modèle serait alors :

$$AM_{ajusté} = \sum_i \omega_i AM_i$$

$$RM_{ajusté} = \sum_i \omega_i RM_i$$

Avec :

$$\omega_i \text{ le facteur de crédibilité associé au modèle et } \sum_i \omega_i = 1$$

Une telle approche présenterait la possibilité de mieux prendre en compte le risque de modèle car on peut remarquer très facilement que $AM_{ajusté} \leq AM$ et $RM_{ajusté} \leq RM$.

Cependant, le problème majeur concerne essentiellement la détermination des facteurs de crédibilité qui peut se révéler être une tâche fastidieuse et devra être l'objet d'une étude complète. Il faudra toutefois veiller à ce que l'apport d'un avis d'expert ne soit pas à son tour une potentielle source significative de risque de modèle.

En accordant la même crédibilité aux modèles de notre classe pour faire simple, nous obtenons :

$$AM_{ajusté} = 0,3611$$

En prenant donc en compte la mesure de l'erreur de modèle, l'assureur devrait provisionner pour sa branche auto corporelle

$$(1 + AM_{ajusté}) \times 1\,046\,823 = 1,3611 \times 1\,046\,823$$

Soit : **1 424 831 FCFA**

Ce qui lui permet quel que soit l'évènement exceptionnel qui surviendrait d'être solvable avec une probabilité de 95% tout en gérant d'une façon optimale ses ressources, raison pour laquelle nous avons ajusté la mesure et pas nous positionner juste par rapport au pire des cas.

Conclusion et perspective

Ce chapitre nous a permis de prendre en compte l'erreur de modèle afin d'évaluer au plus juste la provision d'une branche d'assurance IARD à déroulé long. Car nous avons vu dans la première partie de notre mémoire que de telles branches étaient très sensible au choix du modèle. Ainsi pour la branche Auto corporelle, il en ressort après prise en compte de l'erreur de modèle que l'assureur devrait plutôt provisionner à hauteur de 1 424 831 FCFA au lieu de 1 046 823 FCFA soit un écart relatif de 36%, ce qui correspond pratiquement au coefficient de variation (35%) trouvé par la méthode de Mack. L'assureur pour ses branches à déroulé long devrait donc à défaut de revoir la profondeur de son triangle prendre en compte l'erreur de modèle.

CONCLUSION ET PERSPECTIVES GENERALES

Pour honorer les engagements pris vis-à-vis des assurés et des actionnaires, et respecter les lignes directrices données par l'autorité de contrôle, la bonne évaluation des provisions pour sinistres à payer est primordiale pour l'assureur. Dans ce mémoire, nous avons étudié les méthodes de provisionnement classiques, et avons pu comparer les différents résultats par les biais d'applications numériques sur un triangle de développement. Ceci nous a permis de mettre en avant les disparités des réserves estimées, et par conséquent d'introduire la question de l'existence d'un « *vrai modèle* », ou du moins le plus adaptée à l'historique des sinistres à étudier.

Dans le but d'expliquer ces différences, nous avons comparé les résultats des triangles issus de branches d'assurances différentes. Cette étude comparative a permis de trouver un lien entre la durée de règlements des sinistres d'une branche et les variations des réserves obtenues par différentes méthodes de notre classe de modèles.

Il s'est donc avéré que les écarts relatifs étaient presque nuls pour les branches à déroulé court. L'assureur dans ce cas pouvait donc juste se préoccuper de la validation des hypothèses d'un modèle pour en faire usage.

Pour les branches à déroulé moyen, l'écart relatif n'était plus à négliger. Le choix du modèle devenait donc délicat d'autant plus que les hypothèses de validation des différents modèles n'étaient pas vérifiées entièrement. Pour cela l'assureur ne devrait pas se contenter juste d'une estimation. Il devrait prendre en compte l'information sur l'erreur d'estimation afin de tenir compte de la volatilité de ses réserves.

Les branches à déroulé long quant à elles donnaient des écarts relatifs assez poussés. La problématique du choix du modèle trouvait ici tout son sens. Il était donc question de quantifier l'erreur que commettrait l'assureur en n'utilisant pas le « *vrai modèle* ». Ceci nous a conduit à décrire le risque de modèle par plusieurs approches et en à ressortir les différentes sources de ce risque. Nous nous sommes focalisés sur le risque de spécification, considéré comme sa source principale. Nous avons utilisé la mesure de risque *VaR* et l'approche dite de « modèle de référence » pour quantifier l'erreur de modèle. Nous avons approché l'erreur par rapport au pire des cas avant d'améliorer cette dernière pour prendre en compte tous les modèles de notre classe en utilisant des facteurs de crédibilités dépendant d'avis d'experts et d'avis statistiques. L'application de toute cette méthodologie à notre branche Auto corporelle a nettement amélioré

la provision de cette dernière. Ceci nous amène à croire que le risque de modèle doit être pris en compte par les professionnels du monde assurantiers et les études de ce risque doivent être davantage approfondies.

Comme piste de développement, s'il serait possible de revoir la profondeur de triangle pour certaines branches d'activité (exemple Auto corporelle). Cette profondeur est importante et a une véritable influence sur la robustesse des résultats statistiques. La disparité des résultats de différentes méthodes peut être fortement lié à cette profondeur.

S'il serait aussi possible de se pencher sur la question du choix du seuil délimitant sinistres attritionnels des sinistres graves. Il pourrait être intéressant par des études poussées de chercher un seuil optimal afin d'éviter une grande volatilité comme le constat a été fait en incendie.

Et enfin comme déjà souligné, prendre en compte l'erreur de modèle.

Une voie de prolongation de ce mémoire serait de s'intéresser à plusieurs classes de modèles. Ainsi après avoir évalué l'erreur de modèle dans chacune des classes, utiliser une pondération fonction de la vraisemblance de chaque classe ou d'avis d'experts pour évaluer l'erreur globale. Comme exemple de choix des classes : Bootstrap (gamma, normale, ...), GLM (poisson sur-dispersé, gamma, ...).

BIBLIOGRAPHIE

- ASTESAN E. (1938) Les réserves techniques des sociétés d'assurance contre les accidents d'automobiles, thèse de droit.
- BENJAMEN S, EAGLES L.M. (1986) Reserves in Lloyd's and the London Market, *Journal of the Institute of Actuaries*, 113, 197-256.
- BUSSON E. (2012) Evaluation du risque de provisionnement à un an : Adaptation de la méthode de Merz & Wuthrich à des cas non standards, Mémoire d'actuariat, Université de Strasbourg.
- CHARPENTIER A. (2005) *Mathématique de l'assurance non-vie*, 2, Chapitre 3 : Mesures de risque.
- CHARPENTIER A. (2010) *Statistiques de l'Assurance, partie 1 – assurance non-vie tarification et provisionnement*, Cours des universités de Rennes et Montréal.
- CHARPENTIER A et DUTANG C. (2012) *L'Actuariat avec R, version numérique*.
- CROUHY M, GALAI D, et MARK R. (1998) Model Risk. *Journal of Financial Engineering*, 7, 267–288.
- DAVESNE C. (2015) Etude du risque de modèle dans le cadre d'un modèle interne, Mémoire d'actuariat, ENSAE, Paris.
- DJAGANA O (2020) *Provisionnement non vie*, cours de l'Institut International de Assurances, Yaoundé
- Draper D. (1994) Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society*.
- ENGLAND P.D, VERRALL R.J (1999), Analytic and Bootstrap estimates of prediction errors in claims reserving, *Insurance : Mathematics and Economics*, 25, 281-293
- HABIB I, RIBAN S (2012) Quelle méthode de provisionnement pour des engagements non-vie dans Solvabilité 2, Mémoire d'actuariat, ENSAE, Paris.
- KERKHOF J, MELENBERG B, et SCHUMACHER H. (2002) Model Risk and regulatory capital.
- KOUADIO K E. (2020) *Assurances Dommages*, cours de l'Institut International des Assurances, Yaoundé
- LALLEMENT T. (2014) Le risque de modèle, Mémoire d'actuariat, Université de Strasbourg.
- MACK T. (1993) Distribution-free calculation of the standard error of Chain Ladder reserve estimates, *Astin Bulletin*, 23(2), 213-225.
- MACK T. (1999) The standard error of Chain Ladder reserve estimates : recursive calculation and inclusion of a tail factor: *ASTIN BULLETIN*, 2.
- NELDER J et WEDDERBURN R. (1972) Generalized linear models. *Journal of the Royal Statistical Society, Series A (General)*, 135(3), 370–384.
- RENSHAW A E, VERRALL R J. (1998) A Stochastic Model Underlying the Chain-Ladder Technique, *British Actuarial Journal*, 4, Issue 04, 903-923.
- SABAM M et NGOUFO K L. (2015) Méthodes de provisionnement non-vie : Application sur la branche Incendie, Mémoire d'actuariat, INSEA, Rabat.
- WANG F et SAKUTH A. (2012) Model Risk, *MSA400 Financial Risk*

