Mémoire présenté le :
pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuaires

<table>
<tr><td colspan="2">Par : Guillaume FERRE</td></tr>
<tr><td colspan="2">Titre du mémoire : Tarification de contrats collectifs en assurance santé</td></tr>
<tr><td colspan="2">Confidentialité : ☐ NON ☑ OUI (Durée : ☐ 1 an ☑ 2 ans)</td></tr>
<tr><td colspan="2">Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus.</td></tr>
<tr><td rowspan="2">Membres présents du jury de la filière :
Signature :</td><td>Entreprise :

Nom : AG2R La Mondiale

Signature :</td></tr>
<tr><td>Directeur de mémoire en entreprise

Nom : Marwa HANNAFI
Signature :</td></tr>
<tr><td rowspan="2">Membres présents du jury de l'Institut des Actuaires :
Signature :</td><td>Invité :
Nom :
Signature :</td></tr>
<tr><td>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 :</td></tr>
</table>

Note de synthèse

Tarification de contrats collectifs en assurance santé

Contexte et objectifs :

De nos jours, le marché de l'assurance santé complémentaire est de plus en plus concurrentiel et il est régulièrement soumis à des évolutions législatives. Les acteurs de ce marché doivent donc adapter leurs offres afin de proposer un tarif plus fin et adapté à leurs portefeuilles.

AG2R La Mondiale dispose actuellement d'un outil de tarification qui utilise des normes tarifaires basées sur une méthode couramment utilisée de détermination de la prime pure en fonction du niveau de garantie par acte à laquelle sont appliqués différents coefficients permettant une segmentation en fonction de l'âge, du sexe, de la CSP ou encore de la localisation géographique.

En 2021, AG2R La Mondiale a pris la décision de challenger la méthode utilisée pour la construction de ces normes. Par conséquent, l'objectif de cette étude est, d'une part, d'obtenir de nouvelles normes tarifaires en collectif et, d'autre part, de challenger les dernières normes tarifaires réalisées pour le produit sur-mesure appelé « Simpleo » avec l'utilisation de l'outil « Akur8 » et, ainsi, décider si Akur8 doit être retenu en tant qu'outil cible au sein de l'entreprise ou non.

Méthodologie :

Dans le cadre de ce mémoire, une base de données composée d'une base de bénéficiaires et d'une base de prestations a été utilisée. Une analyse des données a donc été nécessaire au début de l'étude.

Présentation des bases de données utilisées

Pour cette étude, des données du portefeuille d'AG2R La Mondiale intégrant le segment standard collectif avec la gamme « Omega 2 » sur la période du 01/01/2017 au 31/12/2021 ont été utilisées. La date d'arrêté des prestations pour l'étude est le 31 mai 2022. La base de données a été enrichie avec les données des 4 plus grandes CCN du portefeuille en chiffre d'affaires en 2021 que sont Afflec, Aide à domicile, Propreté et Boulangerie artisanale ainsi que celles des 5 plus grandes entreprises en chiffre d'affaires en 2021 dont Cnes, Daher, UES Générale des eaux, Euro-Disneyland et Leroy Somer.

Cette base contient donc plusieurs produits hétérogènes en run-off ou en cours de commercialisation. Il s'agit d'une nouveauté par rapport à l'ancienne méthode de construction des normes tarifaires où la base de données ne contenait que des données du segment standard collectif avec la gamme « Omega 2 ». Un traitement de la base de données a été réalisé avant de procéder à la tarification.

La base des bénéficiaires contient toutes les caractéristiques propres aux bénéficiaires et elle regroupe en moyenne 659 627 bénéficiaires par année :

Année de survenance	Nombre de bénéficiaires	Exposition
2017	723 852	581 773
2018	684 030	556 845
2019	643 841	529 988
2020	620 342	516 707
2021	626 070	510 334
Toutes années confondues	1 178 794	1 025 038

Exposition et nombre de bénéficiaire en fonction de l'année de survenance

La base des prestations regroupe tous les sinistres des assurés dans le secteur de la santé et contient en moyenne 136 500 000 euros de prestations et 17 millions nombre d'actes par année de survenance.

De nombreux retraitements ont été apportés dans la base des bénéficiaires ainsi que dans la base des prestations afin d'avoir une base de données propre et fiable. En effet, certaines variables telles que le sexe présentaient des valeurs manquantes tandis que d'autres variables telles que le montant de frais réels contenaient des valeurs aberrantes. Il a également été nécessaire d'effectuer un travail d'homogénéisation des niveaux de garanties afin de pouvoir comparer plusieurs niveaux de garanties d'un même acte entre eux. En effet, les données provenant de différents produits et s'étalant sur 5 ans, de nombreux niveaux de garanties, s'exprimant de manières différentes, sont présentes au sein de la base de données.

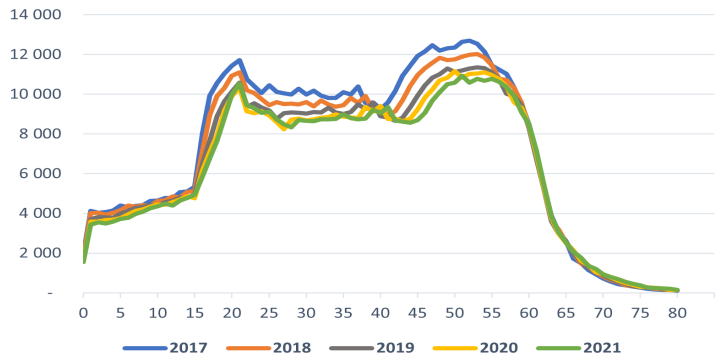
Enrichissement des données

La base de données contient des variables telles que l'âge, le sexe, le type de bénéficiaire (assuré principal, conjoint, majeur protégé ou enfant), le code postal de l'entreprise du salarié ou encore la date d'adhésion au contrat du bénéficiaire. Cette base a été enrichie par la création de plusieurs variables telles que l'exposition du bénéficiaire correspondant à la durée de présence d'un assuré sur une année, la tranche d'effectif de l'entreprise du salarié, le niveau de gamme par poste ainsi que le niveau de gamme du contrat.

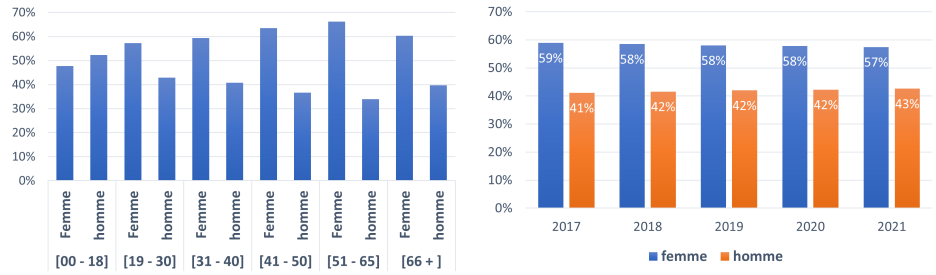
De plus, des variables externes portant sur la densité médicale et le niveau de vie en fonction du département et datant de 2017 ou 2018 ont été récupérées sur le site de l'INSEE et de Score Santé. Il aurait été intéressant d'avoir également la CSP de l'assuré ou encore sa situation familiale mais ces informations n'ont pas pu être récupérées pour l'ensemble des bénéficiaires. Il s'agit d'un axe d'amélioration à l'étude, ces données étant présentes dans les DSN des entreprises.

Analyse des données

La moyenne d'âge du portefeuille est de 35 ans. Les assurés principaux ont une moyenne d'âge de 40 ans, les conjoints de 48 ans, et les enfants de 11 ans. Le portefeuille d'étude est à dominance féminine avec en moyenne 58% de femmes pour 42% d'hommes par année.



Exposition en fonction de l'âge et de l'année de survenance



Répartition de la population en fonction du sexe et de la tranche d'âge à gauche et répartition de la population par sexe et année de survenance à droite

Modélisation

Dans un premier temps, la méthode utilisée lors de la construction des dernières normes tarifaires a été reprise, à savoir l'utilisation de coefficients correcteurs liés à l'âge, au genre et à la région qui sont appliqués à la prime pure déterminée selon le niveau de garantie avec une méthode interne.

Pour challenger cette méthode, une tarification acte par acte a également été choisie. Des modèles GAM ont été utilisés avec Akur8 pour modéliser les actes sur l'année de survenance 2021 car il s'agit de l'année la plus récente et qu'aucun effet « rebond » n'a été constaté suite à la crise sanitaire ayant eu lieu en 2020, mais plutôt un changement de comportement des assurés. De plus, le fait de ne considérer que l'année 2021 permet de ne pas avoir de biais entre les années par rapport à la réforme du 100% santé.

Dans le cadre de cette étude, il a été choisi de modéliser un acte par grand poste de soins en fonction de son importance au sein du poste en nombre de consommant et en montant de prestations. La comparaison avec l'ancienne méthode se fera donc sur 6 actes :

Poste	Acte modélisé
Actes Médicaux	Consultations et visites spécialistes non CAS
Autres prestations	Analyses
Dentaire	Prothèse dentaire
Hospitalisation	Honoraires
Optique	Verres simples adulte
Pharmacie	Pharmacie

La comparaison entre les deux méthodes a été établie selon différents critères d'évaluation tels que la robustesse et la transparence du modèle, la précision des prédictions par rapport à l'observé mais également le gain de temps potentiel.

Conclusion

Dans le cadre de cette étude, la construction des nouvelles normes tarifaires pour des contrats collectifs en santé a été entamé et la méthode utilisée lors de la construction des dernières normes tarifaires a été challengée par l'utilisation de modèles GAM avec l'outil Akur8.

Les résultats permettent de constater des améliorations, dans plusieurs aspects, des normes tarifaires avec l'utilisation des modèles GAM avec Akur8. Cette nouvelle méthode s'avère ainsi pertinente et plus performante que l'approche utilisée lors de la construction des dernières normes tarifaires.

L'ajout de nouvelles variables telles la CSP ou la situation familiale de l'assuré avec les données de la DSN seraient susceptibles d'améliorer les résultats obtenus. Des données externes pourraient aussi être ajoutées avec l'Open Damir. De plus, il serait intéressant de créer des modèles « Fréquence X Coût Moyen » et de les comparer aux modèles « coût total » dans l'avenir. Par ailleurs, les travaux ont été réalisés dans une perspective d'évolution afin d'être facilement ajustables dans le cadre de futures études.

Summary

Pricing of group health insurance contracts

Context and objectives :

Nowadays, the complementary health insurance market is more and more competitive and is regularly subject to legislative changes. The players in this market must therefore adapt their offers in order to propose a more refined rate adapted to their portfolios.

AG2R La Mondiale currently has a pricing tool that uses pricing standards based on an commonly used method of determination of the pure premium according to the level of coverage per act to which various coefficients are applied allowing segmentation according to age, sex, socio-professional category or geographic location.

In 2021, AG2R La Mondiale has decided to challenge the method used to build these pricing standards. Therefore, the objective of this study is, on the one hand, to have new collective tariff standards and, on the other hand, to challenge the last tariff standards realized for the tailor-made product called « Simpleo » with the use of the « Akur8 » tool and, thus, decide whether Akur8 should be retained as a target tool within the company or not.

Methodology :

For this study, a database consisting of a beneficiary base and a claims base was used. Therefore, a data analysis was required at the beginning of the study.

Presentation of the databases used

For this study, data from AG2R La Mondiale's portfolio including the group Standard segment with the « Omega 2 » range over the period from 01/01/2017 to 31/12/2021 were used. The cut-off date for the study is May 31, 2022. The database has been enriched with data from the 4 largest Nationals Collectives Agreements in the portfolio in terms of turnover in 2021, namely Afflec, Aide à domicile, Propreté and Boulangerie artisanale, as well as from the 5 largest companies also in terms of turnover in 2021, including Cnes, Daher, UES Générale des eaux, Euro-Disneyland and Leroy Somer.

This database therefore contains several heterogeneous products, either in run-off or being marketed. This is a change from the previous method of constructing pricing standards, where the database only contained data from the collective standard segment with the « Omega 2 » range. The database was processed prior to pricing.

The beneficiary base contains all the specific characteristics of the beneficiaries and includes an average of 659,627 beneficiaries per year :

Année de survenance	Nombre de bénéficiaires	Exposition
2017	723 852	581 773
2018	684 030	556 845
2019	643 841	529 988
2020	620 342	516 707
2021	626 070	510 334
Toutes années confondues	1 178 794	1 025 038

Exposure and number of beneficiaries by year of occurrence

The claims database contains all the claims of the insured in the health sector and contains an average of 136,500,000 euros of benefits and 17 million procedures per year of occurrence.

Several adjustments were made to the beneficiary base and the claims base in order to have a clean and reliable database. Indeed, some features such as sex had missing values while other variables such as the amount of actual expenses contained absurd values. It was also necessary to homogenize the levels of coverage in order to be able to compare several levels of coverage of the same act. Indeed, the data coming from different products and spread

over 5 years, many levels of coverage, expressed in different ways, are present in the database.

Data Enrichment

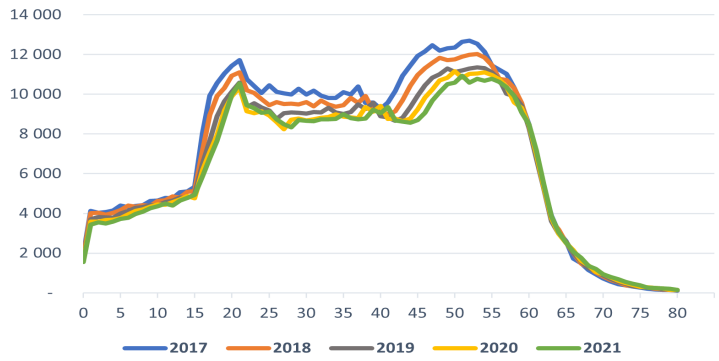
The database contains features such as age, gender, type of beneficiary (main insured, partner, protected adult or child), the postal code of the insured's company or the beneficiary's end date.

This database has been enriched by the creation of several features such as the beneficiary's exposure corresponding to the duration of the insured person's presence over a year, the employee's company's staffing level, the range level per medical post as well as the range level of the contract.

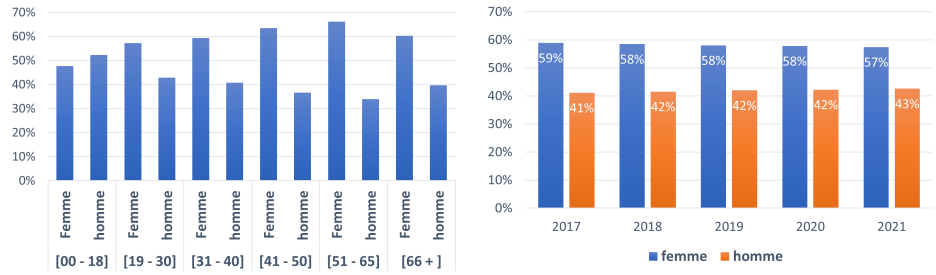
Moreover, external variables on medical density and standard of living according to the department and dating from 2017 or 2018 were retrieved from the INSEE and Score Sante websites. It would have been interesting to also have the insured person's CSP or family situation, but this information could not be retrieved for all beneficiaries. This is an area for improvement that is currently being studied, as these data are present in the companies' DSNs.

Data analysis

The average age of the portfolio is 34 years. The average age of the principal insureds is 38, the average age of the partners is 47, and the average age of the children is 12. The study portfolio is female-dominated with an average of 59% women and 41% men per year.



Exposure by age and year of occurrence



Distribution of the population by gender and age group on the left and distribution of the population by gender and year of occurrence on the right

Modelization

In the first instance, the method used in the construction of the last tariff standards was adopted, i.e. the use of correction coefficients linked to age, gender and region which are applied to the pure premium determined according to the level of cover with an internal method.

To challenge this method, an act by act pricing was also chosen. GAM models were used with the Akur8 tool to model the acts for the year of occurrence 2021 because this is the most recent year and no "rebound" effect was observed following the health crisis that occurred in 2020, but rather a change in policyholder behavior. In addition, considering only the year 2021 avoids any bias between years in relation to the 100% health reform.

For this study, it was decided to model one act by major post according to its importance within the post in terms of the number of consumers and the amount of claims. The comparison with the old method will therefore be made on 6 acts :

Medical post	Modelized act
Medical acts	Non-CAS specialist consultations and visits
Other claims	Analysis
Dental	Dental prosthesis
Hospitalization	Medical fee
Optics	Adult single lenses
Pharmacy	Pharmacy

The comparison between the two methods was then established according to different evaluation criteria such as the robustness and transparency of the model, the accuracy of the predictions compared to the observed but also the potential time saving.

Conclusion

As part of this study, the construction of new tariff standards for group health contracts was initiated and the method used in the construction of the last tariff standards was challenged by testing the use of GAM models with Akur8.

The results show improvements in several aspects of the tariff standards with the use of GAM models with Akur8.

The fact of using new features such as the CSP or the family situation of the insured with data from the DSN would be likely to improve the results obtained. External data could also be added with the Open Damir. Also, it would be interesting to create « Frequency X Average Cost » models and to compare them to « total cost » models in the future. Moreover, the work has been done with an evolutionary perspective in order to be easily adjustable for future studies.

Résumé

En assurance santé, la tarification des complémentaires santé est fréquemment impactée par les évolutions réglementaires et le marché de l'assurance santé complémentaire est de plus en plus concurrentiel. Face à ces enjeux, il faut alors établir une tarification fine, précise et qui puisse être mise à jour rapidement.

Actuellement, le processus de tarification utilisée par AG2R La Mondiale est basé sur une méthode utilisée couramment de détermination de la prime pure en fonction du niveau de garantie par acte à laquelle sont appliqués différents coefficients permettant une segmentation en fonction de l'âge, du sexe, de la CSP ou encore de la localisation géographique.

En 2021, la direction de l'Actuariat a décidé de tester un nouvel outil, sollicitant l'utilisation de modèles additifs généralisés et nommé « Akur8 », pour la construction de jeux de normes tarifaires pour le risque santé.

L'objectif de ce mémoire est, d'une part, d'obtenir de nouvelles normes tarifaires en collectif et, d'autre part, de challenger les dernières normes tarifaires réalisées pour le produit sur-mesure appelé « Simpleo ». Cela permettra également de vérifier si l'utilisation du nouvel outil Akur8 apporte un gain de temps opérationnel dans la création et mise à jour de normes tarifaires santé tout en permettant d'obtenir un modèle robuste, transparent et performant en vitesse d'exécution et en précision.

