

**Mémoire présenté le :**

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

Par : Fanta DIENG

Titre **Modèle Interne Partiel en RC médicale : « Revue de la  
segmentation dans le calcul du SCR primes »**

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

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

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

signature *Entreprise :*

Nom : SHAM

Signature :

*Directeur de mémoire en entreprise :*

Nom : Leila Elijalat

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





---

Modèle Interne Partiel en RC médicale : « Revue de la  
segmentation dans le calcul du SCR primes »

---

**TUTEURS** ELIJALAT Leila

LOISEL EYRAUD Anne

**AUTEUR** DIENG Fanta

Mémoire Actuaire

Année universitaire 2020 - 2021

## Résumé

**Mots clés :** Solvabilité II, SCR (Solvency Capital Requirement), Risque de primes, Value -at-risk, Modèle interne partiel, Segmentation, Responsabilité Civile Médicale (RCM), Modélisation de fréquences, Régression de poisson, Régression linéaire, Imputations, Simulation.

L'assurance non-vie représente l'ensemble des opérations d'assurances qui n'ont pas pour objet la vie de l'assuré. Elle regroupe les assurances de biens, assurances de responsabilité ou de dettes, et des assurances de personnes. Ce qui la différencie de l'assurance vie est la survenance du sinistre, qui est le plus souvent certain en assurance vie, tandis qu'il est juste probable en assurance non-vie. L'une des particularités de l'assurance non-vie est que le coût du sinistre est rarement connu.

En conséquence, il est nécessaire pour l'assureur exerçant ce type d'activités de produire un modèle de sinistralité de son portefeuille lui permettant de simuler à la fois la fréquence de sinistres et le coût de ceux-ci. Les compagnies d'assurance doivent être solvables afin de pouvoir répondre à leurs engagements, à cause des risques qu'elles encourent. Afin de les permettre de fonctionner prudemment, des normes ont été mises en place, notamment la réglementation Solvabilité 2 entrée en vigueur en 2016, succédant au référentiel Solvabilité 1 qui comportait plein de limites. La norme Solvabilité 2 a pour objectif de mieux adapter les fonds propres exigés des compagnies d'assurance et de réassurance aux risques que celles-ci encourent dans leurs activités afin de renforcer la solidité financière de ces entreprises. L'évaluation de ce montant est une phase importante, car cela dépend du profil de risque de l'entreprise ce qui demande alors une bonne maîtrise des différents risques qu'encourt l'entreprise.

À cause de la spécificité de son risque "Responsabilité Civile Médicale", SHAM a dû mettre en place un modèle interne partiel pour pallier les limites des méthodes de la réglementation Solvabilité II dans le calcul des fonds propres, afin d'évaluer son SCR de primes et réserve. Des études sur les hypothèses du modèle ont été nécessaires pour l'amélioration et la pertinence de l'estimation du SCR primes. En effet, ce modèle comporte certaines limites qui ont été identifiées par SHAM et l'audit externe. Certaines hypothèses sont à revoir, c'est le contexte de cette étude.

L'objectif de ce mémoire est d'étudier l'impact d'une nouvelle segmentation des risques afin de calculer le risque de primes dans le cadre du modèle interne partiel de SHAM. Il s'agira de modifier le modèle actuel, par le challenge de certaines hypothèses, et de mesurer les impacts. En d'autres termes, il s'agit de contribuer à l'amélioration du Modèle Interne Partiel, par l'étude de la pertinence d'une nouvelle segmentation d'une classe de risques dans l'évaluation du risque de primes. Ainsi, l'étude suivra le plan suivant :

- Premièrement, nous allons présenter le modèle interne partiel mis en place par SHAM et validé par l'ACPR en janvier 2016. Nous présenterons la méthodologie utilisée pour le calcul du SCR primes, particulièrement la modélisation de la fréquence des sinistres.

- Ensuite, nous présenterons les données après retraitements ainsi que les différentes méthodes de segmentation nécessaires à notre étude.

- Enfin, nous effectuerons différents tests d'impact d'hypothèses alternatives et nous conclurons par une comparaison de la qualité des résultats obtenus avec l'étude de segmentation. Nous vérifierons également la stabilité des différentes méthodes utilisées.

En conclusion, nous déduisons la nécessité de prévoir une modification du calcul du risque de primes dans le cadre de la politique de changement de modèle définie lors de l'approbation du Modèle Interne Partiel par l'ACPR.

## Abstract

**KeyWords :** Premium risk, Solvency 2, SCR (Solvency Capital Requirement), Segmentation, Exposure variable, Partial Internal Model, Simulation, Value-at-Risk, Frequency estimation, Simple linear regression, Fitting distribution, Imputation (allocation).

Non-life insurance is characterised by operations which are not related to the insured's life. It includes property damage insurance, liability or debt insurance, and personal insurance. The difference between non-life insurance and life insurance is the occurrence of loss, which is most often "certain" in life insurance, while it is "probable" in non-life insurance. In Non-Life Insurance claims costs are rarely known. Therefore, it's necessary for the insurer of this line of business to produce some models of its portfolio in order to simulate the frequency of claims and their cost. Insurance companies must be solvent in order to be able to meet their commitments. In order to enable them to build these models, regulations have been created, as we can see the Solvency 2 regulation promulgated in 2016, which succeeds the Solvency 1. The goal of the Solvency 2 regulation is to better evaluate the required capital of insurance and reinsurance companies in order to reinforce their financial soundness. The evaluation of this amount is an important step, it depends on of the company risks profile, which then requires good control of the various risks of the company.

Because of the specificity of its "Medical Liability" risk, SHAM had to set up a partial internal model to cover the limits of the Solvency II regulatory methods in the evaluation of capital requirements, in order to assess its premiums risk and reserves risk. Studies on the model assumptions were necessary to improve the relevance of the premiums SCR estimation. Some limits in the partial internal model were identified by Sham and the external audit. This report aims to study one of these limits.

The aim of this subject is to study the impact of a new segmentation in the premium SCR estimation, within the framework of Sham's partial internal model. This will involve modifying the current model, by challenging some assumptions, and measuring the impacts on the Premiums Risk in order to improve the model. In other words, we will study the relevance of a new segmentation of a class of risks in the evaluation of the premiums SCR.

The thesis will follow the following plan :

- First of all, we will present the partial internal model set up by Sham and validated by the ACPR in January 2016. We will the Premiums SCR methodology, precisely the frequency of claims modeling.
- Then, we will introduce the restated data and segmentations methodologies necessary for our segmentation study.
- Finally, we will perform different impact tests of alternative hypotheses and we will conclude with a comparison of the quality of the results obtained with the segmentation study. We will also check the stability of the different methods used.

In conclusion, we will discuss the need for modification of the premium risk calculation according to the policy of change of model defined during the approval of the Partial Internal Model by the ACPR..

## Remerciements

Avant de commencer cette étude, je souhaite remercier ceux qui m'ont accompagné et aidé à réaliser ce mémoire.

Je tiens à remercier particulièrement Leila Eljalat, ma tutrice en entreprise, qui m'a formée et accompagnée tout au long de cette expérience avec beaucoup de patience et de pédagogie. Aussi, je remercie les membres de l'équipe du Risk Management pour les conseils qu'ils ont pu me prodiguer au cours de ces 13 mois.

Je souhaite également remercier Mme Anne LOISEL EYRAUD, professeur à l'ISFA, pour son écoute et sa disponibilité tout au long de la réalisation de ce mémoire.

Pour finir, je dédie ce mémoire à mes parents, qui me soutiennent et continuent de croire en moi dans mes études en France.

# Table des matières

<b>Résumé</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>Remerciements</b>	<b>3</b>
<b>Introduction</b>	<b>7</b>
<b>1 Présentation Générale</b>	<b>8</b>
1.1 La Société Hospitalière d'Assurance Mutuelle (SHAM)	8
1.2 La Responsabilité Civile Médicale (RCM)	9
1.2.1 Généralités	9
1.2.2 Le Contrat d'assurance Responsabilité Civile Médicale (RCM)	10
1.2.3 Loi About (2002)	11
1.3 La réglementation Solvabilité 2	11
<b>2 Vers un Modèle Interne Partiel pour prendre en compte la spécificité du risque RCM</b>	<b>14</b>
2.1 Pilier 1 Solvabilité 2 : Evaluation du Solvency Capital Requirement (SCR)	14
2.1.1 Différentes méthodes possibles	14
2.1.2 Approche modulaire du calcul du SCR	15
2.1.3 SCR Souscription Non-Vie	17
2.2 Le Modèle Interne Partiel (MIP) SHAM	18
2.2.1 Données utilisées	18
2.2.2 Présentation du modèle	19
2.2.3 Modélisation de la fréquence	20
2.2.4 Modélisation du coût moyen par sinistre	22
2.2.5 Problématique étudiée	23
<b>3 Analyse et exploration des données</b>	<b>27</b>
3.1 Présentation des données	27
3.2 Analyse descriptive des données	29
3.2.1 Analyse globale	29
3.2.2 Fiabilité des données	30
<b>4 Méthodes de segmentations</b>	<b>32</b>
4.1 La classification non supervisée	32
4.1.1 Classification Ascendante Hiérarchique	32
4.1.2 Classification par ré-allocation dynamique (K-means)	32
4.2 La classification supervisée	32
4.2.1 K plus proches voisins (K-NN)	33
4.2.2 Arbre de décision	33
4.2.3 Réseaux neurones	33
<b>5 Segmentations alternatives</b>	<b>34</b>
5.1 Première segmentation testée : Clinique/ Etablissement hospitalier/ Autre	34
5.2 Deuxième segmentation testée : Public/ Privé/ Autre	37

5.2.1	La classe Public : . . . . .	37
5.2.2	La classe Privé : . . . . .	38
5.2.3	La classe Autre : . . . . .	39
5.3	Troisième segmentation testée : Public/ Privé à but lucratif/ Privé à but non lucratif/ Autre . . . . .	41
5.3.1	La classe Privé à but lucratif . . . . .	41
5.3.2	La classe "Privé à but non lucratif" : . . . . .	42
5.4	Segmentations obtenus par méthodes statistiques . . . . .	43
5.4.1	Traitement des données . . . . .	43
5.4.2	Approche Classification Ascendante Hiérarchique (CAH) . . . . .	43
5.4.3	Approche K-Means . . . . .	45
5.4.4	Quatrième segmentation testée . . . . .	46
5.5	Estimation de la fréquence pour l'exercice N+1 . . . . .	47
5.5.1	Modèle de fréquence "méthode moyenne" : Approche par une régression de Poisson . . . . .	47
5.5.2	Modèle de fréquence : Approche par une régression linéaire . . . . .	48
5.5.3	Modélisation de la fréquence sur chaque segmentation proposée . . . . .	49
5.5.4	Tests de validation des modèles . . . . .	53
5.5.5	Deuxième approche : retraitement des MCO . . . . .	54
5.6	Vérification de l'hypothèse d'indépendance des classes RC autre . . . . .	56
5.6.1	Test d'indépendance du Khi 2 . . . . .	57
5.6.2	Rank-Rank plot et Scatter plot . . . . .	57
5.6.3	Diagramme de Kendall ou K-plots . . . . .	58
<b>6</b>	<b>Vérification de la pertinence de la nouvelle segmentation dans le calcul du SCR primes</b>	<b>61</b>
6.1	Simulation du nombre de sinistres pour l'exercice de souscription N+1 . . . . .	61
6.1.1	Algorithme de simulation du nombre de sinistres . . . . .	61
6.1.2	Résultats obtenus sur nos différentes segmentations . . . . .	61
6.2	Détermination du risque de primes . . . . .	62
6.3	Comparaison et Backtesting : impact sur le risque de primes et choix de la segmentation . . . . .	63
6.3.1	Comparaison sans retraitement du groupe MCO X . . . . .	63
6.3.2	Comparaison avec retraitement du groupe MCO X . . . . .	64
6.4	Projection avec la vision 31/12/2019 . . . . .	64
6.5	Etude de sensibilité sur l'hypothèse d'indépendance des classes globales . . . . .	65
<b>7</b>	<b>Recherche d'une variable d'exposition alternative</b>	<b>66</b>
7.1	Etude des variables potentielles . . . . .	66
7.2	Construction de l'indice d'inflation . . . . .	67
7.3	Gestion des données manquantes . . . . .	68
7.3.1	Méthode par la moyenne . . . . .	69
7.3.2	Méthode par la médiane . . . . .	71
7.3.3	Méthode par simulation . . . . .	72
7.4	Modélisation de la fréquence . . . . .	74
7.4.1	Estimation de la fréquence . . . . .	74
7.4.2	Estimation de l'exposition . . . . .	78
7.5	Validation et mesure de l'impact sur le risque de primes . . . . .	80
7.6	Backtesting . . . . .	81
7.7	Tests de sensibilités : Impact des données manquantes de la variable BUDGET . . . . .	81

<b>Conclusion</b>	<b>83</b>
<b>Annexes</b>	<b>84</b>
Annexe 1 : Vérification de l'hypothèse d'indépendance des classes RC autre . . . . .	84
Annexe 2 : Sous-dispersion de la fréquence historique . . . . .	84
Annexe 3 :Quelques tests de sensibilités . . . . .	84
Segmentation 2 : Revue de la modélisation de la fréquence de la classe "Privé" RC autre . . . . .	85
Segmentation 3 bis : Revue de la modélisation de la fréquence de la classe "Privé à but non lucratif"	85
Annexe 3 : Autres résultats sur les différentes modélisations de fréquences . . . . .	86
Segmentation 2 . . . . .	87
Segmentation 3 . . . . .	87
<b>Bibliographies</b>	<b>88</b>

# Introduction

La nouvelle réglementation **Solvabilité II** est entrée en vigueur dans un contexte financier particulièrement tendu. Cette dernière a pour objectif d'assurer la solvabilité et la fiabilité des entreprises exerçant une activité d'assurance ou de réassurance de manière plus précise et davantage axée sur les risques pris par les entreprises que la réglementation Solvabilité I. Ainsi, les entités exerçant une activité d'assurance sont tenues d'élaborer leur bilan en se basant sur cette nouvelle réglementation, et doivent détenir un montant de Fonds Propres minimum qui prend en compte leur activité et le niveau de risque pris. Pour calculer le montant de fonds propres, l'assureur peut utiliser la Formule Standard qui est une approche normative des risques. Cependant, l'application de cette formule ne permettant pas forcément de refléter le profil spécifique de l'entité, d'autres solutions alternatives sont à la disposition de l'assureur : l'utilisation des USP (définis ci-après), ou encore l'élaboration d'un modèle interne.

Ce mémoire a été réalisé dans le cadre de ma formation d'actuaire à l'ISFA (Institut des sciences financières et de l'assurance). Ce projet a été effectué au sein de l'équipe Risk Management de SHAM, qui est un acteur de référence en matière de **Responsabilité Civile Médicale (RCM)**. La Responsabilité Civile Médicale (RCM) est une garantie obligatoire, concernant tout établissement ou entité effectuant une activité en rapport avec la santé. Elle a pour but de couvrir les établissements hospitaliers et les professionnels de santé contre les conséquences pécuniaires de la responsabilité encourue à l'égard des tiers à l'occasion de l'exercice de leur activité. La RCM représente 78% du chiffre d'affaires de SHAM.

Afin de répondre aux exigences de Solvabilité II, SHAM a élaboré un Modèle Interne Partiel pour le calcul de son risque de souscription non-vie afin de prendre en compte le caractère spécifique de son activité prépondérante : la RC Médicale. J'ai intégré l'équipe **Risk Management de SHAM** sous la responsabilité de Madame Leila ELIJALAT. SHAM a mis en place un Modèle Interne pour calculer son risque de primes et son risque de réserve relatifs à la RC Médicale France. J'ai eu l'occasion de travailler sur différentes missions en lien avec ce dernier qui sera présenté dans ce mémoire.

L'objectif principal de cette étude est de travailler sur l'amélioration continue du modèle interne partiel (MIP) de SHAM, de challenger certaines de ses hypothèses, et d'en mesurer l'impact.

Ma mission principale concernait de **la segmentation de risques pour la Responsabilité Civile Médicale (RCM)** dans le cadre du MIP. Dans notre étude, segmenter consiste à répartir les différents contrats dans des classes homogènes. Mon objectif était de proposer une nouvelle segmentation de classes de risques et d'étudier son impact sur l'évaluation du risque de primes.

Dans la première partie de ce mémoire, nous présenterons le contexte de cette étude. Nous verrons aussi la méthodologie utilisée par SHAM pour l'évaluation de son **SCR primes**.

Ensuite, nous procéderons à l'extraction et à l'analyse descriptive de nos données. Une fois nos données traitées, nous détaillerons la recherche d'une segmentation de la classe de risque concernée par l'étude. Pour cela, nous étudierons différentes segmentations, et ferons des analyses complémentaires tout au long de cette étude afin d'expliquer et d'améliorer nos différents modèles.

Puis, nous confronterons nos différentes segmentations testées, mesurerons leur impact sur le risque de primes et procéderons au choix de l'une d'entre elles. Enfin, nous vérifierons la stabilité de la segmentation choisie dans le futur.

*NB : Pour la protection des données de l'entreprise, les résultats présentés dans ce mémoire ont été masqués.*

# 1 Présentation Générale

## 1.1 La Société Hospitalière d'Assurance Mutuelle (SHAM)

Fondée en 1927 par des Directeurs d'hôpitaux, SHAM est une Société d'Assurance Mutuelle spécialisée dans l'assurance et le management des risques de l'ensemble des acteurs de l'offre de soins, comme des établissements publics et privés, organisations et professionnels ; ou encore des filières et coopérations des secteurs de la santé, du social et du médico-social. SHAM est un opérateur de référence français et européen en matière de responsabilité civile, et est devenu récemment Relyens. À travers SHAM, le groupe Relyens est en capacité d'accompagner les établissements de santé, sociaux et médico-sociaux dans leur approche globale de management des risques. La structuration et la mutualisation des ressources, la mise en place d'outils et de méthodes dédiés, le développement de la coopération et du partage d'expériences sont autant d'orientations clé pour assurer la permanence et la qualité des soins apportés aux patients dans un cadre budgétaire contraint, tout en garantissant de bonnes conditions de travail pour les professionnels de santé.

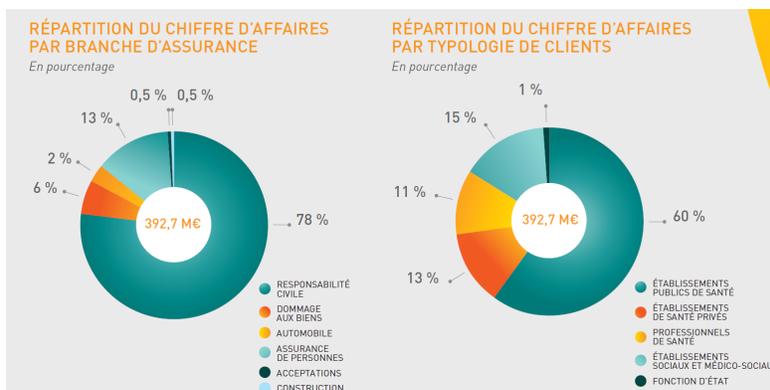
Cela fait quatre-vingt-dix ans que SHAM est un acteur mutualiste spécialisé dans l'assurance et le management des risques, assurant la totalité du parcours patient et des professionnels de santé, capable de couvrir tous les types d'établissements sur un territoire, pour un accompagnement sur le long terme.

Opérateur de référence en matière de Responsabilité Civile Médicale en France, en Espagne, en Italie et en Allemagne, comptant environ 530 collaborateurs répartis dans ces pays et plus de 10 000 sociétaires, SHAM propose à ses clients sociétaires l'ensemble des solutions nécessaires au pilotage pérenne et optimales de leurs risques : identification, évaluation, prévention, couverture, gestion et indemnisation.

Les différentes branches d'assurances présentes à SHAM sont :

- Responsabilité civile
- Dommages aux biens
- Automobile
- Assurance de personnes
- Constructions

Dans les graphiques ci-dessous, nous pouvons voir la répartition du Chiffre d'affaires de SHAM en 2018 par branche d'activité ainsi que par type de clients.



Sur le premier graphique, nous pouvons constater que la branche prépondérante est la RC Médicale, elle représente près de 80% du chiffre d'affaires de SHAM. Sur le second, nous pouvons voir que les établissements publics de santé constituent plus de la moitié du chiffre d'affaires de SHAM.

Dans ce mémoire, nous nous focaliserons uniquement sur la branche Responsabilité Civile Médicale qui sera présentée ci-dessous.

## 1.2 La Responsabilité Civile Médicale (RCM)

### 1.2.1 Généralités

Comme vu dans la partie précédente, SHAM est une compagnie d'assurance couvrant les établissements et professionnels de santé pour divers risques. Dans ce mémoire, je me focaliserai uniquement sur la RC médicale.

La Responsabilité Civile Médicale (RCM) est une garantie obligatoire, concernant tout établissement ou entité effectuant une activité en rapport avec la santé. La RCM a pour but de couvrir les établissements hospitaliers et les professionnels de santé contre les conséquences pécuniaires de la responsabilité encourue à l'égard des tiers à l'occasion de l'exercice de leur activité. Cela couvre les dommages résultants d'atteintes à la personne, par exemple les dommages corporels survenant dans le cadre des activités de prévention, de diagnostic et de soins. Le contrat responsabilité civile médicale est un contrat en base réclamation, c'est-à-dire que la victime doit solliciter expressément l'indemnisation du dommage qui lui a été causé.

L'offre d'assurance en RCM se répartit aujourd'hui entre un nombre limité d'acteurs généralistes, qui constitue une de leurs activités secondaires, et quelques sociétés d'assurance spécialisées, telles que Sham, pour lesquelles elle est une activité principale. La RCM présente des caractéristiques propres en raison d'un déroulement très long, de la coexistence de risques hétérogènes (praticiens et établissements de soins, centres public hospitaliers et cliniques privées,...). Il faut des modèles techniques pour estimer le risque et prédire les sinistres, car ce sont des sinistres à déroulement long, avec des montants importants pour un certain nombre de sinistres, avec plusieurs jugements possibles par les tribunaux intervenant au cours du sinistre.

La nature de la relation entre un médecin et son patient est contractuelle. Ce contrat est défini par l'arrêt Mercier du 20 Mai 1936 de la Cour de Cassation. Dans cet arrêt, il est écrit : « Il se forme entre le médecin et son client un véritable contrat comprenant pour le praticien de l'engagement, sinon, bien évidemment de guérir le malade [...] du moins de lui donner des soins [...] consciencieux et attentifs, et, réserve faite de circonstances exceptionnelles conformes aux données acquises de la science. ». La responsabilité du médecin ne peut être engagée qu'en cas de faute.

On distingue cependant deux grands types de responsabilités :

- ***La responsabilité pénale***

C'est la responsabilité qu'encourt un individu lorsque son comportement fautif est à l'origine du dommage, et ce, afin d'obtenir des sanctions. Concrètement, la victime va porter plainte auprès des services de police ou de gendarmerie afin de déclencher une enquête dirigée par le Procureur susceptible d'aboutir à une mise en examen par le juge d'instruction, puis à une condamnation par le tribunal : peine d'emprisonnement (avec sursis généralement) et/ou amende.

- ***La responsabilité civile***

Il s'agit de la responsabilité de l'institution ou de la personne « civilement responsable », et ce, afin d'obtenir une indemnisation. Concrètement, la victime va user des différentes voies de recours à sa disposition (demande amiable, saisine de la CCI (Commission de Conciliation et d'Indemnisation des accidents médicaux, des affections iatrogènes et des infections nosocomiales), saisine des tribunaux judiciaires ou administratifs) afin que le responsable lui verse des dommages-intérêts.

Contrairement à la responsabilité pénale qui atteint directement chacune des personnes fautives, et ce, quel que soit leur statut (agent public, salarié, libéral), la responsabilité civile n'atteint que celui qui est désigné comme étant « civilement responsable » de la personne fautive. Le « responsable - auteur » (celui qui cause le dommage) peut ne pas être le « responsable - débiteur » (celui qui est tenu de réparer le dommage).

Concernant la responsabilité civile médicale, nous faisons la distinction entre deux types d'établissements de santé : les établissements publics et les établissements privés. Nous verrons que le mécanisme lié à la responsabilité n'est pas forcément le même selon le type de l'établissement.

- ***Les établissements publics***

Ce sont par exemple les centres hospitaliers universitaires. Dans ces types d'établissements, les employés ne sont pas considérés directement comme étant civilement responsables des dommages qu'ils causent à l'occasion de leur activité professionnelle. C'est l'établissement public employeur qui assume cette responsabilité indemnitaire vis-à-vis des victimes. S'il n'y a pas de faute personnelle, c'est l'assureur de l'établissement qui va verser l'indemnisation aux victimes si la responsabilité de son assuré est engagée, sinon l'agent public peut voir sa responsabilité civile engagée. Une faute personnelle est une faute relevant d'un « manquement volontaire et inexcusable à des obligations d'ordre professionnel ou déontologique ». En pratique, la faute personnelle en établissement de santé est très peu retenue par la jurisprudence.

On considère aussi comme « faute personnelle » :

- La fuite d'un médecin lors d'un incendie au bloc opératoire laissant sa patiente anesthésiée sur la table d'opération ;
- Le refus d'un obstétricien d'astreinte de se déplacer pour pratiquer une césarienne en urgence alors même qu'il avait pleinement conscience de la situation de danger imminent ;
- Le silence gardé par un chef de service sur une erreur d'injection commise par un médecin de son service ayant provoqué un choc septique grave.

- ***Les établissements de santé privés***

On distingue dans ces établissements :

- o Le praticien salarié de l'établissement : ce dernier est considéré comme l'agent public, et ne voit sa responsabilité civile individuelle engagée que s'il commet un « abus de fonction », l'établissement employeur étant civilement responsable des fautes de ses préposés. Un abus de fonction en santé, c'est lorsque le praticien agit à des fins étrangères à ses attributions sans autorisation de son employeur.

- o Le praticien exerçant à titre libéral : le praticien exerçant à titre libéral doit obligatoirement souscrire une assurance RC à titre individuel (contrats professionnels de santé). Il engage nécessairement sa responsabilité civile personnelle, l'établissement n'étant responsable que de l'organisation des soins.

La condamnation du professionnel ou de l'établissement de santé donne lieu au versement de dommages et intérêts au profit du patient. Le principe est celui de la réparation intégrale des préjudices subis. En pratique, c'est la compagnie d'assurance du professionnel ou de l'établissement de santé qui procède à ce versement. C'est dans ce contexte que s'inscrit l'assurance responsabilité civile médicale.

### **1.2.2 Le Contrat d'assurance Responsabilité Civile Médicale (RCM)**

Il s'agit dans cette partie de spécifier les différentes garanties couvertes par ce contrat. Le contrat d'assurance responsabilité civile médicale garantit les conséquences pécuniaires des dommages corporels, matériels et immatériels

que l'assuré peut causer à des tiers au cours de son activité. Sur le périmètre français, les garanties couvertes par les différents contrats RC médicales de SHAM sont :

- La responsabilité civile hospitalière
- La responsabilité civile des cliniques
- La responsabilité civile des professionnels
- La responsabilité civile psychiatrique
- La responsabilité civile des dirigeants
- La responsabilité civile recherches biomédicales
- La responsabilité civile des Etablissements d'Hébergement pour Personnes Agées (EHPA)
- La responsabilité civile des établissements accueillant des personnes handicapées ou en difficulté sociale
- La responsabilité civile des Services Départementaux d'Incendie et de Secours (SDIS)

### 1.2.3 Loi About (2002)

L'année de rattachement d'un sinistre correspond à l'année à laquelle est rattaché le sinistre. Souvent, il s'agit de l'année de survenance du sinistre. Cependant, il existe des cas particuliers pour lesquels le rattachement se fait par année de souscription ou d'ouverture du sinistre.

En 2002, la loi About est entrée en vigueur. Cette loi, relative à la responsabilité médicale, modifie le fonctionnement de la RC médicale en instaurant le principe de la base réclamation et la garantie subséquente.

Avant 2002, la RC médicale appliquait un régime en base survenance. La date du fait dommageable (ou fait générateur) du sinistre était l'élément retenu par la jurisprudence de la Cour de cassation pour déterminer la période de garantie prévue au contrat d'assurance RC médicale. Les sinistres étaient rattachés à l'exercice comptable au cours duquel survenait le sinistre. La loi About a rendu caduc ce raisonnement, rendant obligatoire le régime en base réclamation afin de faciliter la détermination de l'assureur impliqué par le sinistre. Ainsi, l'assureur dont le contrat est en cours lors de la réclamation du sinistre par le patient doit prendre en charge le sinistre quelle que soit la date du fait dommageable. Les sinistres sont rattachés à l'exercice comptable au cours duquel la demande de réparation amiable ou contentieuse a été formulée par la victime ou les ayants droit à l'assuré ou à l'assureur. Les sinistres relevant de la garantie subséquente sont donc également rattachés à l'année de réclamation et non à la date d'expiration ou de résiliation de tout ou partie des garanties du contrat.

Cette loi vise notamment à adapter les contrats de responsabilité civile médicale à la spécificité des dommages consécutifs à des accidents médicaux, qui peuvent survenir de nombreuses années après la réalisation des actes de soins. En la matière elle introduit dans le code des assurances une définition de la notion de sinistre : il s'agit du ou des dommages qui engagent la responsabilité de l'assuré et qui résultent d'un ou de faits générateurs ayant la même cause technique, imputables aux activités garanties dans le cadre du contrat et ayant donné lieu à réclamation. Elle définit également la réclamation comme étant une demande en réparation, amiable ou contentieuse, formée par la victime ou ses ayants droit et adressée à l'assuré ou à son assureur. Elle précise que le contrat d'assurance garantit les sinistres pour lesquels la première réclamation est formée pendant le contrat.

Ainsi d'après la réglementation spécifique à la RC Médicale les sinistres sont gérés par année de réclamation (année d'ouverture) depuis 2002.

## 1.3 La réglementation Solvabilité 2

Adoptée par le Parlement européen en 2009, la Directive Solvabilité 2 est un nouveau système pour le secteur de l'assurance. Cette nouvelle norme a été introduite afin de pallier les insuffisances de la Directive Solvabilité 1.

Elle a pour but de :

- Créer un marché unique européen de l'assurance ;
- Renforcer la solidité des assureurs et la sécurité des assurés ;
- Homogénéiser la mesure des fonds propres et le capital requis à leur niveau économique entre les pays européens ;
- Evaluer la solvabilité globale basée sur tous les risques ;
- Inciter les assureurs à mesurer et contrôler leurs risques.

La structure de cette nouvelle norme se présente sous la forme de 3 piliers :

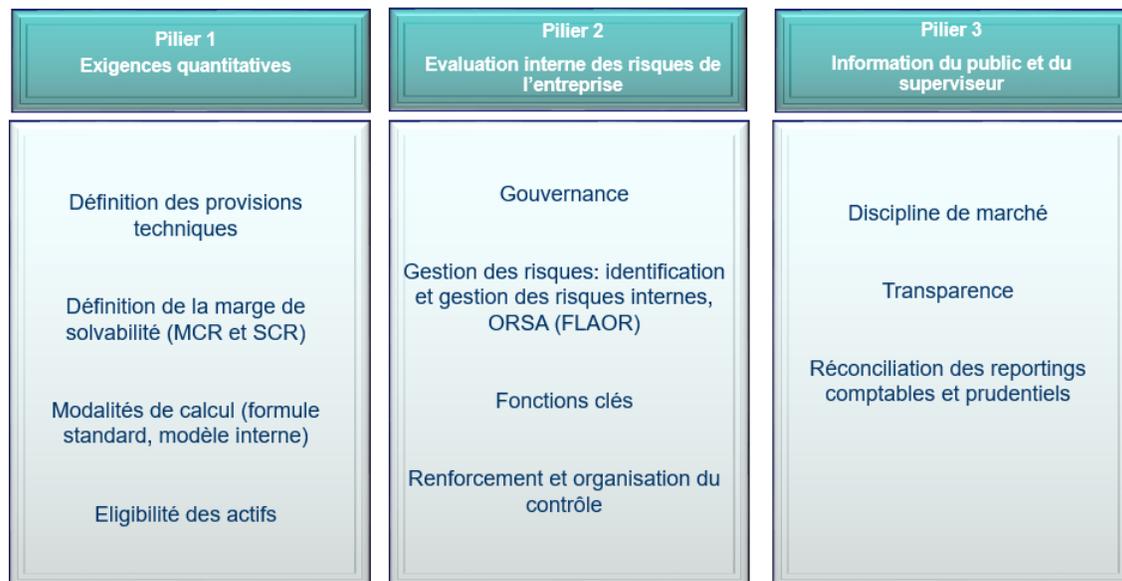


FIGURE 1 – Présentation des piliers de la solvabilité 2

Dans le cadre de cette étude, je me limiterai au pilier 1 de Solvabilité 2. Le pilier 1 de la Solvabilité 2 exige de calculer deux capitaux réglementaires :

- Le MCR (Minimum Capital Requirement), est le minimum de fonds propres requis pour qu'une entreprise d'assurance puisse exercer son activité.
- Le SCR (Solvency Capital Requirement) correspond au niveau de fonds propres permettant à une entreprise d'absorber des pertes imprévues importantes et d'offrir une sécurité raisonnable aux assurés et aux souscripteurs. On doit estimer ce montant chaque année pour l'exercice suivant.

