

**Mémoire présenté devant l'Institut du Risk Management
pour la validation du cursus à la Formation d'Actuaire
de l'Institut du Risk Management
et l'admission à l'Institut des actuaires
le 25/11/2019**

Par : Thibault MONNET

Titre : S'orienter vers un business plan stochastique en assurance IARD

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

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

Membres présents du jury de l'Institut
des actuaires :

Membres présents du jury de l'Institut du
Risk Management :

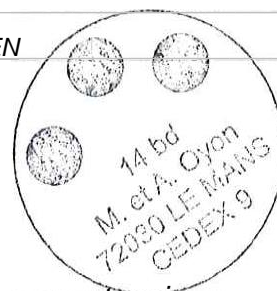
Secrétariat :

Bibliothèque :

Entreprise : COVÉA

Nom : Xavier GUEGUEN

Signature et Cachet :



Directeur de mémoire en entreprise :

Nom : Xavier GUEGUEN

Signature :

Invité :

Nom : Charles DUMARTINET

Signature :

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

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

Signature(s) du candidat(s)

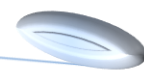


Table des matières

Résumé	4
Abstract	4
Remerciements	5
1 Introduction	6
1.1 Quels sont les objectifs de ce mémoire ?	6
1.2 Mise en relation avec le contexte	6
1.3 Quel périmètre retenir pour cette étude ?	7
2 Modèles préliminaires et éléments de contexte	8
2.1 Le vocabulaire général utilisé dans le métier	8
2.2 Quelques rappels méthodologiques	9
2.3 Comment fonctionne le Business Plan actuel	13
3 Etudier la volatilité des sinistres, un pas vers le stochastique	26
3.1 Quelles méthodes tester et selon quelles conditions ?	26
3.2 Etude des résultats obtenus sur les méthodes de liquidation	33
3.3 Un regard vers les nouvelles approches du marché	42
3.4 Ce qu'il faut retenir de ces tests	43
4 Quelle part d'aléa dans l'atteinte des objectifs commerciaux	45
4.1 Comment rendre le modèle actuel stochastique ?	45
4.2 Vers un modèle simplifié et plus robuste	46
4.3 Les contraintes de la propagation du modèle dans le temps	48
4.4 Quelles conclusions pour ce modèle	52
5 Aller plus loin en ajoutant la volatilité liée à la propagation	54
5.1 Passer d'un portefeuille futur à une charge sinistre espérée	54
5.2 Passer d'un portefeuille à une prime acquise	55
6 Conclusion	56
6.1 Le contexte initial	56
6.2 Les évolutions pour le BP travaillées dans le cadre de ce mémoire	56
6.3 Et pour conclure	67
Annexes	68
Table des illustrations	87

Résumé

Ce mémoire a pour but de matérialiser les travaux de recherche en liquidation stochastique des sinistres connus, leur propagation dans le temps, ainsi que la modélisation des primes émises, dans le cadre de la construction des Business Plans d'assurance non vie de Covéa avec un encadrement des résultats, à des fins tarifaires.

La réassurance est une constituante du Business Plan et les méthodes présentées sont adaptables pour l'inclure dans la plupart des modèles. Il est privilégié une approche en sinistralité totale de l'assureur pour des raisons de délais.

Seront abordés les processus en vigueur, pour la plupart déterministes, ainsi que des pistes d'évolutions, leurs avantages, inconvénients et les résultats obtenus.

Ces travaux se veulent le plus opérationnel possible, afin que l'équipe puisse les mettre en œuvre.

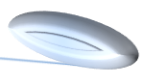
Mots clés : Liquidation stochastique des charges estimées, propagation dynamique des portefeuilles, Chain-Ladder avec bootstrap, encadrement des résultats d'un business plan en assurance non-vie

Abstract

This document is a view of the evolutions COVEA group is able to do in order to improve its business model, from an organization massively determinist in the science of risks, to an approach stochastically based. It will provide intel about the stochastic reserving and calculation of non-life insurance Risk and how it will evolve in the next few years. It will also be dedicated to the premium estimation and fitting, in order to give to the group middle-term visibility (5 years) of its results.

Reinsurance is a part of that work that has been voluntarily aborted due to the timing. It is important to note that all the models could be adapted to treat the reinsurance.

Key words: stochastic approach of claims, time propagation of portfolio, bootstrap Chain-Ladder, confidence interval for business plan indicator in non-life insurance



A ma mère, pour son soutien indéfectible
A mon père, pour sa force, lui qui traverse des moments difficiles
A mes frères et sœurs pour leur capacité à me distraire
Et à celle qui partage ma vie, pour toutes ces qualités et bien d'autres encore

Evidemment, je tiens à remercier mon employeur qui m'a permis de suivre cette formation
Et notamment tous les managers qui m'ont encadré, il y en aura eu cinq sur le cursus
Philippe LECOMTE, Joël DECOSSE, Sandrine BOULET et Charles DUMARTINET
Plus particulièrement Xavier GUEGUEN qui encadre ces travaux depuis 2017

Je remercie également l'équipe Pilotage de la Performance Économique
Et spécifiquement Rémy TEMPELAERE pour son aide sur les données

1 Introduction

Avant d'entrer plus en détail dans le sujet, il semble important de faire état des objectifs, du contexte et de rappeler quelques notions générales sur le groupe Covéa.

1.1 Quels sont les objectifs de ce mémoire ?

Dans le cadre de l'activité de pilotage des résultats à des fins tarifaires, le groupe cherche à se doter d'outils et de méthodes efficaces, capables de répondre aux besoins actuels et futurs, comme par exemple estimer une fourchette d'encadrement des résultats.

A ce jour, les Business Plans (BP) sont réalisés à partir de modèles complets et déterministes. Dit autrement, pour un jeu d'hypothèses donné, le résultat affiché est unique.

On notera que certaines étapes du BP sont réalisées par simulation et mettent en avant des distributions. Cela donnera lieu à un arbitrage pour fixer une valeur unique qui pourra être la moyenne ou un quantile, selon la prudence ou l'aversion aux risques exigée.

L'objectif principal de ce mémoire consiste à étudier comment faire évoluer les modèles, à partir d'une approche simplifiée dans un premier temps. Pour cela divers aspects seront étudiés à commencer par l'encadrement de la charge sinistre à l'ultime pour les exercices passés. Puis la projection des portefeuilles, des primes et de la sinistralité à l'horizon de 5 ans.

Tout au long du mémoire, un autre objectif sera d'intégrer de nouvelles dimensions aux modèles, par exemple pour prendre en considération l'impact de la marque. Si un scénario intègre une tempête, toutes les marques la subissent, avec une intensité plus ou moins forte. Sur le même raisonnement, les perméabilités entre les segments seront étudiées (Auto, Habitation...) ainsi que les transferts intra et inter segments ou garanties.

Ce mémoire donnera également l'opportunité de faire le tour des méthodes actuelles de liquidation, de vérifier leur adéquation aux données disponibles et si possible de positionner Covéa dans ses pratiques par rapport aux autres acteurs du marché.

La vocation première est donc d'adapter les approches récentes testées sur marché, à la réalité du quotidien, tout en restant scientifiquement le plus juste possible et accessible au plus grand nombre.

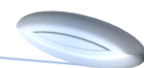
1.2 Mise en relation avec le contexte

Les travaux du présent mémoire s'appuient sur des données du groupe Covéa. Ce groupe se compose de 24 sociétés, dont 17 en non vie, 5 en vie et 2 hybrides. Ces sociétés ont collecté en 2017 une prime de 16,3 milliards d'euros.

Le groupe fait partie des leaders français de l'assurance dommages, 1^{er} sur le marché des particuliers, avec plus de 10 millions de véhicules assurés et près de 8 millions d'habitations en portefeuille. Ces contrats sont gérés et suivis par la Direction IARD, plus particulièrement dans la direction de la Performance Économique pour lequel ce mémoire est constitué.

Comme évoqué précédemment, ces travaux s'inscrivent dans l'amélioration continue des BP et serviront également à définir les tendances tarifaires futures. Par conséquent, la précision sera demandée dans l'encadrement du résultat afin de faciliter la prise de décision.

A noter que le principal indicateur suivi dans le cadre du BP est le ratio combiné, qui intègre



notamment les évolutions de frais de gestion. Historiquement et structurellement, les marques du groupe Covéa ont des approches différentes de ces postes. Par conséquent, l'indicateur sur lequel portera ce mémoire sera le rapport de Sinistralité sur Primes (S/P).

1.3 Quel périmètre retenir pour cette étude ?

Afin de réduire les tests à un périmètre réaliste, seules 2 sociétés seront étudiées elles seront désignées Société A et Société B. En complément, il a été décidé de travailler sur 2 compagnies ayant depuis plusieurs années une même approche de la gestion des risques. En effet, le groupe a pu avoir par le passé des approches différentes dans l'estimation des sinistres. Étant donné que les méthodes classiques de liquidation, qui font référence dans ce mémoire, sont sensibles à toute rupture temporelle, la restriction s'est imposée d'elle-même.

En complément de la restriction sur les sociétés, l'analyse sera réduite à 2 branches d'activité qui seront désignées par la suite comme Automobile (Auto) et Habitation. Les produits de prévoyance, de protection juridique, ou encore les risques d'entreprises et de grands comptes sont exclus de l'étude.

Enfin, pour l'étude de la sinistralité, il a été décidé de regrouper les garanties en 3 blocs, à savoir les risques Corporels ou Graves, les risques Climatiques et l'ensemble des autres risques qui pourront être appelés risques de Masse ou *Attritionnels* par anglicisme. Ces regroupements permettent d'isoler les risques à liquidation très longue – i.e. dont le coût sera incertain pendant de longues années – des risques à liquidation plus courte.

Par ailleurs, la liquidation des risques climatiques peut varier selon l'intensité et la nature de l'évènement. Pour exemple la sécheresse, où la reconnaissance du sinistre est liée à un arrêté de catastrophe naturelle qui peut mettre plus de deux ans à être validé et donc générer des vagues d'ouvertures de dossiers au fil du temps. Ce risque climatique a été isolé car il semble important de prendre en considération les phénomènes météorologiques comme lien de dépendance entre les marques, voire entre les branches.

Concernant la dépendance, sous certaines conditions, il semble a priori raisonnable de dire qu'une même tempête touchera l'ensemble des marques et que, selon l'exposition au risque de la marque face à une typologie d'évènement, l'impact financier diffèrera.

En poussant ce raisonnement plus loin, l'on supposera une dépendance entre la sinistralité Auto et l'Habitation. De fortes crues ou de longues périodes de froid toucheront les deux segments.

Par conséquent, l'une des étapes de l'étude consistera à prendre en considération ces relations et à les intégrer aussi bien dans le calcul des charges, à l'ultime sur les exercices antérieurs, que dans la projection à 5 ans.

2 Modèles préliminaires et éléments de contexte

Dans la suite du document un certain type de vocable pourra être utilisé, il fait référence à des notions internes ou propres au métier, aussi il semble utile d'en définir les principales avant tout autre chose.

De plus, la compréhension du document pourrait nécessiter quelques prérequis, notamment car les méthodes et modèles statistiques usuels et classiques seront considérés comme acquis. Il est donc important, sans entrer dans le détail, de donner au lecteur un rappel théorique et de l'orienter vers des documents plus techniques sur ces sujets s'il désire approfondir le sujet.

Puis, dans un troisième temps, le fonctionnement actuel du Business Plan Covéa IARD sera détaillé.

2.1 Le vocabulaire général utilisé dans le métier

Comme spécifié précédemment, certains mots pourront avoir une signification propre tout au long de ce document. Cette partie se veut en expliciter un maximum afin de rendre la lecture plus fluide et compréhensible.

Deux types de vocable sont différenciés, ceux liés au métier, considérés comme communs ou usuels, et ceux liés au groupe Covéa, dits internes.

2.1.1 Le vocabulaire propre au métier

La **liquidation** est un phénomène de vieillissement du sinistre et donc de son coût. Un sinistre est considéré comme liquidé dès lors qu'il est clos et totalement payé, avec certitude de non réouverture.

Lorsque tous les sinistres d'un même exercice de survenance sont liquidés on dit que sa sinistralité est à l'**ultime** ou vue à la **fin des temps**.

Le **triangle de développement** est un ensemble de données disponibles à un instant t, mises en forme particulière pour simplifier le calcul des visions à l'ultime dans certaines méthodes. NB : cette forme permet des approches matricielles et aussi de visualiser la logique des calculs.

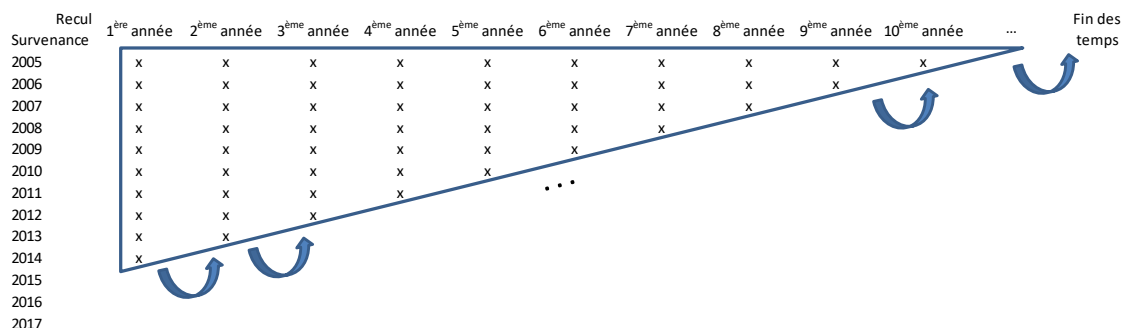
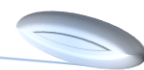


FIGURE 1 - ILLUSTRATION D'UN TRIANGLE DE LIQUIDATION

2.1.2 Le vocabulaire interne à Covéa



Les **résultats économiques** correspondent à la vision la plus réaliste possible des résultats à un instant t donné. Par nature et construction, ils prennent en compte la conjoncture, comme l'occurrence de phénomènes météorologiques (tempête, vague de froid...) ou l'impact de manifestations. Ainsi un phénomène qui serait estimé avec une période de retour de 50 ans serait comptabilisé à 100% s'il a eu lieu et à 0% dans les autres cas.

Les **résultats techniques** sont eux des résultats qui se veulent être un lissage de la vision économique dans le temps. Dit autrement, ils représentent la tendance de fond, parfois dite structurelle, du résultat. Si l'on regarde l'exemple précédent, un phénomène ayant une période de retour de 50 ans serait actualisé par mise en équivalence selon l'exercice puis comptabilisé pour $1/50^{\text{ème}}$ dans la vision technique.

La **mise en équivalence**, très pratiquée en réassurance, traduite de l'anglais *as if*, consiste à actualiser un évènement comme s'il avait lieu à un autre instant t . Par exemple, pour estimer quel serait le nombre de sinistres attendu si la tempête Lothar traversait la France aujourd'hui, à l'identique. Plusieurs étapes sont nécessaires. Selon les cas il faudra notamment prendre en compte l'évolution du portefeuille et des risques assurés, leur localisation, la valeur assurée... Une autre étape consistera à actualiser les coûts des réparations, des pièces comme de la main d'œuvre, sur la base des indices en vigueur et si possible localement.

Le **terme** correspond à l'échéance des contrats. On dit qu'un contrat est termé lorsqu'il est émis pour une nouvelle année. A noter que l'émission pour une nouvelle année s'accompagne d'une revalorisation tarifaire et du droit, dans les 2 mois qui suivent, à la résiliation par le client (Loi Chatel).

Le **risque-année** est une mesure qui vaut 1 pour tout contrat présent une année entière. Par extension, un contrat présent six mois contribuera à la hauteur de $1/2$ risque-année, et 1 contrat présent 1 jour, $1/365^{\text{ème}}$. On définit le portefeuille exposé au risque comme la somme des expositions unitaires, l'unité de mesure est le risque-année.

2.2 Quelques rappels méthodologiques

Dans la construction d'un business plan, notamment pour la projection à 5 ans des S/P, un grand nombre de techniques sont utilisées. Cela va de la liquidation des charges sinistres par diverses approches à la projection de l'activité et des primes acquises, en passant par l'analyse des tendances des Primes Pures par modèles classiques ou autorégressifs sur la sinistralité. Etant donné qu'une grande partie de ces techniques entre dans les bases que doit maîtriser un actuair, il apparaît plus logique de détailler ici quelques notions et de garder les explications approfondies pour les techniques avancées qui sont utilisées pour les tests dans ce mémoire.

2.2.1 La liquidation des sinistres

Le **Chain-Ladder**, qui pourra être noté **CL** par la suite, est une méthode déterministe et classique de liquidation. Elle se décompose aujourd'hui sous plusieurs variantes, un terme plus générique pourrait être Méthode des Facteurs de Développement de l'anglais *Loss Development Factor* ou *Development Factor Methods* telles que nommées dans ResQ – progiciel dédié aux liquidations.

Ces méthodes consistent à s'appuyer sur l'historique connu du développement des triangles pour l'appliquer sur les exercices plus récents. Cela implique, intuitivement, des restrictions d'usage, notamment si les données historiques comportent des ruptures ou des tendances différentes.

Déjà très détaillée dans plusieurs articles, cours ou mémoires, comme dans le mémoire de Selma Jaziri¹, il semble important de retenir que cette méthode s'appuie sur deux hypothèses, l'indépendance des montants cumulés à la fin d'une période, d'un exercice de survenance à l'autre, d'une part. Et d'autre part, l'existence de facteurs de développement tels qu'en tout temps t , l'on puisse définir l'espérance des montants cumulés au temps $t+1$ comme la multiplication de montant cumulé en t par le facteur de développement entre t et $t+1$.

La méthode de **Bornhuetter-Ferguson**, aussi noté **BF** dans la suite du document, a été développée par les scientifiques éponymes en 1972. Sans entrer dans la théorie, car là encore il existe de nombreux documents sur le sujet – à commencer par la publication originale des auteurs dans *The Actuary and IBNR*² – il semble intéressant d'en dire quelques mots.

Cette méthode s'appuie sur l'existence d'une vision a priori de la charge ultime et sur une connaissance de l'état de développement de l'exercice, basée sur des méthodes plus classiques. A noter qu'à partir des méthodes de type CL, on peut déduire à tout instant t , le rapport entre la charge connue en t et celle à l'ultime, l'écart ainsi exprimé en taux permet d'identifier le développement restant.

L'idée du modèle de BF consiste à dire qu'à une date donnée, la charge connue sera complétée par une proportion d'une sinistralité définie a priori, proportion égale à la part restant à développer selon la méthode retenue pour calculer la liquidation. La définition de la charge à l'ultime a priori étant du ressort de l'actuaire, un chapitre y sera consacré dans le présent document.

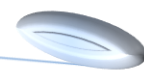
En 1976, Gunnar **Benktander** introduit une méthode³ qui portera son nom, elle mélange le Chain-Ladder et le Bornhuetter-Ferguson, par approche crédibilisée. Dans la pratique c'est une itération de BF avec comme charge a priori celle obtenue à la première itération du BF. Elle permet de donner plus de crédibilité à la méthode de développement classique, elle est donc plus représentative de l'expérience acquise et moins dépendante de l'estimation a priori.

Le **facteur de queue**, ou *tail factor*, est un coefficient permettant de continuer à projeter le triangle au-delà de la dernière année d'expérience. Dit autrement, c'est une estimation du développement restant à parcourir entre la dernière observation disponible et l'ultime.

2.2.2 La projection des tendances

Cette partie s'intéresse aux séries temporelles, et avant de projeter une tendance, la première étape consiste à lisser les données observées, donc à estimer la vision dite technique du résultat. C'est cette vision lissée qui sera projetée. Le lissage peut se faire de plusieurs manières, comme en retirant un point aberrant ou en estimant la moyenne de la série. Cependant, une approche graphique est préconisée pour savoir s'il y a une tendance, une saisonnalité ou tout simplement des points à retirer. Sont listées ici diverses méthodes simples ou vues en cours.

La **régression linéaire** est la méthode la plus commune de lissage, elle permet également la



projection à une échéance plus ou moins longue. Par définition, le temps sera la variable explicative et le modèle définira une pente et le cas échéant une ordonnée à l'origine, composantes de la variable à expliquer. On mesure la qualité du modèle en calculant le coefficient de détermination R^2 , qui estime le carré de la corrélation entre la série originelle et la régression. Plus ce coefficient est proche de 1 et meilleure sera cette corrélation, ce qui rend le modèle plus qualitatif qu'un autre.

On notera que par transformation des données, il est toujours possible de se rapporter à cette forme d'équation, on parle aussi de fonction de lien. Il sera parfois plus pertinent d'étudier les tendances en travaillant sur le logarithme népérien de la variable expliquée ou de la variable explicative, voire parfois des deux, on parle alors, respectivement, de régression **exponentielle**, **logarithmique** ou **puissance**.

La **régression polynomiale**, toujours en vision univariée sur le temps t , est une alternative à la régression linéaire. Il s'agira ici de résoudre des équations quadratiques, cubiques... attention cependant, si sur un instant donné les courbes sont proches, les formes parabolique et hyperbolique de cette méthode peuvent entraîner de très forts mouvements, parfois illogiques au sens métier, sur les projections.

Dans la panoplie d'outils de lissage, nous disposons des modèles de **moyenne mobile**, ces derniers étant plutôt utilisés pour supprimer un effet saisonnier ou cyclique. A noter que ces méthodes peuvent se cumuler aux précédentes.

Enfin, parler de série temporelle sans évoquer les **processus autorégressifs** semble incomplet. L'idée sous-jacente de cette approche est qu'en un instant t , la variable explicative s'explique par la combinaison pondérée d'une à plusieurs valeurs qui auront été prise par le passé.

Une famille générique de modèles combine les processus autorégressifs, les moyennes mobiles ainsi que les fonctions de lien, il s'agit des modèles **ARIMA**, de l'anglais *autoregressive integrated moving average*.

2.2.3 Quelques rappels sur les lois usuelles de probabilité

Les lois de probabilité permettent de décrire mathématiquement l'aléa de certains phénomènes. Au cours de la lecture de ce document, plusieurs lois dites usuelles pourront être rencontrées, il semble important de rappeler ici certaines d'entre elles et de rappeler leurs caractéristiques.

La **loi de Poisson** est une loi discrète permettant de caractériser le comportement d'un nombre d'évènement sur un intervalle qu'il soit spatial et/ou temporel. Elle est régulièrement utilisée en assurance pour approximer un nombre de sinistres. Elle a pour particularité de présenter une variance égale à sa moyenne, ce qui peut restreindre la volatilité et sous-estimer ou surestimer les résultats.

La **loi Binomiale-Négative** est une autre loi discrète. Originellement, elle permet d'estimer un nombre d'essais nécessaires pour obtenir un nombre n de succès (ou d'échec). Par construction, elle peut s'apparenter à une généralisation de la loi de Poisson, et sera également utilisée pour les modèles de nombre. Sa variance étant différente de sa moyenne, elle permet, dans certain cas, un meilleur ajustement que la loi de Poisson.

La **loi Normale**, ou loi de Gauss, est une loi continue et symétrique de référence. Elle sera utilisée notamment pour intégrer un bruit blanc dans les simulations, ce qui permet d'intégrer un aléa. Cette loi, en se référant au théorème central limite et sous certaines conditions, sera la limite de plusieurs autres lois ou processus.

La **loi Log-Normale** est une transformation de la loi Normale. Si une variable aléatoire suit une loi Normale, une nouvelle variable qui serait l'exponentielle de la première suivrait elle, une loi Log-Normale. Cette loi est continue et présente des avantages du fait de sa proximité à la loi Normale. Par construction, elle est asymétrique et a une queue de distribution plutôt épaisse, ce qui en fait une candidate à la distribution de certains risques.

La **loi de Pareto** est loi usuelle pour les queues de distribution épaisse, elle est particulièrement célèbre car très imagée, c'est l'équation mathématique de la règle des 80/20, 20% des dossiers qui prennent 80% du temps, de la charge... Cette règle est testée sur les modèles de risques graves, en effet quelques sinistres peuvent atteindre des montants très importants.

2.2.4 Qu'est-ce qu'un générateur aléatoire de type Monte-Carlo

Les méthodes dites de Monte-Carlo regroupent des méthodes de modélisation de phénomènes aléatoires par simulations itératives parfois dites calculatoires. Il faut savoir que bien d'autres approches sont disponibles dans le panel d'utilisations de ces méthodes, comme l'évaluation d'intégrales ou la résolution de systèmes d'équations... Ces méthodes tirent leur nom des casinos de la cité de Monaco, en référence au hasard.

Il est possible de manipuler ces méthodes dans différents outils, même Excel permet de faire des approches calculatoires, cependant les temps de réponse sont moins performants qu'avec d'autres outils.

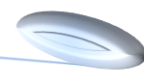
Dans ce mémoire, ces méthodes sont utilisées dans diverses étapes. On notera que le bootstrap qui est une méthode de Monte-Carlo, sera expliqué plus en détail par la suite.

Nota Bene, *bootstrap* est un mot anglais, cependant aucune traduction est suffisamment satisfaisante pour l'utiliser. Ce point sera détaillé plus en avant lors de la présentation de cette méthode.

Maintenant, si l'on s'attarde à la simulation aléatoire, de manière plus classique, l'une des étapes clés va consister à générer un aléa. L'idée est de générer une loi uniforme, dont les valeurs sont comprises entre 0 et 1 (exclus). Cela permet de disposer d'une probabilité d'occurrence d'un évènement. A partir des lois et des paramètres estimés, il devient alors possible de convertir ces probabilités en nombre de sinistres ou en coût unitaire et plus génériquement, on peut construire des échantillons de très grande taille de la loi estimée. Par exemple, il est possible de générer aléatoirement un échantillon de 10 000 valeurs représentant le nombre de sinistres probables pour un exercice. Si pour chacune des valeurs, i.e. nombre de sinistres, de cet échantillon, l'on calcule aléatoirement un montant unitaire de leur coût, on obtient 10 000 charges sinistres annuelles. Cela permet d'obtenir une distribution empirique, par modélisation, de la charge annuelle des sinistres probables.

Sans entrer dans la technicité informatique et mathématique, il est important de parler de la graine ou *seed* lorsque l'on parle de générateurs de type Monte-Carlo. La graine est une mesure mathématique qui se doit de respecter plusieurs conditions et qui permet de générer pseudo-aléatoirement un très grand nombre de valeurs. Il faut savoir qu'un mauvais choix de graine peut entraîner une répétition des valeurs générées, ce qui est clairement inadapté pour générer un aléa.

Il faudra retenir que fixer une valeur comme graine peut s'avérer très pratique. En effet, toute



personne réalisant la même étude, avec les mêmes données donc, pourra reproduire les mêmes résultats. Dans le milieu professionnel, pouvoir retrouver les résultats menant à la prise de décision est primordial.

2.3 Comment fonctionne le Business Plan actuel

Avant d'aller plus loin dans l'étude, il semble important de présenter le business plan et son fonctionnement actuel. Puis de discuter les modèles actuels, leurs avantages et leurs axes d'amélioration. Tout d'abord il est nécessaire de faire un point de contexte.

2.3.1 A quels enjeux ce BP fait-il face ?

Le business plan est une traduction chiffrée de la stratégie du groupe, il s'appuie sur l'échange et la collaboration entre plusieurs entités et spécialistes. Sa réalisation commence dès lors que la stratégie du groupe est fixée, que ce soit sous forme d'objectifs quantitatifs ou encore par des enjeux qualitatifs, comme une meilleure sélection des risques.

Afin de s'assurer de la bonne interprétation de la stratégie qui est définie à huis clos, les équipes sont amenées à définir sous confidentialité, des notions de développement, d'évolutions des coûts des risques comme de fonctionnement, de chiffre d'affaires, de marge et de rentabilité.

Le BP étant réalisé sur le second semestre, les objectifs chiffrés définis pour l'année en cours et surtout pour la première année sont engageants pour les directions commerciales qui devront mettre les moyens nécessaires pour atteindre ces objectifs arrêtés conjointement. Cela nécessite donc, en plus d'estimations chiffrées, basées sur des notions mathématiques et sur l'analyse de la conjoncture, une dose de négociation et d'adhésion. A noter que la décision finale est à la main des directeurs généraux des marques ou de leurs représentants pour certains partenariats.

Il en va de même pour différentes hypothèses qui font partie des modèles, la validation pouvant être à la main de directeurs transverses, notamment sur la gestion des sinistres qui se fait à l'échelle de Covéa et non plus des marques.

Le BP permet donc, une fois constitué, de présenter des trajectoires partagées et engageantes pour l'ensemble des directions et des marques. C'est également un document qui permet aux équipes en charge des diffusions de Solvabilité 2 (pourra être écrit S2) de fournir une note actuarielle et qui leur permet de jouer différents scénarios de choc dans le cadre de l'ORSA.

2.3.2 Quelles sont les composantes du BP Covéa ?

Comme explicité précédemment, le business plan regroupe diverses notions, plus ou moins figées, qui sont à projeter sur cinq années. Seront détaillées ici les diverses composantes et leur imbrication.

2.3.2.1 Pour se donner une vue d'ensemble

Un empereur français avait coutume de dire qu'un bon croquis vaut mieux qu'un long discours, aussi pour prendre en considération ces propos, la première phase sera de détailler les grandes étapes du business plan, pour la partie à la main de la direction IARD, dans sa forme actuelle.

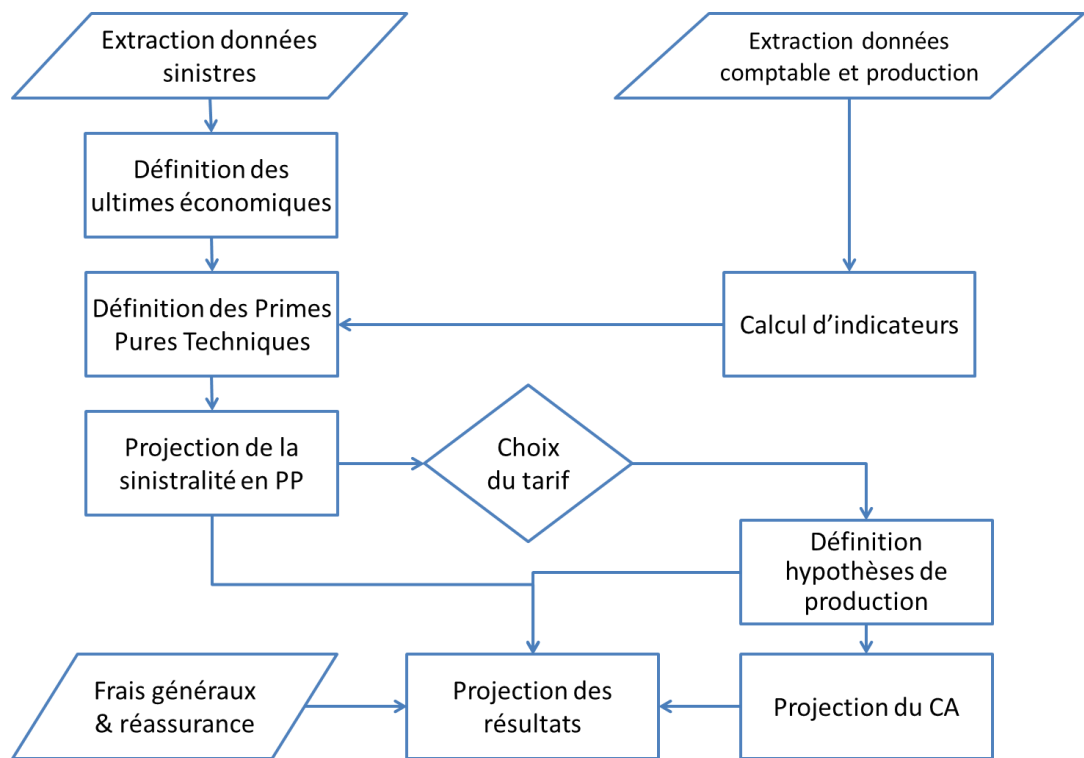


FIGURE 2 - SCHEMA SIMPLIFIE DE LA PROJECTION A 5 ANS DES RESULTATS

2.3.2.2 S'arrêter sur les données pour comprendre les enjeux

Selon les marques, l'historique des données peut remonter jusqu'à 30 ans, sur d'autres il se limite à 8 ou 10 ans. Cela restreint certaines approches combinatoires, en effet avec des historiques de tailles différentes, il est ardu de s'assurer des corrélations entre les marques. Sur les sociétés retenues pour l'étude, l'historique remonte jusqu'à l'exercice 2000 et les dernières valeurs sont vues au 31 décembre 2017, ce qui laisse 18 années d'observation. Bien que des visions mensuelles soient disponibles, il leur aura été préféré des visions annuelles, cela pour simplification.

Sur une période aussi longue, diverses complications apparaissent, comme l'évolution des produits, donc des couvertures et de l'indemnisation, ou encore les évolutions juridiques et réglementaires qui peuvent apporter de fortes variations sur de telles périodes. S'ajoutent aussi le développement des portefeuilles et l'inflation qui peuvent rendre les séries obsolètes.

Toutefois, il semble important de conserver le plus long historique possible. A défaut de l'exploiter pour le calcul du développement, cette profondeur peut servir à définir les facteurs de queue – *tail factor* – qui font défaut à certaines méthodes ou types de données.

Cela amène donc aux types de données utilisées. Historiquement l'équipe en charge du pilotage de la performance économique travaille des triangles de règlements, ou de paiements et des triangles de charges dites "gestionnaires" ou estimées par les gestionnaires, elles sont nettes de recours. A noter que les recours sont comptabilisés, toujours selon les marques, soit lorsqu'ils sont effectifs et perçus, soit lorsqu'ils sont estimés. L'équipe privilégie l'approche en charges estimées par rapport aux règlements, car ces triangles présentent plus rapidement une vision proche de l'ultime et pour les liquidations longues ils sont moins sensibles au facteur de queue. En effet, sur des risques corporels, avec versement d'une rente viagère, les triangles de règlement, même avec 18 exercices de recul, ne suffisent pas pour atteindre la charge à l'ultime sans ajouter un dernier facteur.

Dans l'étude suivante, les sociétés A et B ont toutes les deux la même approche de leur gestion des sinistres et, sauf mention contraire, les triangles étudiés portent sur la charge gestionnaire.

2.3.2.3 Projeter à la fin des temps les sinistres des exercices passés

Cette étape, cruciale, consiste à détailler les méthodes actuellement en vigueur, leurs hypothèses et prérequis.

Trimestriellement, les charges à l'ultime sont calculées avec le progiciel ResQ par des méthodes standardisées de type chain-ladder. Ces travaux sont faits pour toutes les garanties, après regroupement des garanties à faibles enjeux et à liquidations courtes dans un ensemble divers. On s'attardera sur certaines facilités proposées par ResQ, qui permettent de retirer les points jugés atypiques, soit dans leur développement, soit dans leur valeur. Cette opération de nettoyage est réalisée à dire d'expert, par et selon la sensibilité du réalisateur de l'étude.

Il semble intéressant de noter que les méthodes proposées, après nettoyage des triangles, permettent de nuancer la méthode classique du chain-ladder limitant l'utilisation des données à une partie de l'expérience disponible, en définissant les coefficients de développement comme la moyenne des 4 dernières années exploitables. Pour faire référence à ces modèles, on trouve parfois le terme "parallélogramme de liquidation", à noter que les formes géométriques à trous, pour cause de nettoyage des données, n'ont pas de nom.

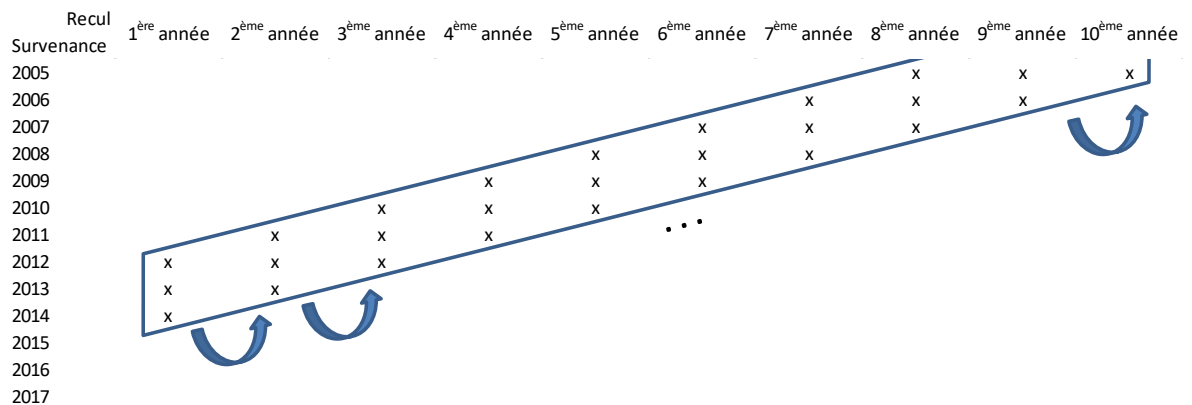


FIGURE 3 - ILLUSTRATION D'UN PARALLELOGRAMME DE LIQUIDATION

Pour les garanties plus atypiques, comme les graves (sinistre excédant un seuil, à compter du premier euro) ou les risques climatiques, d'autres méthodes de liquidation sont appliquées. On retrouvera notamment les méthodes de Bornhuetter-Ferguson ou de Benktander. L'outil ResQ permet d'appliquer un grand nombre de méthodes non détaillées dans ce rapport.

Les méthodes de BF nécessitent des charges de référence, à ce titre des charges techniques sont utilisées. Ces charges sont obtenues en multipliant la prime pure technique (cf. § *L'estimation de la Prime Pure technique*) par l'exposition au risque. On rappellera que l'exposition au risque d'une année close est connue, aussi seule la prime pure technique est à définir.

2.3.2.4 Comment est estimée la Prime Pure technique

Les charges à l'ultime représentent une vision conjoncturelle de la charge annuelle, ainsi elles prennent en compte la réalité économique, comme des émeutes, des tempêtes ou encore la connaissance des dossiers corporels et de leurs jugements. Cependant, cette vision économique présente un défaut, il ne s'agit pas d'une vision projetable, elle peut fortement évoluer d'une année à l'autre selon l'occurrence d'un sinistre grave de 25 millions d'euros ou d'une forte inondation. Afin d'asseoir les projections du business plan sur des bases plus stables, il est

d'usage de calculer une charge dite technique.

Toutefois, la notion de charge présente des défauts lors de sa projection, elle est notamment sensible à l'évolution de l'exposition au risque, autrement dit à la taille du portefeuille. Aussi l'indicateur le plus stable à projeter est la prime pure. Parfois notée PP, en assurance IARD elle se calcule comme la charge divisée par l'exposition, c'est aussi la fréquence multipliée par le coût moyen.

$$PP = \frac{\text{Charge Sinistre}}{\text{Exposition}} = \frac{CS}{\text{Nb sinistres}} \times \frac{\text{Nb sinistres}}{\text{Exposition}} = \text{coût moyen} \times \text{fréquence}$$

Les méthodes utilisées actuellement pour définir les primes pures techniques utilisent l'une ou l'autre des parties de l'équation précédente. Certaines méthodes sont plus poussées que d'autres selon le risque, sa volatilité et l'impact qu'il peut avoir sur les résultats.

2.3.2.4.1 Méthode générique qui s'applique aux garanties de masse

Pour les garanties de masse, l'approche est simplifiée du fait de la rapidité de sa liquidation. Il sera alors possible de travailler directement à partir des visions économiques.

L'objectif consiste alors à lisser dans le temps la prime pure. Pour ce faire il est plus intéressant de distinguer la fréquence et le coût moyen, car l'inflation peut intervenir sur le coût tandis que les évolutions technologiques peuvent réduire la fréquence tout en augmentant le coût des réparations.

On notera que l'étude porte sur des séries temporelles et que les modélisations linéaires sont la principale approche mathématique en vigueur. Ces méthodes consistent à établir des courbes de tendance sur l'historique de fréquence et de coût moyen en vision économique. Parfois, pour neutraliser un point atypique, jugé accidentel, une interpolation linéaire pourra être réalisée.

On notera que le lissage, notamment pour donner moins d'importance à une année atypique, peut entraîner une modification de la moyenne de la série étudiée. L'ensemble des points de la série pourra aussi être retravaillé pour conserver la moyenne initiale. Ce choix est à la main de l'actuaire, selon l'origine de l'écart, s'il est explicable, s'il faut prévoir qu'il puisse se reproduire et qu'il faille donc provisionner sa répétition...

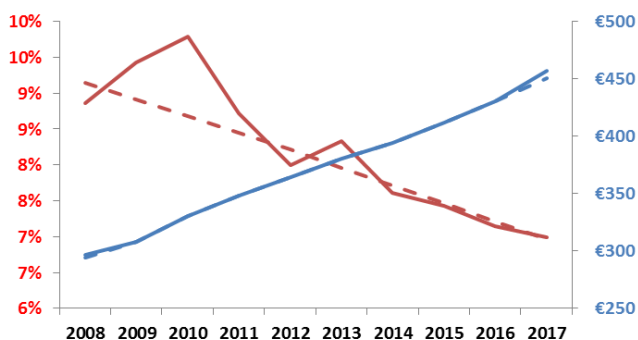
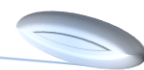


FIGURE 4 - ILLUSTRATION LISSAGE TECHNIQUE FREQUENCE ET COUT MOYEN



2.3.2.4.2 Une méthode plus avancée pour les garanties climatiques en habitation

Dans la pratique, aujourd'hui la prime pure climatique est définie en 3 étapes. Sont modélisées les primes pures en catastrophe naturelle de type sécheresse et hors sécheresse (principalement le risque inondation) et la prime pure des autres risques climatiques, dits TGN, acronyme de Tempête Grêle Neige.

Pour tous ces modèles, l'idée sous-jacente est la même, il s'agit de simuler par des approches de Monte-Carlo des exercices de sinistralité type, selon les lois de nombre et de coût.

