

Mémoire présenté devant le Conservatoire National des Arts et Métiers pour l'obtention du Master Droit Economie Gestion mention Actuariat et l'admission à l'Institut des Actuaires

le mercredi 14 juin 2023

Par : Maxime MECHALI

Titre: Méthodes de provisionnement stochastiques utilisant des modèles bayésiens

Confidentialité : NON OUI Durée : 2 ans

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

Membres présents du jury de l'Institut des Actuaires signatures

M. Michel GERMAIN

M. Davy SENGDY

Entreprise :

Nom :

AXA

Membres présents du jury du Cnam

M. Nathanaël ABECERA

M. Olivier DESMETTRE

M. David FAURE

M. François WEISS

Directeur de mémoire en entreprise :

Nom : Xavier SERVEL

Signature :

Tuteur académique :

Nom : Olivier LOPEZ

Signature :

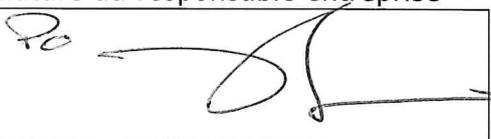
Autorisation de publication et de

mise en ligne sur un site de

diffusion de documents actuariels

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise



Secrétariat

Bibliothèque :

Signature du candidat



Résumé

Le but de ce mémoire a été de développer un modèle bayésien de provisionnement stochastique des réserves à l'ultime d'une part et l'étendre à la mesure de l'incertitude sur l'évaluation des Boni-Mali, ce qui a permis d'évaluer le SCR et le respect du pilier quantitatif de la norme "Solvabilité II". Le modèle bayésien Correlated Chain Ladder (CCL) est présenté et testé sur différents triangles de AXA XL, entité du modèle interne du groupe AXA. Pour simuler les distributions finales en sortie des modèles bayésiens, il est nécessaire de faire appel aux algorithmes dit "Monte Carlo Markov Chain" qui simulent des chaînes de Markov convergeant vers la distribution cible. Les études menées sur des triangles de provisionnement de AXA XL permettent de conclure que le CCL fournit une estimation assez précise du risque de réserve à horizon 1 an d'une ligne d'activité, tout en permettant d'inclure des connaissances extérieures qui ne sont pas présentes dans les données. En cela, le CCL constitue une alternative intéressante aux modèles de provisionnement stochastiques déjà existant comme Merz & Wutrich à laquelle il a été régulièrement comparé dans ce mémoire.

Mots clefs: Modèles bayésiens ; provisionnement stochastique ; Monte Carlo Markov Chain

Abstract

The goal of this paper was to develop a bayesian model of stochastic claims reserving. An ultimate stochastic reserving model is firstly introduced and was extended to the estimation of the Claims Development Result uncertainty. This extension allowed us to evaluate the SCR and check the respect of the quantitative pillar of the "Solvency II" directive. The bayesian model Correlated Chain Ladder (CCL) is presented and tested on different triangles of AXA XL, an entity in the internal model scope of AXA Group. In order to simulate the output distribution of the bayesian model, it is necessary to use "Monte Carlo Markov Chain" algorithms, which simulate Markov chains that converge toward the target distribution. The studies of some reserving triangles of AXA XL have led us to conclude that the CCL method provides a precise estimation of the one-year reserve risk of a line of business, and that embed external knowledge not present in the data. In this way, the CCL is a legitimate alternative to the more traditional stochastic reserving method such as Merz and Wutrich, a method that were often compared to the CCL in this paper.

Keywords: Bayesian models ; stochastic claims reserving ; Monte Carlo Markov Chain

Remerciements

Je tiens à remercier pour commencer Monsieur Olivier Lopez, qui a accepté d'être mon tuteur académique pour ce mémoire et dont l'expertise et les connaissances en statistiques m'ont permis d'acquérir beaucoup de rigueur dans la rédaction de ce mémoire. J'aimerais ajouter que malgré son emploi du temps très chargé, il a pris le temps de relire et m'aider à valider ce rapport.

Je remercie également Monsieur Xavier Servel, mon tuteur d'entreprise à AXA, qui a pris le temps de suivre l'avancement de mon travail et grâce à qui je me suis rendu compte de beaucoup d'incohérences dans mes raisonnements que j'ai pu finalement corriger. D'autres part, j'ai pu également grâce à lui prendre du recul sur le sujet et recentrer les objectifs du mémoire pour une éventuelle utilité à la gestion des risques d'AXA.

Je remercie aussi mes collègues d'AXA Bryan Gautier, Yannick Bessy-Roland, Maxence Pierrat et François Tchangäi, qui ont pris le temps de s'intéresser au sujet et d'avoir des discussions avec moi pour parler de certains détails que j'ai pu oublier. Et également Théo Sermet qui m'a encouragé à faire l'école d'actuariat afin de passer le diplôme de l'Institut des Actuaire.

Enfin je remercie mes autres collègues et anciens collègues qui m'ont soutenu pendant toute la rédaction du mémoire, à savoir Geoffrey Bard, Riad Abdellatif, Maxime Hersent, Hugo Rakotoarimanga, Grégoire Gallot, Mildred Ochoa, Lila Ahmed-Zouaoui, Laura Gourlaouen, Sophie Navarro, Jade Halili qui ont été à mes côtés et m'ont soutenu pendant la rédaction de ce mémoire.

Note de synthèse

Le modèle du Correlated Chain Ladder (CCL) est tout d'abord une méthode de provisionnement bayésien à l'ultime permettant l'estimation, dans un triangle de paiement cumulé $D^{(t)} = (C_{i,j})_{1 \leq i \leq t, 1 \leq j \leq t-i+1}$, des paiements cumulés manquants, et en particulier les paiements à l'ultime $(C_{i,n})_{2 \leq i \leq t}$.

La fonction de vraisemblance du modèle consiste à représenter les paiements cumulés par une loi log-normale dont les paramètres vont dépendre de paramètres eux-mêmes aléatoires. Ces paramètres vont initialement avoir une loi dite *a priori* et les observations du triangle vont permettre à ces lois d'évoluer en loi *a posteriori*. C'est le principe de l'inférence bayésienne qui est résumé par l'équation : Densité *a posteriori* non normalisée \propto Vraisemblance \times Densité *a priori* non normalisée. Ces lois *a posteriori* vont ensuite permettre de simuler des distributions de sorties pour les paiements cumulés.

Le modèle repose sur une structure dite "cross-classified" c'est-à-dire que les paiements cumulés à prédire vont s'écrire sous la forme $C_{ij} = f(\alpha_i, \beta_j)$ avec les α comme paramètres contenant les informations liées aux années d'accident et les β comme paramètres contenant les informations liées aux années de développement.

Le modèle bayésien du CCL consiste à lier les paiements cumulés à des paramètres dont on veut déterminer les lois *a posteriori*. Ce lien est fait via une loi log-normale : $C_{ij} \sim \log \mathcal{N}(\mu_{ij}, \sigma_j)$ pour $i, j \in [1 : t]$. Les observations de certains C_{ij} seront ceux qui vont permettre d'apprendre la loi *a posteriori* des paramètres.

Les détails des lois *a priori* du modèle sont les suivants (avec B_i le montant de prime brut de l'année d'accident i) :

- $\mu_{1j} = \alpha_1 + \log(elt_1) + \beta_j$
- $\mu_{ij} = \alpha_i + \log(elt_i) + \beta_j + \rho[\log(C_{i-1,j}) - \mu_{i-1,j}]$
- $\alpha_i \sim \mathcal{U}[\log(B_i) - \epsilon_i, \log(B_i) + \epsilon_i]$ qui est la loi *a priori* du paramètre α_i ne dépendant que de l'année d'accident.
- $elt_i \sim \log \mathcal{N}(\gamma_i, \delta_i)$ qui est la loi *a priori* de "l'expected loss ratio" de l'année i .
- $\beta_j \sim \mathcal{U}[-\eta_j, 0]$ avec comme contrainte $\beta_t = 0$, qui est une loi *a priori* d'un paramètre ne dépendant que de l'année de développement des paiements.
- $\sigma_j^2 = \sum_{h=j}^t \tau_h$ où $\tau_h \sim \text{Beta}(a, b)$, la loi *a priori* de la volatilité qui est décroissante au fur et à mesure que l'année d'accident s'est développée

- $\rho \sim \mathcal{U}[-1, 1]$ la loi *a priori* du paramètre représentant la corrélation entre année d'accident

Etant donné la complexité des expressions des densités des lois *a posteriori*, il est nécessaire de faire appel à des algorithmes Markov Chain Monte Carlo (MCMC) pour simuler leurs distributions. Il s'agit de méthodes consistant à fabriquer et simuler des processus de Markov $(X_t)_{t \geq 0}$ dont la distribution limite est la distribution cible qui est la loi *a posteriori* à estimer. Une méthode bien populaire et assez simple permettant de construire ces chaînes de Markov est l'algorithme de Metropolis-Hastings. Dans ce mémoire, et étant donné la complexité du modèle, nous utilisons des méthodes dérivées de cet algorithme qui sont le Hamiltonian Monte Carlo et le No-U-Turn Sampler et qui permettent d'accélérer la convergence des chaînes de Markov vers leur distribution cible.

Une fois les simulations des chaînes de Markov et des distributions *a posteriori* des paramètres effectuées, on est en mesure de simuler l'incertitude sur les réserves pour une ligne d'activité, c'est à dire la différence entre les charges ultimes et ce qui a déjà été payé : Réserves = Charge Ultime – Paiements.

Deux lignes d'activités de l'entité AXA XL ont été étudiées. La première est constituée de 10 années d'accidents historiques. Il s'agit d'une ligne que l'on va nommer "ligne A" pour des raisons de confidentialité. La figure 1 est le résultat des simulations des réserves à l'ultime pour cette ligne avec un exemple de paramétrage de lois *a priori* initiales. On retrouve bien la forme de distribution log-normale qui a été introduite dans le modèle plus haut. La deuxième ligne, que l'on va appeler "ligne B", n'a que 5 années d'accident et la méthode CCL a été testée dans le but de voir dans quelles mesures les jugements d'expert, reflétés dans les lois *a priori* des paramètres, vont permettre une estimation pertinente de la distribution des réserves.

Pour répondre aux directives Solvabilité II, et en particulier le pilier quantitatif qui impose une estimation du SCR à horizon 1 an, il est nécessaire d'étendre la méthode CCL à l'estimation de l'incertitude de la déviation des charges ultimes estimés entre deux années calendaires. Autrement dit, on cherche à estimer la volatilité autour des Boni-Mali pour un triangle donné.

L'idée de l'extension à 1 an du CCL est de simuler la dernière diagonale du triangle qui correspond aux paiements de l'année calendaire $t + 1$ et de mettre à nouveau à jour les distributions *a posteriori* grâce aux nouvelles informations apportées par la diagonale simulée, puis de refaire une estimation de la distribution des réserves à l'ultime ce qui va donner ensuite une distribution des Boni-Mali. Cette distribution de Boni-Mali, pour le même exemple de paramétrage de la ligne A, est visible sur la figure 2, et on y retrouve l'hypothèse de lognormalité.

Les principaux avantages de la méthode CCL qui ont été révélés par cette étude sont :

- L'obtention d'une distribution complète et précise pour représenter le risque de

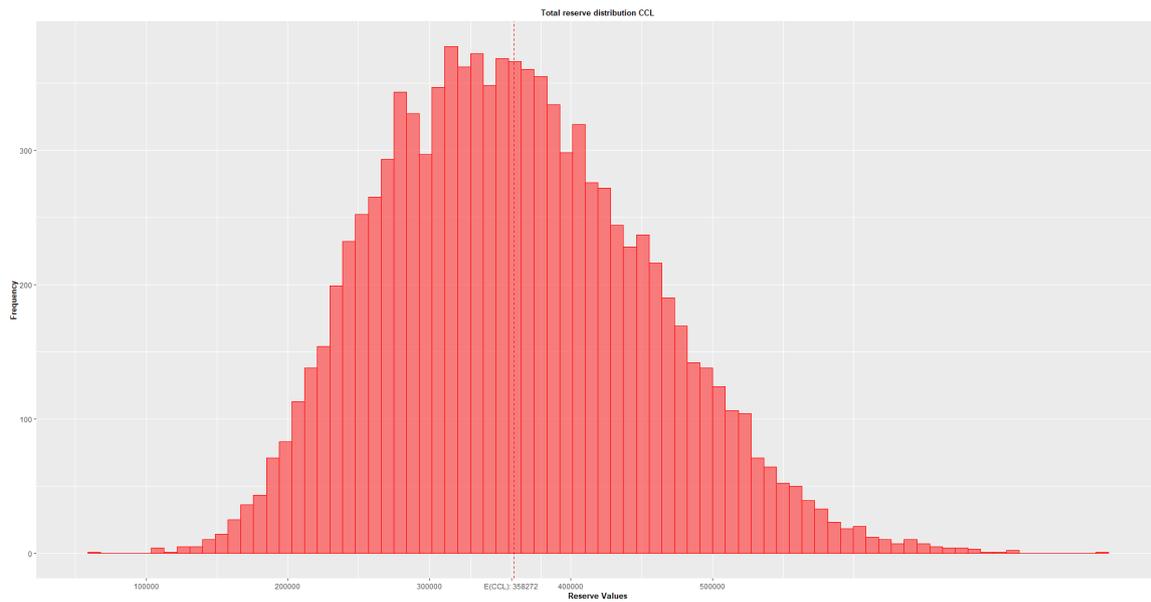


FIGURE 1 – LoB A : Distribution des réserves à l'ultime

réserve à l'ultime et à 1 an

- Une possibilité d'inclusion des jugements d'expert dans les paramétrages pour augmenter artificiellement le volume de données
- Une facile manipulation des paramétrages qui vont permettre d'insérer de la prudence dans les estimations

En contrepartie, les principaux inconvénients sont :

- La méthode est couteuse en temps de calcul
- Le nombre de paramètres à calibrer est élevé et peut donc nécessiter beaucoup de justifications auprès des auditeurs

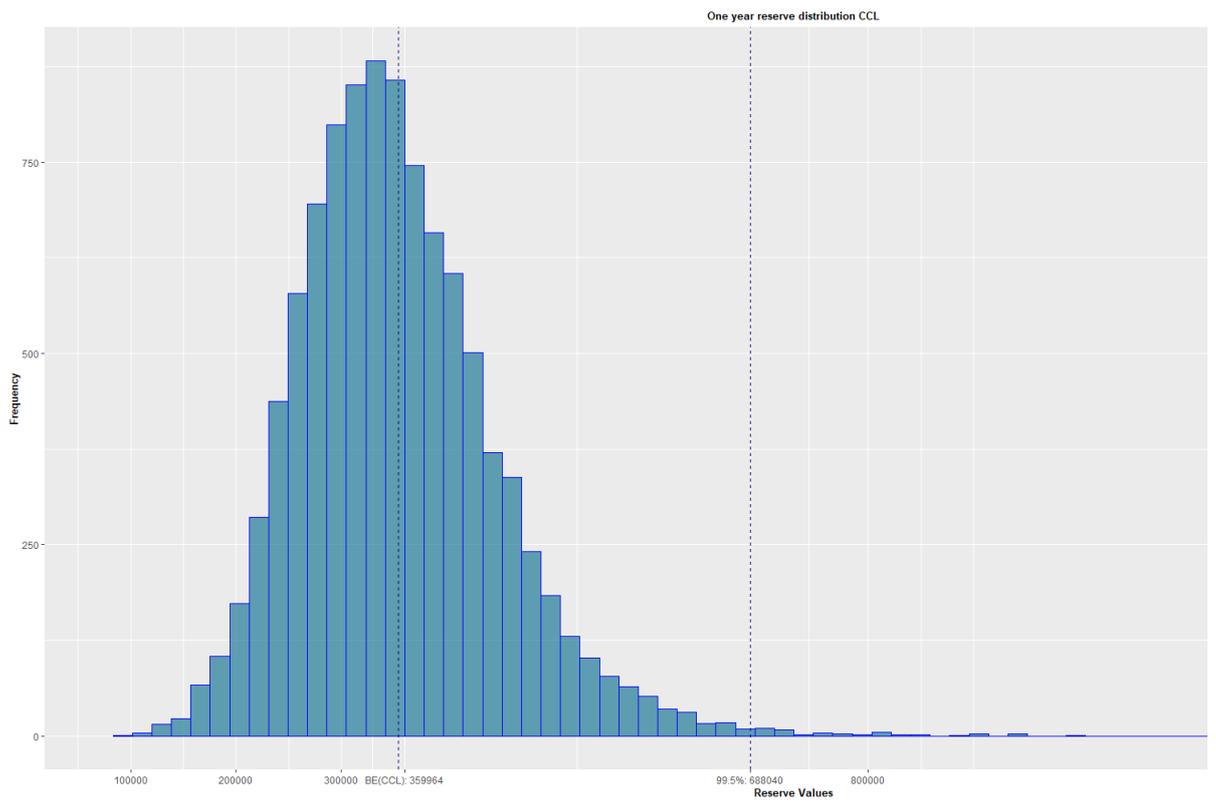


FIGURE 2 – LoB A : Distribution des réserves vision 1 an

Summary

The Correlated Chain Ladder (CCL) is, on the first hand, a reserving methodology that estimates the missing cumulative payment of a reserving triangle $D^{(t)} = (C_{i,j})_{1 \leq i \leq t, 1 \leq j \leq t-i+1}$, and particularly, the ultimate amounts $(C_{i,n})_{2 \leq i \leq t}$.

The likelihood function of the model is supposed to represent the cumulative payments with a lognormal law with parameters that are calibrated with random parameters as well. These parameters initially follow a law called *prior* and the observation in the triangle will allow these law to evolve into *posterior*. This is the concept of bayesian inference which can be summarized with the equation : Posterior \propto Likelihood \times Prior. These *posterior* will then allow us to simulate output distributions representing the cumulative payments.

The model has a "cross-classified" structure. That means that the cumulative payment to be predicted can be written as $C_{ij} = f(\alpha_i, \beta_j)$ with α the parameters containing informations linked to the accident years and β the parameters containing informations linked to the development years.

The model consists in simulating the cumulative payments with a log-normal distribution depending on some parameters : $C_{ij} \sim \log \mathcal{N}(\mu_{ij}, \sigma_j)$ for $i, j \in [1 : t]$. The observations of the C_{ij} enable to determine the *posterior* of the parameters.

The details of the model *prior* are the following (with B_i representing the premium amount of accident year i) :

- $\mu_{1j} = \alpha_1 + \log(elr_1) + \beta_j$
- $\mu_{ij} = \alpha_i + \log(elr_i) + \beta_j + \rho[\log(C_{i-1,j}) - \mu_{i-1,j}]$
- $\alpha_i \sim \mathcal{U}[\log(B_i) - \epsilon_i, \log(B_i) + \epsilon_i]$ the *prior* of the parameter depending on the accident year.
- $elr_i \sim \log \mathcal{N}(\gamma_i, \delta_i)$ the *prior* of the "expecting loss ratio" of the accident year i .
- $\beta_j \sim \mathcal{U}[-\eta_j, 0]$ with the constraint $\beta_t = 0$, the *prior* of the parameter depending on the development year.
- $\sigma_j^2 = \sum_{h=j}^t \tau_h$ where $\tau_h \sim \text{Beta}(a, b)$, the *prior* of the volatility which is decreasing with the development of the accident year.

- $\rho \sim \mathcal{U}[-1, 1]$ the *prior* of parameter representing the correlation between accident year.

Considering the complexity of the expressions of the *posterior*, it is necessary to use Markov Chain Monte Carlo (MCMC) algorithms to simulate these distributions. These methods involves the creation and the simulation of Markov process $(X_t)_{t \geq 0}$ whose limit distribution is the target distribution which is the *posterior* to be estimated. A popular and simple methodology that build these Markov chains is the Metropolis-Hastings algorithm. In this paper, because of the complexity of the model, we use extensions of this algorithm which are Hamiltonian Monte Carlo and the No-U Turn Sampler. These extension allow to accelerate the convergence of the Markov Chain toward the target distribution.

Once the simulation of the Markov chain and the *posterior* is performed, we can simulate the uncertainty around the reserves for a line of business (LoB), that means the difference between the ultimate amount and what has already been paid : Reserves = Ultimate – Payments.

Two LoBs from the AXA XL entity have been studied. The first, for privacy purpose, will be named "LoB A" (Line of Business A) and it contains an historic of 10 accident years. The graph 3 shows the simulation of ultimate reserves for the LoB with an example of initial *prior*. We can observe a lognormal distribution which is coherent with the hypothesis of the model. The second LoB is called "LoB B", which only contains 5 accident years and on which the CCL methodology has been tested in order to see how the expert judgment, reflected in the *prior*, will allow a relevant estimation of the reserve distribution.

In order to respect the Solvency II guidelines, and especially the quantitative pillar which requires an estimation of the SCR on a 1-year view, it is necessary to extend the CCL to the estimation of the reserve deviation from two consecutive calendar years. In other words, we are trying to estimate the volatility around Boni-Mali for a given triangle.

The idea of this 1-year extension is to simulate the last diagonale which corresponds to the payment of the calendar year $t + 1$ and to update the *posterior* distribution thanks to the new informations provided by the new simulated diagonal, and then to redo an estimation of the reserve distribution which will provide the Boni-Mali distribution. The latest distribution, with the same parametrization of the LoB A, is visible on the figure 4. The lognormality hypothesis is still present.

The main benefits of the CCL methodology that the paper has revealed are :

- The obtention of a complete and precise distribution to represent the ultimate and one-year reserve risk

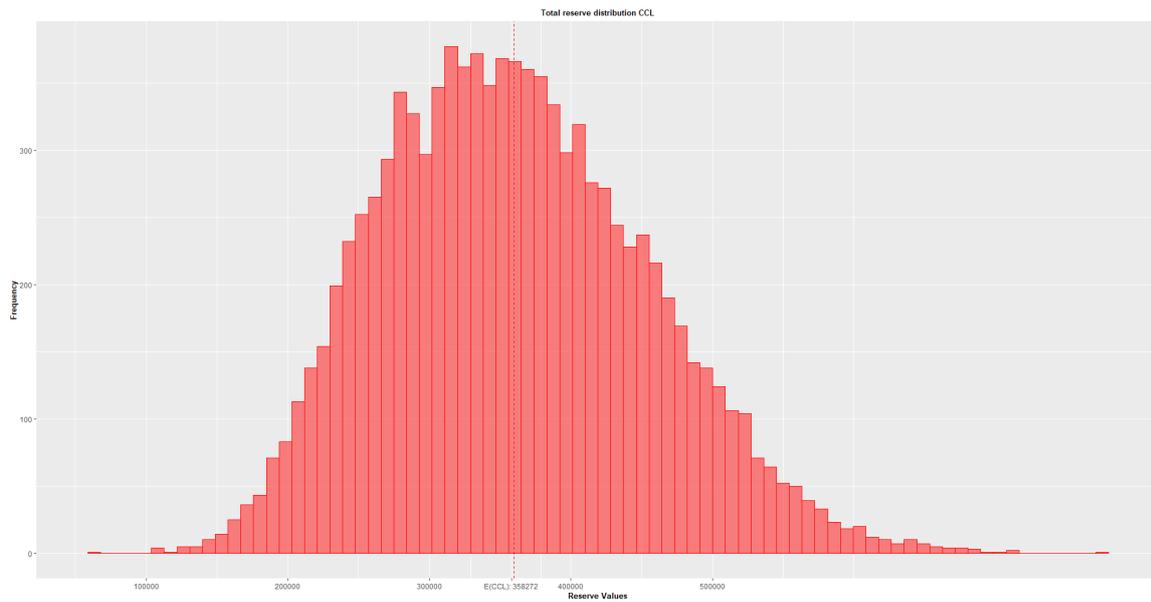


FIGURE 3 – LoB A : Ultimate reserves distribution

- The possibility of an inclusion of the expert judgments inside the parametrization to artificially increase the volume of data
- An easy manipulation of the parametrization that will ensure prudence in the calculation

However, there are some drawbacks :

- The methodology is time-consuming
- The number of parameters to calibrate is very high and may require a lot of justifications

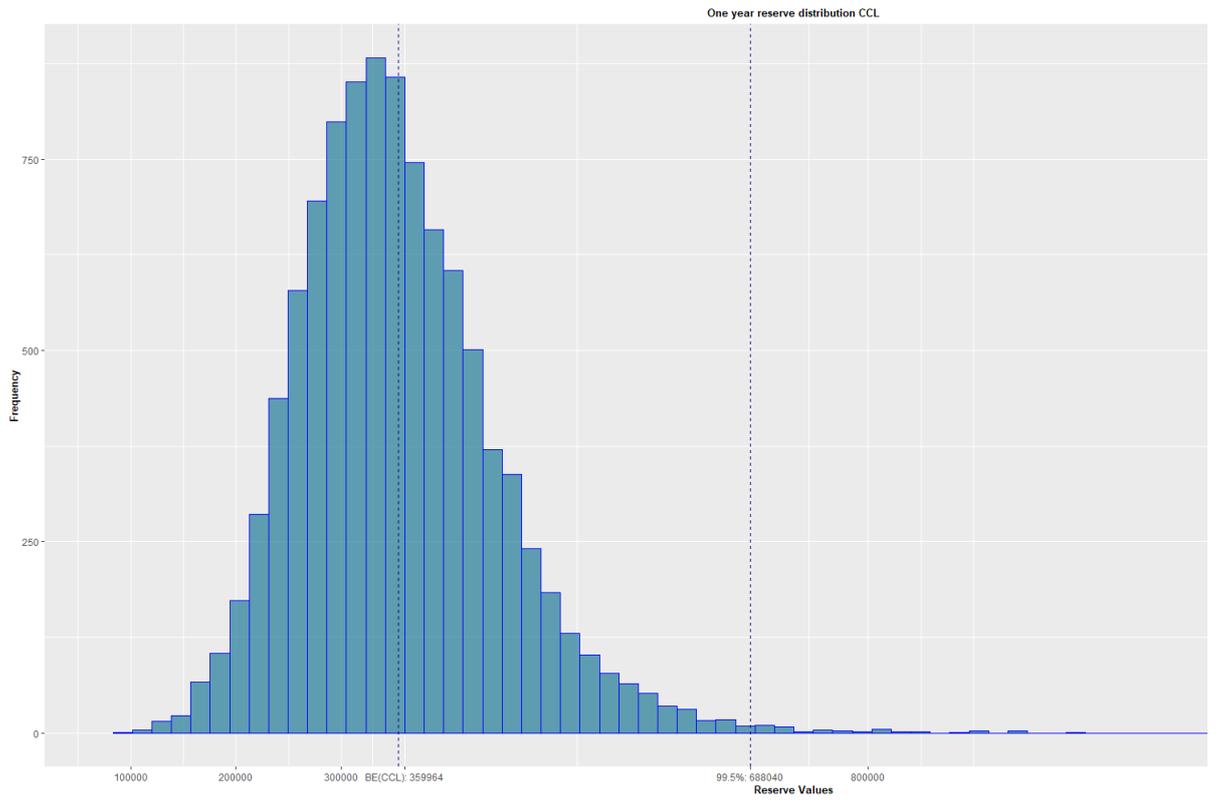


FIGURE 4 – LoB A : One-year view reserves distribution

Table des matières

Résumé	i
Abstract	iii
Remerciements	v
Note de synthèse	vii
Summary	xi
Introduction	1
1 Le risque de réserve en assurance non-vie	5
1.1 Le cadre prudentiel Solvabilité II	5
1.1.1 Objectifs	5
1.1.2 Bref historique de la réglementation	5
1.1.3 Les trois piliers	6
1.1.4 Le Solvency Capital Requirement (SCR)	6
1.2 Provisionnement en assurances non-vie sous les directives Solvabilité II	8
1.2.1 Définition d'une provision technique en assurance non-vie	8
1.2.2 Le SCR réserve	9
1.3 Le Modèle Interne	10
1.3.1 Généralités sur l'évaluation du SCR	10
1.3.2 Le Modèle Interne d'AXA	11
2 Méthodes actuarielles de provisionnement en assurance non-vie	15
2.1 Généralités sur les méthodes de provisionnement	15
2.2 Méthode classique de provisionnement à l'ultime : Modèle de Mack	16
2.2.1 La méthode de provisionnement déterministe du Chain Ladder	16
2.2.2 Extension vers la modèle stochastique de Mack	17
2.3 Méthode classique de provisionnement à 1 an : Modèle de Merz & Wütrich	18
2.4 Limites des méthodologies	21