Dans le cadre de cette étude, nous nous focaliserons sur le calcul du Solvency Capital Requirement (SCR). Tout d'abord, ce dernier est l'outil principal d'appréciation du niveau de solvabilité utilisé par les autorités de contrôle. En effet, le SCR est fondé sur l'exposition aux risques en incorporant les principaux risques liés à l'activité de l'organisme d'assurance à savoir le risque de souscription, le risque de défaut de contrepartie, le risque opérationnel, et le risque de marché. Le calcul du SCR constitue donc un enjeu majeur pour les organismes d'assurance.

A partir du SCR obtenu pour l'année d'exercice en cours, l'assureur en déduit son ratio de solvabilité. Le ratio de solvabilité est le rapport entre les fonds propres détenus par l'entreprise et le SCR. Ce ratio permet de déterminer la capacité de l'entreprise à faire face à ses engagements, il doit être au moins égal à 100%. A la clôture de l'année 2018, le ratio de solvabilité de SHAM est de 170 %. SHAM fait donc face à ses engagements à hauteur de 170%, dépassant le seuil minimal de 100 % requis dans le cadre de la directive européenne Solvabilité 2. Ce

ratio de solvabilité se base sur le Modèle Interne Partiel (MIP) de SHAM, qui prend en compte les caractéristiques particulières des risques de RC médicale portés par le groupe que nous verrons plus en détail par la suite.

## 2 Vers un Modèle Interne Partiel pour prendre en compte la spécificité du risque RCM

### 2.1 Pilier 1 Solvabilité 2 : Evaluation du Solvency Capital Requirement (SCR)

Le SCR mesure le risque d'événement extrême survenant une fois tous les 200 ans à horizon 1 an. Cela correspond à la Value-at-Risk à 99,5% sur un horizon d'un an : si le montant de fonds propres d'un organisme est égal au SCR, l'organisme pourra faire face à ses engagements dans 99,5% des cas, sur un horizon d'un an.

#### 2.1.1 Différentes méthodes possibles

Au vu des enjeux que cela regroupe, l'évaluation du SCR est une phase importante pour l'assureur. Dans la directive Solvabilité 2, l'assureur a à sa disposition plusieurs alternatives lui permettant de calculer son SCR, présentées ci-dessous :

##### — La Formule Standard (FS) :

La Formule Standard est une approche normative des risques. Elle a été élaborée pour être appliquée par l'ensemble des assureurs européens. Son calibrage a été réalisé sur les données de l'ensemble des assureurs européens. Celle-ci peut ne pas être adaptée pour prendre en compte la spécificité du profil de risque de certaines entreprises, d'où l'utilisation des USP (cf. ci-après) ou l'élaboration d'un Modèle Interne. Il est également possible de recourir à des versions ajustées de la Formule Standard ou un Modèle Interne Partiel, permettant de distinguer suivant les branches de l'organisme, ou suivant les risques, Formule Standard et Modèle Interne.

##### — Undertaking Specific Parameter (USP) :

Afin d'améliorer sa visibilité sur son profil de risque, l'assureur a la possibilité de développer la méthodologie des Undertaking Specific Parameters, c'est-à-dire la mise en oeuvre de la Formule Standard, mais dont certains paramètres sont calibrés sur les données historiques propres à l'entreprise. Cette approche, qui reste relativement simple, permet de se rapprocher du profil de risque de l'organisme, mais est soumise à un certain nombre de conditions.

Le principe de l'application des USP est de calculer la volatilité des risques de primes et de réserve, afin de les incorporer au calcul de la Formule Standard.

Pour prendre en compte la spécificité de son risque RC Médicale SHAM avait testé l'utilisation des USP. Par contre, celle-ci fut écartée car non adaptée au risque RC médicale porté par SHAM.

##### — Le Modèle Interne total :

Les compagnies d'assurance peuvent développer un Modèle Interne, afin d'évaluer le capital réglementaire de solvabilité. Le modèle sera alors propre à l'entreprise qui l'a développé et devra être soumis à l'approbation des autorités de contrôle. Le développement d'un Modèle Interne permet de se détacher de la Formule Standard et de calculer le capital réglementaire d'une manière cohérente avec les caractéristiques propres à l'activité exercée et avec l'organisation mise en place.

##### — Le Modèle interne partiel (MIP) :

Certaines compagnies d'assurance élaborent un modèle interne uniquement sur certains risques pour lesquels la Formule Standard n'est pas adaptée, on parle alors de Modèle Interne Partiel. Le modèle interne partiel (MIP) constitue une alternative pertinente à la Formule Standard pour les organismes d'assurance de petite et moyenne taille. En effet, la mise en place d'un modèle interne total peut être très coûteuse en temps et en moyens. La mise

en place d'un MIP se présente alors comme le compromis idéal, car il permet de prendre en compte les spécificités de certains risques tout en gardant la Formule Standard pour le calcul du SCR des autres branches.

J'ai intégré l'équipe Risk Management (RM) du département des risques de SHAM sous la supervision de Leila Eljalat, actuaire en Risk Management. Durant mon alternance, j'ai eu l'occasion de participer à différents travaux en lien avec le modèle interne partiel de SHAM, qui est un outil élaboré par SHAM permettant de calculer le SCR des risques de primes et de réserves concernant la RC en France, afin de prendre en compte la spécificité de son profil de risque. Le but principal de mon alternance était d'effectuer la segmentation des risques dans le calcul du risque de primes. J'ai eu aussi l'occasion de travailler sur d'autres missions annexes qui m'ont permis de mieux comprendre le MIP et de développer des compétences techniques et mes connaissances.

L'équipe Risk Management fait partie du Département des Risques de SHAM et est en charge de la gestion et du suivi des risques de l'assureur du groupe SHAM et de l'ensemble de ses filiales. L'équipe se divise en deux pôles : pôle Passif, pôle Actif du bilan prudentiel de l'entreprise. J'ai été affectée dans le pôle Passif. Composée de cinq collaborateurs, l'équipe RM est notamment en charge du calcul du ratio de Solvabilité 2 incluant la mise à jour du modèle interne partiel, comme le calcul du Solvency Capital Requirement (SCR), ainsi que les calculs liés à l'ORSA.

L'équipe RM travaille en étroite collaboration avec d'autres services comme la comptabilité, le contrôle de gestion, le département technique et actuariat. L'équipe participe aussi à la réalisation des reportings à l'ACPR, qui est l'autorité de contrôle des organismes d'assurance, notamment en charge de valider ou non le MIP, contrôler la solvabilité de l'entreprise, ainsi que d'autres missions de contrôle.

### 2.1.2 Approche modulaire du calcul du SCR

Étant la méthode la plus utilisée par les assureurs, la Formule Standard (FS) est sur une approche normative des risques et repose sur la formule suivante :

$$SCR = BSCR + Adj + SCR_{op}$$

Avec :

- $BSCR$  : Basic Solvency Capital Requirement, il s'agit du capital réglementaire de solvabilité de base
- $Adj$  : Ajustement - effet d'absorption
- $SCR_{op}$  : Le capital réglementaire requis pour couvrir le risque opérationnel

Le calcul du SCR est modulaire. C'est-à-dire qu'on a plusieurs modules de risques, et le calcul s'effectue sur chaque module (un module de risque étant lui-même composé de sous-modules). En effet, le BSCR se décompose lui-même également pour la Formule Standard en plusieurs sous-SCR, comme nous pouvons le voir dans ce graphique suivant :

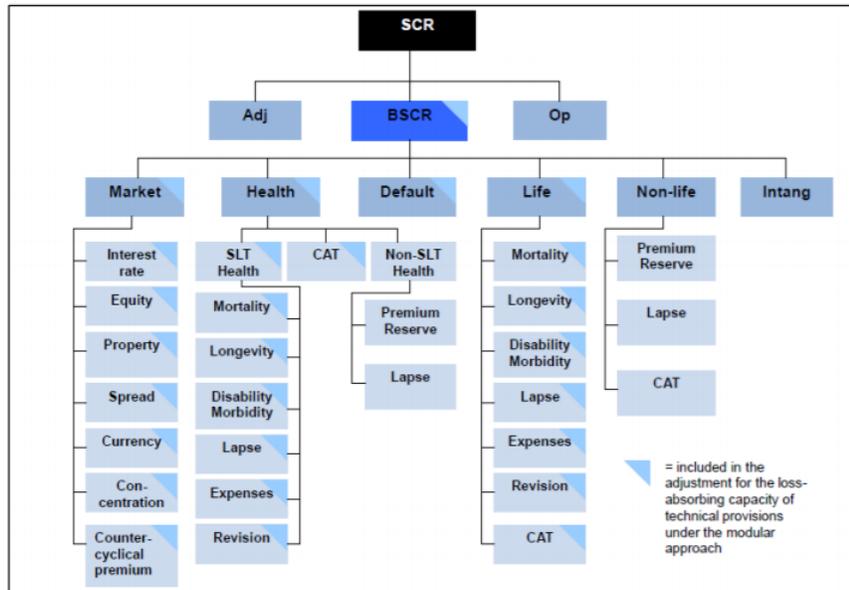


FIGURE 2 – SCR : Approche modulaire

SHAM utilise la Formule Standard pour la plupart de ses risques sauf pour son module souscription non-vie en RC France. En effet à cause de la spécificité de ce risque, il est nécessaire de réaliser des modèles qui permettent de prendre en compte sa particularité. D’où l’élaboration d’un modèle interne partiel. Ainsi, le calcul du SCR au sein de SHAM suit la décomposition suivante :

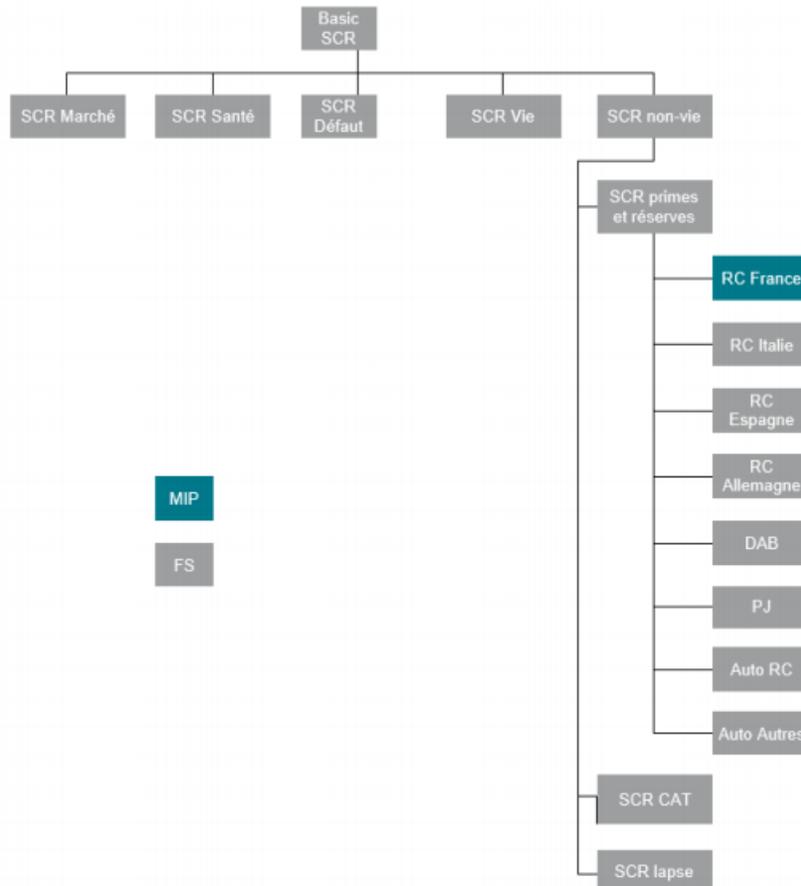


FIGURE 3 – Approche Modulaire du SCR SHAM

Le MIP SHAM que nous présenterons dans la suite de ce mémoire porte uniquement sur les risques de primes et réserves RC France. Pour le reste de ses risques SHAM utilise la Formule Standard.

### 2.1.3 SCR Souscription Non-Vie

Le calcul du SCR non-vie est aussi modulaire. En effet dans ce même module nous avons trois sous-modules de risques qui sont agrégés : le risque de primes et de réserves, le risque de cessation ainsi que le risque de catastrophe en non-vie. Le SCR de chaque module est calculé spécifiquement. En effet :

- Le SCR risque de primes et de réserves en non-vie ( $SCR_{nl,prime,res}$ ) :

Le risque de primes est le risque de sous-estimation de la sinistralité relative aux primes qui seront acquises au cours de l'année à venir.

Le risque de réserves est le risque lié à la nature aléatoire de l'évaluation des sinistres du stock et à leur mauvaise estimation.

Le SCR de ce module par la Formule Standard est donné par :

$$SCR_{nl,prime,res} = 3\sigma_{nl}V_{nl}$$

Avec :

$\sigma_{nl}$  : l'écart-type du risque de primes et de réserves ;

$V_{nl}$  : la mesure de volume pour le risque de primes et de réserves en non-vie. Ce dernier correspond à la somme du volume de risque de réserve  $V_{nl,reserves}$  et du volume du risque de primes  $V_{nl,prime}$ . De plus, le  $V_{nl,prime}$  est composé des cotisations (primes à acquérir au cours des 12 mois à venir + valeur actuelle des

primes futures), tandis que le  $V_{nl,reserves}$  est composé des Best Estimates de sinistres.

— Le SCR risque de cessation en non-vie ( $SCR_{nl,lapse}$ ) :

Dans ce sous-module, il s'agit d'évaluer l'impact qu'aurait une baisse de l'activité en non-vie sur le portefeuille d'assurance. Le risque de cessation est nul pour Sham, car :

- L'émission des primes est annuelle ; l'assuré est couvert sur la période de garantie qu'il paie
- La majorité des contrats RC est souscrite sous forme d'appels d'offres
- La RC est une garantie obligatoire pour les assurés

— Le SCR risque de catastrophe non-vie ( $SCR_{nl,cat}$ ) :

L'objectif principal de la mise en place d'un tel sous-module est de prémunir l'entreprise contre le risque d'un événement catastrophe sur le nouvel exercice de souscription, qui n'est intégré ni dans le risque de primes ni dans le risque de réserves.

Nous obtenons enfin le SCR souscription non-vie en Formule Standard par :

$$SCR_{nl} = \sqrt{\sum CorrNL_{ij} SCR_i SCR_j}$$

où CorrNL représente la matrice des corrélations entre les différents sous-modules du SCR de souscription non-vie.

Dans cette formule, nous pouvons constater la prise en compte de l'effet de diversification de chaque sous-module de risque. En effet, le fait d'allouer des coefficients de corrélation inférieurs à 1 implique que, le capital réglementaire agrégé sera toujours inférieur à la somme des capitaux réglementaires, ce qui va faire apparaître un bénéfice de diversification.

La Formule Standard est une méthode proposée dans la nouvelle réglementation Solvabilité 2 qui permet de calculer le SCR selon le profil de risque de l'assureur. Cependant, elle a été calibrée sur les risques de tous les assureurs européens, parmi lesquels des grands groupes diversifiés, et ces paramètres ne correspondent pas à un risque très spécifique concernant peu d'assureurs comme la RC médicale.

## 2.2 Le Modèle Interne Partiel (MIP) SHAM

Dans la première partie, nous avons constaté la spécificité du risque RC médicale. Lors du début des travaux relatifs à Solvabilité 2, SHAM a constaté une forte volatilité du ratio de Solvabilité 2 notamment due à la mauvaise adéquation de la Formule Standard aux spécificités du risque RC médicale. Plusieurs études ont montré que cette forte volatilité porte essentiellement sur les éléments de SCR relatifs au risque RC.

Différentes études ont alors été menées afin de mieux prendre en compte le risque RC dans la Formule Standard. Certains de ces travaux n'ayant pas été suffisants pour prendre en compte de manière adéquate le risque RC médicale dans la Formule Standard en particulier l'utilisation des USP, SHAM a donc pris la décision de développer un modèle interne partiel.

Dans cette partie, nous nous intéresserons au modèle interne partiel (MIP) en RC médicale mis en place par SHAM et validé par l'ACPR (Autorité de Contrôle Prudentiel et de Résolution) en Janvier 2016. Dans un premier temps, nous allons faire une présentation du modèle et expliquer plus en détail la méthodologie utilisée pour le calcul du SCR risque de primes tout en explicitant le contexte du sujet de ce mémoire, enfin nous verrons en détail la problématique étudiée.

### 2.2.1 Données utilisées

Dans le calcul du risque de primes, les sinistres sont gérés par exercice d'ouverture.

Pour la modélisation de la fréquence des sinistres, seul le nombre de sinistres corporels, relatifs aux garanties délivrées aux sociétaires français (y compris sinistres de subséquent) ayant un coût non nul en fin de 1ère année sur les exercices d’ouverture N-21 à N sont pris en compte dans le modèle. Les données utilisées proviennent des bases de données de sinistres RC fournies par la Direction des systèmes d’informations (un projet qualité de la donnée a été mis en place pour assurer l’exhaustivité, la pertinence et l’exactitude des informations).

Pour l’évaluation du coût moyen, les données issues des simulations du modèle BE sophistiqué au 31/12/N nettes de réassurance sont utilisées. Ces simulations intègrent la même inflation future que celle retenue dans le modèle BE RC sophistiqué. Ce coefficient d’inflation relève d’un jugement d’expert.

Comme les modélisations sont effectuées par année de réclamation (par année d’ouverture), la modélisation de la fréquence des sinistres tient aussi compte de sinistres tardifs. En effet, les sinistres tardifs sont des sinistres pour lesquelles l’année d’ouverture est supérieure à l’année de survenance, or la réglementation sur la Responsabilité Civile Médicale impose depuis de nombreuses années de gérer les sinistres en base réclamation. Enfin, comme les données sont considérées par année d’ouverture, les sinistres tardifs sont pris en compte dans la modélisation de la fréquence. Ainsi, le nombre de sinistres déclarés n’est pas complété par des IBNYR.

### 2.2.2 Présentation du modèle

Le modèle Interne partiel de SHAM a été élaboré pour le calcul du risque de primes et de réserves en souscription non-vie pour la RC France. Dans le MIP, le risque de primes et le risque de réserves sont calculés séparément. Dans cette étude, nous nous limiterons à la méthodologie utilisée pour le calcul du risque de primes.

Tout d’abord, le risque de primes est le risque de sous-estimation de la sinistralité relative aux primes qui seront acquises au cours de l’année à venir. Les cotisations utilisées pour l’évaluation du risque de primes modèle interne partiel sont déterministes. En effet, la quasi-totalité des cotisations RC France est connue dès le 1er janvier de chaque année, il y a donc très peu d’incertitudes sur le montant total des cotisations du nouvel exercice de souscription. Le seul facteur d’incertitude réside dans l’actualisation des assiettes de cotisations des centres hospitaliers pour le nouvel exercice de souscription, les cotisations étant calculées à partir du budget. Le risque de primes est calculé sur la base des cotisations acquise nettes de réassurance N+1, nettes des commissions de réassurance QP plancher.

De plus, pour le 31/12/N l’estimation des cotisations de l’exercice N+1 repose sur un coefficient d’évolution des cotisations entre la date d’extraction de la base de données et le 31/12/N+1.

Le SCR Primes est calculé à la clôture de l’année en cours (31/12/N) pour l’exercice de souscription N+1, et selon le principe indiqué par la Directive Solvabilité 2 : « le capital de solvabilité requis correspond à la valeur en risque des fonds propres de base de l’entreprise d’assurance, avec un niveau de confiance de 99,5% à horizon d’un an ».

Pour le calcul du risque de primes dans son Modèle interne partiel, SHAM utilise une méthode fréquence/coût. C’est-à-dire que nous allons estimer la charge ultime totale de l’exercice N+1 (en vision Best Estimate) à partir du nombre de sinistres simulé pour l’exercice de souscription N+1 multiplié par le coût moyen estimé pour un sinistre. Le coût moyen calibré et simulé est déjà net de réassurance.

Le nombre de sinistres N+1 est obtenu à partir d’un modèle de fréquence et la charge ultime d’un sinistre est obtenue à partir des simulations du modèle BE RC fournies par le Département Technique et Actuariat. Ainsi au 31/12/N le risque de primes de l’année N+1 vaut selon la méthodologie retenue dans le modèle interne partiel de SHAM :

$$SCR_{primes}^{31/12/N} = VAR_{99,5}(Charge_{ultime_{N+1}}) + Frais_{deGestion_{N+1}} - Cotisation_{N+1} - BE_{N+1}$$

Les détails des différents calculs seront expliqués plus tard dans ce mémoire.

La procédure du calcul du SCR du risque de primes peut se résumer sur le schéma suivant :

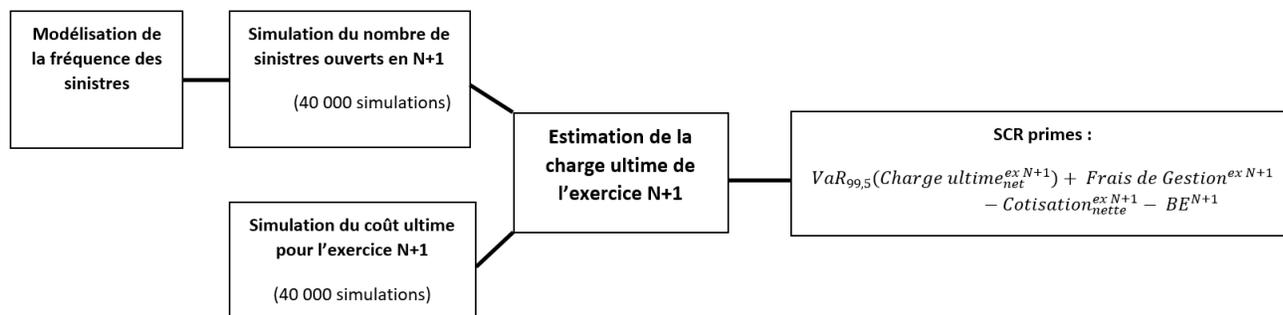


FIGURE 4 – SCHEMA de Synthèse de la méthodologie MIP pour le calcul du SCR Primes

Dans ce mémoire, nous nous concentrerons uniquement sur la partie concernant la modélisation de la fréquence des sinistres, mais nous présenterons néanmoins les impacts de l'étude sur l'évaluation du risque de primes.

### 2.2.3 Modélisation de la fréquence

En considérant que nous sommes l'année N, alors l'estimation du SCR primes pour le risque RC médicale par SHAM au 31/12/N nécessite une estimation préalable du nombre de sinistres qui pourrait être réclamés en N+1. Cependant, le portefeuille de Sham étant en constante évolution, l'utilisation de la fréquence (Nombre de sinistres/Exposition) plutôt que le nombre de sinistres historiques apparaît comme étant plus pertinente. Le choix d'une variable reflétant véritablement l'exposition au risque de nos établissements assurés est donc important, car il impacte directement le montant du SCR primes calculé. Cette variable d'exposition diffère selon le type d'établissement considéré. Cela constitue un challenge dans notre étude.

Dans cette partie, nous présenterons la méthodologie utilisée pour la modélisation de la fréquence. Tout d'abord, la fréquence de sinistres est définie par le quotient entre d'une part le nombre de sinistres corporels et d'autre part l'exposition.

SHAM a réalisé les projections de fréquence uniquement à partir du nombre de sinistres corporels. En effet, les sinistres matériels représentent 50% du nombre de sinistres total déclarés (ayant un coût non nul en fin de 1er inventaire) mais 2% du coût total en fin de 1ère année de l'ensemble de ces sinistres. La charge de sinistres est donc constituée dans sa quasi-totalité des sinistres corporels.

De plus, même si la fréquence pour les sinistres matériels ne sera pas modélisée dans le modèle, les sinistres matériels sont bien inclus dans le risque de primes Sham via le coût moyen par sinistre qui est estimé sur l'ensemble des sinistres matériels et corporels.

La modélisation de la fréquence des sinistres consiste à ajuster une loi de Poisson à la fréquence de sinistres à l'aide d'une régression de Poisson, afin de pouvoir simuler le nombre de sinistres qui seront ouverts en N+1 (car ce que nous recherchons au final c'est la VaR99.5). Dans cette étude, nous nous plaçons dans un premier temps

au 31/12/18 et estimons le risque de primes pour l'exercice de souscription 2019. La méthodologie utilisée suit le cheminement suivant :

— **Etape 1 : Segmentation du portefeuille RC médicale SHAM en classes de risques homogènes, et choix des variables d'exposition pour chaque classe de segmentation**

Avant de modéliser la fréquence des sinistres, il est nécessaire dans un premier temps d'effectuer une segmentation des contrats, c'est-à-dire créer des classes homogènes de contrats. Cette segmentation doit être pertinente pour la modélisation de la fréquence du nouvel exercice de souscription.

Segmenter consiste à regrouper des contrats qui semblent avoir les mêmes caractéristiques. Une fois la segmentation effectuée, ensuite arrive le choix de la variable d'exposition. La variable d'exposition doit être la même pour tous les contrats se retrouvant dans une même classe. De plus, son choix doit être pertinent et refléter au mieux le risque encouru par l'établissement.

Nous définissons la fréquence des sinistres par :

$$\text{fréquence} = \frac{\text{nombre des sinistres}}{\text{exposition}}$$

Nous calculons la fréquence pour chaque classe de contrats sur un historique donné (par exemple sur un historique de 10 ans), ensuite nous représentons graphiquement l'évolution de la fréquence pour chaque classe de contrats.

— **Etape 2 : Choix de l'historique retenu pour l'estimation de la fréquence moyenne de chaque classe de segmentation.**

Une fois les courbes de fréquences obtenues nous faisons des prévisions sur l'évolution de la fréquence dans le futur. Nous allons estimer la fréquence moyenne pour l'exercice de souscription 2019. Pour cela nous allons dans un premier temps choisir les années historiques qui nous semblent les plus pertinentes pour estimer la fréquence du nouvel exercice de souscription. Cela nous permet d'estimer la fréquence moyenne de chaque classe de segmentation.

— **Etape 3 : Ajustement de la fréquence**

Une fois l'historique bien choisi comme vu dans la partie précédente, nous allons définir la méthode d'estimation de la fréquence moyenne de chaque classe de segmentation.

Tout d'abord, la modélisation de la fréquence des sinistres consiste à ajuster une loi de Poisson à la fréquence de sinistres ( $\frac{\text{nb des sinistres}}{\text{exposition}}$ ) à l'aide d'une régression de Poisson, afin de pouvoir estimer le nombre de sinistres qui seront ouverts en N+1. Etant donné que nos données présentent une sous-dispersion pour chacune des classes (Variance < Moyenne) validée par un test de sous-dispersion, choisir une régression de poisson nous semblait plus prudent, puisque cela consiste à retenir l'hypothèse variance = moyenne.

La régression de Poisson est un modèle de prédiction qui s'applique lorsque la variable cible Y est une variable de comptage (nombre d'apparitions d'un événement durant un laps de temps).

Nous rappelons que si Y suit une loi de Poisson de paramètre  $\lambda$ , alors :

$$\text{pour tout } y \text{ dans } \mathbf{N} : \mathbf{P}(Y = y) = \frac{\exp(-\lambda)\lambda^y}{y!}$$

avec :

$$\lambda = \mathbf{E}(Y) = \mathbf{V}(Y)$$

La méthode GLM Poisson approxime Y par  $\lambda$ . En effet nous essayons d'expliquer en moyenne la variable Y. La régression consiste à modéliser le paramètre  $\lambda$  à partir de certaines variables  $X_i$  dites explicatives. la modélisation

est la suivante :

$$\ln \lambda = a_0 + a_1 X_1 + \dots + a_n X_n$$

Avec les  $a_i$  des réels que nous cherchons à estimer.

Ainsi nous avons :

$$\mathbf{E}(Y|X) = \mathbf{V}(Y|X) = \exp(a_0 + a_1 X_1 + \dots + a_n X_n)$$

Dans cette partie nous nous intéressons à la fréquence moyenne. Le paramètre de cette loi de Poisson va correspondre à la fréquence moyenne que l'on cherche à estimer. Afin de calculer cette fréquence moyenne pour l'exercice N+1, nous avons deux possibilités :

- Si nous constatons une stabilité de la fréquence sur l'historique retenu et en faisant l'hypothèse selon laquelle, la nouvelle année de souscription se comportera de la même manière que les années les plus récentes : dans ce cas, nous retenons une fréquence moyenne, pour l'exercice 2019, estimée à partir de la moyenne observée sur l'historique retenu. C'est-à-dire la fréquence simulée en 2019 sera donc égale à la moyenne des fréquences observées sur l'historique retenu.

- Si nous constatons une tendance à la hausse (ou à la baisse) sur l'historique retenu : dans ce cas, nous retenons une fréquence moyenne, pour l'exercice 2019, estimée à partir d'une régression linéaire. Nous allons chercher en premier les paramètres de la régression linéaire, ensuite nous en déduisons la fréquence simulée pour l'année d'exercice N+1.

Nous verrons plus tard dans ce mémoire, comment se modélisent ces deux méthodes.

#### — Etape 4 : Validation des hypothèses retenues pour chaque classe de segmentation

Une fois que les travaux concernant la modélisation de la fréquence ont été effectués, nous réalisons des tests statistiques pour valider les hypothèses retenues pour chaque classe de segmentation (par exemple un Test de Khi 2 et un test de sur/sous-dispersion qui seront présentées plus tard dans ce mémoire).

#### 2.2.4 Modélisation du coût moyen par sinistre

Le calcul du coût moyen est indépendant de la modélisation des fréquences des sinistres. En effet, la partie coût est reprise des calculs stochastiques réalisés pour l'estimation du BE dans le modèle BE sophistiqué, utilisé pour la Formule Standard. Le modèle BE sophistiqué est un modèle stochastique d'estimation du BE RC France faisant partie intégrante du MIP (Modèle Interne Partiel) de SHAM. Le coût moyen par sinistre est obtenu par simulation du modèle BE sophistiqué.

Le modèle BE sophistiqué de SHAM est stochastique et trop complexe pour être recalibré pour l'estimation de la nouvelle année de souscription. Dans son calcul, il n'y a pas de segmentation des classes de risques de la Responsabilité Civile Médicale contrairement à la partie concernant la modélisation de la fréquence des sinistres pour le risque de primes.

De plus, les calculs sont réalisés de manière agrégée pour les sinistres attritionnels et sinistre par sinistre pour les sinistres graves et IBNER. La réassurance s'applique sinistre par sinistre pour les traités XS et est combinée pour les QP et Stop-Loss.

Ainsi, le coût moyen utilisé pour le risque de primes provient du modèle BE sophistiqué en le déduisant à partir du nombre de sinistres total et du BE estimé net de réassurance sur les exercices inférieurs à l'année N (exercice de souscription en cours), toutes classes de risques confondus. Nous faisons l'hypothèse selon laquelle le

coût moyen par sinistre obtenu sur un historique récent peut être utilisé pour estimer la charge de sinistre du nouvel exercice de souscription.

### 2.2.5 Problématique étudiée

Comme vu précédemment l'estimation de la charge de sinistres N+1 repose sur une méthode fréquence x coût. Mais dans le cadre de cette étude nous nous limiterons à la partie modélisation de la fréquence.