De manière générique, l'estimation se fait en 4 étapes. La première étape consiste à estimer le nombre d'évènement par une loi (Poisson ou Binomiale Négative en général), l'hypothèse la plus importante étant que le nombre d'évènements soit stable dans le temps. Cette hypothèse est actuellement remise en cause du fait du réchauffement planétaire et de son accélération. Le pôle des Risques Majeurs de Covéa, spécialisé notamment dans les sciences climatiques, fait actuellement des recherches sur l'impact de ce facteur dans le milieu assuranciel.

La seconde étape va permettre de définir l'intensité des évènements. Ici seront modélisés les nombres de sinistres par évènement, après mise en équivalence. Il s'agit, avant de chercher les paramètres du modèle de la loi, de réestimer le nombre de dossiers qui auraient été ouverts si chaque sinistre avait eu lieu en année n , par exemple 2017. Pour cela, on applique un correctif basé sur l'évolution du portefeuille assuré entre l'exercice de survenance de l'évènement, noté t , et l'exercice technique à estimer, soit ici n (2017). Si l'on note P l'exposition au risque, avec P_i le nombre de risques assurés l'année i et N_t^e le nombre de sinistres de l'évènement e survenu l'année t , alors on obtient le N_n^{e*} nombre de sinistres après mise en équivalence

$$N_n^{e*} = N_t^e \cdot \frac{P_n}{P_t}$$

Après cette mise en équivalence, il est possible de constituer un échantillon d'évènements qui auraient une intensité équivalente à celle de l'année n . Cet échantillon permet d'inférer les paramètres de lois pouvant caractériser la population totale. Là encore, les lois testées seront des lois discrètes. A noter que des lois continues sont testées et qu'elles peuvent être discrétisées lors de la phase n°4.

La troisième étape concerne les coûts, un fois le nombre d'évènements modélisé et son intensité définie, il reste à identifier le coût unitaire de la réparation. Là encore une étape de préparation des données, par mise en équivalence, est nécessaire. Selon les contrats, une indexation peut s'appliquer aux valeurs assurées et également au tarif. Cet indice peut être propre au coût de la construction, en général l'indice de la Fédération Française du Bâtiment ou FFB, ou propre à l'évolution du coût de la vie, en général issue des statistiques de l'INSEE, l'Indice des Prix à la Consommation hors tabac et en France métropolitaine. Parfois d'autres indices s'appliquent ou des composites de ces derniers.

L'objectif de cette mise en équivalence est de redresser le coût unitaire annuel selon l'évolution du coût des réparations pour obtenir un échantillon le plus comparable possible. Il faut garder en tête que si un exercice est très sinistré, il peut se suffire à lui-même pour modéliser le coût unitaire. En général, on préférera intégrer le plus d'expériences possible pour modéliser au mieux la queue de la distribution.

Pour illustrer cela, lors d'une inondation l'indemnisation portera à la fois sur les biens mobiliers

– y compris électroménager, hi-fi, ... – et sur les biens immobiliers, comme les prises électriques, la peinture, le plâtre... Il est donc fort probable que l'inflation passée du coût soit proche d'un indice composite de l'IPC et du FFB, même si le niveau exact est propre à chaque sinistre. A noter que d'un dossier à l'autre, selon les critères et clauses du contrat, ou encore par le type de biens assuré, le coût unitaire varie. Il semble logique qu'une tempête touchant une résidence haut de gamme génèrera un coût très différent que celui généré par une même tempête touchant un logement standard. C'est bien cette diversité que l'on cherche à modéliser après la mise en équivalence.

Dans ces modèles continus, seront privilégiées des lois de Pareto, ou Pareto généralisée, des lois log-normale ou encore de Weibull, Gamma... Pour choisir la loi, l'approche graphique est un bon moyen de réduire le champ des possibles. Notamment en comparant les lois de distribution empirique et théorique, graphiquement ou via le tracé des courbes quantiles théoriques selon les quantiles empiriques – on parle de *Quantile-Quantile Plot* en anglais – qui doivent s'approcher d'une droite. Si l'approche graphique ne suffit pas, il faut étudier les statistiques de test d'adéquation, comme le test d'adéquation du Khi-2 ou le test de Kolmogorov⁴ Smirnov⁵, voire le test de Kuiper⁶. Au regard de la méthodologie utilisée pour définir la loi et ses paramètres, il est recommandé d'utiliser le second test.

Une fois l'ensemble des composantes d'une année climatique estimé – pour rappel la loi du nombre d'évènement, la loi d'intensité de l'évènement et la loi de coût unitaire – il est alors possible de simuler des exercices théoriques, on parle aussi d'exercices types. Pour cela, il suffit de générer aléatoirement S simulations, qui seront chacune une représentation théorique d'un exercice de survenance. Pour chaque simulation, un nombre d'évènements climatiques sera défini selon la loi arrêtée, pour chacun de ces évènements, une intensité sera aléatoirement tirée selon la loi d'intensité et enfin, pour chaque sinistre attendu un coût lui sera affecté. Ainsi pour toutes les S simulations, une charge est obtenue, ce qui permet de définir une charge annuelle moyenne et sa distribution.

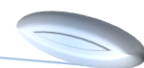
La charge technique ainsi obtenue est divisée par le nombre de contrats en vigueur l'année n , ce qui permet d'obtenir la Prime Pure Technique.

2.3.2.4.3 Une approche très technique pour la RC corporelle grave en auto

Sont considérés comme sinistres graves, ceux dont le montant dépasse 150 000€, soit environ 450 dossiers par an pour Covéa. Tout comme pour le risque climatique, il y a de gros enjeux sur cette garantie et la modélisation se fait par une approche calculatoire. Il y a plusieurs phases, tout d'abord une phase de préparation des données très complète, puis une phase d'analyse pour définir les lois de nombre et de coût et enfin une étape de simulation.

Pour parler rapidement de l'enjeu, au niveau du marché, hors flotte de véhicules – donc hors transport en commun – le plus gros sinistre estimé en vision comptable des réassureurs, est de 43 millions d'euros. Pour Covéa, le plus gros dossier connu à ce jour est estimé, en vision la plus réaliste, à 25 millions d'euros, et annuellement plusieurs sinistres de plus de 8 millions d'euros sont déclarés.

A noter que dans le temps, de très fortes fluctuations peuvent survenir, suite à une décision du tribunal qui serait ou non en la faveur de l'assureur. Il n'est pas rare que certains sinistres soient provisionnés pour leur intégralité, même si l'on espère un partage de responsabilité, voire une non garantie. Cela contribue à complexifier l'estimation de la charge ultime, pour exemple la non linéarité qui ressort des tests réalisés pour l'application des Chain-Ladder.



Dans le cadre de l'étude de la prime pure technique, la première étape va consister à mettre en équivalence l'historique des dossiers, comme s'ils avaient lieu l'année estimée. Pour cela sont actualisés les principaux postes de préjudice, principalement ceux payés en rente, à savoir la tierce personne, les dépenses de santé et la perte de gain professionnel futur.

Ce sujet mériterait à lui seul plusieurs chapitres, d'ailleurs les travaux présentés et utilisés au sein de Covéa s'appuient sur le mémoire d'actuariat de Sabrina SAVARE et Benoit PAIRE⁷. Bien que leur sujet fût basé sur les flottes automobiles, la modélisation des coûts unitaires et le modèle log-log, a été réalisée sur une volumétrie plus importante de victimes et s'adapte très bien aux divers portefeuilles des sociétés du groupe, en assurance automobile.

Il faut retenir que la tierce personne correspond aux dépenses faites pour assurer à la victime des soins et/ou une aide dans la réalisation des tâches de la vie courante. Dans l'estimation de cette charge, sont intégrés le nombre d'heures journalières, le coût horaire et un facteur de rente, basé sur l'espérance de vie de la victime.

Il est courant que les jugements attribuent un nombre d'heures d'assistance supérieur à 24h par jour, ce qui signifie que la victime a besoin de plus d'une personne. Il faut également savoir que certains créneaux de présence sont considérés comme passifs, par exemple la nuit, la tierce personne doit être présente, pour autant ses contributions seront réduites aux sollicitations de la victime.

A partir des éléments en vigueur pour l'année t de survenance étudiée, il est possible d'actualiser le coût horaire des sinistres de l'historique afin de redéfinir le coût des dossiers comme s'ils étaient survenus l'année t .

Enfin, pour la rente, l'autorité des normes comptables impose l'utilisation de la table TD 88-90 et un taux d'actualisation égal à 60% du TME moyen des 24 derniers mois. A noter que depuis 2015, suite à un désengagement du Fonds de Garantie d'Assurance Obligatoire – qui prenait à sa charge la revalorisation des rentes – pour tous les dossiers survenus après le 1^{er} janvier 2013, l'assureur a à sa charge la revalorisation du coût des rentes, l'ANC demandant une provision de 2,25%, révisée à 2% en 2019. On notera que l'exercice qui est fait dans le cadre de cette étude consiste à estimer un coût économique de la rente. L'idée est donc de savoir quel taux servir, après intégration de la revalorisation – qui correspond à l'augmentation du coût de la vie future – et du rendement des actifs, compte tenu des courbes de projections EIOPA ou de l'Institut des Actuaire. A ce jour, compte tenu des taux bas, le taux appliqué est nul. Cela revient à dire que le rendement attendu des placements, sur la durée de vie de chaque sinistre, ou plutôt en moyenne sur l'ensemble des sinistres, compensera l'évolution du coût de la vie. Aussi, le risque économique mesuré dans la rente est un risque viager.

Les pertes de gains professionnels sont, dans la pratique, transigées à l'amiable et versées sous forme de capital. Toutefois, dans l'estimation du coût, le poste est mesuré viagèrement. On notera que ceci pourrait paraître prudent, toutefois il n'est pas rare qu'une fois le capital utilisé, la victime demande une révision du montant. Pour la mise en équivalence, le salaire annuel de la victime est indexé sur l'évolution du salaire médian français de l'INSEE. On notera que les derniers exercices peuvent manquer dans les rapports de l'INSEE, aussi une extrapolation sur des bases linéaires est faite pour définir ces exercices récents.

Pour la rente, en se référant aux jurisprudences, il a été décidé de convertir la rente viagère en rente temporaire jusqu'à 65 ans – âge estimé de la retraite – pour prendre en compte le fait

qu'une fois passé cet âge, la victime aurait perçu une retraite, soit un autre type de préjudice évalué par ailleurs.

Les dépenses de santé futures correspondent aux frais qui seront engagés pour payer les soins de la victime, ici le raisonnement est plus simple, seule la rente est révisée, sur le même principe que pour la tierce personne.

Une fois ces principaux postes mis en équivalence, il en découle le nouveau coût unitaire de chacun des dossiers. On notera que la dernière vision connue de chacun des dossiers est ré-évaluée, aussi leurs montants restent difficilement comparables. C'est pour cela qu'ils sont projetés à l'ultime, sur la base des triangles calculés précédemment. Toutefois, le coefficient de développement qui sera appliqué est le même par ancienneté de sinistre, partant du principe que certains vont peut-être augmenter, d'autres baisser et que nos méthodes actuelles de liquidation se font sur la totalité des sinistres, l'application du développement est la même pour tous les sinistres d'une génération.

Pour les nombres de sinistres, après avoir défini le nombre de dossiers à l'ultime, une actualisation sur le portefeuille est requise, comme pour le risque climatique.

A partir de cette étape, il est possible d'estimer la loi de coût unitaire et la loi de nombre. Pour le nombre, étant donné la variance empirique assez forte au regard de la moyenne, il est souvent privilégié une loi binomiale négative. Pour le coût, la loi mélange log-log, telle que l'ont expliquée mes prédécesseurs – S. Savare et B. Paire – dans leur mémoire, s'applique très bien.

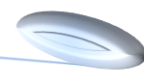
Actuellement les montants unitaires sont capés à 50 M€ dans les simulations. Cela étant, il est possible de voir apparaître ces montants jugés extrêmes. Ce montant se veut être une valeur impossible à atteindre et elle est discutable. Toutefois sans disposer des éléments détaillés du plus gros sinistre du marché (43 M€ en vision réassureur), il est compliqué de le passer dans une vision à l'ultime aux conditions économiques. Pour rappel, cette vision est une estimation du coût réel et elle consiste à appliquer les taux de revalorisation tels que définis par la loi 74-1118 du 27 décembre 1974, en estimant une valeur future sur la période de la rente et à intégrer une performance des rendements des actifs du groupe en fonction de la courbe des taux sans risques.

Néanmoins, il est possible d'estimer que son coût, suite à une révision des taux – en général avantageuse si l'on se positionne en rendement attendu et non au rendement de la courbe EIOPA – sera plus faible et certainement inférieur à 40 millions d'euros. Ce choix d'un coût in fine de 50 millions d'euros comme maximum a été arrêté avec la fonction actuarielle Covéa et sert de borne pour l'instant.

Si l'on revient à la méthodologie, tout comme pour les risques climatiques, la dernière étape va consister à simuler des exercices qui suivent les précédentes modélisations réalisées. Pour commencer, le nombre de sinistres par exercice simulé est tiré, puis pour chacun des n_i sinistres attendus pour la $i^{\text{ème}}$ simulation les charges des deux lois log-normale du modèle log-log sont calculées. Enfin un aléa pour chacun des n_i sinistres est défini sur la base d'une loi uniforme entre 0 et 1. Cet aléa est comparé au paramètre de seuil défini pour la loi mélange log-log. Selon les cas, la loi log-normale des graves classiques ou la loi log-normale des hyper-graves sera appliquée pour définir le montant à retenir.

Il conviendra ensuite de sommer par exercice simulé les montants pour obtenir une répartition de la charge. Il sera alors possible à l'auteur de définir, selon le niveau d'appétence au risque du groupe, le montant à retenir comme estimation technique de la charge.

Nota bene, le choix se fait en prime pure, l'exercice étudié étant clos, l'exposition au risque est connue, la prime pure est obtenue par division de la charge par l'exposition.



2.3.2.5 Définir l'évolution future de la Prime Pure sur la base du passé...

Une fois les primes pures techniques définies, ou pour les garanties de masse les fréquences et coûts moyens techniques arrêtés, ces éléments serviront à définir une tendance du passé qui pourra être extrapolée et proposée comme hypothèse de projection du futur.

On rappellera que le lissage se fait par l'étude de séries chronologiques avec des approches de type linéaire, ainsi, définir les points suivants, en moyenne, est assez simple. Toutefois, le choix est basé sur d'autres éléments, comme l'explication d'une évolution, est-ce une tendance de fond ou est-ce lié à une évolution du portefeuille ? Observons-nous ces tendances sur toutes les marques ? Voici des questions auxquelles il faut répondre pour valider les hypothèses.

On notera qu'étudier les fréquences permet de s'affranchir de l'évolution du portefeuille en volume. Cependant toute déformation de la structure du portefeuille, donc des risques assurés, entraîne des évolutions dans le passé qui ne seront pas nécessairement reproduites dans le futur. Par ailleurs, si l'approche en fréquence ou en prime pure, permet de s'affranchir du développement, l'approche en coût moyen soulève des questions quant à l'inflation passée et donc future.

C'est pour cette raison que ces modèles macro-économiques sont confrontés à des modèles plus raffinés qui, en auto, intègrent le coût des pièces ou de la main d'œuvre et leur évolution. Enfin l'intégration d'études sur la déformation de la structure permet de bien définir s'il faut anticiper une amplification du phénomène ou, a contrario, sa fin.

Ce mémoire se voulait être l'occasion de tester d'autres approches, comme des modèles auto-régressifs, toutefois cette partie aura été mise de côté pour des raisons de temps.

A noter que la valeur déterministe obtenue suite aux modèles actuels et aux échanges avec les experts servira de base à la projection stochastique. L'idée est d'intégrer un bruit blanc autour de la tendance, ce bruit étant différent pour chaque simulation, il induira une volatilité dans le résultat final.

2.3.2.6 ... Pour en déduire des mesures tarifaires adaptées ...

Les mesures tarifaires sont définies afin d'atteindre un objectif de S/P, elles s'appuient donc sur l'évolution attendue de la prime pure. Cependant, selon les produits, il se peut que la mesure tarifaire prise ait un impact plus faible, notamment si le tarif intègre des critères qui évoluent avec le contrat, comme l'âge du véhicule, le coefficient de réduction-majoration – souvent appelé bonus-malus. Ces éléments perturbent le modèle et nécessitent d'être connus.

On notera que la prime peut être indexée sur des statistiques du type IPC (Indice des Prix à la Consommation, INSEE) ou FFB (Fédération Française du Bâtiment). Ainsi une évolution minimale du tarif est obligatoire, sauf si l'assureur décide de blanchir tout ou partie de l'inflation réglementaire. Cette indexation est inscrite aux conditions générales.

Les mesures tarifaires étant révisées à chaque BP, il est envisageable d'intégrer au script informatique des règles de gestion pour corriger la trajectoire dynamiquement soit selon un objectif en S/P à atteindre. L'objectif étant d'intégrer une volatilité aux projections, pour encadrer le résultat. Cette idée a été écartée. Un tel processus serait une sorte d'intelligence artificielle visant à produire des BP dynamiques. Sans entrer dans un débat d'éthique, cela remet en cause l'analyse de toutes les étapes précédentes qui serait automatisée et les décisions

prises par la machine, est-ce là l'avenir de l'actuaire ?

2.3.2.7 ... Qui serviront à adapter les trajectoires commerciales

Cette étape est réalisée conjointement avec la direction commerciale, qui connaît ses capacités de vente, et différentes équipes de la direction IARD qui ont à leur main l'application du tarif, les budgets commerciaux ou encore la connaissance des évolutions des offres ou des produits. Tout arbitrage étant à la main des directeurs, le sujet pouvant être sensible et surtout engageant à tenir une fois validé.

L'objectif est donc de définir les trajectoires de production brute et nette, à savoir les nombres d'affaires nouvelles et d'affaires perdues, en prenant en considération le passé et selon la période de réalisation de l'étude, les tendances connues de l'année.

Là encore plusieurs approches s'allient pour obtenir un consensus. Une approche du chiffre, en décomposant les diverses opérations à des mailles plus fines, comme isoler les résiliations – ou affaires perdues – selon leur nature, est-ce une volonté du client ou de l'assureur. De plus, les départs à l'initiative du client sont décomposés, notamment pour prendre en considération des déformations sur les modèles. Pour exemple les lois Chatel ou Hamon, qui ont donné plus de liberté au consommateur dans la rupture de contrat, modifiant parfois les remboursements.

Une seconde approche d'élasticité au prix, ou de sensibilité du client face à l'évolution des prix, est réalisée, cette approche est enrichie des retours des commerciaux et de retours des clients. En croisant cela avec les effectifs et surtout les efforts à produire sur l'ensemble des secteurs, sachant que les conseillers vendent à la fois des produits IARD et des produits Vie ou Santé, la direction des réseaux veille à la tenue des objectifs et à la définition de ces derniers avec toutes les directions techniques.

A noter que ces hypothèses sont définies de manière déterministe sans approche stochastique.

2.3.2.8 Compiler les étapes pour estimer les futurs résultats techniques

La première étape consiste à définir, à partir des hypothèses de développement, l'évolution des portefeuilles. Dans un modèle déterministe, les portefeuilles d'ouverture, au 1^{er} janvier de chaque exercice, sont obtenus en ajoutant la production nette au portefeuille d'ouverture de l'année précédente.

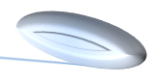
$$Ptf_{01.01.N+1} = Ptf_{01.01.N} + AN_N - AP_N$$

A partir des portefeuilles d'ouverture sont déduits les portefeuilles exposés au risque, on parle aussi de portefeuilles en contrats-année. L'approche macro-économique cherche à approximer cet estimateur, contrairement aux équipes de tarification qui comptent le nombre de jours sur l'année pendant lesquels chaque contrat a été actif.

On prendra les hypothèses suivantes :

- i) les contrats actifs au 01/01/n et présents 01/01/n+1 comptent pleinement
- ii) les contrats résiliés au cours du mois m contribuent pour $m - \frac{1}{2}$
- iii) les contrats activés le mois m, contribuent pour $\frac{1}{2} + 12 - m$

Autrement dit, l'hypothèse consiste à dire que les affaires nouvelles et perdues sont réparties uniformément au fil du mois.



Les trois hypothèses précédentes se traduisent mathématiquement comme :

- i) $\frac{m}{12} (Ptf_0 - \sum_{i=1}^m AP_i)$ et quand $m = 12$: $(Ptf_0 - \sum_{i=1}^{12} AP_i)$
- ii) $\frac{1}{12} \sum_{i=1}^m \frac{2i-1}{2} AP_i$
- iii) $\frac{1}{12} \sum_{i=1}^m \frac{2m+1-2i}{2} AN_i$

Avec Ptf_0 le portefeuille au 1^{er} janvier et AN_i et AP_i les Affaires nouvelles et perdues du mois i .

En regroupant et simplifiant les équations précédentes, on obtient l'équation

$$iv) \quad Ptf_0 + \frac{1}{12} \sum_{i=1}^m \frac{2m+1-2i}{2} (AN_i - AP_i) = \frac{1}{12} \sum_{i=1}^m \frac{Ptf_{i-1} + Ptf_i}{2}, \text{ avec } m = 12.$$

On remarque que l'approximation du portefeuille moyen annuel, en contrats-année, est la moyenne des portefeuilles moyens mensuels :

$$Ptf \text{ Moyen Annuel}_{\text{Contrats Année}} = \text{Moyenne des Ptf Mensuel Moyen} = \frac{1}{12} \sum_{i=1}^{12} \frac{Ptf_{i-1} + Ptf_i}{2}$$

Les hypothèses de production sont données pour l'année, la répartition mensuelle est déduite des précédents exercices, en prenant la part moyenne des 2 ou 3 derniers exercices. Ainsi les portefeuilles peuvent être projetés, ce qui servira à la fois à la définition de la charge sinistre et au calcul de la prime.

La **charge ultime future** est immédiate, c'est le produit de la Prime Pure, telle qu'estimée au chapitre précédent, par l'exposition au risque, en contrats-année.

La définition de la prime demande plus d'étapes, entre autres raisons car l'on distingue deux types de primes pour le business. Tout d'abord on estime la prime émise qui correspond à l'émission annuelle des contrats en vigueur, puis la prime acquise. La prime acquise est la prime prenant en considération l'exposition des contrats sur l'année étudiée. Les notions de développement mensuel serviront pour définir cette prime acquise.

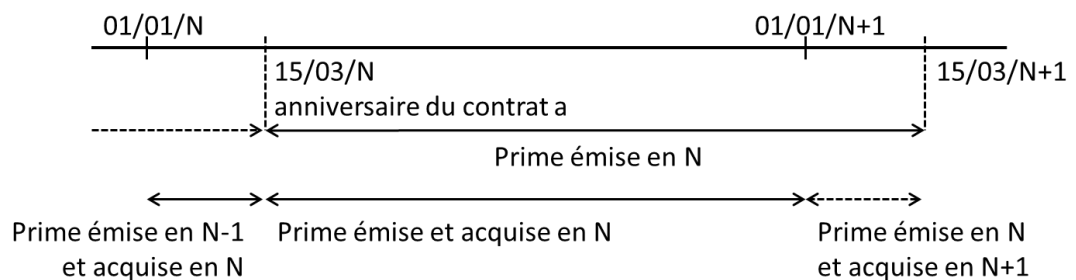


FIGURE 5 - ACQUISITION DES PRIMES DANS UN SYSTEME A EMISSIONS PLURIANNUELLES

Dans un système de fonctionnement où le tarif est émis à date anniversaire des contrats, cela signifie que le contrat se renouvelle de date à date. Ainsi si un contrat commence le 15 mars, la prime sera émise du 15/03/n au 14/03/n+1. La part de prime acquise pour ce même contrat sera, dans le cadre d'un renouvellement tacite, une part du tarif émis l'année précédente qui couvre le risque du 1^{er} janvier au 14 mars avec le tarif en vigueur l'année précédente, à laquelle s'ajoute la part de la prime émise cette année du 15 mars au 31 décembre, avec le tarif en N.

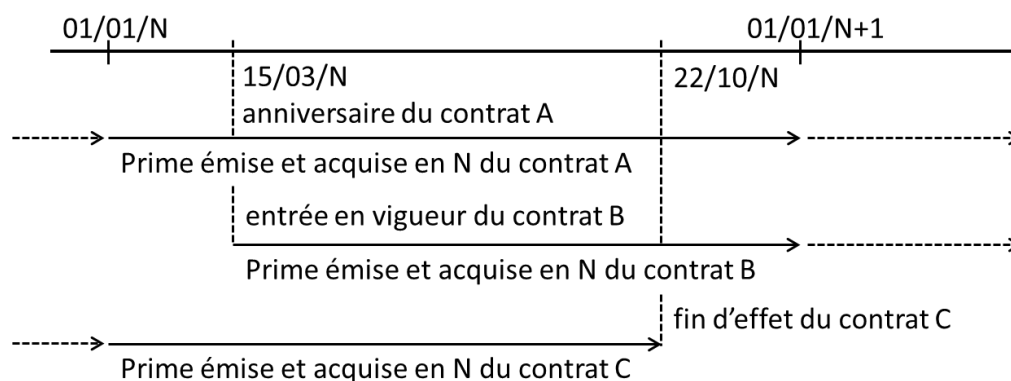


FIGURE 6 - ACQUISITION DES PRIMES DANS UN SYSTEME A EMISSION UNIQUE

Dans un système de prime émise unique, i.e. pour le cas où le prix est fixé jusqu'à une date t – généralement le 1^{er} janvier – alors la prime est émise de la date d'ouverture du contrat jusqu'au terme t de ce contrat. D'une année sur l'autre, si le contrat est inchangé et non mouvementé, alors sa prime sera émise du 01/01/ n au 01/01/ $n+1$ et sans mouvement, alors la prime acquise sera égale à la prime émise.

Comme stipulé en préambule, il y a au sein de Covéa plusieurs sociétés et les deux modes de gestion, ce qui demande donc des approches différentes par marque. Dans ce mémoire, nous étudierons uniquement les primes émises, ce qui permet d'utiliser des modèles comparables.

Enfin, pour définir la prime, il faut poser plusieurs hypothèses en plus de la production, comme l'impact net de la revalorisation des tarifs sur le portefeuille, l'impact de cette évolution tarifaire sur les contrats entrants et les avenants. De plus, seront intégrés au modèle les déformations de la structure et leurs impacts sur la cotisation moyenne. Cet effet structurel sera défini sur les mouvements de production, il sera ainsi propagé au portefeuille au fil des exercices projetés. La dernière hypothèse majeure est le budget commercial et surtout son évolution dans le temps.

Une fois ce modèle construit, il reste un reliquat inexplicé que l'on appellera erreur de modélisation. Il peut s'agir d'un écart lié aux approximations dans le calcul des indicateurs ou à des mouvements de gestion comptable qui ne seraient pas identifiés par le modèle, comme les impayés suite à mise en contentieux... Cette erreur est donc intégrée aux projections afin de corriger au mieux le modèle.

On peut également noter le modèle comme :

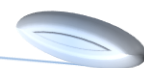
$$P_{n+1} = P_n \times \text{terme}_n + AN_{n+1} \times Cot_{n+1}^{AN} - AP_{n+1} \times Cot_{n+1}^{AP} + AV_{n+1} \times \Delta Cot_{n+1}^{AV} + E_{n+1}$$

Avec :

- P_n la prime émise de l'année n
- terme, l'effet du tarif sur la prime émise
- AN/AP/AV respectivement les affaires nouvelles, perdues et les avenants
- $Cot^{AN/AP}$ la cotisation moyenne des affaires nouvelles et perdues
- ΔCot^{AV} l'évolution de la cotisation avant et après avenant qui peut être négative

On notera que ce modèle est un modèle additif, qui s'appuie sur des hypothèses, voire même des hypothèses issues d'hypothèses. Il donne un résultat unique sans prise en considération de l'erreur d'estimation ni de la réalisation des hypothèses par le passé. Etudier la volatilité et faire en sorte d'appliquer un modèle dynamique sera étudié dans ce mémoire.

2.3.2.9 Intégrer et adapter les éléments externes pour définir la marge



d'exploitation

Pour cette dernière partie, des éléments fournis par d'autres directions sont intégrés au BP afin de constituer tous les indicateurs, jusqu'au Ratio Combiné.

Sans entrer dans les détails, on pourra noter que les frais de gestion sont estimés en trois parties, les frais d'acquisition, les frais de gestion des sinistres et tous les autres frais, notamment les frais d'entretien. Ils sont donc calculés sur la base des portefeuilles, des primes et aussi de l'évolution des fréquences, ainsi les contrôleurs de gestion s'appliquent à utiliser les projections des indicateurs définis par les directions techniques pour estimer une enveloppe globale de frais et la ventiler entre les diverses branches d'activité.

Pour la réassurance, plusieurs composantes sont analysées par programme de réassurance. On détaillera ainsi les primes versées, les charges cédées et les commissions acceptées. Sur certains traités, une Participation aux Bénéfices (PB) est également estimée.

Pour la partie prime, la tâche consiste à projeter les primes payées par exercice comptable en une prime totale par exercice de survenance, en intégrant donc les paiements connus et en utilisant la connaissance des exercices antérieurs pour estimer les appels complémentaires futurs. On parlera de primes cédées économiques. Sur les exercices futurs, cette prime évolue à la même vitesse que le portefeuille, en prime, à laquelle elle se rapporte.

Pour les commissions, elles sont déduites immédiatement par application des textes des traités sur la base des coefficients à appliquer aux primes.

Pour la charge sinistre cédée, selon les traités le passage en vision économique est plus ou moins immédiat, en effet si la gestion est en Quote-Part (QP) il suffit d'appliquer le taux aux sinistres évalués. Pour les autres types de traités, notamment en excédents de charge sinistre (XS) l'enjeu est de définir s'il y a activation du traité. A noter que sur les exercices futurs, c'est l'espérance de la charge cédée du passé qui sera retenue, sous réserve de l'inflation attendue sur l'assiette de garanties à laquelle le programme se réfère. Il faudrait mettre cette partie en relation avec les simulations de Monte-Carlo pour estimer une charge moyenne technique de cession. Ce sujet est en cours de réflexion et sera à mener avec la direction en charge de la réassurance du groupe. En effet, de telles méthodes serraient de bons supports pour les négociations avec les réassureurs.

Enfin, pour les PB, cette charge de sinistres cédée est alors obtenue à partir des postes calculés précédemment.

Une fois le solde de réassurance adapté à une vision éco-technique, il est possible d'estimer un S/P net de réassurance, plus élevé que le brut, sauf cas rare, par exemple d'évènements climatiques de grande ampleur ayant activé des programmes du type cumul de rétention, qui visent à se couvrir d'une année trop riche en évènements climatiques.

Si l'on ajoute les frais de gestion, l'indicateur obtenu est le ratio combiné, qui représente le rapport entre l'ensemble des dépenses et l'ensemble des revenus. Dit autrement, si ce ratio est inférieur à 1, le segment est rentable et génère de la marge, sinon il est déficitaire.

On pourra en déduire la marge attendue à 5 ans, avant inclusion des produits financiers, sous réserve de réalisation de toutes les hypothèses prises pour le Business Plan.

3 Etudier la volatilité des sinistres, un pas vers le stochastique

Cette partie va consister à présenter et débattre des méthodes qui permettent d'induire l'aléa dans les résultats et avant tout à tester si et comment les modèles peuvent s'appliquer.

3.1 Quelles méthodes tester et selon quelles conditions ?

Cette première partie va consister à rappeler les principales méthodes en vigueur sur le marché et l'historique de ces dernières. Ce sera également l'occasion de s'attarder sur les conditions d'utilisation de certaines méthodes, comme le chain-ladder, pour tester si elles sont applicables aux données disponibles. Enfin, une décision quant à l'application de la méthode à retenir sera débattue.

3.1.1 La méthode de référence est le chain-ladder

Pour la plupart des études des garanties dites de masse, l'équipe utilise des méthodes de type chain-ladder, aussi nous vérifierons pour le bloc garantie de masse que les hypothèses du chain-ladder s'appliquent. On rappellera que cette méthode est la référence de départ de l'étude.

Le chain-ladder peut se résumer à 2 hypothèses, tout d'abord il doit y avoir indépendance des exercices de survenance et enfin, il existe un coefficient de développement unique, tel que :

$$E(C_{i,j+1} | C_{i,1}, \dots, C_{i,j}) = f_j \times C_{i,j}$$

Avec $C_{i,j}$ la charge cumulée de l'exercice i , vue en j .

Cette seconde condition se traduit par l'existence d'une droite absolue lorsque l'on trace les différents points $C_{i,j+1}$ en fonction des $C_{i,j}$. Graphiquement la condition se vérifie donc si les points s'alignent sur une droite passant par l'origine et l'on pourra s'appuyer sur le coefficient de détermination du modèle linéaire, le R^2 , qui sera gage de qualité lorsqu'il sera proche de 1.

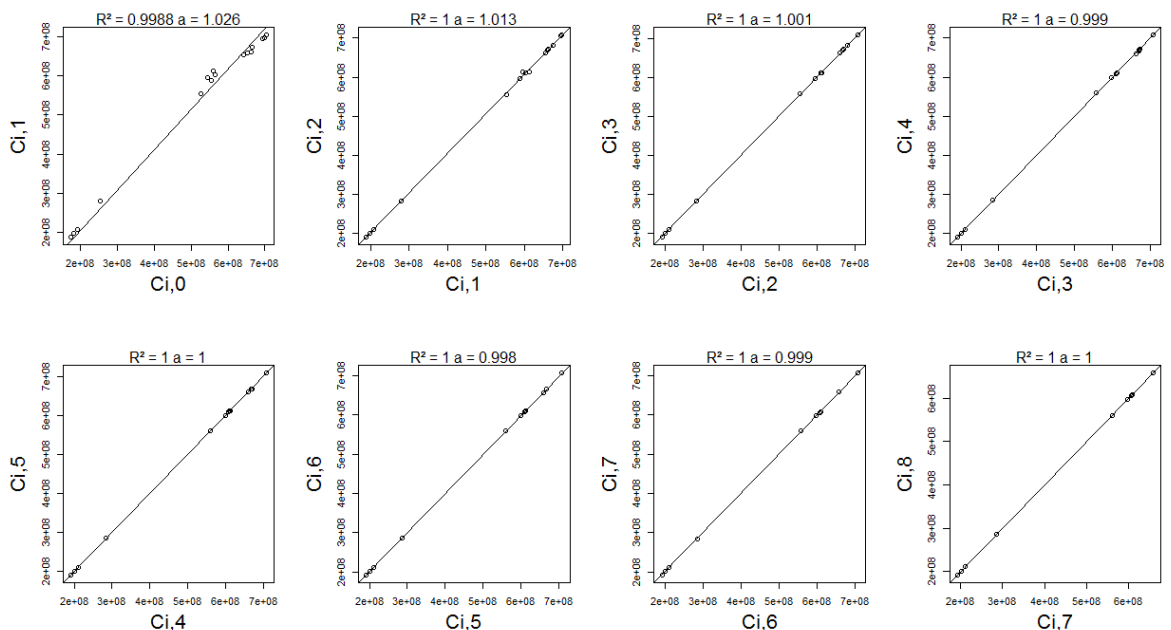
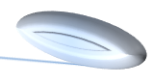


FIGURE 7 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE AUTO DE LA SOCIETE A



Dans le test précédent, l'alignement s'observe dès la première année de développement, et le R^2 est très proche de 1. On notera qu'au terme de l'année de survenance quelques couples sont hors de l'alignement et que cela est encore le cas sur d'autres exercices. Bien qu'imparfaits, ces graphiques laissent penser que la méthode de chain-ladder a du sens et peut s'appliquer. Le constat est le même sur les garanties de masse en auto comme en habitation, pour les deux sociétés. Les graphiques sont disponibles en annexe A.

Si l'on s'arrête sur les couples $(C_{i,0}, C_{i,1})$, ils peuvent parfois suivre une droite ne passant pas par l'origine. Ce constat pose la question de l'intégration d'un correctif dans le calcul du coefficient de passage de l'année de survenance à la première année de développement.

Dans le cas présent, l'équation deviendrait :

$$E(C_{i,1} | C_{i,0}) = f_0 \times C_{i,0} + \beta_0 = f_0 \times (C_{i,0} + \beta_0/f_0) = f_0 \times (C_{i,0} + \alpha_0)$$

De fait, la vision à l'ultime (en $t=u$) serait alors :

$$E(C_{i,u} | C_{i,0} \dots C_{i,u-1}) = (C_{i,0} + \alpha_0) \times \prod_{k=1}^u f_{k-1}$$

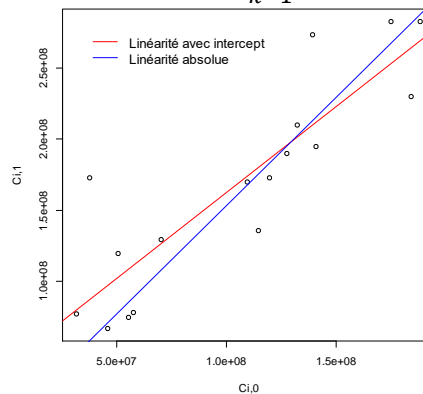


FIGURE 8 - ALIGNEMENT DES COUPLES $(C_{i,0}, C_{i,1})$ GARANTIE GRAVES AUTO DE LA SOCIETE A

A noter que cette correction apportée aux données se base sur l'estimation de l'équation de la droite rouge de la figure viii. Elle possède donc une erreur, cela pourrait être utile plus tard, lors de l'estimation par bootstrap du résultat.

La suite des couples se rapproche très rapidement d'une relation linéaire, aussi cette correction serait à apporter uniquement sur la projection pour l'exercice de survenance. De plus, dans le cas des garanties de masse, passée la première année de développement, la charge est quasi stable.

Si l'on élargit les tests aux garanties graves et climatiques, l'alignement est moins bon, cependant la linéarité reste très satisfaisante. Après 3 exercices, on retrouve une linéarité assez bonne, bien que sur les graves l'on constate régulièrement quelques points hors de la droite.

Dans la suite de l'étude, pour simplifier les calculs et pouvoir appliquer des méthodes plus avancées, il sera supposé que l'ensemble des triangles satisfait à cette hypothèse. L'ensemble des graphiques, pour chacun des triangles de charge, est disponible en annexes.

Pour tester la seconde condition, celle d'**indépendance**, le plus simple est de réaliser un test de corrélation. On rappellera que l'indépendance entre deux variables aléatoires entraîne la nullité de la corrélation, ainsi si ces variables sont corrélées, elles ne peuvent être indépendantes.

A noter que la réciproque est fautive, ainsi prouver que les corrélations sont nulles n'impliquera pas nécessairement l'indépendance, cela permettra de savoir s'il y a lieu de pousser plus en avant les tests.

Pour tester la corrélation, sur des données numériques continues, le test de Pearson est le plus communément utilisé, il sera le premier à être calculé. Une approche avec les méthodes de Kendall ou de Spearman – qui permettent de détecter des corrélations liées à l'ordre et qui sont plus robustes – sera également testée.

Le coefficient de Pearson s'obtient en divisant la covariance de 2 variables par le produit de leurs écarts-types.

$$\rho(C_i, C_j) = \frac{\text{cov}(C_i, C_j)}{\sigma(C_i) \cdot \sigma(C_j)}$$

Avec C_i la variable aléatoire contenant les paiements cumulés pour l'exercice de survenance i . On démontre que si 2 variables sont indépendantes, alors elles sont non corrélées, i.e. $\rho=0$.

Dans les données utilisées, notamment au sein des triangles de charge climatique, les exercices sont corrélés :

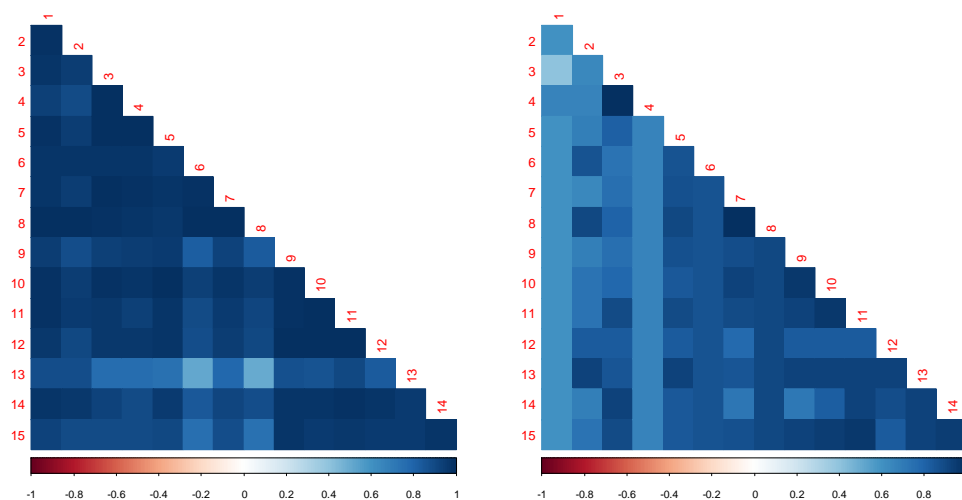


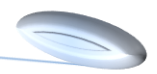
FIGURE 9 – CORRELATIONS DES EXERCICES DU TRIANGLE DES CHARGES CLIMATIQUES HABITATION SOCIETE A (PEARSON – KENDALL)

Tout comme pour les tests précédents, l'ensemble des résultats est disponible en annexe B.

Au regard des résultats, quel que soit le triangle étudié, il est quasi-impossible de conclure à l'indépendance des exercices.

Dans la littérature, un test graphique sur la stabilité des coefficients de développement et leur écart au coefficient moyen est souvent testé, il sert à démontrer la légitimité du test. Ici ce test n'est pas réalisé. En effet, cette variation entre le coefficient de développement moyen et le développement annuel, propre à chaque exercice, est source de volatilité. Dans le cadre d'une projection stochastique, notamment par bootstrap, comme ce sera le cas dans cette étude, cette information est importante.