3	Inférence bayésienne via Monte Carlo Markov Chain	23
3.1	Inférence bayésienne	23
3.1.1	Règle de Bayes	24
3.1.2	Structure hiérarchique	24
3.2	Méthodes de simulations ordinaires	25
3.2.1	Inversion de la fonction de répartition	25
3.2.2	Méthode de rejet	25
3.2.3	L'échantillonnage préférentiel	26
3.3	Markov Chain Monte Carlo	27
3.3.1	Les chaînes de Markov pour l'approximation des distributions	27
3.3.2	L'algorithme de Metropolis-Hastings	28
3.3.3	L'échantillonnage de Gibbs	29
3.3.4	Hamiltonian Monte Carlo	29
3.3.5	Introduction au variant No-U-Turn Sampler	31
3.3.6	Le langage STAN	32
3.4	Décomposition des risques du modèle	32
3.5	Critère de qualité du modèle	33
4	Méthode Correlated Chain Ladder et extension	35
4.1	Historique et contexte de l'étude	35
4.2	CCL à l'ultime	36
4.2.1	Le modèle d'inférence	36
4.2.2	Spécification des paramètres	36
4.2.3	Distribution a posteriori	38
4.3	Extension du modèle CCL à horizon 1 an	39
4.4	Calcul du DIC	41
5	Application du CCL aux triangles de réserve d'AXA XL	43
5.1	A propos du risque de réserve de l'entité AXA XL	43
5.2	Bonnes pratiques de calibration	44
5.3	Étude de la ligne A	45
5.3.1	Preprocessing	45
5.3.2	Paramétrages convergents vers les résultats du modèle de Mack	46
5.3.3	Utilisation des EAxa	54
5.3.4	Étude de la stabilité du modèle	56
5.3.5	Comparaison des critères d'informations	57
5.3.6	Conclusion de l'étude de la ligne A	57
5.4	Étude de la ligne B	58
5.4.1	Preprocessing	59
5.4.2	Quelques essais de paramétrages à forte volatilité	61
5.4.3	Paramétrages retenues : à faible volatilité	62
5.4.4	Conclusion de l'étude de la ligne B	63
	Conclusion	67

A	Justification des méthodes de simulations	69
A.1	Méthode de rejet	69
A.2	Échantillonnage préférentiel	70
A.3	Metropolis Hastings	71
B	Exemple d'implémentation de MCMC	73
C	Implémentation du modèle CCL dans STAN	77
D	Diagnostic des MCMC	81
D.1	Potential Scale Reduction \hat{R}	81
D.2	Effective sample size	81
D.3	Bayesian Fraction of Missing Information (BFMI)	81
E	Stabilité des modèles de Mack et M&W	83

Table des figures

1	LoB A : Distribution des réserves à l'ultime	ix
2	LoB A : Distribution des réserves vision 1 an	x
3	LoB A : Ultimate reserves distribution	xiii
4	LoB A : One-year view reserves distribution	xiv
1.1	Bilan Economique Solvabilité II	7
1.2	Solvabilité II : les 3 piliers	7
1.3	Illustration du concept de Boni-Mali	10
1.4	Structure hiérarchique des risques en Formule Standard	12
1.5	Structure hiérarchique des risques en Modèle Interne à AXA	13
5.1	Visualisation du développement du triangle de la LoB A	46
5.2	LoB A : Distribution des réserves à l'ultime	49
5.3	LoB A : Distribution des réserves vision 1 an	50
5.4	LoB A : Distribution a posteriori du paramètre de corrélation ρ (1)	51
5.5	LoB A : Distribution a posteriori du paramètre de corrélation ρ (2)	52
5.6	LoB A : Distribution a posteriori du paramètre de corrélation ρ , avec utilisation des EAXa	56
5.7	Visualisation du développement du triangle LoB B	60
5.8	LoB B : Loi a posteriori du paramètre α_3 du premier paramétrage à forte volatilité	62
5.9	LoB B : CCL à l'ultime	64
5.10	LoB B : CCL à 1 an	64
5.11	LoB B : Distribution <i>a posteriori</i> du paramètre ρ	65
5.12	Loi log-normales de différents paramètres de volatilité	66
B.1	Autocorrelogramme dans la simulation de la loi normale	74
B.2	Simulation de la loi normale grâce à Metropolis-Hastings	75

Introduction

La groupe internationale français AXA, comme toute entreprise spécialisée dans l'assurance et la réassurance, doit respecter des normes de solvabilité, c'est-à-dire de capacité à rembourser ses clients. En Europe, la norme réglementaire est Solvabilité II et le pilier quantitatif de cette norme peut se formuler de la façon suivante : "L'entreprise exerçant une activité d'assurance doit avoir un niveau de fonds propre nécessaire pour être certaine à 99.5% de pouvoir rembourser tous les sinistres qui pourraient avoir lieu au cours de l'année". Parmi les risques pouvant impacter la solvabilité, il y a le risque en assurance non-vie (ou Property and Casualty, ou P&C) qui concerne les dommages aux bien et la responsabilité. Et plus précisément, ce qui va nous intéresser dans le mémoire est le risque de réserve, c'est-à-dire le risque d'un mauvais provisionnement pour les sinistres futures dû à une mauvaise estimation. AXA a acheté la société bermudienne XL Catlin en 2018, et cette dernière est devenue une entité d'AXA appelée AXA XL. Elle est particulièrement exposée au risque P&C, principalement sur le risque de catastrophe (naturelle ou d'origine humaine) mais aussi au risque de réserve.

Dans le cadre de Solvabilité II, le risque de réserve se traduit par un risque d'avoir une trop grande différence dans l'estimation du provisionnement entre deux années consécutives. Le calcul de ce risque est effectué grâce à des triangles de provisionnement qui vont servir de données d'entrée pour l'estimation. Actuellement, le calcul des provisions, et plus précisément la volatilité estimée des boni-mali (boni dans le cas d'un surprovisionnement et mali à l'inverse), est effectué avec des méthodes inspirées de celles de Merz and Wütrich (M&W). Par abus de langage, on dit que l'on estime la volatilité ou le risque de réserve "à 1 an" pour parler de la volatilité des boni-mali, par opposition du risque de réserve "à l'ultime" pour désigner la volatilité des réserves provisionnées. Le modèle de M&W, initialement introduit en 2008 dans [17], est une extension du modèle de Mack, introduit plus tôt en 1993 dans [11], qui lui avait pour objectif d'estimer des réserves et une volatilité sur ces réserves à l'ultime à partir d'un triangle de provisionnement. Il serait intéressant de les comparer à d'autres méthodes stochastiques et en particulier les méthodes basées sur l'inférence bayésienne. Ces méthodes consistent à créer une fonction liant la variable aléatoire représentant les réserves estimées aux observation d'un triangle de provisionnement, et cette fonction va contenir des paramètres eux-même aléatoires qui auront au début des distributions dites *a priori* puis qui changeront après apprentissage des données d'observation pour devenir des distributions *a posteriori*.

On peut trouver diverses publications scientifiques concernant les méthodes de provisionnement stochastique par différents modèles bayésiens mais principalement pour le provisionnement à l'ultime. On peut citer par exemple les deux éditions des travaux de Glenn Meyers [13] et [14] qui proposent une multitude de modèles utilisant les triangles de paiement cumulés *Incurred* ou *Paid* ainsi que les primes reçues. Aussi il y a l'article [8] qui propose des distributions finales à valeurs extrêmes des réserves, et plus précisément des distributions-t généralisées qui peuvent être des lois de Student, uniforme ou exponentielles selon les paramètres p et q estimés. Et enfin [7] propose des modèles utilisant simultanément les triangles *Incurred* et *Paid* et une extension bayésienne de M&W qui fait rejoindre les analyses des deux triangles en une seule estimation de distribution de réserve à l'ultime. L'idée principale de notre étude est de comprendre les modèles et essayer de les adapter pour une vision à un an, dans le but de répondre au besoin des directives Solvabilité II. Pour cela, nous allons nous concentrer sur la méthode Correlated Chain Ladder proposée par Glenn Meyers dans [13] et son extension proposée dans [6].

L'intégration d'AXA XL dans le modèle interne d'AXA a eu comme conséquence pour le Group Risk Management de revoir la manière d'estimer le risque de réserve au sein des entités du modèle interne en général. Cela est dû au périmètre très varié de cette nouvelle entité et aux particularités des triangles de provisionnement, car cette entité est implantée dans de nombreux pays comme les Etats-Unis, la France, l'Angleterre ou l'Australie et les risques assurés sont principalement des risques à fort impact (aviation, aérospaciale, gestion de crise ...). L'intégration de AXA XL dans le module "Reserve Risk" du modèle interne a engendrée une différente utilisation des calibrages. En effet, cet intégration a développé l'utilisation de Tail Factor ([10]), de la méthode de Bornhuetter Fergusson ([9]) ou de différentes corrélations entre années d'accident, et aussi plus de tests de sensibilité à but comparatifs avec des méthodes Bootstrap ([5]). Il serait intéressant d'essayer d'inclure d'autres méthodes, que ce soit comme méthode officielle de calcul ou comme test de sensibilité à but comparatif pour valider les résultats d'autres méthodologies.

Les problématiques soulevées sont les suivantes :

1. Comment calibrer le modèle bayésien du Correlated Chain Ladder pour représenter de la façon la plus fiable la réalité d'une activité ?
2. Dans une optique Solvabilité II, comment modéliser et estimer les pertes extrêmes des distributions ?
3. Quelles sont les métriques permettant de valider le modèle et dans quelles mesures ce nouveau modèle est plus performant que les autres déjà existant ?

Ainsi, ce mémoire sera consacré pour commencer à l'explication des simulations de Markov Chain Monte Carlo, nécessaire à la méthodologie Correlated Chain Ladder (que l'on dénommera CCL par la suite), puis à la description de la méthodologie CCL et de

son extension à horizon 1 an, et pour finir sur le but recherché, c'est-à-dire l'utilisation de cette dernière méthodologie à certains triangles d'AXA XL. L'originalité du mémoire est qu'il va étudier le modèle bayésien CCL sur les données réelles qui ne sont pas resimulées. Ces données, qui sont des triangles de réserves, auront généralement une profondeur d'historique assez faible, ce qui va rendre intéressant la méthode CCL qui permet d'inclure des jugements d'expert dans la calibration via le choix des distributions *a priori* du modèle.

Chapitre 1

Le risque de réserve en assurance non-vie

Dans ce chapitre, nous allons expliquer dans quel cadre s'inscrivent les méthodes bayésiennes qui font l'objet du mémoire. Elles servent à l'estimation du risque de réserve au sein du modèle interne P&C d'AXA, dans le but de l'estimation du SCR (du risque de réserve puis du groupe) pour respecter la réglementation Solvabilité II. Les notions de Solvabilité II, SCR, risque de réserve et le modèle interne d'AXA seront présentés ici.

1.1 Le cadre prudentiel Solvabilité II

1.1.1 Objectifs

La Directive Européenne Solvabilité II a pour ambition de donner aux assureurs les moyens de mieux garantir leur solvabilité en imposant des exigences quantitatives et qualitatives. Cela permet notamment de préserver le système financier et protéger les clients, assurés et bénéficiaires des contrats d'assurance face à l'incertitude de l'environnement. Cette directive a aussi pour but de construire un marché unique européen de l'assurance en uniformisant les réglementations au sein de l'Union Européenne.

1.1.2 Bref historique de la réglementation

L'évolution de l'environnement économique (concurrence accrue, crises financières, risques émergeant,...) dans laquelle exercent les compagnies d'assurance a rendu nécessaire l'évolution des normes réglementaires qui existaient (Solvabilité I). Dans cet objectif, la commission européenne a procédé à la révision et au durcissement desdites normes avec le nouveau référentiel Solvabilité II, entré en vigueur au 1er janvier 2016 après avoir été adopté en Novembre 2009. Cette directive a été créée dans la continuité des principes énoncés lors du comité de Bâle II pour les banques.

Solvabilité I prévoyait une marge de solvabilité déterminée en fonction de pourcentages sur les primes et les sinistres, et donc la réglementation des assurances passe à des règles plus complexes intégrant le risque. De plus, Solvabilité I n'était pas totalement harmonisée au niveau européen. Enfin, l'ancienne approche n'abordait pas la question de la gouvernance des risques.

1.1.3 Les trois piliers

La norme Solvabilité II repose sur 3 piliers fondamentaux :

- Le pilier I : les exigences quantitatives
- Le pilier II : les exigences qualitatives et les activités de contrôle
- Le pilier III : les exigences en matière d'information prudentielles et de publication

Le pilier I consiste à exiger une évaluation "market-consistent" des actifs et des passifs avec notamment au sein du passif les provisions techniques. Les provisions techniques sont composées du Best Estimate qui est la valeur marché des provisions et d'une marge de risque (ou Risk Margin). Additionné aux provisions techniques, il y a deux exigences en capital : le Solvency Capital Requirement (SCR) et le Minimum Capital Requirement (MCR). Le MCR représente le niveau minimum de fonds propres en dessous duquel l'intervention de l'autorité de contrôle sera automatique. Le SCR représente le capital cible nécessaire pour absorber le choc provoqué par un risque majeur. Un récapitulatif du bilan économique, extrait de [1], est visible sur le schéma 1.1.

Le pilier II appelle à un renforcement de la gouvernance, du contrôle interne et de la gestion des risques au sein de l'entreprise, ainsi qu'une auto-évaluation des besoins de capital (Own Risk and Solvency Assessment ou ORSA). Aussi, l'autorité de contrôle aura le pouvoir de contrôler la qualité des données et des procédures d'estimation et exiger une marge de solvabilité complémentaire.

Le pilier III concerne la communication d'informations au public et aux autorités de contrôle. Il vise à harmoniser au niveau européen les informations publiées par les organismes d'assurance ainsi que celles remises aux superviseurs. En plus des régulateurs, ces informations seront communiquées au grand public (actionnaires et analystes) qui pourra également juger si l'analyse effectuée est fidèle à la réalité.

La schéma 1.2, extrait du site de la mutuelle AMPLI, résume les 3 piliers de la norme SII.

1.1.4 Le Solvency Capital Requirement (SCR)

Le Solvency Capital Requirement, dit SCR, correspond à l'exigence en fond propres à respecter par l'organisme d'assurance sous peine d'intervention de l'ACPR (Autorité de

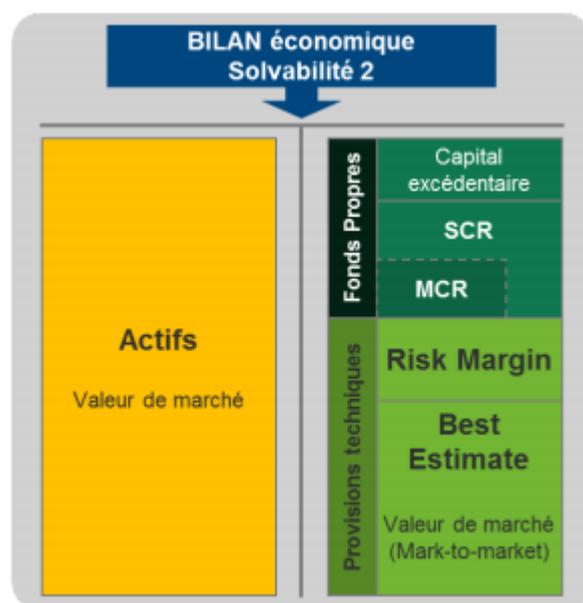


FIGURE 1.1 – Bilan Economique Solvabilité II

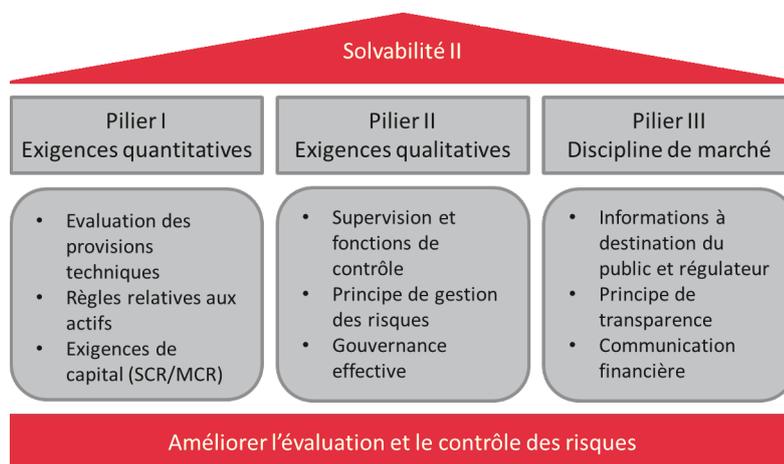


FIGURE 1.2 – Solvabilité II : les 3 piliers

contrôle prudentiel et de résolution) qui est le régulateur français des organismes d'assurance et d'une surveillance renforcée de cette dernière. Il s'agit d'une mesure de risque, qui s'évalue comme le niveau de fonds propre nécessaire pour que l'organisme ai une probabilité de défaillance inférieure à 0.5%, ou autrement dit, pour faire face à une perte extrême bicentenaire (un perte pouvant survenir tous les 200 ans).

Si on formalise le problème en considérant la variable aléatoire X représentant le montant de perte de l'organisme d'assurance, le SCR de cette variable correspond à :

$$\text{SCR}(X) = \text{VaR}_{99.5\%}(X)$$

La notation VaR_α correspond à la *Value at Risk* d'ordre α qui est le quantile d'ordre α du montant de perte défini comme suit (avec F_X la fonction de répartition de la variable aléatoire X) :

$$\text{VaR}_\alpha(X) = \min\{x \in \mathbb{R}, F_X(x) \geq \alpha\}$$

Le SCR est calculé dans le but de définir un niveau de fond propre sur le périmètre d'AXA Group tout entier. Mais il est intéressant d'avoir des mesures du SCR à des niveaux plus bas, comme le SCR calculé sur le montant de perte d'entités comme AXA France, ou bien le SCR de risques spécifiques comme le risque P&C (non-vie) ou plus précisément le risque de réserve qui sera l'objet de notre étude. Nous verrons que généralement, les organismes d'assurances calculent des SCR à des faibles niveaux puis agrègent les résultats pour au final aboutir à une valeur du SCR de toutes les pertes au niveau du groupe.

1.2 Provisionnement en assurances non-vie sous les directives Solvabilité II

1.2.1 Définition d'une provision technique en assurance non-vie

Les provisions techniques sont inscrites au passif du bilan d'une compagnie d'assurance. Ces provisions ont pour origine l'inversion du cycle de production des compagnies d'assurances. En effet, dans une industrie classique, la fabrication du produit est réalisée avant la vente et donc le coût de ce produit est connu. Alors que dans l'industrie de l'assureur, la vente (correspondant à la signature du contrat) vient avant la fabrication du produit, c'est-à-dire le paiement du sinistre dont le montant est inconnu et qu'il faut provisionner.

Les provisions pour sinistres IARD correspondent :

- d'une part aux charges à prévoir pour faire face à la sinistralité déclarée mais non encore réglée et à la sinistralité non encore déclarée mais prévisible des contrats en

cours. Par exemple, dans le cas d'un sinistre automobile déjà déclaré, aux sommes qui seront probablement versées aux bénéficiaires

- d'autre part à une anticipation des prestations futures auxquelles l'assureur devra faire face lorsqu'un engagement prendra effet. Par exemple, l'anticipation des rentes futures dans le cas d'un contrat d'arrêt de travail

1.2.2 Le SCR réserve

Le risque de réserve porte sur l'incertitude du provisionnement des réserves faite par l'organisme d'assurance. Il peut être défini "à l'ultime" comme l'incertitude sur la valeur des réserves qui ont été provisionnées, directement liées aux pertes finales (à l'ultime) de toutes les sinistres passés. Il peut aussi être défini comme l'incertitude sur la déviation des réserves entre deux années calendaires, et c'est ce que l'on appelle le risque de réserve "1 an". On dit "à 1 an" car une différence d'estimation de la charge ultime entre deux années consécutives peut être vu comme une perte survenu au cours de l'année. Ce risque de réserve à 1 an va particulièrement nous intéresser car il est celui qu'il faut évaluer pour vérifier la solvabilité du groupe.

Les déviations des réserves entre deux années calendaires sont appelées "Boni-Mali". Il s'agit d'un Boni lorsque l'écart de l'estimation de la charge ultime (et donc les provisions) à l'année $N + 1$ est plus faible que celle faite à l'année N et un Mali dans le cas contraire. Cela se formalise de la manière suivante, à l'instant $N + 1$ et en considérant les réserves et paiements des années d'accident des années avant l'année N (inclusive) :

$$\begin{aligned} BM(N + 1) &= \text{Estimation Charge Ultime } (N) - \text{Estimation Charge Ultime } (N + 1) \\ &= \text{Réserves } (N) - \text{Réserves } (N + 1) - \text{Paiement } (N + 1) \end{aligned}$$

La figure 1.3, extraite des documentations d'AXA, illustre comment se manifestent les Boni-Mali au fil des années calendaires.

Le SCR Réserve à l'année N se définit comme la différence entre la VaR 99.5% des réserves et du paiement à l'année $N + 1$ et les réserves moyennes estimées à l'année N . En notant R_t et P_t les variables aléatoires correspondant respectivement à la valeur du provisionnement et de paiement à l'instant t , on a :

$$\text{SCR}_{Reserves}(N) = \text{VaR}_{99.5\%}(R_{N+1} + P_{N+1}) - \mathbb{E}(R_N)$$

Plus formellement, en notant comme nous le ferons dans les sections suivantes $D^{(t)}$, l'ensemble des informations disponibles à l'instant t , le SCR correspond à :

$$\text{VaR}_{99.5\%}(-BM(N + 1)) = \text{VaR}_{99.5\%}[\mathbb{E}(R_{N+1} + P_{N+1}|D^{(N+1)}) - \mathbb{E}(R_N|D^{(N)})]$$

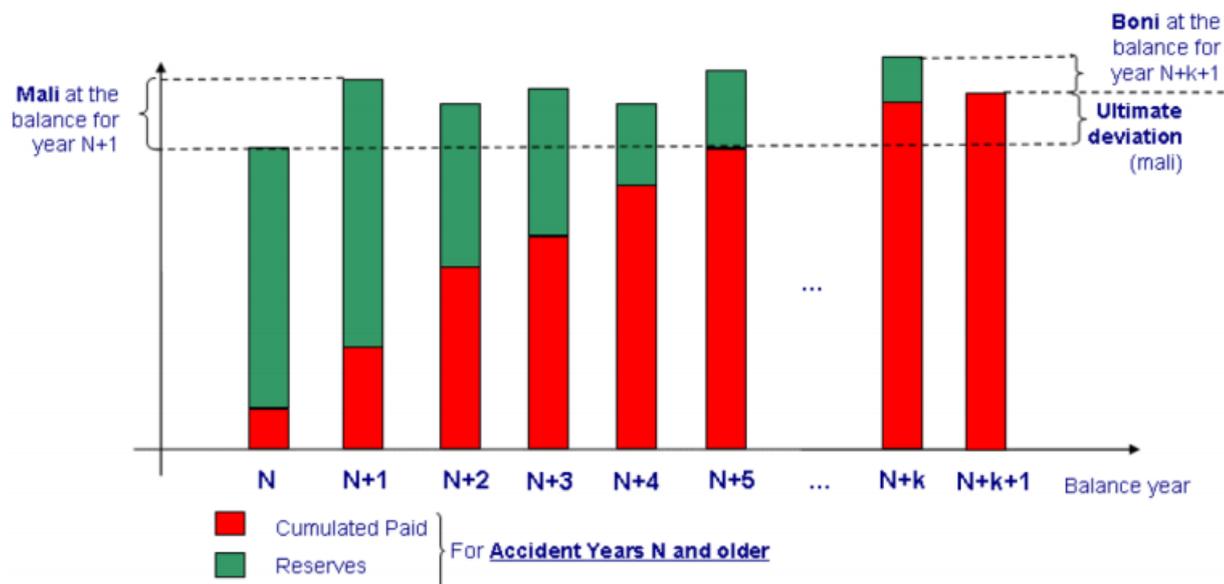


FIGURE 1.3 – Illustration du concept de Boni-Mali

Le premier terme $\mathbb{E}(R_{N+1} + P_{N+1}|D^{(N+1)})$ est aléatoire car l'information $D^{(N+1)}$ n'est pas disponible à l'instant N . Le deuxième terme est déterministe. Ainsi :

$$\text{SCR}_{Reserves}(N) = \text{VaR}_{99,5\%}[\mathbb{E}(R_{N+1} + P_{N+1}|D^{(N+1)})] - \mathbb{E}(R_N|D^{(N)})$$

1.3 Le Modèle Interne

1.3.1 Généralités sur l'évaluation du SCR

Trois approches sont possibles pour évaluer le SCR :

- La formule standard
- Les paramètres spécifiques (USP)
- Le Modèle Interne

La formule standard est l'option par défaut. C'est celle retenue par l'essentiel des acteurs dans l'assurance. Elle est simple et peu coûteuse à mettre en place. Aussi, elle ne nécessite pas de justification auprès des régulateurs (ACPR au niveau national ou EIOPA au niveau européen). Cependant, cette approche peut se révéler coûteuse en capitale du fait des paramètres, parfois trop prudent, proposés par les superviseurs.

Les paramètres USP est une approche intermédiaire entre la formule standard et le modèle interne. Il s'agit de garder le schéma de la formule standard mais en remplaçant les paramètres de marché imposés par les régulateurs par des paramètres spécifiques basés sur le portefeuille de l'organisme. Il s'agit d'un compromis intéressant car moins lourd à faire valider qu'une approche modèle interne.

Le modèle interne est l'option la plus ambitieuse. C'est une méthode propre à l'organisme pour calculer le SCR tout en prenant en compte les particularités de l'entreprise et aboutit généralement à une nette réduction de ce dernier. En revanche, l'utilisation de cette approche est accompagnée de nombreux contrôles de la part des régulateurs. L'importance des coûts de déploiement et de maintien d'un modèle interne pour un organisme d'assurance fait que seul un faible nombre d'acteurs (essentiellement des groupes internationaux comme AXA) ont décidé de l'adopter.

1.3.2 Le Modèle Interne d'AXA

Dans cette section nous allons décrire brièvement comment se structure le calcul du SCR du groupe AXA avec une attention particulière sur le risque d'intérêt dans ce mémoire, à savoir le risque de réserve.

1.3.2.1 Structure hiérarchique d'agrégation

Le SCR se calcul en formule standard avec la structure hiérarchique présente sur la figure 1.4 (que l'on peut apercevoir également sur le site officiel de l'EIOPA) et les SCR sont agrégés et corrélés pour donner le SCR de base avec la formule dite "Var-Covar" :

$$\text{SCR} = \sqrt{\sum_{i,j=1}^n \text{Corr}(i,j) \times \text{SCR}_i \times \text{SCR}_j}$$

Dans la formule précédente, les SCR peuvent représenter les différents SCR de chaque hiérarchie (SCR de base, Santé, Non-Vie ...). Cette formule est juste dans le cadre d'une hypothèse de loi elliptiques pour représenter les sous-risques.

