Mémoire présenté devant le CNAM

pour l'obtention du diplôme du Master Droit économie et gestion mention actuariat

et l'admission à l'Institut des Actuaires

le 23 juin 2021

Par : Franck Lopez

Titre: Provisionnement individuel stochastique appliqué à la réassurance non proportionnelle

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 signatures

H. JEBALI
S. FOATA
H. GIBELLO
L. MARTINI

Membres présents du jury du CNAM

D. FAURE
N. ABECERA
O. DESMETTRE
F. WEISS

Entreprise :

Nom : AON

Directeur de mémoire en entreprise :

Nom : F. RAMAHAROBANDRO

Signature :

Invité :

Nom :

Signature :

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

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

Signature du candidat

Secrétariat

Bibliothèque :

Adresse du cnam

Remerciements

Je tiens tout d'abord à remercier le groupe AON de m'avoir donné l'opportunité de travailler sur ce projet et plus particulièrement Fabien Ramaharobandro, mon tuteur professionnel, pour la proposition du sujet et pour son accompagnement.
J'adresse également mes remerciements à tous les membres de l'équipe actuariat d'AON Reinsurance Solutions pour leur soutien et leur bonne humeur.
Je remercie aussi les personnes qui ont pris le temps de relire mon mémoire et pour leurs remarques constructives.
Enfin, je remercie le CNAM et son corps professoral pour avoir rendu possible ma transition professionnelle.

Résumé

Mots-clefs :
Provisionnement Individuel Stochastique, Réassurance Non-proportionnelle, Responsabilité Civile, Régression Logistique Multinomiale, Modèle Additif Généralisé Linéaire Paramétrique, Régression Quantile par Forêt d'arbre décisionnel.

L'inversion du cycle de production de l'activité d'assurance nécessite la constitution de provisions techniques destinées à couvrir les prestations dues aux assurés. Ces prestations étant par nature incertaines, le niveau des provisions techniques doit être évalué en prenant en compte les informations disponibles au moment de leur constitution. En assurance directe, les provisions à constituer et leur volatilité peuvent être estimées globalement à partir des triangles de règlements agrégés par année de survenance avec des méthodes statistiques classiques (Chain Ladder, Bornhuetter-Ferguson, Mack, Bootstrap, etc.). De la même manière, les assureurs doivent évaluer les provisions cédées aux réassureurs qui viennent au bénéfice des provisions techniques. Les méthodes agrégées d'évaluation des provisions utilisées pour l'assurance sont peu adaptées à la réassurance en excédent de sinistre qui s'applique individuellement et de manière non-proportionnelle, ce qui empêche un calcul des cessions moyennes à partir des charges moyennes. D'une part, l'hétérogénéité des sinistres et de leurs développements imposent l'emploi d'une méthode individuelle. D'autre part, ce calcul doit aussi tenir compte de la volatilité pour un sinistre donné.

Pour répondre à ces deux aspects fondamentaux, un outil de provisionnement à la fois individuel et stochastique a été développé sur R. Cet outil permet la réalisation d'une projection stochastique sinistre par sinistre en cascade jusqu'à l'ultime puis l'estimation des provisions cédées par application de la réassurance au niveau individuel. Chaque pas de temps est décomposé en deux sous-étapes. La première consiste à prédire l'évolution en boni, stable ou mali de la charge sinistre via la calibration d'un modèle de régression logistique multinomiale. La seconde consiste à prédire une charge au décalage suivant au moyen de deux méthodes alternatives, soit par un modèle additif généralisé linéaire paramétrique, soit par un modèle de régression quantile par forêt d'arbres décisionnels. La simulation de multiples développements stochastiques des charges individuelles permet enfin d'obtenir la distribution des provisions brutes puis, après l'application individuelle de la réassurance, la distribution des provisions cédées. Le calcul des tardifs n'est pas traité dans ce mémoire. Seuls les sinistres connus sont développés à l'ultime.

En conclusion, les méthodes développées permettent non seulement d'améliorer l'estimation du *Best Estimate* des provisions brutes et des provisions cédées mais aussi d'obtenir une estimation de leur volatilité et de leur distribution. De plus, l'outil développé est déployé au sein de l'entreprise pour le calcul des *Best Estimate* de cessions et des commutations.

Abstract

Key words :
Stochastic Individual Reserving, Non-proportional Reinsurance, Third Party Liability, Multinomial Logistic Regression, Linear Parametric Generalized Additive Model, Quantile Random Forest Regression.

Insurance activity reverse production cycle requires the constitution of technical reserves to cover benefits due to insured persons. These benefits, uncertain in nature, requires that the level of technical reserves be assessed taking into account the available information at the time of their constitution. In direct insurance, constituted reserves and their volatility can be estimated globally from paid triangles aggregated by accident year with classic statistical method (Chain Ladder, Bornhuetter-Ferguson, Mack, Bootstrap, etc.). In the same way, insurers must assess reserves ceded to reinsurers which come to the benefit of technical reserves. Aggregated methods for reserves evaluation used for insurance are poorly adapted to the excess of loss reinsurance which applies individually and non-proportionally. This core charateristic prevents from calculating average cessions with average loss amounts. In one hand, claims heterogeneity and their developments require the use of an individual method. In the other hand, this calculation must also take into account volatility within a given claim.

To address these two fundamental aspects, an individual and stochastic reserving tool has been developped with R. This tool enables to perform a stochastic projection claim by claim in series to the ultimate and then ceded reserves estimation by application of reinsurance at individual level. Each time step is split in two substeps. The first substep consists in claim evolution prediction (positive, negative or stable) with calibration of a multinomial logistic regression model. The second substep consists in loss amount prediction for next development year. This step can be performed either by a parametric linear generalized additive model or by a quantile random forest regression model. Simulation of multiple stochastic developments for individual loss amounts finally makes it possible to obtain gross reserves distribution, and after individual reinsurance application, ceded reserves distribution. Late reported claims calculation is not covered in this study. Only known claims are developed to the ultimate.

To conclude, methods developed in this study allow not only to improve Best estimate estimation of gross reserves and ceded reserves but also to obtain an estimation for their volatility and probability density. In addition, the developed tool is now deployed in the company for ceded losses best estimate calculation and reinsurance treaties commutations.

Table des matières

Remerciements **1**

Résumé **2**

Abstract **3**

Glossaire des abréviations **5**

Introduction **6**

I Cadre de l'étude **7**

1 Contexte de l'étude **8**
1.1 Type d'assurance étudiée : l'Assurance Dommage 8
1.2 Garantie Etudiée : la Responsabilité Civile 8
1.2.1 Généralités sur la Responsabilité Civile 8
1.2.2 Les garanties d'assurance 9
1.2.3 Les dommages concernés 9
1.3 Provisionnement d'un sinistre 10
1.4 Réassurance 11
1.4.1 Généralités 11
1.4.2 Réassurance en excédent de sinistre (XS) 11
1.5 Les provisions cédées aux réassureurs 12
1.5.1 Un enjeu réglementaire : *Best Estimate* de cession 12
1.5.2 Un enjeu économique : commutations 13

2 Problématique du provisionnement pour la réassurance non proportionnelle **14**
2.1 Provisionnement individuel en assurance directe 14
2.2 Un enjeu spécifique pour la réassurance non proportionnelle 14

II Partie Théorique **17**

1 Préparation des données avant application des méthodes **18**
1.1 Indexation des payés et des suspens 18

2 Méthode Chain Ladder appliquée aux sinistres individuels **19**
2.1 Le modèle Chain Ladder 19
2.1.1 Théorie 19

2.1.2 Les limites de l'utilisation d'une méthode agrégée et déterministe appliquée aux sinistres individuels . . . 21

3 Méthodologie du modèle de Provisionnement Individuel Stochastique 22
3.1 Objectif . . . 22
3.2 Architecture du modèle . . . 22
3.3 Modèle de génération de bonis ou malis . . . 23
3.4 Prédiction de la charge au décalage suivant . . . 23
3.5 Prédiction de la charge à l'ultime en cascade . . . 24
3.6 Remplissage parallèle stochastique des triangles de charges individuelles . . . 24
3.7 Estimation d'une cadence de paiement . . . 24
3.8 Calcul de la distribution des provisions brutes . . . 24
3.9 Calcul des coefficients de stabilisation . . . 25
3.10 Application de la réassurance sur les sinistres individuels . . . 26
3.11 Calcul des provisions cédées . . . 26

4 Calibration et validation des modèles 27
4.1 Objectif . . . 27
4.2 Modèle multinomial d'évolution du sinistre . . . 27
4.3 Modèle de prédiction de la charge au décalage suivant . . . 29
4.3.1 Méthode paramétrique : Modèle Additif Généralisé linéaire paramétrique 29
4.3.1.1 Théorie . . . 29
4.3.1.2 Choix des lois . . . 32
4.4 Méthode non paramétrique : Régression quantile par forêt d'arbres décisionnels . 35
4.4.1 Généralité . . . 35
4.4.2 Théorie . . . 35
4.4.2.1 Régression Quantile . . . 35
4.4.2.2 Arbre de décision . . . 35
4.4.2.3 Algorithme des forêts d'arbres décisionnels (Random Forest) . . 36
4.4.2.4 Régression quantile par forêt d'arbres décisionnels (Quantile Random Forest) . . . 37
4.4.2.5 Choix des hyperparamètres . . . 38
4.5 Validation des modèles . . . 39
4.5.1 Validation sur les données d'apprentissage . . . 39
4.5.2 Validation des modèles sur des données nouvelles : *Back Testing* . . . 40

III Application à un cas pratique 41

1 Développement de l'outil de provisionnement individuel sur R 42
1.1 Enjeux spécifiques à l'entreprise . . . 42
1.2 Développement sur R . . . 42

2 Présentation et retraitement des données **43**
2.1 Présentation du cas étudié 43
2.2 Retraitement des données 44
2.3 Statistiques descriptives 45

3 Modèle Chain Ladder agrégé appliqué aux sinistres individuels **49**
3.1 Calcul des coefficients 49
3.2 Calcul des provisions 49

4 Application des modèles de provisionnement individuel stochastique **51**
4.1 Modèle d'évolution en boni, stable ou mali 51
4.2 Modèle Additif Généralisé linéaire paramétrique 53
4.2.1 Choix des distributions 53
4.2.2 Validation générale sur les moyennes prédites 54
4.2.3 Validation détaillée des distributions 55
4.2.4 Calcul des provisions brutes et cédées 57
4.3 Modèle non-paramétrique : Régression quantile par forêt d'arbres décisionnels (*Quantile Random Forest*) 59
4.3.1 Choix des hyperparamètres 59
4.3.2 Validation générale sur les moyennes prédites 61
4.3.3 Validation détaillée des distributions 61
4.3.4 Calcul des provisions brutes et cédées 63
4.4 Comparaison des modèles 65
4.4.1 Application du *Back Testing* 65
4.4.2 Comparaison des provisions brutes et provisions cédées estimées 67
4.5 Limites des modèles 69

Conclusion **70**

Annexes **72**

Glossaire des abréviations

IARD	Incendie-Accident-Risques Divers
RC	Responsabilité Civile
RCA	Responsabilité Civile Automobile
RCVP	Responsabilité Civile Vie Privée
RCE	Responsabilité Civile Entreprise
PSAP	Provision pour Sinistre A Payer
IBNR	*Incurred But Not Reported*
IBNER	*Incurred But Not Enough Reported*
IBNYR	*Incurred But Not Yet Reported*
XS	En excédent de sinistre
S2	Solvabilité II
BE	*Best Estimate*
CL	*Chain Ladder*
LM	Modèle Linéaire
GLM	Modèle Linéaire Généralisé
GAM	Modèle Additif Généralisé
GB	*Generalized Beta*
GG	*Generalized Gamma*
AIC	*Akaike Information Criterion*
QRF	*Quantile Random Forest*

Introduction

Les activités d'assurance et de réassurance se caractérisent par une inversion du cycle de production, c'est à dire que la prime est encaissée immédiatement tandis que les règlements interviennent plus tard, si un sinistre survient. Ainsi, l'activité d'assurance nécessite la constitution de provisions techniques destinées à couvrir les prestations dues aux assurés. Ces prestations étant par nature incertaines, le niveau des provisions techniques doit être suffisant pour honorer les engagements futurs en prenant en compte les informations disponibles au moment de leur constitution. Sous Solvabilité 2, ces provisions contiennent, en plus de la meilleure estimation des engagements, une marge de risque représentant le coût d'immobilisation des fonds propres requis pour honorer ses engagements.
Dans le cadre d'une activité d'assurance directe, les provisions à constituer peuvent être estimées globalement à partir des triangles de règlements agrégés par année de survenance. Des méthodes statistiques (Chain Ladder, Bornhuetter-Ferguson, etc.) permettent de projeter les règlements futurs pour une année de survenance donnée à partir de l'évolution constatée sur les années de survenance passées en ajoutant pour certaines méthodes des hypothèses de sinistralité a priori. Une estimation de la volatilité autour de la provision moyenne estimée peut être calculée avec différentes méthodes (Mack, Bootstrap, GLM, etc.). Ces méthodes, largement utilisées dans le cadre de l'assurance directe, fournissent généralement des résultats satisfaisants et le recours à une approche individuelle (sinistre par sinistre) peut ne pas apporter d'avantages significatifs en regard de l'investissement. Néanmoins, le recours à une méthode individuelle peut dans certains cas apporter une précision accrue des prédictions, notamment pour les années de survenance les plus atypiques, en prenant mieux en compte l'hétérogénéité des sinistres. Enfin, la dimension stochastique de la projection en assurance dommage directe n'a pas d'effet sur la moyenne des provisions brutes, contrairement à la réassurance non proportionnelle.

La réassurance en excédent de sinistre s'applique par nature sur les sinistres individuels de manière non proportionnelle. Ce type de couverture est commun notamment en réassurance responsabilité civile automobile. De la même manière que les assureurs doivent constituer des provisions techniques pour couvrir les prestations dues aux assurés, ils doivent aussi évaluer les provisions cédées aux réassureurs qui viennent au bénéfice des provisions techniques. Les méthodes agrégées d'évaluation des provisions sont peu adaptées à ce type de réassurance qui s'applique individuellement et de manière non-proportionnelle.

L'objectif de ce mémoire est d'étudier des méthodes de provisionnement individuelles et stochastiques. Ces méthodes doivent permettre la réalisation d'une projection stochastique sinistre par sinistre puis l'évaluation des provisions cédées par application de la réassurance au niveau individuel.

Première partie

Cadre de l'étude

Chapitre 1

Contexte de l'étude

Plusieurs éléments de contexte sont développés dans ce chapitre pour permettre d'aborder la problématique de l'étude par la suite.

1.1 Type d'assurance étudiée : l'Assurance Dommage

Le Code des Assurances dénombre, dans l'article R 321-1, 26 branches d'assurance. Un agrément administratif est la plupart du temps nécessaire pour exercer des opérations d'assurance dans chacune de ces branches.
Selon le mode de gestion des primes, on distingue deux grandes familles d'assurance : les assurances IARD (Incendie-Accident-Risques Divers), et les assurances Vie. L'assurance IARD est gérée en répartition , c'est à dire que les primes des assurés sont affectées au paiement des sinistres au cours de l'exercice tandis que l'assurance vie est gérée en capitalisation, la capitalisation financière et viagère des primes permettant le paiement des sinistres a posteriori.
De plus, selon le préjudice subi, l'indemnisation du sinistre peut être réalisée selon le principe indemnitaire ou forfaitaire. Le principe indemnitaire a pour objectif de compenser une perte (bien matériel, invalidité, etc.) tandis que le principe forfaitaire entraîne le versement des montants fixés par le contrat d'assurance.
Les Assurances IARD peuvent être de deux types : les Assurances de Dommages et les Assurances de Personnes. Les Assurances de Dommages obéissent au principe indemnitaire et peuvent concerner les biens de l'assuré ou sa responsabilité civile (réparation des préjudices causés à un tiers). Ce mémoire s'intéresse au provisionnement des sinistres des Assurances de Dommages et plus spécifiquement aux sinistres graves qui sont généralement liés à la responsabilité civile.

1.2 Garantie Etudiée : la Responsabilité Civile

1.2.1 Généralités sur la Responsabilité Civile

La responsabilité civile (RC) est l'obligation faite à une personne de réparer le préjudice causé à autrui. Le dommage doit être intégralement réparé, par le versement de dommages-intérêts à la victime par le responsable. La responsabilité civile est dite « contractuelle » en cas de mauvaise exécution ou d'inexécution totale ou partielle des obligations nées d'un contrat. Elle est dite « extracontractuelle » lorsque, comme son nom l'indique, son origine n'est pas un contrat. C'est l'article 1382 du Code civil qui prévoit la responsabilité extra-contractuelle simple : « Tout fait quelconque de l'homme, qui cause à autrui un dommage, oblige celui par la faute duquel il est arrivé, à le réparer ».

1.2.2 Les garanties d'assurance

Responsabilité Civile Automobile (RCA)
Instaurée le 27 février 1958, l'assurance Automobile est obligatoire pour tout véhicule terrestre à moteur. Elle couvre la Responsabilité Civile du Conducteur contre les dommages causés aux tiers avec son véhicule (passagers, piétons, automobilistes). Si cette garantie constitue le minimum obligatoire, elle ne couvre ni les dommages corporels du conducteur, ni les dommages aux véhicules.

Responsabilité Civile Vie Privée (RCVP)
La garantie responsabilité civile vie privée couvre les conséquences pécuniaires de la responsabilité civile incombant à l'assuré dans le cadre de sa vie privée en raison des dommages corporels, matériels et immatériels causés à des tiers par : son propre fait ou le fait des personnes dont il doit répondre ; les biens mobiliers ou les animaux dont l'assuré est propriétaire ou gardien.

Responsabilité Civile Entreprise (RCE)
Une entreprise a l'obligation de réparer tout dommage causé à autrui par les personnes ou les biens dont elle répond. Il s'agit de sa responsabilité civile.

1.2.3 Les dommages concernés

Dommages Corporels
Un dommage corporel est un dommage portant atteinte à l'intégrité de la personne, c'est à dire causant des blessures ou le décès de la personne. Les dommages corporels sont assurables avec différents contrats d'assurance de responsabilité ou de personnes. Les dommages corporels représentent uniquement des atteintes à la personne humaine : ils s'opposent aux dommages matériels qui sont des atteintes aux choses ou aux animaux, même s'il peuvent survenir lors d'un même sinistre.
Du fait d'une spécificité légale française, la couverture des dommages corporels est illimitée, c'est à dire que la loi impose aux assureurs la prise en charge intégrale de tout sinistre corporel causé par les personnes assurées, quel qu'en soit le montant.
Certains postes d'indemnisation peuvent être payés sous la forme de rentes viagères (principalement l'Assistance Tierce Personne). Dans ce cas, des règles spécifiques de calcul des provisions mathématiques s'appliquent. Ces règles de calculs font intervenir des tables de mortalité, un taux d'actualisation et un taux de revalorisation encadrés par la loi.
Enfin, les règles à appliquer par le traité de réassurance peuvent être différentes de celles pratiquées sur le contrat d'assurance (tables de mortalité différentes, taux d'actualisation fixé contractuellement, limite sur la revalorisation des rentes, etc.). Il convient dans ce cas de convertir les charges sinistres d'un référentiel à un autre pour la bonne application des traités.

Dommages Matériels
Un dommage matériel est un dommage représentant une atteinte à une chose, un bien ou un animal. Il s'oppose au dommage corporel qui représente une atteinte à la personne humaine. Révisé tous les 5 ans en fonction de l'inflation, le plafond minimal de couverture des dommages

matériels en assurance auto en cas de sinistre est fixé à 1 220 000 €. Très loin, cependant des pratiques du marché, les contrats d'assurance automobile prévoyant des plafonds plus élevés, allant de 50 à 100 M€.

Dommages Immatériels
Un dommage immatériel est un dommage « résultant de la privation de jouissance d'un droit, de l'interruption d'un service rendu par une personne ou par un bien meuble ou immeuble ou de la perte de bénéfice ». Il s'oppose au dommage corporel et au dommage matériel, même s'il peut en être une conséquence.

1.3 Provisionnement d'un sinistre

Provision pour Sinistre A Payer (PSAP)
Entre la survenance d'un sinistre et son paiement définitif, de nombreuses années peuvent s'écouler. Entre ces deux instants de la vie d'un sinistre, la provision pour sinistre à payer (PSAP) est destinée à couvrir l'ensemble des coûts restant à régler au titre des sinistres. Plus précisément, cette PSAP est définie dans l'article R331-6l du Code des Assurance comme la « valeur estimative des dépenses en principal et en frais, tant internes qu'externes, nécessaires au règlement de tous les sinistres survenus et non payés ».
Cette provision peut se décomposer en deux parties principales :

- La provision dossier/dossier : évaluée individuellement par les gestionnaires sinistre en suivant généralement un guide de provisionnement, il s'agit d'une estimation des règlements à venir en fonction des informations disponibles à date. Ces provisions sont réévaluées au fil des nouvelles informations disponibles sur le dossier concerné.
- Les IBNR (*Incurred But Not Reported*) : une provision complémentaire destinée à couvrir l'insuffisance (ou la prudence) des dossiers ouverts (IBNER) ou des dossiers non encore déclarés (IBNYR).
  - IBNER (*Incurred But Not Enough Reported*) : provision pour sinistres insuffisamment provisionnés. Elle constitue un complément aux provisions dossier / dossier si celles-ci s'avèrent insuffisantes (malis) ou trop importantes (bonis).
  - IBNYR (*Incurred But Not Yet Reported*) : provision pour sinistres survenus mais pas encore déclarés (dit tardifs).

Dans le calcul de la PSAP peuvent aussi se retrouver la PFGS (Provision pour Frais de Gestion des Sinistres) et la PRAE (Prévision de Recours A Encaisser). La décomposition de la charge sinistre est présentée dans l'équation 1.1.

$$\begin{aligned} Charge\ Sinistre &= Payés + PSAP \\ Charge\ Sinistre &= Payés + Provision\ d/d + IBNR \end{aligned} \tag{1.1}$$

La provision d/d est connue et dépend des éléments connues à ce jour par le gestionnaire du sinistre. Estimer la charge à l'ultime revient donc à estimer les IBNR. Les méthodes qui sont développées dans ce mémoire ont pour objectif d'estimer précisément les IBNER. Les IBNYR ne sont pas traités dans ce mémoire et nécessiterait , pour leur calcul, le développement d'un modèle de tardif.

1.4 Réassurance

1.4.1 Généralités

L'opération de réassurance est un transfert de risque de l'assureur (la cédante) vers un réassureur qui accepte de porter une partie du risque, cette partie étant définie contractuellement dans un traité de réassurance qui lie la cédante et le réassureur. Cependant, seul l'assureur répond à l'égard de l'assuré, aucun lien contractuel ne lie l'assuré au réassureur. La réassurance permet de réduire la volatilité et de protéger le résultat des pertes extrêmes mais aussi d'accompagner le développement d'un portefeuille. La réassurance permet à la cédante de réduire ses fonds propres nécessaires pour couvrir ses risques grâce à la mutualisation. Le réassureur est lui aussi soumis à des règles de solvabilité et emploie la diversification dans sa souscription pour réduire ses besoins en capitaux ainsi que la rétrocession (transfert du risque vers un autre réassureur). L'opération de réassurance permet ainsi à la cédante de tirer profit de la mutualisation à une plus grande échelle, qui lui est normalement inaccessible.

1.4.2 Réassurance en excédent de sinistre (XS)

Le montant d'un sinistre corporel grave peut représenter plusieurs dizaines de millions d'euros. De plus la garantie illimitée obligatoire peut potentiellement aboutir à des charges sinistres encore plus élevées. Ceci amène les compagnies d'assurance à vouloir se couvrir contre l'intensité exceptionnelle des sinistres individuels. Pour ce faire, une Réassurance en excédent de sinistre (XS) est généralement mise en place.
Un traité de réassurance en excédent de sinistre est une couverture de réassurance qui s'applique individuellement sur la charge sinistre de manière non proportionnelle. Ce type de traité est dit non proportionnel car la charge cédée au réassureur n'est pas proportionnelle à la charge brute du sinistre.

Cette couverture se caractérise par :

- Une priorité : montant à partir duquel l'excédent est pris en charge par le réassureur.

- Une portée (ou capacité) : montant maximal pris en charge par la réassurance pour un sinistre individuel.
- Un nombre de reconstitution de la capacité ou une limite annuelle agrégée (AAL).
- Une franchise annuelle agrégée (AAD).
- Une clause de rente qui spécifie comment les rentes doivent être valorisées dans la charge sinistre affectée au traité.
- Une clause de stabilisation qui définit la règle du partage de l'inflation entre l'assureur et le réassureur par le biais de la multiplication de la priorité et de la portée par un coefficient dépendant de l'inflation et de la cadence de paiement.

La figure 1 illustre le fonctionnement de la réassurance en excédent de sinistre.

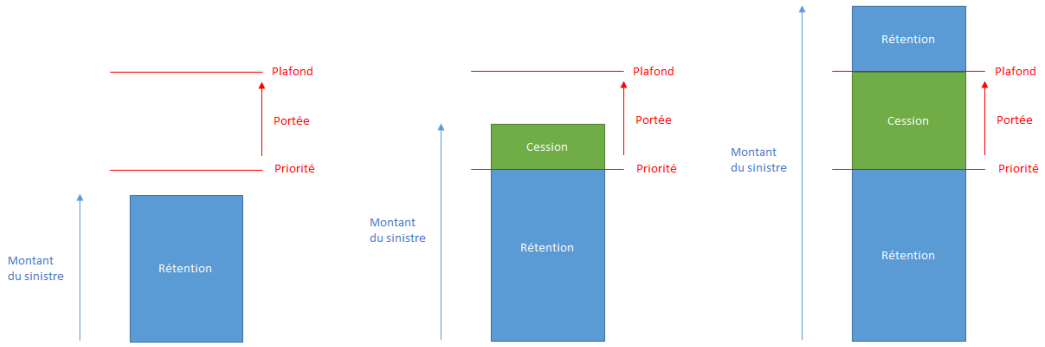


FIGURE 1 – Principe de la réassurance en excédent de sinistre.

1.5 Les provisions cédées aux réassureurs

Les assureurs doivent calculer des provisions techniques brutes pour répondre à des besoins réglementaires : norme comptable française, norme prudentielle Solvabilité II (S2), IFRS, etc. Ces provisions techniques représentent les engagements de l'assureur envers ses assurés. Si l'assureur bénéficie de couvertures de réassurance, les engagements des réassureurs sont admis en représentation des provisions techniques brutes. Le montant estimé des provisions cédées vient au bénéfice des provisions techniques brutes en libérant des fonds propres et en améliorant la solvabilité. L'estimation des provisions cédées est donc importante pour la bonne représentation du résultat technique mais aussi pour le calcul de la solvabilité de l'assureur.

1.5.1 Un enjeu réglementaire : *Best Estimate* de cession

La norme prudentielle Solvabilité II (S2) stipule que la valeur des provisions techniques doit être égale à la somme de la meilleure estimation, le *Best Estimate* (BE) et d'une Marge de risque. La valorisation des BE se fait en sommant les flux futurs (décaissements et encaissements) probabilisés et actualisés (ce qui nécessite d'estimer la date des flux) sans application d'une marge de prudence. Les BE des provisions brutes de réassurance sont inscrits au passif du Bilan. Les BE de cessions sont constitués en représentation à l'actif pour prendre en compte les cessions aux réassureurs et sont ajustés « afin de tenir compte des pertes probables pour défaut de la

contrepartie ». Une meilleure estimation des BE de cessions aura donc des répercussions sur le calcul de la solvabilité de l'assureur. Pour plus de précisions sur les aspects réglementaires S2 des provisions cédées, le lecteur pourra se reporter à la lecture du manuel *Best-Estimate Liabilities* [Taillieu, 2016].