Dans le modèle d'origine validé par l'ACPR, SHAM a décomposé son portefeuille RC médicale, France en 5 classes de contrats :

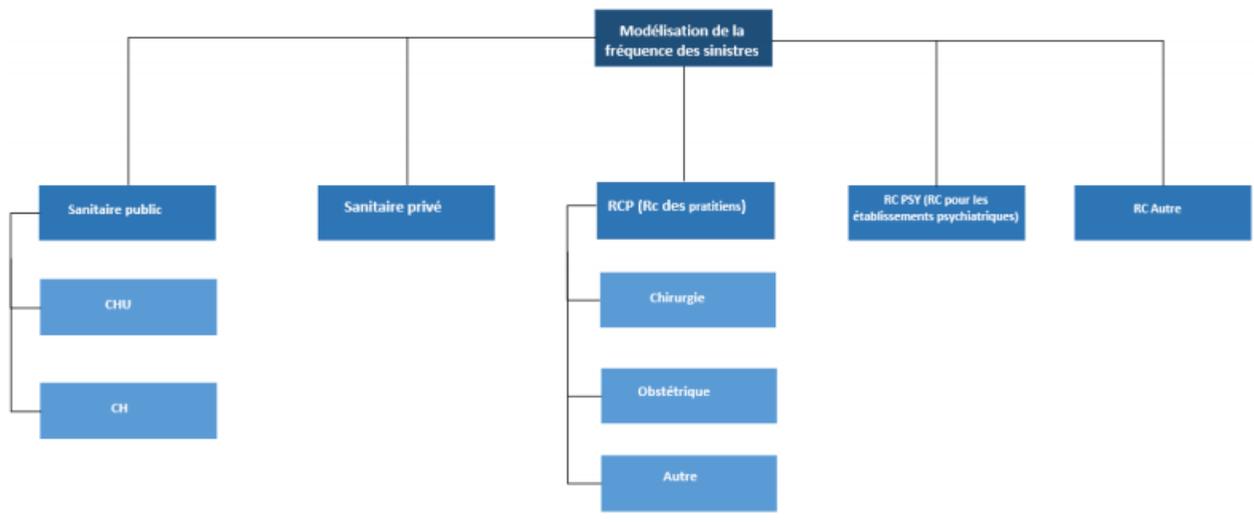


FIGURE 5 – Segmentation du portefeuille RC médicale France SHAM

#### Sanitaires publics :

Dans ce groupe, nous avons les établissements publics à risque MCO (Médecine, chirurgie et obstétrique)/ Médecine. En raison d'évolution de fréquences différentes il a fallu séparer les établissements CHU(Centre Hospitalier Universitaire) et CH (Centre Hospitalier).

#### Sanitaires privés

Dans ce groupe, nous avons les établissements privés MCO/Médecine. Ce segment comprend en particulier les cliniques ou groupements de cliniques.

#### RCP (Responsabilité Civile des praticiens)

Il s'agit de la RC des praticiens qui ont une activité libérale. Dans ce groupe nous distinguons trois types de spécialités : Chirurgie, Obstétrique, Autre.

#### RC PSY (Responsabilité Civile des établissements psychiatriques)

Les contrats constituant ce groupe correspondent aux contrats RC des établissements psychiatriques.

#### RC Autre

Les contrats RC autre regroupent tous les établissements ne pouvant être affectés dans les groupes précédents. Ce groupe est non homogène, est constitué de différents types de contrats et d'établissements. Il commence à prendre

de l'ampleur depuis quelques années. En effet ces contrats ont fortement augmenté depuis la création du Modèle Interne Partiel (MIP), ils représentent aujourd'hui près de 10% des cotisations totales du portefeuille RC France de SHAM contre 6% au moment du dossier de candidature.

Comme nous pouvons le voir sur le graphique suivant les nombres de contrats et de sinistres de la classe RC autre ont fortement augmenté ces dernières années :

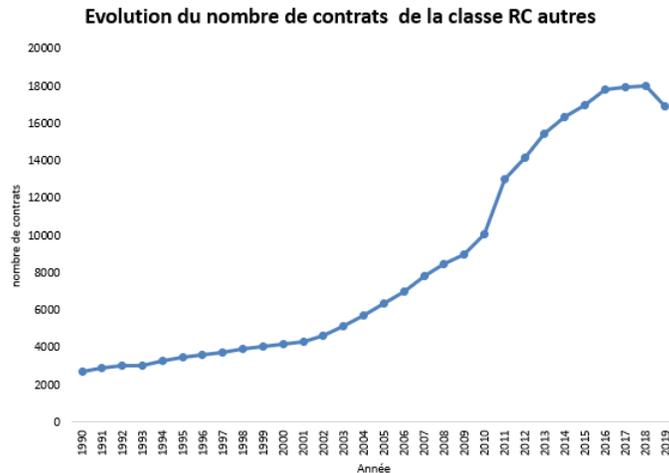


FIGURE 6 – Evolution du nombre de contrats de la classe RC autre

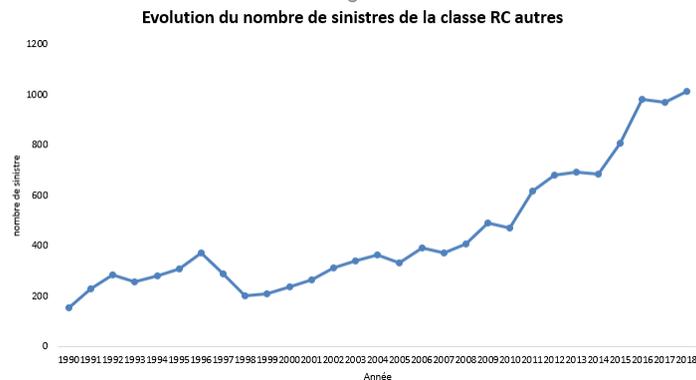


FIGURE 7 – Evolution du nombre de sinistres de la classe RC autre

Nous pouvons constater l'importance de cette classe aujourd'hui. De plus, le backtesting réalisé entre le nombre de sinistres prédits par le modèle et le nombre de sinistres réellement observés (courbe rouge) n'est pas très satisfaisant sur les dernières années. En effet, comme nous pouvons le voir sur le graphique suivant :

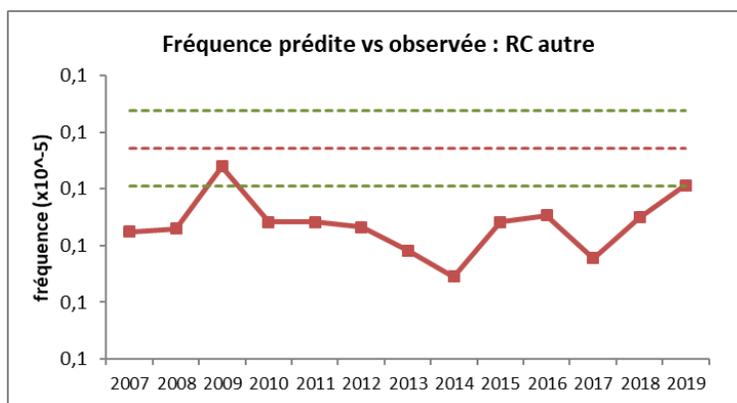


FIGURE 8 – Backtesting de la classe RC autre

Au 31/12/18, le modèle a sur-estimé la fréquence moyenne des sinistres de la classe RC autre pour l'exercice de souscription 2019.

D'où la nécessité de se poser la question suivante :

***Une segmentation plus fine de la classe RC autre permettrait-elle d'améliorer la modélisation de cette classe de risque ?***

Il s'agit ici de vérifier s'il est pertinent de segmenter la classe RC autre en plusieurs sous-segments dans le MIP pour le calcul du risque de primes. Pour cela, nous allons chercher une segmentation pertinente pour les contrats RC autre, modéliser leur fréquence, étudier l'impact de la segmentation sur le risque de primes et vérifier la stabilité de la segmentation retenue et regarder l'impact en matière de backtesting. De plus, des analyses complémentaires seront menées afin d'expliquer nos différents modèles, ainsi que des tests permettant de valider ou non le modèle retenu et d'en vérifier sa fiabilité.

Afin de répondre à la problématique posée, nous allons dans un premier temps présenter les données ayant servi à cette étude et effectuer des premières analyses descriptives. Ensuite, nous allons chercher des segmentations potentielles pour la classe RC autres ainsi que des analyses permettant de justifier chacun des modèles choisis. De plus, nous verrons l'impact de chacune de ces segmentations sur le risque de prime, nous étudierons la stabilité de la segmentation retenue à l'issue de cette étude, en calculant le risque de prime aux deux dernières clôtures 31/12/2018 et 31/12/2019. Et enfin, nous essayerons de challenger la variable d'exposition de la segmentation choisie.

Le schéma suivant résume l'ensemble des travaux menés lors de cette étude.

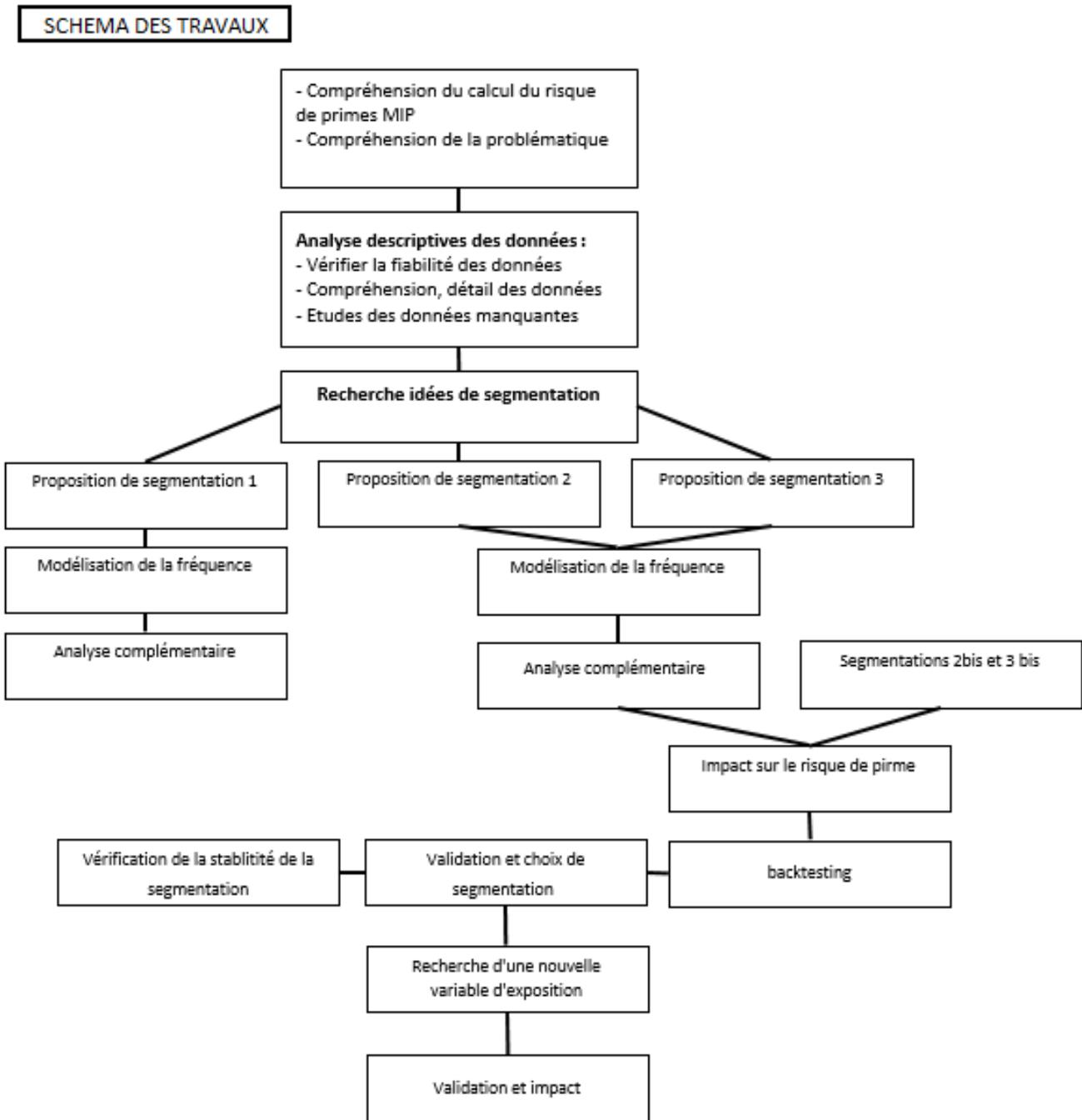


FIGURE 9 – Synthèse de l'étude

### 3 Analyse et exploration des données

Cette étude s'effectuera sur les données ayant servi à la clôture du 31/12/18. Une étude complémentaire sera ensuite effectuée sur celles utilisées au 31/12/2019 afin de vérifier la cohérence et la stabilité du modèle que nous retiendrons au final.

Dans cette partie nous allons dans un premier temps présenter les données utilisées dans le cadre de cette étude. Nous procéderons ensuite à une première analyse descriptive pour voir plus en détail ce que contiennent nos données et les premières segmentations que nous pouvons réaliser.

#### 3.1 Présentation des données

Dans cette partie, nous présenterons le système de gestion des données de SHAM, mais aussi la récupération ainsi que la fiabilité des données. Effectivement, afin d'exploiter un quelconque résultat, il faut s'assurer de la confiance et de la cohérence des données utilisées.

Les données utilisées sont disponibles sous SAS sous forme de tables mises en place par la Direction des Systèmes d'Information et par le Département technique et actuariat.

Sham dispose d'un système de gestion interne des données relatives aux assurés. Cet outil permet de connaître à la fois les différentes caractéristiques du contrat (date de début, situation du contrat, date de résiliation, informations générales, garanties associées) mais aussi l'ensemble des données des établissements clients (données générales, données spécifiques et les sinistres relatifs aux contrats).

Les données sont mises à jour tous les mois, et ces tables permettent d'avoir toutes les informations concernant les différents contrats ainsi que les informations sur l'assuré et la sinistralité. Comme le portefeuille RC autre a augmenté considérablement ces dernières années, l'objet de ce mémoire est donc de rechercher si une nouvelle segmentation de la classe RC autre sera pertinente dans le calcul du risque de primes. Nous allons donc comparer les résultats de cette étude à ceux obtenus à la clôture du 31/12/2018 afin de voir l'impact sur le risque de primes et de justifier le résultat observé.

Afin d'effectuer cette étude les mêmes données ayant servi à la clôture 31/12/2018 seront utilisées dans ce mémoire.

La première phase de cette étude consiste à recréer la base de données. Pour cela, nous utilisons les tables suivantes :

- *Bddprod20190212* : cette table contient les données ayant servi pour le calcul du SCR primes 31/12/18 avant notre étude. Elle regroupe aussi les données caractéristiques concernant les différents contrats, par exemple : leur situation actuelle, le budget pour certains types d'établissement comme les CH et CHU (Centres Hospitaliers, Centres Hospitaliers Universitaires).

- *bddsinsin\_ret\_crci\_connexe\_forfait* : Nous avons aussi les informations concernant les sinistres, exemple : date de survenance, date de déclaration, le nombre de sinistres...

- *tb\_synth\_mmataass\_rc* : contient l'ensemble des informations concernant les matières assurées des établissements ayant un contrat RC chez Sham.

Dans ces tables nous pouvons trouver les données comme :

Le numéro de client, la police du contrat, les dates de début du contrat et de fin du contrat, les informations sur la situation du contrat qui prends en compte les différentes modifications. Nous y trouvons aussi, les informations

sur le type d'établissement, leurs activités, leur statut juridique...

A partir des tables citées ci-dessus, nous allons créer nos bases d'études. Nous procédons ensuite à la nouvelle segmentation du portefeuille. Dans cette étude, seule la classe RC autre nous intéresse. Nous récupérons donc les bases de données correspondantes. Une fois nos bases récupérées nous procédons ensuite à leur analyse.

Dans la classe RC autre, nous avons une grande diversité de contrats, et une grande variété de types d'établissements. Une étude préliminaire est donc nécessaire pour connaître la constitution de nos données.

Nous pouvons voir cette grande diversité dans le tableau ci-dessous :

Type contrat	Branche / contrat	Date de fin
RC transfusion collectif	04/01	1997
RC GTAM	04/02	2004
RC dirigeants / RC dirigeants avec questionnaire	4/20 - 21	
RC pollution	04/50	
RC assurance obligatoire	04/51	
RC hospitalière	4/70-71-73-90-93 Non MCO/Med	
RC CCAS	04/75	
RC recherches biomédicales	04/88	
RC incapables majeurs	04/89	
RC vaccinations	04/95	2007
RC dispensaires	04/96	2000
RC spécifique avec / sans questionnaire	4/97-98	
RC élèves	04/99	
Contrat global : Clinique adaptassur (non MCO / non médecine)	20/01	
RC ESPIC Sham'Assur (non MCO / non médecine)	20/02	
Contrat global[1] : Soins de suite et de réadaptation	20/04	
Contrat global : Etablissement d'hébergement des personnes âgées	20/11	
Contrat global : global médico-social	20/12	
Contrat global : global Sham Adom	20/13	
Contrat global : offres ESMS	20/14	
Contrat global : service départemental d'incendie et de secours	20/15	
Contrat global : Sham Protect' Directeur	20/20	
Contrat global : Partenariat	20/70	

FIGURE 10 – Les branches contrats de la classe RC autre

La colonne branche/contrat du tableau précédent contient une classification spécifique à SHAM permettant d'identifier les différents contrats souscrits. Les contrats de la branche 04 regroupent les contrats RC, les contrats de la branche 20 sont des contrats dits "globaux" qui regroupent à la fois des garanties RC et des garanties multirisques ou autres. Pour ces contrats globaux, nous avons conservé uniquement les garanties RC pour notre étude.

Le but de cette étude est donc d'étudier les différents liens que nous pourrions trouver entre nos contrats afin de les regrouper en différentes classes.

Plus tard, afin de vérifier la stabilité de la segmentation des RC autres, nous allons refaire l'étude pour la clôture du 31/12/2019.

## 3.2 Analyse descriptive des données

Nous allons dans un premier temps vérifier la fiabilité de nos bases de données récupérées. Ensuite, nous procéderons à une analyse descriptive, dans le but d'avoir une vision globale des contrats RC autre. De plus, nous allons présenter les différentes variables pertinentes pouvant servir à la segmentation et procéder à leur analyse.

Comme vu précédemment, nous retrouvons dans la classe RC autre des contrats des branches 4 et 20.

### 3.2.1 Analyse globale

L'analyse descriptive nous permettra de voir plus en détail la répartition des différents types de contrats de la classe RC autre.

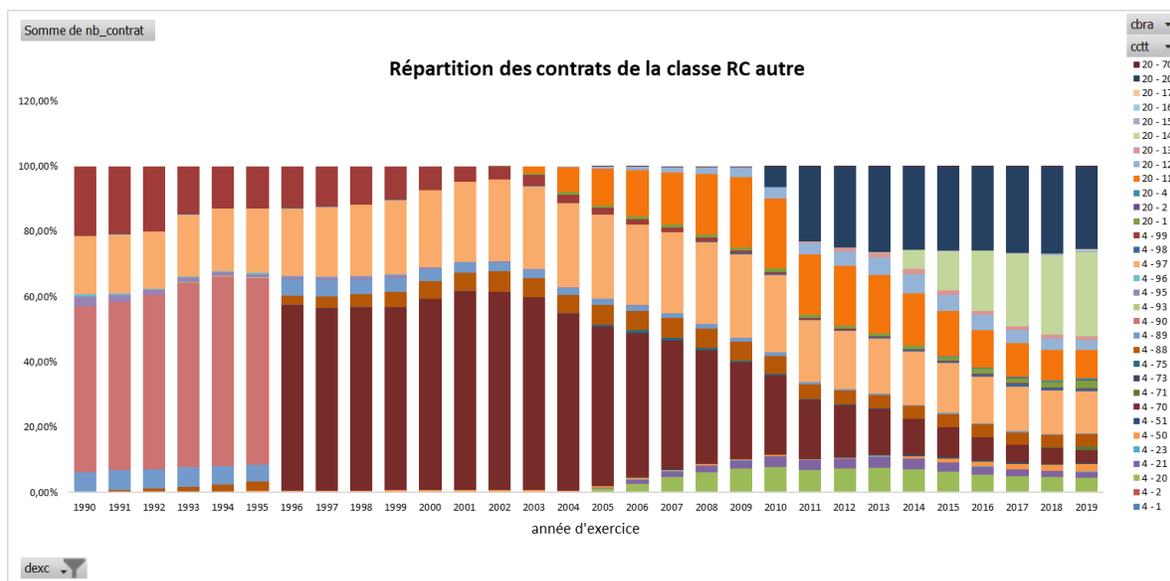


FIGURE 11 – Répartition du nombre de contrats RC autre par type de contrats

Tout d'abord, dans la classe RC autre, nous avons deux branches de contrats, qui sont la branche 4 et la branche 20. Ce graphe permet de voir que la branche 20 commence à prendre de l'ampleur (surfaces bleu foncé et vert clair) alors que la branche 4 a tendance à diminuer. La branche 20 une nouvelle branche qui existe depuis 2003, ce phénomène est dû au fait qu'il y a des migrations de certains contrats de la branche 4 vers la branche 20. Cette migration est à prendre en compte dans la segmentation du portefeuille RC médicale SHAM.

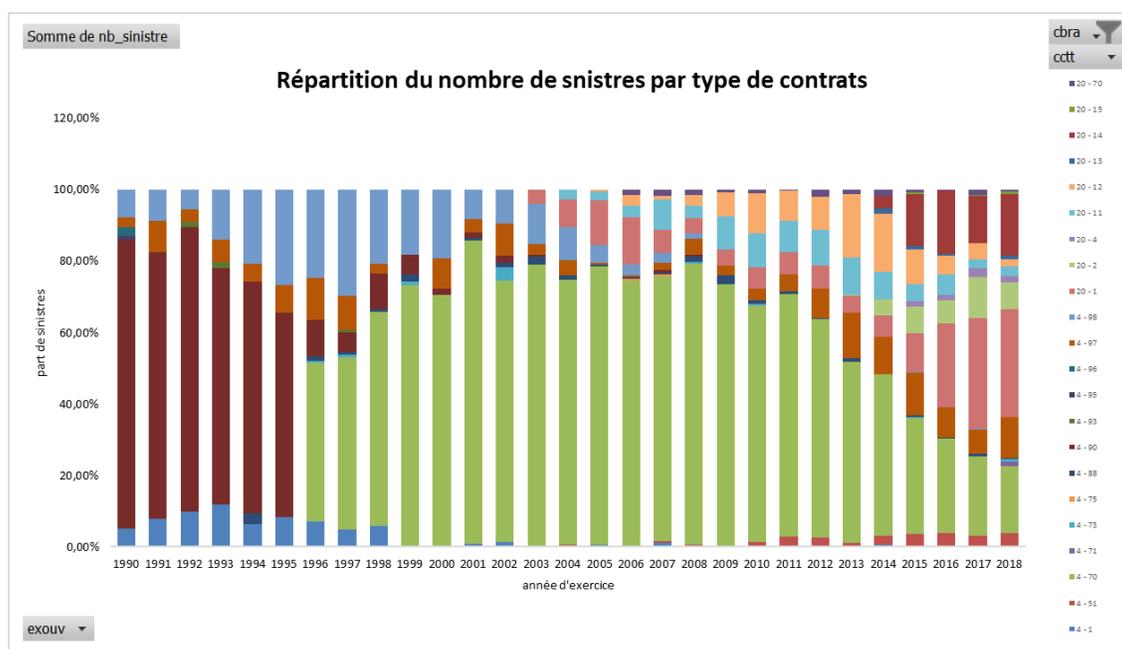


FIGURE 12 – Répartition du nombre de sinistres RC autre par type de contrats

Les deux graphes précédents nous permettent de détecter les contrats qui sont prépondérants dans la classe RC autre. En effet, nous observons que la répartition est assez inégale et l'idée de faire une segmentation par branche n'est pas une bonne idée, car cela ne permettra pas de prendre en compte la migration de certains contrats, comme les 4-70 qui ont migré vers les 20-14.

Nous allons par la suite lister certaines variables qui pourront ou pas nous servir à la segmentation.

Nous disposons de nombreuses variables, mais, pour beaucoup d'entre elles, nous manquons des données pour les utiliser afin d'effectuer la segmentation. Nous en avons sélectionné quelques unes présentées sur le tableau ci-dessous :

Etudes des données entre 2016-2019		
variables	Part de données manquante	Part de données renseignées
Nombre de contrat	0%	100%
Budget	71,07%	28,93%
Statut juridique	0,07%	99,93%
codes branches	0%	100%
codes catégories	0%	100%

Comme nous pouvons le constater sur le tableau précédent, nous sommes assez limités en matière de variables. Et pour certaines nous avons beaucoup de données manquantes. Nous choisissons donc dans un premier temps de conserver le nombre de contrats comme variable d'exposition (qui est celle utilisée dans le modèle d'origine pour la classe RC autre), afin de rechercher une nouvelle segmentation pour la classe RC autre.

### 3.2.2 Fiabilité des données

L'objectif de cette partie est de décrire les différents contrôles et retraitements effectués sur les données afin d'obtenir une base de travail fiable. Ainsi, nous allons étudier dans un premier temps le problème des données manquantes puis la résolution des anomalies en lien avec les processus de gestion.

À chaque extraction, nous faisons des contrôles afin de vérifier si nos données sont correctement sorties. En

effet, notre comparaison s'appuyant sur les données utilisées à la clôture du 31/12/2018, il est important d'avoir par exemple les mêmes nombres de sinistres et de contrats.

La base de données de SHAM est assez complexe à gérer, les données sont souvent mises à jour, et il peut y avoir des changements dans les codifications de certaines variables. Des retraitements automatisés, voire manuels, sont souvent nécessaires.

L'analyse descriptive nous a permis dans un premier temps d'effectuer des premières propositions de segmentations que nous allons étudier dans cette partie. Nous verrons les segmentations proposées, la raison de leur choix ainsi que l'étude et la modélisation de la fréquence des sinistres. De plus, afin de réaliser l'étude des fréquences des sinistres, nous avons conservé dans un premier temps la variable d'exposition d'origine qui est le nombre de contrats. Cette variable a été choisie par défaut, car à l'origine SHAM ne disposait pas de beaucoup d'informations concernant les contrats RC autre, et la variable d'exposition devant être commune et disponible pour tous les contrats d'une même classe. L'objectif dans la suite du mémoire sera donc de challenger cette variable d'exposition.

## 4 Méthodes de segmentations

Le but de la segmentation dans cette étude consiste à créer des classes de risques homogènes. Nous faisons l'hypothèse que les classes sont indépendantes entre elles. La classification consiste à regrouper  $n$  observations en un certain nombre de groupes ou de classes homogènes. Il existe deux principaux types de classifications :

### 4.1 La classification non supervisée

Cette famille de méthodes est utilisée lorsque l'on ne connaît souvent pas le nombre de groupes qui existent dans la population observée. De plus, nous ignorons le groupe auquel appartient chaque observation de la population. Nous voulons alors classer les observations dans des groupes homogènes à partir de différentes variables. Il existe plusieurs familles de méthodes de classification non supervisée. Les plus communes sont :

- Classification hiérarchique ;
- Classification basée sur une densité ;
- Classification basée sur des modèles statistiques/probabilistes, par exemple un mélange de lois normales.

#### 4.1.1 Classification Ascendante Hiérarchique

Il s'agit de regrouper itérativement les individus, en commençant par le bas (les deux individus les plus proches) et en construisant progressivement un arbre, ou dendrogramme, regroupant finalement tous les individus en une seule classe, à la racine. Ceci suppose de savoir calculer, à chaque étape ou regroupement, la distance entre un individu et un groupe ainsi que celle entre deux groupes. Ceci nécessite donc, pour l'utilisateur de cette méthode, de faire un choix supplémentaire : comment définir la distance entre deux groupes connaissant celles de tous les couples d'individus entre ces deux groupes. Le nombre de classes est déterminé a posteriori, à la vue du dendrogramme ou d'un graphique représentant la décroissance de la hauteur de chaque saut, ou écart de distance, opéré à chaque regroupement.

#### 4.1.2 Classification par ré-allocation dynamique (K-means)

Dans ce cas, le nombre de classes  $k$  est fixé a priori. Ayant initialisé  $k$  centres de classes par tirage aléatoire (ou autres procédures), tous les individus sont affectés à la classe dont le centre est le plus proche au sens de la distance choisie (en principe, euclidienne pour cette méthode). Dans une deuxième étape, l'algorithme calcule des barycentres de ces classes qui deviennent les nouveaux centres. Le procédé (affectation de chaque individu à un centre, détermination des centres) est itéré jusqu'à convergence vers un minimum (local) ou un nombre d'itérations maximum fixé.

### 4.2 La classification supervisée

Cette famille de méthodes est utilisée dans le cas où nous connaissons déjà le nombre de groupes qui existent dans la population observée. De plus, nous connaissons le groupe auquel appartient chaque observation de la population. Le but est alors de classer les observations dans les bons groupes à partir de différentes variables.

Avec cette méthode, nous pourrons ensuite utiliser une règle de classification pour prédire les groupes auxquels appartiennent de nouvelles observations. Cette méthode est utilisée par exemple pour identifier si une transaction bancaire est frauduleuse ou pas ; ou reconnaître des chiffres écrits à la main ; ou aussi identifier le type de cancer dont souffre un patient.

Les méthodes supervisées permettent de voir si les contrats de la classe RC autres peuvent être attribués dans les classes déjà formées.

#### 4.2.1 K plus proches voisins (K-NN)

Comme il s'agit d'une classification supervisée, nous connaissons déjà les différentes classes. L'objectif est donc d'attribuer les nouvelles données à leur classe correspondante.

Cette méthode, comme son nom l'indique, consiste à attribuer à chaque nouveau point la même classe que le plus proche des points appartenant au nuage initial.

Même si cette méthode est facile à implémenter, son inconvénient est son temps d'exécution.

#### 4.2.2 Arbre de décision

Un arbre de décision (decision tree) est une structure très utilisée en classification de formes. Son fonctionnement repose sur des heuristiques construites selon des techniques d'apprentissage supervisé (Mezghani, 2015). C'est un classifieur interprétable représenté sous forme d'arbre tel que :

- Les nœuds de l'arbre testent les attributs ;
- Chaque nœud réalise un test portant sur la valeur d'un attribut dont le résultat indique la branche à suivre dans l'arbre ;
- Il y a une branche pour chaque valeur possible de l'attribut testé ;
- Les feuilles spécifient les catégories (deux ou plus).

Contrairement aux K-NN, les arbres de décision fonctionnent facilement sur les données qualitatives tant que le nombre de caractéristiques n'est pas trop grand.

#### 4.2.3 Réseaux neurones

Un réseau de neurones est une méthode de calcul à base de fonctions mathématiques dont la conception est inspirée du fonctionnement des neurones biologiques. En effet, les réseaux de neurones sont une méthode de l'intelligence artificielle optimisée par des méthodes d'apprentissage de type probabiliste.

## 5 Segmentations alternatives

### 5.1 Première segmentation testée : Clinique/ Etablissement hospitalier/ Autre

À l'issue des premières analyses descriptives, nous avons pu identifier trois classes de contrats dans RC autre, qui semblent chacune présenter un niveau de risque différent qui nécessite une analyse particulière. Les groupes formés sont les suivants :

- Les cliniques non MCO / non Médecine : Il s'agit des établissements pour lesquels les contrats se trouvent dans les 20/01 et 20/03.

- Les établissements hospitaliers (EH) non MCO/non Médecine : Dans cette classe nous retrouvons les établissements hospitaliers non MCO et non Médecine. Les contrats se retrouvent dans les branches 4/70-90-93-73 et 20-14.

- Autre : Nous y retrouvons tous les autres contrats ne pouvant pas être affectés dans les deux classes précédentes.

Pour rappel, les établissements à risque MCO / Médecine sont comptabilisés dans les segments "sanitaire public" et "sanitaire privé". Le segment RC autre comprend les établissements publics et privés uniquement non MCO et non médecine.

Le choix de cette segmentation clinique/établissement hospitalier/autre peut se justifier par l'étude des fréquences. En effet, en observant l'évolution de la fréquence de chaque groupe, nous pouvons constater des évolutions différentes. Par conséquent, il pourrait être pertinent de modéliser la fréquence des sinistres séparément pour ces 3 types d'établissements. Concernant la variable d'exposition, par défaut c'est le nombre de contrats.