ANNEXES

Annexe I : éléments de validation des hypothèses et résultats en Auto matérielle

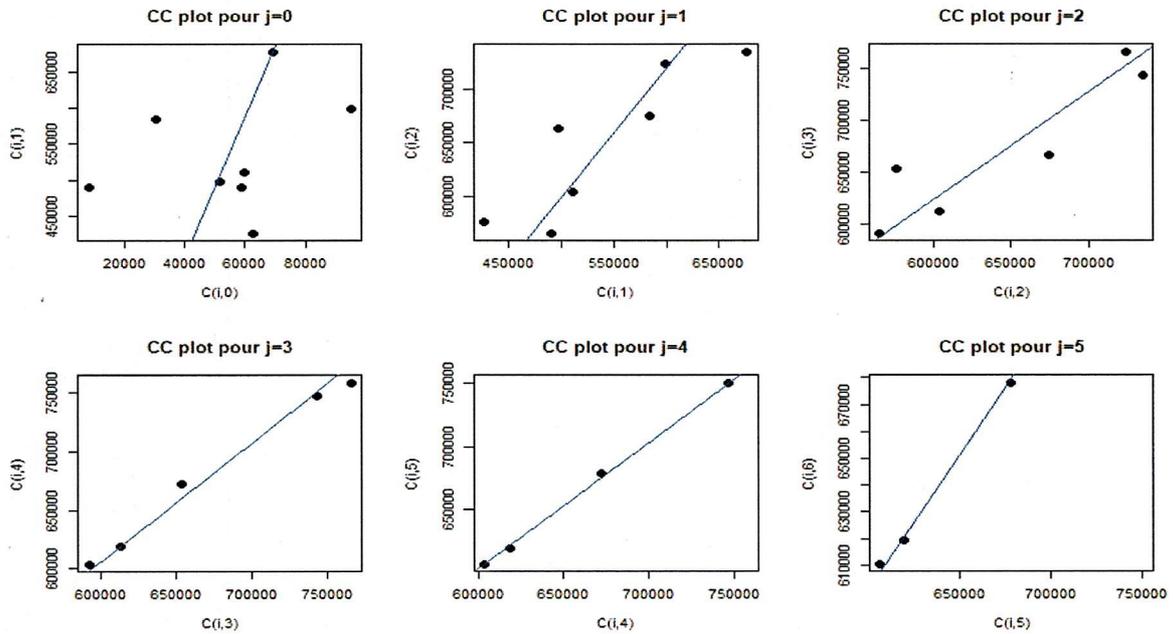


Figure 20_AI : CC-Plots auto matérielle

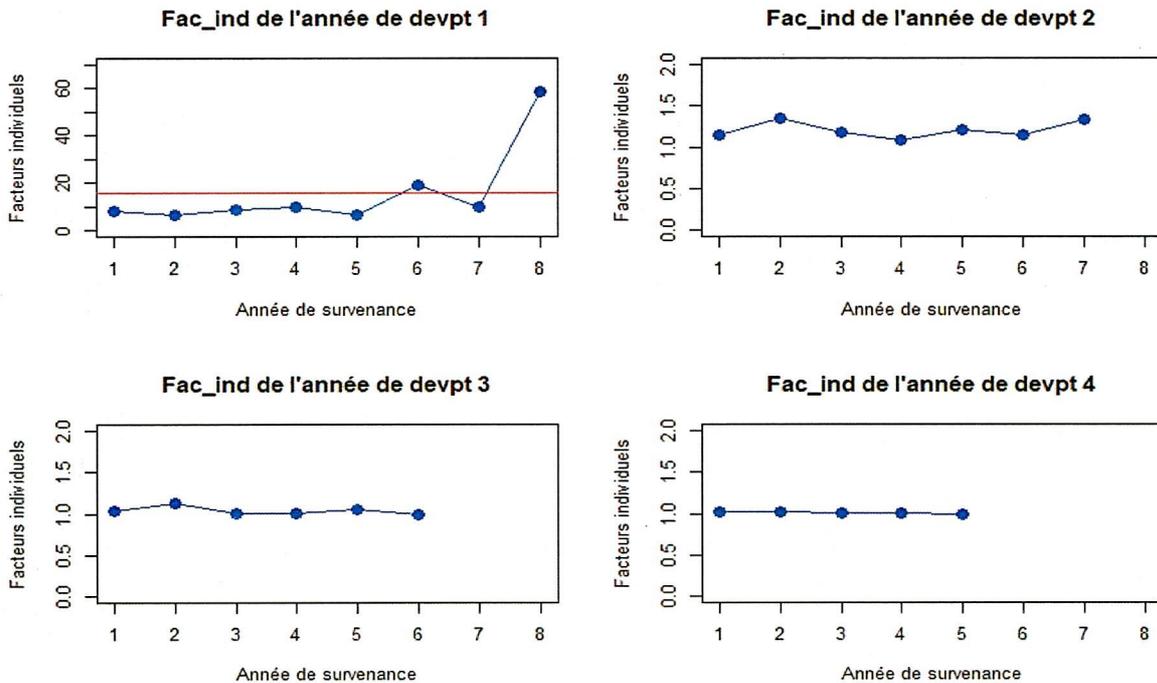


Figure 21_AI : Test d'indépendance des facteurs ($j=0$ à 3) auto matérielle

Tableau 34_AI : Développement du triangle inférieur auto matérielle par la méthode Chain Ladder

	0	1	2	3	4	5	6	7	8
2013	58 695	490 242	565 752	592 151	603 812	606 415	610 581	614 089	614 603
2014	62 725	426 647	575 944	653 567	672 065	678 397	678 097	682 962	683 533
2015	59 890	510 440	604 243	612 740	618 744	619 095	619 334	623 358	623 879
2016	69 397	677 018	734 588	743 339	746 247	750 707	752 325	757 213	757 846
2017	95 059	599 113	723 623	765 587	757 673	761 617	763 258	768 218	768 860
2018	30 774	584 687	674 699	666 689	672 857	676 360	677 818	682 222	682 792
2019	52 017	497 629	663 678	690 237	696 623	700 250	701 759	706 319	706 909
2020	8 361	489 810	587 720	611 239	616 895	620 106	621 443	625 481	626 003
2021	27 317	267 313	320 748	333 583	336 670	338 422	339 152	341 355	341 641
fac_estimé	9,7858	1,1999	1,0400	1,0093	1,0052	1,0022	1,0065	1,0008	1

Tableau 35_AI : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	614 603	614 603	0	0,00%
2014	683 533	682 962	571	0,11%
2015	623 879	619 334	4 545	0,85%
2016	757 846	750 707	7 139	1,34%
2017	768 860	757 673	11 187	2,10%
2018	682 792	666 689	16 103	3,02%
2019	706 909	663 678	43 231	8,11%
2020	626 003	489 810	136 193	25,54%
2021	341 641	27 317	314 324	58,94%
Total	5 806 066	5 272 771	533 295	100%

