

**Mémoire présenté le :**

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

Par : Oskar Laverny

Titre Provisionnement et risque de réserve en assurance construction

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

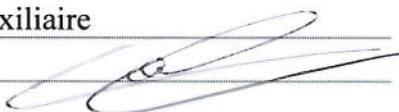
*Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus*

*Membres présents du jury de l'Institut  
des Actuaires*

Signature

Entreprise :

Nom : L'auxiliaire

Signature : 

Nom : Maxime LENFANT

Signature : 

Invité :

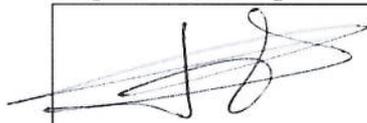
Nom :

Signature :

*Membres présents du jury de l'ISFA*

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

Signature du responsable entreprise



Signature du candidat





L'Auxiliaire  
50 Cours Franklin Roosevelt  
69006 Lyon

I.S.F.A  
50 Avenue Tony Garnier  
69007 Lyon



---

# Provisionnement et risque de réserve en assurance construction

---

Oskar LAVERNY

*Sous la direction de Pierre THEROND et Maxime LENFANT*

Le 10 septembre 2018



## Résumé

Après avoir précisé les spécificités de l'assurance construction française et analysé les problématiques posées par les provisions supplémentaires spécifiques à cette branche d'assurance non-vie, nous prendrons le temps de rappeler les modèles de bases, déterministes comme stochastiques, utilisés en provisionnement non-vie. Puis, ayant transcrit les modèles principaux dans un cadre mathématique plus large, nous présenterons une adaptation de ces modèles au cadre tri-dimensionnel de l'assurance construction française, comprenant l'estimation de variabilité à horizon quelconque. Enfin, par plusieurs approches différentes, nous dégagerons des estimateurs du risque de réserve en normes Solvabilité II dans ce cadre, et nous concluons par une analyse de ces estimateurs et une comparaison avec la formule standard sur un portefeuille.



### **Abstract**

After having specified the specificities of the French builder's insurance and analyzed the problems posed by additional specific reserves to this line of business, we recall basic models, deterministic and stochastic, used in non-life insurance. Having transcribed the main models in a larger mathematic framework, we present new models shaped for the tri-dimensional issue of French construction insurance, including the estimation of variability in every point of view. Then, through several different approaches, we derive Solvency II reserve risk estimators in those models, and conclude with an analysis of these estimators on a certain portfolio, shaped toward the standard formula.



# Remerciements

Je remercie chaleureusement toutes les personnes qui m'ont accompagné dans la rédaction de ce mémoire. En premier lieu l'Auxiliaire et tout particulièrement Jérémie Garrot, pour m'avoir donné l'opportunité de réaliser ce travail au sein de la Division Actuariat de l'Auxiliaire.

Merci également à Pierre Théron et à Maxime Lenfant pour l'encadrement, le temps investi, et les nombreuses discussions passionnantes sur le sujet.

Ich bedanke mich herzlich bei Lukas Hahn für den spannenden Austausch rund um das Braun-Modell.

Thanks to Markus Gesmann for the discussions about Mack's bootstrap and the future of the ChainLadder R package.

Enfin, merci à Isabelle pour le soutien inconditionnel dont elle a su faire preuve.

*À mes parents,*



# Sommaire

<b>Introduction</b>	<b>1</b>
<b>1 Cadre général de l'étude</b>	<b>3</b>
1.1 Cadre réglementaire et principes généraux de l'assurance construction française	3
1.1.1 Historique de l'assurance construction	3
1.1.2 Cadre réglementaire actuel de l'assurance construction	5
1.1.3 Les garanties de l'assurance construction	8
1.1.4 Provisions spécifiques à l'assurance construction	9
1.2 L'assurance construction sous Solvabilité II	10
1.2.1 Principe de <i>Best estimate</i>	11
1.2.2 Principe de <i>Risk Margin</i>	11
1.2.3 Minimums réglementaires de capitaux propres	12
1.3 Solvabilité II et le risque de réserve	12
1.3.1 Principe de risque à un an	12
1.3.2 Risque de réserve en non-vie	13
1.3.3 Paramètres propres à l'entreprise	14
1.4 Données, cahier des charges et problématique	15
<b>2 Provisionnement non-vie : de l'algorithme aux modèles complets</b>	<b>17</b>
2.1 Algorithme et modèles de base	18
2.1.1 Algorithme <i>Chain-ladder</i>	18
2.1.2 Modèle de Mack	19
2.1.3 Modèle de Merz-Wüthrich	20
2.2 Réécriture du modèle de Mack	22
2.2.1 Retour sur les (quasi-)Modèles Linéaires Généralisés	23
2.2.2 Modèle de Poisson sur-dispersé	24
2.2.3 Modèle binomial négatif et approximation normale	26
2.2.4 Conclusion	28
2.3 Le Bootstrap en provisionnement	28
2.3.1 Brefs rappels sur le Bootstrap	29
2.3.2 Dans le cadre des quasi-Modèles Linéaires Généralisés	29
2.3.3 Sous les hypothèses du quasi-Modèles Linéaires Généralisés de Mack	30
2.3.4 Principe de <i>zonage</i> des quasi-Modèles Linéaires Généralisés	31
2.4 Quelques extensions multivariées	32
2.4.1 Deux approches : Modèle de Schnauss et Modèle de Braun	33
2.4.2 Modèle de Braun	34
2.4.3 Bootstrap dans le modèle de Braun : Extension	35
2.5 Ajustements classiques	36
2.5.1 Bornhuetter-Fergusson : intégration d'un <i>a-priori</i> bayésien	37
2.5.2 Stabilisation des coefficients de passage et curve-fitting	37
2.5.3 Vision stochastique de ces ajustements	38

2.5.4	Dans un cadre Bootstrap . . . . .	39
<b>3</b>	<b>Assurance construction : un problème tri-dimensionnel</b>	<b>41</b>
3.1	Les trois dimensions de l'assurance construction . . . . .	41
3.1.1	Méthode règlementaire . . . . .	43
3.1.2	Modélisation en deux dimensions en un temps . . . . .	44
3.1.3	Modélisation en deux dimensions et en deux temps . . . . .	44
3.2	Extension des modèles existants . . . . .	45
3.2.1	Vision poissonnienne . . . . .	45
3.2.2	Recherche d'un ordre 2 . . . . .	47
3.2.3	Impossibilité des modèles récursifs . . . . .	48
3.3	Distributions complète par bootstrap . . . . .	49
3.3.1	Cas poissonien . . . . .	49
3.3.2	Cas de la modélisation en deux temps . . . . .	49
3.3.3	Conclusion : risque de réserve en assurance construction . . . . .	51
<b>4</b>	<b>Sigma de réserve : recherche d'un estimateur</b>	<b>53</b>
4.1	Méthodes règlementaires . . . . .	53
4.1.1	Méthode 1 : Log-normalité . . . . .	53
4.1.2	Méthode 2 : Merz-Wüthrich . . . . .	58
4.2	Bootstrap récursif multivarié : le modèle de base. . . . .	62
4.2.1	Principes du modèle . . . . .	62
4.2.2	Vérification des hypothèses . . . . .	65
4.2.3	Résultats . . . . .	66
4.3	Bootstrap récursif multivarié : raffinements . . . . .	70
4.3.1	Modèle de charges et modèle de règlements . . . . .	71
4.3.2	Zonage en capitalisation . . . . .	71
4.3.3	Stabilisation des coefficients de passage . . . . .	72
4.3.4	Rééchantillonnage paramétrique ou non-paramétrique . . . . .	72
4.3.5	Exclusion de résidus . . . . .	74
4.3.6	Bornhuetter-Fergusson en capitalisation . . . . .	74
4.4	Comparaison des résultats des différentes méthodes . . . . .	75
	<b>Conclusion</b>	<b>79</b>
	<b>Annexes</b>	<b>81</b>
	<b>Bibliographie</b>	<b>98</b>
	<b>Table des figures</b>	<b>100</b>
	<b>Liste des tableaux</b>	<b>101</b>

# Introduction

Les entreprises d'assurance sont des acteurs économiques confrontées à un problème de provisionnement courant mais complexe à gérer. En effet, pour couvrir les sinistres futurs de leurs assurés, les assureurs doivent mettre de l'argent de côté, et s'attachent donc à modéliser le coût (futur, donc aléatoire) des sinistres de leurs assurés le plus précisément possible. Plusieurs variables telles que le nombre de sinistres, les règlements effectués, les provisions déjà passées pour chaque sinistre par les gestionnaires sont autant de données que les services actuariels doivent analyser et projeter pour répondre à la problématique de provisionnement.

Dans le cadre de Solvabilité II, il devient nécessaire de mettre en place un provisionnement à l'ordre 2, qu'il soit par défaut calqué sur la formule standard ou non, c'est à dire la construction d'estimateurs de variance des provisions ou des sinistres futurs. En effet, la valeur moyenne des sinistres futurs ne suffira que rarement à les couvrir réellement si les risques ne sont pas symétriques. De plus, la connaissance du processus de sinistralité en tout temps futur peut être utile pour comprendre les risques auxquels sont soumis l'entreprise.

Enfin, on ne redira jamais assez qu'en provisionnement, la connaissance de la distribution complète des sinistres futurs est une forme de *graal*. En effet, cette distribution donne non seulement les deux premiers moments mais elle donne aussi beaucoup d'informations sur la nature du risque encouru.

L'algorithme principal de provisionnement non-vie est l'algorithme *Chain-ladder*. Plusieurs modèles stochastiques s'attachent à fournir des estimateurs d'ordre 1, d'ordre 2, et même des distributions complètes pour cet algorithme, nous en étudierons plusieurs dans le chapitre 2.

En assurance construction, cependant, la nature tri-dimensionnelle des données de sinistres pousse à plus d'inventivité, et la spécificité du problème fait qu'il n'existe que peu de solutions fournissant des résultats satisfaisants. Si le chapitre 1 présentera ces spécificités de l'assurance construction française, nous nous attacherons dans le chapitre 3 à étendre à ce cas par plusieurs moyens certains des modèles phares en provisionnement non-vie décrits au chapitre 2.

Le cadre réglementaire donné par la formule standard de Solvabilité II n'est pas parfaitement adapté à l'assurance construction *à la française*, notamment au niveau des facteurs de risques utilisés sur le risque de réserve. Nous appliquerons alors les méthodes décrites dans les deux parties précédentes à un portefeuille d'assurance construction et mesurerons les écarts qui peuvent survenir entre la formule standard et les différents estimateurs proposés dans ce mémoire, avant de conclure.



# Chapitre 1 Cadre général de l'étude

Ce premier chapitre s'attachera à définir certains concepts nécessaires à la compréhension de l'étude menée tout au long de ce mémoire. Au travers des différentes sources à notre disposition, nous essaierons d'obtenir une vision historique de ce qu'a pu être l'assurance construction en France, puis de saisir les problématiques actuelles liées à cette activité, qu'elles soient de l'ordre de l'analyse financière (données de marché, informations règlementaires, différents produits), d'ordre légal (problèmes de responsabilités, loi Spinetta, convention CRAC, double détente) ou d'ordre actuariel (provisions spécifiques, réglementations, risque de réserve).

Ce chapitre s'appuie majoritairement sur les textes légaux et règlementaires français et européens (notamment [PARLEMENT-EUROPÉEN et CONSEIL-DE-L'EUROPE 2009], [COMMISSION-EUROPÉENNE 2014] et [Code-des-assurance]), ainsi que sur les études précédentes réalisées par [HAEGEL 2009], [CAMBON 2010], [DEPLANTE 2010], [LENFANT 2010], [HAZAN 2011], [BUSSON 2012], [MICHAUD 2012], [LE et VIROT 2013], [COLLET 2013], [DEGIOVANNI 2014], [NATAF 2014], [LAGNAOUI 2015], [CASTRO 2016]. Pour plus d'exhaustivité, notamment sur les sujets historiques, légaux et règlementaires, le lecteur pourra consulter ces sources.

## 1.1 Cadre réglementaire et principes généraux de l'assurance construction française

Nous commencerons par dresser le tableau historique de ce qu'a pu être l'assurance construction, avant de s'intéresser à son cadre réglementaire actuel, aux garanties proposées, ainsi qu'aux provisions spécifiques de cette assurance.

### 1.1.1 Historique de l'assurance construction

**Code de Hammurabi** Gravé sur une Stèle de plus de deux mètres de haut en écriture cunéiforme (voir fig. 1.1.1), le *code de Hammurabi* est un écrit royal datant du 18ème siècle avant J-C, qui se présente sous la forme d'un texte de loi découpé en articles. Bien que n'étant pas le plus vieux texte de loi écrit en Mésopotamie antique retrouvé à ce jour<sup>1</sup>, c'est l'un des plus complets.

Il contient entre autre une idée d'assurance maritime pour les voyages en mer dangereux (préfigurant la version génoise de l'histoire<sup>2</sup>), et nous pouvons notamment en tirer l'extrait suivant qui nous donne une idée de ce que pouvait être la responsabilité du constructeur à l'époque (voir [HAMMURABI 2011] ou encore [DOMINIQUE CHARPIN 2003]) :

---

1. Le Code d'Ur-Nammu datant du 21ème siècle avant J-C est à ce jour le plus vieux texte de loi retrouvé, mais il est plus court et couvre moins de sujets que le code de Hammurabi. Voir [https://fr.wikipedia.org/wiki/Code\\_d%27Ur-Nammu](https://fr.wikipedia.org/wiki/Code_d%27Ur-Nammu) pour plus de détails.

2. L'assurance à la "grosse aventure" existait déjà sous les romains mais relevait plus du domaine pré-bancaire. Ce sont les génois qui instituèrent au milieu du 14ème siècle l'assurance maritime au sens propre



FIGURE 1.1.1 – Photo de la stèle du code de Hammurabi, visible au musée du Louvre. Par Mbzt, CC BY 3.0

“§229. Si un entrepreneur construit une maison pour quelqu’un mais ne l’achève pas convenablement, et si cette maison s’écroule et tue son propriétaire, alors l’entrepreneur sera mis à mort.

§230. Si elle tue le fils du propriétaire le fils de l’entrepreneur sera mis à mort.

§231. Si elle tue l’esclave du propriétaire, alors il devra payer esclave pour esclave au propriétaire de la maison.

§232. Si elle détruit des biens, il devra dédommager pour tout ce qui a été détruit, et dans la mesure où il n’a pas construit convenablement la maison qu’il a bâtie et qu’elle s’est écroulée, il doit la reconstruire à ses propres frais.

§233. Si un entrepreneur construit une maison pour quelqu’un, même s’il ne l’a pas achevée; si par la suite les murs semblent pencher, l’entrepreneur doit les consolider à ses frais.”

La construction de bâtiments était donc déjà à l’époque un sujet légal et une problématique importante, la responsabilité du constructeur pouvant être mise en cause en cas de vice de construction entraînant des dommages matériels ou humains. On comprend bien qu’il est normal que le constructeur soit tenu responsable, cependant on a ici une responsabilité à vie : La responsabilité du constructeur est maintenue jusqu’à la mort du propriétaire de la maison. L’intensité des condamnations est aussi notable.

Quelques siècles plus tard, on observe la naissance de la notion de responsabilité décennale du constructeur dans le droit romain. C’est le code civil napoléonien qui, s’inspirant du droit latin en 1804, crée la notion de garantie décennale française dans son article 1792 et apporte jusqu’à nous ce principe. Voici les différentes versions historiques de l’article qui nous intéresse, fournies par [CRIMINOCORPUS 2014] :

*Code civil - Section III - Article 1792 - Version du 7 mars 1804*<sup>3</sup>

“Si l’édifice construit à prix fait, périt en tout ou en partie par le vice de la construction, même par le vice du sol, les architectes et entrepreneurs en sont responsables pendant dix ans.”

*Code civil - Section III - Article 1792 - Version du 3 janvier 1967*<sup>4</sup>

3. Texte source : Code civil des Français, édition originale et seule officielle, à Paris, de l’imprimerie de la République, An XII, 1804.

4. Texte source : Loi n° 67-3 du 3 janvier 1967 relative aux ventes d’immeubles à construire ou en cours de construction et à l’obligation de garantie à raison des vices de construction.

“Si l'édifice périclète en tout ou en partie par le vice de la construction, même par le vice du sol, les architectes, entrepreneurs et autres personnes liées au maître de l'ouvrage par un contrat de louage en sont responsables pendant dix ans.”

*Code civil - Section III - Article 1792 - Version du 4 janvier 1978*<sup>5</sup>

“Tout constructeur d'un ouvrage est responsable de plein droit, envers le maître ou l'acquéreur de l'ouvrage, des dommages, même résultant d'un vice du sol, qui compromettent la solidité de l'ouvrage ou qui, l'affectant dans l'un de ses éléments constitutifs ou l'un de ses éléments d'équipement, le rendent impropre à sa destination. Une telle responsabilité n'a point lieu si le constructeur prouve que les dommages proviennent d'une cause étrangère.”

Cette version, toujours en vigueur actuellement, date de la **loi Spinetta**. Cette loi pose la responsabilité décennale des constructeurs, qui est l'objet de la majorité des contrats d'assurance dont nous traiterons ici. La forme obligatoire de l'assurance construction n'est par contre arrivée que très tard, la version 1804 du code civil ne pose en effet pas d'obligation d'assurance quelconque.

### 1.1.2 Cadre réglementaire actuel de l'assurance construction

Cette section s'intéressera à l'activité d'assurance construction contemporaine, et plus particulièrement à son cadre juridique et réglementaire. Nous verrons tout d'abord en quoi l'assurance construction est une assurance non-vie, puis nous nous intéresserons aux différents intervenants de l'activité de construction, et nous aborderons enfin les spécificités de cette assurance *via* la loi Spinetta et la convention CRAC.

#### 1.1.2.1 Assurances non-vie, Assurances de biens et Assurances de responsabilités

**L'assurance non-vie** Elle regroupe l'ensemble des activités d'assurance qui n'ont pas pour objet d'assurer la vie de personnes, mais d'assurer des biens, qu'ils soient matériels (voitures, maisons, outils...) ou immatériels (notamment les responsabilités).

En échange d'une prime payée par le souscripteur à l'assureur, l'assureur accepte de dédommager au bénéficiaire les frais engendrés par une éventuelle réalisation du sinistre. Obligatoire ou facultative, une assurance permet à une personne morale ou physique d'être sûre qu'elle n'encaissera pas de pertes trop importantes pour tout état futur de la nature prévu par les clauses contractuelles. À l'intérieur de l'assurance non-vie, on distingue :

- **Les assurances de biens**, ayant pour objet un élément de l'actif du patrimoine de l'assuré. Elle viennent garantir une détérioration possible, un vol, en somme une perte de valeur des possessions de l'assuré. Un exemple simple est une assurance contre le vol, ou l'assureur promet le remboursement ou remplacement des biens subtilisés.
- **Les assurances de responsabilité** viennent quant à elles garantir la réparation juridique d'éventuels dommages causés par l'assuré à un tiers. Elles viennent donc garantir l'assuré contre ses éventuelles erreurs. Un exemple courant est celui de la Responsabilité civile automobile : si l'assuré se voit juridiquement contraint de dédommager un autre conducteur suite à un accident, l'assureur s'engage à rembourser à sa place.

Les régimes juridiques appliqués à ces deux types d'assurance ne sont pas les mêmes<sup>6</sup>. Dans le cadre de ce mémoire, nous parlerons d'assurance construction sous sa forme responsabilité civile décennale, appartenant aux assurances de responsabilités.

5. Texte source : Loi n° 78-12 du 4 janvier 1978 relative à la responsabilité et à l'assurance dans le domaine de la construction.

6. Pour plus d'informations sur la distinction, se reporter au code des assurances [**Code-des-assurance**]

### 1.1.2.2 Intervenants en construction

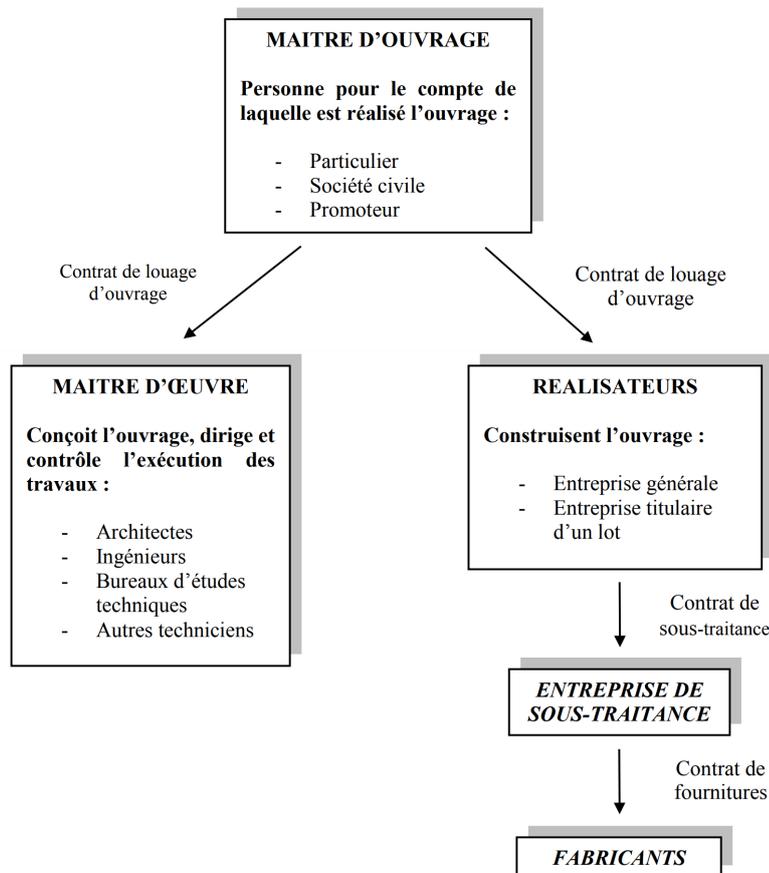


FIGURE 1.1.2 – Principaux intervenants de l'assurance construction [CAMBON 2010]

Durant l'activité de construction, plusieurs personnes physiques ou morales sont impliquées. Nous nous intéresserons ici aux principaux intervenants de cette activité et nous exposerons par un schéma les liens juridiques entre eux. Pour plus de détails, voir par exemple [HAZAN 2011] mais aussi [NATAF 2014].

**Le maître d'ouvrage** Personne physique ou personne morale de droit public ou privé, le maître d'ouvrage est le propriétaire du chantier, son commanditaire : il en définit l'objectif, le calendrier et le budget (c'est lui qui finance l'ouvrage). Sa fonction dans l'activité d'assurance est définie par [Code-des-assurance] Il représente dans le processus de construction les futurs propriétaires du bâtiment, et définit donc les besoins que le bien doit remplir en amont de la construction. Il est lié au maître d'œuvre, i.e la personne qui conçoit et exécute l'ouvrage par un contrat dit de *louage d'ouvrage*.

Sa responsabilité peut être engagée après livraison du bien si un dommage lié à la construction survient. C'est cette responsabilité qui était déjà présente en Mésopotamie et dans la Rome antique. Nous verrons plus tard que *l'assurance Dommage-ouvrage obligatoire* est prévue pour couvrir cette responsabilité via un préfinancement initial sans recherche de responsabilité.

**Le maître d'œuvre** Le maître d'œuvre est la personne qui conçoit l'ouvrage, qui dirige et supervise l'exécution des travaux sur ordre du maître d'ouvrage. Ce peut-être une personne

physique ou morale de droit public ou privé. Il est chargé des questions techniques et économiques relatives au chantier, mais peut également assister le maître d'ouvrage dans ses tâches : choix des intervenants, choix des matériaux, fixation des délais de livraison, etc. Souvent, le maître d'œuvre est un architecte, ingénieur ou technicien, ou un cabinet d'étude.

La responsabilité du maître d'œuvre peut être engagée par le maître d'ouvrage en répercussion de la sienne si le dommage au bien construit est de son fait. Le maître d'œuvre pourra ensuite se retourner contre les constructeurs, et eux-mêmes contre les sous-traitants. Leurs assureurs respectifs pourront ensuite utiliser le mécanisme de subrogation.

**Les constructeurs et sous-traitants** Les constructeurs sont souvent des entreprises, mais peuvent aussi être des personnes physiques. Ils sont liés au maître d'œuvre par des contrats de travaux ou de louage d'ouvrage. Trois cas se distinguent :

- **Le constructeur général** : Contractuellement lié au maître d'œuvre pour l'entièreté des travaux et constructions.
- **Le constructeur titulaire d'un lot** : Engagé par un contrat avec le maître d'œuvre pour la réalisation d'une tâche précise et déterminée sur le chantier.
- **Le sous-traitant** : Engagé par un contrat avec une entreprise générale ou titulaire d'un lot pour la réalisation d'une partie ou de toutes les prestations normalement exécutées par elle.

Leurs responsabilités peuvent être mises en œuvre si il y a des défauts dans la construction. L'assurance prévue pour les couvrir s'appelle la *Responsabilité civile décennale obligatoire*.

**L'assureur** Les différents acteurs prennent tous des risques dans l'activité de construction qu'il convient d'assurer. Notamment, si les constructeurs mettent en jeu leurs assurances *Responsabilité civile décennale*, le maître d'ouvrage est aussi concerné de par la loi Spinetta.

L'assureur, qui peut être multiple, est donc un acteur important dans l'activité de construction.

### 1.1.2.3 L'obligation d'assurance

**1928 : Police globale de chantier** En 1928 l'apparition de la police globale de chantier permet un meilleur encadrement de l'assurance construction alors que les lois précédentes (notamment l'article 1792) n'encadraient que la responsabilité civile du constructeur et ne disait rien sur les assurances correspondantes. L'assurance est toujours facultative.

**1940 : Obligation d'assurance décennale** Pour éviter les problèmes liés à l'absence de couverture des constructeurs et donc de remboursement des dettes générées par la responsabilité décennale en cas de faillite, une obligation d'assurance a été mise en place par la loi du 31 décembre 1940. Cette garantie dure 10 ans et commence à la livraison du bien, i.e la fin du chantier et non au début de la construction.

Cette obligation ne couvrait pas tous les chantiers et en 1978 lors de la mise en place de la loi Spinetta, seuls 60% des constructeurs et 50% des maîtres d'œuvre étaient couverts par ces assurances. Ce fut l'une des raisons de mise en place de la loi Spinetta, qui régla ce problème entre autres choses comme nous l'apprend [DEGIOVANNI 2014].

**1978 : Loi Spinetta** Parue au journal officiel le 4 janvier 1978, la loi Spinetta corrige comme nous l'avons vu précédemment l'article 1792 du code civil pour lui donner sa forme actuelle. Elle pose **une présomption de responsabilité des constructeurs** de plein droit, pour une durée de 10 ans (donnant son nom *décennale* à la responsabilité) à compter de la réception des travaux.

La loi Spinetta instaure un système dit à **double détente**, de par l'obligation d'assurance en dommage-ouvrage du maître d'ouvrage et en responsabilité civile décennale des constructeurs. L'assurance dommage ouvrage du maître d'ouvrage est en effet censée couvrir la totalité des sinistres le temps que les responsabilités soient départagées entre les constructeurs. Puis, l'assureur en dommage ouvrage se retourne contre les responsables des dommages qui remboursent via leurs assurances responsabilité civile décennale.

Ce mécanisme à double détente a été essentiellement mis en place pour éviter le sur-accident, *i.e* le fait que l'état du bien se dégrade lors du temps perdu par la recherche des responsabilités. En permettant de régler les sinistres plus rapidement, il réduit également leurs coûts.

**1983 : Convention CRAC** La convention de règlement de l'assurance construction est une convention entre les différents assureurs dommages-ouvrages et/ou responsabilité civile décennale datant de 1983 et ayant pour but d'accélérer les règlements des sinistres en obligeant les assureurs en dommage ouvrage signataires à régler dans un délais de trois mois. Les recours exercés ensuite permettent à l'assureur en dommage ouvrage de récupérer environ 45% de ses règlements<sup>7</sup>. Les assureurs en responsabilité ont un délai de deux mois pour se mettre d'accord sur le partage de la responsabilité, à défaut un partage provisoire de la charge des recours sera mis en place. La convention CRAC met aussi en place une expertise unique, *dans le but de réduire les coûts externes associés aux sinistres*, notamment les frais.

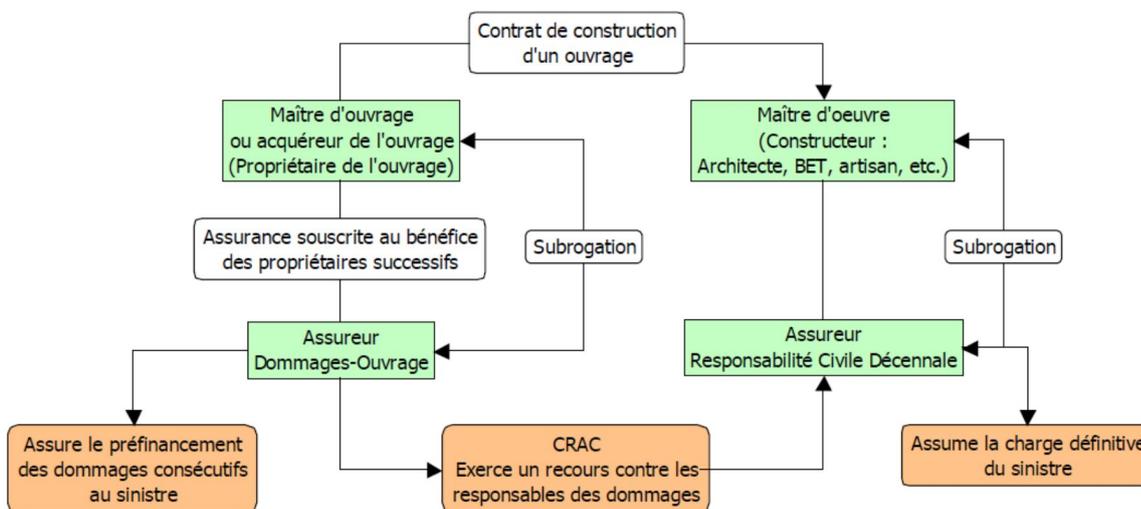


FIGURE 1.1.3 – Fonctionnement en double détente via la convention CRAC [COLLET 2013]

### 1.1.3 Les garanties de l'assurance construction

Lors de notre étude, nous utiliserons les données de sinistres en assurance construction qui correspondent aux trois garanties décrites ci-après, et seulement à celles-ci. Ces trois garanties

7. Logiquement, les recours devraient être totaux, seulement la recherche de responsables est parfois trop coûteuse voire impossible. De plus, la CRAC met en place un ticket modérateur.

ont été choisies car elles font partie de la *Line of Business* 8<sup>8</sup> dans le référentiel Solvabilité II dont nous parlerons à la section suivante, et qui sera notre cadre de modélisation pour la suite de ce mémoire.

**Garantie *Responsabilité civile générale*** Cette assurance couvre la responsabilité du chef d'entreprise en cas de dommages extérieurs à l'ouvrage. Si l'on distingue généralement les assurances responsabilité civile avant et après livraison, les risques couverts sont entre autres les suivants :

- La responsabilité du fait des matériaux et engins de chantier en action de travail,
- La responsabilité à l'égard des préposés, les dommages environnementaux,
- La responsabilité civile incendie,
- La responsabilité du fait des produits défectueux,
- La responsabilité civile loueur de matériel et d'engins de chantier,
- La responsabilité civile transporteur d'engins de chantier,
- La responsabilité vol des préposés,
- La responsabilité des dommages aux existants ne relevant pas de l'assurance obligatoire,
- La responsabilité du fait des matériaux et engins de chantiers pris en location.

**Garantie *Responsabilité civile décennale obligatoire*** Cette assurance est obligatoirement souscrite par tout constructeur d'un ouvrage pour couvrir sa responsabilité civile décennale. La responsabilité garantie est celle recherchée, après réception, sur le fondement des articles 1792, 1792-2 et 1792-4 du Code civil pour la durée de 10 ans fixée par l'article 1792-4-1 du Code civil. Elle est alors garantie :

- Pour les ouvrages à la réalisation desquels l'assuré a participé en qualité de loueur d'ouvrage,
- Pour les ouvrages à la réalisation desquels l'assuré a participé en qualité de sous-traitant.

**Garantie *Responsabilité civile décennale complémentaire*** Cette assurance comporte des garanties qui sont généralement adossées à la *Responsabilité civile décennale obligatoire*. On y trouve majoritairement :

- La garantie de bon fonctionnement
- La garantie des dommages immatériels consécutifs
- La garantie des dommages aux existants
- La garantie de responsabilité du sous-traitant en cas de dommages de nature décennale

#### 1.1.4 Provisions spécifiques à l'assurance construction

En règle générale en assurance non-vie, il est demandé aux assureurs de provisionner les sinistres survenus, qu'ils soient connus ou non.

Ainsi, pour un contrat donné, l'assureur touche la prime en début d'année<sup>9</sup>, et provisionne en fin d'année la charge qu'il estime devoir encore payer pour les sinistres qui ont été déclarés

---

8. Mais la *Line of Business* 8 contient également d'autres garanties, dont nous ne parlerons pas car elles ne sont pas présentes dans notre portefeuille.

9. Les contrats sont en effet souvent annuels

dans l'année, mais aussi pour certains sinistres dits *tardifs*, c'est-à-dire survenus dans l'année mais dont l'assureur n'a pas encore connaissance<sup>10</sup>.

Cependant en assurance construction, le caractère décennal de la couverture vendue oblige l'assureur à procéder différemment et à provisionner des sinistres futurs : en effet, la garantie vendue a l'assuré couvre des sinistres qui surviendront longtemps après la date d'arrêt comptable correspondant au versement de la prime.

Ainsi, nous avons dans cette branche d'assurance française une gestion spécifique, dite **gestion en capitalisation**, qui consiste à :

- Provisionner bien entendu les sinistres déjà survenus à leur valeur future attendue,
- Mais également provisionner les sinistres qui se manifesteront tout au long de la garantie, peut-être dans une dizaine d'année<sup>11</sup>, et dont la charge future doit, elle aussi, être prise en compte (et donc estimée).

Ce mode de gestion spécifique comporte donc deux provisions bien distinctes : si la première, la provision pour sinistres à payer, est classique et commune à toutes les branches d'assurance non-vie, la seconde, la provision pour sinistres non-encore manifestés, est spécifique à l'assurance construction française.

Sur les garanties Responsabilités civiles présentées au point précédent, seule la *Responsabilité civile décennale obligatoire* fait l'objet de cette gestion en capitalisation. Cette spécificité peut paraître anodine, mais elle a des conséquences sur la modélisation des charges futures non négligeables. Nous aborderons ce problème de bien des manières lors de ce mémoire.

## 1.2 L'assurance construction sous Solvabilité II

Dans toute activité d'assurance, l'assureur a l'obligation légale, réglementaire et prudentielle de passer des provisions correspondantes au risque souscrit. En assurance non-vie classique française, ces provisions sont globalement dotées au sein de la **provision pour sinistres à payer**. Cette provision, définie à l'article R331-6 du [**Code-des-assurance**], correspond aux sinistres déjà survenus dans les périodes de garanties terminées à la date d'inventaire.

En *Responsabilité civile décennale obligatoire*, cependant, les assureurs ont obligation de gérer le risque en capitalisation. Ce mode de gestion les oblige à doter une provision spécifique, la **provision pour sinistres non-encore manifestés**. Cette provision, définie à l'article R331-17 du [**Code-des-assurance**], représente le coût ultime des sinistres qui surviendront dans le futur sur la période de garantie décennale pour les contrats.

En normes françaises, une méthode est proposée par l'Autorité des normes comptables et par l'article A331-21 du [**Code-des-assurance**] pour calculer ces provisions spécifiques (nous la discuterons à la section 3.1.1 de ce mémoire). En normes Solvabilité II, par contre, une évaluation en *Best estimate* est demandée, et les assureurs sont libres des méthodes de calcul.

En norme Solvabilité II, les provisions pour sinistres, dites *provisions techniques*, doivent faire l'objet d'une évaluation selon des principes économiques bien spécifiques. Nous prendrons le temps lors de cette section de décrire brièvement ces principes, tout en référant les lecteurs intéressés aux spécifications [COMMISSION-EUROPEENNE 2014; PARLEMENT-EUROPEEN et CONSEIL-DE-L'EUROPE 2009].

---

10. le cas caricatural est le sinistres survenu entre Noël et le jour de l'an pour un arrêté de comptes au 31/12/N, et déclaré à l'assureur qu'en janvier.

11. Ou plus, selon la durée des chantiers et la rapidité de mise en cause de l'assureur dans les dossiers

### 1.2.1 Principe de *Best estimate*

Le *Best estimate*, c'est la valeur économique actuelle probable des risques futurs auxquels est soumise l'entreprise. Plus formellement, l'article 77 de la directive [PARLEMENT-EUROPÉEN et CONSEIL-DE-L'EUROPE 2009] relatif au calcul des provisions techniques en normes Solvabilité II énonce en ses points 1, 2 et 3 :

“§1. La valeur des provisions techniques est égale à la somme de la meilleure estimation et de la marge de risque respectivement décrites aux paragraphes 2 et 3.

§2. La meilleure estimation correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risques pertinents. Le calcul de la meilleure estimation est fondé sur des informations actualisées et crédibles et des hypothèses réalistes et il fait appel à des méthodes actuarielles et statistiques adéquates, applicables et pertinentes. La projection en matière de flux de trésorerie utilisée dans le calcul de la meilleure estimation tient compte de toutes les entrées et sorties de trésorerie nécessaires pour faire face aux engagements d'assurance et de réassurance pendant toute la durée de ceux-ci. La meilleure estimation est calculée brute, sans déduction des créances découlant des contrats de réassurance et des véhicules de titrisation. Ces montants sont calculés séparément, conformément à l'article 81.

§3. La marge de risque est calculée de manière à garantir que la valeur des provisions techniques est équivalente au montant que les entreprises d'assurance et de réassurance demanderaient pour reprendre et honorer les engagements d'assurance et de réassurance.”

Les provisions techniques se segmentent donc en *Best estimate* et en *Risk Margin*. Notons que le *Best estimate* dont il est question ici correspond à la somme des flux futurs actualisés, que ces flux viennent de sinistres déjà manifestés ou non. Ainsi, le périmètre couvert par le *Best estimate* correspond à celui de la provision pour sinistres à payer et de la provision pour sinistres non-encore manifestés, à *hypothèses prudentielles égales*.

### 1.2.2 Principe de *Risk Margin*

Les provisions techniques se segmentent donc en *Best estimate* et marge pour risque.

Si le *Best estimate* se conçoit comme la meilleure estimation de la valeur actuelle probable des engagements futurs, la *Risk Margin* est conçu comme la valeur d'échange (portage du coût du risque à l'ultime) d'un portefeuille. En cela, elle constitue un surplus de prudence au bilan imposé par le régulateur européen. En effet, dans la plupart des cas, la valeur moyenne ne suffira pas à couvrir le risque<sup>12</sup>.

La *Risk Margin*, notée *RM*, se calcule souvent par une approche de coût du capital (c'est en effet l'approche recommandée en formule standard par la directive [PARLEMENT-EUROPÉEN et CONSEIL-DE-L'EUROPE 2009], et le règlement délégué [COMMISSION-EUROPÉENNE 2014]).

Si les provisions techniques semblent suffire à couvrir le risque pris par l'assureur, la réglementation impose aussi des limites de capitaux propres que l'assureur doit détenir pour encaisser d'éventuelles pertes, et ainsi garantir sa solvabilité.

Nous référons aux spécifications de Solvabilité II pour plus de détails, ou aux nombreux mémoires d'actuaire sur le sujet<sup>13</sup>.

12. Si l'on considère un risque symétrique, le coût du risque dépassera le *Best estimate* environ une fois sur deux. Il n'est pas acceptable du point de vue du régulateur d'une entreprise d'assurance ait une chance sur deux de ne pas pourvoir à ses engagements.