Mots clés : assurance santé collective, complémentaire santé, tarification, modèle additif généralisé, segmentation.

Abstract

In health insurance, the pricing of complementary health insurance is frequently impacted by regulatory changes and the complementary health insurance market is increasingly competitive. It is therefore necessary to establish a precise and accurate pricing system that can be updated quickly.

Currently, the pricing process used by AG2R La Mondiale is based on an commonly used method of determination of the pure premium according to the level of coverage per act to which various coefficients are applied allowing segmentation according to age, sex, socio-professional category or geographic location.

In 2021, the Actuarial Department has decided to test a new tool, soliciting the use of generalized additive models and called « Akur8 », for the construction of sets of tariff standards for health risk.

The objective of this study is, on the one hand, to have new tariff standards in collective and, on the other hand, to challenge the last tariff standards realized for the tailor-made product called « Simpleo ». This will also allow to verify if the use of the Akur8 brings operational time savings in the creation and updating of health tariff standards while allowing to obtain a model that is robust, transparent and efficient in terms of execution speed and accuracy.

Keywords : group health insurance, complementary health insurance, pricing, generalized additive model, segmentation.

Remerciements

En premier lieu, je souhaite remercier AG2R La Mondiale qui m'a donné l'opportunité d'intégrer leur entreprise dans le cadre d'une alternance.

Je remercie également l'équipe de la Direction de l'Actuariat d'AG2R La Mondiale pour l'accueil chaleureux qui m'a été réservé ainsi que pour leur bienveillance tout au long de mon alternance.

En particulier, je tiens à remercier Stéphanie Laccam, manager de l'équipe Santé, et Marwa Hannafi, mon tuteur au sein de l'entreprise, qui m'ont accompagnés et encadrés durant cette étude et qui ont pris le temps de me fournir de précieux conseils dans la rédaction de mon mémoire. Je remercie également Marie Buisson, membre de l'équipe Santé, ainsi que Marwa Hannafi et Stéphanie Laccam qui m'ont fait confiance en acceptant ma candidature et en me permettant de rejoindre leur équipe.

Je remercie également François-Henri Toutain, Jean Ferrero ainsi qu'Ellen Olympio, pour leurs précieux conseils ainsi que leur expertise métier qui m'ont beaucoup aidé dans la réalisation de mon mémoire.

Par ailleurs, je remercie l'ensemble de l'équipe pédagogique de l'ISUP pour la qualité de la formation qui m'a permis de réaliser ce mémoire. Je remercie Maud Thomas, mon référent ISUP, qui m'a permis de prendre le recul nécessaire sur mon mémoire grâce à ses conseils.

Enfin, je tiens à remercier ma famille pour leur soutien tout au long de mes études et particulièrement lors de la rédaction de ce mémoire.

Table des matières

Lexique

AT/MP : Accidents du Travail et Maladies Professionnelles,
BAD : Branche de l'aide, de l'accompagnement et des services à domicile,
BR : Base de remboursement retenue par l'assurance maladie obligatoire,
CCN : Convention Collective Nationale,
Coût moyen : Il correspond au montant total des remboursements complémentaires divisé par le nombre d'actes,
Covid-19 : Coronavirus disease 2019,
FR : Frais réels engagés par le bénéficiaire,
Fréquence : Fréquence annuelle de survenance d'un sinistre par bénéficiaire, c'est le nombre d'actes total d'une population considérée sur le nombre de bénéficiaires de cette population,
Nombre d'actes : Donnée quantitative d'une ligne de décompte qui peut correspondre par exemple aux jours d'hospitalisation,
PMSS : Plafond Mensuel de la Sécurité sociale,
PSAP : Provisions pour Sinistres A Payer,
Reste à Charge : Montant restant à la charge de l'assuré,
RG : Régime Général,
RL : Régime Local Alsace Moselle,
RO : Régime Obligatoire,
RSS : Remboursement Sécurité sociale = montant remboursé par l'assurance maladie obligatoire et calculé par l'application du taux de remboursement légal en vigueur à la base de remboursement,
TM : Ticket Modérateur : partie de la base de remboursement non prise en charge par l'assurance maladie obligatoire (TM = BR – RSS),
TNS : Travailleurs Non-Salariés,
URSSAF : Union de Recouvrement des Cotisations de Sécurité Sociale et d'Allocations Familiales.

Introduction

En France, le secteur de l'assurance santé est régulièrement impacté par les évolutions législatives et réglementaires. Ces impacts ont pu être constatés récemment avec les réformes du contrat responsable, de la résiliation infra-annuelle ou encore du « 100% Santé » qui vient transformer le système de santé en France. Le contexte sanitaire a également eu un impact suite à l'apparition de la pandémie liée au Covid-19 avec notamment une sous-consommation notable sur certains postes de soins en 2020. De plus, le marché de l'assurance santé complémentaire est soumis à une concurrence de plus en plus forte.

Ainsi, la tarification des complémentaires santé est impactée par ces évolutions et il faut alors en tenir compte lors des travaux de tarification. Dans ce contexte, la nécessité de posséder une tarification précise, adaptée à son portefeuille ainsi qu'aux différents changements règlementaires et qui puisse être mise à jour rapidement est plus que jamais au cœur du métier d'Actuaire.

Les normes tarifaires utilisées actuellement au sein d'AG2R La Mondiale afin de tarifer les contrats en assurance santé sont basées sur une méthode couramment utilisée de détermination de la prime pure en fonction du niveau de garantie par acte à laquelle sont appliqués différents coefficients permettant une segmentation en fonction de l'âge, du sexe, de la CSP ou encore de la localisation géographique.

Toutefois, ces normes tarifaires sont construites sur des données datant de 2016-2017 et elles ne sont mises à jour que tous les 3 ans. Ainsi, un nouvel outil de tarification, nommé Akur8, va être testé pour la construction de jeux de normes tarifaires pour les trois risques santé, prévoyance et dépendance au sein de la direction de l'actuariat d'AG2R La Mondiale. Ce mémoire se

focalise sur les normes tarifaires pour le risque santé.

L'objectif de ce mémoire est, en premier lieu, de posséder de nouvelles normes tarifaires en collectif avec des données récentes. Dans un second temps, ce mémoire cherche à challenger les dernières normes tarifaires réalisées pour le produit sur-mesure appelé « Simpleo » en utilisant les modèles additifs généralisés avec l'outil Akur8. Cette étude va permettre d'évaluer cette nouvelle méthode sur plusieurs aspects tels que la performance avec des métriques telles que le Gini, la précision des normes tarifaires par rapport à la population visée, le gain de temps potentiel ou encore la robustesse et la transparence du modèle utilisé.

Pour réaliser ces travaux, plusieurs bases de données sont mises à disposition et recensent tous les bénéficiaires présents au sein du portefeuille standard collectif d'AG2R, des 4 plus grandes CCN et des 5 plus grandes entreprises en gestion directe en chiffre d'affaires ainsi que leurs sinistres survenus sur la période 2017-2021. Les données sont arrêtées à fin mai 2022.

Une tarification par acte a été choisie étant donné que suffisamment de données étaient disponibles sur la plupart des actes. Ainsi, la méthode utilisée lors de la construction des dernières normes tarifaires va être challengée par l'utilisation de modèles GAM avec Akur8. Une comparaison des deux méthodes sera réalisée sur certains actes. En effet, il a été choisi de modéliser un acte par poste afin de pouvoir observer et comparer les résultats sur chacun des postes. Les actes modélisés ont été sélectionnés selon leurs importances au sein du poste en nombre de consommants et en montant de prestations.

La restitution de ces travaux s'organise en 5 parties afin de répondre à la problématique de construction et d'amélioration des normes tarifaires pour des produits collectifs d'assurance santé.

Dans un premier temps, le système de santé en France sera présenté. Il y sera détaillé l'assurance santé, la Sécurité sociale ainsi que la complémentaire santé.

Dans une deuxième partie, l'aspect théorique de la tarification en santé sera étudié en présentant également les modèles de tarification utilisés en assurance santé.

La troisième partie du mémoire sera consacrée à la présentation de la base de données ainsi que des premières statistiques descriptives. En effet, une analyse statistique des données a été réalisée afin d'analyser la population de cette étude ainsi que la consommation des bénéficiaires selon plusieurs variables.

Dans une quatrième partie, il sera présenté les modèles de tarification utilisées dans ce mémoire. La méthode de tarification utilisée dans le cadre de la construction des dernières normes tarifaires va ainsi être analysée avant de présenter la nouvelle méthode testée sollicitant l'utilisation de modèles GAM avec l'outil Akur8.

Enfin, dans une dernière partie, la construction des jeux de normes tarifaires par acte avec les résultats de ces modèles et la comparaison entre les différentes méthodes seront présentés.

Chapitre 1

Contexte : Le système de santé en France

1.1 Présentation de l'assurance santé

Afin de comprendre toutes les spécificités ainsi que les raisons de ce mémoire, il est essentiel de bien analyser le contexte de l'assurance santé.

1.1.1 Le risque santé

Il s'agit d'un risque provenant du fait qu'un assuré va engager des dépenses pour se soigner lorsque son état physique et/ou moral se détériore. Le risque santé est couvert par une partie obligatoire avec l'Assurance Maladie et par la couverture complémentaire.

L'assurance santé peut être classée dans la catégorie des assurances de personnes. En effet, les assurances de personnes sont les assurances qui protègent les assurés contre les risques affectant directement leur intégrité.

1.1.2 L'économie de la santé

L'économie de la santé est particulièrement spécifique. D'une part, le professionnel de santé n'est pas un producteur essayant de minimiser ses coûts. D'autre part, il y a une incertitude de comportement et de lien entre le professionnel de santé et le patient. Cette incertitude ainsi que les situations

d'asymétrie d'information entre assureur et assuré et entre malades et médecins induisent une limite des marchés concurrentiels pour le financement de la couverture du risque maladie. Ils justifient également l'existence de dispositifs pour réguler la relation patient-médecin.

Par ailleurs, il existe un risque moral en économie de la santé. Pour une même pathologie, un assuré va généralement dépenser plus qu'un non-assuré. En effet, un assuré aura tendance à ne pas être autant attentif au prix des soins qu'un non-assuré et à adopter un niveau de consommation plus élevé car il va se faire rembourser une partie de ces soins.

Pour avoir une vision plus précise de l'économie de la santé, il est intéressant d'analyser les facteurs de l'offre et de la demande de cette économie.

Facteurs de la demande

Un accroissement de la demande des soins de santé est constaté et il peut être expliqué par plusieurs facteurs. Dans un premier temps, le vieillissement de la population est un de ces facteurs et est généralement estimé par l'accroissement de la part des plus de 60 ans dans l'ensemble de la population :

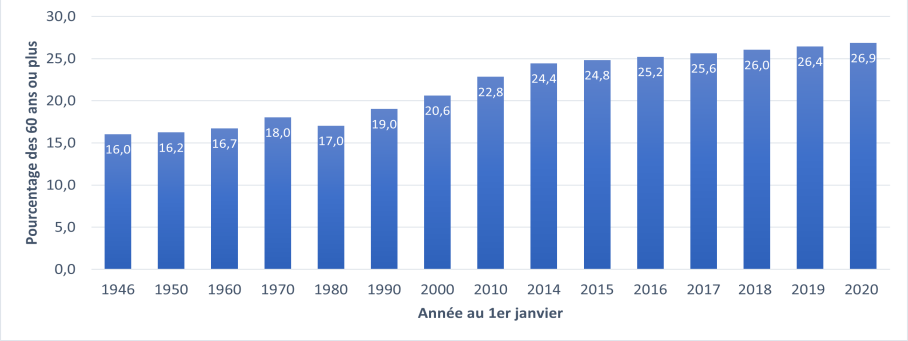


FIGURE 1.1 – Part des plus de 60 ans au sein de la population en France métropolitaine (Source : Insee)

Les conséquences du vieillissement sur l'augmentation de la dépense de santé sont complexes, ils résultent à la fois de phénomènes de génération, de l'allongement de l'espérance de vie et de la stagnation de l'espérance de vie en bonne santé. En effet, les nouvelles générations de personnes âgées

consomment relativement de plus en plus de soins et sur une durée plus longue.

Autre facteur explicatif de l'accroissement de la demande, le niveau de vie. En effet, à mesure que les revenus augmentent, les consommateurs exigent davantage de soins de santé.

Facteurs de l'offre

En parallèle, une augmentation de l'offre est constatée. Elle peut être expliquée par deux principaux facteurs.
Le premier concerne le **progrès médical**. L'application des découvertes scientifiques implique l'utilisation de techniques de pointe engendrant des répercussions sur l'économie de la santé. L'utilisation de techniques nouvelles créent en effet des alternatives pouvant majorer les coûts ou bien apporter une possibilité de traitement qui n'existait pas auparavant.

Le deuxième facteur est la **demande induite** qui correspond à la mesure dans laquelle un médecin peut recommander et imposer une prestation médicale différente de celle que le patient choisirait s'il détenait la même information que lui. La relation entre le médecin et le patient implique que ce dernier délègue au médecin son pouvoir de décision. Les médecins poursuivant l'objectif d'un revenu-cible peuvent inciter les patients à consommer davantage de soins.

Inflation médicale

Les facteurs de l'offre et de la demande, qui ont été analysés précedemment, permettent d'expliquer le phénomène de l'inflation médicale. En effet, l'inflation médicale correspond à l'augmentation, chaque année, du coût des dépenses de santé. En général, cette inflation concerne l'hospitalisation, le tarif d'une consultation ou encore les prix des médicaments.

En moyenne, l'inflation médicale se situe entre 6% et 8% mais elle peut être différente selon le pays concerné. Dans le rapport « Global Medical Trends Survey Report » de Willis Tower Watson en 2021, les chiffres de l'inflation médicale en fonction de la zone géographique par année sont répertoriés dans le tableau suivant :

Zone géographique	2017	2018	2019	2020	2021
Asia pacific	7,1 %	6,8 %	7,5 %	6,2 %	8,5 %
Europe	4,4 %	4,3 %	5,6 %	4,2 %	5,8 %
Latin america	11,1 %	11,6 %	10,8 %	9,0 %	13,6 %
MiddleEast/Africa	8,5 %	7,9 %	8,7 %	8,7 %	10,0 %
North America	11,0 %	11,1 %	5,6 %	2,8 %	7,1 %
Global	6,7 %	6,5 %	7,2 %	6,8 %	8,1 %

Tableau 1.1 - Inflation médicale annuelle en fonction de la zone géographique (Source : Willis Tower Watson : Global Medical Trends Survey Report)

Chaque année, les frais médicaux et l'inflation médicale augmentent plus rapidement que l'inflation normale. Ce fait s'explique par les points vus précédemment, c'est-à-dire le progrès technologique, la surfacturation des fournisseurs de santé ou encore l'augmentation et le vieillissement de la population. L'inflation médicale est donc un phénomène à prendre en considération dans la tarification de contrats d'assurance.

1.1.3 Système beveridgien et système bismarckien

Traditionnellement en Europe, deux systèmes d'assurance sont mis en opposition.

Les systèmes d'assurance sociale Bismarckiens

C'est le chancelier allemand Otto Bismarck, qui mit en place en 1883 un système de protection sociale pour les travailleurs et leurs familles. Dans le cadre du système Bismarckien, les prestations sont proportionnelles aux cotisations : c'est le principe contributif.

Les principales caractéristiques du système bismarckien sont :

- Le remboursement de prestations données par des praticiens librement choisis
- L'ouverture de droits aux prestations sur la base de la qualité de cotisant de l'intéressé
- Le caractère obligatoire et national de l'assurance

Les systèmes nationaux de santé Beveridgiens

En ce qui concerne le système beveridgien, il a été introduit par l'économiste anglais William Beveridge dans un rapport datant de 1942. Constatant que le système d'assurance maladie s'est développé sans réelle cohérence, il propose de le refonder sur plusieurs principes qui seront les fondements du système "beveridgien". Ces principes fondateurs sont les suivants :

- Universalité : couverture pour tous les citoyens contre tous les risques
- Uniformité : une aide identique pour tous sous la forme de prestations espèces
- Unicité : tous les risques sont couverts par un système unique

Les deux types de systèmes sont toutefois de moins en moins différents qu'à l'origine. En France, par exemple, la Sécurité sociale française se caractérise par un système mixte essayant de prendre les avantages des deux modèles.

1.2 La Sécurité sociale

Le système actuel de Sécurité sociale a pour objectif de couvrir l'ensemble des actifs, qui peuvent être toutefois sous différents régimes, des conséquences d'événements ou de situations diverses qui surviennent tout au long de la vie.

1.2.1 Présentation globale

L'ordonnance du 4 octobre 1945 instaure un système de Sécurité sociale basé sur un réseau coordonné de caisses. Composée de cinq branches, elle accompagne l'ensemble des actifs tout au long de la vie.

Les branches de la Sécurité sociale

Pour mettre en œuvre cette solidarité nationale et assurer le bon fonctionnement du service public, le régime général de la Sécurité sociale, qui concerne plus de 80% de la population française, se décompose en 5 grandes branches :

- **La branche Famille** aide les familles dans leur vie quotidienne et développe la solidarité envers les personnes vulnérables. Cette branche

est gérée par la Caisse Nationale d'Allocations Familiales (CNAF) qui verse de très nombreuses allocations.

- **La branche Maladie** assure la prise en charge des dépenses de santé des assurés et garantit l'accès aux soins. Son champ d'action couvre non seulement la maladie mais aussi la maternité, l'invalidité et le décès. Elle est gérée par la Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS) et couvre environ 4 Français sur 5.
- **La branche Accidents du travail-maladies professionnelles** gère les risques professionnels auxquels sont confrontés les travailleurs : accidents du travail, accidents de trajet et maladies professionnelles. À cet égard, elle indemnise les victimes et fixe la contribution respective des entreprises au financement du système. La branche AT-MP est gérée par la Caisse nationale d'assurance maladie (Cnam) et par les caisses d'assurance retraite et de la santé au travail (Carsat) en région.
- **La branche Retraite** verse les pensions aux retraités de l'Industrie, des services et du commerce. Elle suit les salariés tout au long de leur carrière et les aide à préparer leur retraite. La branche Retraite joue également un rôle en matière de prospective et de recherche sur le vieillissement afin d'apporter un éclairage aux pouvoirs publics.
- **La branche Autonomie**, créée en janvier 2021, gère les dépenses liées à l'autonomie des personnes âgées et des personnes handicapées. La gestion de cette cinquième branche est confiée à la Caisse nationale de solidarité pour l'autonomie (CNSA).