Ainsi, malgré la dépendance observée sur les données, nous appliquerons les liquidations par chain-ladder comme méthode de référence. Toutefois, l'absence de lien linéaire entre la charge constatée de deux exercices continus poussera à tester, voire à privilégier d'autres approches, comme celle de Bornhuetter-Ferguson.



Par ailleurs, il est également possible de discuter de l'indépendance des exercices. En effet, si mathématiquement les séries sont corrélées, la notion de dépendance au sens de la survénance est, elle discutable. Au sein d'un exercice, le nombre de sinistre est prédictible, cela signifie qu'il existe bien des relations modélisables. Toutefois, dire que le 24 mars sera une journée avec un nombre de sinistre connu, définissable et projetable dans le temps est faux. Dans les triangles de liquidation, sont étudiés les développements des charges sinistres entre 2 dates de clôture d'exercice. Par conséquent, il est logique de retrouver des mouvements proportionnels et donc corrélés d'un exercice à l'autre. Si l'on étudie un pas plus court, journalier ou hebdomadaire, la saisonnalité et donc la corrélation devrait s'amoindrir. Par exemple, les fortes tempêtes de fin d'année 1999 ont pu avoir un développement très atypique dans la charge à N+1, l'ampleur et la date d'occurrence étant suffisamment exceptionnelles pour perturber le modèle.

3.1.2 Une autre méthode fait référence, celle de Bornhuetter-Ferguson

On rappellera que l'hypothèse de cette méthode consiste à dire que les sinistres inconnus à un instant t, vont se développer selon une charge sinistre espérée a priori. Cela implique de définir, par une méthode de type chain-ladder – si elle est adaptée – une cadence de liquidation type. A partir de là, il est possible d'ajouter à la charge connue des sinistres, une part de la charge théorique aussi appelée, au sein de Covéa, charge technique.

La définition de la charge à l'ultime par cette méthode s'appuie donc sur un chain-ladder classique, pour estimer la cadence a priori à l'instant t. Les conditions ont été vérifiées au paragraphe précédent.

Sans s'attarder sur la méthode de détermination et de simulation de la charge technique, définie dans la première partie, il est toutefois acceptable de dire que l'occurrence des sinistres suit une loi binomiale négative et que le coût s'apparente à 2 lois log-normales qui se chevauchent. Ainsi, nous pouvons en déduire que les conditions d'application de la méthode sont vérifiées.

Si l'on s'attarde sur la méthode en elle-même, elle présente également des restrictions. Ainsi, si la cadence de développement estimée est inférieure ou égale à 1 – dit autrement, si la charge connue est jugée supérieure à la charge ultime – alors la méthode devient inapplicable.

Mathématiquement, on peut rapporter le BF à

$$E(C'_{i,u}) = E(C_{i,u} | C_{i,0} \dots C_{i,t-1}, C_{.,t} \dots C_{.,u-1}) \times \frac{1}{\prod_{k=t}^u \hat{f}_{k-1}} + E(C_{i,u}^*) \times \left(1 - \frac{1}{\prod_{k=t}^u \hat{f}_{k-1}}\right)$$

Avec $E(C_{i,u})$ la charge ultime estimée par méthode chain-ladder et $E(C_{i,u}^*)$ la charge technique estimée a priori.

Si l'on définit la cadence de développement comme le produit des facteurs de développement f alors, la formule s'apparente à une pondération entre 2 ultimes.

Par conséquent :

$$\left(\frac{1}{\prod_{k=t}^u f_{k-1}}\right) \in [0, 1] \Leftrightarrow \prod_{k=t}^u f_{k-1} \geq 1$$

En termes plus explicites, si l'on escompte des recours ou un dégonflement de la charge, cette baisse de la charge ne peut s'assimiler à une pondération entre deux estimations de la charge.

Dans le cadre de la méthode de Bornhuetter et Ferguson, le calcul de la charge a priori, ou technique a été légèrement enrichi par rapport à l'existant.

On rappellera que dans le sous-chapitre III-B-4-c) *Une approche très technique pour la RC corporelle grave en auto*, le sinistre maximum évalué par simulation était capé à 50 M€. Cela correspond à un montant du scénario le plus grave imaginable au sens économique pour Covéa. Même si cette restriction à un montant de 50 M€ lors de la simulation limite la charge totale estimée, l'évènement reste probable.

Une évolution proposée lors des recherches pour ce mémoire consiste à tronquer, au sens statistique du terme, la loi de modélisation pour faire en sorte qu'elle s'éteigne à 50 M€ et que ce montant soit donc inatteignable.

D'un point de vue théorique, cela consiste à tronquer les deux lois log-normale du modèle. On illustre ci-dessous la matérialisation d'une troncature d'une loi Normale :

$$P(X \leq x \mid a \leq X \leq b) = \frac{F(x) - F(a)}{F(b) - F(a)}$$

Avec X la variable aléatoire modélisée – ici le coût unitaire d'un sinistre corporel – et F la fonction de répartition de la loi qui modélise X, soit dans le cas des modèles corporels la loi mélange de type log-log.

Dans la pratique, la troncature est appliquée lors de la simulation, en 3 étapes :

- Estimation des paramètres de la loi
- Estimation de p_{\min} et p_{\max} :
 - $p_{\min} = 1 - P(X \leq \text{plancher} \mid \text{loi } X)$
 - $p_{\max} = P(X \leq \text{plafond} \mid \text{loi } X)$
- Simulation à partir de la loi inverse et d'un tirage aléatoire qui sera compris entre p_{\min} et p_{\max}

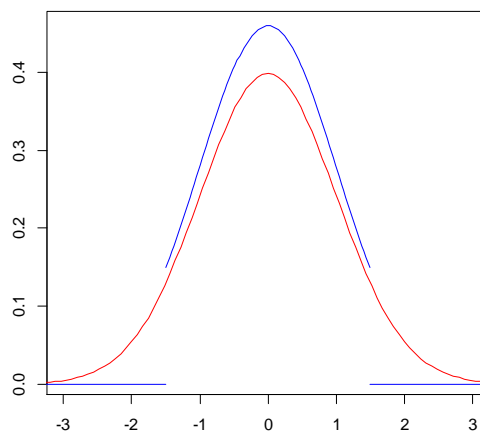


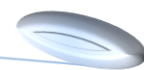
FIGURE 10 - EXEMPLE DE TRONCATURE SUR UNE LOI NORMALE ENTRE -1,5 ET +1,5

L'aire sous la courbe est identique, ici la troncature est volontairement très forte, pour illustrer le phénomène. L'application retenue se limite à décaler tout au plus 0,02% de la densité alors que l'exemple en tronque 13,4%.

3.1.3 Mack démontre qu'il est possible d'estimer une volatilité

Les méthodes de Thomas Mack⁸⁻⁹, Munich Re, consistent à repartir d'un Chain-Ladder ou d'une approche de Bornhuetter-Ferguson et à en estimer les moments d'ordre 1 et 2, à partir desquels il est possible de calculer un intervalle de confiance, par l'application du théorème de la limite centrale.

Ce sont les premières approches stochastiques des liquidations, pour le chain-ladder, cela date



de 1993. L'application au modèle de Bornhuetter-Ferguson a été plus tardive (2008). Cette méthode a été testée dans le présent document. En effet, les conditions d'application du chain-ladder ont été validées précédemment et donc la généralisation de Mack est possible. On notera que l'intervalle de confiance obtenu est assez grand, ce qui peut être problématique.

3.1.4 Passer d'une liquidation univariée à une approche multivariée

Introduit en 2004 par Braun¹⁰, le modèle original est parfois appelé chain-ladder bi-varié et s'applique sur des portefeuilles avec dépendance. L'idée est de prendre en considération les corrélations existantes entre les deux portefeuilles et leurs liquidations dans la détermination des facteurs de développement et pour le calcul de la volatilité.

Dans le logiciel R, qui est le principal outil utilisé pour ce mémoire, on trouve dans la librairie ChainLadder plusieurs méthodes, dont deux approches multivariées.

La première approche, dit modèle de Chain-Ladder Multivarié (MCL en anglais), consiste à liquider chaque triangle par rapport à son passé, en intégrant des corrélations. Cette méthode permet de récupérer l'erreur quadratique moyenne (selon plusieurs options). C'est un prérequis à l'estimation de la volatilité, car les résidus de la variance peuvent être isolés et projetés. Ce qui a un fort intérêt dans le cadre d'une projection avec intervalle de confiance des résultats. Pour rappel, afin d'induire une volatilité dans le Business Plan, il est important de pouvoir estimer une "erreur" dans les projections. Cette méthode s'appuie sur les travaux de Prohl et Schmidt¹¹ de 2005, eux-mêmes basés sur les travaux de Braun précédemment abordés.

La seconde approche disponible, ou le modèle Chain-Ladder Multivarié Généralisé de Zhang¹², est une extrapolation du modèle précédent. Dans son approche, Zhang intègre des coefficients d'interception afin d'améliorer l'adéquation du modèle à la réalité. Toutefois, comme son prédécesseur, le modèle estime les paramètres en utilisant la méthode SUR, de l'anglais *seemingly unrelated regression*. Démontré par Zellner¹³ en 1962, cette approche est couramment utilisée en économétrie, en finance ou encore en assurance. Cette méthode présente une forte sensibilité aux points extrêmes, on parle en anglais d'*outliers*. Cela nécessite donc de bien nettoyer ses données avant d'appliquer ces outils, ou de mettre en place une pondération adaptée.

3.1.5 Le bootstrap peut aussi servir à induire une dispersion empirique

Le bootstrap est une méthode de simulation qui tire son nom de l'anglais. Plusieurs origines sont couramment citées, notamment celle du baron de Münchhausen. Si l'on revient à la base et donc la définition d'origine, les *bootstraps* sont des anneaux de cuir ou de tissus, cousus sur le rebord des bottes qui permettent de tirer dessus afin d'enfiler les dites bottes. Une expression courante du début du XIX^{ème} siècle serait à l'origine du nom de la méthode, "*to pull oneself over a fence by one's bootstraps*". Cette expression peut se traduire par *se faire soi-même passer au-dessus d'une clôture à l'aide de ses propres chausse-bottes*, ce qui peut signifier se donner à soi-même l'impulsion initiale pour réussir à aller plus loin. Le bootstrapping, toujours de l'anglais, action de faire du bootstrap, s'appuie sur cette idée. A partir des données observées et en procédant des tirages aléatoires avec remise, il devient possible de simuler bien plus de données que l'on en disposait initialement.

Dans la pratique, après avoir estimé un indicateur empirique – comme la moyenne d'une série – l'idée est de construire plusieurs séries en repartant de l'indicateur simulé, puis pour chaque point simulé à lui ajouter une erreur d'approximation aléatoire, basée sur les données connues, par tirage avec remise.

Cette méthode peut s'utiliser pour énormément de choses et notamment pour la définition des

coefficients de développement d'un chain-ladder. Pour l'expliquer simplement, la méthode va consister à générer des pseudo-triangles de liquidation, basés sur les données empiriques. La méthode permet de générer presque infiniment des ultimes basés sur ces données. Partant de ces résultats, il est possible de constituer une distribution empirique de l'ultime, donc un intervalle de confiance.

L'idée sera de partir d'un triangle de données connues, d'estimer chacun des coefficients de développement et de faire la moyenne par exercice de développement. Jusqu'à cette étape, il s'agit d'une approche par chain-ladder classique. On calculera alors, pour chaque exercice de développement, un vecteur d'erreur qui sera l'écart à la moyenne de chaque coefficient annuel. On peut alors générer une multitude de liquidations théoriques, en déformant les coefficients de développement initiaux du chain-ladder. Cette déformation va consister à partir de la valeur moyenne et à lui ajouter une valeur d'erreur issue du vecteur empirique. Pour chaque liquidation théorique et chaque exercice, on effectuera un tirage aléatoire de l'erreur, avec remise.

3.1.6 D'autres approches, dits modèles à variables explicatives, existent

Ces approches de régressions cherchent à modéliser les valeurs annuelles de développement en tout point $X_{i,j}$. Autrement dit, on va s'atteler à estimer, par des méthodes de régression, le développement d'un temps à un autre.

Ces méthodes se basent sur les travaux de Renshaw et Verrall¹⁴ et se regroupent en deux familles. Le modèle Log-Normal et le Modèle Linéaire Généralisé GLM, en anglais *Generalized Linear Models* soit GLM. Cette seconde famille utilise plusieurs fonctions dans son approche de modélisation, ainsi que des modèles additifs ou multiplicatifs.

3.1.6.1 Le modèle log-normal

Le principe de ce modèle a été évoqué en préambule, il s'agit de rechercher les paramètres qui s'ajustent le mieux aux données. Ici les données à expliquer sont les incréments annuels selon l'exercice de survenance. Pour rappel, ces montant incrémentationaux sont notés $X_{i,j}$ tandis que les montant cumulés sont notés $C_{i,j}$.

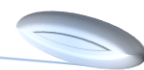
Etant donné les écarts, les données sont transformées au préalable via une fonction log. On notera que les données disponibles, de par leur nature, sont parfois à évolution négative entre 2 temps d'observation. Sans nettoyer ces données, la méthode est inapplicable.

3.1.6.2 Le modèle linéaire généralisé

Cette famille de modèles s'appuie sur une fonction de lien, au regard des méthodes testées pour calculer un développement, la fonction est logarithmique, ce qui impose des restrictions. La plus forte, est la positivité des développements annuels. En effet, le logarithme est défini sur l'intervalle $]0, +\infty[$, soit des valeurs strictement positives. On notera que cette condition peut se réduire à la positivité du total de l'expérience observée, soit la somme des développements unitaires entre deux périodes qui doit être positive ou nulle. Toutefois, les données peuvent parfois ne pas s'y conformer, ce qui est problématique dans le cadre de ce mémoire. On rappelle que l'objectif principal est de pouvoir estimer la volatilité au quotidien, pas uniquement dans des cas restreints.

Si l'on s'arrête sur la méthode de Poisson sur-dispersée, méthode utilisant une fonction de lien logarithmique (i.e. non applicable à des données pouvant avoir un développement négatif entre deux temps t), il est bon de retenir ses avantages et limites.

Dans ces approches, on va chercher à modéliser, par un modèle linéaire multivarié, le point



suivant du développement en fonction des points déjà connus. Dans le cas d'une loi de Poisson surdispersée, cela implique qu'il existe un paramètre Φ tel que si l'on divise la variable à expliquer par ce paramètre, il existe un paramètre λ' qui suit une loi de Poisson.

On peut en déduire simplement des moyennes et variances pour les calculs. Cette méthode a été particulièrement travaillée car elle permet d'afficher une charge à l'ultime identique à celle issue d'un chain-ladder, et de lui apporter un intervalle de confiance. Enfin, cette approche présente une précision plus fine que l'approche de Mack sur les garanties dommages, en RC le gain est faible, voire inexistant.

La principale contrainte reste son application, impossible sur les données et dans les pratiques actuelles du service, du fait de la fonction de lien.

Dans les GLM, il existe d'autres fonctions de lien, comme la fonction identité, qui va permettre de tester une approche gaussienne. Toutefois, la connaissance a priori des données laisse à penser que ces résultats ne seront pas exploitables. Les quelques tests effectués sont non concluants.

3.2 Etude des résultats obtenus sur les méthodes de liquidation

Avant de s'attarder sur les résultats, se pose la question de ce que l'on affiche. Historiquement et à des fins tarifaires, les projections à l'ultime sont agrégés par segment, ici automobile et habitation. Toutefois, la somme est faite par hypothèse d'indépendance des triangles. Pour les espérances, cela est logique, toutefois, pour les variances cela est plus problématique. En effet, on démontre aisément (cf. annexe E) que les covariances entre les populations sommées s'intègrent à la définition de la variance globale.

Les premiers résultats seront sur la base d'une hypothèse d'indépendance, puis les méthodes décrites précédemment, qui permettent d'intégrer cette contrainte, seront présentées, bien que la liquidation multiple pose plusieurs questions. Elles seront débattues dans ce mémoire, à des fins d'applications concrètes.

3.2.1 Des résultats sous hypothèse d'indépendance qui font référence

Les premiers résultats sur lesquels l'on va s'attarder sont une comparaison entre la méthode de Mack et celle du Bootstrap du chain-ladder. Comme expliqué ci-après, la comparaison se fait sur la somme de plusieurs triangles, post liquidation et par segment.

Le premier constat sera de dire que les espérances de charge ultimes moyennes diffèrent d'une méthode à l'autre. Toutefois, ce constat s'atténue avec le recul, en effet, avant 2014, les écarts sont minimes.

Le second constat sera sur l'intervalle de confiance lui-même. Il est graphiquement 3 fois plus faible par bootstrapping que par application des méthodes de Mack, sur les exercices récents. D'un point de vue opérationnel, cela fait nettement pencher la balance pour la méthode de bootstrap.

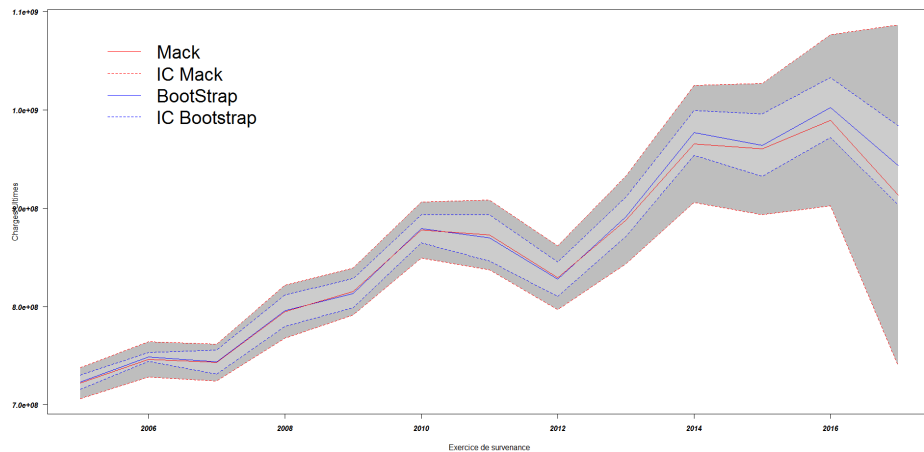


FIGURE 11 - INTERVALLES DE CONFIANCE A 95% SUR LES METHODES DE MACK ET DE BOOTSTRAP – SOCIETE A - TOTAL

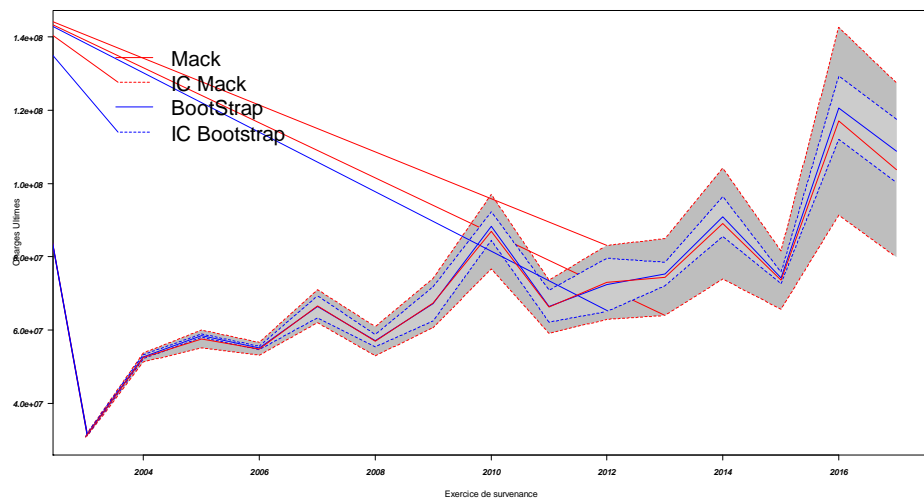


FIGURE 12 - INTERVALLES DE CONFIANCE A 95% SUR LES METHODES DE MACK ET DE BOOTSTRAP – SOCIETE B - TOTAL

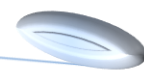
L'ensemble des graphiques par segment et marque est disponible en annexe F, les constats sont les mêmes.

Après étude et quelle que soit la méthode, les intervalles se réduisent avec le recul, ce qui est à la fois logique et rassurant. En effet, plus l'on a de recul sur les sinistres d'un exercice et plus il est aisé d'estimer la charge à la fin des temps.

Avant de passer aux approches avec intégration de variance, il faut noter que les résultats des méthodes à variable explicatives, même sur un jeu de données différents (triangle de paiement) sont inexploitable. Outre les problématiques liées à la définition d'un facteur de queue pour les garanties à liquidation très longues, les données sont sujettes à des évolutions négatives, suite à une évolution juridique, la fin d'un provisionnement, la démonstration d'une fraude... C'est pour ces raisons qu'ils sont absents de cette partie.

Enfin, les méthodes de référence, notamment le Chain-Ladder, qui sont sans estimation d'une variance, donc sans intervalles de confiance, sont également omises sciemment. Toutefois, l'on pourra rappeler que les résultats issus de chain-ladder et d'approche de Bornhuetter-Ferguson sont identiques à ceux des méthodes de Mack. Plus précisément, on devrait dire que les méthodes de Mack ont, par construction, une espérance identique aux méthodes de référence.

3.2.2 Des résultats où l'on intègre les dépendances



Dans cette partie, les résultats ont donc une valeur plus proche de la réalité, cependant plusieurs questions se posent. Dans ce mémoire, il a été volontairement étudié 3 garanties par segment (sinistres graves, climatiques et de masse), ainsi que 2 segments (automobile et MRH) par marque (société A et B) du groupe Covéa.

On rappellera que l'un des objectifs de ce business plan est de définir le tarif. Actuellement, le tarif est réalisé par segment, voire par garantie au sein de ce segment et par marque. La première envie est donc de reconstituer, comme dans la partie précédente, les charges à l'ultime telles qu'elles sont exploitées. Il faudra s'en abstenir, comme nous tâcherons de le montrer par la suite.

Avant de s'attarder sur les résultats, seront discutés les regroupements faisables et logiques. Puis, les résultats obtenus et dans un dernier temps, des approches pour reconstituer les agrégats aux formats exploitables pour les besoins du groupe, tout en conservant au mieux les éléments mis en évidence lors de cette étude.

3.2.2.1 Quels regroupements appliquer à nos triangles ?

Deux approches ont été testées, une étude de classification et une approche par corrélation. Ces tests ont été appliqués sur les triangles de données dans un premier temps et sur les triangles de développement, on parlera aussi de *d-triangle*, dans un second temps.

3.2.2.1.1 Classifier les triangles

Les méthodes de classification sont nombreuses et l'on retrouve le terme pour la première fois dans les travaux de Tryon¹⁵ en 1939. La classification ascendante hiérarchique, ou CAH, est une méthode qui consiste à calculer la distance entre deux points ou nuages de points, puis à regrouper les points les plus proches. Cela met donc en œuvre deux notions, celle de la distance et celle d'estimation de nuages de points les plus proches.

Concernant la distance, diverses méthodes de calcul de la distance existent. La plus connue devant être la distance euclidienne qui fonctionne très bien dans des espaces orthonormés. Une autre distance intéressante est celle la distance de Manhattan, qui consiste à diviser une zone en bloc, comme la ville éponyme, et à définir le chemin le plus court entre les deux points selon cette grille.

Pour cette étude, la distance euclidienne a été préférée. Pour pallier à la problématique d'orthogonalité entre les variables, toutes numériques, un travail de réduction de la variance va être effectué. La méthode appliquée est une analyse en composantes principales, ou ACP, qui permet de projeter les variables, comme les points, sur des axes. Chaque axe résumant une partie, plus ou moins conséquente, de la variance des données étudiées. Plusieurs documents traitent du sujet, notamment les travaux de Jolliffe¹⁶. On pourra retenir que la méthode est ancienne, elle a été introduite par Karl Pearson¹⁷ en 1901. Il faut retenir que l'approche est à la fois géométrique et statistique, qu'elle permet de résumer sur moins de dimensions une part forte de la variance contenue dans les variables étudiées par projection sur des axes, axes qui sont orthogonaux entre eux. Par conséquent, si l'on projette des données ayant une unité commune, les axes deviennent orthonormés, ce qui rend la classification, par distance euclidienne efficiente.

Une fois les données centrées, réduites, résumées par une ACP, il est possible de commencer la classification. La dernière étape consiste à définir la méthode de regroupement. Il en existe plusieurs, les citer et faire référence à leurs auteurs pourrait être long, aussi seule la méthode retenue sera présentée ici. Il s'agit de la méthode de Ward¹⁸. Le critère de Ward s'appuie sur l'inertie et le rapport entre l'inertie interclasse et l'inertie totale. A noter que l'inertie totale est la somme des inerties interclasse et intra-classe. Le rapport entre inertie interclasse et inertie totale est un estimateur de qualité des regroupements. Quand ce rapport tend vers 1, alors les points regroupés sont très proches car la quasi-totalité de l'inertie du système est contenue entre les classe et non à l'intérieur. La méthode de Ward va consister à agréger les classes a et b qui minimisent la diminution de l'inertie interclasse, donc à conserver la qualité des regroupements.

$$Inertie(a) + Inertie(b) = Inertie(a \cup b) - \frac{m_a \cdot m_b}{m_a + m_b} \cdot d^2(a, b)$$

Avec m le centre de gravité des classes a et b et d la distance entre ces classes. L'objectif de la méthode consiste à minimiser la partie en bleu à droite de l'équation.

Cette méthode a été appliquée aux triangles, après centrage et réduction des données, par triangle. Elle a également été appliquée sur les d-triangles sans centrer ni réduire, les valeurs étant toute sur une même norme. On obtient alors les résultats présentés à la page suivante.

Dans cette étude, il faut considérer les triangles comme les points et les dates d'observation et exercice de survenance comme les variables, étant donné que l'on observe jusqu'à 18 années, le modèle se compose de plus de 150 variables.

L'ACP permet de réduire à 7 le nombre de variables qui représentent environ 98,5% de la volatilité présente dans les données.

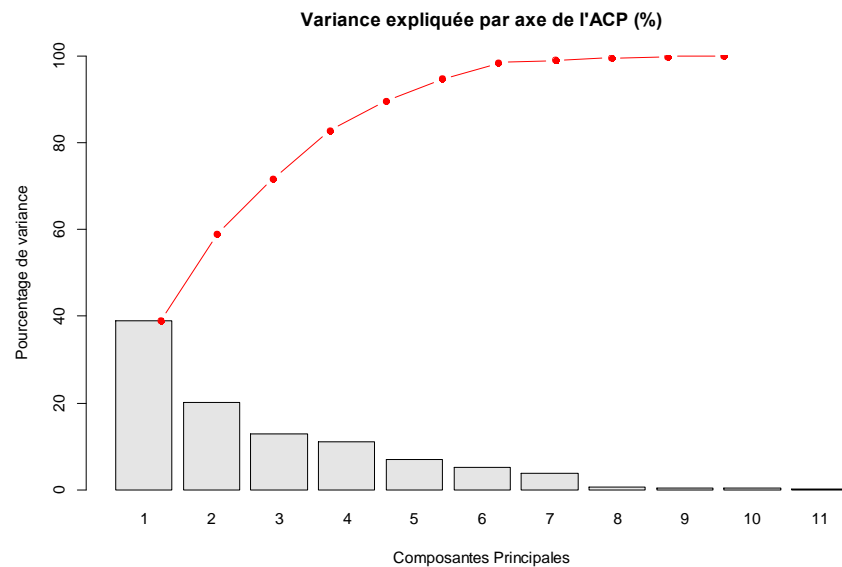


FIGURE 13 - RESULTATS DE L'ACP : CONSTRUCTION DES AXES RESUMANT LA VARIANCE

En appliquant la CAH sur ces 7 premiers axes, le résultat obtenu est assez flagrant. Très rapidement les garanties Climatiques et de Masse, par segment Automobile et Habitation se regroupent pour former 4 classes indépendantes. Les triangles de sinistres graves sont eux plus atypiques.

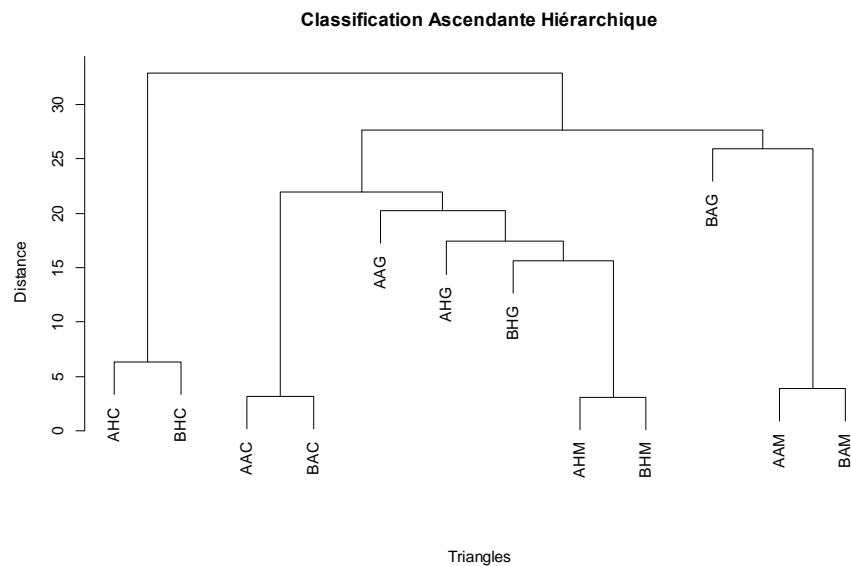
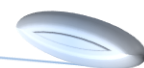


FIGURE 14 - RESULTAT DE LA CAH SUR LES TRIANGLES DE CHARGE

Aide à la lecture :

- la première lettre du trigramme est celle de la société A ou B,
- la seconde fait référence au segment, Automobile ou Habitation,
- enfin la dernière fait référence à la garantie de Masse, Climatique ou Grave.

Sur le second cas testé, soit l'étude des coefficients de développement, l'ACP permet de réduire les dimensions jusqu'à 6 ou 8 axes, soit respectivement 96% ou 99% de la variance expliquée. La CAH appliquée aux 8 axes permet d'obtenir des blocs différents et pourtant proches de ceux obtenus sur les triangles de charge.

On distingue des blocs, avec les garanties de Masse en Habitation qui sont très vite regroupées pour les 2 sociétés, tout comme les garanties de Masse en Auto. Puis les garanties Climatique en Auto qui se greffent à ce bloc de garanties de masse. Le tout formant très rapidement 1 seul et même groupe. A part, un second groupe se forme, composé des garanties Climatiques en Habitation des 2 marques. Enfin, les garanties Graves sont très éloignées entre elles et des blocs précédemment formés.

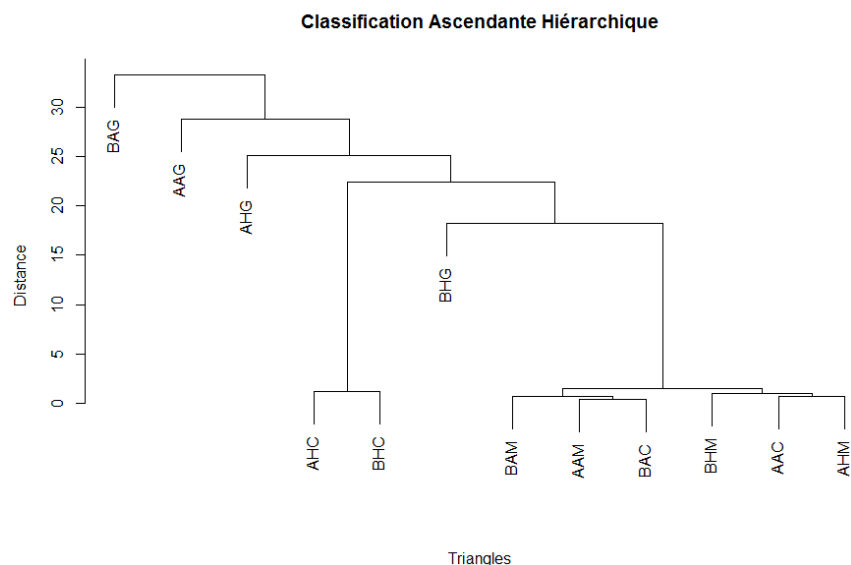


FIGURE 15 - RESULTAT DE LA CAH SUR LES TRIANGLES DE DEVELOPPEMENT

La première conclusion que l'on tirera de ces tests est que les regroupements entre les marques sont à faire avant les regroupements au sein d'un segment Auto ou Habitation. Ce qui est clairement différent des pratiques actuelles.

Avant d'aller plus en avant dans les conclusions, les corrélations entre les triangles doivent être étudiées.

3.2.2.1.2 Quelles corrélations retenir pour les triangles

La théorie de la corrélation ayant déjà été développée précédemment (*IV A 1 - La méthode de référence est le chain-ladder*), seul les résultats seront présentés.

En annexe G les représentations des triangles les uns en fonction des autres est disponible. On pourra y voir que certains nuages de points ont des formes remarquables qui dénotent de la présence de corrélations fortes.

Encore une fois, deux types de triangles sont étudiés, les corrélations entre les charges, centrées et réduites, et les triangles de coefficients de développement. A noter que le second test présente un intérêt particulier car il met en avant des notions proches de celles utilisées dans le modèle de Braun et ses dérivés, présentés précédemment.

Enfin, sur les schémas de résultats, des classifications ont été ajoutées, elles utilisent le même principe que celui expliqué au chapitre précédent, à savoir des CAH, sur les corrélations.

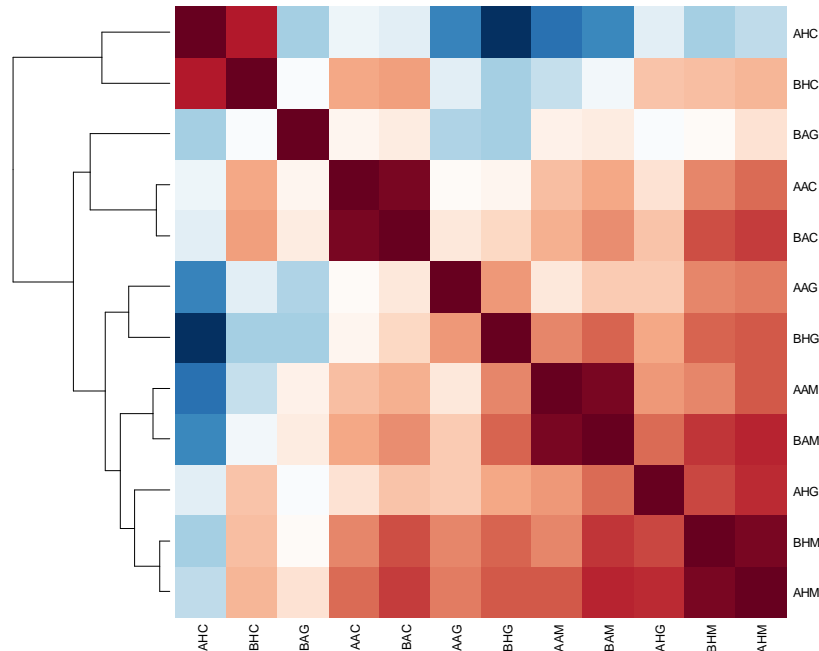


FIGURE 16 - CORRELATION ET CLASSIFICATION DES TRIANGLES DE CHARGE

Les résultats sont proches de ceux obtenus lors de la première approche, avec rapidement des classes par garanties de sinistres, Masse et Climatiques des marques, et avec la séparation sur les segments Auto et Habitation. On peut également voir que des corrélations existent selon que la couleur tende vers le rouge ou le bleu, respectivement ces corrélations sont positives ou négatives.

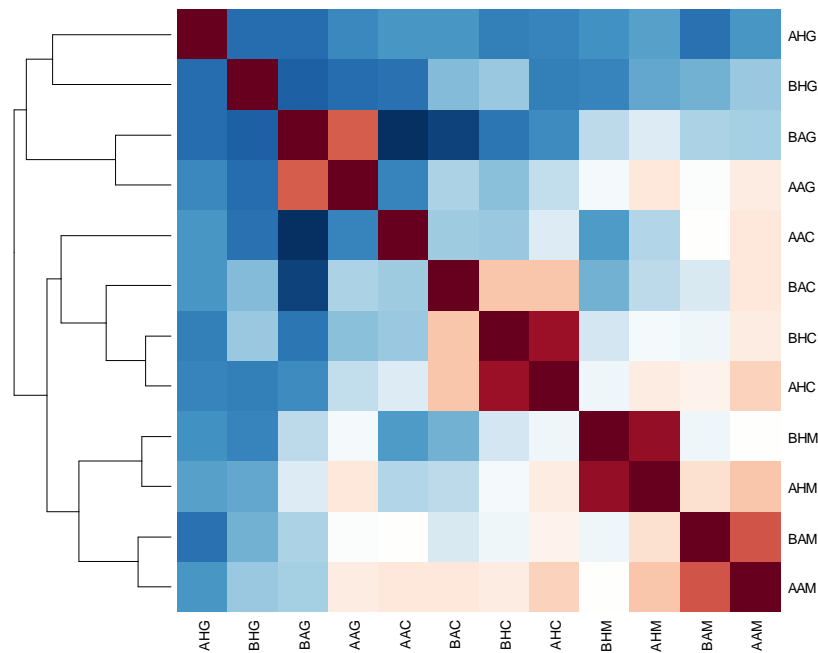
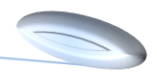


FIGURE 17 - CORRELATION ET CLASSIFICATION DES D-TRIANGLES

L'interprétation est la même sur les triangles de coefficients de développement, toutefois les corrélations sont majoritairement négatives, bien que positives et voisine de 1 sur les premières classes constituées. On retrouve donc les classes de climatiques en habitation, ainsi qu'en auto, de masse en Auto et en Habitation et enfin les graves, d'abord en auto, puis en habitation...

On retiendra, là encore, que le type de garanties prédomine sur la société qui porte le risque, cela est donc une forte préconisation pour regrouper les typologies de risque par garanties pour la liquidation multivariée, alors que les regroupements se faisaient jusqu'alors par société.

A noter que si l'on souhaite couper à 3 classes, alors cette dernière approche préconise de regrouper les segments avant les garanties. Toutefois réduire le nombre de triangles à trois, comme pourrait l'indiquer les différentes classifications effectuées, semble trop réducteur. En effet la perte de significativité serait trop importante.

$$\text{Taux d'information} = \frac{\text{inertie intergroupe}}{\text{inertie totale}} = \frac{\text{Inertie}(gr_1 \cup gr_2)}{\sum \text{Inertie}(gr_i \cup gr_j)}$$

3.2.2.2 Les résultats intéressants des approches multivariées

Avant de s'attarder sur les résultats des différentes méthodes multivariées, il semble important de répreciser le périmètre. En effet, le chapitre précédent laisse sous-entendre que des regroupements de triangles seraient nécessaires avant liquidation. Après clarification du périmètre étudié, les résultats obtenus seront détaillés.

3.2.2.2.1 Choix du périmètre

On restera sur le périmètre historique, à savoir une vision par marque et segment de marché. En effet, rechercher des solutions opérationnelles suite aux éléments précédemment énoncés, notamment la prédominance de la liquidation à la garantie, prendrait plus de temps. A noter qu'initialement ces éléments étaient acquis et le mémoire ne se voulait pas de les remettre en

cause.

A noter que 4 approches par chain-ladder multivariés vont être testées, soit une par marque et segment de marché. Cela reste insuffisant pour tester les interactions externes, comme l'incidence du triangle d'une autre marque ou d'un autre segment de marché sur le résultat. Toutefois, cela répond au besoin premier, à savoir livrer un résultat par marque et segment de marché, pour aider à la prise de décision lors du renouvellement des tarifs.

Initialement des approches multimarque et mono-segment, ainsi que mono-marque et multi-segment devaient être testées. Au regard des résultats de l'étude des corrélations, ces approches ont été remises à plus tard car les éléments mis en avant soulèvent de nouvelles questions.

3.2.2.2 Résultats

Parmi les méthodes énumérées précédemment dans les approches multivariées, on distinguera les familles MCL, ou *Multivariate Chain-Ladder*, du GMCL *General Multivariate Chain-Ladder*. Pour ces deux familles, seront testées des approches avec et hors l'ajout de coefficient d'interception, qui permettent de corriger les trajectoires.

Par ailleurs, sur la famille MCL, une approche par régression indépendante des coefficients de liquidation des triangles sera testée, à titre comparatif avec la méthode de Mack. Cette méthode intègre uniquement les corrélations inter-triangles lors du regroupement final, contrairement à l'approche de Mack vue précédemment.

Sur la figure ci-dessous, on notera que les deux approches ont une espérance identique tandis que leurs intervalles de confiance diffèrent, du fait des corrélations entre les triangles. De manière assez logique, un sinistre grave qui sortirait de la garantie grave, viendrait enrichir la charge de la garantie de masse, aussi la volatilité ne se somme pas et les corrélations entraînent une réduction de l'intervalle de confiance.