13. Par exemple et dans le désordre [BEAUNE 2016; BOUGNON, BARRUEL et ZEC 2016; BUSSON 2012; DEGIOVANNI 2014; DEPLANTE 2010; CAMBON 2010; CASTRO 2016; HAEGEL 2009]

### 1.2.3 Minimums règlementaires de capitaux propres

Les capitaux propres se conçoivent comme la différence entre la valeur de l'actif de l'assureur et la valeur de ses engagements, matérialisée par les provisions techniques, plus d'éventuels capitaux hors-bilan reconnaissables sous Solvabilité II.

Si l'on note  $A$  la valeur de l'actif, et que l'on segmente le passif de l'assureur en fonds propres  $FP$  et provisions techniques  $PT = BE + RM$ , on a par équilibre du bilan que :

$$FP = A - BE - RM.$$

En exercice Solvabilité II, il convient de disposer d'un montant de fonds propres suffisant pour ne pas faire faillite dans 199 cas sur 200 en moyenne à horizon un an, le *Solvency Capital Requirement*, noté  $SCR$ . Ainsi, on peut définir en première approche le *Solvency Capital Requirement* comme un quantile à 99.5% de la distribution des fonds propres de l'année prochaine, soit en notant  $FP_1$  les fonds propres de l'année prochaine,

$$SCR = VaR_{99.5\%}(FP_1)$$

Enfin, la réglementation pose aussi un *Minimum Capital Requirement*, noté  $MCR$ , comme étant le niveau de fonds propres en dessous duquel la situation devient critique. Si l'assureur passe sous certains seuils de capitaux, fonctions du *Solvency Capital Requirement* ou du *Minimum Capital Requirement*, le régulateur se doit d'intervenir.

## 1.3 Solvabilité II et le risque de réserve

L'architecture interne de Solvabilité II impose aux assureurs de calculer leur *Solvency Capital Requirement* annuellement, afin de pouvoir tracer et garantir leur solvabilité à horizon un an. Ainsi, sous Solvabilité II, la quantité de fonds propres disponibles pour l'assureur se doit d'être supérieure au *Solvency Capital Requirement*, et l'on utilise pour mesurer le degré de suffisance des fonds propres un *ratio de solvabilité*, servant à exprimer les fonds propres de l'assureur en fonction du montant de fonds propres pensé comme requis pour couvrir le risque qu'ils portent.

Se pose alors rapidement la question des techniques de calcul de ces quantités. Solvabilité II propose aux assureurs l'utilisation d'une *formule standard* pour ce calcul, que nous allons discuter brièvement ici. La spécification de cette formule se fait tout au long des [PARLEMENT-EUROPEEN et CONSEIL-DE-L'EUROPE 2009], et nous référons donc le lecteur curieux à cette source, mais également aux nombreux mémoires disponibles sur le sujet.

### 1.3.1 Principe de risque à un an

Le *Solvency Capital Requirement* se conçoit comme un quantile. La formule standard propose alors un principe d'explosion en différents modules et sous-modules du risque porté par un assureur, ainsi qu'une agrégation des quantiles de ces modules en *Solvency Capital Requirement* global représentant le quantile de leur somme. Le graphe d'organisation de cette agrégation est disponible en figure 1.3.1.

Supposons que l'on dispose à un certain niveau de l'arbre des *Solvency Capital Requirement* de  $n$  sous catégories, notés  $\mathbf{q} = \{q_1, \dots, q_n\}$ <sup>14</sup>. Supposons également que l'on dispose de la matrice

14. La notation  $q$  est utilisée car ces *Solvency Capital Requirement* sont fondamentalement des quantiles

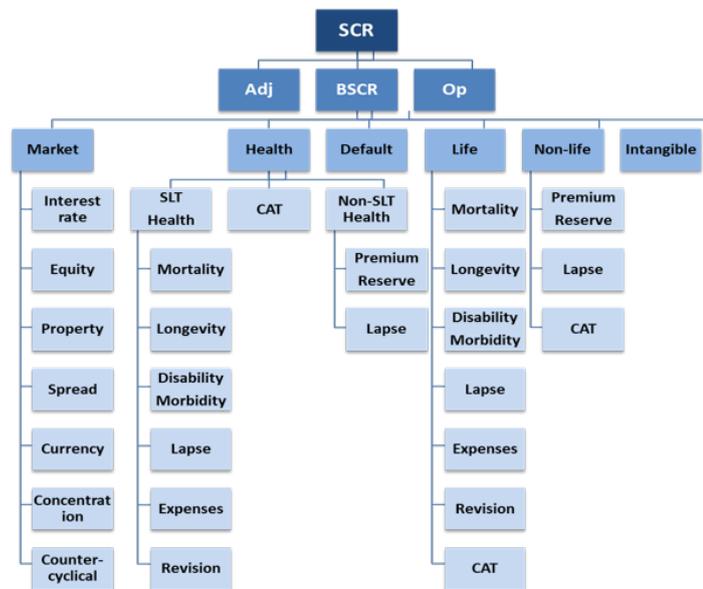


FIGURE 1.3.1 – Schéma d’agrégation des risques unitaires en sous-modules et modules de risque pour le calcul du *Solvency Capital Requirement*

de corrélation entre les variables aléatoires représentant les risques de ces sous-catégories  $P = (\rho_{i,j})_{i,j \in \{1, \dots, n\}}$ . La formule standard propose alors de calculer le *Solvency Capital Requirement* de la somme des risques marginaux,  $q$  comme étant :

$$q = \sqrt{\mathbf{q}P\mathbf{q}},$$

Soit une agrégations de quantiles équivalente à une hypothèse elliptique sur le vecteur de risques sous-jacents.

### 1.3.2 Risque de réserve en non-vie

Un des sous-modules de risques de la formule standard nous intéresse tout particulièrement ici de part la nature de notre étude : le risque de réserve en non-vie, inclus dans le module non-vie.

La formule standard propose de calculer le *Solvency Capital Requirement* associé à ce risque, le quantile de la distribution de ses pertes, de la manière décrite par les actes délégués, reprise par exemple dans [BOUGNON, BARRUEL et ZEC 2016, p. 21]. Notons  $V$  le volume du portefeuille,  $\sigma$  l’estimation de la volatilité du portefeuille non-vie, et  $\rho$  une fonction d’écart-type dépendant de  $\sigma$  de la manière suivante :

$$\rho(\sigma) = \frac{\exp\left(\text{VaR}_{99.5\%}(\mathcal{N}(0, 1))\sqrt{\ln(\sigma^2 + 1)}\right)}{\sqrt{\sigma^2 + 1}} - 1 \approx 3\sigma$$

Le paramètre  $\sigma$  est calculé selon la formule d’agrégation précédente avec une matrice de corrélation entre les risques sous-jacents préfixée par les actes délégués et des coefficients de variation de chacun des sous-modules de risques  $\{\sigma_1, \dots, \sigma_n\}$  également préfixé par le texte.

Si les coefficients de variation des différents sous-modules de risques sont fixés par le texte, il y a néanmoins la possibilité d’en modifier certains et de les calculer sur la base des données

propres à l'entreprise. On parle alors de paramètres propres à l'entreprise, ou *Undertaking specific parameter*.

### 1.3.3 Paramètres propres à l'entreprise

Certains des paramètres du modèle de la formule standard peuvent potentiellement être recalculés en interne par l'entreprise chaque année. La description proposée ici de ces principes s'appuie entre autre sur [POUDRET et BEAUNE 2016].

C'est le cas notamment des coefficients de variations  $\sigma_i$  de certains risques : les risques de révision en santé, de primes et de réserves en santé non-SLT, de révision en vie et de primes et de réserves en non-vie.

C'est ici le risque de réserve en non-vie qui nous intéressera tout au long de ce mémoire. Le coefficient de variation du risque de réserve est calculé, toujours par la même technique d'agrégation, en segmentant le portefeuille non-vie de l'assureur en *Line of Business*. La formule standard définit donc des coefficients de variation par *Line of Business* pour le risque de prime comme pour le risque de réserve, dont les valeurs sont données par le texte, repris à la table 1.3.1

TABLE 1.3.1 – Écart-types standard pour les risques de primes et de réserve en non-vie

Line of business	Écart-type pour le risque de primes brut du segment	Écart-type pour le risque de de réserve du segment
Responsabilité civile automobile	10%	9%
Autre assurance des véhicules à moteur	8%	8%
Maritime, aérienne et transport	15%	11%
Incendie et autres dommages aux biens	8%	10%
<b>Responsabilité civile générale</b>	<b>14%</b>	<b>11%</b>
Crédit et cautionnement	12%	19%
Protection juridique	7%	12%
Assistance	9%	20%
Pertes pécuniaires diverses	13%	20%
Réassurance accidents non proportionnelle	17%	20%
Réassurance maritime, aérienne et transport non proportionnelle	17%	20%
Réassurance dommages non proportionnelle	17%	20%

La *Line of Business* à laquelle sont rattachés les risques que nous traiterons ici est la *Line of Business* Responsabilité civile générale, en gras ci-dessus.

Le coefficient de variation du risque de réserve non-vie pour la *Line of Business* Responsabilité civile générale ainsi définie par la formule standard sera notre point central de réflexion tout au long de ce mémoire. Comme nous ne traiterons que de ce risque spécifique, nous utiliserons pour cette quantité la dénomination simple de  $\sigma$ .

C'est de ce paramètre que la formule standard propose aux assureurs un recalcul sur la base de leurs données, selon une procédure appelée *Undertaking specific parameter*. C'est ce paramètre que nous essaierons de calibrer selon plusieurs méthodes sur les données à notre disposition.

## 1.4 Données, cahier des charges et problématique

Pour cette étude, nous disposons de données de sinistres au format agrégé : règlements, provisions dossier/dossier, recours... Chacune de ces données est disponible au niveau de segmentation suivant :

- Année de survenance du sinistre concerné
- Année comptable de la donnée
- Le cas échéant la date réglementaire d'ouverture de chantier à laquelle le sinistre est rattaché.

De plus, nos données sont segmentées par Groupes de risques homogènes, nous disposons de trois Groupes de risques homogènes dans la *Line of Business* sur laquelle nous travaillons ici :

1. la *Responsabilité civile générale* (RCG), gérée en répartition
2. la *Responsabilité civile décennale complémentaire* (RCDC), gérée en répartition
3. la *Responsabilité civile décennale obligatoire* (RCDO), gérée en capitalisation.

La branche gérée en capitalisation, la *Responsabilité civile décennale obligatoire*, est donc la seule à disposer de l'information de date d'ouverture de chantier. La géométrie de ces données est analysée et décrite plus en profondeur en section 3.1.

Nous disposons d'un historique de 27 ans (1991 – 2017) pour ces données.

Nous considérerons que les données à l'intérieur d'un même Groupe de risques homogènes constituent un seul sous-jacent, c'est-à-dire qu'il n'y a pas de disparités entre les risques sous-tendant ces triangles : l'homogénéité des Groupes de risques homogènes ne sera pas ici remise en cause.

Le processus stochastique d'intérêt pour nous ici correspond aux réserves de la *Line of Business* composée de ces trois Groupes de risques homogènes au cours du temps. Ces réserves comportent la provision pour sinistres à payer, mais aussi la provision pour sinistres non-encore manifestés (sur la *Responsabilité civile décennale obligatoire*), créée pour faire face à l'obligation de couverture décennale et à la gestion en capitalisation.

La problématique principale de cette étude sera de construire des estimateurs complets de la charge future de la *Line of Business*, c'est-à-dire des distributions empiriques de cette variable. À défaut, les moments d'ordre un et deux de cette même variable seront recherchés. Dans tous les cas, nous chercherons la pertinence et les propriétés des estimateurs que nous proposerons et des estimateurs que nous utiliserons.

En effet, en régime Solvabilité II, la connaissance de la variance des provisions à un an, c'est-à-dire du processus de provision pris en un certain temps futur (un an), est primordiale pour calculer le risque de réserve, notre objectif.

Le calibrage par l'EIOPA dans les exercices préparatoires de Solvabilité II du sigma de réserve de cette *Line of Business* à été effectué au niveau européen, et fut calibré sur l'ensemble des risques de la *Line of Business* et pas uniquement sur le risque construction français. Le mode de fonctionnement de la décennale française étant bien particulier (unique en Europe), nous observerons alors la non-cohérence de cette spécificité avec le modèle global.



# Chapitre 2 Provisionnement non-vie : de l'algorithme aux modèles complets

Dans ce chapitre, nous nous attacherons à présenter brièvement les principales méthodes utilisées en provisionnement non-vie. Nous nous concentrerons sur quelques modèles, et n'irons pas dans les détails de chacun d'entre eux.

Les modèles de provisionnement non-vie sont nombreux, et les actuaires et praticiens en ont développés beaucoup, avec de nombreuses variantes au cours du temps. Pour un aperçu plus large de ce qui existe, se référer à [ENGLAND et VERRALL 2002; CHARPENTIER et DENUIT 2004] ou plus largement à [ADAMS 2017].

Avant de parler des quelques modèles qui nous intéressent dans le cadre de notre étude, notamment *Chain-ladder*, Mack, Merz-Wüthrich et leurs extensions, nous commencerons par donner quelques définitions et notations que nous partagerons tout au long de cette étude.

Durant tout ce chapitre, nous considérerons  $X$  **un triangle de paiements incrémentaux**, dont les valeurs  $X_{i,j}$  sont indexées par une année de rattachement  $i \in \{1, \dots, n\}$  et une année de développement  $j \in \{1, \dots, m\}$ , dont une représentation serait :

$$\begin{array}{c|ccc}
 & j = 1 & \dots & j = m \\
 \hline
 i = 1 & \dots & \dots & \dots \\
 \dots & \dots & X_{i,j} & \dots \\
 i = n & \dots & & 
 \end{array}$$

Pour plus de simplicité, nous considérerons ici uniquement les triangles avec autant d'années de rattachement que d'années de développement, *i.e*  $m = n$ . Cependant la plupart des résultats proposés s'étendent à  $m \geq n$ .

$X_{i,j}$  représente les paiements ou charges afférentes aux sinistres survenus l'année  $i$  et payés l'année  $i + j - 1$ . La nature de l'information importe peu pour le développement théorique des méthodes de projections qui suivent, mais importera par la suite pour les applications pratiques : souvent, les praticiens considèrent en parallèle des triangles de nombres, de charges, de règlements, de recours, etc. Pour plus de simplicité mais sans perte de généralité, nous utiliserons par défaut l'appellation de *charge incrémentale* pour  $X_{i,j}$

L'année d'observation est par principe la dernière année de rattachement des charges,  $n$ . On a alors de manière évidente que l'ensemble des informations disponibles au moment du calcul est :

$$\mathcal{D}_n = \{X_{i,j}, \forall i, j \mid i + j \leq n + 1\},$$

et que l'ensemble des charges futures est par conséquent :

$$\bar{\mathcal{D}}_n = \{X_{i,j}, \forall i, j \mid i + j > n + 1\}.$$

On définit ensuite le **triangle cumulé**  $C$  de la manière suivante :  $\forall i, j \in \{1, \dots, n\}$ ,

$$C_{i,j} = \sum_{p=1}^j X_{i,p}$$

On notera bien évidemment que, de la même manière que pour le triangle incrémental, une partie de ces informations n'est pas supposée connue. Un des points cruciaux pour l'activité d'assurance est d'arriver à estimer, qualifier, comprendre, les éléments du triangles inférieur.

Pour tout  $i \in \{1, \dots, n\}$ , on appelle **Ultimes** les valeurs de  $C_{i,n}$  et l'on appelle **Réserves, Provisions** ou *incurred but not reported* (IBNR) selon la nature du triangle<sup>1</sup> les valeurs  $R_i = C_{i,n} - C_{i,i}$

## 2.1 Algorithme et modèles de base

### 2.1.1 Algorithme *Chain-ladder*

L'algorithme de *Chain-ladder* est un des algorithmes de provisionnement les plus connus et appliqués, place acquise sûrement grâce à sa facilité d'utilisation, sa robustesse, sa simplicité de principe. La présentation qui en est faite ici est courte mais suffisante pour en comprendre le fonctionnement, et saisir les enjeux traités *supra*.

On définit les **coefficients de passage individuels** comme étant les  $F_{i,j}$  tels que  $\forall i, j$ ,

$$F_{i,j} = \frac{C_{i,j+1}}{C_{i,j}} \quad (2.1)$$

On définit les **coefficients de passage** comme étant les  $\hat{f}_j$  tels que  $\forall j \in \{1, \dots, n-1\}$ ,

$$\hat{f}_j = \frac{\sum_{i=1}^{n+1-j} C_{i,j+1}}{\sum_{i=1}^{n+1-j} C_{i,j}} \quad (2.2)$$

Ces coefficients de passage servent comme leur nom l'indique, à *passer* d'une colonne à l'autre pour développer et compléter le triangle de la manière suivante :  $\forall i, j \mid i + j > n + 1$ ,

$$\hat{C}_{i,j} = C_{i,j-1} \hat{f}_j$$

On complète alors diagonale après diagonale le triangle inférieur, jusqu'à obtenir les estimations des Ultimes, ainsi que les réserves :  $\forall i \in \{1, \dots, n\}$ ,

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

$$\hat{R}_i = \hat{C}_{i,n} - C_{i,i}$$

---

1. Pour un triangle de charges, on parlera d'*incurred but not reported*, alors que pour un triangle de règlements ce seront des provisions ou des réserves.

Cet algorithme étant à la fois simple et élégant, de nombreux modèles de provisionnement stochastiques s'attachent à garder cette estimation à l'ordre 1, et, sous les bonnes hypothèses, l'estimation fournie est *de bonne qualité*.

Les sections suivantes visent à présenter quelques modèles plus formels sur les incréments  $X_{i,j}$  (ou de manière équivalente, sur les cumulés  $C_{i,j}$ ) englobant *Chain-ladder*, c'est-à-dire rendant ces estimateurs les meilleurs selon certains critères.

### 2.1.2 Modèle de Mack

Le modèle de Thomas Mack (*c.f.* [MACK 1993; MACK 1994; MACK et RE 1994; MACK 1999], ou encore [BUCHWALDER et al. 2006]) est le jeu d'hypothèses stochastique englobant *Chain-ladder* le plus connu et le plus utilisé par les praticiens. Les hypothèses du modèle sont :

**M1** Il existe des coefficients de passage  $f_j, \forall j \in \{1, \dots, n-1\}$  tels que :

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

**M2**  $\forall i \neq j, \{C_{i,1}, \dots, C_{i,n}\} \perp \{C_{j,1}, \dots, C_{j,n}\}$ ,

**M3**  $\forall i, j \in \{1, \dots, n\}, \exists \sigma_j^2 \mid \text{Var}(C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}) = C_{i,j} \sigma_j^2$

La seconde hypothèse suppose simplement l'indépendance entre les différentes années de rattachement. Sous **M1-2**, [MACK 1993] montre que les estimateurs de *Chain-ladder* des coefficients de passage (donnés par (2.2)) sont **non biaisés et non-corrélés** (mais que leurs carrés sont négativement corrélés), et sont les meilleurs estimateurs possibles dans cette classe. On note de plus que les ultimes et les réserves correspondants sont eux-mêmes **non-biaisés**.

Le modèle de Mack s'attache ensuite, via une troisième hypothèse de variance, à dériver les moments d'ordre 2 des ultimes. Sous **M1-3**, Mack donne en effet les estimateurs suivants des paramètres de variances :  $\forall j \in 1, \dots, n-1$ ,

$$\hat{\sigma}_j^2 = \frac{1}{n-j-1} \sum_{i=1}^{n+1-j} C_{i,j} (F_{i,j} - \hat{f}_j)^2 \quad (2.3)$$

Remarquons que ces estimateurs sont encore une fois **non-biaisés**. L'estimation du dernier paramètre de variance  $\sigma_n^2$  pose problème : en effet il ne serait avec cette formule qu'appuyé sur une observation. Différentes options sont alors proposées par Mack :

- $\hat{\sigma}_n^2 = 0$ , *e.g.* si le triangle est déjà complètement développé la dernière année, ou bien
- $\hat{\sigma}_n^2 = \min\{\frac{\hat{\sigma}_{n-1}^4}{\hat{\sigma}_{n-2}^2}, \hat{\sigma}_{n-2}^2, \hat{\sigma}_{n-3}^2\}$ , correspondant à une extrapolation par régression log-linéaire.

On définit l'**erreur quadratique moyenne de prédiction**, ou *mean square error of prediction* d'un estimateur  $\hat{X}$  d'une variable aléatoire  $X$  comme :

$$mse_p(\hat{X}) = \mathbb{E}((X - \hat{X})^2)$$

Mack montre que l'erreur quadratique moyenne de prédiction des provisions  $msep(\hat{R}_i) = mse_p(\hat{C}_{i,n}) = \mathbb{E}(\hat{C}_{i,n} - C_{i,n} \mid \mathcal{D}_n)$  peut s'estimer  $\forall i$  par :

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

Quant à l'estimation de l'erreur quadratique moyenne des réserves totales, Mack propose :

$$\widehat{mse}p \left( \sum_{i=1}^n \hat{R}_i \right) = \sum_{i=2}^n \left\{ \widehat{mse}p(\hat{R}_i) + \hat{C}_{i,n} \left( \sum_{j=i+1}^n \hat{C}_{j,n} \right) \sum_{k=n+1-i}^{n-1} \frac{2\hat{\sigma}_k^2/\hat{f}_k^2}{\sum_{p=1}^{n-k} C_{n,k}} \right\} \quad (2.4)$$

[MACK 1999] ajoute à ce modèle la formule récursive suivante, permettant d'implémenter facilement les calculs :

$$\widehat{mse}p(\hat{R}_i) = C_{i,n}^2 \sum_{k=n+1-i}^{n-1} \frac{\widehat{mse}p(\hat{F}_{i,k}) + \widehat{mse}p(\hat{f}_k)}{\hat{f}_k^2}$$

où :

- $\widehat{mse}p(\hat{F}_{i,k}) = \hat{\sigma}_k^2/\hat{C}_{i,k}$  est une estimation de  $\mathbb{V}(F_{i,j} \mid C_{i,n}, \dots, C_{i,k})$
- $\widehat{mse}p(\hat{f}_k) = \hat{\sigma}_k^2/\sum_{j=1}^{n-k} C_{i,k}$  est une estimation de  $\mathbb{V}(\hat{f}_k) = \sigma_k^2/\sum_{j=1}^{n-k} C_{j,k}$

Remarquons que cette formule était donnée dans un cadre d'hypothèses plus général. Se référer à [MACK 1999] pour le détail et les preuves associées à ces estimateurs.

Notons que dans toute cette partie,  $\hat{C}_{i,j}$  peut faire référence à  $C_{i,j}$  si celui-ci est dans  $\mathcal{D}$ , ou bien à son estimateur *Chain-ladder* sinon.

Ainsi, sous des hypothèses généralement acceptables, le modèle de Mack nous donne une estimation de la variance des provisions dite à *l'ultime*. Le modèle de Merz-Wüthrich vient compléter cela par une estimation de l'erreur commise à *un an*, après l'avoir définie.

### 2.1.3 Modèle de Merz-Wüthrich

Michael Merz et Mario Wüthrich développent en 2007-2008<sup>2</sup>, en s'appuyant sur le modèle de Mack, une estimation du risque à un an. C'est cette version qui est résumée ici pour la compréhension du lecteur, pour plus de détails et les preuves correspondantes nous renvoyons aux références sus-citées.

Rappelons que l'ensemble des informations disponibles au temps  $n$  est noté  $\mathcal{D}_n$ . De manière évidente, nous noterons  $\mathcal{D}_{n+1}$  l'ensemble des informations disponibles l'année suivante. Cela correspond simplement au fait que  $\sigma(\mathcal{D}_{n+1})$  est un grossissement de la filtration  $\sigma(\mathcal{D}_n)$ .

#### 2.1.3.1 Hypothèses

Merz-Wüthrich se placent dans des hypothèses plus contraignantes que celles de Mack, mais les englobants :

**MW1** Les  $C_{i,j}$  sont indépendants pour des rattachements  $i$  différents.

2. *c.f* [WÜTHRICH, MERZ et LYSENKO 2009],[MERZ et WÜTHRICH 2007b] pour les premières versions du modèle, et [MERZ et WÜTHRICH 2008a] pour une version "simplifiée, modifiée et commentée"

**MW2**  $(C_{i,j})_{j \geq 1}$  sont des processus markoviens et il existe  $f_j, j \geq 1$  et  $\sigma_j, j \geq 1$  tels que  $\forall i, j$  on ait :

$$\mathbb{E}(C_{i,j+1} | C_{i,j}) = f_j C_{i,j} \text{ et } \mathbb{V}(C_{i,j+1} | C_{i,j}) = \sigma_j^2 C_{i,j}$$

Ces hypothèses englobant celles de Mack, les estimateurs proposés dans la précédente section sont toujours valides et conservent leurs propriétés. Comme nous travaillons ici avec 2 filtrations, nous noterons par un exposant  $n$  les estimations faites au temps  $n$  et par un exposant  $n+1$  celles faites au temps  $n+1$ . Ainsi, si  $\hat{f}_j^n$  correspond à l'estimateur  $\hat{f}_j$  précédent (donné par (2.2)), on a :

$$\hat{f}_j^{n+1} = \frac{\sum_{i=1}^{n+2-j} C_{i,j+1}}{\sum_{i=1}^{n+2-j} C_{i,j}}$$

### 2.1.3.2 Définitions

Merz-Wüthrich définissent le **True Claim development result**, ou plus simplement *CDR*, comme étant la différence entre l'estimation faite au temps  $n+1$  des ultimes moins celle faite au temps  $n$  : le *CDR* pour l'année de rattachement  $i$  pour l'année comptable  $(n; n+1]$  est donné par :

$$CDR_i(n+1) = \mathbb{E}(C_{i,n} | \mathcal{D}_n) - \mathbb{E}(C_{i,n} | \mathcal{D}_{n+1})$$

Et le *CDR* global est donné par  $\sum_{i=1}^n CDR_i(n+1)$ . Notons que son espérance sachant  $\mathcal{D}_n$  est nulle.

D'autre part, ils définissent le **Claim development result observable**, noté ici  $\widehat{CDR}$ , servant à approximer le *Claim development result* (qui n'est pas observable) de la manière suivante :

$$\widehat{CDR}_i(n+1) = \hat{C}_{i,n}^n - \hat{C}_{i,n}^{n+1}$$

Merz-Wüthrich dérivent ensuite sous ces hypothèses des estimateurs des erreurs de prédiction quadratiques conditionnelles suivantes :

- $msep_{\widehat{CDR}_i(n+1)|\mathcal{D}_n}(0) = \mathbb{E}((\widehat{CDR}_i(n+1) - 0)^2 | \mathcal{D}_n)$
- $msep_{CDR_i(n+1)|\mathcal{D}_n}(\widehat{CDR}_i(n+1)) = \mathbb{E}((CDR_i(n+1) - \widehat{CDR}_i(n+1))^2 | \mathcal{D}_n)$

Si la première donne une vision prospective intéressante de la qualité du provisionnement, la seconde en donne une vision rétrospective.

### 2.1.3.3 Estimateurs

Merz-Wüthrich dérivent des estimateurs pour ces deux erreurs, qu'elles soient par année de rattachement ou agrégées. Nous reprenons ici leurs résultats. Définissons tout d'abord :

$$\begin{aligned}
\bullet \quad \hat{\Delta}_{i,n}^n &= \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^n)^2}{\sum_{p=1}^{i+1} C_{p,j}} + \sum_{j=n-i+1}^n \left( \frac{C_{n-j,j}}{\sum_{p=1}^{n-j+2} C_{p,j}} \right)^2 \frac{\hat{\sigma}_j^2 / (\hat{f}_j^n)^2}{\sum_{p=1}^{n-j+1} C_{p,j}} \\
\bullet \quad \hat{\Phi}_{i,n}^n &= \sum_{j=n-i+1}^n \left( \frac{C_{n-j,j}}{\sum_{p=1}^{n-j+2} C_{p,j}} \right)^2 \frac{\hat{\sigma}_j^2 / (\hat{f}_j^n)^2}{C_{n-j+1,j}} \\
\bullet \quad \hat{\Lambda}_{i,n}^n &= \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^n)^2}{\sum_{p=1}^{i+1} C_{p,j}} + \sum_{j=n-i+1}^n \left( \frac{C_{n-j,j}}{\sum_{p=1}^{n-j+2} C_{p,j}} \right)^2 \frac{\hat{\sigma}_j^2 / (\hat{f}_j^n)^2}{\sum_{p=1}^{n-j+1} C_{p,j}} \\
\bullet \quad \hat{\Psi}_i^n &= \frac{\hat{\sigma}_{n-i}^2 / (\hat{f}_{n-i}^n)^2}{C_{i,n-i+1}} \\
\bullet \quad \hat{\Gamma}_{i,n}^n &= \hat{\Phi}_{i,n}^n + \hat{\Psi}_i^n \\
\bullet \quad \hat{\Xi}_{i,n}^n &= \hat{\Phi}_{i,n}^n + \frac{\hat{\sigma}_{n-i+1}^2 / (\hat{f}_{n-i+1}^n)^2}{\sum_{p=1}^{i+1} C_{p,i}}
\end{aligned}$$

Puis, les estimateurs des erreurs quadratiques de prédictions précédentes pour une seule année de rattachement sont données par :

$$\begin{aligned}
\bullet \quad \widehat{Var}(CDR_i(n+1) | \mathcal{D}_n) &= (\hat{C}_{i,n}^n)^2 \hat{\Psi}_i^n \\
\bullet \quad \widehat{mse}_{\widehat{CDR}_i(n+1)|\mathcal{D}_n}(0) &= (\hat{C}_{i,n}^n)^2 (\hat{\Gamma}_{i,n}^n + \hat{\Delta}_{i,n}^n) \\
\bullet \quad \widehat{mse}_{\widehat{CDR}_i(n+1)|\mathcal{D}_n}(\widehat{CDR}_i(n+1)) &= (\hat{C}_{i,n}^n)^2 (\hat{\Phi}_{i,n}^n + \hat{\Delta}_{i,n}^n)
\end{aligned}$$

Enfin, les estimateurs agrégés sont donnés par :

$$\begin{aligned}
\bullet \quad \widehat{mse}_{\sum_{i=1}^n CDR_i(n+1)|\mathcal{D}_n}(\widehat{CDR}_i(n+1)) &= \sum_{i=1}^n \widehat{mse}_{\widehat{CDR}_i(n+1)|\mathcal{D}_n}(\widehat{CDR}_i(n+1)) \\
&\quad + 2 \sum_{k>i>1} \hat{C}_{i,n}^n \hat{C}_{k,n}^n (\hat{\Phi}_{i,n}^n + \hat{\Lambda}_{i,n}^n) \\
\bullet \quad \widehat{mse}_{\sum_{i=1}^n \widehat{CDR}_i(n+1)|\mathcal{D}_n}(0) &= \sum_{i=1}^n \widehat{mse}_{\widehat{CDR}_i(n+1)|\mathcal{D}_n}(0) \\
&\quad + 2 \sum_{k>i>1} \hat{C}_{i,n}^n \hat{C}_{k,n}^n (\hat{\Xi}_{i,n}^n + \hat{\Lambda}_{i,n}^n)
\end{aligned} \tag{2.5}$$

Le modèle de Merz-Wüthrich a l'avantage d'apporter ces formules fermées. Cependant ses hypothèses sont très fortes<sup>3</sup>, et nous essayerons dans la suite du chapitre d'arriver à des résultats équivalents sous des hypothèses plus faibles.

## 2.2 Réécriture du modèle de Mack

Cette section vise à décrire quelques Modèles Linéaires Généralisés englobants *Chain-ladder*, et vise ensuite à prouver la conjecture faite par [ENGLAND et VERRALL 2002], point 7.6.5, selon laquelle le modèle de Mack correspond à une modélisation jointe d'un Modèle Linéaire

3. Notamment la seconde, **MW2**

Généralisé binomial négatif avec approximation normale et d'une variance dépendant uniquement des colonnes. De plus, nous étendrons ce résultat à des hypothèses plus faibles et redémontrerons la formule de Mack de cette manière.

Les développements stochastiques du modèle *Chain-ladder* ont été beaucoup traités dans les années 1980-2000. Notons entre autre la *discussion* entre [MACK 1994], [VERRALL 2000], [MACK et VENTER 2000],[VERRALL et ENGLAND 2000] visant à déterminer quel était le modèle d'ordre 2 englobant *Chain-ladder* le plus pertinent. [ENGLAND et VERRALL 2002] clarifie et clôt la discussion, et c'est cet aboutissement que nous allons d'abord décrire puis analyser et compléter ici.

Pour cela, après avoir rappelé ce qu'étaient les Modèles Linéaires Généralisés et quasi-Modèles Linéaires Généralisés, nous commencerons par traiter le cas du modèle de Poisson sur-dispersé, qui est le Modèle Linéaire Généralisé le plus classique en provisionnement non-vie.

### 2.2.1 Retour sur les (quasi-)Modèles Linéaires Généralisés

Un Modèle Linéaire Généralisé, développé par [WEDDERBURN 1974], est un modèle pour une **variable réponse**  $Y$  s'appuyant sur des **covariables explicatives**  $X = (X_1, \dots, X_p)$ . Il s'écrit de manière traditionnelle comme :

$$g(\mu) = g(\mathbb{E}(Y)) = \beta^T X$$

,

avec :

- $g$ , la *fonction lien*
- $\beta$  un paramètre de dimension  $p$
- $Y$ , une variable aléatoire appartenant à la famille exponentielle, c'est-à-dire :
  - sa densité s'écrit sous la forme  $f(y) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\}$ , ou :
  - $\theta, \phi$  sont respectivement le *paramètre canonique* et le *paramètre de dispersion*
  - $b$  et  $c$  déterminent la "forme" de la distribution, en particulier :
    - $\mathbb{E}(Y) = b'(\theta)$ , et
    - $\mathbb{V}(Y) = b''(\theta)a(\phi)$ .

Souvent, la convention  $a(\phi) = \phi/\omega$  est prise pour pouvoir pondérer les observations. Avec ou sans cette convention, nous avons un lien entre la variance et l'espérance, que l'on exprime par la *fonction variance* :

$$V(\mu) = b''(\theta) = b''(b'^{-1}(\mu))$$

Un tel modèle peut, dans certains cas, avoir une solution explicite pour le paramètre  $\beta$ , mais en règle générale on cherche les paramètres via une maximisation de la vraisemblance du modèle (ou, plus couramment, de son log). Cela consiste, pour  $n$  observations  $(x_i, y_i)$  de la réponse et des covariables, à maximiser :

$$\ln(\mathcal{L}(\theta)) = \ln\left(\prod_{i=1}^n f(y_i)\right) = \sum_{i=1}^n \ln(f(y_i))$$

Cette maximisation se fait en général par l'algorithme de Newton-Raphson, puisque la dérivée seconde est explicite, bien que d'autres possibilités existent.

L'estimateur obtenu est alors un **estimateur du maximum de vraisemblance**, ou *maximum likelihood estimator*, MLE, qui a certaines propriétés intéressantes :

- le MLE est asymptotiquement efficace,
- il est asymptotiquement non-biaisé,
- il est consistant,
- il atteint la borne de Cramér-Rao (variance minimum dans sa classe),
- il peut cependant être biaisé.

Les Modèles Linéaires Généralisés sont de très bons modèles appréciés par les statisticiens, mais ils requièrent une spécification complète de la distribution de la variable réponse. Souvent, ce niveau d'hypothèse n'est pas possible, et une extension classique des Modèles Linéaires Généralisés sont les modèles dits quasi-Modèles Linéaires Généralisés : au-lieu de supposer  $Y$  dans la famille exponentielle, nous nous donnons une hypothèse d'ordre 2 uniquement, la forme de la fonction variance  $V(\mu)$ .

Le souci principal est que dans ce modèle plus général, il n'y a pas de vraisemblance disponibles pour la variable réponse. On utilise alors comme cible de l'optimisation la fonction de **log-quasi-vraisemblance**, ou plus simplement  $Q$ , définie comme :

$$Q(\mu, Y) = \sum_{i=1}^n n Q_i(\mu_i, y_i) = \sum_{i=1}^n n \int_{y_i}^{\mu_i} \frac{y_i - t}{\phi_i V(t)} dt$$

Notons que pour certains cas simples de fonction variance, typiquement  $V(\mu) = \mu$  ou bien  $V(\mu) = 1$ , correspondant respectivement à la fonction variance d'une loi de Poisson ou d'une loi normale auxquelles nous nous intéressons *supra*, la fonction de quasi-vraisemblance correspond exactement aux vraisemblances respectivement Poisson ou normal. Cette simplification est en fait disponible pour toutes les fonctions variances de la famille exponentielle<sup>4</sup>.

Rappelons enfin que les **Résidus de Pearson** d'un modèle Modèle Linéaire Généralisé ou quasi-Modèle Linéaire Généralisé sont définis, en omettant les indices, par :

$$r = \frac{(Y - \mu)}{\sqrt{\phi/\omega V(\mu)}} \quad (2.6)$$

De nombreux documents sont disponibles sur les Modèles Linéaires Généralisés, le plus complet et abordable est [McCULLAGH et NELDER 1989]. Ce sera tout au long de cette section notre référence, et la plupart des concepts abordés ici y sont décrits, illustrés et commentés.<sup>5</sup>

### 2.2.2 Modèle de Poisson sur-dispersé

Dans le cadre du provisionnement non-vie, un des Modèles Linéaires Généralisés le plus utilisé est le quasi-Modèle Linéaire Généralisé Poisson sur-dispersé, ou **Over-Dispersed Poisson**

4. C'est d'ailleurs ce qui explique que la plupart des implémentations disponibles pour les Modèles Linéaires Généralisés dans les logiciels de statistiques sont en fait implémentés directement sous des hypothèses de quasi-Modèle Linéaire Généralisé, sauf pour certaines statistiques.

5. Notamment le point 9.2.2 de [McCULLAGH et NELDER 1989] décrit le choix de cette fonction de quasi-vraisemblance pour les quasi-Modèle Linéaire Généralisé.

**Model.** Il est notamment étudié par [ENGLAND et VERRALL 1998 ; ENGLAND et VERRALL 2006].

En considérant comme dans la section précédente le triangle d'incrémentes  $X_{i,j}$ , le Modèle Linéaire Généralisé Poisson sur-dispersé a pour hypothèses :

**ODP1** Les incréments  $X_{i,j}$  sont indépendants

**ODP2**  $\exists x, y \in \mathbb{R}^n \mid \forall i, j \in \{1, \dots, n\}, \mathbb{E}(X_{i,j}) = x_i y_j$

**ODP3**  $\exists \phi > 0 \mid \forall i, j \in \{1, \dots, n\}, \mathbb{V}(X_{i,j}) = \phi \mathbb{E}(X_{i,j})$

**ODP4**  $\sum_{j=1}^n y_j = 1$

On remarque d'emblée que la structure de la variance donnée dans **ODP3** pose une contrainte de positivité sur l'espérance des incréments. Contrairement à ce qu'on pourrait penser, ce modèle n'est pas restreint aux incréments positifs, mais au fait que les sommes des lignes et que les sommes des colonnes du triangle incrémental soient toutes positives.

Notons de plus qu'à l'ordre 1, ce modèle *englobe* bien l'algorithme de *Chain-ladder*. En effet les équations de quasi-vraisemblance du modèle s'écrivent :

$$\forall i \in \{1, \dots, n\}, \mathcal{Q}_i^1 : \sum_{j=1}^{n+1-i} x_i y_j = \sum_{j=1}^{n+1-i} X_{i,j}$$

$$\forall j \in \{1, \dots, n\}, \mathcal{Q}_j^2 : \sum_{i=1}^{n+1-j} x_i y_j = \sum_{i=1}^{n+1-j} X_{i,j}$$

ce qui correspond aux **marginal sum equations** très en vogue dans les années 80-90. [MACK 1991] a montré (appendix A) que ces équations aboutissaient à l'estimateur de *Chain-ladder*. Pour une preuve plus longue et plus descriptive, *c.f* [SCHMIDT et WÜNSCHE 1998].

Il est notamment précisé dans ces preuves que le modèle n'est pas bien spécifié sans la quatrième hypothèse car il aurait alors une infinité de solutions en multipliant tous les  $x_i$  par un réel non nul et en divisant tout les  $y_j$  par ce même réel. Le choix de la somme des  $y_j$  est donc arbitraire, mais comme prendre  $\sum_{j=1}^n y_j = 1$  permet d'obtenir les ultimes par ligne directement dans les paramètres  $x_i$ , ce choix est souvent fait.