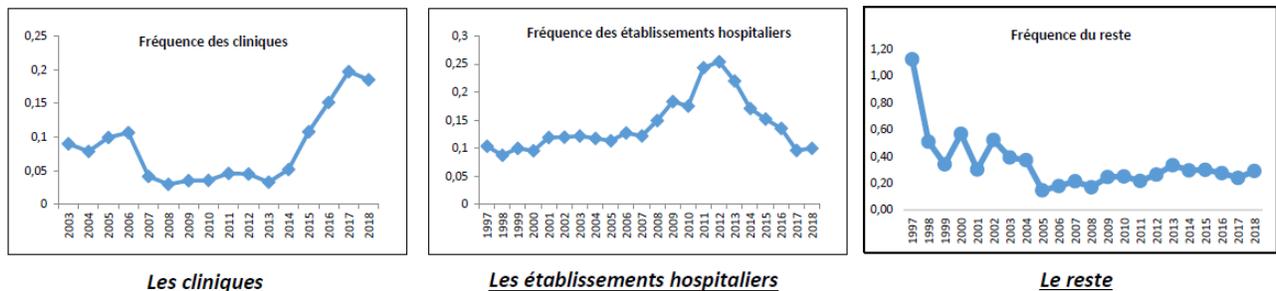


FIGURE 13 – Segmentation 1 : Courbes des fréquences historiques

#### — Clinique :

Une analyse approfondie des cliniques a été nécessaire pour comprendre pourquoi nous observons une forte hausse de la fréquence à partir de 2015. Dans cette analyse, nous avons regardé les catégories de clients qui influençaient le plus la fréquence des sinistres. Pour cela nous avons dans un premier temps comparé parmi les clients toujours présents durant ces quatre dernières années leur nombre de sinistres par rapport à leur nombre de contrats. Ensuite, nous avons calculé la fréquence des sinistres de chaque année pour chacun des clients que nous avons comparé à la fréquence moyenne. La fréquence moyenne est le rapport entre le nombre de sinistres à l'année et le nombre de contrats sur l'année. Cela nous a permis d'identifier les clients qui avaient une grande influence sur le nombre de sinistres.

Après cette étude nous avons regardé ensuite l'évolution de la fréquence avec ou sans les contrats ayant un fort impact sur le nombre de sinistres.

L'une des raisons pour laquelle nous observons cette forte hausse de la fréquence est la présence d'un (ou plusieurs) établissement(s) MCO privés dans ce groupe. Cette présence est due aux manques d'informations fournies lors de la souscription des contrats. En effet, ces établissements auraient dû être classés dans le groupe "Sanitaires privé", mais cela n'a pas été le cas car ils ont été identifiés dans le système de gestion avec une codification ne permettant pas de les identifier comme des établissements MCO. Afin de repérer les MCO dans le groupe RC autre nous avons recherché des variables dans d'autres bases de données permettant de connaître les activités des établissements et de les classer en MCO ou non MCO de manière indirecte. Nous avons pu extraire les MCO qui se trouvaient dans la classe RC autre dont un groupement de cliniques MCO que nous nommerons "groupe MCO X" dans la suite de l'étude. Afin de voir les autres facteurs qui impactent cette forte hausse de la fréquence, une analyse des contrats résiliés en 2014 et de ceux qui sont entrés en 2014 permet de comprendre le phénomène observé comme l'illustre le graphique suivant :

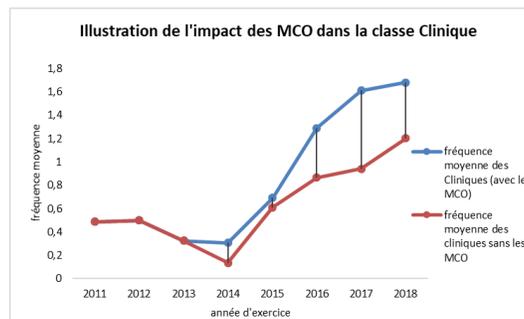


FIGURE 14 – Impact des MCO dans la fréquence des cliniques

Sur le graphique précédent nous constatons que sans les établissements MCO, la fréquence redevient cohérente.

— **Établissement hospitalier (EH) :**

Comme pour les cliniques nous avons effectué les mêmes études pour les établissements hospitaliers dans les RC autres. Nous avons fait l'hypothèse que les établissements hospitaliers étaient regroupés dans les branches/contrats les contrats 4/70-90-93-73 (cf. tableau p22). De plus en 2014 à cause de la migration de certains contrats de la branche 4 vers les 20/14 (dont les 4/70 par exemple), des retraitements étaient nécessaires pour récupérer tous les contrats 4/70 se trouvant dans les 20/14, afin de ne pas perdre des informations.

— **Autre :**

Ce sous-groupe regroupe tous les autres contrats dans le groupe RC autre ne pouvant être affectés ni dans les cliniques ni dans les établissements hospitaliers.

Pour l'analyse de chaque groupe nous avons effectué des contrôles pour vérifier la cohérence des données.

Finalement, cette segmentation ne sera pas retenue dans la suite de cette étude pour les raisons suivantes :

1- Tout d'abord nous avons un problème d'identification : du fait de la migration de certains contrats, certes nous avons réussi à détecter les anciens 4/70 devenus 20/14 par exemple mais nous ne sommes pas en mesure de détecter les nouveaux contrats dans les 20/14 qui sont des 4/70, et il serait difficile de les reclasser dans les classes cliniques et établissements hospitaliers.

2- Ensuite nous observons une mauvaise répartition des contrats dans les différentes classes. En effet nous avons plus de 70% des contrats qui se retrouvent dans la catégorie Autre. Il nous reste peu de contrats dans les catégories cliniques et établissements hospitaliers, comme nous pouvons le constater sur le tableau suivant :

Part de contrats par classe (RC autre)			
année	Clinique	EH	Autre
2010	1,07%	24,41%	74,52%
2011	0,90%	17,98%	81,12%
2012	0,93%	15,80%	83,27%
2013	0,95%	13,97%	85,08%
2014	1,19%	12,53%	86,28%
2015	1,10%	11,91%	86,99%
2016	1,52%	11,88%	86,61%
2017	1,56%	11,18%	87,26%
2018	1,51%	10,59%	87,89%
2019	2,25%	9,88%	87,87%

FIGURE 15 – Segmentation 1 : Répartition des contrats

Nous ne poussons donc pas plus loin l'analyse de la fréquence des différentes classes de cette segmentation, puisque nous avons constaté qu'elle est pas exploitable.

## 5.2 Deuxième segmentation testée : Public/ Privé/ Autre

Dans cette segmentation, nous avons classé les contrats selon le statut juridique des établissements. Les classes obtenues sont les suivantes : Public, Privé et Autre. Nous pouvons noter que cette segmentation coïncide avec celle retenue pour le segment sanitaire (Privé/Public).

Cette classification peut se justifier sur la comparaison des fréquences historiques de chacune des classes. En effet, nous pouvons observer sur le graphique suivant que les niveaux de fréquence des sinistres sont très différents pour chacune des classes.

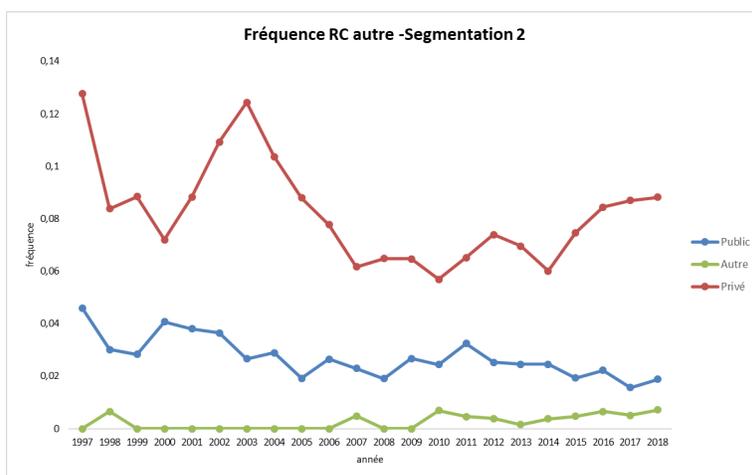


FIGURE 16 – Courbes des fréquences historiques Segmentation 2

Ce graphique permet de montrer la forte sinistralité des établissements privés par rapport aux autres.

Tout d’abord, cette segmentation nous paraît mieux que la précédente. En effet, les classes semblent pertinentes, de plus nous pouvons facilement identifier les différents contrats pour chacune des classes. Enfin, la répartition des contrats est homogène comme nous pouvons le voir sur le tableau ci-dessous :

Répartition du nombre de contrat par classe			
année	Public	Privé	Autre
2011	36%	40%	24%
2012	34%	40%	26%
2013	32%	40%	27%
2014	31%	42%	26%
2015	31%	42%	26%
2016	30%	43%	26%
2017	30%	44%	27%
2018	30%	45%	26%

FIGURE 17 – Segmentation 2 : Répartition des contrats

### 5.2.1 La classe Public :

Dans cette classe, nous retrouvons les établissements publics hospitaliers et territoriaux non MCO, non-médecine. Il est à noter que la majorité des établissements publics sont des établissements hospitaliers.

En analysant en détail cette classe, notre but était d’observer s’il n’y avait pas d’éventuels établissements ou groupes d’établissements qui avaient un impact prépondérant dans la fréquence globale. Sur le tableau suivant,

nous pouvons voir que malgré l'allure de la courbe de fréquence historique, la fréquence moyenne semble stable :

annee	2 011	2 012	2 013	2 014	2 015	2 016	2 017	2 018
Nombre de contrats moyen par client	1,0547	1,0561	1,0707	1,0797	1,0878	1,0898	1,0933	1,0807
Maximum des nombres de contras par client	6	6,0000	6,0000	6,0000	6,2658	8,2596	7,5753	7,0000
fréquence moyenne	0,062527	0,048727	0,047897	0,047972	0,037221	0,042759	0,030198	0,036193

FIGURE 18 – Analyse complémentaire de la classe Public

Le nombre de contrats augmente de plus en plus dans cette classe, et la baisse de la fréquence peut s'expliquer par l'arrivée de contrats générant peu de sinistres.

### 5.2.2 La classe Privé :

Dans cette classe nous y retrouvons les établissements privés à but lucratif et ceux qui sont à but non lucratif non MCO et non médecine.

Tout d'abord l'évolution du nombre de contrats de cette classe est croissante. En effet, la souscription des cliniques dans ce portefeuille a commencé à partir de 2003 et elle s'est fortement développé ces dernières années, de plus l'historique de fréquence est vraiment exploitable à partir de 2007, comme nous pouvons le voir sur le tableau suivant :

annee	nombre de contrats
2003	504
2004	616
2005	750
2006	904
2007	1096
2008	1267
2009	1447
2010	1577
2011	1718
2012	1880
2013	2075
2014	2302
2015	2368
2016	2554
2017	2604
2018	2610

FIGURE 19 – Evolution du nombre de contrats de la classe "Privé"

En 2016 la fréquence augmente fortement : cela est dû à l'arrivée de contrats qui génèrent une forte sinistralité par rapport à la moyenne. En effet, en comparant la fréquence de chaque établissement à la fréquence moyenne par année nous avons pu détecter la présence du "groupe MCO X". Ce dernier a un niveau de fréquence élevé (aux environs de 3,00) par rapport à la fréquence moyenne du groupe privé (aux alentours de 0,15), comme nous pouvons le voir sur ce tableau :

Groupe MCO X			
annee	2016	2017	2018
nombre de contrats	21	26	29
nombre de sinistres	60	88	81
freq moy	2,85	3,43	2,79

FIGURE 20 – Groupe MCO X

En retirant ce groupe MCO dans la classe "Privé" comme nous pouvons le remarquer sur le graphique suivant, nous revenons à des niveaux de fréquences stables.

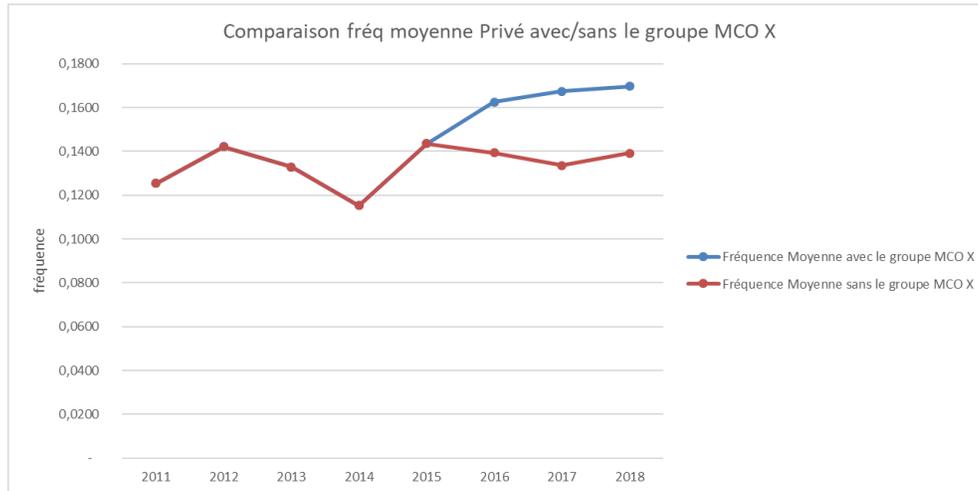


FIGURE 21 – Analyse complémentaire de la classe "Privé"

Dans la suite de ce mémoire, nous effectuerons une segmentation 2 bis en réaffectant le groupe MCO X dans le segment "Sanitaire privé", ce qui nous permettra d'étudier l'impact sur le risque de prime et nous pourrions vérifier s'il y a une amélioration au niveau du backtesting. Cependant, nous devons conserver dans un premier temps ce groupe MCO X dans le segment RC autre, car c'est dans ce segment qu'il a été comptabilisé pour le calcul du risque de primes 31/12/18. Ce n'est qu'à la clôture 31/12/19 que la correction a pu être effectuée dans le calcul du risque de primes. Par conséquent, si nous voulons calculer l'impact marginal de notre nouvelle segmentation au 31/12/18, nous devons le faire à périmètre constant.

### 5.2.3 La classe Autre :

Nous y retrouvons les personnes physiques (par exemple pour les contrats RC des dirigeants) et les établissements dont le statut juridique est non renseigné.

En effectuant les mêmes analyses, nous avons pu constater que dans cette classe, nous avons des contrats spéciaux qui génèrent rarement de sinistres.

annee	nombre de contrat	nombre de sinistre
2003	100	0
2004	95	0
2005	94	0
2006	100	0
2007	105	1
2008	111	0
2009	110	0
2010	223	3
2011	995	9
2012	1183	9
2013	1338	4
2014	1359	10
2015	1415	13
2016	1476	19
2017	1510	15
2018	1428	20

FIGURE 22 – Evolution des nombres de contrats et de sinistres de la classe Autre

En conclusion de cette partie, cette segmentation semble pertinente et exploitable pour modéliser la fréquence du segment RC autre pour l'exercice de souscription N+1.

### 5.3 Troisième segmentation testée : Public/ Privé à but lucratif/ Privé à but non lucratif/ Autre

Cette segmentation dérive de la précédente. Nous décidons ici de séparer les établissements privés en deux. D'une part nous aurons les établissements "Privé à but lucratif" et d'autre part ceux qui sont "Privé à but non lucratif". Le but de cette segmentation est de voir si les deux types d'établissements ont une sinistralité similaire. En effet, nous cherchons à améliorer la segmentation 2, afin de la rendre plus précise.

Le graphique suivant nous permet de comparer l'évolution des fréquences de ces deux classes.

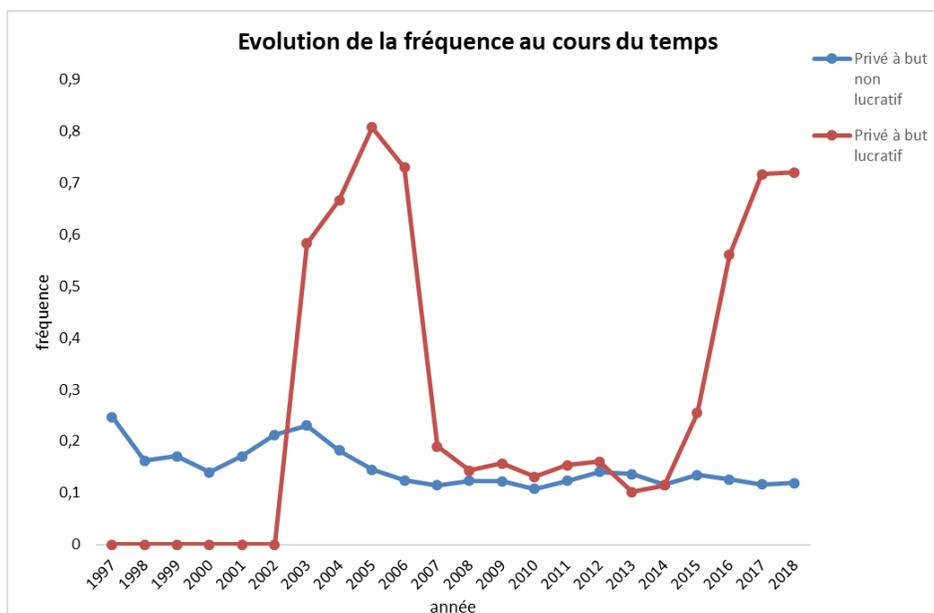


FIGURE 23 – Segmentation 3 : Courbes de fréquences des établissements privés

Nous observons une grande différence entre les deux types d'établissements privés surtout à partir de 2015. Entre 2008 à 2014 les fréquences des deux types d'établissements semblent assez proches et nous avons un décrochage de la fréquence des établissements privés à but lucratif à partir de 2015. Nous constatons que la fréquence pour les établissements privés à but non lucratif se stabilise tandis que pour la classe "privé à but lucratif" nous observons une forte hausse à partir de 2015 et une stabilité entre 2017 et 2018. Ce graphique nous permet de justifier la raison de la séparation de ces deux classes.

#### 5.3.1 La classe Privé à but lucratif

Nous avons comparé la fréquence des établissements par rapport à la fréquence moyenne de cette classe : comme pour la segmentation précédente, le groupe MCO X vient augmenter fortement la fréquence moyenne. Nous avons retiré ce dernier dans cette classe et nous pouvons constater la différence sur le graphique suivant :

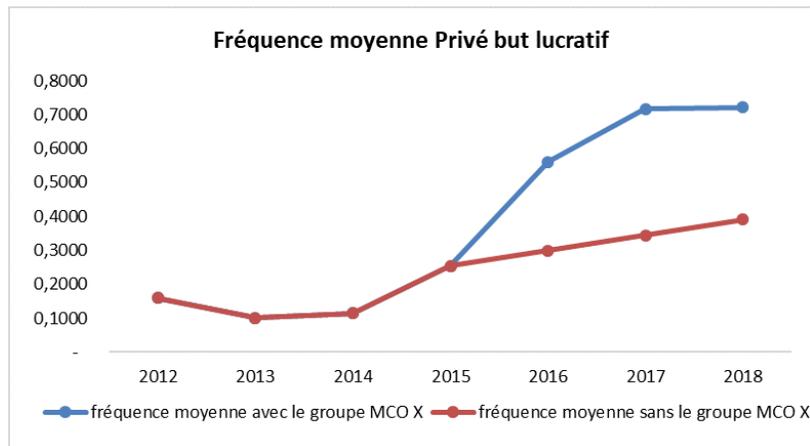


FIGURE 24 – Illustration de l’impact des MCO dans la classe privé à but lucratif

Nous observons une tendance croissante de la fréquence moyenne lorsque nous retirons le groupe MCO X. Cette croissance s’explique par divers facteurs. Tout d’abord, cela s’explique par l’entrée d’établissements ayant une fréquence croissante, et aussi la résiliation de certains contrats qui généraient peu de sinistres.

### 5.3.2 La classe "Privé à but non lucratif" :

L’évolution du nombre de contrats est croissante et nous observons une fréquence stable, ce qui signifie que le nombre de sinistres augmente proportionnellement au nombre de contrats.

Comme pour la segmentation 2, nous ferons dans la suite une segmentation 3bis dans laquelle nous réaffecterons le groupe MCO X, et étudierons l’impact sur le risque de primes.

D’autres segmentations ont été testées mais non retenues car non concluantes. Les segmentations 2 et 3 semblent les plus pertinentes : nous choisissons donc de continuer la suite de l’étude avec ces 2 segmentations. Comme expliqué précédemment, nous réaliserons 2 calculs successifs : un premier calcul en conservant le groupe MCO X dans le segment RC autre et un second calcul en réaffectant le groupe MCO X au segment sanitaire privé.

## 5.4 Segmentations obtenus par méthodes statistiques

Dans cette partie nous allons rechercher des segmentations par méthodes statistiques. Nous appliquerons les méthodes suivantes :

- Classification ascendante hiérarchique (CAH)
- K-Means

### 5.4.1 Traitement des données

Nous allons segmenter les établissements selon leur statut juridique.

La liste des types de "statut juridique" est la suivante :

- NC : Non connu
- personne physique
- prive à but lucratif
- prive à but non lucratif (y compris associatif)
- public hospitalier
- public territorial

Pour ce faire, nous calculons l'évolution annuelle de la fréquence des sinistres de chaque type de statut juridique. Le but est donc de segmenter en regroupant les établissements partageant des niveaux de fréquences similaires.

### 5.4.2 Approche Classification Ascendante Hiérarchique (CAH)

Dans cette partie nous allons chercher des segmentations potentielles par la méthode Classification Ascendante Hiérarchique. Nous testons deux valeurs de K (K=3, K=4).

Nous nous basons sur le critère de Ward. Cette méthode consiste à minimiser la somme des carrés des distances, et repose sur la notion d'inertie inter-classe (entre les classes) et intra-classe (dans une même classe). Le critère de Ward consiste à regrouper les classes de telle sorte que l'inertie inter-classe soit maximale et l'inertie intra-classe soit minimale. Ainsi, ce critère permet aussi d'obtenir une partition plus agrégée. Les formules des deux types d'inerties dépendent de la distance euclidienne. Nous avons les formules suivantes :

Soit  $G = \{e_i : i = \{1 : n\}\}$  un groupe d'individus, de centre de gravité  $g$ , partitionné en  $k$  classes d'effectifs  $n_1, n_2, \dots, n_k$  qu'on appellera  $G_1, G_2, \dots, G_k$  qui ont pour centres de gravité  $g_1, g_2, \dots, g_k$  alors :

$$\text{l'inertie inter-classe est égale à : } I_e = \frac{1}{n} \sum_{i=1}^k n_i \times d(g_i, g)^2$$

$$\text{l'inertie intra-classe est égale à : } I_a = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} d(e_j, g_i)^2$$

Avec  $d(x,y)$ , la distance euclidienne entre deux variables.

Les résultats sont obtenus grâce à l'implémentation sous R (cf annexe).

**Pour  $k=3$  :**

Ainsi les résultats obtenus par CAH pour K=3 sont :

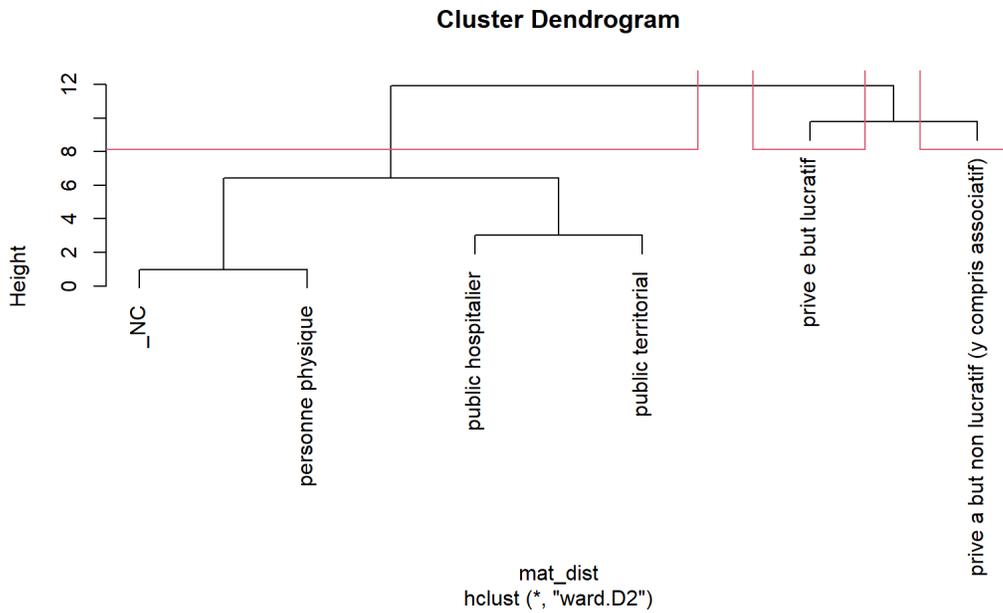


FIGURE 25 – Résultat CAH k=3

Nous observons un découpage de 3 groupes qui sont :

- Privé à but lucratif
- Privé à but non lucratif
- Autre : Public, Personne physique, NC

**Pour  $k=4$  :**

Lorsque nous choisissons  $k=4$ , nous avons les groupes suivants formés :

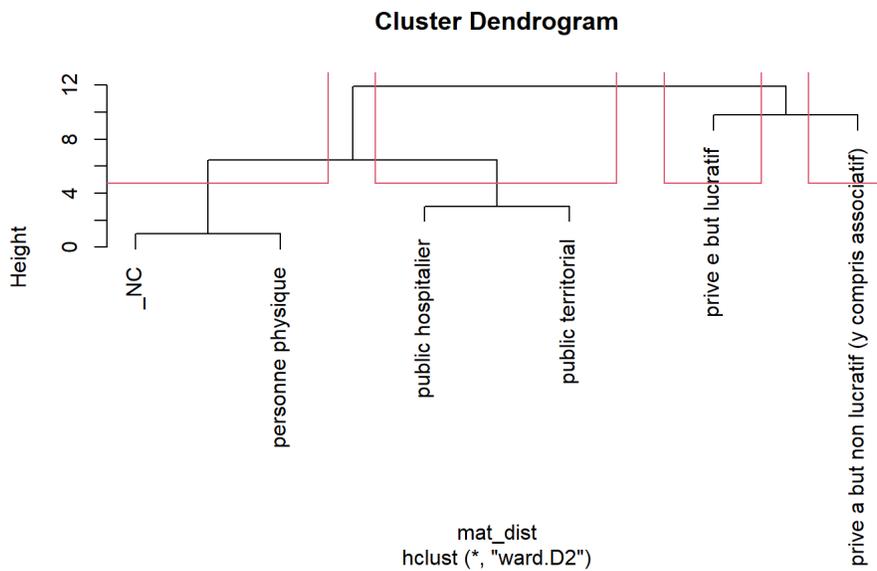


FIGURE 26 – Dendrogramme résultats de la segmentation par CAH  $k=4$

Dans cette figure les groupes d'établissements formés par la méthode CAH sont :

```
> print(sort(groupe.cah))
                _NC
                1
    personne physique
                1
    prive e but lucratif
                2
    prive a but non lucratif (y compris associatif)
                3
                public hospitalier
                4
                public territorial
                4
> |
```

FIGURE 27 – Liste des groupes formés selon la méthode CAH

Nous allons ensuite rechercher les segmentations par la méthode K-means.

### 5.4.3 Approche K-Means

*k=3* :

Les résultats obtenus par la méthode K-means sont :

```
Clustering vector:
                _NC
                2
    personne physique
                2
    prive e but lucratif
                3
    prive a but non lucratif (y compris associatif)
                1
                public hospitalier
                2
                public territorial
                2

Within cluster sum of squares by cluster:
[1] 0.00000 25.87883 0.00000
    (between_SS / total_SS = 82.2 %)

Available components:
[1] "cluster"      "centers"      "totss"      "withinss"    "tot.withinss"
[6] "betweenss"    "size"        "iter"      "ifault"      "
```

FIGURE 28 – Résultats segmentation par K-means k=3

Nous observons les mêmes groupes formés avec la méthode CAH pour K=3. La statistique "between\_SS / total\_SS" mesure à quel point la classification trouvée par la méthode K-means est optimale (bonne), elle doit être proche de 1. Dans notre cas nous avons 82,2% ce qui nous permet de valider le modèle.

*K=4* :

Le résultat obtenu par la méthode k-means avec K=4 est le suivant :

```
Clustering vector:
                _NC
                1
    personne physique
                1
    privé e but lucratif
                3
    privé a but non lucratif (y compris associatif)
                4
                public hospitalier
                2
                public territorial
                2

Within cluster sum of squares by cluster:
[1] 0.4852353 4.5950312 0.0000000 0.0000000
    (between_ss / total_ss = 96.5 %)

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss"
[6] "betweenss"    "size"         "iter"         "ifault"
```

FIGURE 29 – Résultats de la classification par K-means k=4

Ainsi les groupes formés sont les suivants :

- Public
- Privé à but non lucratif
- Privé à but lucratif
- Autre : Personne physique et NC

Il semblerait que les méthodes CAH et K-means coïncident dans cette étude.

En conclusion de cette partie, il semblerait que nous ayons deux segmentations potentielles, qui sont :

- Segmentation A : Privé lucratif, Privé non lucratif, Autre
- Segmentation B : Public, Privé lucratif, Privé non lucratif, Autre

La segmentation B correspond à la troisième segmentation testée dans la partie précédente. Cependant nous avons trouvé une nouvelle segmentation potentielle, "Segmentation A", que nous étudierons dans la suite de ce mémoire.

#### 5.4.4 Quatrième segmentation testée

Dans cette partie nous allons étudier la segmentation suivante :

- Privé à but lucratif
- Privé à but non lucratif
- Autre : Public, Personne Physique, NC

Comme les classes Privés lucratif et Privé non lucratif ont été étudiés dans les autres parties, nous nous focaliserons uniquement sur la classe Autre.

L'évolution de la fréquence des sinistres de cette classe est la suivante :

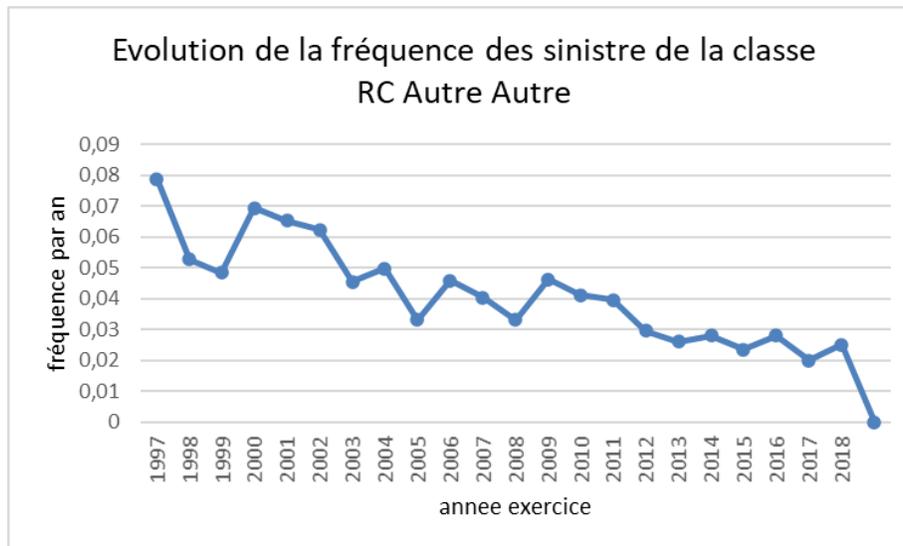


FIGURE 30 – Fréquence de la classe RC Autre Autre

La fréquence a une tendance décroissante. Cette segmentation ne nous semble pas pertinente, car la fréquence des établissements Publics domine sur la fréquence globale. De plus, les classes formées ne sont pas homogènes. En conclusion nous n'intégrera pas cette segmentation dans la suite de cette étude.

## 5.5 Estimation de la fréquence pour l'exercice N+1

La loi choisie pour modéliser nos nombres de sinistres est une loi de Poisson. Cette hypothèse sera vérifiée par des tests de sensibilités. Dans cette première partie, nous avons appliqué la même méthodologie pour les différentes classes de chacune des deux segmentations. Afin d'estimer la fréquence des sinistres pour l'exercice 2019, nous allons utiliser une approche par régression.

### 5.5.1 Modèle de fréquence "méthode moyenne" : Approche par une régression de Poisson

Comme nous l'avons vu dans la partie 2 de ce mémoire, l'estimation de la fréquence des sinistres repose sur une régression de Poisson. Dans ce paragraphe nous cherchons à modéliser la fréquence des sinistres dans le cas où nous observons une stabilité ou nous n'avons pas de phénomènes linéaires. En notant Y la variable modélisant le nombre de sinistres nous savons que :

$$frequency = \frac{Y}{Exposition}$$

L'exposition est supposée connue.

Donc :

$$Y = \text{fréquence} \times Exposition$$

Ainsi, la modélisation de la fréquence par une méthode moyenne, repose sur le principe suivant :

$$\ln(\mathbf{E}(nb_{sinistre})) = \log(Exposition) + \beta$$

Avec  $\beta$  un coefficient à déterminer.

Nous obtenons ainsi :

$$\mathbf{E}(nb_{sinistre}) = \exp(\beta) \times Exposition$$

$$\begin{aligned}\mathbf{E}(Frequency) \times Exposition &= \exp(\beta) \times Exposition \\ \mathbf{E}(Frequency) &= \exp(\beta)\end{aligned}$$

Donc notre paramètre  $\lambda$  à estimer est :

$$\lambda = \exp(\beta) = \mathbf{E}(Frequency)$$

On utilise la méthode des moments pour l'estimation de  $\lambda$ , nous obtenons ainsi la moyenne empirique de la fréquence des sinistres sur l'historique choisis.