Tableau 36_AI : Test de significativité des paramètres additifs modèle London Chain en auto matérielle

Paramètre testé (constante)	Valeur	p-valeur	Significativité
a ₀	481875	0,00052	Pertinent
a ₁	259685	0,0387	Pertinent
a ₂	123173	0,377	Non Pertinent
a ₃	68000	0,14424	Non Pertinent
a ₄	-12336	0,49413	Non Pertinent
a ₅	30257	0,476	Non Pertinent

Tableau 37_AI : Développement du triangle inférieur auto matérielle par la méthode London-Chain

	0	1	2	3	4	5	6	7	8
2013	58 695	490 242	565 752	592 151	603 812	606 415	610 581	614 089	614 603
2014	62 725	426 647	575 944	653 566	672 065	678 397	678 097	682 962	683 533
2015	59 890	510 440	604 243	612 740	618 744	619 095	619 334	623 018	623 538
2016	69 397	677 018	734 588	743 339	746 247	750 707	746 791	753 037	753 667
2017	95 059	599 113	723 623	765 587	757 673	763 438	758 943	765 433	766 072
2018	30 774	584 687	674 699	666 689	673 543	677 298	676 724	681 562	682 131
2019	52 017	497 629	663 678	686 959	691 954	696 149	694 717	699 916	700 501
2020	8 361	489 810	612 215	643 242	652 247	655 493	655 912	660 331	660 883
2021	27 317	508 171	625 430	654 468	662 443	665 933	665 876	670 496	671 056
Pente, fj	0,96261	0,71973	0,84949	0,90828	1,0239	0,95448	1,0201	1,0008	
Constante, aj	481875	259685	123173	68000	-12336	30257	-8763,6	0	

Tableau 38_AI : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	614 603	614 603	0	0,00%
2014	683 533	682 962	571	0,06%
2015	623 538	619 334	4 204	0,48%
2016	753 667	750 707	2 960	0,34%
2017	766 072	757 673	8 399	0,95%
2018	682 131	666 689	15 442	1,75%
2019	700 501	663 678	36 823	4,17%
2020	660 883	489 810	171 073	19,37%
2021	671 056	27 317	643 739	72,89%
Total	6 155 984	837 529	883 211	100%

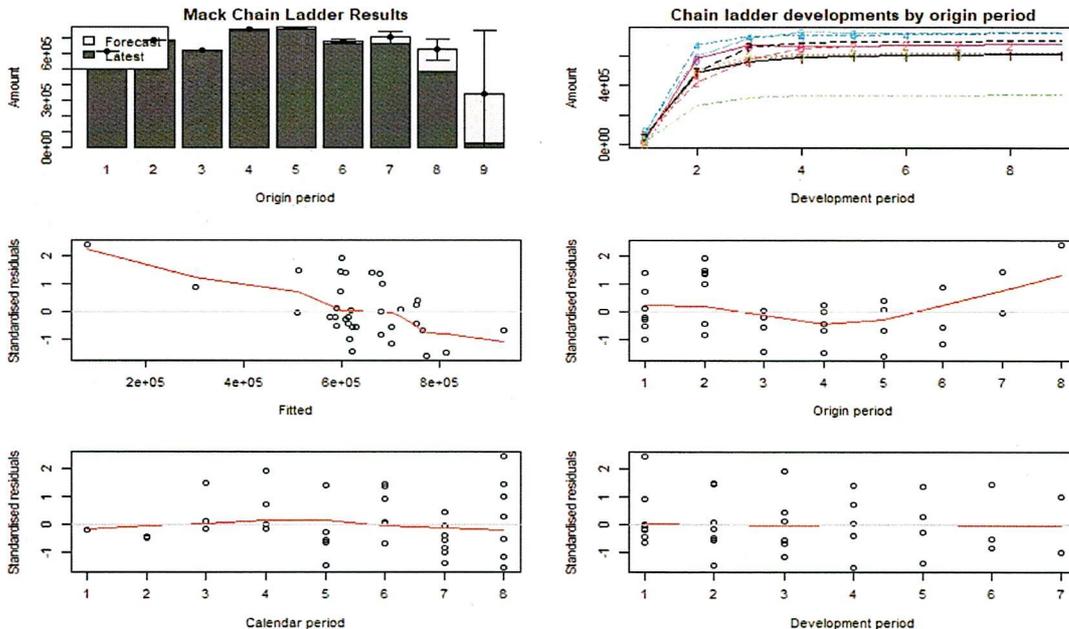


Figure 22_AI : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder

Tableau 39_AI : Résultats du modèle de Mack en auto matérielle

Année de survénance	Provision	$s\hat{e}p(\hat{R})$	Loi log-normale		Loi normale	
			Borne_inf	Borne_sup	Borne_inf	Borne_sup
2013	0	0				
2014	571	215	412	693	150	992
2015	4545	802	4 215	4 753	2 973	6 117
2016	7139	3385	4 335	9 599	504	13 774
2017	11187	4511	7 722	13 939	2 345	20 029
2018	16103	12016	5 420	30 731	- 7 448	39 654
2019	43231	38639	10 198	101 879	- 32 501	118 963
2020	136193	66519	80 461	186 127	5 816	266 570
2021	314324	409736	27 323	1 339 642	- 488 759	1 117 407
Total	533 295	418 758	163 613	1 075 276	- 287 470	1 354 060

Tail factor

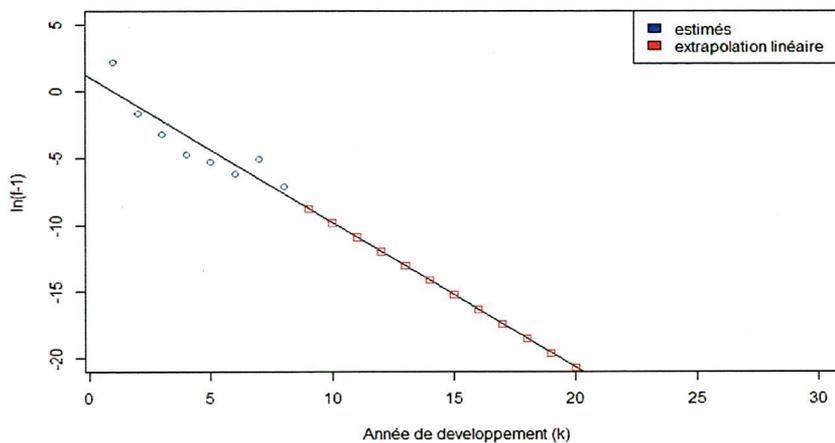
Figure 23_AI : Evolution des $\ln(f_j - 1)$ et prédiction par un modèle linéaire : branche auto matérielle

Tableau 40_AI : Provision auto matérielle après prise en compte du tail factor

tail factor	1,0002				
Année de survénance	Charge ultime Chain Ladder	Charge ultime après tail factor	Dernier paiement	Provision	% de la provision totale
2013	614 603	614750	614603	147	0,03%
2014	683 533	683696	682962	734	0,14%
2015	623 879	624028	619334	4 694	0,88%
2016	757 846	758028	750707	7 321	1,37%
2017	768 860	769044	757673	11 371	2,13%
2018	682 792	682956	666689	16 267	3,04%
2019	706 909	707079	663678	43 401	8,12%
2020	626 003	626153	489810	136 343	25,50%
2021	341 641	341723	27317	314 406	58,80%
Total	5 806 066	5 807 457	5 272 771	534 686	100%

Annexe II : Eléments de validation des hypothèses et résultats en branche Incendie