De plus, **la branche Recouvrement** (Urssaf), à la différence des autres, ne gère pas un risque mais collecte les cotisations et contributions sociales pour les redistribuer au bénéfice des autres branches. La branche Recouvrement assure ainsi la gestion de la trésorerie de la Sécurité sociale. Elle est gérée par l'Agence centrale des organismes de Sécurité sociale (Acoss).

Dans le cadre de ce mémoire, il a été fait le choix de se focaliser sur la branche Maladie.

Les graphiques ci-dessous synthétisent la répartition des recettes et des dépenses par branche pour l'ensemble des régimes obligatoires de base :

FIGURE 1.2 – Répartition des recettes et des dépenses par branche (Source : PLFSS 2020)

En ce qui concerne le financement de la Sécurité sociale, son budget est encadré par la loi de « financement de la Sécurité sociale » (LFSS). Ses ressources se composent principalement des cotisations sociales qui représentent environ 58% des recettes, de la Contribution Sociale Généralisée (CSG) qui s'applique sur différents revenus et est prélevée à la source sur les revenus ainsi que des impôts, taxes et autres contributions sociales.

Les principes fondateurs de la Sécurité sociale ont été conservés depuis sa création à l'après-guerre même si son système a été amené à évoluer.

1.2.2 Histoire de la Sécurité sociale

Afin de bien comprendre son fonctionnement, il est pertinent de connaître ses origines et comment la Sécurité sociale telle qu'elle existe aujourd'hui s'est forgée.
L'apparition de la Sécurité sociale dès la fin du 19e siècle est liée à la première loi d'assurance sociale sur les accidents du travail. Depuis, elle a beaucoup évolué pour s'adapter aux besoins.

Les premières réformes à la fin du 19e siècle

Le premier système complet d'assurances sociales a été créé en Allemagne à l'initiative du Chancelier Bismarck entre **1881 et 1889**.

En 1898, la loi du 8 avril assure la protection contre les accidents du travail

des salariés de l'industrie en France. Le régime de responsabilité civile est alors modifié en reconnaissant la responsabilité de l'employeur.

Les réformes du 20e siècle

Le 5 avril 1928, le premier système complet et obligatoire d'assurances sociales pour les risques maladie, maternité, invalidité, vieillesse et décès a été créé au bénéfice des salariés de l'industrie.

En 1945, les ordonnances des 4 et 19 octobre permettent la création du système de Sécurité sociale en France sur le modèle « bismarckien » ainsi que la refonte du système des assurances sociales des années trente.

En 1946, la loi du 22 août étend les allocations familiales à pratiquement toute la population et la loi du 30 octobre 1946 intègre la réparation des accidents du travail à la Sécurité sociale.

Enfin, la loi du **27 juillet 1999** crée la Couverture maladie universelle (CMU) qui est un dispositif permettant à une personne qui réside en France depuis au moins trois mois et qui n'est pas couverte par un régime obligatoire d'assurance maladie de bénéficier de la Sécurité sociale française pour ses dépenses de santé. La CMU a été remplacée en 2016 par la protection universelle maladie (PUMA) qui a assoupli les conditions pour bénéficier d'une prise en charge des frais de santé pour une personne qui ne travaille pas.

Réformes récentes des années 2000

En 2004, la loi du 13 août portant sur la réforme de l'Assurance maladie amène la création du médecin traitant, la création du dossier médical personnel, la réforme de la gouvernance de l'Assurance maladie et du système de santé, la promotion des médicaments génériques, l'aide à l'acquisition d'une couverture complémentaire ainsi que la responsabilisation des assurés sociaux par la création du forfait d'1 euro.

En 2015, l'Assemblée nationale adopte le projet de loi santé instaurant la généralisation du tiers payant à tous les assurés.

En 2018, le régime social des indépendants est rattaché au régime géné-

ral de la Sécurité sociale.

La réforme 100% Santé

C'est également en 2018 que la réforme 100% Santé a été lancée par le gouvernement d'Emmanuel Macron. Cette réforme permet d'améliorer l'accès à des soins de qualité et de renforcer la prévention.

Depuis le 1er janvier 2021, tous les Français bénéficiant d'une complémentaire santé responsable ou de la complémentaire santé solidaire ont accès à des soins et à un large choix d'équipements en audiologie, optique et dentaire qui sont pris en charge à 100% par la Sécurité sociale et les complémentaires.

Suite à cette réforme, plusieurs paniers de soins sont proposés pour chaque poste : un panier sans reste à charge appelé « panier 100% santé », un panier maitrisé avec des tarifs plafonnés et un panier libre avec des tarifs qui ne sont pas encadrés. Plus concrètement, l'assuré bénéficiant d'un contrat responsable pourra choisir entre le panier 100% santé et le panier libre. Il faut donc prendre en considération l'impact de cette réforme dans la tarification.

Le système français de Sécurité sociale se caractérise donc aujourd'hui par une protection contre les risques sociaux généralisée à l'ensemble de la population qui est le fruit d'une longue histoire.

1.2.3 Les régimes de base

En 1945, une Sécurité sociale unique devait être instaurée. Cependant, certaines professions ou corps sociaux avaient déjà mis en place leur propre système de protection sociale et ils ont alors préféré le conserver.

Par conséquent, bien que tous les Français bénéficient de la Sécurité sociale, tout le monde n'est pas couvert par le même régime ni de la même façon. Cela dépend de sa situation personnelle ainsi que du secteur dans lequel l'assuré travaille. Ainsi, il existe le régime général et d'autres régimes de Sécurité sociale, selon le secteur d'activité auquel l'assuré est rattaché.

Le régime général

Le régime général concerne les salariés du secteur privé ainsi que les travailleurs indépendants et couvre environ 80% de la population française. Il est composé de 5 branches, qui couvrent les grands risques et gèrent le recouvrement des cotisations.

Le régime agricole

Le régime spécifique agricole accompagne les exploitants, les salariés agricoles et les entreprises agricoles. Il couvre environ 5% de la population française.

Le Régime agricole (Mutualité sociale Agricole) est constitué d'une seule entité qui gère à la fois les prestations d'assurance maladie, accidents du travail et maladies professionnelles, retraite et famille. La Mutualité sociale Agricole gère elle-même le recouvrement de ses cotisations.

Les régimes spéciaux

Les régimes spéciaux propres aux salariés des grandes entreprises publiques. Les mêmes risques sont couverts mais d'une manière différente pour la majorité. Ils regroupent les fonctionnaires, la SNCF, EDF-GDF, les employés et clercs de notaires, les mines ou encore les cultes. Ces régimes spéciaux sont au nombre de 27 et couvrent 7% de la population française.

Le régime Alsace-Moselle

Dans le cadre de la tarification santé, il convient de prendre en compte l'existence d'un régime local d'assurance maladie en Alsace-Moselle.

En 1884, l'Alsace et la Moselle, alors annexées par l'Allemagne, ont bénéficiés des réformes de Bismarck concernant la protection sociale. En 1945, lors de la création de la Sécurité sociale, les Alsaciens et les Mosellans ont refusés de rejoindre le régime général, préférant leur régime. Ils bénéficient donc depuis d'un régime spécifique, géré de façon autonome depuis 1995.

Le régime Alsace-Moselle offre des remboursements plus importants que le régime général. En contrepartie, les assurés de ce régime payent des montants

de cotisation plus élevés que dans le reste de la France. Les complémentaires prennent donc en charge une part moins importante des remboursements. Ces meilleurs remboursements permettent aux assurés de la région Alsace-Moselle de souscrire une mutuelle moins chère que les assurés du régime général. En effet, la complémentaire santé sera moins sollicitée pour compenser les remboursements d'une Sécurité sociale plus avantageuse.

1.2.4 Le déficit de la Sécurité sociale

En France, le déficit de la Sécurité sociale est un sujet récurent depuis plus d'une vingtaine d'années. Ce déficit implique un besoin de financement complémentaire qui est assouvi par emprunt, contribuant à la dette de la Sécurité sociale appelée communément « trou de la sécu ».

En 2021, le déficit de la Sécurité sociale n'a jamais été aussi élevé depuis 2001 tandis qu'en 2018, le déficit avait diminué en s'élevant à 1,2 milliards d'euros. Effectivement, face au choc sanitaire du Covid-19, le déficit des comptes de la Sécurité sociale en 2021 a atteint 34,6 milliards d'euros d'après un rapport la Commission des comptes de la Sécurité sociale en septembre 2021.

La branche maladie est la branche qui a vu son déficit s'accroître le plus en 2021 du fait notamment des mesures prises pour faire face à la pandémie. La crise sanitaire a donc entrainée des conséquences financières très importantes sur l'Assurance maladie avec une perte de 30 milliards d'euros en 2020.

En 2022, le déficit de la Sécurité sociale va se réduire à 16,8 milliards d'euros selon un rapport de la Commission des comptes de la Sécurité sociale, soit 4,8 milliards de moins que prévu. Cependant, majoritairement en raison de la crise sanitaire, la branche maladie est encore largement déficitaire (-19,7 milliards).

En effet, sur le graphique prévisionnel ci-dessous, il est constaté que le déficit de la Sécurité sociale diminuera au fil des années. De plus, le décicit de la branche maladie devait baisser de 34,6 Milliards à 21,6 Milliards. Selon le PLFSS, le remboursement de la dette sociale doit être assuré par le prolongement de la Caisse d'amortissement de la dette sociale (CADES) jusqu'en 2033 :

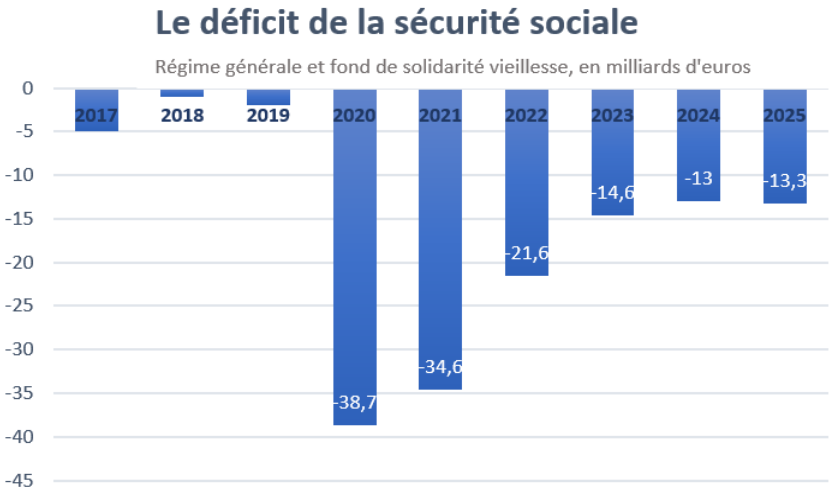


FIGURE 1.3 – Le déficit prévisionnel de la Sécurité sociale (Source : PLFSS 2022)

1.2.5 Les postes de consommation de soins et biens médicaux

Après avoir expliqué la Sécurité sociale et son fonctionnement, il est intéressant d'appréhender son activité à travers les différents postes de consommation de soins et biens médicaux.

La branche santé se divise en plusieurs postes que l'on peut regrouper dans des catégories :

- Les actes médicaux : les consultations chez le généraliste ou chez un spécialiste, les auxiliaires médicaux, etc.
- La pharmacie : les médicaments, les vaccins, etc.
- L'hospitalisation : les frais de séjour, les honoraires , les forfait hospitalier, la chambre particulière, les frais accompagnant, etc.
- L'optique : les lentilles, les lunettes, la chirurgie de l'œil, etc.
- Les soins dentaires : les prothèses dentaires, les implants dentaires, la parodontologie, l'orthodontie, etc.
- Les soins audio : les prothèses auditives, etc.

- Des soins et prestations divers : les frais de transports, les cures thermales, l'orthopédie, les médecines douces, la maternité, les frais obsèques, etc.

Pour la consommation de soins et biens médicaux, ces postes sont classés différemment :

- Les soins hospitaliers
- Les soins ambulatoires
- Les transports de malades
- Les médicaments
- Les autres biens médicaux

La répartition de la CSBM en 2020 par poste est représentée sur le graphique ci-dessous :

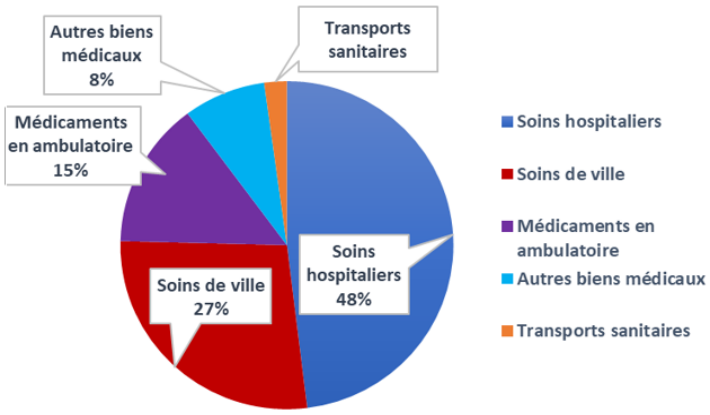


FIGURE 1.4 – Répartition de la consommation de soins et biens médicaux en 2020 en pourcentage (Source : DREES)

Les postes "soins hospitaliers" et "soins de ville" représentent plus de 70% de la CSBM (consommation de soins et biens médicaux) en montant de frais réels engagés. De plus, le poste "Transport sanitaires" est le moins représenté dans la CSBM.

Le graphique ci-dessous indique que la part des soins hospitaliers dans la CSBM a reculé entre 1985 et 2000. A l'inverse, les dépenses de transports sanitaires ont très fortement augmentés (de 4% à 7% de la CSBM) du fait notamment de l'accroissement rapide de leurs prix depuis 1985 (+3% en moyenne d'après DREES) :

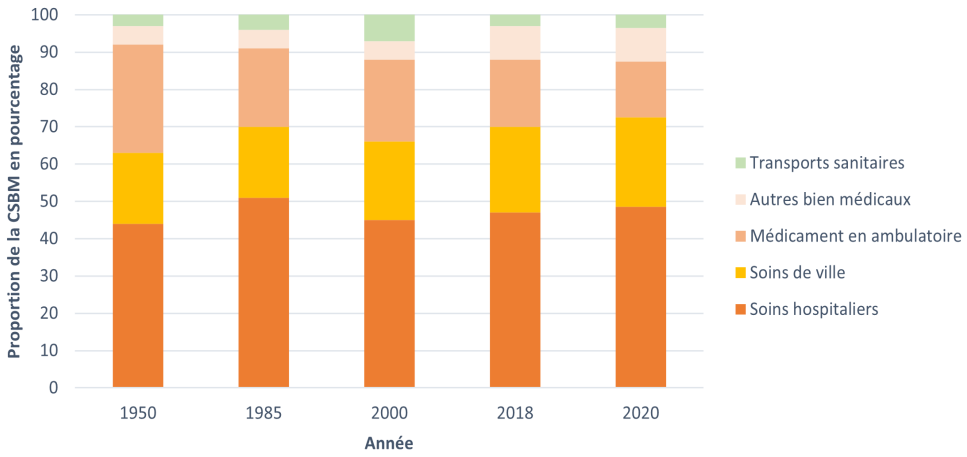


FIGURE 1.5 – Répartition de la CSBM de 1950 à 2020 (Source : DREES)

De même, la part des médicaments a nettement diminué entre 2000 et 2020 (-7 points) alors que les autres biens médicaux, les transports sanitaires et les soins de ville ont vus leurs parts s'accroîtrent légèrement.

Et en 2020, la CSBM a connu une restructuration due à la crise sanitaire de la Covid avec notamment une hausse de la part des soins hospitaliers (+1,5 points).

1.2.6 Le mécanisme de remboursement des frais de santé

Dans cette partie, les mécanismes de remboursement des assurés lors de la consommation d'un acte médical seront présentés. La répartition de la prise en charge du remboursement entre le régime de base et la complémentaire sera également détaillée.

Le schéma suivant illustre comment se décompose le remboursement d'un acte :

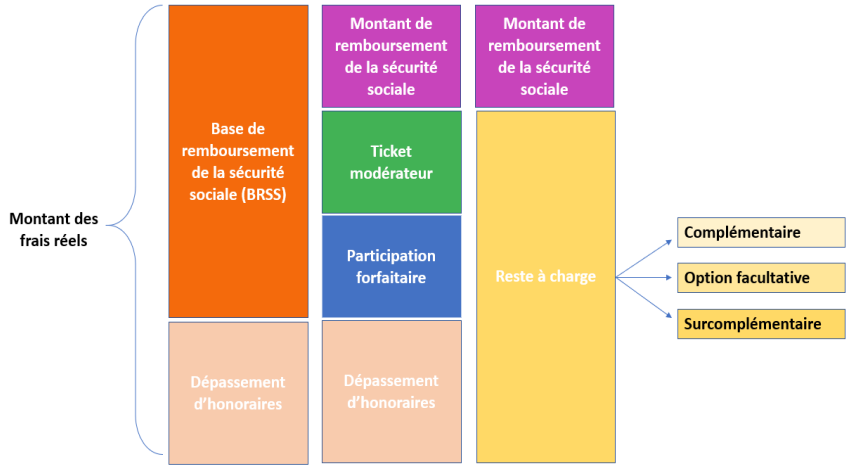


FIGURE 1.6 – Schéma récapitulatif du remboursement d'un acte

Le principe de remboursement

Pour chaque acte médical, l'Assurance Maladie fixe :

- une base de remboursement de la Sécurité sociale (BR) qui correspond à son tarif de référence exprimé en euros. Il s'agit du montant limite et non des frais réels, soit le montant dépensé par l'assuré.
- un taux de remboursement qu'elle applique à ce tarif de référence exprimé en pourcentage et il détermine le montant que l'assurance maladie devra rembourser à l'assuré.

Les composantes du remboursement d'un acte

Le montant de remboursement de la Sécurité sociale est calculé à partir du taux de remboursement qui varie en fonction des actes et médicaments, de la santé de l'assuré, du respect ou non du parcours de soins coordonnés. En effet, les personnes souffrant d'affection longue durée (ALD) ont des taux de remboursement correspondant à 100% de la base de remboursement de la Sécurité sociale sur les actes liés à leur affection.

En France, afin de préserver le système de santé, une participation forfaitaire d'1 € est demandée si l'individu est âgé de plus de 18 ans depuis le 1er janvier de l'année en cours. Elle s'applique aux consultations et actes réalisés par un médecin, ainsi qu'aux examens radiologiques et analyses de biologie médicale.