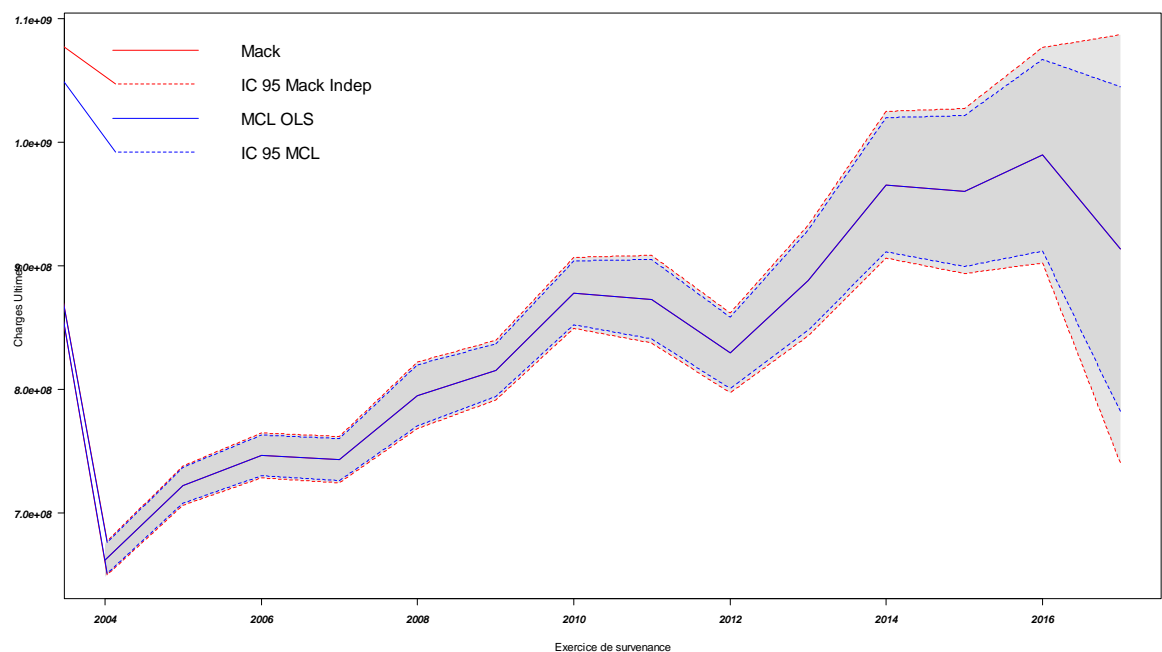


FIGURE 18 - COMPARAISON INDEPENDANCE TOTALE ULTIMES MACK ET MULTI-CHAIN-LADDER – SOCIETE A - TOTAL

Si l'on s'attarde sur les autres méthodes testées, notamment la GMCL (cf. figure page suivante), en les triant par taille d'intervalle de confiance, soit en donnant de l'importance à la précision, il s'avère que les espérances sont très volatiles.

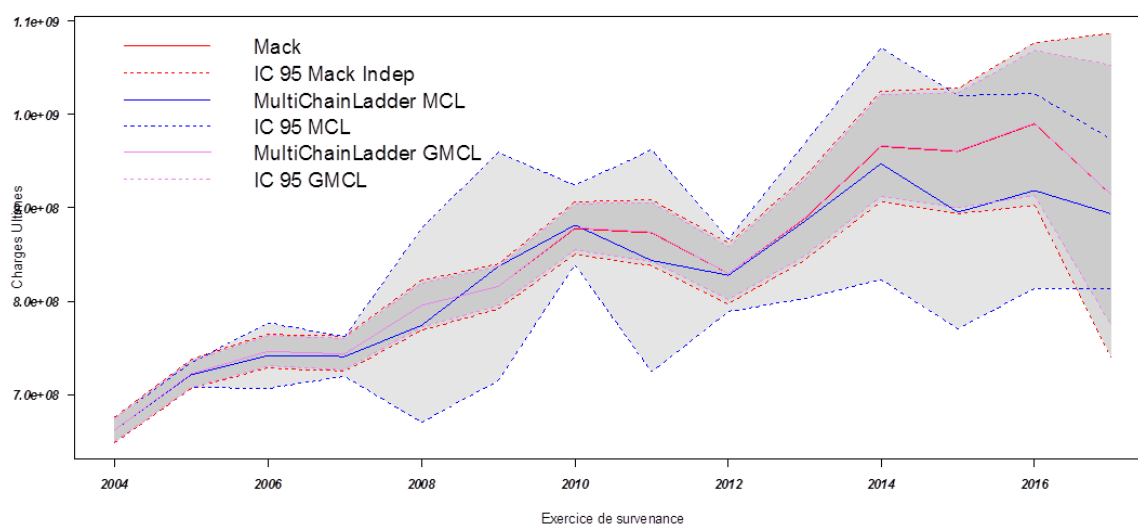


FIGURE 19 - COMPARAISON INDEPENDANCE TOTALE ULTIMES MACK, MCL ET GMCL – SOCIETE A - TOTAL

Le tableau ci-dessous résume les résultats, sur la dernière année observée.

Zoom 2017	Mack	MCL - SUR	MCL - OLS	GMCL-int	GMCL	MCL+int
Ultime	913,3	913,5	913,3	951,8	893,0	882,2
IC-	739,9	774,1	781,9	828,0	812,9	807,5
IC+	1 086,8	1 053,0	1 044,8	1 075,7	973,0	956,9
Taille IC	346,8	279,0	262,9	247,6	160,2	149,3

On pourra noter que le modèle MCL avec la méthode SUR (*seemingly unrelated regression*) est proche de la méthode indépendante que ce soit en termes d'espérance ou d'IC.

On notera que les méthodes avec intégration des coefficient d'interception, soit l'application de droite non absolue pour estimer les coefficients de liquidation entre deux temps t et $t+1$, ont des intervalles plus resserrés et des espérances plus faibles. On parle ici des méthodes GMCL, qui par définition intègre les *intercepts* et le MCL avec *intercepts* (MCL+int).

Les fortes disparités observées entre ces modèles sur un seul exercice sont déroutantes, on voit clairement ici que des points atypiques peuvent perturber la modélisation. De plus, la notion de modèles multivariés implique que ces points atypiques peuvent être jugés, de manière univariée, comme adaptés, seulement la combinaison multidimensionnelle de ces points – i.e. leur valeur sur les différents triangles – peuvent être eux atypiques.

On notera toutefois que l'intervalle est toujours plus petit dans les modèles multivariés que dans l'approche de Mack sous hypothèse d'indépendance totale. Cela laisse donc à penser que les corrélations sont négatives et tendent à réduire la volatilité du système.

3.3 Un regard vers les nouvelles approches du marché

Depuis quelques années, les recherches sur les cadences de développement sont approfondies, pour des besoins réglementaires et aussi pour des besoins de précision.

3.3.1 Modèle de Braun et ses dérivées

Une thèse belge, *A Robust General Multivariate Chain Ladder Method* – publiée fin juillet 2018, par Kris Peremans, Stefan Van Aelst et Tim Verdonck – s'appuie sur les travaux de Zhang et les enrichis d'un estimateur, le *MM-estimator*, permettant de diminuer la sensibilité de modèles face aux points extrêmes ou aberrants.

Par ailleurs, Marc Linde et Lukas J. Hahn travaillent actuellement sur des approches combinatoires dérivées du modèle de Braun, en intégrant plusieurs portefeuilles, en combinant des chain-ladder et des modèles additifs et enfin en réalisant des simulations par bootstrap de ces méthodes. Des présentations sont disponibles, en anglais sur internet, ainsi que de la littérature plus technique¹⁹.

Les résultats de ces récents modèles semblent intéressants et ils répondent aux exigences opérationnelles de ce mémoire. Toutefois, ces éléments sont très récents et lors de la phase initiale de définition de l'état de l'art, en 2017, ces documents n'existaient pas encore. Ils ont été lus, voire rédigés pour certains, lors de la phase rédaction du mémoire, à savoir fin 2018 et début 2019, en cherchant à approfondir le sujet. L'écriture et l'adaptation des codes en R pour tester cette approche a donc été mise en attente et pourra faire l'objet de recherches complémentaires lors de prochains travaux.

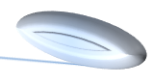
3.3.2 Approches de liquidation au sinistre

S'appuyant sur des méthodes de statistiques modernes et utilisant toute la puissance des ordinateurs actuels, des méthodes appelées génériquement *data-science* vont être utilisées pour estimer une liquidation au sinistre, par mimétisme de l'expérience passée. Parmi ces méthodes, se retrouvent des approches par modèle linéaire généralisé, avec un facteur d'enchaînement. On retrouvera également des approches de type forêt aléatoire, parfois modifiées ou altérées. Enfin les réseaux de neurones devraient également être de bons prédicteurs des situations observées.

Sur le marché, lors de la journée IARD de l'Institut des Actuaire, en mars 2018, les travaux du mémoire d'actuariat de Damien FABRE RUDELL²⁰ sur le sujet ont été présentés. Les premières conclusions sont mitigées, car les résultats sont souvent aussi précis que le chain-ladder. Cependant ce sont encore les débuts des usages.

A ce jour, ces approches sont en phase de laboratoire au sein de Covéa, via son partenariat européen Eurapco. A noter que Benoit LEBRUN a étudié, dans le cadre de sa formation et de son mémoire de Data-Science, l'application de ces approches aux données Covéa. Il est encore tôt pour parler de conclusion, toutefois les premiers retours semblent montrer que ces nouvelles méthodes seraient complémentaires des méthodes traditionnelles.

On peut également les voir comme des passerelles avec d'autres entités du groupe, comme de l'aide à la détermination, semi-automatique, des montants forfaitaires d'ouverture.



3.4 Ce qu'il faut retenir de ces tests

Si l'on revient sur les résultats des différentes méthodes présentées, il semble important de dissocier deux phases, tout d'abord le choix opérationnel d'une méthode et enfin la question des corrélations et des approches priorisant la garantie à la marque.

3.4.1 Le choix d'une méthode se base sur le besoin initial

Différentes méthodes ont été tantôt testées, tantôt exclues, parfois approfondies et optimisées. Ces méthodes présentent des avantages et des inconvénients qui sont des aides pour définir la ou les méthodes à retenir dans une optique de déploiement au quotidien. Le tableau suivant récapitule les avantages et inconvénients des méthodes testées, informations obtenues dans la littérature ou par expérience.

Méthodes		Avantages	Inconvénients
Approche déterministe	Chain-Ladder	Méthode de référence Simple à mettre en œuvre	Sensible aux valeurs extrêmes Hypothèses d'application difficiles à respecter
	Bornhuetter-Ferguson	Méthode courante pour certains risques Mise en œuvre rapide et simple	S'appuie sur une valeur à priori - disponible chez Covéa Des limites d'application
Approche par modélisation	Mack	Résultat identique au modèle d'origine (CL ou BF) Permet d'avoir un encadrement du résultat	Idem Chain-Ladder ou Bornhuetter-Ferguson Intervalle de confiance large
		La valeur moyenne est identique au Chain-Ladder	Sensible aux valeurs extrêmes
	Braun*	Modèle avec volatilité et capable de gérer plusieurs triangles et leurs corrélations	Hypothèses d'application difficiles à vérifier Uniquement pour le Chain-Ladder
		Sous condition (modèle de poisson surdispersé) le résultat est identique au Chain-Ladder	Des restrictions d'utilisation très fortes
GLM	Intervalle de confiance plus fin que Mack sur certaines garanties	Rien dans la littérature ouvre des possibilités d'étude sur des multiples triangles, sauf peut être les copules (très complexes)	
Approche par simulation	Bootstrap	Un intervalle de confiance plus précis que Mack Une mise en place simple (sous R) Une méthode applicable sur tout type de données	Des résultats différents des méthodes déterministes Des temps de traitement relativement long N'intègre pas les corrélations entre triangles

FIGURE 20 - TABLEAU RECAPITULATIF DES METHODES ETUDIEES

A partir de ces éléments, il apparaît plus intéressant d'utiliser les approches de Chain-Ladder "bootstrapée", toutefois on notera que selon la précision souhaitée, les temps de calcul peuvent être plus longs que sur les méthodes alternatives. Ainsi 10 000 triangles générés par bootstrap, pour obtenir une répartition empirique, vont prendre un temps machine de quelques minutes. Par contre, passer à 1 000 000 de cas, va prendre près de deux heures par triangle. Cependant, la précision obtenue est plus fine. Sur la société A, en Auto, l'intervalle de confiance se resserre de 12 M€, soit 1,3% de moins, tandis que le scénario central augmente de 260 k€, soit 0,03%. Ces tests de sensibilité ont été réalisés uniquement sur 1 société et 1 segment de marché, avant d'arrêter une décision il faudrait généraliser l'étude.

Dans la pratique les garanties à traiter sont bien plus nombreuses et cette solution, qui semble plus intéressante devrait être écartée, car trop coûteuse en temps. Une alternative serait de travailler sur un serveur dédié et de réécrire les scripts pour permettre d'exploiter l'intégralité des cœurs de calcul de la machine.

Un autre inconvénient majeur de cette méthode est qu'elle ne permet pas de prendre en considération les corrélations entre les triangles.

Au regard des tests réalisés sur les approches de Chain-Ladder multivariées, la conclusion est de dire que la variance totale est plus faible que la somme des variances propre à chaque triangle (application de la méthode de Mack pour ce calcul).

Etant donné que l'intervalle de confiance obtenu par bootstrap est plus fin que celui issu de l'approche multivariée, on est en droit de penser que l'approche par bootstrap est un majorant

de la réalité, ce serait d'ailleurs le majorant le plus faible.

Partant de ce principe, il semble que les approches par Chain-Ladder bootstrapée satisfont les restrictions établies initialement. Elles permettent également d'obtenir des résultats rapidement ou, avec du développement et des calculs informatiques, plus lentement et plus précisément.

Cette approche est donc à prioriser, ainsi que ses extensions, comme un BF bootstrapé présenté dans les travaux de A. Le Tesson, A. Lenain, S. Samba et J. Ung²¹.

3.4.2 De l'intérêt de se réorganiser pour traiter les résultats par garantie

Concernant les études de corrélation et la prédominance des garanties dans le regroupement, cela pose bien des questions, notamment méthodologique. Si demain les charges sinistres à l'ultime devaient être calculée par garantie au global Covéa, se poserait alors la question de la descente vers les marques, pour la tarification et pour le pilotage de la rentabilité.

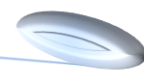
Les conclusions précédentes ont été faites sur la base d'un regroupement traditionnel des garanties par marque et segment de marché. En effet, il faut avant tout changement confirmer le constat en réalisant l'étude sur l'ensemble des marques du groupe.

Ensuite, si les conclusions demeurent les mêmes, alors il faudra définir une méthode pour passer d'une définition stochastique des résultats par garanties à un regroupement par marque et segment de marché.

Concernant le passage de l'indicateur moyen d'un ultime global à un ultime plus traditionnel, il y a plusieurs approches plus ou moins simple. La première, qui est certainement la plus simple, consiste à mutualiser entre les marques la charge latente. Dit autrement, il suffit d'appliquer le développement obtenu sur la charge globale aux charges initiales détaillées par segment. Cela a l'avantage d'être simple et rapide, avec l'inconvénient d'un manque de précision, surtout si certaines marques ou segments de marché ont des développements plus atypiques. Autre point à surveiller, le poids des marques, qui peut changer dans le temps et donc induire une latence différente.

Une alternative pourrait être de réaliser une approche crédibilisée du phénomène, en intégrant le portefeuille comme pondération. Bien que la méthode de crédibilité²² soit plus utilisée pour définir le prochain point, en prenant en considération l'indépendance d'un segment par rapport à la globalité de la population, elle permet d'obtenir des indicateurs permettant de pondérer une méthode indépendante et traditionnelle à l'approche nouvelle proposée dans ce mémoire, à savoir une vision globale par garantie.

Cela laisse encore des questions sur la volatilité de ces résultats, car s'il est possible de présenter des méthodes pour la moyenne, la question de la variance ou de calculs type Monte-Carlo reste en suspens sur cette partie.



4 Quelle part d'aléa dans l'atteinte des objectifs commerciaux

Cette partie cherche à intégrer au modèle déterministe de la production une volatilité. Initialement prévue en trois phases – la première basée sur l'expérience du passé et la seconde sur l'impact d'indicateurs exogènes et enfin une dernière partie visant à réunir les deux premières – il a été décidé de se concentrer uniquement sur la première phase, la seconde sera présentée pour une ouverture et des recherches ultérieures.

4.1 Comment rendre le modèle actuel stochastique ?

Il existe énormément de possibilités pour rendre le modèle à la fois stochastique ou plus interdépendant. En effet, les hypothèses sont aujourd'hui déterministes, il est donc possible de les étudier pour extrapoler des jeux stochastiques.

Une première approche envisagée serait de savoir si les hypothèses se réalisent, plus ou moins justement en fonction du recul, i.e. est-ce qu'à 1 an, 2 ou 5 ans, elles ont une probabilité plus ou moins importante de se réaliser.

Partant de ce principe, avec des études de changement d'état, introduisant des processus de Markov²³, il serait possible de rechercher les états possibles en chaque temps t de la projection. Cela pourrait également se rapprocher du modèle de Black-Scholes²⁴, avec notamment une recherche d'une probabilité équivalente et surtout, l'intégration d'un mouvement brownien²⁵.

A noter que l'ensemble des notions citées précédemment a été abordé en cours et étant donné que ces méthodes ont été écartées par manque de temps, elles ne seront pas plus détaillées.

Une autre approche consisterait à rechercher des variables explicatives exogènes au système actuel, que l'on appellera de marché, comme le prix de l'énergie, son évolution, ou encore l'inflation... Outre la complexité dans la récupération de la donnée brute et les questions sur les indicateurs à construire (moyenne annuelle, moyenne des prix les plus hauts/bas...) se pose la question de la projection de ces critères dans le futur. Dans la pratique, projeter des variables endogènes ou exogènes revient au même, soit l'on se base sur des éléments déterministes, comme l'inflation attendue par la Banque de France, soit l'on émet des hypothèses sur ces variables dans le futur.

Par conséquent, la recherche de variables exogènes a été mise de côté, bien que cette solution puisse être une très bonne approche pour générer des jeux d'hypothèses cohérentes d'une marque à l'autre. Le temps imparti pour réaliser ces travaux étant court, la priorité a été donnée à une méthode très opérationnelle.

Enfin, la méthode retenue, et qui sera détaillée par la suite, consiste à partir du modèle existant, déterministe, et à regarder s'il est robuste. Si toutes les variables qui le compose sont influentes et enfin à analyser les résidus de ce modèle retravaillé. En effet, si les résidus sont gaussiens, il devient possible de projeter le modèle dans le temps, tout en lui injectant une incertitude, on parlera de bruit, sur la base d'une loi normale. Cette méthode s'appuie sur des techniques de traitement du signal, l'écart entre le résultat modélisé et le résultat obtenu, les résidus, pouvant s'apparenter à un signal, que l'on réinjectera. A noter qu'en cas de non normalité des résidus, il est possible soit d'identifier la loi sous-jacente, soit d'appliquer à nouveau une méthode de bootstrap.

4.2 Vers un modèle simplifié et plus robuste

Comme expliqué précédemment, cette partie va consister à rechercher un modèle plus robuste et simple que le modèle existant.

Pour la suite, il a été décidé de travailler l'impact de la production sur le portefeuille, en risques-année. Ainsi, les formules actuelles complètes et exactes, vont être revisitées. On commencera par une approche linéaire, en injectant l'ensemble des opérations de production qui constitue le modèle actuel et la variable elle-même au temps t-1 précédent. Les opérations de production sont les affaires nouvelles, les affaires perdues par motif (siège ou client) et par prise d'effet (immédiat ou au terme).

$$Ptf_t = Ptf_{t-1} + \alpha \cdot AN_t + \beta \cdot AP_t^{s^i} + \gamma \cdot AP_t^{s^f} + \delta \cdot AP_t^{c^i} + \varepsilon \cdot AP_t^{c^f} + \phi$$

Avec les paramètres :

- $\alpha, \beta, \gamma, \delta$ et ε des coefficients d'ajustement selon le type d'opération
- AN les affaires nouvelles attendues ou connues pour l'exercice t
- AP les affaires perdues attendues ou connues pour l'exercice t selon que l'origine du motif soit le siège ^s ou le client ^c et que l'effet soit immédiat ⁱ ou fin de contrat ^f
- ϕ l'erreur d'estimation du modèle linéaire

Cette première approche est réalisée sur l'ensemble du périmètre de l'étude (Auto et Habitation des marques A et B), indépendamment pour chaque croisement marque/segment. le premier constat est que l'exercice 2015 est atypique. Cela s'explique par l'évolution juridique, avec la loi dite Hamon²⁶ qui modifie radicalement la proportion entre prise d'effet immédiate et à date anniversaire du contrat.

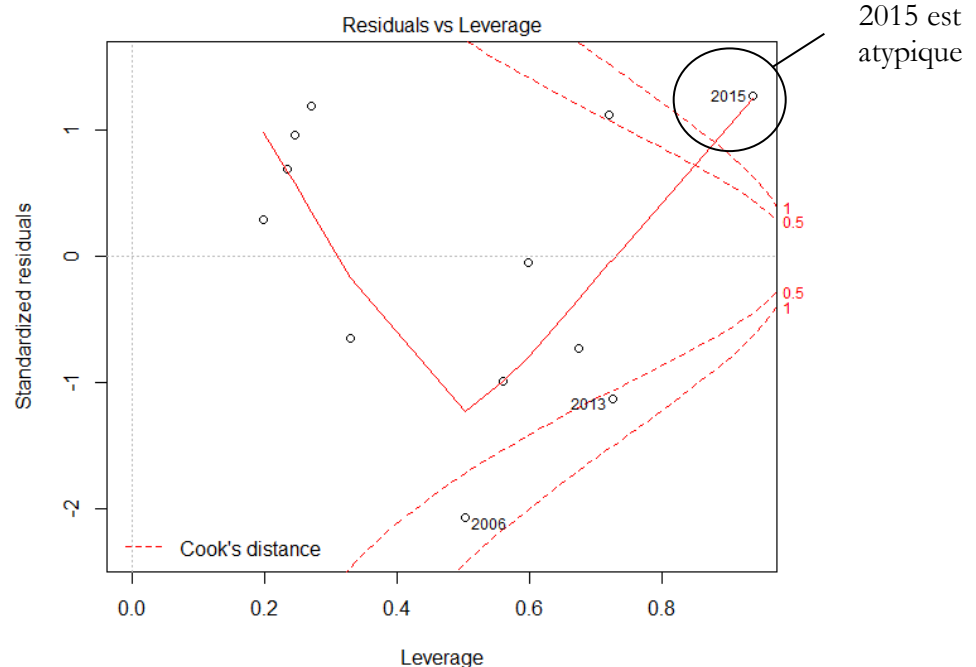
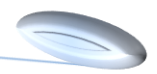


FIGURE 21 - RESIDUS STANDARDISES DU MODELE LINEAIRE COMPLET - SOCIETE A AUTOMOBILE

Fort de ce constat, deux options sont possibles, recommencer la modélisation en retirant ce point ou chercher un modèle plus générique qui permet de tolérer un point atypique comme 2015.

Après plusieurs expériences, il semble plus important de se concentrer sur la création d'un



modèle. On s'attardera sur le modèle linéaire généralisé et notamment sur l'option *stepwise selection* de ce dernier. La traduction de cette option serait une procédure pas à pas, qui va consister à partir d'un modèle et à l'enrichir ou à lui retirer des variables.

Dans le cas où l'on partirait d'un modèle réduit, voire minimaliste, la méthode consiste à tester une par une l'ajout d'une variable et mesure son impact sur le modèle, puis à retenir celle apportant le plus d'information et à répéter l'opération jusqu'à ce que le gain soit trop infime par rapport à la variabilité induite. Cette procédure est appelée en anglais *forward*, car elle va vers l'avant.

L'autre option consiste à partir d'un modèle complet, ou connu, et à tester l'impact du retrait de chaque variable une à une, puis à retirer les variables présentant un intérêt moindre. Cette procédure en marche arrière, appelée *backward*, s'adapte bien à notre situation.

Ce modèle aura la faveur des tests, car il est plus simple à expliquer.

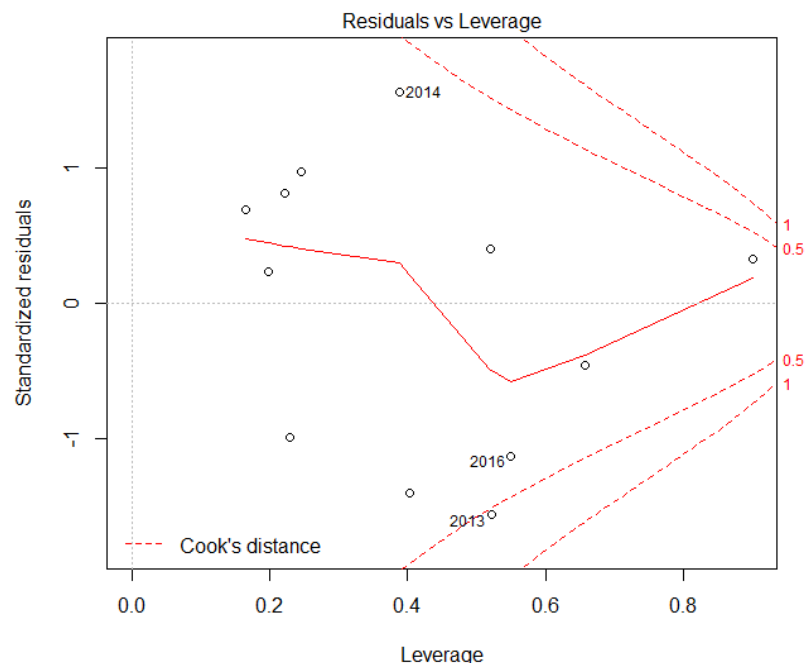


FIGURE 22 - RESIDUS STANDARDISES DU MODELE LINEAIRE GENERALISE *BACKWARD* - SOCIETE A AUTOMOBILE

Là encore, les tests sont réalisés sur l'ensemble du périmètre et l'on pourra constater que l'exercice 2015 est bien souvent dans le domaine acceptable, comme on peut le voir sur la figure précédente. Toutefois, le constat principal est que les critères retenus diffèrent d'un croisement marque/segment à l'autre. Cela exclu donc ce type de modèle, car pour une bonne opérationnalité il est important d'avoir le même modèle sur toutes les marques et segments de marché de Covéa.

Pour pallier à ce phénomène, les notions de marque et de segment ont été intégrées comme facteurs dans les données, il devient possible de les utiliser comme variable explicative dans le modèle.

Le modèle ainsi obtenu présente évidemment des approximations par marque et segments, au sens où il sera moins précis que les modèles individuels présentés ci-dessus. Pour juger de la qualité de modèle, deux indicateurs existent, le critère d'information d'Akaike²⁷ (AIC) et le critère d'information Bayésien²⁸ (BIC). Le modèle à retenir étant celui dont le score est le plus faible.

Si l'on s'attarde sur le modèle retenu avec l'option pas à pas en marche arrière, *backward*, les critères ajoutés au modèle individuel sont la marque et le segment de marché, ils sont ajoutés de manière multiplicative, au sens où c'est le croisement d'une marque ou d'un segment avec les variables initiales qui permettra de définir le portefeuille théorique.

Evidemment, appliquer de telles méthodes sous-entend l'existence de corrélations ou à défaut une interdépendance dans le développement des portefeuilles des marques. Intuitivement, on modélise ici une sensibilité face au marché de ces marques, bien que les hypothèses soient endogènes au système et par définition indépendante du marché.

On retrouve dans ce modèle un exercice 2015 atypique sur tous ces croisements marque et segment de marché. Il devient alors important de débattre du retrait de ce point de la création du modèle. L'atypisme est très certainement lié à un phénomène externe, une rupture. Comme énoncé précédemment, 2015 coïncide avec la mise en place de la loi Hamon. Cette loi ayant donné plus de liberté au consommateur dans le choix de l'assureur et surtout dans la rupture de son contrat, sous réserve de respecter le devoir de s'assurer. L'étude a posteriori de cette évolution a montré qu'il aura fallu 2 ans de transition pour que les nouvelles tendances se stabilisent, soit 2015 l'année de la transition et 2016 une année qui était asymptotiquement proche de l'état final. Dans ces conditions il semble donc préférable de retirer le point 2015 du modèle.

4.3 Les contraintes de la propagation du modèle dans le temps

Le modèle retenu est donc une combinaison linéaire, prenant en compte les interactions entre les variables – dont le segment de marché et la marque – et permettant de minimiser le critère d'Akaike. L'étape suivante consiste à projeter les résultats dans le temps, en intégrant le jeu d'hypothèses défini par la direction commerciale.

NB : d'autres approches pourraient se passer des hypothèses commerciales, soit en réalisant une analyse en série chronologique de leur répartition, soit en intégrant des variables explicatives endogènes ou exogènes (prix de l'énergie, inflation...). Ces options seront à étudier ultérieurement.

Dans le modèle, chaque exercice fait référence au précédent, c'est une forme de modèle auto-régressif. Afin d'encadrer le résultat, il est possible de récupérer les paramètres des résidus du modèle et d'en déduire une erreur de prédiction. L'étude des résidus, par le test de Shapiro²⁹ montre que les résidus sont gaussiens de moyenne nulle et avec un écart type de l'ordre de 5000 unités, ici des contrats. Selon la marque et le segment de marché, cela peut être important.

Pour la projection, il serait possible d'encadrer le premier exercice projeté à partir des résidus du modèle initial. La problématique va être de définir les points suivants, étant donné que chaque exercice calculé par modélisation s'appuie sur le précédent, il faut donc intégrer ce biais dans les exercices futurs. La méthodologie va donc consister à générer consécutivement des scénarios dépendants, dans lesquels chaque exercice simulé – on parlera de nœud – engendrera à son tour plusieurs nouvelles simulations filles. Cf. illustration page suivante.

Il convient de fixer n le nombre de simulations nécessaires à la prédiction d'un intervalle de confiance sur un exercice, afin d'obtenir une précision satisfaisante. Si l'on fixe n à 200 pour générer un intervalle de confiance à 95% par répartition empirique en excluant donc les 2,5% extrêmes, soit les 5 simulations les plus extrêmes à chaque borne autour du scénario moyen. On notera que $n = 40$ est le nombre le plus petit de simulations nécessaires car il permet d'exclure la simulation la plus extrême.

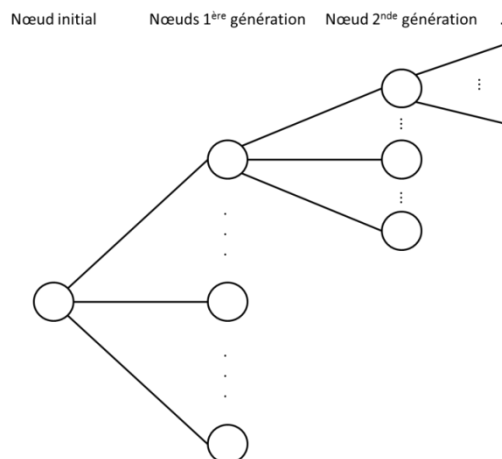


FIGURE 23 - PROPAGATION TEMPORELLE AVEC MEMOIRE : LA SIMULATION PAR NŒUD

Par construction du modèle de propagation, pour chacune des n simulations de premier ordre, il faudra générer n simulations de second ordre. Puis, pour la troisième génération n nouvelles valeurs pour chacune des valeurs obtenues au second tirage, et ainsi de suite.

Avec ce raisonnement, pour le business plan interne projetant les résultats sur une période de 5 ans, il conviendrait de calculer plus de n^5 simulations – pour être exact : $n + n^2 + n^3 + n^4 + n^5$. Soit plus de 320 milliards de cas, en fixant n à 200.

Si l'on considère les limites de calculs des ordinateurs actuels et les contraintes d'espace disque, cette méthode ne peut être exploitée. En effet stocker les résultats, pour les analyser et les exploiter, nécessite pour chaque valeur calculée de lui affecter son nœud, son nœud père et les valeurs obtenues par le calcul pour les 4 croisements marque-segment, soit 6 informations. Les identifiants des nœuds sont à stocker sur un entier allant jusqu'à 200, soit un octet chacun. On rappellera que le stockage binaire de l'information par les ordinateurs, permet avec un octet de mémoriser un chiffre allant jusqu'à 2^8 , soit 256. S'ajoute à ces 2 octets les valeurs à stocker, ces dernières comptent ici des contrats, soit un chiffre de plusieurs millions, soit 3 voire 4 octets pour chacune d'elle. L'espace de stockage approximatif et minimal pour stocker les résultats serait de 4,8 To (téraoctet), pour 15 octets de stockage et 200 simulations. De manière minimaliste, réduire n à 40 revient à stocker un peu moins de 2 Go (gigaoctet) d'information, pour une qualité moindre. Malgré cette baisse importante liée au passage de 200 à 40 nœuds par sous-arbre, il est jugé que la perte de qualité reste trop importante. Par ailleurs, dans la pratique les outils ont besoin de plus d'espace encore, pour stocker, adresser et calculer.

Au regard de ces éléments il a été décidé de limiter à 3 ans la projection par cette méthode, ainsi, toujours avec un nœud à 200 ramifications, le nombre de simulation maximale passe de 320 milliard à 8 millions. Avec cette restriction, l'espace de stockage devrait être de 120 Mo.

La question du stockage étant partiellement résolue, il semble utile de se pencher sur la seconde restriction, celle des temps de calcul. La solution minimaliste, à savoir une profondeur de 3 ans, avec un nombre de simulations de 40 par nœud prend 6 minutes pour 64 000 calculs. A noter que la méthode nécessite pour chaque nœud de recharger les hypothèses de l'exercice et d'y ajouter le résultat du nœud père, puis de réaliser la modélisation linéaire multiple sur ce nouvel échantillon et d'y ajouter un aléa gaussien conforme aux résidus observés par le passé. Passer à 200 simulations, ce qui reste très faible pour la qualité de l'estimation, surtout en première année, va donc demander de réaliser 8 millions de fois le calcul, soit 125 fois plus. On notera que l'augmentation sur 3 ans correspond au cube de l'augmentation de première année. En partant du principe que le temps de calcul est linéaire, on peut estimer la durée à 46 000 secondes, soit approximativement 12 heures et 45 minutes. Pour passer à 1 000 scénarii, la durée est estimée à plus 66 jours.

Ces temps de réponses restent insatisfaisants dans le cadre d'une étude pluriannuelle, une alternative est donc proposée par la suite. A noter que ces temps de réponse sont ceux obtenus sur les machines mise à disposition de l'équipe. En travaillant sur un serveur dédié il serait possible d'obtenir une puissance de calcul plus importante. De plus, le langage R permet de travailler en mode *multi-core* ou multi-cœur en français. Cela signifie que le langage permet de faire travailler la machine sur l'ensemble de ses cœurs de calcul physique et cela de manière simultanée. Si le gain en temps peut aller jusqu'à un facteur 4x sur les machines individuelles (quadri-core) il pourrait être bien plus important sur un serveur correctement configuré. Enfin, cette évolution nécessite de repenser l'intégralité du code et de maîtriser le langage R. Ce qui ajoute à l'outil une complexité d'utilisation dans la maintenance opérationnelle, sans parler du coût informatique pour déployer une solution de type serveur ayant plusieurs téraoctets de capacité de stockage.

La solution proposée va consister à donner une plus grande importance aux exercices les plus proches du nœud initial. Cette importance est donnée en nombre de simulations. Il est alors possible de gagner en nombre de calcul tout en perdant en précision. Cependant, il devient possible de simuler un premier exercice avec 1 000 points différents, soit une précision accrue. A l'instar de météo France, il est alors concevable d'attribuer un indice de confiance, en plus de l'intervalle, du fait de cette perte de précision. Ce point restera à creuser, même s'il est important d'avoir en tête que le résultat encadré sera de moins en moins fiable à chaque génération temporelle. Par exemple, simuler un premier exercice avec 1 000 cas possibles, pour ces cas de première génération, 100 cas seront simulés en deuxième génération et pour la troisième génération 10. Enfin, chaque nœud pourrait être projeté "certainement" sur les générations 4 et 5, sans ajouter de nouvel aléa. On arrive alors à calculer un intervalle de confiance empirique jusqu'à 3 ans ou jusqu'à 5 ans sans aléa, qui prend en considération le champ de possibles.

Initialement, l'idée était de partir sur un modèle avec 1 000 simulations pour le premier nœud, 200 pour les seconds nœuds et 40 pour les troisièmes. Cela aurait permis des approches différentes dans la présentation des résultats. Avec 40 simulations par nœud, issues de la seconde génération, il deviendrait possible de calculer des statistiques empiriques par nœud et d'estimer ainsi les intervalles de confiance moyens à 95%. Soit des résultats plus robustes. Une telle simulation prend près de 20 heures, ce qui semble long pour seulement 2 marques et 2 segments de marché. La méthode proposée dans le paragraphe précédent prend environ 2 heures et 30 minutes, ce qui est plus acceptable.

On notera qu'il est possible de réaliser 5 000 simulations en première année, puis 200 en seconde et 1 seule en troisième (quatrième et cinquième) pour environ la même durée que la méthode retenue. Ce pourrait être une solution à approfondir, car elle permet une bonne approximation de la première année, tout en obtenant une seconde année avec de la précision. Par contre, la troisième année devient "certaine", ce qui rend sa lecture moins pertinente.

NB : lorsque l'on parle ici d'une projection certaine, c'est au sens où elle s'appuie sur un résultat connu sans ajouter un nouvel aléa dans le passage à l'exercice suivant.

Sur la page suivante, sont présentés les résultats obtenus pour les 3 méthodes. La première méthode, avec 40 simulations par exercices ou méthode minimaliste. La méthode retenue, avec 1 000 simulations en première année, puis 100 en seconde et 10 en dernière, ou la méthode intermédiaire. Ou encore la méthode qui surexposition du premier exercice (N+1) au détriment du troisième exercice. On voit immédiatement que la première année est très grossièrement estimée sur le modèle minimaliste, presque correctement sur le scénario intermédiaire et bien sur le dernier scénario.

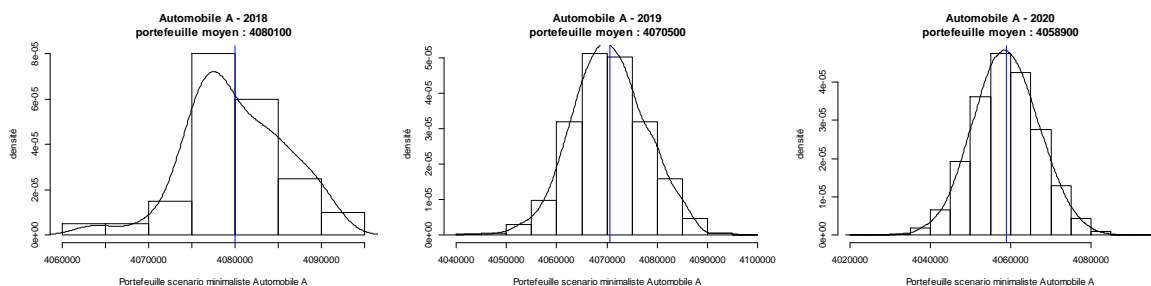
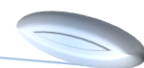


FIGURE 24 - PROPAGATION DU SCENARIO MINIMALISTE POUR LE PORTEFEUILLE DE LA SOCIETE A, SEGMENT AUTO

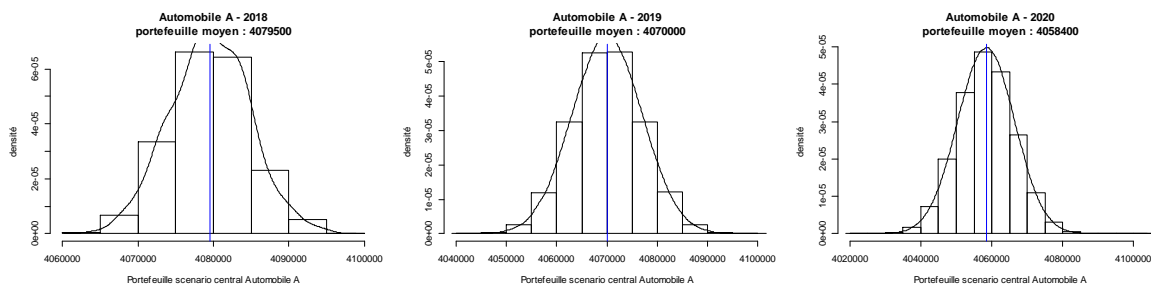


FIGURE 25 - PROPAGATION DU SCENARIO INTERMEDIAIRE POUR LE PORTEFEUILLE DE LA SOCIETE A, SEGMENT AUTO

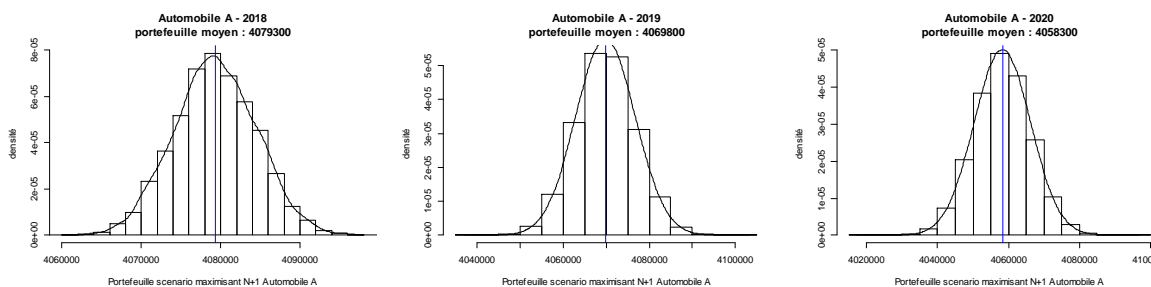


FIGURE 26 - PROPAGATION DU SCENARIO OPTIMISANT N+1 POUR LE PORTEFEUILLE DE LA SOCIETE A, SEGMENT AUTO

L'Intervalle de Confiance, aussi noté IC, diffère selon les scénarios testés. Cela est assez peu visuel sur les graphiques précédents, on s'attardera sur le tableau suivant. La taille de l'intervalle de confiance a été définie comme le rapport entre l'étendue à 95% – i.e. l'écart maximal entre les simulations fixant les bornes de l'IC – et la valeur moyenne des simulations de l'exercice.