1.5.2 Un enjeu économique : commutations

La commutation en réassurance consiste au rachat par le réassureur de ses engagements envers la cédante. Celle-ci intervient pour solder la relation qui lie l'assureur et le réassureur généralement sur les traités anciens avec un nombre limité de sinistres en cours. Le transfert du risque se fait donc du réassureur vers l'assureur, le réassureur versant alors une prime de commutation qui doit refléter la valeur des engagements ainsi qu'une marge de risque. Dans ce cadre, la bonne estimation des provisions cédées permet l'estimation d'une prime de commutation qui correspond précisement au risque transféré.

Chapitre 2

Problématique du provisionnement pour la réassurance non proportionnelle

Le contexte ayant été fixé, la problématique du provisionnement qui se pose spécifiquement à la réassurance non proportionnelle est développée dans ce chapitre.

2.1 Provisionnement individuel en assurance directe

Dans le cadre du provisionnement en assurance directe, les méthodes classiques agrégées de type Chain-Ladder sont adaptées la plupart du temps pour fournir une estimation des provisions brutes moyennes et de leur volatilité. Néanmoins de nombreux mémoires ([Beneteau, 2004], [Barbaste, 2017], [Quiquet, 2019]) et articles ([Antonio et Plat, 2012], [ASTIN, 2017], [Baudry et Y.Robert, 2017]) ont exploré les avantages de l'utilisation de méthodes de provisionnement individuel en assurance directe en utilisant des méthodes paramétriques ou non-paramétriques. Ils concluent généralement que les méthodes individuelles permettent d'augmenter la précision des prédictions en prenant mieux en compte l'hétérogénéité des sinistres, surtout pour les sinistres à durée de vie longue et les années de sinistralité atypiques. La projection individuelle est généralement faite de manière déterministe avec un grand nombre de variables explicatives. En effet, l'objectif est, dans ce cas, d'estimer au mieux la provision moyenne de chaque sinistre et la composante aléatoire de la variable à prédire n'est généralement pas étudiée. Pourtant, cette composante aléatoire est déterminante en réassurance non proportionnelle.

2.2 Un enjeu spécifique pour la réassurance non proportionnelle

Tout d'abord les sinistres qui touchent la réassurance non proportionnelle, notamment la réassurance en excédent de sinistre, ne sont pas de même nature que ceux qui touchent principalement l'assurance directe. En effet, ils sont moins nombreux et plus graves en raison de la sélection qui se fait par l'application d'une franchise. De plus, leur évolution est généralement plus longue et des bonis ou des malis importants peuvent être constatés au cours de la vie du sinistre sur des dizaines d'années. Ces éléments qualitatifs indiquent qu'un développement individuel des sinistres revêt une importance particulière enrRéassurance non proportionnelle.

De manière plus détaillée, en réassurance proportionnelle, il est possible d'estimer la cession moyenne d'un sinistre en estimant sa charge sinistre moyenne à l'ultime sur laquelle on applique

ensuite une quote-part (eq. 2.1).

$$
\begin{aligned}
Cession_Moyenne &= \int_0^\infty C.f(C).QP.dC \\
Cession_Moyenne &= (\int_0^\infty C.f(C).dC) \times QP \\
Cession_Moyenne &= Charge_Moyenne \times QP
\end{aligned}
\tag{2.1}
$$

Ce type de réassurance est donc adaptée à une méthode agrégée pour estimer les provisions cédées puisque le lien entre charge et cession est linéaire. Ce n'est pas le cas pour la réassurance non proportionnelle et plus spécifiquement pour la réassurance en excédent de sinistre définie par l'équation (eq. 2.2). La non linéarité du lien qui existe entre la sinistralité brute et les cessions empêche de raisonner de manière agrégée.

$$
\begin{aligned}
Cession_Moyenne &= \int_0^\infty max(C - priorité, 0).f(C).dC \\
Cession_Moyenne &\neq max((\int_0^\infty .f(C).dC) - priorité, 0)
\end{aligned}
\tag{2.2}
$$

Les figures 2 et 3 illustrent graphiquement cette idée. Sur la première figure, la charge ultime projetée avec la méthode Chain Ladder aboutit à une charge en dessous de la priorité et donc à des cessions nulles tandis que la modélisation de différents scénarios potentiels d'évolution de la charge sinistre aboutit à une cession non nulle. Sur la seconde figure, la charge est initialement au dessus de la priorité et est projetée de la même manière que pour le premier sinistre ce qui aboutit à une cession encore plus élevée tandis que la modélisation individuelle permet d'appliquer une évolution différente du sinistre précédent (dans ce cas globalement à la baisse) et fait aussi apparaître des cas où la charge sinistre passe en dessous de la priorité, baissant d'autant plus les cessions.

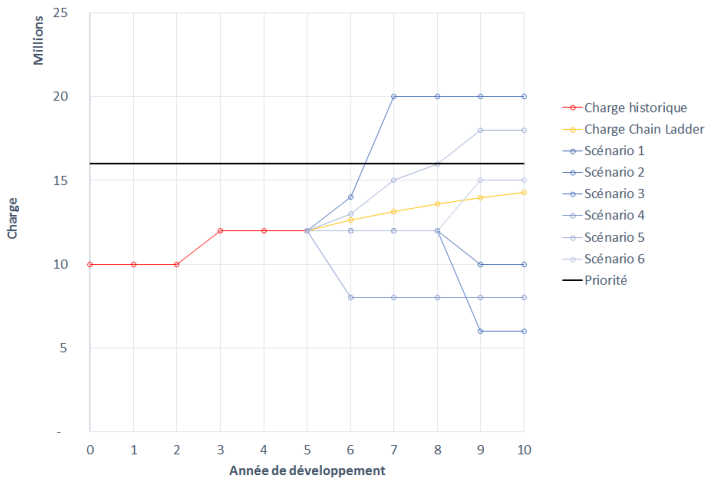


FIGURE 2 – Illustration de l'intérêt d'un provisionnement individuel stochastique en réassurance non proportionnelle : sinistre initialement en dessous de la priorité.

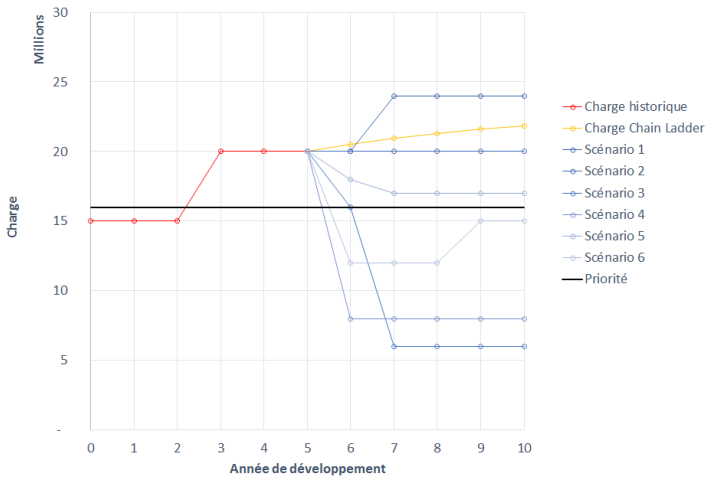


FIGURE 3 – Illustration de l'intérêt d'un provisionnement individuel stochastique en réassurance non proportionnelle : sinistre initialement au dessus de la priorité.

Deux biais principaux de la méthode agrégée sont illustrés par cet exemple. Le premier biais consiste à ne pas tenir compte des caractéristiques individuelles du sinistre, c'est à dire l'hétérogénéité de la sinistralité, pour sa projection ce qui entraîne un biais sur la charge ultime brute individuelle et donc sur la provision cédée individuelle. Au brut, si les provisions à l'ultime des petits sinistres sont systématiquement sous estimées tandis que celles des gros sinistres sont surestimées, l'agrégation des provisions individuelles aura tendance à compenser les biais induit par l'hétérogénéité des sinistres. Ce n'est plus le cas pour la réassurance en excédent de sinistre qui sera uniquement impactée par les biais des sinistres supérieurs à la priorité. Le deuxième biais consiste à ne pas tenir compte des différents scénarios d'évolutions possibles pour un même sinistre individuel ce qui, là aussi, induit un écart sur la provision cédée individuelle et donc sur la provision cédée totale calculée. Le développement d'une méthode de provisionnement qui est à la fois individuelle et stochastique permet de répondre conjointement à ces deux biais.

Deuxième partie

Partie Théorique

Chapitre 1

Préparation des données avant application des méthodes

La première étape, avant d'utiliser une méthode de projection à l'ultime, est une étape de préparation des données en leur appliquant les retraitements nécessaires à l'application des méthodes. En effet, les montants présents dans la base sinistre sont issus d'années de survenance différentes vues à différents arrêtés. Ces montants ne sont donc pas forcément directement comparables. Cela peut être du à l'inflation mais aussi à des évolutions des règles de provisionnement d'une survenance à l'autre (taux d'actualisation des rentes par exemple). Dans un contexte de tarification, le référentiel de calcul doit être celui du traité de réassurance à renouveler. Dans le cadre d'un provisionnement, le référentiel de calcul choisi est arbitraire. Il doit avant tout permettre d'appliquer correctement la méthode de projection à l'ultime pour ensuite revenir aux référentiels historiques.

1.1 Indexation des payés et des suspens

Pour l'application des différentes méthodes de développement de la charge sinistre, on considère que les charges sont, relativement aux années de survenance, indépendantes et exprimées dans une base monétaire homogène. C'est pourquoi une indexation des données de base est réalisée avant l'application des méthodes de développement. Toutes les années de survenance et leurs flux correspondants (Payés décumulés et Suspens) sont translatés dans le temps pour que les années de survenance coïncident avec une année de référence unique (eq. 1.1). L'indice utilisé pour cette transformation doit inclure l'inflation passée et future.

$$Flux\ Indexé = Flux \times \frac{Indice(Année\ Référence + Décalage)}{Indice(Arrêté)} \tag{1.1}$$

Une fois que le triangle des charges individuelles est complété jusqu'à l'ultime, la transformation inverse est appliquée pour revenir aux montants exprimés dans les référentiels historiques (eq. 1.2). En effet, les traités de réassurances sont associés à des années de survenance et s'appliquent sur les montants historiques.

$$Flux = Flux_Indexé \times \frac{Indice(Arrêté)}{Indice(Année_Référence + Décalage)} \tag{1.2}$$

Chapitre 2

Méthode Chain Ladder appliquée aux sinistres individuels

2.1 Le modèle Chain Ladder

Le modèle Chain Ladder et ses dérivés est largement utilisé en Assurance. Il s'agit initialement d'une méthode heuristique (approche intuitive) qui a par la suite été confirmée dans un formalisme mathématique plus rigoureux. En effet, les résultats de cette méthode correspondent non seulement au maximum de vraisemblance mais peuvent aussi être représentés dans le cadre plus général d'un modèle linéaire généralisé (GLM) [Taylor et McGuire, 2016].

2.1.1 Théorie

Le principe de base du modèle Chain-Ladder est de projeter des montants dans le futur à partir des accroissements constatés dans le passé. Les facteurs d'accroissement qui permettent de passer d'un pas de temps au suivant sont appelés facteurs de développement (f). Soit un sinistre i survenu l'année $surv$, la charge au décalage d est notée $x_{surv,i}(d)$. Le dernier arrêté de la base de donnée est noté A_{diag}. Le décalage correspondant dépend de la survenance et vaut donc $A_{diag} - surv$. Enfin, pour l'application correcte du modèle Chain-Ladder, les années de survenance doivent être indépendantes entre elles. L'équation 2.1 détaille la méthodologie Chain Ladder.

$$\hat{f}(d) = \frac{\sum_{surv,i} x_{surv,i}(d+1) \times 1_{\{d<A_{diag}-surv\}}}{\sum_{surv,i} x_{surv,i}(d) \times 1_{\{d<A_{diag}-surv\}}} \tag{2.1}$$

Une autre formulation [Taylor et McGuire, 2016] permet de se placer dans le cadre plus général d'une régression linéaire de $f_{surv,i}(d)$ pondérée par des poids $w_{surv,i}(d)$ (eq. 2.2). Les poids représentent alors l'importance relative de la charge individuelle de sorte que la somme pondérée des facteurs de développement individuels soit égale au facteur de développement agrégé par année de survenance. Les deux formulations sont équivalentes, la deuxième ayant l'avantage de permettre le calcul d'une volatilité associée à une distribution via un GLM par exemple.

$$
\begin{aligned}
\hat{f}_{surv,i}(d) &= \frac{x_{surv,i}(d+1)}{x_{surv,i}(d)} \\
w_{surv,i}(d) &= \frac{x_{surv,i}(d)}{\sum_i x_{surv,i}(d)} \times 1_{\{d<A_{diag}-surv\}} \\
\hat{f}(d) &= \sum_{surv,i} w_{surv,i}(d) \times \hat{f}_{surv,i}(d)
\end{aligned}
\tag{2.2}
$$

Dans une approche classique, la projection est réalisée sur les montants agrégés par année de survenance (eq. 2.3).

$$
x_{surv}(d) = \sum_{i}^{N_{surv}} x_{surv,i}(d) \tag{2.3}
$$

La projection à l'ultime de la charge sinistre agrégée par survenance est réalisée ensuite en cascade en appliquant l'équation 2.4 d'un décalage à l'autre (fig. 4). On note $diag_{surv} = A_{diag} - surv$ le décalage correspondant au dernier arrêté pour l'année de survenance $surv$.

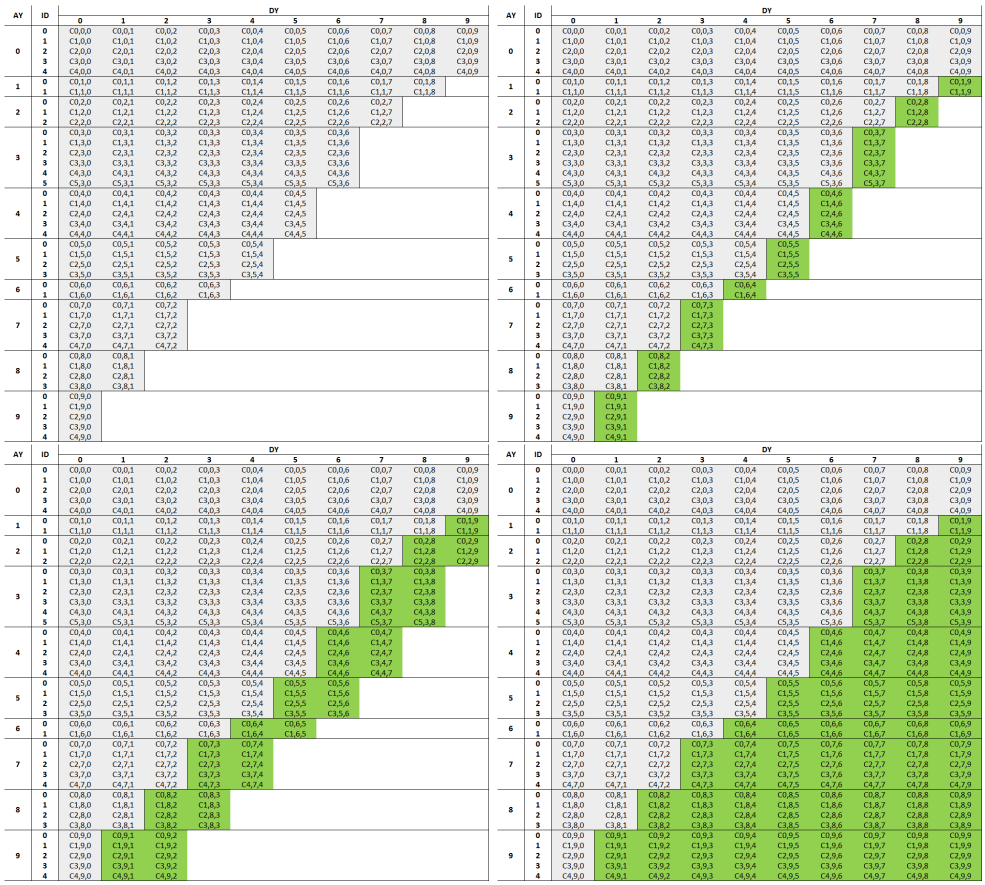


FIGURE 4 – Remplissage en cascade d'un triangle individuel

$$
\begin{aligned}
x_{surv}(d) &= \sum_{i}^{N_{surv}} x_{surv,i}(d) \\
x_{surv}(d+1) &= x_{surv}(d) \times f(d), \quad (d \geq diag_{surv}) \\
x_{surv}(ultime) &= x_{surv}(diag_{surv}) \times \prod_{d=diag_{surv}} f(d)
\end{aligned}
\tag{2.4}
$$

Pour obtenir une estimation approximative des charges ultimes individuelles les coefficients de passage sont appliqués directement aux charges individuelles (eq. 2.5).

$$
\begin{aligned}
x_{surv,i}(d+1) &= x_{surv,i}(d) \times f(d), \quad (d \geq diag_{surv}) \\
x_{surv,i}(ultime) &= x_{surv,i}(diag_{surv}) \times \prod_{d=diag_{surv}} f(d)
\end{aligned}
\tag{2.5}
$$

L'agrégation par année de survenance des charges ultimes individuelles aboutit aux charges ultimes agrégées précédentes (eq. 2.6). L'application individuelle de la méthode Chain-Ladder, à défaut d'être une véritable méthode individuelle, est cohérente avec la méthode agrégée.

$$
x_{surv}(ultime) = \sum_{i}^{N_{surv}} x_{surv,i}(ultime) \tag{2.6}
$$

2.1.2 Les limites de l'utilisation d'une méthode agrégée et déterministe appliquée aux sinistres individuels

Avec cette méthode, le développement d'un sinistre ne dépend que de sa maturité et est réalisé de manière déterministe. Les limites de l'utilisation d'une méthode agrégée et déterministe ont été introduites dans la partie I au chapitre 2 p. 14. La méthode Chain Ladder simple est peu adaptée à la projection individuelle des sinistres graves qui ont une volatilité importante. De plus, cette volatilité impacte significativement le calcul des cessions en réassurance non proportionnelle ce qui doit conduire à des biais dans le calcul des provisions cédées.

Chapitre 3

Méthodologie du modèle de Provisionnement Individuel Stochastique

3.1 Objectif

L'objectif est de prédire, pour chaque sinistre individuel, l'ensemble des développements probables de la charge sinistre en se basant sur l'état actuel connu du sinistre (décalage et charge). La connaissance de cet ensemble doit permettre d'appliquer la réassurance non proportionnelle de manière individuelle et stochastique. La construction de cette méthode s'inspire en partie des travaux de [Taylor et McGuire, 2016] qui développent une méthode de projection stochastique paramétrique mais agrégée et du rapport [ASTIN, 2017] qui développe une méthode de projection individuelle non-paramétrique mais déterministe.

3.2 Architecture du modèle

L'architecture générale du modèle est présentée ci-dessous. Chaque étape de calcul est présentée de manière plus détaillée dans les sections suivantes.

Départ : Triangle des charges individuelles (payés et suspens)
1. Indexation du triangle des charges individuelles (payés décumulés et des suspens)
2. Calibration du modèle multinomial d'évolution des sinistres (boni, stable, mali)
3. Calibration du modèle de prédiction de la charge au décalage suivant
4. Simulation des N scénarios :
   4.1. Projection individuelle stochastique du triangle t jusqu'à l'ultime
      a. Départ : Charge individuelle $x_i(diag)$ au dernier arrêté
      b. Tirage de l'évolution au décalage suivant (boni, stable ou mali)
      c. Tirage de la charge au décalage suivant en fonction de l'évolution et de la charge
      d. Répétition des étapes b et c en cascade jusqu'à l'ultime
   4.2. Calcul des règlements individuels à partir de la cadence des règlements
   4.3. Calcul des provision brutes
   4.4. Application des traités de réassurance : calcul des cessions
   4.5. Calcul des provision cédées
5. Calcul d'une distribution des provisions brutes et cédées à l'ultime

3.3 Modèle de génération de bonis ou malis

En observant l'évolution des charges sinistres corporels d'un arrêté à un autre, plusieurs observations peuvent être faites. Tout d'abord, une grande partie des sinistres garde une charge stable d'une année à l'autre. Ensuite, pour les cas où la charge est réévaluée, les bonis ne sont pas de même nature que les malis. En effet, un boni ne peut pas excéder la charge du sinistre tandis qu'un mali n'a théoriquement pas de borne supérieure. Il est donc justifié de commencer par séparer les données en trois sous ensembles (boni, stable et mali).

Ainsi, la première étape à accomplir afin de pouvoir développer la charge d'un sinistre d'une année sur l'autre est d'évaluer la probabilité d'évolution d'un sinistre en fonction de son état actuel. Un modèle multinomiale à trois variables (boni, stable et mali) est indiqué pour accomplir cette première étape. Soit E la variable aléatoire catégorielle qui peut prendre les valeurs nominales "boni", "stable" ou "mali". E suit une loi multinomiale avec trois issues possibles (eq. 3.1).

$$E \sim \mathscr{E}(X, d) \tag{3.1}$$

Les paramètres de cette loi sont les probabilités de chacune des issues (eq. 3.2).

$$\mathbb{P}(E = e|(X = x, D = d)) = \begin{cases} pBoni(x, d) & si \quad e = "boni" \\ pStable(x, d) & si \quad e = "stable" \\ pMali(x, d) & si \quad e = "mali" \end{cases} \tag{3.2}$$

3.4 Prédiction de la charge au décalage suivant

La charge au décalage suivant suit une loi de probabilité qui dépend de la charge, du décalage et de l'évolution prédite du sinistre (eq. 3.3).

$$X_{d+1} \sim \mathscr{L}(x_d, d, e) \tag{3.3}$$

Cette loi peut être décomposée en trois parties selon l'évolution du sinistre en boni, stable ou en mali (eq. 3.4 et 3.5).

$$\begin{aligned} X_{d+1}|E = "boni" &\sim \mathscr{B}(x_d, d) \\ X_{d+1}|E = "stable" &= x_d \\ X_{d+1}|E = "mali" &\sim \mathscr{M}(x_d, d) \end{aligned} \tag{3.4}$$

$$X_{d+1} \sim \begin{cases} \mathscr{B}(x_d, d) & avec\ probabilité\ pBoni(x_d, d) \\ x_d & avec\ probabilité\ pStable(x_d, d) \\ \mathscr{M}(x_d, d) & avec\ probabilité\ pMali(x_d, d) \end{cases} \tag{3.5}$$

Les lois de probabilités $\mathscr{B}$ et $\mathscr{M}$ doivent être calibrées à partir de deux sous ensembles de la base de données constitué respectivement des bonis et des malis. Cette calibration peut être réalisée

par plusieurs méthodes et est développée dans le chapitre 4.

3.5 Prédiction de la charge à l'ultime en cascade

A l'aide de la calibration précédente, pour une charge et un décalage donnés, le tirage aléatoire d'une charge au décalage suivant peut être effectué. Pour cela, un tirage de l'évolution du sinistre est tout d'abord réalisé. Une fois l'évolution fixée, la loi de probabilité sachant l'évolution est utilisée pour tirer aléatoirement une charge au décalage suivant. Ce processus peut alors être répété en cascade jusqu'à l'ultime.

3.6 Remplissage parallèle stochastique des triangles de charges individuelles

Afin d'obtenir une bonne estimation de toutes les évolutions possibles des charges sinistres individuelles, le processus en cascade décrit précédemment doit être répété un grand nombre de fois. Ce processus sera donc appliqué en parallèle sur un grand nombre de triangles de charge individuelle. Ces matrices simulées de charges individuelles constituent alors toutes les issues probables du triangle de charges individuelles d'origine projeté à l'ultime.

3.7 Estimation d'une cadence de paiement

L'estimation d'une cadence de paiement est nécessaire pour deux raisons. Premièrement, le calcul des provisions *Best Estimate* est effectué en actualisant les règlements futurs. Deuxièmement, l'application de la réassurance en excédent de sinistre se fait après stabilisation de la priorité et de la portée. Cette stabilisation permet de corriger les bornes des traités de réassurance en fonction de l'inflation observée. En pratique, le calcul du coefficient de stabilisation nécessite la connaissance des règlements.

Pour calculer une cadence de paiement, une approche agrégée a été choisie. Celle-ci consiste à calculer, pour chaque décalage, la part de la somme des règlements observés sur la somme des charges à l'ultime calculées (eq. 3.6). Cette cadence agrégée par triangle est ensuite appliquée à l'ensemble des charges individuelles projetées afin d'obtenir une matrice des règlements individuels.

$$Cadence_d = \frac{\sum_{i=1}^{Nsinistre} Règlements_{i,d}}{\sum_{i=1}^{Nsinistre} Charge_{i,ultime}} \tag{3.6}$$

3.8 Calcul de la distribution des provisions brutes

A partir du jeu de matrices simulées des charges individuelles, la distribution des provisions brutes peut être calculée. Pour ce faire, la provision de chaque sinistre i appartenant à la simulation s doit tout d'abord être calculée en effectuant la somme actualisée des règlements futurs.