Ainsi, en considérons les fréquences observées sur un historique de taille N, nous obtenons :

$$\hat{\lambda} = \frac{1}{N} \sum Frequency_i$$

### 5.5.2 Modèle de fréquence : Approche par une régression linéaire

La régression linéaire suppose un lien de proportionnalité entre la fréquence des sinistres et l'année d'exercice de souscription. Le modèle est :

La fréquence pour la i-ème année d'exercice de souscription est :

$$Frequency_i = a_0 + a_1 \text{annee\_exercice}_i + \epsilon_i$$

Avec :

$a_0$  et  $a_1$  des réels représentant les paramètres que nous cherchons à estimer.

$\text{annee\_exercice}_i$  : l'année d'exercice de souscription i.

$\epsilon_i$  : est une variable aléatoire que nous supposons gaussienne et représente le terme d'erreur.

Hypothèse du modèle :

Les termes d'erreurs  $\epsilon_i$ ,  $i=1, \dots, n$  supposés gaussiens, sont aussi supposés centrés, homoscédastiques et non corrélés entre eux.

H2 :  $\mathbf{E}(\epsilon_i) = 0$ , pour tout  $i=1, \dots, n$

H3 :  $\mathbf{V}(\epsilon_i) = \sigma^2$  pour tout  $i=1, \dots, n$

H4 :  $\mathbf{Cov}(\epsilon_i, \epsilon_j) = 0$  pour  $i \neq j$ .

L'estimation des paramètres  $a_0$  et  $a_1$  se fait par la méthode des moindres carrés ordinaires (MCO) Nous rappelons que la méthode s'appuie sur le principe suivant :

Nous cherchons à minimiser la somme des carrés des résidus de la régression où les résidus sont définis par :

$$\hat{\epsilon}_i = Frequency_i - \hat{Frequency}_i$$

où  $\hat{Frequency}_i$  est l'estimateur de  $Frequency_i$ , obtenu par :

$$\hat{Frequency}_i = \hat{a}_0 + \hat{a}_1 \text{annee\_exercice}_i$$

Ainsi les estimateurs des paramètres du modèle s'obtiennent en résolvant le problème suivant :

$$(\hat{a}_0, \hat{a}_1) = \mathbf{Argmin}(\sum (Frequency_i - (\hat{a}_0 + \hat{a}_1 \text{annee\_exercice}_i))^2)$$

Ce qui nous donne :

$$\hat{a}_1 = \frac{\frac{1}{n} \sum (\text{annee\_exercice}_i - \text{annee\_exercice}_n)(Frequency_i - \hat{Frequency}_n)}{\frac{1}{n} \sum (\text{annee\_exercice}_i - \text{annee\_exercice}_n)^2}$$

$$\hat{a}_0 = \bar{y}_n - \hat{a}_1 \text{annee\_exercice}_n$$

Avec :

$$annee\_ex_n = \frac{1}{n} \sum annee\_exercice_i$$

$$Freq_n = \frac{1}{n} \sum Freq_i$$

Grâce au théorème de Gauss-Markov, nous pouvons montrer facilement que les estimateurs MCO sont à variance minimale parmi la classe des estimateurs linéaires sans biais.

### 5.5.3 Modélisation de la fréquence sur chaque segmentation proposée

Nous allons dans cette partie appliquer les différentes méthodologies vues précédemment, afin d'estimer la fréquence des classes RC Autre de chaque segmentation.

#### 4.5.3.1. Segmentation 2

Pour chacune des classes, nous allons étudier la fréquence des sinistres des années antérieures à 2019, afin de simuler celle de 2019.

— **Public :**

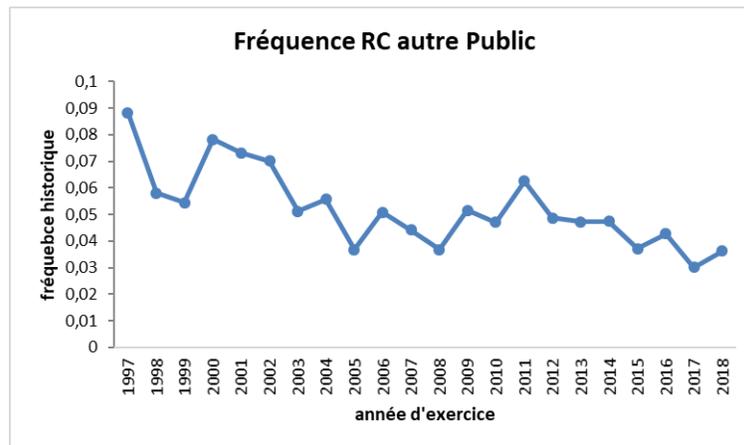


FIGURE 31 – Segmentation 2 : Fréquence historique des établissements Publics (RC autre)

Nous pouvons constater une tendance décroissante, suivie d'une légère stabilité entre 2015 et 2018. Pour modéliser la fréquence moyenne de l'année 2019, nous choisissons l'historique 2015-2018 : en effet nous sommes à peu près sur le même niveau de fréquence. A cause de la forme de la courbe nous pouvons penser à utiliser une régression linéaire mais en observant la tendance de la fréquence de ce groupe, ce choix nous paraît peu prudent : en effet nous risquons de sous-estimer la fréquence en 2019. Nous choisissons donc la méthode moyenne pour plus de prudence.

— **Privé :**

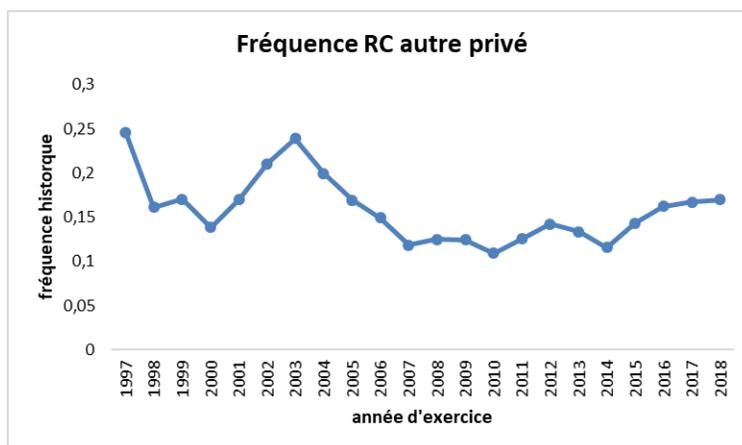


FIGURE 32 – Segmentation 2 : Fréquence historique des établissements Privés (RC autre)

La présence du groupe MCO X fait augmenter considérablement la fréquence de ce groupe, en effet nous observons une rupture de la fréquence entre 2014 et 2016, ensuite une quasi-stabilité entre 2016 et 2018. Ainsi, nous pouvons choisir l'historique 2016-2018. Nous allons estimer la fréquence moyenne pour l'exercice 2019 à partir de la moyenne des fréquences entre 2016 et 2018. En effet nous constatons une stabilisation à partir de 2016 : une étude de sensibilité nous permettra de comparer avec une régression linéaire.

— Autre :

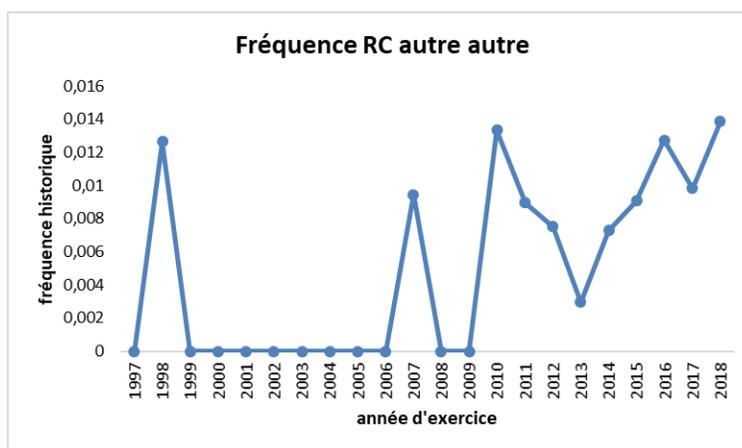


FIGURE 33 – Segmentation 2 : Fréquence historique des établissements Autres

Les fréquences historiques observées pour cette classe sont irrégulières. Cette classe regroupe des contrats qui génèrent peu de sinistres, ce sont en général des contrats spéciaux dont les sinistres sont rares comme par exemple le "Contrat global : Sham Protect' Directeur" couvrant les directeurs des établissements hospitaliers. De plus, nous avons jusqu'à 2010 très peu de contrats. Donc, il suffit de quelques sinistres pour que la fréquence soit élevée. Nous avons donc choisi d'estimer la fréquence moyenne pour l'exercice 2019 à partir de la fréquence en 2018, car cela nous semble plus prudent.

Ainsi pour chacune des classes nous avons estimé la fréquence moyenne pour l'exercice 2019 par la moyenne de l'historique choisi pour chacune. Pour cela nous allons effectuer pour chaque classe une régression de Poisson afin d'obtenir le paramètre Lambda correspondant à la fréquence moyenne simulée pour 2019.

Nous avons obtenus les résultats suivants :

Contrats		Années sélectionnées	Loi	Estimation fréquence	Paramètre lambda
RC Autre	Public	2015-2018	Poisson	moyenne	0,036602
	Privé	2016-2018	Poisson	moyenne	0,166547
	Autre	2018-2018	Poisson	moyenne	0,013848

FIGURE 34 – Segementation 2 : Résultats de la modélisation de la fréquence RC autre

Nous allons comparer les fréquences moyennes estimées pour l'exercice 2019 par rapport à nos données historiques. Nous pouvons observer cela sur les graphiques suivants :

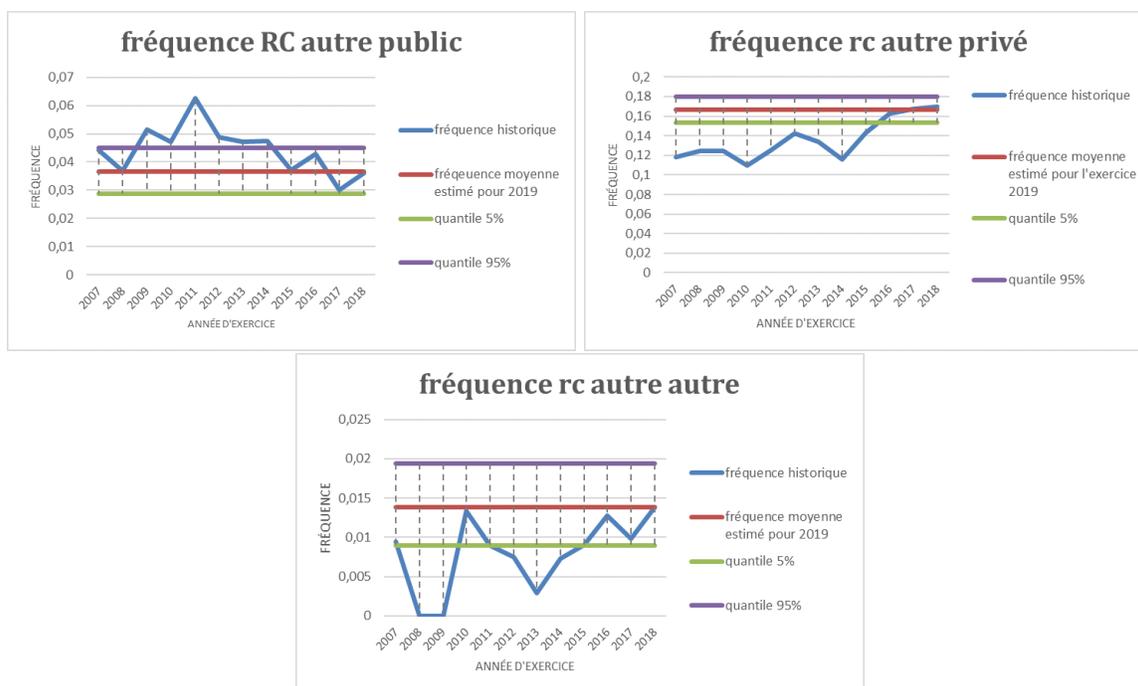


FIGURE 35 – Comparaison fréquences historiques et estimations

#### 4.5.3.2. Segmentation 3

Comme expliqué dans la présentation des différentes segmentations testées, celle-ci est un dérivé de la segmentation 2. Donc la modélisation de la fréquence pour les classes Publiques et Autres est les mêmes que précédemment. Nous allons nous focaliser dans cette partie que celle des classes "Privé à but lucratif" et "privé à but non lucratif".

— **Privé à but lucratif :**

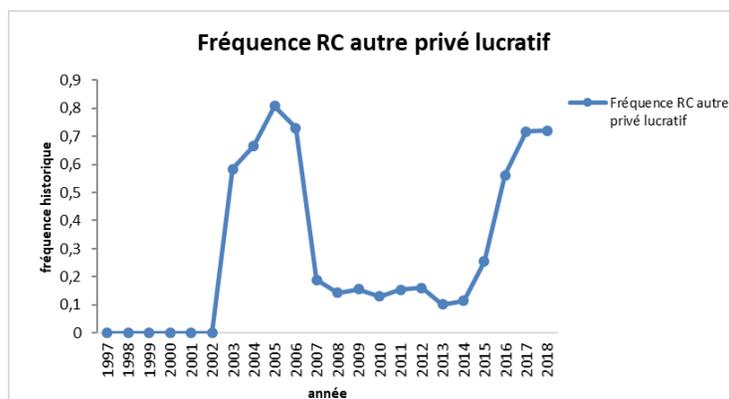


FIGURE 36 – Evolution de la fréquence historique de la classe Privé à but lucratif

Nous observons, malgré la particularité de l'évolution de la fréquence, une stabilité entre 2017 et 2019. Nous estimons donc la fréquence moyenne de cette classe pour l'exercice 2019 par la moyenne de l'historique 2017-2018.

— Privé à but non lucratif :

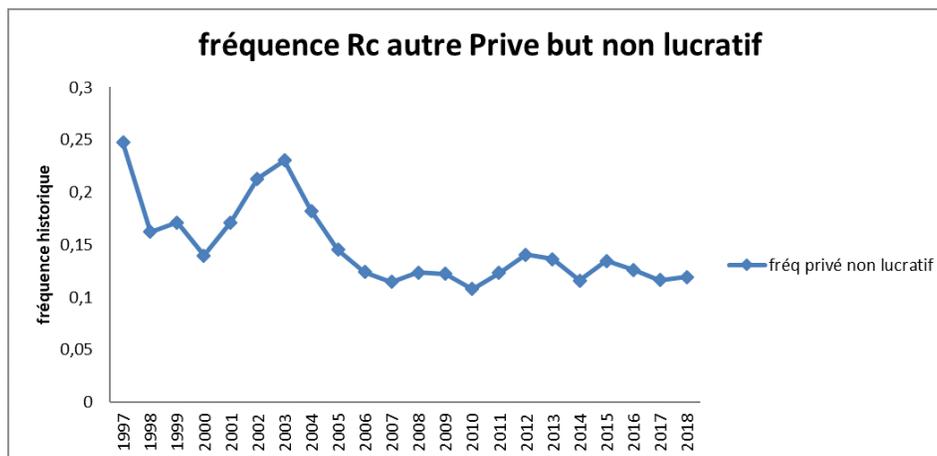


FIGURE 37 – Evolution de la fréquence historique de la classe Privé à but non lucratif

Comme nous constatons une stabilité à partir de 2007, l'estimation de la fréquence moyenne de cette classe pour 2019 se fera à partir de la moyenne de l'historique 2011-2019.

Ainsi, nous obtenons les résultats suivants pour cette segmentation.

Contrats		Années sélectionnées	Loi	Estimation fréquence	Paramètre lambda
<b>RC Autre</b>	<b>Public</b>	<b>2015-2018</b>	<b>Poisson</b>	<b>moyenne</b>	<b>0,036601788</b>
	<b>Privé à but non lucratif</b>	<b>2011-2018</b>	<b>Poisson</b>	<b>moyenne</b>	<b>0,125989814</b>
	<b>Privé à but lucratif</b>	<b>2017-2018</b>	<b>Poisson</b>	<b>moyenne</b>	<b>0,719682966</b>
	<b>Autre</b>	<b>2018-2018</b>	<b>Poisson</b>	<b>moyenne</b>	<b>0,013847736</b>

FIGURE 38 – Segmentation 3 : Résultats de la modélisation de la fréquence

Dans ce tableau nous pouvons remarquer la grande différence entre les fréquences moyennes pour 2019 des classes privés. Ce modèle semble être meilleur que le précédent, car il est plus précis. En effet, nous pouvons observer

sur le tableau ci-dessus la disparité des fréquences de sinistres selon le type d'établissement.

#### 5.5.4 Tests de validation des modèles

Dans cette partie nous cherchons à valider nos modèles. Pour cela nous aurons recours à des tests statistiques que nous présenterons dans un premier temps.

Nous utiliserons un test d'adéquation du Khi 2 pour valider nos modèles.

Dans ce test nous allons vérifier si les nombres de sinistres suivent une loi de Poisson de paramètre la fréquence moyenne simulée pour l'exercice de souscription 2019 multipliée par l'exposition. Pour cela nous simulons les nombres de sinistres sur l'historique choisi et nous effectuons le test de Khi 2. Nous cherchons à valider nos modèles de fréquences en confrontant simulations et observations.

L'hypothèse nulle du test est la suivante :

$H_0$  : Le nombre de sinistres suit une loi de Poisson de paramètre la fréquence moyenne estimée pour 2019 multiplié par l'exposition.

Et la statistique est :

$$\chi^2 = \sum \frac{(n_i - m_i)^2}{m_i}$$

Avec :

$n_i$  : Le nombre réel de sinistres pour l'année i.

$m_i$  : Le nombre de sinistres simulé pour l'année i.

De plus, nous rappelons que la statistique du test suit une loi du Khi 2 à n-1 degrés de liberté (n étant le nombre d'années) sous l'hypothèse nulle. Ainsi l'hypothèse  $H_0$  est validée si la p-valeur est supérieur à 0.05. Nous avons réalisé ce test sur chacune des classes des deux segmentations nous avons obtenus les résultats suivants.

Contrats		Années sélectionnées	Loi	Test Khi 2	p-valeur
RC Autre	Public	2015-2018	Poisson	accepté	0,27541873
	Privé	2016-2018	Poisson	accepté	0,81578815
	Autre	2018-2018	Poisson	accepté	1

FIGURE 39 – Segmentation 2 : validation du modèle

Contrats		Années sélectionnées	Test Khi 2	p-valeur
RC Autre	Public	2015-2018	accepté	0,423
	Privé à but non lucratif	2011-2018	accepté	0,183
	Privé à but lucratif	2017-2018	accepté	1,000
	Autre	2018-2018	accepté	1,000

FIGURE 40 – Segmentation 3 : validation du modèle

Pour chacune des segmentations, les modèles de chaque classe semblent convenir comme nous pouvons le voir sur les résultats obtenus.

### 5.5.5 Deuxième approche : retraitement des MCO

Suite à l'analyse complémentaire nous avons pu expliquer les phénomènes observés sur les courbes de fréquences. Cela est principalement dû à la présence d'un établissement MCO dans le groupe RC autre. Dans cette partie nous cherchons donc à refaire la modélisation de la fréquence pour l'exercice 2019 sur les deux segmentations, sans le groupe MCO X dans la classe RC autre. Nous verrons aussi l'impact du groupe MCO X dans la Classe "Sanitaire Privé".

#### 5.4.5.1 Segmentation 2bis

Dans cette segmentation, seule la classe "RC Autre privé" va subir un changement. Pour les autres classes la modélisation restée inchangée.

Sans le groupe MCO X, la fréquence de la classe Privée se stabilise. Donc pour modéliser la fréquence moyenne pour l'exercice 2019, nous effectuons une régression par moyenne.

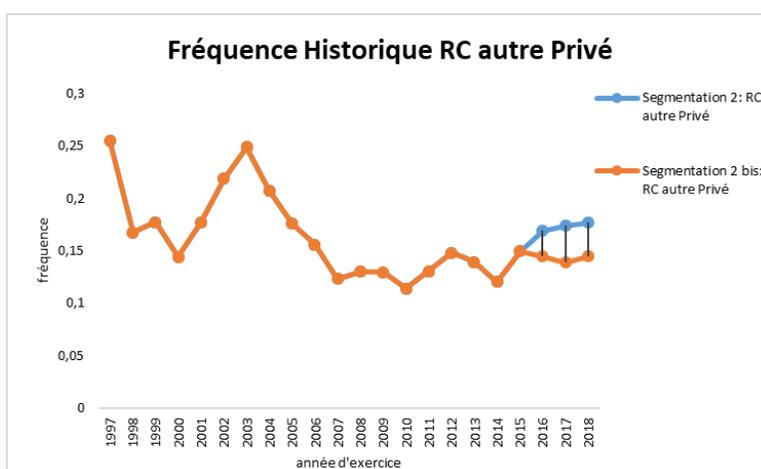


FIGURE 41 – Impact de MCO X dans la fréquence de la classe Privé RC autre

Sur le graphique pouvons voir que la présence du groupe MCO X dans ce groupe a tendance à pousser la fréquence à la hausse.

Contrats		Années sélectionnées	Loi	Estimation fréquence	Paramètre lambda	Test Khi 2	p-valeur
RC Autre	Public	2015-2018	Poisson	moyenne	0,036602	accepté	0,423
	Prive	2015-2018	Poisson	moyenne	0,138767	accepté	0,830
	Autre	2018-2018	Poisson	moyenne	0,013848	accepté	1,000

FIGURE 42 – Segmentation 2bis : Résultats de la modélisation de la fréquence RC autre

#### 5.4.5.2 Segmentation 3bis

Dans cette segmentation nous allons retirer le groupe MCO X de la classe privé à but lucratif de la segmentation 3. Les autres classes restent inchangées. En effet, nous avons vu dans la partie 4.4.1 que la présence du groupe MCO X augmentait fortement la fréquence du privé à but lucratif. Nous cherchons donc à estimer la fréquence RC autre Privé à but lucratif sans le groupe MCO X. Tout d'abord, sur la courbe suivante, nous observons une tendance croissante à partir de 2016.

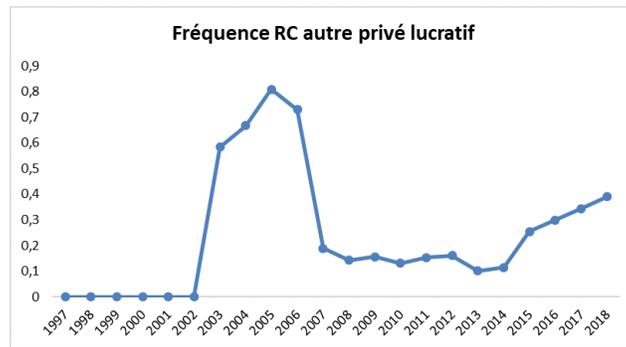


FIGURE 43 – Segmentation 3bis : Fréquence historique de la classe privé lucratif

Nous allons donc effectuer une régression linéaire (définie dans la partie 4.5.2) pour modéliser la fréquence de cette classe.

Afin de valider le modèle nous avons effectué un test de Khi 2 et calculer le coefficient de détermination (défini ci-après).

— Coefficient de détermination  $R^2$  :

C'est l'indicateur principal de la qualité de l'ajustement. Il permet de mesurer l'adéquation du modèle aux données. On le définit de la manière suivante :

$$R^2 = \frac{\text{Variabilité expliquée}}{\text{Variabilité totale}}$$

Le  $R^2$  indique la part de nos observations expliquées par le modèle. Plus le coefficient est proche de 1, plus le modèle choisi est adéquat.

Nous attirons l'attention sur le fait que, l'ensemble des régressions mises en œuvre étant basées sur un petit nombre d'observations, cela constitue une limite à l'interprétation des niveaux de  $R^2$  et de la qualité des régressions qu'on peut en déduire.

— Test de  $\chi^2$  :

Il s'agit du même test décrit dans la partie 4.5.4.

Les résultats obtenus sont :

Contrats		Années sélectionnées	Loi	Estimation fréquence	Paramètre lambda	Test Khi 2	p-valeur	param a	param b	freq simulé
RC Autre	Public	2015-2018	Poisson	moyenne	0,036602	accepté	0,423			<b>0,036602</b>
	Privé non lucratif	2015-2018	Poisson	moyenne	0,125990	accepté	0,183			<b>0,125990</b>
	Privé lucratif	2015-2018	Poisson	reg linéaire	-	accepté	1,000	0,045	-91,4069849	0,43572024
	Autre	2018-2018	Poisson	moyenne	0,013848	accepté	1,000			<b>0,013848</b>

FIGURE 44 – Segmentation 3bis : Résultats de la modélisation de la fréquence RC autre

Nous avons aussi calculé le  $R^2$  qui vaut 0.998 (99,8% de la variance des données sont expliquées par le modèle) ce qui est proche de 1, et le test de Khi2 est aussi acceptée : nous validons donc le modèle.

De plus sur le graphique suivant nous pouvons aussi valider le choix d'une régression linéaire afin d'estimer la fréquence RC autre "Privé à but lucratif" :

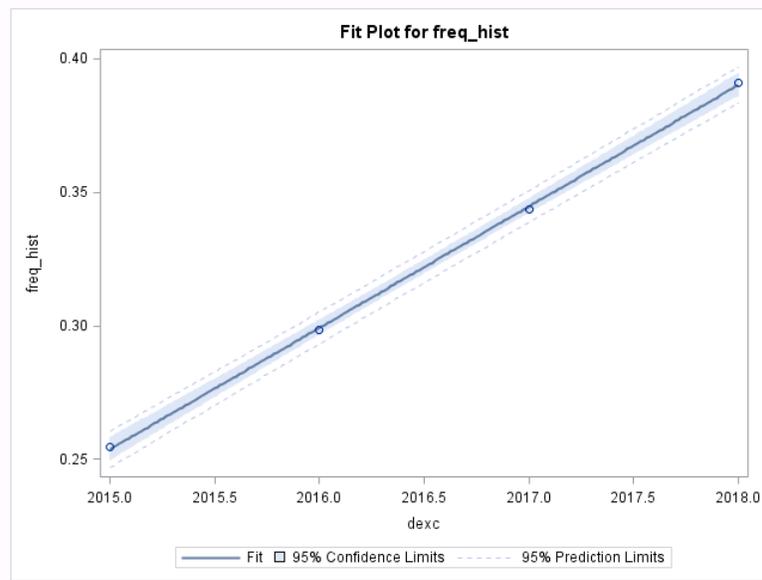


FIGURE 45 – Ajustement par régression linéaire

En effet, le modèle prédit bien et les observations sont confondues.

#### 5.4.5.3 Impact du groupe MCO X sur la fréquence des Sanitaires Privés

Nous avons reclassé le groupe MCO X dans la classe Sanitaire privé. Sa fréquence moyenne est dans le même ordre de grandeur que celle de cette classe, mais l'impact sur la fréquence globale est significatif sur les 2 dernières années, en effet nous avons une légère baisse de la fréquence et une quasi-constance de cette dernière. Nous obtenons les résultats suivants :

Contrats	Années sélectionnées	Loi	Estimation fréquence	Paramètre a	Paramètre b	Test Khi 2	R <sup>2</sup>	fréquence simulée	écart freq
Sanitaire privé MCO/Med (sans ALMAVIVA)	2017-2018	Poisson	régression lin.	0,0550833	-107,77036	accepté	0,999	3,442863741	-0,09176009
Sanitaire privé MCO/Med (avec ALMAVIVA)	2017-2018	Poisson	régression lin.	0,006023382	-8,810104805	accepté	0,998200941	3,351103655	

FIGURE 46 – Impact du groupe MCO X dans la modélisation de la fréquence des Sanitaires Privés

Nous avons des  $R^2$  proches de 1 : nous validons ainsi les modèles utilisés.

Nous en concluons que, même si le groupe MCO X fait baisser la fréquence de la classe Sanitaire privé, nous restons sur les mêmes ordres de grandeur et l'écart est négligeable. En effet nous obtenons 3.35 contre 3.44 (sans MCO X).

## 5.6 Vérification de l'hypothèse d'indépendance des classes RC autre

Dans la méthode du risque de primes par le MIP de SHAM, les classes de segmentations sont supposées indépendantes en termes de fréquence de sinistres. Dans cette nous allons vérifier cette hypothèse sur les nouvelles classes formées pour chaque segmentation du RC autre. Nous utiliserons quelques tests statistiques.

Dans cette partie, nous présenterons uniquement la démarche effectuée pour la segmentation 2. Les résultats de celle de segmentation 3 seront présentés en annexe.

Pour les études menées pour vérifier l'indépendance de nos classes, nous avons utilisé les données de fréquences.

### 5.6.1 Test d'indépendance du Khi 2

Le test d'indépendance du Khi 2 permet de vérifier l'absence de lien entre deux ou plusieurs variables aléatoires. Les variables sont dites indépendantes lorsqu'il n'existe aucun lien statistique entre elles, dit autrement, la connaissance de l'une ne permet en aucune manière de se prononcer sur la connaissance de l'autre. Ainsi pour deux variables aléatoires X et Y, l'hypothèse nulle ( $H_0$ ) de ce test est la suivante : "les deux variables X et Y sont indépendantes". Si la p-value  $\leq 0.05$  on rejette l'hypothèse  $H_0$ .

Nous avons réalisé ce test sur nos données et nous avons obtenu le résultat suivant :

```
> chisq.test(as.matrix(cbind(Public,Prive,Autre)))  
  
Pearson's Chi-squared test  
  
data: as.matrix(cbind(Public, Prive, Autre))  
X-squared = 22.491, df = 42, p-value = 0.9941
```

FIGURE 47 – Résultat du test d'indépendance de khi 2

Nous observons que la p-value est supérieure à 0.05. Nous en déduisons que nos classes sont indépendantes.

### 5.6.2 Rank-Rank plot et Scatter plot

Les graphiques des rank-rank plots et scatter plots nous permettent de visualiser les formes de dépendance entre deux variables. Dans notre étude, nos variables sont nos fréquences historiques par classe. Nous observons :

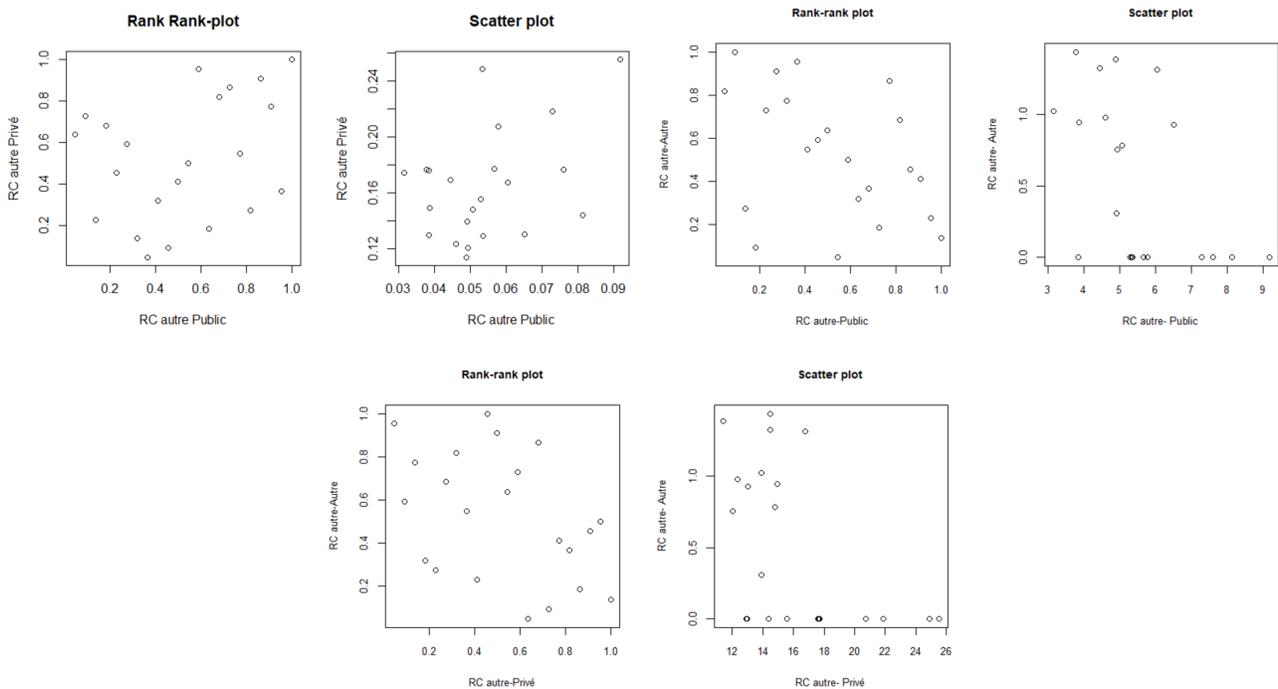


FIGURE 48 – Visualisation des rank-rank plots et des skatter plots

Les points ne sont pas concentrés sur une droite ou dans un coin. Les points sont dispersés : on observe une

indépendance entre les classes.