Ensuite, des dépassements d'honoraires qui ne sont pas pris en charge par l'Assurance Maladie peuvent s'y ajouter. L'assurance complémentaire santé intervient donc pour prendre en charge totalement ou partiellement le ticket modérateur et les dépassements d'honoraires. Pour compléter ces remboursements, l'assuré peut également choisir de souscrire à une option facultative ou à une surcomplémentaire. Enfin, le reste à charge correspond à ce que l'assuré devra payer de sa poche.

1.3 Le régime complémentaire

En complément des remboursements des régimes obligatoires d'assurance maladie, les différentes structures d'assureurs proposent des contrats d'assurance complémentaire santé.

Ces garanties couvrent totalement ou partiellement la partie restant à la charge de l'assuré et ont pour but de réduire ce reste à charge suite au remboursement de l'assurance maladie. Certaines complémentaires prennent également en charge des prestations qui ne sont pas remboursées par l'Assurance Maladie telles que les médecines douces, les implants dentaires ou encore la chirurgie de la myopie.

En France, de nombreux acteurs sont sollicités par ce système de remboursement des soins.

1.3.1 Les acteurs

Le marché de la complémentaire santé en France est représenté par différents types de structures : les mutuelles, les sociétés d'assurance et les institutions de prévoyance. En 2019, le marché de la complémentaire santé se répartissait de la façon suivante :

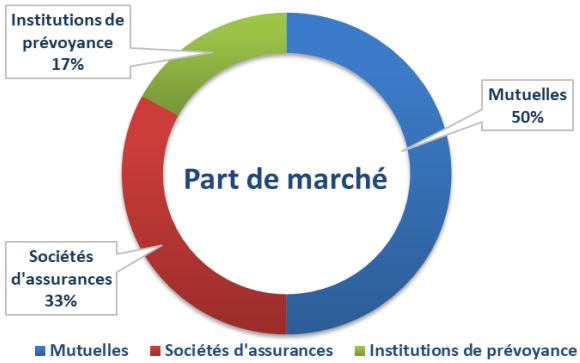


FIGURE 1.7 – Répartition du marché de la complémentaire santé en 2019 en pourcentage des cotisations collectées (Source : Fonds CSS))

Les mutuelles

Représentant 50% du marché de la complémentaire, elles sont aussi appelées les " mutuelles 45 " car elles se sont développées en 1945 avec la création de la Sécurité sociale.

Régies par le Code de la Mutualité, les mutuelles sont des sociétés de personnes à but non lucratif qui se financent essentiellement par le biais des cotisations de leurs membres.

Leur mission est d'organiser pour et avec leurs adhérents les réponses aux besoins sociaux qu'ils expriment. L'activité des mutuelles 45 se situe exclusivement dans les produits d'assurance de personne. Les remboursements maladie représentent 67% de l'activité totale selon l'institut national de recherche et documentation en économie de la santé (IRDES).
Elles sont dépendantes du contrôle du Ministère des Affaires sociales et au contrôle de la commission de contrôle des mutuelles et des institutions de prévoyance.

Les sociétés d'assurances

Les compagnies d'assurances représentent 31% des parts du marché en 2020. Il s'agit du deuxième plus grand acteur du marché.

Régies par le Code des Assurances, les sociétés d'assurance peuvent exis-

ter sous la forme de société anonyme avec un statut de société commerciale ou bien de sociétés d'assurances mutuelles qui sont des sociétés à but non lucratif.

Elles ont un but lucratif et les actionnaires sont les assurés. Ainsi, les bénéfices réalisés sont redistribués aux assurés. De plus, les sociétés d'assurance peuvent avoir recours à des intermédiaires tels que des courtiers ou des agents pour diffuser leurs contrats d'assurance.

Les institutions de prévoyance

Les institutions de prévoyance représentent 17% du marché de la complémentaire santé.

Régies par le code de la Sécurité sociale, les institutions de prévoyance sont des organismes paritaires à but non lucratif. Elles gèrent essentiellement les contrats collectifs souscrits par les entreprises pour leurs salariés. AG2R la mondiale fait partie de ces institutions de prévoyance.

Elles proposent des garanties qui peuvent prendre la forme de capitaux décès, de rentes pour les conjoints survivants, de rentes pour l'éducation des enfants, de remboursements de frais de soins en santé ou encore de rentes en cas de dépendance.
Etant donné que ce sont des organismes à but non lucratif, les institutions de prévoyance n'ont pas d'actionnaires à rémunérer. Ainsi, ils investissent leurs bénéfices dans l'objectif d'améliorer leurs garanties ainsi que leur activité sociale.

1.3.2 Les différents type de contrats

En assurance santé, il existe, d'une part, les contrats individuels qui concernent les particuliers et, d'autre part, les contrats collectifs qui sont souscrits par les entreprises pour leurs employés.

Les contrats collectifs

Un contrat collectif est établi sur la base d'une relation tripartite. Une personne morale ou un chef d'entreprise souscrit le contrat auprès de l'un

des trois types d'assureurs. Les contrats de groupe peuvent être facultatifs ou obligatoires.

La plupart des régimes d'entreprise sont des régimes collectifs à adhésion obligatoire. Depuis le 1er janvier 2016, les entreprises du secteur privé doivent proposer à tous leurs salariés une couverture santé complémentaire. L'Accord National Interprofessionnel (ANI) du 14 juin 2013 est à l'origine de cette règlementation. L'employeur doit financer au moins 50% du montant total de la cotisation. Une complémentaire obligatoire d'entreprise peut être mise en place soit par accord collectif, soit par référendum au sein de l'entreprise, soit par décision unilatérale du chef d'entreprise.

Le fait que les nouveaux arrivants de l'entreprise soient obligés de souscrire à la complémentaire permet ainsi d'éviter l'antisélection. En effet, lorsqu'un assuré a le choix de s'assurer ou de ne pas s'assurer en fonction de ses besoins, il y a « antisélection ». Par exemple, si l'assuré a des problèmes de vues, il va privilégier les assurances avec une garantie en Optique.

L'entreprise peut également prévoir des garanties optionnelles à adhésion facultative pour les salariés qui souhaiteraient bénéficier de garanties améliorées en plus du contrat obligatoire. Le salarié décide alors d'adhérer au régime ou non.

Les contrats individuels

Les contrats individuels apportent une solution aux personnes qui ne se considèrent pas suffisamment couvertes ou ne sont pas couvertes par un contrat d'entreprise, ou aux anciens salariés après leur départ en retraite. Ils permettent de choisir spécifiquement le niveau de couverture qu'on souhaite avoir. En contrepartie, les contrats individuels bénéficient de tarifs moins avantageux que les complémentaires santé collectives.

Dans le cadre de ce mémoire, la tarification a été établie sur des contrats collectifs.

1.3.3 Le contrat responsable

Afin de remédier au déficit de l'Assurance maladie, la réforme de 2006 a créé la notion de contrat « responsable ». Ils sont applicables, entre autres, aux contrats complémentaires santé individuels souscrits ou renouvelés depuis le 1er avril 2015.

L'objectif d'un contrat responsable est de responsabiliser les patients afin de réduire le déficit de l'Assurance maladie. Le contrat responsable encourage à consulter un médecin traitant unique et déclaré et à respecter le parcours de soins coordonnés. Ce parcours consiste à confier à un médecin traitant la coordination des soins pour votre suivi médical. Il garantit une meilleure prise en charge des dépenses de santé.

L'ensemble des assurés de la Sécurité sociale âgés de 16 ans ou plus est concerné par le parcours de soin. Son objectif est de réduire les consultations dispensables et les examens redondants et donc les dépenses de santé. Pour répondre à cet objectif, le contrat responsable impose des garanties minimales.

Le panier de soins minimal initialement instauré propose la prise en charge du ticket modérateur en intégralité sur les actes remboursées par l'Assurance maladie ainsi que la totalité du forfait journalier hospitalier.

Depuis 2015, un contrat responsable doit respecter certaines obligations supplémentaires :

- L'organisme assureur ne doit recueillir aucune information médicale auprès de l'assuré
- Les cotisations ou les primes ne sont pas fixées en fonction de l'état de santé de l'assuré
- La prise en charge de deux actes de prévention parmi une liste définie le 8 juin 2006. Le choix d'AG2R La Mondiale s'est porté sur le détartrage et le vaccin DTP dans le but de toucher une population plus large.

Plus récemment, le décret du 11 janvier 2019 a modifié le cahier des charges des contrats responsables. La réforme « 100% santé » a fixé de nouveaux niveaux de prise en charge de ces contrats sur des postes de dépense tels que l'optique, le dentaire et l'auditif.

En effet, toute complémentaire santé collective basée sur un contrat responsable est désormais concernée par la réforme 100% santé. Pour les contrats responsables, le reste à charge est alors supprimé pour certains soins et équipements optiques, dentaires et auditifs.

De plus, afin de réduire les dépassements d'honoraires, la convention médicale du 26 août 2016 remplace le contrat d'accès aux soins (CAS) par l'option pratique tarifaire maitrisée (OPTAM). Par conséquent, les dépassements d'honoraires peuvent être pris en charge dans la limite de 100% de la base de remboursement pour les médecins non adhérents à l'OPTAM et sans limite pour les médecins adhérant à l'OPTAM. La différence de prise en charge entre OPTAM et non OPTAM doit au moins être égale à 20 % de la base de remboursement.

1.3.4 Les différentes façons d'exprimer la garantie

Enfin, en assurance santé, il existe plusieurs manières d'exprimer les garanties d'un contrat d'assurance :

- En pourcentage du Ticket Modérateur : dans le cadre d'un remboursement « 100% du TM ».
- En pourcentage des Frais réels : l'assuré se fera rembourser la différence entre ses frais et le remboursement de la SS plus la participation forfaitaire dans la limite d'un plafond.
- En pourcentage du remboursement de la Sécurité sociale (RSS)
- En pourcentage de la base de remboursement de la Sécurité sociale
- Remboursement à caractère forfaitaire pouvant être un montant en euro ou un pourcentage du PMSS correspondant au plafond de la Sécurité sociale qui est utilisé comme base de calcul de certaines prestations sociales.

Ces garanties peuvent s'exprimer "y compris Sécurité sociale" ou bien "en plus de la Sécurité sociale". De plus, ces garanties ont parfois un plafond maximal annuel ou journalier et des franchises. Les niveaux de garanties doivent être pris en considération lors de la tarification d'un contrat complémentaire.

Chapitre 2

Tarification d'un contrat d'assurance santé complémentaire

La théorie ainsi que les méthodes usuelles du processus de la tarification d'un contrat d'assurance santé complémentaire vont être présentés dans cette partie.

2.1 Principes de la Tarification générale en Santé

La tarification d'un contrat d'assurance santé s'effectue en deux étapes. Dans un premier temps, la prime pure technique payée par l'assuré afin de financer les sinistres survenus est déterminée. Dans un second temps, il faut procéder à la tarification commerciale en tenant compte des chargements et des taxes.

Concernant la tarification technique, il est nécessaire de considérer tous les facteurs qui peuvent influencer la tarification tels que les spécificités du contrat, les garanties, les caractéristiques de l'entreprise et de chaque assuré comme l'adresse des bénéficiaires ou l'âge moyen des assurés. En assurance santé, la prime est généralement calculée selon un modèle « Fréquence X Coût Moyen ».

2.2 Le modèle « Fréquence X Coût Moyen »

La prime pure payée par l'adhérent sur une période est l'estimation de sa consommation moyenne sur cette même période. Pour déterminer cette prime pure, la méthode « Fréquence X Coût Moyen » est couramment utilisée. Dans cette partie, il va être détaillé en quoi consiste théoriquement cette méthode. Premièrement, on pose S le montant de prestations de l'assuré sur une période donnée et π la prime pure. L'objectif est de minimiser l'erreur quadratique moyenne : $MSE = E((S - \pi)^2)$

Pour cela, deux hypothèses doivent être établies :

- Soit $X_i, i \in \mathbb{N}$ la suite des prestations indépendantes et identiquement distribuées.
- Indépendance entre le nombre de prestations et les coûts des prestations

Ainsi, la charge de prestations de l'adhérent peut être décomposée en montants de prestations individuelles :

$S = \sum_{i=1}^{N} X_i$ avec N la variable aléatoire du nombre de sinistres.

Il faut désormais déterminer l'espérance de la charge sinistre :

$$E(S) = E(E(S|N)) = E(E(\sum_{i=1}^{n} X_i|N)) \tag{2.1}$$

Une forme simple de l'espérance de la charge de prestations peut être déduite avec les hypothèses qui ont été prises :

$$E(S) = E(N)E(X_1) \tag{2.2}$$

La montant de prestation moyen est ainsi égal au produit du nombre moyen de prestations et du coût moyen d'une prestation. Il est donc possible de calculer la prime pure assez simplement avec cette méthode et c'est la raison pour laquelle elle est très souvent utilisée.

Cependant, cette méthode ne permet pas d'adapter le montant de la prime aux caractéristiques des assurés. En effet, elle permet seulement de déterminer une prime pure mutualisée que tous les assurés vont payer. Par conséquent, les assureurs peuvent ensuite appliquer différents coefficients correctifs

pour adapter leurs tarifs selon les caractéristiques de l'assuré.
Les différents coefficients correctifs appliqués peuvent être liés à l'âge, le sexe, la CSP ou encore la localisation de l'assuré. C'est à l'assureur de choisir le niveau de segmentation appliqué en fonction d'une politique commerciale s'appuyant sur le principe de solidarité ou sur une personnalisation du tarif.

Par ailleurs, cette méthode présente certaines limites dont notamment :

- l'hypothèse d'indépendance et d'identique distribution des $(X_i), i \in \mathbb{N}$ n'est pas toujours vérifiée. Par exemple, il est possible que le coût varie au cours du temps du fait d'augmentations des tarifs médicaux ;
- l'indépendance entre la fréquence et le coût moyen n'est pas toujours vraie. En effet, en assurance santé, un assuré très bien couvert et donc générant des coûts élevés sera plus enclin à avoir une forte fréquence de consommation. Dans ce cas là, le coût et la fréquence seraient donc positivement corrélés.

Cette méthode est régulièrement mise en pratique lors de la tarification de contrats en santé. Par ailleurs, dans le cadre de la tarification en assurance santé, des modèles de machine learning tels que les modèles linéaires généralisés ou les modèles additifs généralisés sont utilisés. Ces modèles vont être présentés dans la partie suivante.

2.3 Les modèles GAM

Cette partie présente, dans un premier temps, la théorie du modèle linéaire classique et du modèle linéaire généralisé avant de présenter en détail les modèles additifs généralisés.

2.3.1 Le modèle linéaire

En régression linéaire, un ajustement linéaire par la méthode des moindres carrés est calculé pour un ensemble de variables X appelés prédicteurs, afin de prévoir une variable dépendante Y. L'équation d'une régression linéaire avec n prédicteurs permettant de prévoir une variable dépendante Y est de

la forme :

$$Y_j = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \epsilon i, i \in [1; n], \beta_i \in R^2 \tag{2.3}$$

Y_j représente les valeurs prévues de la variable dépendante, les coefficients β_i sont les paramètres inconnus du modèle et les quantités ϵ_i sont des variables aléatoires. Elles viennent du fait que les points ne sont jamais parfaitement alignés sur une droite. Elles sont appelées les erreurs (ou bruits) et elles sont supposées aléatoires. Dans le but de pouvoir interpréter correctement et pertinemment le modèle, les hypothèses suivantes doivent être prises :

- $\forall$i, E(ϵ_i)=0 ;
- $\forall$i, Var(ϵ_i)= θ^2 (homoscédasticité des erreurs) ;
- $\forall$i,j, i $\neq j, cov(\epsilon_i,\epsilon_j)$=0 ;

Par ailleurs, il faut toujours prendre implicitement l'hypothèse selon laquelle les x_i ne sont pas tous égaux. Les erreurs sont donc supposées centrées, de même variance et non corrélées entre elles.

L'objectif du modèle linéaire général est d'estimer $\beta_1, \beta_2, .., \beta_n$ sur la base des observations $y_1, ..., y_n$ des variables $Y_1, ..., Y_n$, et de $x_1, ..., x_n$ de façon à expliquer au mieux les variables Y_i en fonction des x_i. Pour estimer les coefficients β_i, les estimateurs des moindres carrés sont généralement utilisés.

Ces hypothèses impliquent que la variable réponse suit une distribution normale. En effet, $\epsilon \sim \mathcal{N}(0, In\sigma^2)$ implique $Y \sim \mathcal{N}(X\beta, In\sigma^2)$. En pratique, ce n'est pas toujours le cas en assurance santé où les coûts des sinistres sont asymétriques et donc non gaussien.

2.3.2 Les modèles linéaires généralisés (GLM)

Les modèles linéaires généralisés (GLM) sont une généralisation des modèles de régression linéaire qui permettent de modéliser une relation non-linéaire entre la variable à expliquer et les variables explicatives. Ils permettent de s'affranchir de l'hypothèse de normalité de Y_i et donc de prendre en compte des variables non nécessairement normales mais appartenant à la famille exponentielle linéaire.

En assurance santé comme dans de nombreux autres domaines de l'assurance, le Modèle Linéaire Généralisé est une méthode de tarification régulièrement utilisée. Ces modèles ont été mis en place pour résoudre certains problèmes des modèles linéaires.

Définition et formalisation du GLM

L'équation du modèle s'exprime de la façon suivante :

$$Y_i = \beta_0 + \sum_{j=1}^{n} \beta_j \times x_{ij} + \epsilon_{i,j} \in [1; n] \tag{2.4}$$

Où :

- Y_i est la variable à expliquer ;
- $X_i, i \in [1; n]$ sont les variables explicatives ;
- $\beta_i, i \in [1; n]$ sont les paramètres du modèle ;
- $\epsilon_i, i \in [1; n]$ est l'erreur (ou bruit) du modèle ;

Il existe, pour tout $i \in [1; n]$, s'il remplit les conditions suivantes :
$E(\epsilon_i) = 0$, $Var(\epsilon_i) = \theta^2$, $Cov(\epsilon_i, \epsilon_j) = 0 \ \forall$ j $\in [1; n]$ et $i \neq j$.