Taille de l'intervalle de confiance

Scénario	2018	2019	2020
Minimaliste	0,39%	0,60%	0,67%
Retenu	0,42%	0,56%	0,65%
Surexposition N+1	0,41%	0,56%	0,65%

FIGURE 27 - COMPARAISON DE LA TAILLE DES IC SELON LA METHODE DE PROPAGATION APPLIQUEE, PORTEFEUILLE AUTO SOCIETE A

On notera qu'en première année la propagation minimaliste – soit le modèle qui génère 40 estimations par nœud – permet d'obtenir un intervalle assez faible en taux, il est plus grand en valeur absolue car la moyenne de 40 points est plus forte que celle des modèles à 1 000 ou 5 000 simulations.

Par ailleurs, il est délicat de se faire une idée sur la propagation temporelle, car en rejouant le scénario avec une graine différente, ou successivement (car les aléas générés diffèrent) on

obtient des résultats différents.

Dans cette dernière figure, comme abordé plus tôt, on constate que plus l'on s'éloigne du point d'origine et plus l'imprécision et donc la taille de l'intervalle augmente. Toutefois, le fait de réduire fortement ou totalement le nombre de simulations d'année en année semble avoir peu d'impact sur cet effet. A noter qu'il faudrait lancer des scénarii plus long pour débattre correctement de ce point.

Pour rappel, la méthode avec surexposition N+1, consiste à générer 5 000 nœuds en première année puis 200 en seconde et 1 seul en troisième année.

En termes de précision, sur la première année, il est possible de comparer la variance observée dans les simulations à celle du modèle d'origine. Id est, la variance de l'échantillon comparée à celle des résidus du modèle initial. En rapportant cet écart à la variance des résidus, on obtient un indicateur de la précision de la simulation. Ainsi, la méthode minimaliste est peu précise et celle, non retenue, avec 5 000 estimations l'est bien plus :

Scénario	2018
Minimaliste	4,7%
Retenu	2,0%
Surexposition N+1	0,8%

FIGURE 28 - ESTIMATION DE LA PRECISION DES MODELES A 1 AN

Il serait intéressant de généraliser cette méthode aux exercices suivants, toutefois un biais s'intègre dès lors que l'on utilise un nœud père différent, au sens où la précision mesurée ainsi se calcule entre un nœud père et l'ensemble de ses fils.

L'ensemble des éléments mis en avant laisse à croire que le dernier modèle serait le meilleur. Son défaut principal, outre les quelques 3 heures nécessaires à son calcul, est qu'il perd en précision sur la 3^{ème} année, et cela uniquement par construction. En effet, aucune incertitude propre au 3^{ème} exercice n'est intégrée dans cette propagation.

4.4 Quelles conclusions pour ce modèle

Sans entrer dans les retours faits a posteriori, notamment sur l'atterrissage 2018 et la qualité de l'estimation, car ce sujet reste encore très confidentiel, il faut noter que cette approche aurait été un meilleur estimateur du résultat obtenu que la méthode en place lors de la réalisation des business plan en 2017.

On s'attardera toutefois sur les limites de la méthodologie, et particulièrement sur la société B et son portefeuille habitation. En effet, le modèle de prédiction retenu semble moins pertinent et adapté que le modèle initialement calculé pour cette marque et ce segment.

Pour la visualisation, les graphiques ci-après montrent que le modèle initial permettait d'envisager une croissance. Tandis que le modèle retenu au final, modèle identique pour toutes les marques et les segments de marché, affiche-lui un recul du portefeuille.

Pour un gain de temps le modèle spécifique a été propagé à 1 an seulement, avec un nombre de simulations identique au modèle retenu. On constate clairement qu'en première année la tendance est inversée. Ce phénomène remet en question le modèle unique, bien que les intervalles de confiance se recourent.

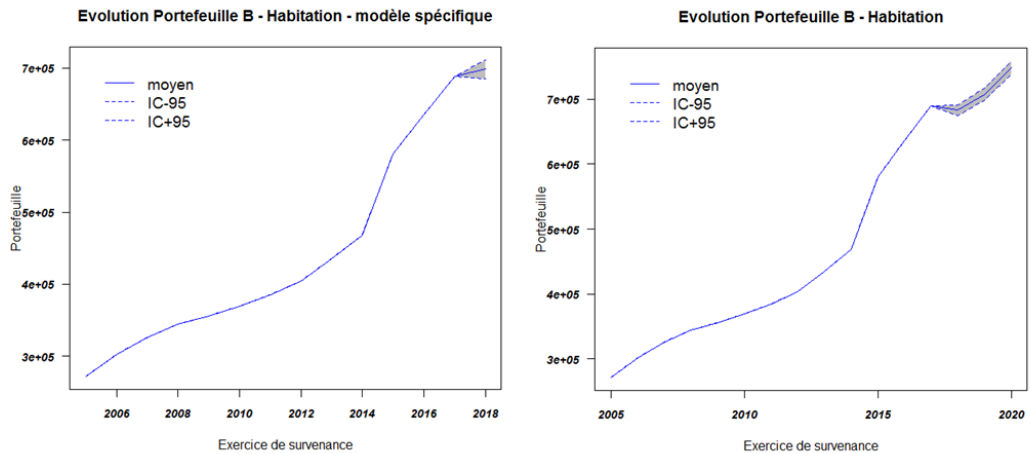
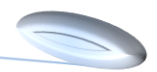


FIGURE 29 - COMPARAISON MODELE SPECIFIQUE ET MODELE GENERIQUE

Il serait intéressant d'approfondir le sujet en généralisant à toutes les marques de Covéa et de prendre le temps d'étudier les résidus du modèle GLM par segment de marché et par marque. En toute logique, ayant intégré ces facteurs dans le modèle, les résidus en sont décorrélés. Toutefois, étudier plus en détails ce modèle, une fois enrichi de toutes les marques, pourrait aider à décider de la meilleure approche à avoir.

Il semble, malgré toutes les réserves émises, que ce modèle soit plus efficient que l'actuel et qu'il répond aux conditions de réalisation du Business Plan, notamment dans la définition des hypothèses commerciales avec la direction éponyme.

A compter de cette étape, il reste à intégrer à cette projection du portefeuille la charge des sinistres et les primes acquises.

5 Aller plus loin en ajoutant la volatilité liée à la propagation

Sujet partiellement abordé lors de l'intégration de la volatilité liée à la charge sinistre connue, notamment lors des tests de corrélations entre marque et segment – qui seront qualifiées par la suite de dimensions spatiales – Ici l'accent sera donné à la propagation temporelle.

La principale question que l'on souhaite aborder est la corrélation entre les éléments de l'équation et leur volatilité.

Dit autrement, les erreurs d'estimation sont-elles indépendantes et sommables ? ou bien au contraire dépendantes et ayant des effets accentuant les autres phénomènes ?

Ou encore, l'erreur induite par le modèle de contrat génère-t-il une amplification de l'erreur induite par la liquidation des sinistres ? Et si oui dans quelle mesure ?

5.1 Passer d'un portefeuille futur à une charge sinistre espérée

Dans le modèle actuel, le lien est direct, il est obtenu en deux temps. Une fois les portefeuilles projetés à 5 ans, il convient de les multiplier par la prime pure technique estimée pour le futur. On en déduit immédiatement la charge sinistre.

Dans la pratique, comme évoqué au chapitre III-A-4, la prime pure des années future est estimée par analyse de série chronologique. Une fois les résultats économiques arrêtés et lissés, il convient de les projeter sur la base des tendances du passé. Dans ces projections s'intègre, par discussion avec des experts ou sur la base d'études spécifiques, un correctif de tendance pour déformation probable du risque dans le temps. Par exemple, si un nouveau produit est commercialisé, que ce dernier vise les résidences principales de type maison, la déformation du portefeuille va entraîner une déformation du coût.

Par manque de temps, la transition de cette partie vers un modèle moins déterministe a été laissée de côté. On notera que l'approche serait semblable à celle de la propagation du portefeuille.

L'évolution future peut être considérée comme une tendance définie sur le passé, à laquelle s'ajoute une incertitude, ou un bruit blanc, par définition gaussien.

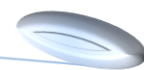
Une alternative serait d'estimer un modèle auto-régressif (ARMA ou ARIMA), sur la série étudiée, par marque, segment de marché et garantie. Partant de là, les résidus du modèle, s'ils sont gaussiens, peuvent être réintégrés. En cas de répartition de ces résidus non normale, il faut étudier s'il y aurait une fonction de lien à considérer ou une autre loi sous-jacente.

Puis, tout comme les portefeuilles sont projetés, avec un système de nœud de père en fils, ajouter à chaque simulation l'évolution théorique attendue, basée sur un à plusieurs exercices passés, puis moduler cette valeur sur la base du bruit existant dans le signal d'apprentissage.

Il resterait alors une inconnue à traiter, à savoir, comment se comporte la prime pure et son encadrement, sachant que, en moyenne, la propagation temporelle implique une évolution de $x\%$.

Pour cela, plusieurs solutions sont à étudier. Tout d'abord, étudier s'il est possible d'approcher la prime pure 2017 (par exemple) par une loi qui permettrait de retrouver une densité et une répartition proche de celle obtenue par les méthodes de bootstrap. Ainsi, il devient possible pour chaque nœud d'estimer une prime pure théorique avec une volatilité de liquidation. Puis de lui injecter une évolution temporelle, sur la base des prédictions arrêtées pour les BP, et de la volatilité mesurée (par exemple en réinjectant des résidus par bootstrap) et conditionnée à son héritage (le chemin parcouru dans la succession de nœuds) et à l'aléa propre au nœud fils.

Dans le cas où il serait impossible de trouver une loi pour encadrer la prime pure, il est alors plus délicat d'obtenir, par simulation un Intervalle de confiance.



5.2 Passer d'un portefeuille à une prime acquise

Dans le modèle actuel, le lien est indirect, chaque composante du portefeuille contribue également à la prime, toutefois ce modèle semble trop complexe pour être modélisé de manière itérative, telle que le fut le portefeuille précédemment.

Une première solution consiste à se ramener à un système plus simple, sur les mêmes bases que dans le chapitre précédent sur la projection de la charge. Il est possible de résumer le chiffre d'affaires à un portefeuille et une cotisation unitaire. L'évolution du portefeuille a été projetée dans la partie précédente, il convient donc de projeter la cotisation unitaire.

La cotisation unitaire évolue en fonction de deux paramètres, les mesures prises en tarification et la déformation liée au portefeuille. La mesure tarifaire est arrêtée sur la base de l'évolution de la charge sinistre espérée, plus précisément de la prime pure des sinistres projetée, donc un élément calculé dans le modèle. Par conséquent, il y a une interaction forte entre la propagation des sinistres vue à l'étape précédente et la cotisation moyenne.

Concernant l'effet de structure – ou encore l'effet résiduel de l'évolution de la cotisation après retrait de l'impact du tarif – il est possible d'apprendre du passé pour en déduire un modèle par étude chronologique. Ce modèle peut être, avec ou sans tendance, linéaire ou non... Il devient alors possible de le réinjecter dans la projection future, en plus du tarif, avec un aléa basé sur les erreurs de modélisation, ou directement, sans modèle, comme cela se fait avec le bootstrap.

Toutefois, des éléments externes peuvent venir modifier cette déformation et donc rendre un modèle mathématique inopérant. On retrouve comme éléments perturbateur le lancement d'une offre, l'arrivée sur le marché de SUV au détriment de véhicules familiaux, ou encore des contraintes réglementaires, comme la DDA (Directive sur la Distribution d'Assurance) qui agit sur l'équipement et le prix. Une autre piste, plus visible chez des bancassureurs, est la baisse des taux, qui permet aux consommateurs d'obtenir des prêts plus importants et donc de s'acheter des biens (voiture/appartement/maison...) à valeur plus importante, ce qui entraîne des prix unitaires d'assurance plus élevés.

Etant donné ces considérations, il semble complexe de projeter jusqu'à 5 ans des bruits autour du tarif constant. Il semble opportun de prévoir une extinction dans les modèles de ces effets. Pour preuve, on pourra parler de l'impact qui est exponentiel chaque année. Ainsi la multiplication des effets peut conduire à des hausses très marquées de la prime acquise. Ce phénomène est à surveiller dans un contexte de projection. Il serait dangereux de propager une forte hausse des primes du fait d'une déformation structurelle observée dans le passé, liée à une conjoncture à un instant t , et qui sera finie ou dont la fin sera programmée à l'horizon du BP.

Pour la réalisation, deux approches sont alors possibles. La première consiste à rester sur un processus voisin de l'existant, en injectant des hypothèses tarifaires sur la base des projections des sinistres. L'inconvénient dans la réalisation va être les temps d'exécution, il faudra attendre d'avoir réalisé la projection de la charge sinistre avant de propager la prime, tout en projetant la prime en respectant les simulations de sinistralité et de portefeuille réalisées précédemment.

La seconde option pourrait consister à développer un modèle autonome, qui ajuste donc la mesure tarifaire dynamiquement, selon l'évolution de la sinistralité constatée d'un nœud à l'autre. Cela revient donc à faire automatiquement ce qui est fait d'une année sur l'autre, en avance de phase. In fine, un tel modèle permettrait de récupérer l'évolution moyenne de la sinistralité et son encadrement, ainsi que l'effort tarifaire à prévoir, et son encadrement, afin d'obtenir un résultat, lui aussi encadré. Dans cette seconde approche, la machine calcule énormément de scénarii. Il faut donc rester vigilant et le rôle de l'actuaire évolue vers un rôle de réflexion, d'analyste et de critique des résultats, et non plus celui de réalisateur de calculs.

6 Conclusion

Pour résumer les présents travaux et définir les évolutions importantes réalisées pour Covéa, il semble approprié de commencer par rappeler le point de départ, puis dans un second temps de présenter un résumé des méthodes mises en œuvre et les résultats obtenus. Enfin cette conclusion traitera des gains potentiels pour Covéa et pour la profession avant d'exposer les suites à donner à ces travaux.

6.1 Le contexte initial

Avant tout, on rappellera la motivation première à l'origine de ces travaux. Le business plan stochastique permet d'apporter à Covéa une vision probabiliste des résultats. Outre l'encadrement de la trajectoire, cela apporte la possibilité de définir – pour un indicateur comme le ratio combiné – la probabilité de dépasser un seuil de résultat à horizon 3 à 5 ans. On notera également que les résultats du BP sont aujourd'hui recalculés dès lors que les évolutions de certains paramètres ne suivent pas le scénario initial, déterministe. Cela serait donc un gain supplémentaire que de pouvoir disposer d'une vision encadrée des résultats. A l'aune de ces constats et hypothèses, il semblait opportun d'étudier les possibilités pour orienter le business plan vers une réalisation de plus en plus stochastique.

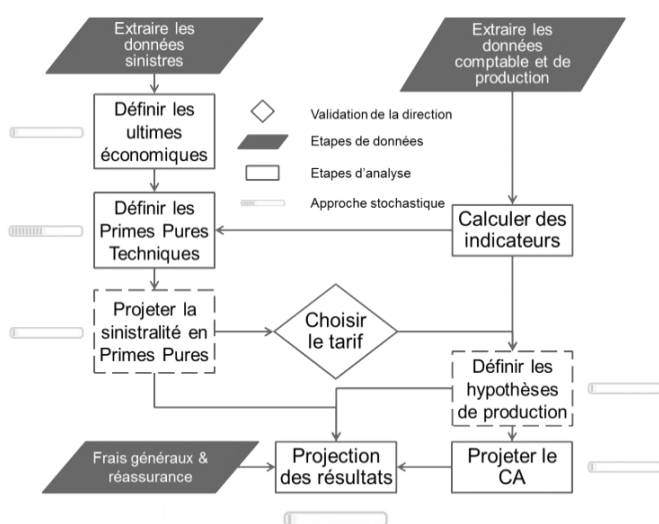


FIGURE 30 - LE BP COVEA ET SA PART D'ETAPE STOCHASTIQUE

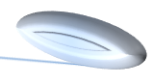
Comme l'illustre la figure ci-contre, bien que la majorité du BP soit aujourd'hui réalisée de manière déterministe, il est à noter que les primes pures techniques sont partiellement calculées par approche stochastique.

Ces approches sont réalisées pour les sinistres corporels de plus de 150 k€ et pour les sinistres climatiques, attendants ou non aux catastrophes naturelles. Pour ces périmètres, les lois sous-jacentes au nombre de sinistres et à leur coût unitaire sont étudiées. Puis des simulations par approche de Monte-Carlo sont effectuées pour construire des exercices types.

Malgré cette approche probabiliste, seule la valeur centrale sera retenue pour la suite du business plan.

6.2 Les évolutions pour le BP travaillées dans le cadre de ce mémoire

Dans ce mémoire, 3 parties ont été approfondies pour intégrer une volatilité dans le résultat ou améliorer la connaissance que l'on en avait. On retrouve ainsi, la modulation des simulations des primes pures techniques avec troncature des lois de coûts, une proposition de projection à 5 ans des portefeuilles de manière stochastique, permettant de déduire un chiffre d'affaire et enfin la liquidation stochastique de la charge des sinistres par approche en triangle.



6.2.1 Prime Pure Technique : caper le modèle de coût unitaire

Si l'on s'attarde sur les propositions faites sur les primes pures techniques en RC corporelle Auto, la méthode utilisée pour borner les sinistres graves ($\geq 150\text{k€}$ en Auto) au-delà d'un seuil historique a été révisée. Initialement la borne était fixée à 50M€, et tout sinistre simulé dépassant ce seuil se voyait affecter une valeur plafonnée à 50M€. Etant donné les lois de coût et de nombre modélisées, ainsi que le nombre de tirages aléatoires effectués, cette valeur plafond pouvait être simulée plusieurs fois, parfois dans le même exercice. Etant donné que l'indicateur final retenu est la moyenne – et l'on connaît la sensibilité de cette dernière face à des extrêmes forts – il semblait utile de retravailler cette méthode.

La première étape a donc été de discuter le seuil des 50M€. On rappellera qu'au sein de Covéa, après normalisation des résultats, i.e. transposition des paramètres constituant les dépenses à la valeur actuelle, le plus gros sinistre connu est de 25M€. On sait également qu'à ce jour sur le marché, en vision réassureur le plus gros sinistre est estimé à 43M€ – calculé sur des taux d'actualisation différents de ceux utilisés pour normaliser les données en interne Covéa. D'après les informations à notre disposition, notamment les traités en vigueur, il est cohérent de penser que le taux d'actualisation appliqué à ce dossier est de 1,50%. La question se pose de la revalorisation de la rente attenante, et de la normalisation appliquée... Sachant que pour les études internes le taux d'actualisation utilisé permet de compenser la revalorisation, soit 2,25%, il semblerait que cette plus haute valeur marché devrait être inférieure à 43M€, autour de 35M€. Le montant de 50M€ semble donc encore non atteint, sur le marché, depuis l'apparition de l'automobile et de son assurance.

Si l'on s'attarde à présent à la seconde phase du correctif, il s'agit de revoir l'étendu de la loi modélisée pour que sa valeur la plus forte simulée soit donc de 50M€. Si l'on étudie la répartition du coût unitaire normalisé des sinistres corporels de la société A, cf. courbe suivante, on peut la modéliser par une loi mélange log-log.

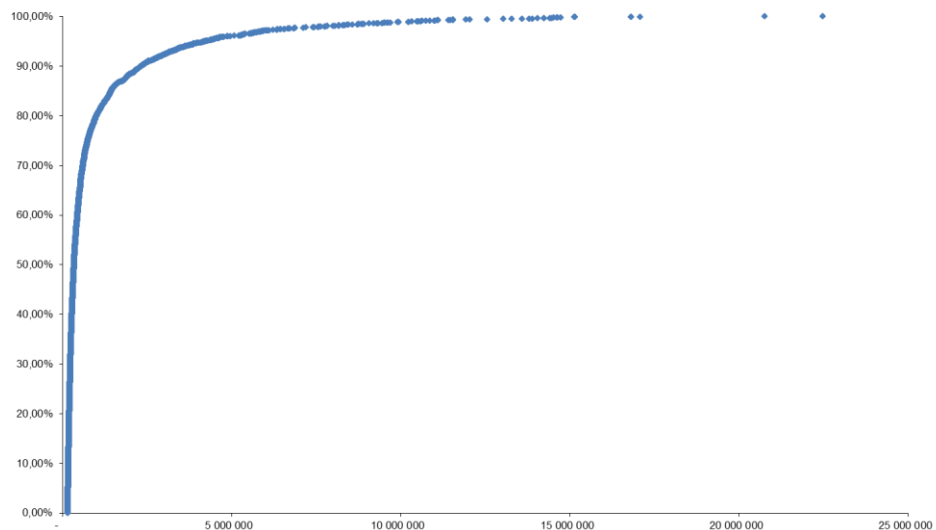
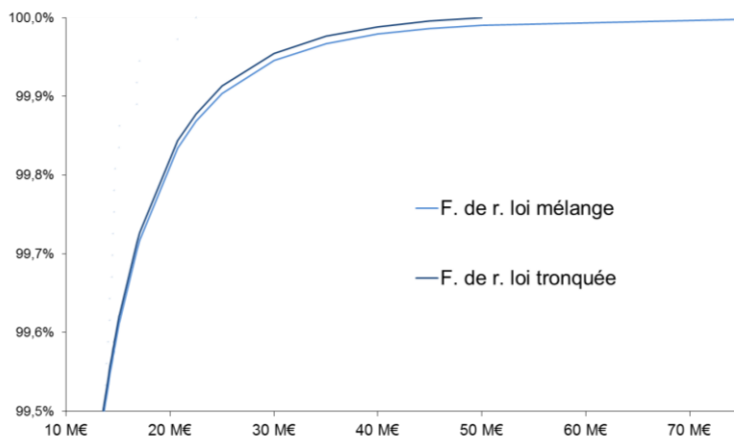


FIGURE 31 - REPARTITION EMPIRIQUE DES SINISTRES DE PLUS DE 150K€

Le mélange de lois sous-jacentes nous permet d'estimer à environ 1 cas sur 5 000 la survenance d'un sinistre de 50M€. Pour passer d'une probabilité en évènement à une période de retour, on s'appuiera sur la distribution annuelle des évènements. Avec en moyenne légèrement moins de 100 évènements par an, on estime la période de retour d'un sinistre de 50M€ à 52 ans, sous hypothèse de stationnarité, pour la société A.

On notera qu'à l'échelle de Covéa, sous hypothèse de conservation de la loi de coût, cette période de retour serait inférieure à 10 ans et estimée entre 2 et 3 ans pour le marché.

Ces constats étayant donc la proposition initiale visant à revoir la distribution en appliquant une loi tronquée, au sens probabiliste. Ci-dessous les fonctions de répartition de la loi mélange et de la loi mélange tronquée.



La loi proposée permet de conserver des probabilités proches jusqu'à 99,5%. La figure ci-contre met en avant l'extinction plus rapide de la loi tronquée.

De plus, cette méthode donne des estimations des périodes de retours plus en phase avec ce qui est connu aujourd'hui (cf. tableau ci-après).

FIGURE 32 - REPARTITION MODELISEE DES SINISTRES GRAVES AU-DELA DE 10M€

On note que le choix d'un plafond à 50M€ reste contestable, bien qu'il ait été discuté précédemment il pourrait être revu à la hausse au fil du temps, pour intégrer l'expérience et l'inflation. Ce montant et cette méthodologie ont été soumis à la fonction actuarielle de Covéa pour validation.

Périodes de retour estimées en années		Montant du sinistre					
		25 M€	30 M€	35 M€	40 M€	45 M€	50 M€
Loi mélange	Société A	5	9	15	23	36	52
	Covéa	1	2	3	5	7	10
	Marché	1/4	4/9	1	1	2	3
Loi mélange tronquée	Société A	6	11	21	43	112	∞
	Covéa	1	2	4	9	22	∞
	Marché	2/7	1	1	2	6	∞

FIGURE 33 - ESTIMATION DES PERIODES RETOUR SELON LE MONTANT DU SINISTRE

NB : le calcul de la période de retour se lit comme l'inverse de la probabilité qu'un évènement dépasse le seuil fixé divisé par le nombre annuel moyen de sinistres, sur la base des travaux de Gumbel.

6.2.2 Prime Pure Technique : choix du taux d'actualisation des rentes

Une fois la notion de plafond de modélisation traitée, un autre point important à aborder est la normalisation de l'historique. Plus particulièrement les hypothèses prises pour les taux. L'article 143-2 du règlement de l'ANC relatif aux comptes annuels des entreprises d'assurance, précise le taux de revalorisation et la table de mortalité à appliquer pour le provisionnement comptable. Depuis le 11/12/2018 ce taux de revalorisation des rentes, dans le cadre du désengagement du FGAO, a été revu de 2,25% à 2%, cf. règlement n° 2018-08. Cela pose donc question. Aujourd'hui à l'heure de taux historiquement bas, est-il concevable de maintenir les hypothèses économiques à long termes à 2,25% ?

A noter que la question pourrait être de savoir si l'hypothèse initiale, i.e. la compensation de la revalorisation par le taux d'actualisation, est toujours possible dans le contexte actuel des taux et la baisse de la revalorisation.

On pourra pour cela s'appuyer sur les courbes EIOPA et IA, en prenant l'espérance de vie de la table TD 88-90 – table préconisée par l'ANC – de la victime type en sinistres corporels majeurs. On pourra alors estimer le rendement moyen au-dessus du taux sans risque à réaliser sur la durée de vie moyenne de l'individu type. Nota bene, on pourrait critiquer la méthode en mettant en avant que l'individu type coûte en général moins cher à indemniser que les profils majeurs, comme un enfant de moins de 5 ans par exemple. Toutefois, selon la table de mortalité l'espérance de vie pour une victime jeune est plus longue, ce qui donne plus de temps pour compenser l'écart et réduit la surperformance à enregistrer sur ces profils plus coûteux. Surtout que la courbe des taux utilisée remonte au-dessus du taux de revalorisation...

Des travaux ont été réalisés à l'été 2017, pour valider l'hypothèse d'un taux d'actualisation de 2,25% permettant de compenser la revalorisation et de raisonner sur un risque viager uniquement. Pour cela l'étude s'appuie sur la victime type en termes de sexe et d'âge, soit un homme de 42 ans.

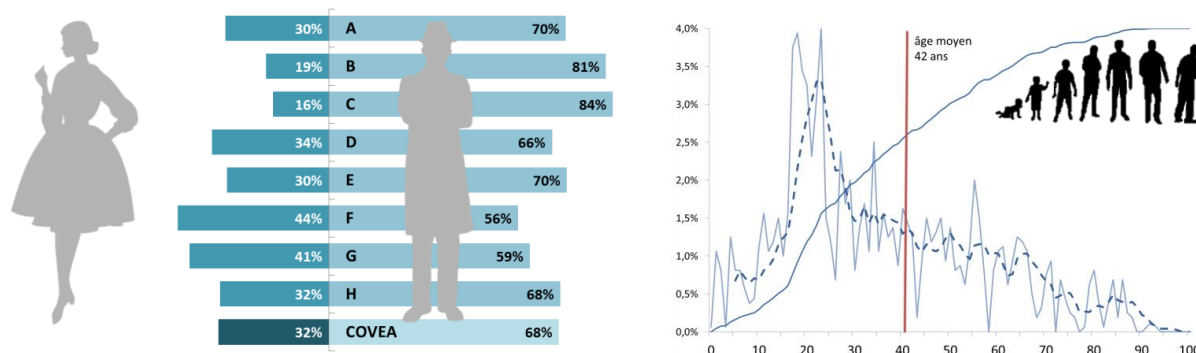


FIGURE 34 - PROFIL DE LA VICTIME TYPE POUR LES SINISTRES CORPORELS IMPORTANTS CHEZ COVEA

A la lecture de la table en vigueur dans les normes comptables, soit la TD 88-90, l'espérance de vie de notre individu type serait de 34 ans. Le choix de la table se discute or il n'existe pas de table sur le marché permettant d'estimer l'espérance de vie d'un accidenté de la route ou d'un accidenté grave, toutes causes confondues. Partir de la norme comptable fait aujourd'hui encore sens au sein de la profession.

Dans ces conditions, nous cherchons à calculer l'écart entre un euro revalorisé à 2,25% par an et ce même euro placé au taux sans risque entre le 01/01/2017 et le 31/12/2051 (soit 34 ans plus tard).

L'écart obtenu entre ces 2 montants sera traduit sous forme d'un taux fixe de rendement complémentaire sur la période, on l'appellera taux de surperformance.

$$\text{Revalorisation}(1\text{€ } 2,25\% \text{ } 34 \text{ ans}) = 1 \times (1 + 2,25\%)^{34}$$

$$\begin{aligned} \text{Placement}(1\text{€ } EIOPA/IA \text{ } 34 \text{ ans}) \\ = 1 \times (1 + 0,25\%) \times (1 + 0,25\%) \times \dots \times ((1 + 2\%)) \end{aligned}$$

$$\begin{aligned} \text{Hypothèse}(1\text{€ } EIOPA/IA \text{ } \Psi \text{ } 34\text{ans}) \\ = 1 \times (1 + 0,25\% + \Psi) \times (1 + 0,25\% + \Psi) \times \dots \times ((1 + 2\% + \Psi)) \\ = \text{Revalorisation}(1\text{€ } 2,25\% \text{ } 34 \text{ ans}) \end{aligned}$$

La résolution de ce système permet d'estimer le sur-rendement moyen Ψ sur la période qui serait de 129 bp (+1,29%) par an si l'on s'appuie sur la courbe EIOPA et de 121 bp selon la courbe IA. A noter qu'avec des tables de mortalité plus récentes, l'espérance de vie augmente et donc ce rendement lui diminue.

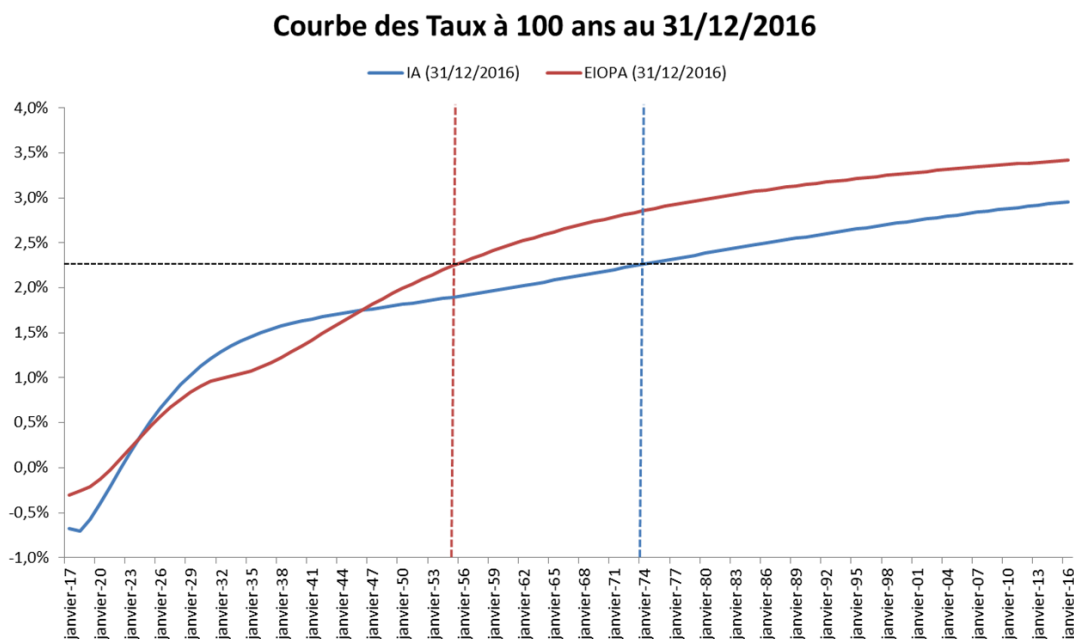


FIGURE 35- ESTIMATION DE LA DATE DE DEPASSEMENT DU TAUX DE REVALORISATION A 2,25% SUR LES COURBES DES TAUX SANS RISQUE EIOPA/IA A FIN 2016

Cet objectif a été soumis aux directions appropriées pour avoir leur avis. Au regard des performances observées sur les exercices récents, cela apparaissait crédible. Le contexte ayant évolué entre l'été 2017 et aujourd'hui, il semble donc important de revoir ces calculs et de revalider les scénarios.

Il est donc possible de résumer les travaux d'optimisation des méthodes stochastiques en vigueur à ce jour chez Covéa par la mise en place d'un cap à 50 M€ et par une étude sur le taux d'actualisation à utiliser, étude évidemment à actualiser du fait de l'évolution de la réglementation et des courbes des taux.

6.2.3 Projection du portefeuille à 5 ans

La seconde partie traitée dans ce mémoire pour s'orienter vers un BP stochastique au sein de Covéa est la projection des portefeuilles. Cette partie se décompose en deux phases, la projection à 1 an et la projection au-delà de cette première année.

6.2.3.1 Projection à 1 an

Concernant la projection à 1 an, on se souviendra que les travaux s'appuient sur un algorithme auto-récursif d'estimation du portefeuille. Plusieurs modèles ont été réalisés, d'abord un par marque et segment de marché, puis un modèle unique intégrant la marque et le segment de marché comme facteur.

$$Expo^N \sim (Marque + Produit) * (Expo^{N-1} + AN + AP^{si} + AP^{cf} + AP^{sf})$$

Les affaires perdues suite à décision du client et à effet immédiat n'impactent pas le modèle, tout comme appartenir à la société A ou au produit Auto. Pour ces 2 dernières informations, étant donné que ce sont les modalités les plus représentées cela semble logique, leur effet étant absorbé par les autres paramètres.

Si l'on s'attarde sur les prévisions à 1 an réalisées à partir de ces éléments et que l'on observe le résultat obtenu, on retrouve le graphique suivant.

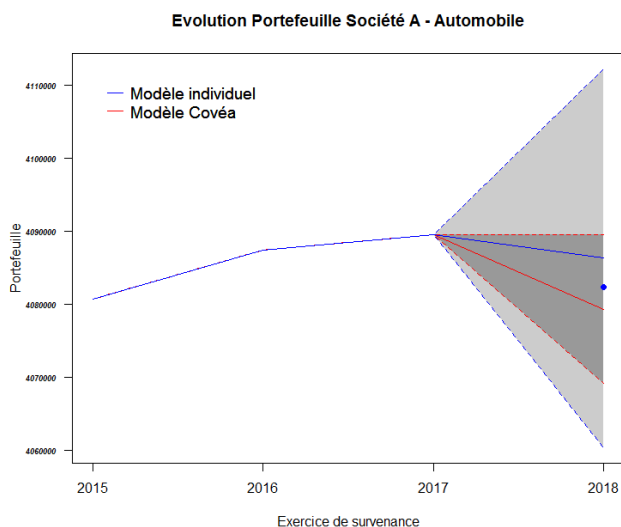


FIGURE 36 - COMPARAISON DES PREVISIONS A 1 AN MODELE COVEA VS MODELES INDIVIDUELS ET ATERRISSAGES REELS

Ce graphique permet de voir que les 2 modèles permettent d'estimer la valeur réelle, qui se situe à l'intérieur des intervalles de confiance (en gris, ici à 95%).

En annexe, la figure 70 permet de visualiser tous les graphiques par marque et segment de marché. On notera que l'atterrissage réel 2018 est compris dans les intervalles de confiance sauf pour la société B en Auto. Les ambitions commerciales étaient trop fortes et aucun modèle ne savait le prévoir.

On rappellera que l'un des objectifs premiers de ces travaux était de passer d'une vision marque à une vision groupe, aussi le modèle global Covéa sera privilégié pour la suite. A noter que les résultats permettent de prendre cette décision, car la taille de l'IC est plus faible et tout aussi fiable.

Enfin, on rappellera que des variables exogènes au système auraient pu être intégrées, notamment des indicateurs démographiques ou une sensibilité au prix de l'énergie. Un tel changement imposerait de prendre des hypothèses sur un domaine autre que l'assurance ce qui semble encore plus complexe.

6.2.3.2 Hypothèses du modèle

Dans le cadre du modèle de portefeuille projeté à 5 ans, diverses hypothèses sont prises. A commencer par le nombre d'affaires nouvelles escomptées pour l'année et le nombre d'affaires perdues. On notera que sur les affaires perdues une distinction est faite entre l'origine (siège ou client) et la prise d'effet (immédiat ou fin de période contractuelle).

Ces hypothèses sont construites sur la base de sources diverses, principalement l'analyse des indicateurs dans le temps et la disponibilité des forces de ventes sur les segments de marché, compte tenu des objectifs fixés sur l'ensemble du périmètre.

Initialement, cette partie devait intégrer une analyse des réalisations à 1 et 2 ans des hypothèses. Il s'agissait de comparer une série longue d'hypothèses à leur réalisation et d'étudier l'existence d'une structure dans l'écart observé. Plusieurs éléments ont fait que cette étude a été retirée, le premier étant le nombre d'exercices d'hypothèses disponible. Le processus BP ayant été modifié, dans ses dates de réalisation notamment, les hypothèses d'atterrissage de l'exercice en

cours n'étaient plus comparables. En effet, avant 2016, ces hypothèses de production étaient fixées en février, elles sont depuis fixées en septembre. Cela signifie qu'en septembre 2018 on fixe l'atterrissage 2018 et la trajectoire 2019 – 2023. Avant ces chiffres étaient fixés en février, soit 7 mois plus tôt, ce qui change énormément la qualité de l'estimation à 1 an et donc à 2 ans, puisque la première inconnue est mieux « estimée ».

A noter qu'au-delà de 2 ans, une étude peut se faire en utilisant les probabilités de faire défaut, sachant qu'au temps $t-1$ il y a eu 1 à n défaut. L'idée était donc de montrer ce type de lien à 2 ans et de généraliser par la suite. Cette étude pourra être réalisée ultérieurement avec un alternant ou lors d'un prochain sujet de mémoire d'actuariat.

Pour revenir à ce qui a été réalisé dans les présents travaux, on notera qu'il a été décidé de conserver les hypothèses prises avec les experts sans les challenger. A partir des modèles linéaires construits sur le passé et de ces hypothèses, on peut donc projeter les portefeuilles à 1 an, tel que représenté sur la figure précédente. Pour la projection à plus long terme une autre approche a été retenue.

6.2.3.3 Projection au-delà de la première année

La dernière partie de cette étude consiste à projeter le portefeuille au-delà de la première année tout en mesurant l'incertitude induite par la trajectoire du portefeuille. Ici, l'idée fut de simuler un aléa autour du résultat du modèle fonction de la structure des résidus obtenus par le passé. On peut générer ainsi un nombre de résultats probables à 1 an. Pour chacun de ces résultats, on pourra simuler une nouvelle vague de portefeuille à $n+2$, et ainsi de suite.

Il reste à définir le nombre de simulations par nœud nécessaire à l'estimation d'un intervalle de confiance à 95%. Le nombre minimal de simulations pour répondre à cela est de 40, ce qui revient à retirer la simulation minimale et la simulation maximale.

Pour nous aider à fixer ce seuil, on pourra rappeler que l'on cherche à estimer une valeur moyenne, issue du modèle GLM, dont on connaît l'écart type. Si l'on pose n le nombre de simulations, à partir d'un niveau de confiance fixé α et d'une marge d'erreur e , la théorie de l'échantillonnage nous apprend que pour estimer une moyenne, la taille de l'échantillon est :

$$n = \frac{z^2 \sigma^2}{e^2}$$

Etant donné que la moyenne que l'on cherche à estimer est connue et définie par une régression linéaire généralisée, on dispose d'un écart type. Toutefois, la dimension multiple du modèle rend cette méthode peu exploitable.

Toujours en se basant sur les méthodes d'échantillonnage, lorsque l'on souhaite estimer une proportion P dans une population, on pourra utiliser l'équation suivante :

$$n = \frac{z^2 \cdot P \cdot (1 - P)}{e^2}$$

Cette équation est majorée pour $P = 50\%$. Dans le cas d'une précision à 95% (soit $z = 1,96$ loi Normale) et d'une marge d'erreur de 5%, n s'estime à 384 et à 196 si la marge d'erreur est de 7%.

Etant donné que pour les simulations il faut un multiple de 40, on proposera 200 simulations, soit une approche impliquant une marge d'erreur de 7%. Toutefois, on se heurte ici à des limites informatiques, en effet à 5 ans on parle de plus de 321 milliards de calculs.

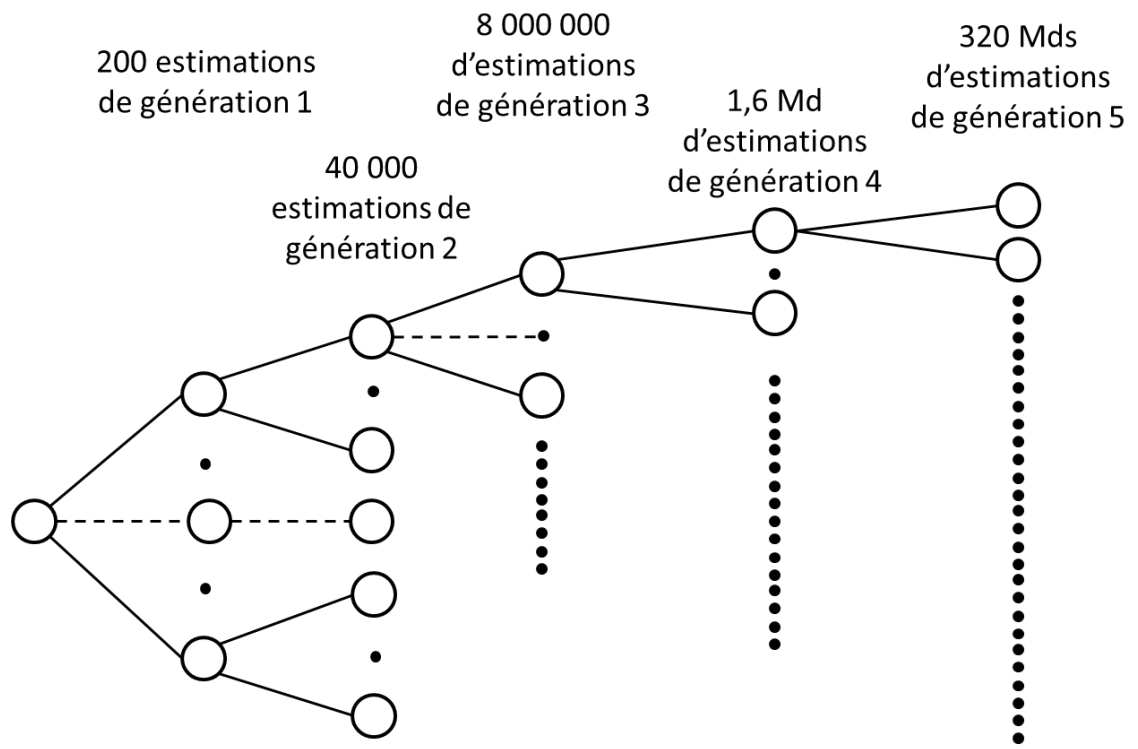


FIGURE 37 - PROPAGATION A 5 ANS - 200 SIMULATIONS PAR NŒUD

S'ajoute à la complexité informatique une notion de définition d'un intervalle de confiance. Dans cette approche où chaque nœud génère son propre futur aléatoire, chaque nœud permet d'obtenir des bornes d'un intervalle de confiance probable. En seconde génération nous disposons donc de 200 bornes inférieures et supérieures. On pourrait alors définir l'intervalle comme la moyenne de ces bornes, cependant cette moyenne aurait elle-même un intervalle de confiance. A 5 ans cette notion se complexifie encore plus.