L'ensemble des paramètres  $y_j$  s'appelle **la cadence de développement** du triangle. Cette cadence est en bijection avec les coefficients de passage du modèle *Chain-ladder*  $f_j$  ( en considérant que la somme des  $y_j$  est connue) via le système suivant :

$$\forall i \in \{1, \dots, n-1\}, f_j = \frac{\sum_{p=1}^{j+1} y_p}{\sum_{p=1}^j y_p} \quad (2.7)$$

Notons que [VERRALL 2000] traite la question soulevée par Mack du nombre de paramètres supérieur au nombre de paramètres de l'algorithme de *Chain-ladder*. En effet, la vraisemblance du modèle, ou sa quasi-vraisemblance, peut être découpée en deux parties : une partie estimant le triangle inférieur et une deuxième partie, conditionnelle sur le triangle inférieur, estimant les cadences de développement. La seconde partie de cette vraisemblance est celle qui nous intéresse si nous souhaitons reproduire l'estimation faite par Mack, en prenant des hypothèses d'espérance et de variance seulement conditionnelles.

Bien que l'estimation des ultimes soit la même quelle que soit l'approche, l'erreur associée à cette estimation est par contre différente : dans le cas où l'on conditionne sur les observations, on les considère comme données, fixes, sans erreur. Au contraire, pour le cas où la vraisemblance est prise dans son entier, l'erreur des observations n'est plus nulle. Le nombre de paramètres est aussi très différent dans les deux cas.

Remarquons finalement que, dans le modèle de Poisson sur-dispersé, si nous supprimons l'hypothèse de variance, la contrainte de positivité disparaît et nous obtenons tout simplement un modèle uniquement d'ordre 1 *englobant Chain-ladder*.

### 2.2.3 Modèle binomial négatif et approximation normale

Le modèle binomial négatif, dérivé par [VERRALL 2000] depuis le modèle de Poisson sur-dispersé en conditionnant sur les observations pour supprimer les paramètres de lignes (voir aussi [ENGLAND et VERRALL 2002]), est un quasi-Modèle Linéaire Généralisé sur les incréments ou sur les cumulés conditionnels. Son développement passe par un reparamétrage du modèle Poisson sur-dispersé pour utiliser cette vraisemblance conditionnelle uniquement, reparamétrage que Verrall appelle *récuratif*. Ses hypothèses sont :

**ONB1** Les lignes du triangle sont indépendantes

**ONB2**  $\exists \mathbf{f} \mid \forall i, j \in \{1, \dots, n\}, \mathbb{E}(C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}) = f_j C_{i,j}$

**ONB3**  $\mathbb{V}(C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}) = f_j(f_j - 1)C_{i,j}$

Remarquons que ce modèle est un quasi-Modèle Linéaire Généralisé *conditionnel* : la vraisemblance du modèle complet, tout comme sa quasi-vraisemblance, sont données conditionnellement aux observations de la partie supérieure du triangle.

Ce modèle englobe bien *Chain-ladder* (les prédictions à l'ordre 1 sont identiques), et a l'avantage (sous cette forme) d'être paramétré directement par les coefficients de passage. Cependant, comme le modèle de Poisson, l'hypothèse de variance est invalidée si la somme marginale d'une colonne est négative. C'est pour résoudre ce problème que [ENGLAND et VERRALL 2002] propose une approximation normale de ce modèle, dont les hypothèses sont :

**ONBNA1** Les lignes du triangle sont indépendantes

**ONBNA2**  $\exists \mathbf{f} \mid \forall i, j \in \{1, \dots, n\}, \mathbb{E}(C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}) = f_j C_{i,j}$

**ONBNA3**  $\exists \phi \mid \forall i, j \in \{1, \dots, n\}, \mathbb{V}(C_{i,j+1} \mid C_{i,1}, \dots, C_{i,j}) = \phi_j C_{i,j}$

Ce dernier modèle a un paramètre de dispersion qui n'est pas commun à l'ensemble des observations (mais simplement commun à chaque année de développement, comme l'était la variance du modèle précédent), comme c'est en général le cas dans les Modèles Linéaires Généralisés. Sa résolution fait donc appel à la *modélisation jointe* (*c.f* [MCCULLAGH et NELDER 1989], chapitre 10), qui peut se voir de la manière suivante ([ENGLAND et VERRALL 2002]) dans ce cas particulier :

En posant  $w_{i,j} = C_{i,j-1}$ , correspondants à des poids *a priori*, le modèle peut s'écrire :

$$\mathbb{E}\left(\frac{C_{i,j}}{w_{i,j}}\right) = f_j \text{ et } \mathbb{V}\left(\frac{C_{i,j}}{w_{i,j}}\right) = \phi_j / C_{i,j-1}$$

ou encore :

$$\mathbb{E}(F_{i,j}) = f_j \text{ et } \mathbb{V}(F_{i,j}) = \phi_j / C_{i,j-1}$$

La procédure de modélisation jointe consiste à prendre des valeurs initiales arbitraires pour les paramètres  $\phi_j$ , et à ajuster le modèle ci-dessus avec des poids  $W_{i,j} = \frac{w_{i,j}}{\phi_j}$  inversement proportionnels à la dispersion.

La statistique de dispersion peut-être choisie parmi deux candidats : les résidus de Pearson ou les résidus de déviance. Notons que dans un modèle à variance normale, comme c'est le cas ici, les deux formes sont équivalentes. Nous utiliserons la notation des résidus de Pearson. Dans notre cas particuliers, les résidus donnés par (2.6) deviennent :

$$r_{i,j} = \frac{F_{i,j} - \hat{f}_j}{\sqrt{\phi_j / \omega_{i,j} V(\hat{f}_j)}} = \sqrt{W_{i,j}} (F_{i,j} - \hat{f}_j)$$

En effet, dans notre cas normal,  $V(\hat{f}_j) = \phi_j / C_{i,j-1} V(\hat{f}_j)$ .

Le modèle de dispersion s'exprime alors comme (cf [GOULD 2008] ou [VERRALL 2000] ou encore [ENGLAND et VERRALL 2002], ou plus à la racine du problème [MCCULLAGH et NELDER 1989] chapitre 10) :

$$\mathbb{E}(r_{i,j}^2) = \phi_j \text{ et } \mathbb{V}(r_{i,j}^2) = 2\phi_j^2$$

Remarquons que le quasi-Modèle Linéaire Généralisé pour la dispersion correspond à une spécification gamma. Cependant l'hypothèse de variance de ce modèle est sans intérêt et peut-être supprimée : un autre choix pour la fonction variance serait sans effet sur l'estimation d'ordre 1 des paramètres de dispersion.

En ajustant ce modèle, on obtient facilement que les  $\phi_j$  sont les moyennes par colonnes des  $r_{i,j}^2$ , i.e :

$$\hat{\phi}_j = \frac{1}{n-j} \sum_{j=1}^{n+1-j} C_{i,j} (F_{i,j} - f_j)^2$$

Cependant, cet estimateur n'est pas satisfaisant car il n'est pas ajusté par le degré de liberté du modèle : en effet, ([MCCULLAGH et NELDER 1989], point 10.5.3) le modèle de dispersion ne prend pas en compte l'estimation précédente des  $f_j$ , un par colonne<sup>6</sup>. De cette manière, il nous faut *corriger* nos résidus par :

$$r_{i,j} = \sqrt{\frac{n-j}{n-j-1}} C_{i,j} (F_{i,j} - \hat{f}_j)$$

ce qui donne pour l'estimation des paramètres de dispersions :

$$\hat{\phi}_j = \frac{1}{n-j-1} \sum_{j=1}^{n+1-j} C_{i,j} (F_{i,j} - f_j)^2$$

Notons enfin que cette estimation n'est pas possible sur la dernière colonne du triangle, au dernier développement : il n'y a en effet qu'une seule observation.

6. Quant à l'ajustement pour le Kurtosis normalisé ([MCCULLAGH et NELDER 1989], point 10.5.1) de l'hypothèse de variance du modèle de dispersion, dans notre cas normal  $(1 + \rho_4/2) = 1$  et cet ajustement n'est donc pas nécessaire.

## 2.2.4 Conclusion

En regardant précisément les hypothèses du modèle précédent, l'approximation normale du modèle binomial négatif, on voit qu'elles sont équivalentes aux hypothèses prises par Mack, en posant  $\phi = \sigma^2$ .

Les deux modèles arrivant aux mêmes estimateurs des  $f_j$  et des  $\sigma_j$ , nous sommes donc en mesure d'affirmer l'équivalence entre les deux modèles, supputée par [ENGLAND et VERRALL 2002]. En effet, les réserves exprimées par [ENGLAND et VERRALL 2002] sur les *ajustements* de l'hypothèse de variance faits par Mack sont en fait explicables dans la théorie des quasi-Modèles Linéaires Généralisés :

- le choix des résidus de pearson ou des résidus de déviance pour la construction du modèle de dispersion n'a pas d'influence dans un cadre normal,
- la renormalisation pour le nombre de paramètres déjà estimés pour la modélisation des facteurs de développement proposée par Mack est nécessaire dans un cadre de modélisation jointe,
- l'ajustement des résidus pour le Kurtosis normalisé a été oublié par Mack comme par England et Verral, mais il n'est pas influent si l'hypothèse de variance correspond à une hypothèse normale (l'ajustement est alors sans effet).

De cette manière, l'estimateur de Mack est un estimateur du quasi-maximum de vraisemblance dans un cadre de modélisation jointe, conditionnelle et pondérée, avec hypothèse de variance normale.

Cette conclusion nous fournit un cadre théorique pour le modèle de Mack qui permettra de justifier non seulement les résidus que nous emploierons dans le bootstrap récursif, mais aussi de comprendre très facilement comment sont construits les modèles de bootstrap, qu'ils visent des statistiques à l'ultime, à un an, ou simplement les coefficients de passage. Ce cadre a en effet deux principaux intérêts :

- le passage par les Modèles Linéaires Généralisés nous permet l'utilisation d'outils théoriques déjà existants pour les Modèles Linéaires Généralisés sur le modèle de Mack, notamment le bootstrap des résidus du modèle.
- le passage par les Modèles Linéaires Généralisés simplifie les implémentations futures : la plupart des logiciels de statistiques, R y compris, apportent de nombreux outils pour travailler avec et autour de Modèles Linéaires Généralisés, et beaucoup moins pour travailler avec le modèle de Mack.<sup>7</sup>

## 2.3 Le Bootstrap en provisionnement

Le bootstrap est une technique statistique introduite par Efron en 79 [EFRON 1978 ; EFRON 1979 ; EFRON 1980] avec la puissance de calcul en augmentation. L'idée est de s'intéresser à la distribution d'estimateurs.

De très bonnes références sont disponibles sur les techniques de bootstrap en statistique générales, par exemple [EFRON 1994], ainsi qu'en provisionnement non-vie, par exemple [PINHEIRO, PAULO J. R., ANDRADE E SILVA, JOÃO M. et CENTENO, MARIA DE LOURDES 2001 ; TEE, KÄÄRIK et VIIN 2017 ; ZBOŇ 2014].

---

7. La librairie *ChainLadder* de R, voir par exemple [GESMANN 2009 ; GESMANN 2017], permet d'appliquer des modèles de provisionnement non-vie, mais n'est pas assez permissive sur les hypothèses prises et trop restrictive sur les formes de modèles proposées.

En particulier, dans le modèle de Mack, les simulations par bootstrap peuvent permettre de facilement obtenir des estimations pour le risque à un an ou pour le risque à l'ultime sans passer par les formules de Mack ou de Merz-Wüthrich. [BOUMEZOUED et al. 2011] (parmi d'autres) démontre d'ailleurs une convergence vers la formule fermée de Merz-Wüthrich.

Après quelques brefs rappels sur le bootstrap en général, nous nous attacherons dans cette section à décrire la procédure générale de bootstrap dans un cadre de provisionnement par Modèles Linéaires Généralisés, et nous verrons le cas particulier du modèle de Mack. Les techniques de bootstrap présentées ici permettent une grande souplesse dans les hypothèses du modèle (cadre quasi-Modèle Linéaire Généralisé), et surtout une extensivité des résultats : distribution du *Claim development result*, distribution des ultimes, calcul de risque de réserve à un an...).

Dans la prochaine section, nous aborderons une procédure générale (et le cadre théorique correspondant) de synchronisation des résidus de plusieurs triangles permettant d'étendre ces résultats au cadre multivarié.

### 2.3.1 Brefs rappels sur le Bootstrap

Le principe du bootstrap est le suivant : à partir d'un échantillon statistique de variables aléatoires indépendantes et identiquement distribuées  $\mathbf{X} = (X_1, \dots, X_n)$ , on échantillonne avec remise  $B$  nouveaux *pseudo-échantillons* indexés par  $b \in \{1, \dots, B\}$  :  $\mathbf{X}^{(b)} = (X_1^{(b)}, \dots, X_n^{(b)})$ .

Supposons qu'une statistique, i.e une fonction  $f$  du vecteur aléatoire  $X$  nous intéresse. La *distribution bootstrap* de cette statistique est  $\{f(\mathbf{X}^{(b)}), \forall b \in \{1, \dots, B\}\}$ . Cette distribution bootstrap a une moyenne empirique, une variance empirique, des quantiles, et une distribution de probabilité empirique.

Cette technique simple est très puissante : à partir d'un échantillon statistique supposé *i.i.d* on peut créer des lois empiriques pour des statistiques complexes. En provisionnement non-vie, l'application de ce principe est classique et souvent utilisé pour retrouver des distributions empiriques des réserves, des ultimes ou des paiements futurs.

### 2.3.2 Dans le cadre des quasi-Modèles Linéaires Généralisés

En provisionnement non-vie, nous disposons de variables aléatoires (incrément, cumul,...) qui ne sont pas *i.i.d*. Cependant, si nous prenons le cadre d'une modélisation quasi-Modèle Linéaire Généralisé, par exemple le modèle poissonien ou le modèle de Mack, les résidus de Pearson du modèle constituent une matière de choix pour un bootstrap.

En effet, il existe une **relation bijective** entre les résidus et n'importe quelle information du triangle, sachant, bien sûr, les paramètres choisis pour le modèle (et les poids). Par exemple, dans un modèle de Mack :

$$r_{i,j} = \sqrt{W_{i,j}}(F_{i,j} - \hat{f}_j) \Leftrightarrow F_{i,j} = \hat{f}_j + \frac{r_{i,j}}{\sqrt{W_{i,j}}} \quad (2.8)$$

La donnée des  $F_{i,j}$  (et des poids) étant suffisante pour construire le modèle de Mack, la donnée des résidus est donc elle aussi suffisante pour reconstruire le modèle. Ce principe s'applique à tout Modèle Linéaire Généralisé ou quasi-Modèle Linéaire Généralisé.

Un Bootstrap peut être **paramétrique**, c'est-à-dire que la loi des résidus est connue (ou approchée), et que les nouveaux résidus sont simulés directement de cette loi, ou **non-**

**paramétrique**, c'est-à-dire que les résidus sont tirés avec remise parmi les résidus empiriquement observés.

La procédure de bootstrap en provisionnement non-vie peut ainsi se concevoir de la manière suivante :

- sur le triangle original, calculer le paramétrage d'un modèle quasi-Modèle Linéaire Généralisé au choix.
- rééchantillonner les résidus du modèle et en reconstruire des pseudo triangles, ou trapèzes selon les cas :
  - dans le cas où l'on recherche une statistique à date, rééchantillonner les résidus sur le triangle supérieur.
  - dans le cas où l'on recherche une statistique à horizon  $a$  années,  $a \in \{1, \dots, n\}$ , rééchantillonner des résidus supplémentaires sur les  $a$  prochaines diagonales.
  - dans le cas où l'on recherche une statistique à l'ultime, *i.e*  $a = n$ , rééchantillonner les résidus sur l'ensemble du rectangle.
- par inversion de la formule des résidus (2.8), reconstituer les pseudo-triangles ou pseudo-trapèzes.
- calculer la statistique d'intérêt pour chaque échantillon bootstrap.

Dans des cas d'un bootstrap paramétrique, certains raccourcis peuvent être utilisés si l'on connaît la loi de la statistique d'intérêt,<sup>8</sup> ou la loi d'autres éléments que les résidus qui s'en rapprochent plus, en simulant simplement depuis ces distribution. Cependant le calcul de telles distributions n'est pas toujours aisé et demande des hypothèses supplémentaires.

### 2.3.3 Sous les hypothèses du quasi-Modèles Linéaires Généralisés de Mack

Sous les hypothèses du modèle de Mack, la procédure de bootstrap est simple à mettre en œuvre. Ce bootstrap particulier s'appelle dans la littérature *récuratif* [VERRALL 2000]. Il peut être décrit de la manière suivante :

- sur le triangle original, calculer les paramètres du modèle de Mack  $\hat{f}_j, \hat{\sigma}_j^2$  et enfin les résidus de Pearson  $r_{i,j}$ , via (2.6).
- rééchantillonner ces résidus, noter  $r_{i,j}^{(b)}$ ,  $b \in \{1, \dots, B\}$  ces échantillons, reconstruire via l'inversion de leur formule (2.8)  $B$  pseudo-triangle de coefficients de passage individuels :

$$F_{i,j}^{(b)} = r_{i,j}^{(b)} \sqrt{\frac{\hat{\sigma}_j^2}{C_{i,j}}} + \hat{f}_j$$

- calculer pour chacun de ces triangles les coefficients de passage  $f_j^{(b)}$  en prenant soin de pondérer par les  $C_{i,j}$  du triangle d'origine et non pas les  $C_{i,j}^{(b)}$ .<sup>9</sup>

Nous avons ici le choix dans le modèle de Mack de l'année de vision que nous souhaitons pour nos statistiques bootstrap. Si nous souhaitons simplement obtenir des distributions des provisions à l'ultime, on notera  $a = 0$  la différence de temps entre la date de calcul et la date d'évaluation de nos statistiques. Dans le cadre de statistiques à un an, par exemple le *Claim development result*, on posera  $a = 1$ . Pour des statistiques à  $k$  années, on posera  $a = k$ , et la procédure continuera ainsi :

8. Notamment des raccourcis permettant de récupérer la variance à l'ultime sans rééchantillonner sur le triangle inférieur, voir VERRALL 2000

9. Cela s'explique et se justifie par le fait que les  $C_{i,j}$  du triangle original sont des *poids* dans le quasi-Modèle Linéaire Généralisé de Mack. Ces poids sont *a priori* : on travaille ici conditionnellement aux observations.

- simuler pour chaque pseudo-triangle la prochaine diagonale en multipliant la dernière diagonale cumulée obtenue  $C_{i,n+1-i}^{(b)} \forall i$  par des réalisations de moyenne  $f_j^{(b)}$  et de variance  $\frac{\sigma_{n+1-i}^2}{C_{i,n+1-i}^{(b)}}$  :
  - dans un cadre de bootstrap paramétrique, prendre des lois  $\mathcal{N}(f_j^{(b)}, \frac{\sigma_{n+1-i}^2}{C_{i,n+1-i}^{(b)}})$
  - dans un cadre de bootstrap non-paramétrique, tirer avec remise de nouveaux résidus  $r_{i,n+2-i}^{(b)}$  puis remplacer les lois normales précédentes par  $r_{i,n+2-i}^{(b)} \frac{\sigma_{n+1-i}^2}{C_{i,n+1-i}^{(b)}} + f_j^{(b)}$
- si besoin (si  $a \geq 2$ ), simuler de même les  $a - 1$  diagonales suivantes.
- récupérer les incréments  $X_{i,n+1-i}^{(b)}$  correspondants aux diagonales simulées dans chaque pseudo-triangle
- appliquer l'algorithme de *Chain-ladder* aux pseudo-trapezoides constitués du triangle d'origine et de chacune de ces nouvelles diagonales pour obtenir des ultimes.
- récupérer sur chaque triangle les statistiques d'intérêt.

Le *Claim developpement result*, par exemple, se calcule alors pour chaque simulation comme la différence entre l'ultime vu au temps  $t$  (fixe) obtenu par *Chain-ladder* et l'ultime (différents selon les simulations) vu l'année suivante, additionné des incréments simulés. Il y a cependant un choix à faire sur la nature de ces incréments simulés :

- la simulation de nouvelles diagonales par une loi normale est cohérente avec les hypothèses du modèle de Mack et permet dans le cadre Merz-Wüthrich une convergence vers le CDR.
- d'un autre côté, l'utilisation de nouveaux résidus empiriques permet de garder des hypothèses plus faibles dans le modèle.

Cette seconde possibilité consistant à rééchantillonner les résidus de la partie supérieure du triangle dans les  $a$  diagonales inférieures, est présentée et analysée dans [TAYLOR 2000, points 6.5, 7.3 et 11.4]. Taylor y discute notamment la qualité de l'estimateur obtenu et conclut à une qualité équivalente à l'estimation classique sous des hypothèses plus faibles<sup>10</sup>. La première possibilité, consistant à rééchantillonner par des lois paramétriques, est déjà proposée dans [ENGLAND et VERRALL 2002] et est développée dans le cadre du modèle poissonien par [DIERS 2010].

Ré-intuitée mais non prouvée par [HAHN 2017], cette méthodologie s'avère pratique car elle demande moins d'hypothèses qu'un ajustement normal, et est donc plus souple. Repris par [CARRATO 2016] entre autres, elle gagnerait à être plus répandue car elle présente des avantages non-négligeables. La seule nécessité pour cette technique d'estimation est que les résidus du modèle soient *i.i.d.*, ce qui est une des conditions de validité de base pour un modèle quasi-Modèle Linéaire Généralisé (et donc pour le modèle de Mack).

### 2.3.4 Principe de *zonage* des quasi-Modèles Linéaires Généralisés

Les techniques de bootstrap classiques dans le cadre quasi-Modèles Linéaires Généralisés proposent de rééchantillonner avec remise l'ensemble des résidus du modèle dans l'ensemble des emplacements possibles. C'est effectivement la meilleure chose à faire car cela augmente le nombre de valeurs et permet une meilleure approximation de la distribution bootstrap des statistiques d'intérêt.

10. Cette référence de Taylor est la plus vieille que nous ayons trouvée qui propose de prendre en compte l'erreur de processus par un rééchantillonnage directement depuis les résidus.

Cependant, dans certains cas, notamment si l'hypothèse de variance du modèle est trop restrictive, on peut observer une hétéroscédasticité des résidus.

Une des hypothèses des quasi-Modèles Linéaires Généralisés étant l'homoscédasticité des résidus, une telle constatation peut être désastreuse : mélanger des résidus à variance très différente pourrait biaiser l'estimation.

Considérons dans la plage rectangulaire contenant le triangle supérieur (connu) et le triangle inférieur (inconnu), de taille donc  $n^2$ , un ensemble  $\mathcal{Z}$  de **zones de rééchantillonnage** arbitraires, partitionnant la plage, représentées par des ensembles de couples de coordonnées  $i, j$ , qui sont telles que :

$$\forall z \in \mathcal{Z}, r_{i,j} \mid (i, j) \in z \text{ sont homoscédastiques.} \quad (2.9)$$

Alors, il est tout à fait possible de rééchantillonner ces résidus (à l'intérieur de chaque zone, donc) pour obtenir des pseudo-triangles correspondants au phénomène sous-jacent.<sup>11</sup>

Pour des triangles complètement déroulés et même "trop grands" par rapport au risque sous-jacent, les derniers résidus seront négligeables (car la charge incrémentale dans les dernières colonnes sera presque nulle), et l'utilisation d'une zone spécifique à la queue du triangle évitera un rééchantillonnage croisé entre les premières colonnes et celles-ci. : on se sert de ces zones pour s'assurer que les résidus à l'intérieur de chacune d'elles sont bien homoscédastiques.

D'un point de vue théorique, ce zonage correspond à la construction de différents quasi-Modèle Linéaire Généralisé aux hypothèses différentes, un par zone, conditionnés les uns aux autres si nécessaire. Il est donc fondamental que les paramètres estimés n'influent pas sur plusieurs zones. La taille minimum d'une zone est donc clairement la colonne si le modèle sous-jacent est un modèle de Mack.

Notons enfin que dans un cadre synchronisé, dont nous parlerons dans la prochain section, ces zones permettrons de définir la structure de dépendance entre les différents triangles.

## 2.4 Quelques extensions multivariées

Nous proposons dans cette section quelques extensions multivariées, c'est-à-dire avec plusieurs triangles, des modèles précédents. En effet, dans le cadre de notre étude nous serons amenés à étudier plusieurs triangles de liquidation de manière conjointe, sans pouvoir supposer, de part le cadre théorique fixé ou de part la contradiction avec les données, l'indépendance entre les différentes branches. Après avoir discuté plusieurs approches possibles, nous en présenterons une que nous utiliserons dans les applications suivantes, puis nous nous intéresserons aux différents bootstrap lui correspondants.

Dans cette section nous considérerons  $p$  triangles de même taille, notés  $X^{(1)}, \dots, X^{(p)}$  pour leurs versions incrémentales et  $C^{(1)}, \dots, C^{(p)}$  pour leurs versions cumulés. On note :

- $C_{i,j}^{(p)}$  la charge cumulée correspondante à l'année de rattachement  $i$ , l'année de développement  $j$  et le triangle  $p$
- $F_{i,j}^{(p)}$  les coefficients de passage individuels correspondants, selon la formule (2.1)
- $\mathbf{C}_{i,j}$  le vecteur des  $C_{i,j}^{(k)}, k \in \{1, \dots, p\}$
- $\mathbf{F}_{i,j}$  le vecteur des  $F_{i,j}^{(k)}, k \in \{1, \dots, p\}$

11. Il est important dans ce cas de s'assurer que les zones définies permettent de rééchantillonner des résidus partout où il est nécessaire de le faire.

—  $\Delta_{i,j} = \text{diag}(\mathbf{C}_{i,j})$  la matrice ayant pour diagonale les  $C_{i,j}^{(p)}$

Suivant Prohl-schmidt [PROHL et SCHMIDT 2005], la dernière notation est juste utilisée pour simplifier les écritures. On remarque en effet que :

$$\mathbf{C}_{i,j} = \Delta_{i,j} \mathbf{1}$$

où  $\mathbf{1}$  représente le vecteur unité, et que :

$$\mathbf{C}_{i,j+1} = \Delta_{i,j} \mathbf{F}_{i,j+1}$$

### 2.4.1 Deux approches : Modèle de Schnauss et Modèle de Braun

Le modèle de Schnauss [SCHMIDT et SCHNAUS 1996] est une extension au cadre bivarié du modèle de Mack, et donc de l'estimateur de *Chain-ladder*. [PROHL et SCHMIDT 2005], en construisant la forme multivariée de ce modèle, montre que les estimateurs optimaux dans ce cadre ne sont pas les mêmes que les estimateurs marginaux de différents modèles de Mack, à l'ordre deux bien sûr, mais aussi des différents algorithmes de *Chain-ladder* marginaux à l'ordre 1.

Le modèle de Schnauss consiste à généraliser (stricto-sensu) à plusieurs triangles le cadre d'hypothèses de Mack. Les hypothèses du **Modèle Multivarié de Schnauss** sont les suivantes<sup>12</sup> :

**MMS1**  $\forall j \in \{1, \dots, n-1\}, \exists \Phi_j$  vecteur aléatoire tel que,  $\forall i \in \{1, \dots, n\}$ ,

$$\mathbb{E}(\mathbf{C}_{i,j+1}) = \Delta_{i,j} \Phi_j$$

**MMS2**  $\forall k \in \{1, \dots, n-1\}, \exists \Sigma_k$  matrice aléatoire semi définie positive telle que,  $\forall i, j \in \{1, \dots, n\}$ ,

$$\text{Cov}(\mathbf{C}_{i,k}, \mathbf{C}_{j,k}) = \Delta_{i,k}^{1/2} \Sigma_k \Delta_{j,k}^{1/2} \mathbf{1}_{i=j}.$$

Notamment, sous ces hypothèses, [PROHL et SCHMIDT 2005] montrent que l'on a :

$$\mathbb{E}(\mathbf{F}_{i,j}) = \Phi_j \text{ et } \text{Cov}(\mathbf{F}_{i,k}, \mathbf{F}_{j,k}) = \Delta_{i,k}^{-1/2} \Sigma_k \Delta_{j,k}^{-1/2} \mathbf{1}_{i=j}.$$

Sous ces nouvelles hypothèses, Prohl et Schmidt trouvent également des estimateurs des ultimes et des réserves marginales et agrégées, et résolvent donc (négativement) le problème de non additivité du modèle de Mack en concluant à la non additivité des estimateurs de *Chain-ladder*, mais en fournissant une version multivariée de l'algorithme de *Chain-ladder* qui reste optimale sous les hypothèses multivariées du modèle de Schnauss. L'estimateur principal (non biaisé et meilleur dans sa catégorie) est le suivant :

$$\hat{\mathbf{F}}_k = \left( \sum_{j=1}^{n+1-k} \Delta_{j,k}^{1/2} \Sigma_k^{-1} \Delta_{j,k}^{1/2} \right)^{-1} \left( \sum_{j=1}^{n+1-k} \Delta_{j,k}^{1/2} \Sigma_k^{-1} \Delta_{j,k}^{1/2} \mathbf{F}_{j,k} \right)$$

Les marges de cet estimateur ne sont malheureusement pas égales aux estimateur marginaux de Mack. On notera également que Prohl et Schmidt concluent que l'optimalité des algorithmes

12. Les questions de mesurabilité des estimateurs ont été omises pour simplifier les expressions, voir [PROHL et SCHMIDT 2005] pour un exposé plus extensif.

de *Chain-ladder* marginaux dans le cadre multivarié est équivalente à la non corrélation entre les différents triangles, c'est-à-dire au fait que les  $\Sigma_k$  soient toutes des matrices diagonales.

De part leurs travaux, ils disqualifient le modèle proposé par [BRAUN 2004], et son extension multivariée, qui consistait à garder comme estimateurs marginaux l'algorithme de *Chain-ladder*. Cette technique d'estimation est possible, mais il est démontré que l'estimateur utilisé n'est plus optimal dans les hypothèses fixées.

Notons de plus que l'extension multivariée des estimateurs de Merz-Wüthrich correspondante au modèle de Schnauss existe, elle est donnée par [MERZ et WÜTHRICH 2008b].

Cependant le modèle de Schnauss a du mal à percer et beaucoup de praticiens continuent d'utiliser le modèle de Braun, ce sera aussi notre cas. La raison est que la différence entre la somme des provisions univariées et la provision multivariée est complexe à gérer dans un cadre pratique. Les lecteurs intéressés par ces estimateurs optimaux sont invités à regarder les références.

## 2.4.2 Modèle de Braun

**Le modèle de Braun**, dû à [MACK 1991] dans le cadre univarié, à [BRAUN 2004] dans le cadre bivarié, et à [PROHL et SCHMIDT 2005] dans le cadre multivarié, est un cas particulier du modèle de Schnauss qui consiste à considérer les hypothèses suivantes :

$$\mathbf{B1} : \forall i \neq j, \{\mathbf{C}_{i,1}, \dots, \mathbf{C}_{i,n}\} \perp \{\mathbf{C}_{j,1}, \dots, \mathbf{C}_{j,n}\}$$

$\mathbf{B2} \forall j \in \{1, \dots, n-1\}, \exists \mathbf{f}_j$  des coefficients de passage vectoriels, tels que :

$$\mathbb{E}(\mathbf{C}_{i,j+1}) = \Delta_{i,j} \mathbf{f}_j$$

$\mathbf{B3} \forall k \in \{1, \dots, n-1\}, \exists \Sigma_k$  matrice aléatoire semi définie positive telle que,  $\forall i, j \in \{1, \dots, n\}$ ,

$$\text{Cov}(\mathbf{C}_{i,k}, \mathbf{C}_{j,k}) = \Delta_{i,k}^{1/2} \Sigma_k \Delta_{j,k}^{1/2} \mathbb{1}_{i=j}.$$

Sous ces hypothèses, [BRAUN 2004] dérive des estimateurs pour les erreurs de prédiction correspondantes aux estimateurs marginaux de *Chain-ladder* (qui ne sont pas optimaux sous ces hypothèses, rappelons-le), en prenant en compte les corrélations (éléments non-diagonaux des matrices  $\Sigma_j$ ) entre les différents triangles.

Les estimateurs proposés par Braun de ces corrélations sont cohérents, sous les hypothèses fixées, avec les modèles de Mack marginaux, mais n'ont pas beaucoup de propriétés (par exemple on ne sait pas s'ils sont non biaisés). De plus, les estimateurs de Braun ne sont optimaux que si l'ensemble des matrices  $\Sigma_k$  sont diagonales, comme prohl-schmidt l'ont démontré.

Ces estimateurs sont donnés par des modèles de Mack univariés pour les coefficients de passage et les paramètres de variances. Braun ajoute alors les estimateurs de corrélation donnés par,  $\forall a \neq b \in \{1, \dots, p\}, \forall j \in \{1, \dots, n-2\}$ ,

$$\hat{\rho}_j^{(a,b)} = \frac{1}{n-j-1 - \left(\omega_j^{(a,b)}\right)^2} \sum_{i=1}^{n+1-j} \sqrt{C_{i,j}^{(a)} C_{i,j}^{(b)}} \left(F_{i,j}^{(a)} - f_j^{(a)}\right) \left(F_{i,j}^{(b)} - f_j^{(b)}\right) \quad (2.10)$$

avec :

$$\omega_j^{(a,b)} = \frac{\left( \sum_{i=1}^{n+1-j} \sqrt{C_{i,j}^{(a)} C_{i,j}^{(b)}} \right)^2}{\left( \sum_{i=1}^{n+1-j} C_{i,j}^{(a)} \right) \left( \sum_{i=1}^{n+1-j} C_{i,j}^{(b)} \right)}$$

Avec ce nouvel estimateur et les estimations marginales de Mack, l'ensemble des paramètres nécessaires pour calculer l'erreur de la provision totale (conçue comme la somme des provisions marginales) sont disponibles.

Bien que Braun ait développé une formule fermée pour cette erreur, à l'ultime comme à date, nous nous intéressons plutôt ici au cas du bootstrap dans ce modèle. Le lecteur intéressé pourra se référer à [BRAUN 2004] pour la formule.

Notons finalement que des estimateurs du risque à un an existent pour le modèle de Braun, ils sont notamment étudiés par [MERZ et WÜTHRICH 2007a; HUERGO, HEBERLE et MERZ 2010; APPERT-RAULLIN et al. 2013; HEBERLE, HUERGO et MERZ 2008; JOCHEN HEBERLE, LUIS HUERGO et MICHAEL MERZ 2011; HEBERLE 2014]. Des estimateurs à  $a$  années sont proposés par [LINDE 2016], nous conseillons cependant la lecture de [LUKAS J. HAHN 2017] sur ce sujet.

### 2.4.3 Bootstrap dans le modèle de Braun : Extension

Dans le modèle de Braun, l'entière structure de dépendance entre les triangles est captée par les facteurs de corrélations entre eux, un par colonne.

Ce qui signifie qu'à l'intérieur de chaque colonne, les vecteur

$$\mathbf{r}_{i,j} = \left( r_{i,j}^{(1)}, \dots, r_{i,j}^{(p)} \right)$$

des résidus des modèles de Mack marginaux sont *i.i.d* et sont donc candidats au bootstrap. Si l'on choisit la normalité des résidus, les seuls paramètres de dépendance nécessaire pour la simulation paramétrique des résidus de l'année prochaine sont les corrélations, estimées par Braun via (2.10), et le modèle est donc complet.

C'est l'approche proposée par [BOUMEZOUED et al. 2011], qui est encore aujourd'hui l'une des plus pertinente pour reproduire l'erreur à un an de Merz-Wüthrich dans le modèle de Braun. Mais on peut être plus ambitieux en gardant les résidus empiriques et un utilisant le concept de zones définies plus haut.

En effet, de manière plus générale, s'il l'on choisit de fixer la structure de dépendance interne aux vecteurs de résidus (disons par une copule quelconque) sur des **zones de rééchantillonnage** communes aux triangles, les vecteurs de résidus à l'intérieur d'une zone sont alors supposés *i.i.d*<sup>13</sup> et sont donc candidats au bootstrap. Si Braun considère que chaque colonne est une "zone" différente, l'extension que nous proposons généralise à un zonage commun quelconque.

Prenons l'ensemble de zone de rééchantillonnage défini par l'équation (2.9), et considérons les mêmes zones sur chacun des triangles. Posons comme hypothèse que les vecteurs de résidus à l'intérieur de chaque zone ont une structure de dépendance identique (et sont donc *i.i.d*).

En bootstrappant alors les vecteurs de résidus (à l'intérieur de chaque zone), on obtient des jeux de pseudo-triangles ayant la même structure de dépendance que le jeu de triangles initial.

13. En effet, les résidus marginaux sont déjà supposés *i.i.d* dans chacun des quasi-Modèle Linéaire Généralisé sous-jacents à leur calculs.

Si l'on considère que les résidus des modèles de Mack sont bien homoscédastiques, le choix des zones ne se fait donc que par les structures de dépendances entre elles.

Cette **synchronisation** des bootstrap marginaux permet, dans un cadre de Modèles Linéaires Généralisés marginaux, de conserver les structures de dépendances définies entre les triangles.

De plus, on observe que, pour obtenir des résultats à  $a$  années,  $a \in \{1, \dots, n - 1\}$ , il suffit de rééchantillonner les résidus (de manière paramétrique ou non) sur les  $a$  prochaines diagonales inférieures tout en gardant cette structure de dépendance entre eux.

Pour les mêmes raisons que celles décrites au point 2.3.4, les zones permettront de gérer une probable hétéroscédasticité des résidus dans la queue des triangles, pour des triangles complètement déroulés et même "trop grands" par rapport au risque sous-jacent.

Remarquons enfin que les modèles sous-jacents marginaux n'ont pas à être tous des modèles de Mack : pour peu qu'ils soient des quasi-Modèles Linéaires Généralisés, les résidus fournis par les modèles sont par hypothèses centrés, réduits, et *i.i.d* par zones. On pourrait donc tout à fait choisir un modèle binomial négatif donné par **ONB** pour un des triangles, un modèle de Mack donné par **M** pour un autre, etc. Tout le raisonnement proposé ici est valable à condition que les modèles soient des quasi-Modèles Linéaires Généralisés.

Enfin, le choix d'une copule paramétrique par zone est également possible : la simulation sera alors paramétrique pour la structure de dépendance. C'est notamment le choix qui est fait par [BOUMEZOUED et al. 2011] sous l'hypothèse normale.

## 2.5 Ajustements classiques

Dans cette section, nous discuterons quelques ajustements classiques faits aux modèles précédents, après avoir posé la problématique principale les justifiant. Notons qu'il en existe bien d'autres possibles que nous n'évoquerons pas ici.

L'estimateur fourni par la méthode de *Chain-ladder* est connu pour être très volatile sur les derniers rattachements, car l'information disponible sur ces années est trop incertaine, et très sensible aux derniers développements, car le nombre de points disponibles est également trop faible. Ces deux zones de sensibilité de l'estimateur sont représentées à la figure 2.5.1.

Pour la sensibilité aux derniers développements, les méthodes usuelles consistent à stabiliser les coefficients de passage, soit en les fixant directement à 1 à partir d'un certain temps, soit en les stabilisant via des méthodes dites de *curve fitting* (c'est à dire des ajustements paramétriques sur la cadence), par exemple en utilisant [BENJAMIN et EAGLES 1997]

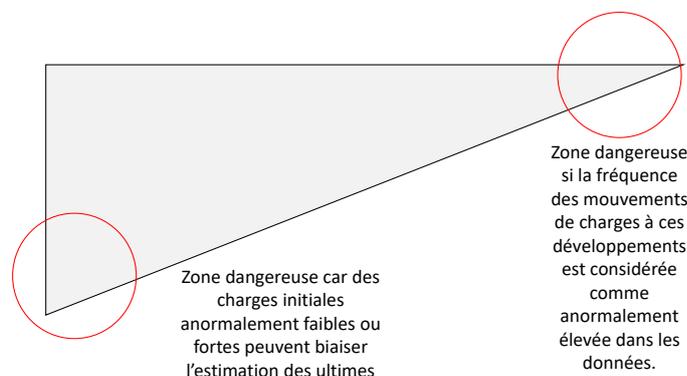


FIGURE 2.5.1 – Zones sensibles pour une estimation via *Chain-ladder*

Pour la sensibilité aux premières charges observées des dernières années de rattachement, l'utilisation d'un *a priori* bayésien permet bien souvent de contourner le problème.