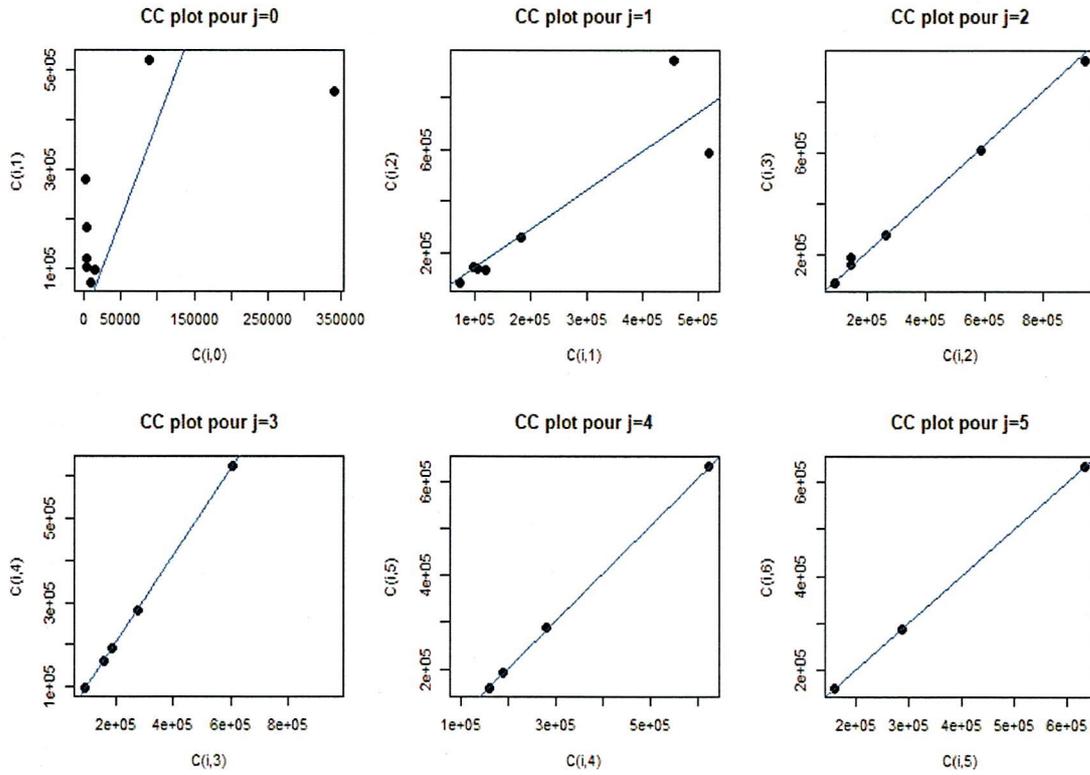


Figure 24_AII : CC-Plots incendie

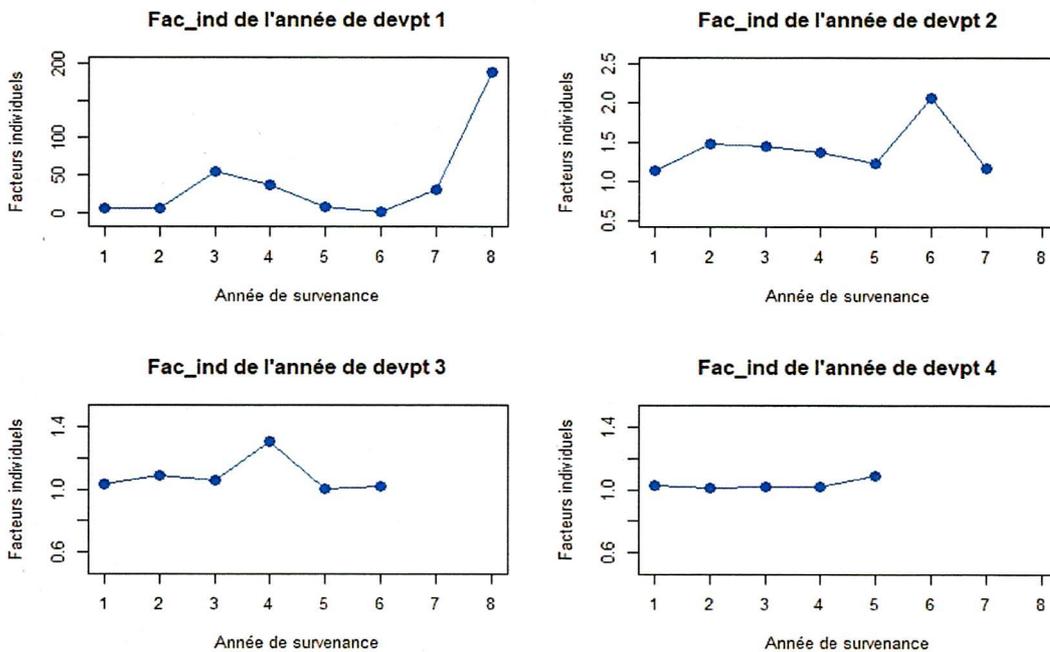


Figure 25_AII : Test d'indépendance des facteurs (j=0 à 3) incendie

Tableau 41_AII : Développement du triangle inférieur incendie par la méthode Chain Ladder

	0	1	2	3	4	5	6	7	8
2013	89 157	519 526	586 741	607 789	624 451	632 464	634 585	639 519	659 538
2014	15 522	98 289	145 767	158 921	160 758	161 352	161 543	161 543	166 600
2015	3 322	182 335	262 620	277 112	282 405	288 175	288 175	289 961	299 038
2016	2 843	105 123	143 076	186 556	190 381	192 292	192 703	193 897	199 967
2017	8 540	73 768	90 721	90 841	98 858	100 138	100 352	100 974	104 135
2018	339 941	456 775	940 415	960 064	985 958	998 724	1 000 858	1 007 060	1 038 586
2019	3 958	120 479	139 816	147 031	150 996	152 951	153 278	154 228	159 056
2020	1 474	278 796	413 664	435 010	446 743	452 527	453 494	456 304	470 588
2021	16 921	66 812	99 133	104 248	107 060	108 446	108 678	109 352	112 775
fac_estimé	3,9485	1,4838	1,0516	1,0270	1,0129	1,0021	1,0062	1,0313	1

Tableau 42_AII : Tableau récapitulatif des provisions en incendie par la méthode Chain Ladder

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	659 538	659 538	0	0,00%
2014	166 600	161 543	5 057	1,22%
2015	299 038	288 175	10 863	2,62%
2016	199 967	192 292	7 675	1,85%
2017	104 135	98 858	5 277	1,27%
2018	1 038 586	960 064	78 522	18,95%
2019	159 056	139 816	19 240	4,64%
2020	470 588	278 796	191 793	46,30%
2021	112 775	16 921	95 854	23,14%
Total	3 210 283	2 796 003	414 280	100%

Tableau 43_AII : Test de significativité des paramètres additifs modèle London Chain en incendie

Paramètre testé	Valeur	p-valeur	Significativité
a ₀	170000	0,019	Pertinent
a ₁	-16600	0,8592	Non Pertinent
a ₂	17200	0,16	Non Pertinent
a ₃	878	0,78	Non Pertinent
a ₄	-479	0,82	Non Pertinent
a ₅	-863	0,42	Non Pertinent

Tableau 44_AII : Développement du triangle inférieur en incendie par la méthode London-Chain

	0	1	2	3	4	5	6	7	8
2013	89 157	519 526	586 741	607 789	624 451	632 464	634 585	639 519	659 538
2014	15 522	98 289	145 767	158 921	160 758	161 352	161 543	161 543	166 600
2015	3 322	182 335	262 620	277 112	282 405	288 175	288 175	289 496	298 558
2016	2 843	105 123	143 076	186 556	190 381	192 292	192 300	192 621	198 651
2017	8 540	73 768	90 721	90 841	98 858	99 810	99 399	98 751	101 842
2018	339 941	456 775	940 415	960 064	983 647	997 402	1 001 055	1 009 812	1 041 423
2019	3 958	120 479	139 816	157 564	162 168	164 035	163 915	163 940	169 072
2020	1 474	278 796	417 891	436 779	447 986	453 990	455 183	458 246	472 591
2021	16 921	187 331	275 333	293 636	301 458	305 342	305 862	307 367	316 989
Pente, f _j	1,0214	1,5586	1,0041	1,0236	1,0145	1,0045	1,0104	1,0313	
Constante, a _j	170047	-16641	17174	877,79	-479,06	-862,56	-1685	0	