### 5.6.3 Diagramme de Kendall ou K-plots

La méthode du K-plot est fondée sur les rangs, inspirée par la notion de droite de Henry, et aussi liée à la transformation intégrale de probabilité.

En considérant deux paires de variables aléatoires  $(X_i, Y_i)$ ,  $i = 1, \dots, n$  nous construisons les K-plots de la manière suivante :

Etape 1 : Construction des  $H_i$  définis par :

$$H_i = \frac{1}{n-1} \#\{j \neq i : X_j \leq X_i, Y_j \leq Y_i\}$$

Etape 2 : On ordonne les  $H_i$  définis précédemment. Nous formerons donc des  $H_{(i)}$ .

Etape 3 : Nous traçons les paires  $(W_{i:n}, H_{(i)})$ , avec  $W_{i:n}$  l'espérance de la  $i$ -ème statistique d'ordre dans un échantillon aléatoire de taille  $n$  de  $K_0$ .  $K_0$  correspond à la fonction de répartition de la variable  $H$  définie par :

$$H = C(F(X), F(Y))$$

où  $C(\cdot)$  est la copule de  $X$  et  $Y$  définie selon le théorème de Sklar.

Le K-Plot est fonction des rangs des observations. Le K-Plot est issu de la transformation intégrale de probabilité multi-variée et découle, sur le plan graphique, du principe de la droite de Henry ou "Q-Q-Plot". De ce fait, l'absence d'indépendance se manifeste dans le K-Plot par la présence d'une courbure caractéristique de la copule sous-jacente au modèle. En d'autres termes, si les points sont alignés sur la droite linéaire, cela signifie que nous sommes en situation d'indépendance. Sinon, nos variables sont dépendantes.

Nous avons réalisé ce test sur nos données et nous obtenons les résultats suivants :

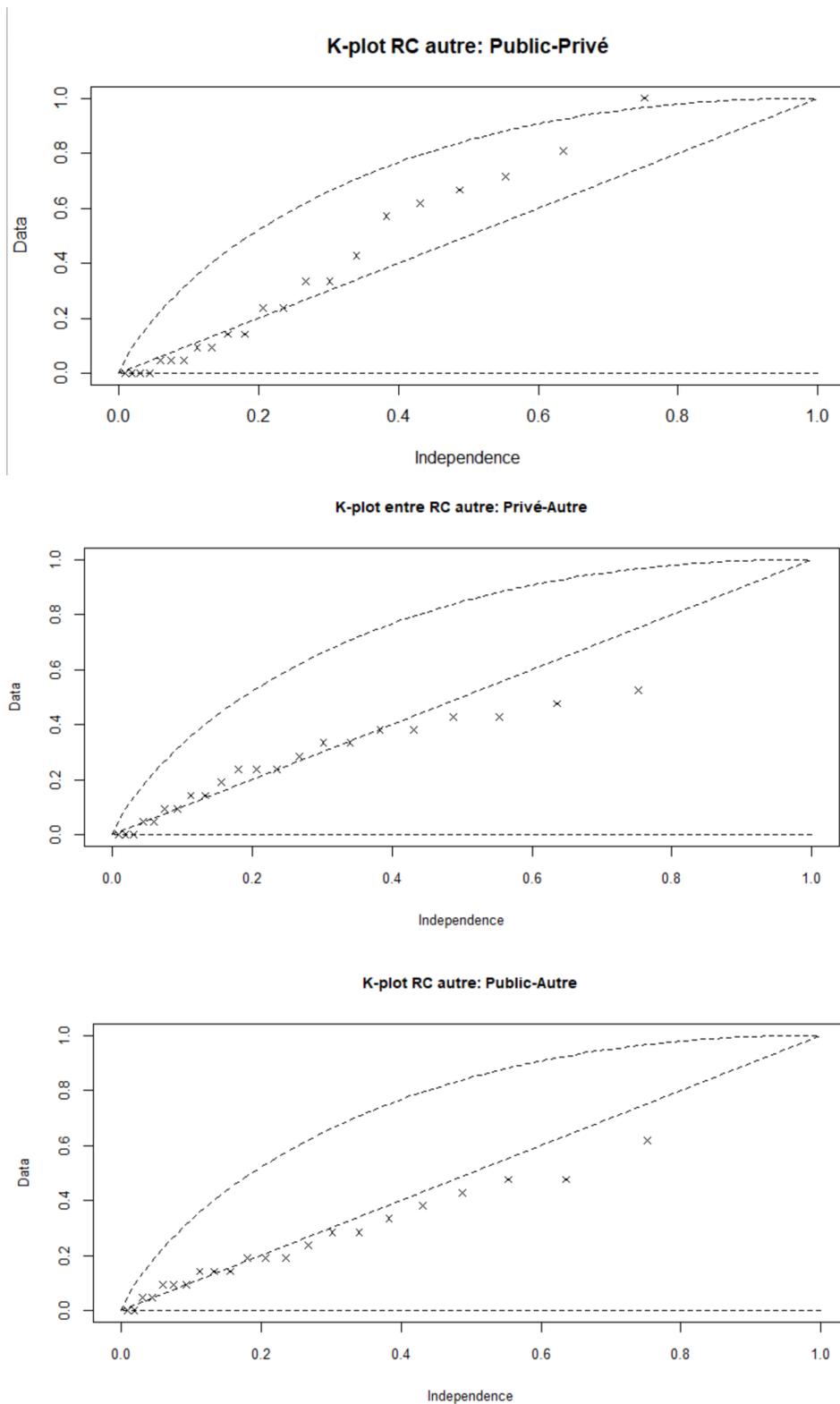


FIGURE 49 – Visualisation des K-plot

Pour chaque graphique, nous observons que les points sont alignés sur la droite. Nous en déduisons que les données sont indépendantes.

Nous avons montré que l'hypothèse d'indépendance entre les classes de contrats est vérifiée. Nous avons obtenu les mêmes conclusions pour la segmentation 3 (cf Annexe).

## 6 Vérification de la pertinence de la nouvelle segmentation dans le calcul du SCR primes

Nous allons maintenant essayer de choisir entre nos segmentations testées. Pour cela nous mesurerons l'impact sur le risque de primes et aussi nous ferons un backtesting entre les segmentations proposées et le modèle utilisé à la clôture du 31/12/2018.

### 6.1 Simulation du nombre de sinistres pour l'exercice de souscription N+1

Pour nos différentes segmentations nous avons choisi de retenir pour valeur d'exposition le nombre de contrats. Afin d'estimer l'exposition pour l'année 2019, nous allons sommer les nombres de contrats entrés toujours présents en 2019 sans tenir en compte de tous les futurs rentrés en 2019. Nous faisons des retraitements sur la variable nombre de contrats afin d'obtenir l'exposition. Ainsi :

$$\text{Exposition} = \text{nombre de contrats}$$

#### 6.1.1 Algorithme de simulation du nombre de sinistres

Nous savons que la fréquence est le rapport entre le nombre de sinistres et l'exposition. Suivant, la méthodologie choisie, nous avons estimé pour chacune des classes des segmentations la fréquence moyenne de l'exercice de souscription 2019.

Nous avons choisi une loi de poisson pour modéliser le nombre de sinistres. Ainsi le paramètre de la loi des sinistres est :

$$\hat{\lambda} = \text{freq\_moy}_{(n+1)} \cdot \text{Exposition}_{(n+1)}$$

Une fois le paramètre estimé nous utilisons une méthode d'inversion pour simuler le nombre de sinistres pour l'exercice de souscription 2019. Dans notre étude, cette méthode repose sur l'inversion de la fonction de répartition d'une loi de Poisson de paramètre  $\hat{\lambda}$ .

La méthode d'inversion est l'un des algorithmes les plus utilisés pour simuler des variables aléatoires d'une loi.

Soit  $U$  une variable aléatoire suivant une loi uniforme sur  $[0,1]$ .

Propriété :

Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F(x)=P(X \leq x)$ , dont l'inverse généralisé (appelé quantile) est défini par  $F^{-1}(u) = \inf\{x : F(x) \geq u\}$ . Alors  $F^{-1}(U)$  suit la même loi que  $X$ .

Inversement, si  $F$  est continue, alors  $F(X)$  suit une  $U([0,1])$ .

Nous avons effectué 40000 simulations pour simuler le nombre de sinistres de l'exercice 2019, en utilisant la méthode d'inversion.

Les résultats obtenus de ces simulations sont détaillés dans les annexes.

#### 6.1.2 Résultats obtenus sur nos différentes segmentations

Nous avons les résultats suivants :

Nous constatons que nous simulons plus de sinistres dans la segmentation 3 par rapport au 2. Ensuite, nous observons que les rapports variance/ moyenne du nombre de sinistres simulés sont quasiment égaux à 1.

Modèles	Type d'établissement	Estimation de l'exposition pour 2019	Nombre de sinistres simulé pour 2019	Ecart-type nb sin simulé 2019	Var/moy freq historique	Var/moy nb_sin sim
Segmentation 2	Public	1 559	57	8	0,40%	101,8%
	Privé	2 457	409	20	0,85%	100,2%
	Autre	1 445	20	4	0,59%	100,0%
	<b>Total</b>	<b>5 461</b>	<b>486</b>	-	-	-
Segmentation 3	Public	1 559	57	8	0,40%	101,8%
	Privé à but lucratif	214	154	12	29,26%	100,0%
	Privé à but non lucratif	2 243	282	17	0,99%	101,4%
	Autre	1 445	20	4	0,59%	100,0%
	<b>Total</b>	<b>5 461</b>	<b>513</b>	-	-	-

FIGURE 50 – Nombre de sinistres simulés pour 2019 par classe RC autre (avec MCO X)

Modèles	Type d'établissement	Estimation de l'exposition pour 2019	Nombre de sinistres simulé pour 2019	Ecart-type nb sin simulé 2019	Var/moy freq historique	Var/moy nb_sin sim
Segmentation 2 bis	Public	1 559	57	8	0,40%	101,8%
	Privé	2 420	336	18	0,90%	100,3%
	Autre	1 445	20	4	0,59%	100,0%
	<b>Total</b>	<b>5 424</b>	<b>413</b>	-	-	-
Segmentation 3 bis	Public	1 559	57	8	0,40%	101,8%
	Privé à but lucratif	177	77	9	25,34%	100,0%
	Privé à but non lucratif	2 243	282	17	0,99%	101,4%
	Autre	1 445	20	4	0,59%	100,0%
	<b>Total</b>	<b>5 424</b>	<b>436</b>	-	-	-

FIGURE 51 – Nombre de sinistres simulés pour 2019 par classe RC autre(sans MCO X)

Dans ces deux segmentations nous simulons moins de sinistres par rapport aux segmentations précédentes 2 et 3 : cela est dû à l'absence du groupe MCO X dans la classe RC autre.

## 6.2 Détermination du risque de primes

La méthode retenue pour le calcul du risque de primes repose sur une hypothèse de proportionnalité entre la charge totale des sinistres et le nombre de sinistres corporels. Des tests de linéarité ont été réalisés afin de vérifier cette hypothèse.

Le risque de primes de SHAM se calcule de la façon suivante :

$$SCR\_primes^{31/12/N} = VaR_{99,5}(Charge\_ultime_{nette}^{N+1}) + Frais\_Gestion^{N+1} - Cotisation^{N+1} - BE^{N+1}$$

### — Détermination de la charge ultime pour l'exercice de souscription N+1

Considérons une variable aléatoire X de fonction de répartition F et  $\alpha$  un réel dans  $[0; 1]$ . Alors la Value-at-Risk de niveau de confiance  $\alpha$  est le quantile d'ordre  $(1 - \alpha)$  de la distribution de X, c'est-à-dire que, pour par exemple un échantillon de 10000 valeurs d'une variable aléatoire, pour obtenir la Var à 99.5 nous allons ordonner nos données, et cela va correspondre à la 9950-ème valeur.

Nous cherchons premièrement à estimer la VaR 99.5 de la charge ultime nette pour l'exercice de souscription N+1. Nous avons obtenu dans la partie précédente des simulations du nombre de sinistres pour l'exercice N+1. Les simulations du coût ultime par sinistre ne sont pas modifiées par rapport au modèle utilisé à la clôture du 31/12/2018.

Une fois le nombre de sinistres simulé pour chaque classe de contrats, nous simulons ensuite la charge ultime correspondante pour chacune. La charge ultime de la i-ème classe de contrats pour l'année d'ouverture N+1 est :

$$Charge\_ultime_{nette,i}^{N+1} = Coût\_Moyen\_par\_sinistre_{nette,i}^{N+1} \times Nb\_sinistres_i^{N+1}$$

$$Charge\_ultime_{nette}^{N+1} = \sum Charge\_ultime_{nette,i}^{N+1}$$

A chaque simulation du coût moyen par sinistre nous faisons correspondre un nombre de sinistres de manière aléatoire. Nous ne détaillerons pas la méthode utilisée pour simuler le coût moyen d'un sinistre : notons que dans la méthodologie nous supposons l'indépendance entre le nombre de sinistres et le coût moyen d'un sinistre.

Enfin nous obtenons 40000 simulations de la charges ultimes. Nous en déduisons ainsi le  $SCR\_primes^{N+1}$ .

### 6.3 Comparaison et Backtesting : impact sur le risque de primes et choix de la segmentation

Nous avons vu dans les parties précédentes des segmentations potentielles pour la classe RC autre. Nous allons dans cette partie les confronter, regarder l'impact sur le risque de primes et déterminer celle qui modélise le mieux la classe RC autre.

#### 6.3.1 Comparaison sans retraitement du groupe MCO X

Nous allons dans un premier temps comparer les nombres moyens de sinistres simulés par rapport aux nombres de sinistres réels obtenus.

Contrats	Type d'établissement	nb de sinistres réelles en 2019			nb de sinistres simulé pour 2019		
		Version clôture	Segmentati on 2	Segmentati on 3	Version clôture	Segmentati on 2	Segmentati on 3
RC Autre	Public		48	48		57	57
	Privé (seg 2)		459			409	
	Privé (seg 3)			183			154
	Privé non lucratif(seg 3)			276			282
	Autre		20	20		20	20
	Total	527	527	527	475	486	513
<b>Total(globa</b>			<b>7 143</b>		<b>7 728</b>	<b>7 739</b>	<b>7 766</b>

FIGURE 52 – Backtesting sur le nombre moyen de sinistres simulé 2019 (avec MCO X)

Sur ce tableau nous pouvons constater que les deux segmentations RC autres améliorent le modèle utilisé à la clôture. En effet, le nombre moyen de sinistres simulés par nos deux segmentations pour la classe RC Autre est plus proche au nombre de sinistres réels observé en 2019, que celui de la clôture. La segmentation 3 nous semble meilleure que la première car elle apparaît plus précise et plus prudente.

Ainsi les résultats du risque de primes obtenus sont :

Risque de prime	Version Clôture (31/12/2018)	Segmentation 2	Segmentation 3
	90 997 768,23	91 377 343,56	91 716 200,52
<b>Ecart risque de prime</b>	-	379 575,33	718 432,30
<b>Variation du risque de prime</b>	-	0,42%	0,79%

FIGURE 53 – Résultats des risques de primes pour 2019 (avec MCO X)

### 6.3.2 Comparaison avec retraitement du groupe MCO X

Contrats	Type d'établissement	nb de sinistres réelles en 2019			nb de sinistres simulé pour 2019		
		Version clôture	Segmentation 2 bis	Segmentation 3 bis	Version clôture	Segmentation 2 bis	Segmentation 3 bis
RC Autre	Public		48	48		57	57
	Privé (seg 2)		356			336	
	Privé (seg 3)			83			77
	Privé non lucratif(seg 3)			273			282
	Autre		20	20		20	20
	Total		424	424	424	475	413
<b>Total (globale)</b>			<b>7 143</b>		<b>7 874</b>	<b>7 812</b>	<b>7 835</b>

FIGURE 54 – Backtesting sur le nombre de sinistres simulé pour 2019 (sans MCO X)

Pour comparer les résultats obtenus avec les segmentations bis avec ceux obtenus à la clôture, nous allons recalculer les résultats de la clôture en enlevant le groupe MCO X et aux résultats réels obtenus.

Pour les segmentations 2bis et 3bis, nous obtenons respectivement 413 et 436 sinistres simulés en moyenne, alors que, avec la segmentation retenue à la clôture, nous avons simulé 474 sinistres en moyenne. Nous observons que le nombre de sinistres simulés par les segmentations 2bis et 3bis sont plus proches de la réalité que celui du modèle de la clôture.

La segmentation 3bis semble plus prudente que la segmentation 2bis, car cela nous évite de sous-estimer le nombre de sinistres.

Risque de prime	Version Clôture	Segmentation 2 bis	Segmentation 3 bis
	90 997 768,23	92 502 929,84	92 732 131,51
Ecart risque de prime	-	1 505 161,62	1 734 363,28
Variation du risque de prime	-	1,654%	1,906%

FIGURE 55 – Résultats des risques de primes pour l'exercice 2019 (sans MCO X)

Nous choisissons les segmentations 3bis, en effet, elles semblent plus prudentes que les autres. Etant données que notre base de comparaison a été effectuée en incluant le groupe MCO X dans RC Autre, nous retiendrons alors dans la suite la segmentation 3.

## 6.4 Projection avec la vision 31/12/2019

Dans le modèle ayant servi pour calculer le risque de primes à la clôture du 31/12/2019, le groupe MCO X a été intégré dans la classe sanitaire privé. Comme nous souhaitons effectuer des comparaisons sur le même périmètre, nous choisissons donc la segmentation 3bis. Les résultats obtenus pour le risque de primes pour l'exercice 2020 sont les suivants :

Risque de prime	Version Clôture	Segmentation 2 bis	Segmentation 3 bis
		81 800 536	81 040 933
Ecart risque de prime	-	759 604	296 471
Variation du risque de prime	-	-0,929%	-0,362%

FIGURE 56 – Résultats des risques de primes pour l'exercice de souscription 2020

Cependant nous ne pouvons pas faire de backtesting sur l'année 2020 car nous n'avons pas tous les données de sinistres à notre disposition. Nous doutons fortement d'avoir un nombre de sinistres assez élevé notamment à cause du COVID 19.

## 6.5 Etude de sensibilité sur l'hypothèse d'indépendance des classes globales

Dans la méthodologie utilisée pour l'évaluation du SCR primes SHAM, nous supposons que toutes les classes de contrats sont indépendantes, c'est-à-dire afin d'obtenir le nombre de sinistres globale pour le risque de primes 2019, nous faisons la somme directe entre les nombres de sinistres simulés des différentes classes. Dans cette partie, nous testons l'agrégation des nombres de sinistres simulés avec une dépendance à 100% entre les classes. La dépendance à 100% signifie que les coefficients de corrélations entre nos classes sont quasiment égaux à 1.

Pour mesurer l'impact d'une dépendance de 100% entre les classes de contrats, nous agrégeons les sinistres simulés pour l'exercice 2019 entre les différentes classes de contrats de la manière suivante :

- les simulations produisant un faible nombre de sinistres pour chacune des classes sont agrégées entre elles.
- les simulations produisant un nombre élevé de sinistres pour chacune des classes sont agrégées entre elles.

Nous avons effectué ce test sur la segmentation 3bis. Le risque de primes calculé pour l'exercice de souscription 2019, à partir d'une hypothèse de dépendance de 100% entre les classes de contrats, s'élève à **91.2M€** (contre **92.7M€** avec une hypothèse d'indépendance).

Les résultats obtenus découlent de la combinaison, entre le nombre de sinistres simulés pour 2019 et le coût individuel 2019, correspondant à la VaR99.5. La VaR99.5 du test de dépendance à 100% (91.2M€) résulte d'un nombre de sinistres 2019 inférieur à celui du scénario central combiné à un coût individuel supérieur à celui du scénario central.

## 7 Recherche d'une variable d'exposition alternative

Suite à l'étape concernant la recherche d'une nouvelle segmentation pour la classe RC Autre nous cherchons maintenant une nouvelle variable d'exposition pour celle retenue. Pour l'étude suivante nous excluons le groupe MCO X.

Avant de commencer nos différentes études nous allons extraire et vérifier la fiabilité de nos données. Nous listerons les potentielles variables d'exposition et nous choisirons les plus pertinentes afin de pouvoir simuler le risque de primes.

### 7.1 Etude des variables potentielles

L'ensemble des variables susceptibles d'être une variable d'exposition est décrit ci-après :

- Le Budget : le montant alloué par l'Etat à l'établissement par contrat et par année sur l'ensemble des matières assurées des établissements à but non lucratif.
- EPRD : le montant alloué par l'Etat à l'établissement par contrat et par année sur l'ensemble des matières assurées des établissements à but non lucratif.
- Chiffre d'Affaires (CA) : qui représente le chiffre d'affaires annuel par contrat et par année sur l'ensemble des matières assurées des établissements à but lucratif.
- Nombre de lits : qui représente le nombre total de lits par contrat et par année de l'établissement client.
- Nombre de Séjours dans l'établissement : Une nouvelle variable qui a été créé pour comptabiliser les entrées

Dans un premier temps, nous allons effectuer une pré-sélection des variables. Pour cela, nous utilisons une méthode GLM dont l'objectif sera de déterminer les variables significatives pour expliquer le nombre de sinistres de chacune des classes.

Le modèle s'écrirait de la manière suivante :

Pour chaque classe de contrats  $i$  ;

Nombre de sinistres de la classe  $i = constante + Budget_i + EPRD_i + CA_i + NombreLits_i + Sejours_i + \dots$

Cette méthode est intéressante et facile à réaliser. Il s'agit de trier nos différentes variables et choisir celle qui peuvent potentiellement être pertinentes pour refléter l'exposition aux risques des établissements "RC Autre".

Malheureusement, contenu du manque de données pour certaines variables, nous ne pouvons pas utiliser cette méthode permettant de sélectionner les variables les plus pertinentes permettant d'expliquer le caractère à risque des établissements de chacune des classes. Etant donné que nous n'avons pas beaucoup de variables disponibles cette méthode ne nous semblait pas utilisable.

Le nombre de lits et le nombre de séjours d'un établissement de santé peuvent constituer des variables pertinentes pour refléter l'exposition au risque de l'établissement. En effet, si l'on considère deux établissements avec un nombre de lits (respectivement un nombre de séjours) très différents, celui avec plus de lits (respectivement plus de séjours) a une probabilité plus élevée d'avoir des sinistres que l'autre. Cependant, dans la classe "RC Autre" il nous manque beaucoup de données et le nombre de lits (respectivement le nombre de séjours) n'est pas souvent renseigné pour les établissements correspondants. De plus, concernant le nombre de séjours, il faudra distinguer deux situations : les séjours ayant une longue durée et ceux avec une courte durée. Selon la situation l'intensité du risque exposé n'est pas le même. Même si elles sont pertinentes, compte tenu de notre situation, utiliser ces deux variables peut potentiellement fausser le modèle de fréquence des sinistres par une mauvaise estimation de celle-ci.

En conclusion les variables qui restent et qui paraissent pertinentes à nos différentes études sont : le nombre de contrats, le Budget, EPRD et le chiffre d'affaires. Comme pour le modèle expliqué dans les parties précédentes, nous avons pris le nombre de contrats comme variables d'exposition. L'objectif dans la suite de cette section est de tester les variables Budget, EPRD et le Chiffre d'affaires.

Le budget est un montant annuel alloué par l'Etat aux établissements de santé publics et aux établissements privés à but non lucratif. En effet, les établissements doivent, en fonction du montant de ce budget, programmer leur activité et sont donc limités dans leurs dépenses. L'EPRD (État de prévision de recettes et de dépenses) a été introduit en 2006 en remplacement du budget. Le schéma précédent est inversé, c'est-à-dire que les établissements doivent désormais déterminer leurs recettes en fonction de leurs activités pour ensuite prévoir leurs dépenses. Au fur et mesure des années, le budget a été remplacé par l'EPRD dans les bases de données de SHAM. Ainsi le Budget et l'EPRD joue le même rôle. Dans la suite nous allons considérer la variable BUDGET égale au budget ou à l'EPRD.

Compte tenu des données disponibles, nous allons modéliser la fréquence de sinistres pour les établissements "RC Autre Public" en utilisant comme variable d'exposition BUDGET, puis la fréquence de sinistres de la classe "RC Autre Privé Non Lucratif" en utilisant aussi le BUDGET comme variable d'exposition.

Une fois nos données extraites et présentées nous procédons ensuite à leur analyse.

## 7.2 Construction de l'indice d'inflation

Le Budget et le Chiffre d'affaires sont naturellement impactés par le phénomène d'inflation. Il est donc nécessaire et important d'utiliser une variable d'exposition isolée de tout effet exogène, notamment du phénomène d'inflation, dans le but d'analyser le profil de la fréquence des sinistres.

Dans cette étude nous nous intéresserons au retraitement de l'inflation impactant le chiffre d'affaires et le BUDGET des établissements RC autres. Pour cela, nous allons construire des indices d'inflation de nos variables BUDGET et du CA à partir des informations présentes dans notre portefeuille.

Pour le calcul du risque de primes 31/12/18 (resp 31/12/2019) nous allons convertir l'exposition des années inférieures à 2018 (resp. en 2019) en euros 2018 (resp. euros 2019) afin de neutraliser l'effet d'inflation.

Dans une première étape, nous allons déterminer l'évolution annuelle du BUDGET et du chiffre d'affaires (CA) total des établissements présents dans notre portefeuille pour chaque année N. Pour cela, nous calculons un taux d'évolution du CA et du BUDGET des établissements présents à la fois l'année N et l'année N-1. En effet, pour observer véritablement l'évolution du BUDGET ou CA total, il est nécessaire de se baser sur les établissements présents durant deux années consécutives. Ainsi le taux d'évolution du Budget et du CA pour chaque année N s'obtient par :

$$\text{Taux Evolution pour l'année N} = \frac{\sum_{i=1}^{Nb\text{etabl}}(\text{Budget}N)}{\sum_{i=1}^{Nb\text{etabl}}(\text{Budget}(N-1))} - 1$$

Avec Nb établ. les établissements présents à la fois en N et N-1. La même formule est valable pour le CA, mais dans la suite seule la variable BUDGET sera étudiée.

Dans une seconde étape, il faut construire un indice reflétant l'inflation présente dans le BUDGET et le CA des établissements présents dans notre portefeuille. Nous calculons les indices en base 100, et elles suivent la relation suivante :

$$\text{Indice Budget N} = \text{Indice Budget (N-1)} \times (1 + \text{Taux d'évol Budget N})$$

Le graphique suivant représente l'évolution de l'indice d'inflation du Budget construite pour la période (2005-2018) :

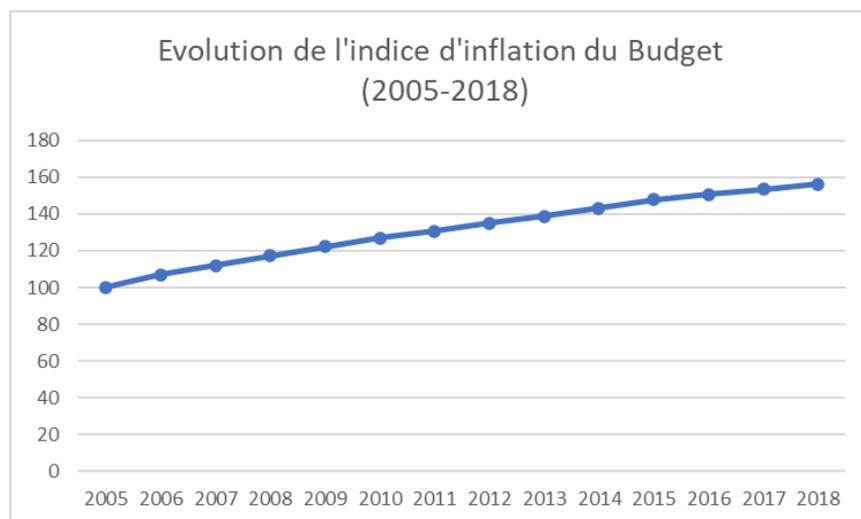


FIGURE 57 – Evolution de l'indice d'inflation du Budget sur (2005-2018)

Sur le graphique ci-dessus nous remarquons que l'indice d'inflation du BUDGET que l'on a construit présente une croissance stable avec une pente qui a tendance à diminuer. Cela est cohérent avec le fait que le BUDGET provient d'une allocation de l'Etat qui souhaite de plus en plus diminuer les dépenses en matière de santé et qui diminue depuis plusieurs années le taux d'évolution des BUDGETS alloués aux établissements de santé.

Le budget des établissements va être recalculé en €2018 pour chacun des exercices de souscription, à partir de l'indice d'évolution des budgets, afin d'exclure l'effet de l'inflation dans le calcul de la fréquence de sinistres. Ainsi, le budget en € 2018 de l'année de souscription N est obtenu de la manière suivante :

$$\text{BUDGET } \text{€ } 2018 \text{ N} = \text{BUDGET}(\text{€ COURANT}) \text{ N} * \frac{\text{Indice Inflation Budget 2018}}{\text{Indice Inflation Budget N}}$$

Pour retrouver le Chiffre d'affaires en € 2018, nous utiliserons la même méthodologie.

### 7.3 Gestion des données manquantes

La problématique de données manquantes est souvent rencontrée par l'actuaire. Avant de commencer les différentes études il faut identifier le type de donnée manquante. Il en existe plusieurs présentés ci-dessous :

- MCAR (missing completely at random) : il s'agit ici d'une donnée manquante de façon complètement aléatoire si la probabilité d'absence est la même pour toutes les observations. Cette probabilité ne dépend donc que de paramètres extérieurs indépendants de cette variable.
- MAR (Missing at random) : ce cas est peu courant. Il arrive lorsque les données ne manquent pas de façon complètement aléatoire ; si la probabilité d'absence est liée à une ou plusieurs autres variables observées, on parle de missingness at random (MAR). Il existe des méthodes statistiques appropriées qui permettent d'éviter de biaiser l'analyse.
- MNAR (Missing not at random) : La donnée est manquante de façon non aléatoire (MNAR) si la probabilité d'absence dépend de la variable en question. Les données MNAR induisent une perte de précision (inhérente à tout cas de données manquantes) mais aussi un biais qui nécessite le recours à une analyse de sensibilité.

Ayant observée beaucoup de données manquantes, nous allons chercher des méthodes pour pallier cela. Enlever les données manquantes n'est pas la solution la plus adéquate car nous perdrons plus d'information pour la suite de notre étude.

Tout d'abord nous allons observé l'évolution de nos données budgets qui comprends aussi l'EPRD.

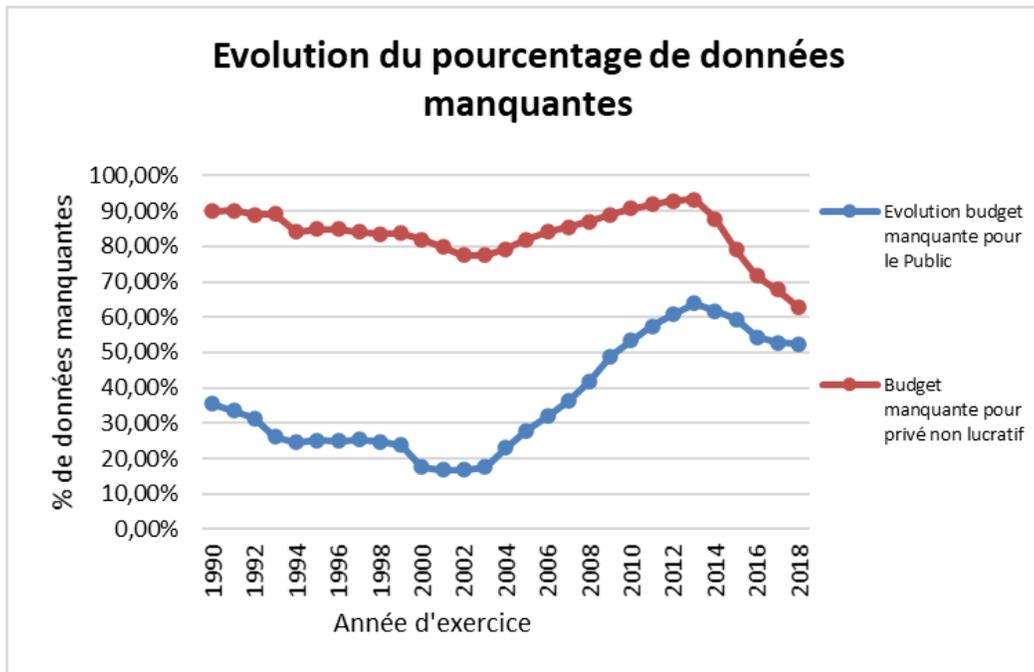


FIGURE 58 – Evolution du pourcentage de données manquantes

Nous remarquons qu'il nous manque beaucoup de budgets manquants pour les établissements, public et privés non lucratifs. Cela est dû à un manque de renseignement des informations dans la base de données principales. Il en est de même pour les établissements Privés Lucratifs et autres, Pour les établissement à but lucratif, le chiffre d'affaires permet de connaître le profil de l'entité, mais malheureusement comme nous pouvons le voir ci-dessus nous sommes confrontés à un nombre élevé de données manquantes. Nous aboutissons à la même conclusion pour la Classe "Autre".