Nous décrirons ici brièvement deux méthodes s'attachant à résoudre ces problèmes : Bornhuetter-Fergusson pour la volatilité des derniers rattachements, et la stabilisation des coefficients pour la volatilité des derniers développements.

### 2.5.1 Bornhuetter-Fergusson : intégration d'un *a-priori* bayésien

La méthode de Bornhuetter-Fergusson, ayant pour origine [BORNHUETTER et FERGUSSON 1972] et décrite entre autre par [MAXIME LENFANT 2010], est un principe de crédibilité Bulhmanien consistant à calculer l'ultime, pour un certain rattachement noté  $i$ , comme un mélange entre :

- un ultime *a priori* par rattachement, noté  $\hat{U}_i^{a.p}$ , souvent venu d'une méthode de loss-ratio moyens pondérés<sup>14</sup>.
- l'ultime donné par la méthode de *Chain-ladder*, noté  $\hat{C}_{i,n}$  comme précédemment.

Le paramètre du mélange est en général pris comme étant la cadence de développement (cumulée) correspondant à l'année, soit, selon (2.7),  $\sum_{j=1}^i y_j$ . On pose alors comme ultime donné par la méthode de Bornhuetter-Fergusson :

$$\hat{C}_{i,n}^{BF} = \left( \sum_{j=1}^i y_j \right) \hat{C}_{i,n} + \left( 1 - \left( \sum_{j=1}^i y_j \right) \right) \hat{U}_i^{a.p} \quad (2.11)$$

Pour la construction de l'ultime *a priori*, la méthode des loss-ratio moyens pondérés propose :

- de construire des loss-ratio *a priori* entre les sinistre ultimes de *Chain-ladder* et des primes ultimes
- de calculer un loss-ratio moyen pondéré avec des poids  $\mathbf{p}$  à déterminer
- d'obtenir les  $\hat{U}_i^{a.p}$  comme le produit du loss-ratio moyen pondéré et de la prime ultime correspondant à l'année de rattachement  $i$

Autrement dit, le couple Bornhuetter-Fergusson et loss ratio moyens pondéré demande comme données exogènes au triangle des primes ultimes et une pondération entre les loss-ratio permettant de définir le loss-ratio moyen pondéré.

### 2.5.2 Stabilisation des coefficients de passage et curve-fitting

Supposons qu'un coefficient de passage individuel des derniers développements soit seul à différer de 1 de tous les coefficients de passage individuels l'entourant. Typiquement, la queue du triangle incrémental problématique ressemblerait à :

14. La méthode des loss-ratio pondérés consiste à utiliser pour les années problématiques l'information des primes ultimes, que l'on multiplie par la moyenne pondérée (par des poids à déterminer) des loss-ratio des années précédentes.

$X_{i,j}$	...	...	...	...	$n-1$	$n$
1	...	0	0	0	10	0
2	...	0	0	0	0	
3	...	0	0	0		
4	...	0	0			
5	...	0				
...	...					

Comment penser dans ces conditions que la charge de 10 se reproduira avec probabilité  $1/2$  les années suivantes ? C'est pourtant ce que préconise l'algorithme de *Chain-ladder* et ce que suppose les modèle stochastiques l'englobants. Ainsi, de nombreuses méthodes existent pour décrédibiliser cette valeur et ne pas prendre en compte une probabilité d'apparition de  $1/2$ .

Si des modèles de crédibilité *horizontale* sur les coefficients de passage du triangle, par analogie avec la méthode de crédibilité *verticale* qu'est Bornhuetter-Fergusson, sont tout à fait envisageables (en utilisant par exemple des coefficients de passage de marché, fournis par le régulateur), on leurs préférera souvent des ajustements paramétriques de la fin de la cadence.

Par exemple, un ajustement paramétrique tel que celui proposé à la figure 2.5.2 peut être utilisé. Plusieurs types d'ajustements paramétriques existent, qui diffèrent majoritairement par la forme de la fonctionnelle proposée<sup>15</sup>.

L'ajustement paramétrique le plus trivial est de stabiliser la cadence à 1 à partir d'un certain nombre d'années fixé à l'avance, ce qui donne donc un modèle biaisé et une variance plus faible. C'est cette modification de la méthode *Chain-ladder* que nous appellerons ici **stabilisation**.

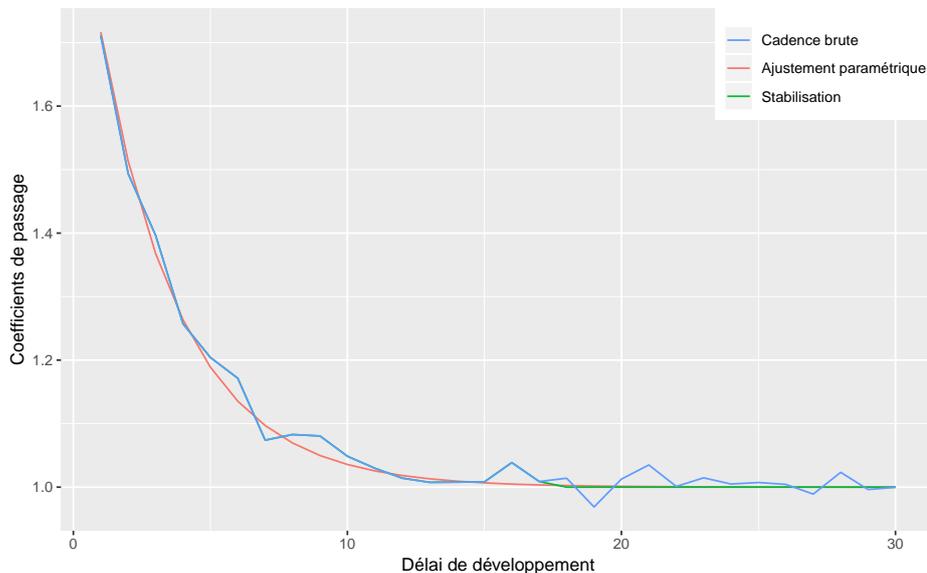


FIGURE 2.5.2 – Cadences de développements brutes et ajustées

### 2.5.3 Vision stochastique de ces ajustements

La méthode Bornhuetter-Fergusson possède plusieurs développements stochastiques divergeant. Certains, notamment [MACK 2008] mais aussi [ALAI, MERZ et WÜTHRICH 2009; ALAI, MERZ et WÜTHRICH 2011; SCARTH 2015; SCHMIDT 2008], en ajoutant un principe de crédibilité bullmanien au modèle **ODP**, proposent des estimations bayésiennes des ultimes, et

<sup>15</sup>. Des ajustement exponentiels, weibull, etc. sont classiquement utilisés

notamment des erreurs quadratiques de prédiction à l'ultime. Cette vision poissonnienne du problème permet de construire des estimateurs du risque à un an dans ce cadre, développés par [ALAI, MERZ et WÜTHRICH 2011]. On peut également noter le travail de [SALUZ 2014]

Une autre vision, développée notamment par [LINDA et KUBANOVA 2012], considérant un modèle de Mack pour la partie *Chain-ladder* de l'estimateur et un modèle stochastique quelconque, indépendant, pour les primes, permet de dégager une estimation de la volatilité sous des hypothèses ressemblants plus à  $\mathbf{M}$ . Cette méthode de construction du modèle sera celle que nous utiliserons dans ce mémoire. Nous référons donc à [LINDA et KUBANOVA 2012] pour l'optimalité des estimateurs.

Quant à la stabilisation des coefficients, si l'on considère de même une vision bayésienne du problème, les techniques développées pour Bornhuetter-Fergusson sont également valables. Dans un cadre de *curve-fitting* paramétrique, les paramètres estimés pour l'ajustement font également partis des paramètres du modèle. Ainsi la construction d'estimateurs par formules fermées des variances à l'ultime ou à  $a$  années se complique.

Du fait de la complexification du modèle qu'implique ces différents ajustements, la construction d'estimateurs des erreurs à  $a$  années est plus complexe. Cependant, les principes de bootstrap développés dans les deux sections précédentes permettent bien souvent de contourner le problème.

#### 2.5.4 Dans un cadre Bootstrap

Si l'on souhaite inclure des ajustements tels qu'un Bornhuetter-Fergusson ou une stabilisation des coefficients de passage dans une modélisation de type bootstrap, plusieurs possibilités s'offrent à nous.

Si l'ajustement est conçu ou peut se concevoir comme un raisonnement de crédibilité bayésienne, basé sur des données externes au triangle supposé indépendantes de celui-ci, alors la méthodologie générale consiste à :

- bootstrapper le quasi-Modèle Linéaire Généralisé choisi pour le triangle, pour reconstruire des pseudo-triangles comme précédemment
- bootstraper indépendamment les données externes (primes ultimes dans le cadre d'un Bornhuetter-Fergusson), un même nombre de fois.
- calculer l'estimateur sur des paires d'échantillons.

On comprend bien que si les données externes ne sont pas indépendantes du triangle, par exemple un triangle de primes correspondant au triangle de charges, les bootstraps doivent être synchronisés pour conserver une structure de dépendance.

Pour les ajustements ne pouvant pas se concevoir sous l'angle bayésien, tels que les *curve-fitting*, le problème est plus complexe. Pour la stabilisation des coefficients cependant, il existe une astuce dans un modèle type  $\mathbf{M}$  (conditionnel d'une colonne sur l'autre) consistant à :

- ajuster le modèle uniquement sur les colonnes correspondantes aux coefficients de passage ne faisant pas l'objet de la stabilisation<sup>16</sup>
- utiliser les résidus de ce modèle pour le bootstrap, sans rééchantillonner sur la queue stabilisée

Si on suppose un modèle de *curve-fitting* n'accordant aucune confiance aux dernières données du triangle, en prenant par exemple une cadence de marché pour les derniers développements,

16. Par exemple sur un triangle ou  $n = 20$ , l'on ne prendrais les 15 premières colonnes

ou en stabilisant à 1 inconditionnellement, l'estimateur de ces développements est alors indépendant du reste du triangle.

Pour une stabilisation à 1 des facteurs de développements, on peut le considérer déterministe (de variance nulle), et le bootstrap de cette partie du triangle n'est plus nécessaire quelle que soit l'année  $a$  voulue (à condition de ne pas utiliser les résidus de la partie stabilisée pour le bootstrap). Pour une cadence de marché, stochastique, un bootstrap approprié devrait être mis en place.

# Chapitre 3 Assurance construction : un problème tri-dimensionnel

Après avoir donné dans le chapitre précédent des rappels quant au provisionnement classique, nous nous attacherons ici à traiter le cas de l'assurance construction. Ce cas bien particulier, par le rattachement des sinistres à leurs dates réglementaires d'ouverture de chantier respectives et non pas à leur date de survenance, demande une provision supplémentaire, la provision pour sinistres non-encore manifestés. Elle vient compléter la provision pour sinistres à payer pour couvrir les sinistres non-encore survenus. Si son calcul est complexe, il est souvent approximé par des hypothèses simplificatrices.

Dans un cadre *comptes sociaux*, les techniques de calcul de la provision pour sinistres non-encore manifestés sont fixées par la réglementation et s'appuient sur des coefficients réglementaires appliqués aux vecteurs de primes et de sinistres. Dans un cadre *comptes économiques*, ici Solvabilité II, la modélisation est plus complexe.

Sur la base de la discussion faite sur la méthode de *Chain-ladder* ainsi que sur les modèles stochastiques l'englobant développés dans le chapitre précédent, nous nous attacherons ici à construire des modèles de provisionnement pour ce nouveau type de données, après avoir défini statistiquement la situation.

## 3.1 Les trois dimensions de l'assurance construction

En assurance construction, nous disposons d'observations de charges, règlements, nombres, ou autres informations incrémentales  $X$  indexées selon trois dimensions :

- l'année de la date réglementaire d'ouverture de chantier, notée  $i \in \{1, \dots, I\}$
- le délai de survenance du sinistre, noté  $j \in \{1, \dots, J\}$
- le délai de développement, noté  $k \in \{1, \dots, K\}$

Ainsi,  $X_{i,j,k}$  représente la charge (incrémentale) enregistrée l'année d'inventaire  $i + j + k - 2$  pour un sinistre survenu l'année  $i + j - 1$  pour un chantier ayant ouvert l'année  $i$ . Pour la simplicité d'exposition on pose  $n = I = J = K$ .

Ces données se présentent traditionnellement sous forme d'une pyramide, que l'on peut par exemple observer à la figure 3.1.

On remarque immédiatement que le cumul des sinistres pour la construction d'une "pyramide" cumulée n'est pas trivial : en effet, on pourrait cumuler les sinistres par délai de développement ou par délai de survenance.

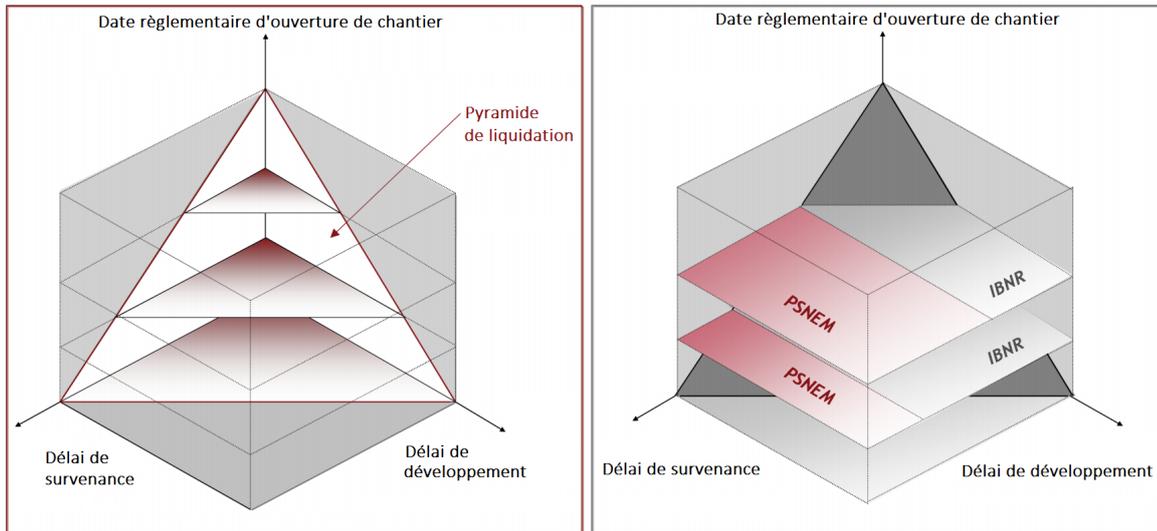


FIGURE 3.1.1 – Représentation de la pyramide de données disponibles en assurance construction. *En bas à droite, les projections nécessaires pour compléter le cube et calculer les deux provisions*[DEPLANTE 2010].

Pour éviter toute méprise, nous noterons :

$$C_{i,j,k} = \sum_{p=1}^k X_{i,j,p} \text{ et } D_{i,j,k} = \sum_{p=1}^j X_{i,p,k}, \quad (3.1)$$

représentant respectivement les cumuls par délai de développements et par délai de survénance.

Nous introduisons de plus des notations pour les sommes marginales de la pyramide :

$$\begin{aligned} X_{\bullet,j,k} &= \sum_{i=1}^{n+2-j-k} X_{i,j,k}, & X_{i,\bullet,\bullet} &= \sum_{j=1}^{n+1-i} \sum_{k=1}^{n+2-i-j} X_{i,j,k}, \\ X_{i,\bullet,k} &= \sum_{j=1}^{n+2-i-k} X_{i,j,k}, & X_{\bullet,j,\bullet} &= \sum_{i=1}^{n+1-j} \sum_{k=1}^{n+2-i-j} X_{i,j,k}, \\ X_{i,j,\bullet} &= \sum_{k=1}^{n+2-i-j} X_{i,j,k}, & X_{\bullet,\bullet,k} &= \sum_{i=1}^{n+1-k} \sum_{j=1}^{n+2-i-k} X_{i,j,k}, \end{aligned} \quad (3.2)$$

Ces sommes marginales sont l'analogie des sommes marginales en provisionnement non-vie dans les triangles classiques vues au chapitre précédent, notamment étudiées par [MACK 1991] et [SCHMIDT et WÜNSCHE 1998] parmi d'autres.

Remarquons que l'ensemble des sommes marginales par date réglementaire d'ouverture de chantier,  $X_{\bullet,j,k}$ , représente un triangle en survénance-développement classique, mais où les sinistres ne sont plus indexés par leur année de survénance calendaire mais par leur *délai de survénance*, ce qui est très différent.

Enfin, on pose les triangles en deux dimensions agrégés suivants :

$$\begin{aligned}
 Y_{\bullet,j,k} &= \sum_{i=1}^{n+2-j-k} X_{i,j+1-i,k}, \\
 Y_{i,\bullet,k} &= \sum_{j=1}^{n+2-i-k} X_{i,j,k+1-j}, \\
 Y_{i,j,\bullet} &= \sum_{k=1}^{n+2-i-j} X_{i,j,k},
 \end{aligned} \tag{3.3}$$

représentants respectivement les triangles agrégés en survenance par délai de développement, en date règlementaire d'ouverture de chantier par délai de survenance plus délai de développement, et en date règlementaire d'ouverture de chantier par délai de survenance.

Le but des différents modèles que nous proposerons ici sera d'estimer la partie inférieure de la pyramide, c'est-à-dire les charges futures, pour en déduire les ultimes, les réserves, et toutes autres statistiques intéressantes. Nous prendrons soin de bien séparer les méthodes de complétion de la pyramide et les modèles stochastiques sous-jacents à ces méthodes.

### 3.1.1 Méthode règlementaire

La question tri-dimensionnelle de l'assurance construction étant une spécificité française, les provisions spécifiques à l'assurance construction, les PSNEM, sont définies aux articles 143-13 et 143-14 de [AUTORITÉ-DES-NORMES-COMPTABLES 2015] et de [AUTORITÉ-DES-NORMES-COMPTABLES 2016]. Selon l'Autorité des normes comptables, les provisions afférentes aux garanties décennales ne peuvent, pour chaque année d'ouverture de chantier, être inférieures à la somme des éléments suivants :

- le coût total des sinistres qui se sont manifestés jusqu'à la date de l'inventaire, constitué des règlements, des provisions dossier/dossier (sinistres déclarés) et d'une estimation du coût des sinistres non-encore déclarés.
- une estimation du coût des sinistres non-encore manifestés et qui devraient se manifester d'ici à l'expiration de la période de prescription décennale.

Selon l'article 143-14, l'assureur doit calculer pour chaque exercice d'ouverture de chantier les quantités  $A_n$  et  $B_n$  suivantes (où  $n$  représente l'âge de la date règlementaire d'ouverture de chantier :  $n = 1$  représente alors la dernière année disponible.)

- $A_n$  représente le coût total estimé dossier / dossier des sinistres déjà survenus rattachés à la date règlementaire d'ouverture de chantier d'âge  $n$
- $B_n$  représente le montant des primes (émises et restantes à émettre), nettes d'annulation et de frais d'acquisition, correspondantes aux garanties de la date règlementaire d'ouverture de chantier d'âge  $n$ .

Enfin, l'estimation des sinistres non-encore manifestés est égale à  $\max(a_n A_n, b_n B_n)$ , pour des coefficients  $a_n$  et  $b_n$  qui sont donnés dans le texte, repris à la table 3.1.1.

TABLE 3.1.1 – Coefficients règlementaires pour les calcul des PSNEM en compte sociaux

$n$	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$a_n$	0	0	3,4	2	1,4	1	0,7	0,5	0,35	0,25	0,20	0,15	0,10	0,05
$b_n$	1	1	0,95	0,85	0,75	0,65	0,55	0,45	0,35	0,25	0,20	0,15	0,10	0,05

Cette méthode règlementaire permet donc de calculer les provisions afférentes à l'assurance construction en deux étapes :

- les provisions relatives aux sinistres déjà survenus, la provision pour sinistres à payer, d'une part,
- les provisions relatives aux sinistres non encore survenus, la provision pour sinistres non-encore manifestés, d'autre part.

Le principal souci de cette méthode est qu'elle ne donne qu'une estimation des provisions et ne complète en rien la partie inférieure de la pyramide. De plus, aucun modèle stochastique ne l'englobant n'est proposé.

### 3.1.2 Modélisation en deux dimensions en un temps

Si on considère le triangle agrégé par date d'ouverture de chantier et délai de règlement plus délai de manifestation donné en (3.3) on se rend compte qu'une projection de ce triangle par une méthode d'ordre 1 minimum donne des ultimes, des provisions, des cadences de règlements, ou tout autre statistiques intéressantes, correspondant bien au besoin de provisionnement global d'une branche en capitalisation.

Cette méthode consiste plus précisément à appliquer *Chain-ladder* au triangle agrégé  $Y_{i,\bullet,k}$ . La provision trouvée contient alors à la fois la provision pour sinistres à payer mais également la provision pour sinistres non-encore manifestés. Elle correspond à l'ensemble des paiements futurs pour les chantiers couverts.

Cette méthode a un avantage principal : elle permet, via l'extension de *Chain-ladder* par un modèle d'ordre 2 ou plus au choix, de construire les moments d'ordre 2 ou plus des estimateurs fournis. Il suffit de reprendre ce qui a été fait dans la partie 2<sup>1</sup>.

Cependant son principal défaut est le fait qu'elle ne donne pas la provision pour sinistres à payer et la provision pour sinistres non-encore manifestés séparément l'une de l'autre. D'autre part une vision prenant pour origine la date règlementaire d'ouverture de chantier et pour développement la somme du délai de développement et du délai de survenance est "grossière" et mélange dans une même cadence des phénomènes qui sont pourtant très différents : la manifestation des sinistres et leurs développements. Cette solution est donc gênante économiquement.

### 3.1.3 Modélisation en deux dimensions et en deux temps

Dans un cadre de calcul de bilan économique, les provisions en assurance construction se calculent généralement en utilisant les deux autres triangles agrégés de la manière suivante :

1. tout d'abord, le triangle  $Y_{\bullet,j,k}$  est utilisé pour calculer la provision pour sinistres à payer, la provision afférente aux sinistres déjà manifestés dans le portefeuille.
2. puis en utilisant la cadence de développement donnée par (2.7) sur ce triangle, les données incrémentales du triangle  $Y_{i,j,\bullet}$  sont *passées à l'ultime*.
3. enfin, ce triangle vu à l'ultime est projeté sur les survenances futures pour obtenir la provision pour sinistres non-encore manifestés.

---

1. Notamment, les estimateurs de variabilité à un an de Merz-Wüthrich classiques sont disponibles dans ce cadre

Plus formellement, une fois le triangle  $Y_{\bullet,j,k}$  complété selon un modèle d'ordre 1 au choix (par exemple, *Chain-ladder*), on récupère sa cadence de développement  $y$  via (2.7). Rappelons que cette cadence correspond au pourcentage de développement des sinistres chaque année. Ainsi, en posant  $\forall i, j \in \{1, \dots, n\}$  :

$$Z_{i,j} = Y_{i,j,\bullet} \left( \sum_{p=1}^{i+j-1} y_p \right), \quad (3.4)$$

$Z$  représente bien le triangle en date règlementaire d'ouverture de chantier par délai de survenance ou les montant ont été ramenés à l'ultime, c'est-à-dire développés jusqu'au bout.

Enfin, en projetant  $Z$  par une technique (d'ordre 1 minimum) au choix, on obtient des provisions qui correspondent à la charge ultime des sinistres futurs dans chaque date règlementaire d'ouverture de chantier, c'est-à-dire la provision pour sinistres non-encore manifestés.

Cette technique calcule donc la provision pour sinistres non-encore manifestés conditionnellement à la provision pour sinistres à payer.

Notons qu'aucune formule fermée sous des hypothèses par exemple de Mack ne permet de calculer ici la variance de la provision, et encore moins la variance à un an des provisions. Cependant, une vision bootstrap permet de construire un modèle (un cadre théorique) pour cette méthode. De plus, ce modèle permet d'aller beaucoup plus loin en terme de variance des provisions à l'ultime comme à un an, cette approche sera développée dans la suite de ce chapitre.

Après avoir donc pris le temps d'étudier les trois méthodes classiques d'estimation de provisions dans le cadre de la construction, nous étudierons dans la section suivante la problématique de provisionnement directement dans le cadre tri-dimensionnel via l'approche par Modèles Linéaires Généralisés.

## 3.2 Extension des modèles existants

L'algorithme de *Chain-ladder* classique repose sur la définition de coefficients de passage que l'on construit d'ordinaire sur les triangles cumulés. L'extension à une troisième dimension de ce principe n'est pas triviale, et nous proposons donc de changer notre manière de voir les choses. Nous commencerons par étudier une analogie poissonnienne du problème, puis nous parlerons de l'impossibilité d'une vision récursive et enfin nous rechercherons une estimation bootstrap des ultimes basée sur la vision poissonnienne.

### 3.2.1 Vision poissonnienne

Rappelons que [MACK 1991] a montré que les estimateurs donnés par l'algorithme de *Chain-ladder* sont équivalents à la spécification par sommes marginales suivante :

$$\sum_{i=1}^{n+1-j} C_{i,j} = \sum_{i=1}^{n+1-j} x_i y_j \quad \forall j \in \{1, \dots, n\}$$

$$\sum_{j=1}^{n+1-i} C_{i,j} = \sum_{i=1}^{n+1-j} x_i y_j \quad \forall i \in \{1, \dots, n\}$$

correspondant aux équations de quasi-vraisemblance du modèle de Poisson vu au chapitre précédent<sup>2</sup>. Ces équations donnent des estimateurs pour les paramètres de lignes  $x$  et les paramètres de colonnes  $y$  qui sont équivalents à *Chain-ladder*, via la formule de passage (bijective) (2.7). Le modèle quasi-Poisson correspondant se définit à l'ordre 1 par :

$$\mathbb{E}(X_{i,j}) = e^{\alpha_i + \beta_j} = x_i y_j$$

pour des paramètres  $x_i = e^{\alpha_i}$  et  $y_j = e^{\beta_j}$  correspondant chacun à un effet ligne et un effet colonne.

Ainsi, nous proposons naturellement l'**extension poissonnienne** suivante :

**EP1** Les incréments  $X_{i,j,k}$  sont indépendants

**EP2**  $\exists x, y, z \in \mathbb{R}^n \mid \forall i, j, k \in \{1, \dots, n\}, \mathbb{E}(X_{i,j,k}) = x_i y_j z_k = e^{\alpha_i + \beta_j + \gamma_k}$

**EP3**  $\sum_{j=1}^n y_j = 1$

**EP4**  $\sum_{k=1}^n z_k = 1$

pour des paramètres  $x = e^{\alpha}$ ,  $y = e^{\beta}$  et  $z = e^{\gamma}$  correspondant respectivement aux *effets dates règlementaires d'ouverture de chantier*, *effets délai de manifestation* et *effets délai de développement*.

L'ajout de **EP3** et **EP4** permet une identification du modèle ainsi qu'une interprétation naturelle des paramètres :

- $x$  donne les ultimes par date règlementaire d'ouverture de chantier,
- $y$  donne les cadences de manifestations,
- $z$  donne les cadences de développement.

Les équations de quasi-vraisemblance de ce modèle s'écrivent alors comme les équations de sommes marginales suivantes (en termes cumulés) :

$$\begin{aligned} \forall i \in \{1, \dots, n\}, \mathcal{Q}_i^1 : X_{i, \bullet, \bullet} &= \sum_{j=1}^{n+1-i} \sum_{k=1}^{n+2-i-j} x_i y_j z_k, \\ \forall j \in \{1, \dots, n\}, \mathcal{Q}_j^2 : X_{\bullet, j, \bullet} &= \sum_{i=1}^{n+1-j} \sum_{k=1}^{n+2-i-j} x_i y_j z_k, \\ \forall k \in \{1, \dots, n\}, \mathcal{Q}_k^3 : X_{\bullet, \bullet, k} &= \sum_{i=1}^{n+1-k} \sum_{j=1}^{n+2-i-k} x_i y_j z_k, \\ \sum_{j=1}^n y_j &= 1 \text{ et } \sum_{k=1}^n z_k = 1, \end{aligned}$$

Notons que ces équations correspondent à un quasi-Modèle Linéaire Généralisé log-poissonien. Son principal défaut est que l'hypothèse de variance sous-jacente au modèle implique des contraintes de positivité sur les sommes marginales, qui peuvent être considérées comme trop fortes<sup>3</sup>. L'utilisation du *log* comme fonction de lien pose le même problème.

2. Mais aussi des autres quasi-Modèles Linéaires Généralisés reproduisant *Chain-ladder*, moyennant des reparamétrages

3. La plupart des données assurentielles française en garanties décennales comportent des recours (notamment les triangles de charges), souvent sur les développements les plus tardifs, ce qui contribue fortement à la création d'incrément négatifs voir même de sommes marginales négatives sur ces queues de triangles.

Cependant, cette extension est *cohérente* avec la méthode de *Chain-ladder* et le modèle **ODP** dans le sens où les estimateurs ne changent pas si l'on retire une dimension aux sommes marginales : aplatir la pyramide donnera la méthode de *Chain-ladder* et comme modèle l'englobant le quasi-Modèle Linéaire Généralisé poissonien. De plus, de manière analogue au quasi-Modèle Linéaire Généralisé poissonien, il y a restriction sur les sommes marginales.

Pour résoudre ce problème, notre première idée fût de passer en vision *réursive* ce modèle, afin d'obtenir le même type d'analogie que celle du chapitre précédent avec un modèle normal sur les coefficients de passage individuels. En effet, notre but ici est la construction d'un modèle d'ordre 2 pour pouvoir déduire des statistiques de variance sur les provisions, notamment à un an dans un cadre Solvabilité II. Cependant, nous verrons rapidement qu'un tel passage en récursif du problème n'est pas possible.

### 3.2.2 Recherche d'un ordre 2

L'hypothèse de variance canonique de la spécification poissonienne posée *supra* est de la forme :

$$\mathbb{V}(X_{i,j,k}) = \phi \mathbb{E}(X_{i,j,k})$$

Via une modélisation jointe, il est possible de différencier le paramètre de dispersion  $\phi$  sur les covariables, c'est-à-dire de construire une hypothèse de variance se rapprochant légèrement plus de celle de Mack. Par exemple, on pourrait utiliser :

$$\mathbb{V}(X_{i,j,k}) = \phi_{i,j,k} \mathbb{E}(X_{i,j,k}) \text{ et } \mathbb{E}(r_{i,j,k}^2) = \phi_{j,k} \text{ ou } \phi_k,$$

c'est-à-dire une modélisation du facteur de dispersion sur les délais de développement ou sur les délais de survenance ou les deux, ou encore n'importe quel regroupement de délais de survenance et de délais de développement.

Cette vision donnée par l'extension quasi-Poisson de l'algorithme de *Chain-ladder* implique des contraintes de positivité très fortes sur les sommes marginales du modèle si l'on souhaite obtenir un cadre stochastique à l'algorithme suffisamment complet pour fournir une estimation d'ordre 2 des différentes statistiques d'intérêt.

Les données avec lesquelles nous travaillerons dans le prochain chapitre ne satisfont pas ces contraintes de positivité, et ce modèle n'est donc pas applicable dans notre étude de cas. Cependant, un modèle poissonien avec modélisation jointe de la variance serait à notre avis le modèle étendant *Chain-ladder* au cadre de la construction le plus proche de ce que peut être un *Chain-ladder* classique.

D'autre part, avec des données négatives, certaines valeurs des paramètres  $x, y, z$  du modèle précédent seront négatives. Ainsi, la spécification de variance poissonienne n'est plus envisageable. Si on choisit une **Extension normale** du modèle, le modèle d'ordre 2 correspondant serait alors :

**EN1** Les incréments  $X_{i,j,k}$  sont indépendants

**EN2**  $\exists x, y, z \in \mathbb{R}^n \mid \forall i, j, k \in \{1, \dots, n\}, \mathbb{E}(X_{i,j,k}) = x_i y_j z_k$

**EN3**  $\exists \phi_{i,j,k} \mid \forall i, j, k \in \{1, \dots, n\}, \mathbb{V}(X_{i,j,k}) = \phi_{i,j,k}$ .

Une hypothèse de modélisation jointe pour  $\phi$  est encore une fois possible. Cependant contrairement à l'approche récursive, les paramètres de dispersion correspondent ici à la variance

des incréments et non plus à la variance des coefficients de passage, et ne sont donc plus homogènes par date réglementaire d'ouverture de chantier ou par survenance.

Une autre possibilité est de *forcer* le modèle poissonien via la méthodologie suivante :

- remplacer le jeu de données incrémentales par sa valeur absolue, tout en conservant la probabilité (pondéré par les valeurs absolues) qu'une case soit négative par délai de développement,
- construire un modèle poissonien sur les incréments ainsi positifs,
- calculer par bootstrap les statistiques dont on a besoin en tirant une binomiale pour le signe des incréments futurs.

La suite de ce chapitre s'attachera alors à chercher des modèles d'ordre 2 moins contraints sur la positivité des incréments.

### 3.2.3 Impossibilité des modèles récurrents

Pour abolir cette contrainte de positivité, dans le chapitre précédent, la méthode utilisée fut de passer le modèle en *récurrent*, suivant la méthodologie décrite dans [ENGLAND et VERRALL 2002]. Le modèle résultant, par l'introduction de poids *a priori*, s'attache alors à modéliser directement les facteurs de développement de *Chain-ladder*.

Or, nous avons ici dans notre pyramide deux "sens de développement". Choisir des poids *a priori*, ce serait choisir l'un des deux sens comme étant prépondérant sur l'autre. La nature du phénomène modélisé nous indique que les sinistres se manifestent avant de se développer, mais ne nous indique pas si l'information du délai de survenance est prépondérante par rapport à celle de l'âge du sinistre pour la détermination de la charge.

Ainsi, la normalisation de la variable réponse par des poids ne semble pas possible dans nos contraintes tri-dimensionnelles. On pourrait cependant normaliser l'ensemble des réponses par la taille de la date réglementaire d'ouverture de chantier donnée par un estimateur de primes, et travailler alors sur les loss-ratios incrémentaux par date réglementaire d'ouverture de chantier. Cependant cette information de primes étant elle aussi stochastique nous ne ferions que repousser le problème.

Si on considère que le passage d'un développement à l'autre doit se faire avant le passage d'une survenance à l'autre, on arrive *stricto-sensu* au modèle en deux temps proposé à la section précédente<sup>4</sup>. A l'inverse, choisir de forcer le développement de la pyramide sur l'axe des survenances avant l'axe des développements reviendrait à projeter en premier lieu des données non-homogènes.

Nous concluons donc à l'existence d'un modèle adapté au provisionnement en trois dimensions et cohérent avec les modèles classiques reprenant la méthode *Chain-ladder* : un modèle poissonien. Cette extension contraint les charges incrémentales à être positives, et n'est de ce fait pas applicable sur notre jeu de données<sup>5</sup>. Cependant une application a été faite par [DEPLANTE 2010] en supprimant tout simplement ces charges négatives gênantes. Des extensions de l'algorithme chain-ladder original sont également possible dans le cadre de données négatives, voir par exemple [ENRIQUE DE ALBA 2002; ENRIQUE DE ALBA et MARCO A. RAMÍREZ CORZO 2005], ou encore [XIA et SCOLLNIK 2015].

Nous avons donc à notre disposition plusieurs estimateurs d'ordre 1 et quelques estimateurs d'ordres 2 pour les provisions spécifiques de l'assurance construction, certains donnant la provision pour sinistres non-encore manifestés, d'autre la provision pour sinistres à payer et d'autres les deux indistinctement. Nous allons prendre le temps de construire des estimateurs

4. Cela se montre aisément via les spécifications par sommes marginales

5. plus de 10% du volume de charge de nos données sont problématiques

de la variance de ces provisions et des *Claim development result* correspondants, c'est-à-dire leur erreurs à un an, en passant par des modèle de bootstrap.

### 3.3 Distributions complète par bootstrap

Afin d'obtenir des distributions complètes des estimateurs que nous recherchons, que ce soit à l'ultime ou à un an, la méthode la plus efficace<sup>6</sup> sur un plan théorique est le bootstrap. Nous présentons donc dans cette section plusieurs bootstrap adaptés au cas du provisionnement en assurance construction : tout d'abord nous observerons le cas poissonien déjà développé dans la section précédente, puis nous étudierons le cas de la modélisation en deux temps développée au début du chapitre.

#### 3.3.1 Cas poissonien

Dans le cas poissonien, sous **EP**, le fondement théorique du modèle est un Modèle Linéaire Généralisé, ou quasi Modèle Linéaire Généralisé, avec une hypothèse de dispersion non-constante et donc une éventuelle modélisation jointe<sup>7</sup>.

Dans ce cadre, l'extension bootstrap la plus appropriée au cadre du provisionnement non-vie est le bootstrap des résidus de Pearson (donnés par la formule (2.6)) de ce Modèle Linéaire Généralisé.

Lors de la modélisation jointe, le paramètre de dispersion estimé sur les covariables se doit d'être normalisé, comme nous l'avons déjà vu dans le cas normal du Modèle Linéaire Généralisé de Mack, par le nombre de paramètres déjà utilisés pour décrire les données auxquelles il se rattache.

Cependant, un ajustement supplémentaire est nécessaire pour le kurtosis du modèle, (qui n'avait aucun effet dans le cas normal, alors que le kurotis normalisé est  $\rho_4 = \lambda^{-1}$  pour une loi de Poisson). Cet ajustement a été décrit dans le chapitre précédent (se reporter à [MCCULLAGH et NELDER 1989, chapitre 10] pour le détail).

La principale différence avec le Modèle Linéaire Généralisé construit dans la partie précédente est que ce modèle n'est pas pondéré, et qu'il a un jeu de covariables supplémentaires.

Nos données ne correspondant pas aux hypothèses de ce modèle, nous ne pouvons malheureusement pas le tester. Nous nous rabattons alors vers une modélisation en deux temps pour la construction de notre estimateur.

#### 3.3.2 Cas de la modélisation en deux temps

La modélisation en deux dimensions et en deux temps de la partie précédente correspond à :

- une méthode *Chain-ladder* inconditionnelle sur le triangle en survenance/développement,  $Y_{\bullet,j,k}$ ,
- une méthode *Chain-ladder* conditionnelle à la précédente sur le triangle en date règlementaire d'ouverture de chantier par survenance vue à l'ultime,  $Z_{i,j}$ ,

6. "efficace" au sens où ces méthodes fournissent d'important résultats pour peu d'hypothèses, en contrepartie il est vrai d'un temps de calcul important.

7. Se reporter au chapitre 2 pour plus de détails sur ces modèles.

Si nous choisissons pour la première méthode un quasi-Modèle Linéaire Généralisé (par exemple Mack), nous pouvons utiliser le bootstrap adapté à ce modèle pour obtenir des états du monde où les statistiques conditionnantes le deuxième triangle<sup>8</sup> sont disponibles.

Ainsi, dans chacun des ces états du monde, il convient d'utiliser un second bootstrap sur le second triangle pour obtenir une modélisation jointe de la provision pour sinistres à payer et de la provision pour sinistres non-encore manifestés.