Enfin, ce SCR de base est ajusté avec le SCR des risques opérationnels et d'autres ajustements :

$$\text{SCR Final} = \text{SCR de base (BSCR)} + \text{SCR Opérationnel} - \text{Ajustements}$$

Dans le modèle interne d'AXA on parle de STEC (désignant le Short Term Economic Capital) au lieu du SCR mais la signification est la même. La structure hiérarchique du modèle interne d'AXA présente quelques similarités avec celle de la Formule Standard (voir figure 1.5) mais a une structure légèrement différente et une méthode d'agrégation différente pour certains sous-risques. Sans rentrer dans les détails, ce qui est décrit sur

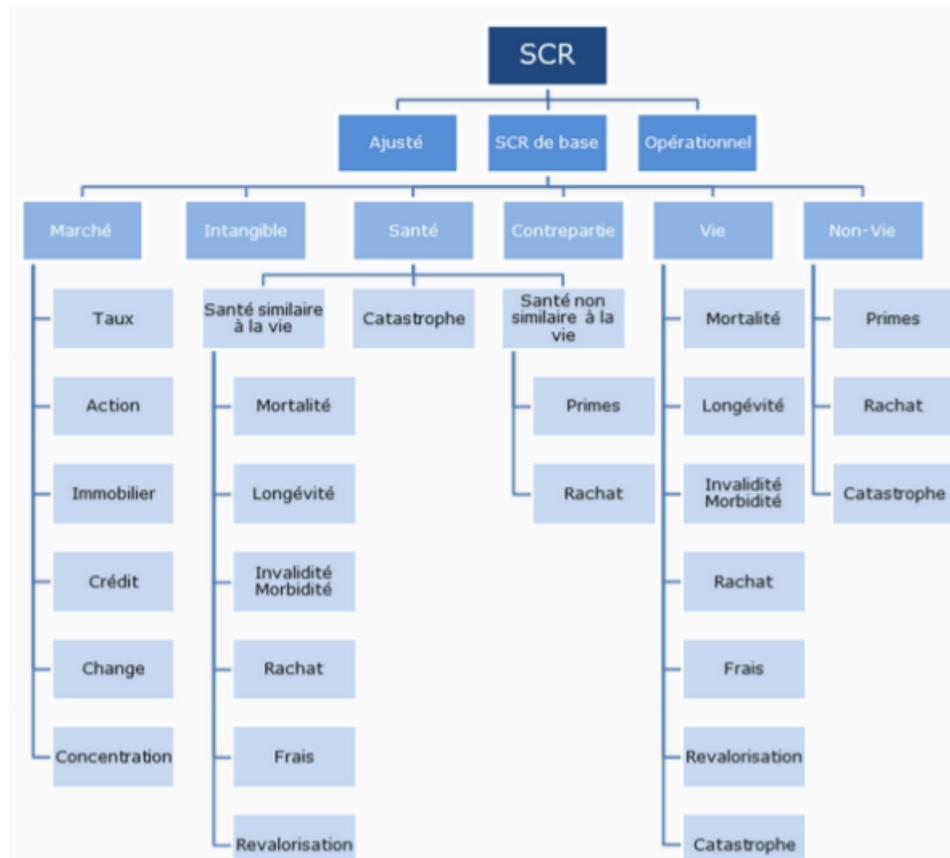


FIGURE 1.4 – Structure hiérarchique des risques en Formule Standard

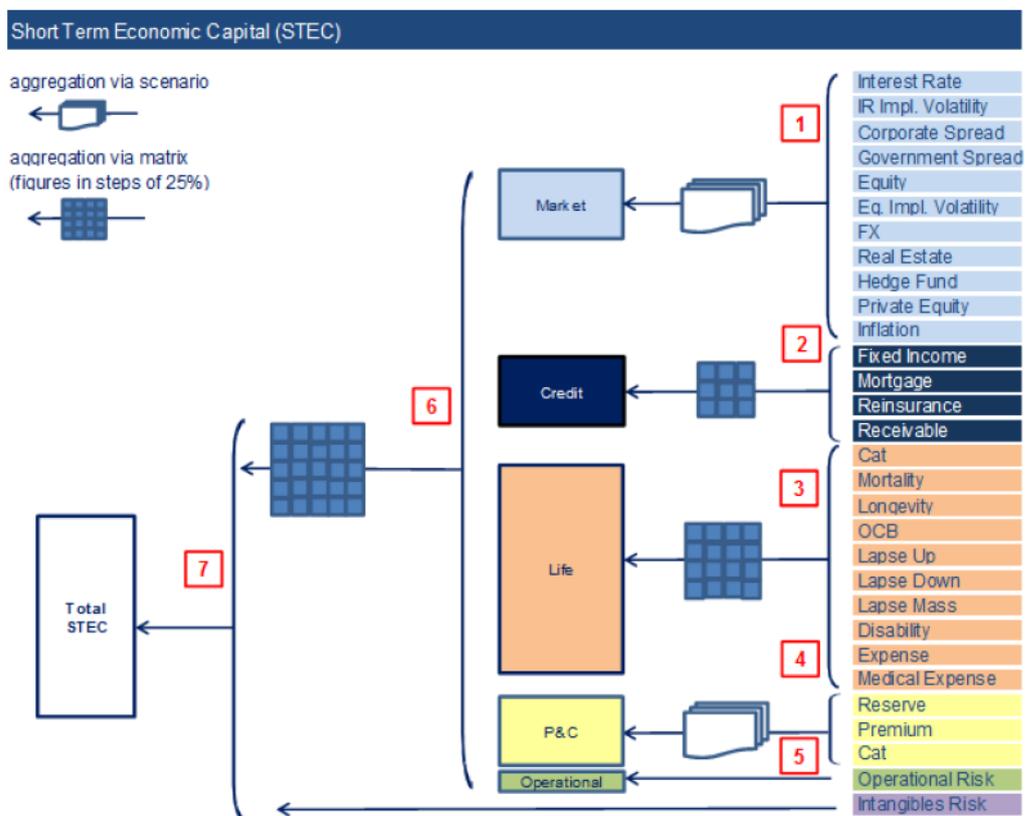


FIGURE 1.5 – Structure hiérarchique des risques en Modèle Interne à AXA

le schéma comme "aggregation via matrix" correspond à une méthode Var-Covar décrite plus haut. La méthode "aggregation via scenario" désigne le fait de simuler une distribution de perte pour les sous-risques et agréger ces vecteurs de pertes en les ordonnant pour créer de la corrélation et puis en les sommant.

1.3.2.2 Évaluation du SCR réserve dans le modèle interne d'AXA

Intéressons nous maintenant au risque qui va être au coeur du mémoire c'est à dire le risque de réserve, qui est, dans le modèle interne d'AXA, un des trois sous-risques du risque P&C.

Il y a plusieurs étapes d'agrégation pour aboutir à la distribution de perte réserve du groupe :

1. L'agrégation des risques de réserve des LoB (Line of Business) au sein d'une entité : on calcul le SCR de l'entité en utilisant la formule Var-Covar

$$SCR_{ent} = \sqrt{\sum_{i,j=1}^n \text{Corr}(LoB_i, LoB_j) \times SCR_{LoB_i} \times SCR_{LoB_j}}$$

Puis on crée un vecteur de distribution de perte réserve de l'entité de loi log-normale passant par la moyenne des réserves et le SCR de l'entité.

2. L'agrégation des risques de réserve inter-entité : il s'agit d'un réordonnement des vecteurs de pertes réalisées grâce à des copules (outil mathématique permettant de créer des lois uniformes multivariées) et qui permet d'introduire une corrélation entre ces vecteurs de pertes réserve d'entités qui seront ensuite sommés pour avoir un vecteur de pertes réserve groupe.

Dans ce mémoire, nous allons voir comment évaluer la distribution de perte réserve et le SCR réserve d'une ligne d'activité. Dans la Formule Standard, on a :

$$\text{SCR}_{\text{Reserve,LoB}} = \text{Best Estimate} \times 3 \times \sigma_{\text{LoB}}$$

Le facteur σ_{LoB} mesure la volatilité des Boni-Mali, et l'EIOPA propose des valeurs par défaut de ces paramètres. Les USP proposent de calculer des paramètres plus adaptés à la ligne d'activité.

Le modèle interne d'AXA commence d'abord par définir un nouvel ensemble de lignes d'activité qui n'est pas forcément le même qu'en Formule Standard. Ensuite, le SCR est généralement évalué à partir des volatilités calculées à partir de triangles de provisionnement et la méthode de Merz & Wütrich que l'on verra dans la section suivante. Dans certains cas un peu moins fréquents, les volatilités peuvent être évaluées grâce à une méthode Benchmark en utilisant des volatilités de triangles publics (dans le cas d'un faible volume de données) ou avec la méthode dite de Borhuetter-Fergusson version stochastique.

Le calcul de la distribution réserve pour une ligne d'activité se fait en plusieurs étapes qui sont décrites plus en détails dans la section 5.1 :

1. Estimation de provisions Best Estimate grâce à des méthodes dossier-dossier, Chain Ladder ou autre jugements d'experts incluant parfois une marge de prudence. Ces provisions sont appelés EAXa.
2. Calcul de volatilité par méthode quantitative actuarielle (M&W) ou par méthode Benchmark.
3. "Fitting" d'une loi log-normale de moyenne égale aux EAXa et de volatilité égale à celle calculée à l'étape précédente pour calculer le SCR et la distribution de perte.

Cette façon de procéder possède plusieurs avantages. Elle permet d'une part d'avoir une granularité plus fine en séparant mieux les lignes d'activité que ce qui est fait dans la formule standard. Et d'autre part elle permet d'inclure une hypothèse de loi log-normale sur la distribution de perte qui est en général plus prudente car elle estime un SCR généralement plus élevé.

Chapitre 2

Méthodes actuarielles de provisionnement en assurance non-vie

Ici, nous allons présenter les méthodes statistiques de provisionnement les plus populaires qui sont les modèles de Mack et de Merz & Wütrich. Elles se basent sur l'utilisation de triangles de réserve. L'idée est de comprendre ce qui est déjà fait par AXA et d'étudier les limites de ces méthodologies. Cela va permettre de justifier la recherche de nouvelles méthodes d'estimation du risque de réserve.

2.1 Généralités sur les méthodes de provisionnement

Les provisions techniques sont destinés à régler intégralement un sinistre. On doit pouvoir indemniser les sinistres survenus avant la date n . Entre le moment de la déclaration du sinistre et sa clôture, le provisionnement peut être révisé et il peut y avoir des paiements. Les méthodes actuarielles se basent sur les cadences de paiements. On estime que les cadences de paiement suivent une certaine régularité qui va permettre de prédire des paiements futurs.

Pour bien visualiser les cadences de paiement et établir des prédictions à partir de celles-ci, on utilise un triangle de provisionnement comme en 2.1. Les colonnes représentent les années d'accident et les lignes les années de développement. Il s'agit d'un triangle de paiement cumulé donc chaque case correspond à l'accumulation des paiements de la même ligne. Par exemple, sur la deuxième ligne et troisième colonne, le montant 19 364 correspond à l'accumulation des paiements des sinistres ayant eu lieu en 2012 pour la troisième année de paiement (donc à la fin de l'année 2014). Aussi, il faut faire la distinction entre les triangles de paiement cumulé dit *Paid*, qui correspondent aux triangles des paiements seuls pour la ligne d'activité, et les triangles *Incurred* qui incluent en plus de ce qui a été payé quelques estimations de la charge sinistre.

AY/DY	1	2	3	4	5	6	7	8	9	10
2011	15 102	25 750	28 400	29 888	35 444	39 379	39 412	40 455	42 567	44 570
2012	9 077	15 078	19 364	24 203	30 816	32 548	35 000	36 281	36 684	
2013	12 118	27 736	36 093	42 520	41 724	44 016	46 991	50 842		
2014	8 187	26 257	38 535	45 788	49 381	54 015	56 823			
2015	36 111	66 995	72 797	75 059	85 428	86 500				
2016	10 812	73 384	74 548	79 243	94 429					
2017	18 666	72 685	89 038	109 861						
2018	21 782	120 196	130 017							
2019	21 415	54 967								
2020	22 616									

TABLE 2.1 – Format d’un triangle de paiement cumulé

Nous allons chercher à estimer :

- Le coût final des sinistres à payer survenus l’année i
- Les provisions pour sinistre à payer
- L’incertitude sur les provisions pour sinistre à payer
- Les changements de l’estimation de la charge ultime, c’est-à-dire les Boni-Mali

2.2 Méthode classique de provisionnement à l’ultime : Modèle de Mack

2.2.1 La méthode de provisionnement déterministe du Chain Ladder

Considérons le triangle théorique de paiement cumulé 2.2 ayant n années d’accident et n années de développement et où on note $C_{i,j}$ le paiement cumulé de l’année d’accident i et de développement j . On note $D^{(n)}$ l’information à la date n contenue dans ce triangle. On fait l’hypothèse que la première année d’accident est entièrement développée et que chaque année d’accident est entièrement développée au bout de n années.

AY/DY	1	2	...	n-1	n
1	$C_{1,1}$	$C_{1,2}$...	$C_{1,n-1}$	$C_{1,n}$
2	$C_{2,1}$	$C_{2,2}$...	$C_{2,n-1}$	
...			
n	$C_{n,1}$				

TABLE 2.2 – Triangle de paiement cumulé

Le Chain Ladder consiste à estimer les $C_{i,j}$ qui ne sont pas disponible avec l’information $D^{(n)}$ grâce à des facteurs de développement, aussi appelés coefficient de passage, f_j obtenus avec la formule, pour j allant de 1 à $n - 1$:

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

Les paiements cumulés non disponible vont alors être estimés, pour $j > n - i + 1$, par :

$$\hat{C}_{i,j} = C_{i,n-i+1} \prod_{h=n-i+1}^{j-1} \hat{f}_h$$

Notamment, on arrive à estimer les paiements à l'ultime $\hat{C}_{i,n}$, puis les réserves par année de survenance $\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i+1}$ puis les réserves total $\hat{R} = \sum_{i=2}^n \hat{R}_i$.

2.2.2 Extension vers la modèle stochastique de Mack

Le modèle de Mack, introduit en 1993 dans l'article [11], est un modèle stochastique qui consiste à construire de la volatilité autour des résultats du Chain Ladder.

Le méthode est fondée sur 3 hypothèses :

- (H1) Les années d'accident sont indépendantes, c'est-à-dire que les $C_{i,j}$ et $C_{k,m}$ sont indépendant pour $i \neq k$
- (H2) Il existe des facteurs f_j tels que $\mathbb{E}[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = f_j C_{i,j}$ pour $1 \leq j \leq n - 1$
- (H3) $\forall i = 1, \dots, n, \forall j = 1, \dots, n-1$, il existe des paramètres σ_j tel que : $\text{Var}[C_{i,j+1}|C_{i,1}, \dots, C_{i,j}] = \sigma_j^2 C_{i,j}$

Il est important de noter que les \hat{f}_j introduis dans la section précédente sont des estimateurs sans biais des f_j présents dans (H2).

Sous ces hypothèses, Mack arrive à obtenir une formule fermée pour la variance de la charge ultime par année d'accident et la charge ultime totale. La première valeur d'intérêt ici est l'erreur quadratique moyenne (ou MSEP pour Mean Square Error of Prediction) de la charge ultime par année de survenance, conditionnée à la connaissance des données passées :

$$MSEP(\hat{C}_{i,n}) = \mathbb{E}[(\hat{C}_{i,n} - C_{i,n})^2 | D^{(n)}]$$

Sachant que $\hat{R}_i = \hat{C}_{i,n} - C_{i,n-i+1}$, on a $MSEP(\hat{R}_i) = MSEP(\hat{C}_{i,n})$.

On a $\sigma_j^2 = \frac{1}{n-j} \sum_{i=1}^{n-j} C_{i,j} (\frac{C_{i,j+1}}{C_{i,j}} - f_j)^2, \forall 1 \leq j \leq n-2$ et $\sigma_{n-1}^2 = \min\{\frac{\sigma_{n-2}^4}{\sigma_{n-3}^2}, \sigma_{n-2}^2, \sigma_{n-3}^2\}$.

Theorem 2.2.1. Les formules fermées obtenues sont, d'une part pour l'incertitude sur le provisionnement par année d'accident :

$$MSEP(\hat{R}_i) = \hat{C}_{i,n}^2 \sum_{j=n-i+1}^{n-1} \frac{\sigma_j^2}{f_j^2} \left(\frac{1}{\hat{C}_{i,j}} + \frac{1}{\sum_{k=1}^{n-j} C_{k,j}} \right), \forall i = 2, \dots, n$$

et d'autre part pour l'incertitude sur le provisionnement total :

$$MSEP(\hat{R}) = \sum_{i=2}^n [MSEP(\hat{R}_i) + \hat{C}_{i,n} \left(\sum_{k=i+1}^n C_{k,n} \right) \left(\sum_{j=n-i+1}^{n-1} \frac{2\sigma_j^2}{f_j^2 \sum_{l=1}^{n-j} C_{l,j}} \right)]$$

Le détail de la démonstration de ces formules est disponible dans l'article original [11] de Mack.

2.3 Méthode classique de provisionnement à 1 an : Modèle de Merz & Wütrich

La méthode de Merz & Wütrich (M&W) est une méthode d'estimation de la variance de la variation calendaire de la charge sinistre ultime estimée. Elle peut être vue comme une extension à 1 an du modèle de Mack car cette dernière s'appuie sur les résultats de la formule fermée de la MSEP de Mack.

Les hypothèses du modèle de M&W sont les mêmes que celles de Mack (c'est-à-dire (H1), (H2) et (H3) avec les notations précédentes), avec toutefois l'ajout de la condition de Markov (H4) : $(C_{i,j})_{j \geq 0}$ est un processus de Markov.

On garde la notation $D^{(n)} = \{C_{i,j}; i \in [1 : n], j \in [1 : n], i + j \leq n + 1\}$ introduite précédemment et on introduit également $D^{(n+1)} = \{C_{i,j}; i \in [1 : n], j \in [1 : n+1], i + j \leq n + 2\} = D^{(n)} \cup \{C_{i,n-i+1}; i \in [2 : n]\}$ correspondant à l'information contenu dans le triangle à la date $t = n + 1$ qui est la même qu'à la date n mais avec une nouvelle diagonale et sans ajouter de ligne car on ne considère que les réserves sur les années passées. Subséquemment, les estimations faites aux dates $t = n$ et $t = n + 1$ vont être différentes et il est donc nécessaire d'introduire les notations suivantes des estimateurs pour faire la différence :

En $t = n$,

$$\hat{f}_j^{(n)} = \frac{\sum_{i=1}^{n-j} C_{i,j+1}}{\sum_{i=1}^{n-j} C_{i,j}} \text{ et } R_i^{(n)} = C_{i,n} - C_{i,n-i}$$

En $t = n + 1$,

$$\hat{f}_j^{(n+1)} = \frac{\sum_{i=1}^{n-j+1} C_{i,j+1}}{\sum_{i=1}^{n-j+1} C_{i,j}} \text{ et } R_i^{(n+1)} = C_{i,n} - C_{i,n-i+1}$$

Nous avons évoqué le concept de Boni-Mali dans les sections précédentes qui est le même que celui de Claim Development Result (CDR) évoqué dans [17]. Nous allons

reprendre les notations de l'article pour plus de cohérence et donner la définition du CDR réel pour l'année d'accident i pour $i \in [2 : n]$ pour l'année calendaire $n + 1$:

$$CDR_i(n+1) = \mathbb{E}[R_i^{(n)}|D^{(n)}] - (P_{i,n-i+1} + \mathbb{E}[R_i^{(n+1)}|D^{(n+1)}]) = \mathbb{E}[C_{i,n}|D^{(n)}] - \mathbb{E}[C_{i,n}|D^{(n+1)}]$$

On a noté ici $P_{i,n-i+1}$ le paiement incrémentale durant l'année calendaire $n + 1$, et ces CDR s'aggrègent pour donner un CDR total égale à $\sum_{i=2}^n CDR_i(n + 1)$.

La formule de $CDR_i(n+1)$ permet de déduire facilement que $\mathbb{E}[CDR_i(n+1)|D^{(n)}] = 0$ ce qui se traduit par le fait qu'avec les informations à la date $t = n$, on ne prévoit ni de faire des Boni, ni des Mali.

Les facteurs de développement f_j n'étant pas connu, le CDR réel n'est pas observable et on définit alors le CDR observable qui n'est autre qu'un estimateur du CDR réel par :

$$C\hat{D}R_i(n + 1) = \hat{R}_i^{(n)} - (P_{i,n-i+1} + \hat{R}_i^{(n+1)}) = \hat{C}_{i,n}^{(n)} - \hat{C}_{i,n}^{(n+1)}$$

Notons qu'avec les connaissances à $t = n + 1$, le paiement cumulé $P_{i,n-i+1}$ est connu et ne se note pas comme un estimateur.

La première quantité évaluée dans l'article [17] est :

$$MSEP_{C\hat{D}R_i(n+1)|D^{(n)}}(0) = \mathbb{E}[(C\hat{D}R_i(n + 1) - CDR_i(n + 1))^2|D^{(n)}]$$

Il s'agit de la MSEP d'un point de vu prospectif de solvabilité. Elle quantifie l'incertitude de considérer les estimations de Boni-Mali comme étant nulles. Pour préserver sa solvabilité, l'organisme doit évaluer les possibles déviations des CDR par rapport à 0. Ainsi cette quantité va présenter un intérêt pour l'évaluation du SCR.

La deuxième quantité évaluée est :

$$MSEP_{CDR_i(n+1)|D^{(n)}}(C\hat{D}R_i(n + 1)) = \mathbb{E}[(C\hat{D}R_i(n + 1) - CDR_i(n + 1))^2|D^{(n)}]$$

Cette quantité permet de savoir à quel point les CDR observables peuvent différer des CDR réels. Cette vision dite "rétrospective" permet de séparer ce qui est purement aléatoire (c'est-à-dire la partie "risque de process" décrit dans la section 3.4) de l'incertitude des estimations de paramètres.

Nous allons maintenant établir une estimation de ces deux quantités, pour une seule année d'accident d'abord, puis son agrégation pour toute année d'accident. La démonstration, assez calculatoire de ces formule est disponible en annexe de l'article original [17].

Introduisons les différentes notations, dans le but de simplifier les formules des théorèmes qui vont suivre :

$$S_j^{(n)} = \sum_{i=1}^{n-j} C_{i,j}$$

$$\hat{\sigma}_j^2 = \frac{1}{n-j} \sum_{i=1}^{n-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - \hat{f}_j \right)^2 \text{ (la même volatilité que dans la section précédente)}$$

$$\hat{\Delta}_{i,n}^{(n)} = \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^{(n)})^2}{S_{n-i}^{(n)}} + \sum_{j=n-i+1}^{n-1} \left(\frac{C_{n-j,j}}{S_j^{(n+1)}} \right)^2 \frac{\hat{\sigma}_j^2 / (\hat{f}_j^{(n)})^2}{S_j^{(n)}}$$

$$\hat{\Phi}_{i,n}^{(n)} = \sum_{j=n-i+1}^{n-1} \left(\frac{C_{n-j,j}}{S_j^{(n+1)}} \right)^2 \frac{\hat{\sigma}_j^2 / (\hat{f}_j^{(n)})^2}{C_{n-j,j}}$$

$$\hat{\Psi}_i^{(n)} = \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^{(n)})^2}{C_{i,n-i}}$$

$$\Gamma_{i,n}^{(n)} = \hat{\Phi}_{i,n}^{(n)} + \hat{\Psi}_i^{(n)}$$

$$\hat{\Lambda}_{i,n}^{(n)} = \frac{C_{i,n-i}}{S_{n-i}^{(n+1)}} \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^{(n)})^2}{S_{n-i}^{(n)}} + \sum_{j=n-i+1}^{n-1} \left(\frac{C_{n-j,j}}{S_j^{(n+1)}} \right)^2 \frac{\hat{\sigma}_j^2 / (\hat{f}_j^{(n)})^2}{S_j^{(n)}}$$

$$\hat{\Xi}_{i,n}^{(n)} = \hat{\Phi}_{i,n}^{(n)} + \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^{(n)})^2}{S_{n-i}^{(n+1)}}$$

Theorem 2.3.1. Les MSEP conditionnels pour l'année d'accident i sont estimés par les formules :

$$M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(0) = (\hat{C}_{i,n}^{(n)})^2 (\Gamma_{i,n}^{(n)} + \hat{\Delta}_{i,n}^{(n)})$$

$$M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(C\hat{D}R_i(n+1)) = (\hat{C}_{i,n}^{(n)})^2 (\hat{\Phi}_{i,n}^{(n)} + \hat{\Delta}_{i,n}^{(n)})$$

Un constat que l'on peut faire immédiatement est que :

$$M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(0) = M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(C\hat{D}R_i(n+1)) + \hat{C}_{i,n}^{(n)2} \hat{\Psi}_i^{(n)}$$

Cela illustre bien la décomposition de $M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(0)$ comme un risque de process (représenté par la quantité $M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(C\hat{D}R_i(n+1))$) et d'un risque de paramétrage.

Theorem 2.3.2. Les MSEP conditionnels pour l'ensemble du triangle sont estimés par les formules :

$$M\hat{S}EP_{\sum_{i=2}^n C\hat{D}R_i(n+1)|D^{(n)}}(0) = \sum_{i=2}^n M\hat{S}EP_{C\hat{D}R_i(n+1)|D^{(n)}}(0) + 2 \sum_{k>i>1} \hat{C}_{i,n}^{(n)} \hat{C}_{k,n}^{(n)} (\hat{\Xi}_{i,n}^{(n)} + \hat{\Lambda}_{i,n}^{(n)})$$

$$\begin{aligned} M\hat{S}EP_{\sum_{i=2}^n CDR_i(n+1)|D^{(n)}}(C\hat{D}R_i(n+1)) &= \sum_{i=2}^n M\hat{S}EP_{CDR_i(n+1)|D^{(n)}}(C\hat{D}R_i(n+1)) \\ &+ 2 \sum_{k>i>1} \hat{C}_{i,n}^{(n)} \hat{C}_{k,n}^{(n)} (\hat{\Psi}_{i,n}^{(n)} + \hat{\Lambda}_{i,n}^{(n)}) \end{aligned}$$

2.4 Limites des méthodologies

La premier défaut de la méthodologie est qu'elle ne produit pas de distribution des CDR estimés pour les années d'accident ou le triangle mais seulement une mesure de volatilité (la moyenne étant nulle). Il est donc nécessaire de faire une hypothèse supplémentaire sur la distribution des CDR (généralement une loi log-normale).

Mais le principal inconvénient de la méthode est la nécessité du respect de toutes les hypothèses (de (H1) à (H4)) évoquées plus haut ce qui peut parfois être contraignant. En particulier, les hypothèses d'indépendance entre année d'accident et l'hypothèse (H2) de l'unicité de facteurs par année de développement ont pu être testés dans le modèle interne d'AXA et les résultats ne sont pas toujours satisfaisant.

Ainsi, il semble logique de s'intéresser à d'autre méthodes d'estimation du risque de réserve à partir des triangles pour palier aux défauts de ces méthodologies.

Chapitre 3

Inférence bayésienne via Monte Carlo Markov Chain

Le modèle pour l'estimation du risque de réserve qui fait l'objet du mémoire est un modèle qui s'appuie sur les statistiques bayésiennes.

Dans ce chapitre, nous allons d'abord nous intéresser à la théorie bayésienne qui constitue la base des formules qui apparaîtront dans nos approches. Nous verrons que les distributions de probabilité à prédire seront parfois difficiles à simuler et donc, après une brève description des méthodes de simulations classiques, de la nécessité de l'utilisation des méthodes basés sur les Markov Chain Monte Carlo (MCMC). L'avant dernière section décrira comment se décompose le risque d'erreur suite à l'estimation. La dernière section présentera une mesure la qualité d'un modèle bayésien qui sera utilisé par la suite dans nos études.

3.1 Inférence bayésienne

L'incertitude dans nos problèmes est décrite par une distribution de probabilité. En particulier, certaines distributions peuvent être paramétrées par un paramètre lui-même incertain et modélisé également par une distribution de probabilité.

La distribution de ce paramètre (ou de cet ensemble de paramètres) est d'abord appelée distribution *a priori* du fait qu'elle a été formulée avant d'avoir pris connaissance des données d'observation du problème. Le terme de distribution *a priori* est propre aux statistiques bayésiennes car seuls celles-ci utilisent des paramètres comme des variables aléatoire. L'utilisation des observations va ensuite permettre de déduire une distribution *a posteriori* sur ce même paramètre.

3.1.1 Règle de Bayes

Formalisons maintenant le problème de l'inférence bayésienne. Notons X une variable (ou un vecteur) aléatoire observable qui a pour paramètre une variable (ou un vecteur) aléatoire Θ dont on va chercher à déterminer la distribution. Les observations des valeurs de X vont nous permettre d'obtenir des informations sur la distribution de Θ .

Si on note :

- π_{Θ} la densité de probabilité *a priori* de Θ . C'est la distribution du paramètre Θ supposée sans tenir compte des observations de X .
- f la fonction de vraisemblance qui sert de lien entre les observations de X et le paramètre aléatoire Θ . Pour chaque observation x de X , $f(x, \theta)$ représente à quel point il est vraisemblable d'observer $X = x$ lorsque $\Theta = \theta$. Plus précisément, c'est la densité de probabilité de la variable $X|\Theta = \theta$ évalué en x .
- $\pi_{\Theta}(\cdot|X = x)$ la densité de probabilité *a posteriori* qui est la densité de probabilité mis-à-jour de Θ après observation de $X = x$.
- La loi marginale (ou densité de probabilité) de X évaluée en l'observation x , que l'on note $m_X(x)$. Comme X est une variable aléatoire de paramètre aléatoire Θ , cette quantité vaut $\int_{\Theta} f(x, \theta)\pi_{\Theta}(\theta)d\theta$.