A travers une fonction de lien g, les modèles linéaires généralisés (GLM) permettent de modéliser la relation entre la variable réponse Y et les variables explicatives X :

$$g(E[Y_i|X_i]) = X_i^t \beta \tag{2.5}$$

Les composantes d'un GLM

Les GLM sont caractérisés par trois composantes :

➢ La composante aléatoire qui se définit par la distribution de probabilité de la variable réponse Y.
L'espérance de la variable réponse est celle qui doit être expliquée. Sa loi de probabilité doit appartenir à la famille des lois exponentielles. Ainsi, sa densité s'exprime de la façon suivante :

$$f_{Y_i}(y_i, \theta, \phi) = \frac{exp(y_i\theta_i - b(\theta_i))}{a(\phi)} + c(y_i, \phi) \tag{2.6}$$

avec :

— $\theta i \in \mathbb{R}$: paramètre canonique de la moyenne ;
— $\phi \in \mathbb{R}$: paramètre de dispersion ;
— a fonction définie sur $\mathbb{R}$ non nulle ;
— b fonction définie sur $\mathbb{R}$ deux fois dérivable ;
— c fonction définie sur $\mathbb{R}^2$

Les lois binomiales, de Bernoulli, de Poisson, Normale, Gamma ou Gaussienne inverse sont alors éligibles.

➢ La composante déterministe qui se définit par une fonction linéaire des variables explicatives. La composante déterministe relie le paramètre η aux variables explicatives X :

$$g(\eta) = \beta^t X = \beta_1 \times X_1 + ... + \beta_p \times X_p \tag{2.7}$$

➢ La fonction de lien qui exprime une relation fonctionnelle entre l'espérance mathématique de Y notée η et les variables explicatives X. Il s'agit du lien entre la composante aléatoire et la composante déterministe et il précise la nature de la relation entre l'espérance de la variable réponse et la combinaison linéaire constituée par les variables explicatives.
Soit g la fonction de lien. Elle est définie telle que :

$$g(\mu_i) = \beta_0 + \beta_1 \times x_{i1} + ... + \beta_p \times x_{ip} = x_i'\beta \tag{2.8}$$

où $\mu_i = E(Y_i)$ doit être monotone et dérivable.

Il existe de nombreuses fonctions de lien et les plus couramment utilisées sont :
— Le lien identité : $g(\mu) = \mu$
— Le lien logarithmique : $g(\mu) = ln(\mu)$: Il s'agit alors d'un modèle multiplicatif pour expliquer une variable positive
— Le lien inverse : $g(\mu) = \frac{1}{\mu}$
— La fonction de lien logit : $g(\mu) = log(\frac{\mu}{1-\mu})$. Elle modélise le logarithme du rapport des chances.

Ainsi, le modèle linéaire généralisé diffère du modèle linéaire sur deux aspects majeurs :

— la distribution de la variable dépendante peut être non-normale et elle ne doit pas être nécessairement continue. Elle peut être binomiale par exemple ;

— les valeurs de la variable dépendante sont déterminées à partir d'une combinaison linéaire des variables prédictives qui sont liées à la variable dépendante par une fonction de liaison. Plus précisément, dans le modèle linéaire général, une variable réponse Y est associée de façon linéaire aux valeurs des variables X alors que la relation dans le modèle linéaire généralisé s'exprime de la manière suivante : $Y = g(b_0 + b_1 \times X_1 + ... + b_m \times X_m)$ où g représente une fonction.

Les modèles linéaires généralisés permettent donc de conserver la simplicité des modèles linéaires tout en autorisant une forme plus générale. Toutefois, la procédure d'estimation n'est efficace que si la vraie loi conditionnelle appartient à cette famille exponentielle.

2.3.3 Les modèles additifs généralisés (GAM)

Les Modèles Additifs Généralisés, constituent une généralisation du Modèle Linéaire Généralisé qui est lui-même un cas particulier du modèle linéaire.
Ils offrent la possibilité de choisir parmi une large gamme de distributions pour la variable dépendante telle que la distribution normale, Gamma ou Poisson. Ils permettent aussi de choisir parmi différentes fonctions de liaison pour déterminer les effets des variables prédictives sur la variable dépendante :

- Liaison Log : f(z) = log(z)
- Liaison inverse : f(z) = $\frac{1}{z}$
- Liaison identité : f(z) = z
- Liaison Logit : f(z)=log($\frac{z}{1-z}$)

Les modèles additifs généralisés représentent une combinaison des modèles additifs et des modèles linéaires généralisés. Ils prennent la forme suivante :

$$gi(\mu_Y) = S_i(f_i(X_i)) \tag{2.9}$$

L'objectif des modèles additifs généralisés consiste à maximiser la qualité de la prévision d'une variable dépendante Y à partir de distributions, en estimant des fonctions non-paramétriques des variables prédictives qui sont liées à la variable dépendante par une fonction de liaison.

Pour simplifier, l'objectif n'est pas d'estimer des paramètres simples comme les poids de la régression dans une régression multiple mais de rechercher plutôt, dans les modèles additifs généralisés, une fonction généraliste non-spécifiée qui permet de lier les valeurs prévues Y aux valeurs prédictives.

La plupart des modèles additifs généralisés utilisent les splines. Succinctement, une spline est une fonction polynomiale définie par morceaux. A l'inverse de l'interpolation polynomiale qui ajuste une courbe à travers tous les points de données à la fois, l'interpolation spline rapproche une courbe entre chaque paire proche de points de données et ajoute toutes les courbes ensemble pour créer l'approximation finale. Les splines sont généralement utilisées pour définir les fonctions de lissage dans un GAM.

Les Modèles Additifs Généralisés sont très flexibles et permettent d'obtenir un excellent ajustement en présence de relations non-linéaires et de bruit important dans les variables prédictives. Cependant, du fait de cette flexibilité, il y a un risque de sur-ajuster les données.
En effet, il faut faire attention à ne pas appliquer un modèle trop complexe aux données afin de produire un bon ajustement qui ne sera peut être pas obtenu lors de l'étape de validation. L'enjeu est donc d'obtenir un bon équilibre entre un ajustement satisfaisant sur les données et la complexité ajoutée par l'utilisation des modèles additifs généralisés.

De manière générale, il est souvent préférable d'utiliser un modèle simple et facile à comprendre pour prévoir de nouvelles observations plutôt qu'un modèle complexe difficile à interpréter. En assurance santé, les GAM peuvent être utilisés pour la tarification afin de considérer l'impact des variables explicatives sur le coût moyen et la fréquence des sinistres.

2.4 Validation du modèle

Lorsque la création du modèle est terminée, il faut évaluer sa qualité. Or, la qualité d'un modèle se mesure sur ses performances mais également sur sa complexité. En effet, un modèle très performant mais complexe et difficilement interprétable ne peut pas être considéré comme meilleur à un modèle moins performant mais plus simple.
Il faut donc également des métriques qui pénalisent la complexité du modèle et donc le nombre de variables utilisées. Détaillons certaines métriques utilisées pour évaluer un modèle :

- **Le coefficient de Gini** : le coefficient de Gini représente deux fois l'aire entre la courbe de Lorenz et la diagonale. La courbe de Lorenz est expliquée dans la partie consacrée à la présentation du nouvel outil de modélisation Akur8. Un coefficient de Gini plus élevée indique de meilleures prédictions alors qu'une valeur proche de 0 signifie que les prédictions étaient comparables au hasard.
- **le coefficient de détermination** : il permet de mesurer la justesse de l'estimation et il s'exprime de la manière suivante :

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{2.10}$$

  Plus le R^2 est proche de 1, meilleure est la qualité de l'ajustement. En effet, ce coefficient est le rapport de la somme des carrés expliquée à la somme des carrés totale. Un R^2 proche de 1 signifie ainsi une perte minime d'information dans la modélisation.

  Cet estimateur a cependant le défaut de tendre systématiquement vers 1 avec l'ajout de variables explicatives, il lui sera donc préféré le coefficient de détermination ajusté suivant, pénalisé par le nombre p de variables explicatives du modèle :

$$R_a^2 = 1 - \frac{n-1}{n-p-1}\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{2.11}$$

- **La déviance standardisée** : le modèle estimé est comparé au modèle saturé, c'est-à-dire le modèle possédant autant de paramètres que d'observations et estimant donc exactement les données.

La déviance est définie à partir de la log-vraisemblance de ces deux modèles :
$D = 2(log(Lsat) - log(L))$

➢ **La statistique du Khi-Deux de Pearson standardisée** : un test de X^2 permet de comparer les valeurs observées Y_i à leurs valeurs prédites par le modèle. La statistique du test s'écrit :

$$X^2 pearson = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{Var(\hat{\mu}_i)^2} \tag{2.12}$$

➢ **L'erreur quadratique moyenne** (Mean Square Error (MSE)) : la MSE se calcule comme la moyenne des carrés des écarts entre la valeur théorique et la valeur prédite :

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.13}$$

avec :
- n le nombre d'observations
- y_i la valeur observée pour la i-ème observation
- $\hat{y}_i$ la valeur prédite pour la i-ème observation

On a également la Root Mean Square Error (RMSE) qui est la racine carrée de la MSE.

➢ **L'erreur absolue moyenne (Mean Absolute Error (MAE))** : la MAE est donnée par la formule :

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|^2 \tag{2.14}$$

La MAE et la MSE sont des erreurs plus fréquemment utilisées dans le but d'évaluer les performances d'un modèle. Le principal avantage de la MAE est d'être plus interprétable que la MSE. La MSE accorde plus d'importance aux grandes erreurs de prédiction, ce qui peut être utile si l'on souhaite éviter d'avoir des prédictions trop éloignées des valeurs observées.

Il existe également des métriques permettant de prendre en compte la complexité du modèle telles que l'AIC, le BIC et le R^2 ajusté. Ces erreurs sont utilisées afin de juger de la qualité des modèles obtenus. Il est conseillé de prendre en considération plusieurs erreurs pour évaluer des modèles.

Chapitre 3

Analyse des données

La première étape de cette étude concerne le traitement et l'analyse des données. En effet, dans le domaine de l'assurance, les données représentent un enjeu majeur. Des valeurs manquantes ou aberrantes, des erreurs ou des différences dans le renseignement des données peuvent être présentes au sein d'une base de données.

Le traitement des données est donc primordial afin d'avoir une base de données propre, fiable et utilisable pour ce mémoire mais également pour l'entreprise. En plus du retraitement des données, il faut également effectuer une phase d'analyse et de compréhension des données avant de procéder aux étapes suivantes.

3.1 Présentation des bases de données utilisées

Pour cette étude, des données du portefeuille d'AG2R La Mondiale intégrant le segment standard collectif avec la gamme « Omega 2 » sur la période du 01/01/2017 au 31/12/2021 ont été utilisées. La base de données a été enrichie avec les données des 4 plus grandes CCN du portefeuille en chiffre d'affaires en 2021 que sont Afflec, Aide à domicile, Boulangerie artisanale et Propreté ainsi que celles des 5 plus grandes entreprises en chiffre d'affaires en 2021 dont Cnes, Daher, UES Générale des eaux, Euro-Disneyland et Leroy Somer. Cet ajout permet d'avoir une base de données plus importante et donc d'obtenir des analyses plus robustes.

Cette base contient donc plusieurs produits hétérogènes, en run-off ou en cours de commercialisation. Il s'agit d'une nouveauté par rapport à l'ancienne méthode de construction des normes tarifaires où la base de données ne contenait que des données du segment standard collectif avec la gamme « Omega 2 ». Par ailleurs, au cours de ces 5 années, l'assurance santé a connu des évolutions suite à plusieurs réformes réglementaires. Un traitement de la base de données est donc nécessaire avant de procéder à la tarification.

Dans cette partie, les données utilisées et l'ensemble des traitements effectués sur la base de données pour pouvoir l'exploiter vont être présentées.

3.1.1 Base des bénéficiaires

La base des bénéficiaires contient l'ensemble des personnes assurées par AG2R La Mondiale sur les années 2017 à 2021. En effet, les données sont issues du portefeuille standard collectif, sur-mesure et CCN. La base de données est constituée de variables correspondant aux caractéristiques du bénéficiaire telles que le sexe, le type de bénéficiaire, le code postal de l'assuré ou encore la date de naissance.

Pour les besoins de l'étude, il a été nécessaire d'ajouter certaines variables à la base de données :

- **L'exposition :** il s'agit de la durée d'observation exprimée en année associée à chaque bénéficiaire par année de survenance. Concrètement, l'exposition correspond au ratio entre la durée de couverture de l'assuré et la période considérée (une année dans ce cas). Il est important de la calculer afin de corriger la fréquence de consommation de l'assuré par sa durée de couverture. Par exemple, un assuré couvert entre janvier et juin aura une exposition de 0,5. Cette variable a été calculée à partir des dates de début et de fin d'adhésion. Le bénéficiaire qui est présent de 2017 à 2021 a donc une exposition par année de présence.
- **L'âge :** calcul en millésime (différence d'année) à partir de la date de naissance du bénéficiaire et de la date de survenance : Âge = Année de survenance - Année de naissance.
- **Le nombre de bénéficiaire :** calcul du nombre d'identifiant de bénéficiaire différent pour un même salarié, salarié inclus dans le comptage.

- **La tranche d'effectif de l'entreprise :** calcul du nombre de bénéficiaires au sein d'un même contrat collectif pour les segments standard collectif et Conventionnel (CCN). Pour le segment sur-mesure, il s'agit du nombre de bénéficiaires par entreprise car une même entreprise peut avoir plusieurs contrats collectifs. La variable « Tranche_entre » peut alors être créée :

Tranche d'effectif de l'entreprise	Tranche_entre
Entre 0 et 10 bénéficiaires	0-10
Entre 10 et 50 bénéficiaires	10-50
Entre 50 et 100 bénéficiaires	50-100
Entre 100 et 1000 bénéficiaires	100-1000
Entre 1000 et 5000 bénéficiaires	1000-5000
Entre 5000 et 10 000 bénéficiaires	5000-10 000
Entre 10 000 et 100 000 bénéficiaires	10 000-100 000
Plus de 100 000 bénéficiaires	100 000+

- **La catégorie de gestion personnelle du bénéficiaire :** Cette variable a été créée à partir des codes et libellés de la catégorie de gestion personnelle.
- **Le secteur d'activité de l'entreprise du salarié :** Cette variable était disponible au sein de la base de données mais elle disposait de trop de modalités différentes et était donc inexploitable. Des regroupements ont donc été effectués afin de réduire le nombre de secteurs d'activités à 8 contre presque 200 initialement :

Secteur d'activité de l'entreprise
Services Financiers et administratif & Recherche-développement scientifique
Restauration et agro-alimentaires
Santé et actions sociales
Commerce de détail en fruits et légumes, epicerie, crémerie
Santé et actions sociales
Energie, Industrie et travaux de réparation et de batiments
Activités sportives, récréatives et de loisirs & Arts et cinéma
Nettoyage courant des bâtiments
Commerce, transport et autres services

La base des bénéficiaires a ensuite été restructurée de façon à ce que chaque ligne représente un bénéficiaire sur une année. La variable année prenant les valeurs de 2017 à 2021 a donc été ajoutée. Par conséquent, un bénéficiaire présent de 2018 à 2020 sera présent sur 3 lignes différentes : une ligne pour chaque année avec son âge et son exposition au cours de cette année. Lorsque la base des bénéficiaires sera fusionnée avec celle des prestations, les différents montants de remboursement correspondant à chacun des actes seront en colonne.

Ainsi, le tableau ci-dessous est un récapitulatif du contenu de la base de données des bénéficiaires :

Variable	Description
typer	Segment : standard, CCN, sur-mesure
idctr	Identifiant du contrat individuel
dtnaiass	Date de naissance de l'assuré
cosexass	Sexe de l'assuré
idben	Identifiant du bénéficiaire
dtnaiben	Date de naissance du bénéficiaire
cosexben	Sexe du bénéficiaire
tyben	Type du bénéficiaire : AD : Adulte, CJ : Conjoint, EN : Enfant, MP : majeur protégé
lbreg	Libellé du régime du bénéficiaire : Régime Général, Régime Local
copro	Code du produit
NOSECSOC	Numéro de Sécurité sociale
COOPT	Code de l'option
COETACTR	Code de l'état du contrat : C=en cours, R=résilié
COMOTRES	Code du motif de résiliation éventuel
IDPER	Identifiant de l'adhérent principal
IDPERJUR	Identifiant de personnalité juridique de l'entreprise de l'assuré
Secteur_activite	Libellé du groupe de l'activité principale exercée par l'entreprise de l'assuré
code_postal_ass	Code postal de l'assuré
Dept_ass	Département de l'assuré
Dept_ent	Département de l'entreprise de l'assuré

ADCODPOS	Code postal de l'entreprise de l'assuré
date_new_fin_ass	Date de fin d'adhésion de l'assuré, prenant la valeur 31/12/9999 si aucune date de résiliation n'est encore prévue
date_new_deb_ass	Date de début d'adhésion de l'assuré
date_new_fin_ben	Date de fin d'affiliation du bénéficiaire, prenant la valeur 31/12/9999 si aucune date de résiliation n'est encore prévue
date_new_deb_ben	Date de début d'affiliation du bénéficiaire
Cocatper	code de la catégorie de gestion personnelle de l'assuré
Categorie_pop	catégorie de gestion personnelle de l'assuré
dtsurann	Année de survenance
exposition	Exposition du bénéficiaire
annaiben	Année de naissance du bénéficiaire
age_ben	Âge du bénéficiaire
nb_ben	Nombre de bénéficiaires différents pour un même salarié
Tranche_entre	Tranche d'effectif de l'entreprise de l'assuré

Tableau 3.1 : Récapitulatif des variable de la base des bénéficiaires

Ajout de variables externes

Afin d'enrichir la base de données et de pallier le manque de données sur certains départements dans le zonier, des variables externes ont été intégrées à la base de données.

Ces variables ont été insérées au sein de la base par code postal ou département de l'assuré. Ces données datant de 2017 ou 2018 ont été récupérées sur les sites de l'INSEE, ECOSANTE et SCORE SANTE.

Il y a notamment des variables telles que la population de plus de 60 ans 2018 (INSEE), le taux d'équipement en services de santé 2017 (INSEE), l'indice de vieillissement 2015 (ECOSANTE) ou encore le nombre de médecins, généralistes, spécialistes et établissement de santé public en 2017 (SCORE SANTE).

3.1.2 Base des prestations

La base des prestations a été divisée par type d'acte. En effet, elle contient les différents remboursements effectués de 2017 à 2021 pour chacun des 38 actes. La base des prestations regroupe 136 600 128 prestations représentant 924 138 707 euros.