La procédure de ce bootstrap est alors la suivante :

- obtenir les résidus de Pearson  $\mathbf{r}$  du quasi-Modèle Linéaire Généralisé pondéré de Mack sur le premier triangle  $X_{\bullet,j,k}$ .
- tirer avec remise de nouveaux résidus  $\mathbf{r}^{(b)}$ , pour  $b \in \{1, \dots, B\}$ , par zone si besoin, sur une plage correspondant au triangle supérieur et aux  $a$  diagonales suivantes.
- reconstruire les triangles cumulés  $\mathbf{C}^{(b)}$  et y calculer de nouveaux coefficients de passage  $\mathbf{f}^{(b)}$ , vus au temps  $a$ .
- calculer sur chacun de ces nouveaux triangles la provision pour sinistres à payer.
- se servir de ces coefficients de passage pour calculer le triangle vu à l'ultime et au temps  $a$   $Z_{i,j}^{(b)}$  en fonction de  $Y_{i,j,\bullet}$  et de la nouvelle cadence  $\mathbf{y}^{(b)}$  via la relation (2.7) et la relation (3.4).
- calculer les résidus de Pearson du modèle de Mack sur ces nouveaux triangles, les ré-échantillonner sur le triangle supérieur et les  $a$  diagonales inférieures  $\beta$  fois, et calculer la provision pour sinistres non-encore manifestés correspondante dans chaque simulation.

Cette procédure de bootstrap imbriquée donne alors une distribution jointe de la provision pour sinistres à payer et de la provision pour sinistres non-encore manifestés, vue à  $a$  années.

Plusieurs remarques :

1. Les remarques faites en 2.4.3 quant au rééchantillonnage des résidus sur la partie inférieure du triangle de manière non-paramétrique complète ou de manière paramétrique ainsi que sur les possibilités de zonage sont également valables ici, dans les deux parties de la modélisation.
2. Cette procédure donne une variance du *Claim development result* sur la partie correspondant à la provision pour sinistres à payer qui est cohérente et convergente avec la formule (2.5) sous les hypothèses **MW**, voir [BOUMEZOUED et al. 2011].
3. Ce modèle donne pour la provision pour sinistres non-encore manifestés une distribution conditionnelle à la provision pour sinistres à payer ainsi qu'aux données disponibles, ce qui est pris en compte par le fait que les bootstrap soient *nestés* (imbriqués).
4. Le nombre de simulation utilisées dans les deux bootstraps, respectivement  $B$  et  $\beta$ , sont indépendants l'un de l'autre, et le nombre d'état du monde finaux est  $B\beta$ . Par la construction du modèle, chaque état du monde a une probabilité d'apparition  $1/B\beta$ . On peut ainsi prendre  $\beta = 1$  sans perte de signal.
5. Enfin, la synchronisation dans le modèle de Braun du bootstrap en survenance-inventaire avec les autres Groupes de risques homogènes inclus dans la *Line of Business* est tout à fait possible, et donnera donc une distribution jointe des différentes provision pour sinistres à payer de la *Line of Business* et de la provision pour sinistres non-encore manifestés, ce qui résout nos problématique de variance des provisions totales de la *Line of Business* et de risque de réserve.

---

8. C'est-à-dire la cadence de développement du premier triangle, permettant de passer le second l'ultime via (3.4)

Une application de ce bootstrap est proposée dans le prochain chapitre, au point 4.2. Un graphique de la version multivariée proposée dans la dernière remarque est également disponible à la figure 4.2.1.

### 3.3.3 Conclusion : risque de réserve en assurance construction

Dans le cadre du modèle de Poisson comme dans le cadre du modèle de bootstrap imbriqués, l'obtention du sigma de réserve de la provision pour sinistres à payer, de la provision pour sinistres non-encore manifestés, ou total, peut donc se faire par bootstrap directement dans un modèle global en trois dimensions, ou par bootstrap imbriqué dans un modèle en deux temps.

Ainsi, nous disposons donc d'estimateurs bootstrap de la distribution du *Claim development result* de la *Line of Business* dans le cadre de l'assurance construction. Mais nous disposons également d'estimateurs sous forme de distribution complète pour toutes les statistiques nécessaires et calculables à partir de nos données, vues en  $b, \forall b \in \{0, \dots, a\}$ . En effet, l'exhaustivité des procédures bootstrap est notable.

Dans le prochain chapitre, nous proposons une application des ces différents estimateurs à un portefeuille d'exemple, avant de comparer les résultats avec la formule standard et de conclure sur l'adéquation du sigma de réserve de la formule standard au cadre de l'assurance construction.



# Chapitre 4 Sigma de réserve : recherche d'un estimateur

Après avoir analysé en détails les problématiques de provisionnement intrinsèques à l'assurance construction, rappelé les bases en provisionnement non-vie et conçu les modèles classiques dans un cadre plus général, nous avons pris le temps de développer des modèles spécifiques à cette branche d'assurance qui permettraient de dégager des estimateurs du sigma de réserve de la *Line of Business* 8 pour une société d'assurance soumise à du risque décennal.

Ce chapitre, par la mise en pratique de ces modèles sur un même portefeuille d'exemple (décrit au point 1.4), s'attachera à comparer ces différents modèles et ces différents estimateurs.

Nous commencerons par traiter les deux méthodes règlementaires, proposées par les spécifications Solvabilité II, puis nous étudierons la méthode de bootstrap déjà esquissée au point 3.3.2.

## 4.1 Méthodes règlementaires

Le calcul par une entreprise d'un sigma de réserve propre en tant qu'*Undertaking specific parameter* à l'intérieur de la formule standard est normalisé par [COMMISSION-EUROPÉENNE 2014, annexe XVII, points C. et D.]. Ces méthodes de calculs, appelées ici **méthodes règlementaires**, sont décrites, commentées et appliquées dans cette section. On pourra se référer à [FABIEN BESSEYRE 2015 ; COMMISSION-EUROPÉENNE 2014] pour plus de détails sur ces méthodes.

### 4.1.1 Méthode 1 : Log-normalité

**Le Modèles théorique de la méthode 1** La première méthode donnée par le règlement délégué consiste à supposer (en plus d'une hypothèse log-normale) une relation linéaire entre les deux variables suivantes :

- le *Best estimate* de provisions établi à la fin de l'exercice pour les sinistres qui étaient à payer en début d'exercice dans le segment, majoré des paiements effectués durant l'exercice pour les sinistres qui étaient à payer en début d'exercice dans le segment, noté ici  $y_t$
- le *Best estimate* de la provision pour sinistres à payer dans le segment au début de l'exercice, noté ici  $x_t$ ,

et ce pour toutes les années  $t \in \{1, \dots, T\}$ . On notera ici :

- $\sigma$ , le coefficient de variation du risque de réserve, dit aussi *sigma de réserve*, déjà aperçu en section 1.3

- $\delta \in [0, 1]$ , le paramètre de mélange,
- $\beta$ , le *loss-ratio* attendu,
- $\gamma = \ln(\frac{\sigma}{\beta})$ .

Le **modèle log-normal** proposé par le règlement délégué consiste sous ces notations en les hypothèses suivantes :

- MLN1**  $\mathbb{E}(y_t) = \beta x_t$ , *i.e* les pertes sont proportionnelles aux primes, de coefficient de proportionnalité  $\beta$  constant.
- MLN2**  $\mathbb{V}(y_t) = \sigma^2 ((1 - \delta)\bar{x}x_t + \delta x_t^2)$ , où  $\bar{x} = \sum_{t=1}^T x_t$
- MLN3**  $y_t \sim \log\mathcal{N}(\mu_t, \omega_t^2)$ ,  $y_t$  suit une loi log-normale.

La méthode 1 consiste alors à chercher un estimateur du maximum de vraisemblance pour  $\sigma$  sous ces hypothèses. La vraisemblance de ce modèle est alors :

$$\mathcal{L}^1(y_1, \dots, y_T, \beta, \sigma) = \prod_{t=1}^T \frac{1}{y_t \omega_t \sqrt{2\pi}} \exp \left\{ -\frac{(\ln(y_t) - \mu_t)^2}{2\omega_t} \right\} \quad (4.1)$$

La méthode de calcul du paramètre  $\sigma$  se base alors sur une maximisation du logarithme de cette vraisemblance sous les hypothèses précédentes. Notons que le paramètre sigma peut être extrait de la log-vraisemblance. En effet, si l'on note :

$$\pi_t(\hat{\delta}, \hat{\gamma}) = \frac{1}{\ln \left( 1 + e^{2\hat{\gamma}} \left( (1 + \hat{\delta}) * \frac{\bar{x}}{x_t} + \hat{\delta} \right) \right)}$$

et que l'on estime  $\hat{\delta}, \hat{\gamma}$  en maximisant la log-vraisemblance, [ROOS 2015] a prouvé que cela revenait à minimiser la fonction proposée par le règlement délégué [COMMISSION-EUROPÉENNE 2014], donnée par :

$$\sum_{t=1}^T \pi_t(\hat{\delta}, \hat{\gamma}) \left( \ln \left( \frac{y_t}{x_t} \right) + \frac{1}{2\pi_t(\hat{\delta}, \hat{\gamma})} - \frac{\frac{T}{2} + \sum_{t=1}^T \pi_t(\hat{\delta}, \hat{\gamma}) \ln \left( \frac{y_t}{x_t} \right)}{\sum_{t=1}^T \pi_t(\hat{\delta}, \hat{\gamma})} \right)^2 - \sum_{t=1}^T \ln \left( \pi_t(\hat{\delta}, \hat{\gamma}) \right), \quad (4.2)$$

on retrouve alors notre estimateur du maximum de vraisemblance pour  $\sigma$  en  $\hat{\sigma}(\hat{\delta}, \hat{\gamma})$ , défini comme :

$$\hat{\sigma}(\hat{\delta}, \hat{\gamma}) = \exp \left\{ \hat{\gamma} + \frac{\frac{T}{2} + \sum_{t=1}^T \pi_t(\hat{\delta}, \hat{\gamma}) \ln \left( \frac{y_t}{x_t} \right)}{\sum_{t=1}^T \pi_t(\hat{\delta}, \hat{\gamma})} \right\}. \quad (4.3)$$

Comme nous disposons d'un historique  $T > 15$  années, ce qui pour la *Line of Business* 8 nous donne une crédibilité suffisante pour calculer l'intégralité du sigma de réserve, la pondération entre le paramètre de la formule standard et le paramètre propre à l'entreprise n'est ici pas nécessaire. Nous référons aux études précédentes (notamment [ROUCHATI 2016; GAUVILLE 2017]) et à la formule standard pour plus de détails sur ce point.

Enfin, le règlement délégué propose le débiaisement suivant :

$$\hat{\sigma} = \hat{\sigma}(\hat{\delta}, \hat{\gamma}) \sqrt{\frac{T+1}{T-1}}$$

Notons que, comme l'a remarqué [ROOS 2015, appendice A], ce facteur de débiaisement n'est pas optimal sous les hypothèses du modèle, mais un facteur de débiaisement optimal (rendant effectivement l'estimateur du sigma non-biaisé sous les hypothèses du modèle) peut y être calculé.

Pour plus de détails sur le théorie sous-jacente à ce modèle, notamment les preuves relatives à ces estimateurs, nous référons le lecteur intéressé à [BEAUNE 2016] ou encore [ROOS 2015].

**Données** Nous disposons de chroniques des meilleures estimations  $x_t$  et  $y_t$ , sur un historique de  $T$  années comptables (2003 à 2017). Ces données sont présentées dans la table 4.1.1.

TABLE 4.1.1 – Best estimates en inputs de la méthode 1

	$x_t$	$y_t$
2003	66 281 747	73 458 786
2004	79 728 778	81 755 943
2005	87 220 735	91 577 290
2006	99 252 117	90 161 115
2007	97 195 578	90 016 965
2008	97 170 784	96 136 986
2009	102 926 985	99 800 389
2010	103 651 373	101 391 232
2011	103 427 394	108 978 523
2012	111 734 686	121 446 257
2013	124 496 933	134 535 438
2014	136 618 520	138 101 695
2015	137 630 436	143 049 047
2016	145 205 664	145 905 864
2017	148 643 067	153 886 797

**Vérifications des hypothèses du modèle** Le modèle a fondamentalement deux hypothèses : l'hypothèse de log-normalité de la chronique  $y_t$ , ainsi que l'hypothèse de linéarité entre  $y_t$  et  $x_t$ .

Afin d'étudier la log-normalité de  $y_t$ , nous proposons d'utiliser un test de Shapiro-Wilk. Ce test classique, construit par Shapiro et Wilk en 1965 est décrit dans l'article original [SHAPIRO et WILK 1965]. La statistique de test et la  $p$ -value du test sont disponibles dans la table 4.1.2.

TABLE 4.1.2 – Tests de log-normalité de Shapiro-Wilk

	Statistique	p-value
$x_t$	0,9516855	0,5513798
$y_t$	0,9385680	0,3647053

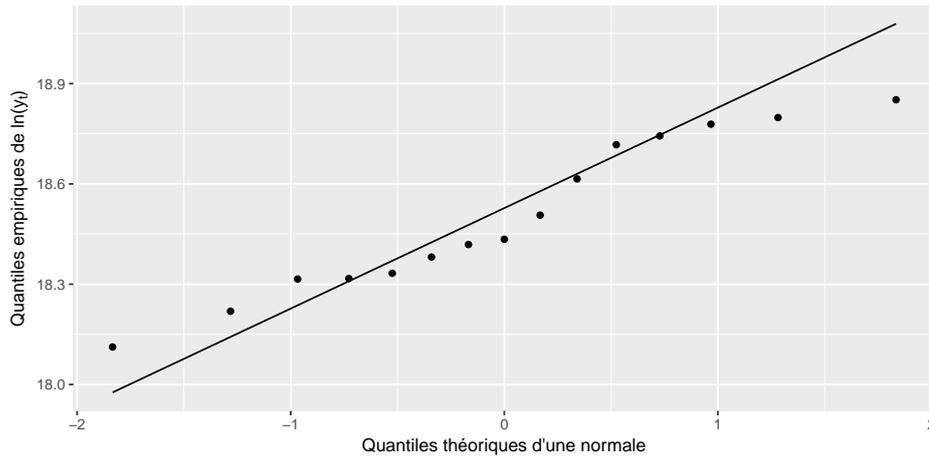
Nous proposons également d'effectuer un test de normalité Kolmogorov-Smirnov. Notons qu'il est équivalent de tester la normalité du log que la log-normalité directement. Les résultats de ce test sont disponibles dans la table 4.1.3

TABLE 4.1.3 – Tests de log-normalité de Kolmogorov-Smirnov

	Statistique	p-value
$x_t$	0,1523436	0,8273300
$y_t$	0,1547377	0,8134446

Ces deux tests confirment bien, en rejetant leurs hypothèses nulles respectives, la log-normalité à la fois pour  $y_t$  mais aussi (ce qui n'est pas demandé dans les hypothèses) pour  $x_t$ .

Pour s'en assurer, nous proposons finalement un quantile-quantile plot pour  $y_t$ <sup>1</sup>, donné en figure 4.1.1.

FIGURE 4.1.1 – Quantile-Quantile plot log-normal des  $y_t$ 

L'alignement des points n'est pas parfait, mais le peu de données disponibles nous empêche de rejeter l'hypothèse de log-normalité. Au vu des résultats des tests, nous choisissons donc d'accepter l'hypothèse de log-normalité des  $y_t$ .

L'hypothèse de linéarité sur les  $y_t$  peut se transcrire en une hypothèse de normalité de résidus donnés par :

$$r_t = \frac{\ln(y_t) - \hat{\mu}_t}{\hat{\sigma}_t},$$

avec  $\mu_t$  et  $\sigma_t$  les paramètres de la loi normale sous-jacente<sup>2</sup>.

Les résidus obtenus par cette transformation sont donc censés suivre un bruit blanc fort, c'est-à-dire un processus gaussien à auto-corrélations nulles. Si la normalité des résidus a déjà été montrée (car elle est équivalente à la log-normalité des  $y_t$ ), les tests correspondants à l'hypothèse de bruit-blanc (faible, *i.e* non-gaussien) sont disponibles à la figure 4.1.2.

On constate bien sur la figure 4.1.2 que les résidus du modèle satisfont l'hypothèse de bruit-blanc. La normalité étant déjà acquise par les tests précédents, nous acceptons donc l'hypothèse de bruit-blanc fort pour ces résidus.

Enfin, nous proposons de calculer pour chaque valeur prédite des  $y_t$  un intervalle de confiance sous les hypothèses du modèle (log-normalité, espérance et variance estimées par le modèle).

1. Pour les lois log-normales, l'usage est de passer au log et de tracer les quantiles-quantiles plot de la loi normale sous-jacente, afin d'obtenir des points moins concentrés sur un bord du graphique et donc un graphique plus lisible

2. On a par définition de la loi log-normale la méthode des moments suivante pour trouver les paramètres : (voir par exemple wikipedia pour une preuve)  $\mathbb{E}(y_t) = e^{\mu_t + \sigma_t^2/2}$  et  $\text{Var}(y_t) = (e^{\sigma_t^2} - 1)e^{2\mu_t + \sigma_t^2}$

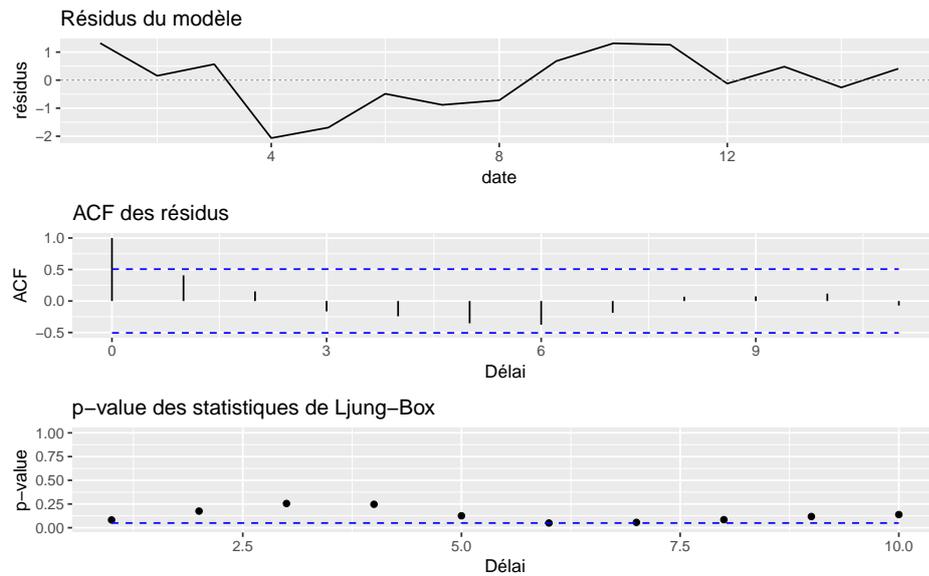


FIGURE 4.1.2 – Test de modélisation d'un bruit blanc fort sur les résidus

La droite de régression correspondante au modèle ainsi que cet intervalle de confiance sont représentés sur la figure 4.1.3.

Le fait que les données soient à l'intérieur de l'intervalle de confiance du modèle sur la figure 4.1.3 appuie encore l'acceptation des hypothèses.

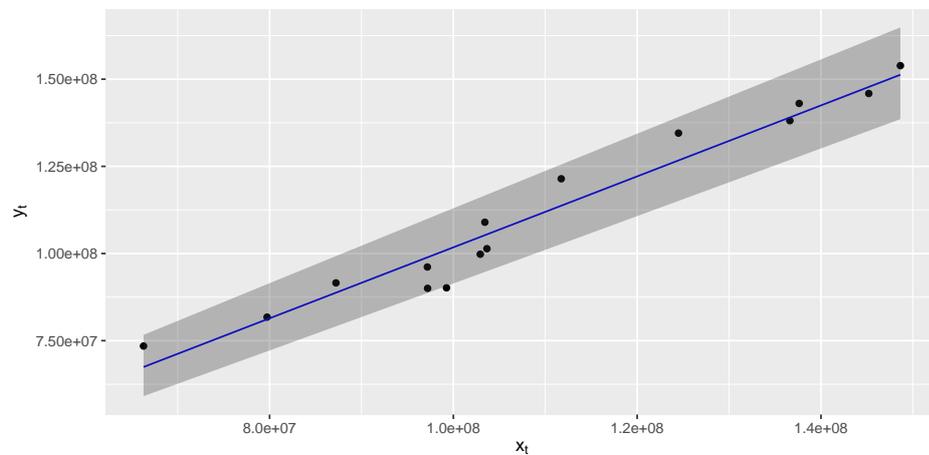


FIGURE 4.1.3 – Graphique de la relation linéaire entre  $y_t$  et  $x_t$  (l'intervale de confiance est calculé sous les hypothèses du modèle à  $\alpha = 5\%$ )

**Résultats** Les hypothèses du modèle ayant été dûment vérifiées, nous allons nous attacher à estimer les différents paramètres de ce modèle. L'application de la méthode du maximum de vraisemblance est faite par l'algorithme de Newton-Raphson, qui converge dans notre cas très vite car les données ne sont pas nombreuses. On obtient alors les estimateurs donnés à la table 4.1.4.

TABLE 4.1.4 – Résultat de l'optimisation de la méthode 1

	$\hat{\gamma}$	$\hat{\delta}$	$\hat{\sigma}(\hat{\delta}, \hat{\gamma})$	$\hat{\sigma}$
Estimations faites	-2,96175	0	5.26434%	5.62782%

On remarque tout d'abord que  $\hat{\delta} = 0$ , ce qui correspond au fait que  $\mathbb{V}(y_t) = \sigma^2 \bar{x}x_t$ .

Cette première estimation du sigma de réserve est satisfaisante car elle est très haut niveau (elle s'intéresse directement aux *Best estimate*), et prend donc en compte les spécificités de l'assurance construction telles qu'elles sont prises en compte dans le processus de provisionnement de l'entreprise.

C'est donc là une méthode efficace pour juger de la qualité du provisionnement de l'entreprise, mais elle est très générale et ne prend que peu de données en entrée ce qui fait qu'elle reste "grossière".

Nous allons dans les sections suivantes développer des méthodes plus précises et plus spécifiques à notre problématique.

### 4.1.2 Méthode 2 : Merz-Wüthrich

La seconde méthode proposée par la réglementation consiste à appliquer le modèle de Merz-Wüthrich aux données de règlements agrégées au niveau *Line of Business*. Cette méthode n'est pas applicable dans notre cas car nous disposons de données rattachées par années de survenance, pour les garanties gérées en répartition, et de données rattachées par années d'ouverture de chantier, pour les garanties décennales obligatoires gérées en capitalisation. La construction d'un unique triangle pour la *Line of Business* n'est donc pas possible.

Cependant, si l'on considère les variables aléatoires représentant des *Claim development result* de sous-portefeuille, et que l'on connaît la corrélation entre ces variables aléatoires, on peut calculer la variance de la somme via :

$$\text{Var}(\mathbf{X}_{\bullet}) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i,j \in \{1, \dots, n\}, i \neq j} \text{Cov}(X_i, X_j) \quad (4.4)$$

Il est donc tout à fait possible de calculer via la méthode 2 les écart-types des *Claim development result* de différents sous-portefeuilles et de les agréger, sous hypothèses de corrélation entre eux à définir. Cela se fait alors par application de Merz-Wüthrich à des triangles n'ayant pas tous le même type de rattachement : les données sont rattachées par inventaires pour les garanties classiques (*Responsabilité civile générale* et *Responsabilité civile décennale complémentaire*) et sont rattachées par année d'ouverture de chantier en *Responsabilité civile décennale obligatoire*.

L'application de cette entorse à la méthode règlementaire suppose :

- que l'on dispose d'une segmentation de la *Line of Business* en sous-portefeuilles sur lesquels les hypothèses du modèle de Merz-Wüthrich soient (marginale)ment vérifiées
- que l'on dispose d'une information de corrélation entre les différents *Claim development result* des sous-portefeuilles de la *Line of Business*
- que les provisions soient additives

Si la vérification des hypothèses de Merz-Wüthrich sur les différents triangles est possible, l'information de corrélation ne pourra pas être fournie par ce modèle. Nous utiliserons alors une hypothèse simplificatrice pour agréger les estimateurs d'écart-types obtenus. Après avoir trouvé une estimation plus fine de ces corrélations à la section suivante, nous reviendrons sur la fin du calcul pour compléter cet estimateur.

**Différentes visions** Plusieurs visions sont possibles pour notre modèle :

- une vision par Groupe de risques homogènes, avec trois triangles : un pour la *Responsabilité civile décennale complémentaire*, un pour la *Responsabilité civile générale*, tout deux rattachés par survenance, et un pour la *Responsabilité civile décennale obligatoire* rattaché par date réglementaire d'ouverture de chantier
- une vision par rattachement, deux triangles : un triangle par survenance agrégeant *Responsabilité civile décennale complémentaire* et *Responsabilité civile générale*, et un par date réglementaire d'ouverture de chantier pour la *Responsabilité civile décennale obligatoire* comme précédemment.

Nous donnerons les résultats et appliquerons le modèle pour ces deux visions. Cela demande donc de vérifier les hypothèses de Merz-Wüthrich sur 4 triangles différents.

**Principes et hypothèses du modèle** Pour chacun des triangles proposés, les différentes hypothèses du modèle de Merz-Wüthrich doivent être vérifiées. Nous proposons alors d'inclure les différents graphiques de vérifications des hypothèses des modèles de Mack sous-jacents aux différents modèles et des hypothèses de séries temporelles.

Pour des questions de brièveté d'exposition, les planches de graphiques des modèles de Mack ainsi que leurs analyses sont disponibles en annexe A. Les commentaires de ces planches de graphiques sont sensiblement les mêmes d'un triangle à l'autre : certaines valeurs (notamment dans la queue du triangle) sont considérées comme aberrantes par le modèle, ce qui est normal étant donné le peu d'informations disponibles sur ces cases, et nous observons également des estimations des coefficients de passage très stables sur les quatre triangles.

L'analyse des résidus des différents modèles (permettant de valider l'hypothèse de Merz-Wüthrich) ne montre aucune tendance<sup>3</sup> et valide donc les hypothèses faites par ces modèles et conforte notre analyse. Ainsi, nous sommes en mesure de calculer sur chacun de ces triangles les statistiques des modèles de Mack et de Merz-Wüthrich leur correspondants.

**Résultats** Les différents modèles de Mack nous donnent des estimations des ultimes, des *incurred but not reported*, ainsi que des variabilités à l'ultime. Ces résultats sont disponibles à la table 4.1.5.

Notons une fois de plus la non-additivité de la méthode *Chain-ladder* : Si les charges observées de la *Responsabilité civile décennale complémentaire* et de la *Responsabilité civile générale* ont bien pour somme la charge observée du triangle total en répartition, ça n'est pas le cas pour les charges ultimes et les provisions, comme nous l'avons expliqué à la section 2.4.1.

De même, on remarque que la variance de Mack diminue, ce qui montre que l'addition des deux triangles stabilise légèrement l'estimation.

TABLE 4.1.5 – Résultats des différents modèles de Mack

	Aggrégation par Groupe de risques homogènes			Aggrégation par rattachement	
	RCDC	RCG	RCDO	Repartition	Capitalisation
Charge Observée	24 525 986	111 030 855	170 308 988	135 556 841	170 308 988
Ultimes	26 173 845	107 871 774	263 932 932	134 387 186	263 932 932
IBNR	1 647 859	-3 159 082	93 623 945	-1 169 655	93 623 945
Variance de Mack	1 632 659	4 177 480	10 873 376	4 833 702	10 873 376

3. Sauf pour le triangle de *Responsabilité civile décennale complémentaire* où les résidus sont tendancieux et clairement non normaux. C'est des quatre triangles le plus faible en terme de volume et donc le plus erratique.

Les résultats des modèles de Merz-Wüthrich correspondants sur ces différents triangles sont disponibles à la table 4.1.6.

Remarquons que la variance à un an est bien inférieure à la variance à l'ultime, alors que l'approximation proposée dans [MERZ et WÜTHRICH 2008a] n'a pas été appliquée<sup>4</sup>.

TABLE 4.1.6 – Résultats des différents modèles de Merz &amp; Wuthrich

	Aggrégation par Groupe de risques homogènes			Aggrégation par rattachement	
	RCDC	RCG	RCDO	Repartition	Capitalisation
IBNR à un an	1 647 859,1	-3 159 082	93 623 945	-1 169 655	93 623 945
Variance du <i>Claim développement result</i>	769 545,4	2 031 042	5 792 741	2 393 266	5 792 741
Variance des ultimes (Mack)	1 632 659,1	4 177 480	10 873 376	4 833 702	10 873 376

Enfin, en ajoutant les PDD, nous arrivons aux résultats décrits à la table 4.1.7, nous donnant finalement une estimation du sigma de réserve fournie en 4.1.8.

TABLE 4.1.7 – Calcul des *Best estimate*

	Aggrégation par Groupe de risques homogènes			Aggrégation par rattachement	
	RCDC	RCG	RCDO	Repartition	Capitalisation
Provision dossier dossier	7 207 387	32 133 678	50 203 159	39 341 065	50 203 159
IBNR	1 647 859	-3 159 082	93 623 945	-1 169 655	93 623 945
Provision finale	8 855 247	28 974 596	143 827 104	38 171 410	143 827 104

TABLE 4.1.8 – Sigma de réserves des différents triangles

Aggrégation par Groupe de risques homogènes			Aggrégation par rattachement	
RCDC	RCG	RCDO	Repartition	Capitalisation
8.69028%	7.00973%	4.02757%	6.26979%	4.02757%

Des graphiques des *ibnr* par années de rattachement et des *Claim développement result* par années de rattachement sont respectivement disponibles en annexe B. Ces deux graphiques montrent bien la part prédominante de la *Responsabilité civile décennale obligatoire* dans le volume de l'estimation finale.

C'est un des soucis de cette méthode, car la *Responsabilité civile décennale obligatoire* prise dans cette géométrie est particulièrement stable et ne reflète probablement que mal la réalité de nos données, car elle est trop agrégée et mélange dans sa cadence deux phénomènes bien différents. Sa cadence étant également très longue (le temps de construction du chantier, le temps avant la survenance du sinistre, plus le temps de recherche des responsabilités), sa variabilité à un an est mécaniquement faible.

Enfin, un graphique général des résidus des différents triangles, permettant d'apprécier la précision des modèles, est également disponible en annexe B

**Problématique d'agrégation** Nous avons déjà fait remarquer que les écart-types des deux ou trois *Claim développement result* (respectivement selon que l'on travaille par rattachement

4. Cette approximation permet d'obtenir une forme de l'estimateur de la variance du *Claim développement result* se rapprochant de la forme obtenu par Mack en l'équation (2.4). Elle a pour défaut de forcer la variance à un an à être inférieure à la variance à l'ultime, ce qui a déjà été montré comme n'étant pas toujours le cas.

ou par Groupe de risques homogènes) peuvent s'agréger avec une simple hypothèse de corrélation, quelle que soit la structure de dépendance entre les risques sous-jacents, via la formule (4.4). En effet, le conditionnement de  $X$  par  $Y$  permet une construction simple de la variance de  $X + Y$ .

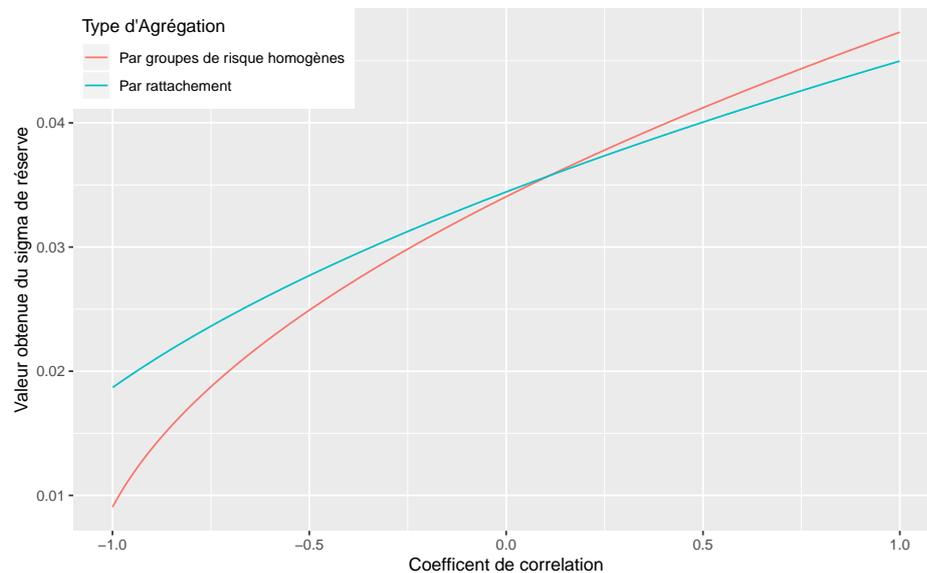


FIGURE 4.1.4 – Valeurs obtenues du sigma de réserve global en fonction du coefficient de corrélation

N'ayant pour l'instant pas d'information de corrélation entre ces variables aléatoires, nous proposons un graphique d'agrégation pour toute corrélation  $\rho \in [-1; 1]$  dans le cas par rattachement comme dans le cas par Groupes de risques homogènes<sup>5</sup>, disponible à la figure 4.1.4.

On observe sur ce graphique très logiquement que le sigma de réserve est très faible quel que soit le choix de la méthode d'agrégation, ce qui est majoritairement dû au triangle de *Responsabilité civile décennale obligatoire* rattaché par date réglementaire d'ouverture de chantier. En effet ce triangle, en mélangeant les effets de survenances futures et les effets de développement des sinistres survenus, "lisse" le bruit sous-jacent et est donc très stable.

On obtient ainsi un résultat qui va d'un peu moins de 2% pour des *Claim development result* négativement corrélés, à presque 5% pour des *Claim development result* fortement corrélés.

**Modèle de charge et modèle de règlement.** Nous avons utilisé pour cette méthode des données de charges, c'est-à-dire les règlements effectués plus les provisions dossier-dossier proposées par les gestionnaires sinistres. L'information apportée par les gestionnaires sinistres tend à stabiliser les bonis-malis futurs *par principe*, et il est donc souhaitable que les triangles de charge soient plus stables que les triangles de règlements. Avec des triangles de règlements, nous obtenons les résultats de la table 4.1.9

5. Dans le cas par Groupes de risques homogènes, nous considérons un seul coefficient de corrélation commun aux trois provisions à un an afin de simplifier le graphique. La même agrégation aurait très bien pu être faite avec n'importe quelle matrice semi-définie positive.

TABLE 4.1.9 – Sigma de réserves des différents triangles de règlements

Aggrégation par Groupe de risques homogènes			Aggrégation par rattachement	
RCDC	RCG	RCDO	Repartition	Capitalisation
16.22019%	6.48798%	6.88821%	6.05788%	6.88821%

Les modèles de règlements sont donc clairement plus volatiles que les modèles de charges, ce que nous observerons également dans la prochaine section. L'effet est flagrant en capitalisation, il s'exprime pleinement car la cadence du triangle est très longue.

Cette méthode, contrairement à la précédente, ne prend pas du tout en compte les choix de modélisation utilisés pour la construction des *Best estimate* de l'entreprise. Ainsi, si elle permet des comparaisons entre différentes entreprises par exemple, elle reste trop neutre pour observer réellement la volatilité spécifique de l'entreprise.

Pour mieux prendre en compte les besoins et les techniques de provisionnement de l'entreprise dont on calcule le sigma de réserve, nous proposons alors dans les sections suivantes une alternative construite sur les principes d'*Actuary in the box*, par bootstrap sur des quasi-Modèles Linéaires Généralisés.

## 4.2 Bootstrap récursif multivarié : le modèle de base.

Cette section a pour objet de présenter le modèle principal de ce mémoire. Il s'agit d'un bootstrap synchronisé (sous les hypothèses **B**) à un an sur les différents triangles représentant les Groupes de risques homogènes inclus dans la *Line of Business*, ainsi qu'une extension de ce principe à la capitalisation.

Les différents points théoriques abordés dans le chapitre 2 et le chapitre 3 seront utilisés dans la première partie de cette section pour construire le modèle, puis dans la seconde partie nous nous attacherons à présenter les données à notre disposition et à vérifier les hypothèses du modèle, puis nous présenterons les résultats obtenus et comparerons avec les résultats précédents, avant de conclure. Dans la prochaine section nous étudierons l'impact de certains choix de modélisation sur le résultat.

### 4.2.1 Principes du modèle

La modélisation d'une assurance construction française gérée en capitalisation dans le cadre Solvabilité II pose certains problèmes, notamment la prise en compte de la capitalisation pour la construction d'un *Undertaking specific parameter* sur la *Line of Business* correspondante.

En capitalisation, peu d'estimateurs de la variance des provisions à un an (et encore moins à  $a$  années) existent. En effet, ce cas est trop particulier pour que des méthodes type Merz-Wüthrich soient développées. La plupart des études précédentes que nous avons à disposition sur le sujet se contentent d'utiliser le *framework* fourni par Merz-Wüthrich sur des triangles agrégés par date réglementaire d'ouverture de chantier et par délai de survenance + délai de développement, correspondant à la notation  $Y_{i,\bullet,k}$  introduite en (3.3). C'est d'ailleurs l'approche que nous avons proposée dans la section 4.1.2.

Cette approche a un soucis fondamental : elle ne fait pas la différence entre les survenances futures et les développements de sinistres déjà manifestés, les traitant comme un seul phénomène. C'est un souci de type économique : l'information captée par la cadence du triangle  $Y_{i,\bullet,k}$  est trop brouillée, trop mélangée pour que l'estimateur soit satisfaisant. En réduisant

le nombre de paramètres du modèle, nous avons en effet annihilé la variance intrinsèque du sous-jacent. C'est pourquoi nous proposons ici une approche plus complète de modélisation, prenant en compte les trois dimensions de l'assurance construction.

Notons cependant que quelques prémisses existent, notamment [DEPLANTE 2010, point 9.4] qui propose une modélisation quasi-Poisson (sans modélisation jointe, *i.e* avec un paramètre de dispersion unique) correspondante à notre modèle **EP** développé dans le chapitre précédent, mais ne pousse le modèle que jusqu'aux provisions et pas jusqu'au risque à un an. Or il est évident qu'un tel modèle peut facilement (via un bootstrap des résidus sur les cases correspondantes à l'année prochaine, paramétrique ou non) servir à calculer le risque à un an, nous en avons parlé au point 3.2.3. Nous avons déjà cependant identifié que nos données présentaient des sommes marginales négatives disqualifiant ce type d'approche.

A contrario, suivant [BOUMEZOUED et al. 2011], les calculs de risque à un an se prêtent très bien à l'approche bootstrap, du moins dans le cadre de provisionnement classique, comme nous l'avons montré à la section 2.3 de ce mémoire.

Rappelons de plus que la méthode de calcul de la provision pour sinistres non-encore manifestés en normes économiques, développée au point 3.1.3 de ce mémoire, donne une vision de la provision pour sinistres non-encore manifestés *conditionnellement* à la provision pour sinistres à payer. C'est ce conditionnement que nous allons exploiter ici pour développer un estimateur du risque à un an de la provision globale de la *Line of Business*.

De même, la présence de plusieurs triangles de liquidation conjoints, intrinsèque à la définition en Groupes de risques homogènes de notre *Line of Business*, est lui aussi problématique. Si des estimateurs Merz-Wüthrich du risque à un an existent dans le modèle de Braun (d'hypothèses **B**) comme dans le modèle de Schnauss (d'hypothèses **MMS**), l'approche par bootstrap synchronisé est ici plus intéressante de par le besoin de conditionnement dû à la gestion de capitalisation.

Notons que les *Claim development result* marginaux de l'approche bootstrap proposée ici (en survenance-inventaire) sont démontrés comme étant convergents vers les résultats de Merz-Wüthrich si nous choisissons un bootstrap de type paramétrique avec simulation de lois normales, comme nous l'avons montré au point 2.4.3.

Enfin, la vision Modèles Linéaires Généralisés des différents modèles nous permettra de facilement *découper* la cadence de manifestation par un zonage tel que défini au point 2.4.3, ainsi que d'inclure des ajustements type Bornhuetter-Fergusson ou stabilisation des coefficients de passage dans la modélisation. L'ensemble de ces ajustements seront présentés *supra*.

Le modèle que nous proposons de mettre en œuvre sur nos données, appelé *supra* *modèle de base*, est représenté par un schéma synthétique disponible en figure 4.2.1. Il consiste en les étapes suivantes :

- les trois triangles de *Responsabilité civile décennale obligatoire*, *Responsabilité civile décennale complémentaire* et *Responsabilité civile générale*, pris tout trois en survenance inventaire, feront l'objet d'un bootstrap dans un modèle de Braun **B**, à horizon  $a$  années.
- sur chaque échantillon, la cadence de développement obtenue (à  $a$  années) sur le triangle de *Responsabilité civile décennale obligatoire* sera utilisée pour passer à l'ultime le triangle en capitalisation via (3.4).
- la provision pour sinistres non-encore manifestés sera ensuite bootstrapée à  $a$  années dans le modèle de Mack **M**, conditionnellement au bootstrap précédent.