Puis la provision totale par simulation est obtenue en sommant les provisions par sinistre pour une simulation donnée. Enfin, l'ensemble des provisions totales par simulation constitue une distribution des provisions. Ces étapes sont présentées dans l'équation 3.7.

$$\begin{aligned} Provision_{s,i} &= \sum_{d=diag+1}^{ultime} \frac{Règlements_{s,i,d}}{(1+t_{actu})^d} \\ Provision_s &= \sum_{i=1}^{Nsinistres} Provision_{s,i} \end{aligned} \tag{3.7}$$

3.9 Calcul des coefficients de stabilisation

Les traités en excédent de sinistre comportent généralement une clause de stabilisation qui permet de garantir que les bornes du traité restent stables relativement à l'inflation des coûts des sinistres. L'inflation est mesurée par un indice qui est fixé dans le traité. L'application de la stabilisation consiste à calculer un coefficient de stabilisation à appliquer aux bornes du traité (priorité et portée). Ce coefficient de stabilisation est défini comme la somme des règlements et suspens divisée par la somme des règlements et suspens stabilisés. Les étapes de calculs sont détaillés dans les équations 3.8 et 3.9.

$$\begin{aligned} Reglement_Stab_{i,j} &= Reglement_i \times \frac{I_{surv}}{I_{surv+j}} \\ Suspens_Stab_{i,j} &= Suspens_i \times \frac{I_{surv}}{I_{surv+j}} \end{aligned} \tag{3.8}$$

$$CoefStab_{i,d} = \frac{Reglements_Cumules + Suspens}{Reglements_Cumules_Stab + Suspens_Stab}$$

$$CoefStab_{i,d} = \frac{\sum_{j=0}^{d} Reglement_{i,j} + Suspens_{i,d}}{\sum_{j=0}^{d} Reglement_i \times \frac{I_{surv}}{I_{surv+j}} + Suspens_{i,d} \times \frac{I_{surv}}{I_{surv+d}}} \tag{3.9}$$

A l'ultime, l'équation précédente se simplifie en l'absence de suspens (eq. 3.10).

$$CoefStab_{i,ult} = \frac{\sum_{j=0}^{ult} Reglement_{i,j}}{\sum_{j=0}^{ult} Reglement_i \times \frac{I_{surv}}{I_{surv+j}}} \tag{3.10}$$

Dans la pratique, un paramètre supplémentaire rentre en compte dans le calcul du coefficient de stabilisation. Il s'agit du seuil de déclenchement de la stabilisation. En effet, il est généralement spécifié dans les traités qu'en dessous de ce seuil (généralement 10%), les règlements et les suspens ne sont pas stabilisés. Ce seuil s'applique sur l'indice lui-même et sera pris en compte dans la suite de l'étude. La transformation de l'équation 3.10 est décrite dans l'équation 3.11.

$$CoefStab_{i,ult} = \frac{\sum_{j=0}^{ult} Reglement_{i,j}}{\sum_{j=0}^{ult} Reglement_i \times \left\{ \begin{array}{ll} \frac{I_{surv}}{I_{surv+j}} & si \ \frac{I_{surv+j}}{I_{surv}} - 1 \geq seuil \\ 1 & sinon \end{array} \right\}} \tag{3.11}$$

Finalement, le coefficient de stabilisation calculé individuellement est appliqué aux bornes du traité avant application à la charge sinistre (eq. 3.12).

$$\begin{aligned} Priorité_Stab &= Priorité \times CoefStab_{s,i,d} \\ Portée_Stab &= Portée \times CoefStab_{s,i,d} \end{aligned} \tag{3.12}$$

3.10 Application de la réassurance sur les sinistres individuels

Pour une simulation s, un sinistre i et un décalage d donné, on dispose désormais des règlements passés et futurs ainsi que des coefficients de stabilisation associés. Afin d'obtenir les provisions cédées, les cessions cumulées puis décumulées nominales (sans actualisation) doivent être calculées dans un premier temps (eq. 3.13).

$$\begin{aligned} Cessions_Cumulées_{s,i,d} &= min(max(Règlements_Cumulés - Priorité_Stab, 0), Portée_Stab) \\ Cessions_{s,i,d} &= Cessions_Cumulées_{s,i,d} - Cessions_Cumulées_{s,i,d-1} \end{aligned} \tag{3.13}$$

3.11 Calcul des provisions cédées

Les provisions cédées s'obtiennent en réalisant la somme actualisée des cessions décumulées (eq. 3.14).

$$\begin{aligned} Provision_Cédée_{s,i} &= \sum_{d=diag+1}^{ultime} \frac{Cessions_{s,i,d}}{(1+t_{actu})^d} \\ Provision_Cédée_s &= \sum_{i=1}^{Nsinistres} Provision_Cédée_{s,i} \end{aligned} \tag{3.14}$$

Chapitre 4

Calibration et validation des modèles

4.1 Objectif

Comme cela a été présenté dans le chapitre 3, afin de pouvoir développer la charge d'un sinistre au décalage suivant, des calibrages de lois de probabilité doivent être effectués. Le premier calibrage concerne les lois de probabilité prédisant l'évolution du sinistre (boni, stable ou mali) tandis que le second concerne les lois de probabilité $\mathscr{B}$ et $\mathscr{M}$ qui prédise la charge au décalage suivant en prenant en compte son état actuel et son évolution.

4.2 Modèle multinomial d'évolution du sinistre

Cette section s'appuie en partie sur le cours de [Hua et al., 2010] sur la régression logistique multinomiale. La variable à prédire, ici le type d'évolution du sinistre, est catégorielle (boni, stable et mali). Une régression logistique binomiale est indiquée pour traiter les variables catégorielle à deux issues. Cette méthode fournit la probabilité d'appartenance de la variable à une catégorie. La régression linéaire s'effectue sur la transformation logit de cette probabilité (eq. 4.1). Ce modèle suppose donc que la relation entre la transformation logit de la probabilité p d'appartenance à une catégorie et les variables explicatives est linéaire.

$$\begin{aligned} logit(p) &= ln(\frac{p}{1-p}) = \beta_0 + \beta_1 x + \beta_2 d \\ p &= \frac{exp(\beta_0 + \beta_1 x + \beta_2 d)}{1 + exp(\beta_0 + \beta_1 x + \beta_2 d} \end{aligned} \tag{4.1}$$

Une régression logistique binomiale peut être représentée par un réseau de neurones à une couche (fig. 5).

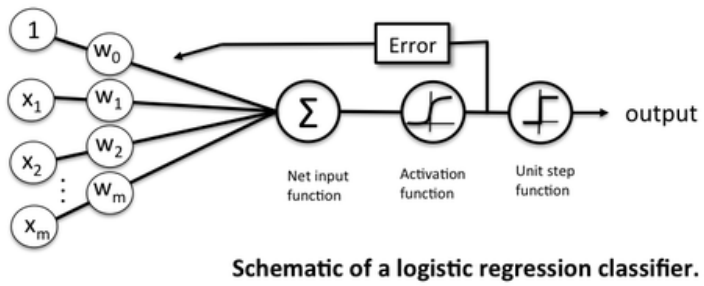
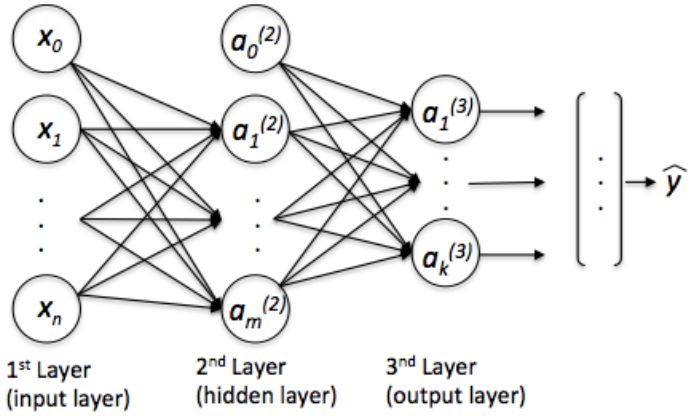


Figure 5 – Illustration d'une régression logistique binomiale représentée par un réseau de neurones à une couche [Rachka., 2020].

La régression logistique multinomiale est une extension de la régression logistique binomiale permettant d'étendre au delà de deux le nombre d'issues possibles pour la variable catégorielle. Elle peut être réalisée par un réseau de neurones avec une couche cachée (fig. 6). Soit y le résultat de la régression via le réseau de neurone pour un échantillon (x,d) donné. Ce résultat doit être transformé via la fonction softmax qui est une généralisation de l'équation eq. 4.1 au cas multinomial. Cette transformation permet d'obtenir les probabilités que l'échantillon de réponse y appartiennent aux différentes classes i. Cette transformation est décrite par l'équation 4.2.



Schematic of a multi-layer perceptron.

FIGURE 6 – Illustration d'un réseau de neurones avec une couche cachée [Rachka., 2020].

$$P(E = i|y) = \Phi_{softmax}(y_i) = \frac{e_i^y}{\sum_{m=1}^{M} e_m^y} \tag{4.2}$$

4.3 Modèle de prédiction de la charge au décalage suivant

Les lois $\mathscr{B}$ et $\mathscr{M}$ peuvent être calibrées sur les données sinistres disponibles à l'aide de différentes méthodes. Ces méthodes consistent toutes à réaliser une régression de la charge (ou d'une transformation de celle-ci) en fonction de la charge au décalage précédent et du décalage actuelle. On note Y la variable prédite. Plusieurs manières de définir cette variable sont présentées dans l'équation 4.3. Elle peut être définie par la charge sinistre, son accroissement absolu ou bien son accroissement relatif.

$$\begin{aligned} & Y_i = X_{i,d+1} \\ ou \quad & Y_i = X_{i,d+1} - X_{i,d} \\ ou \quad & Y_i = X_{i,d+1}/X_{i,d} - 1 \end{aligned} \tag{4.3}$$

Ces transformations permettent de placer la variable d'origine dans le cadre d'application de la méthode choisie. En effet, certaines lois de probabilité requièrent que la variable prédite soit comprise dans l'intervalle $[0, +\infty]$ ou bien $[0, 1]$, ou $]0, 1[$, etc.

4.3.1 Méthode paramétrique : Modèle Additif Généralisé linéaire paramétrique

Cette section s'appuie sur les travaux de [Rigby et Stasinopoulos, 2010] qui développent une approche générale des régressions paramétriques et semi-paramétriques dénommée *GAMLSS* (*Generalized Additive Models for Location, Scale and Shape*).

4.3.1.1 Théorie

La première méthode étudiée est dite paramétrique car la régression est réalisée en postulant une loi de probabilité théorique puis en estimant ses paramètres en fonction des variables explicatives. Une approche simpliste consiste à réaliser une régression linéaire simple (LM). L'équation 4.4 présente les caractéristiques principales d'un modèle LM tandis que la figure 7 illustre sur un exemple les résultats du modèle.

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_{i,d} + \beta_2 d + \epsilon_{i,d} \\ \epsilon_{i,d} &\sim \mathscr{N}(0, \sigma) \end{aligned} \tag{4.4}$$

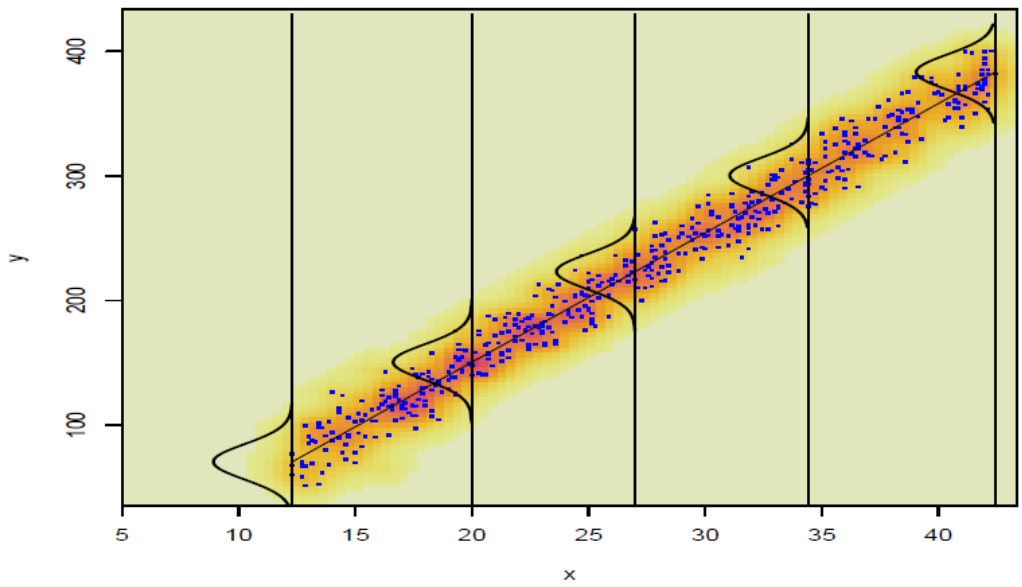


FIGURE 7 – Exemple d'une régression linéaire simple [Rigby et Stasinopoulos, 2010]

Une première limitation de la régression linéaire simple est que la variable prédite doit avoir une relation strictement linéaire avec les variables explicatives. De plus, la variable prédite est supposée suivre une loi normale de variance constante. La variance peut tout de même être variable en affectant des poids aux valeurs explicatives mais la distribution reste normale.

Le modèle linéaire généralisé (GLM) permet de s'affranchir en partie des limitations de la régression linéaire simple. Tout d'abord, la distribution normale de la variable prédite est remplacée par une famille de distribution notée EDF (Exponential Dispersion Family) tandis que le lien entre la variable prédite et les variables explicatives est défini par une fonction de lien monotone "g" qui permet la non linéarité. L'équation 4.5 présente les caractéristiques principales d'un modèle GLM tandis que la figure 8 illustre sur un exemple les résultats du modèle.

$$\begin{aligned} Y_i &\sim EDF(\mu_{i,d}, \Phi) \\ g(\mu_{i,d}) &= \beta_0 + \beta_1 x_{i,d} + \beta_2 d \end{aligned} \tag{4.5}$$

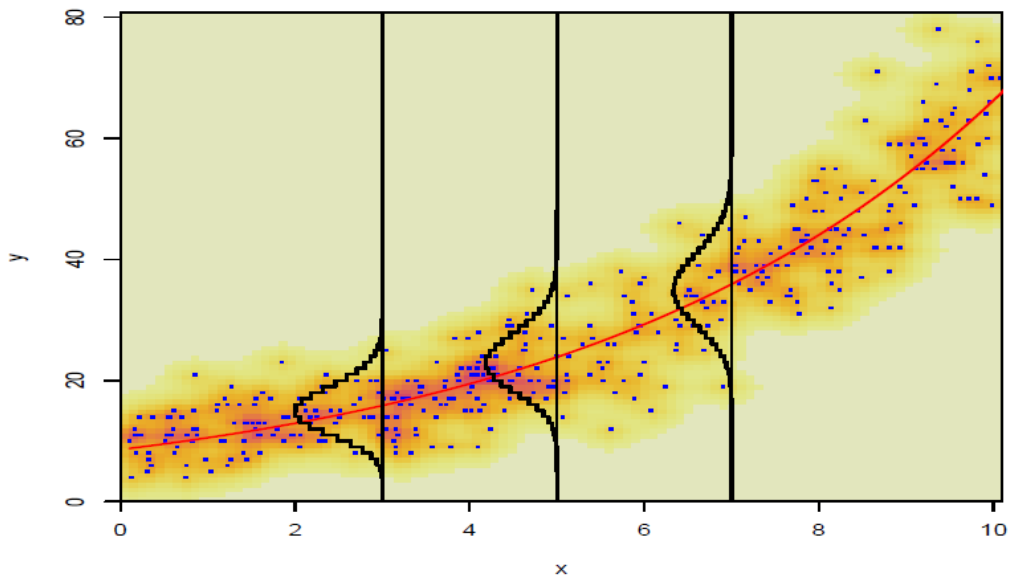


FIGURE 8 – Exemple d'une régression linéaire généralisée [Rigby et Stasinopoulos, 2010]

Les lois faisant parties de la famille exponentielle permettent de couvrir un grand nombre de cas avec des variances non constantes et la fonction de lien "g" permet de décrire des relations non linéaires. Néanmoins, la famille de loi exponentielle ne permet pas de décrire toutes les distributions possibles de la variable prédite.

Le modèle additif généralisé (GAM) permet d'élargir le cadre d'application d'une régression paramétrique en s'affranchissant de l'hypothèse de distribution appartenant à la famille exponentielle en la substituant par une famille plus générale de distribution. De plus, la régression peut s'effectuer non seulement sur l'espérance mais aussi sur d'autres paramètres comme la variance permettant une paramétrisation plus fine du modèle de régression. L'équation 4.6 présente les caractéristiques principales d'un modèle GAM tandis que la figure 9 illustre sur un exemple les résultats du modèle.

$$\begin{aligned} \theta_{i,d} &= \mu_{i,d}, \sigma_{i,d}, \nu_{i,d}, \tau_{i,d} \\ Y_i|\theta_{i,d} &\sim D(\theta_{i,d}) \end{aligned} \tag{4.6}$$

La moyenne μ ainsi que l'écart type σ possèdent leurs propres paramètres de régressions β et leur propre fonction de lien g (eq. 4.7).

$$\begin{aligned} g1(\mu_{i,d}) &= \beta_{1,0} + \beta_{1,1}x_{i,d} + \beta_{1,2}d \\ g2(\sigma_{i,d}) &= \beta_{2,0} + \beta_{2,1}x_{i,d} + \beta_{2,2}d \end{aligned} \tag{4.7}$$

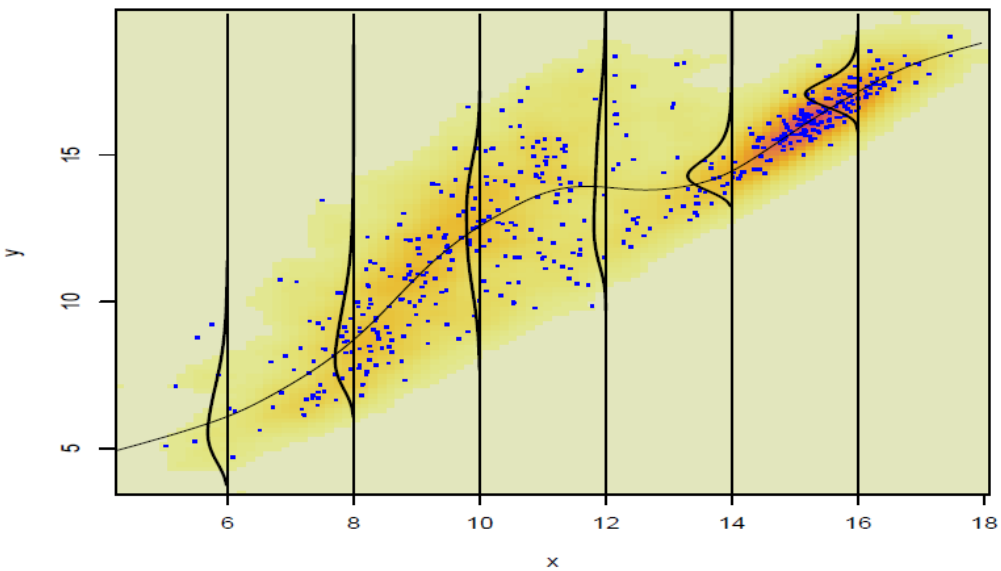


FIGURE 9 – Exemple d'une régression additive généralisée [Rigby et Stasinopoulos, 2010]

Cette formulation est celle d'un modèle GAM linéaire paramétrique. En effet, le cadre général du modèle GAM permet l'ajout d'un terme additif ("A" de GAM) non paramétrique aux expressions des g_k ce qui aboutit à un modèle semi-paramétrique. De plus, des combinaisons non linéaires des paramètres peuvent être utilisées ce qui aboutit à un modèle non-linéaire semi-paramétrique. Néanmoins, seule la première formulation linéaire paramétrique du GAM est utilisée dans la suite de l'étude.
Pour l'ensemble des modèles présentés, les paramètres β sont déterminés par maximisation de la log-vraisemblance. La fonction de vraisemblance représente la probabilité d'observer les données d'apprentissage avec les paramètres du modèle. La log-vraisemblance est, comme son nom l'indique, le log de la fonction de vraisemblance. Sa maximisation permet d'obtenir, pour un modèle donné, les paramètres les plus vraisemblables du modèle pour générer les données d'apprentissage.

4.3.1.2 Choix des lois

Le modèle GAM permet d'utiliser une grande variété de loi de probabilité. Etant donné la partition en boni et mali de la base d'apprentissage, deux lois de probabilité doivent être choisies. Le critère de sélection AIC (Akaike Information Criterion) permet de faire ce choix (eq. 4.8).

$$AIC = -2log(\hat{L}) + 2 \times df \tag{4.8}$$

$\hat{L}$ représente la vraisemblance calculée du modèle pour les données d'apprentissage et df représente le nombre de degrés de liberté du modèle. Le critère AIC doit être minimal, il s'agit donc d'un compromis entre la log-vraisemblance du modèle et son nombre de degrés de liberté. En effet, augmenter le nombre de degrés de liberté de la loi choisie permet de calibrer un modèle au plus proche des données d'apprentissage mais peut conduire à un sur-apprentissage en faisant perdre au modèle sa qualité de généralisation. C'est pourquoi le nombre de degrés de liberté est pénalisé.

Loi des bonis

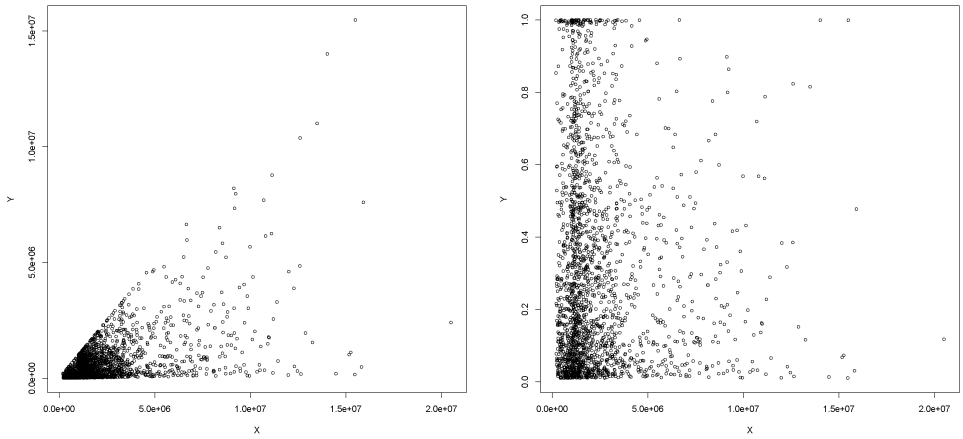
Soit $y_{i,d}$ le boni du sinistre i au décalage d, $x_{i,d}$ sa charge actuelle et $x_{i,d+1}$ sa charge au décalage suivant ($x_{i,d+1} < x_{i,d}$), un boni est défini par l'équation 4.9.

$$y_{i,d} = x_{i,d} - x_{i,d+1} \tag{4.9}$$

La variable $Y_{i,d}$ appartient à $]0, X_{i,d}]$, le boni ne pouvant excéder la charge actuelle du sinistre. $Y_{i,d}$ est donc borné à gauche et à droite avec une borne supérieure variable. Afin de rendre la borne supérieure constante, on normalise la variable prédite en divisant celle-ci par la charge actuelle $x_{i,d}$ (eq. 4.10).

$$y_{i,d} = \frac{x_{i,d} - x_{i,d+1}}{x_{i,d}} \tag{4.10}$$

La figure 10 représente un exemple de bonis avant et après normalisation.



(a) Avant normalisation. (b) Après normalisation.

FIGURE 10 – Exemple de bonis avant et après normalisation.

$Y_{i,d}$ appartient désormais à l'intervalle $]0, 1]$ ce qui permet d'utiliser l'ensemble des lois à support $]0, 1]$ (fig. 13).

Distributions	Family	Parameters	Skewness	Kurtosis
Beta	BE	2	(both)	
Logit normal	LOGITNO	2	(both)	
Generalized beta type 1	GB1	4	(both)	(both)

FIGURE 11 – Lois continues de support (0,1) [Rigby et Stasinopoulos, 2010].

Loi des malis

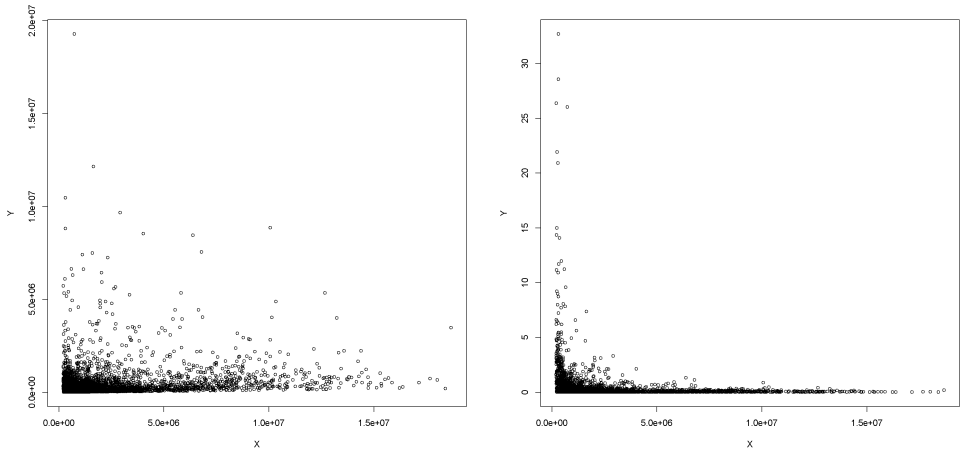
Soit $y_{i,d}$ le mali du sinistre i au décalage d, $x_{i,d}$ sa charge actuelle et $x_{i,d+1}$ sa charge au décalage suivant ($x_{i,d+1} > x_{i,d}$), un mali est défini par l'équation 4.11.

$$y_{i,d} = x_{i,d+1} - x_{i,d} \tag{4.11}$$

La variable $Y_{i,d}$ appartient à $]0, \infty[$, un mali pouvant théoriquement être sans borne supérieure. De la même manière que pour les bonis, il peut être utile de définir la variable prédite comme la variation relative de la charge plutôt que la variation absolue de celle-ci en divisant celle-ci par la charge actuelle $x_{i,d}$ (eq. 4.12.

$$y_{i,d} = \frac{x_{i,d+1} - x_{i,d}}{x_{i,d}} \tag{4.12}$$

La figure 12 représente un exemple de malis avant et après transformation.



(a) Avant normalisation. (b) Après normalisation.

FIGURE 12 – Exemple de malis avant et après normalisation.

Le choix d'une loi peut ensuite être fait parmi les lois à support $]0, \infty[$ (fig. **??**).

Distributions	family	no parameters	skewness	kurtosis
Box-Cox Cole and Green	BCCG	3	both	-
Box-Cox Power Exponential	BCPE	4	both	both
Box-Cox-t	BCT	4	both	lepto
Exponential	EXP	1	(positive)	-
Gamma	GA	2	(positive)	-
Generalized Beta type 2	GB2	4	both	both
Generalized Gamma	GG-GG2	3	positive	-
Generalized Inverse Gaussian	GIG	3	positive	-
Inverse Gaussian	IG	2	(positive)	-
Log Normal	LOGNO	2	(positive)	-
Log Normal family	LNO	2 + (1 fixed)	positive	
Reverse Generalized Extreme	RGE	3	positive	-
Weibull	WEI-WEI3	2	(positive)	-

FIGURE 13 – Lois continues à support $(0, \infty)$ [Rigby et Stasinopoulos, 2010].

4.4 Méthode non paramétrique : Régression quantile par forêt d'arbres décisionnels

Cette section s'appuie principalement sur les travaux de [Meinshausen, 2006] qui développent une approche des régressions quantiles par forêt d'arbres décisionnels.

4.4.1 Généralité

La deuxième méthode étudiée est dite non-paramétrique car la régression est réalisée sans faire d'hypothèse de famille de lois de probabilité paramétriques pour la variable prédite. En effet, plutôt que de supposer une distribution théorique de la variable prédite, il est possible de réaliser directement une régression sur les quantiles de la variable à prédire.

4.4.2 Théorie

4.4.2.1 Régression Quantile

Une régression quantile cherche à estimer la fonction de répartition de la variable prédite Y conditionnellement aux variables explicatives X et D. Cette fonction est définie par l'équation 4.13.

$$F(y|(X = x, D = d)) = \mathbb{P}(Y \leq y|(X = x, D = d)) \tag{4.13}$$

La fonction quantile $Q_\alpha(x, d)$ se définit comme la valeur minimale des y pour lesquels la fonction de répartition conditionnelle $F(y|(X = x, D = d))$ est supérieure ou égale à α (eq. 4.14).

$$Q_\alpha(x, d) = inf\{y : F(y|(X = x, D = d)) \geq \alpha\} \tag{4.14}$$

Une régression sur les quantiles permet d'obtenir la distribution de la variable prédite.

4.4.2.2 Arbre de décision

Les arbres de décision (AD) sont une catégorie d'arbres utilisée dans l'exploration de données et en informatique décisionnelle. Ils emploient une représentation hiérarchique de la structure des données sous forme des séquences de décisions (tests) en vue de la prédiction d'un résultat ou d'une classe. Chaque individu (ou observation), qui doit être attribué(e) à une classe, est décrit(e) par un ensemble de variables qui sont testées dans les nœuds de l'arbre. Les tests s'effectuent dans les nœuds internes et les décisions sont prises dans les nœuds terminaux appelés feuilles. Un exemple illustrant la construction d'un arbre de décisions est (fig. 14).