Pour répondre à la problématique initiale, notamment la mise en œuvre rapide, différents tests ont été réalisés. L'idée est de diminuer le nombre de simulations tout en conservant un résultat qualitatif et donc d'obtenir un temps de calcul raisonnable. La solution retenue consiste à éteindre les branches en diminuant à chaque génération le nombre de fils calculé. Après plusieurs tests de performance, pour conserver un intervalle de confiance proche de celui issu du modèle il faut réaliser 1 000 simulations de première génération. Puis pour obtenir des temps de calculs satisfaisant, chaque génération successive simulera dix fois moins de fils. Soit 100 estimations par nœud de seconde génération et enfin 10 en troisième. A noter qu'au-delà il faudrait simuler 1 solution par branche. Cela n'a pas été testé car l'on proposera de maintenir l'IC estimé en 3^{ème} génération autour du scénario central, i.e. le résultat du modèle linéaire généralisé, pour les exercices 4 et 5.

On gardera donc en mémoire que cette approche de projection des portefeuilles, nouvelle pour Covéa, permet d'envisager un encadrement des résultats de la production, tout en conservant le processus en vigueur de définition des hypothèses avec les experts. De plus, cette méthode reste facilement adaptable à la projection des primes pures techniques et donc d'intégrer un aléa dans la trajectoire de tarification.

6.2.4 Liquidation stochastique de la charge des sinistres

Afin de clore les travaux réalisés, il faut à présent parler de la projection à l'ultime de la charge des sinistres. Les travaux réalisés reprennent les méthodes existantes sur le marché depuis plusieurs années en termes d'approches stochastiques. Parmi les méthodes testées, on retrouvera l'approche de Mack, l'approche de Braun, ses dérivées très intéressantes pour l'étude de l'impact des corrélations dans la liquidation et enfin les modèles linéaires généralisés, notamment la méthode de Poisson surdispersée. Enfin, une approche par bootstrap a aussi été traitée, elle sera détaillée dans un second temps.

6.2.4.1 Approches par modélisation

On retrouvera ici les méthodes de liquidation stochastiques dites modélisées, i.e. s'appuyant sur des équations mathématiques pour définir un intervalle de confiance.

La méthode des GLM a été étudiée, avec pour objectif l'implémentation de copules visant à réguler les corrélations entre les triangles selon les segments de marché et les marques. Etant donné les restrictions imposées par les objectifs, notamment le non travail des triangles, et la fonction de lien logarithmique imposée par ce modèle, aucun résultat n'a pu être exploité.

Concernant les approches de la famille de Braun, la forte sensibilité de certaines méthodes aux valeurs extrêmes pose problème dans l'exploitation des résultats obtenus. En effet, la figure 66 (cf. annexe page 83) montre que des exercices sur lesquels nous disposons de plus de recul sont parfois plus volatiles que des exercices plus récents. Ces méthodes sont donc difficiles à défendre et mériteraient certainement un nettoyage plus fin des données. Il est toutefois plus qu'intéressant de noter que les résultats obtenus lors de l'agrégat des garanties sont moins volatiles lorsque l'on prend en considération les corrélations (cf. figure 64 et 65 en annexe pages 81-82).

Enfin les méthodes de Mack appliquées via Chain-Ladder et Bornhuetter-Ferguson, présentent des résultats cohérents. Toutefois, si l'on prend le cas de la société A, sur le segment de marché Automobile, on obtient les résultats suivants pour les exercices 2012 à 2017, estimés à fin 2017.

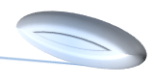
	2012	2013	2014	2015	2016	2017
Charge Ultime estimée	829 M€	888 M€	965 M€	960 M€	989 M€	913 M€
Étendue de l'IC	7,8%	10,1%	12,3%	13,9%	17,6%	38,0%

On notera ici la taille de l'intervalle de confiance à 95%, qui parfois dépasse le tiers du résultat. Il est possible de questionner l'indice de confiance souhaité, le seuil de 95% pourrait être réduit, pour estimer un intervalle à 90% ou 80%.

Maintenant, l'un des objectifs était de pouvoir estimer le ratio combiné futur avec une probabilité d'obtenir des résultats supérieurs à un seuil. Il semble intéressant de limiter au plus la marge d'erreur présente dans chaque étape composant ce résultat. On préférera explorer d'autres approches de liquidation.

6.2.4.2 Approche calculatoire : le bootstrap

Dans cette partie on étudie la liquidation par *bootstrap*. Cette méthode est autosuffisante, au sens où elle utilise l'aléa observé dans chaque triangle pour constituer un très grand nombre de visions à l'ultime possible. A partir des simulations il devient possible de définir un scénario moyen et de l'encadrer. On notera toutefois que le *bootstrap* a l'inconvénient de ne pas intégrer les corrélations existantes entre les triangles. Toutefois, les résultats obtenus donnent des intervalles de confiance empiriques plus resserrés.



Des tests ont également montré que plus le nombre de liquidations aléatoires par *bootstrap* était important et plus l'intervalle de confiance se réduisait. Toutefois, avec des temps de traitement de plusieurs dizaines d'heures, contre quelques minutes, on préconisera de simuler 10 000 triangles à partir de l'original et non 1 million de triangle, le gain en précision ayant été jugé trop faible.

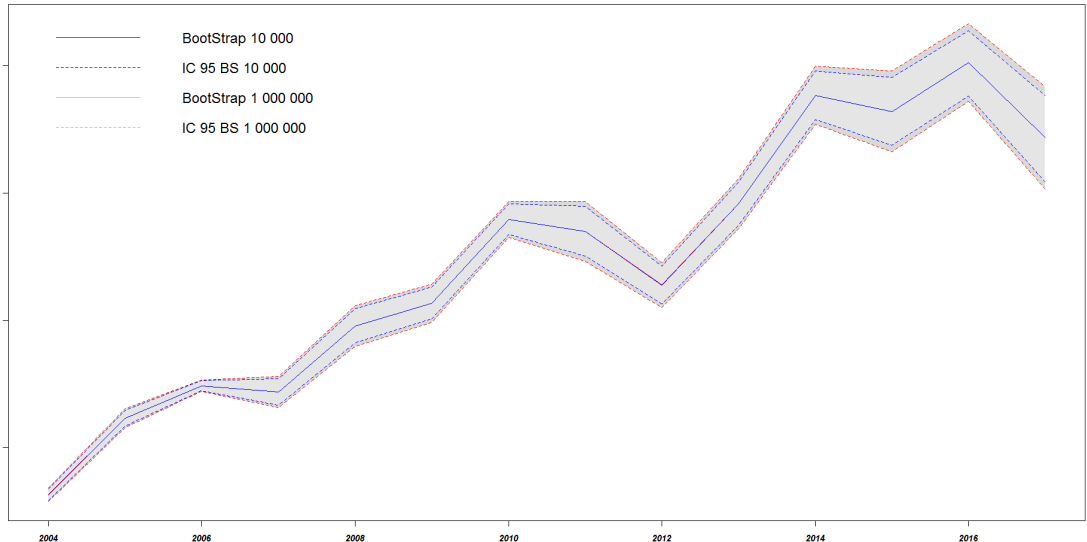


FIGURE 38 - IMPACT DU NOMBRE DE SIMULATIONS SUR L'INTERVALLE DE CONFIANCE POUR LE BOOTSTRAP

De plus, il faut garder en mémoire ce que nous ont enseigné les modèles dérivés de Braun, à savoir que la prise en compte des corrélations lors de l'agrégat des garanties permet de réduire la volatilité totale, c'est ce qu'illustre bien la figure suivante.

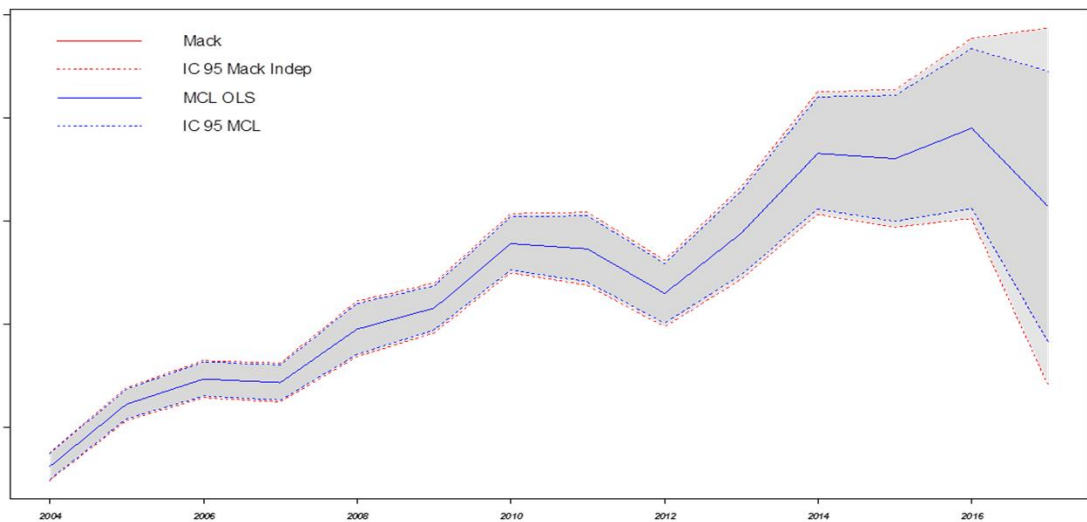


FIGURE 39 - IMPACT DES CORRELATIONS SUR L'INTERVALLE DE CONFIANCE - CUMUL DES GARANTIES DE LA SOCIETE A EN AUTOMOBILE

Ainsi, les approches de liquidation par *bootstrap*, sous hypothèse d'indépendance pour l'agrégat des garanties, seraient des majorants des intervalles de confiance réels. Suite à ces travaux, il est donc préconisé d'utiliser des méthodes de *bootstrap*, sous réserve de validation sur l'ensemble des marques et des segments de marché de Covéa.

6.2.4.3 Comparaison et regroupement des triangles par ressemblance

Enfin, on pourra reparler ici des tests réalisés sur la comparaison des triangles, par approche de segmentation. Après avoir calculé une distance entre les triangles bruts ou directement sur les cadences individuelles de développement, les populations ont été regroupées. Pour cela l'analyse a été réalisée en deux temps.

Tout d'abord, on effectue une analyse des composantes principales (ACP), en prenant comme individus les triangles et comme variables le croisement des exercices de survenance et d'observation. On obtient plusieurs axes orthogonaux entre eux, donc indépendants, résumant la volatilité expliquée par les exercices de survenance et d'observation. Par simplification, les données sont centrées et réduites avant de calculer l'ACP. Les résultats ont été illustrés par la figure 13 (page 36). A ce stade, pour chaque triangle on dispose de nouvelles variables indépendantes, qui sont projetables sur des axes orthonormés.

Il devient alors possible de positionner les individus – nos triangles par marque et segment de marché – sur ces axes.

Dans un second temps, une méthode de classification ascendante hiérarchique (CAH), permet de regrouper les individus les plus proches. Le choix de la classification et surtout de la distance est simplifié par l'orthonormalité des axes qui pousse à retenir la distance euclidienne et donc d'identifier les individus les plus proches.

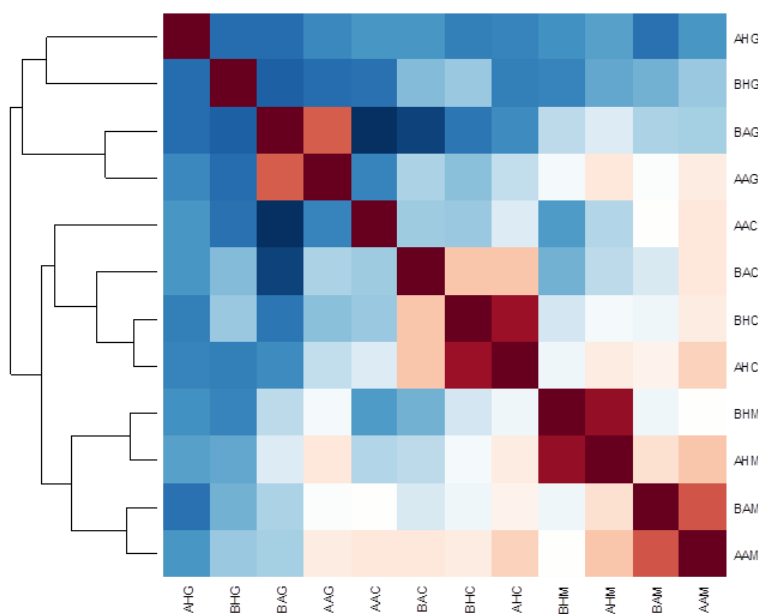
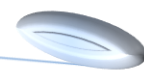


FIGURE 40 - CLASSIFICATION DES TRIANGLES DE COEFFICIENTS DE DEVELOPPEMENT

Rappel de lecture de la figure 40, la première lettre correspond à la société (A ou B), la seconde au segment de marché (A – Auto ou H – Habitation) et la dernière correspond à la garantie (C – Climatique, G – Graves et M – Masse).

De ces travaux il a été possible de mettre en évidence que la garantie a une très forte incidence sur la distance entre les triangles, puis en second on retrouve généralement le segment de marché. Dans la figure précédente, les triangles de sinistres climatiques en Habitation se regroupent en premier. Viennent ensuite les triangles des garanties de Masse, en Habitation puis en Auto. Enfin on retrouvera un regroupement des triangles graves en Auto. Pour la suite, le climatique en Auto va progressivement rejoindre les sinistres climatiques. En parallèle les triangles de sinistres de masse se regroupent...



Cette étude permet d'identifier les triangles les plus proches, il serait intéressant de les regrouper pour les étudier comme des blocs, cela pourrait réduire la volatilité globale dans la projection à l'ultime.

A ce jour, aucune suite n'a été donnée à ce constat. En effet, modifier la façon de projeter à l'ultime et de regrouper les triangles impose de refondre plus profondément la modèle actuel. De plus, cela pose des questions sur la répartition a posteriori. Si l'on définit une charge ultime climatique Covéa, comment est-il possible de réaffecter cette charge entre les sociétés du groupe sans trop mutualiser la charge et en conservant une logique tarifaire ?

6.3 Et pour conclure

Au regard de tous ces éléments, il me semble important de préciser que ces travaux exploratoires ont permis d'identifier les approches à privilégier tant pour choisir une méthode de liquidation stochastique que pour définir une trajectoire de portefeuille à 5 ans. Maintenant, il est prévu que l'équipe s'approprie les travaux et généralise les tests sur toutes les marques du groupe Covéa et que ces tests soit étendus aux autres segments de marché, à une maille plus fine, à savoir les garanties dans leur ensemble, non plus uniquement les blocs utilisés pour ces travaux.

L'objectif est simple, il s'agit de valider la qualité des modèles afin de pouvoir décider d'une mise en place pérenne, d'une approche plus moderne de la science des sinistres et de leur liquidation, tout comme dans la projection des portefeuilles. Ce serait également un premier pas vers un modèle interne Covéa.

Enfin, il me paraît important de rappeler que ces travaux devront servir de référence au quotidien, pour rester garant d'une activité alliant exigence opérationnelle et qualité des résultats, tout en livrant un encadrement attendu par les décideurs.

Je préciserai encore que ces travaux ont été réalisés avec des contraintes opérationnelles, à savoir qu'il doit être simple et rapide de mettre en place et d'exploiter les méthodes utilisées. Cela a donc parfois limité certaines options, complexes dans leur mise en place ou leur compréhension.

Par ailleurs, je préconise une veille active sur les méthodes en cours de documentation et également de prendre le temps de les adapter aux données, en travaillant avec des stagiaires ou des alternants. En effet, les théories étudiées par Marc Linde qui mettent à profit les approches généralisées de Braun et le *bootstrap* pourraient être une solution très intéressante, surtout s'il est possible d'y intégrer les recherches en cours sur la robustesse des modèles de Braun.

Le sujet est encore très riche et demande à être toujours plus affiné. [Ces recherches peuvent d'ailleurs servir de base pour des travaux sur un modèle interne, ou être complétées sur la projection à 5 ans de la sinistralité afin de présenter une probabilité d'avoir un ratio combiné supérieur à un seuil, seuil à fixer par les décideurs.](#)

Annexes

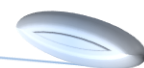
A - Données

Triangle de liquidation de la société B, garantie climatique habitation, vision cumulée et incrémentale.

Clim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
2000	371730	544116	638394	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086	661086
2001	897492	987798	1070774	1071617	1071617	1076323	1079358	1079358	1079358	1079358	1079358	1079358	1079358	1079358	1079358	1079358	1079358	1079358	1079358
2002	1186324	1857195	2213341	2376471	2491302	2588948	2554871	2635342	2754347	2816202	2837575	2835132	2833881	2839160	2839160	2839160	2839160	2839160	2838616
2003	1797950	3489218	4689576	5584191	6257146	6874848	7134404	7268379	7248339	7131397	7064454	7061252	6970177	6970178	6909022				
2004	749857	796450	798229	850113	907193	907193	920800	920800	920442	920442	920442	920442	920442	920442	920442				
2005	957059	1069372	1091152	2183908	2556426	2810016	2808390	2914696	3000517	2980552	2977513	2978653	2970183						
2006	1328497	1516808	1790917	2036523	2065070	2319785	2400151	2399229	2300561	2337146	2353562	2332200							
2007	1678817	1859255	2267389	2512818	2484965	2520436	2551550	2484586	2482269	2511968	2514128								
2008	2595981	3815273	3893210	3964798	4043798	4059904	4131593	4133820	4133820	4133694									
2009	8745584	9460526	9917035	10258438	10347960	10431520	10394326	10466369	10476308										
2010	8179783	8742822	8923794	9039474	9092594	9068233	9542720	9468831											
2011	2786777	4780739	6255852	6693222	6827102	7335267	7242750												
2012	4424568	5077903	5693690	5809396	5911547	5852900													
2013	6031627	7028837	7182368	7212010	7241798														
2014	8847552	9201628	9494765	9231131															
2015	5559275	6350361	6008806																
2016	12613984	12224044																	
2017	7258771																		

Clim	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
2000	371730	172387	94277	22693	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2001	897492	90307	82975	843	0	4707	3035	0	0	0	0	0	0	0	0	0	0	0	0
2002	1186324	670871	356146	163131	114830	97646	-34077	80471	119005	61855	21373	-2443	-1251	5279	0	-544			
2003	1797950	1691268	1200358	894615	672955	617702	259556	133976	-20040	-116942	-66943	-3202	-91074	1	-61156				
2004	749857	46593	1779	51884	57080	0	13607	0	-358	0	0	0	0	0	0				
2005	957059	112314	21779	1092756	372518	253590	-1626	106306	85821	-19965	-3039	1139	-8470						
2006	1328497	188311	274109	245606	28548	254715	80366	-922	-98668	36585	16416	-21362							
2007	1678817	180438	408134	245429	-27853	35471	31114	-66965	-2316	29698	2161								
2008	2595981	1219291	77957	71588	79000	16106	71689	2227	0	-126									
2009	8745584	714942	456509	341403	89522	83560	-37194	72043	9939										
2010	8179783	563039	180972	115680	53120	-24361	474487	-74389											
2011	2786777	1993962	1475113	437370	133880	508166	-92518												
2012	4424568	653335	615788	115705	102152	-58648													
2013	6031627	997209	153531	29642	29788														
2014	8847552	354077	293136	-263634															
2015	5559275	791086	-341555																
2016	12613984	-389939																	
2017	7258771																		

FIGURE 41- TRIANGLE DE CHARGE DES SINISTRES CLIMATIQUE SOCIETE B EN HABITATION



B – Tests Chain-Ladder

B.1 – Linéarité des coefficients

Tests réalisés pour chaque triangle soit 12 figures, société (2) x segment marché (2) x garantie (3).

Pour rappel on compare graphiquement et à l'aide du R^2 , le coefficient de détermination de la régression linéaire, on le souhaite le plus proche de 1. La pente est également donnée, le a , de l'équation $y = ax$. Pour que le chain-ladder s'applique, la relation linéaire doit être absolue, au sens où la droite doit passer par l'origine.

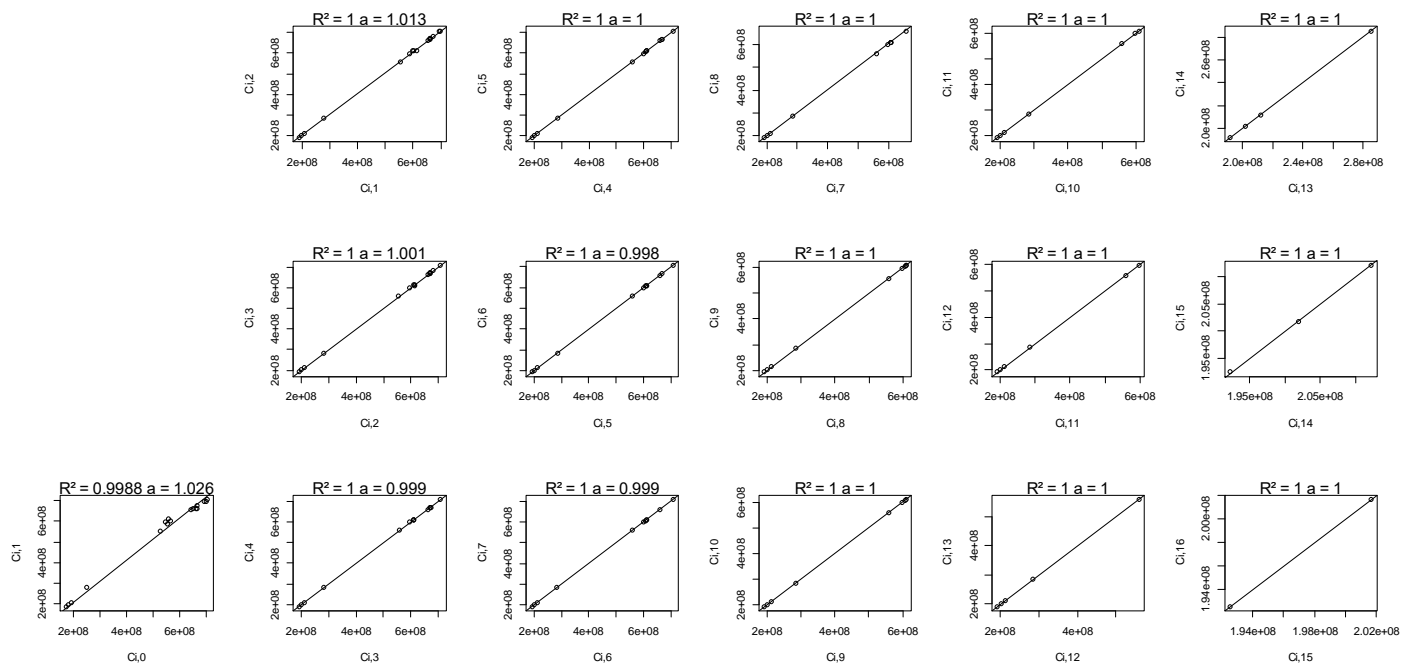


FIGURE 42 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE AUTO DE LA SOCIÉTÉ A

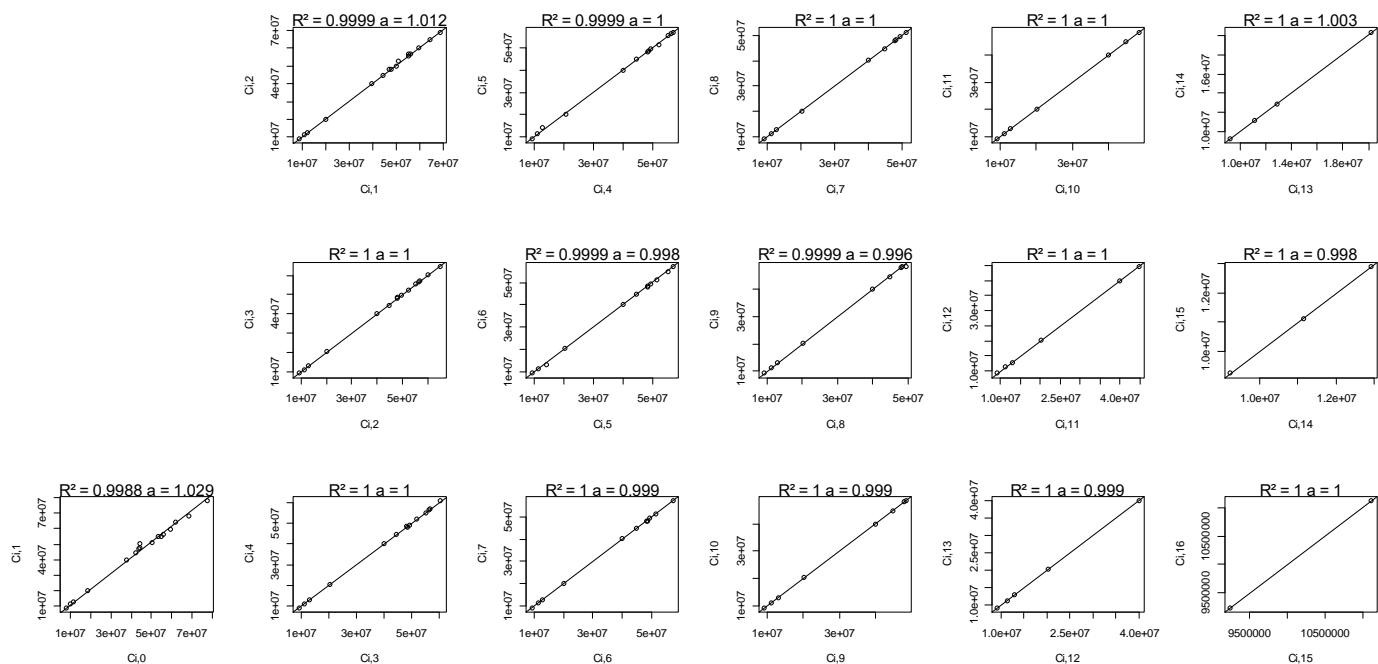


FIGURE 43 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE AUTO DE LA SOCIETE B

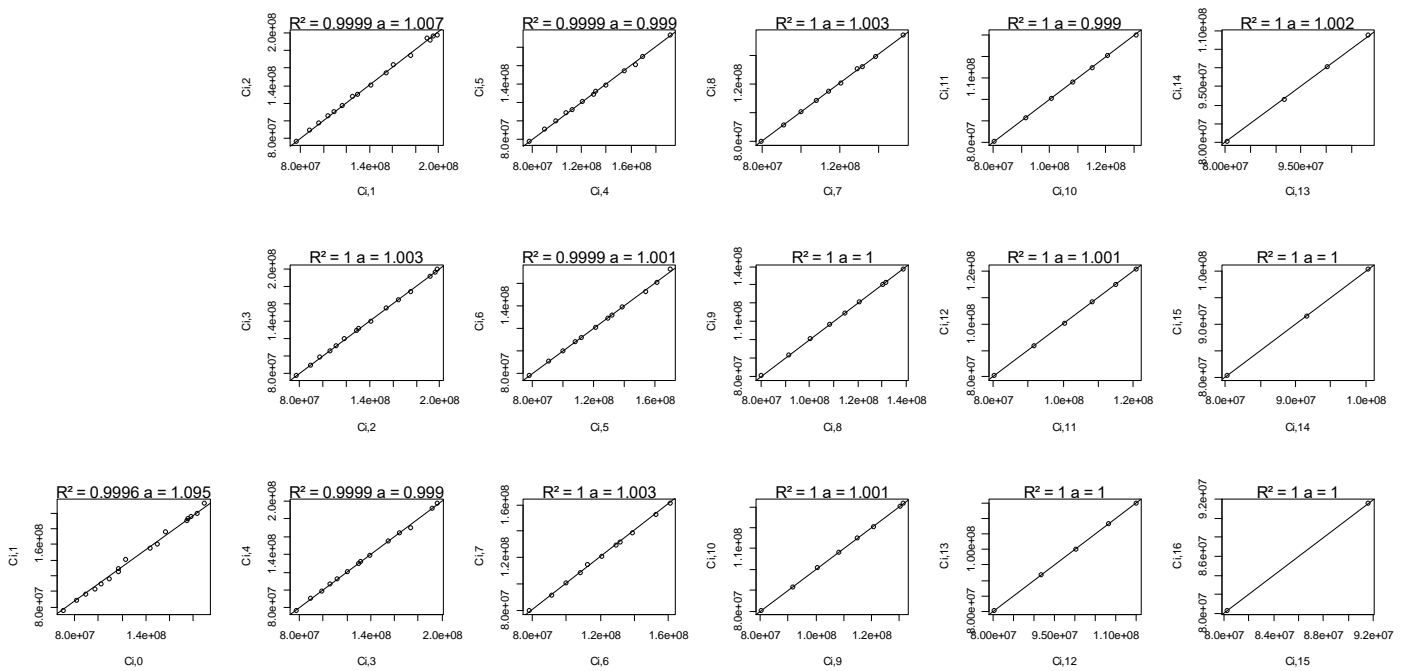


FIGURE 44 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE HABITATION DE LA SOCIETE A

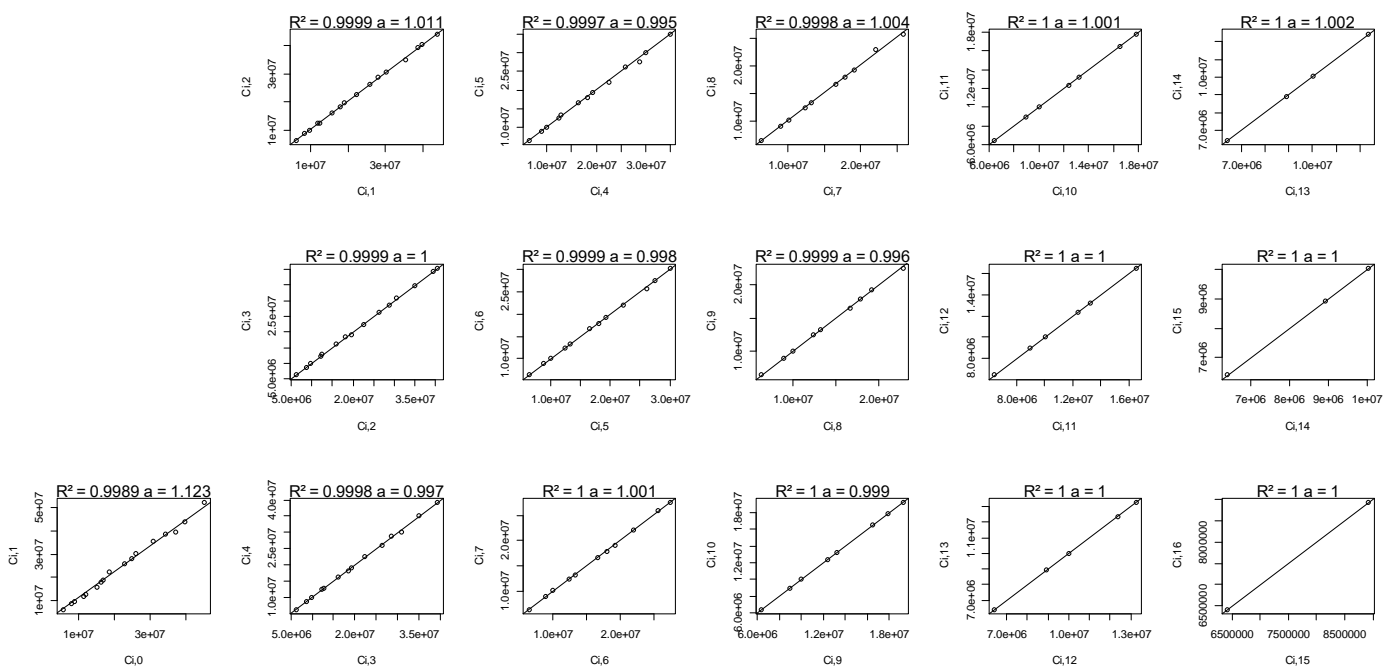


FIGURE 45 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE HABITATION DE LA SOCIETE B

On notera que sur les garanties de masse, le chain-ladder peut s'appliquer, aussi bien en auto, qu'en habitation, sur la société A et la société B.

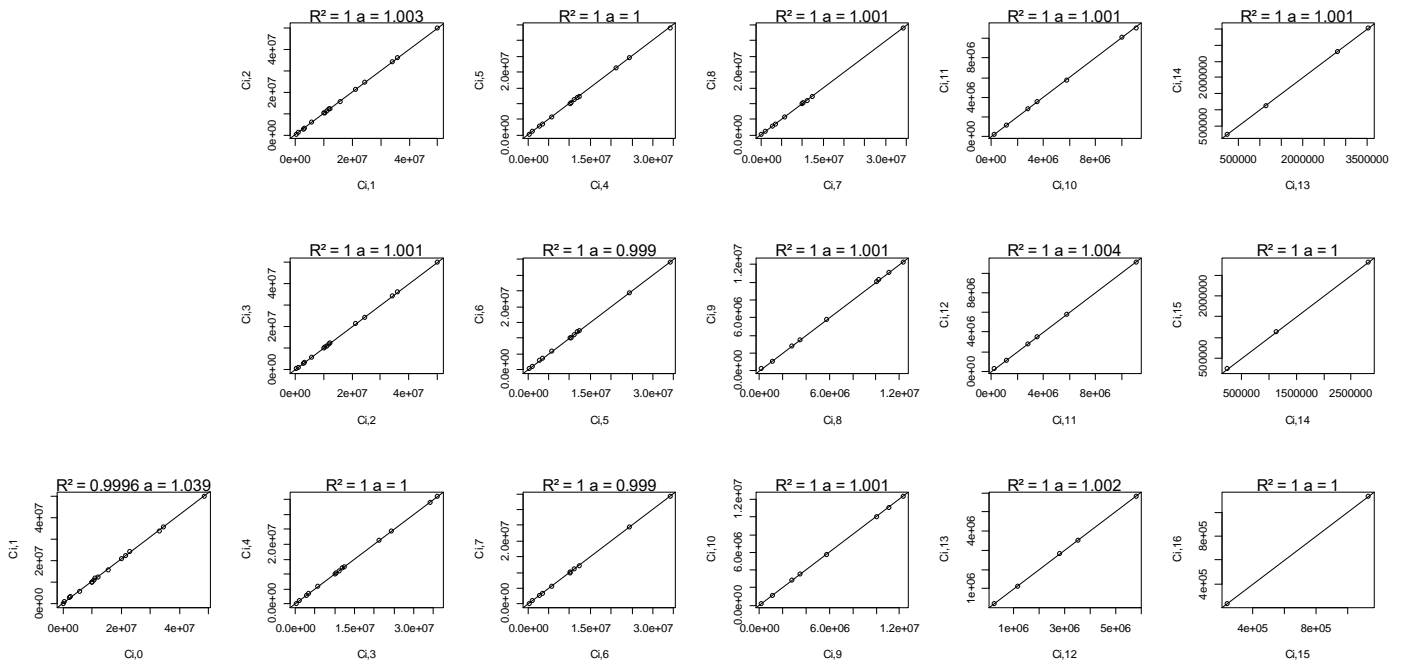
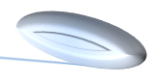


FIGURE 46 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE AUTO DE LA SOCIETE A

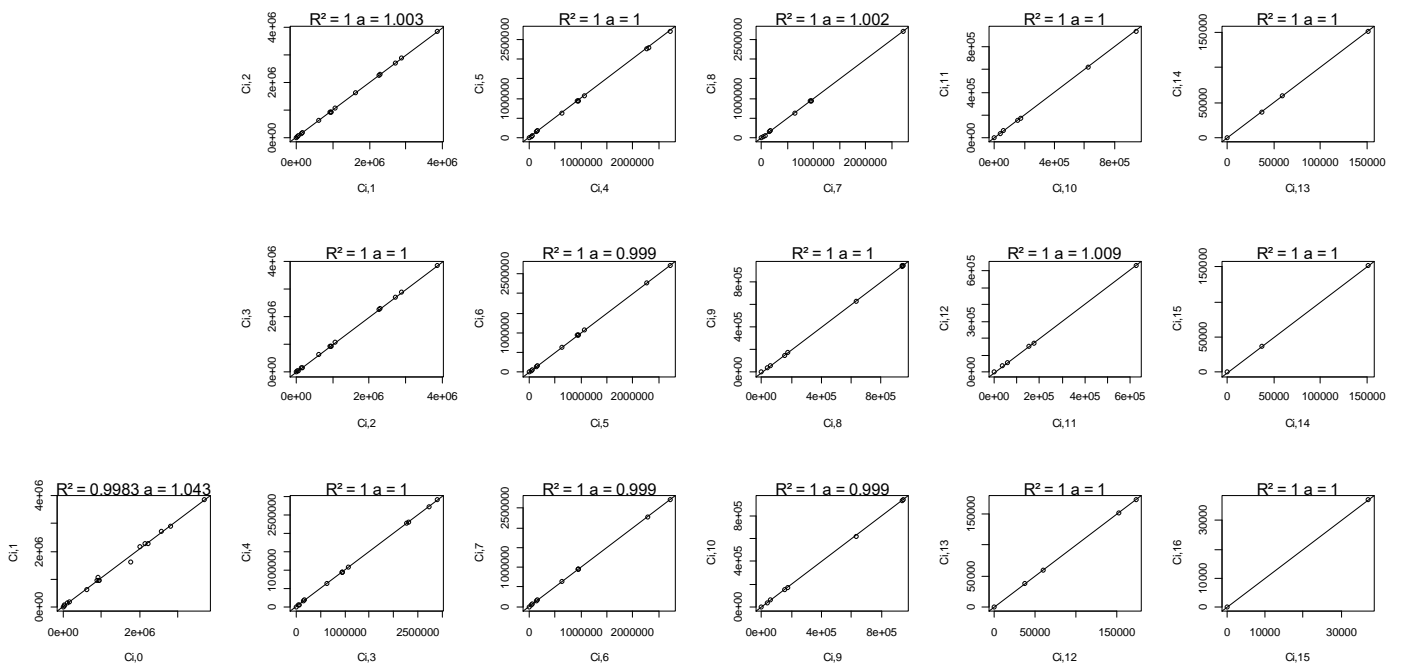


FIGURE 47 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE AUTO DE LA SOCIETE B

Tout comme pour les garanties de masse, les garanties climatiques en auto respectent la condition du chain-ladder.