La formule de Bayes pour les densité de probabilité donne, pour toute observation x et pour toute valeur θ du support de Θ :

$$\pi_{\Theta}(\theta|X = x) = \frac{f(x, \theta)\pi_{\Theta}(\theta)}{m_X(x)}$$

La quantité au dénominateur est une constante de normalisation qui n'a pas besoin d'être connu pour simuler la distribution *a posteriori*. En effet, la plupart des algorithmes de simulations que l'on va présenter par la suite ne nécessitent pas de densité de probabilité normalisées. Par conséquent, la loi *a priori* n'a pas besoin d'être normalisée non plus et on peut avoir une équation de Bayes de cette forme :

Densité *a posteriori* non normalisée = Vraisemblance \times Densité *a priori* non normalisée

Enfin, il est important de noter que la fonction de vraisemblance et la distribution *a priori* sont définis par jugement d'expert et ce sont ces fonctions qui définissent le modèle.

3.1.2 Structure hiérarchique

Dans nos analyses, on sera amené à étudier des distributions de paramètre dont on va fixer une loi *a priori*. Cependant, cette distribution sur le paramètre peut également dépendre d'un autre paramètre que l'on nomme dans le jargon *hyperparamètre*. Cet hyperparamètre peut également être aléatoire et aura également une distribution *a priori*.

En anglais on parle de *hyperprior* pour désigner la distribution *a priori* de l'hyperparamètre, et pour le distinguer du *prior* qui est la distribution du paramètre du premier niveau de la structure hiérarchique.

On peut généraliser la structure à un nombre plus élevé de niveaux et être amené à parler de *hyperhyperprior* puis *hyperhyperhyperprior*... même si (heureusement) nous n'aurons pas à en arriver là dans nos études.

3.2 Méthodes de simulations ordinaires

Le premier objectif de ce chapitre est de présenter les méthodes classiques de simulation et leurs limites, ce qui va justifier la nécessité de l'utilisation des MCMC. Le deuxième objectif est de préparer le lecteur à la compréhension de l'intuition derrière les méthodes MCMC. Enfin, le troisième objectif est de présenter des méthodes qui pourront également servir de benchmark par rapport aux algorithmes MCMC.

Nous supposons, dans chacune de sections suivantes, être en mesure de simuler une loi uniforme sur $[0; 1]$.

3.2.1 Inversion de la fonction de répartition

La méthode de simulation par inversion de la fonction de répartition est la méthode la plus simple et la plus intuitive pour simuler une variable aléatoire réelle X . La fonction de répartition est définie par $F : x \mapsto \mathbb{P}(X \leq x)$ et son inverse généralisé par

$$F^{-1} : u \mapsto \inf \{x \in \mathbb{R}, F(x) \geq u\}$$

Une solution pour simuler la variable aléatoire X est de simplement simuler une loi uniforme U sur $[0, 1]$ et d'appliquer l'inverse généralisé $X = F^{-1}(U)$.

Cette méthode d'apparence simple a deux principaux défauts. Le premier est qu'elle ne permet pas de simuler des lois multivariées. Le deuxième est que, connaissant la densité de probabilité, on doit réaliser numériquement une intégration pour obtenir la fonction de répartition, puis une inversion de fonction, ce qui peut être coûteux en temps de calcul.

3.2.2 Méthode de rejet

L'idée de la méthode de rejet est de simuler une distribution q , différente de la distribution cible π (pas nécessairement normalisée) et de corriger la différence en acceptant ou rejetant certaines simulations. On suppose ici être en mesure de simuler la distribution q .

Soit \mathbb{X} le support des distributions q et π . On suppose que pour tout $x \in \mathbb{X}$, on a $\frac{\pi(x)}{q(x)} < M$ pour un certain $M < \infty$.

Algorithm 1 Méthode de rejet

1. Simuler X suivant la loi de q
2. Avec une probabilité

$$\alpha = \frac{\pi(X)}{Mq(X)}$$

renvoyer X , sinon, reprendre l'étape 1

La distribution simulée à partir de cet algorithme est exactement la distribution π . Une preuve du fonctionnement de cette méthode dans le cas $\mathbb{X} = \mathbb{R}$ est disponible en annexe A.1. Dans la démonstration, on peut également constater que la probabilité d'accepter une simulation est de $1/M$ à chaque étape. Pour que l'algorithme soit plus rapide, il est intéressant de choisir M le plus petit possible et donc une distribution q "proche" de π . Le choix de q n'est pas évident dans le cas multidimensionnel, car il est difficile de trouver une distribution proche de π sur plusieurs axes simultanément lorsque la dimension est élevée.

3.2.3 L'échantillonnage préférentiel

La technique de l'échantillonnage préférentiel est légèrement différente de celles présentées précédemment car elle n'a pas pour but de générer des échantillons d'une densité de probabilité π mais plutôt de calculer sa moyenne $\mathbb{E}_\pi(X) = \int x\pi(x)dx$ et même pour toute fonction f suffisamment régulière $\mathbb{E}_\pi(f(X)) = \int f(x)\pi(x)dx$, si $X \sim \pi$. Elle servira surtout à évaluer certaines méthodes d'échantillonnage en comparant les moyennes. On parle d'échantillonnage préférentiel car il s'agit de donner plus d'importance à des échantillons que d'autres dans le calcul de la moyenne. Plus précisément, à partir d'une distribution q que l'on sait simuler, si on prend une réalisation x de $X \sim q$, on va lui donner l'importance suivante dans le calcul de la moyenne : $w = \pi(x)/q(x)$. Un des principaux avantages de cette méthode est qu'elle ne nécessite pas de borne supérieure pour la fonction $x \mapsto \frac{\pi(x)}{q(x)}$.

L'algorithme fonctionne comme décrit dans l'algorithme 2 avec une distribution π pas nécessairement normalisée.

Algorithm 2 Échantillonnage préférentiel

1. Simuler X_1, \dots, X_n suivant la loi de q pour obtenir des observations x_1, \dots, x_n .
2. Attribuer des poids $w_i = \pi(x_i)/q(x_i)$ à chaque simulation.
3. Calculer $\hat{\mathbb{E}}(f(X))_\pi = \frac{\sum_{i=1}^n w_i f(x_i)}{\sum_{i=1}^n w_i}$

La démonstration du fonctionnement de l'échantillonnage préférentiel se trouve en annexe A.2.

3.3 Markov Chain Monte Carlo

Souvent, il est difficile de trouver une bonne approximation q de π . Dans ce genre de cas la méthode la plus préférable et qui ne nécessite pas beaucoup d'information consiste à mettre à jour un processus de Markov $(X_t)_{t \geq 0}$ jusqu'à ce qu'après un certain nombre d'étapes, soit suffisamment proche de la distribution cible. Ces méthodes sont communément appelées Markov Chain Monte Carlo (MCMC).

3.3.1 Les chaînes de Markov pour l'approximation des distributions

Le principe des MCMC consiste d'abord à initialiser une chaîne de Markov par un X_0 suivant une loi π_0 pas nécessairement proche de la distribution cible π (contrairement à la méthode de rejet ou l'échantillonnage préférentiel). Un algorithme MCMC va générer une séquence X_0, X_1, \dots, X_t avec chaque X_t qui va être déterminé à partir de X_{t-1} et d'une possible autre variable aléatoire indépendante mais pas des éléments précédents de la séquence X_{t-2}, \dots . Ainsi, le processus $(X_t)_{t \geq 0}$ est une chaîne de Markov. La distribution de X_t sachant X_{t-1} va dépendre de X_{t-1} mais pas de de l'indice t . Cette distribution sera noté $K(X_t \rightarrow \cdot)$ et est appelée "noyaux de transition".

Les algorithme MCMC produisent une séquence (X_0, \dots, X_T) qui permet une approximation de π , dans le sens où, pour des fonctions $h : \mathbb{X} \mapsto \mathbb{R}$ suffisamment régulières et $X \sim \pi$:

$$\text{MC}(T, K, h) = \frac{1}{T+1} \sum_{t=0}^T h(X_t) \xrightarrow[T \rightarrow \infty]{\mathbb{P}} \mathbb{E}[h(X)] \quad (3.1)$$

et que après un nombre suffisant d'itération T_1 , on aura $X_t \sim \pi$ pour $t \geq T_1$.

L'équation 3.1 est une loi des grands nombres, ou "théorème ergodique" pour les chaînes de Markov. La quantité $\text{MC}(T, K, h)$ va être un estimateur de $\mathbb{E}[h(X)]$.

Contrairement aux algorithmes précédents, les échantillons successifs de $(X_t)_{t \geq 0}$ vont être corrélés. Il sera alors courant d'utiliser des autocorrélogrammes pour étudier la corrélation entre états successifs.

Il y aura plusieurs choix à faire pour savoir quels valeurs de (X_0, \dots, X_T) seront choisis pour représenter des simulations de π :

1. Le "burn in" qui est le nombre d'itération à effectuer pour considérer que la chaîne de Markov a suffisamment convergé et que les X_t suivent bien la distribution limite.
2. La "thinning" qui est le nombre d'itération entre deux choix de X_t successif de la chaînes, pour avoir des simulations le moins corrélées possibles et presque indépendantes.

3.3.2 L'algorithme de Metropolis-Hastings

L'algorithme de Metropolis-Hastings (MH) est la méthode la plus commune des MCMC. Si on considère une distribution cible π à simuler, l'algorithme MH va construire une chaîne de Markov réversible X_0, X_1, \dots, X_n de distribution stationnaire π .

La construction de la chaîne se fait par itération successive. On simule le premier élément de la chaîne selon une distribution initiale π_0 . Si on suppose avoir atteint l'état X_{t-1} , on simule une proposition d'un nouvel état X^* pour X_t via la "proposal distribution" ou "proposal kernel" que l'on note $q(X_{t-1} \rightarrow \cdot)$ qui dépend de X_{t-1} . Pour cela, on suppose bien sûr d'être en mesure de simuler le "proposal kernel".

L'état X^* va ensuite être ou ne pas être accepté pour être égale à X_t . Dans le cas où il n'est pas accepté, on aura $X_t = X_{t-1}$. La probabilité d'acceptation est la suivante :

$$\alpha = \min\left(1, \frac{\pi(X^*)q(X^* \rightarrow X_{t-1})}{\pi(X_{t-1})q(X_{t-1} \rightarrow X^*)}\right)$$

L'algorithme 3 fournit une description détaillée de la procédure pour Metropolis-Hastings.

Algorithm 3 Algorithme de Metropolis-Hastings

1. Simuler $X_0 \sim \pi_0$
 2. A l'itération t (état courant X_{t-1}), simuler $X^* \sim q(X_{t-1} \rightarrow \cdot)$
 3. Calculer $\alpha = \min\left(1, \frac{\pi(X^*)q(X^* \rightarrow X_{t-1})}{\pi(X_{t-1})q(X_{t-1} \rightarrow X^*)}\right)$
 4. Simuler $U \sim U[0; 1]$
 - (a) Si $U < \alpha$, alors $X_t = X^*$
 - (b) Sinon, $X_t = X_{t-1}$
 5. Recommencer à l'étape 2 avec l'itération suivante
-

Généralement, la "proposal distribution" est une distribution symétrique, ce qui réduit la probabilité d'acceptation à la formule simplifiée : $\alpha = \min\left(1, \frac{\pi(X^*)}{\pi(X_{t-1})}\right)$. Une "proposal distribution" classique est la loi normale centrée en X_{t-1} et d'une variance / matrice de covariance judicieusement choisi. Ce cas particulier s'appelle le "Random Walk Metropolis Hastings".

En pratique, on essaye si possible de choisir :

- Une distribution initiale π_0 "proche" de la distribution cible pour que la chaîne de Markov converge rapidement.
- Un "proposal kernel" q d'une variance suffisamment grande pour proposer des états différents du précédent mais pas trop grande pour que ces états soient acceptés.

La preuve du bon fonctionnement de l'algorithme de Métropolis-Hasting se trouve en annexe A.3. Un exemple pour illustrer l'algorithme est fourni en annexe B. Il est très simple et montre comment simuler une loi normale avec l'algorithme MH.

Cet exemple est le seul de ce mémoire où l'on va manuellement coder des algorithmes MCMC. Les modèles que l'on va étudier seront trop complexes et pour cela nous allons faire appel au logiciel STAN qui va simuler les chaînes de Markov. Notre travail sera plutôt de choisir les bons paramètres (burn-in, thinning, nombre de chaînes ...).

3.3.3 L'échantillonnage de Gibbs

Dans le cas des distributions multivariées, l'algorithme de Metropolis-Hastings tel que décrit plus haut suppose de choisir à chaque itération un nouvel échantillon dans l'espace multi-dimensionnel. Quand le nombre de dimensions est élevé, trouver le pas du déplacement peut être difficile, car chaque dimension peut se comporter très différemment, or le pas du saut doit être adapté à chaque dimension pour éviter un mélange trop lent. Une approche alternative qui peut mieux fonctionner dans de telles situations est l'échantillonnage de Gibbs. Cette approche consiste à choisir un nouvel échantillon pour chaque dimension séparément des autres. Elle est utilisable en particulier lorsque la distribution multivariée est composée d'un ensemble de variables aléatoires conditionnées par un petit ensemble d'autres variables, ce qui est typiquement le cas dans les modèles bayésiens hiérarchiques. Les variables individuelles sont échantillonnées une à une et chaque variable est conditionnée sur la valeur la plus récente de celles dont elle dépend.

L'algorithme 4 décrit le fonctionnement de l'échantillonnage de Gibbs dans le cas simplifié du balayage systématique où les éléments du vecteur aléatoire sont mis-à-jour un par un.

3.3.4 Hamiltonian Monte Carlo

3.3.4.1 Cas particulier : Algorithme MALA

Un variant classique de l'algorithme MH est le "Metropolis Adjusted Langevin Algorithm" ou MALA. Il est applicable lorsque la distribution cible est définie sur \mathbb{R}^d , et que le gradient $\nabla \log(x)$ peut être évalué pour tout x . La "proposal distribution" sera, à l'instant t , $q(X_{t-1} \rightarrow \cdot) \equiv \mathcal{N}(X_{t-1} + (h\Sigma/2)\nabla \log \pi X_{t-1}, h\Sigma)$. Le paramètre h est un scalaire positif représentant la taille du pas dans l'itération et Σ est une matrice de covariance. On note la similarité avec le "Random Walk MH", sauf que cette fois-ci le centre n'est pas X_{t-1} mais $X_{t-1} + (h\Sigma/2)\nabla \log \pi X_{t-1}$. Autrement dit, le centre suit la gradient de la distribution cible.

Algorithm 4 Echantillonnage de Gibbs

1. On dispose d'une simulation du vecteur aléatoire $X^{(t)}$ de dimension m à l'itération t
2. La première composante du vecteur $X^{(t+1)}$ est simulée selon la loi conditionnel $\pi(x_1|X_2^{(t)}, \dots, X_m^{(t)})$
3. La deuxième composante du vecteur $X^{(t+1)}$ est simulée selon la loi conditionnel $\pi(x_2|X_1^{(t+1)}, X_2^{(t)}, \dots, X_m^{(t)})$
4. ...
5. La i -ème composante du vecteur $X^{(t+1)}$ est simulée selon la loi conditionnel $\pi(x_i|X_1^{(t+1)}, \dots, X_{j-1}^{(t+1)}, X_{j+1}^{(t)}, \dots, X_m^{(t)})$
6. ...
7. On obtient alors une simulation de $X^{(t+1)}$ qui va servir pour la prochaine itération.

3.3.4.2 Cas général : Algorithmme HMC

La présentation du cas particulier MALA nous permettra d'avoir une meilleure appréhension de la démarche de l'Hamiltonian Monte Carlo (dit "HMC" par la suite). La distribution cible est toujours définie sur \mathbb{R}^d et on définit une extension de cette distribution cible $\bar{\pi}(x, p) = \pi(x)\mathcal{N}(p; 0, M)$ dans $\mathbb{R}^d \times \mathbb{R}^d$. La matrice M est appelée la matrice de masse. Nous avons connaissance de l'état X_{t-1} de l'étape $t - 1$ et la variable P_{t-1} est échantillonnée selon la loi $\mathcal{N}(0, M)$. On imagine ensuite la paire (X_{t-1}, P_{t-1}) comme une particule de position X , d'impulsion P et de masse M (on ne tient pas compte du fait que la M n'est pas un scalaire). On interprète $-\log \pi$ comme son energie potentiel et $p^T M^{-1} p / 2$ comme son énergie cinétique. On définit la position initiale $q(0) = X_{t-1}$ de la particule ainsi que son impulsion $p(0) = P_{t-1}$, et la fonction Hamiltonienne $H(q, p) = -\log \pi(q) + p^T M^{-1} p / 2$. On résout ensuite l'équation de la trajectoire de la particule qui évolue en fonction de son énergie en résolvant les équations hamiltonienne :

$$\frac{dx_i}{dt} = \frac{\partial H}{\partial p_i}$$

et

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial x_i}$$

Cette résolution se fait en L étape de "l'intégrateur de Verlet" de pas ϵ c'est-à-dire que pour $l = 1, \dots, L$:

$$\begin{aligned} p(l + 1/2) &= p(l) + \frac{\epsilon}{2} \nabla \log \pi(q(l)), \\ q(l + 1) &= q(l) + \epsilon M^{-1} p(l + 1/2), \\ p(l + 1) &= p(l + 1/2) + \frac{\epsilon}{2} \nabla \log \pi(q(l + 1)) \end{aligned}$$

Après ces L étapes, nous obtenons des proposition $X^* = q(L)$ et $P^* = p(L)$ de futurs états de la chaînes de Markov qui seront acceptés ou non comme remplacement de (X_{t-1}, P_{t-1}) , selon une certaine probabilité. Notons que (X^*, P^*) est une fonction de (X_{t-1}, P_{t-1}) qui dépend de L et ϵ et que l'on peut noter $(X^*, P^*) = \Phi_{L,\epsilon}(X_{t-1}, P_{t-1})$. En définissant la probabilité d'acceptation $\alpha_{\text{HMC}}(X_{t-1}, P_{t-1}) = \min(1, \frac{\bar{\pi}(X^*, P^*)}{\bar{\pi}(X_{t-1}, P_{t-1})})$ et en imposant que $(X_t, P_t) = (X^*, P^*)$ avec une probabilité $\alpha_{\text{HMC}}(X_{t-1}, P_{t-1})$ et $(X_t, P_t) = (X_{t-1}, -P_{t-1})$ avec une probabilité $1 - \alpha_{\text{HMC}}(X_{t-1}, P_{t-1})$, on construit une chaîne de Markov $\bar{\pi}$ invariante (voir la démonstration dans l'article [18]). On peut également voir une chaîne de Markov $(X_t)_{t \geq 0}$ laissant π invariant car π est une distribution marginales de $\bar{\pi}$ sur le composant X .

L'algorithme 5 résume le passage de l'état à l'instant $t - 1$, X_{t-1} de la chaîne vers l'état à l'instant t , X_t .

Algorithm 5 Hamiltonian Monte Carlo

1. Tirer aléatoirement P_{t-1} selon $\mathcal{N}(0, M)$.
 2. Réaliser L itérations de l'intégrateur de Verlet de pas ϵ et obtenir $(X^*, P^*) = \Phi_{L,\epsilon}(X_{t-1}, P_{t-1})$.
 3. Avec une probabilité $\alpha_{\text{HMC}}(X_{t-1}, P_{t-1})$, poser $X_t = X^*$ et $X_t = X_{t-1}$.
-

On remarque que l'algorithme MALA décrit plus haut correspond au cas particulier du HMC où $L = 1$, $h = \epsilon^2$ et $\Sigma = M^{-1}$.

Pour résumer, le HMC est un algorithme MCMC qui évite les déplacements aléatoires du Metropolis-Hastings classique en prenant en compte les informations du gradient du premier ordre. Cela permet une convergence plus rapide pour les distributions cibles définies sur des hautes dimensions.

3.3.5 Introduction au variant No-U-Turn Sampler

On note que l'algorithme HMC est sensible au paramètre du pas ϵ et au nombre d'itération L dans l'intégrateur de Verlet. En effet, si L est trop petit, l'état proposé va prendre une direction trop aléatoire, et si L est trop grand, il y aura un trop grand coût en temps de calcul.

L'extension No-U-Turn Sampler (NUTS), présentée en détail dans [12], a été imaginée pour éliminer la nécessité de fixer ces paramètres en entrée de l'algorithme. Au lieu d'utiliser un paramètre L fixé, NUTS utilise un algorithme récursive qui permet d'explorer une large panel de candidat pour la distribution cible et arrête l'exploration lorsque celle-ci tourne en rond. D'autre part, l'algorithme NUTS intègre une méthode qui permet d'adapter la taille du pas ϵ . Cette méthode est une méthode d'optimisation

stochastique et plus spécifiquement une adaptation de l'algorithme "primal-dual" de [15].

Nous n'allons pas ici entrer dans les détails de l'algorithme NUTS qui est assez complexe et qui nécessiterait beaucoup de descriptions. Mais le lecteur intéressé pourra étudier l'article [12] pour avoir tous les détails de la méthode.

3.3.6 Le langage STAN

STAN est un langage de programmation utilisé pour spécifier un modèle d'inférence bayésienne. Il écrit en C++. Il est accessible à travers différentes interfaces et en particulier sur R grâce à la librairie "rstan".

STAN implémente les méthodes MCMC basées sur l'utilisation de gradient parmi lesquelles se trouvent l'algorithme HMC et le NUTS.

Dans nos futures simulations de chaînes de Markov, nous allons utiliser l'algorithme NUTS qui est l'algorithme par défaut de STAN.

3.4 Décomposition des risques du modèle

Dans un modèle bayésien, il y a deux types d'incertitude. La première est due à la variance du modèle qui impose des distributions plus ou moins volatiles. L'autre est dû au fait que les paramètres sont estimés avec plus ou moins de précisions.

La décomposition des risques du modèle va se baser sur le théorème très célèbre de la variance total :

Theorem 3.4.1. Si X et Y sont deux variables aléatoires sur un même espace probabilisé et que Y est de variance finie, alors

$$\text{Var}(Y) = \mathbb{E}_X[\text{Var}(Y|X)] + \text{Var}_X(\mathbb{E}[Y|X])$$

Démonstration. On a :

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 \\ &= \mathbb{E}[\mathbb{E}[Y^2|X]] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 \\ &= \mathbb{E}[\text{Var}(Y|X) + (\mathbb{E}[Y|X])^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 \\ &= \mathbb{E}[\text{Var}(Y|X)] + (\mathbb{E}[(\mathbb{E}[Y|X])^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2) \\ &= \mathbb{E}[\text{Var}(Y|X)] + \text{Var}(\mathbb{E}[Y|X]) \end{aligned}$$

□

Dans notre cas, on va considérer une variable aléatoire X conditionnée par un paramètre lui-même aléatoire Θ :

$$\text{Var}(X) = \mathbb{E}_{\Theta}[\text{Var}(X|\Theta)] + \text{Var}_{\Theta}(\mathbb{E}[X|\Theta])$$

La quantité à gauche de l'équation est le "risque total". Le premier terme du côté droit de l'équation est appelé "risque de process" car il représente la variance moyenne des sorties du modèle. Le deuxième terme à droite est appelé "risque de paramètre" car il représente la variance due aux multiples possibilités de paramètres de la distribution *a posteriori*.

En pratique, dans les modèles que nous étudierons, la valeur du risque totale dépendra principalement du risque de paramètre qui lui va changer en fonction du paramétrage que l'on choisit. Selon que le paramétrage *a priori* est plus ou moins volatile, la volatilité de la quantité d'intérêt (les paiements cumulés ou les CDR) sera plus ou moins élevée.

3.5 Critère de qualité du modèle

Nous allons nous intéresser à un critère qui permet d'évaluer si le modèle est approprié par rapport aux données d'observations : le Deviance Information Criterion. Il s'agit d'une généralisation du critère d'information d'Akaike.

On définit la déviance par $D(\theta) = -2\log(f(x|\theta)) + C$, où x est la donnée d'observation, θ est le paramètre inconnu, $f(x|\theta)$ la fonction de vraisemblance du modèle et $C = 2\log(f(x))$ une constante qui n'a pas d'importance car elle s'annule lors des calculs de comparaison des modèles.

Il y a plusieurs définition du DIC et nous allons nous intéresser à celle proposée dans [16] :

$$DIC = \mathbb{E}(D(\Theta)) - \Delta D(\mathbb{E}(\Theta))$$

où $\mathbb{E}(D(\Theta))$ est une mesure d'adéquation du modèle (qui est ici l'opposé de la moyenne de la vraisemblance et qui mesure bien le bon "fit" des données au modèle) et $\Delta D(\mathbb{E}(\Theta)) = \mathbb{E}(D(\Theta)) + D(\mathbb{E}(\Theta))$ estime le nombre de paramètres effectifs du modèle (moins il est élevé, meilleur est le modèle car le risque de sur-apprentissage sera réduit).

Ainsi, le modèle avec le DIC le plus faible est celui qui "fit" mieux les données.

Dans les études de modèles bayésiens, il est intéressant d'évaluer le DIC mais cette valeur n'est généralement pas déterminante dans le choix du modèle. En pratique, elle est plus intéressante pour comparer différents modèles que pour comparer un choix de paramétrage *a priori* qui est surtout choisi par jugement d'expert. Dans ce mémoire, il n'y a qu'un seul modèle et seul le paramétrage diffère.

Chapitre 4

Méthode Correlated Chain Ladder et extension

Après avoir passer en revue les méthodes classiques d'estimation du risque de réserve et la théorie mathématique des modèles bayésiens, on peut enfin s'intéresser au modèle Correlated Chain Ladder qui peut apporter une solution pour éviter les hypothèse trop contraignante de M&W, en plus de ses autres avantages que l'on verra dans la suite du mémoire.

4.1 Historique et contexte de l'étude

Le modèle Correlated Chain Ladder est initialement un modèle de provisionnement de réserves à l'ultime introduit par l'actuaire américain Glenn Meyers dans [13]. Meyers s'est basé sur les triangles de réserves publics de Casual Acturial Society pour constater que généralement les modèle classiques de Mack et Bootstrap Overdispersed Poisson (Bootstrap ODP) avaient tendance à sous-estimer la variabilité des réserves. Pour faire ce constat il a fallu un nombre élevé de triangle développés suffisamment pour connaître les ultimes de plusieurs années d'accident.

Le but de ce mémoire ne va pas être de donner suite au travail de Glenn Meyers mais plutôt de voir comment appliquer la méthodologie qu'il a développé avec les données d'AXA, sans connaître la réalité des résultats, et de les comparer avec les autres méthodologie de provisionnement.

L'article [6] plus récent va proposer une extension à 1 an du modèle CCL ce qui va particulièrement nous intéresser dans le cadre des estimations des risques pour respecter la norme Solvabilité II.

4.2 CCL à l'ultime

Dans ce chapitre, nous allons décrire le modèle CCL que l'on va utiliser dans ce mémoire qui peut être très légèrement différent de ceux décrit dans [13] et [6] à cause de choix différents de certaines distributions *a priori* plus faciles à calibrer ou plus adaptées aux données d'AXA.

4.2.1 Le modèle d'inférence

On considère un triangle de paiement cumulé :

$$D^{(t)} = \{C_{ij}; i \in [1 : t], j \in [1 : t], i + j \leq t + 1\}$$

Le but est de trouver une distribution des C_{ij} inconnu et pour cela nous allons utiliser un modèle bayésien avec différents paramètres. Le modèle repose sur une structure dite "cross-classified" c'est-à-dire que les paiements cumulés à prédire vont s'écrire sous la forme $C_{ij} = f(\alpha_i, \beta_j)$ avec les α comme paramètres contenant les informations liées aux années d'accident et les β comme paramètres contenant les informations liées aux années de développement.