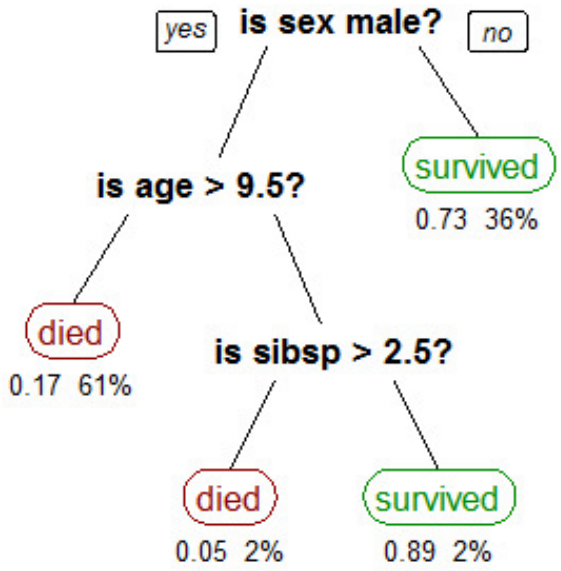


FIGURE 14 – Exemple d'arbre de décision : survie des passagers du Titanic [Crucianu, 2020]

.

Un arbre de décision est construit à partir des observations d'origine (eq 4.15).

$$(Y_i, X_i, D_i), i = 1, ..., n \tag{4.15}$$

Pour un nouveau jeu de variables explicatives (X=x,D=d), une seule feuille sera atteinte. On définit ω_i, le poids de l'observation originale i dans le calcul de la prédiction (eq. 4.16).

$$\omega_i(x, d) = \frac{1_{i \in Feuille(x,d)}}{\sum_{j=1}^{n} 1_{j \in Feuille(x,d)}} \tag{4.16}$$

La valeur prédite est obtenue en calculant la moyenne des observations Y_i contenues dans cette feuille (eq. 4.17).

$$\mu(x, d) = \sum_{i=1}^{n} \omega_i(x, d) Y_i \tag{4.17}$$

Un arbre de décision est très sensible au sur-apprentissage et des changements légers dans les données d'apprentissage peuvent conduire à des différences de prédictions importantes.

4.4.2.3 Algorithme des forêts d'arbres décisionnels (Random Forest)

L'algorithme des forêts d'arbres décisionnels est un algorithme d'apprentissage supervisé qui effectue un apprentissage sur de multiples arbres de décision entraînés sur des sous-ensembles de données légèrement différents. Il a pour but de réduire la variance de l'estimateur, en d'autres termes de corriger l'instabilité des arbres de décision (fig. 15).

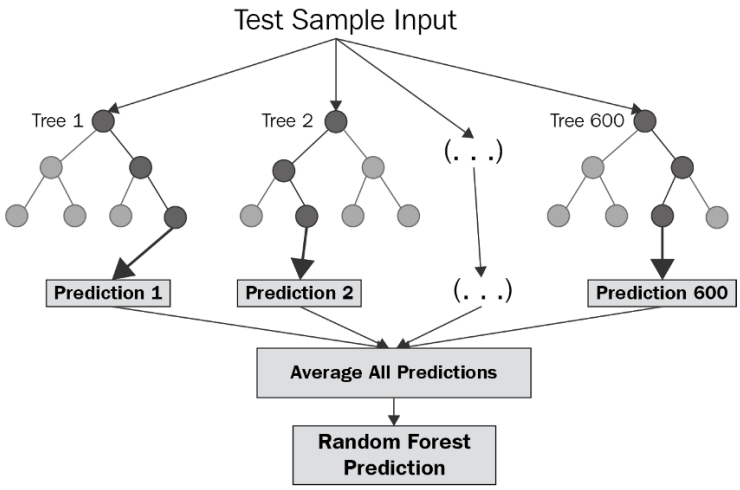


FIGURE 15 – Illustration d'une forêt d'arbres décisionnels.

Pour ce faire, le principe du *bootstrap* est utilisé en créant de nouveaux échantillons par tirage aléatoire avec remise d'observations i dans l'échantillon d'origine de taille N. Un grand nombre d'arbres de décision (K arbres) sont créés à partir de ces sous-ensembles. Les estimateurs ainsi obtenus sont moyennés (lorsque les données sont quantitatives, cas d'un arbre de régression) ou utilisés pour un vote à la majorité (pour des données qualitatives, cas d'un arbre de classification). En pratique la prédiction de l'espérance de la variable prédite est obtenue en effectuant la moyenne des observations originales mais, contrairement à un arbre de décision unique, les poids utilisés pour réaliser cette moyenne sont calculés en réalisant une moyenne des poids $\omega_i(x, d, t)$ de chaque arbre t (eq. 4.18).

$$\omega_i(x, d) = \frac{1}{K} \sum_{t=1}^{K} \omega_i(x, d, t) \tag{4.18}$$

4.4.2.4 Régression quantile par forêt d'arbres décisionnels (Quantile Random Forest)

L'algorithme des forêts d'arbres décisionnels prédit une valeur moyenne en effectuant la moyenne des prédictions de chaque arbre mais il est aussi possible d'utiliser l'ensemble des valeurs prédites afin d'obtenir une prédiction de la distribution de la variable à prédire. L'espérance $\mathbb{E}(Y|(X = x, D = d))$ est remplacé par $\mathbb{E}(1_{\{Y \leq y\}}|(X = x, D = d))$ en utilisant les mêmes poids ω que pour l'algorithme des forêts d'arbres décisionnels (eq. 4.19 et 4.20).

$$F(y|(X = x, D = d)) = \mathbb{P}(Y \leq y|(X = x, D = d)) = \mathbb{E}(1_{\{Y \leq y\}}|(X = x, D = d)) \tag{4.19}$$

$$\hat{F}(y|(X = x, D = d)) = \sum_{i=1}^{n} \omega_i(x, d) 1_{\{Y_i \leq y\}} \tag{4.20}$$

On obtient finalement les quantiles désirés en remplaçant $F(y|(X = x, D = d))$ par $\hat{F}(y|(X = x, D = d))$ dans l'équation 4.12.

4.4.2.5 Choix des hyperparamètres

La calibration d'une forêt d'arbres aléatoire, bien qu'étant non-paramétrique, nécessite de fixer des hyperparamètres qui contrôlent la forme et la taille de la forêt. Les paramètres à fixer sont le nombre d'arbres de la forêt et la taille minimale des feuilles de la forêt (*nodesize*). La taille minimale des feuilles définit le nombre d'observations que doit au minimum contenir chaque feuille d'un arbre. Ce paramètre limite la profondeur d'un arbre en empêchant une sélection trop fine des individus par découpage en nœuds successifs.
L'algorithme de construction d'une forêt aléatoire utilise un échantillon aléatoire des données d'apprentissage (environ 2/3) pour construire chaque arbre. Cet échantillon appelé *In Bag* est différent pour chaque arbre de la forêt. Les individus qui ne sont pas dans cet échantillon constitue l'échantillon *Out Of Bag* OOB. Ainsi pour chaque individu de la base d'apprentissage, une fraction des arbres de la forêt n'a pas utilisé cet individu dans leur construction. La prédiction de ces arbres pour cet individu constitue ce que l'on appelle l'erreur OOB. Le calcul de l'ensemble des erreurs OOB permet d'estimer les performances de la forêt en fonction des hyperparamètres choisis sans avoir à procéder à une validation croisée plus coûteuse en temps. Une fonction d'erreur ou fonction de perte traditionnellement utilisée est l'erreur quadratique moyenne (*Root Mean Square Error*, RMSE) qui mesure la différence entre les valeurs observées et les valeurs calculées (eq. 4.21).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} y_i - \hat{y}_i}{N}} \tag{4.21}$$

Dans notre cas, les quantiles observés et prédits sont des grandeurs d'intérêt principales. Or, la RMSE s'applique sur la valeur moyenne des prédictions et ne permet pas de rendre compte de la bonne estimation de la distribution des prédictions autour de la moyenne. La fonction de perte « Pinball » ou fonction de perte quantile permet de mesurer des écarts entre les quantiles observés et les quantiles prédits. Cette fonction $\mathscr{L}_\alpha$ est définie en fonction de la variable observée y et du quantile prédit $\hat{q}_\alpha$ pour la probabilité α (eq. 4.22).

$$\mathscr{L}_\alpha(y, \hat{q}_\alpha) = \begin{cases} \alpha(y - \hat{q}_\alpha) & si \quad y \geq \hat{q}_\alpha \\ (\alpha - 1)(y - \hat{q}_\alpha) & sinon \end{cases} \tag{4.22}$$

Une valeur plus faible de cette fonction pour un quantile donné signifie une meilleure précision pour la prédiction de ce quantile. La figure 16 illustre graphiquement les résultats de cette fonction selon l'écart de prédiction. Plus α est petit, plus les écarts positifs sont pénalisés et inversement plus α est grand, plus les écart négatifs sont pénalisés.

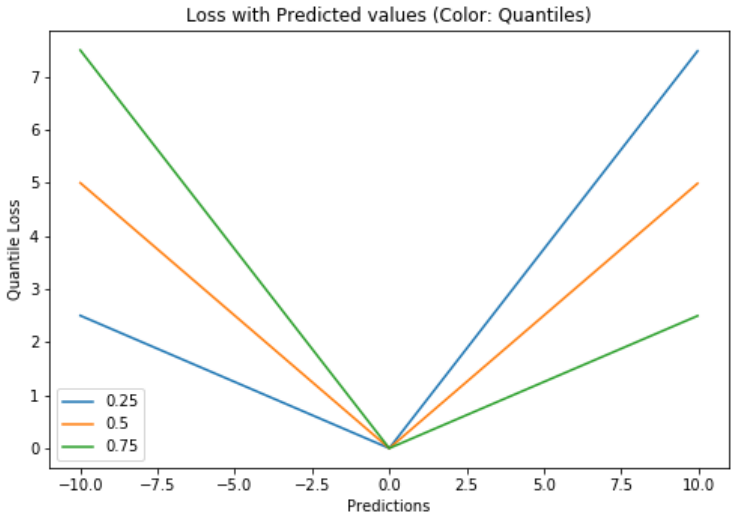


FIGURE 16 – Illustration de la fonction de perte quantile [Grover, 2018].

4.5 Validation des modèles

4.5.1 Validation sur les données d'apprentissage

La variable prédite Y_{id}, pour un sinistre i au décalage d, peut être décomposée en deux termes : une valeur moyenne et un résidu aléatoire centré autour de la valeur moyenne (eq. 4.23).

$$Y_{id} = \mu_{id} + \mathscr{E}_{id} \tag{4.23}$$

Par définition, l'espérance de Y_{id} est égale à μ_{id} et l'espérance du résidu $\mathscr{E}_{id}$ est nulle (eq. 4.24).

$$\begin{aligned} \mathbb{E}(\mathscr{E}_{id}) &= 0 \\ \mathbb{E}(Y_{id} &= \mu_{id}) \end{aligned} \tag{4.24}$$

La prédiction moyenne de cette variable par le modèle est notée $\hat{Y}_{id}$ (eq. 4.25).

$$\hat{Y}_{id} = \hat{\mu}_{id}(x, d) \tag{4.25}$$

Enfin l'erreur de prédiction est obtenue en calculant l'écart entre la valeur expérimentale et sa prédiction par le modèle (eq. 4.26). L'erreur sur la moyenne $(\mu_{id} - \hat{\mu}_{id}(x, d))$ est appelée erreur de paramétrage tandis que le terme de fluctuation $\mathscr{E}_{id}$ autour de la moyenne est appelé erreur de processus. En pratique, la fonction μ_{id} est une approximation de la véritable fonction qui détermine l'espérance de la variable aléatoire. Ce troisième terme d'erreur est appelé l'erreur de modèle et n'est pas traité par la suite (eq. 4.26).

$$
\begin{aligned}
e_{id} &= Y_{id} - \hat{Y}_{id} \\
e_{id} &= (\mu_{id} - \hat{\mu}_{id}(x,d)) + \mathscr{E}_{id} \\
e_{id} &= \textit{erreur de paramétrage} \ + \ \textit{erreur de processus}
\end{aligned}
\tag{4.26}
$$

La validation des modèles consiste donc à s'assurer que l'erreur de paramétrage soit la plus faible possible et que la calibration de l'erreur de processus, aléatoire par nature, soit fidèle à l'expérience.

4.5.2 Validation des modèles sur des données nouvelles : *Back Testing*

Après avoir validé que les modèles reproduisent correctement les données d'apprentissage, la capacité de prédiction des différents modèles doit être évaluée sur des données inconnues. Dans le cadre de séries temporelles, les données inconnues sont les évolutions futures des charges individuelles. La méthode du *Back Testing* consiste à exclure les années les plus récentes pour la phase d'apprentissage puis à prédire les années exclues avec les différents modèles. Les écarts observés entre les valeurs prédites et observées permettent alors de mesurer les performances des différents modèles. La figure 17 illustre le processus de *Back Testing* pour différentes profondeurs, sur un triangle comportant à l'origine 15 années de survenance. A mesure que la profondeur du *Back Testing* augmente, la quantité de données de référence disponible pour la comparaison diminue.

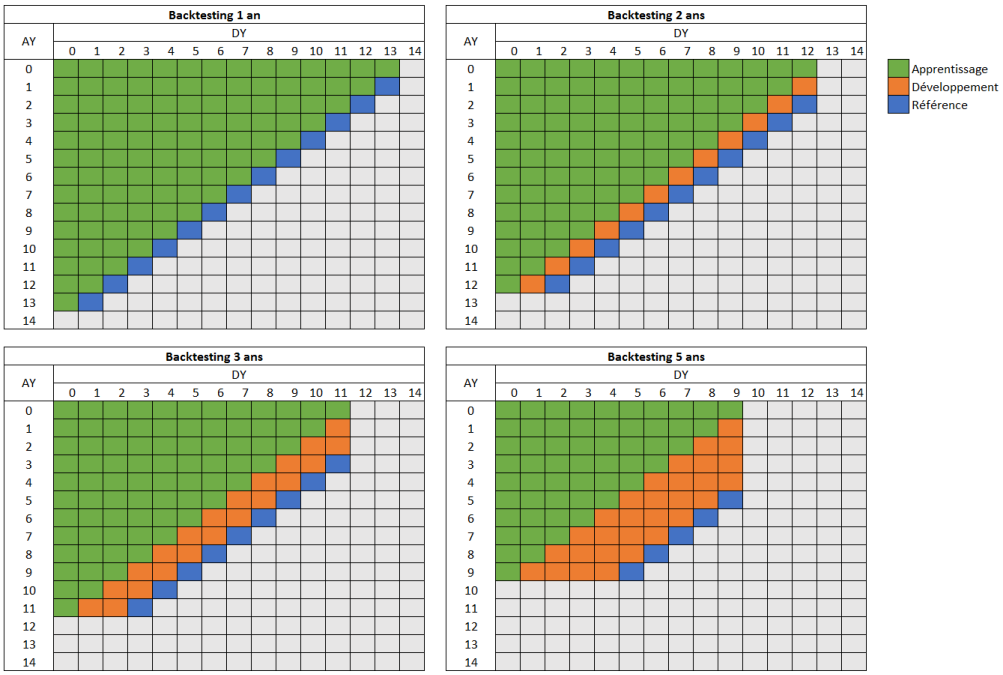


FIGURE 17 – Illustration du *Back Testing*.

Troisième partie

Application à un cas pratique

Chapitre 1

Développement de l'outil de provisionnement individuel sur R

L'application des concepts théoriques développés dans la partie précédente a nécessité le développement d'un outil sur R présenté ci-après.

1.1 Enjeux spécifiques à l'entreprise

Le mémoire s'est effectué au sein de l'entreprise AON, dans le service Analytics de l'entité Reinsurance Solutions. Son activité principale, le courtage en réassurance, nécessite le recours à des modélisations statistiques sur des périmètres variés. La motivation initiale du mémoire est donc de répondre aux enjeux spécifiques de l'entreprise en permettant l'utilisation des méthodes développées pour la réalisation des études de *Best Estimate* de cession ainsi que les commutations. C'est pourquoi au delà de l'aspect théorique, un outil complet permettant la réalisation de ces objectifs a été développé sur R.

1.2 Développement sur R

L'outil développé reprend l'architecture détaillée dans la partie théorique. Les *packages* R principaux utilisés sont les suivants :

- *nnet* pour la régression logistique multinomiale,
- *GAMLSS* pour le modèle Additif Généralisé linéaire paramétrique,
- *quantregForest* pour la régression quantile par forêt d'arbres décisionnels.

L'outil prend en entrée une base de sinistres individuels comportant les payés et suspens par année de survenance et par arrêté puis les tranches de réassurance définies par année de survenance. Il fournit en sortie les provisions brutes et cédées par traité et par réassureur pour l'ensemble des simulations réalisées.

Chapitre 2

Présentation et retraitement des données

Ce chapitre présente le cas étudié sur lequel sont appliqués les différentes méthodes de calcul des provisions brutes et cédées.

2.1 Présentation du cas étudié

La base sinistre étudiée est une base de responsabilité civile comportant principalement des sinistres de la branche Responsabilité Civile Automobile (> 90%), le reste en Responsabilité Civile Vie Privée (RCVP). Le tableau 2.1 présente les caractéristiques du cas étudié.

Table 2.1 – Présentation du cas étudié.

Années de survenance	2005-2019
Seuil de modélisation	1 M
Nombre de sinistres au dessus du seuil	1481
Tranche réassurance 1	4 M XS 6 M
Tranche réassurance 2	ILL XS 10 M
Seuil de stabilisation	1.1

L'indice utilisé pour la stabilisation des bornes des traités de réassurance et l'indexation des payés et suspens est l'indice INSEE des salaires pour la santé humaine (identifiant 001567402 remplacé en 2018 par 010562727). Pour les valeurs futurs de l'indice une hypothèse d'évolution de 1,5% par an est prise en accord avec les évolutions passées récentes de l'indice. L'évolution de l'indice est présentée sur la figure 18.

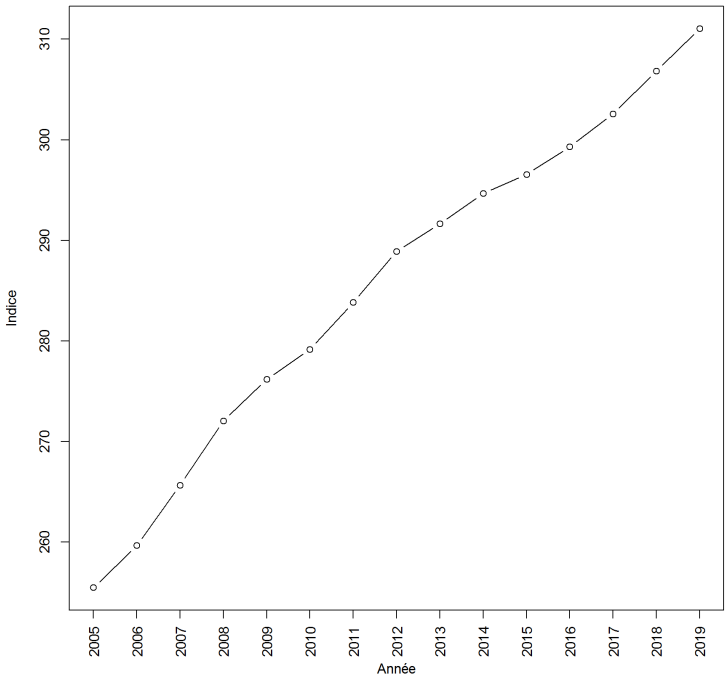


FIGURE 18 – Evolution de l'indice utilisé pour l'indexation et la stabilisation

2.2 Retraitement des données

La valeur relative du seuil de communication évolue au cours du temps en raison de l'inflation, il est donc nécessaire de relever ce seuil après l'indexation afin de garantir l'exhaustivité dans le temps de la statistique sinistre au delà du seuil de modélisation choisi. La figure 19 illustre ce propos en affichant la valeur effective du seuil de communication avant et après indexation. La dernière année de survenance est prise comme année de référence pour l'indexation. Le seuil de modélisation doit être a minima supérieur à la valeur maximale du seuil de communication indexé.

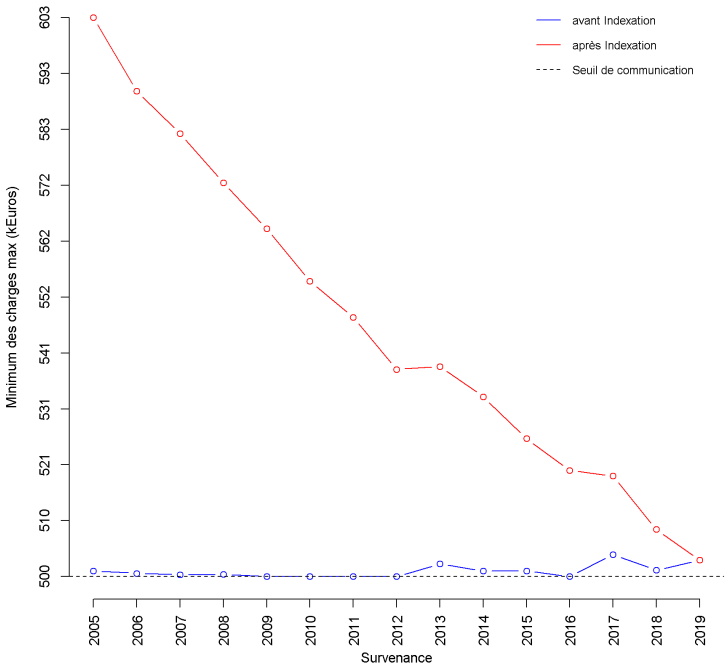
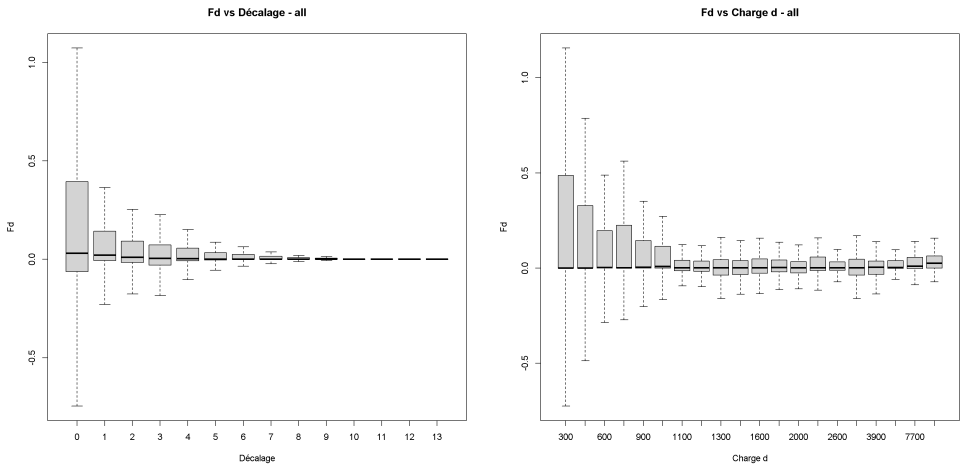


FIGURE 19 – Effet de l'indexation sur le seuil de communication

2.3 Statistiques descriptives

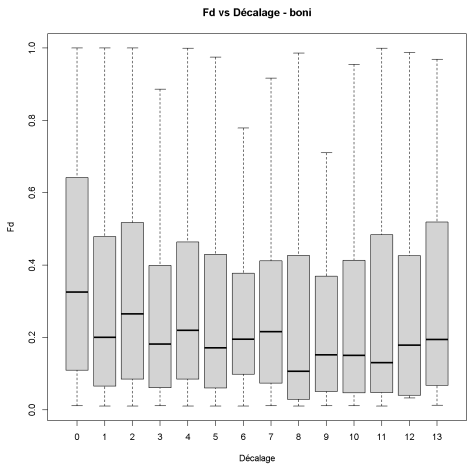
L'observation préliminaire de l'évolution des charges sinistre individuelles en fonction des variables explicatives permet de confirmer les choix de modélisation. Tout d'abord la figure 20 présente deux *boxplots* du facteur de développement individuel (défini ici comme la croissance relative de la charge individuelle) en fonction du décalage et de la charge. Le diagramme en fonction du décalage montre que la volatilité du facteur de développement décroît fortement en fonction du décalage. Ceci s'explique principalement par le fait que la probabilité d'avoir un boni ou un mali diminue fortement avec le nombre d'années suivant la survenance. Le choix de segmenter la méthode en une partie prédisant l'évolution du sinistre et en une autre prédisant son montant se justifie par cet aspect. Le diagramme en fonction de la charge montre une plus grande volatilité du facteur de développement pour les charges faibles ce qui est logique étant donné que ce facteur est défini relativement à la charge. De plus, les charges faibles évoluent principalement en mali tandis que l'évolution des charges élevées est plus équilibrée entre boni et mali.



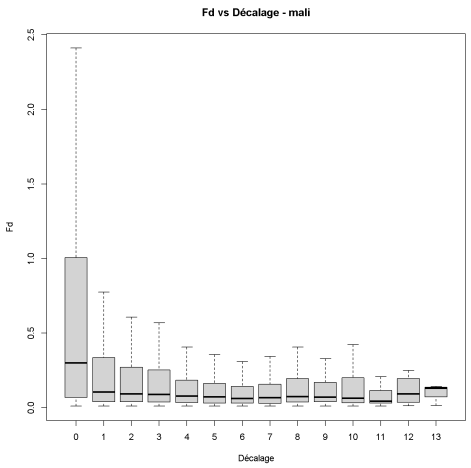
(a) En fonction du décalage. (b) En fonction de la charge.

FIGURE 20 – *Boxplots* du facteur de développement individuel.

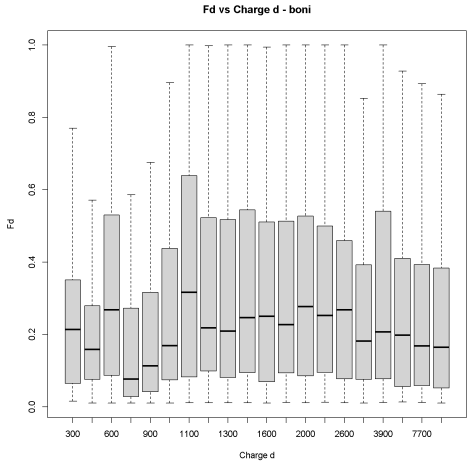
La figure 21 présente des *boxplots* du facteur de développement individuel en fonction du décalage ou de la charge pour les bonis et les malis seulement. La volatilité des bonis est stable en fonction du décalage tandis que la médiane décroît légèrement. La volatilité et la médiane des malis décroissent fortement sur les premiers décalages pour se stabiliser par la suite. De plus, la volatilité et la médiane des bonis sont plutôt stables en fonction de la charge. Les fluctuations observées pour les charges faibles s'expliquent par le peu de bonis présents dans la base pour ces charges. Enfin, la volatilité et la médiane des malis décroissent fortement en fonction de la charge.



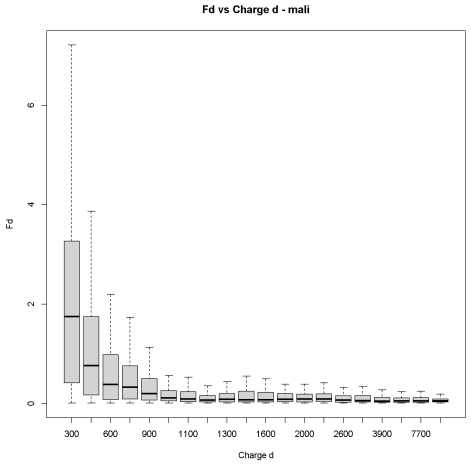
(a) Bonis en fonction du décalage.