## 4.2.2 Vérification des hypothèses

Les premières hypothèses à vérifier sont les hypothèses de quasi-Modèles Linéaires Généralisés pour les modèles de Mack sous-jacents au modèle de Braun.

Comme les modèles sont des modèles linéaires, nous pouvons simplement chercher, comme le préconisent Merz-Wüthrich, un *trend* dans nos résidus, par années calendaires, par années de développements, ou par années de rattachements. Les graphiques des résidus sur ces 3 covariables possibles sont disponibles à la figure 4.2.2. Nous attirons tout particulièrement l'attention sur les années calendaires, les méthodes de *link-ratio* supposant cet effet nul il serait très gênant d'observer des résidus non-aléatoires sur cet axe.

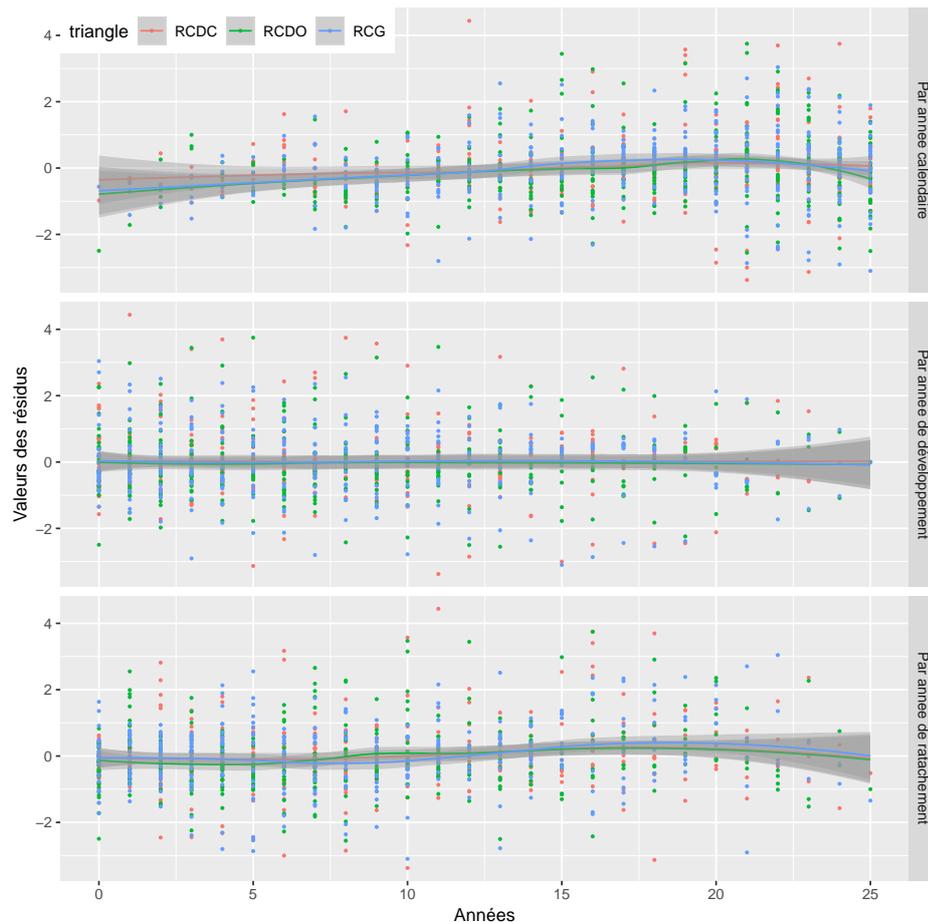


FIGURE 4.2.2 – Résidus observés par années calendaires, par délais de développements et par années de rattachements sur les triangles

Au vu de ces graphiques, nous pouvons accepter la non-existence de *trend* dans nos modèles.

La validation d'un quasi-Modèle Linéaire Généralisé passe normalement par l'observation des *p-value* et des statistiques de Fisher des paramètres. En effet les *p-value* et les statistiques de Fisher associés à l'estimation par maximum de vraisemblance sont toujours valables dans le cadre d'un pseudo-maximum de vraisemblance. Toutes ces statistiques sont disponibles en annexe C.

Les *p-value* sont très bonnes, sauf sur la *Responsabilité civile décennale complémentaire* qui est le Groupe de risques homogènes le plus volatile que nous ayons.

Les graphiques d'appréciation de la modélisation jointe sont également disponibles aux figures en annexe C.

L'analyse de ces planches de graphiques nous pousse, selon la même méthodologie que précédemment, à accepter les hypothèses des modèles de Mack sur les triangles en survenance-inventaire, avec une légère réserve sur la *Responsabilité civile décennale complémentaire*.

Observons d'autre part les différentes corrélations entre les colonnes des triangles proposées par Braun en (2.10), disponibles à la figure 4.2.3.

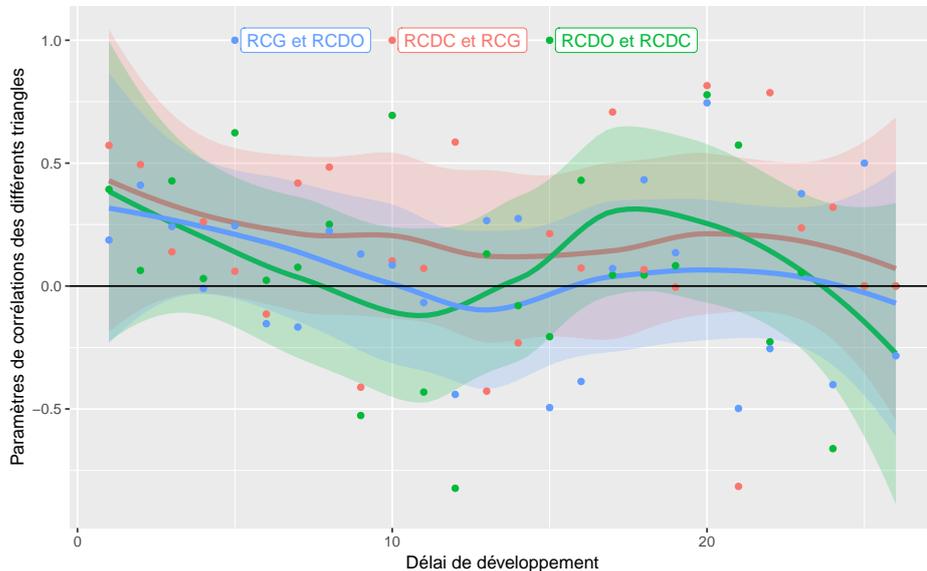


FIGURE 4.2.3 – Paramètres de corrélations de Braun des triangles en survenance-inventaire par délai de développement

On remarque sur la figure 4.2.3 que les trois triangles du modèle de Braun sont *globalement* positivement corrélés, même si l'effet est assez faible et que l'estimateur est très volatile. On remarque que les charges d'ouvertures sont très corrélées entre elles, et que cette corrélation s'amenuise au fur et à mesure du développement des triangles.

Comme nous comptons ici utiliser un bootstrap non-paramétrique, l'étude de cette corrélation n'apporte que peu d'informations à la validation de nos hypothèses. Elle sera cependant utile plus tard, lors des extensions du modèle que nous proposerons à la section suivante.

Enfin, les dernières hypothèses à valider sont les hypothèses des modèles de Mack sur les triangles de *Responsabilité civile décennale obligatoire* en date règlementaire d'ouverture de chantier par survenance ramenés à l'ultime, conditionnels aux triangles de *Responsabilité civile décennale obligatoire* en survenance-inventaire bootstrapés. La validation des hypothèses des triangles dans chaque état de la nature créé par le rééchantillonnage est cependant complexe à mettre en œuvre.

### 4.2.3 Résultats

Les hypothèses du modèle ayant été vérifiées, nous proposons ci-dessous quelques résultats et commentaires de l'application.

Tout d'abord comme nous sommes dans une problématique bootstrap, il convient de regarder le Burn-in des estimateurs cibles. Nous proposons alors d'observer le burn-in des corrélations des différentes provisions entre elles, mais aussi des sigma de réserves obtenus, ainsi que de la variance des provisions à la figure 4.2.4.

Notons que nous avons tiré 10 000 échantillons bootstrap en répartition, et seulement 1, conditionnel, en capitalisation. En effet, la probabilité conditionnelle de chaque rééchantillonnage

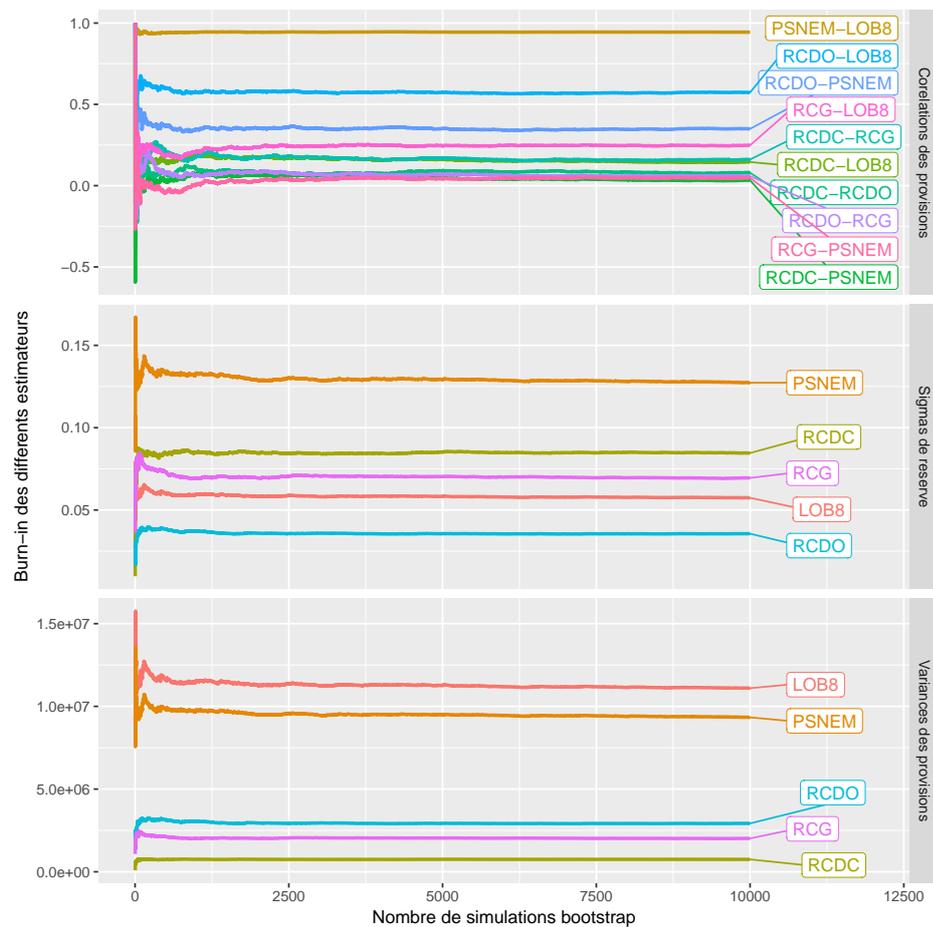


FIGURE 4.2.4 – Burn-in des corrélations des IBNR à un an, de leurs variances et des sigmas de réserves correspondants

en capitalisation étant la même, et le nombre de rééchantillonnages en répartition étant déjà élevé, il n'est pas nécessaire de tirer plus d'échantillons dans le bootstrap imbriqué. On observe en effet à la figure 4.2.4 une grande stabilité des estimateurs et une convergence forte à partir de 2000 tirages. Des résultats équivalents ont été observés sur des nombres de tirages plus importants.

Le bootstrap permet également d'observer de manière simple des intervalles de confiance pour les coefficients de passage des différents triangles. Les cadences de développement bootstrapées (vues à la date de calcul), avec leurs intervalles de confiance à 99%, sont disponibles à la figure 4.2.5.

Si l'écart-type des estimations est satisfaisant, on observe cependant un peu de bruit sur la queue qui pourrait faire l'objet d'une stabilisation tel qu'il en fut question au point 2.5.2 de ce mémoire. Nous en discuterons dans la section suivante.

Enfin, le bootstrap dans le modèle de Braun donne des résultats (d'ordre 1) intéressants en répartition, disponibles à la table 4.2.1, durant la lecture desquels il faut prendre en compte la Provision Dossier-Dossier disponible à la table 4.2.2.

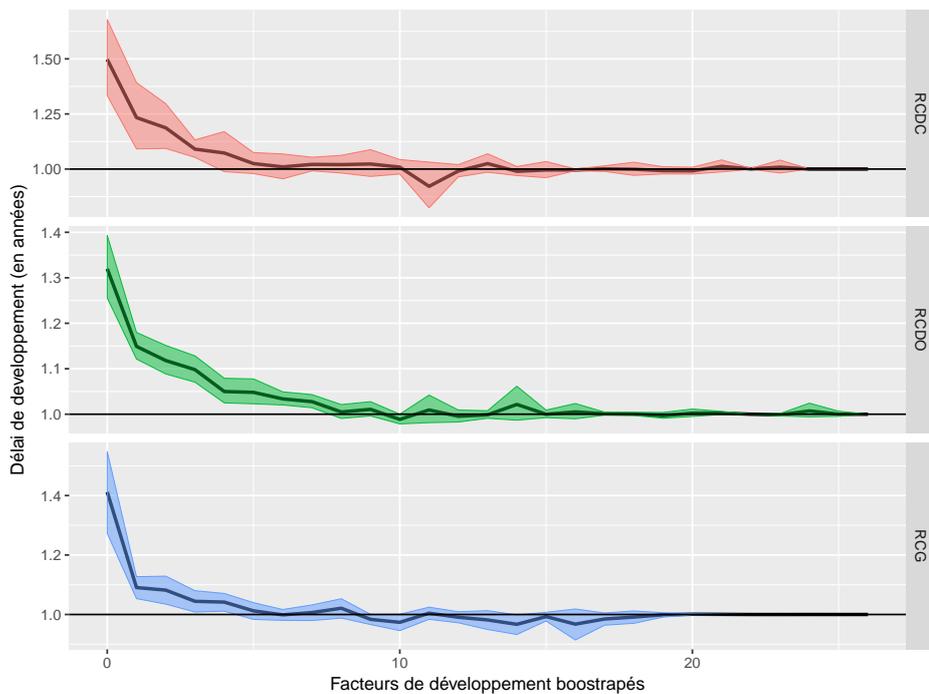


FIGURE 4.2.5 – Facteurs de développement obtenus par le bootstrap synchronisé des Groupes de risques homogènes, partie répartition

TABLE 4.2.1 – Résultats moyens du Bootstrap synchronisé - Totaux par Groupe de risques homogènes en répartition

	RCDC	RCDO	RCG	Tot
Ultimes $n$	26 173 845	217 221 080	107 871 774	<b>351 266 698</b>
IBNR $n$	1 647 859	32 053 923	-3 159 082	<b>30 542 700</b>
Ultimes $n + 1$	26 173 856	217 191 652	107 903 099	<b>351 268 607</b>
IBNR $n + 1$	1 647 870	32 024 496	-3 127 756	<b>30 544 609</b>

TABLE 4.2.2 – Provisions dossiers des différents GRH

	RCDC	RCDO	RCG
Provisions Dossier	7 207 387	50 144 583	32 133 678
Dossier			

Puis, le bootstrap synchronisé permet d'avoir une vision stochastique des *incurred but not reported* des différents triangles et de la provision pour sinistres non-encore manifestés. Ces densités sont représentées à la figure 4.2.6, où les traits noir verticaux correspondent aux valeurs déterministes cette année des *incurred but not reported* ou provision pour sinistres non-encore manifestés. Notons la prépondérance de la variance de la provision pour sinistres non-encore manifestés dans la (future) variance totale.

Enfin, les résultats principaux du modèle et finalement le sigma de réserve sont disponibles aux tableaux 4.2.3, 4.2.4 et 4.2.5.

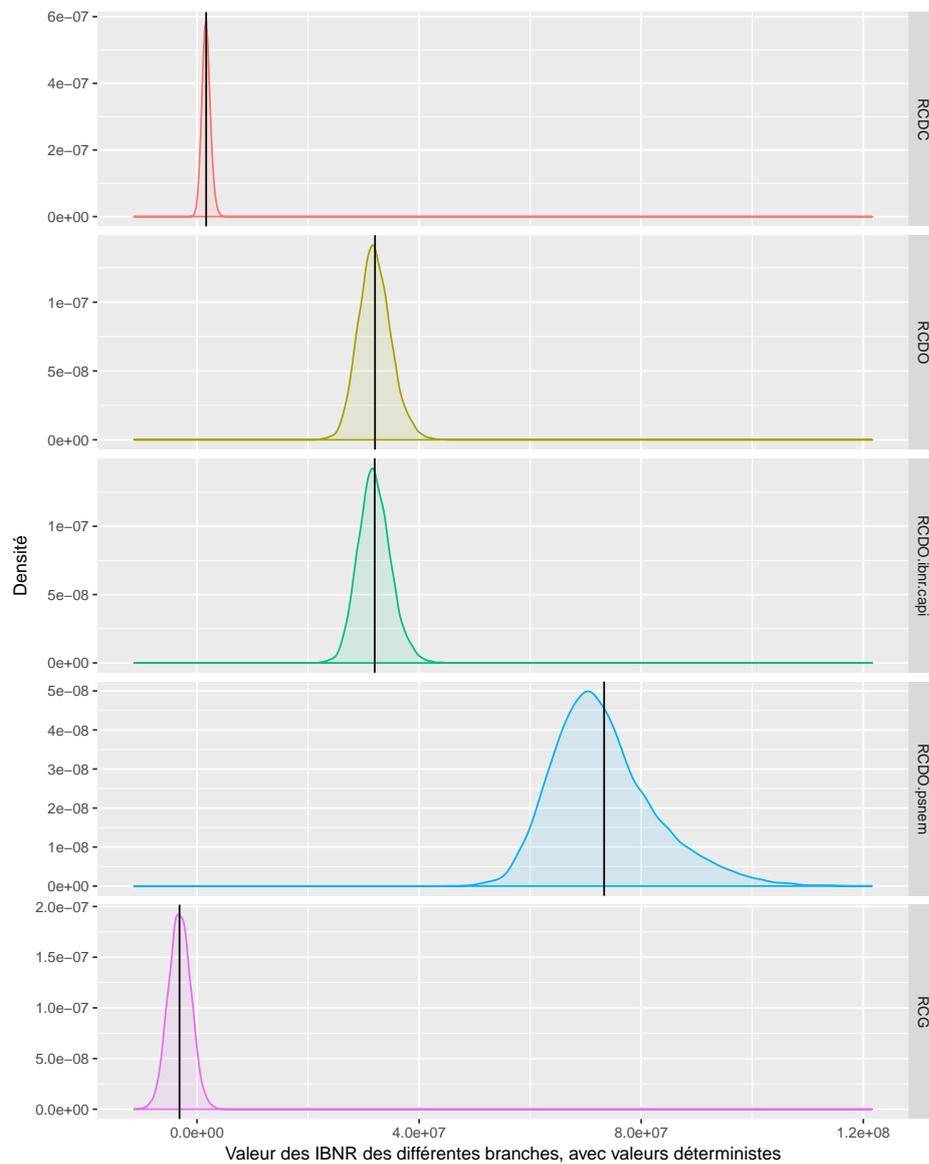


FIGURE 4.2.6 – Densité des IBNR à un an, comparées à leurs valeurs déterministes cette année.

TABLE 4.2.3 – Corrélations des CDR, des Provisions et/ou des IBNR selon les branches

	Psap - RCDC	Psap - RCG	Psap - RCDO	Psnem - RCDO
Psap - RCDC	1,0000	0,1609	0,0788	0,0330
Psap - RCG	0,1609	1,0000	0,0580	0,0477
Psap - RCDO	0,0788	0,0580	1,0000	0,3481
Psnem - RCDO	0,0330	0,0477	0,3481	1,0000

TABLE 4.2.4 – Statistiques finales du Bootstrap

	RCDC	RCG	RCDO	PSNEM	<b>LOB8</b>
IBNR (ou PSNEM)	1 647 870	-3 127 756	31 967 620	73 329 345	<b>103 817 079</b>
PDD	7 207 387	32 133 678	50 144 583	0	<b>89 485 648</b>
Provisions	8 855 258	29 005 921	82 112 203	73 329 345	<b>193 302 727</b>
Ecartype	749 284	2 015 667	2 923 819	9 335 510	<b>11 098 220</b>

TABLE 4.2.5 – Sigma de réserve obtenus par le bootstrap synchronisé

	Ecartype	Moyenne	Sigma de réserve
RCDC	749 284,2	8 855 258	<b>8.4615%</b>
RCDO	2 923 819,1	82 112 203	<b>3.5608%</b>
RCG	2 015 666,7	29 005 921	<b>6.9492%</b>
PSNEM	9 335 509,9	73 329 345	<b>12.7309%</b>
<b>LOB</b>	<b>11 098 219,6</b>	<b>193 302 727</b>	<b>5.7414%</b>

Un graphique montrant la convergence sur les marginales du bootstrap vers la formule de Merz-Wüthrich est disponible à la figure 4.2.7. Nous avons déjà parlé de cette convergence et nous profitons ici de l'application du modèle pour la vérifier empiriquement.

Nous pouvons également vérifier la cohérence (et avoir une idée du bruit du bootstrap) entre les résultats de Merz-Wüthrich sur les triangles de *Responsabilité civile décennale complémentaire* et de *Responsabilité civile générale*, disponibles à la table 4.1.8, et ceux obtenus par le bootstrap disponibles à la table 4.2.5.

Une telle comparaison sera aussi possible entre le modèle de règlement de la méthode 2 et le modèle de règlement du bootstrap, que nous développerons dans la section suivante.

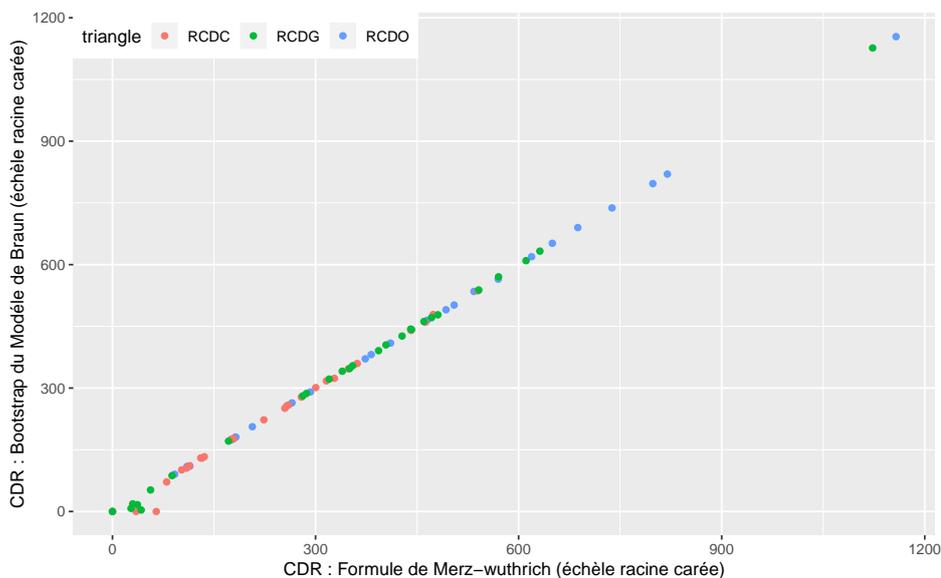


FIGURE 4.2.7 – Comparaison des CDR obtenus par année d'origine sur les triangles et des formules fermées de Merz-Wüthrich

### 4.3 Bootstrap récursif multivarié : raffinements

Si le **modèle de base** développé au point précédant est simple à comprendre et à mettre en œuvre, il ne reflète que peu les techniques de calculs de provisions utilisées dans le cadre du calcul des *Best estimate* par la majorité des entreprises confrontées au problème. En effet, l'utilisation par les entreprises d'un modèle *Chain-ladder* pur à l'ordre 1 est rare. Afin de mieux comprendre la volatilité réelle des provisions de la *Line of Business*, nous proposons ici quelques ajustements et raffinements du modèle de base qui permettront de prendre en compte les estimations faites par l'entreprise à l'ordre 1 dans la construction de l'ordre 2. Nous comparerons ensuite les résultats obtenus selon les différentes modalités.

### 4.3.1 Modèle de charges et modèle de règlements

Le modèle de base est appliqué sur les triangles de charges. On peut également l'appliquer directement aux triangles de règlements. En effet la spécification du modèle de bootstrap, tout comme la spécification de l'extension à la capitalisation, est indépendante de la donnée présente dans les triangles<sup>6</sup>.

L'utilisation d'un modèle de charge demande l'ajout de la provision dossier-dossier aux provisions finales estimées. Si cela permet la prise en compte du travail des gestionnaires sinistres et donne donc un meilleur estimateur des variations de provisions à un an de l'entreprise, le modèle de règlements correspond mieux aux spécifications Solvabilité II de la méthode 2 étudiée *supra*, au point 4.1.2.

En théorie, ces deux modèles devraient converger vers la même estimation, à l'ordre 1, mais le modèle de règlements devrait être plus dispersé que le modèle de base (modèle de charges). On s'attend donc à observer des sigma de réserve plus élevés, mais correspondants moins à la réalité des provisions techniques de l'entreprise, dans le cas des modèles de règlements.

Nous comparerons donc les résultats d'un modèle basé sur les règlements par rapport aux modèles de base.

### 4.3.2 Zonage en capitalisation

La figure 4.3.1 représente la cadence de manifestations bootstrap de la *Responsabilité civile décennale obligatoire* dans le modèle de base. On observe un effet classique, dû à la définition de la garantie décennale, nous donnant une cadence avec une forme caractéristique.

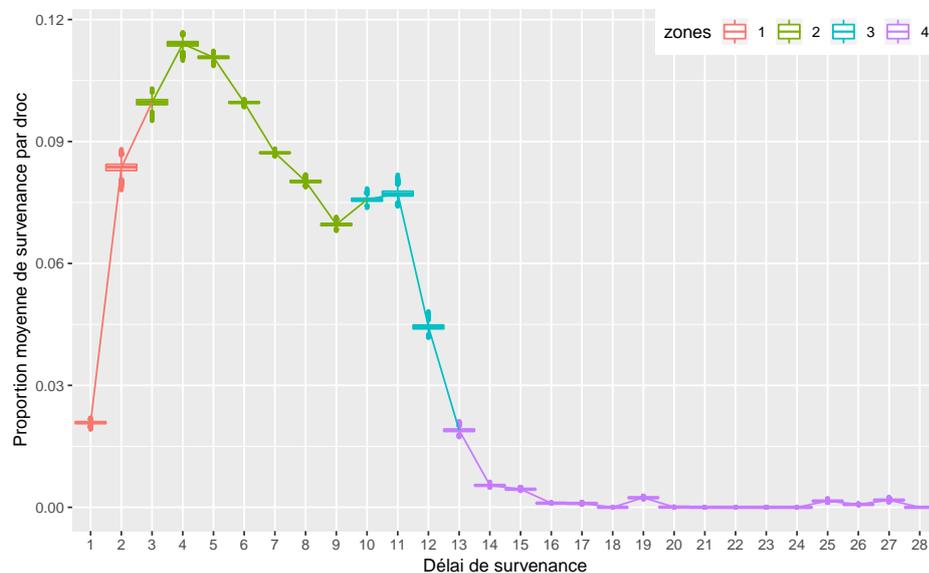


FIGURE 4.3.1 – Boxplots des cadences de manifestation bootstrap de la RCDO - Zonées selon les natures de risques.

On observe sur ce graphique plusieurs zones :

- une première zone de montée en charge dans les premières années de survenance, correspondante à la période de construction des chantiers. Les sinistres survenant dans cette zone sont spécifiques : ils sont immédiatement remarqués, ou très vite, et déclarés rapidement après la réception du chantier.

6. Une extension aux triangles de primes ou de nombres serait également envisageable

- une second zone, plus longue, correspondant au rythme de croisière de la garantie décennale
- un pic en fin de garantie décennale, correspondant au fait que, la garantie arrivant à son terme, on refait une vérification de la construction *au cas où*.
- puis une décroissance exponentielle en fin de garantie.

On observe également sur le graphique que la variabilité de la cadence de manifestation bootstrappée est différente selon les zones. Nous proposons alors une possibilité de zoner le triangle lors du bootstrap, selon les principes développés en 2.3.4. La partition par colonne proposée (correspondante aux couleurs sur le graphique), construite pour correspondre à une vision économique de la cadence de manifestation, est alors :

$$\mathcal{Z} = \{j \leq 2, j \in \{3, \dots, 9\}, j \in \{10, \dots, 12\}, j \geq 13\} \quad (4.5)$$

Étant donné le caractère de modélisation jointe de notre modèle, nos résidus sont par construction homoscedastiques par colonne. Nous nous attendons donc à un effet négligeable de cet ajustement.

### 4.3.3 Stabilisation des coefficients de passage

On observe aux graphiques 4.2.5 et 4.3.1 les cadences de développements respectives des 4 triangles avec lesquels nous travaillons. Pour la plupart d'entre elles, la queue du triangle est trop bruitée et nous proposons alors une stabilisation à 1 des cadences.

L'impact de cette stabilisation sur le bootstrap est simplement l'exclusion des résidus correspondants du rééchantillonnage, et la stabilisation inconditionnelle des pseudo-coefficients de passage à 1.

Pour les triangles en répartition, la synchronisation du bootstrap nous oblige à choisir une année de stabilisation identique pour les trois triangles. Au vu des cadences, et surtout au vu de leurs variabilités, nous proposons une stabilisation au bout de 18 années de développement.

Pour le triangle en capitalisation, nous proposons une stabilisation au bout de 16 années de développement. Ces deux stabilisations seront évaluées indépendamment l'une de l'autre.

Le choix du nombre d'années est calquée sur les avis d'experts utilisés lors du calcul des *Best estimate* de l'entreprise.

### 4.3.4 Rééchantillonnage paramétrique ou non-paramétrique

Si jusqu'à maintenant nous avons utilisé la version non-paramétrique des rééchantillonnages, une convergence vers la formule de Merz-Wüthrich est également possible via une technique paramétrique, consistant à simuler des lois normales centrées réduites plutôt que de bootstrapper des résidus sur les prochaines diagonales. Les démonstrations relatives à ces principes ont été évoquées en 2.3, et sont plus largement décrites dans [BOUMEZOUED et al. 2011].

Cette version paramétrique du bootstrap a les implications suivantes :

- de la structure de dépendance entre les résidus des différents triangles, seule la corrélation est retenue (hypothèse de vecteur gaussien pour les résidus), par colonne.
- l'hypothèse de loi transforme les différents quasi-Modèles Linéaires Généralisés en Modèles Linéaires Généralisés complets, cette hypothèses doit être vérifiée

Les graphiques représentant les densités des résidus des différents triangles, ainsi que les quantiles-quantiles plot de ces résidus sont disponibles aux figures 4.3.2 et 4.3.3.

L'hypothèse de normalité des résidus est discutable, mais comme nous ne recherchons que fondamentalement un moment d'ordre 2 et pas une distribution complète, l'ajustement que cette option propose nous semble suffisant.

Sous cette modalité, nous tirons alors les nouveaux résidus selon des loi normales multivariés centrées réduites construites selon les corrélations par colonnes exposées à la figure 4.2.3, et calculées dans le modèle **B** selon l'équation (2.10).

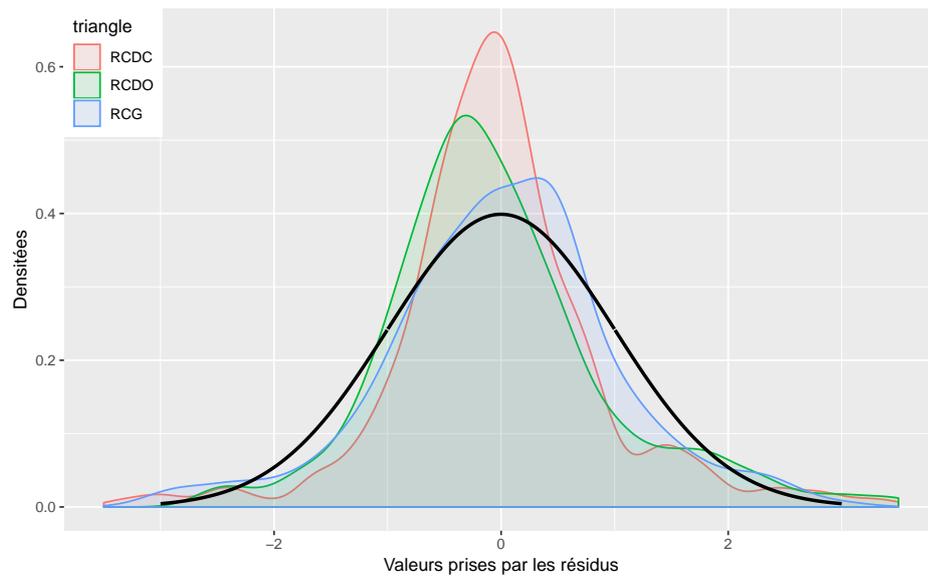


FIGURE 4.3.2 – Densité des résidus des modèles de Mack, avec (en noir) une loi normale centrée réduite

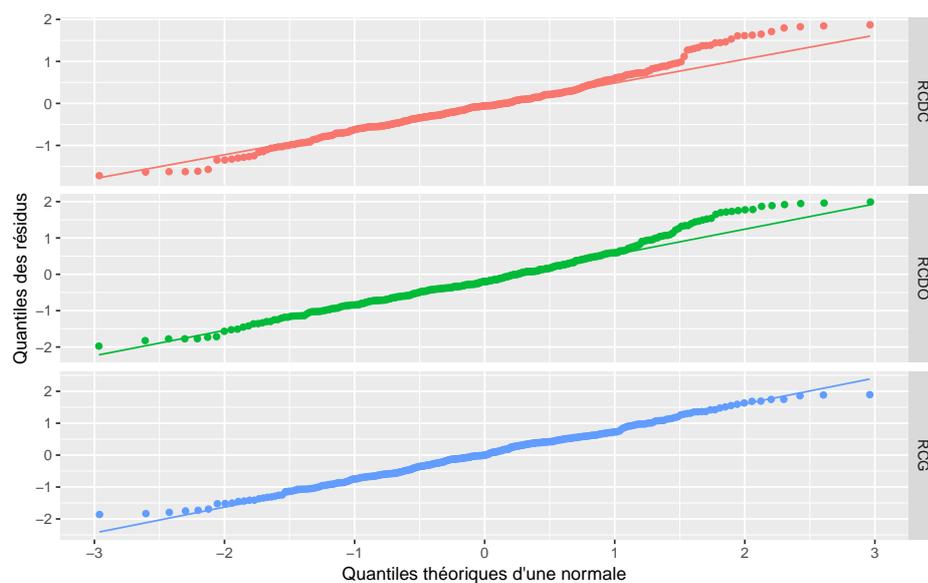


FIGURE 4.3.3 – Quantile-Quantile plot des résidus contre une loi normale centrée réduite

### 4.3.5 Exclusion de résidus

Lors de la modélisation par quasi-Modèle Linéaire Généralisé, certains résidus du modèle peuvent sembler suspects.

En effet, des valeurs aberrantes et/ou influentes peuvent poser problème, surtout dans un cadre de bootstrap.

Si c'est le cas, nous pouvons tout simplement les exclure du rééchantillonnage, comme nous les aurions exclus du calcul des paramètres lors d'une modélisation par quasi-Modèle Linéaire Généralisé. Nous référons à [LO 2011] sur ce sujet.

Nous avons alors conçu une modalité de bootstrap où tous les résidus dont la valeur absolue est supérieure ou inférieure au quantile à 97,5% d'une loi normale centrée réduite (soit environ 1,96) seront exclus. Nous excluons ainsi, sous hypothèse normale, environ 5% des résidus.

En effet, l'analyse des résidus en survénance-inventaire montre que ce choix permet d'exclure des résidus aberrants, et l'analyse fine des triangles en droc-survenance ne peut se faire, de par la structure du modèle, qu'échantillon bootstrap par échantillon bootstrap.

### 4.3.6 Bornhuetter-Ferguson en capitalisation

En capitalisation, comme le montre la figure 4.3.1, les premières années de survénance ne représentent qu'une très faible partie de la charge finale.

C'est pourquoi un ajustement Bornhuetter-Ferguson est souvent utilisé sur les dernières dates réglementaires d'ouverture de chantier afin de décrédibiliser l'estimation *Chain-ladder* de la charge ultime et d'introduire une information de primes.

Notons que l'esprit est semblable à ce qui est proposé dans la réglementation comptable française, décrite au point 3.1.1 : les primes sont considérées plus représentatives du devenir de la charge rattachée à une date réglementaire d'ouverture de chantier récente que les sinistres. La mise en œuvre est par contre bien différente.

Nous proposons alors, dans chaque échantillon bootstrap du triangle en capitalisation, de calculer et utiliser un ajustement Bornhuetter-Ferguson, selon la méthodologie de modification du modèle décrite au point 2.5.1. L'information extrinsèque au triangle est alors composée :

- des primes ultimes, considérées comme déterministes<sup>7</sup>.
- d'un vecteur de poids pour la pondération des loss-ratio.

Nous choisissons deux vecteurs de poids différents : si le premier propose d'appliquer aux 5 dernières dates réglementaires d'ouverture de chantier un loss-ratio moyen calculé sur les 6 dates réglementaires d'ouverture de chantier précédentes, le second propose d'appliquer aux 3 dernières dates réglementaires d'ouverture de chantier un loss-ratio moyen calculé sur les 6 dates réglementaires d'ouverture de chantier précédentes.

L'option "(6/5)" sera probablement plus stable que l'option "(6/3)", car les dates réglementaires d'ouverture de chantier plus récentes sont moins développées et donc plus volatiles.

---

7. Notons que la volatilité des primes est alors négligée, mais qu'elle est négligeable : les ultimes de primes sont atteints au bout d'un an avec une incertitude très faible

## 4.4 Comparaison des résultats des différentes méthodes

Les différentes possibilités de paramétrage des ajustements fournissent tous des sigma de réserve entre 3% et 7% environ. La combinatoire des possibilités porte le nombre de modèles à 192, ce qui est élevé et ne permet pas une analyse fine de chacun d'entre eux. Cependant, on peut essayer de comprendre quels sont les ajustements les plus influents sur le résultat via une analyse de la variance de l'estimateur.

Nous proposons alors deux visualisations de ces résultats. La première sous forme d'arbre de régression (CART) où les covariables sont les différents ajustements et paramétrages possibles<sup>8</sup>. L'arbre est *pruné* pour une meilleure visualisation des paramètres les plus influents, la variable réponse étant les sigma de réserve obtenus. Cet arbre est disponible à la figure 4.4.1.