La base permet d'avoir accès à :

- L'identifiant et les caractéristiques du bénéficiaire ;
- L'année de survenance du sinistre et les différents montants de remboursements associés au sinistre : frais réels, base de remboursement de la Sécurité sociale, montant remboursé par la Sécurité sociale, remboursement de la complémentaire, le reste à charge de l'assuré ainsi que le remboursement éventuel d'une autre mutuelle.

Le tableau ci-dessous récapitule les informations présentes au sein de la base des prestations :

Variable	Description
NOCTR	Numéro de contrat collectif
IDCTR	Identifiant du contrat individuel
IDBEN	Identifiant du bénéficiaire de la prestation
DTSOI	Date de survenance du sinistre
DTCOM	Date de règlement du sinistre
IDGAR	identifiant de la garantie
LBGAR	Libellé de la garantie
DTSURANN	Année de survenance du sinistre
NBQTE	Nombre d'actes
GAMME	Gamme du produit
MTREM	Montant de la prestation (part complémentaire d'AG2R La Mondiale)
MTFRR	Montant des frais réels
MTPAIBAS	Base de remboursement du régime obligatoire (RO)
MTBAS	Montant de remboursement du régime obligatoire (RO)
MT1MU	Montant de remboursement de la première mutuelle
COPRO	Code du produit
COOPT	Code de l'option

NOELT	Numéro d'élement de remboursement liquidé

Tableau 3.3 : Récapitulatif des variables de la base des prestations

D'autre part, il est indispensable de procéder à un regroupement d'actes. Les données pourront ainsi être observées et les bases regroupées plus simplement. Un regroupement d'actes a ainsi été effectué :

— par grands postes : il y a 6 grands postes : Actes médicaux, Autres prestations, Hospitalisation, Dentaire, Pharmacie et Optique.

— par acte : l'objectif est de déterminer les primes pures par acte donc les actes similaires sont regroupés entre eux.

Le regroupement utilisé est présenté dans le tableau suivant :

Poste	**Acte**
Actes médicaux	Actes de chirurgie et technique Actes de chirurgie et technique CAS Consultations et visites généralistes Consultations et visites généralistes CAS Consultations et visites spécialistes Consultations et visites spécialistes CAS Médecine douce
Optique	Chirurgie réfractive Monture Lentilles acceptées ou refusées par le RO verres simples adulte verres simples enfant verres multifocaux ou progressifs adulte verres multifocaux ou progressifs enfant
Autres prestations	Actes d'imagerie médicale, radiologie, et échographie CAS Analyses Autres Auxiliaires médicaux Cure Thermale acceptée par le RO (honoraire et traitement) Cure Thermale acceptée par le RO (Transport et Hébergement)

	Honoraires Honoraires CAS Maternité Prothèses non dentaire Transport remboursé par la SS
Dentaire	Implantologie Inlay-Cores Inlay-Onlay Orthodontie Parodontologie Prothèses dentaires remboursées par le RO Prothèses dentaires non remboursées par le RO Soins dentaires
Hospitalisation	Chambre particulière Forfait hospitalisation Frais d'accompagnement Frais de séjour
Pharmacie	Pharmacie

Tableau 3.4 : Répartition des actes par grand poste

Le choix a été fait de ne pas procéder à des regroupements plus larges afin d'obtenir une tarification fine.

3.2 Analyse et retraitements des données

De nombreux retraitements ont été effectués sur la base des bénéficiaires ainsi que sur la base des prestations. En assurance santé, il est courant de trouver des erreurs dans les bases de données si elles n'ont pas été analysées auparavant. Il faut alors repérer ces erreurs et les corriger si possible, sinon les supprimer pour ne pas biaiser les analyses.

3.2.1 Base des bénéficiaires

Dans un premier temps, des retraitements ont été effectués sur la base des bénéficiaires afin de sélectionner le périmètre d'étude. Ainsi, dans le cadre de cette étude, il a été décidé d'écarter les bénéficiaires correspondant aux

lois Evin et aux travailleurs non-salariés (TNS) et de ne conserver que les bénéficiaires issus des catégories suivantes :

- Cadres
- Non-cadres
- Ensemble du personnel
- Ayants-droits
- Suspendus
- Portabilité

Dans le cadre de cette étude, la base a également été restreinte aux assurés appartenant au régime général de la Sécurité sociale. Les assurés des autres régimes tels que le régime Alsace-Moselle ont donc été écartés de cette étude. Au total, les bénéficiaires n'appartenant pas au régime général représentaient environ 11% de la population d'étude. Les assurés des autres régimes ainsi que des catégories « loi Evin » et « TNS » feront partie d'une étude ultérieure qui ne sera pas traitée dans ce mémoire.

De plus, comme il a été dit précédemment, la base de données comporte 5 années de survenance : 2017 à 2021. Ce choix a été fait car cette étude cherche à établir des modèles de tarification par année de survenance. Or, l'année 2020 est atypique suite à la crise sanitaire liée au Covid-19.

Sur l'année 2021, on observe des consommations de frais de santé plus importantes que les années précédentes liées à la réforme du 100% santé et à un changement de comportement. Il est donc nécessaire d'étudier plusieurs années dont la consommation n'a pas été affecté par cette pandémie. Tous les bénéficiaires ayant une date de fin d'adhésion inférieure au 01/01/2017 ont alors été écartés de la base des bénéficiaires.

Par ailleurs, le choix a été fait de conserver uniquement les bénéficiaires ayant des âges qui semblaient cohérents par rapport à la cible visée :

- Les adultes (salariés) ayant un âge compris entre 16 et 80 ans
- Les conjoints ayant un âge compris entre 18 et 80 ans
- Les enfants ayant un âge compris entre 0 et 26 ans
- Les Majeurs protégés ayant un âge compris entre 18 et 100 ans

De même, les bénéficiaires ayant souscrit une option facultative ou une surcomplémentaire sont écartés de l'étude afin de ne pas biaiser les résultats avec le phénomène d'antisélection. Ainsi, la base d'étude ne contient que les contrats à adhésion collective obligatoire. Ces bénéficiaires représentaient environ 8% de l'exposition totale de la base de données pour chaque année de survenance, correspondant à 10% des prestations totales. Il s'agit majoritairement d'options facultatives car les surcomplémentaires ne représentent qu'environ 1% des prestations et de l'exposition totale de la base de données.

De plus, un problème de doublons des salariés dans la base a été identifié. La variable « idben » n'identifie pas de manière unique un bénéficiaire mais plutôt les contrats du bénéficiaire car cet identifiant change si le contrat correspond aux garanties de base ou aux garanties optionnelles. L'objectif est donc d'avoir un identifiant du bénéficiaire unique afin d'obtenir des résultats non biaisés pour l'étude. Pour dédoublonner les bénéficiaires, une table de correspondance qui contient la variable « idben_gold » correspondant à l'identifiant unique du bénéficiaire a été utilisée.

Dans un second temps, du fait du manque de renseignements ou d'anomalies dans certaines variables, certaines manipulations ont été nécessaires sur les variables suivantes :

- **Le sexe :** la variable sexe présente quelques valeurs manquantes. Il a donc été nécessaire de compléter ses valeurs manquantes avec son numéro de Sécurité sociale. En effet, le premier chiffre, différent de 0, correspond au sexe de la personne (1 = H, 2 = F, 7 = H, 8 = F).
- **La date de fin et de début du bénéficiaire :** certains cas où la date de début du bénéficiaire était supérieure à sa date de fin d'adhésion ont été identifiés. Etant donné le peu de cas concerné et le faible montant de prestation correspondant à ces bénéficiaires, ces cas ont été supprimés de la base des bénéficiaires.
- **L'exposition :** cette variable correspondant à la durée passée par l'individu sur l'année de survenance, les lignes avec une exposition du bénéficiaire négative ou nulle pour une année de survenance ont alors été supprimées.

En moyenne, par année de survenance, l'ensemble de ces retraitements concerne 23% des bénéficiaires correspondant à 21% des prestations totales. Finale-

ment, la base totale regroupe 1 178 794 personnes et le détail par année de survenance est indiqué dans le tableau suivant :

Année de survenance	Nombre de bénéficiaires	Exposition
2017	723 852	581 773
2018	684 030	556 845
2019	643 841	529 988
2020	620 342	516 707
2021	626 070	510 334
Toutes années confondues	1 178 794	1 025 038

Tableau 3.5 : Exposition et nombre de bénéficiaire en fonction de l'année de survenance

3.2.2 Base des Prestations

Le traitement de la base des prestations débute par la suppression des sinistres ayant des frais réels nuls. Ensuite, certains actes de la base de données ayant des frais réels négatifs ont été identifiés au sein de la base de données. Il s'agit en général de régularisations purement comptables : par exemple, si un acte a été entré deux fois par erreur, une troisième ligne est créée avec le même montant en négatif.

Ainsi, d'un point de vue comptable, les lignes s'annulent. Néanmoins, dans le cadre de cette étude, il n'est pas possible de les conserver afin de ne pas surestimer la fréquence. Les lignes concernées ont donc été écartées en supprimant la ligne négative ainsi que la ligne positive correspondant à l'erreur.

De plus, une variable calculant le reste à charge final a été ajoutée :

$$\text{Reste à charge} = \text{Mtfrr} - \text{Mtbas} - \text{Mt1mu} - \text{Mtrem}$$

Avec :

- Mtfrr = Montant de frais réels
- Mtbas = Montant remboursé par la Sécurité sociale
- Mt1mu = Montant remboursé par la 1ère mutuelle
- Mtrem = Montant remboursé par AG2R La Mondiale

A la suite de ce calcul, certaines lignes de prestation ayant un reste à charge

négatif ont été identifiées. Ces anomalies peuvent avoir plusieurs explications possibles. En effet, il peut y avoir des remboursements de la Sécurité sociale ne correspondant pas au taux de la Sécurité sociale multiplié par la BR. Il peut aussi y avoir un remboursement de la complémentaire ou une base de remboursement de la Sécurité sociale supérieure ou égale aux frais réels.

Enfin, le remboursement de la complémentaire peut être égal aux frais réels alors que la Sécurité sociale a effectué un remboursement. Le choix de supprimer ces cas-là a été fait compte tenu de la faible proportion qu'ils représentaient (moins de 0,03%). Toutefois, les cas concernant les actes « Cure » et « Maternité » ne sont pas des anomalies car des forfaits indépendant de la dépense et limités aux frais réels sont versés pour ces deux actes. Ainsi, les données ont été conservées lorsqu'il s'agissait de ces deux actes.

Construction d'une base historisée « as-if »

La base de données rassemble l'ensemble des sinistres entre le 01/01/2017 et le 31/12/2021. Or, il faut prendre en compte le fait que les coûts sur les exercices précédents ne sont plus représentatifs des coûts actuels. L'idée initiale était de retraiter la base de données à disposition afin de recréer une base « as-if », c'est à-dire un historique de prestations actualisé.

L'objectif initial était de normaliser les prestations de 2017 à 2020 afin de prendre en compte la dérive entre ces années et 2021 par poste pour 2020 et globalement pour les autres années.

Ce choix n'a pas été retenu car le détail de la dérive par poste pour chaque année n'était pas connu. De plus, la dérive par poste ne correspondrait pas à la dérive pour chaque acte d'un même poste. Ainsi, cette méthode conduirait à des biais dans le montant de prestations final. Afin de régler ce problème de dérive des prestations, des modèles de tarification par année de survenance ont été réalisés séparément.

L'année 2021 a finalement été retenue pour la modélisation des actes car il s'agit de l'année la plus récente et qu'il n'y a pas eu d'effet « rebond » constaté suite à la crise sanitaire ayant eu lieu en 2020.

Estimation des PSAP

La base des sinistres recense les sinistres payés du 1er janvier 2017 au 30 décembre 2021 et extrait en juin 2022 avec une date d'arrêté à fin mai 2022. Ainsi, étant donné qu'en assurance santé les durées de déclaration d'un sinistre et de remboursement sont relativement courtes (moins de 3 mois), les montants de PSAP sont négligeables et il n'est pas nécessaire de les prendre en compte dans le cadre de ce mémoire.

Une estimation des PSAP par année de survenance est déterminée par le service décisionnel d'AG2R La Mondiale qui mets à disposition les données nécessaires aux différentes études actuarielles. Le tableau ci-dessous indique le ratio des PSAP restantes par année de survenance à fin mai 2022 en gestion directe :

Année de survenance	Montant de prestations	PSAP	Ratio PSAP
2019	186 031 140 €	0 €	0,0%
2020	168 813 560 €	337 627 €	0,2%
2021	193 064 275 €	3 282 093 €	1,7%

Tableau 3.6 : Proportion de PSAP par année de survenance

En effet, le domaine de la santé est un secteur de l'assurance dans lequel la vitesse de règlement, aussi appelée cadence de règlement, est assez rapide comparé au domaine de la prévoyance.

Ainsi, les sinistres survenus entre 2020 et 2021 sont très majoritairement pris en compte dans la base de données puisque celle-ci se termine en 2021 et les données ont été récupérées en juin 2022 avec une date d'arrêté à fin mai 2022.

De plus, l'utilisation de plus en plus généralisée du tiers-payant a tendance à réduire encore plus la durée entre la déclaration et la durée de remboursement d'un sinistre.

3.3 Garanties

Les produits au sein de la base de données ont des niveaux des garanties proposés différents et la tarification appliquée ne sera pas la même selon ce niveau de garantie. Le comportement d'un assuré peut varier selon le niveau de garantie. En effet, un assuré ayant un niveau de garantie plus élevé sera amené à avoir une fréquence de consommation ainsi qu'un coût moyen plus élevé qu'un assuré ayant une garantie de base. Lors de la construction de nouvelles normes tarifaires, il faut donc tenir compte des niveaux de garanties proposés.

Les garanties sont des variables essentielles afin de mener à bien une tarification. Au sein de l'entreprise AG2R La Mondiale, les niveaux de garanties ne font pas l'objet d'une codification dans le système d'information homogène selon les produits et rend cette information difficilement exploitable directement. Ces informations ont donc été récupérées en se rapprochant de la direction de la souscription qui a accès aux grilles de garantie des différents produits de cette étude.

Toutefois, constatant un grand nombre de niveau de garanties exprimés différemment, une étude a été réalisée afin de les harmoniser. Ainsi, des correspondances ont été créées pour n'avoir que des garanties s'exprimant en « Frais Réels », « Ticket Modérateur », « BR », « Forfait Journalier Hospitalier » et « EUROS » pour chaque acte étudié. Certains niveau de garanties sont exprimés par jour, par semestre ou encore par année civile. D'autres présentent des plafonds comme en implantologie pour certaines options : « forfait par implants avec max 3 par ans ».

Ce travail a permis d'avoir toutes les garanties sous la même expression. La différence de niveau remboursé par la complémentaire santé lorsque le médecin est adhérent ou non adhérent au CAS/OPTAM a également été prise en compte. Cette différence s'est accentuée depuis la mise en vigueur des contrats responsables.

Un travail préliminaire a ensuite été effectué afin de faire le lien entre les grilles de garanties et chaque acte de remboursement. Les produits ayant connus des évolutions au fil des années, entrainant la modification des grilles de garanties, il a donc fallu prendre en compte les dates de validité des grilles

de garanties. Des grilles de garanties ont ainsi été créées pour les joindre à la base de données :

module	Garantie	Rbst_AD1
base 1	Frais de séjour	220% BR
base 1	Honoraires (actes de chirurgie, actes d'anesthésie, autres honoraires)	200% BR
base 1	Honoraires CAS	220% BR
base 1	Honoraires	200% BR
base 1	Chambre particulière	70 €/jour
base 1	Forfait hospitalisation	100% FJH
base 1	Frais d'accompagnement	35 €/jour
base 1	Transport remboursé par la SS	100% BR
base 1	Consultations et visites généralistes	175% BR
base 1	Consultations et visites généralistes CAS	195% BR
base 1	Consultations et visites spécialistes	200% BR
base 1	Consultations et visites spécialistes CAS	220% BR
base 1	Actes de chirurgie et technique médicaux	200% BR
base 1	Actes de chirurgie et technique CAS	220% BR
base 1	Actes de chirurgie et technique	200% BR
base 1	Actes d'imagerie médicale, radiologie, et échographie	150% BR
base 1	Actes d'imagerie médicale, radiologie, et échographie CAS	170% BR
base 1	Auxiliaires médicaux	100% BR
base 1	Analyses	100% BR
base 1	Pharmacie remboursable autres produits	100% BR
base 1	Prothèses auditives	100% BR
base 1	Autres Prothèses et appareillages	100% BR
base 1	Maternité	300 €
base 1	Cure Thermale acceptée par le RO (honoraire et traitement)	100% BR
base 1	Cure Thermale acceptée par le RO (Transport et Hébergement)	250 €
base 2	Frais de séjour	270% BR
base 2	Honoraires (actes de chirurgie, actes d'anesthésie, autres honoraires)	200% BR
base 2	Honoraires CAS	270% BR
base 2	Honoraires	200% BR

FIGURE 3.1 – Exemple de grille de garantie

Dans la grille de garantie ci-dessus, la variable « Rbst_AD1 » spécifie le niveau de garantie et la variable « module » permet de différencier les garanties du produit de base à celles des options. Il peut y avoir jusqu'à 5 niveau d'options pour le Dentaire, 4 pour l'Optique et 4 pour les autres postes.

La variable « module » a été créée car certains produits présentent différentes options et il a donc fallu déterminer le niveau de l'option correspondant pour chaque acte de remboursement. Pour ce faire, deux nouvelles variables provenant des bases de données de l'infocentre d'AG2R La Mondiale ont été

intégrées permettant respectivement de savoir si le produit est une base ou une option et de connaitre le niveau de l'option.

À la suite de ce travail, les niveaux de garanties ont pu être associés à chaque acte remboursé en fusionnant les grilles de garanties avec la base de données par module et garantie. Il a également été important d'ajouter une autre variable indiquant s'il s'agit d'une option facultative ou obligatoire ou encore d'une surcomplémentaire.

En fonction du niveau de garantie pour chaque acte, les garanties ont ensuite été classées en 4 niveaux :

Niveau 1	**Niveau 2**	**Niveau 3**	**Niveau 4**
entrée de gamme	milieu de gamme	haut de gamme	très haut de gamme

Une grille de décision a été créée en fonction des niveaux de garanties présents pour chaque acte dans la base afin de lui attribuer un niveau de gamme. Cette grille de décision est présente dans l'annexe de ce mémoire.