(b) Malis en fonction du décalage.



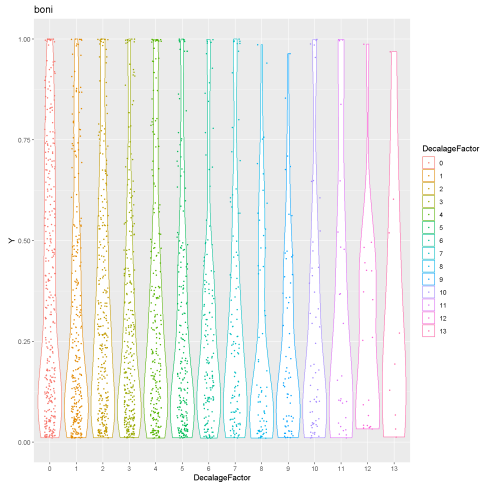
(c) Bonis en fonction de la charge.



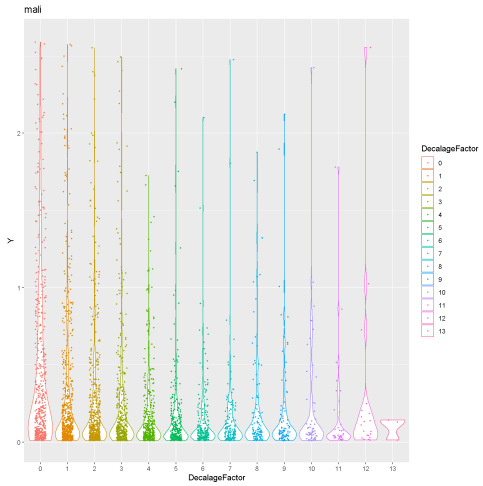
(d) Malis en fonction de la charge.

FIGURE 21 – *Boxplots* du facteur de développement individuel pour les bonis et malis.

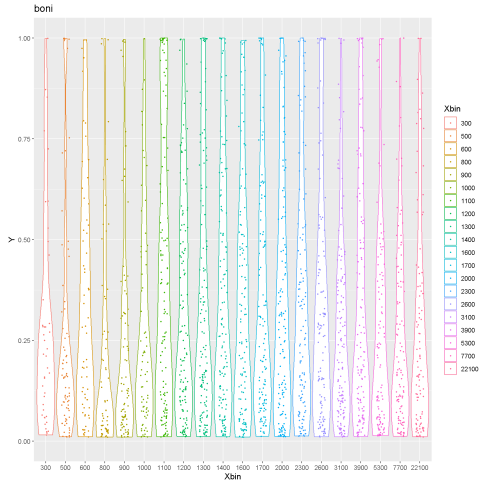
Afin d'observer en détail les densités de probabilité empiriques des facteurs de développement individuels en fonction des variables explicatives, des *violinplots* sont tracés. Ces diagrammes permettent l'observation sur une même figure des points de données et de la distribution de probabilité en fonction d'une variable explicative. La figure 22 présente des *violinplots* du facteur de développement individuel en fonction du décalage ou de la charge pour les bonis et les malis. Ce sont ces distributions que les modèles développés par la suite doivent reproduire tout en les généralisant à des données inconnues.



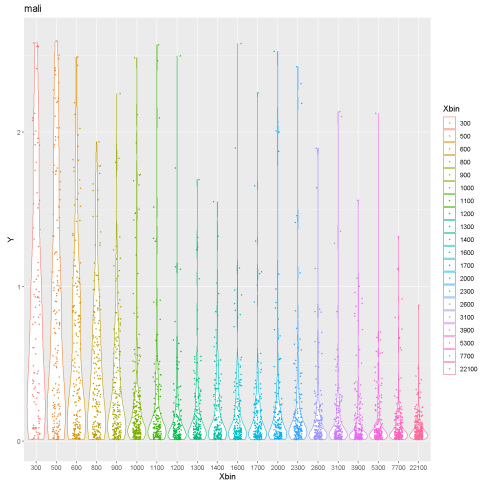
(a) Bonis en fonction du décalage.



(b) Malis en fonction du décalage.



(c) Bonis en fonction de la charge.



(d) Malis en fonction de la charge.

FIGURE 22 – Violinplots du facteur de développement individuel pour les bonis et malis.

Chapitre 3

Modèle Chain Ladder agrégé appliqué aux sinistres individuels

La première application pratique du calcul des provisions cédées pour le cas étudié est réalisée avec un modèle Chain Ladder agrégé appliqué aux sinistres individuels. Cette méthode simple sert de base de comparaison.

3.1 Calcul des coefficients

La méthode Chain Ladder présentée dans la partie théorique, chapitre. 2 est appliquée sur le triangle de charges individuelles. La figure 23 présentent les facteurs de développements (f) obtenus ainsi que l'évolution type de la charge jusqu'à l'ultime. Cette évolution représente pour chaque décalage l'écart de charge par rapport à la charge à l'ultime. Cette évolution est définie par l'équation 3.1.

$$Evolution(d) = \frac{1}{\prod_{k=d}^{ultime} f(k)} \tag{3.1}$$

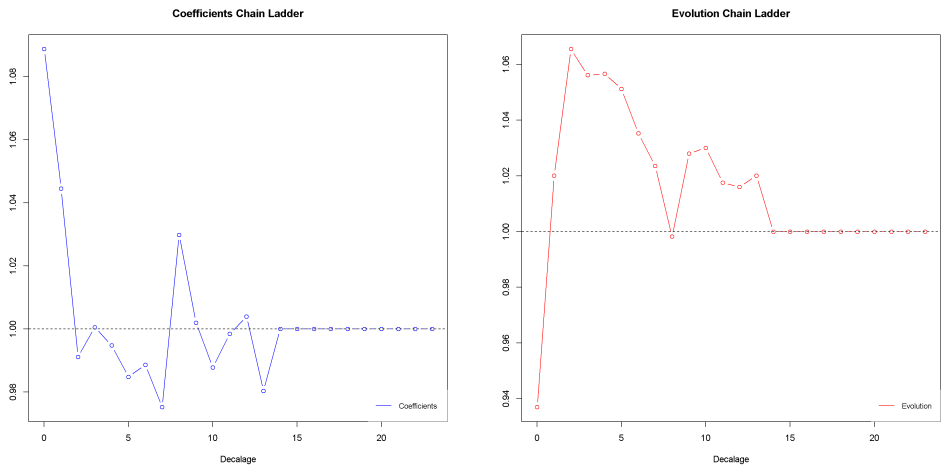


Figure 23 – Coefficients Chain Ladder et évolution de la charge en fonction du décalage.

3.2 Calcul des provisions

Les coefficients calculés précédemment sont appliqués individuellement en cascade pour remplir le triangle de charge individuel. La figure 24 présente les provisions brutes obtenues par année

de survenance ainsi que les provisions cédées (après application individuelle de la réassurance). Sur ces figures sont aussi représentées les provisions sans projection pour comparaison.

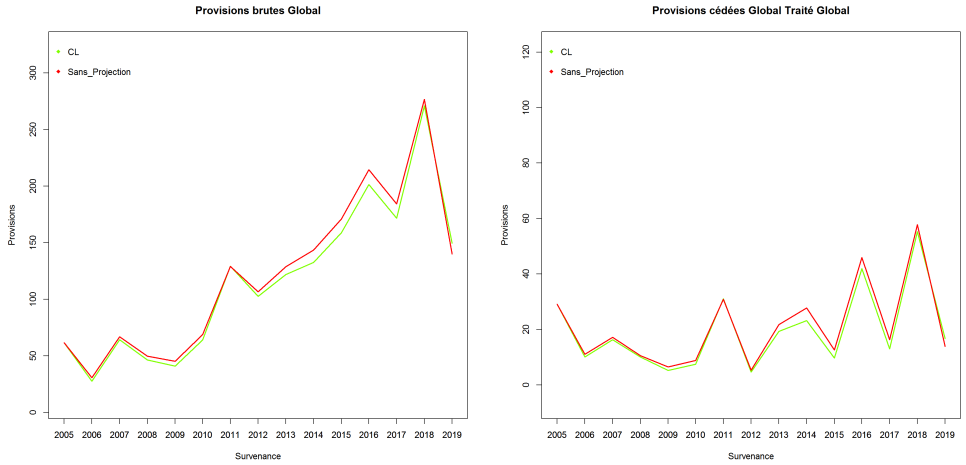


FIGURE 24 – Provisions brutes et cédées estimées en fonction de l'année de survenance avec la méthode Chain Ladder.

Chapitre 4

Application des modèles de provisionnement individuel stochastique

Ce chapitre présente les résultats de l'application des modèles de provisionnement individuel stochastique. Les différentes étapes de calcul sont développées, validées puis comparées entre les méthodes.

4.1 Modèle d'évolution en boni, stable ou mali

Comme présenté dans la partie théorique, la première étape consiste à calibrer le modèle d'évolution de la charge sinistre en boni, stable ou mali. La figure 25 représente l'ensemble des évolutions observées sur la base sinistre. La définition d'un stable est soumise à une marge de tolérance de $+/-1\%$ sur l'évolution de la charge afin de filtrer les évolutions mineures qui ne sont pas liées à une évolution significative du sinistre.

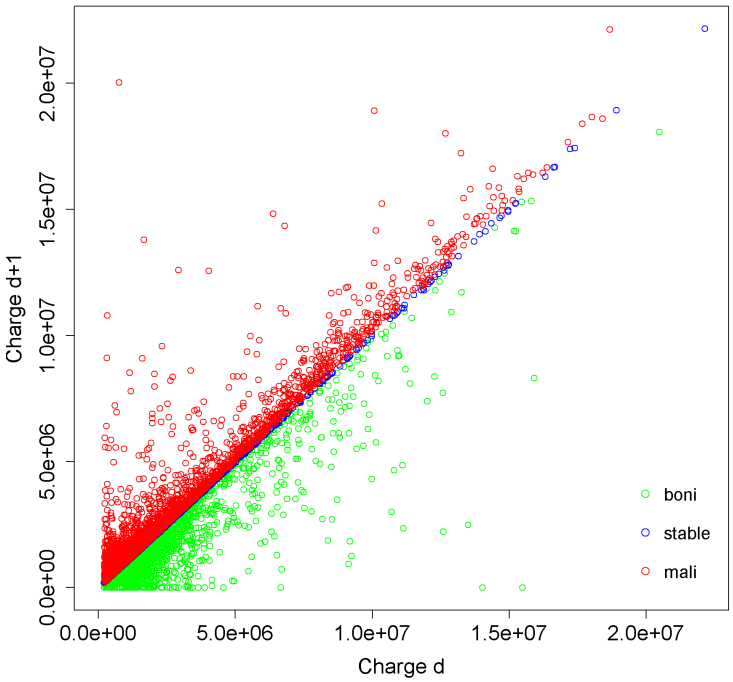
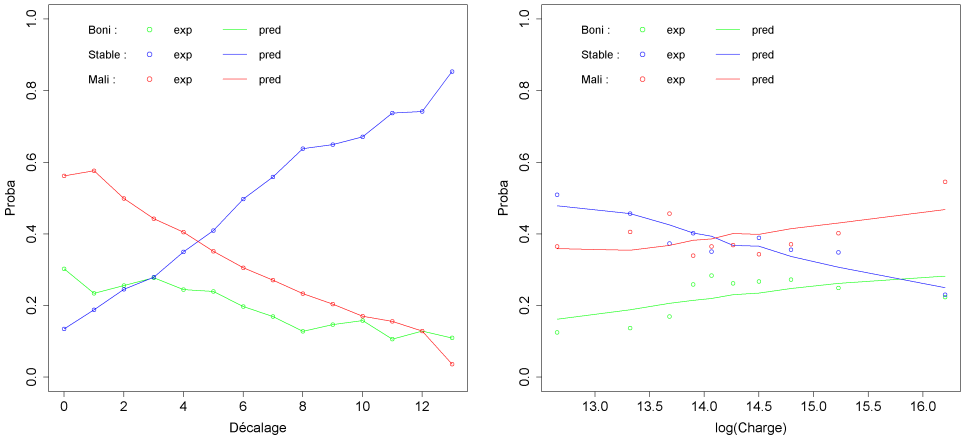


Figure 25 – Visualisation des bonis, stables et malis.

L'application du modèle multinomial permet d'obtenir les probabilités d'évolution en fonction de la charge et du décalage (fig. 26). Sans surprise, la probabilité d'évolution en boni ou en mali diminue avec la maturité du sinistre. De plus, les sinistres de charge élevée ont plus de chance de voir leur charge évoluer que ceux plus petits. Néanmoins, l'effet de la variable décalage est le plus significatif sur l'évolution d'un sinistre. Enfin, l'adéquation entre prédiction et expérience est bonne au niveau global.



(a) Fonction du décalage. (b) Fonction de la charge.

FIGURE 26 – Probabilité d'un boni, stable ou mali.

En annexe 4.5, la figure 48 présente les probabilités d'évolution en boni, stable ou mali en fonction de la charge pour des décalages spécifiques. L'adéquation entre prédiction et expérience reste bonne au niveau détaillé.

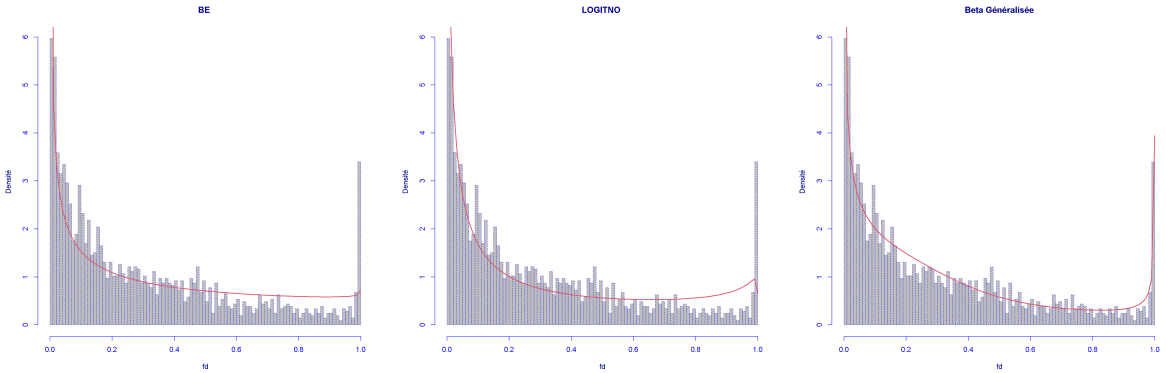
4.2 Modèle Additif Généralisé linéaire paramétrique

4.2.1 Choix des distributions

Comme présenté dans la partie II au chapitre 4, le choix des distributions de probabilité pour les lois représentant les bonis ou les malis est effectué en calculant les critères AIC pour chaque distribution et en sélectionnant les lois qui obtiennent le score le plus faible. Ce critère est calculé sur l'ensemble des données de sinistre sans distinction de décalage ou de charge. Les résultats obtenus pour les bonis sont présentés dans le tableau 4.1 et la figure 27. Le choix se porte sur une loi Beta Généralisée qui obtient le score le plus faible avec le plus grand nombre de degrés de liberté.

Table 4.1 – Choix de la distribution des bonis.

Lois	Param.	AIC
GB1	4	-1478
BE	2	-1146
LOGITNO	2	-1142



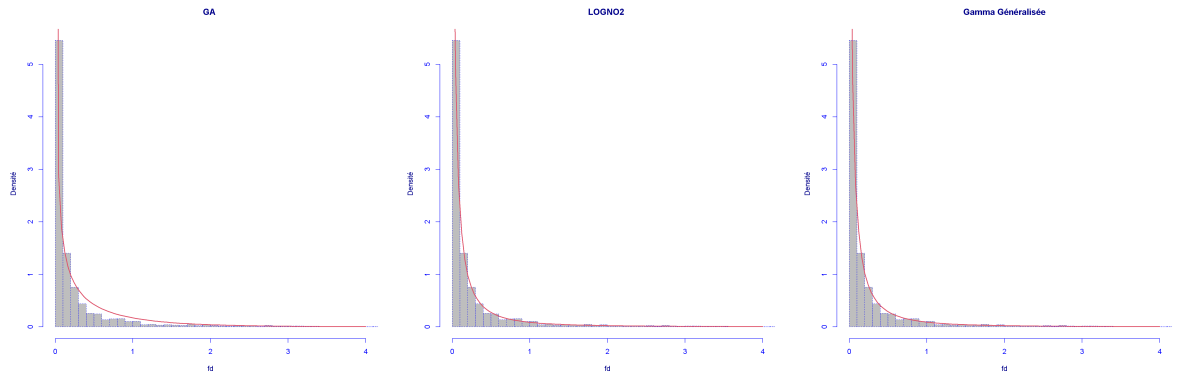
(a) Beta. (b) Logit-Normale. (c) Beta Généralisée.

Figure 27 – Choix de la distribution des bonis.

Les résultats obtenus pour les malis sont présentés dans le tableau 4.2 et la figure 28. Le choix se porte sur une loi Gamma Généralisée qui obtient le score le plus faible avec, elle aussi, le plus grand nombre de degrés de liberté.

Table 4.2 – Choix de la distribution des malis.

Lois	Param.	AIC
GG	3	-3097
LOGNO2	2	-3078
GA	2	-1834

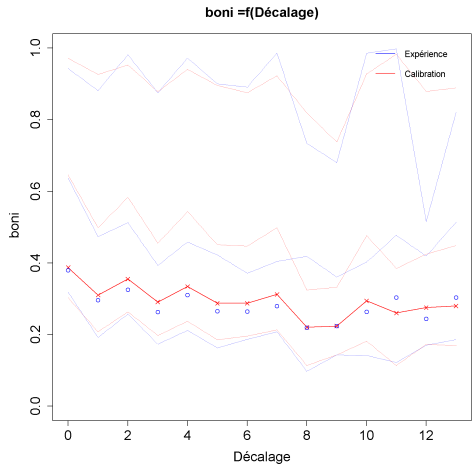


(a) Gamma. (b) Log-Normale. (c) Gamma Généralisée.

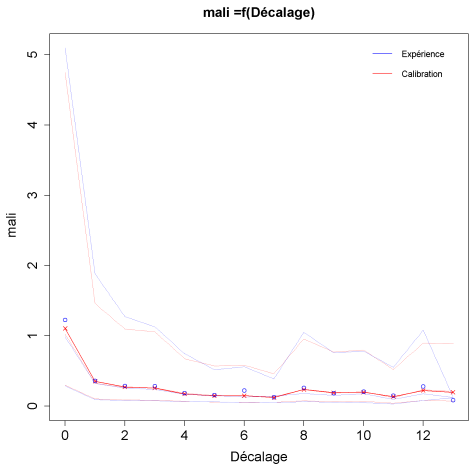
FIGURE 28 – Choix de la distribution des malis.

4.2.2 Validation générale sur les moyennes prédites

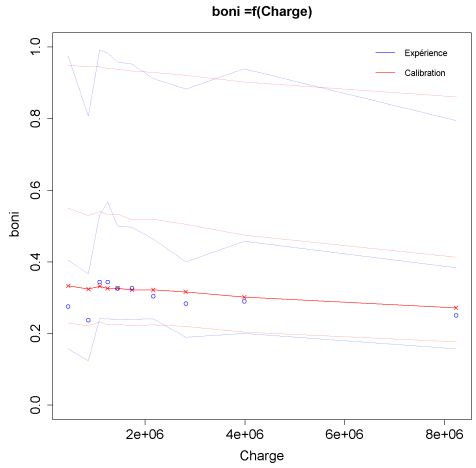
La première validation du modèle se fait au niveau global en comparant les moyennes prédites ainsi que quelques quantiles (50%, 75% et 95%) avec leurs équivalents expérimentaux (fig. 29). Une bonne adéquation se dégage de ces figures aussi bien sur les moyennes que sur les quantiles, pour les bonis et les malis.



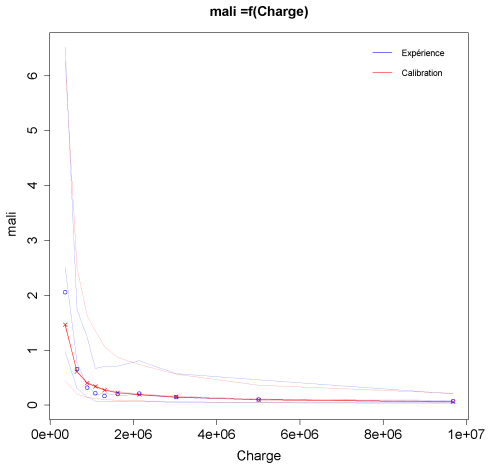
(a) Bonis par décalage



(b) Malis par décalage



(c) Bonis par charge



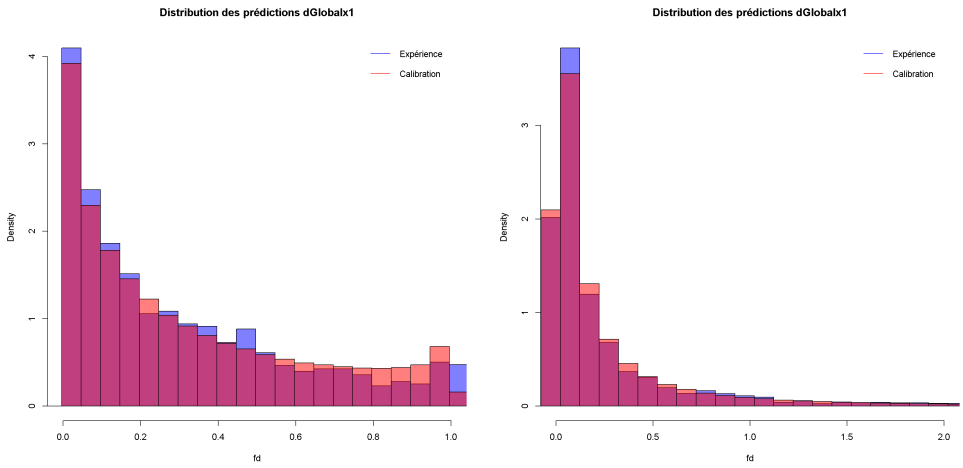
(d) Malis par charge

FIGURE 29 – GAM - Validation globale de la calibration des bonis et malis, moyenne et quantiles (50%, 75% et 95%).

En annexe 4.5, le détail de la comparaison des moyennes prédites en fonction de la charge pour une sélection de décalage est présenté sur les figures 49 et 50.

4.2.3 Validation détaillée des distributions

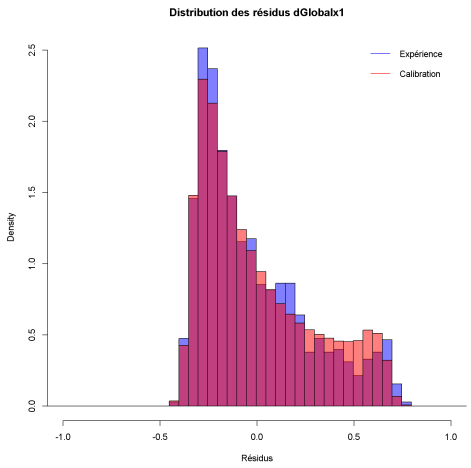
La seconde validation s'effectue sur la comparaison des distributions de la variable prédite et de ses résidus. La figure 30 présente la comparaison de la distribution de la variable prédite expérimentale et calibrée.



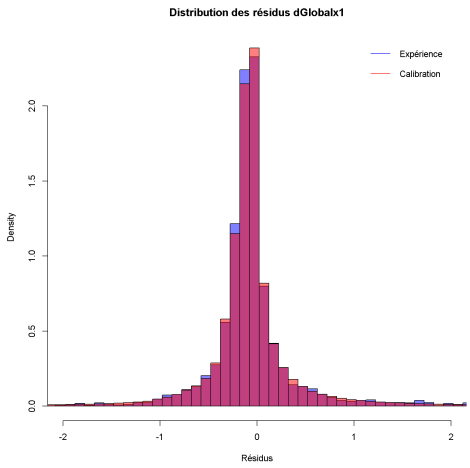
(a) boni (b) mali

FIGURE 30 – GAM - Comparaison de la distribution des prédictions avec l'expérience.

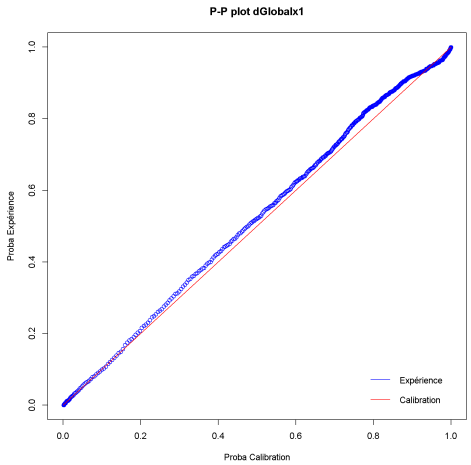
La figure 31 présente la distribution des résidus de la variable prédite expérimentale et calibrée ainsi que le diagramme probabilité-probabilité associé. L'adéquation est bonne sur l'ensemble des données.



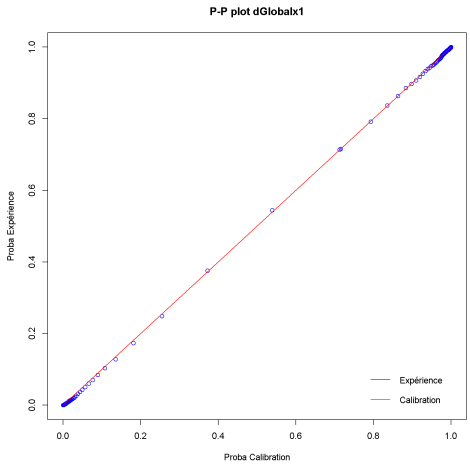
(a) Bonis - Distribution des résidus



(b) Malis - Distribution des résidus



(c) Bonis - diagramme PP



(d) Malis - diagramme PP

FIGURE 31 – GAM - Comparaison de la distribution des résidus avec l'expérience.

Enfin, en annexe 4.5, les figures 51, 52 et 53 présentent la distribution des résidus expérimentale et calibrée ainsi que le diagramme probabilité-probabilité associé pour des décalages spécifiques (0, 5 et 10). Les adéquations, bien que moins bonnes qu'au niveau global, restent satisfaisantes.

4.2.4 Calcul des provisions brutes et cédées

Les lois calibrées précédemment sont appliquées individuellement en cascade pour remplir stochastiquement plusieurs centaines de versions du triangle de charges individuelles. La figure 32 présente les provisions brutes et cédées par année de survenance tandis que la figure 33 présente la distribution des provisions brutes et cédées totales obtenues grâce aux développements stochastiques des triangles de charges individuelles. Sur ces figures sont aussi représentées les provisions sans projection ainsi que celles projetées avec la méthode Chain Ladder pour comparaison.