Tableau 45_AII : Tableau récapitulatif des provisions auto matérielle par la méthode Chain Ladder

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	659 538	659 538	0	0,00%
2014	166 600	161 543	5 057	0,80%
2015	298 558	288 175	10 383	1,65%
2016	198 651	192 292	6 359	1,01%
2017	101 842	98 858	2 984	0,47%
2018	1 041 423	960 064	81 359	12,93%
2019	169 072	139 816	29 256	4,65%
2020	472 591	278 796	193 795	30,80%
2021	316 989	16 921	300 068	47,69%
Total	3 425 264	837 529	629 261	100%

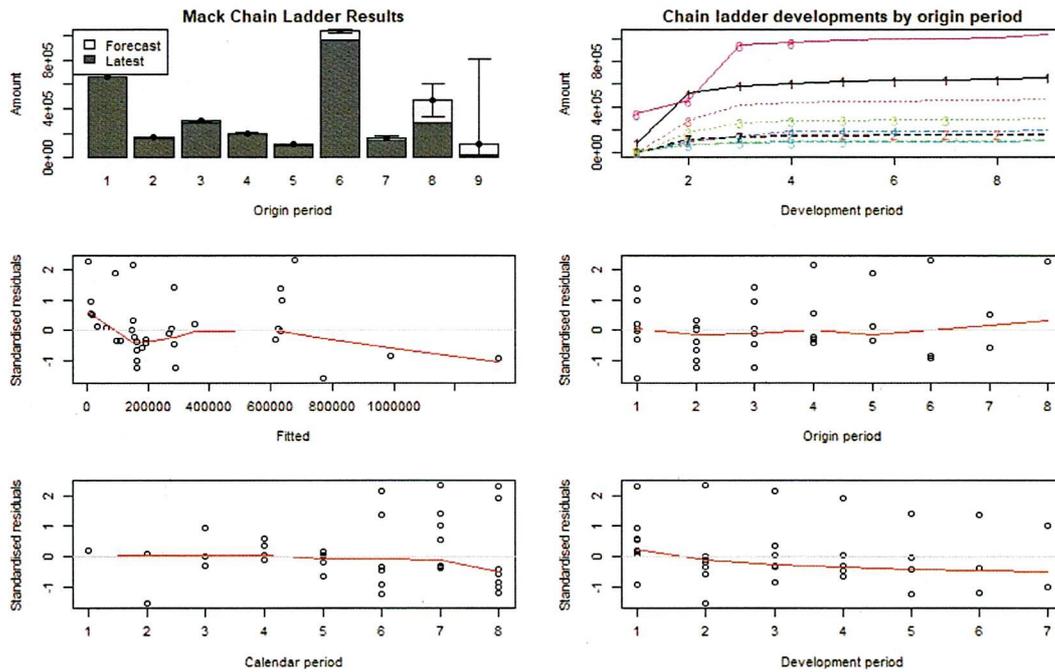


Figure 26_AII : Validation des hypothèses de Mack en incendie : cadence de règlements presque régulière, distribution aléatoire des résidus

Tableau 46_AII : Résultats du modèle de Mack en incendie

Année de survenance	Provision	$\widehat{sep}(\hat{R})$	Loi log-normale		Loi normale	
			Borne_inf	Borne_sup	Borne_inf	Borne_sup
2013	0	0				
2014	5 057	77	5 054	5 059	4 906	5 208
2015	10 863	1 806	10 158	11 304	7 323	14 403
2016	7 675	1 510	6 990	8 113	4 715	10 635
2017	5 277	1 510	4 348	5 920	2 317	8 237
2018	78 522	15 062	71 843	82 776	49 000	108 044
2019	19 240	19 769	3 268	55 099	19 507	57 987
2020	191 793	134 174	71 981	343 110	71 188	454 774
2021	95 854	699 210	5	32 611 318	-1 274 598	1 466 306
Total	414 281	713 577	13 968	3 097 569	- 984 329	1 812 891

Tail factor

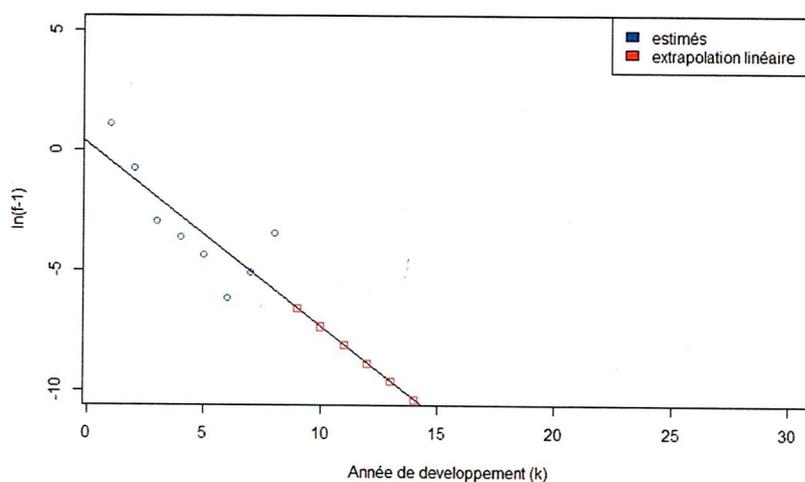
Figure 27_AII : Evolution des $\ln(f_j - 1)$ et prédiction par un modèle linéaire en incendie

Tableau 47_AII : Provision incendie après prise en compte du tail factor

tail factor	1,0026				
Année de survenance	Charge ultime Chain Ladder	Charge ultime après tail factor	Dernier paiement	Provision	% de la provision totale
2013	659 538	661257	659538	1 719	0,41%
2014	166 600	167034	161543	5 491	1,30%
2015	299 038	299817	288175	11 642	2,75%
2016	199 967	200488	192292	8 196	1,94%
2017	104 135	104406	98858	5 548	1,31%
2018	1 038 586	1041292	960064	81 228	19,22%
2019	159 056	159470	139816	19 654	4,65%
2020	470 588	471815	278796	193 019	45,67%
2021	112 775	113069	16921	96 148	22,75%
Total	3 210 283	3 218 648	2 796 003	422 645	100%