Le modèle, ou plus précisément la fonction de vraisemblance du modèle bayésien, s'écrit :

- $C_{ij} \sim \log \mathcal{N}(\mu_{ij}, \sigma_j)$ pour $i, j \in [1 : t]$
- $\mu_{1j} = \alpha_1 + \log(elt_1) + \beta_j$
- $\mu_{ij} = \alpha_i + \log(elt_i) + \beta_j + \rho[\log(C_{i-1,j}) - \mu_{i-1,j}]$

Un autre modèle développé par Meyers est le "Leveled Chain Ladder" qui est ce même modèle mais sans l'introduction du terme $\rho[\log(C_{i-1,j}) - \mu_{i-1,j}]$ qui permet de mettre en évidence des dépendances entre les années d'accident.

On peut déjà constater un point très intéressant du modèle CCL par rapport au modèle de Mack : il n'y a pas nécessité d'hypothèse d'indépendance entre deux années d'accident.

4.2.2 Spécification des paramètres

En notant B_i les primes brut de l'année d'accident i , on explicite les distributions *a priori* :

- $\alpha_i \sim \mathcal{U}[\log(B_i) - \epsilon_i, \log(B_i) + \epsilon_i]$
- $elt_i \sim \log \mathcal{N}(\gamma_i, \delta_i)$
- $\beta_j \sim \mathcal{U}[-\eta_j, 0]$ avec comme contrainte $\beta_t = 0$
- $\sigma_j^2 = \sum_{h=j}^t \tau_h$ où $\tau_h \sim \text{Beta}(a, b)$
- $\rho \sim \mathcal{U}[-1, 1]$

4.2.2.1 Explication du paramétrage des μ

Le choix des α_i permet de ramener le centre de la distribution des μ_{ij} et donc des C_{ij} autour de la valeur des primes pour chaque années d'accident. Les elr_i sont les "expected loss ratio" qui permettent d'ajuster encore le centre de la distribution des μ_{ij} en fonction de ce qui est attendu comme rapport prime/pertes et la volatilité de ce dernier rapport. Les β_j sont toujours à support négatifs et représentent de niveau de développement des pertes. Généralement ρ est choisi pour avoir une distribution non-informative de moyenne nulle comme nous le verrons plus tard.

Autrement dit, nous avons $\mu_{ij} = \alpha_i + \log(elr_i) + \beta_j + \rho[\log(C_{i-1,j}) - \mu_{i-1,j}]$ avec :

- $\mathbb{E}(\alpha_i) = \log(B_i)$
- $\mathbb{E}(\log(elr_i)) = \log(\text{Loss Ratio}_i)$ avec une certaine hypothèse de Loss Ratio
- $\beta_j < 0$ et qui va *a posteriori* se rapprocher de 0 au fil des années de développement
- $\mathbb{E}(\rho[\log(C_{i-1,j}) - \mu_{i-1,j}]) = 0$

4.2.2.2 Explication du paramétrage des σ

Le choix de la calibration des σ permet d'avoir des distributions pour les C_{ij} de moins en moins volatiles en fonction des années de développement (on a $\sigma_1 > \dots > \sigma_t$). En effet, plus les pertes sont développées, moins le volume des pertes restantes à payer est élevé et donc la volatilité est faible.

4.2.2.3 Explication du paramétrage ρ

Le paramètre ρ est généralement choisi avec une distribution non-informative. Il s'agit d'un paramètre de corrélation comme on va le montrer dans le théorème suivant.

Theorem 4.2.1. Sachant les paramètres $\alpha_i, \beta_j, \sigma_j, elr_i, \rho$ et le développement $C_{i-2,j}$, le coefficient de corrélation linéaire entre $\log C_{ij}$ et $\log C_{i-1,j}$ est ρ .

Démonstration. On suppose que les $\alpha_i, \beta_j, \sigma_j, elr_i, \rho$ et le développement $C_{i-2,j}$ connus et qu'ils ne sont donc pas considérés comme des variables aléatoires. Il en va donc de même pour $\mu_{i-1,j}$.

Le coefficient de corrélation linéaire entre $\log(C_{ij})$ et $\log(C_{i-1,j})$ est donné par :

$$\mathbf{C} = \mathbb{E}\left[\left(\frac{\log(C_{ij}) - \mathbb{E}[\log(C_{ij})]}{\sigma(\log(C_{ij}))}\right)\left(\frac{\log(C_{i-1,j}) - \mathbb{E}[\log(C_{i-1,j})]}{\sigma(\log(C_{i-1,j}))}\right)\right)$$

Sachant les paramètres $\alpha_i, \beta_j, \sigma_j, \rho$ et le développement $C_{i-2,j}$, on a clairement $\sigma(\log(C_{ij})) = \sigma(\log(C_{i-1,j})) = \sigma_j$ et $\mathbb{E}[\log(C_{i-1,j})] = \mu_{i-1,j}$.

D'autre part :

$$\begin{aligned}
\mathbb{E}[\log(C_{ij})] &= \mathbb{E}[\mathbb{E}[\log(C_{ij})|\mu_{ij}]] \\
&= \mathbb{E}[\mu_{ij}] \\
&= \alpha_i + elr_i + \beta_j + \rho(\mathbb{E}[\log(C_{i-1,j})] - \mu_{i-1,j}) \\
&= \alpha_i + elr_i + \beta_j
\end{aligned}$$

Ainsi on a :

$$\begin{aligned}
\mathbf{C} &= \mathbb{E}\left[\left(\frac{\log(C_{ij}) - \mathbb{E}[\log(C_{ij})]}{\sigma(\log(C_{ij}))}\right)\left(\frac{\log(C_{i-1,j}) - \mathbb{E}[\log(C_{i-1,j})]}{\sigma(\log(C_{i-1,j}))}\right)\right)] \\
&= \mathbb{E}\left[\left(\frac{\log(C_{ij}) - (\alpha_i + elr_i + \beta_j)}{\sigma_j}\right)\left(\frac{\log(C_{i-1,j}) - \mu_{i-1,j}}{\sigma_j}\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\log(C_{ij}) - (\alpha_i + elr_i + \beta_j)}{\sigma_j}\right)\left(\frac{\log(C_{i-1,j}) - \mu_{i-1,j}}{\sigma_j}\right) \middle| C_{i-1,j}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{\log(C_{ij}) - (\alpha_i + elr_i + \beta_j)}{\sigma_j}\right) \middle| C_{i-1,j}\right]\left(\frac{\log(C_{i-1,j}) - \mu_{i-1,j}}{\sigma_j}\right)\right] \\
&= \mathbb{E}\left[\left(\frac{\rho(\log(C_{i-1,j}) - \mu_{i-1,j})}{\sigma_j}\right)\left(\frac{\log(C_{i-1,j}) - \mu_{i-1,j}}{\sigma_j}\right)\right] \\
&= \rho \mathbb{E}\left[\frac{(\log(C_{i-1,j}) - \mu_{i-1,j})^2}{\sigma_j^2}\right] \\
&= \rho (\text{car } \log(C_{i-1,j}) \sim \mathcal{N}(\mu_{i-1,j}, \sigma_j))
\end{aligned}$$

□

4.2.3 Distribution a posteriori

Une fois le modèle et les distributions *a priori* des paramètres définies, on simule les distributions *a posteriori* grâce à l'outil d'inférence STAN. Plus concrètement dans notre étude, la librairie *rstan* de R va permet de faire la connexion avec un fichier ".stan" dans lequel va être décrit le modèle et les spécifications des paramètres en langage STAN.

Les codes présents en annexe C donnent un exemple de spécification de paramètre en STAN et un code R permettant de simuler, à partir du fichier ".stan", une distribution *a posteriori* des paramètres. Plus précisément, le code permet de simuler, grâce à l'algorithme NUTS, 8 chaînes de Markov avec chacune un "burn-in" de 2500 simulations (nombre jugé nécessaire pour que la chaîne converge vers la distribution d'équilibre), un "thinning" de 10 (pour avoir des simulations indépendantes) et un nombre total de simulation de 15000. Un rapide calcul permet de déduire qu'il y a bien $8 * \frac{15000-2500}{10} = 10000$ simulations. Les indicateurs décrit en annexe D permettront de valider ces choix.

On introduit la notation suivante, pour $k \in \{1, \dots, K\}$, avec K le nombre de simulation (dans notre étude $K = 10000$) :

$$\Theta^{(k)} = (\{\alpha_i\}_{i=1}^t, \{elr_i\}_{i=1}^t, \{\beta_j\}_{j=1}^{t-1}, \{\sigma_j\}_{j=1}^t, \rho)_k$$

Ces ensembles de paramètres représentent chacun une simulation de la distribution *a posteriori* des paramètres.

Enfin, à partir de chaque simulation k des paramètres, on peut en déduire une simulation des paiements cumulés de la partie inconnu du triangle de développement :

$$C_{ij}^{(k)} \sim \log \mathcal{N}(\mu_{ij}^{(k)}, \sigma_j^{(k)})$$

On en déduit, pour tout k , et puisque l'on fait l'hypothèse que la première année d'accident est entièrement développée, une simulation des montants de réserves :

$$R^{(k)} = \sum_{i=2}^t C_{it}^{(k)} - \sum_{i=2}^t C_{i,t-i}^{(k)}$$

Ainsi, nous parvenons à obtenir une distribution empirique de K simulations des montants de réserve.

4.3 Extension du modèle CCL à horizon 1 an

Ce chapitre a pour but de présenter une extension du modèle CCL dans le but de calculer le SCR du risque de réserve :

$$SCR_{Reserve} = VaR_{99,5\%} \left(\sum_{i=2}^t P_{i,t-i+1} + R^{D^{(t+1)}} \right) - R^{D^{(t)}}$$

Ici, $R^{D^{(t)}}$ désigne le Best Estimate (BE) des provisions à la date t . Le BE des provisions à la date $t + 1$, $R^{D^{(t+1)}}$, est une variable aléatoire car il dépend des paiements à la date $t + 1$ qui sont aléatoires. Ces paiements sont les $P_{i,t-i+1}$ dans la formule.

L'idée générale de l'extension à horizon un an va être de simuler des diagonales de paiements de l'année $t + 1$ grâce aux ensembles de paramètres $\Theta^{(k)}$, et d'utiliser ces diagonales simulées pour ré-évaluer la vraisemblance de ces paramètres $\Theta^{(k)}$ ce qui nous permettra de déduire des simulations de $R^{D^{(t+1)}}$.

Les étapes de la procédures sont les suivants :

1. Nous avons à disposition les ensembles de paramètres $\Theta^{(k)}$ et on va en déduire les $\mu_{ij}^{(k)}$. Notons que $\mu_{ij} = \alpha_i + elr_i + \beta_j + \rho[\log(C_{i-1,j}) - \mu_{i-1,j}]$ et que si $C_{i-1,j}$ n'est pas connu, on le simule avec la loi log-normale de paramètres $\mu_{i-1,j}$ et σ_j (et si $\mu_{i-1,j}$ n'est pas connu, idem une avec $i - 2$, etc ...). On sauvegarde ensuite les paramètres

μ dans une matrice, de taille $t \times t$ pour chaque simulation k , $\mathbf{M}^k = [\mu_{ij}^k]$. De même, on sauvegarde les paramètres σ dans un vecteur de taille t qui est $\sigma^k = [\sigma_j^k]$.

2. On simule ensuite, grâce aux paramètres, pour chaque k , S simulations de la diagonale de l'année calendaire $t + 1$: $C_{i,t-i+1}^{(k,s)}$. Le nombre total de simulation est alors $K \times S$ et chaque simulation est un "triangle/trapèze/polygone" $T^{(k,s)}$ composé du triangle $D^{(t)}$ et de la diagonale des $C_{i,t-i+1}^{(k,s)}$.
3. A partir des $T^{(k,s)}$, nous allons mettre à jour les $\Theta^{(k)}$ pour obtenir leur densité *a posteriori* après connaissance de la dernière diagonale simulée. Plus précisément, on calcul d'abord la vraisemblance (avec $\Phi_L(x|\mu, \sigma)$ densité d'une variable aléatoire log-normale) :

$$L[T^{(k,s)}|\Theta^{(k)}] = \prod_{C_{ij}^{(k,s)} \in T^{(k,s)}} \Phi_L(C_{ij}^{(k,s)}|\mu_{ij}^{(s,k)}, \sigma_j^{(s,k)})$$

Ensuite, la distribution *a posteriori* s'écrit

$$\mathbb{P}(\Theta^{(k)}|T^{(k,s)}) = \frac{L[T^{(k,s)}|\Theta^{(k)}] \times \mathbb{P}(\Theta^{(k)})}{\sum_{l=1}^K L[T^{(l,s)}|\Theta^{(l)}] \times \mathbb{P}(\Theta^{(l)})}$$

Chaque simulation $\Theta^{(k)}$ est équiprobable donc $\mathbb{P}(\Theta^{(1)}) = \dots = \mathbb{P}(\Theta^{(K)})$. On a finalement :

$$\mathbb{P}(\Theta^{(k)}|T^{(k,s)}) = \frac{L[T^{(k,s)}|\Theta^{(k)}]}{\sum_{l=1}^K L[T^{(l,s)}|\Theta^{(l)}]}$$

Ainsi, nous arrivons à obtenir S différentes distributions *a posteriori* de Θ , chacune en fonction des nouvelles diagonales.

4. On calcul ensuite $\mathbb{E}[C_{ij}^{(k)}] = \exp(\mu_{ij}^{(k)} + \frac{\sigma_j^{2(k)}}{2})$ pour les C_{ij} inconnus et on en déduit le Best Estimate $R_k^{D^{(t+1)}}$. Ce BE est estimé simplement par $R_k^{D^{(t+1)}} = \sum_{i=2}^t \mathbb{E}[C_{i,t}^{(k)}] - \mathbb{E}[C_{i,t-i+2}^{(k)}]$ mais peut aussi être estimé avec des taux forward appliqués sur les incréments de paiement.
5. A partir de $R_k^{D^{(t+1)}}$ et de la distribution *a posteriori* $\mathbb{P}(\Theta^{(k)}|T^{(k,s)})$ qui comprend la connaissance de la nouvelle diagonale, on calcul la distribution des réserves évaluées à l'année $t + 1$. On a pour tout s :

$$R_s^{D^{(t+1)}} = \sum_{k=1}^K R_k^{D^{(t+1)}} \times \mathbb{P}(\Theta^{(k)}|T^{(k,s)})$$

Ainsi, nous avons estimé une distribution des des valeurs moyennes des réserves résiduels.

6. D'autre part, pour chaque simulation s , il y a à disposition K simulations de diagonale de l'année $t + 1$. Il s'agit des $C_{i,t-i+1}^{(k,s)}$ qui nous permettent de déduire les

incréments de paiement $P_{i,t-i+1}^{(k,s)}$. On choisi aléatoirement une de ces K valeurs pour en déduire enfin les obligations pour l'année suivante :

$$A_s^{1an} = \sum_{i=1}^t P_{i,t-i+1}^{(s)} + R_s^{D^{(t+1)}}$$

Nous avons finalement à disposition S simulations des obligations futures A_s^{1an} .

7. Le $SCR_{Reserve} = VaR_{99.5\%}(\sum_{i=2}^t P_{i,t-i+1} + R^{D^{(t+1)}}) - R^{D^{(t)}}$ se déduit ensuite comme la différence du quantile 99.5% empirique des A_s^{1an} et du BE à l'année t $R^{D^{(t)}}$.

Cette extension du modèle à horizon 1 an nécessite autant voir plus de temps de calcul que les simulations de chaînes de Markov de la partie précédente. En effet, l'étape 3 qui consiste à évaluer $K \times S$ valeurs de vraisemblance est coûteuse en temps de calcul. Nous choisirons dans cette étude $K = S = 10000$ simulations.

Une autre remarque importante est de définir bonne manière de faire le calcul du BE de l'année t . Il y a deux possibilités (on a toujours $\mathbb{E}[C_{ij}^{(k)}] = \exp(\mu_{ij}^{(k)} + \frac{\sigma_j^{2(k)}}{2})$) :

- Pour tout k , $R_k^{D^{(t)}} = \sum_{i=2}^t \mathbb{E}[C_{i,t}^{(k)}] - \mathbb{E}[C_{i,t-i+1}^{(k)}]$ puis $\hat{R}^{D^{(t)}} = \frac{1}{K} \sum_{k=1}^K R_k^{D^{(t)}}$. Cette approche est consistante avec la manière de calculer les $R_k^{D^{(t+1)}}$ mais a le défaut de ne pas utiliser les vraies valeurs de la diagonal des $C_{i,t-i+1}$.
- Pour tout k , $R_k^{D^{(t)}} = \sum_{i=2}^t \mathbb{E}[C_{i,t}^{(k)}] - C_{i,t-i+1}$ puis $\hat{R}^{D^{(t)}} = \frac{1}{K} \sum_{k=1}^K R_k^{D^{(t)}}$. Cette approche permet d'avoir des valeurs proches entre la moyenne des réserves à l'ultime estimée avec l'approche CCL à l'ultime et le BE.

Ces deux approches donnent un BE proche mais qui peut différer légèrement à cause du fait qu'il y a parfois un léger sous-apprentissage des paramètres qui n'arrivent pas à "fitter" les données d'observation (en effet l'hypothèse de lognormalité des paiements cumulés n'est pas toujours pertinente). On a alors parfois une différence entre les vraies observations $C_{i,t-i+1}$ et les moyennes $\frac{1}{K} \sum_{k=1}^K \mathbb{E}[C_{i,t-i+1}^{(k)}]$. Il sera important dans nos études de vérifier que ces deux valeurs coïncident au moins partiellement. Par défaut, nous définirons le BE à l'année t comme celui calculé avec la deuxième approche.

4.4 Calcul du DIC

Lors de l'étude d'un triangle, plusieurs paramétrages de loi *a priori* peuvent être testés et une façon de comparer la pertinence des paramétrages est de calculer leur DIC. Nous allons dans ce chapitre donner un estimateur du DIC pour le modèle CCL.

Nous avons vu dans le chapitre précédent la définition du DIC qui équivaut ici à

$$DIC = 2\mathbb{E}(D(\Theta)) - D(\mathbb{E}(\Theta))$$

On ne tient pas compte de constante dans la formule de la deviance pour simplifier le problème.

On a d'une part l'estimateur :

$$\hat{\mathbb{E}}(D(\Theta)) = -2\frac{1}{K} \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^{n-i+1} \log(f(C_{ij}|\mu_{ij}^{(k)}, \sigma_j^{(k)}))$$

où $f(C_{ij}|\mu_{ij}^{(k)}, \sigma_j^{(k)})$ est la densité de probabilité de la loi log-normale de paramètre $\mu_{ij}^{(k)}, \sigma_j^{(k)}$ évaluée en C_{ij} .

D'autre part, on a la deviance de l'estimateur :

$$D(\hat{\mathbb{E}}(\Theta)) = -2 \sum_{i=1}^n \sum_{j=1}^{n-i+1} \log(f(C_{ij}|\sum_{i=1}^K \mu_{ij}^{(k)}, \sum_{i=1}^K \sigma_j^{(k)}))$$

Ainsi :

$$DIC = -2\left(\frac{2}{K} \sum_{k=1}^K \sum_{i=1}^n \sum_{j=1}^{n-i+1} \log(f(C_{ij}|\mu_{ij}^{(k)}, \sigma_j^{(k)})) - \sum_{i=1}^n \sum_{j=1}^{n-i+1} \log(f(C_{ij}|\sum_{i=1}^K \mu_{ij}^{(k)}, \sum_{i=1}^K \sigma_j^{(k)}))\right)$$

Chapitre 5

Application du CCL aux triangles de réserve d'AXA XL

Maintenant que la théorie de la méthodologie Correlated Chain Ladder a été introduite, nous allons voir comment elle peut s'appliquer sur des triangles de réserves de certaines lignes d'activité d'AXA. L'étude se fera principalement sur des branches longues (qui se développent lentement) et peu développées car elles nécessitent plus de jugements d'expert et moins de temps de calcul. En effet, le CCL peut transcrire les jugements d'expert à travers le paramétrage ce qui semble rendre la méthodologie attrayante dans ce cas là.

5.1 A propos du risque de réserve de l'entité AXA XL

Le processus d'estimation du risque de réserve du groupe AXA inclut une estimation du risque de réserve pour tout le périmètre d'exposition de la division AXA XL, qui comprend toutes les lignes de business modélisées dans le modèle interne Non-Vie. La division AXA XL comprend les sous-entités XL Catlin, AXA Corpportate Solution et AXA Art. Le risque de réserve en assurance Vie d'AXA XL n'est pas étudié dans ce mémoire.

La méthodologie standard d'estimation du risque de réserve des lignes d'activité des entités d'AXA consiste à d'abord à évaluer un montant de réserve pour chaque année d'accident du triangle avec des études dossier-dossier, la méthodologie Chain-Ladder, des jugements d'expert ou autres. Ce montant de réserve est appelé E_{Axa} . Ensuite, on estime par des méthodes statistique la volatilité autour de ces valeurs d' E_{Axa} . Cette volatilité, dite MSEP (pour Mean Square Error of Prediction), pour le risque à l'ultime ou à 1 an, va permettre de déduire des CoV (pour Coefficient of Variation) avec $CoV = MSEP/E_{Axa}$. Dans le cas du risque à 1 an, on fait une hypothèse de loi log-normale pour la distribution de pertes (c'est-à-dire la distribution des boni-mali) et on estime le SCR par la formule suivante :

$$SCR_{LoB} = E_{Axa} \times [VaR_{99,5\%}(\log \mathcal{N}(\mu, \sigma)) - 1]$$

Avec :

$$\mu = -\frac{1}{2} \ln[1 + \text{CoV}^2]$$

$$\sigma = \sqrt{\ln[1 + \text{CoV}^2]}$$

L'idée ici était d'introduire une loi log-normale $\log \mathcal{N}(\mu, \sigma)$ de moyenne 1 et dont la volatilité est égale à au coefficient de variation CoV. On multiplie ensuite cette loi par le EAXa. Cette méthode diffère légèrement de celles qui est réellement utilisées par AXA, notamment du fait de l'introduction de corrélations entre années d'accident, mais c'est celle-ci que l'on va comparé avec les méthodes CCL. Plus précisément, la méthode qui va servir pour l'étude comparative consistera à calculer les MSEF avec la méthode de M&W et la moyenne ne sera pas faite basées sur les EAXa mais avec les résultats du Chain Ladder.

5.2 Bonnes pratiques de calibration

A la différence de l'étude dans l'article [6], les données des triangles d'AXA XL ne sont pas re-simulées pour suivre une loi log-normale. Il va donc être nécessaire d'être prudent sur le choix des paramètres *a priori* car l'algorithme aura plus de difficulté à apprendre sur les données d'observation.

Il est en général assez difficile de bien paramétrer les distributions *a priori* dans le monde log-normale car elles dépendent beaucoup de l'ordre de grandeur des données du triangle. Pour simplifier le problème, nous allons toujours utiliser le même triangle à un coefficient multiplicatif près pour nous permettre d'avoir un triangle qui a pour ordre de grandeur des développement à l'ultime d'environ 100 000 euros par années d'accident.

Enfin, il n'est pas toujours aisé de réussir une bonne calibration dès le premier essai de paramètres. On constate en effet parfois des pathologies dans l'apprentissage avec les MCMC ou une trop importante incohérence entre les paramètres appris et les données d'observation.

Les résultats qui seront présentés dans les sections suivantes ne feront pas toujours sens. En effet, nous allons présenter un certain nombres de calibrations y compris celles qui donnent des résultats incohérents ou faux, en particulier sur la première ligne (LoB A). Cela permet au lecteur de comprendre la démarche qui permet d'aboutir à une bonne calibration.

Dans les sections suivantes, nous calibrerons les *elr* pour qu'ils aient une loi *a priori* log-normale ayant une certain moyenne m égale au loss ratio que l'on suppose et un certain paramètre de volatilité σ . On va en déduire le paramètre μ pour que $\mathbb{E}(\log \mathcal{N}(\mu, \sigma)) = m$, grâce à la formule $m = \exp(\mu + \sigma^2/2)$.

5.3 Étude de la ligne A

La ligne d'activité A désigne une partie de l'activité d'assurance qui fournit une compensation financière aux employés en cas de sinistres ayant eu lieu sur le lieu de travail, et plus particulièrement ici, dans le milieu du bâtiment aux Etats-Unis. Les principales couvertures de cette ligne d'activité concernent les dommages aux biens, l'auto, la responsabilité civile générale, la responsabilité professionnelle, l'assurance excédentaire (qui couvre les pertes au-dessus de l'assurance primaire), ainsi que les indemnités.

Les méthodes bayésiennes peuvent apporter des résultats intéressants pour l'étude de cette ligne d'activité. En effet, même si le triangle de réserve semble assez développé, elle ne contient que peu d'années d'accidents et de données donc les jugements d'experts auront beaucoup d'importance dans cette étude et ces hypothèses d'experts pourront se retranscrire à travers les lois a priori du modèle. De plus, le fait de ne contenir que peu d'années d'accidents implique un plus faible nombre de paramètres à estimer et donc un temps de simulation des MCMC plus court. Nous constaterons que le temps de simulation des MCMC (la partie CCL à l'ultime uniquement donc) est d'un peu moins de 2h en général. La partie "extension du modèle à 1 an" nécessite un temps de calcul d'environ 1h.

5.3.1 Preprocessing

Pour préserver la confidentialité des données d'AXA, nous n'allons pas présenter le triangle d'origine et nous allons directement étudier le triangle qui suit le même *payment pattern* mais à un facteur multiplicatif et quelques ajouts de légers bruits près : il s'agit du tableau 5.1 (on utilise les notations AY pour désigner les Accident Year / années d'accident et DY pour désigner les Development Year / années de développement).

AY/DY	1	2	3	4	5	6	7	8	9	10	Prime
2011	938	25 899	28 394	27 257	35 331	35 379	35 412	35 455	35 567	33 683	28 886
2012	3 603	15 078	19 754	29 203	31 816	31 548	31 000	31 281	31 684		64 969
2013	11 419	27 736	36 093	42 494	41 724	44 016	46 991	50 842			112 747
2014	6 637	26 257	38 535	45 788	49 381	54 015	57 823				153 271
2015	6 216	66 995	72 797	75 059	85 428	80 400					197 451
2016	10 732	73 384	74 548	79 243	96 886						216 649
2017	12 619	72 685	89 038	110 465							270 193
2018	11 782	120 196	132 067								296 012
2019	11 415	52 967									340 052
2020	24 794										347 523

TABLE 5.1 – Triangle de paiement cumulé LoB A

On visualise ensuite le développement des paiements cumulés du triangle sur le graphique 5.1.

Cette visualisation des développements par années d'accident permet de constater

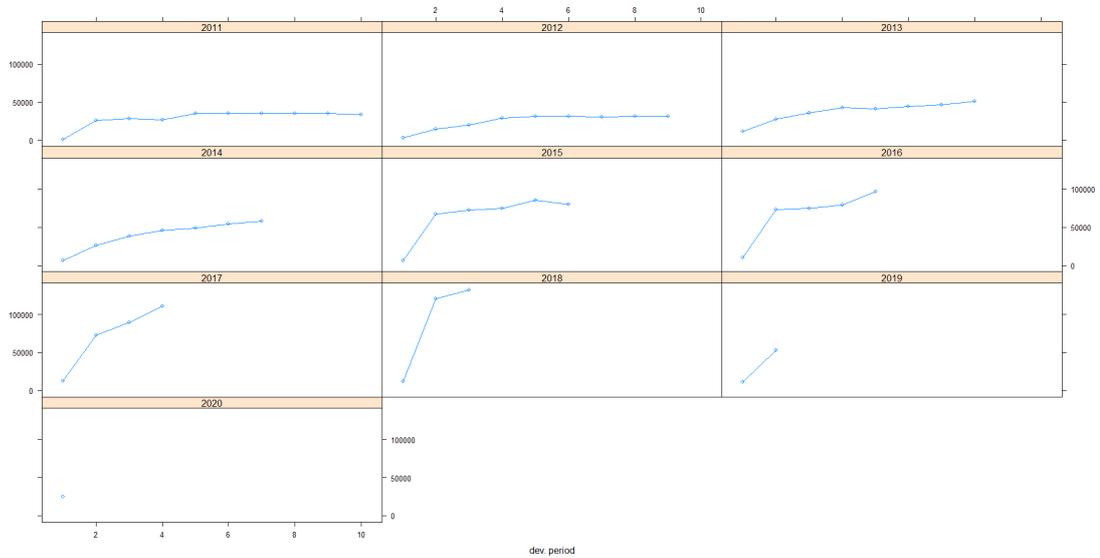


FIGURE 5.1 – Visualisation du développement du triangle de la LoB A

certain points :

- Les années d'accident 2011 et 2012 semblent être les seules à être entièrement développées (cependant, nous verrons que les experts ont ajoutés des EAXa par prudence)
- Il y a des incréments négatifs sur certaines années
- Aucun pattern de développement ne semble pouvoir être déterminé précisément (certaines années se développent plus vite que d'autres)

5.3.2 Paramétrages convergents vers les résultats du modèle de Mack

La particularité générale des paramétrages suivants est que l'on va donner des distributions *a priori* sur les paramètres qui vont permettre d'avoir une moyenne de paiement cumulé à l'ultime égale à celle que donnerait le modèle Mack/Chain-Ladder. L'apprentissage des données va ensuite faire que cette moyenne ne sera plus exactement la même avec les distributions *a posteriori*. En d'autres termes, dans ces paramétrages, nous faisons confiance au modèle de Mack.