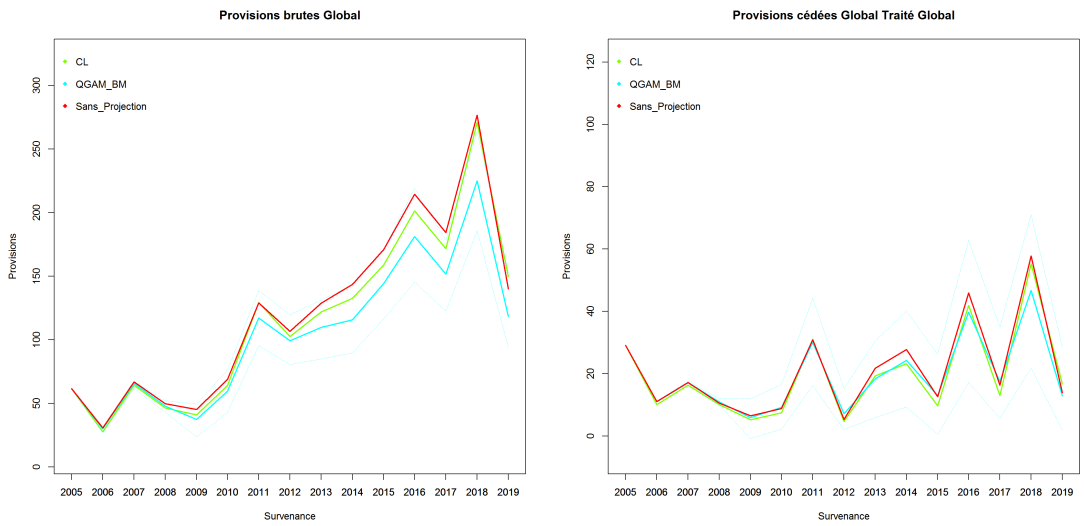


FIGURE 32 – GAM - Provisions estimées par survenance et quantiles 5% et 95%.

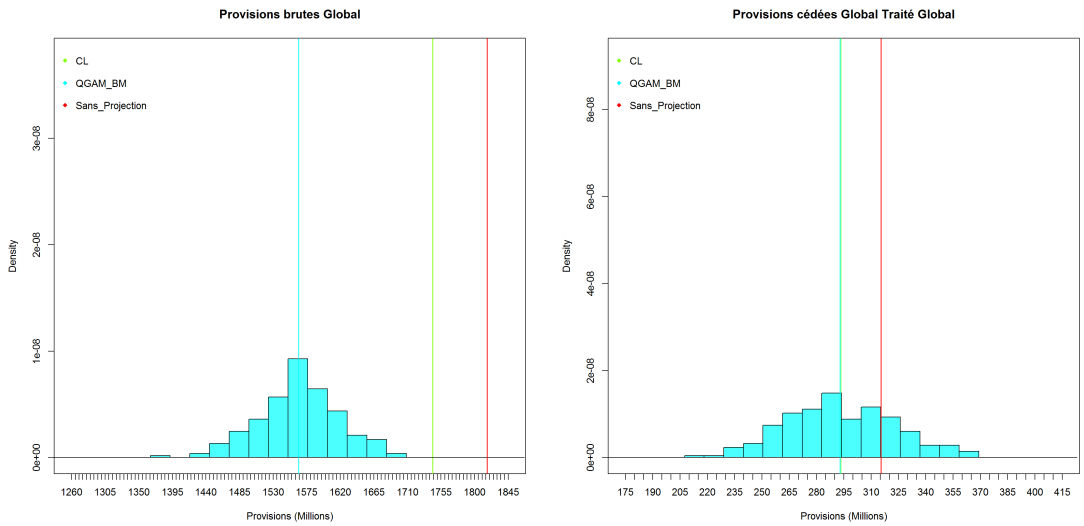
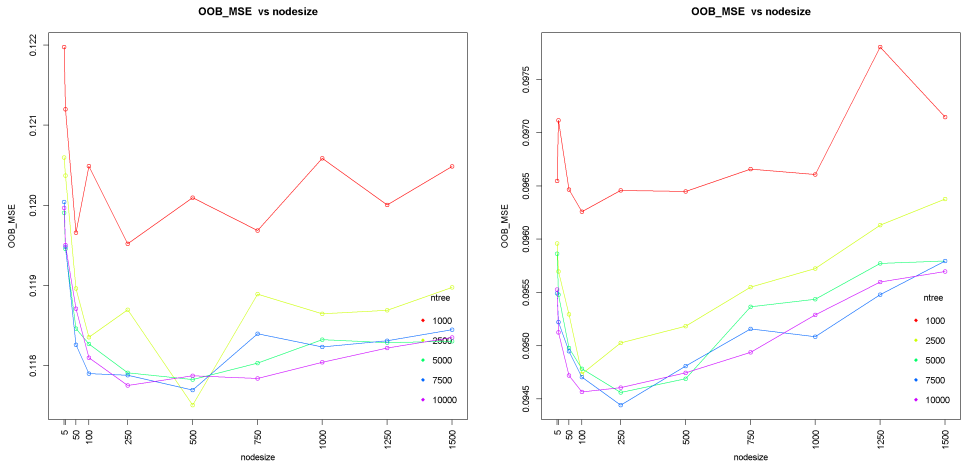


FIGURE 33 – GAM - Provisions totales estimées.

4.3 Modèle non-paramétrique : Régression quantile par forêt d'arbres décisionnels (*Quantile Random Forest*)

4.3.1 Choix des hyperparamètres

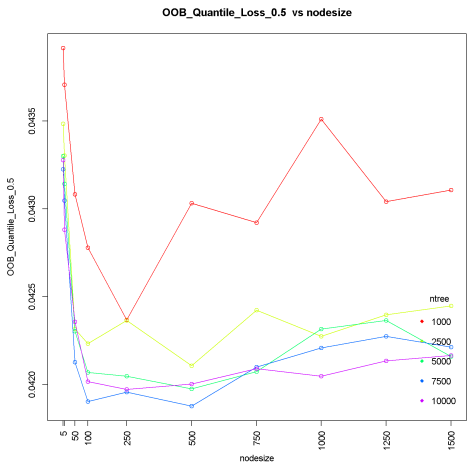
Comme présenté dans la partie théorique, la calibration d'une forêt d'arbres aléatoire, bien qu'étant non-paramétrique, nécessite de fixer des hyperparamètres qui contrôlent la forme et la taille de la forêt. La figure 34 présente les erreurs quadratiques moyennes et les fonctions de pertes quantiles calculées pour différentes configurations des hyperparamètres *ntree* (nombre d'arbre dans la forêt) et *nodesize* (nombre minimal d'individu que chaque noeud doit contenir). Pour le choix du nombre d'arbre, les gains sont faibles à partir de 5000 arbres. Pour le *nodesize*, l'optimum se trouve autour de 250.



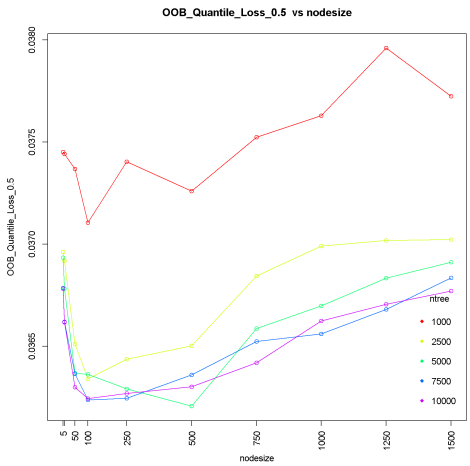
(a) Bonis par décalage (b) Malis par décalage

FIGURE 34 – QRF - Choix des hyperparamètres de la forêt aléatoire.

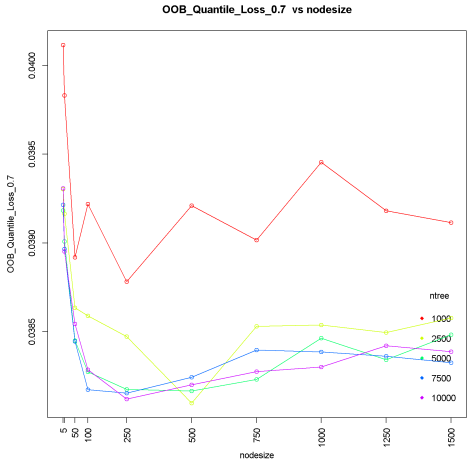
La figure 35 représente les fonctions de pertes quantiles calculées pour les mêmes configurations d'hyperparamètres. Les résultats détaillés par quantile confirme le choix d'un nombre d'arbre supérieur ou égal à 5000 et un *nodesize* autour de 250. Dans la suite de l'étude, les calculs seront réalisés avec 10 000 arbres et un *nodesize* de 250.



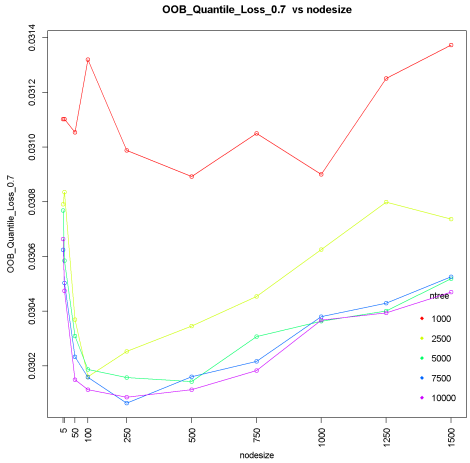
(a) Bonis 50%



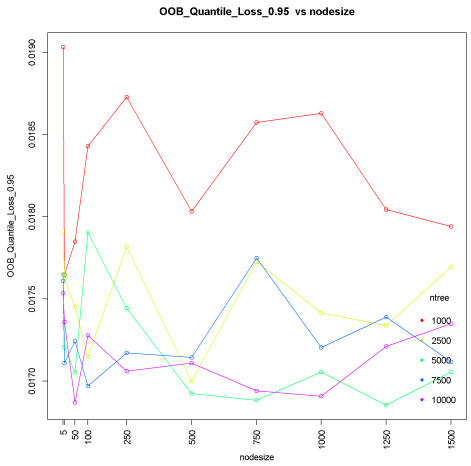
(b) Malis 50%



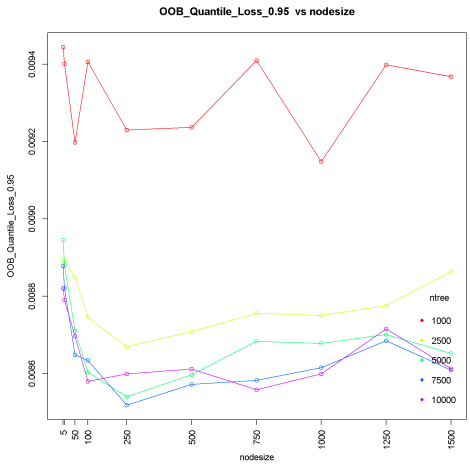
(c) Bonis 70%



(d) Malis 70%



(e) Bonis 95%

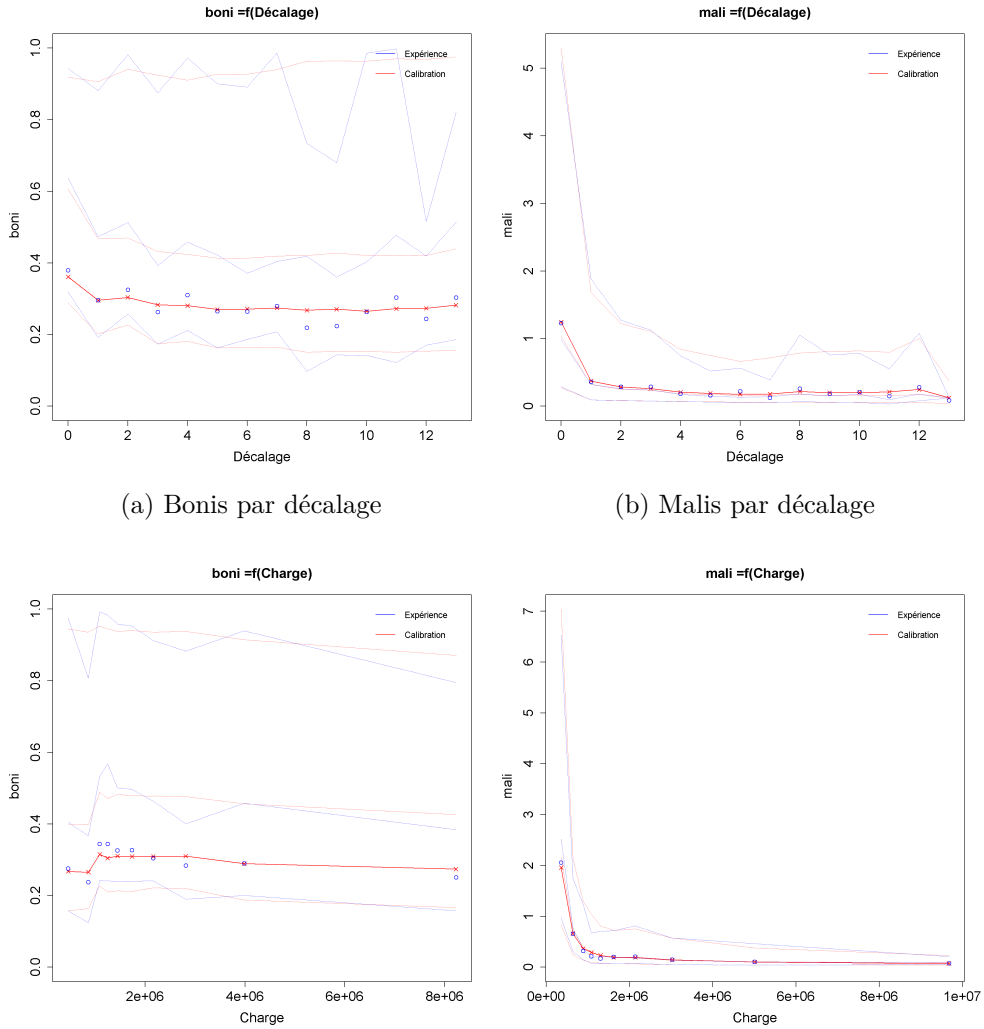


(f) Malis 95%

FIGURE 35 – QRF - Choix des hyperparamètres de la forêt aléatoire.

4.3.2 Validation générale sur les moyennes prédites

La première validation du modèle se fait au niveau global en comparant les moyennes prédites ainsi que quelques quantiles (50%, 75% et 95%) avec leurs équivalents expérimentaux (fig. 36). Une bonne adéquation se dégage de ces figures aussi bien sur les moyennes que sur les quantiles, pour les bonis et les malis.



(a) Bonis par décalage

(b) Malis par décalage

(c) Bonis par charge

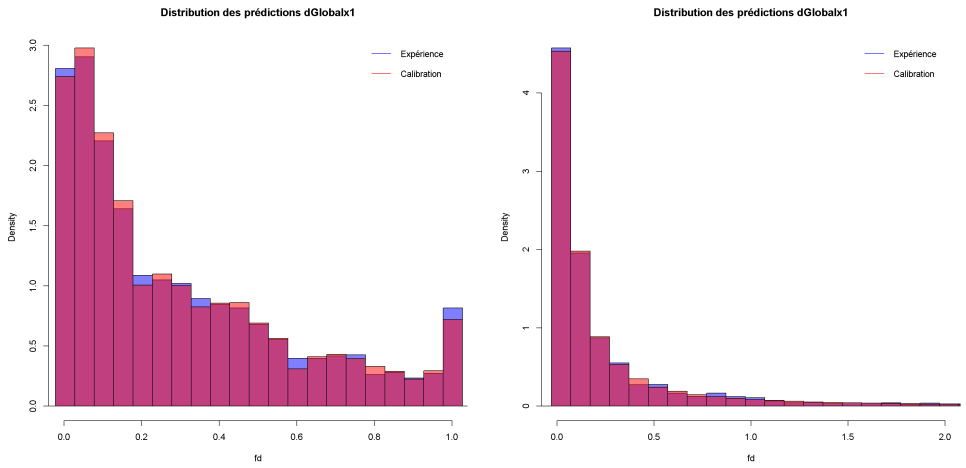
(d) Malis par charge

FIGURE 36 – QRF - Validation globale de la calibration des bonis et malis, moyenne et quantiles (50%, 75% et 95%).

En annexe 4.5, le détail de la comparaison des moyennes prédites en fonction de la charge pour une sélection de décalage est présenté sur les figures 54 et 55.

4.3.3 Validation détaillée des distributions

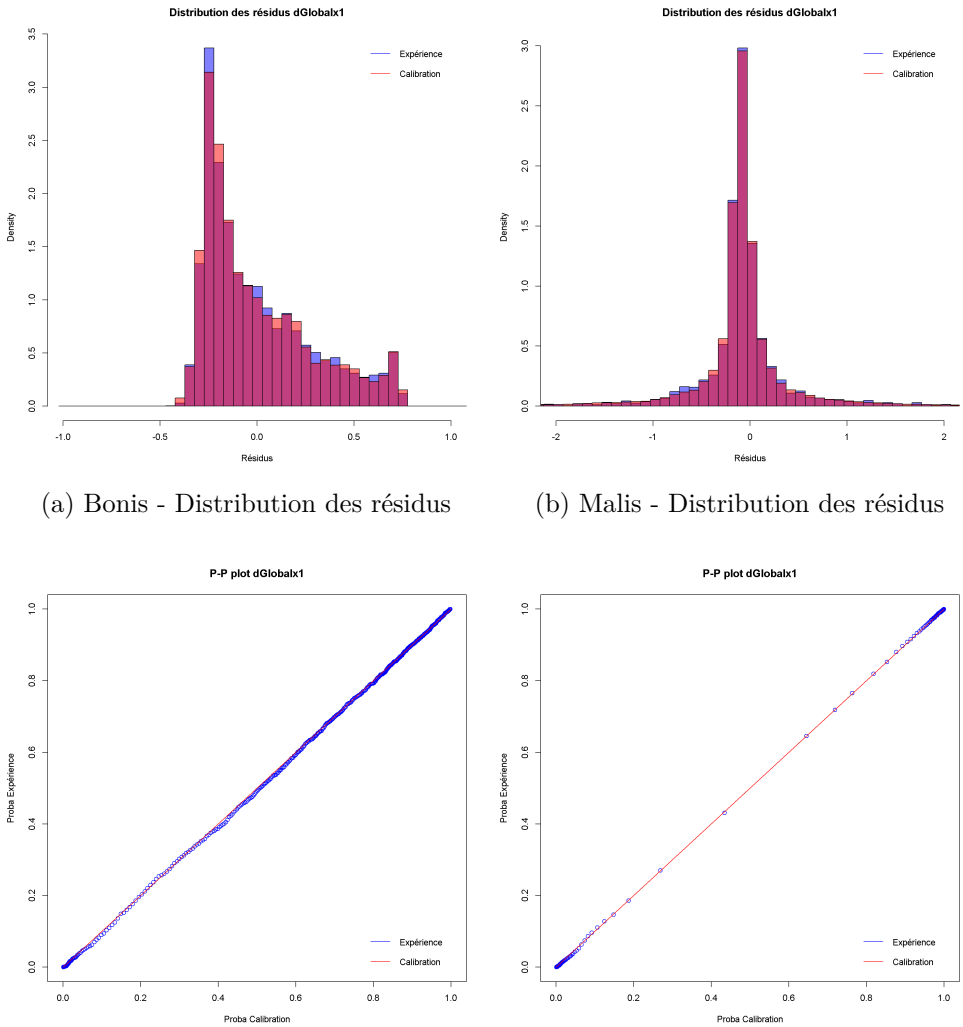
La seconde validation s'effectue sur la comparaison des distributions de la variable prédite et de ses résidus. La figure 37 présente la comparaison de la distribution de la variable prédite expérimentale et calibrée.



(a) boni (b) mali

FIGURE 37 – QRF - Comparaison de la distribution des prédictions avec l'expérience.

La figure 38 présente la distribution des résidus de la variable prédite expérimentale et calibrée ainsi que le diagramme probabilité-probabilité associé. L'adéquation est bonne sur l'ensemble des données.



(a) Bonis - Distribution des résidus

(b) Malis - Distribution des résidus

(c) Bonis - diagramme PP

(d) Malis - diagramme PP

FIGURE 38 – QRF - Comparaison de la distribution des résidus avec l'expérience.

Enfin, en annexe 4.5, les figures 56, 57 et 58 présentent la distribution des résidus expérimentale et calibrée ainsi que le diagramme probabilité-probabilité associé pour des décalages spécifiques (0, 5 et 10). Les adéquations, bien que moins bonne qu'au niveau globale, restent satisfaisantes.

4.3.4 Calcul des provisions brutes et cédées

Les modèles calibrés précédemment sont appliqués individuellement en cascade pour remplir stochastiquement plusieurs centaines de versions du triangle de charge individuel. La figure 39 présente les provisions brutes et cédées par année de survenance tandis que la figure 40 présente la distribution des provisions brutes et cédées totales obtenues grâce aux développements stochastiques des triangles de charges individuelles. Sur ces figures sont aussi représentées les provisions sans projection ainsi que celles projetées avec la méthode Chain Ladder pour comparaison.

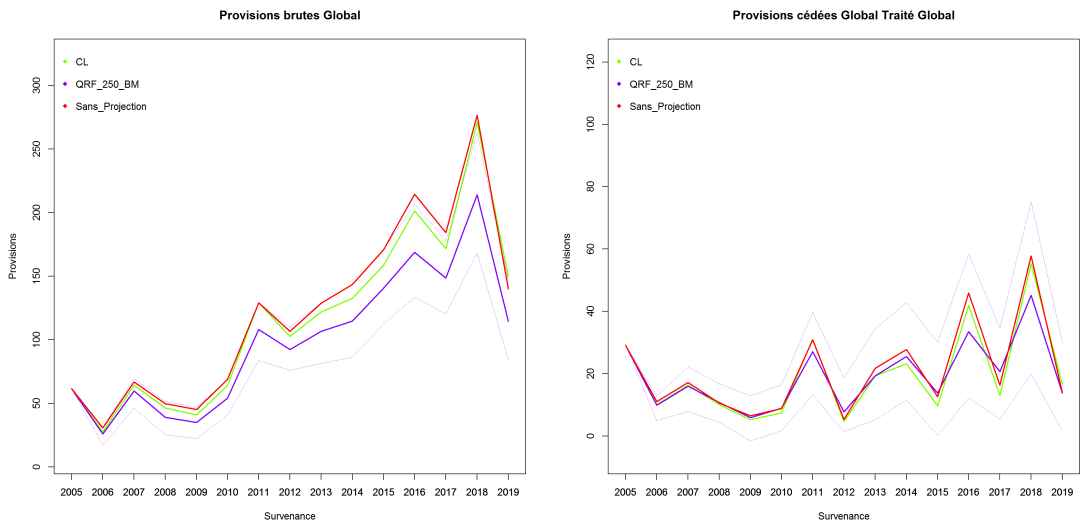


FIGURE 39 – QRF - Provisions estimées par survenance et quantiles 5% et 95%.

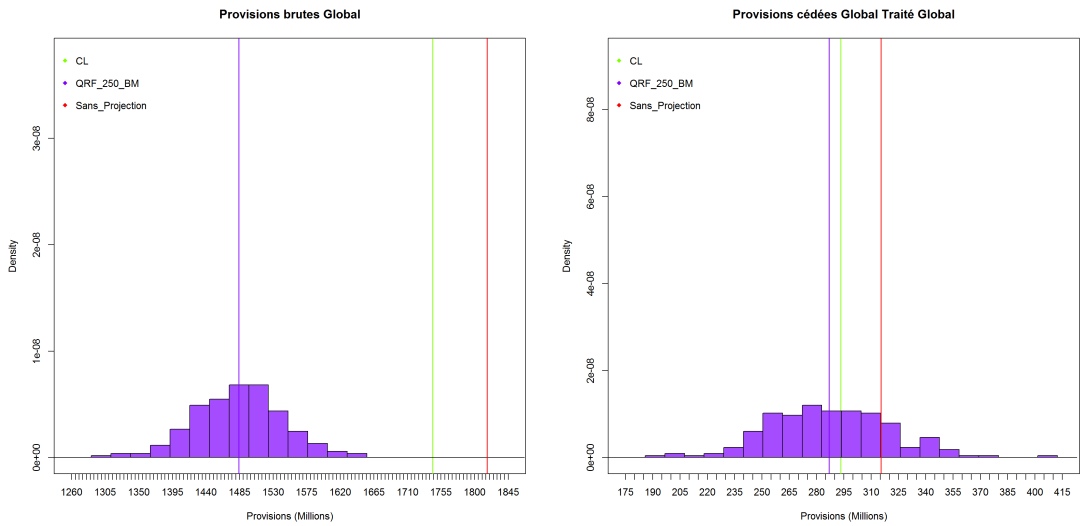


FIGURE 40 – QRF - Provisions totales estimées.

4.4 Comparaison des modèles

4.4.1 Application du *Back Testing*

Comme présenté dans la partie théorique au chapitre 4, la capacité prédictive des modèles peut être comparée en réalisant un *Back Testing*. La profondeur du *Back Testing* est fixée à 1/3 de la base de données soit 5 années en arrière sur un total de 15 années. Les écarts entre les quantiles observés et prédits via la fonction de perte quantile sont estimés. La figure 41 présente les résultats de la fonction de perte quantile en fonction du quantile et de la méthode. Les méthodes individuelles obtiennent des écarts plus faibles que la méthode Chain-Ladder, la différence étant plus importante pour les quantiles les plus extrêmes. De plus, la mesure de ces écarts ne permet pas de distinguer significativement l'une ou l'autre des deux méthodes individuelles étudiées.

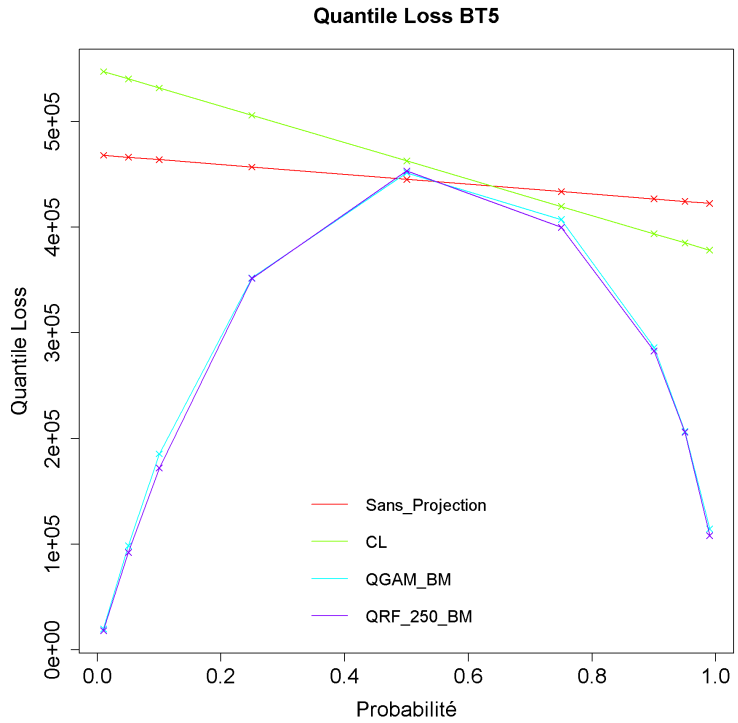


FIGURE 41 – Fonction de pertes quantile en fonction du quantile et la méthode utilisée.

Afin de compléter la validation par quantiles, le calcul des RMSE pour une série consécutive de tranche XS de portée 2 M€ est réalisé pour chaque méthode et présenté sur la figure 42. Là encore, les méthodes individuelles obtiennent des écarts plus faibles que la méthode Chain-Ladder.