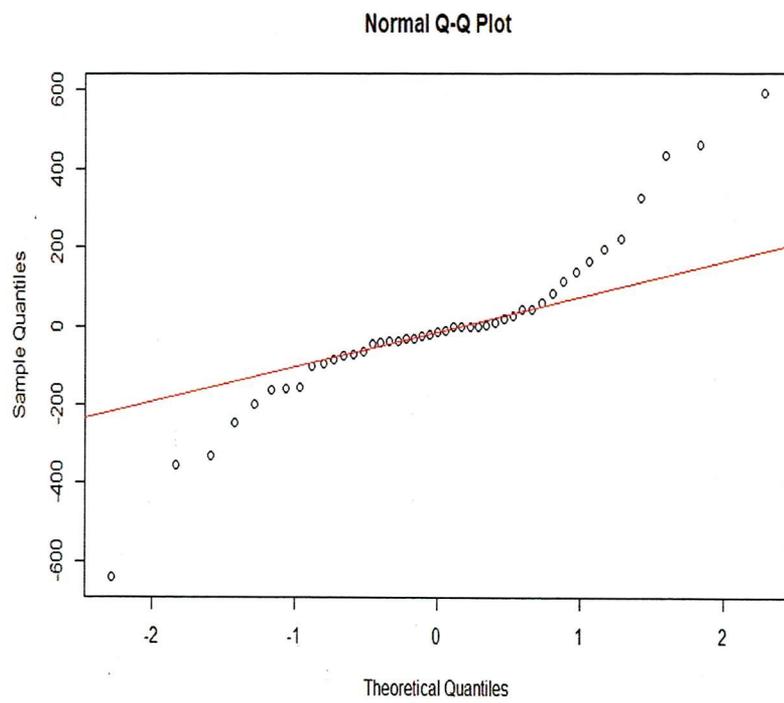


Figure 28_AII : QQ-plot GLM quasi-poisson incendie

Annexe III : Eléments de validation des hypothèses et résultats en branche Santé

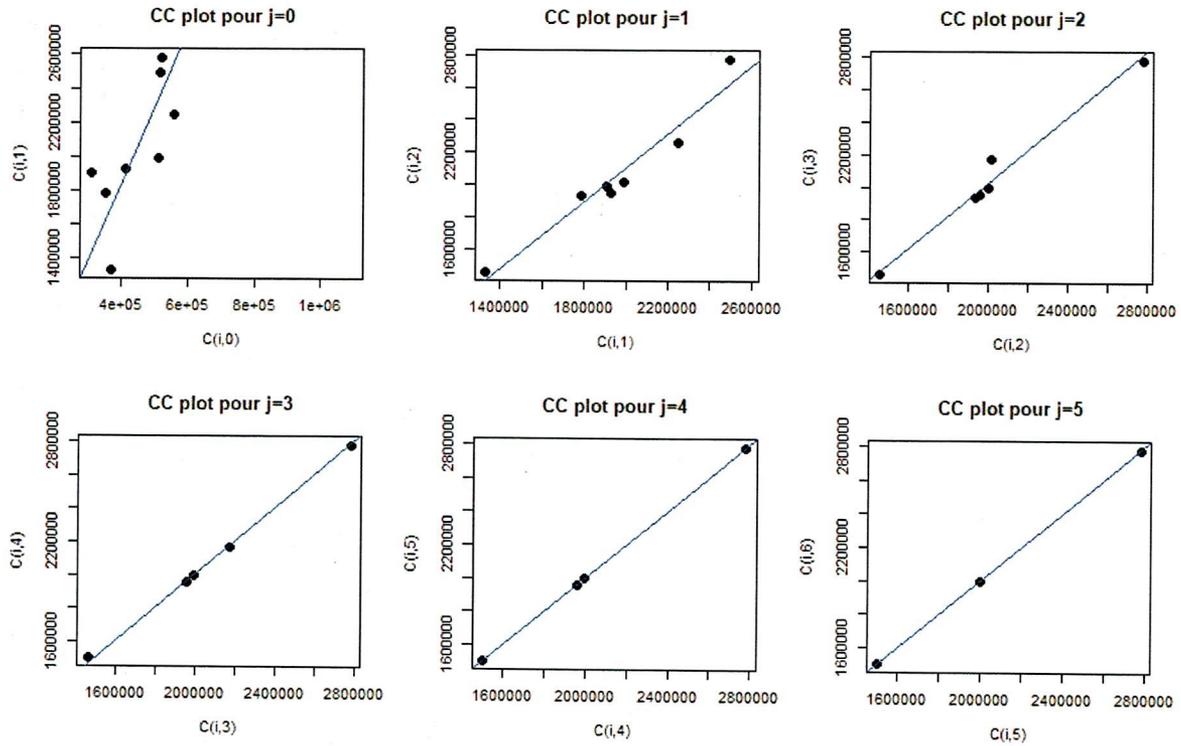


Figure 29_AIII : CC-Plots santé

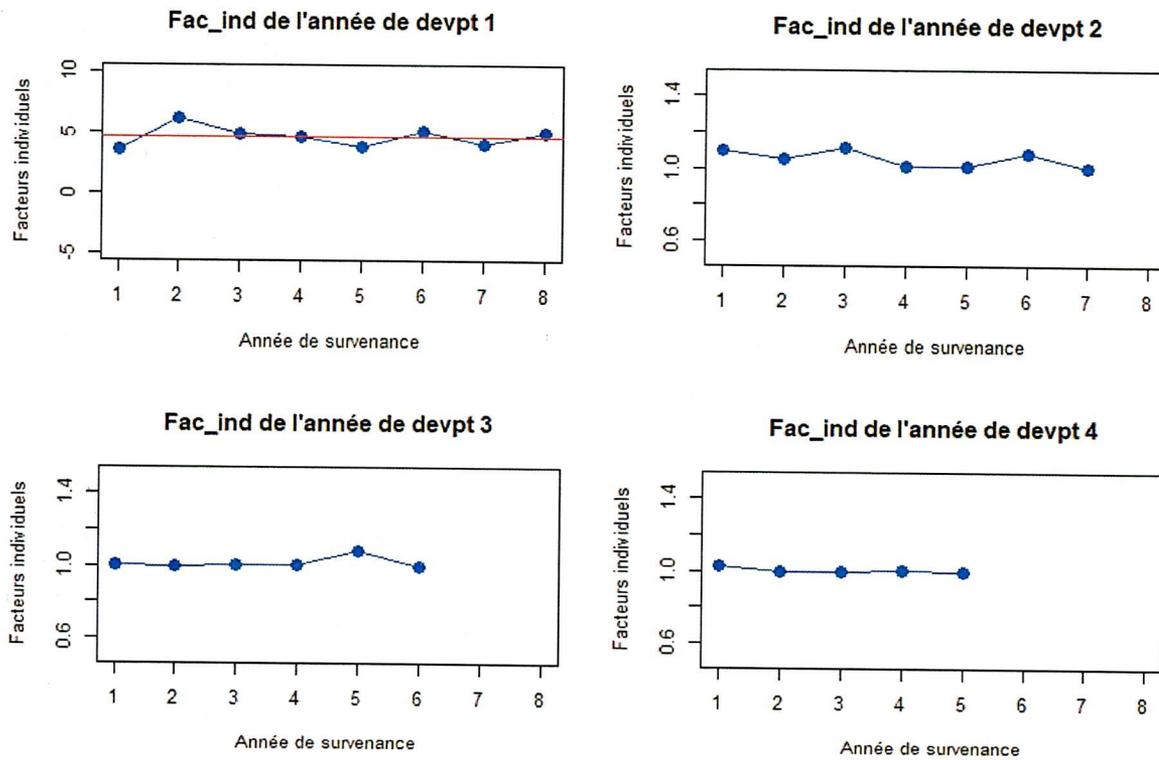


Figure 30_AIII : Test d'indépendance des facteurs (j=0 à 3) santé

Tableau 48_AIII : Développement du triangle inférieur santé par la méthode Chain Ladder

	0	1	2	3	4	5	6	7	8
2013	368 657	1 331 146	1 456 556	1 462 620	1 504 817	1 504 817	1 504 817	1 504 817	1 504 817
2014	309 417	1 904 717	1 994 607	1 994 962	1 994 962	1 997 255	1 997 255	1 997 255	1 997 255
2015	514 821	2 492 389	2 775 961	2 777 902	2 778 412	2 778 412	2 778 412	2 778 412	2 778 412
2016	411 386	1 924 848	1 953 327	1 955 148	1 958 687	1 958 687	1 958 687	1 958 687	1 958 687
2017	512 126	1 990 958	2 013 566	2 170 227	2 170 227	2 170 831	2 170 831	2 170 831	2 170 831
2018	352 172	1 784 640	1 935 415	1 935 415	1 944 054	1 944 595	1 944 595	1 944 595	1 944 595
2019	556 063	2 246 587	2 264 076	2 295 219	2 305 463	2 306 105	2 306 105	2 306 105	2 306 105
2020	519 973	2 581 961	2 717 565	2 754 945	2 767 242	2 768 013	2 768 013	2 768 013	2 768 013
2021	1 095 691	5 025 348	5 289 278	5 362 033	5 385 966	5 387 466	5 387 466	5 387 466	5 387 466
fac. estimé	4,5865	1,0525	1,0138	1,0045	1,0003	1,0000	1,0000	1,0000	1