Dans la suite nous nous intéresserons qu'aux classes "RC autre Public" et "RC autre Privé à but non lucratif".

Dans un premier temps nous procédons à la gestion des données manquantes. Pour cela nous avons plusieurs alternatives :

- Attribuer aux établissements correspondants un budget calculé comme la moyenne des budgets connus, la médiane des données observées, ou à partir d'une méthode par simulation.
- Supprimer les établissements correspondants, ou leur attribuer un budget nul.

Tout d'abord, si nous supprimons les établissements pour lesquels les données budgets ne sont pas renseignées nous risquons de perdre beaucoup d'informations dans la suite de cette étude.

Dans la suite nous allons dans un premier temps tester des méthodes permettant de compléter les budgets manquants.

### 7.3.1 Méthode par la moyenne

Comme les classes formées sont supposées homogènes, nous allons supposer que les établissements qui composent chacune des classes ont les mêmes profils de risque, et nous allons combler les données manquantes par la moyenne des données observées non nulles.

Nous comparons sur le graphique suivant les budgets complétés et les budgets réels pour les établissements "RC Autre Public" et "RC Autre Privé à but Non Lucratif".

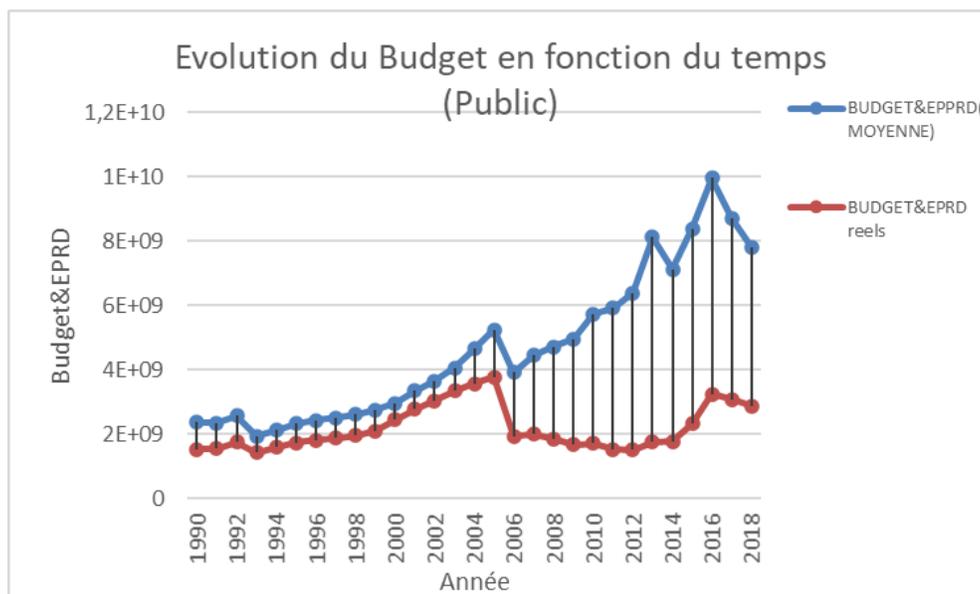


FIGURE 59 – Evolution du budget des établissements Publics au cours du temps

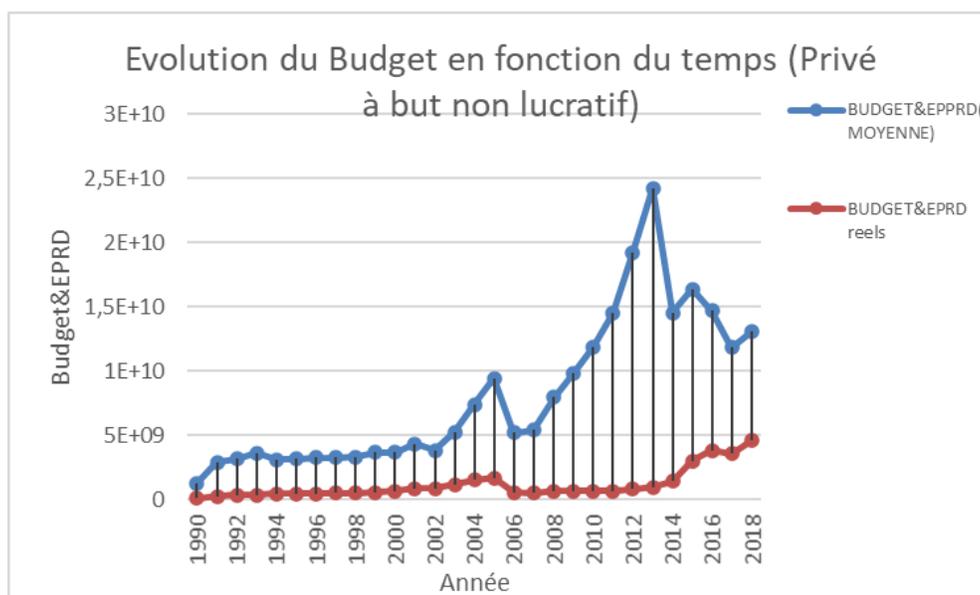


FIGURE 60 – Evolution du budget des établissements Privé à but non Lucratif

Sur les deux graphiques précédents les courbes rouges correspondent aux données réelles disponibles, et les courbes bleues correspondent aux données complétées. La différence des deux courbes se justifie à cause des données manquantes. Pour les établissements "Public", lorsque les données sont complétées, le budget a une évolution plutôt croissante. Il en est de même pour les établissements "Privé à but non lucratif".

Une fois que les données manquantes des budgets complétées nous allons ensuite ramener le budget en € courant 2018.

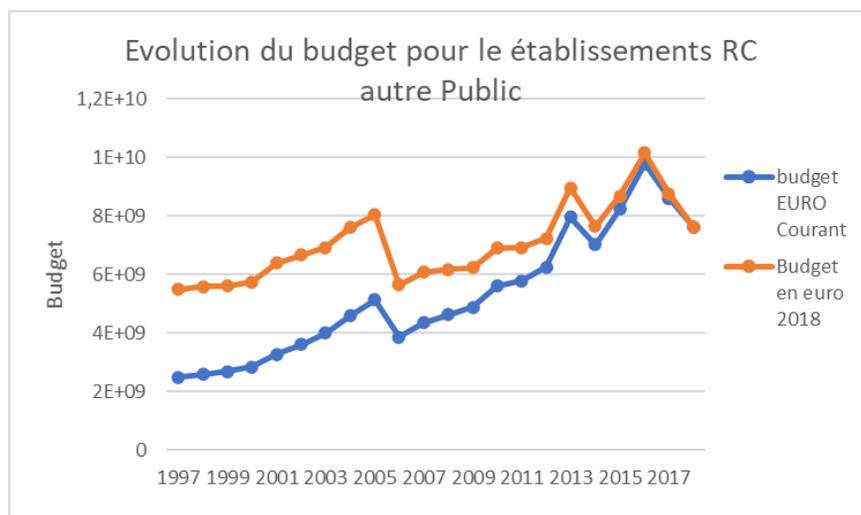


FIGURE 61 – Evolution du budget (inflation) des établissements RC Autre Public

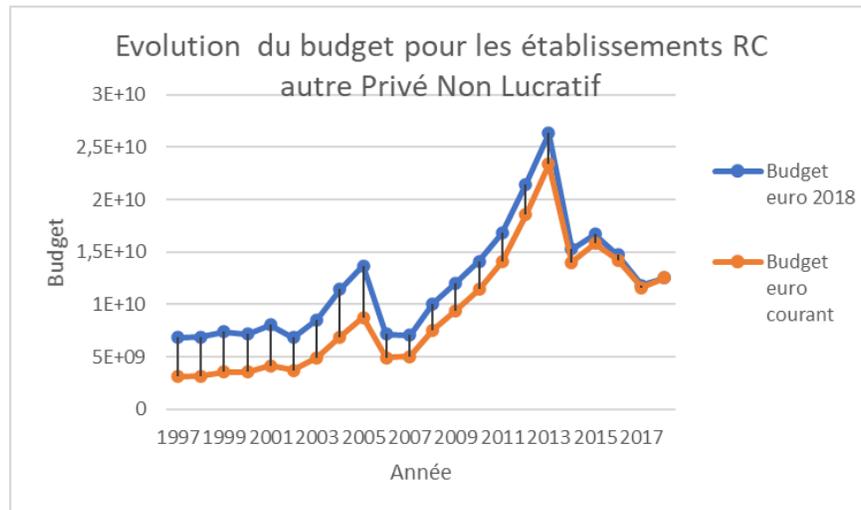


FIGURE 62 – Evolution du budget (inflation) des établissements RC Autre Privé à but non lucratif

### 7.3.2 Méthode par la médiane

L'imputation des données manquantes par la médiane des données observées est la méthode la plus utilisée. Tout d'abord, la médiane d'un ensemble d'observations est une valeur  $x$  qui permet de couper l'ensemble des données en deux parties égales : mettant d'un côté une moitié des valeurs, qui sont toutes inférieures ou égales à  $x$  et de l'autre côté l'autre moitié des valeurs, qui sont toutes supérieures ou égales à  $x$ . Donc choisir la médiane pour compléter les données manquantes semble être judicieux.

Ainsi le budget obtenu après retraitement pour les différents types d'établissements est le suivant :

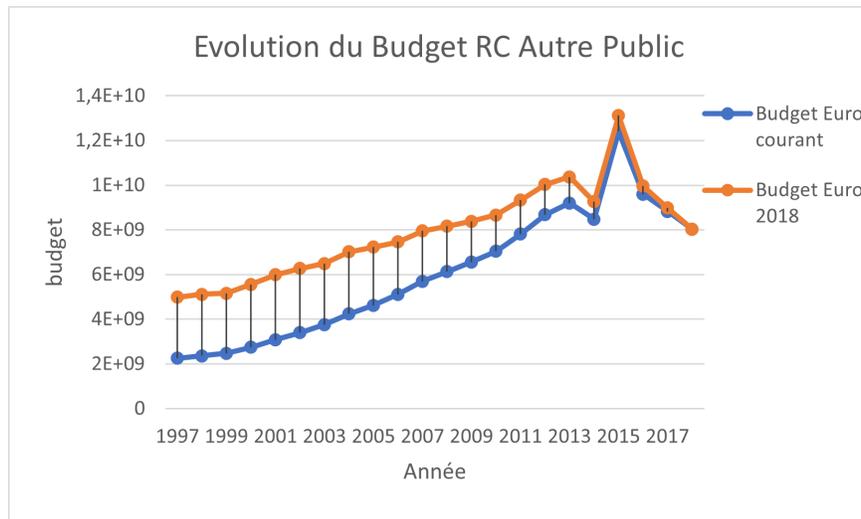


FIGURE 63 – Evolution du budget RC Autre Public après retraitement

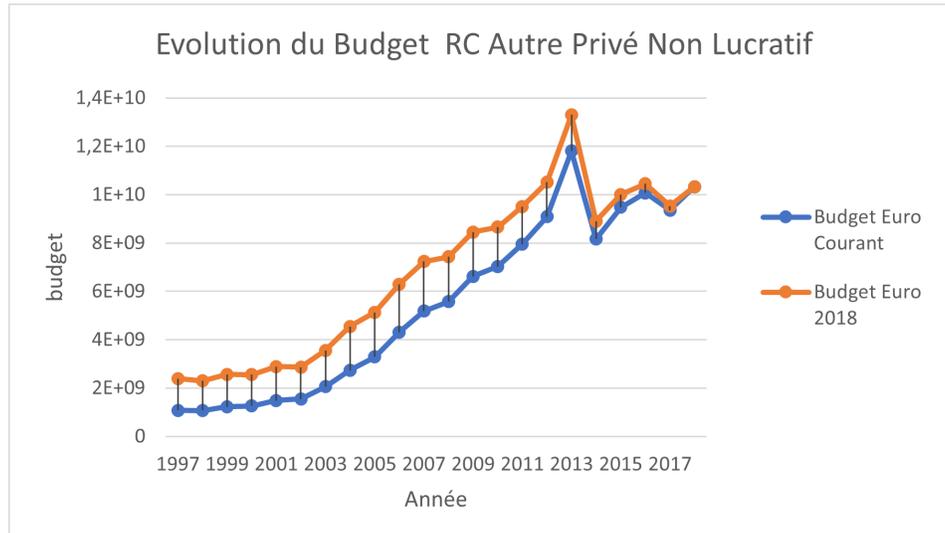


FIGURE 64 – Evolution du budget RC Autre Privé à but Non Lucratif après retraitement

### 7.3.3 Méthode par simulation

Dans cette partie la méthode utilisée pour l'imputation de données manquantes consiste à effectuer des simulations de lois. Tout d'abord, nous allons déterminer la distribution de nos données BUDGET. La distribution sera choisie de telle sorte que les paramètres de loi varient pour chaque année. Cela se traduit par :

Pour chaque année  $N$ , la variable  $BUDGET_N$  suit une loi de fonction de répartition  $F_N$ .

Ainsi, nous allons construire la fonction  $F_N$  par les étapes suivantes :

- Ordonner les données pour chaque année
- Pour chaque donnée  $X_{Ni}$ , nous comptons le nombre de valeurs inférieures strictement à  $X_{Ni}$
- Enfin nous allons calculer pour chaque donnée  $X_{Ni}$ ,  $F_N(X_{Ni})$  qui correspond au rapport entre le nombre de valeurs strictement inférieures à  $X_{Ni}$  et le nombre total des valeurs observés.

Une fois les calculs effectués nous vérifions que nous avons :

$$(F_N(X_{NNbtalissements})) = 1$$

Car en effet, le maximum des données observées pour chaque année est  $X_{NNbtalissements}$ , donc :

$$(F_N(X_{NNbtalissements})) = \sum_{i=1}^{Nb \text{ talissements}} (F_N(X_{Ni}) = 1$$

En appliquant cette méthodologie à nos données il semblerait que pour chaque année N, le BUDGET suit une loi uniforme sur [ Minimum Budget N ; Maximum Budget N ].

Les courbes de la fonction de répartition ont l'allure suivante :

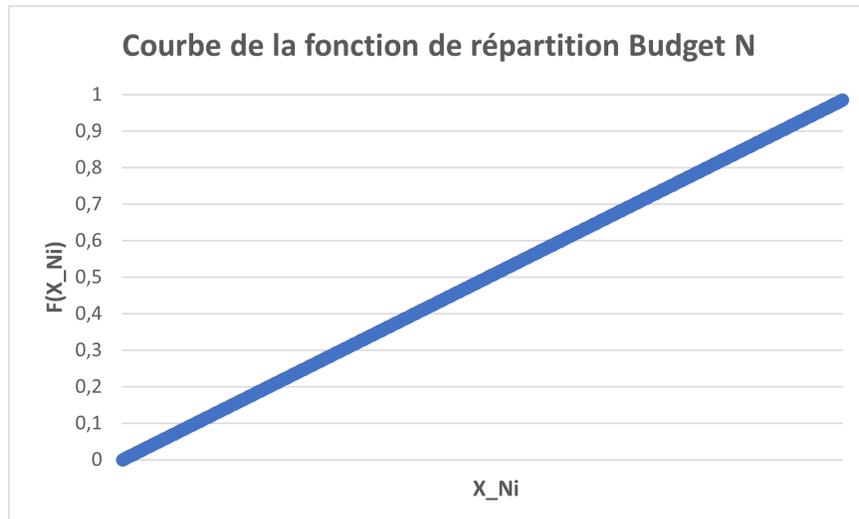


FIGURE 65 – Fonction de répartition du Budget N

La courbe de la fonction de répartition correspond à celle d'une loi uniforme : nous avons effectué un test de Kolmogorov-Smirnov qui nous a permis de valider cette hypothèse.

Ensuite, nous allons simuler pour chaque année 10000 Budgets, nous calculerons la moyenne empirique (Monte Carlo) et nous compléterons les données manquantes par cette moyenne. Enfin, nous calculons le budget après inflation et nous observons les courbes suivantes :

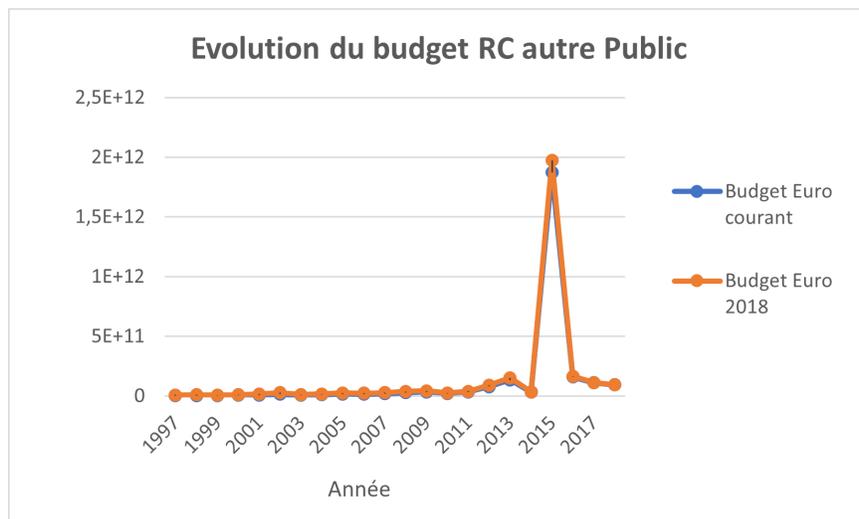


FIGURE 66 – Evolution du Budget RC Autre Public après retraitement

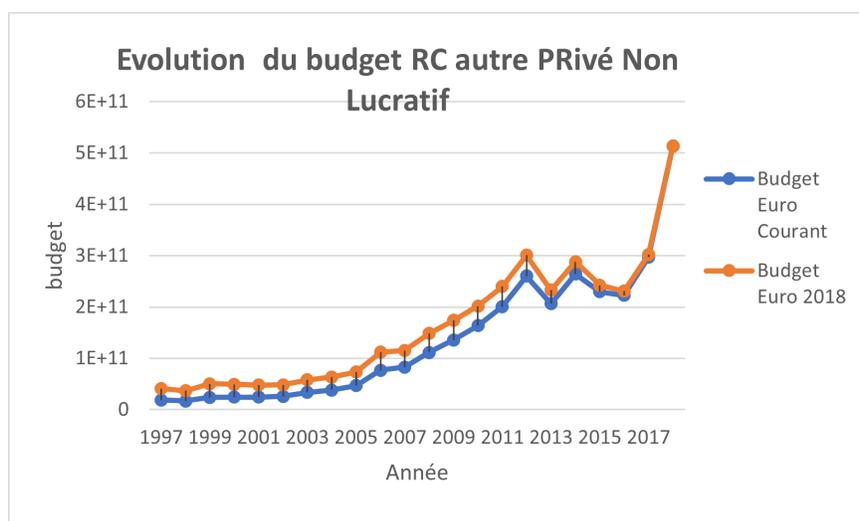


FIGURE 67 – Evolution du Budget RC Autre Privé Non Lucratif après retraitement

Une fois le Budget en € 2018 calculé, nous allons estimer la fréquence de sinistres pour l'année de souscription 2019.

## 7.4 Modélisation de la fréquence

Nous allons utiliser les mêmes méthodes présentées plus haut.

### 7.4.1 Estimation de la fréquence

Nous allons réaliser l'estimation de la fréquence des établissements RC autres "Public" et "Privé à but non-lucratif" en prenant comme variable d'exposition le BUDGET.

#### 7.4.1.1 Les établissements "RC Autre Public"

Pour chacune des méthodes d'imputation nous estimerons la fréquence N+1 (2019) des établissements RC Autre Public. Nous pouvons observer l'allure de la fréquence de sinistres pour chaque méthode d'imputation.

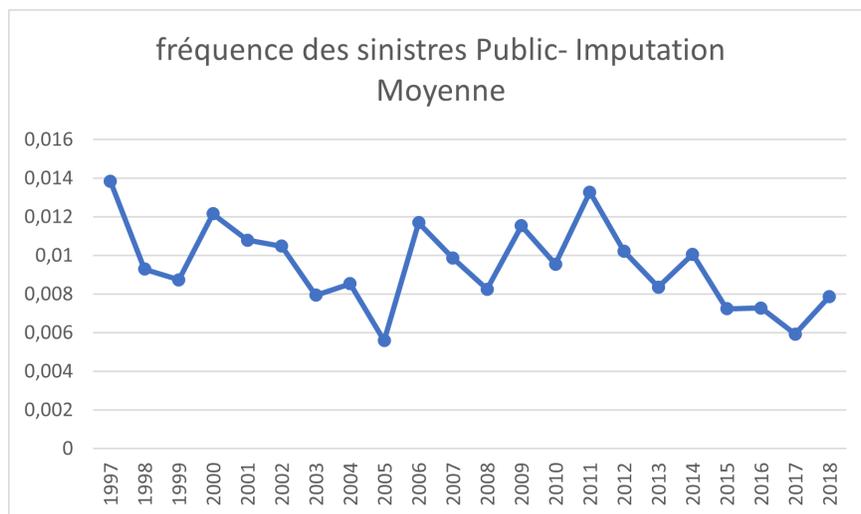


FIGURE 68 – Fréquence RC Autre Public (Budget, Imputation Moyenne)

Nous constatons une stabilité entre 2015-2016 puis une baisse en 2017, enfin une hausse en 2018. Nous pouvons estimer la fréquence de 2019 par moyenne sur l'historique 2015-2018, ou prendre uniquement la fréquence 2018, et nous supposons que la fréquence aura une évolution constante entre 2018-2019. Nous obtenons les résultats suivants :

$$\text{Fréquence RC Autre Public 2019 (Moyenne)} * 10^6 = 0,00787.$$

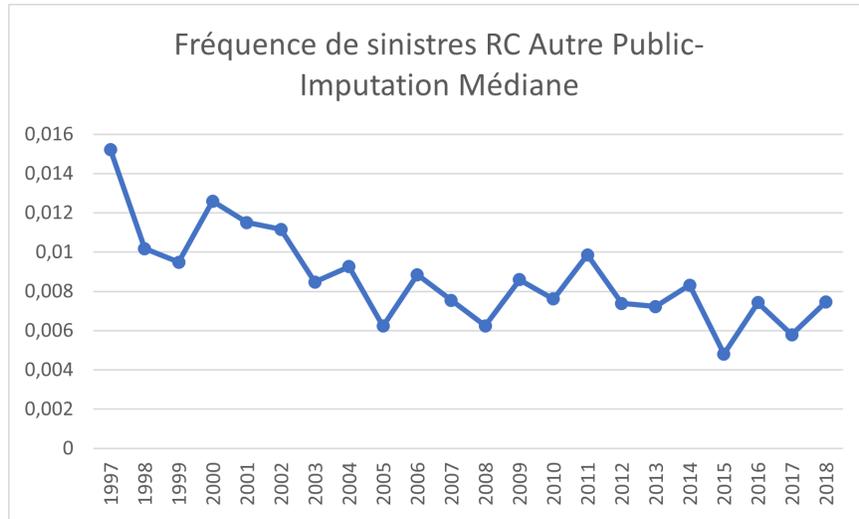


FIGURE 69 – Fréquence RC Autre Public (Budget, Imputation Médiane)

Même si le graphique de fréquence présente beaucoup d'irrégularités, nous remarquons que les valeurs sont autour du même niveau. Nous allons choisir l'historique 2018. Nous obtenons les résultats suivants :

$$\text{Fréquence RC Autre Public 2019 (Médiane)} * 10^6 = 0,00746.$$

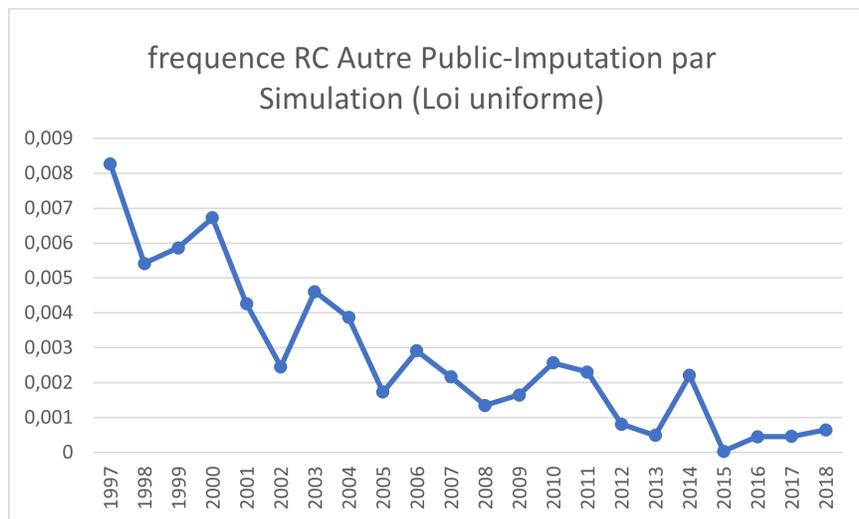


FIGURE 70 – Fréquence RC Autre Public (Budget, Imputation Simulation)

Nous observons une croissance quasi-linéaire sur l'historique 2016-2018, même si la pente s'observe plus sur le segment 2017-2018. Nous allons tester les deux méthodes d'estimation de la fréquence de sinistres, une régression linéaire entre 2015-2018 et une méthode moyenne sur l'historique 2016-2018. Les résultats obtenus sont les suivants :

La régression linéaire donne les résultats suivants :

Comme nous pouvons le voir, les statistiques  $R^2$  et Fisher nous permettent de valider notre modèle.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.755e-01  1.110e-01  -3.384  0.0773 .
annee        1.864e-04  5.503e-05   3.387  0.0772 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0001231 on 2 degrees of freedom
Multiple R-squared:  0.8516,    Adjusted R-squared:  0.7773
F-statistic: 11.47 on 1 and 2 DF,  p-value: 0.0772