5.3.2.1 Premier paramétrage

On reprend les notations des paramètres présentes dans le chapitre 4.

Tout d'abord, nous allons considérer la première année comme totalement développée et sans incertitude donc $\alpha_1 = 0$ et $\log(elt_1) = 0.154$ (c'est-à-dire le loss ratio $\log(33683/28886)$).

Variable	log-moyenne	log-écart-type
elr_3	-0.84	0.000005
elr_4	-0.99	0.001
elr_5	-0.87	0.008
elr_6	-0.77	0.025
elr_7	-0.73	0.035
elr_8	-0.51	0.05
elr_9	-1.43	0.08
elr_{10}	-0.36	0.1

TABLE 5.2 – LoB A : Loi a priori des expected loss ratio

Nous allons presque considérer la même chose pour la deuxième année avec toujours $\alpha_2 = 0$ mais avec un petit degrés de liberté sur l'expected loss ratio : $elr_2 \sim \log \mathcal{N}(-0.77, 0.000005)$. Ainsi, la moyenne de la distribution du Loss Ratio $\mathbb{E}(elr_2) = \exp(-0.77 + \frac{0.000005^2}{2})$ serait au Loss Ratio donné par le modèle de Mack et avec un peu de volatilité autour.

Pour les autres années, nous allons utilisé $\alpha_i \sim \mathcal{U}[-0.5, 0.5]$ avec une volatilité croissante autour du loss ratio de Mack qui sera perçue dans les elr (voir tableau 5.2). On note que le 0.5 choisi dans la loi des α_i a été déterminé en fonction de l'ordre de grandeur des paiement cumulé dans le "monde lognormal".

Le choix pour les β_j sera de prendre une loi uniforme $\mathcal{U}[-3, 0]$ et $\beta_{10} = 0$.

Les τ_h qui vont permettre de réduire les volatilités σ_j au fur et à mesure que j va augmenter vont suivre une loi Beta(1, 7). Les τ_h suivront alors une distribution Beta de moyenne 1/8.

Dans ce premier essai, nous allons définir une distribution *a priori* non-informative pour le coefficient de corrélation ρ , c'est-à-dire $\rho \sim \mathcal{U}[-1, 1]$.

Les diagnostics de convergence (voir Annexe D pour comprendre la signification des mesures) sont positifs :

- Les n_{eff} sont tous proches de 10000 : nous avons donc environ 10000 simulations indépendantes
- Le degrés de convergence \hat{R} est toujours égale à 1 ou très proche de 1 : la convergence a bien eu lieu et les sorties sont stables
- Aucune des 10000 itérations ne s'est terminée par une divergence
- La mesure du BFMI est bonne : pas de comportement pathologique lors de la phase adaptative de la chaîne de Markov et la distribution *a posteriori* a bien été explorée

Les n_{eff} pour chaque paramètres sont présentés dans la table 5.3.

Le constat que l'on peut faire sur les distributions *a posteriori* des paramètres est que les premiers α et elr ne sont pas volatile. En effet, les premières années d'accident sont

Paramètre	n_{eff}	Paramètre	n_{eff}	Paramètre	n_{eff}
α_1	-	β_1	9534	elr_1	-
α_2	-	β_2	9544	elr_2	9608
α_3	9005	β_3	9002	elr_3	10293
α_4	9578	β_4	9786	elr_4	9855
α_5	9727	β_5	9436	elr_5	9895
α_6	10032	β_6	9051	elr_6	9764
α_7	9385	β_7	10206	elr_7	9950
α_8	9772	β_8	10035	elr_8	9868
α_9	10034	β_9	10072	elr_9	9750
α_{10}	10234	ρ	9729	elr_{10}	9998

TABLE 5.3 – LoB A : Nombre de simulations effectives du premier paramétrage

Statistique	CCL	Mack
Moyenne	359 806	313 470
Ecart-type	93 667	116 405
CoV	26%	37%
Médiane	353 996	-
Quantile 99.5%	621 115	-

TABLE 5.4 – LoB A : Résultat à l'ultime

déjà presque entièrement développées. Au contraire, les premiers β sont plus volatile, car ce sont sur les premières années que les développements se font plus vite.

La distribution *a posteriori* des paramètres permet, à partir du modèle de déduire la distribution des réserves à l'ultime avec cette méthode CCL, représentée sur la figure 5.2.

Les résultats du modèle CCL, comparés au modèle standard de Mack, sont présent dans le tableau 5.4.

Les résultats de la vue à 1 an sont présent dans le tableau 5.5 et sont comparés aux résultats du modèle de Merz & Wutrich sur lequel on "fit" une loi log-normale à partir des résultats de moyenne et écart-type. La valeur Best Estimate 1 est calculée à partir des l'estimation de la diagonale due aux paramètres, et la valeur du Best Estimate 2 est calculée grâce à l'utilisation des vraies valeurs de la diagonale du triangle. Le SCR est calculé à partir du Best Estimate 2 dans le cas du CCL 1 an et avec la moyenne de le cas M&W. Nous pouvons faire le constat que les Best Estimates 1 et 2 sont proches donc le modèle "fit" assez bien les données.

La distribution des réserves vision 1 an est représentée sur le graphique 5.3.

Enfin, nous pouvons constater que les données affirment une indépendance entre les années d'accident avec un coefficient ρ de distribution *a posteriori* de moyenne -4.6%. La distribution de ρ est donnée sur le graphique 5.4.

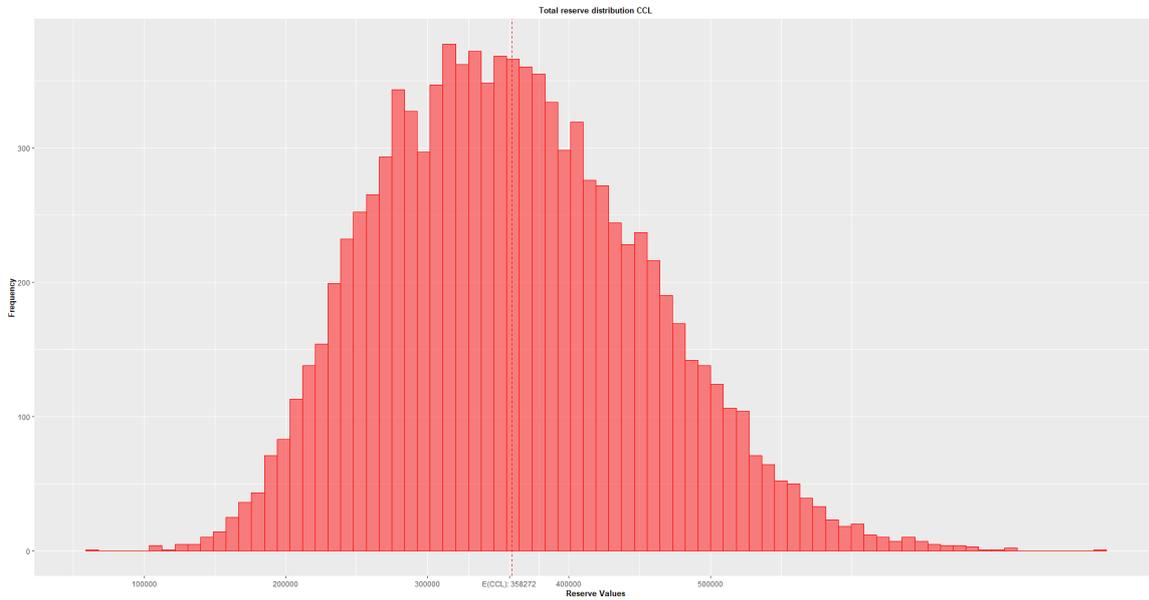


FIGURE 5.2 – LoB A : Distribution des réserves à l'ultime

Statistique	CCL 1 an	M&W avec hypothèse log-normale
Moyenne	353 881	314 603
Best Estimate 1	358 272	-
Best Estimate 2	359 964	-
Ecart-type	95 149	106 382
CoV	27%	37%
Médiane	343 424	298 025
Quantile 99.5%	688 040	695 563
SCR	328 076	380 960
SCR/BE	91%	121%

TABLE 5.5 – LoB A : Résultat à 1 an

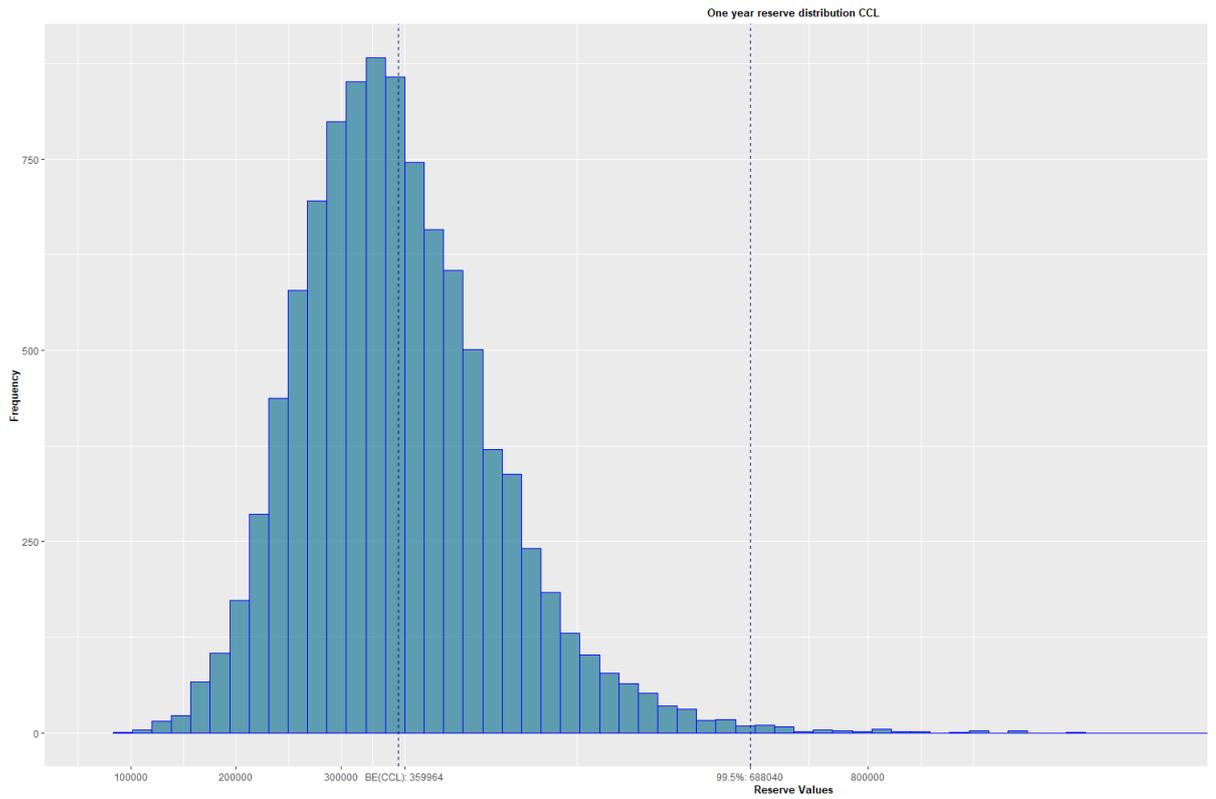


FIGURE 5.3 – LoB A : Distribution des réserves vision 1 an

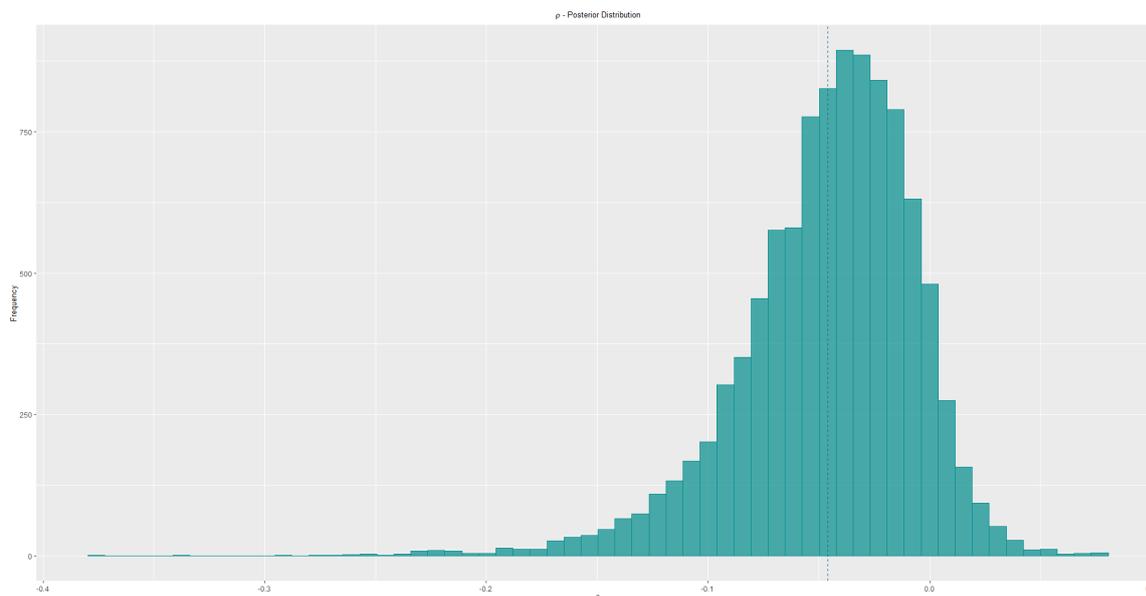


FIGURE 5.4 – LoB A : Distribution a posteriori du paramètre de corrélation ρ (1)

5.3.2.2 Deuxième paramétrage

Ce deuxième paramétrage est exactement le même que le premier à la différence près que l'on ajoute la contrainte d'avoir $\rho \geq 0$. Et donc la loi *a priori* de ce paramètre est $\rho \sim \mathcal{U}[0; 1]$.

Comme dans la calibration précédente, on ne retrouve aucune pathologie de convergence dans les simulations MCMC et les simulations sont bien indépendantes.

Le principal constat qui a été fait est que la distribution *a posteriori* de ρ toujours positive par contrainte mais proche de 0 (voir figure 5.5). La moyenne *a posteriori* de ce paramètre est 1.3%.

Très peu de différences sont observées sur les distributions des réserves à l'ultime et à 1 an. Elles ont en effet la même forme et des statistiques similaires. Une légère différence sur quantile 99.5% de la distribution à 1 an peut être observé : il est de 722 329 contre 688 040 dans le paramétrage précédente. Cette dernière différence vient de l'ajout de la contrainte de corrélation qui augmente donc le coefficient de corrélation entre année d'accident ce qui impacte la volatilité de la distribution et cet impact est particulièrement significatif sur les quantiles élevés.

5.3.2.3 Troisième paramétrage

Nous allons cette fois chercher à être prudent dans notre estimation et pour ce faire nous allons d'une part augmenter la taille du support des distributions des paramètres (pour plus de volatilité) et un coefficient de corrélation de loi *a priori* concentrée autour

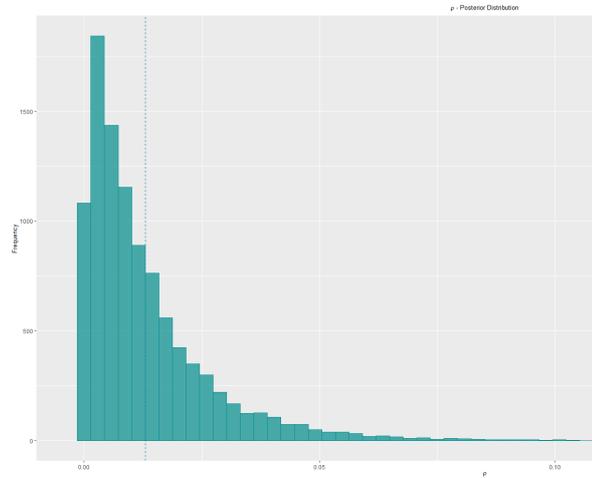


FIGURE 5.5 – LoB A : Distribution a posteriori du paramètre de corrélation ρ (2)

Paramètre	n_{eff}
α_3	8739
ρ	4905
τ_1	5692
τ_{10}	7784

TABLE 5.6 – LoB A : Nombre de simulations effectives du troisième paramétrage

d'une corrélation élevée.

Le choix des distributions a priori sont les suivants :

- $\alpha_1 = \alpha_2 = 0$ et $\alpha_i \sim \mathcal{U}[-0.8, 0.8]$ pour les autres années d'accident
- $\beta_j \sim \mathcal{U}[-4, 0]$ pour j différent de 10.
- $\rho \sim \text{Beta}(5, 5)$ donc un coefficient de corrélation de moyenne 0.5.
- Les lois des *elr* sont celles du premier paramétrage mais avec le double de leurs écart type et leur moyenne ajustés à ces écart type

Les résultats des MCMC ne sont pas satisfaisant. En effet, il y a un certain nombre de paramètres qui ont un $n_{eff} < 9000$ (on choisi ce seuil de façon arbitraire). Ces paramètres et leur nombres de simulations effectives sont présentés dans le table 5.6.

De plus, il y a un écart trop conséquent entre les différentes façon de calculer le Best Estimate, ce qui prouve que les données n'ont pas bien été apprises.

Par la suite, nous allons donc nous concentrer sur l'amélioration des lois *a priori* des paramètres du tableau qui n'ont pas un bon nombre de simulations effectives. Après deux autres essais qu'on ne présente pas ici, on peut conclure que la pathologie vient du

Statistique	CCL 4ème paramétrage	CCL 1er paramétrage	Mack
Moyenne	383 670	359 806	313 470
Ecart-type	133 566	93 667	116 405
CoV	35%	26%	37%
Médiane	367 066	353 996	-
Quantile 99.5%	783 737	621 115	-

TABLE 5.7 – LoB A : Résultat à l’ultime 4ème paramétrage et comparaison

Statistique	CCL 1 an 4ème param.	CCL 1 an 1er param.	M&W
Moyenne	374 156	353 881	314 603
Best Estimate 1	376 871	358 272	-
Best Estimate 2	383 999	359 964	-
Ecart-type	123 391	95 149	106 382
CoV	33%	27%	37%
Médiane	354 584	343 424	298 025
Quantile 99.5%	845 157	688 040	695 563
SCR	461 158	328 076	380 960
SCR/BE	120%	91%	121%

TABLE 5.8 – LoB A : Résultat à 1 an 4ème paramétrage et comparaison

comportement de ρ qui influe sur les distributions de tous les autres paramètres du fait qu’il est présent dans chaque μ_{ij} . Plus précisément, la loi *a priori* $Beta(5, 5)$ que suit ρ a une moyenne de 0.5 et un quantile 0.5% de 0.15 ce qui est bien au-dessus de ce qui est appris avec une distribution *a priori* non informative (la moyenne *a posteriori* est proche de 0%). La loi $Beta(5, 5)$ n’est donc pas une bonne loi *a priori* pour ρ car elle est trop éloignée de la réalité des données.

5.3.2.4 Quatrième paramétrage

Ce quatrième et dernier paramétrage va toujours avoir pour but de donner une distribution prudente des réserves à l’ultime et à 1 an mais avec des distributions *a priori* qui apprennent bien les données.

La paramétrage est le même que le précédent sauf pour les β_j qui vont suivre une loi uniforme entre -3 et 0 et pour ρ (et c’est le point le plus important) qui va avoir la distribution non-informative $\mathcal{U}[-1; 1]$.

Après MCMC, tous les paramètres ont un nombre de simulations effectives proche de 10000 et un \hat{R} très proche de 1.

Les statistiques des distributions des réserves à l’ultime (resp. à 1 an) et comparée au modèle de Mack (resp. M&W) et au premier paramétrage sont visible sur le tableau 5.7 (resp. 5.8).

Enfin, le paramètre ρ a une distribution *a posteriori* similaire au premier paramétrage et avec presque la même moyenne (ici elle est de -4.7% contre -4.6% au premier

AY	EAXa	Estimation de l'ultime	Loss Ratio
2011	8 391	42 074	145.7%
2012	6 592	38 276	58.9%
2013	15 063	65 906	58.5%
2014	32 002	89 825	58.6%
2015	57 694	138 093	69.9%
2016	70 444	167 329	77.2%
2017	89 464	199 929	74.0%
2018	124 138	256 204	86.6%
2019	194 215	247 182	72.7%
2020	249 234	274 026	78.9%

TABLE 5.9 – LoB A : Loss Ratio déterminé par jugement d'expert

paramétrage).

Cette augmentation de la volatilité par rapport au premier paramétrage, visible sur la valeur de l'écart-type, de la CoV et du quantile 99.5%), vient de l'augmentation de la volatilité des paramètres. En effet, la formule du chapitre 3.4 indiquant que Risque Total = Risque de process + Risque de paramètre permet de comprendre que l'augmentation du risque total vient de l'augmentation du risque de paramètre.

5.3.3 Utilisation des EAXa

Pour finir, nous n'allons pas faire confiance au modèle de Mack mais plutôt faire confiance aux experts qui ont estimé un EAXa par année d'accident. Rappelons que les EAXa sont des réserves estimées pour chaque année d'accident par des jugements d'expert et que souvent il y a un sur-provisionnement pour des raisons de prudence. La somme, par année d'accident, des EAXa et du paiement cumulé de la dernière année de développement (sur la dernière diagonale), donne une estimation du paiement cumulé à l'ultime. Ce dernier permet de déduire le Loss ratio qui est simplement rapport de ce dernier paiement cumulé et des primes reçues. Les valeurs sont résumées dans le table 5.9.

Cette fois-ci, nous allons utiliser un cinquième paramétrage de loi *a priori* très volatile, car on constate déjà intuitivement, à cause d'un provisionnement trop prudent, que les données n'indiquent pas une convergence des paiements cumulés vers la somme des EAXa et des paiement cumulés et donc que le Loss Ratio utilisés comme centre des loi *a priori* ne sont pas cohérent avec le modèle. On peut s'attendre à ce que les données "tirent" la moyenne des réserves à l'ultime vers des résultats proches de ceux du Chain Ladder (qui est de 313 470) mais que les moyennes des paramètres des lois attirent plutôt la moyenne des réserves à l'ultime vers la somme des EAXa (qui est de 847 235).

Le paramétrage est à forte volatilité : les α suivent une loi uniforme en -1.2 et 1.2, les β suivent une loi uniforme allant de -4 à 0 (sauf le dernier qui est toujours nul), et des σ dont le paramètre de volatilité varie de 0.02 (pour l'année d'accident la plus ancienne)

Statistique	CCL	Mack et EAxa
Moyenne	517 138	847 235
Ecart-type	211 061	116 405
CoV	41%	14%
Médiane	477 813	-
Quantile 99.5%	1 176 410	-

TABLE 5.10 – LoB A : Résultat à l’ultime, avec jugement d’expert

Statistique	CCL 1 an	M&W avec hypothèse log-normale et EAxa
Moyenne	487 092	847 235
Best Estimate 1	500 483	-
Best Estimate 2	517 768	-
Ecart-type	162 950	106 382
CoV	33%	13%
Médiane	456 214	840 634
Quantile 99.5%	1 154 125	1 961 963
SCR	636 357	1 114 728
SCR/BE	123%	132%

TABLE 5.11 – LoB A : Résultat à 1 an, avec jugement d’expert

à 0.1 (année d’accident la plus récente). Le paramètre de corrélation ρ suit toujours la même loi uniforme non-informative qui va de -1 à 1.

Les résultats de convergence sont assez satisfaisant : les \hat{R} sont tous égaux à 1 et seuls peu de paramètres ont des n_{eff} inférieur à 9000 et tous sont supérieurs à 8500. Nous voyons dans les tableaux 5.10 et 5.11, que les résultats de la moyenne des réserves à l’ultime et que les BE des réserves à 1 an estimés par les 2 méthodes (avec et sans l’utilisation des diagonales observées) sont proches.

Les tableaux 5.10 et 5.11 illustrent, comme prévu, le fait que les EAxa tendent à surestimer les réserves par rapport au modèle de Mack/Chain Ladder. Comme la forte volatilité des lois des paramètres *a priori*, l’incertitude sur les réserves à l’ultime et à 1 an l’est aussi, et ainsi la valeur des quantiles extrêmes est beaucoup plus élevée. L’étude comparative se fait avec un modèle de Mack ayant pour moyenne le EAxa et la même volatilité que celle calculé par le modèle de Mack. En ce qui concerne M&W, une loi log-normale est "fittée" et le SCR est calculé comme dans la section 5.1. L’idée est de comparer la méthode CCL avec la méthodologie du groupe AXA qui utilise aussi les EAxa. On constate ici que l’hypothèse log-normale de M&W implique un SCR plus élevé que celui du CCL malgré un écart-type plus faible.

Les courbes de distributions à l’ultime et à 1 an sont assez similaire à celles des anciens paramétrages, mais avec beaucoup plus de volatilité. Cette grande volatilité se remarque bien sur la distribution *a posteriori* du coefficient de corrélation ρ (voir figure 5.6) qui, bien que toujours centrée, est beaucoup plus volatile.

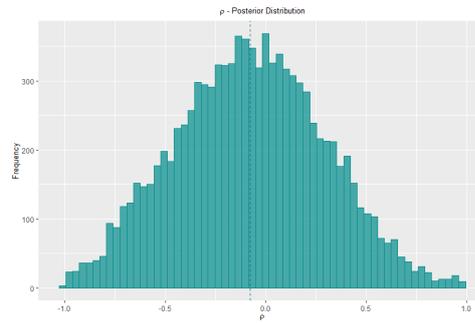


FIGURE 5.6 – LoB A : Distribution a posteriori du paramètre de corrélation ρ , avec utilisation des EAXa

Statistique	Standard	Avec changement de graine	Avec bruit blanc
Moyenne	359 806	357 149	358 284
Ecart-type	93 667	93 592	93 367
CoV	26%	26%	26 %
Médiane	353 996	349 949	351 994
Quantile 99.5%	621 115	615 380	615 980

TABLE 5.12 – Résultat des tests de stabilité sur le CCL à l’ultime

5.3.4 Étude de la stabilité du modèle

Nous allons vérifier la stabilité de la méthode de deux manières.

D’une part, on change la graine de simulation et on compare les résultats des deux graines. Si le modèle est stable, les résultats doivent être assez proches.

D’autre part, on choque les triangle avec un faible bruit blanc et on vérifie que les résultats sont proche avec ou sans bruit blanc. On choisit ici d’ajouter aux valeurs présente sur le triangle de Réserves (qui sont de l’ordre de quelques dizaine de milliers) un bruit blanc défini par une loi uniforme entre -20 et 20. Cela permet de légèrement modifier les valeurs sans changer leur ordre de grandeur ni modifier la structure du développement du triangle.

Pour cette étude de sensibilité, nous allons utiliser le premier paramétrage avec convergence vers les résultats du modèle de Mack.

Les statistiques des distributions des réserves à l’ultime (resp. à 1 an) standard, avec changement de graine et avec choc des valeurs avec un bruit blanc sont visible sur le tableau 5.12 (resp. 5.13). Pour avoir un ordre d’idée de la valeur de ces variations, on présente en annexe E les sensibilité aux changement de graine et aux bruits blancs des modèles de Mack et M&W. Le modèle de Mack et de M&W étant déterministe, un changement de graine ne change pas le résultat. On conclut que le modèle CCL est un peu sensible au changement de graine et au bruit blanc, surtout sur les quantiles élevés, mais l’ordre de grandeur des résultats reste le même.