La moyenne pondérée par les prestations des niveaux de gamme par acte au sein d'un poste a permis de déterminer le niveau de gamme à la maille du poste. Cette variable va permettre d'identifier s'il y a des différences de consommation en fonction du niveau de garantie proposé au niveau du poste puis au niveau du contrat :

Poste	Niveau du poste	Poids du poste	Niveau du poste pondéré	Niveau de gamme du contrat
Actes médicaux	2	0,24	0,48	Somme des niveaux des postes pondérés = 2 = Milieu de gamme
Optique	2	0,43	0,86	
Dentaire	3	0,33	0,99	

FIGURE 3.2 – Exemple du calcul du niveau de gamme du contrat

3.4 Fusion des bases de données

Une fois les deux bases de données sur les prestations et sur les bénéficiaires créées, elles peuvent désormais être fusionnées. Ainsi, la base des prestations a été fusionnée avec la base des bénéficiaires par numéro de contrat, identifiant du bénéficiaire et année de survenance.

La base est ensuite agrégée par bénéficiaire et année de survenance avec la somme des différents montants de remboursement correspondant à chacun des actes en colonne. La base obtenue contient alors une ligne par bénéficiaire et année de survenance correspondant à environ 4 millions de lignes.

Le processus de constitution de la base de données est récapitulé par le schéma suivant :

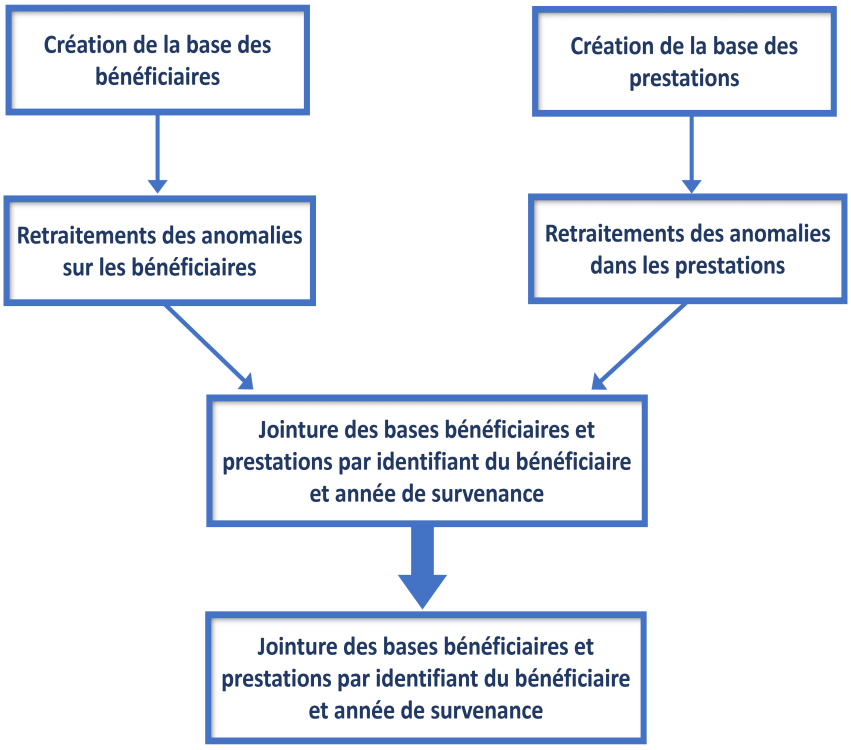


FIGURE 3.3 – Schéma récapitulatif du processus de création de la base de données

3.5 Description et étude des données

Avant de procéder à la tarification, une étude statistique du portefeuille a été effectuée. Cette partie vise à étudier les caractéristiques de la population étudiée et analyser sa consommation entre janvier 2017 et décembre 2021. Ce travail permet d'appréhender et d'interpréter au mieux les résultats qui découleront des modèles prédictifs de tarification.

3.5.1 Analyse de la population des bénéficiaires

L'étude des caractéristiques de la population va être réalisée dans un premier temps. La répartition de la population en fonction de l'âge par année de survenance est présentée ci-dessous :

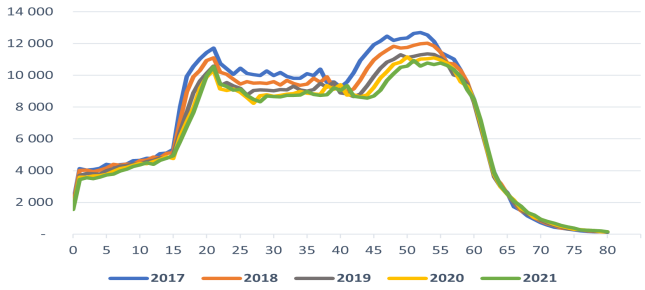


FIGURE 3.4 – Exposition en fonction de l'âge et de l'année de survenance

La population de cette étude est majoritairement âgée entre 18 et 60 ans tandis que les âges extrêmes sont moins représentés. Cette analyse peut être affinée avec la pyramide des âges des bénéficiaires toutes années confondues :

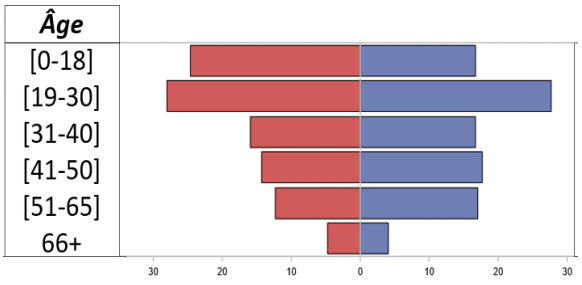


FIGURE 3.5 – Pyramides des âges des bénéficiaires

La pyramide des âges permet de constater que la tranche d'âge de 19 à 30 ans est la plus représentée au sein de la population. A l'inverse, la tranche d'âge la moins représentée est celle des plus de 66 ans.

L'âge moyen du portefeuille d'étude est de 35 ans et l'âge moyen par type de bénéficiaire de la population est détaillé dans le tableau ci-dessous :

Type de bénéficiaire	Âge moyen
Salarié	40
Conjoint	48
Enfant	11
Majeur Protégé	28

Tableau 3.7 : Âge moyen par type de bénéficiaire

Il est aussi intéressant de connaître la répartition de la population en fonction du sexe et de la tranche d'âge du bénéficiaire :

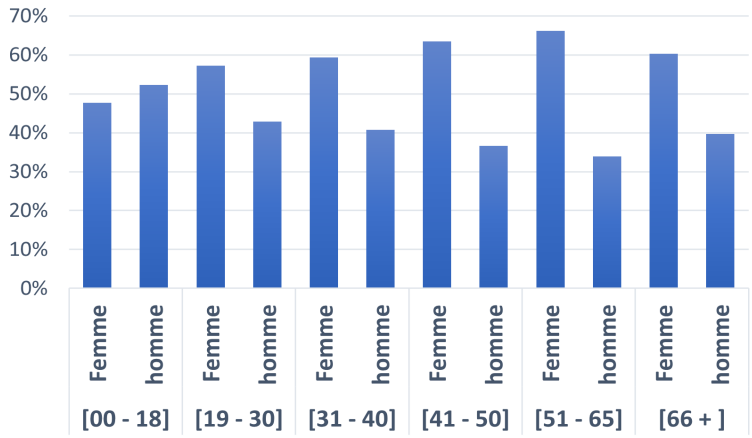


FIGURE 3.6 – Répartition de la population en fonction du sexe et de la tranche d'âge

La population est majoritairement composée de femmes pour chaque tranche d'âge mises à part les tranches d'âge extrêmes de 0 à 18 ans et des plus de 66 ans. Cette différence s'accentue avec l'âge entre 19 ans et 65 ans.

De même, le portefeuille est à prédominance féminine pour chaque année de survenance avec en moyenne 58% de femme pour 42% d'homme :

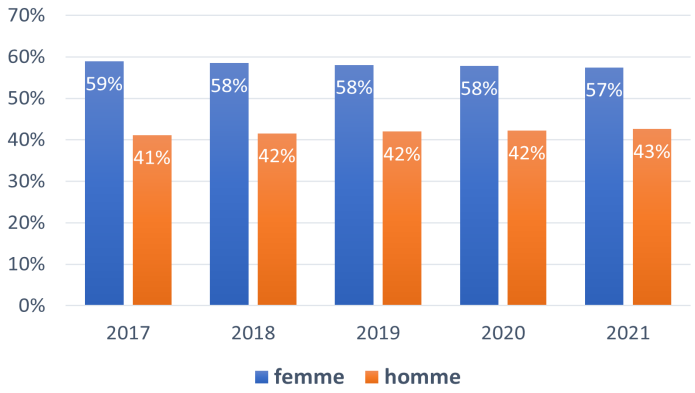


FIGURE 3.7 – Répartition de la population par sexe et année de survenance

Le portefeuille est composé de 10 gammes de produits différentes faisant partit des segments standard collectif, CCN et sur-mesure :

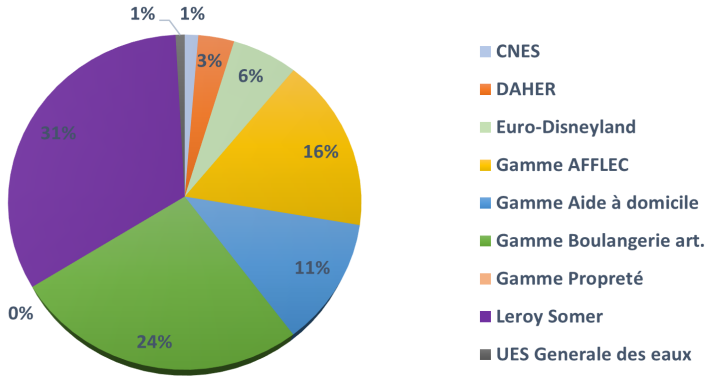


FIGURE 3.8 – Répartition de la population par gamme en 2021

La répartition pour l'année de survenance 2021 est représentée ci-dessus et elle est équivalente pour chaque année de survenance. La gamme « Propreté » est la plus représentée tandis que les entreprises CNES et Leroy Somer sont

minoritaires au sein de la population.

Les CCN représentent une grande partie du portefeuille avec 77% suivis par le sur-mesure avec 16% et le standard collectif avec 7% :

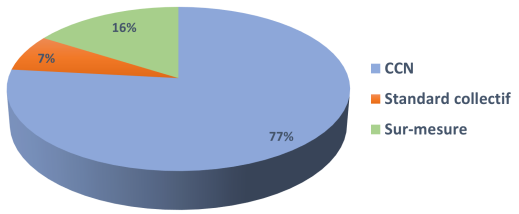


FIGURE 3.9 – Répartition de la population par segment en 2021

La répartition de la population en fonction de la catégorie de gestion personnelle pour toutes années confondues va maintenant être étudiée :

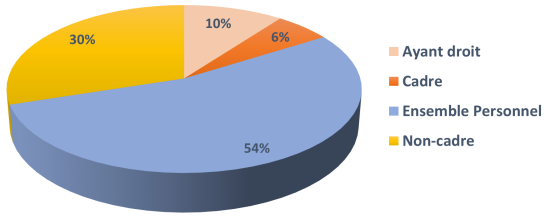


FIGURE 3.10 – Répartition de la population par catégorie de gestion personnelle toutes années confondues

La population d'étude se compose de bénéficiaires appartenant majoritairement à la catégorie « Ensemble personnel » suivie respectivement par les « Non-cadre », « Cadre » et « Ayant droit ». Les poids des autres catégories ne sont pas significatifs au sein de la population d'étude. La catégorie « Ayant droit » corresponds aux bénéficiaires affiliés aux contrats du salariés et qui peuvent aussi bien être des cadres ou des non-cadres.

De même, les assurés appartenant à la catégorie « Ensemble personnel » peuvent être aussi bien des cadres que des non-cadres. La catégorie de gestion personnelle n'est donc pas renseignée de façon optimale pour pouvoir créer une variable correspondant à la CSP du bénéficiaire à partir de la catégorie de gestion personnelle.

Par ailleurs, la population se compose majoritairement d'adulte à 79% ainsi que de 13% d'enfants, 5% de conjoints et 0,03% de majeurs protégés. Le nombre de majeurs protégés est donc dérisoire au sein du portefeuille :

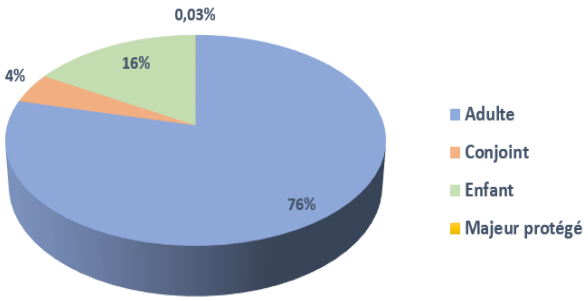


FIGURE 3.11 – Répartition de la population par type de bénéficiaire

De plus, la répartition de la population en fonction du secteur d'activité est affichée dans le graphique ci-dessous :

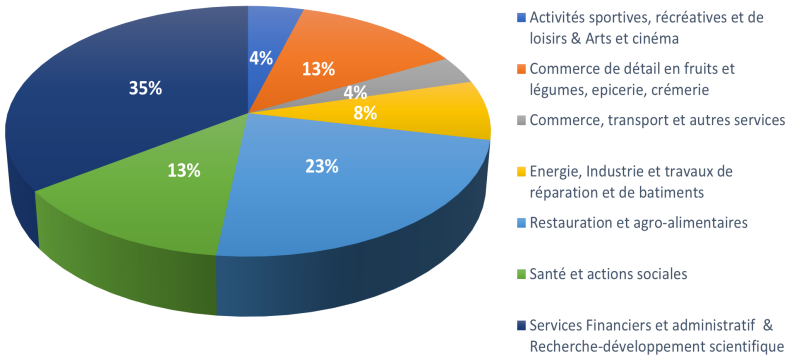


FIGURE 3.12 – Répartition de la population par secteur d'activité

Une grande partie de la population (35%) travaille dans des entreprises liées au secteur « services financiers et administratifs & Recherche-développement scientifique » suivis par le secteur « santé et actions sociales » qui représente 23% de la population. Les secteurs d'activités les moins représentés au sein de la population d'étude sont « activités sportives, récréatives et de loisirs & Arts et cinéma » ainsi que le secteur « commerce, transport et autres services » avec 4% de la population chacun.

Enfin, en observant la répartition de la population par tranche d'effectif de l'entreprise, il peut être constaté que la majorité des entreprises ont entre 0 et 50 salariés :

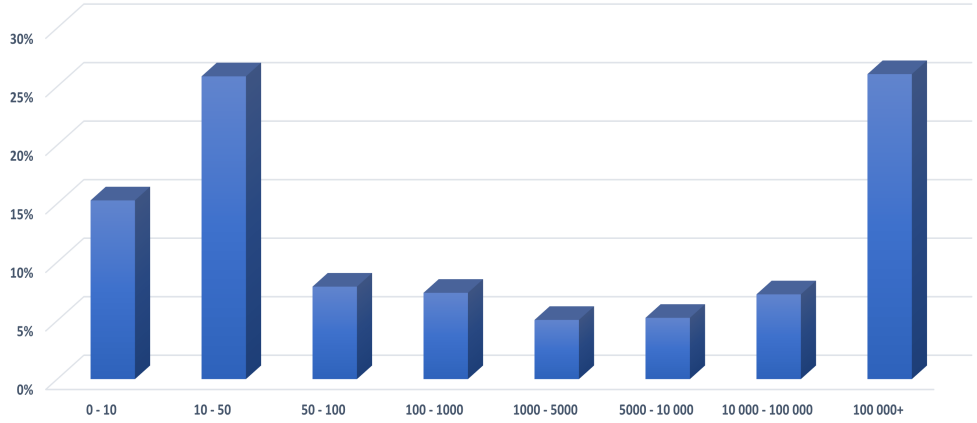


FIGURE 3.13 – Répartition de la population en fonction de la tranche d'effectif de l'entreprise

Il y a effectivement un nombre conséquent de bénéficiaires travaillant dans des petites entreprises entre 10 et 50 salariés ainsi que dans des grandes entreprises à plus de 100 000 salariés. Parmi la population étudiée, une seule entreprise possède plus de 100 000 salariés tandis qu'il y a environ 51 000 entreprises entre 0 et 10 salariés et environ 15 000 entre 10 et 50. En revanche, les tranches de 1000 à 5000 et de 5000 à 10 000 salariés sont moins représentées au sein de la population et elles concernent respectivement 11 et 5 entreprises.

3.5.2 Analyse de la base des prestations

Après avoir analysé les caractéristiques de la population de cette étude, les prestations vont être étudiées. Tout d'abord, la répartition des montants de prestations par poste et année de survenance est représentée ci-dessous :

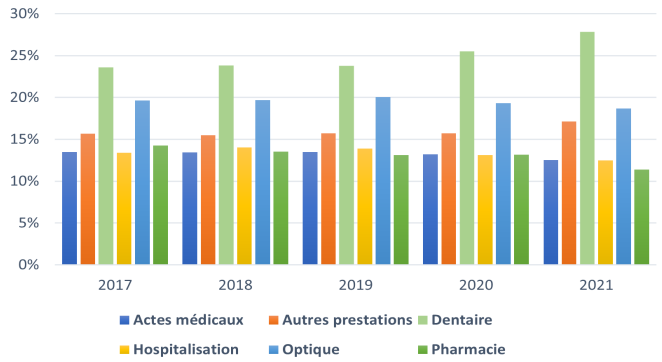


FIGURE 3.14 – Répartition des prestations par poste et année de survenance

Les postes ayant les montants de remboursement les plus importants sont le Dentaire et l'Optique. A contrario, les postes Pharmacie et Hospitalisation sont ceux ayant une somme de prestations annuelle globalement plus faible que les autres.

A l'inverse des prestations, le nombre d'acte est bien supérieur pour le poste Pharmacie pour chaque année de survenance :

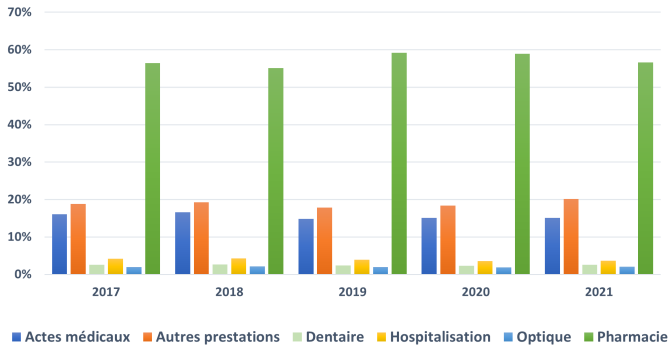


FIGURE 3.15 – Répartition des actes par poste et année de survenance

Pour les postes Dentaire et Optique, le nombre d'acte annuel est plus petit par rapport aux autres postes alors que la somme de prestations annuelle était plus grande.