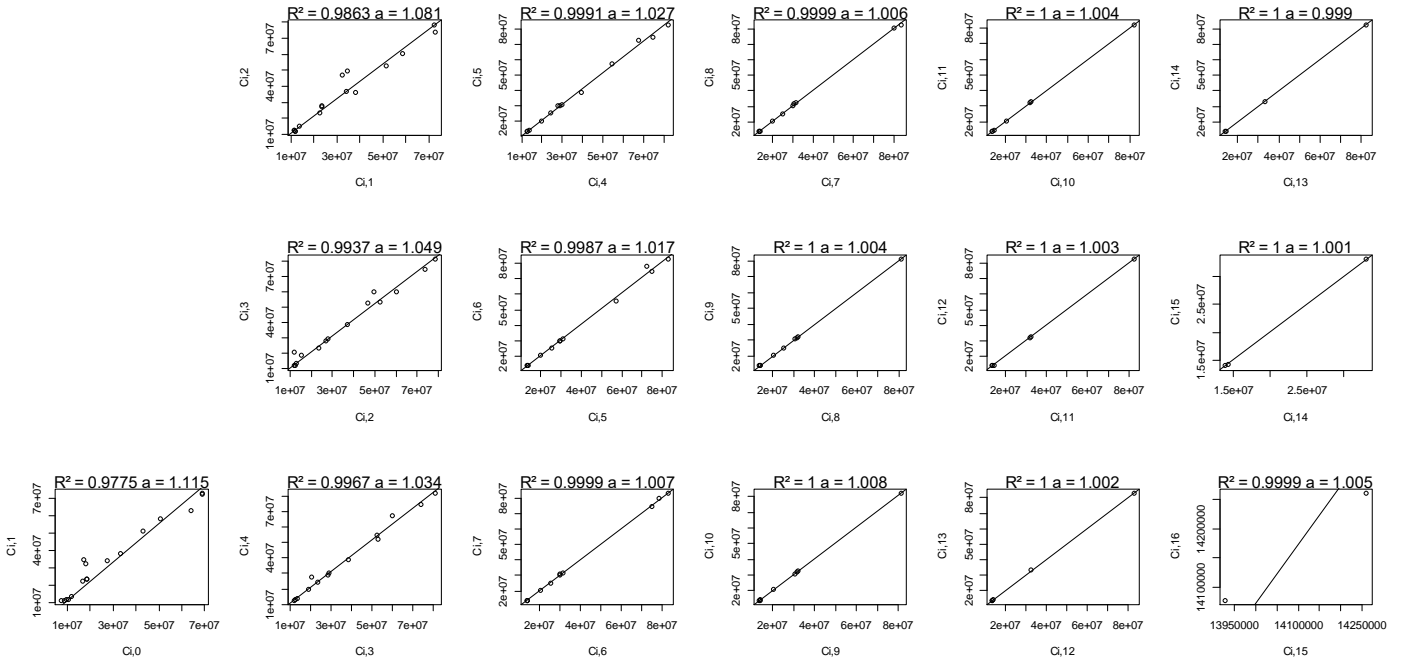


FIGURE 48 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE HABITATION DE LA SOCIETE A

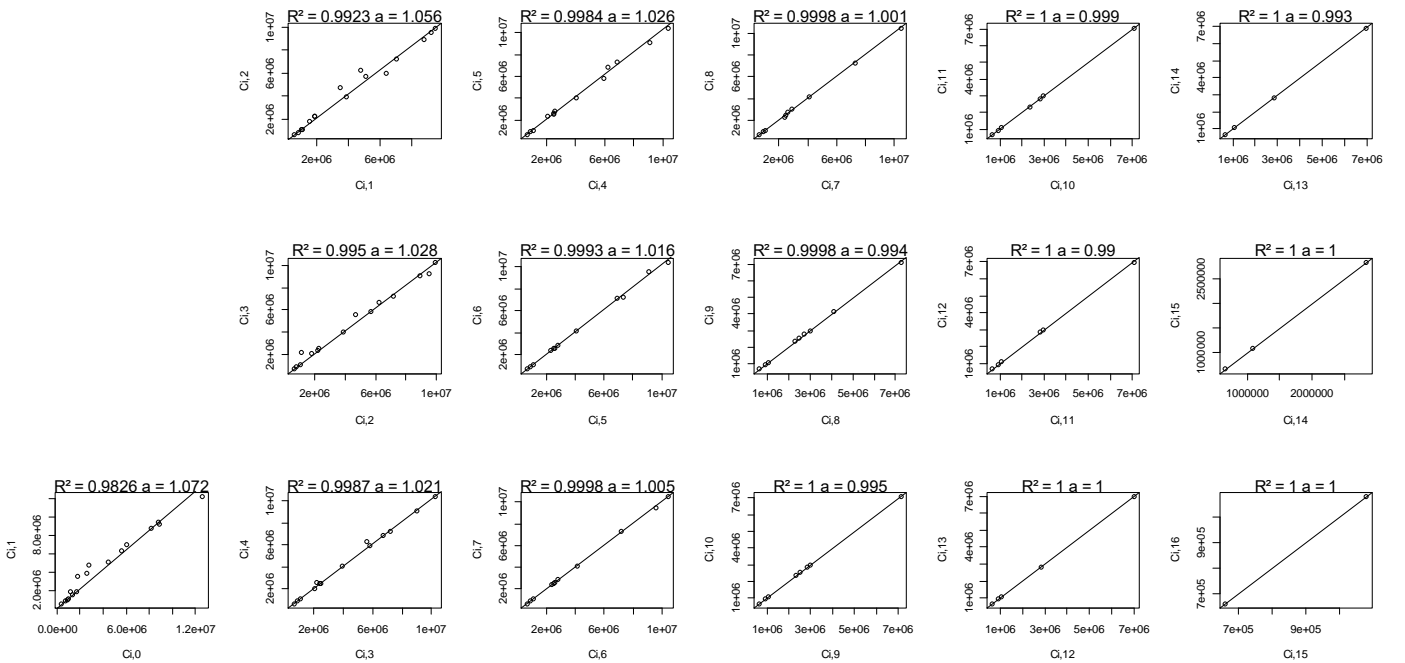


FIGURE 49 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE HABITATION DE LA SOCIETE B

Pour les garanties climatiques en habitation, la qualité de la linéarisation baisse, on le constate au travers du R^2 qui est plus faible les 2 premières années. Graphiquement, l'on distingue des points hors de la droite.

Il sera important de s'assurer que la linéarité absolue est un bon modèle, si le modèle linéaire standard, i.e. avec une ordonnée à l'origine, alors un décalage du triangle pourra être étudié, comme expliqué dans le mémoire.

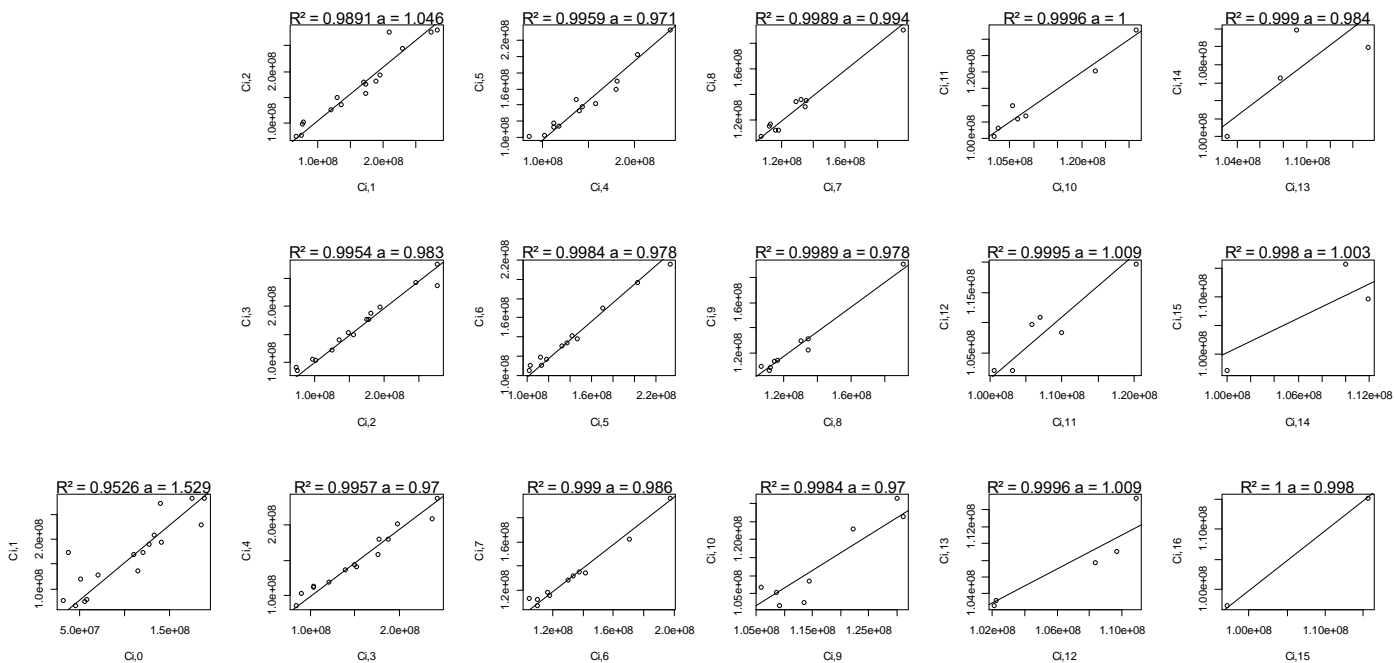
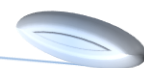


FIGURE 50 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES AUTO DE LA SOCIETE A

Si la qualité de la régression était moindre pour les garanties climatiques en habitation, elle l'est bien moins encore pour les sinistres graves. On notera également que même après plusieurs années, la pente est encore instable. De plus, les points aberrants sont assez nombre en première année.

Avec un R^2 de l'ordre de 0.95 en première année, le modèle linéaire absolu reste plus qu'acceptable. On se reportera au texte principal pour prendre connaissance des tests réalisés sur la linéarité classique, demandant un ajustement des coefficients de développement afin de calculer l'ultime.

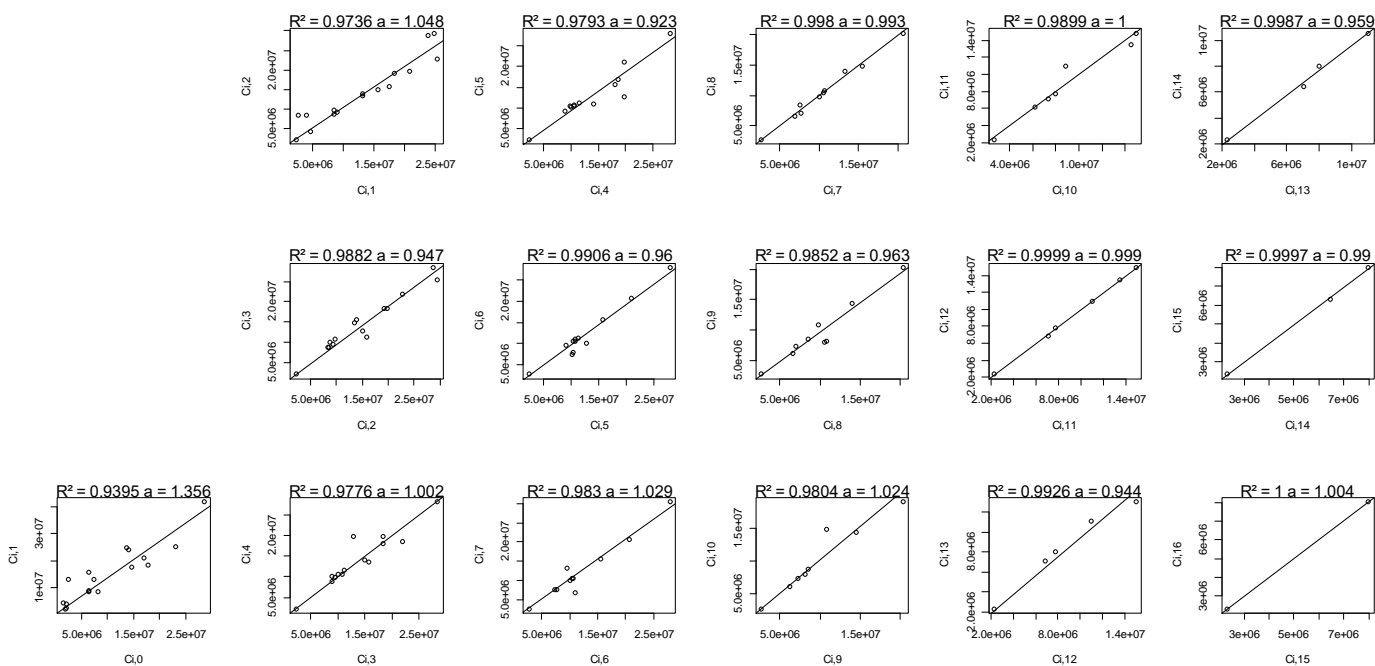


FIGURE 51 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES AUTO DE LA SOCIETE B

Le constat réalisé sur la société A est identique sur la société B, avec un coefficient de développement de l'ordre de 0.94. On notera que les approches de type Bootstrap intégreront les écarts visible sur ces graphes, en taux.

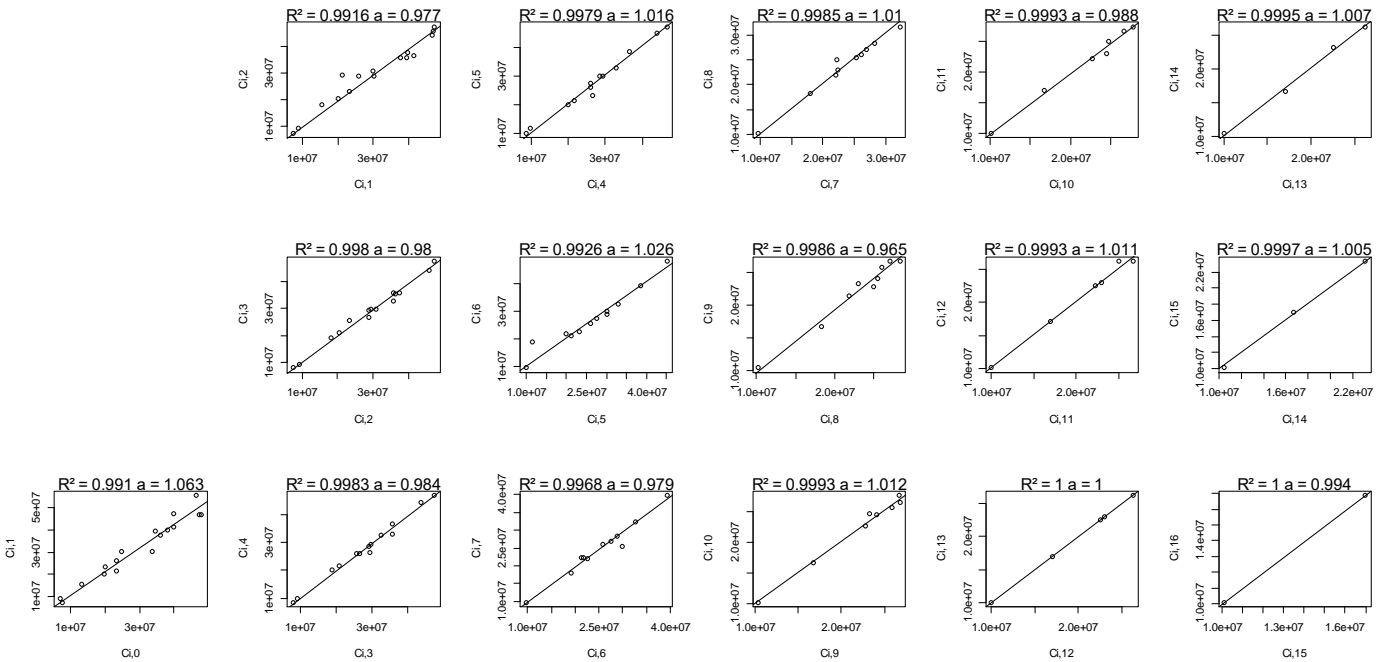


FIGURE 52 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES HABITATION DE LA SOCIETE A

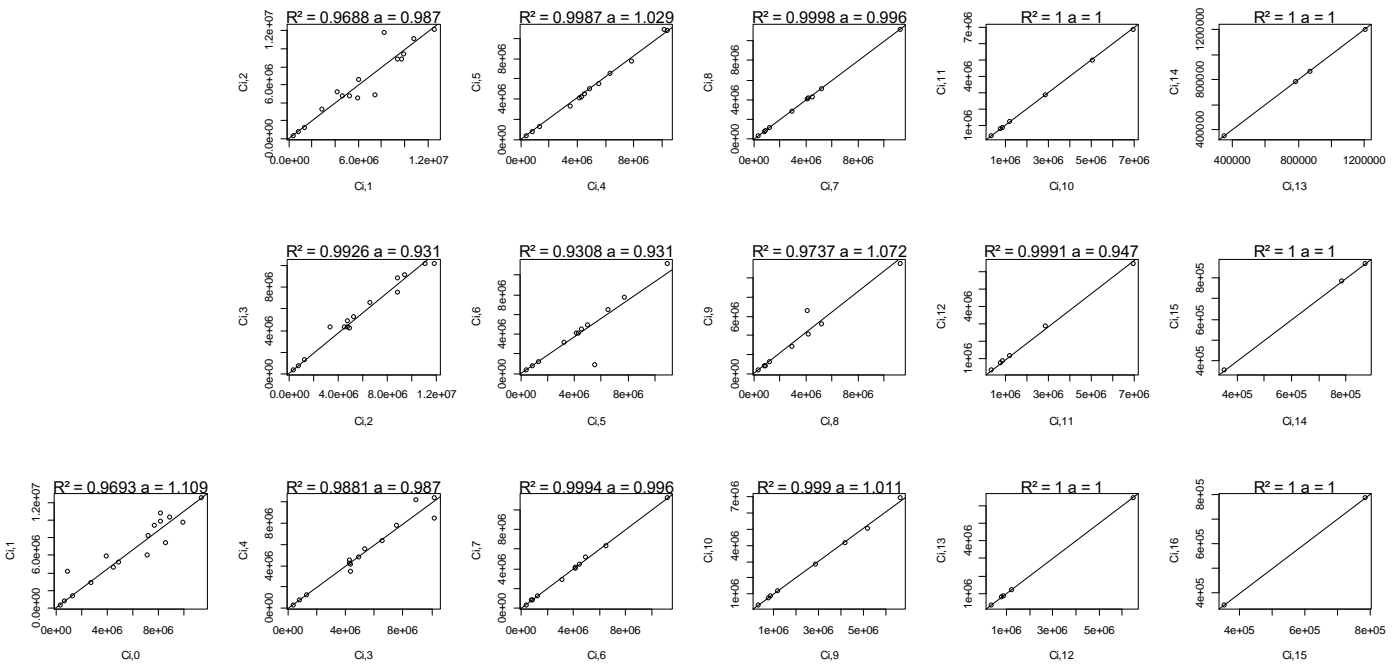
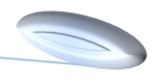


FIGURE 53 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES HABITATION DE LA SOCIETE B

Si les sinistres graves en auto laissent imaginer une liquidation moins linéaire entre deux instants d'observation, cela semble différent pour l'habitation. On notera toutefois que de nombreux points restent hors de la droite.



B.1 – Indépendance par test de corrélation

Les résultats des tests à suivre montrent tous que, les triangles comportent des dépendances entre les exercices, du fait des corrélations qui existent entre chaque série, testée deux à deux. Tout comme pour la linéarité, il y a 12 illustrations, la taille aura permis de les regrouper par marque et segment de marché.

Auto Société A - Masse

Auto Société A - Climatique

Auto Société A - Graves

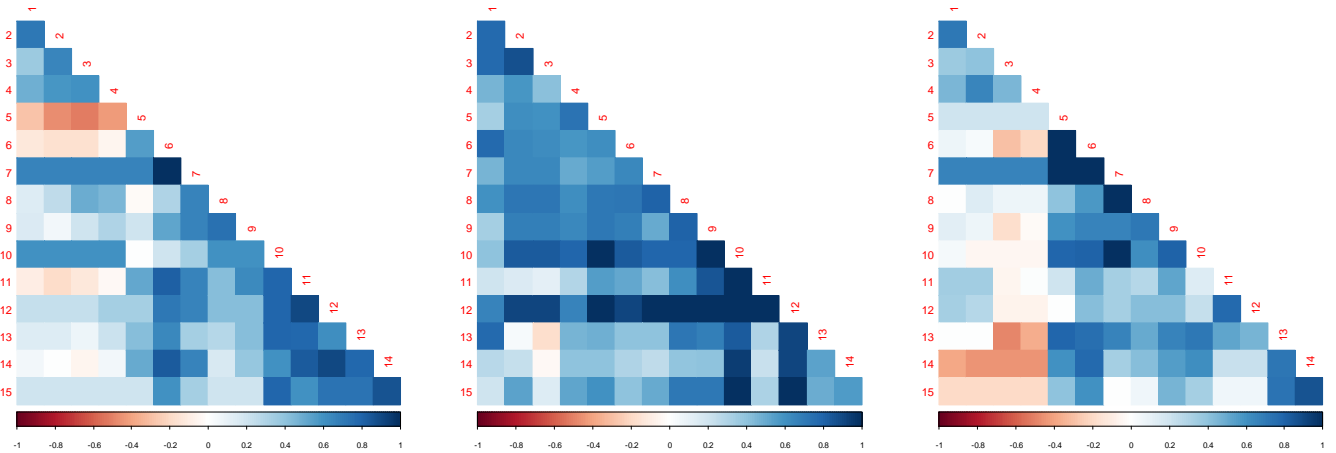


FIGURE 54 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE A, SEGMENT AUTO

Auto Société B - Masse

Auto Société B - Climatique

Auto Société B - Graves

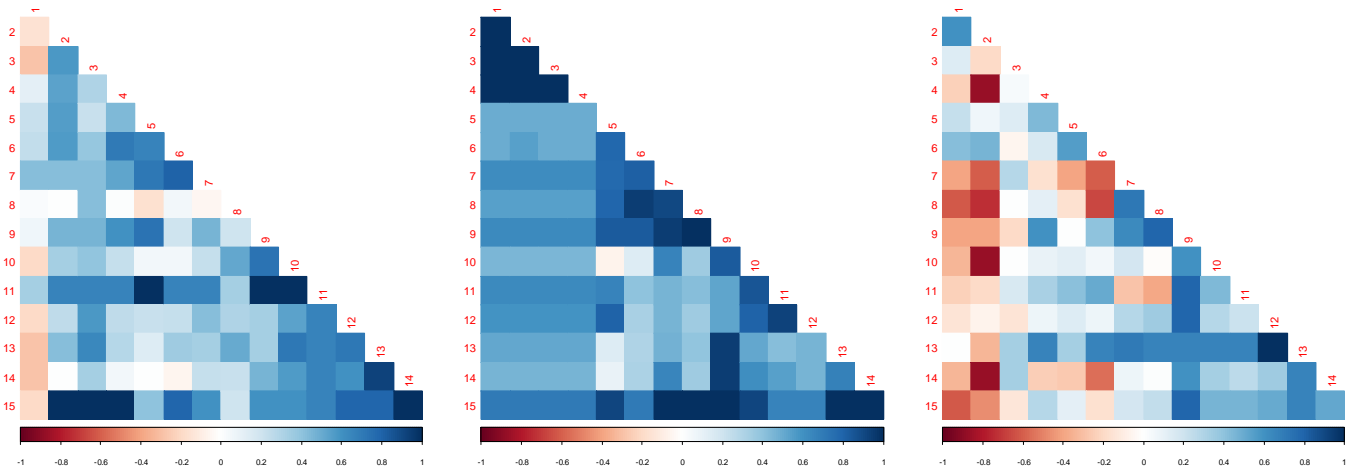


FIGURE 55 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE B, SEGMENT AUTO

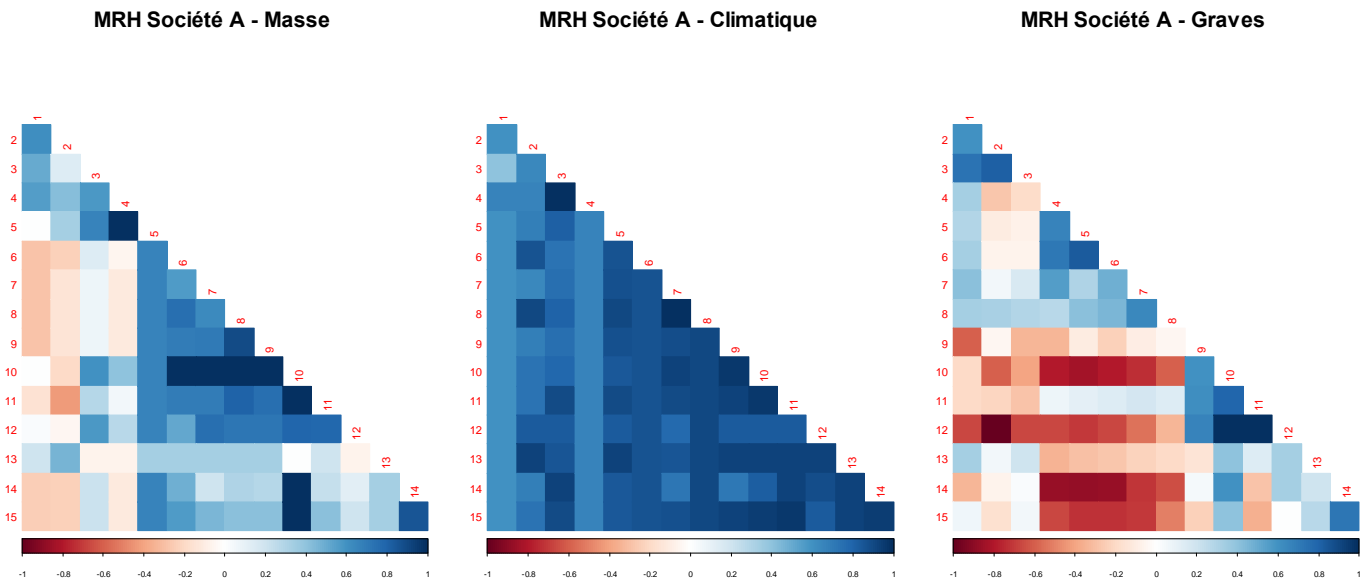


FIGURE 56 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE A, SEGMENT HABITATION

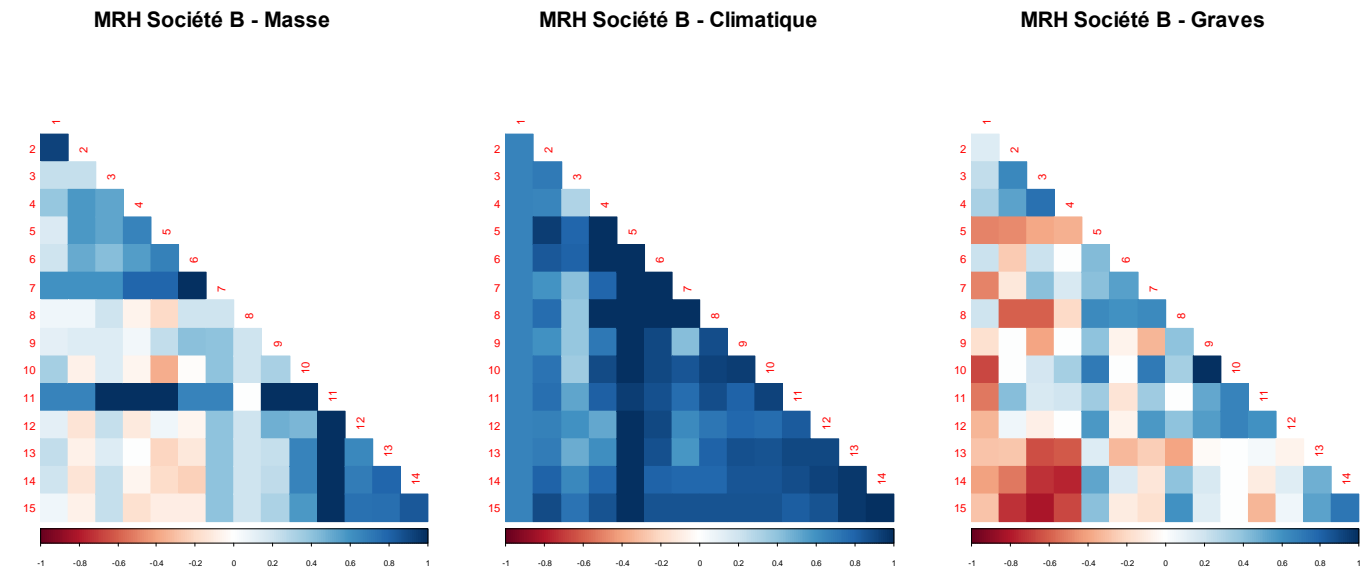


FIGURE 57 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE B, SEGMENT HABITATION

Si certaines corrélations sont nulles, le constat est plutôt à une majorité de coefficients différents de 0. Il devient donc impossible de se prononcer sur les conditions d'application du Chain-Ladder.

A noter qu'une étude de ces triangles, après retrait de la contribution du portefeuille, en prime pure donc, pourrait limiter les corrélations.

C – Tests Bornhuetter Ferguson

Il a été démontré dans le texte principal que la méthode de Bornhuetter-Ferguson était limitée aux exercices pour lesquels le développement atteint serait inférieur à 1. Dans le cas présent, à savoir des triangles de charge totale, ce point est discutable, toutefois la charge ultime a priori étant définie grâce à une méthodologie robuste visant à calculer une charge ultime technique, ignorant la conjoncture, il a été préféré conserver ces limites d'application.

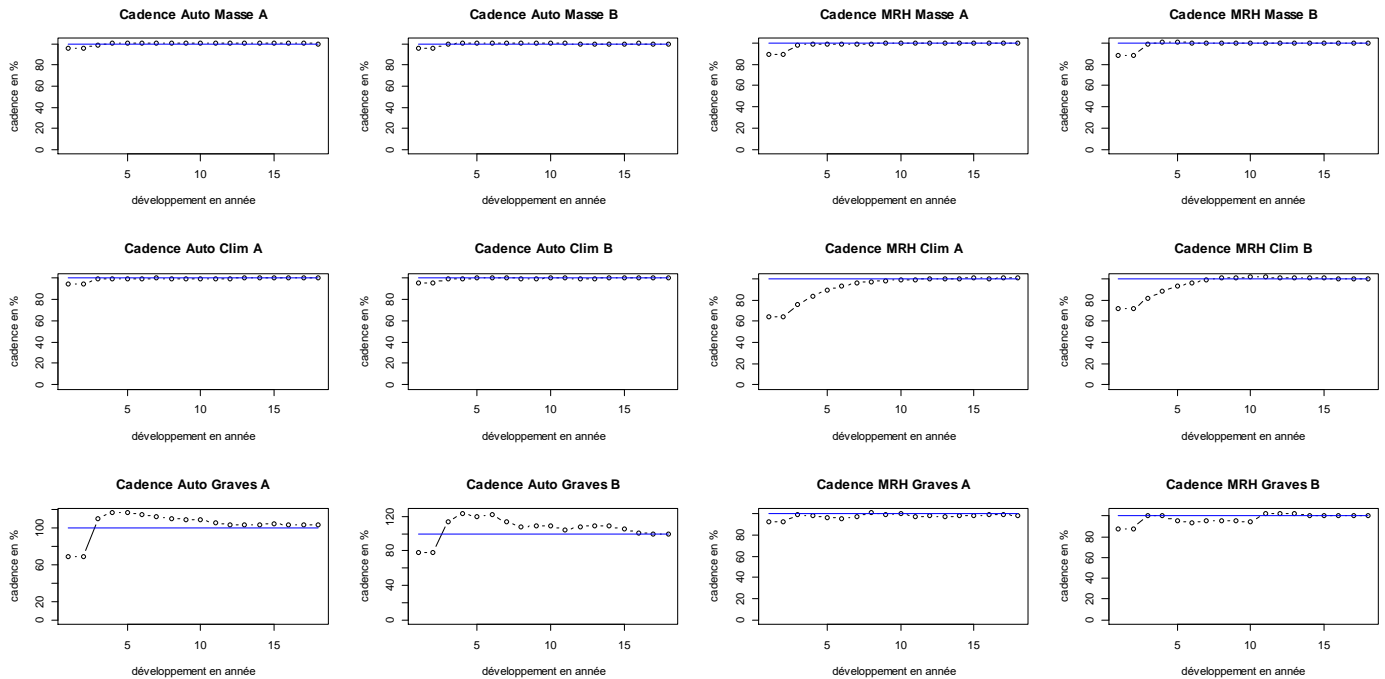
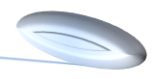


FIGURE 58 - CADENCE DE DEVELOPPEMENT

Dans la figure précédente, les cadences de chaque triangle ont été représentées. Pour rappel, on définit la cadence comme l'inverse du coefficient de développement. Ainsi l'on obtient une courbe, si cette dernière dépasse 100%, alors la méthode de BF devient inapplicable, pour les points au-dessus de 100%.

On notera que les sinistres graves, en Auto, sont au-dessus du seuil à compter du 3^{ème} point, soit 2 ans après l'exercice de survenance. A l'inverse, les charges climatiques en MRH restent très en dessous de la courbe pendant au moins 5 ans. Il devient intéressant, dans un cas comme dans l'autre, d'intégrer cette méthode dans les calculs des ultimes, en prenant en considération ces limites.

D – Tests de Mack

Les tests réalisés pour savoir s'il est possible d'appliquer la méthode de Mack, sur du Chain-Ladder, comme sur du Bornhuetter Ferguson, sont les mêmes que les tests de ces méthodes source.

Il a donc été montré dans les annexes précédentes l'intégralité des points.

E – Covariance / Variance : calcul

L'on cherche à démontrer ici que les variances de deux variables aléatoires ne peuvent être sommées sans prendre en considération un correctif, contrairement aux espérances. NB : il s'agit de démontrer la non nullité de la covariance de deux variables aléatoires.

Soit X et Y deux variable aléatoire, on démontre que :

- $E(X + Y) = E(X) + E(Y)$
Par propriété de l'espérance.
- $V(X + Y) = E[(X + Y)^2] - E^2(X + Y) = E(X^2 + Y^2 + 2XY) - [E(X) + E(Y)]^2$
 $V(X + Y) = E(X^2) + E(Y^2) + 2E(XY) - [E^2(X) + E^2(Y) + 2E(X)E(Y)]$
 $V(X + Y) = V(X) + V(Y) + 2Cov(XY)$
avec $Cov(XY) = E(XY) - E(X)E(Y)$

On notera que dans le cas de deux variables indépendantes, alors $E(XY) = E(X)E(Y)$. Sous ces conditions, la variance devient la somme des variances $V(X + Y)$ devient la somme des variances de X et Y.

F – Graphes IC Mack vs BootStrap

Seront présentés ici les graphiques des méthodes de liquidation de Mack et par bootstrap, les résultats sont cumulés par marque et segment de marché. A noter que sur le segment auto, pour la société A, le bootstrap a été réalisé par deux fois. Avec une première simulation de 10 000 itérations, comme pour les autres segments et une seconde avec 1 000 000 de cas. Seul ce graphique-là présentera 2 méthodes de bootstrap.

A noter que les temps de calcul sont de plusieurs heures par triangle ainsi projeté, pour un gain de précision jugé limité.

On notera également que les scénarii de bootstrap sont indépendants, ils sont réalisés par triangle. Les IC respectifs de chaque triangle sont définies par triangle et l'on définira l'IC global comme la somme des IC empirique. Une autre approche serait de sommer les résultats des scénarii (x par triangle) pour obtenir des montants cumulés et en déduire un IC empiriquement.

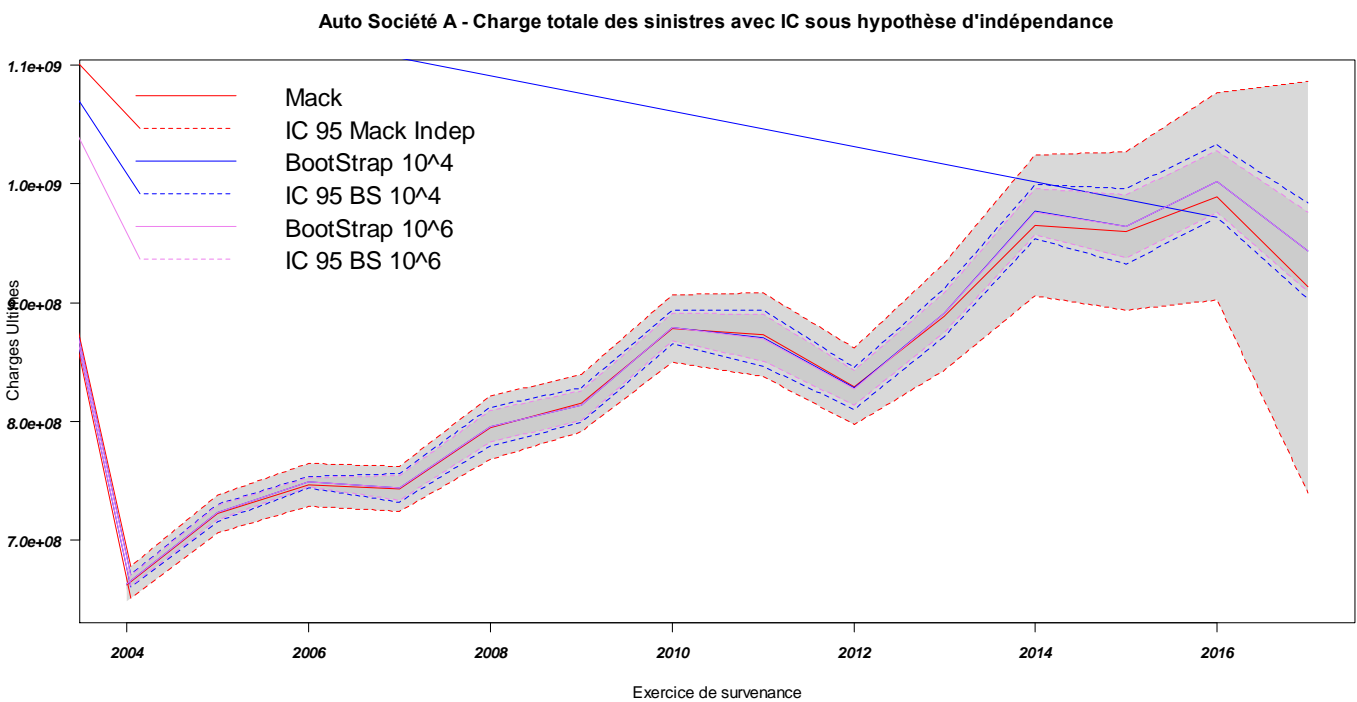


FIGURE 59 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE A, MARCHE DE L'AUTO

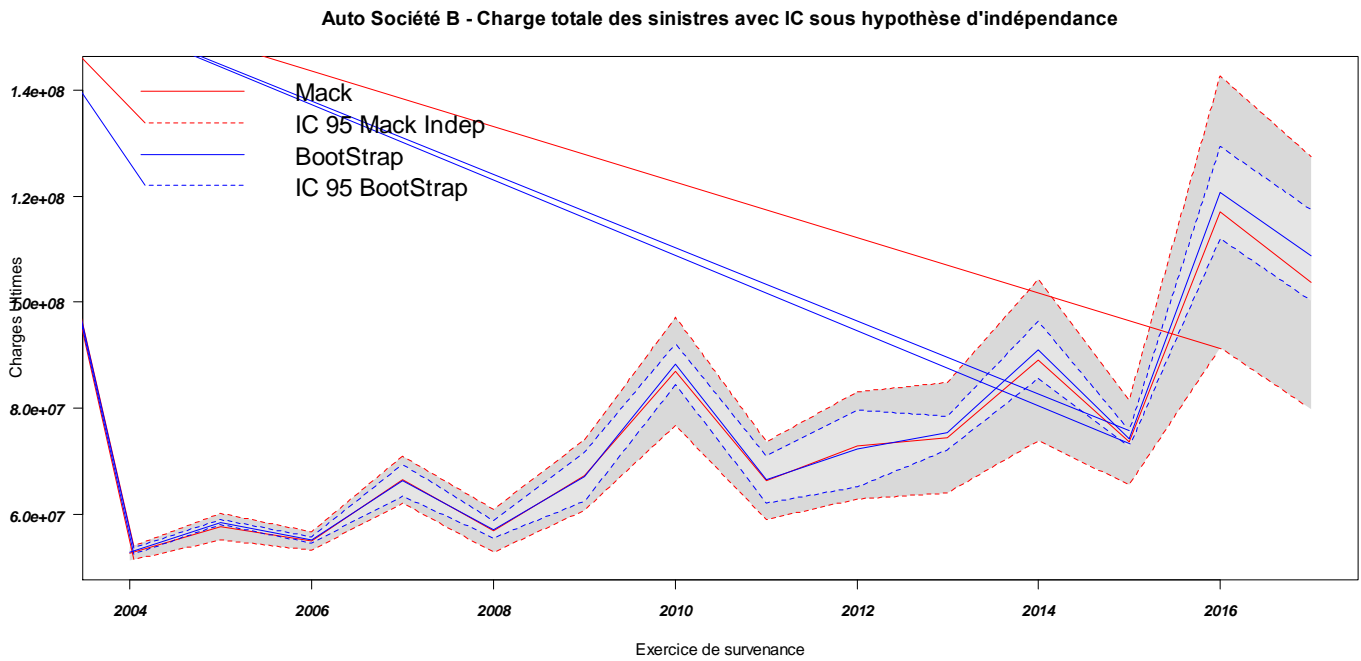
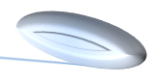


FIGURE 60 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE B, MARCHÉ DE L'AUTO

On notera que sur une société plus petite (B, figure XLVII) la sensibilité face à l'incertitude d'estimation de la charge ultime, augmente en taux.

Malgré cet écart, il ressort que la méthode de bootstrap est moins volatile.

Par ailleurs, la très faible variance en 2015 s'explique par un exercice avec une sinistralité très réduite sur les sinistres graves.