Statistique	Standard	Avec changement de graine	Avec bruit blanc
Moyenne	353 881	347 542	348 673
Best Estimate 1	358 272	356 118	357 057
Best Estimate 2	359 964	357 523	358 586
Ecart-type	95 149	96 139	96 977
CoV	27%	28 %	28%
Médiane	343 424	337 621	338 348
Quantile 99.5%	688 040	692 946	701 066
SCR	328 076	335 423	342 480
SCR/BE	91%	94%	96%

TABLE 5.13 – Résultat des tests de stabilité sur le CCL à 1 an

Paramétrage	Estimation DIC
Premier	1121.618
Premier avec changement de graine	1121.800
Premier avec bruit blanc	1121.738
Deuxième	1124.946
Troisième	1150.151
Quatrième	1123.863
Cinquième (utilisation des EAxa)	1123.769

TABLE 5.14 – LoB A : Estimation du DIC

5.3.5 Comparaison des critères d'informations

On peut vérifier la pertinence du critère d'information DIC présenté précédemment. On construit pour cela le tableau 5.14 avec l'estimation faite du DIC pour chacun des paramétrages et des sensibilité présenté aux paragraphes précédents. Rappelons qu'en théorie, plus le DIC est faible, meilleur est le modèle.

En observant le tableau, on peut constater la pertinence du DIC. En effet, nous ne voyons que très peu de différence entre la valeur du DIC avec ou sans changement de graine, ou avec ou sans bruit blanc, ce qui montre sa stabilité. Aussi, nous avons vu que le 3ème paramétrage était beaucoup moins approprié (problème de convergence, mauvais apprentissage des données ...), et celui-ci a bien un DIC significativement plus élevé.

5.3.6 Conclusion de l'étude de la ligne A

L'étude de la LoB A nous a appris plusieurs choses. Tout d'abord, elle nous a permis de nous rendre compte de la nécessité d'un bon paramétrage, suffisamment volatile pour que la distribution *a posteriori* soit dans l'intervalle exploré par la distribution *a priori*. Ensuite, elle nous a permis de constater que la volatilité de la distribution *a priori* influe beaucoup la valeur du risque total, dans le sens où si celle-là est grande, le quantile 99.5% de la distribution des réserves à l'ultime et à 1 an sera grande. Les études de sensibilité

nous ont permis de nous convaincre de la pertinence de l'utilisation du modèle CCL. Enfin, on a pu constater que le DIC était un indicateur pertinent pour mesurer la qualité du modèle, dans le sens où il permet de décrire si le modèle est adapté aux données d'observation.

Nous avons utilisé divers paramétrages, chacun fournissant des estimations intéressantes du niveau de risque de réserve (sauf le troisième qui n'a pas convergé). Le premier paramétrage peut être vu comme un paramétrage assez peu prudent. Les calibrations chez AXA autorisent rarement une hypothèse de corrélations négatives, quelque soit le niveau, donc on peut voir le deuxième paramétrage comme une extension du premier plus prudente et qui respecte le guide des calibrations d'AXA, même si celle-ci est risquée pour une utilisation sur d'autres lignes car peut entraîner une non-convergence en cas de corrélation a posteriori négative. Le quatrième est paramétrage est un paramétrage prudent. Enfin, le cinquième est sans doute le plus intéressant car il est à la fois prudent et tient compte des avis d'expert sur l'estimation des EAx. Il y a dans ce cinquième paramétrage deux couches de prudence : la première venant des choix conservateurs de EAx, le deuxième venant de la grande volatilité des lois *a priori* qui augmentent le risque total en augmentant le risque de paramètre. On pourrait aussi réfléchir à calculer le SCR à partir de la somme des EAx au lieu des Best Estimates de la méthode CCL 1 an.

5.4 Étude de la ligne B

L'assurance paramétrique est fondée sur l'utilisation de paramètres corrélés à un dommage ou une perte pour un assuré. Une large variété de risques peuvent être couverts par ce type d'assurance, comme par exemple l'impact climatique sur les performances d'une entreprise ou sur les récoltes. Au sein d'AXA Corporate Solutions (sous-groupe d'AXA XL, appartenant déjà à AXA avant le rachat et qui assurait les entreprises), l'équipe d'assurance paramétrique a développé une expertise sur les risques climatiques et travaillé avec plusieurs acteurs dans différents secteurs à travers le monde : agriculture, énergies renouvelables, construction, transports, loisirs et textiles, mais également avec des institutions internationales et des gouvernements.

La ligne B contient encore moins d'années d'accident que la ligne A et donc les méthodes bayésiennes auront sans doute encore plus d'importance pour retranscrire les jugements d'experts à travers les lois a priori du modèle. Il est de plus encore plus facile de faire converger les MCMC du fait du peu de paramètres du modèle. En effet, la simulation des chaînes de Markov ne prend que quelques minutes cette fois. La partie "extension du modèle à 1 an" nécessite en revanche plus de temps mais toujours moins que pour la ligne d'activité A.

5.4.1 Preprocessing

Le triangle de développement de la LoB B est donné dans le tableau 5.15 à un coefficient multiplicatif et un bruit près, toujours pour préserver la confidentialité d'AXA. Les primes et les EAXa sont également présents pour déterminer les futures paramètres des lois *a priori*.

AY/DY	1	2	3	4	5	Prime	EAXa
2016	0	2 404	15 313	17 689	18 952	149 422	707
2017	6 709	37 345	62 212	80 560		211 763	736
2018	5 799	39 736	128 872			273 532	-6 522
2019	55 844	97 375				257 929	-2 084
2020	20 297					297 828	69 538

TABLE 5.15 – Triangle de paiement cumulé LoB B

Une visualisation du développement du triangle est donnée sur la figure 5.7

Cette visualisation des développements par années d'accident et du triangle de paiement cumulé permet de constater certains points :

- Aucune des années d'accident ne semble être développée (même si la première l'est presque)
- Il y a des EAXa négatif, donc les experts estiment des remboursements pour certaines années d'accidents
- Aucun pattern de développement ne semble pouvoir être déterminer précisément (certaines années se développement plus vite que d'autres)
- Il y a peu d'années d'accident donc on peut déjà s'attendre à une volatilité élevée autour des estimations

La méthodologie Correlated Chain Ladder telle que décrite plus haut ne prend pas en compte les réserves ni la déviation des réserves sur 1 an pour la première année d'accident. Cependant, même si cette première année d'accident n'est pas complètement développée et qu'il y a encore de l'incertitude, on peut négliger son impact sur la distribution ultime et à 1 an de toute la ligne d'activité car les EAXa estimés par les experts sont faible pour cette année d'accident.

Avant de se lancer dans l'étude avec la méthode CCL, il est important de noter que les modèles de Mack et M&W ne sont pas pertinents dans notre étude, d'une part car on ne peut pas affirmer que la première année est complètement développée et que même si celle-ci l'était les pattern de développement ne se généraliseraient pas aux autres années d'accident, et d'autre part car le volume de donnée disponible est trop faible ce qui engendre trop d'incertitude. En effet, la moyenne des réserves donnée par le modèle de Mack est de 4.62×10^5 et la volatilité de 7.94×10^8 (et donc une CoV de $1,72 \times 10^5\%$) et les ordres de grandeur sont les même pour les estimations de volatilité à 1 an de M&W.

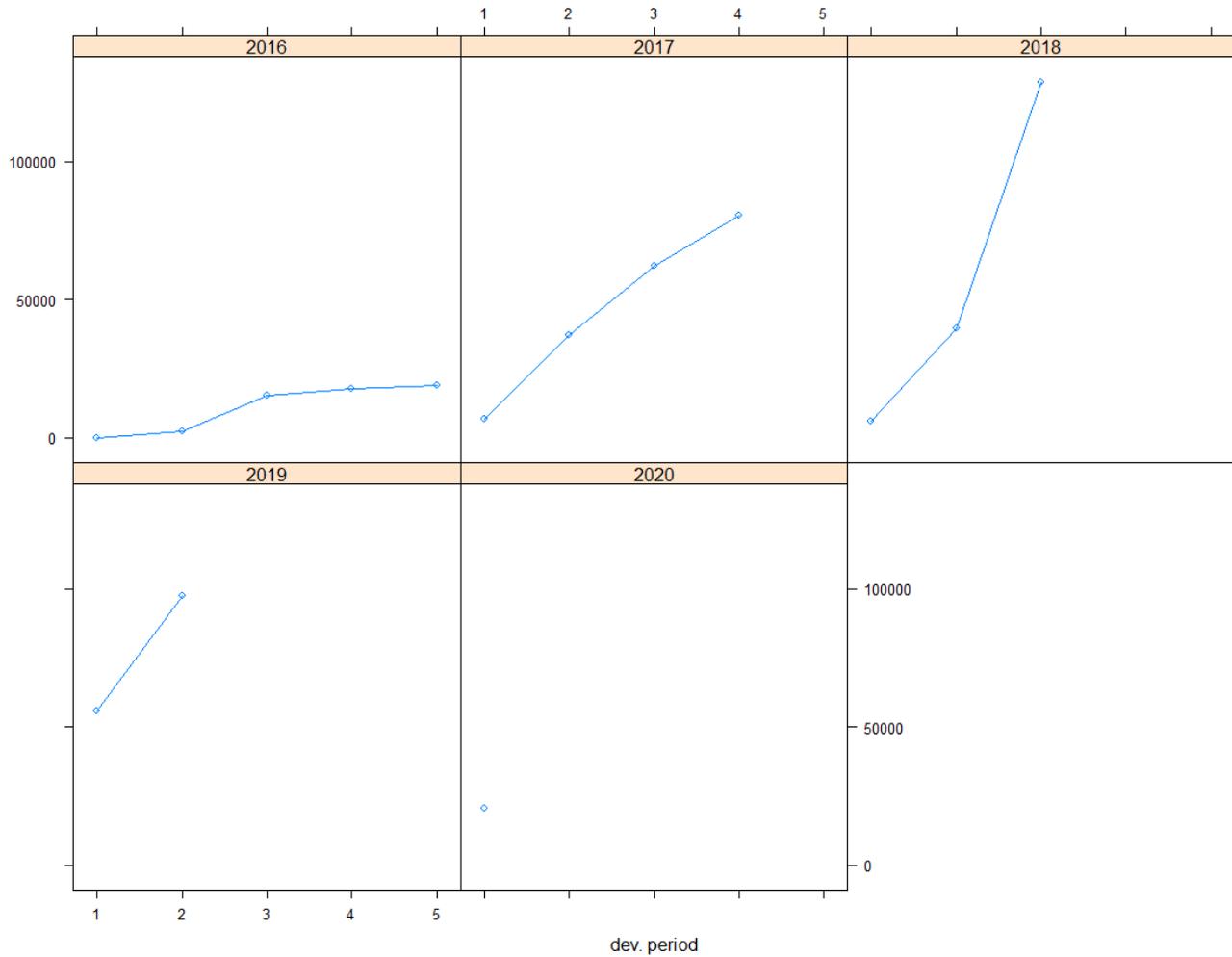


FIGURE 5.7 – Visualisation du développement du triangle LoB B

Variable	log-moyenne	log-écart-type
elr_1	-2.32	0.1
elr_2	-1.26	0.1
elr_3	-1.10	0.3
elr_4	-1.30	0.3
elr_5	-1.50	0.5

TABLE 5.16 – LoB B : Loi a priori des expected loss ratio pour paramétrages volatile

Paramétrage	α	β	Moyenne CCL Ultime	Ecart-type CCL Ultime	DIC
1	$\mathcal{U}[-1; 1]$	$\mathcal{U}[-4; 0]$	3.63×10^5	4.19×10^5	431
2	$\mathcal{U}[-3; 3]$	$\mathcal{U}[-7; 0]$	3.99×10^6	3.08×10^6	412
3	$\mathcal{U}[-3; 3]$	$\mathcal{U}[-7; 3]$	3.76×10^6	2.91×10^6	414
4	$\mathcal{U}[-10; 10]$	$\mathcal{U}[-15; 0]$	2.54×10^9	2.82×10^9	389

TABLE 5.17 – LoB B : Résultats du CCL à l'ultime pour paramétrages volatile

5.4.2 Quelques essais de paramétrages à forte volatilité

Le faible volume de donnée nous permet de tester rapidement plusieurs paramétrages avec divers lois *a priori* plus ou moins volatiles, car les temps de convergence des chaînes de Markov sont courts. Ici, nous allons essayer des paramétrages à forte volatilité, donc non-informatifs, pour voir ce que les données du triangle suggèrent comme résultat. Cependant, les lois *a priori* des elr seront centrée sur les Loss Ratio calculés à partir des EAxa.

Les 4 paramétrages que l'on propose ont tous des elr qui suivent des lois log-normales dont les paramètres sont visibles dans le tableau 5.16. Les paramètres τ qui permettent de déterminer les σ suivront toujours une loi $Beta(1, 7)$ et $\rho \sim \mathcal{U}[-1, 1]$.

Nous allons ensuite jouer sur l'amplitude des distributions uniforme des α et des β . Les résultats de la distribution des réserves à l'ultime du CCL sont présent dans le tableau 5.17. Rappelons que la somme des EAxa est de 62 375. Dans le tableau, nous voyons que laisser de la liberté aux loi des paramètres les encourage à donner une moyenne très élevé, ce qui n'est pas surprenant car la loi uniforme des paramètre se situe dans un monde logarithme donc un α de loi uniforme entre -3 et 3 multipliera le résultat par un nombre entre $exp(-3) \approx 0.05$ ou $exp(3) \approx 20$ ce qui est très volatile. Ce qui est intéressant de constater, c'est que les lois *a posteriori* des paramètres α va être focalisé sur les grandes valeurs comme on peut le voir, par exemple, sur le graphique 5.8, qui représente la distribution *a posteriori* du paramètre α_3 (de la 3ème année d'accident donc) pour le premier paramétrage. On observe des distributions similaires sur les autres α et ceci est vrai pour les 4 paramétrages. Aussi, le DIC diminue au fur et à mesure que l'amplitude des lois *a priori* augmente, ce qui incite à penser que les données sont "fittées" avec des distributions toujours en dehors des limites de celles proposées.

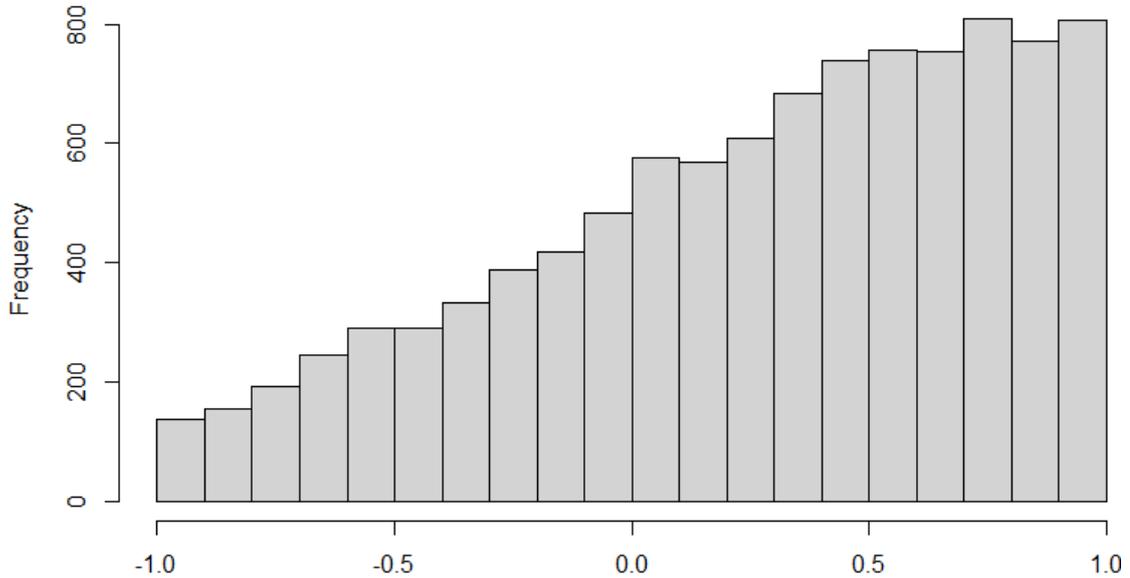


FIGURE 5.8 – LoB B : Loi a posteriori du paramètre α_3 du premier paramétrage à forte volatilité

Variable	log-moyenne	log-écart-type
elr_1	-2.03	0.01
elr_2	-0.96	0.03
elr_3	-0.81	0.05
elr_4	-1.00	0.07
elr_5	-1.20	0.09

TABLE 5.18 – LoB B : Loi a priori des expected loss ratio pour paramétrages peu volatiles

5.4.3 Paramétrages retenues : à faible volatilité

Le paramétrage de loi *a priori* que l'on va retenir ne sera pas celui qui donne le DIC le plus faible mais celui qui va proposer des amplitudes de lois *a priori* suffisamment restrictif pour que les moyennes des réserves convergent vers un résultat assez proche de celui donné par les experts via les EAXa. Ainsi, nous allons proposer un paramétrage à faible volatilité.

Le fait d'avoir des EAXa négatifs montre qu'il est possible que le développement d'une année d'accident se fasse avec des incréments négatifs. Il est donc judicieux d'autoriser les β_j à prendre des valeurs positives. Les paramètres σ et ρ suivent les mêmes lois que dans la section précédente. Les elr que l'on va utiliser sont dans le tableau 5.18 et ces derniers sont comme d'habitude supposés avec une moyenne égale au Loss Ratio visé.

Premièrement, on propose un choix assez prudent avec les $\alpha_i \sim \mathcal{U}[-0.3; 0.3]$ et

Statistique	Option 1	Option 2	Option 3
Moyenne	221 432	208 088	171 124
Ecart-type	277 963	263 406	208 407
CoV	126%	127%	122%
Médiane	161 181	150 772	129 802
Quantile 99.5%	1 419 483	1 396 041	1 012 171
DIC	441.80	444.25	482.08

TABLE 5.19 – LoB B : Résultat CCL à l’ultime

Statistique	Option 1	Option 2	Option 3
Moyenne	229 558	208 231	176 796
Best Estimate 1	-273 189	-42 616	30 123
Best Estimate 2	224 333	211 823	173 285
Ecart-type	1 546 257	887 986	767 134
CoV	674%	426 %	434 %
Médiane	-154 759	-11 313	24 060
Quantile 99.5%	8 931 754	5 022 286	2 971 772
SCR	8 707 421	4 810 463	2 798 487
SCR/BE	3 882 %	2 271%	1 615%

TABLE 5.20 – LoB B : Résultat CCL à 1 an

$\beta_j \sim \mathcal{U}[-3; 2]$ (Option 1). Deuxièmement, on peut aussi proposer un paramétrage plus centré sur les valeurs de Loss Ratio visés avec les $\alpha_i \sim \mathcal{U}[-0.1; 0.1]$ et $\beta_j \sim \mathcal{U}[-2.5; 0.5]$ (Option2). Enfin, on fait les calculs avec une 3ème option de paramétrage qui est la même que la deuxième mais avec $\tau_h \sim \text{Beta}(1, 17)$ pour avoir une moyenne des τ *a priori* beaucoup plus petite et donc des σ moins volatiles. En effet, la moyenne d’une variable aléatoire de loi Beta(a, b) est $a/(a + b)$.

Les résultats de la méthode CCL avec les 3 options à l’ultime sont présent dans le tableau 5.19 et ceux à 1 an sont présents dans le tableau 5.20.

Les formes de la distribution de perte pour cette même option à l’ultime et à 1 an sont visibles respectivement sur les graphiques 5.9 et 5.10. On peut également constater grâce au graphique 5.11 que la distribution *a posteriori* de ρ est assez volatile (comparé à celle de la LoB B). Cette volatilité est probablement la conséquence d’un faible volume de donnée qui empêche une estimation précise d’une corrélation entre année d’accident.

5.4.4 Conclusion de l’étude de la ligne B

Etant donné le très faible volume de donnée, il est difficile de considérer la méthode CCL comme la méthode officielle pour estimer le risque de réserves à l’ultime et à 1 an. Mais cette dernière peut être intéressante comme benchmark pour remettre en question

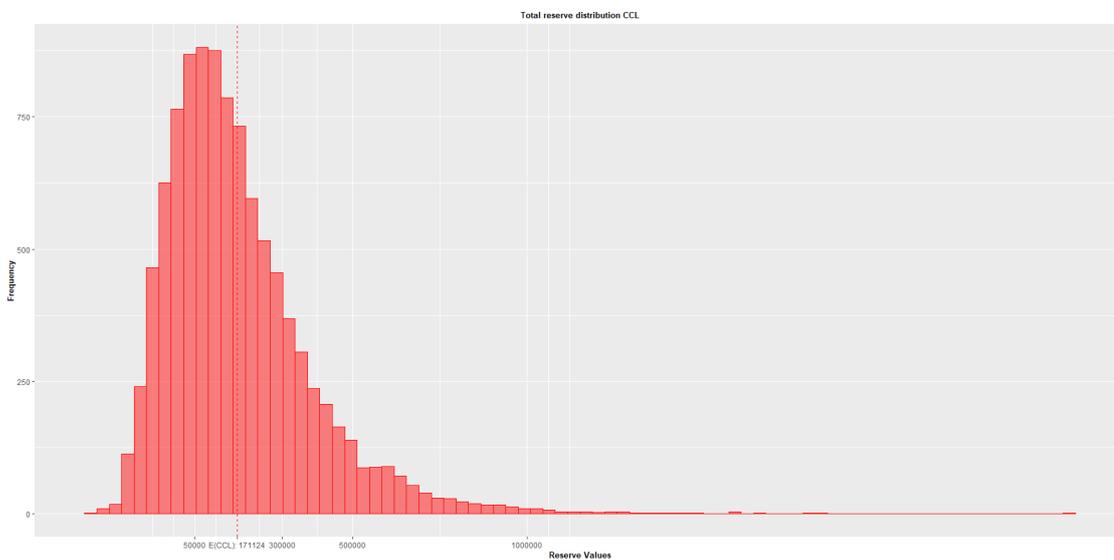


FIGURE 5.9 – LoB B : CCL à l'ultime

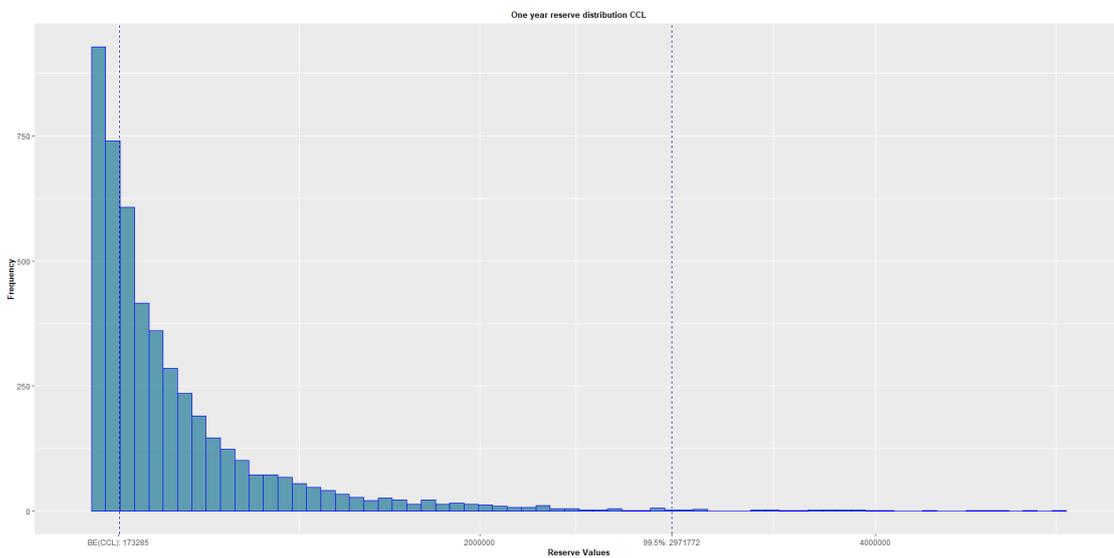


FIGURE 5.10 – LoB B : CCL à 1 an

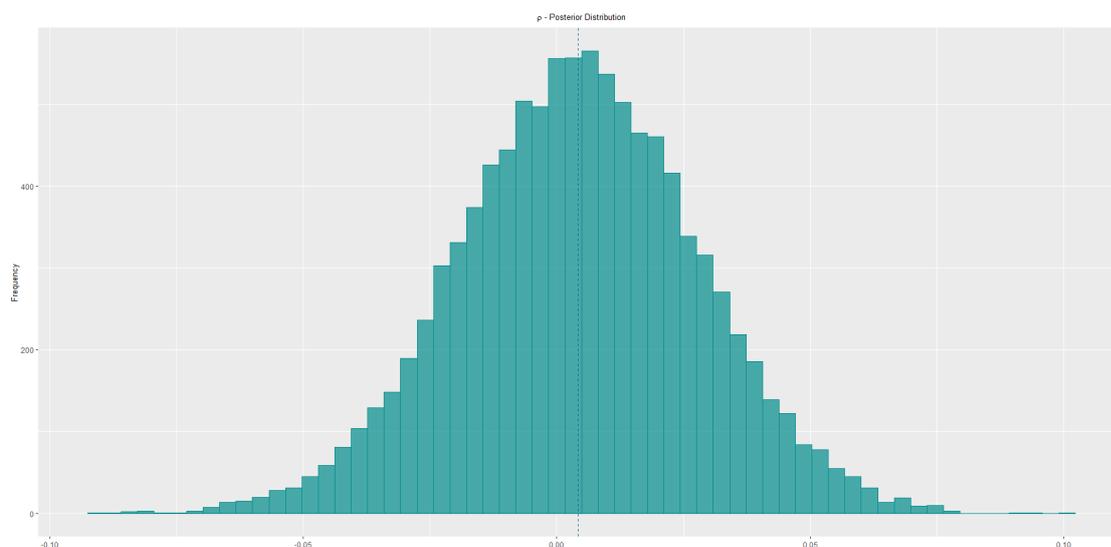


FIGURE 5.11 – LoB B : Distribution *a posteriori* du paramètre ρ

les avis des experts sur le choix de la volatilité.

Sur cette ligne d'activité les experts ont estimé des EAXa négatif, alors que le pattern de paiement se dirigeait vers des réserves à l'ultime positif. La volatilité sur les résultats vient en partie de cette inadéquation. Si les experts sont certains des réserves à provisionner (c'est-à-dire les EAXa) alors ils peuvent ignorer les résultats sur les moyennes (ultime et 1 an) de la méthode CCL.

D'autre part, cette étude de la LoB B qui suit l'étude de la LoB A nous permet de nous rendre compte de la forme des distributions des réserves à l'ultime et à 1 an. On peut dire que celles-ci ressemblent beaucoup à une loi log-normale ce qui n'est pas absurde car il s'agit d'une somme, presque indépendante (car $\rho \approx 0$), de lois log-normales et (d'après [4]) cette somme peut s'apparenter à une loi log-normale. Une autre fait intéressant est que cette distribution "quasi-log-normale" est beaucoup plus volatile dans le cas de la ligne B et nous pouvons bien le constater en comparant les graphiques des LoB A et B. Le graphique 5.12 illustre comment se comporte une distribution log-normale en fonction de sa volatilité et nous permet de catégoriser la LoB B comme ayant une distribution log-normale à forte volatilité et inversement pour la LoB A.