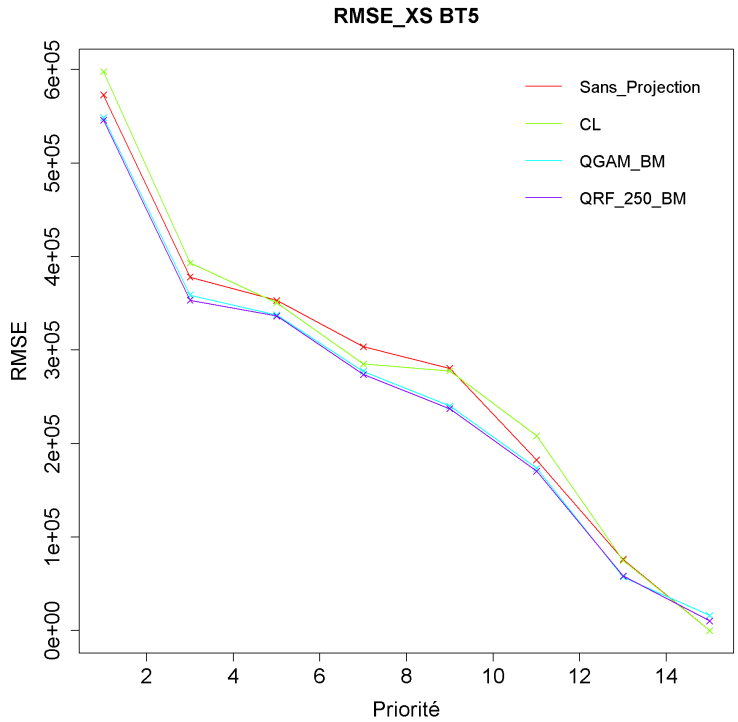
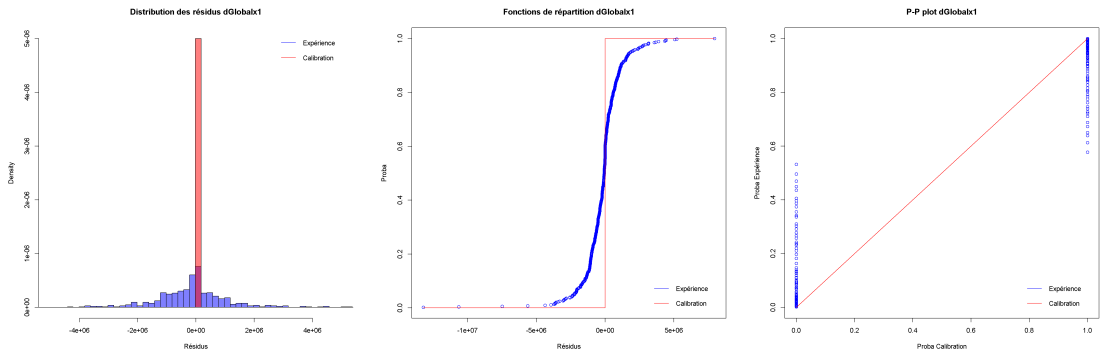


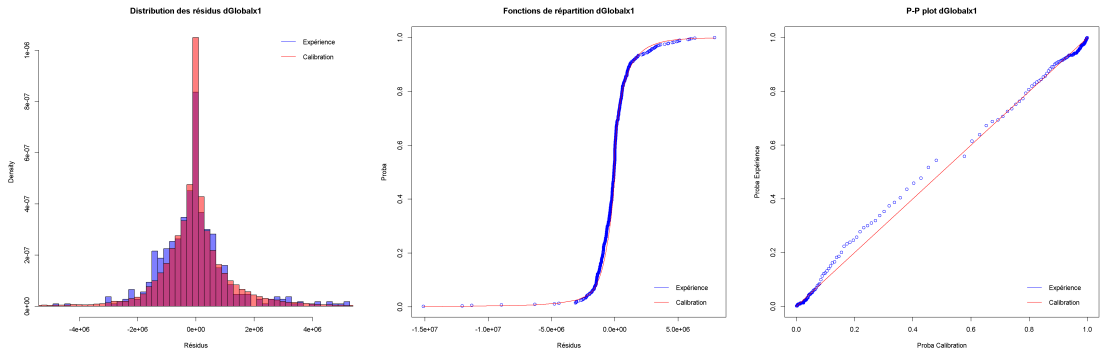
FIGURE 42 – RMSE en fonction de la tranche XS et de la méthode utilisée.

Les distributions des résidus, les fonctions de répartition et les diagrammes probabilité-probabilité des variables observées en fonction des variables prédites sont tracés pour comparer les méthodes entre elles. Les figures 43, 44 et 45 présentent les distributions, fonctions de répartition et diagramme probabilité-probabilité des résidus respectivement pour les méthodes Chain-Ladder, GAM et QRF.



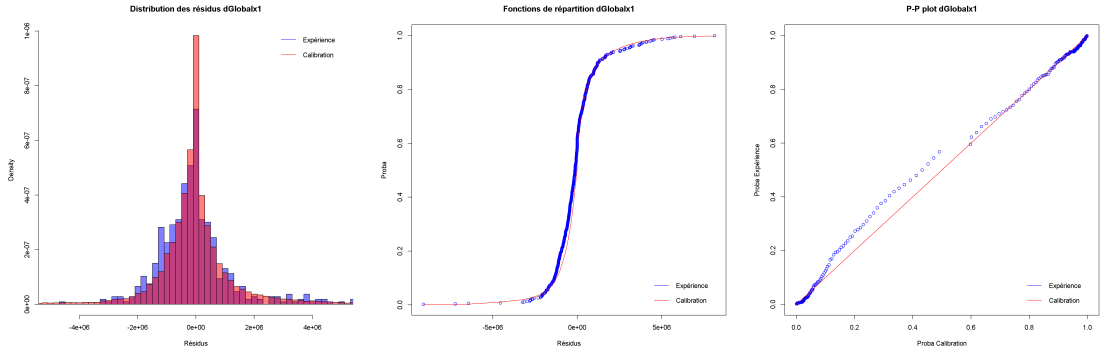
(a) Distribution. (b) Fonction de répartition. (c) Diagramme PP.

FIGURE 43 – Distribution, fonction de répartition et diagramme probabilité-probabilité des résidus Chain-Ladder.



(a) Distribution. (b) Fonction de répartition. (c) Diagramme PP.

FIGURE 44 – Distribution, fonction de répartition et diagramme probabilité-probabilité des résidus GAM.



(a) Distribution. (b) Fonction de répartition. (c) Diagramme PP.

FIGURE 45 – Distribution, fonction de répartition et diagramme probabilité-probabilité des résidus QRF.

4.4.2 Comparaison des provisions brutes et provisions cédées estimées

La figure 46 présente une comparaison des provisions brutes et des provisions cédées globales estimées pour l'ensemble des modèles étudiés.

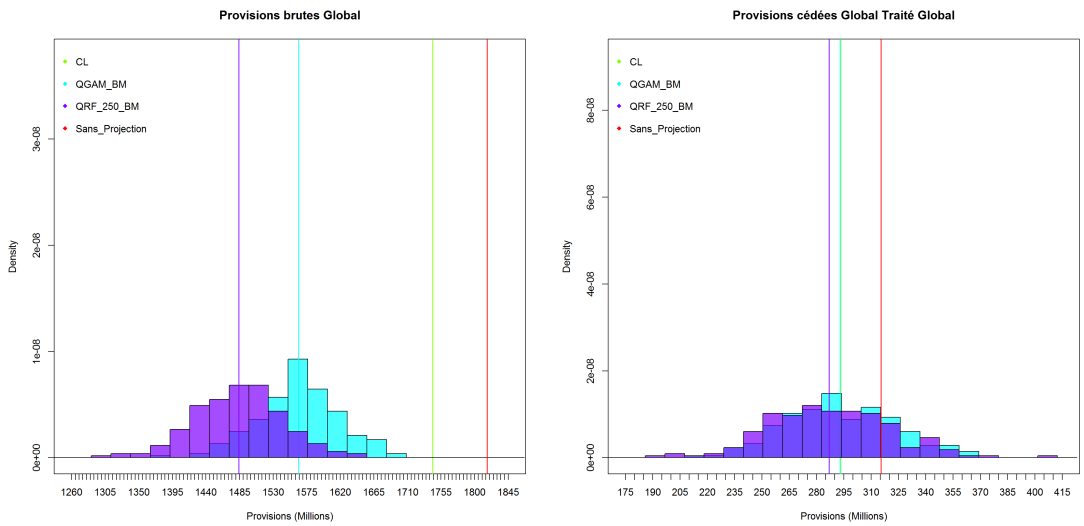
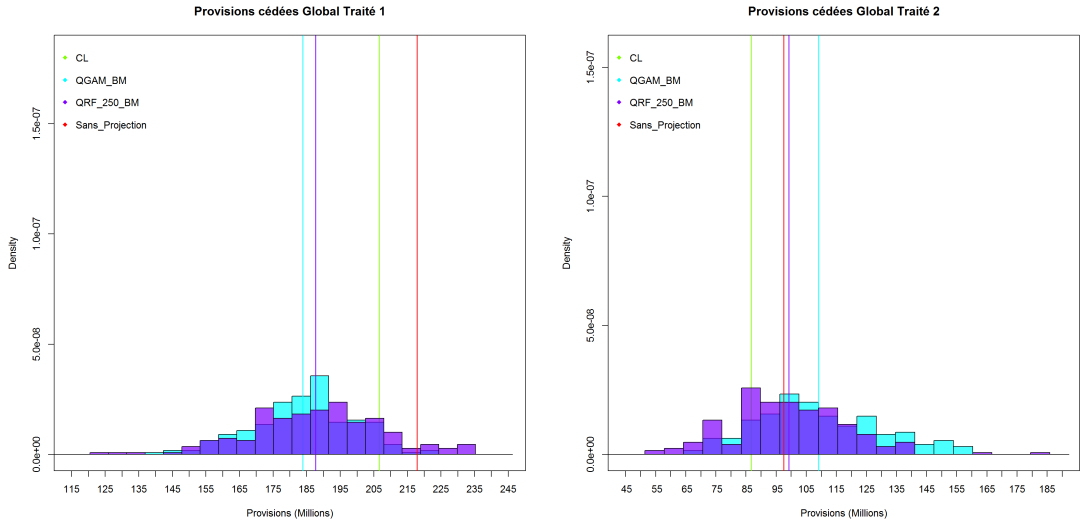


FIGURE 46 – Comparaison des provisions brutes et provisions cédées globales estimées.

La figure 47 présente une comparaison des provisions cédées par traité estimées pour l'ensemble des modèles étudiés. Les écarts entre les méthodes individuelles et la méthode Chain Ladder ne sont pas constants et dépendent des bornes du traité ce qui confirme une fois de plus la nécessité de projeter les charges de manière individuelle et stochastique.



(a) Traité 1 : 4 M XS 6 M.

(b) Traité 2 : ILL XS 10 M.

FIGURE 47 – Comparaison des Provisions cédées pour les deux traités.

4.5 Limites des modèles

Bien que les méthodes développées représentent une amélioration importante pour le calcul des provisions cédées en réassurance non proportionnelle, un certain nombre de limite peut être identifié :

- Ces méthodes sont basées sur les données passées et supposent donc une relative homogénéité des données dans le temps.
- Le volume de données nécessaire pour la calibration d'un modèle individuel qui prédit les quantiles est plus important que pour la calibration d'un modèle agrégé. Ces méthodes seront moins adaptées à un cas disposant de trop peu de données.
- L'année de projection maximale (ultime) est limitée par l'ancienneté des sinistres utilisés pour l'apprentissage.
- L'utilisation d'un nombre restreint de variables explicatives (charge et décalage) peut s'avérer insuffisant pour prédire les évolutions d'un sinistre individuel.
- Enfin, l'utilisation d'une cadence de paiement et d'une inflation future fixe conduit à des coefficients de stabilisation déterministes ce qui peut réduire la volatilité des provisions calculées.

Conclusion et perspectives

Ce mémoire a permis de développer un outil de provisionnement individuel stochastique adapté à l'application de la réassurance non-proportionnelle. Cet outil peut être utilisé notamment pour la réalisation d'étude de *Best Estimate* de cession ainsi que des commutations. L'inadéquation des modèles de provisionnement agrégés appliqué à la réassurance en excédent de sinistre a pu être surmontée en développant une méthode qui décompose la projection d'un sinistre d'une année à l'autre en une étape de prédiction de l'évolution en boni, stable ou mali (régression logistique multinomiale), suivi d'une prédiction de la charge au décalage suivant. Pour la seconde étape, deux méthodes de régression ont été proposées, une paramétrique (modèle additif généralisé linéaire paramétrique) et une non-paramétrique (régression quantile par forêt d'arbres décisionnels). Ces méthodes reproduisent et généralisent les données d'apprentissage de manière satisfaisante. La validation par *Back Testing* confirme l'intérêt des méthodes développées pour prédire la distribution individuelle des charges projetées. Enfin l'utilisation de ces méthodes pour projeter à l'ultime un grand nombre de triangle de charge individuelle aboutit à des visions précises des distributions des provisions brutes et cédées ce qui remplit les objectifs de l'étude.
Au delà des objectifs initiaux, les perspectives d'évolutions sont les suivantes :

- L'ajout de variables explicatives (en plus du décalage et de la charge) pourrait être exploré afin d'améliorer les prédictions (âge victime, présence d'une rente, état du dossier, etc.).
- La réalisation de la calibration des modèles sur un périmètre étendu (groupe de cédante ou marché complet) pourrait améliorer le pouvoir de généralisation des modèles. Dans ce cas, l'ajout du nom de la cédante comme variable explicative pourrait permettre de distinguer la partie générale et la partie spécifique dans la prédiction des évolutions d'un sinistre.
- Un modèle de tardifs pourrait être ajouté pour compléter les provisions des années de survenance récentes.
- La prise en compte d'une cadence de paiement stochastique pourrait avoir un effet sur les coefficients de stabilisation et l'actualisation des flux futurs.
- Enfin, la méthode pourrait être appliquée à d'autres domaines que les *Best Estimate* de cession et les commutations : tarification, autres branches longues comme la construction.

Bibliographie

[Taillieu, 2016] F. Taillieu et al., Manuel « Best Estimate Liabilities Non-vie », 2016, Institut des Actuaires

[Beneteau, 2004] G. Beneteau, « Modèle de provisionnement sur données détaillées en assurance non-vie », 2004, Mémoire d'actuariat

[Barbaste, 2017] M. Barbaste, « Une méthode de provisionnement individuel par apprentissage automatique », 2017, Mémoire d'actuariat

[Quiquet, 2019] J. Quiquet, « Méthode d'estimation de la charge ultime en RC corporelle automobile basée sur des données individuelles », 2019, mémoire d'actuariat

[Antonio et Plat, 2012] K. Antonio et R. Plat, « Micro-level stochastic loss reserving for general insurance », 2012, Scandinavian Actuarial Journal

[ASTIN, 2017] B. Harej et al., « Individual Claim Development with Machine Learning », 2017, ASTIN Report

[Baudry et Y.Robert, 2017] M. Baudry et C. Y.Robert, « Non-parametric individual claim reserving in insurance », 2017, Article Chaire DAMI

[Taylor et McGuire, 2016] G. Taylor et G. McGuire, « Stochastic loss reserving using generalized linear models », 2016, Casualty Actuarial Society

[Hua et al., 2010] C. Hua et al., « Companion to BER 642 : Advanced Regression Methods », 2020, R-Book University of Alabama

[Rachka., 2020] S. Rachka, « What is the relation between Logistic Regression and Neural Networks and when to use which ? », 2020, Machine Learning FAQ, `https://sebastianraschka.com/faq/docs/logisticregr-neuralnet.html`

[Rigby et Stasinopoulos, 2010] B. Rigby et M. Stasinopoulos, « A flexible regression approach using GAMLSS in R », 2010, University of Athens

[Meinshausen, 2006] N. Meinshausen, « Quantile Regression Forests », 2006, Journal of Machine Learning Research

[Crucianu, 2020] M. Crucianu et al., « Apprentissage, réseaux de neurones et modèles graphiques », 2020, Cours CNAM RCP209

[Grover, 2018] P. Grover, « 5 Regression Loss Functions All Machine Learners Should Know », `https://heartbeat.fritz.ai/5-regression-loss-functions-all-machine-learners-should-know-4fb140e9d4b0`

Annexes

Annexe 1

Modèle d'évolution : la figure 48 présente les probabilités d'évolution en boni, stable ou mali en fonction de la charge pour des décalages spécifiques. L'adéquation entre prédiction et expérience reste bonne au niveau détaillé.

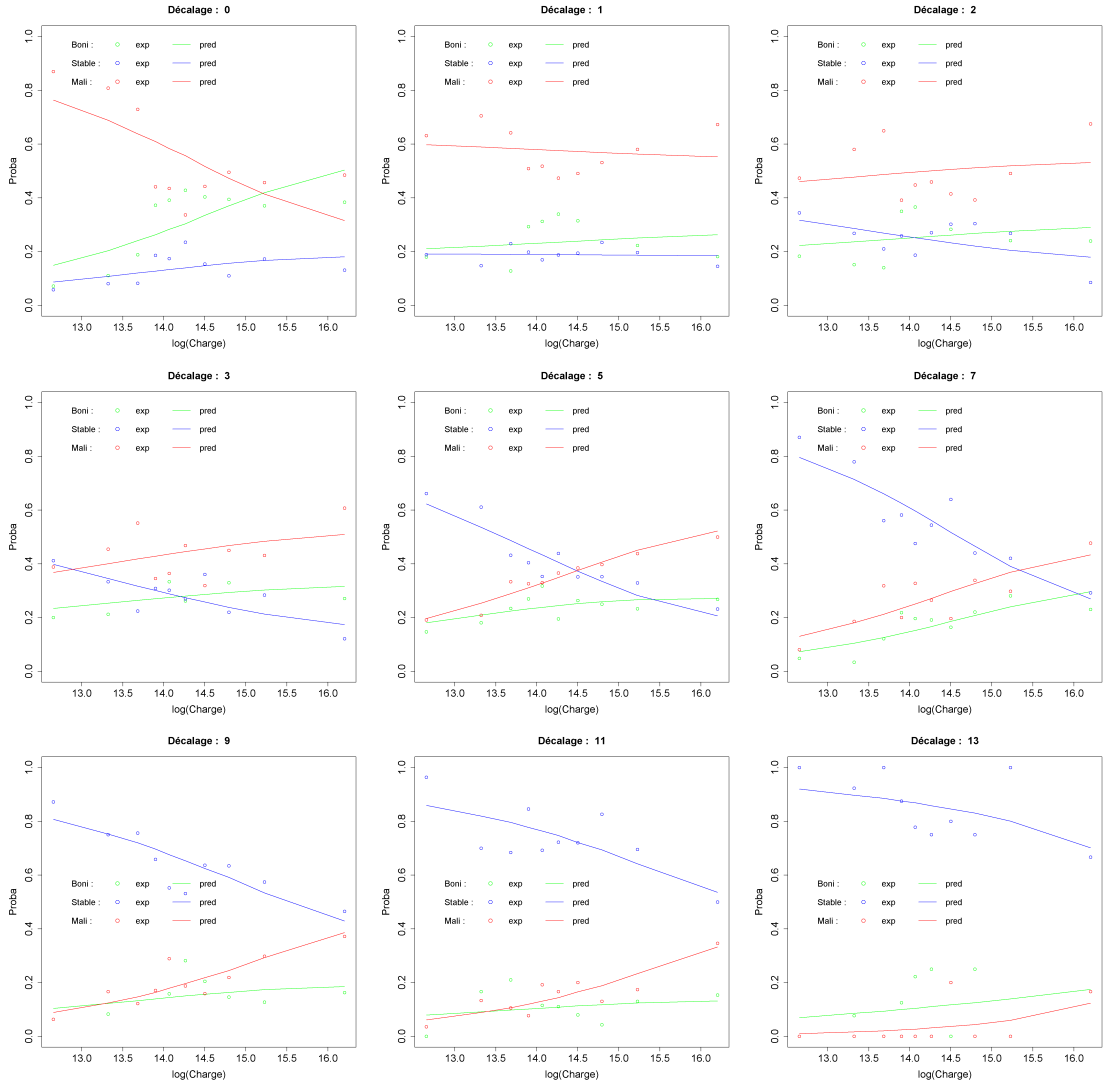


FIGURE 48 – Probabilité d'un boni, stable ou mali en fonction du décalage et de la charge pour une sélection de décalage.

Annexe 2

Méthode GAM : le détail de la comparaison des moyennes prédites en fonction de la charge pour une sélection de décalage est présenté sur les figures 49 et 50.

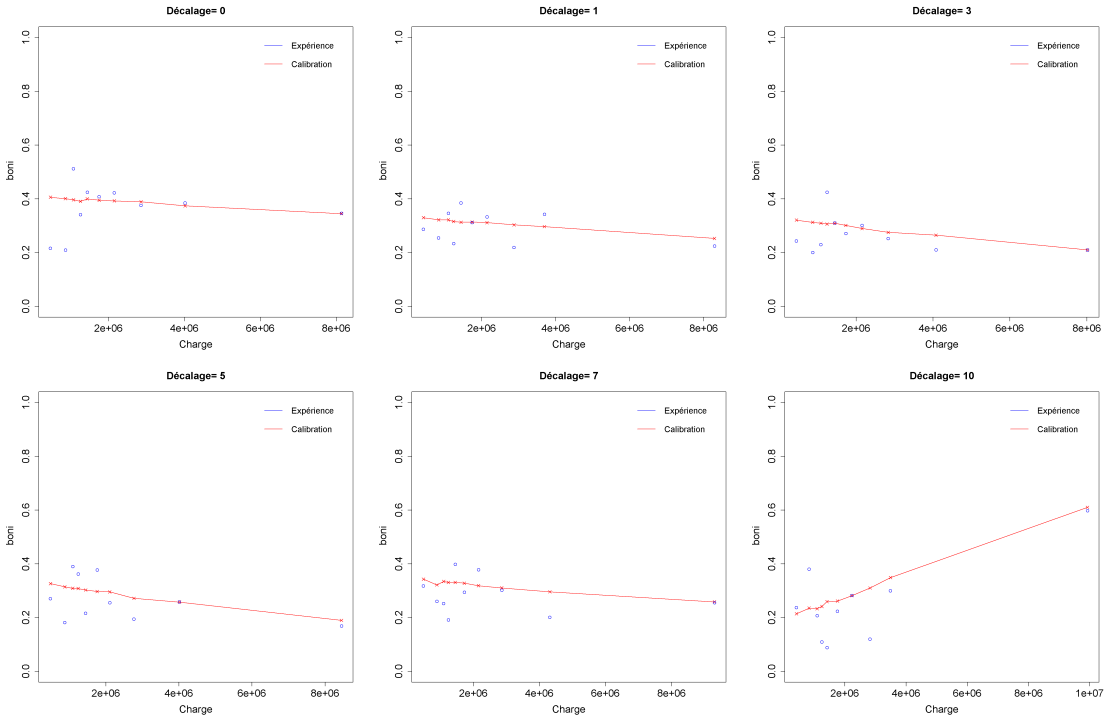


FIGURE 49 – GAM - Validation par décalage de la calibration pour les bonis.

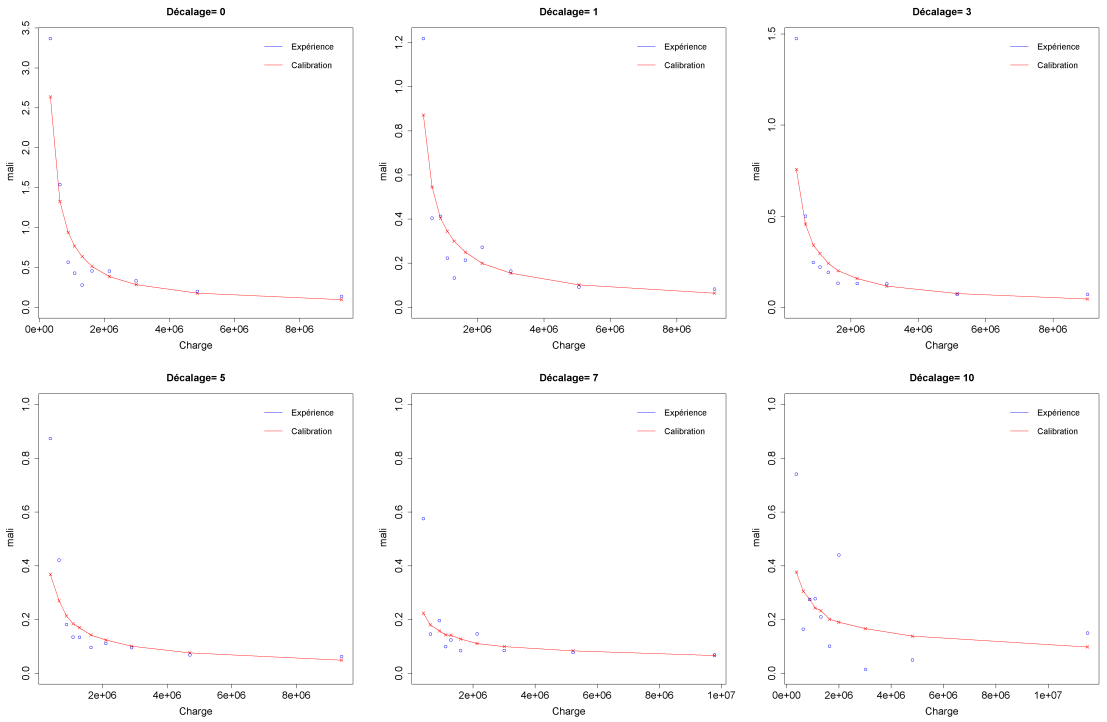
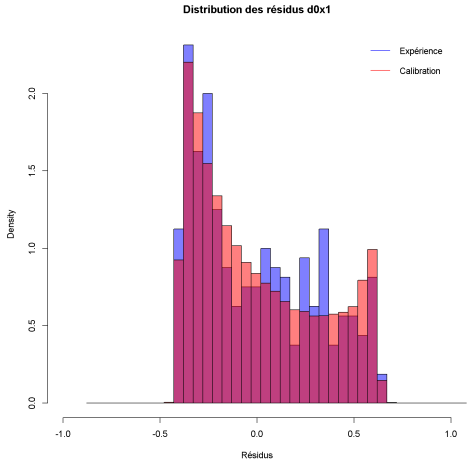


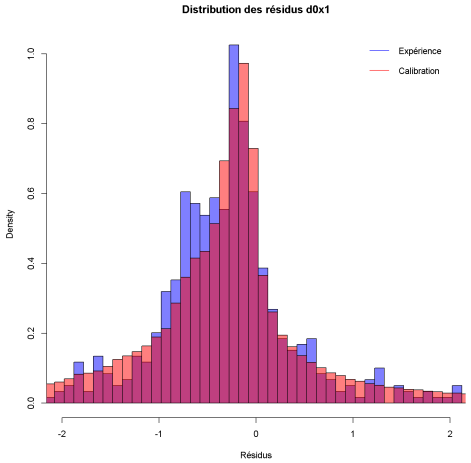
FIGURE 50 – GAM - Validation par décalage de la calibration pour les malis.

Annexe 3

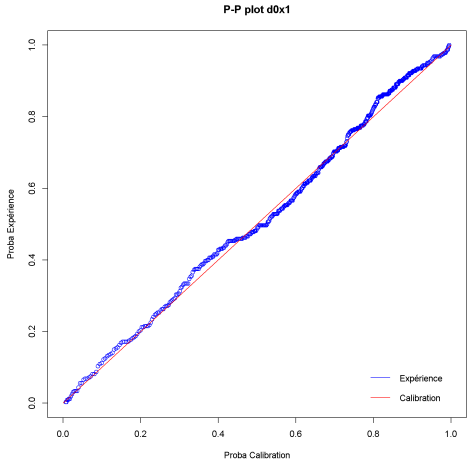
Méthode GAM : les figures 51, 52 et 53 présentent la distribution des résidus expérimentale et calibrée ainsi que le diagramme probabilité-probabilité associé pour des décalages spécifiques (0, 5 et 10). Les adéquations, bien que moins bonnes qu'au niveau globale, restent satisfaisantes.