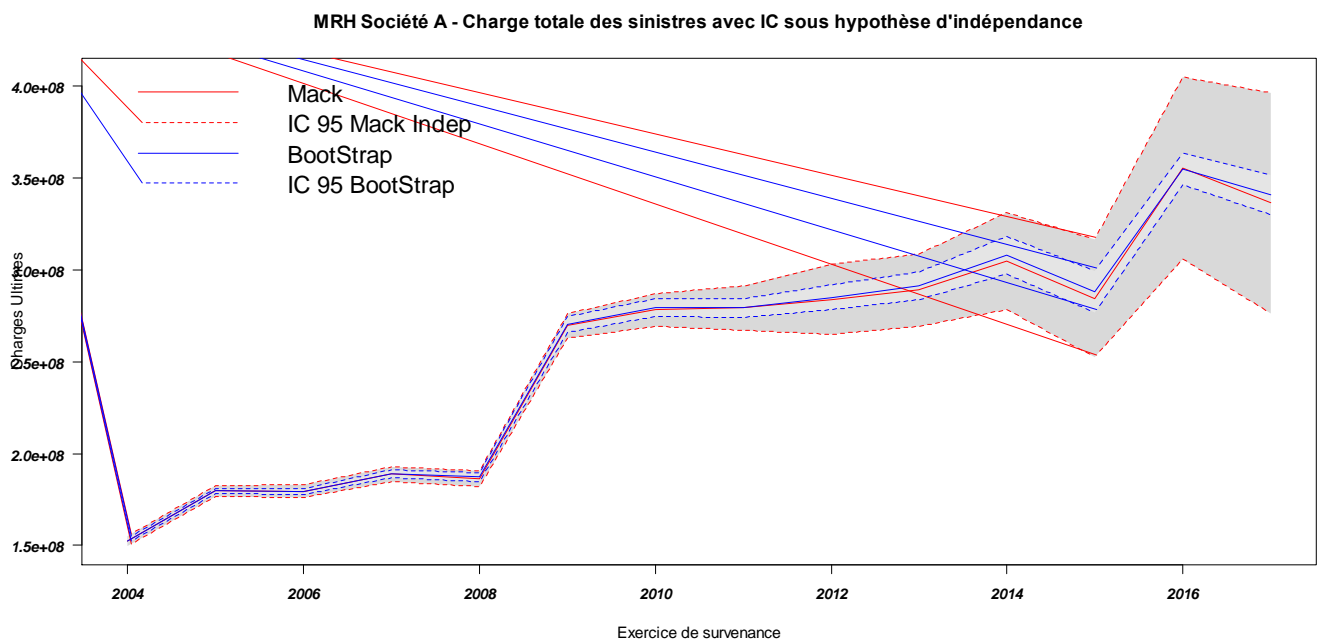


FIGURE 61 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE A, MARCHÉ DE L'HABITATION

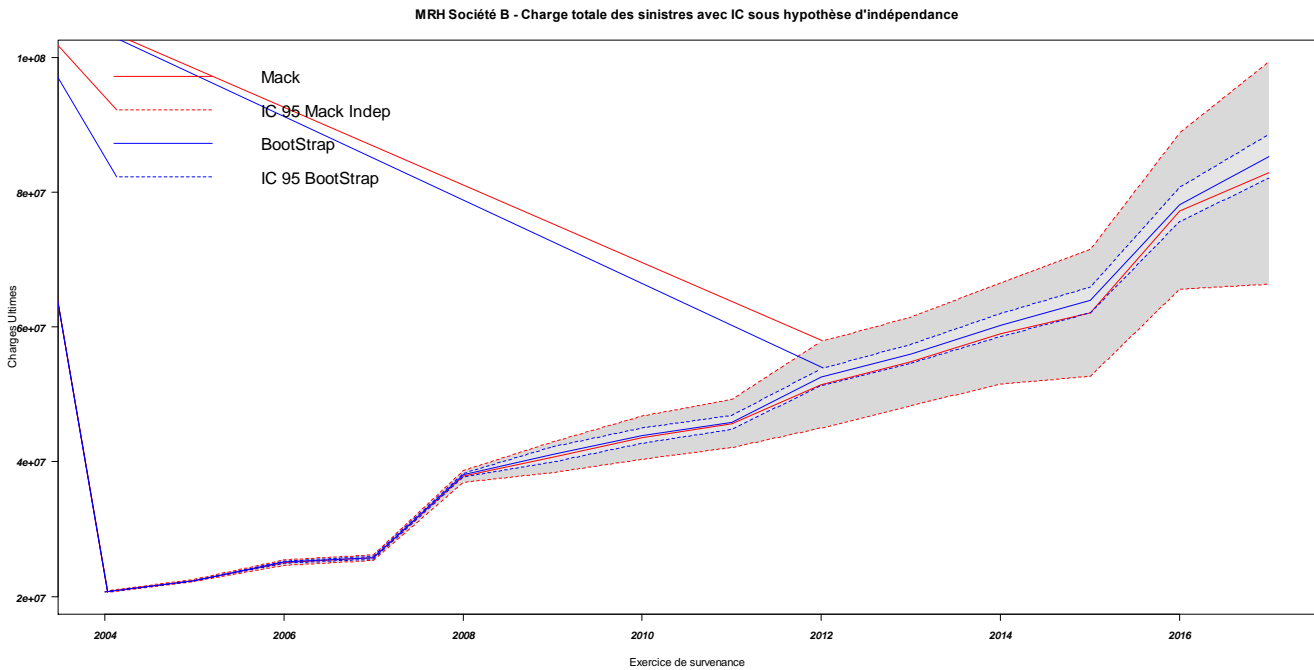


FIGURE 62 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE B, MARCHE DE L'HABITATION

A noter que ces courbes de charges sont à rapprocher du portefeuille qui se développe en parallèle dans des proportions moindre. Si l'on raisonne en Prime Pure, la tendance reflèterait l'inflation des coûts (fréquence et coût unitaire).

On peut noter que l'IC est faible en habitation, surtout avec les approches de bootstrap. Cela vient de la moindre volatilité des graves dans le poids global. L'importance des sinistres climatiques peut générer une incertitude, cependant une fois l'exercice clos, les évènements ont eu lieu et contrairement aux sinistres graves en auto, la probabilité de changer d'état est nulle. Il reste toutefois le cas de la sécheresse, qui met jusqu'à 3 ans avant d'atteindre une pleine connaissance, ce phénomène est absorbé, ou absorbable, grâce à des approches de type Bornhuetter-Ferguson, qui peuvent être également retravaillée par bootstrap.

G – Représentation des nuages de points : triangle a en fonction du triangle b

Il s'agit ici de visualiser les corrélations entre les triangles, de manière rapide et graphique.



FIGURE 63 - MATRICE DES GRAPHIQUES DES TRIANGLES LES UNS EN FONCTION DES AUTRES

Par exemple, la linéarité entre les triangles AHG et AHM, soit respectivement la société A, en habitation, sur les garanties Grave et Masse, montre qu'il existe une corrélation forte.

Il en est de même entre les triangles AAC et BAC, soit la sinistralité climatique en auto des sociétés A et B.

On note également des liens plus confus entre AHC et BAG, ou à l'inverse très net entre AHM et BHM.

Cela illustre ce qui est démontré dans le rapport, à savoir qu'il y a une prédominance dans le regroupement des garanties, voir des segments de marché, face à la société porteuse du risque. Il existe donc bien des phénomènes de marché et une sensibilité des sociétés face à ce marché.

H – Résultats des approches multivariées à la marque x segment marché

Pour des questions de temps et d'analyse, seules la société A, en Auto et MRH, a été testée.

Les approches testées donnent des résultats volatils d'un exercice à l'autre et des espérances parfois très différentes. Il serait intéressant de s'attarder plus sur les causes, en testant la présence de points atypiques et en les pondérant. Il serait également fort enrichissant de tester les nouvelles méthodes émergents sur le marché.

Cela pourra être réalisé par l'équipe performance économique au fil du temps, tout comme l'élargissement aux autres sociétés du groupe.

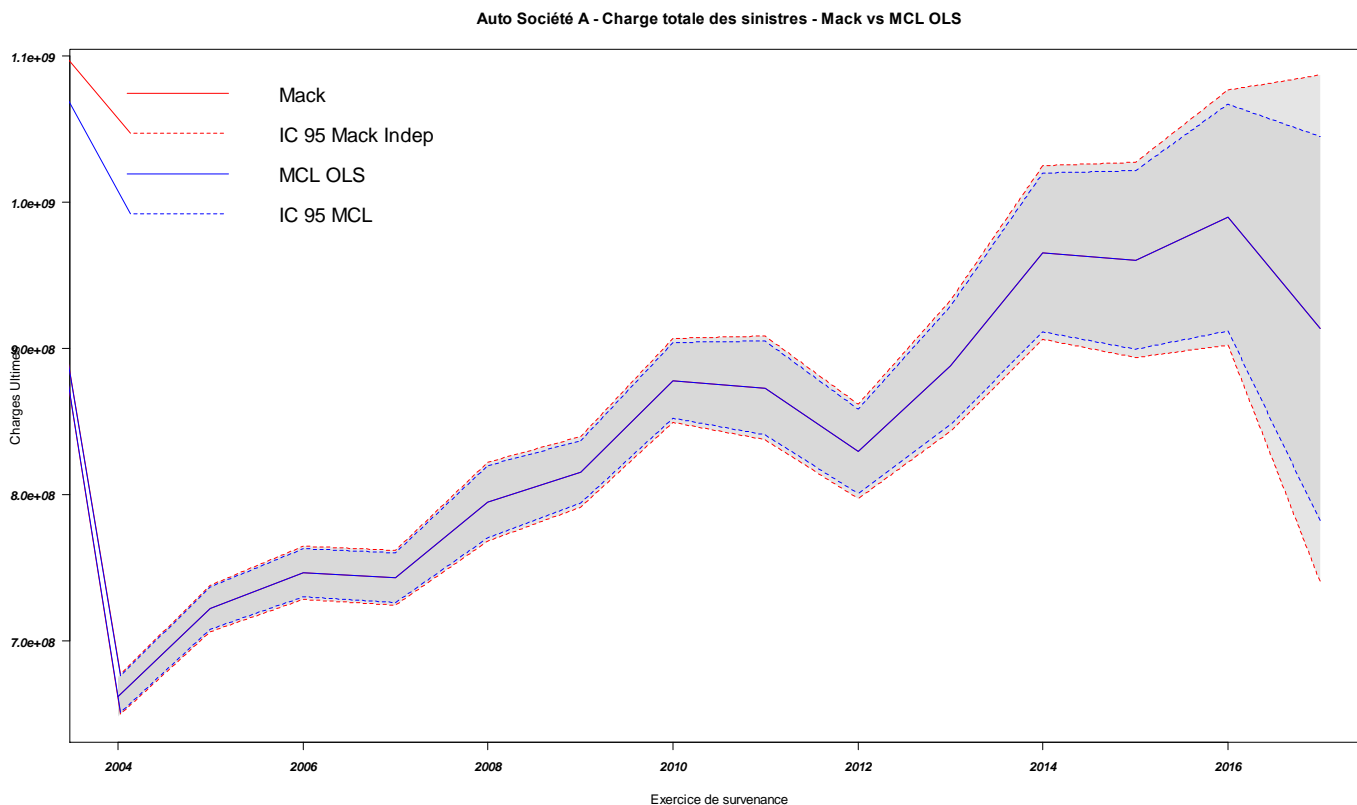


FIGURE 64 - PASSER D'UNE VISION INDEPENDANTE A UNE VISION DEPENDANTE DANS LA SOMME DES VARIANCES, SOCIETE A AUTOMOBILE

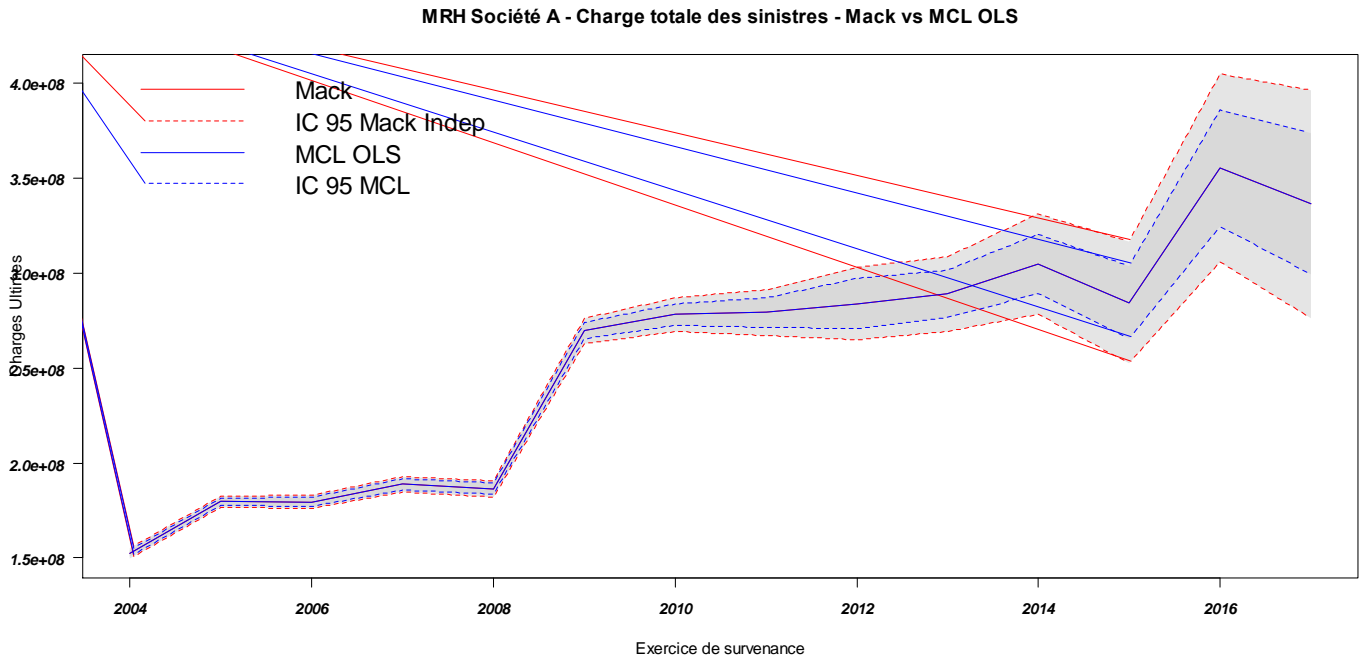
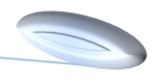


FIGURE 65 - PASSER D'UNE VISION INDEPENDANTE A UNE VISION DEPENDANTE DANS LA SOMME DES VARIANCES, SOCIETE A HABITATION

On note que les corrélations ont un impact à la baisse sur l'encadrement des résultats.

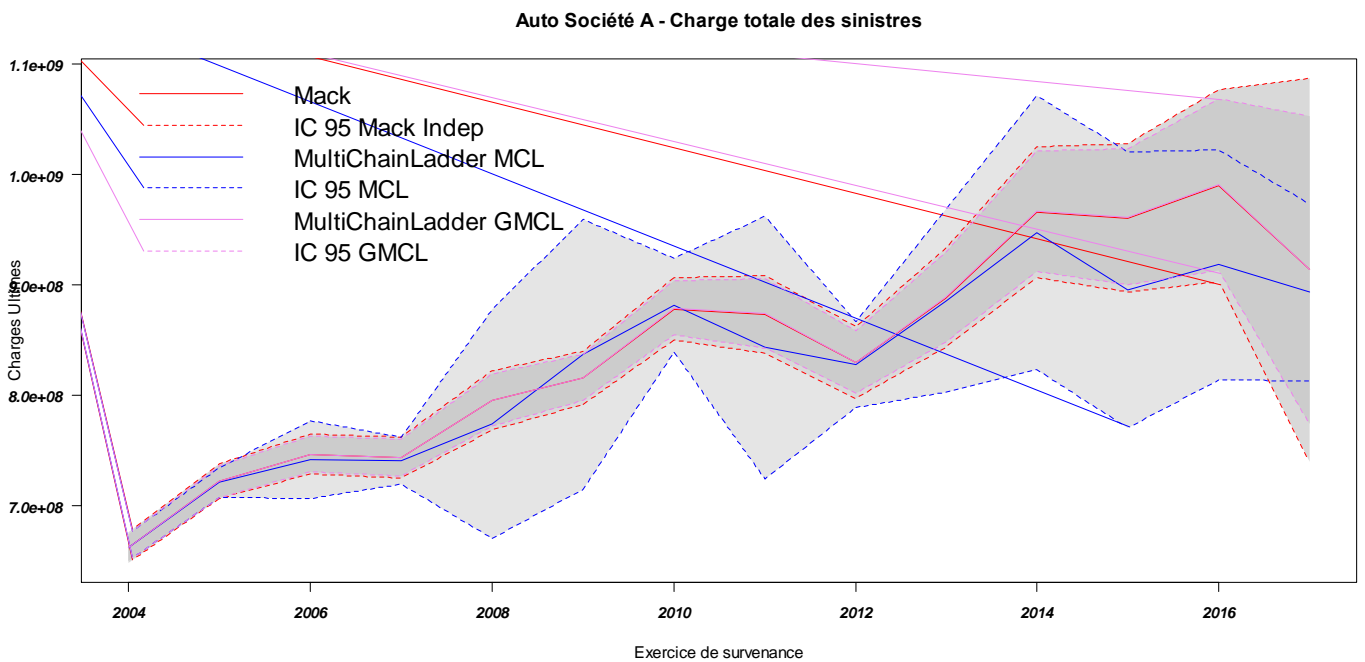


FIGURE 66 - COMPARAISON DES APPROCHES MULTIVARIEES SANS INTERCEPTS, SOCIETE A, AUTOMOBILE

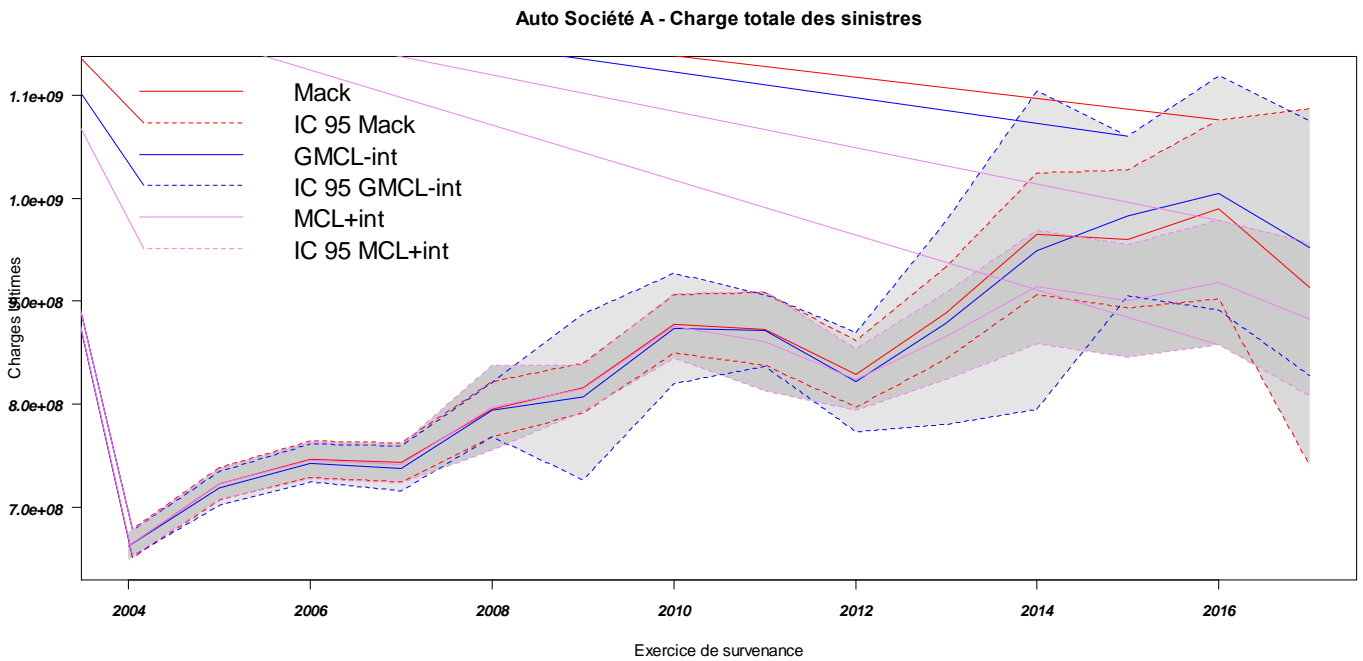
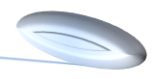


FIGURE 67 - COMPARAISON DES APPROCHES MULTIVARIEES AVEC INTERCEPTS, SOCIETE A, AUTOMOBILE

A retenir, les méthodes multivariée apporte une disparité forte, portée par les points extrêmes, sur certains exercices. De plus, les valeurs moyennes sont très éloignées des approches traditionnelles, ce qui laisse planer un doute fort quant à leur véracité.

Comme avancé dans le mémoire, on leur préférera les approches bootstrapées, bien que ces dernières soient certainement fausses car elles ignorent les corrélations.



I - Résultats des projections du portefeuille par propagation en nœuds

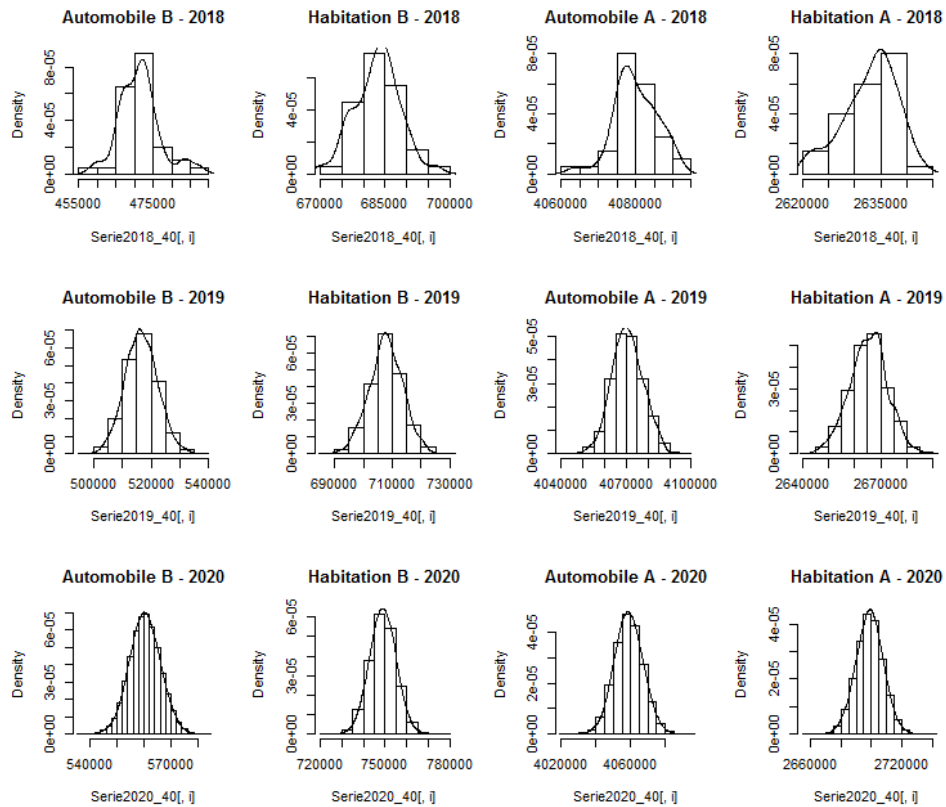


FIGURE 68 - DENSITE DES PROJECTIONS A 3 ANS MODELE MINIMALISTE

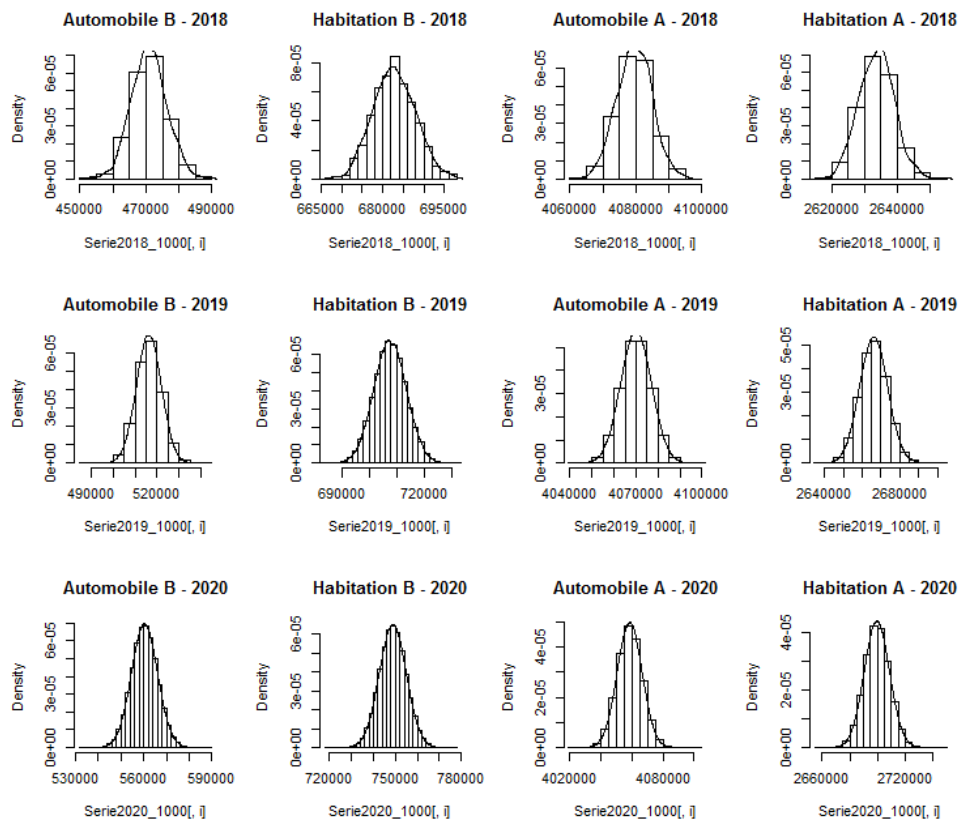


FIGURE 69 - DENSITE DES PROJECTIONS A 3 ANS MODELE RETENU

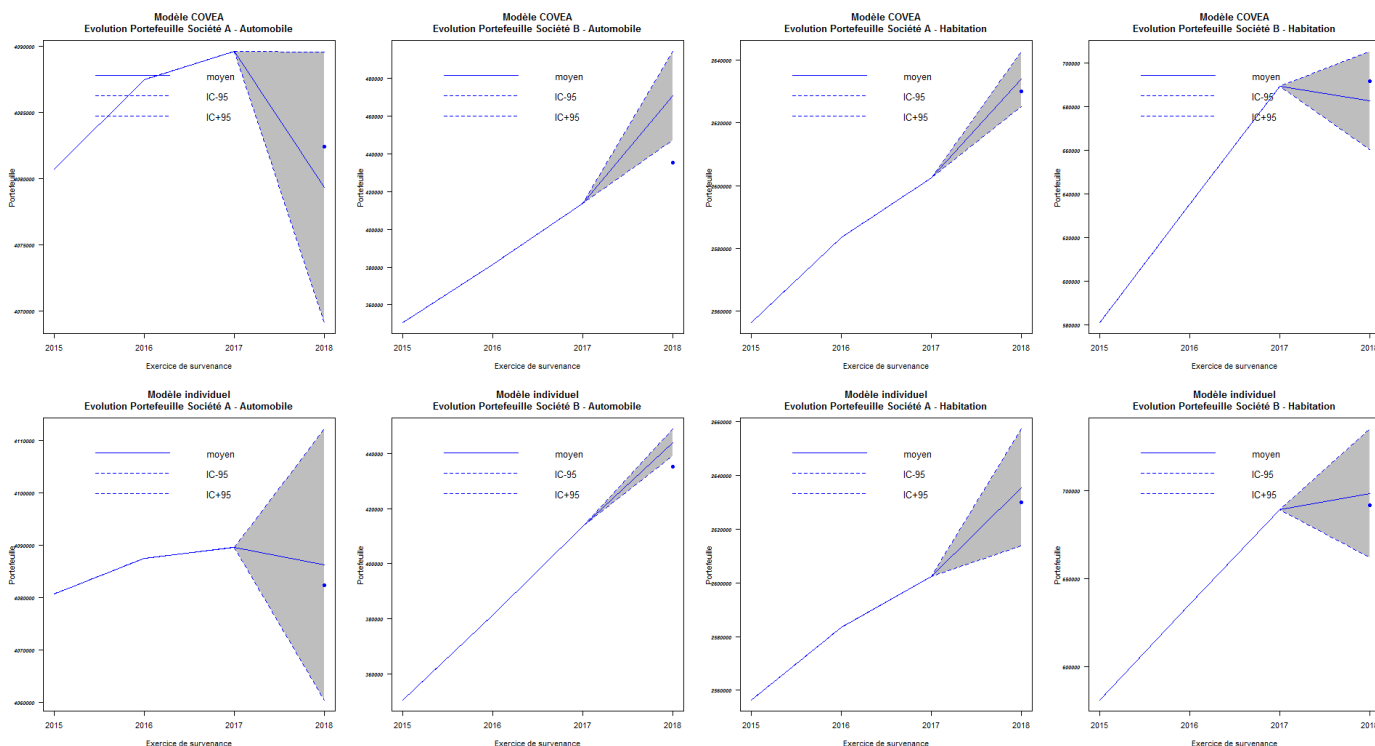


FIGURE 70 – COMPARAISON DES PREVISION A 1 AN MODELE COVEA VS MODELES INDIVIDUELS ET ATERRISSAGES REELS

On note ici que l'incertitude induite par le modèle Covéa est plus grande que celle des modèles individuels. De même, on note que pour le produit automobile de la société B, aucun modèle n'envisageait la performance obtenue. Cela s'explique par des objectifs jugés trop ambitieux et aussi par un ralentissement du marché, avec des acteurs comme les bancassureurs récupérant une part toujours plus importante des affaires.

De manière générale, il faut également noter que les modèles individuels ont des résultats plus proches de la réalité. Ils sont toutefois coûteux en temps de construction. Etant donné que le modèle Covéa a des résultats satisfaisant, notamment car son intervalle de confiance intègre la variable à prédire, et qu'il est plus intéressant de disposer d'un modèle unique, il sera préconiser de généraliser cette étude à toutes les marques et segments de marché pour valider le modèle unique.

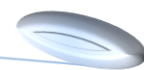
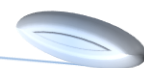


Table des illustrations

FIGURE 1 - ILLUSTRATION D'UN TRIANGLE DE LIQUIDATION	8
FIGURE 2 - SCHEMA SIMPLIFIE DE LA PROJECTION A 5 ANS DES RESULTATS	14
FIGURE 3 - ILLUSTRATION D'UN PARALLELOGRAMME DE LIQUIDATION	15
FIGURE 4 - ILLUSTRATION LISSAGE TECHNIQUE FREQUENCE ET COUT MOYEN	16
FIGURE 5 - ACQUISITION DES PRIMES DANS UN SYSTEME A EMISSIONS PLURIANNUELLES	23
FIGURE 6 - ACQUISITION DES PRIMES DANS UN SYSTEME A EMISSION UNIQUE	24
FIGURE 7 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE AUTO DE LA SOCIETE A	26
FIGURE 8 - ALIGNEMENT DES COUPLES $(C_{i,0}, C_{i,1})$ GARANTIE GRAVES AUTO DE LA SOCIETE A	27
FIGURE 9 - CORRELATIONS DES EXERCICES DU TRIANGLE DES CHARGES CLIMATIQUES HABITATION SOCIETE A (PEARSON - KENDALL)	28
FIGURE 10 - EXEMPLE DE TRONCATURE SUR UNE LOI NORMALE ENTRE -1,5 ET +1,5	30
FIGURE 11 - INTERVALLES DE CONFIANCE A 95% SUR LES METHODES DE MACK ET DE BOOTSTRAP - SOCIETE A - TOTAL	34
FIGURE 12 - INTERVALLES DE CONFIANCE A 95% SUR LES METHODES DE MACK ET DE BOOTSTRAP - SOCIETE B - TOTAL	34
FIGURE 13 - RESULTATS DE L'ACP : CONSTRUCTION DES AXES RESUMANT LA VARIANCE	36
FIGURE 14 - RESULTAT DE LA CAH SUR LES TRIANGLES DE CHARGE	37
FIGURE 15 - RESULTAT DE LA CAH SUR LES TRIANGLES DE DEVELOPPEMENT	38
FIGURE 16 - CORRELATION ET CLASSIFICATION DES TRIANGLES DE CHARGE	38
FIGURE 17 - CORRELATION ET CLASSIFICATION DES D-TRIANGLES	39
FIGURE 18 - COMPARAISON INDEPENDANCE TOTALE ULTIMES MACK ET MULTI-CHAIN-LADDER - SOCIETE A - TOTAL	40
FIGURE 19 - COMPARAISON INDEPENDANCE TOTALE ULTIMES MACK, MCL ET GMCL - SOCIETE A - TOTAL	41
FIGURE 20 - TABLEAU RECAPITULATIF DES METHODES ETUDIEES	43
FIGURE 21 - RESIDUS STANDARDISES DU MODELE LINEAIRE COMPLET - SOCIETE A AUTOMOBILE	46
FIGURE 22 - RESIDUS STANDARDISES DU MODELE LINEAIRE GENERALISE <i>BACKWARD</i> - SOCIETE A AUTOMOBILE	47
FIGURE 23 - PROPAGATION TEMPORELLE AVEC MEMOIRE : LA SIMULATION PAR NŒUD	49
FIGURE 24 - PROPAGATION DU SCENARIO MINIMALISTE POUR LE PORTEFEUILLE DE LA SOCIETE A, SEGMENT AUTO	51
FIGURE 25 - PROPAGATION DU SCENARIO INTERMEDIAIRE POUR LE PORTEFEUILLE DE LA SOCIETE A, SEGMENT AUTO	51
FIGURE 26 - PROPAGATION DU SCENARIO OPTIMISANT N+1 POUR LE PORTEFEUILLE DE LA SOCIETE A, SEGMENT AUTO	51
FIGURE 27 - COMPARAISON DE LA TAILLE DES IC SELON LA METHODE DE PROPAGATION APPLIQUEE, PORTEFEUILLE AUTO SOCIETE A	51
FIGURE 28 - ESTIMATION DE LA PRECISION DES MODELES A 1 AN	52
FIGURE 29 - COMPARAISON MODELE SPECIFIQUE ET MODELE GENERIQUE	53
FIGURE 30 - LE BP COVEA ET SA PART D'ETAPE STOCHASTIQUE	56
FIGURE 31 - REPARTITION EMPIRIQUE DES SINISTRES DE PLUS DE 150K€	57
FIGURE 32 - REPARTITION MODELISEE DES SINISTRES GRAVES AU-DELA DE 10M€	58
FIGURE 33 - ESTIMATION DES PERIODES RETOUR SELON LE MONTANT DU SINISTRE	58
FIGURE 34 - PROFIL DE LA VICTIME TYPE POUR LES SINISTRES CORPORELS IMPORTANTS CHEZ COVEA	59
FIGURE 35 - ESTIMATION DE LA DATE DE DEPASSEMENT DU TAUX DE REVALORISATION A 2,25% SUR LES COURBES DES TAUX SANS RISQUE EIOPA/IA A FIN 2016	60
FIGURE 36 - COMPARAISON DES PREVISIONS A 1 AN MODELE COVEA VS MODELES INDIVIDUELS ET ATERRISSAGES REELS	61
FIGURE 37 - PROPAGATION A 5 ANS - 200 SIMULATIONS PAR NŒUD	63
FIGURE 38 - IMPACT DU NOMBRE DE SIMULATIONS SUR L'INTERVALLE DE CONFIANCE POUR LE BOOTSTRAP	65
FIGURE 39 - IMPACT DES CORRELATIONS SUR L'INTERVALLE DE CONFIANCE - CUMUL DES GARANTIES DE LA SOCIETE A EN AUTOMOBILE	65
FIGURE 40 - CLASSIFICATION DES TRIANGLES DE COEFFICIENTS DE DEVELOPPEMENT	66
FIGURE 41 - TRIANGLE DE CHARGE DES SINISTRES CLIMATIQUE SOCIETE B EN HABITATION	68
FIGURE 42 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE AUTO DE LA SOCIETE A	69
FIGURE 43 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE AUTO DE LA SOCIETE B	70
FIGURE 44 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE HABITATION DE LA SOCIETE A	70
FIGURE 45 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE DE MASSE HABITATION DE LA SOCIETE B	70
FIGURE 46 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE AUTO DE LA SOCIETE A	71
FIGURE 47 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE AUTO DE LA SOCIETE B	71
FIGURE 48 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE HABITATION DE LA SOCIETE A	72
FIGURE 49 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ GARANTIE CLIMATIQUE HABITATION DE LA SOCIETE B	72
FIGURE 50 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES AUTO DE LA SOCIETE A	73
FIGURE 51 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES AUTO DE LA SOCIETE B	73
FIGURE 52 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES HABITATION DE LA SOCIETE A	74
FIGURE 53 - ALIGNEMENT GRAPHIQUE DES COUPLES $(C_{i,j}, C_{i,j+1})$ SINISTRES GRAVES HABITATION DE LA SOCIETE B	74
FIGURE 54 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE A, SEGMENT AUTO	75
FIGURE 55 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE B, SEGMENT AUTO	75
FIGURE 56 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE A, SEGMENT HABITATION	76
FIGURE 57 - ETUDE DES CORRELATIONS ENTRE LES ANNEES DU TRIANGLE DE LA SOCIETE B, SEGMENT HABITATION	76
FIGURE 58 - CADENCE DE DEVELOPPEMENT	77

FIGURE 59 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE A, MARCHE DE L'AUTO	78
FIGURE 60 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE B, MARCHE DE L'AUTO	79
FIGURE 61 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE A, MARCHE DE L'HABITATION	79
FIGURE 62 - COMPARAISON DES METHODES SOUS HYPOTHESE D'INDEPENDANCE, SOCIETE B, MARCHE DE L'HABITATION	80
FIGURE 63 - MATRICE DES GRAPHIQUES DES TRIANGLES LES UNS EN FONCTION DES AUTRES	81
FIGURE 64 - PASSER D'UNE VISION INDEPENDANTE A UNE VISION DEPENDANTE DANS LA SOMME DES VARIANCES, SOCIETE A AUTOMOBILE	82
FIGURE 65 - PASSER D'UNE VISION INDEPENDANTE A UNE VISION DEPENDANTE DANS LA SOMME DES VARIANCES, SOCIETE A HABITATION	83
FIGURE 66 - COMPARAISON DES APPROCHES MULTIVARIEES SANS <i>INTERCEPTS</i> , SOCIETE A, AUTOMOBILE	83
FIGURE 67 - COMPARAISON DES APPROCHES MULTIVARIEES AVEC <i>INTERCEPTS</i> , SOCIETE A, AUTOMOBILE	84
FIGURE 68 - DENSITE DES PROJECTIONS A 3 ANS MODELE MINIMALISTE	85
FIGURE 69 - DENSITE DES PROJECTIONS A 3 ANS MODELE RETENU	85
FIGURE 70 - COMPARAISON DES PREVISION A 1 AN MODELE COVEA VS MODELES INDIVIDUELS ET ATERRISSAGES REELS	86



Bibliographie

- ¹ JAZIRI S. (2011) Méthodes de provisionnement non-vie et risque de réserve à un an, ISFA Lyon
- ² BORNHUETTNER R.L. & FERGUSON R.E. The Actuary and IBNR
- ³ BENKTANDER G. An Approach to Credibility in Calculating I BNR for Casualty Excess Reinsurance. The Actuarial Review, April 1976
- ⁴ KOLMOGOROV A. Sulla determinazione empirica di una legge di distribuzione, Ist. Ital. Attuari. n°4 (1933), p 83–91.
- ⁵ SMIRNOV N. Approximate distribution laws for random variables, constructed from empirical data, Uspekhi Mat. Nauk n°10 (1944) p. 179–206
- ⁶ KUIPER N. H. Tests concerning random points on a circle, Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, Series A 63 (1960) p 38–47
- ⁷ SAVARE S. & PAIRE B. (2012) Charge ultime nette de réassurance en RC corporelle : 2 modèles stochastiques pour les flottes automobiles, CEA
- ⁸ MACK, T. (1993), Distribution-free calculation of the standard error of chain ladder reserve estimates, Astin Bull., 23, p213-225
- ⁹ MACK, T. (2008), The prediction error of Bornhuetter-Ferguson, Casualty Actuarial Society E-Forum, Fall 2008, p222-240
- ¹⁰ BRAUN, C. (2004), The prediction error of the chain-ladder method applied to correlated run-off triangles. Astin Bull., 34, p399-423
- ¹¹ PROHL, C. Schmidt K.D (2005), Multivariate chain-ladder, Dresdner Schriften zur Versicherungsmathematik.
- ¹² ZHANG, Y. (2010), A general multivariate chain-ladder model, Insurance: Mathematics and Economics, n°46, p. 588-599
- ¹³ ZELLNER, A. (1962), An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, journal of the American statistical association, 57, p348-368.
- ¹⁴ RENSHAW & VERRALL (1998), A stochastic model underlying the chain-ladder technique, British Actuarial Journal, Vol. 4, p903-923
- ¹⁵ TRYON, R.C. (1939) Cluster Analysis: Correlation Profile and Orthometric (Factor) Analysis for the Isolation of Unities in Mind and Personality. Edwards Brothers, Ann Arbor.
- ¹⁶ JOLLIFFE, I. (2002) Principal Component Analysis, 2nd edition, Springer-Verlag.
- ¹⁷ PEARSON, K. (1901) On Lines and Planes of Closest Fit to Systems of Points in Space, Philosophical Mag., vol. 2, n° 6, p. 559–572
- ¹⁸ WARD, J. (1963) Hierarchical Grouping to Optimize an Objective Function, jnl of the American Statistical Ass., 58(301), 236-244
- ¹⁹ Co-auteurs, LINDE M (2013-2016), The Multi-Year Non-Life Insurance Risk, Faculté de mathématique et d'économie, ULM
- ²⁰ FABRE RUDELLE D. (2018), Apport des méthodes d'apprentissage statistique pour le provisionnement individuel en assurance non-vie, ISUP
- ²¹ LE TESSON A., LENAIN A., SAMBA S et UNG J. (2014) Estimation de l'erreur de prédiction dans le cas de l'utilisation d'une combinaison de méthodes pour le calcul de provisions en assurance IARD, EURIA Brest, pages 41-42 et 52-53
- ²² LONGLEY-COOK L.H. (1962), an introduction to credibility theory, Casualty Actuarial Society, vol 88, p194-221
- ²³ FRECHET M. (1938), *Les événements en chaîne dans le cas d'un nombre infini d'états possibles*, Rech. théo. modernes calcul des proba., tome II
- ²⁴ BLACK F & SCHOLLES M (1973), *The pricing of Options and Corporate Liabilities*, Journal of Political Economy 81, n°3, p637-654
- ²⁵ BROWN R. (1828), *A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies.*, Philosophical Magazine 4, p161-173
- ²⁶ La loi n°2014-344 du 17 mars 2014 relative à la consommation, dite loi Hamon du nom du ministre de l'économie en place à cette date
- ²⁷ AKAIKE H. (1973), Information theory and an extension of the maximum likelihood principle, 2nd international symposium on information theory, p 267-281
- ²⁸ GIDEON E. Schwarz (1978), Estimating the dimension of a model, Annals of Statistics, vol 6, p 461–464
- ²⁹ SHAPIRO S.S. (1965) An analysis of variance test for normality (complete samples), Biometrika, vol. 52, p 591-611