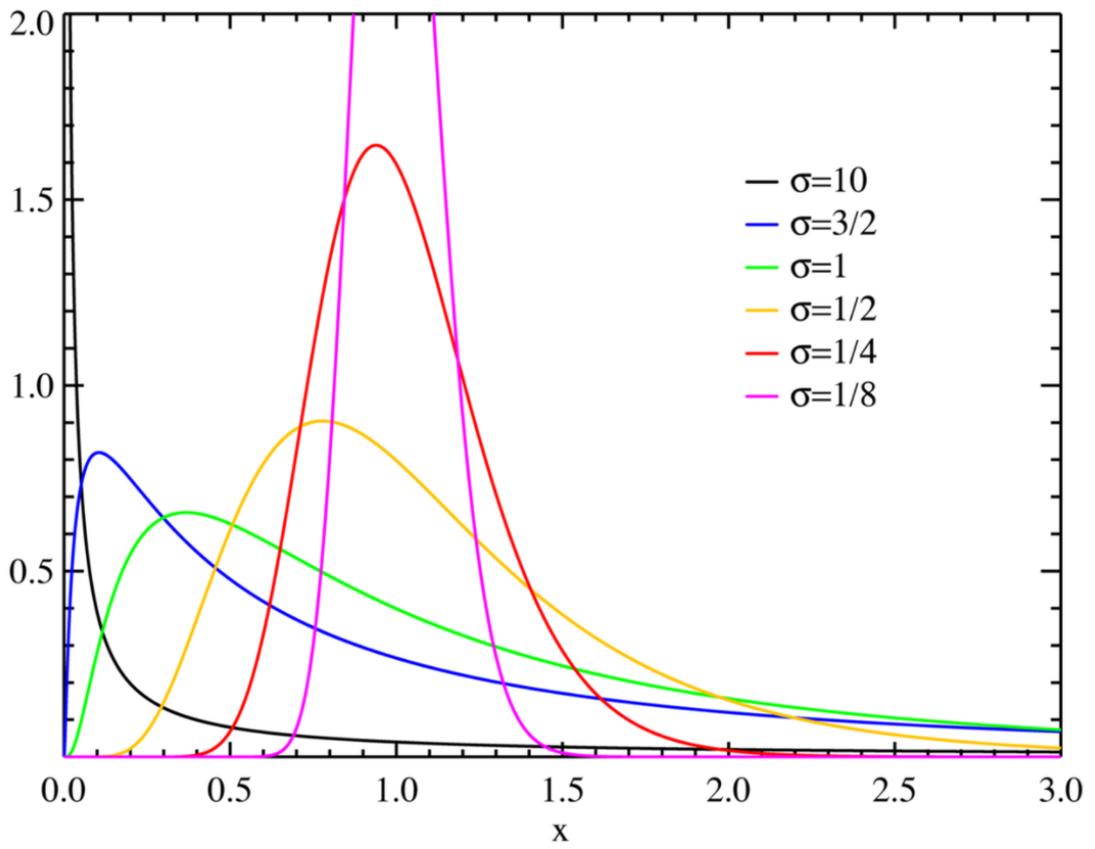


FIGURE 5.12 – Loi log-normales de différents paramètres de volatilité

Conclusion

L'utilisation des statistiques bayésiennes et de la méthode CCL sur les triangles de réserve permet d'obtenir une distribution complète pour les pertes à l'ultime ou pour les boni-mali pour toute la ligne d'activité. On peut alors facilement calculer les statistiques (moyenne, écart-type, quantiles ...) sur cette distribution résultante et en avoir une visualisation précise, en particulier aux alentours des valeurs extrêmes. Nous avons vu que, subséquentement aux hypothèse log-normales du modèle, les distributions à l'ultime et à 1 an des Reseves des lignes d'activité suivent également une loi proche d'une loi log-normale. La lognormalité des distributions de pertes est une hypothèse classique en actuariat (comme nous l'avons vu, c'est avec cette hypothèse que AXA estime son risque de réserve) et elle revient dans la méthode CCL.

D'autre part, la facile manipulation et la grande variété des hypothèses possibles sur les lois a priori permet de tenir compte des connaissances de l'expert à propos de l'activité concerné. Et souvent d'ailleurs, des avis d'expert sont disponibles à propos d'un triangle concerné. Dans nos études, ils se sont traduit par la présence de EAXa, qui sont les Best Estimate des pertes à l'ultime auxquels on ajoute une marge d'erreur, et aussi par la présence de certaines volatilité autour de ces estimations. En revanche, nous ne pouvons pas "forcer" certains paramètres à se rapprocher d'une certaine loi (comme choisir une corrélation positive et proche de 0.5 entre année d'accident) car les chaînes de Markov ne convergeraient pas. Il est également important de se rendre compte que les lois *a priori* ont une forte influence sur la distribution finales, particulièrement quand le volume de données est faible (comme dans la LoB B), et qu'il faut donc attacher une attention particulière à la sélection de ces lois. Ajoutons à cela que la fonction de vraisemblance du modèle peut être manipulée également de façon assez souple et a l'avantage de permettre de s'affranchir de certaines hypothèses, parfois assez restrictives d'autres modèles comme celui de Mack (indépendance entre année d'accident, unicité des facteurs de développement par années de développement ...).

Un autre avantage des statistiques bayésienne en général est qu'elle permet d'obtenir beaucoup de simulations d'une prédiction malgré un nombre d'observation assez faible. Ceci est due à l'utilisation des lois a priori qui augmente artificiellement le volume de données. Comme les données sont souvent peu nombreuse dans un triangle de perte, soit car il n'y a pas eu beaucoup d'années d'historique, soit car les premières années d'acci-

dent sont trop récentes et pas assez développées, l'utilisation de la théorie bayésienne est particulièrement adaptée au problème. Dans les deux ligne d'activité étudiées et qui ont peu d'historique, la méthode CCL a permis d'obtenir une estimation assez pertinente des distributions de pertes réserve, en particulier pour la LoB A. Les résultats du CCL sur la LoB B peuvent difficilement être utilisés dans le cadre de Solvabilité II, et on privilégiera pour cette LoB une méthode dite "Benchmark" utilisant les volatilités des triangles publiques similaire à cette activité. Néanmoins le CCL peut être un bon test de sensibilité pour des études comparatives sur les résultats des calibrations de la LoB B.

On peut cependant retenir deux inconvénients de la méthode CCL. Le premier est qu'elle est très coûteuse en temps de calcul et que le nombre de paramètre à estimer augmente très vite avec la profondeur du triangle. Le deuxième est qu'elle nécessite de définir un grand nombre de paramètre ce qui peut poser problème car cela nécessite beaucoup de justifications de la part de ceux qui font la calibration. Autrement, il n'est pas toujours facile de retranscrire les jugements d'expert en calibration de paramètres. Il serait alors nécessaire de rédiger un bon "guide" de calibration qui s'appliquent à toutes les lignes utilisant la méthodologie et cela impliquerait une revue complète du calcul du risque de réserve.

En conclusion, on peut retenir que la méthode CCL permet une autre approche alternative pertinente et stable de l'estimation de la distribution de perte des réserves à l'ultime et à 1 an, à condition d'une part de bien connaître la ligne d'activité étudiée pour bien définir les calibrations et que la profondeur du triangle étudié ne soit pas trop grande pour contrôler le temps de calcul.

Annexe A

Justification des méthodes de simulations

A.1 Méthode de rejet

Nous cherchons à simuler une variable Y de support $\mathbb{X} = \mathbb{R}$ et de distribution π . Nous utilisons pour cela une distribution dominante et de même support que l'on sait simuler q . On tire X suivant la loi q . Indépendamment, on tire une variable aléatoire uniforme U sur $[0, 1]$.

On définit l'évènement $\{X \text{ est accepté}\} = \left\{U \leq \frac{\pi(X)}{Mq(X)}\right\}$

Lemma A.1.1. Si A est une variable aléatoire de support $[0, 1]$ et U une variable aléatoire de loi uniforme sur $[0, 1]$ indépendante de A . Alors

$$\mathbb{P}(U \leq A) = \mathbb{E}(A)$$

Démonstration. Notons f_A et f_U les densités de probabilité de A et U . Par indépendance, on a $\forall a, u \in [0, 1], f_{(A,U)}(a, u) = f_A(a)f_U(u)$ et donc :

$$\begin{aligned}\mathbb{P}(U \leq A) &= \iint_{\substack{u \leq a \\ (a,u) \in [0,1]^2}} f_{(A,U)}(a, u) da du \\ &= \int_{a \in [0,1]} \left(\int_{u \in [0,a]} f_U(u) du \right) f_A(a) da \\ &= \int_{a \in [0,1]} \mathbb{P}(U \leq a) f_A(a) da \\ &= \int_{a \in [0,1]} a f_A(a) da \\ &= \mathbb{E}(A)\end{aligned}$$

□

Lemma A.1.2. La probabilité d'accepter une simulation de X est $\frac{1}{M}$

Démonstration.

$$\begin{aligned}
\mathbb{P}(\{X \text{ est accepté}\}) &= \mathbb{P}(U \leq \pi(X)/(Mq(X))) \\
&= \mathbb{E}[\pi(X)/(Mq(X))] && \text{(d'après le lemme 1)} \\
&= \frac{1}{M} \int_{-\infty}^{\infty} \frac{\pi(x)}{q(x)} q(x) dx \\
&= \frac{1}{M} \int_{-\infty}^{\infty} \pi(x) dx \\
&= 1/M
\end{aligned}$$

□

Theorem A.1.3. La procédure de rejet décrite ci-dessus permet de générer Y suivant la loi π .

Démonstration. Pour tout $y \in \mathbb{R}$ et $\epsilon > 0$:

$$\begin{aligned}
\mathbb{P}(Y \in (y, y + \epsilon)) &= \mathbb{P}(X \in (y, y + \epsilon) | X \text{ est accepté}) \\
&= \frac{\mathbb{P}(X \text{ est accepté} | X \in (y, y + \epsilon)) \mathbb{P}(X \in (y, y + \epsilon))}{\mathbb{P}(X \text{ est accepté})} \\
&\approx \frac{\mathbb{P}(U \leq \pi(X)/(Mq(X)) | X \in (y, y + \epsilon)) (\epsilon q(y))}{1/M} \\
&\approx \epsilon M q(y) \mathbb{P}(U \leq \pi(y)/(Mq(y))) \\
&= \epsilon M q(y) \pi(y) / (Mq(y)) \\
&= \epsilon \pi(y)
\end{aligned}$$

Les approximations ci-dessus correspondent au cas $\epsilon \rightarrow 0$. Ainsi par un argument classique de passage à la limite on en déduit le résultat.

□

A.2 Échantillonnage préférentiel

Theorem A.2.1. Soit $X \sim \pi$, $X_1, \dots, X_n \sim q$ et $f : E \mapsto \mathbb{R}^k$ continue, où E désigne un ensemble incluant le support image de X . π^* est une distribution non normalisée de normalisation π .

L'estimateur $\mu_n = \frac{\sum_{i=1}^n W_i f(X_i)}{\sum_{i=1}^n W_i}$, avec $W_i = \pi^*(X_i)/q(X_i)$ converge vers $E[f(X)]$ avec une probabilité 1.

Démonstration. D'une part, d'après la loi des grands nombre :

$$\frac{1}{n} \sum_{i=1}^n W_i f(X_i) \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}_q\left(\frac{\pi^*(X_1)}{q(X_1)} f(X_1)\right)$$

Il en est de même pour le dénominateur :

$$\frac{1}{n} \sum_{i=1}^n W_i \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \mathbb{E}_q\left(\frac{\pi^*(X_1)}{q(X_1)}\right)$$

Nous avons alors

$$\mu_n \xrightarrow[n \rightarrow +\infty]{\text{p.s.}} \frac{\mathbb{E}_q\left(\frac{\pi^*(X_1)}{q(X_1)} f(X_1)\right)}{\mathbb{E}_q\left(\frac{\pi^*(X_1)}{q(X_1)}\right)}$$

D'autre part nous avons :

$$\begin{aligned} \mathbb{E}_\pi[f(X)] &= \int f(x)\pi(x)dx \\ &= \int \left(f(x) \frac{\pi(x)}{q(x)}\right) q(x)dx \\ &= \mathbb{E}_q\left[f(X_1) \frac{\pi(X_1)}{q(X_1)}\right] = \mathbb{E}_q\left[f(X_1) \frac{\int \pi^*(x)dx}{q(X_1)}\right] \\ &= \frac{\mathbb{E}_q\left[f(X_1) \frac{\pi^*(X_1)}{q(X_1)}\right]}{\int \frac{\pi^*(x)}{q(x)} q(x)dx} \\ &= \frac{\mathbb{E}_q\left(\frac{\pi^*(X_1)}{q(X_1)} f(X_1)\right)}{\mathbb{E}_q\left(\frac{\pi^*(X_1)}{q(X_1)}\right)} \end{aligned}$$

D'où le résultat du théorème. □

A.3 Metropolis Hastings

On va admettre, pour la démonstration du bon fonctionnement de l'algorithme MH, les résultats sur les chaînes de Markov. Le lecteur intéressé pourra consulter des éléments de cours sur les chaînes de Markov ou consulter simplement la page wikipédia [2] sur les chaînes de Markov à temps discret.

Considérons le processus à temps discret $(X_t)_{t \geq 0}$ produit par l'algorithme Metropolis-Hastings. On garde les mêmes notations que dans la section 3.3.2.

Theorem A.3.1. $(X_t)_{t \geq 0}$ est une chaîne de Markov apériodique, récurrente positive et irréductible.

Démonstration. Par construction, $(X_t)_{t \geq 0}$ est une chaîne de Markov car l'état à l'instant t n'est généré qu'à partir de la connaissance de l'état à l'instant $t-1$. Elle est irréductible et récurrente positive car à chaque itération nous avons une probabilité strictement positive

de d'atteindre un point du support de la "proposal distribution", qui est le même que l'espace d'état de la chaîne. Un état X' va voir une probabilité d'acceptation strictement inférieur à 1, donc la chaîne va rester dans l'état X' avec une probabilité positive et donc sa période est 1. Dans les chaîne de Markov irréductible toute les états ont la même période (ici égale à 1), donc la chaîne est apériodique. \square

Une conséquence directe de ce théorème est que $(X_t)_{t \geq 0}$ va converger vers une unique distribution stationnaire. Il reste alors à montrer que cette distribution stationnaire est bien la distribution cible π . Pour cela, on va se servir du théorème suivant :

Theorem A.3.2. La chaîne $(X_t)_{t \geq 0}$ est réversible par rapport à π .

Démonstration. Il suffit, pour prouver le théorème de montrer que π satisfait l'équation d'équilibre :

$$\pi(X_{t-1})K(X_{t-1} \rightarrow X_t) = \pi(X_t)K(X_t \rightarrow X_{t-1})$$

Le cas $X_t = X_{t-1}$ est trivial.

Si $X_t \neq X_{t-1}$:

D'autre part nous avons :

$$\begin{aligned} \pi(X_{t-1})K(X_{t-1} \rightarrow X_t) &= \pi(X_{t-1})q(X_{t-1} \rightarrow X_t)\alpha(X_{t-1}, X_t) \\ &= \pi(X_{t-1})q(X_{t-1} \rightarrow X_t)\min\left(1, \frac{\pi(X_t)q(X_t \rightarrow X_{t-1})}{\pi(X_{t-1})q(X_{t-1} \rightarrow X_t)}\right) \\ &= \min(\pi(X_{t-1})q(X_{t-1} \rightarrow X_t), \pi(X_t)q(X_t \rightarrow X_{t-1})) \end{aligned}$$

En décomposant de la même manière la partie droite de l'équation d'équilibre, on arrive finalement au même résultat ce qui conclut la démonstration. \square

Annexe B

Exemple d'implémentation de MCMC

Nous allons ici illustrer l'algorithme de Metropolis-Hasting pour simuler une loi normale centrée réduite $\mathcal{N}(0, 1)$ de densité de probabilité $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Notons que ce n'est pas la meilleure méthode pour simuler cette loi. On préférera pour cela l'algorithme de Box-Muller.

La fonction Kernel sera ici la loi uniforme $\mathcal{U}[-1; 1]$ et donc la fonction qui va déterminer si on accepte ou non l'état suivant sera :

$$\alpha(x, y) = \exp\left(-\frac{x^2 - y^2}{2}\right)$$

Nous faisons comme si nous ne connaissions pas la moyenne (égale à 0) et nous allons partir de l'état initial $s_0 = -5$.

Le code R serait alors le suivant :

```
f = fonction(x){exp(-x^2/2)/sqrt(2*pi)}
AcceptanceFunction = fonction(x,y){exp(-(x^2-y^2)/2)}
nSim = 1e6

s = -5
X = vector(length = nSim)

for(n in 1:nSim){

  propState = runif(n = 1, min = s-1, max = s+1)
  u = runif(n = 1)
  a = AcceptanceFunction(propState,s)

  if(u<a){
```

```

    s = propState
  }

  X[n] = s
}

burnIn = 100
thinning = 30

IndexUsed = seq(from = burnIn, to = nSim, by = thinning)
Simulations = X[IndexUsed]

acf(X)
hist(Simulations)

```

Nous vérifions grâce à la fonction *acf* du code l'autocorrélation de la chaîne de Markov que le "thinning" de 30 est suffisant pour créer des simulations indépendantes grâce à la figure B.1. D'autre part la figure B.2 donne la représentation graphique de la distribution simulée de la loi normale grâce à la méthode MCMC de Metropolis-Hastings. On constate que c'est bien une loi normale centrée réduite.

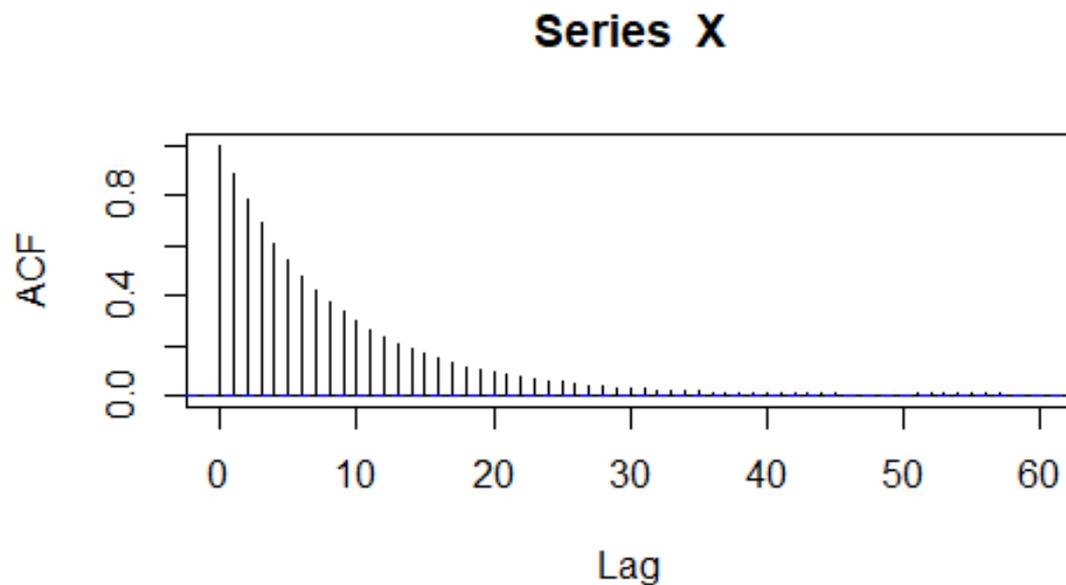


FIGURE B.1 – Autocorrelogramme dans la simulation de la loi normale

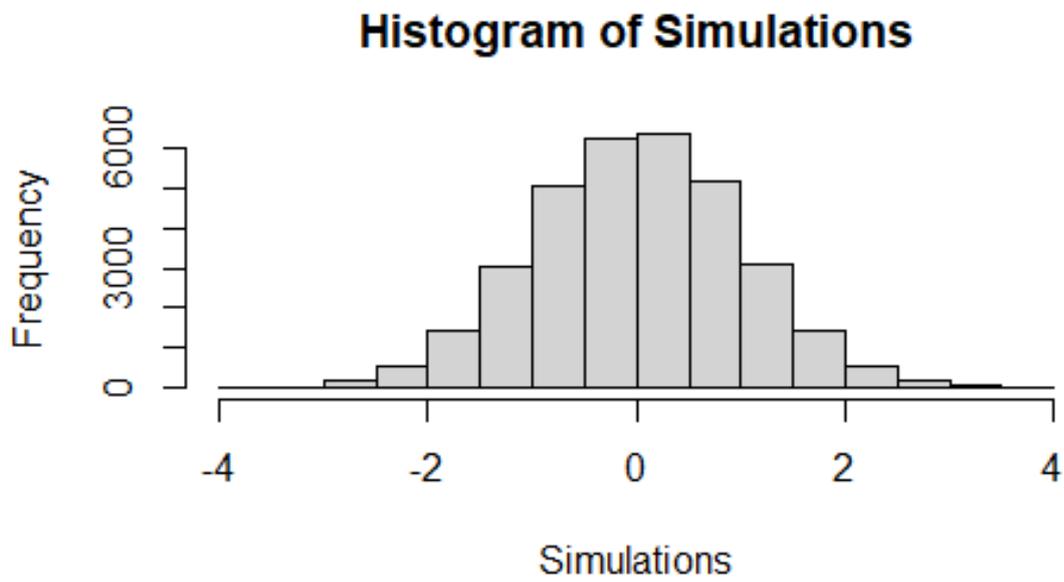


FIGURE B.2 – Simulation de la loi normale grâce à Metropolis-Hastings

Annexe C

Implémentation du modèle CCL dans STAN

```
//##### stan.model #####
data{

  int <lower=1> ndata; // number of rows with data
  int<lower=0, upper=1> originlid[ndata];
  int<lower=1> obs;
  real logprem[ndata];
  real logloss[ndata];
  int<lower=1> origin[ndata]; // origin period
  int<lower=1> dev[ndata]; // development period
}

transformed data{
  int norigin = max(origin);
  int ndev = max(dev);
}
parameters{
  real <lower = -0.5, upper = 0.5> ralpha[norigin - 2];
  real <lower=-3, upper = 0> rbeta[ndev - 1];
  real <lower=0.001> elr[norigin-1];
  //real<lower=0> a[n_dev];
  real<lower=-1, upper=1> rho;
  real <lower=0, upper=1> a[ndev];
}

transformed parameters{
  real mu[ndata];
  real alpha[norigin];
```

```

real beta[ndev];
real sig2[ndev];
real loglossratio[norigin];
//real sig[n_dev];
//real <lower=-1, upper=1> rho;

//rho = -2*rho_r + 1;

//real sigma_trasf[n_dev] = sig2[n_dev];

alpha[1] = 0;
alpha[2] = 0;

for (i in 3:norigin){
  alpha[i] = ralpha[i-2];
}
for (i in 1:(ndev - 1)){
  beta[i] = rbeta[i];
}

beta[ndev] = 0;

for (i in 1:ndev){
  sig2[i] = sum(a[i:ndev]);
}

loglossratio[1] = 0.15363793;

for (i in 2:norigin){

  loglossratio[i] = log(elr[i-1]);
}

mu[1] = logprem[1] + loglossratio[1] + beta[dev[1]];
for (i in 2:ndata){
  mu[i] = logprem[i] + loglossratio[origin[i]] + alpha[origin[i]] + beta[dev[i]] +
  rho*(logloss[origin[i-1]] - mu[origin[i-1]]);
}
}

```

```

model{

  // log loss ratio priors

  elr[2] ~ lognormal(-0.84314719, 0.000005);
  elr[3] ~ lognormal(-0.98540370, 0.001);
  elr[4] ~ lognormal(-0.87179709, 0.008);
  elr[5] ~ lognormal(-0.77147737, 0.025);
  elr[6] ~ lognormal(-0.73145308, 0.035);
  elr[7] ~ lognormal(-0.51355728, 0.05);
  elr[8] ~ lognormal(-1.43059559, 0.08);
  elr[9] ~ lognormal(-0.35926847, 0.1);

  for(i in 1:(norigin - 2)){
    ralpha[i] ~ uniform(-0.5,0.5);
  }
  rbeta ~ uniform(-3,0);
  a ~ beta(1,7);
  rho ~ uniform(-1,1);

  for (i in 1:(ndata)){
    logloss[i] ~ normal(mu[i], sqrt(sig2[dev[i]]));
  }

}

```


Annexe D

Diagnostic des MCMC

D.1 Potential Scale Reduction \hat{R}

A l'équilibre, la distribution des échantillons doit être la même partout dans la même chaîne et dans des chaînes différentes. Une façon de vérifier que cette distribution est bien une distribution d'équilibre est d'une part de regarder la moyenne de différentes chaînes et de vérifier l'égalité des moyennes, et d'autre part de couper une chaîne en deux et de vérifier que la moyenne est la même dans chaque partie de la chaîne. La valeur \hat{R} est le résultat d'un test statistique d'égalité des moyennes.

D.2 Effective sample size

La taille de l'échantillon effectif ou "Effective sample size" (ESS) peut s'interpréter comme une estimation du nombre de simulations indépendantes d'une chaîne de Markov.

Considérons une chaîne de Markov stationnaire qui a atteint son état d'équilibre X_1, X_2, \dots et dont la distribution d'équilibre a une densité f , une moyenne μ et un écart-type σ . Notons $\rho_t = \frac{1}{\sigma^2} \int_{\mathbb{R}} (X_n - \mu)(X_{n+t} - \mu) f(x) dx$ l'autocorrélation pour le décalage t (qui ne dépend pas ici de n par stationnarité).

L'ESS de N simulations d'une chaîne de Markov d'autocorrélation ρ_t s'écrit :

$$N_{eff} = \frac{N}{\sum_{t=-\infty}^{\infty} \rho_t}$$

D.3 Bayesian Fraction of Missing Information (BFMI)

La description de cette mesure est assez technique. Elle est détaillée dans l'article [3].

Annexe E

Stabilité des modèles de Mack et
M&W

Statistique Mack	Standard	Avec changement de graine	Avec bruit blanc
Moyenne	313 470	313 470	313 330
Ecart-type	116 405	116 405	116 710
CoV	37%	37%	37%
Statistique M&W avec LN	Standard	Avec changement de graine	Avec bruit blanc
Moyenne	314 603	314 603	314 468
Ecart-type	106 382	106 382	106 742
CoV	34%	34%	34%
Médiane	298 025	298 025	297 780
Quantile 99.5%	695 563	695 563	697 126
SCR	380 960	380 960	382 658
SCR/BE	121%	121%	122%

TABLE E.1 – Résultat des tests de stabilité du modèle de Mack et M&W

Bibliographie

- [1] Insurance speaker wavestone. <https://www.insurancespeaker-wavestone.com/2014/03/solvabilite-2-enjeux-et-contraintes/>.
- [2] Markov chain wikipédia page. https://en.wikipedia.org/wiki/Markov_chain.
- [3] Michael Betancourt. Diagnosing Suboptimal Cotangent Disintegrations in Hamiltonian Monte Carlo. 2016.
- [4] Maurin Michel El Faouzi. Sur la loi de sommes de variables log-normales : application à la fiabilité de temps de parcours routiers.
- [5] P. D. England and R. J. Verrall. Analytic and bootstrap estimates of prediction errors in claims reserving. *Insurance : Mathematics and Economics*, 25 :281–293, 1999.
- [6] Carnevale Giulio Ercole and Clemente Gian Paolo. A Bayesian Internal Model for Reserve Risk : An Extension of the Correlated Chain Ladder. *Risks*, 2020.
- [7] Robert Kohnd Gareth W. Peters, Alice X.D. Dongc. A copula based Bayesian approach for paid–incurred claims models for non-life insurance reserving. *CAS MONOGRAPH SERIES*, (1), 2019.
- [8] S.T. Boris Choy Jennifer S.K. Chan and Udi E. Makov. Robust Bayesian Analysis of Loss Reserves Data using the Generalized-t Distribution.
- [9] Bornhuetter R. L. and Ferguson R. E. The Actuary and IBNR. *Casualty Actuarial Society*, 1972.
- [10] T. Mack. The standard error of chain ladder reserve estimates : Recursive calculation and inclusion of a tail factor. *Astin Bulletin*, 29 :361–366, 1999.
- [11] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. 1993.
- [12] Andrew Gelman Matthew D. Hoffman. The No-U-Turn Sampler : Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Department of Statistics Columbia University*, 2011.
- [13] G. Meyers. Stochastic Loss Reserving using Bayesian MCMC Models. *CAS MONOGRAPH SERIES*, (1), 2015.
- [14] G. Meyers. Stochastic Loss Reserving using Bayesian MCMC Models 2nd Edition. *CAS MONOGRAPH SERIES*, (1), 2019.

- [15] Y. Nesterov. Primal-dual subgradient methods for convex problems. *Mathematical programming*, 2009.
- [16] Nicola G. ; Carlin Bradley P. ; van der Linde Angelika Spiegelhalter, David J. ; Best. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 2002.
- [17] M. WÜTHRICH and MERZ. Modelling the Claims Development Result for Solvency Purposes. *Casualty Actuarial Society E-Forum*, page 542–568, 2008a.
- [18] Martin J. Wainwright Bin Yu Yuansi Chen, Raaz Dwivedi. Fast mixing of Metropolized Hamiltonian Monte Carlo : Benefits of multi-step gradients.