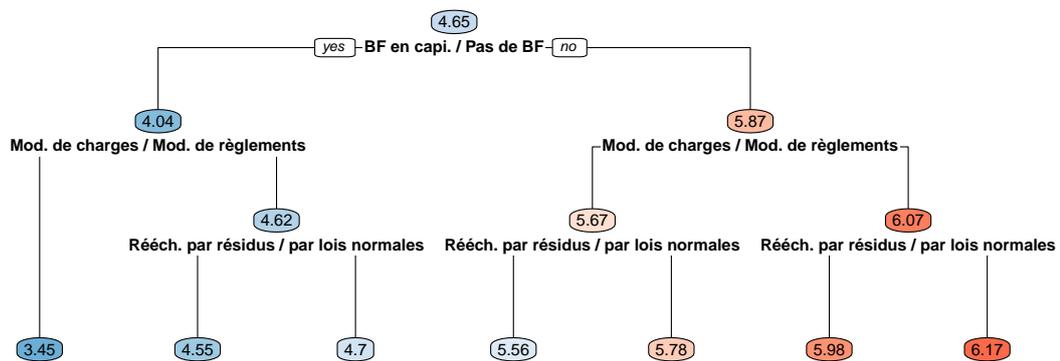


FIGURE 4.4.1 – Sigma de réserve moyen obtenu, classification selon les paramétrages les plus influents

TABLE 4.4.1 – Estimations du sigma de réserve sous les différents paramétrages et ajustements proposés (en %)

	RCDC	RCG	RCDO	PSNEM	LOB8
<b>Modèle de base (MdB.)</b>	<b>8,46</b>	<b>6,95</b>	<b>3,56</b>	<b>12,73</b>	<b>5,74</b>
MdB. + Bornhuetter-Fergusson(6,5)	8,54	6,93	3,56	4,60	<b>3,26</b>
MdB. + Bornhuetter-Fergusson(3,5)	8,63	6,96	3,61	5,75	<b>3,66</b>
MdB. + Utilisation des triangles de règlements	16,08	6,35	5,80	11,46	<b>5,99</b>
MdB. + Rééchantillonnage par loi normales	8,57	7,07	3,58	12,80	<b>5,77</b>
MdB. + Exclusions des résidus	8,56	6,85	3,52	12,42	<b>5,56</b>
MdB. + Zonage des résidus en capitalisation	8,63	6,99	3,58	12,87	<b>5,74</b>
MdB. + Stabilisation en répartition	8,06	6,74	3,43	12,74	<b>5,73</b>
MdB. + Stabilisation en capitalisation	8,61	7,09	3,58	12,95	<b>5,73</b>

La seconde visualisation permet de voir le sigma de réserve associé à chacune des options (binaires) d'ajustement et de paramétrage par rapport au modèle de base, le modèle naïf développé dans la précédente section. Ces résultats sont disponibles à la table 4.4.1. Notons que les effets croisés des paramétrages ne sont ici pas exposés.

Notons que l'on obtient les mêmes informations par Groupe de risques homogènes que sur la *Line of Business* entière, le bootstrap fournissant des *Claim development result* marginaux pour chaque triangle.

8. Cela constitue une *ANalysis Of VAriance*

Notons également que certains résultats sont censés, par principe, être nuls : la variation des résultats de la partie répartition en cas de modification du Bornhuetter-Fergusson ou de la stabilisation en capitalisation doit être nulle. On observe donc un effet "précision" du bootstrap : d'un lancement à l'autre, le même sigma de réserve théorique peut varier jusqu'à 0,03%. Observons également que les différents triangles ne subissent pas de la même manière ce bruit : le sigma de réserve du triangle de *Responsabilité civile décennale complémentaire*, le triangle avec le moins de volume, peut varier jusqu'à 0,2% d'un lancement à l'autre.

Enfin, l'on observe :

- une influence importante du Bornhuetter-Fergusson en capitalisation, quel que soit son paramétrage. Les primes utilisées pour le calcul des ultimes *a priori* n'étant pas stochastiques, la stabilisation est forte. Cependant, la variabilité des primes est presque négligeable, et les premières dates règlementaires d'ouverture de chantier d'un triangle de *Responsabilité civile décennale obligatoire* en capitalisation sont régulièrement peu crédibles : la cadence de manifestation nous montre que ces dates règlementaires d'ouverture de chantier ne contiennent qu'une faible partie de la (future) charge ultime et sont donc sujettes à une grande variabilité.
- une très forte différence entre les triangles de règlements et les triangles de charges, montrant l'utilité des provisions dossier-dossier en terme de variabilité du *Claim development result*.
- un effet plus faible qu'espéré des triangles de règlements sur la volatilité de la provision pour sinistres non-encore manifestés : le triangle en capitalisation étant à l'ultime, il devrait (théoriquement) être le même entre les modèles de règlements et les modèles de charges.
- un effet négatif de la stabilisation en capitalisation : l'exclusion des résidus (souvent presque nuls) de la partie stabilisée du triangle augmente la probabilité de tirer des résidus plus importants sur le reste du triangle et augmente donc la variabilité de l'estimation du *Claim development result*.
- un effet presque nul de la stabilisation en répartition : les résidus de la partie stabilisée sont en effet "cohérents" avec les autres.
- un effet du zonage négligeable : en effet, nos résidus sont déjà renormalisés à l'ordre 1 (ils sont centrés) et à l'ordre 2 (ils sont homoscedastiques), et les séparer en plusieurs groupes n'a pas beaucoup d'influence.

En choisissant les mêmes ajustements que ceux faits dans la méthode 2 de la section précédente, c'est-à-dire en prenant le modèle de base, on retrouve les mêmes résultats sur les triangles en survenance-inventaire, mais une gestion différente de la capitalisation qui fait augmenter légèrement la volatilité à un an.

D'autre part, si l'on choisit une modélisation s'approchant de celle faite dans les *Best estimate* utilisés pour calibrer la méthode 1 (*i.e* un modèle de charges, avec un Bornhuetter-Fergusson(6,5) en capitalisation, et des stabilisations à 18 années en répartition et 16 années en capitalisation), on obtient les résultats décrits aux tableaux 4.4.2, 4.4.3 et 4.4.4.

TABLE 4.4.2 – Résultats moyens du Bootstrap synchronisé - Totaux par Groupe de risques homogènes en répartition, modèle raffiné

	RCDC	RCDO	RCG	Tot
Ultimes $n$	26 027 757	215 544 072	108 605 467	<b>350 177 296</b>
IBNR $n$	1 501 771	30 376 915	-2 425 388	<b>29 453 299</b>
Ultimes $n + 1$	26 026 508	215 609 558	108 564 060	<b>350 200 126</b>
IBNR $n + 1$	1 500 522	30 442 401	-2 466 795	<b>29 476 128</b>

TABLE 4.4.3 – Statistiques finales du Bootstrap, modèle raffiné

	RCDC	RCG	RCDO	PSNEM	LOBs
IBNR (ou PSNEM)	1 500 522	-2 466 795	30 425 959	79 732 352	<b>109 192 038</b>
PDD	7 207 387	32 133 678	50 144 583	0	<b>89 485 648</b>
Provisions	8 707 910	29 666 882	80 570 542	79 732 352	<b>198 677 686</b>
Ecartype	703 272	2 001 715	2 802 451	3 708 236	<b>6 429 693</b>

TABLE 4.4.4 – Sigma de réserve obtenus par le bootstrap synchronisé, modèle raffiné

	Ecartype	Moyenne	Sigma de réserve
RCDC	703 272,4	8 707 910	<b>8.0762%</b>
RCDO	2 802 450,8	80 570 542	<b>3.4783%</b>
RCG	2 001 714,8	29 666 882	<b>6.7473%</b>
PSNEM	3 708 235,9	79 732 352	<b>4.6509%</b>
<b>LOB</b>	<b>6 429 693,0</b>	<b>198 677 686</b>	<b>3.2362%</b>

On observe une stabilité plus importante du *Claim development result* que ce qui est annoncé dans la méthode 1. Notons tout d'abord que les modèles sont très différents, ce qui explique ces divergences, mais également que le modèle proposé ici est appuyé sur plus de données que celui de la méthode 1 (qui n'utilise que 13 points), et est donc plus robuste.



# Conclusion

Après avoir pris le temps d'expliciter les spécificités de l'assurance construction française, son cadre théorique, pratique et réglementaire, nous avons étudié différents modèles et méthodes classiques en provisionnement non-vie, et nous avons adapté certains principes à ce cas particulier pour construire des estimateurs du sigma de réserve en normes Solvabilité II.

Dans le chapitre 2, nous avons formalisé l'angle d'analyse des modèles de provisionnement fourni par la théorie des quasi-Modèles Linéaires Généralisés. Nous avons montré que le modèle de Mack pouvait se construire comme un quasi-Modèle Linéaire Généralisé, confirmant les intuitions d'[ENGLAND et VERRALL 2002]. Nous avons ensuite discuté différents modèles estimant la variabilité à  $a$  années, en cadre univarié comme multivarié, et nous avons également explicité les estimateurs bootstrap correspondants, suivant [BOUMEZOUED et al. 2011].

Cherchant à étendre la méthode Chain-Ladder au cadre tri-dimensionnel de l'assurance construction, nous avons montré que les modèles récursifs ne pouvaient pas être étendus à ce cadre. Cependant nous avons étendu à ce cadre le modèle poissonien (avec une éventuelle modélisation jointe de la variance), sous lequel des estimateurs de variabilité à  $a$  années existent.

Puis, en exploitant la méthode de provisionnement en deux temps de l'assurance construction, nous avons développé un modèle permettant d'estimer la variabilité à  $a$  années d'un portefeuille comportant de l'assurance construction et d'autres assurances classiques, via une extension du modèle de Braun. Nous avons également construit des extensions (stochastiques) correspondants à quelques raffinements du modèle : approches Bornhuetter-Fergusson et stabilisation des coefficients de passage notamment.

Si la problématique de l'assurance construction française sous régime Solvabilité II était bien identifiée comme une mauvaise adéquation de la formule standard à la variabilité à un an, nous avons montré en estimant sous différentes hypothèses et avec différents ajustements la réalité effective de cette divergence sur un portefeuille. Nous confirmons donc, au moins empiriquement, l'intuition originale de ce mémoire selon laquelle le sigma de réserve en assurance construction est plus faible que celui proposé par la formule standard.

L'approche consistant à rééchantillonner des résidus de quasi-Modèle Linéaire Généralisé synchronisés est une approche très souple et qui permet de nombreux ajustements de la modélisation. Pour certains ajustements, par exemple un *curve-fitting* par approximation fonctionnelle, ou un Bornhuetter-Fergusson utilisant des loss-ratio moyens pondérés basés sur des primes non indépendantes, la question du bienfondé des résidus utilisés se pose et pourrait faire l'objet de prochaines analyses. Des modèles quasi-Modèle Linéaire Généralisé plus généraux, exploitants les notions de conditionnement, pourront être construits pour obtenir des résidus correspondants aux hypothèses prises.

Enfin, l'utilisation du modèle de Schnauss donnerait de meilleurs estimateurs sous problématique d'agrégation. S'il existe des estimateurs du risque à un an dans le modèle de Schnauss, les extension bootstrap ne sont, à notre connaissance, pas encore d'actualité. Une telle modélisation bootstrap permettrait une extension à la capitalisation.



# Annexes



# Annexe A Analyse des modèles de Mack de la méthode 2

Graphiques de validation du triangle de *Responsabilité civile décennale obligatoire*  
 Le triangle de *Responsabilité civile décennale obligatoire* rattaché en date réglementaire d'ouverture de chantier est celui des quatre triangles comportant le plus de volume. Ses résidus du modèle de Mack montrent une stabilité du modèle en fonction des valeurs des coefficients de passage, des résidus standardisés presque tous dans l'intervalle  $[-2, 2]$ , avec un effet de levier faible, et cependant une légère tendance. Ils sont de plus dissymétriques et *right-skewed*.

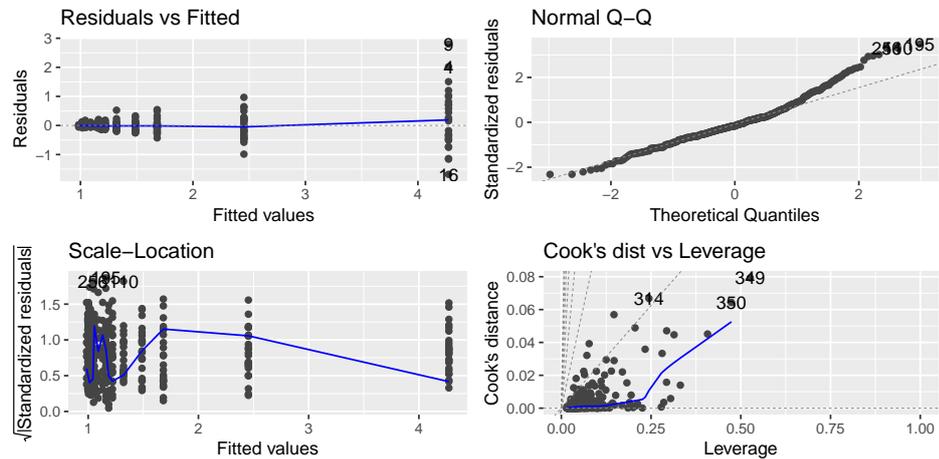


FIGURE A.0.1 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDO (Capitalisation)

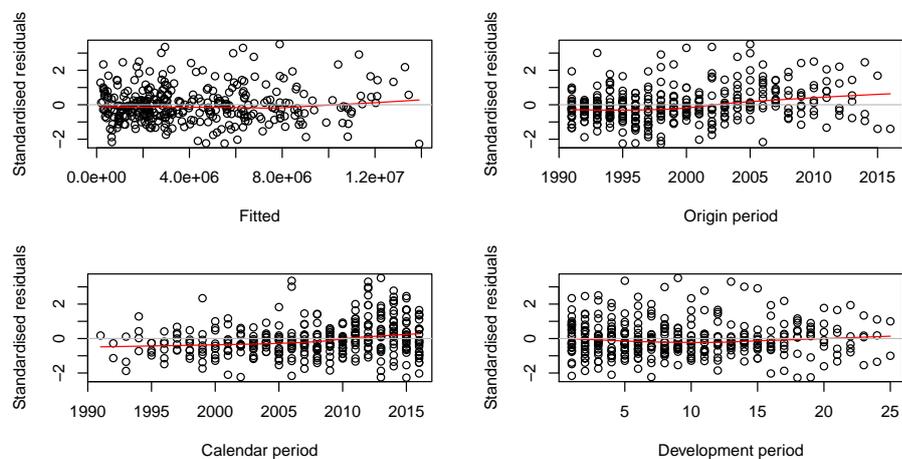


FIGURE A.0.2 – Planches de résidus du modèle linéaire pondéré de mack en RCDO

**Graphiques de validation du triangle de *Responsabilité civile générale*** Pour ce triangle, nous observons sur les graphiques d'analyse du modèle de Mack disponibles ci-dessous une stabilité des résidus du modèle en fonction des valeurs des coefficients de passage, une distance de Cook acceptable car suffisamment faible, des résidus standardisés avec un effet de levier faible (les lignes de la distance de Cook n'apparaissent même pas sur le graphique) ainsi qu'un *qqplot* acceptable sauf dans les extrêmes. Il y a toujours une tendance dans les résidus selon les années calendaires. Nous choisissons cependant d'accepter les hypothèses de Mack pour le triangle de *Responsabilité civile générale*.

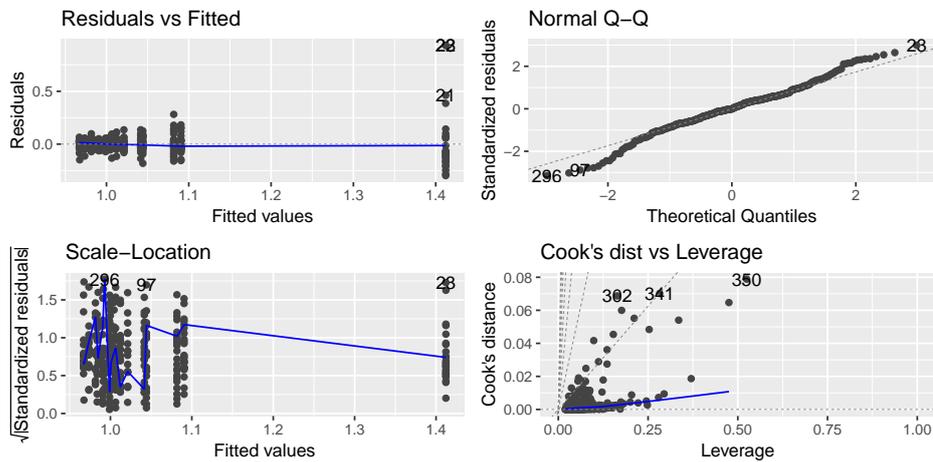


FIGURE A.0.3 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCG (Repartition)

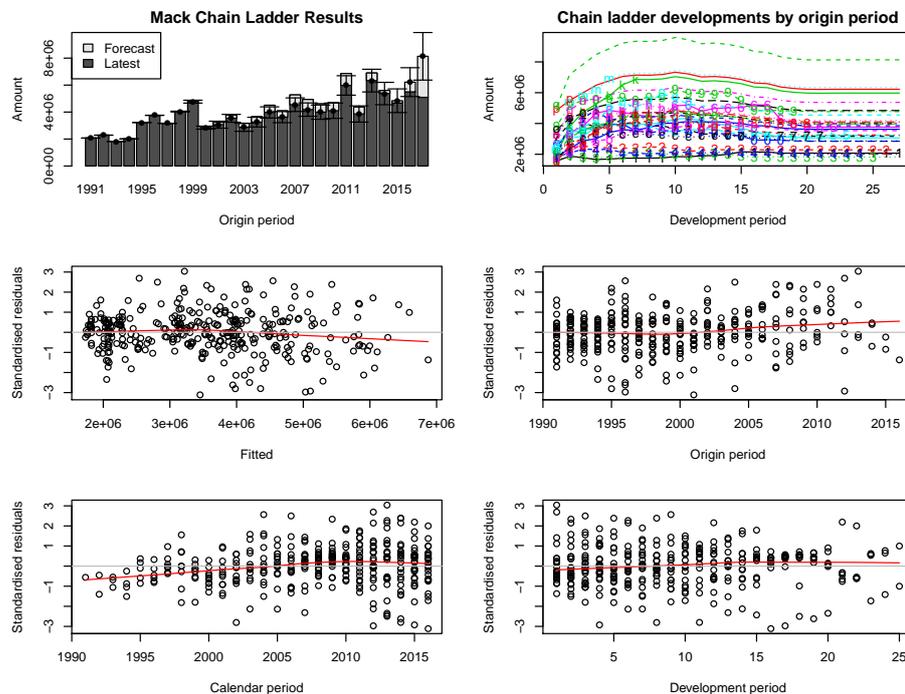


FIGURE A.0.4 – Planches de résidus du modèle linéaire pondéré de mack en RCG

Graphiques de validation du triangle de *Responsabilité civile décennale complémentaire* Pour ce triangle, nous observons sur les graphiques d'analyse du modèle de Mack des résidus très stables, des distances de Cook acceptables et se comportants bien en fonction de l'effet de levier des observations. Le seul souci possible ou élément notable est la non-normalité claire des résidus, qui sont de plus légèrement asymétriques, avec une queue lourde à droite. Ces résidus non-normaux ne disqualifient par le modèle de Merz-Wuthrich, mais il est important de noter ce point.

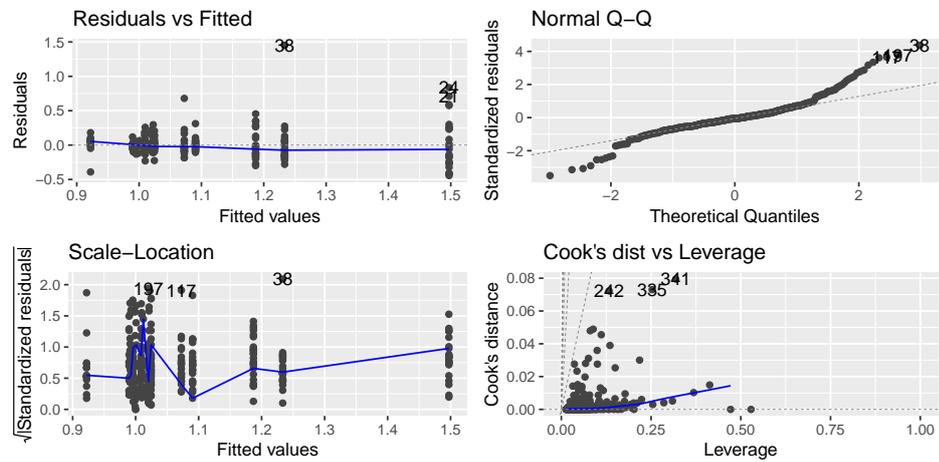


FIGURE A.0.5 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDC (Repartition)

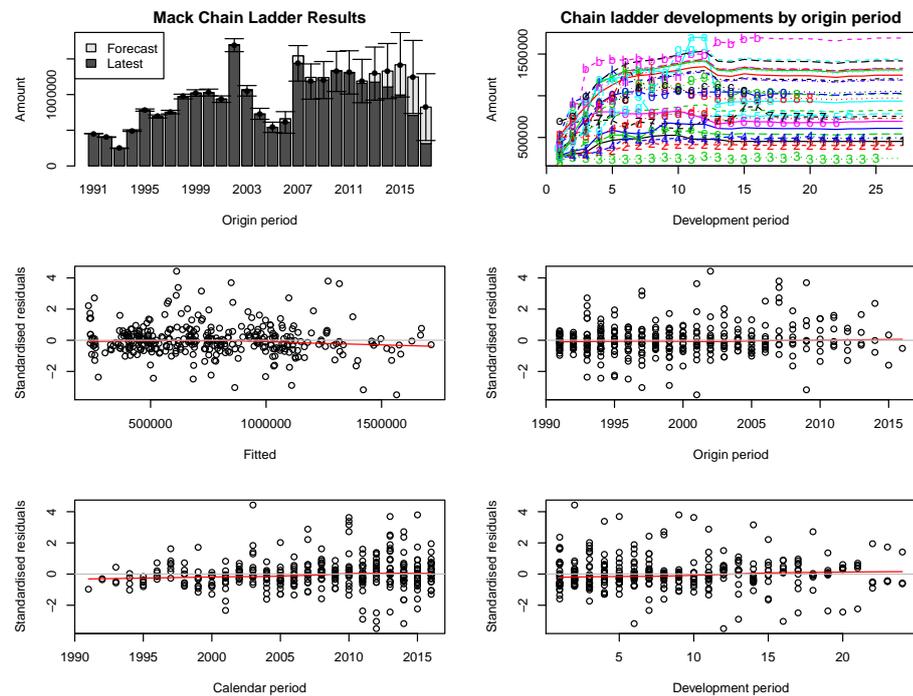


FIGURE A.0.6 – Planches de résidus du modèle linéaire pondéré de mack en RCDC

**Graphiques de validation du triangle de Répartition (Total)** L'analyse des graphiques correspondants au modèle total de répartition permet de se rendre compte que l'ajout de la *Responsabilité civile décennale complémentaire* et de la *Responsabilité civile générale* stabilise le modèle par augmentation simple des montants de charges. Ainsi, on observe alors des résidus très satisfaisants, peu dissymétriques, et sans trop d'*outliers*. Notons également leur stabilité sur les années calendaires.

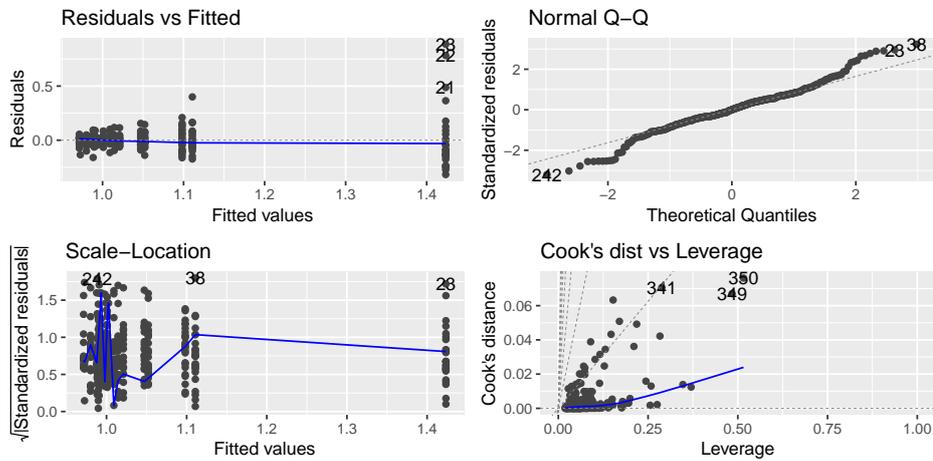


FIGURE A.0.7 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en Repartition (Total)

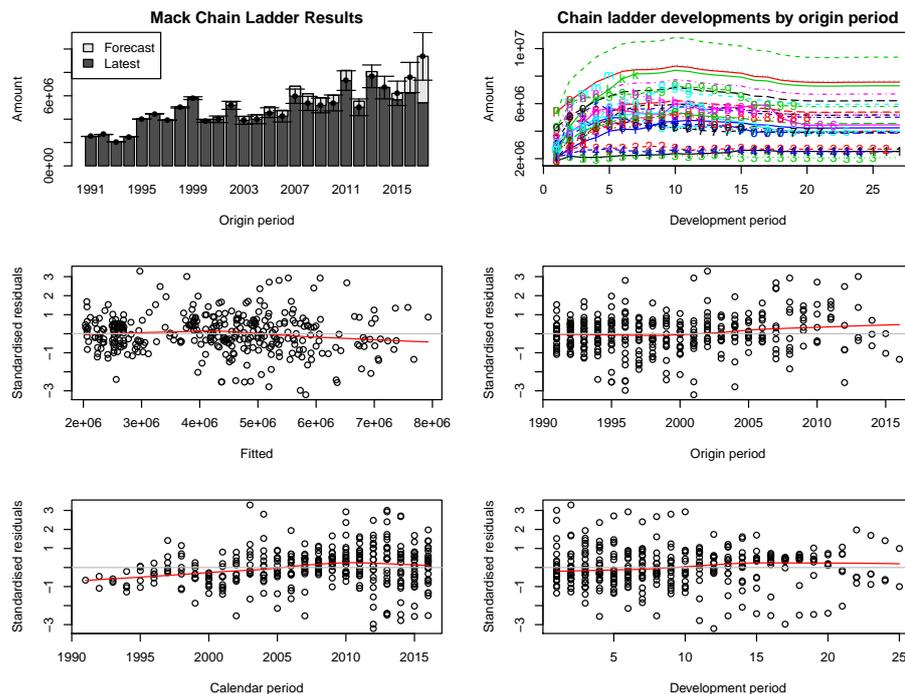


FIGURE A.0.8 – Planches de résidus du modèle linéaire pondéré de mack en répartition (au global)

# Annexe B Graphiques supplémentaire des résultats de la méthode 2

Nous proposons ici trois graphiques :

- le premier montre l'évolution des écarts-types des *Claim development result* par années d'origine sur les trois triangles. On remarque que le triangle de répartition (total) est clairement dirigée par la *Responsabilité civile générale*, ce qui est normal car son volume est plus important que celui de la *Responsabilité civile décennale complémentaire*.
- le second montre les *incurred but not reported* estimés par chain-ladder sur les différents triangles. Notons la prépondérance de la *Responsabilité civile décennale obligatoire*. Notons également le fait que les 10 derniers développements n'apportent presque pas de charges.
- enfin, le dernier propose d'observer sur un même graphique par années de développement l'ensemble des résidus du modèle. Il permet de voir l'effet "taille des triangles" dans notre estimation : nos triangles étant grands, les hypothèses sur nos résidus sont plus facilement acceptables.

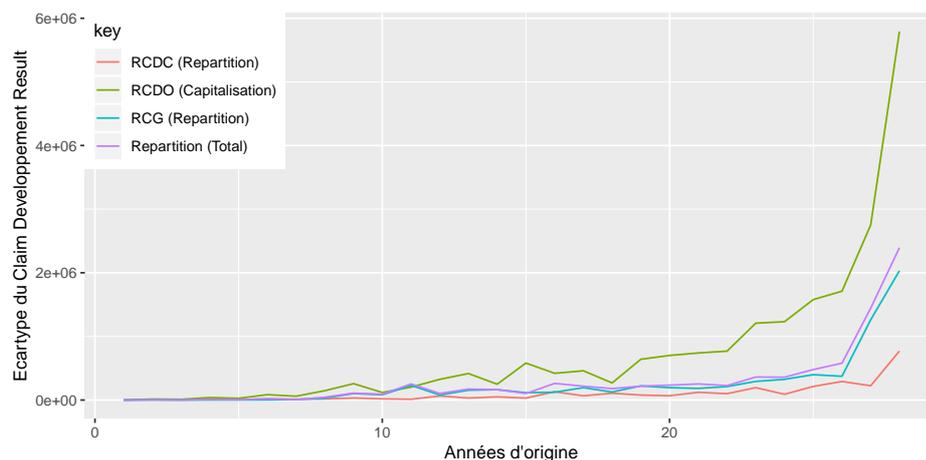


FIGURE B.0.1 – CDR estimés par Merz & Wuthrich sur les cinq triangles dont nous disposons - Par années d'origine.

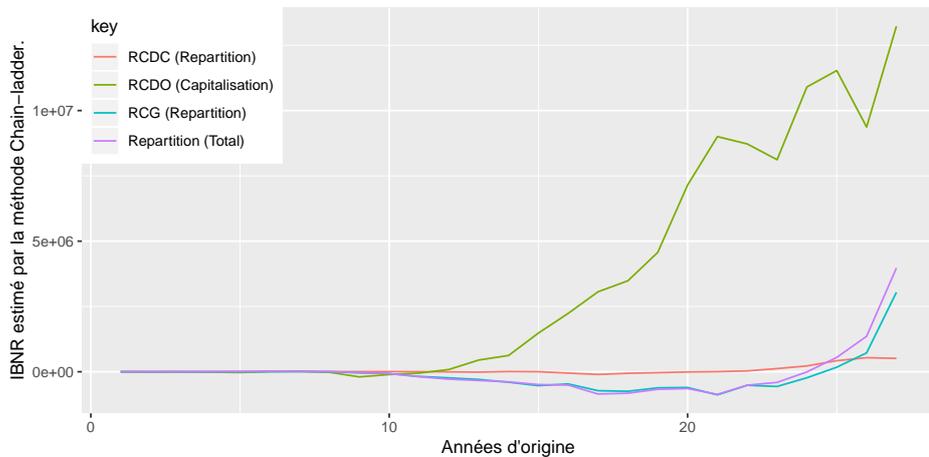


FIGURE B.0.2 – IBNR estimés par Chain-ladder sur les cinq triangles dont nous disposons - Par années d'origine.

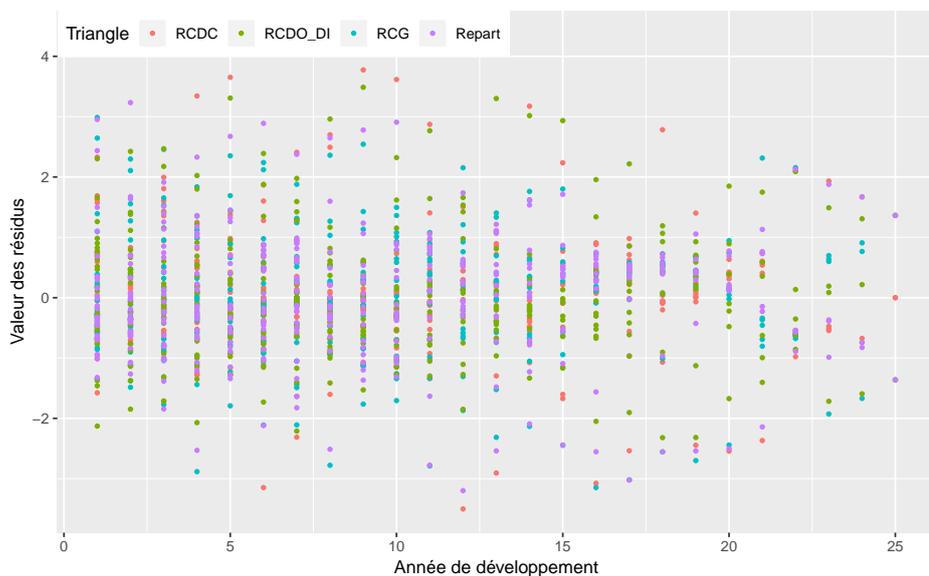


FIGURE B.0.3 – Résidus des différents modèles de Mack par années d'inventaire.

# Annexe C Analyse des modèles de Mack du bootstrap synchronisé

**Paramètres des différentes modélisations jointes** Les paramètres des différentes modélisation jointes sont disponibles à la table C.0.1. On observe facilement sur cette table que :

- les *p-value* des paramètres d'ordre 1 sont presque nulles.
- les *p-value* des paramètres d'ordre 2 sont un peu moins bonnes mais reste satisfaisantes.
- seul le modèle de la *Responsabilité civile décennale complémentaire* pêche légèrement et n'est pas entièrement valide, ce qui s'explique par les forts effets de volume.

Ces résultats concluent à un bien-fondé de notre modélisation selon les outils classiques d'analyse des Modèles Linéaires Généralisés. Nous verrons juste après que les outils plus traditionnels de validation d'un modèle de Mack contredisent légèrement ce point de vue.

TABLE C.0.1: Résultats des quasi-Modèles Linéaires Généralisés en modélisation jointe

<i>j</i>	Modèle des coefficients de passage individuels				Modèle de dispersion				
	$\hat{f}_j$	<i>s.e.</i> ( $\hat{f}_j$ )	statistic	p.value	$\hat{\phi}_j$	<i>s.e.</i> ( $\hat{\phi}_j$ )	statistic	p.value	
<b>RCDC</b>									
1	1,50	0,07	20	0	47 926	12 732	3,76	0,00	
2	1,23	0,06	19	0	53 124	12 984	4,09	0,00	
3	1,19	0,04	27	0	27 878	13 252	2,10	0,04	
4	1,09	0,02	64	0	4 773	13 537	0,35	0,72	
5	1,07	0,04	27	0	26 765	13 841	1,93	0,05	
6	1,02	0,02	50	0	7 158	14 167	0,51	0,61	
7	1,01	0,02	42	0	9 616	14 517	0,66	0,51	
8	1,02	0,01	77	0	2 667	14 894	0,18	0,86	
9	1,02	0,02	59	0	4 274	15 302	0,28	0,78	
10	1,02	0,03	40	0	8 897	15 746	0,57	0,57	
11	1,01	0,01	73	0	2 383	16 230	0,15	0,88	
12	0,92	0,04	21	0	22 797	16 762	1,36	0,17	
13	0,99	0,01	85	0	1 430	17 351	0,08	0,93	
14	1,02	0,02	59	0	2 946	18 006	0,16	0,87	
15	0,99	0,01	118	0	630	18 741	0,03	0,97	
16	1,00	0,02	66	0	1 669	19 574	0,09	0,93	
17	1,00	0,00	602	0	18	20 530	0,00	1,00	
18	1,00	0,01	195	0	143	21 640	0,01	0,99	
19	1,00	0,01	84	0	628	22 953	0,03	0,98	
20	0,99	0,01	155	0	146	24 538	0,01	1,00	
21	0,99	0,01	166	0	101	26 504	0,00	1,00	
22	1,01	0,01	98	0	232	29 033	0,01	0,99	
23	1,00	0,00	1 173	0	1	32 460	0,00	1,00	
24	1,01	0,01	104	0	97	37 482	0,00	1,00	

TABLE C.0.1: Résultats des quasi-Modèles Linéaires Généralisés en modélisation jointe extit(suite)

$j$	Modèle des coefficients de passage individuels				Modèle de dispersion			
	$\hat{f}_j$	$s.e(\hat{f}_j)$	statistic	p.value	$\hat{\phi}_j$	$s.e(\hat{\phi}_j)$	statistic	p.value
25	1,00	0,00	89 430	0	0	45 906	0,00	1,00
26	1,00	0,00	65 013	0	0	64 920	0,00	1,00
<b>RCDO</b>								
1	1,32	0,03	44	0	73 115	8 154	8,97	0,00
2	1,15	0,01	91	0	15 839	8 316	1,90	0,06
3	1,12	0,01	82	0	19 605	8 487	2,31	0,02
4	1,10	0,01	88	0	16 954	8 670	1,96	0,05
5	1,05	0,01	89	0	15 546	8 865	1,75	0,08
6	1,05	0,01	86	0	16 061	9 073	1,77	0,08
7	1,03	0,01	162	0	4 195	9 297	0,45	0,65
8	1,03	0,01	166	0	3 755	9 539	0,39	0,69
9	1,00	0,01	154	0	3 857	9 800	0,39	0,69
10	1,01	0,01	147	0	3 903	10 084	0,39	0,70
11	0,99	0,00	212	0	1 625	10 395	0,16	0,88
12	1,01	0,01	78	0	11 371	10 736	1,06	0,29
13	1,00	0,01	175	0	1 977	11 113	0,18	0,86
14	1,00	0,00	281	0	707	11 532	0,06	0,95
15	1,02	0,02	64	0	12 063	12 003	1,00	0,32
16	1,00	0,00	296	0	487	12 537	0,04	0,97
17	1,01	0,01	142	0	1 764	13 149	0,13	0,89
18	1,00	0,00	719	0	59	13 860	0,00	1,00
19	1,00	0,00	676	0	56	14 701	0,00	1,00
20	1,00	0,00	379	0	137	15 716	0,01	0,99
21	1,00	0,00	305	0	177	16 975	0,01	0,99
22	1,00	0,00	751	0	26	18 595	0,00	1,00
23	1,00	0,00	7 881	0	0	20 790	0,00	1,00
24	1,00	0,00	1 049	0	9	24 006	0,00	1,00
25	1,01	0,00	225	0	142	29 401	0,00	1,00
26	1,00	0,01	143	0	0	41 579	0,00	1,00
<b>RCG</b>								
1	1,41	0,06	23	0	214 993	26 369	8,15	0,00
2	1,09	0,02	68	0	19 982	26 891	0,74	0,46
3	1,08	0,02	53	0	32 731	27 446	1,19	0,23
4	1,04	0,02	68	0	19 156	28 036	0,68	0,49
5	1,04	0,01	81	0	13 017	28 666	0,45	0,65
6	1,01	0,01	83	0	11 584	29 341	0,39	0,69
7	1,00	0,01	123	0	4 702	30 065	0,16	0,88
8	1,01	0,01	87	0	8 888	30 846	0,29	0,77
9	1,02	0,01	72	0	12 604	31 692	0,40	0,69
10	0,98	0,01	134	0	3 249	32 610	0,10	0,92
11	0,97	0,01	83	0	7 419	33 614	0,22	0,83
12	1,00	0,01	115	0	3 702	34 716	0,11	0,92
13	0,99	0,01	126	0	2 767	35 935	0,08	0,94
14	0,98	0,01	74	0	7 289	37 291	0,20	0,85
15	0,97	0,01	70	0	7 094	38 814	0,18	0,86
16	0,99	0,01	157	0	1 304	40 540	0,03	0,97
17	0,97	0,02	44	0	14 046	42 519	0,33	0,74
18	0,98	0,01	116	0	1 858	44 819	0,04	0,97
19	0,99	0,01	115	0	1 566	47 537	0,03	0,97
20	1,00	0,00	338	0	149	50 820	0,00	1,00
21	1,00	0,00	1 056	0	13	54 891	0,00	1,00
22	1,00	0,00	2 961	0	1	60 130	0,00	1,00
23	1,00	0,00	27 914	0	0	67 228	0,00	1,00
24	1,00	0,00	14 003	0	0	77 628	0,00	1,00
25	1,00	0,00	81 592	0	0	95 075	0,00	1,00
26	1,00	0,00	56 246	0	0	134 456	0,00	1,00

**Graphiques de validation du triangle de *Responsabilité civile générale*** En numérotant les 10 graphiques dans le sens de la lecture (de gauche à droite puis de haut en bas), on observe :

- des résidus stables sur les valeurs ajustées des coefficients de passage, ainsi sur les années de développement aux graphiques 1 et 8.
- des résidus non stables sur les années calendaires, et plus généralement sur les délais de développement aux graphiques 9 et 10.
- un *qqplot* rejetant leur normalité de manière catégorique.

Malgré une légère tendance dans les résidus, nous acceptons le modèle. Il faudra tenir compte de cette tendance dans nos analyses.