Tableau 49_AIII : Tableau récapitulatif des provisions santé par la méthode Chain Ladder

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	1 504 817	1 504 817	0	0,00%
2014	1 997 255	1 997 255	0	0,00%
2015	2 778 412	2 778 412	0	0,00%
2016	1 958 687	1 958 687	0	0,00%
2017	2 170 831	2 170 227	604	0,01%
2018	1 944 595	1 935 415	9 180	0,20%
2019	2 306 105	2 264 076	42 029	0,93%
2020	2 768 013	2 581 961	186 052	4,11%
2021	5 387 466	1 095 691	4 291 775	94,75%
Total	22 816 181	18 286 541	4 529 640	100%

Tableau 50_AIII : Test de significativité des paramètres additifs du modèle London Chain en santé

Paramètre testé	Valeur	p-valeur	Significativité
a ₀	645156	0,284	Non Pertinent
a ₁	-19986	0,933	Non Pertinent
a ₂	35321	0,829	Non Pertinent
a ₃	68671	0,12	Non Pertinent
a ₄	935	0,8	Non Pertinent
a ₅	5,38e-10	0,63	Non Pertinent

Tableau 51_AIII : Développement du triangle inférieur santé par la méthode London Chain

	0	1	2	3	4	5	6	7	8
2013	368 657	1 331 146	1 456 556	1 462 620	1 504 817	1 504 817	1 504 817	1 504 817	1 504 817
2014	309 417	1 904 717	1 994 607	1 994 962	1 994 962	1 997 255	1 997 255	1 997 255	1 997 255
2015	514 821	2 492 389	2 775 961	2 777 902	2 778 412	2 778 412	2 778 412	2 778 412	2 778 412
2016	411 386	1 924 848	1 953 327	1 955 148	1 958 687	1 958 687	1 958 687	1 958 687	1 958 687
2017	512 126	1 990 958	2 013 566	2 170 227	2 170 227	2 170 781	2 170 781	2 170 781	2 170 781
2018	352 172	1 784 640	1 935 415	1 935 415	1 948 586	1 949 179	1 949 179	1 949 179	1 949 179
2019	556 063	2 246 587	2 264 076	2 290 982	2 293 956	2 294 488	2 294 488	2 294 488	2 294 488
2020	519 972	2 581 961	2 723 993	2 749 189	2 739 024	2 739 478	2 739 478	2 739 478	2 739 478
2021	1 095 691	4 075 088	4 310 815	4 330 112	4 274 612	4 274 797	4 274 797	4 274 797	4 274 797
Pente, f _j	3,1304	1,0628	0,99628	0,97132	0,99982	1	1	1	1
Constante, a _j	645156	-19986	35321	68671	935,31	5,377E-10	0	0	0

Tableau 52_AIII : Tableau récapitulatif des provisions santé par la méthode London Chain

Année de survenance	Charge ultime	Dernier paiement	Provision	% de la provision totale
2013	1 504 817	1 504 817	0	0,00%
2014	1 997 255	1 997 255	0	0,00%
2015	2 778 412	2 778 412	0	0,00%
2016	1 958 687	1 958 687	0	0,00%
2017	2 170 781	2 170 227	554	0,02%
2018	1 949 179	1 935 415	13 764	0,41%
2019	2 294 488	2 264 076	30 412	0,90%
2020	2 739 478	2 581 961	157 517	4,66%
2021	4 274 797	1 095 691	3 179 106	94,02%
Total	21 667 894	837 529	3 381 353	100%

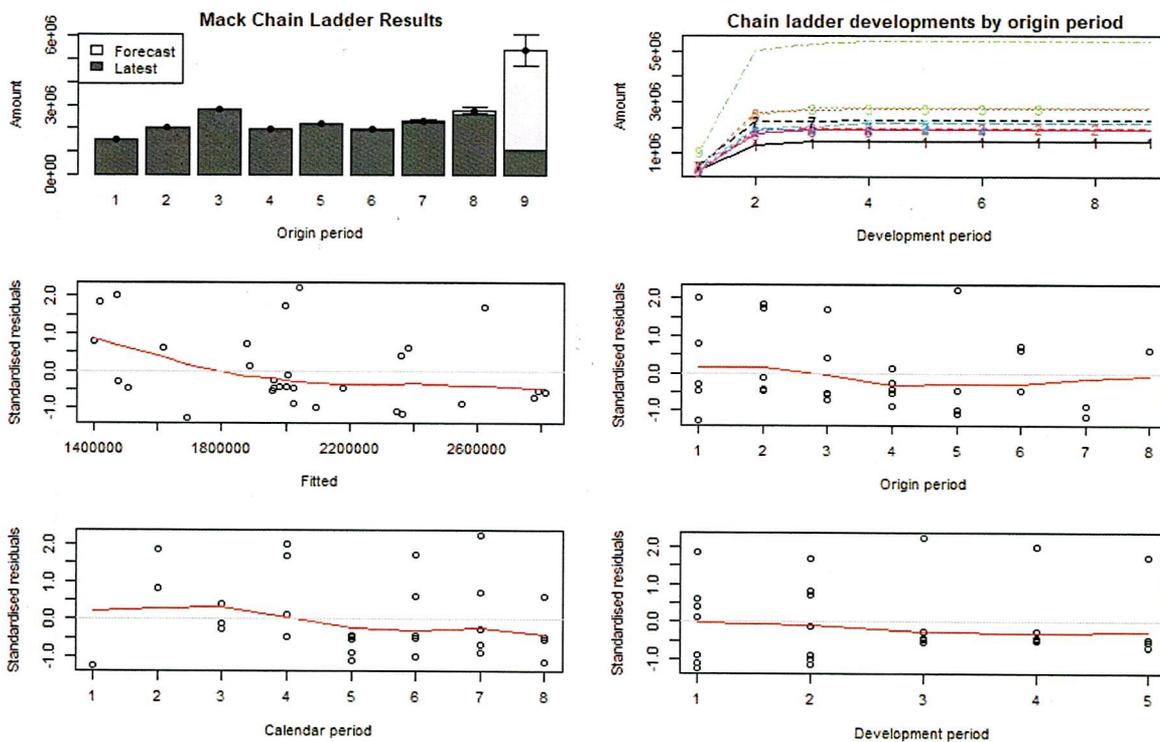


Figure 31_AIII : Validation des hypothèses de Mack en santé : cadence de règlements presque régulière, distribution aléatoire des résidus

Tableau 53_AIII : Résultats du modèle de Mack en santé

Année de survenance	Provision	$sep(\hat{R})$	Loi log-normale		Loi normale	
			Borne_inf	Borne_sup	Borne_inf	Borne_sup
2013	0	0				
2014	0	50			-	99
2015	0	66			-	129
2016	0	50			-	97
2017	604	1 353	7	8 295	-	2 047
2018	9180	24 212	56	189 619	-	38 276
2019	42029	78 088	1 067	371 912	-	111 024
2020	186052	142 733	59 591	365 672	-	93 704
2021	4291775	672 005	4 043 520	4 446 262	2 974 645	5 608 905
Total	4 529 640	707 075	4 269 165	4 691 685	3 143 774	5 915 506

Tableau 54_AIII : Provision santé après prise en compte du tail factor

tail factor	1,0002				
Année de survenance	Charge ultime Chain Ladder	Charge ultime après tail factor	Dernier paiement	Provision	% de la provision totale
2013	1 504 817	1505178	1504817	361	0,01%
2014	1 997 255	1997734	1997255	479	0,01%
2015	2 778 412	2779077	2778412	665	0,01%
2016	1 958 687	1959156	1958687	469	0,01%
2017	2 170 831	2171351	2170227	1 124	0,02%
2018	1 944 595	1945061	1935415	9 646	0,21%
2019	2 306 105	2306658	2264076	42 582	0,94%
2020	2 768 013	2768676	2581961	186 715	4,12%
2021	5 387 466	5388757	1095691	4 293 066	94,66%
Total	22 816 181	22 821 648	18 286 541	4 535 107	100%