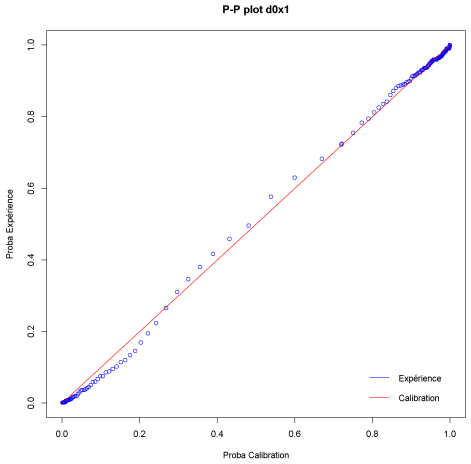
(a) Bonis, d=0 - Distribution des résidus



(b) Malis, d=0 - Distribution des résidus

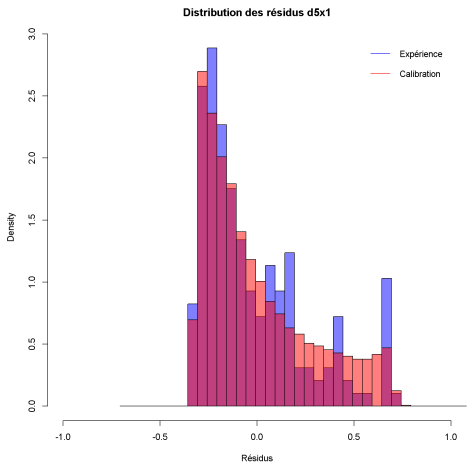


(c) Bonis, d=0 - diagramme PP

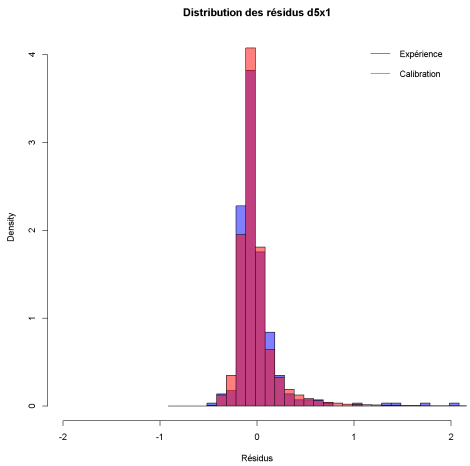


(d) Malis, d=0 - diagramme PP

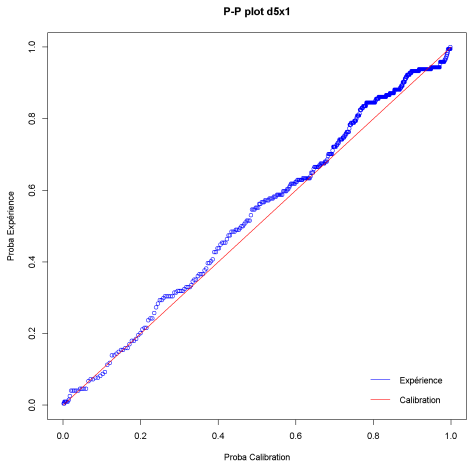
FIGURE 51 – GAM - Comparaison de la distribution des résidus avec l'expérience pour décalage=0



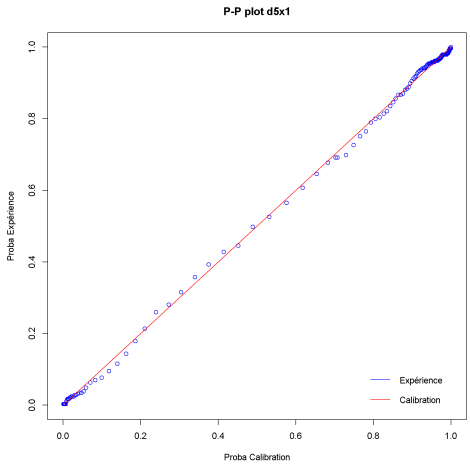
(a) Bonis, d=5 - Distribution des résidus



(b) Malis, d=5 - Distribution des résidus

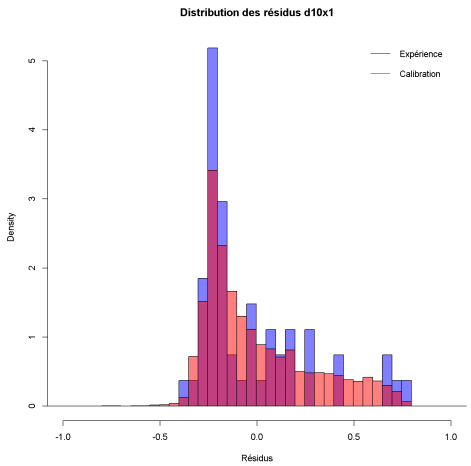


(c) Bonis, d=5 - diagramme PP

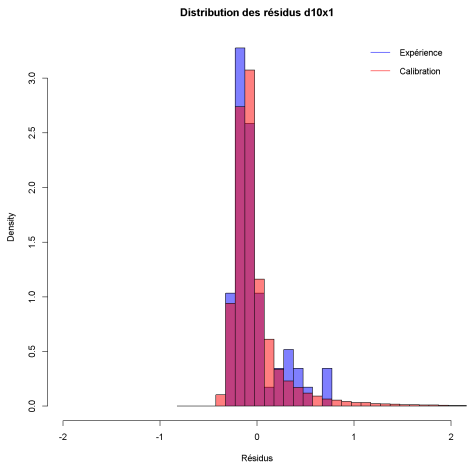


(d) Malis, d=5 - diagramme PP

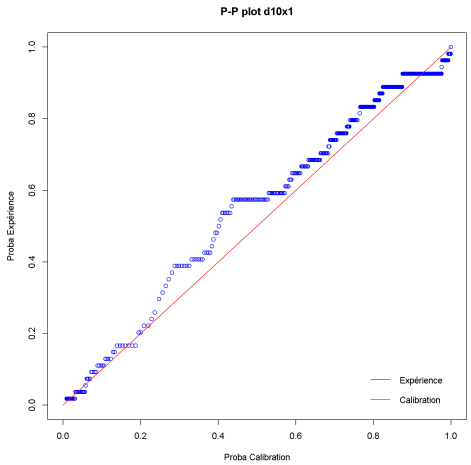
FIGURE 52 – GAM - Comparaison de la distribution des résidus avec l'expérience pour décalage=5



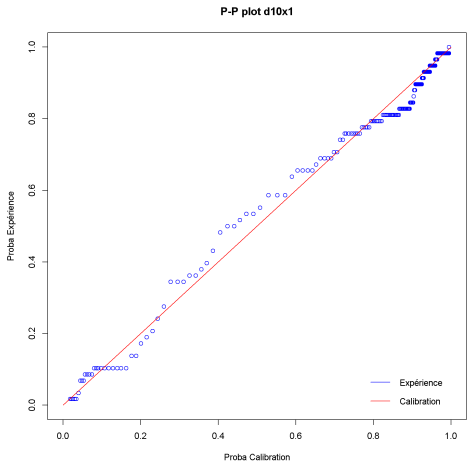
(a) Bonis, d=10 - Distribution des résidus



(b) Malis, d=10 - Distribution des résidus



(c) Bonis, d=10 - diagramme PP



(d) Malis, d=10 - diagramme PP

FIGURE 53 – GAM - Comparaison de la distribution des résidus avec l'expérience pour décalage=10

Annexe 4

Méthode QRF : le détail de la comparaison des moyennes prédites en fonction de la charge pour une sélection de décalage est présenté sur les figures 54 et 55.

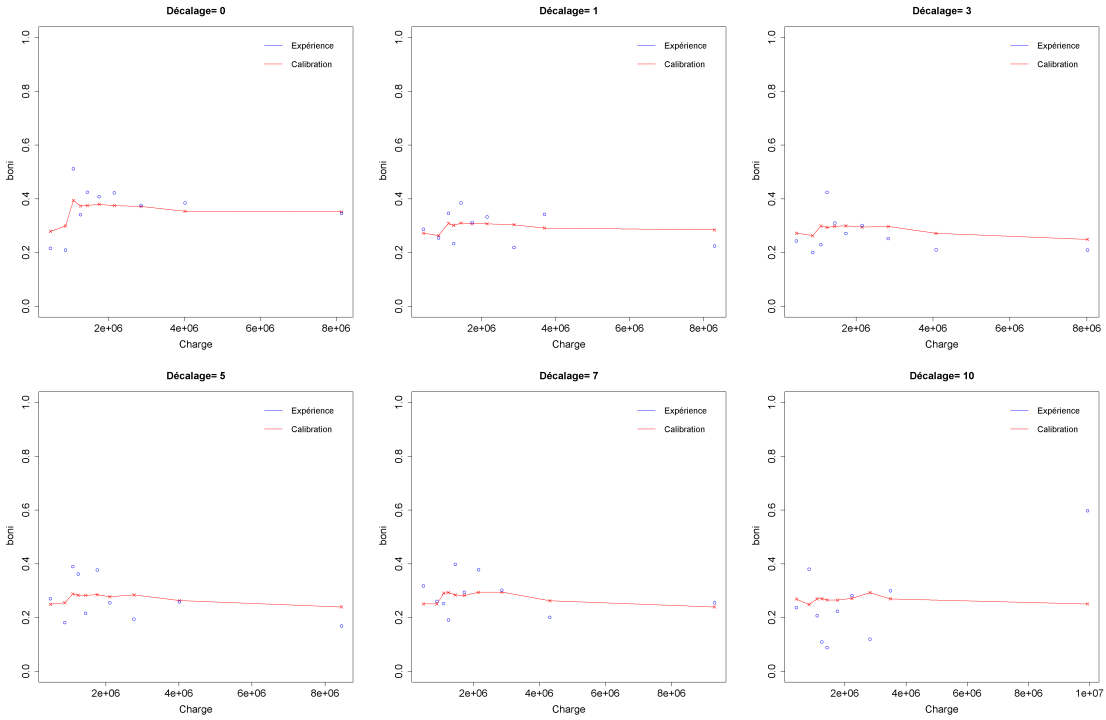


FIGURE 54 – QRF - Validation par décalage de la calibration pour les bonis.

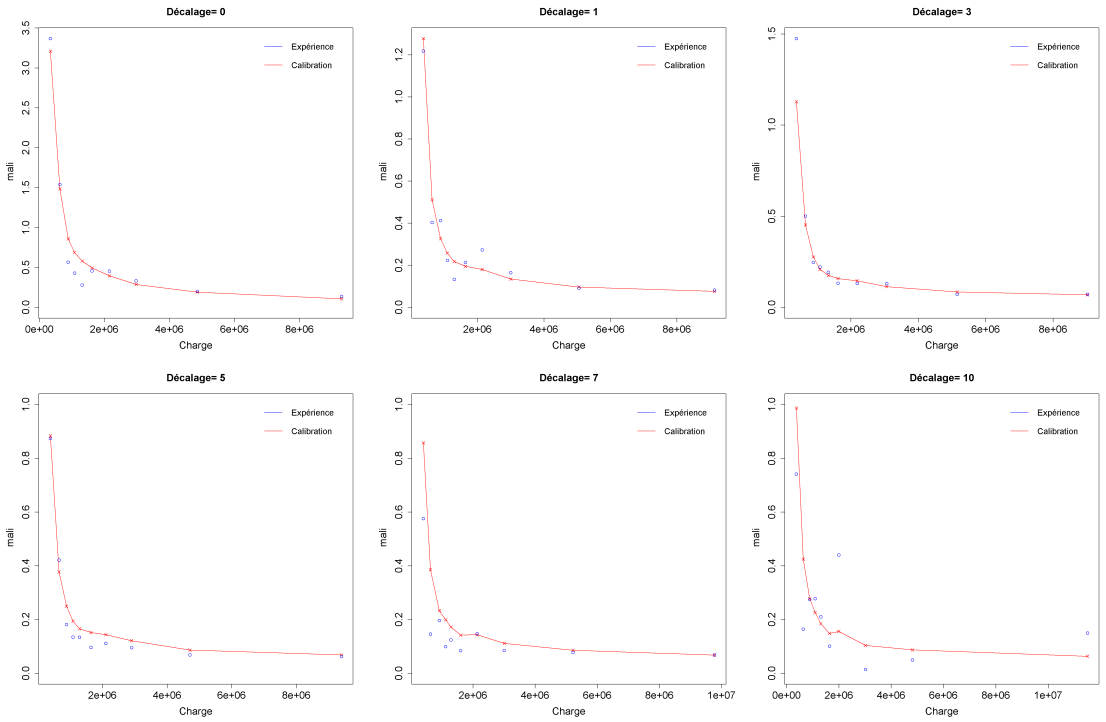
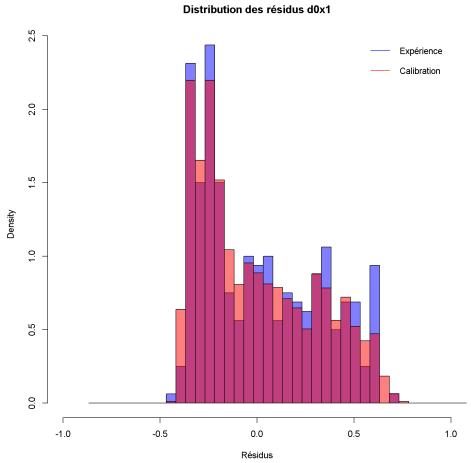


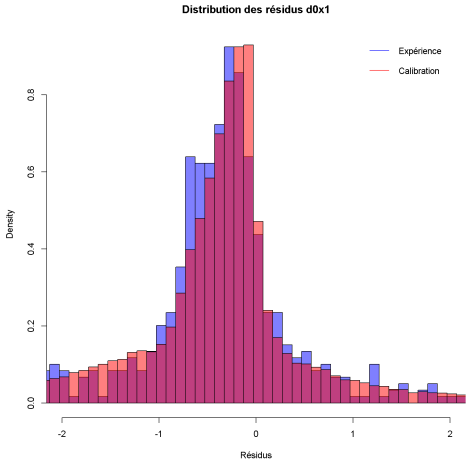
FIGURE 55 – QRF - Validation par décalage de la calibration pour les malis.

Annexe 5

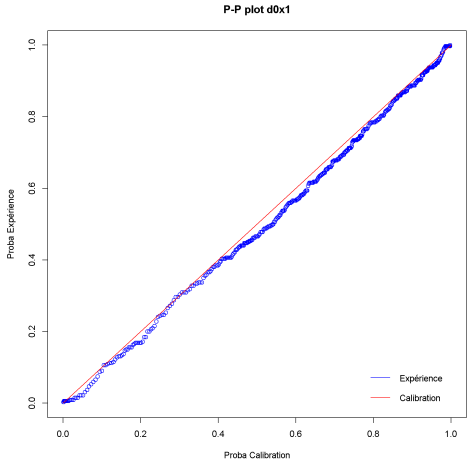
Méthode QRF : les figures 56, 57 et 58 présentent la distribution des résidus expérimentale et calibrée ainsi que le diagramme probabilité-probabilité associé pour des décalages spécifiques (0, 5 et 10). Les adéquations, bien que moins bonnes qu'au niveau globale, restent satisfaisantes.



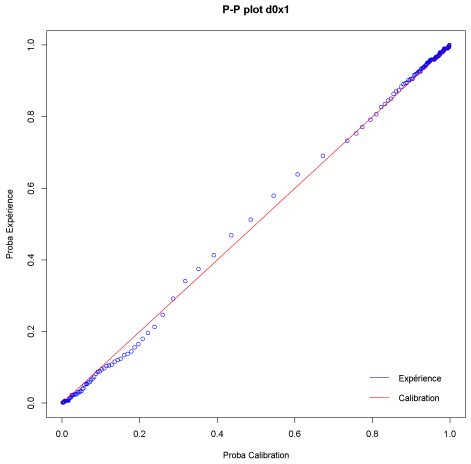
(a) Bonis, d=0 - Distribution des résidus



(b) Malis, d=0 - Distribution des résidus

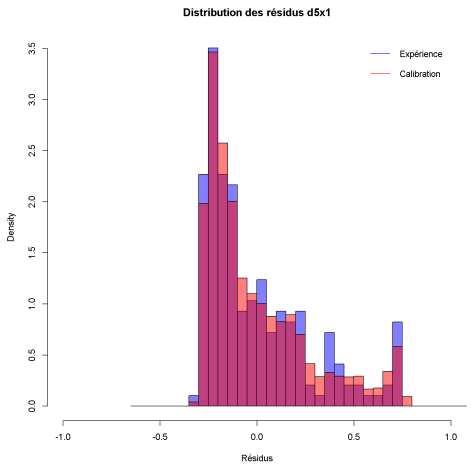


(c) Bonis, d=0 - diagramme PP

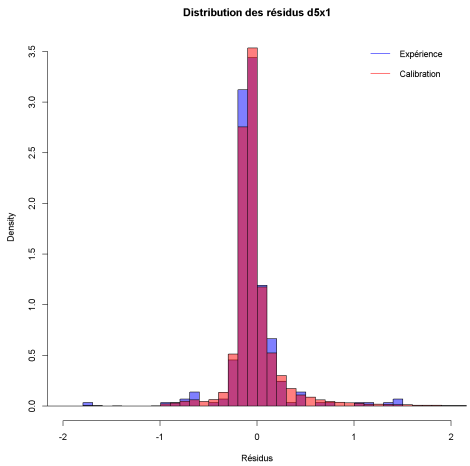


(d) Malis, d=0 - diagramme PP

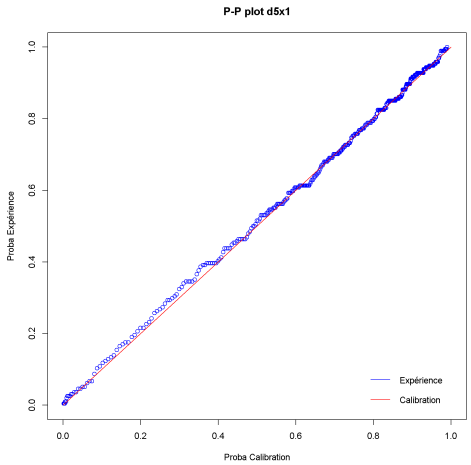
FIGURE 56 – QRF - Comparaison de la distribution des résidus avec l'expérience pour décalage=0



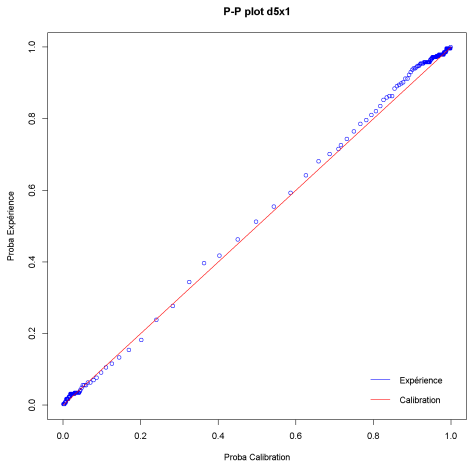
(a) Bonis, d=5 - Distribution des résidus



(b) Malis, d=5 - Distribution des résidus

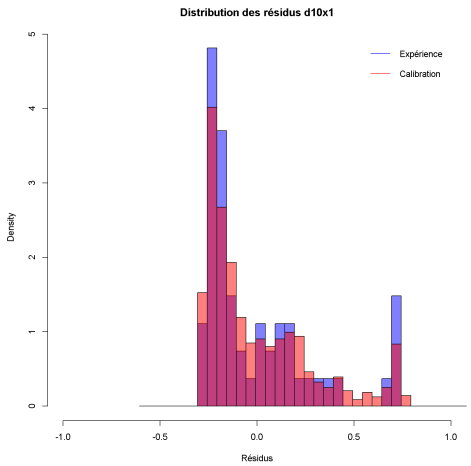


(c) Bonis, d=5 - diagramme PP

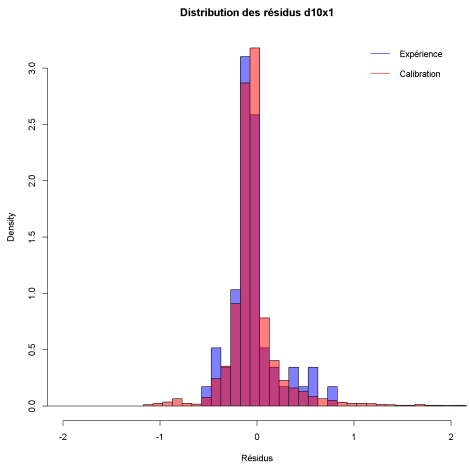


(d) Malis, d=5 - diagramme PP

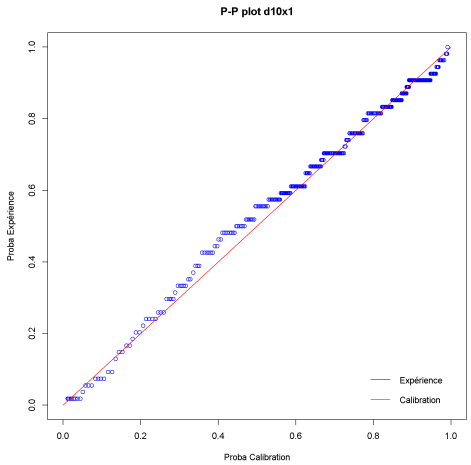
FIGURE 57 – QRF - Comparaison de la distribution des résidus avec l'expérience pour décalage=5



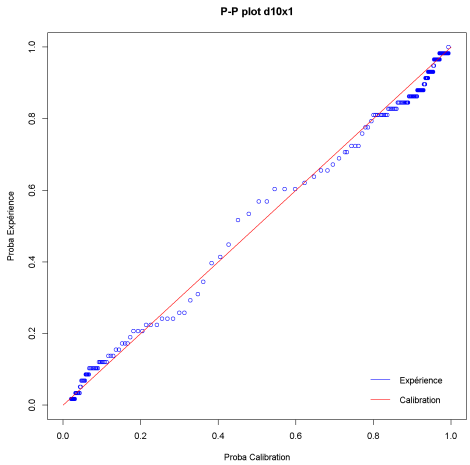
(a) Bonis, d=10 - Distribution des résidus



(b) Malis, d=10 - Distribution des résidus



(c) Bonis, d=10 - diagramme PP



(d) Malis, d=10 - diagramme PP

FIGURE 58 – QRF - Comparaison de la distribution des résidus avec l'expérience pour décalage=10

Table des figures

1 Principe de la réassurance en excédent de sinistre. 12

2 Illustration de l'intérêt d'un provisionnement individuel stochastique en réassurance non proportionnelle : sinistre initialement en dessous de la priorité. 15

3 Illustration de l'intérêt d'un provisionnement individuel stochastique en réassurance non proportionnelle : sinistre initialement au dessus de la priorité. 16

4 Remplissage en cascade d'un triangle individuel 20

5 Illustration d'une régression logistique binomiale représentée par un réseau de neurones à une couche [Rachka., 2020]. 27

6 Illustration d'un réseau de neurones avec une couche cachée [Rachka., 2020]. . . . 28

7 Exemple d'une régression linéaire simple [Rigby et Stasinopoulos, 2010] 30

8 Exemple d'une régression linéaire généralisée [Rigby et Stasinopoulos, 2010] . . . 31

9 Exemple d'une régression additive généralisée [Rigby et Stasinopoulos, 2010] . . . 32

10 Exemple de bonis avant et après normalisation. 33

11 Lois continues de support (0,1) [Rigby et Stasinopoulos, 2010]. 33

12 Exemple de malis avant et après normalisation. 34

13 Lois continues à support $(0, \infty)$ [Rigby et Stasinopoulos, 2010]. 34

14 Exemple d'arbre de décision : survie des passagers du Titanic [Crucianu, 2020] . 36

15 Illustration d'une forêt d'arbres décisionnels. 37

16 Illustration de la fonction de perte quantile [Grover, 2018]. 39

17 Illustration du *Back Testing*. 40

18 Evolution de l'indice utilisé pour l'indexation et la stabilisation 44

19 Effet de l'indexation sur le seuil de communication 45

20 *Boxplots* du facteur de développement individuel. 46

21 *Boxplots* du facteur de développement individuel pour les bonis et malis. 47

22 Violinplots du facteur de développement individuel pour les bonis et malis. . . . 48

23 Coefficients Chain Ladder et évolution de la charge en fonction du décalage. . . . 49

24 Provisions brutes et cédées estimées en fonction de l'année de survenance avec la méthode Chain Ladder. 50

25 Visualisation des bonis, stables et malis. 51

26 Probabilité d'un boni, stable ou mali. 52

27 Choix de la distribution des bonis. 53

28 Choix de la distribution des malis. 54

29 GAM - Validation globale de la calibration des bonis et malis, moyenne et quantiles (50%, 75% et 95%). 55
30 GAM - Comparaison de la distribution des prédictions avec l'expérience. 56
31 GAM - Comparaison de la distribution des résidus avec l'expérience. 57
32 GAM - Provisions estimées par survenance et quantiles 5% et 95%. 58
33 GAM - Provisions totales estimées. 58
34 QRF - Choix des hyperparamètres de la forêt aléatoire. 59
35 QRF - Choix des hyperparamètres de la forêt aléatoire. 60
36 QRF - Validation globale de la calibration des bonis et malis, moyenne et quantiles (50%, 75% et 95%). 61
37 QRF - Comparaison de la distribution des prédictions avec l'expérience. 62
38 QRF - Comparaison de la distribution des résidus avec l'expérience. 63
39 QRF - Provisions estimées par survenance et quantiles 5% et 95%. 64
40 QRF - Provisions totales estimées. 64
41 Fonction de pertes quantile en fonction du quantile et la méthode utilisée. 65
42 RMSE en fonction de la tranche XS et de la méthode utilisée. 66
43 Distribution, fonction de répartition et diagramme probabilité-probabilité des résidus Chain-Ladder. 66
44 Distribution, fonction de répartition et diagramme probabilité-probabilité des résidus GAM. 67
45 Distribution, fonction de répartition et diagramme probabilité-probabilité des résidus QRF. 67
46 Comparaison des provisions brutes et provisions cédées globales estimées. 68
47 Comparaison des Provisions cédées pour les deux traités. 68
48 Probabilité d'un boni, stable ou mali en fonction du décalage et de la charge pour une sélection de décalage. 72
49 GAM - Validation par décalage de la calibration pour les bonis. 73
50 GAM - Validation par décalage de la calibration pour les malis. 73
51 GAM - Comparaison de la distribution des résidus avec l'expérience pour décalage=0 74
52 GAM - Comparaison de la distribution des résidus avec l'expérience pour décalage=5 75
53 GAM - Comparaison de la distribution des résidus avec l'expérience pour décalage=10 76
54 QRF - Validation par décalage de la calibration pour les bonis. 77
55 QRF - Validation par décalage de la calibration pour les malis. 77
56 QRF - Comparaison de la distribution des résidus avec l'expérience pour décalage=0 78
57 QRF - Comparaison de la distribution des résidus avec l'expérience pour décalage=5 79
58 QRF - Comparaison de la distribution des résidus avec l'expérience pour décalage=10 80