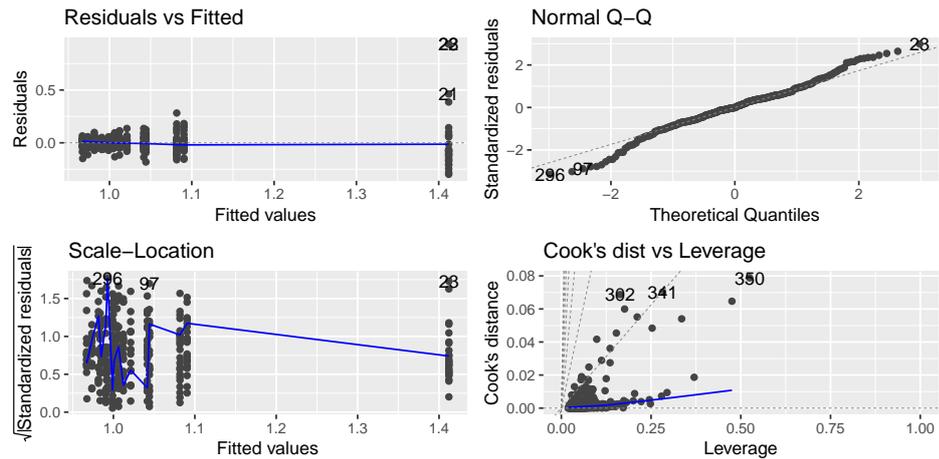


FIGURE C.0.1 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCG

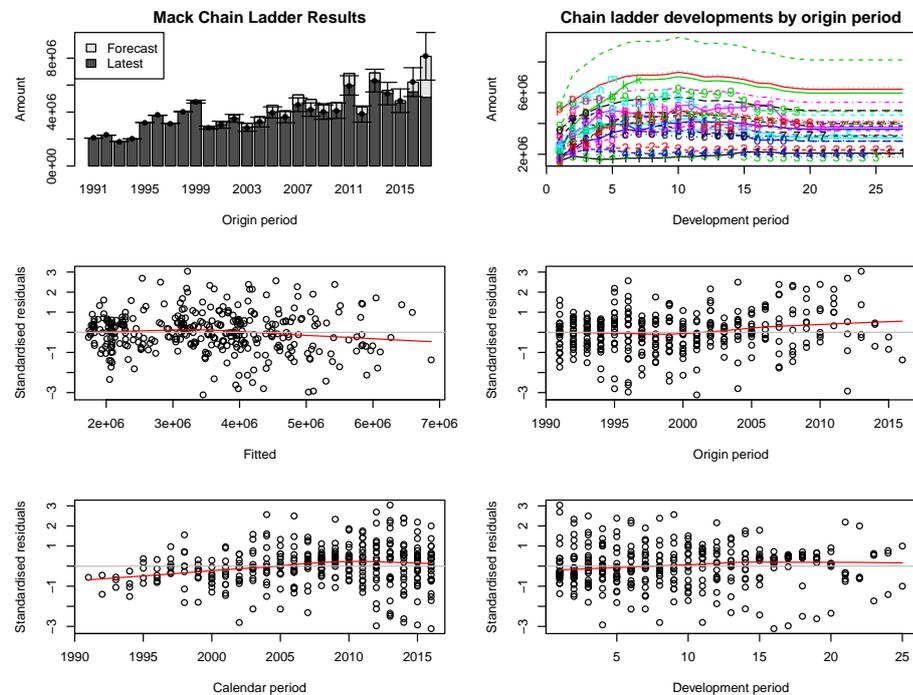


FIGURE C.0.2 – Planches de résidus du modèle linéaire pondéré de mack en RCG

Graphiques de validation du triangle de *Responsabilité civile décennale complémentaire* Les résidus de ce modèle sont bons, n'ont pas de tendances que ce soit sur les valeurs fittées, sur les années calendaires d'origine ou de développement. Nous notons cependant une forte asymétrie et anormalité de ces résidus, qui ne contredisent pas nos hypothèses non-paramétriques de quasi-Modèle Linéaire Généralisé. Nous choisissons d'accepter les hypothèses de Mack sur ce triangle.

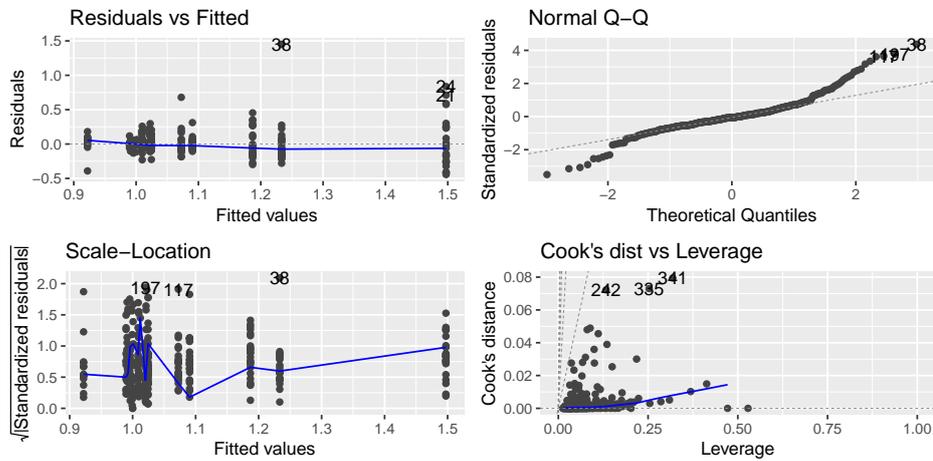


FIGURE C.0.3 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDC

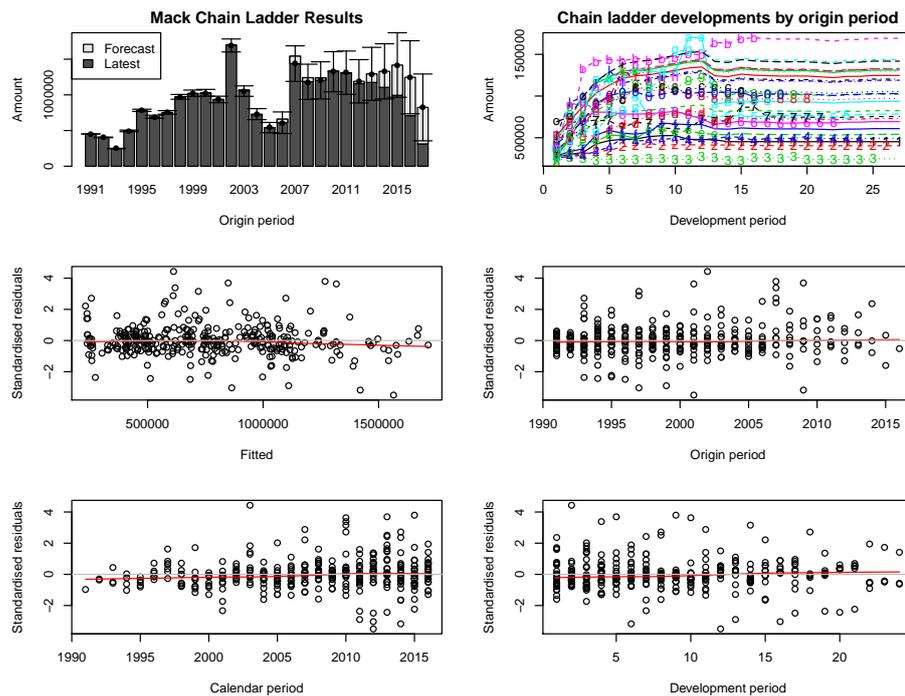


FIGURE C.0.4 – Planches de résidus du modèle linéaire pondéré de mack en RCDC

Graphiques de validation du triangle de *Responsabilité civile décennale obligatoire en répartition*. À part quelques valeurs influentes (structurellement influentes, de par la forme en triangle des données), les résidus de ce modèle ne portent pas de tendances visibles. Le modèle est donc accepté. Nous notons une de fois de plus la contradiction avec des hypothèses normales.

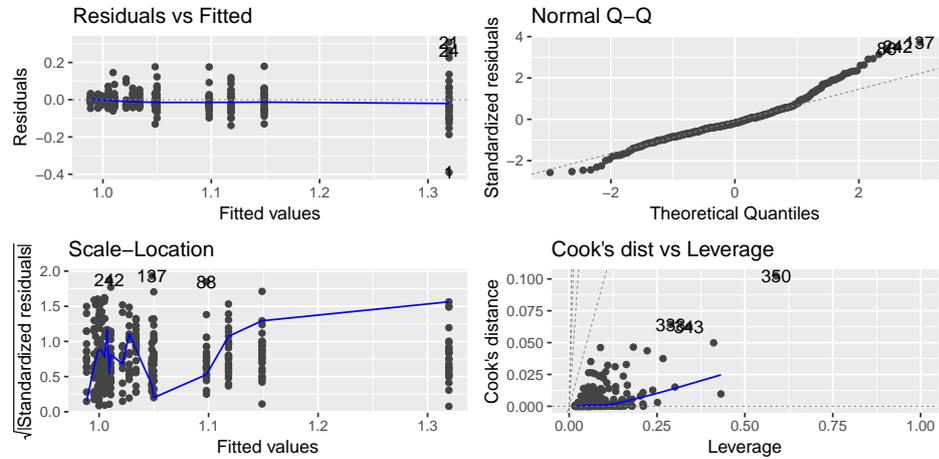


FIGURE C.0.5 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDO

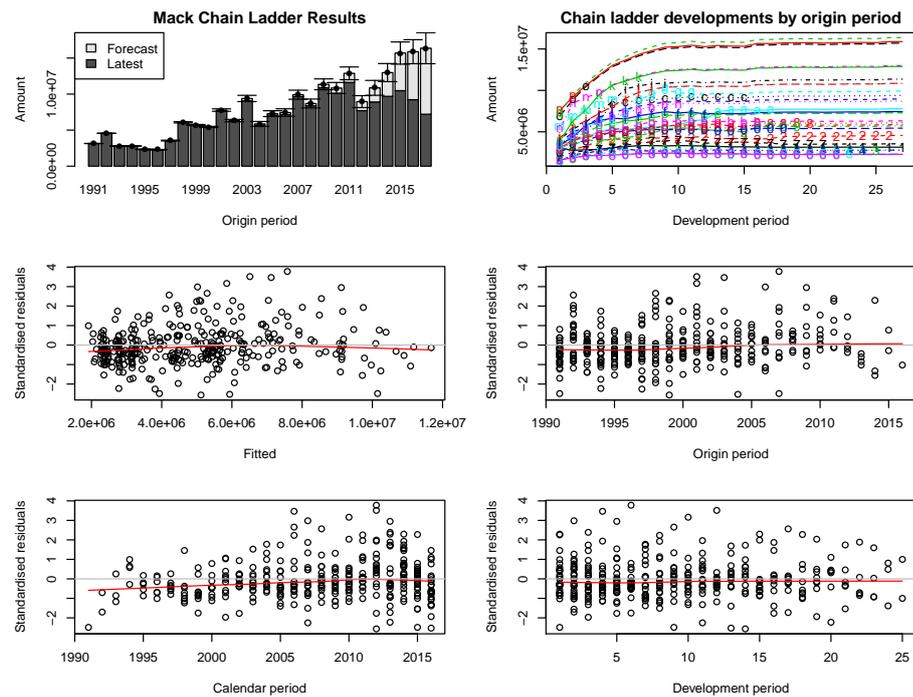


FIGURE C.0.6 – Planches de résidus du modèle linéaire pondéré de mack en RCDO (répartition)

Graphiques de validation du triangle de *Responsabilité civile décennale obligatoire en capitalisation (à l'ultime)* Ce triangle est spécial car passé à l'ultime et donc *conditionnel* à la modélisation précédente du triangle en répartition. Cependant les hypothèses du modèle que nous faisons ici sont les mêmes que précédemment. Nous notons une tendance des résidus à l'augmentation sur les périodes de rattachement, ce qui montre des *trend-test* positifs sur les dates de rattachement indiquant un (possible) allongement des cadences de manifestation au cours des dates règlementaires d'ouverture de chantier. Cependant l'effet reste négligeable et nous acceptons donc les hypothèses du quasi-Modèle Linéaire Généralisé de Mack sur ce triangle.

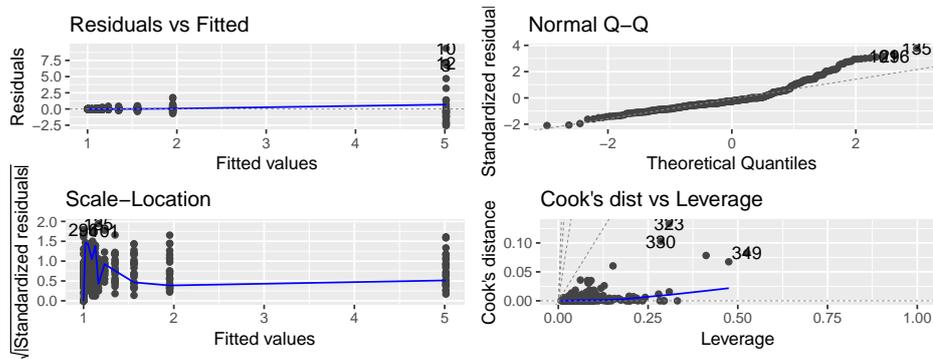


FIGURE C.0.7 – Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDO en capitalisation (à l'ultime)

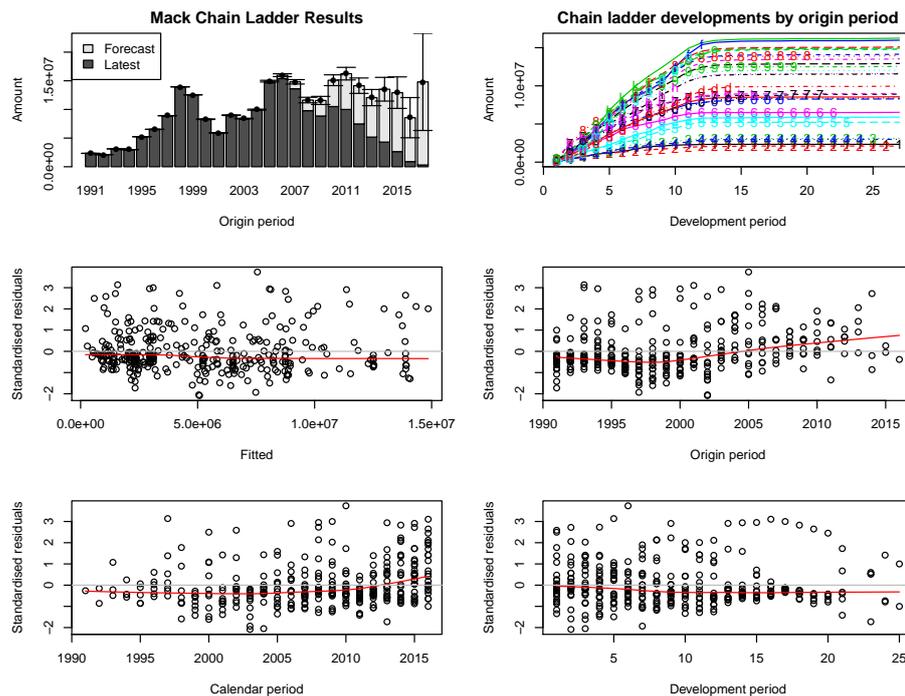


FIGURE C.0.8 – Planches de résidus du modèle linéaire pondéré de mack en RCDO en capitalisation (à l'ultime)

# Bibliographie

- ADAMS, Karen HS (2017). « Chain Ladder Reserving Methods for Liabilities with Per Occurrence Limits ». In : *Casualty Actuarial Society E-Forum, Fall 2013*, p. 112.
- ALAI, D. H., M. MERZ et M. V. WÜTHRICH (2009). « Mean Square Error of Prediction Un the Bornhuetter-Fergusson Claims Reserving Method ». en. In : *Annals of Actuarial Science* 4.01, p. 7–31. ISSN : 1748-4995, 1748-5002. DOI : 10/bf4n97.
- (2011). « Prediction Uncertainty in the Bornhuetter-Ferguson Claims Reserving Method: Revisited ». en. In : *Annals of Actuarial Science* 5.01, p. 7–17. ISSN : 1748-4995, 1748-5002. DOI : 10/b3vpq9.
- APPERT-RAULLIN, Yannick et al. (2013). « One-Year Volatility of Reserve Risk in a Multivariate Framework ». en. In : p. 35.
- AUTORITÉ-DES-NORMES-COMPTABLES (2015). *REGLEMENT N° 2015-11 Du 26 Novembre 2015 Relatif Aux Comptes Annuels Des Entreprises d'assurance*.
- (2016). *RÈGLEMENT N° 2016-12 Du 12 Décembre 2016 Modifiant Le Règlement ANC N° 2015-11 Du 26 Novembre 2015 Relatif Aux Comptes Annuels Des Entreprises d'assurance*.
- BEAUNE, Lucile (2016). « Estimation des paramètres de volatilité propres aux assureurs non-vie ». fr. Mémoire d'Actuariat.
- BENJAMIN, S. et L. M. EAGLES (1997). « A Curve Fitting Method and a Regression Method ». In :
- BORNHUETTER, R L et R E FERGUSSON (1972). « The Actuary and IBNR ». en. In : p. 15.
- BOUGNON, Nathalie, Guillaume BARRUEL et Nicolas ZEC (2016). « Pilotage du risque de souscription non vie sous Solvabilité 2 ». fr. In : p. 150.
- BOUMEZOUE, Alexandre et al. (2011). « One-Year Reserve Risk Including a Tail Factor: Closed Formula and Bootstrap Approaches ». In : *arXiv preprint arXiv:1107.0164*.
- BRAUN, Christian (2004). « The Prediction Error of the Chain Ladder Method Applied to Correlated Run-Off Triangles ». en. In : p. 19.
- BUCHWALDER, Markus et al. (2006). « The Mean Square Error of Prediction in the Chain Ladder Reserving Method (Mack and Murphy Revisited) ». en. In : *ASTIN Bulletin* 36.02, p. 521–542. ISSN : 0515-0361, 1783-1350. DOI : 10/b74pwm.
- BUSSON, Etienne (2012). « Evaluation Du Risque de Provisionnement à 1 an : Adaptation de La Méthode de Merz & Wutrich à Des Cas Non Standards ». Mémoire d'Actuariat.
- CAMBON, Adrien (2010). « Elaboration d'un Modèle Interne Partiel Concernant Le Risque de Souscription Non-Vie Pour Tenir Compte Es Spécificités d'une Société Spécialisée Dans Les Branches Longues ». Mémoire d'Actuariat. ISUP et SMABTP.
- CARRATO, Alessandro (2016). « A Practitioner's Introduction to Stochastic Reserving ». en. In : p. 75.
- CASTRO, Ana Catarina Teixeira (2016). « Valuation of Non-Life Claims Provision and the Capital Requirement for Reserve Risk ». PhD Thesis. Instituto Superior de Economia e Gestão.
- CHARPENTIER, Arthur et Michel DENUIT (2004). « Mathématiques de l'assurance Non-Vie ». In : *Economica*.
- COLLET, Ischia (2013). « Run off En Construction ». Mémoire d'Actuariat.

- COMMISSION-EUROPÉENNE (2014). *Règlement Délégué (UE) 2015/25 de La Commission u 10/10/2014*.
- CRIMINOCORPUS (2014). *Législation Historique Sur l'assurance Construction*. Criminocorpus.
- DEGIOVANNI, Etienne (2014). « Modélisation Du Risque de Provision Engendré Par Les Sinistres Graves de l'assurance Construction ». Mémoire d'Actuariat. I.S.F.A et ACTUARIS.
- DEPLANTE, Brice (2010). « Provisionnement et Tarrification En Dommage Ouvrage ». Mémoire d'Actuariat.
- DIERS, Dorothea (2010). « Stochastic Re-Reserving in Multi-Year Internal Models - An Approach Based on Simulations ». In :
- DOMINIQUE CHARPIN (2003). *Hammu-Rabi de Babylone*. Presses universitaires de France.
- EFRON, B (1978). « Computers and the Theory of Statistics : Thinking the Unthinkable ». In :
- EFRON, B. (1979). « Bootstrap Methods : Another Look at the Jackknife ». In : DOI : 10.1214/aos/1176344552.
- EFRON, Bradley (1980). « Nonparametric Estimates of Standard Error : The Kackknife, the Bootstrap, and Other Methods ». In :
- (1994). *An Introduction to the Bootstrap*.
- ENGLAND, P. D. et R. J. VERRALL (2006). « Predictive Distributions of Outstanding Liabilities in General Insurance ». en. In : *Annals of Actuarial Science* 1.02, p. 221–270. ISSN : 1748-4995, 1748-5002. DOI : 10/bfczvn.
- ENGLAND, P.D. et R.J. VERRALL (2002). « Stochastic Claims Reserving in General Insurance ». en. In : *British Actuarial Journal* 8.03, p. 443–518. ISSN : 1357-3217, 2044-0456. DOI : 10/csqxh7.
- ENGLAND, PETER et RICHARD VERRALL (1998). « STANDARD ERRORS OF PREDICTION IN CLAIMS RESERVING: A COMPARISON OF METHODS WORKSHOP ». en. In : p. 20.
- ENRIQUE DE ALBA (2002). « Claims Reserving When There Are Negative Values in the Development Triangle ». en. In : p. 16.
- ENRIQUE DE ALBA et MARCO A. RAMÍREZ CORZO (2005). « Bayesian Claims Reserving When There Are Negative Values in the RunOff Triangle ». In :
- FABIEN BESSEYRE (2015). « Le Calibrage Des Undertaking Specific Parameters En Prévoyance ». Mémoire d'Actuariat.
- GAUVILLE, Rémi (2017). « Projection Du Ratio de Solvabilité : Des Méthodes de Machine Learning Pour Contourner Les Contraintes Opérationnelles de La Méthode Des SdS ». Mémoire d'Actuariat.
- GESMANN, Markus (2009). « London R-User Group Meeting Claims Reserving in R The ChainLadder Package ». en. In : p. 35.
- (2017). « Generalized Linear Models in R ». en. In : p. 42.
- GOULD, Ingrid Langfeldt (2008). « Stochastic Chain-Ladder Models in Nonlife Insurance ». In : p. 114.
- HAEGEL, Thomas (2009). « Solvabilité II - Impact de La Nouvelle Directive Sur Une Société d'assurance Non-Vie Spécialisée Dans l'assurance Construction : Application et Calibrage de La Formule Standard ». Mémoire d'Actuariat.
- HAHN, Lukas (2017). « Multi-Year Non-Life Insurance Risk of Dependent Lines of Business in the Multivariate Additive Loss Reserving Model ». en. In : *Insurance: Mathematics and Economics* 75, p. 71–81. ISSN : 01676687. DOI : 10/gbsbbs.
- HAMMURABI (2011). *Traduction En Français Du Code d'Hammurabi*.
- HAZAN, Laura (2011). « Mise En Place d'un Modèle Interne Partiel Portant Sur Le Risque de Souscription Non Vie Des Garanties Décennales (Assurance Construction). » Mémoire d'Actuariat. Dauphine et SMABTP.

- HEBERLE, Jochen (2014). « Verteilungsfreie Verfahren in der Schadenreservierung ». de. In : p. 156.
- HEBERLE, Jochen, Luis HUERGO et Michael MERZ (2008). « Measuring the One-Year Reserve Risk Uncertainty in Modern Solvency Tests for Several Correlated Lines of Business ». en. In : p. 32.
- HUERGO, Luis, Jochen HEBERLE et Michael MERZ (2010). « Bootstrapping the Chain-Ladder Method for Several Correlated Run-off Portfolios ». en. In : *Zeitschrift für die gesamte Versicherungswissenschaft* 98.5, p. 541–564. ISSN : 0044-2585, 1865-9748. DOI : 10.1007/S12297-009-0078-2.
- JOCHEN HEBERLE, LUIS HUERGO et MICHAEL MERZ (2011). « Bootstrapping Methoden Zur Bestimmung Des Einjährigen Reserverisikos Für Solvenz betrachtungen Bei Mehreren Abhängigen Run-Off-Portfolios ». In :
- LAGNAOUI, Majid Joseph (2015). « Machine Learning et Nouvelle Classe de Problèmes Actuariels : Une Illustration Par l'étude de Cas ». Mémoire d'Actuariat.
- LE, Phi-hung et Gaël VIROT (2013). « Méthode d'évaluation Stochastique Des Provisions Techniques d'une Entreprise d'assurance Dommages Par Une Approche Ligne à Ligne ». Mémoire d'Actuariat.
- LENFANT, Maxime (2010). « Variabilité Des Provisions et Pertinence Des Données : Quelle Méthodologies ? » Mémoire d'Actuariat.
- LINDA, Bohdan et Jana KUBANOVA (2012). « Bootstrap Application of the Bornhuetter-Ferguson Method ». en. In : p. 6.
- LINDE, Marc (2016). « Multi-Year Non-Life Insurance Risk for Correlated Loss Portfolios Under Chain Ladder Model Assumptions ». en. In : *SSRN Electronic Journal*. ISSN : 1556-5068. DOI : 10/gdqbg8.
- LO, Joseph (2011). « Hypothesis Testing and Resampling Techniques ». en. In : p. 51.
- LUKAS J. HAHN (2017). *Multi-Year Non-Life Insurance Risk for Dependant Lines of Business*.
- MACK, Thomas (1991). « A Simple Parametric Model for Rating Automobile Insurance or Estimating IBNR Claims Reserves ». en. In : *ASTIN Bulletin* 21.01, p. 93–109. ISSN : 0515-0361, 1783-1350. DOI : 10.2143/AST.21.1.2005403.
- (1993). « Distribution-Free Calculation of the Standard Error of Chain Ladder Reserve Estimates ». In : *Astin bulletin* 23.2, p. 213–225.
- (1994). « Wich Stochastic Model Is Underlying the Chain Ladder Method ? » In :
- (1999). « The Standard Error of Chain Ladder Reserve Estimates: Recursive Calculation and Inclusion of a Tail Factor ». en. In : *ASTIN Bulletin* 29.02, p. 361–366. ISSN : 0515-0361, 1783-1350. DOI : 10/c7nvcz.
- (2008). « The Prediction Error of Bornhuetter-Ferguson ». en. In : p. 21.
- MACK, Thomas et Munich RE (1994). « Measuring the Variability of Chain Ladder Reserve Estimates ». In : p. 82.
- MACK, Thomas et Gary VENTER (2000). « A Comparison of Stochastic Models That Reproduce Chain Ladder Reserve Estimates ». In : *Insurance: mathematics and economics* 26.1, p. 101–107. DOI : 10.1016/S0167-6687(99)00039-6.
- MAXIME LENFANT (2010). *Variabilité Des Provisions et Pertinence Des Données : Quelle Méthodologie ?* DUAS - UDS.
- MCCULLAGH, P et J.A NELDER (1989). *Generalized Linear Models*.
- MERZ, M. et M. V. WÜTHRICH (2007a). « Prediction Error of the Chain Ladder Reserving Method Applied to Correlated Run-off Triangles ». en. In : *Annals of Actuarial Science* 2.01, p. 25–50. ISSN : 1748-4995, 1748-5002. DOI : 10.1017/S1748499500000245.
- MERZ, Michael et Mario V. WÜTHRICH (2007b). « Prediction Error of the Chain Ladder Reserving Method Applied to Correlated Run-off Triangles ». In : *Annals of Actuarial Science* 2.1, p. 25–50. DOI : 10.1017/S1748499500000245.
- (2008a). « Modeling the Claims Development Result for Solvency Purposes ». In :

- MERZ, Michael et Mario V. WÜTHRICH (2008b). « Prediction Error of the Multivariate Chain Ladder Reserving Method ». en. In : *North American Actuarial Journal* 12.2, p. 175–197. ISSN : 1092-0277, 2325-0453. DOI : 10/gdsk3b.
- MICHAUD, Clémence (2012). « La Mesure Du Risque de Provisionnement à Horizon Un an et Assurance Non-Vie ». Mémoire d'Actuariat.
- NATAF, Ruben (2014). « Les Méthodes de Provisionnement En Assurance de Construction ». Mémoire d'Actuariat. ISUP et MAZARS ACTUARIAT.
- PARLEMENT-EUROPÉEN et CONSEIL-DE-L'EUROPE (2009). *Directive Solvabilité 2 2009/138/CE Du Parlement Européen et Du Conseil*.
- PINHEIRO, PAULO J. R., ANDRADE E SILVA, JOÃO M. et CENTENO, MARIA DE LOURDES (2001). « Bootstrap Methodology in Claim Reserving ». In :
- POUDRET, Benjamin et Lucile BEAUNE (2016). « Solv2 : Les Undertaking Specific Parameters En IARD ». In : *INFOTECH* 39.
- PROHL, Carsten et Klaus D SCHMIDT (2005). « Multivariate Chain–Ladder ». en. In : p. 15.
- ROOS, Mordehai (2015). « Utilisation des Undertaking Specific Parameters avec une base de données santé ». fr. In : p. 106.
- ROUCHATI, Faries (2016). « Projection Du Ratio de Solvabilité d'un Assureur Retraite : Les Méthodes Paramétriques (Curve Fitting et Least Square Monte Carlo) Peuvent-Elles Se Substituer à La Méthode Des Simulations Dans Les Simulations ? » Mémoire d'Actuariat.
- SALUZ, Annina (2014). « Stochastic Loss Reserving with Emphasis on the Bornhuetter-Ferguson Method ». en. Thèse de doct. ETH Zurich. DOI : 10.3929/ethz-a-010358837.
- SCARTH, Robert (2015). « A Stochastic Bornhuetter- Ferguson Model ». en. In : p. 24.
- SCHMIDT, Klaus D (2008). « Bornhuetter–Ferguson as a General Principle of Loss Reserving ». en. In : p. 14.
- SCHMIDT, Klaus D. et Anja SCHNAUS (1996). « An Extension of Mack's Model for the Chain Ladder Method ». en. In : *ASTIN Bulletin* 26.02, p. 247–262. ISSN : 0515-0361, 1783-1350. DOI : 10/bqmr93.
- SCHMIDT, Klaus D. et Angela WÜNSCHE (1998). « Chain Ladder, Marginal Sum and Maximum Likelihood Estimation ». In : *Blätter der DGVFM* 23.3, p. 267–277. DOI : 10.1007/BF02808289.
- SHAPIRO, S S et M B WILK (1965). « An Analysis of Variance Test for Normality (Complete Samples) ». en. In : p. 22.
- TAYLOR, Gregory Clive (2000). *Loss Reserving: An Actuarial Perspective*. en. OCLC: 878850166. New York : Springer Science+Business Media. ISBN : 978-1-4613-7070-3.
- TEE, Liivika, Meelis KÄÄRIK et Rauno VIIN (2017). « On Comparison of Stochastic Reserving Methods with Bootstrapping ». en. In : *Risks* 5.1, p. 2. ISSN : 2227-9091. DOI : 10/gdqbg9.
- VERRALL, R. J. et P. D. ENGLAND (2000). « Comments on:“A Comparison of Stochastic Models That Reproduce Chain Ladder Reserve Estimates”, by Mack and Venter ». In : *Insurance: mathematics and economics* 26.1, p. 109–111. DOI : 10.1016/S0167-6687(99)00040-2.
- VERRALL, Richard J. (2000). « An Investigation into Stochastic Claims Reserving Models and the Chain-Ladder Technique ». In : *Insurance: mathematics and economics* 26.1, p. 91–99. DOI : 10.1016/S0167-6687(99)00038-4.
- WEDDERBURN, Robert WM (1974). « Quasi-Likelihood Functions, Generalized Linear Models, and the Gauss—Newton Method ». In : *Biometrika* 61.3, p. 439–447.
- WÜTHRICH, Mario V., Michael MERZ et Natalia LYSENKO (2009). « Uncertainty of the Claims Development Result in the Chain Ladder Method ». en. In : *Scandinavian Actuarial Journal* 2009.1, p. 63–84. ISSN : 0346-1238, 1651-2030. DOI : 10/bwzfgw.
- XIA, Michelle et David P. M. SCOLLNIK (2015). « A Bayesian Mixture Model Accounting for Zeros and Negatives in the Loss Triangle ». en. In : *International Journal of Statistics and Probability* 4.2. ISSN : 1927-7040, 1927-7032. DOI : 10/gdqbhc.
- ZBOŇ, Lenka (2014). « Generalized Linear Models in Reserving Risk ». en. In : p. 62.

# Table des figures

1.1.1	Photo de la stelle du code de Hammurabi, visible au musée du Louvre. Par Mbzt, CC BY 3.0 . . . . .	4
1.1.2	Principaux intervenants de l'assurance construction [CAMBON 2010] . . . . .	6
1.1.3	Fonctionnement en double détente via la convention CRAC [COLLET 2013] . . . . .	8
1.3.1	Schéma d'agrégation des risques unitaires en sous-modules et modules de risque pour le calcul du <i>Solvency Capital Requirement</i> . . . . .	13
2.5.1	Zones sensibles pour une estimation via <i>Chain-ladder</i> . . . . .	36
2.5.2	Cadences de développements brutes et ajustées . . . . .	38
3.1.1	Représentation de la pyramide de données disponibles en assurance construction. <i>En bas à droite, les projections nécessaires pour compléter le cube et calculer les deux provisions</i> [DEPLANTE 2010]. . . . .	42
4.1.1	Quantile-Quantile plot log-normal des $y_t$ . . . . .	56
4.1.2	Test de modélisation d'un bruit blanc fort sur les résidus . . . . .	57
4.1.3	Graphique de la relation linéaire entre $y_t$ et $x_t$ (l'intervalle de confiance est calculé sous les hypothèses du modèle à $\alpha = 5\%$ ) . . . . .	57
4.1.4	Valeurs obtenues du sigma de réserve global en fonction du coefficient de corrélation . . . . .	61
4.2.1	Schéma synthétique du bootstrap effectué . . . . .	64
4.2.2	Résidus observés par années calendaires, par délais de développements et par années de ratachements sur les triangles . . . . .	65
4.2.3	Paramètres de corélations de Braun des triangles en survenance-inventaire par délai de développement . . . . .	66
4.2.4	Burn-in des corélations des IBNR à un an, de leurs variances et des sigmas de réserves correspondants . . . . .	67
4.2.5	Facteurs de développement obtenus par le bootstrap synchronisé des Groupes de risques homogènes, partie répartition . . . . .	68
4.2.6	Densité des IBNR à un an, comparées à leurs valeurs déterministes cette année. . . . .	69
4.2.7	Comparaison des CDR obtenus par année d'origine sur les triangles et des formules fermées de Merz-Wuthich . . . . .	70
4.3.1	Boxplots des cadences de manifestation bootstrapées de la RCDO - Zonées selon les natures de risques. . . . .	71
4.3.2	Densité des résidus des modèles de Mack, avec (en noir) une loi normale centrée réduite . . . . .	73
4.3.3	Quantile-Quantile plot des résidus contre une loi normale centrée réduite . . . . .	73
4.4.1	Sigma de réserve moyen obtenu, classification selon les paramétrages les plus influents . . . . .	75
A.0.1	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDO (Capitalisation) . . . . .	83

A.0.2	Planches de résidus du modèle linéaire pondéré de mack en RCDO . . . . .	83
A.0.3	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCG (Repartition) . . . . .	84
A.0.4	Planches de résidus du modèle linéaire pondéré de mack en RCG . . . . .	84
A.0.5	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDC (Repartition) . . . . .	85
A.0.6	Planches de résidus du modèle linéaire pondéré de mack en RCDC . . . . .	85
A.0.7	Planches d'analyse des résidus du modèle linéaire pondéré de mack en Re- partition (Total) . . . . .	86
A.0.8	Planches de résidus du modèle linéaire pondéré de mack en répartition (au global) . . . . .	86
B.0.1	CDR estimés par Merz & Wuthrich sur les cinq triangles dont nous disposons - Par années d'origine. . . . .	87
B.0.2	IBNR estimés par Chain-ladder sur les cinq triangles dont nous disposons - Par années d'origine. . . . .	88
B.0.3	Résidus des différents modèles de Mack par années d'inventaire. . . . .	88
C.0.1	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCG	91
C.0.2	Planches de résidus du modèle linéaire pondéré de mack en RCG . . . . .	91
C.0.3	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDC	92
C.0.4	Planches de résidus du modèle linéaire pondéré de mack en RCDC . . . . .	92
C.0.5	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDO	93
C.0.6	Planches de résidus du modèle linéaire pondéré de mack en RCDO (répartition)	93
C.0.7	Planches d'analyse des résidus du modèle linéaire pondéré de mack en RCDO en capitalisation (à l'ultime) . . . . .	94
C.0.8	Planches de résidus du modèle linéaire pondéré de mack en RCDO en capi- talisation (à l'ultime) . . . . .	94

# Liste des tableaux

1.3.1	Écart-types standard pour les risques de primes et de réserve en non-vie . .	14
3.1.1	Coefficients règlementaires pour les calcul des PSNEM en compte sociaux .	43
4.1.1	Best estimates en inputs de la méthode 1 . . . . .	55
4.1.2	Tests de log-normalité de Shapiro-Wilk . . . . .	55
4.1.3	Tests de log-normalité de Kolmogorov-Smirnov . . . . .	56
4.1.4	Résultat de l'optimisation de la méthode 1 . . . . .	57
4.1.5	Résultats des différents modèles de Mack . . . . .	59
4.1.6	Résultats des différents modèles de Merz & Wuthrich . . . . .	60
4.1.7	Calcul des <i>Best estimate</i> . . . . .	60
4.1.8	Sigma de réserves des différents triangles . . . . .	60
4.1.9	Sigma de réserves des différents triangles de règlements . . . . .	62
4.2.1	Résultats moyens du Bootstrap synchronisé - Totaux par Groupe de risques homogènes en répartition . . . . .	68
4.2.2	Provisions dossiers dossiers des différents GRH . . . . .	68
4.2.3	Corrélations des CDR, des Provisions et/ou des IBNR selon les branches . .	69
4.2.4	Statistiques finales du Bootstrap . . . . .	69
4.2.5	Sigma de réserve obtenus par le bootstrap synchronisé . . . . .	70
4.4.1	Estimations du sigma de réserve sous les différents paramétrages et ajuste- ments proposés (en %) . . . . .	75
4.4.2	Résultats moyens du Bootstrap synchronisé - Totaux par Groupe de risques homogènes en répartition, modèle raffiné . . . . .	76
4.4.3	Statistiques finales du Bootstrap, modèle raffiné . . . . .	77
4.4.4	Sigma de réserve obtenus par le bootstrap synchronisé, modèle raffiné . . . .	77
C.0.1	Résultats des quasi-Modèles Linéaires Généralisés en modélisation jointe . .	89

\_\_\_\_\_ **Provisionnement et risque de réserve en assurance construction** \_\_\_\_\_  
*Adaptation des modèles standards de provisionnement non-vie au cadre tri-dimensionnel de l'assurance construction, avec point de vue à un an*

Après avoir précisé les spécificités de l'assurance construction française et analysé les problématiques posées par les provisions supplémentaires spécifiques à cette branche d'assurance non-vie, nous prendrons le temps de rappeler les modèles de bases, déterministes comme stochastiques, utilisés en provisionnement non-vie. Puis, ayant transcrit les modèles principaux dans un cadre mathématique plus large, nous présenterons une adaptation de ces modèles au cadre tri-dimensionnel de l'assurance construction française, comprenant l'estimation de variabilité à horizon quelconque. Enfin, par plusieurs approches différentes, nous dégagerons des estimateurs du risque de réserve en normes Solvabilité II dans ce cadre, et nous conclurons par une analyse de ces estimateurs et une comparaison avec la formule standard sur un portefeuille.

*Mots-clefs : Assurance Construction Française, Risque de Réserve, Chain-ladder, Mack, Merz-Wuthrich, Quasi-GLM, Modélisation jointe*

\_\_\_\_\_ **Reserving and reserving risk in french builder's insurance** \_\_\_\_\_  
*Reshaping of standard models in non-life reserving for specific tri-dimensional case of the french builder's insurance, with one year point of vue.*

After having specified the specificities of the French builder's insurance and analyzed the problems posed by additional specific reserves to this line of business, we recall basic models, deterministic and stochastic, used in non-life insurance. Having transcribed the main models in a larger mathematic framework, we present new models shaped for the tri-dimensional issue of French construction insurance, including the estimation of variability in every point of view. Then, through several different approaches, we derive Solvency II reserve risk estimators in those models, and conclude with an analysis of these estimators on a certain portfolio, shaped toward the standard formula.

*Keywords : French Builder's insurance, Reserve risk, Chain-ladder, Mack, Merz-Wuthrich, Quasi-GLM, Joint modeling*