Normal Q-Q Plot

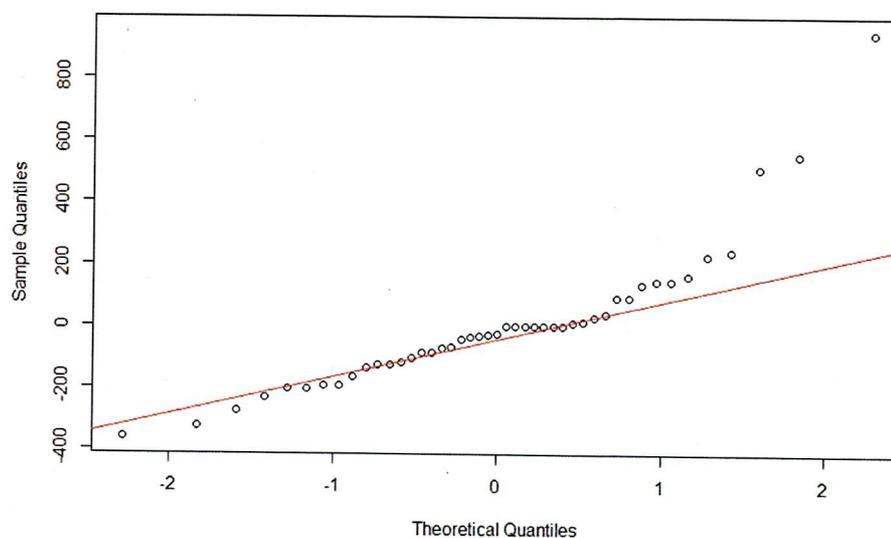


Figure 32_AIII : QQ-plot GLM quasi-poisson santé

Tableau 55_AIII : Résultat du test de Shapiro-Wild des résidus GLM Poisson en santé

Test	Statistique	p-value
Shapiro-Wild	0,838	1,82e-05

Tableau 56_AIII : Résultats du modèle GLM en santé

	Provision	CoV	Intervalle de confiance à 95%	
			Borne_inf	Borne_sup
GLM Poisson	4 529 641	10,67%	3 582 130	5 477 152
GLM quasi-poisson	4 529 641	26,63%	2 165 767	6 893 515

Table de Matières

Dédicace	i
REMERCIEMENTS	ii
Liste des abréviations	iv
Liste des figures	v
Liste des tableaux	vi
Résumé	viii
Abstract	ix
Introduction générale.....	1
Partie I : Évaluation des provisions pour sinistres à payer.....	3
Chapitre 1 : Cadre de l'Etude et Notations	4
I- Le code CIMA et le provisionnement IARD	4
II- Généralités sur la gestion des sinistres	5
1- Cycle de vie d'un sinistre	5
2- Charge ultime	6
3- Triangles de liquidation.....	7
4- Best Estimate des provisions	8
III- Structure d'Accueil.....	8
1- Présentation d'AXA Cameroun.....	8
2- Organisation des périmètres chez AXA Cameroun.....	9
a- Catégorie de sinistres.....	9
b- Sous branches	9
Chapitre 2 : Méthodes classiques de provisionnement.....	11
I- Méthodes déterministes	11
1- Chain-Ladder.....	11
a- Cadre théorique	11
b- Cadre pratique	14
c- Critiques de la méthode Chain-Ladder.....	20
2- Chain-Ladder pondérée	20
a- Cadre théorique	20
b- Cadre pratique branche auto corporelle.....	22
3- London-Chain.....	23
a- Cadre théorique	23
b- Avantage du London-Chain	24

c-	Cadre pratique : branche auto corporelle.....	24
II-	Méthodes stochastiques	26
	Intérêt.....	26
1-	Modèle de Mack (1993)	26
a-	Cadre théorique	26
b-	Cadre pratique	31
c-	Limites de la méthode de Mack.....	33
2-	Notion de Tail factor : Gestion de la liquidation incomplète des triangles.....	33
a-	Extrapolation des facteurs de développement	33
b-	Application	34
3-	Méthode de régression log-normale	35
a-	Présentation du modèle.....	35
b-	Application	36
c-	Les limites du modèle Log-normal.....	40
4-	Modèle Linéaire Généralisé (GLM).....	40
a-	Présentation du modèle.....	41
b-	Cadre pratique	43
5-	Distributions prédictives par convolution.....	48
a-	Présentation	48
b-	Cadre pratique	49
c-	Limite de la distribution prédictive par convolution	49
6-	Bootstrap GLM.....	49
a-	Présentation	49
b-	Cadre pratique	51
c-	Limites de l'approche Bootstrap	54
	Chapitre 3 : Analyse des Résultats	55
I-	Choix de la méthode.....	55
II-	Comparaison des méthodes, analyse des résultats et perspectives	55
1-	Analyse des résultats des sous branches d'une branche	55
a-	Les deux sous branches de la branche Automobile ont des comportements totalement différents.....	56
b-	L'adéquation des données à un modèle avant l'application de ce dernier à une branche est primordiale	57
c-	La branche Auto corporelle est à déroulé long et Auto matérielle à déroulé moyen.....	57
2-	Analyse comparée des résultats des branches : Auto corporelle, Incendie et Santé.....	57
a-	La branche Santé est à déroulé court, Incendie à déroulé moyen et Auto corporelle à déroulé long.....	58

b-	Les branches à déroulé court ne sont presque pas exposées au risque de modèle.....	58
c-	Le choix du modèle impacte moyennement la provision pour les branches à déroulé moyen.....	58
d-	Les branches à déroulé long sont très sensibles au choix du modèle.....	59
Partie II :	LE RISQUE DE MODELE	60
Chapitre 4 :	Description du risque de modèle.....	62
I-	Sources de risque de modèle	62
1-	Modèle sorti du contexte	63
2-	Modèle incorrect.....	63
a-	Solutions analytiques fausses	63
b-	Hypothèses sur le processus stochastique fausses	63
c-	Mauvais choix de variables	63
d-	Mauvaises hypothèses sur la distribution des paramètres	63
e-	Mauvaises hypothèses sur la nature des marchés.....	64
f-	La non prise en compte des coûts de trading.....	64
g-	Mauvaises hypothèses sur les sous-jacents des actifs	64
3-	L'implémentation du modèle erronée.....	64
a-	Méthodes numériques inappropriées.....	64
b-	Nombre de simulations insuffisant.....	64
c-	Choix des données.....	65
4-	Le modèle mal calibré	65
a-	Choix des méthodes statistiques utilisées.....	65
b-	Incertitude des estimations	65
c-	Gestion des données aberrantes.....	65
d-	Calibration des modèles	65
e-	Jugements subjectifs.....	66
5-	Mauvais traitement des données.....	66
a-	Horizon de temps et fréquence des données.....	66
b-	Erreur dans la base de données.....	66
6-	Mauvaise mise en application	66
a-	Hypothèses passées plus valides	66
b-	Réutilisation des modèles	66
II-	Quantification de l'erreur de modèle.....	68
1-	Le modèle	68
2-	Mesures de risque	69
a-	Définition :.....	69

b- Quelques propriétés des mesures de risque	69
c- La « Value-at-Risk » (Var)	70
3- Approches de quantification de l'erreur de modèle.....	71
a- L'approche « <i>worst case</i> ».....	71
b- L'approche bayésienne	73
Chapitre 5 : Evaluation de l'erreur de modèle en provisionnement non vie.....	78
I- Evaluation de l'erreur de modèle	78
II- Quantification de l'erreur de modèle.....	78
Conclusion et perspectives générales	82
Bibliographie.....	84
ANNEXES.....	I