>
```

FIGURE 71 – Résultat de la régression linéaire pour la fréquence RC Autre Public

Ainsi la fréquence des sinistres RC Autre Public estimée pour 2019 est de  $(10^6)$  : 0,0001864. Nous vérifierons les résidus s'ils sont centrés.

#### 7.4.1.2 Les établissements "RC Autre Privé Non Lucratif"

Nous effectuons la même démarche que précédemment. Nous présenterons les différentes fréquences obtenues par les différents budgets retraités avec imputation des données manquantes.

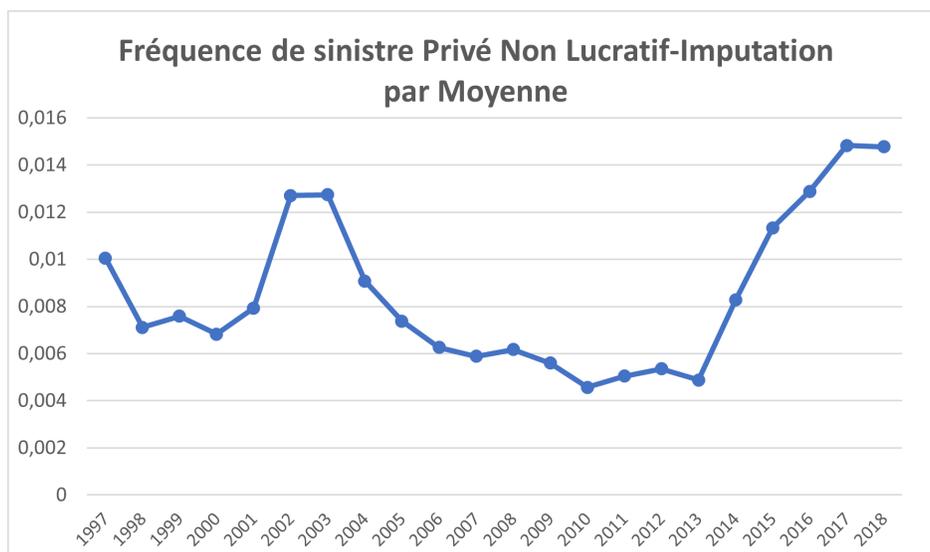


FIGURE 72 – Fréquence RC Autre Privé Non Lucratif (Budget, Imputation Moyenne)

Nous observons une forte croissance de la fréquence entre 2014-2017 puis une stabilité entre 2017 et 2018. Nous choisissons donc la méthode moyenne des fréquences sur l'historique 2017-2018. Les résultats obtenus sont :

Fréquence RC Autre Privé Non Lucratif 2019 (Moyenne) \*  $10^6$  = 0,15.

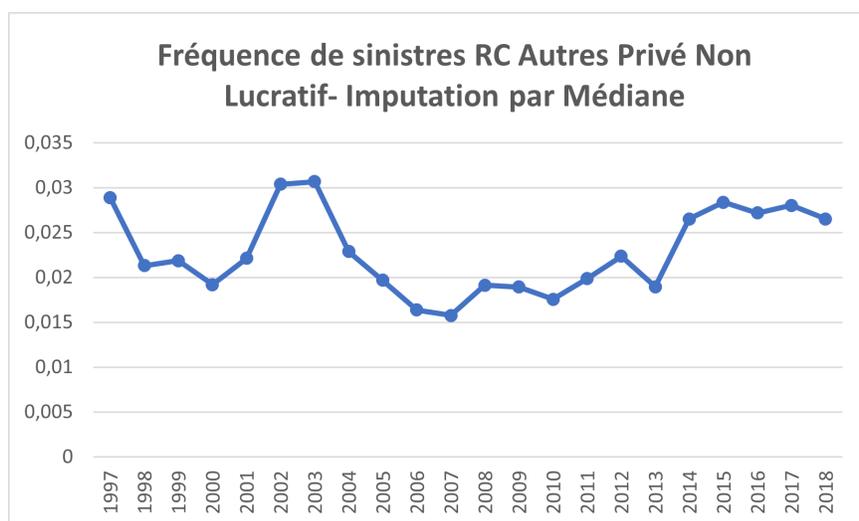


FIGURE 73 – Fréquence RC Autre Privé Non Lucratif (Budget, Imputation Médiane)

Sur ce graphique nous avons une tendance croissante quasi-linéaire entre 2007 et 2015, ensuite nous pouvons une quasi-stabilité entre 2015 et 2018. Nous utilisons la méthode moyenne des fréquence sur l'historique 2015-2018. Les résultats obtenus sont les suivants :

Fréquence RC Autre Privé Non Lucratif 2019 (Médiane) \*  $10^6=0,0075$ .

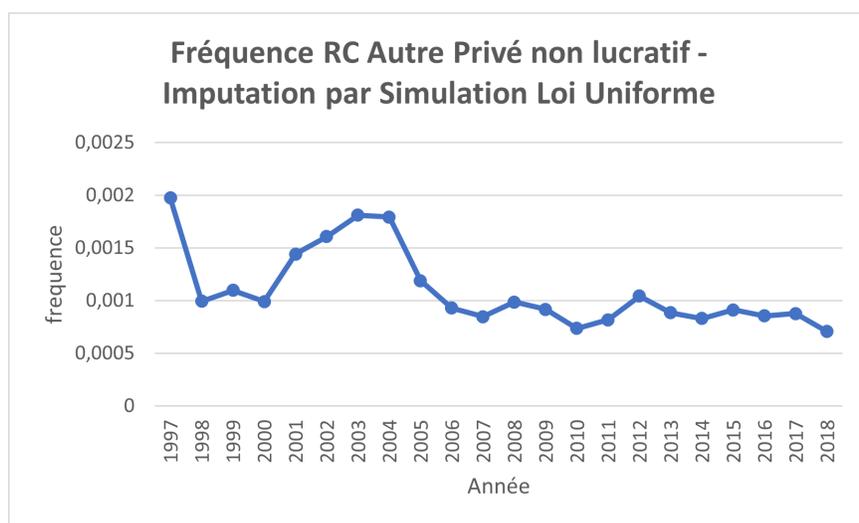


FIGURE 74 – Fréquence RC Autre Privé Non Lucratif (Budget, Imputation Simulation Loi uniforme)

Nous constatons une stabilité sur l'historique 2007 et 2017. Il est intéressant de remarquer que nous avons la même tendance de fréquence que lorsque nous calculons la fréquence avec le nombre de contrats comme variable d'exposition. Par contre nous avons une légère baisse en entre 2017-2018. Pour plus de prudence nous allons estimer la fréquence en faisant la moyenne sur l'historique 2018-2018.

Nous obtenons les résultats suivants :

Fréquence RC Autre Privé Non Lucratif 2019 (Loi Uniforme) \*  $10^6=0,00071$ .

Une fois les fréquences de sinistres estimées pour l'année de souscription 2019, nous allons déterminer l'exposition 2019.

### 7.4.2 Estimation de l'exposition

Nous cherchons dans cette partie à calculer l'exposition de l'année 2019. Cette dernière est différente du Budget 2018 car elle est affectée par le phénomène d'inflation. Nous allons donc dans cette partie estimer le budget 2019 en utilisant un indice spécifique pour les établissements de santé, qui est l'indice ONDAM.

Dans un premier temps, l'objectif national des dépenses d'assurance-maladie (ONDAM) est un montant prévisionnel établi annuellement pour les dépenses d'assurance-maladie en France. C'est un indicateur de maîtrise des dépenses de santé.

L'ONDAM a été créé en 1996 dans le cadre du plan Juppé. Cet objectif concerne les soins de ville, d'hospitalisation dans les établissements privés ou publics, mais aussi dans les centres médico-sociaux. Il est décomposé en sept sous-objectifs qui correspondent à des catégories de dépenses particulières du fait notamment de leurs modalités d'engagement et de régulation.

D'après la Cour des comptes, les dépenses relevant de l'ONDAM prépondérantes parmi celles des branches maladie et accidents du travail-maladies professionnelles ont atteint, selon des données encore provisoires, 200,3 Md€ en 2019. Pour la dixième année consécutive, l'objectif en montant fixé par les lois de financement de la sécurité sociale a été respecté.

Les dépenses ont augmenté de 2,6 % par rapport à 2018, contre 2,5 % prévus en Loi de financement de la sécurité sociale 2019 (LFSS). Les dépenses de soins de ville sont restées dynamiques (+2,7 % contre +2,4 % en 2018), malgré la stagnation ou la baisse de certaines dépenses (médicaments nets des remises, honoraires des médecins généralistes). Les dépenses d'indemnités journalières, de prestations d'auxiliaires médicaux et de transport ont accéléré par rapport à 2018 et la hausse de celles de dispositifs médicaux est restée soutenue. Les économies liées à la maîtrise médicalisée ont été réalisées seulement pour moitié.

Sur le graphique suivant nous pouvons observer l'évolution de l'ONDAM au cours du temps sur (2009-2019) :

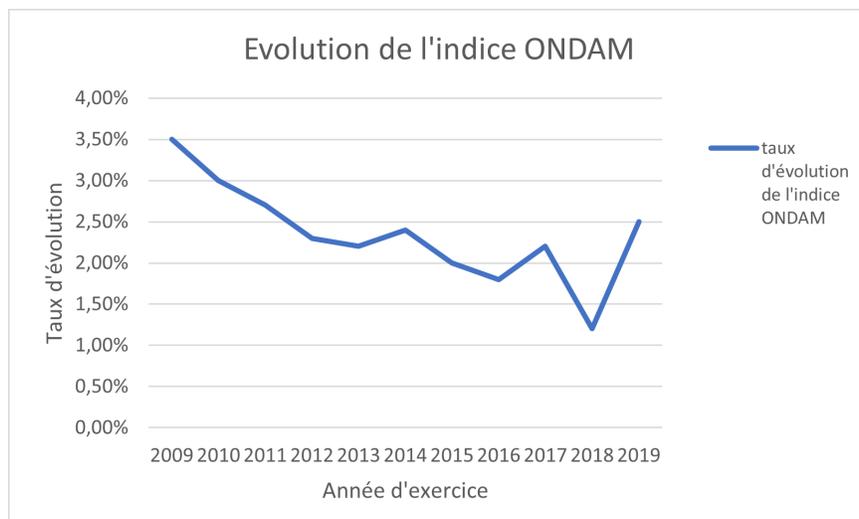


FIGURE 75 – Evolution de l'ONDAM sur 2009-2019

Nous observons sur le graphique précédent que l'indice d'ONDAM connaît une forte baisse entre 2016 et

2018, pour augmenter en 2019.

Dans la suite, nous considérons que le budget alloué par l'Etat en 2019 pour chaque établissement correspond au budget alloué en 2018 auquel on rajoute le montant additionnel dû à l'augmentation annuelle des dépenses en matière de santé. L'indice ONDAM permet de tenir compte de cette augmentation.

On obtient le budget des établissements en 2016 grâce à la formule suivante :

$$\text{BUDGET 2019} = \text{BUDGET 2018} (1 + \text{Indice ONDAM})$$

Pour chacune des méthodes d'imputation nous allons calculer l'exposition BUDGET 2019 pour chacune des classes.

<b>Exposition 2019 RC autre Public</b>	
<b>Imputation Moyenne</b>	468 576 839
<b>Imputation Médiane</b>	493 983 759
<b>Imputation par simulation</b>	5 688 381 167

FIGURE 76 – Estimation de l'exposition RC Autre Public 2019

<b>Exposition 2019 RC Autre Privé Non Lucratif</b>	
<b>Imputation Moyenne</b>	1 140 044 185
<b>Imputation Médiane</b>	635 430 461
<b>Imputation par simulation</b>	23 837 591 144

FIGURE 77 – Estimation de l'exposition RC Autre Privé Non Lucratif 2019

Enfin, nous obtenons le nombre de sinistres simulés pour l'année 2019 de chacune de nos groupes d'établissements.

Ainsi les sinistres estimés pour 2019 sont :

<b>Nombre de sinistre simulé pour 2019 RC autre Public</b>	
<b>Imputation Moyenne</b>	62
<b>Imputation Médiane</b>	62
<b>Imputation par simulation</b>	82

FIGURE 78 – Estimation du nombre de sinistres RC Autre Public 2019

Nous pouvons constater que, lorsque nous complétons les données par imputation par moyenne pour les établissements RC Autre Public, cela revient à peu près à faire une imputation par médiane, d'où l'obtention du même nombre de sinistres simulés pour 2019, l'impact sur le budget étant moins importante.

<b>Nombre de sinistre simulé pour 2019 RC autre Privé Non Lucratif</b>	
<b>Imputation Moyenne</b>	281
<b>Imputation Médiane</b>	289
<b>Imputation par simulation</b>	281

FIGURE 79 – Estimation du nombre de sinistres RC Autre Privé Non Lucratif 2019

## 7.5 Validation et mesure de l'impact sur le risque de primes

Une fois le nombre de sinistres simulés, nous allons effectuer des tests statistiques nous permettant de valider nos différents estimations. Les résultats sont présentés ci-dessous :

RC Autre Public	Méthode d'imputation	Historique choisi	Méthode choisie	P-value	R <sup>2</sup>	Test khl Z/Fisher
	Moyenne	2018-2018	Moyenne	0,61707508	/	Accepté
	Médiane	2018-2018	Moyenne	0,69110222	/	Accepté
	Loi Uniforme	2015-2018	Régression linéaire	/	0,773	Accepté

FIGURE 80 – Validation du Modèle de fréquence RC Autre Public

RC Autre Privé Non Lucratif	Méthode d'imputation	Historique choisi	Méthode choisie	P-value	Test khl 2
	Moyenne	2018-2018	Moyenne	0,7139858	Accepté
	Médiane	2014-2018	Moyenne	0,10597713	Accepté
	Loi Uniforme	2018-2018	Moyenne	0,85539672	Accepté

FIGURE 81 – Validation du Modèle de fréquence RC Autre Privé Non Lucratif

Pour chaque situation, nous pouvons valider la méthodologie utilisée pour estimer la fréquence des sinistres RC Autre Public et RC Autre Privé, comme nous pouvons le constater sur les deux tableaux ci-dessus.

Nous nous intéresserons dans la suite sur l'impact de cette nouvelle variable d'exposition dans le risque de primes, plus précisément sur la sinistralité des contrats RC Autre. Le tableau suivant nous permet de comparer le modèle de la Segmentation 3bis avec le nombre de contrats comme variable d'exposition pour chacune des quatre classes, et la nouvelle segmentation avec le Budget comme variable d'exposition des classes "RC Autre Public" et "RC Autre Privé".

RC Autre	Segmentation 3bis	Segmentation 3bis, nouvelle variable d'exposition (Imputation par moyenne)	Segmentation 3bis, nouvelle variable d'exposition (Imputation par médiane)	Segmentation 3bis, nouvelle variable d'exposition (Imputation par simulation Loi uniforme)
		Nombre de sinistres simulé pour 2019		
Public	57	62	62	82
Privé à but non Lucratif	282	281	289	281
Privé à but lucratif	77	77	77	77
Autre	20	20	20	20
<b>Total</b>	<b>436</b>	<b>439</b>	<b>448</b>	<b>460</b>
<b>différence</b>	-	<b>0,77%</b>	<b>2,75%</b>	<b>5,50%</b>

FIGURE 82 – Comparaison du nombre de sinistres RC Autre en 2019

Le nombre de sinistres simulés pour les classes "Privé à but lucratif" et "Autre" reste inchangé car nous avons gardé la même variable d'exposition. Pour la classe "Public", le nombre de sinistres simulés est le même pour les variables d'exposition budget complétées par les différentes méthodes d'imputation. En ce qui concerne la classe "Privé à but non lucratif", les différences ne sont pas significatives sauf pour les sinistres simulés avec le complément des budgets manquants obtenu par imputation par médiane.

L'impact n'est pas significatif. En effet, sur le total du nombre de sinistres, l'augmentation n'est pas très importante. Nous restons sur les mêmes ordres de grandeurs pour chacune des classes de la segmentation 3 bis.

Nous n'avons pas les données pour recalculer le risque de primes 31/12/2018, mais comme le nombre de sinistres est proportionnel au SCR Primes, nous pouvons supposer que le taux d'évolution de ce dernier est proportionnel aux taux d'évolution des sinistres simulés pour chaque situation.

## 7.6 Backtesting

Dans cette partie, nous allons comparer les nombres de sinistres simulés aux nombres de sinistres réels observés.

RC Autre	Segmentation 3bls	Segmentation 3bls, nouvelle variable d'exposition (Imputation par moyenne)	Segmentation 3bls, nouvelle variable d'exposition (Imputation par médiane)	Segmentation 3bls, nouvelle variable d'exposition (Imputation par simulation Loi uniforme)	RC Autres
		Nombre de sinistres simulé pour 2019			
Public	57	62	62	82	48
Privé à but non Lucratif	282	281	289	281	273
Privé à but lucratif	77	77	77	77	83
Autre	20	20	20	20	20
<b>Total</b>	<b>436</b>	<b>439</b>	<b>448</b>	<b>460</b>	<b>424</b>
<b>écart par rapport au réel</b>	<b>2,83%</b>	<b>3,62%</b>	<b>5,66%</b>	<b>8,49%</b>	

FIGURE 83 – Backtesting sur le nombre de sinistres RC Autres 2019

Nos différents modèles surestiment les nombres de sinistres, pour les établissements Public et Privé Non Lucratif.

Pour les établissements privés, le modèle où l'exposition est le nombre de contrats prédit le mieux le nombre de sinistres. Par contre, pour le privé non lucratif, il semble que les modèles avec comme variables d'exposition le Budget retraité des valeurs manquantes avec la méthode d'imputation par moyenne et avec la méthode de simulation avec une loi uniforme prédisent mieux le nombre de sinistres de cette classe.

Globalement, le modèle avec comme variable d'exposition le nombre de contrats prédit mieux le nombre de sinistres pour la classe RC Autre.

## 7.7 Tests de sensibilités : Impact des données manquantes de la variable BUDGET

Dans cette partie, nous verrons l'impact des données manquantes dans le calcul du risque de primes. Nous n'allons pas prendre en compte les données manquantes, car nous souhaiterons vérifier si l'imputation des données budgets dans notre étude est nécessaire.

Les mêmes calculs présentés dans les parties précédentes sont utilisés dans la suite. Nous obtenons les résultats suivants :

RC Autre	Segmentation 3bis	Segmentation 3bis, nouvelle variable d'exposition (imputation par moyenne)	Segmentation 3bis, nouvelle variable d'exposition (imputation par médiane)	Segmentation 3bis, nouvelle variable d'exposition (imputation par simulation Loi uniforme)	Segmentation 3bis nouvelle variable d'exposition sans imputation des données manquante	RC Autres
		Nombre de sinistres simulé pour 2019				
Public	57	62	62	82	62	48
Privé à but non lucratif	282	281	289	281	281	273
Privé à but lucratif	77	77	77	77	77	83
Autre	20	20	20	20	20	20
Total	436	439	448	460	440	424
écart par rapport au réel	2,83%	3,62%	5,66%	8,49%	3,77%	

FIGURE 84 – Comparaison du nombre de sinistres RC Autre Simulés pour 2019

Nous constatons qu'en réalité les données manquantes n'ont qu'un léger impact sur le calcul du risque de primes. Mais les techniques d'imputation restent intéressantes, car elles permettent de compenser les données et paraissent réalistes si la variable d'exposition est le BUDGET. Mais, comme le budget subit un phénomène d'inflation qui est calculé à partir des données du portefeuille de SHAM, nous n'avons pas pris en compte tous les établissements pour calculer cet indice.

Comme nous pouvons l'observer, il semblerait que la variable d'exposition "Nombre de contrats" soit celle qui permet d'avoir des résultats plutôt proches de la réalité, pour la clôture 31/12/2018.

Il existe d'autres méthodes d'imputation de données manquantes, comme la méthode des k plus proches voisins ou kNN (k- nearest neighbours) qui consiste à remplacer la valeur manquante par la moyenne de ses k plus proches voisins. Avec cette méthode d'imputation nous aurions obtenu des résultats proches de ceux présentés précédemment.

## Conclusion

Le but principal de ce mémoire est d'étudier l'impact d'une segmentation alternative de la classe RC Autre dans le calcul du SCR risque de primes de SHAM. En effet, en raison d'un mauvais backtesting du modèle utilisé à la clôture du 31/12/2018, une étude de segmentation du segment RC Autre semblait intéressante afin d'améliorer le modèle.

Nous avons commencé par l'extraction et l'exploration de nos données. Une fois nos données extraites, nous avons procédé à l'analyse descriptive afin de pouvoir proposer des segmentations potentielles. Ensuite, nous avons procédé à la segmentation de la classe RC Autre. Nous en avons testé plusieurs, mais nous avons présenté uniquement celles qui nous semblaient pertinentes dans ce mémoire. Pour chacune des segmentations retenues, nous avons modélisé les fréquences moyennes pour l'exercice 2019 de chacune des sous-classes. Des études complémentaires ont permis d'améliorer nos modèles proposés tels que le retraitement des MCO.

Une fois que la phase concernant la modélisation de la fréquence des sinistres a été effectuée, nous avons ensuite estimé le SCR primes SHAM de chaque segmentation pour l'exercice 2019. Nous avons confronté les segmentations testées en effectuant un backtesting pour l'année 2019. Nous en sommes arrivés à la conclusion suivante : la segmentation ("Public", "Privé lucratif", "Privé non lucratif", "Autre") en excluant le groupe MCO X nous paraît pertinente pour modéliser la fréquence RC autre dans le calcul du risque de primes et pourrait être utilisée dans le futur dans le MIP SHAM, cependant, il serait intéressant de faire un nouveau backtesting sur l'année 2020 afin de vérifier la stabilité de la segmentation retenue.

Dans une deuxième partie, nous cherchions à challenger la variable d'exposition de la segmentation retenue. Nous avons fait face à de nombreuses limites. Tout d'abord, nous n'avions pas assez de variables pertinentes pouvant être de potentielles variables d'exposition. Nous avons ainsi choisi la variable BUDGET pour le reste de l'étude. De plus, nous avons pour chacune des classes plus 50% de données manquantes en 2017, par exemple. Nous avons trouvé des méthodes permettant de résoudre la problématique de données manquantes.

Pour chacune des méthodes d'imputation utilisées nous simulons le nombre de sinistres 2019 afin d'étudier l'impact sur le risque de primes. Nous avons ensuite effectué un backtesting afin de comparer nos différents modèles et conclure sur cette partie. Finalement, il semble que le nombre de contrats soit la variable d'exposition la plus fiable permettant de modéliser la fréquence des sinistres des classes RC Autre. En ce qui concerne la nouvelle segmentation, elle est pertinente dans l'évaluation du risque de primes SHAM. L'étude nous a permis de mieux comprendre enfin les limites de la segmentation d'origine.

Dans quelques années, il serait intéressant de reconsidérer le "nombre de séjours" comme variable d'exposition ou comme variable explicative pour la tarification par exemple. Les établissements de santé étant touchés par la crise du coronavirus, durant le début de cette pandémie, ces derniers ont dû enregistrer des nombres de séjours supérieurs à leur moyenne annuelle, certains ne disposant pas assez de place pour accueillir tous les malades. Dans cette configuration, le nombre de séjours dépendra du nombre de places disponibles comprenant le nombre de lits. Cette variable dépendra aussi de la probabilité qu'une crise sanitaire se réalise. Il faudra aussi séparer les types de séjours pour plus de précision.

Lors de cette étude nous avons aussi rencontré comme limite le recalibrage du coût moyen par sinistre, en effet ce dernier est calculé à partir d'un modèle BE sophistiqué, si nous devrions recalibrer ce coût moyen, cela nécessiterait aussi le recalibrage du modèle BE utilisé qui est trop complexe.

## Annexes

### Annexe 1 : Vérification de l'hypothèse d'indépendance des classes RC autre pour la segmentation 3

Nous avons effectué un test de Khi deux d'indépendance, nous avons obtenus le résultat suivant :

```
> chisq.test(as.matrix(cbind(Public,Prive_lucratif,Prive_non_lucratif,Autre)))

Pearson's Chi-squared test

data:  as.matrix(cbind(Public, Prive_lucratif, Prive_non_lucratif, Autre))
X-squared = 37.708, df = 63, p-value = 0.9952
```

FIGURE 85 – Test d'indépendance sur les classes RC autres de la segmentation 3

Comme nous pouvons le voir, nous avons une p-value proche de 1, nous pouvons donc valider l'hypothèse d'indépendance entre nos classes.

### Annexe 2 : Sous-dispersion de la fréquence historique

En comparant la variance des fréquences historiques et l'espérance de ces dernières, nous avons obtenus les résultats suivants :

		Freq 5%	Freq 95%	Var/moy Fréquence historique
Segmentation 2	Public	2,89E-02	4,49E-02	0,40%
	Privé	1,53E-01	1,80E-01	0,85%
	Autre	9,00E-03	1,94E-02	0,59%
Segmentation 3	Public	2,89E-02	4,49E-02	0,40%
	Privé à but lucratif	6,26E-01	8,18E-01	29,26%
	Privé à but non lucratif	1,13E-01	1,38E-01	0,99%
	Autre	9,00E-03	1,94E-02	0,59%
		Freq 5%	Freq 95%	Var/moy Fréquence historique
Segmentation 2 bis	Public	2,89E-02	4,49E-02	0,40%
	Privé	1,26E-01	1,51E-01	0,90%
	Autre	9,00E-03	1,94E-02	0,59%
Segmentation 3 bis	Public	2,89E-02	4,49E-02	0,40%
	Privé à but lucratif	3,56E-01	5,20E-01	25,34%
	Privé à but non lucratif	1,13E-01	1,38E-01	0,99%
	Autre	9,00E-03	1,94E-02	0,59%

FIGURE 86 – Sous-dispersion de la fréquence historique

Comme nous pouvons le voir, le quotient Variance, Espérance est inférieure à 1 ce qui permet de valider l'hypothèse de sous-dispersion de la fréquence historique.

## Annexe 3 : Quelques tests de sensibilités

### Segmentation 2 : Revue de la modélisation de la fréquence de la classe "Privé" RC autre

Nous avons testée une régression linéaire sur l'historique 2016-2018, les résultats du modèle sont les suivants :

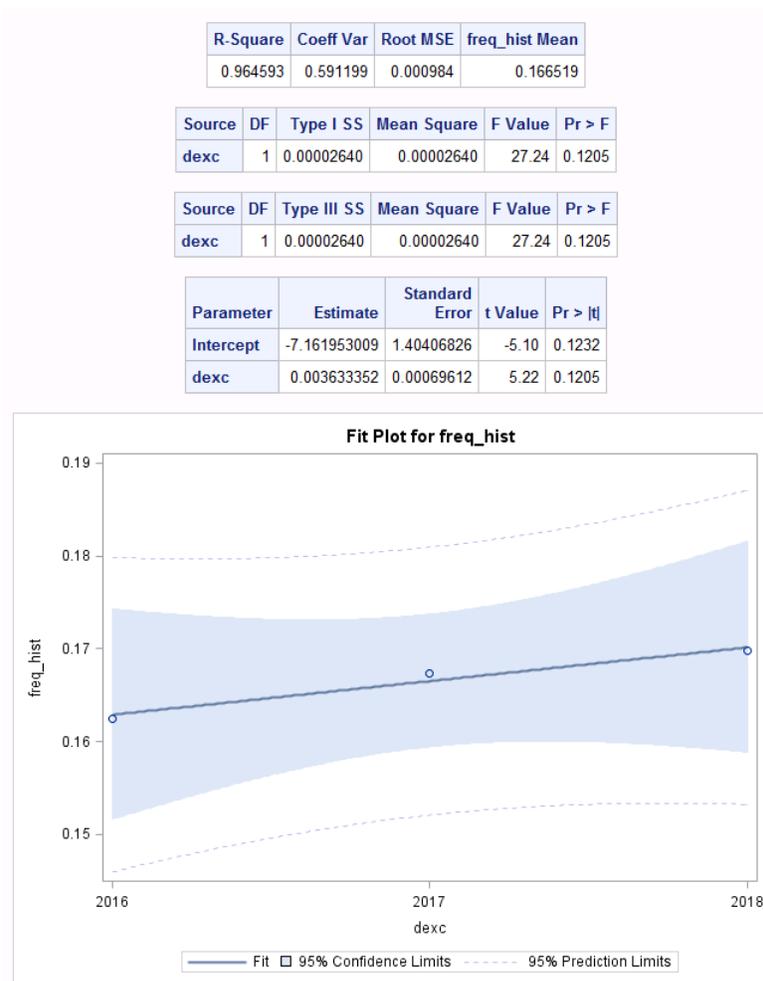


FIGURE 87 – Résultats de la régression linéaire

Le modèle s'ajuste assez bien. Ainsi en comparant nos deux modélisations, nous avons :

Segmentation 2	Fréquence simulé 2019	Nombre de sinistre simulé pour 2019	Var /moy fréq hist	Var/moy nb sin	P-value/RA2
Modèle par moyenne	1,665E-01	409	0,85%	100,2%	0,815788152
Modèle par régression Linéaire	1,738E-01	427	0,85%	100,5%	0,992808858

FIGURE 88 – Etude de sensibilité sur la modélisation de la fréquence des privés RC autre

Avec la régression linéaire nous simulons plus de sinistres, car la tendance était croissante. Nous obtenons un risque de primes de **91.46M €** au lieu de **91.38 M €**, l'écart est assez faible.

### Segmentation 3 : Revue de la modélisation de la fréquence de la classe "Privé à but non lucratif" RC autre

Dans les résultats suivants, nous pouvons voir que l'ajustement du modèle est assez faible :

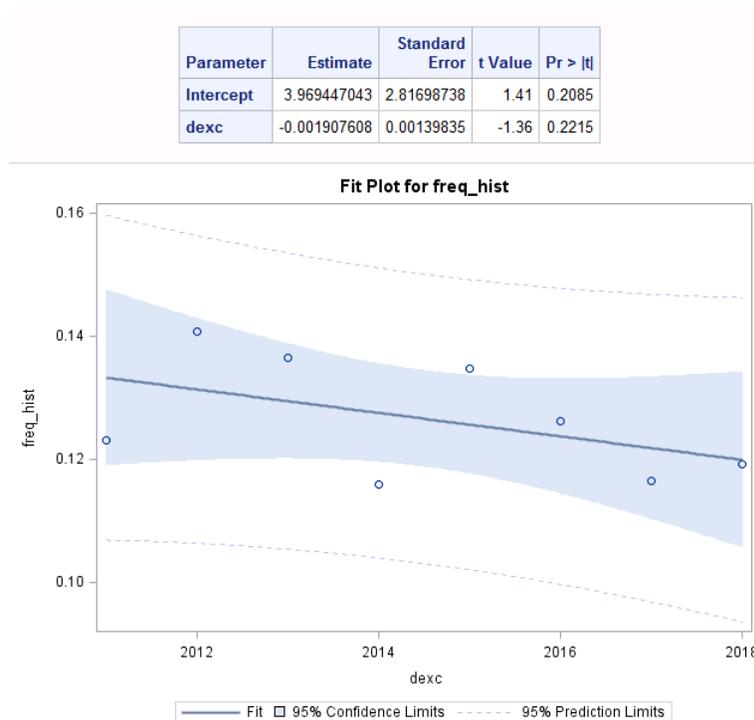


FIGURE 89 – Résultats de la régression linéaire

Nous avons effectué aussi une régression linéaire afin de mesurer l'impact dans l'estimation du nombre de sinistres pour l'exercice N+1. Nous avons obtenu le résultat suivant :

Segmentation 3	Fréquence simulé 2019	Nombre de sinistre simulé pour 2019	Var /moy fréq hist	Var/moy nb sin	P-value/R^2
Modèle par moyenne	1,257E-01	282	0,99%	100,0%	0,182980033
Modèle par régression Linéaire	1,181E-01	265	0,99%	100,4%	0,392

FIGURE 90 – Etude de sensibilité sur la modélisation de la fréquence des Privés non licratifs

Comme la tendance est décroissante, nous simulons un nombre de sinistres inférieur au modèle sans régression linéaire. De plus, nous obtenons un risque de **92.35M €**, contre **92.73M €** obtenu avec une régression par moyenne. L'impact demeure assez faible. La méthode par moyenne paraît plus prudente.

## Annexe 3 : Autres résultats sur les différentes modélisations de fréquences

### Segmentation 2

Analysis Variable : nb_sin_sim							
type_contrat	type_etablismt	N Obs	N	Mean	Std Dev	Minimum	Maximum
PSY		40000	40000	75.6665500	8.7031779	44.0000000	118.0000000
RCP	Autre	40000	40000	97.4732250	9.9124235	58.0000000	141.0000000
	Chirurgie	40000	40000	555.3518750	23.6282768	453.0000000	648.0000000
	Obstetrique	40000	40000	194.1715000	13.9744828	139.0000000	255.0000000
RC_autre	Autre	40000	40000	20.0107250	4.4906975	5.0000000	41.0000000
	Prive	40000	40000	409.1743750	20.2581719	327.0000000	488.0000000
	Public	40000	40000	57.1124250	7.5995940	27.0000000	91.0000000
San_prive		40000	40000	1201.47	34.4666441	1066.00	1344.00
San_public	CH	40000	40000	2931.82	54.4415849	2725.00	3146.00
	CHU	40000	40000	2198.25	46.8315091	2001.00	2402.00

FIGURE 91 – Simulation des nombres de sinistres pour 2019

### Segmentation 3

Analysis Variable : nb_sin_sim							
type_contrat	type_etablismt	N Obs	N	Mean	Std Dev	Minimum	Maximum
PSY		40000	40000	75.6665500	8.7031779	44.0000000	118.0000000
RCP	Autre	40000	40000	97.4732250	9.9124235	58.0000000	141.0000000
	Chirurgie	40000	40000	555.3518750	23.6282768	453.0000000	648.0000000
	Obstetrique	40000	40000	194.1715000	13.9744828	139.0000000	255.0000000
RC_autre	Autre	40000	40000	20.0107250	4.4906975	5.0000000	41.0000000
	Prive	40000	40000	142.8750500	11.9740207	94.0000000	189.0000000
	Prive_N_L	40000	40000	282.4223750	16.9094625	213.0000000	359.0000000
	Public	40000	40000	57.0688000	7.5697754	27.0000000	92.0000000
San_prive		40000	40000	1201.47	34.4666441	1066.00	1344.00
San_public	CH	40000	40000	2931.82	54.4415849	2725.00	3146.00
	CHU	40000	40000	2198.25	46.8315091	2001.00	2402.00

FIGURE 92 – Simulation des nombres de sinistres pour 2019

## Références

- [1] Imputations de données manquantes. <https://www.math.univ-toulouse.fr/besse/Wikistat/pdf/st-scenar-appidm.pdf>.
- [2] L'ondam 2019 a été respecté se félicite la cour des comptes. <https://www.egora.fr/actus-pro/hopitaux-cliniques/59893-l-ondam-2019-a-ete-respecte-se-felicite-la-cour-des-comptes>.
- [3] Projet de financement de la sécurité sociale. [https://www.economie.gouv.fr/files/files/PDF/2018/dp\\_plfss2019.pdf](https://www.economie.gouv.fr/files/files/PDF/2018/dp_plfss2019.pdf), 2018.
- [4] Romain RESPRIGET Anis TRABELSI. "régression de poisson". <https://perso.univ-rennes1.fr/valerie.monbet/ExposesM2/2013/RegressionPoissonAR.pdf>.
- [5] MISERAY Antoine. *Analyse du coût des sinistres dans le cadre de la modélisation ligne à ligne d'un portefeuille d'assurance automobile*. PhD thesis, ISFA, 2015.
- [6] Anthony DERIEN. *Thèse : Solvabilité 2 : une réelle avancée ?* PhD thesis, Ecole Doctorale Sciences Economiques et de Gestion, 2010.
- [7] Lebarbier E. and Mary-Huard T. *Classification non supervisée*. PhD thesis, AgroParisTech.
- [8] Marcin FEDOR and Julien MOREL. *Value at Risk en assurance : recherche d'une méthodologie à long terme*. PhD thesis, Université Paris Dauphine, ISFA.
- [9] Pierre GUILLEMINET and Ying LI. *Intégrer l'allocation d'actifs comme paramètre au modèle interne d'une compagnie d'assurance vie*. PhD thesis, ENSAE, Paris Tech, 2014.
- [10] Inspection générale des affaires sociales Inspection générale des finances. *La proposition de la maîtrise de l'ondam (2013-2017)*. [https://www.economie.gouv.fr/files/rapport\\_ondam\\_gas\\_igf.pdf](https://www.economie.gouv.fr/files/rapport_ondam_gas_igf.pdf), 2012.
- [11] Julien JAQUES. *Classification Supervisée*. 2018.
- [12] Emmanuel KOENING. *Méthode de segmentation, de mesure d'incertitude et d'agrégation en Provisionnement non-vie*. PhD thesis, ISFA, 2012.
- [13] Esterina Masiello. *Cours : Inférence statistique des copules, Compléments du cours d'Assurance non-vie*. PhD thesis, ISFA.
- [14] Maïmouna COULIBALY N'ga, Ahou. *Modélisation de la fréquence et du coût moyen des sinistres en Assurance Automobile du particulier avec une approche temporelle et l'intégration de variables explicatives*. PhD thesis, ISUP, 2010.
- [15] Gambette Philippe. "classification supervisée et non supervisée". <http://igm.univ-mlv.fr/gambette/EnsInfoling/CM3-2014.pdf>.
- [16] YANG Qianglong. *Segmentation du risque en assurance non-vie dans le cadre de l'élaboration du modèle interne partiel de Coface sur le risque de souscription*. PhD thesis, UFR, Strasbourg, 2013.
- [17] Ando Ratsimbazafy. *Modèle interne partiel en RC médicale : Revue de la variable d'exposition des établissements sanitaires privés utilisée pour l'évaluation du risque de primes*. PhD thesis, ISFA, 2016.
- [18] Adrien VOGT. *Modélisation stochastique en assurance non vie : calcul de l'exigence de capital réglementaire sous Solvency II*. PhD thesis, UFR de mathématique de informatique DUAS, 2010.
- [19] Tremblayl Véronique. *Analyse de données et apprentissage statistique en R*. 2020.

[16] [9] [17] [6] [8] [15] [4] [13] [7] [14] [18] [12] [19] [11] [10] [3] [2] [1] [5]