L'étude du coût moyen d'un acte par poste confirme cette observation :

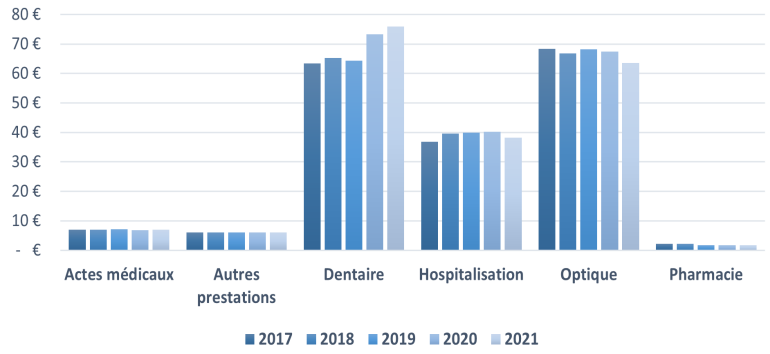


FIGURE 3.16 – Coût moyen par acte en fonction du poste par année de survenance

Ainsi, les postes Dentaire, Optique et Hospitalisation sont les plus coûteux tandis que la Pharmacie est le poste le moins onéreux.

Voici l'analyse de la consommation par poste et par année de survenance :

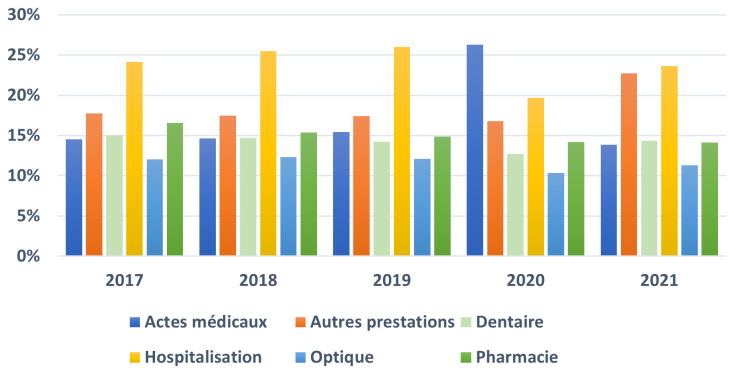


FIGURE 3.17 – Répartition de la consommation par poste et année de survenance

Les postes Hospitalisation et Autres prestations sont généralement ceux étant les plus consommés chaque année.

Il y a une hausse notable de la consommation en Actes médicaux en 2020 par rapport aux autres années ainsi qu'en Autres prestations en 2021.

La forte hausse sur 2020 des actes médicaux est directement liée à la crise sanitaire pendant laquelle il y a eu des confinements durant l'année 2020. La hausse des autres prestations sur 2021, elle, est liée à la réforme du 100% santé avec une meilleure prise en charge des audioprothèses.

De plus, il est remarqué que le poste Hospitalisation a été moins consommé qu'habituellement en 2020 durant la période où la crise sanitaire liée au Covid-19 a fait son apparition.

Par ailleurs, l'analyse de la consommation annuelle par bénéficiaire en fonction de l'âge et du sexe indique que les femmes consomment plus que les hommes de 18 à 50 ans avant que cela s'inverse légèrement :

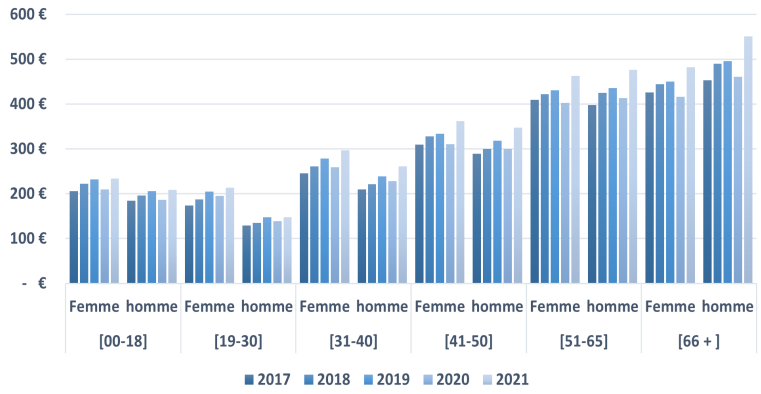


FIGURE 3.18 – Consommation annuelle par tranche d'âge, sexe et année de survenance

Il peut aussi être remarqué que la consommation annuelle augmente linéairement par tranche d'âge à partir de 19 ans.

Il est également intéressant d'étudier les statistiques descriptives relatives aux segments de la population de l'étude :

Segment	Année	Proportion d'homme	Exposition	Proportion de béneficiaire par segment	Montant de prestations	Nombre d'actes	Cout moyen par bénéficiaire	Cout moyen par acte
Conventionnel	2017	38%	447 131	77%	128 683 460 €	19 800 652	288 €	6 €
	2018	39%	421 486	76%	128 961 621 €	18 809 797	306 €	7 €
	2019	39%	399 196	75%	129 029 506 €	20 569 790	323 €	6 €
	2020	39%	390 607	76%	117 476 718 €	18 923 491	301 €	6 €
	2021	40%	391 528	77%	137 587 071 €	20 579 818	351 €	7 €
Standard collectif	2017	51%	66 609	11%	29 072 929 €	4 113 126	436 €	7 €
	2018	51%	55 526	10%	25 362 376 €	3 523 677	457 €	7 €
	2019	51%	47 255	9%	21 602 922 €	3 477 962	457 €	6 €
	2020	51%	41 238	8%	17 460 250 €	2 922 517	423 €	6 €
	2021	51%	36 672	7%	18 023 907 €	2 869 523	491 €	6 €
Sur-Mesure	2017	53%	68 034	12%	30 132 576 €	3 578 960	443 €	8 €
	2018	53%	79 833	14%	34 016 770 €	3 953 461	426 €	9 €
	2019	53%	83 537	16%	35 398 712 €	4 611 695	424 €	8 €
	2020	53%	84 861	16%	33 876 592 €	4 277 007	399 €	8 €
	2021	53%	82 134	16%	37 453 297 €	4 589 207	456 €	8 €
Total	2017	42%	581 773	100%	187 888 964 €	27 492 738	323 €	7 €
	2018	43%	556 845	100%	188 340 767 €	26 286 935	338 €	7 €
	2019	43%	529 988	100%	186 031 140 €	28 659 447	351 €	6 €
	2020	43%	516 707	100%	168 813 560 €	26 123 015	327 €	6 €
	2021	44%	510 334	100%	193 064 275 €	28 038 548	378 €	7 €

FIGURE 3.19 – Tableau récapitulatif des statistiques par segment et année de survenance

La proportion d'homme est assez faible pour le segment conventionnel qui représente à lui seul environ 76% des bénéficiaires par année. C'est la raison pour laquelle il y a une proportion d'homme à seulement 43% pour la population d'étude.

En 2021, un coût moyen par bénéficiaire plus grand que celui de 2020 et 2019 est constaté. Le niveau de consommation élevé observé en 2021 correspond à un changement de comportement des consommateurs couplé avec l'effet récurrent du 100% santé et non à un effet « rebond » court terme suite à la crise sanitaire ayant eu lieu en 2020. Les analyses des premiers mois de 2022 confortent d'ailleurs cette analyse.

De plus, le coût moyen par bénéficiaire est assez faible pour le segment conventionnel et relativement proche entre le segment standard collectif et le sur-mesure avec une légère supériorité pour celui du segment standard

collectif. Cette différence est due au fait d'avoir des secteurs d'activités surreprésentés avec les 4 CCN ajoutées dans la base d'étude :

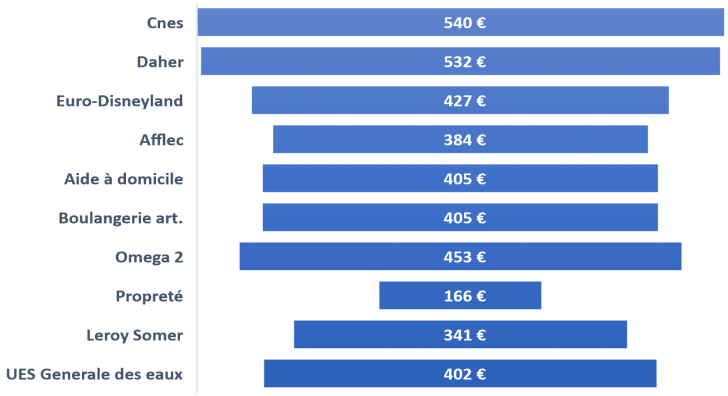


FIGURE 3.20 – Coûts moyens annuels par bénéficiaire en fonction de la gamme

Les coûts moyens annuels par bénéficiaire sont effectivement très bas pour les CCN « Afflec », « Aide à domicile », « Boulangerie art. » et particulièrement pour la CCN « Propreté » avec un coût moyen annuel par bénéficiaire de 166 euros.

D'autre part, l'analyse de la répartition du remboursement de la complémentaire par poste et année de survenance a été faites :

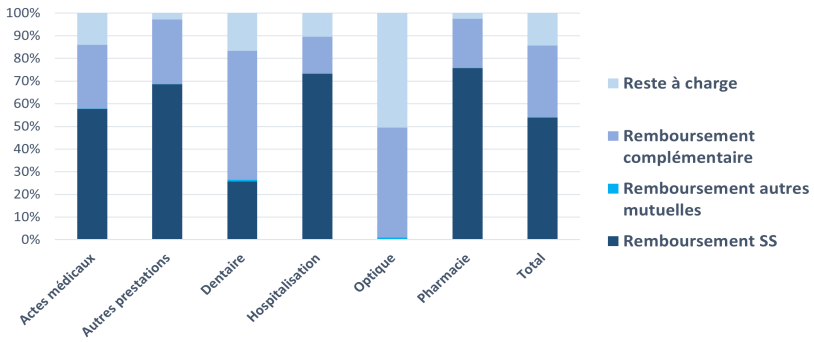


FIGURE 3.21 – Répartition du remboursement entre les différents acteurs par poste en 2021

Deux types de poste sont alors constatés :

— les postes dont les actes sont bien remboursés avec les complémentaires et la Sécurité sociale et n'ayant pas ou peu de dépassements d'honoraires : Actes médicaux, Autres prestations, Hospitalisation et Pharmacie. Les postes Pharmacie et Hospitalisation bénéficient, en effet, de meilleurs remboursements de la part de la Sécurité sociale.

— les postes avec dépassements d'honoraires dont les actes sont moins bien remboursés et qui engendrent un reste à charge plus élevé : Dentaire, Optique

La répartition de la population ayant consommé est représentée dans les graphiques par poste et par âge ci-dessous :

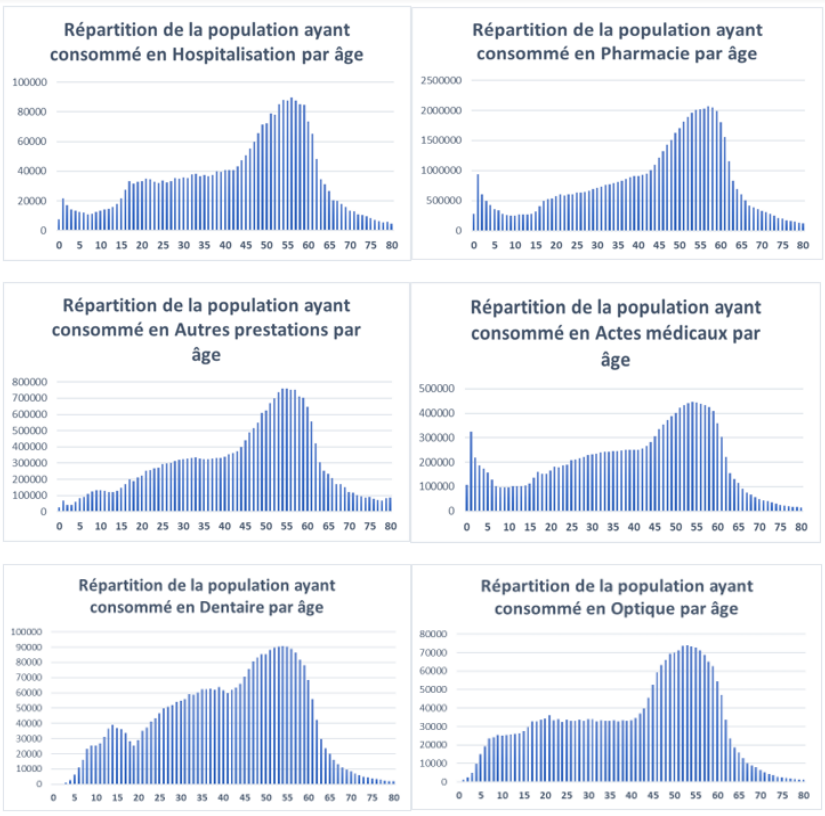


FIGURE 3.22 – Répartition de la population ayant consommé par poste et âge

Les bénéficiaires ayant eu recours à de l'Hospitalisation ont pour la plupart entre 0 et 5 ans pour les enfants et en majorité entre 50 et 62 ans pour les adultes. C'est le cas également en Pharmacie. La consommation en Pharmacie est très importante chez les nouveau-nés. Durant l'enfance et l'adolescence, la consommation est assez faible puis elle augmente à partir de 18 ans.

Pour chaque poste, la consommation augmente assez fortement entre 43 et 62 ans. Il est également observé qu'en Dentaire, il y a un pic de bénéficiaires consommant entre 12 et 18 ans correspondant à l'âge de l'adolescence où les appareils dentaires comme les bagues sont généralement plus demandés que pour d'autres tranches d'âge.

La répartition de la population n'ayant pas consommée en fonction de l'âge va maintenant être étudiée en comparaison :

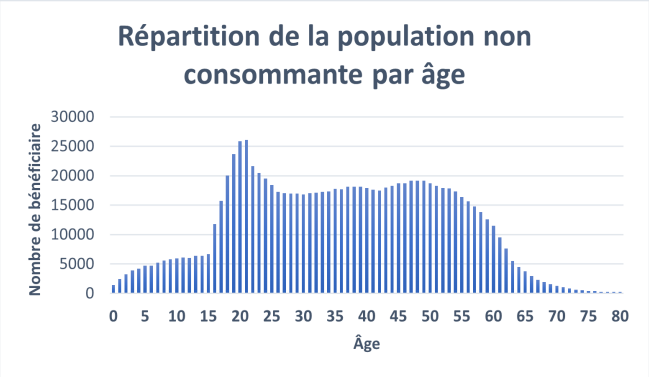


FIGURE 3.23 – Répartition de la population non consommante par âge

La population n'ayant pas consommée est plutôt jeune par rapport à l'ensemble de la population et elle est majoritairement présente sur la tranche d'âge de 15 à 25 ans. On peut en conclure que la consommation n'augmente pas forcément de manière linéaire avec l'âge selon le poste de soins et que l'âge est une variable très importante dans le cadre de la tarification en assurance santé.

Désormais, la partie sur le traitement et l'analyse des données est achevée et la prime pure par acte peut alors être modélisée.

Chapitre 4

Présentation des méthodes de tarification utilisées

4.1 Présentation de la tarification utilisée au sein d'AG2R La Mondiale

Afin de répondre aux appels d'offres sur la mise en place de contrats santé au sein de grandes entreprises (plus de 50 têtes) ou d'entreprises relevant de certaines branches de convention nationales collectives, la direction de la souscription utilise un outil interne conçu sous Excel par la direction de l'actuariat.

Cet outil permet de donner le tarif client TTC à partir des caractéristiques démographiques de l'entreprise et du niveau de couverture choisi. Pour obtenir le tarif TTC, l'outil calcule la prime pure par acte, à laquelle différents coefficients correcteurs selon les caractéristiques de l'entreprise sont appliqués. Cet outil se base sur des normes tarifaires qui ont été réalisées par la direction de l'actuariat et dont les travaux ont débutés en 2018. Le processus de construction de ces dernières normes tarifaires va être décrit dans cette partie.

4.1.1 La base de données

Les données utilisées proviennent du portefeuille Standard d'AG2R La Mondiale pour les survenances 2016-2017.

La base des assurés est agrégée par assuré de manière à ce que chaque ligne corresponde à un assuré avec les informations suivantes : sa CSP, son âge, sa localisation géographique, la garantie et le régime de l'assuré. Cette base contient 57 593 lignes et représente 640 268 assurés, proratisés sur leur durée de présence sur la période étudiée (2016-2017), ce qui représente l'exposition.

La base des sinistres utilisée contient 7 983 685 nombre d'actes. La base répertorie les différents montants alloués aux sinistres pour chaque acte de soins tels que les frais réels du sinistre, les montants remboursés par AG2R La Mondiale, les remboursements de la Sécurité sociale, les remboursements éventuels des autres mutuelles ainsi que les restes à charge de l'assuré.

4.1.2 Retraitement de la base des prestations sur la dérive

Les prestations de 2016 ont été normalisées par grand poste de soins afin de prendre en compte la dérive entre 2016 et 2017. Les montants de la dérive par poste qui ont été utilisés sont présentés dans le tableau suivant :

Poste	Derive2016
Hospitalisation	92,59%
Soins Courants	100,97%
Orthopédie et autres appareillages	90,51%
Pharmacie	101,32%
Dentaire	99,02%
Bien-être / Prévention	106,70%
Optique	98,63%
Autres prestations	101,16%

FIGURE 4.1 – Dérive annuelle en fonction du poste

4.1.3 Normalisation de la base sur l'âge

L'âge étant différent par module d'actes et par formule, il est nécessaire de normaliser la base afin d'appréhender au mieux le coût du risque selon le niveau de garantie. Les coefficients des âges ont ainsi été calculés en fonction des garanties et par grand postes de soins (Hospitalisation/Soins Courants, Dentaire, Optique). Les postes Pharmacie, Actes médicaux et Autres prestations sont rassemblés dans le grand poste "Soins Courant".
Pour cette étude, les primes pures ont été normalisées à un âge moyen de

44 ans afin d'être comparable. Cet âge a été choisi proche de l'âge moyen du portefeuille observé qui est de 42 ans. Par conséquent, chaque coefficient normalisé correspond au coefficient attribué à l'âge concerné :

$$C_i = \frac{PP_{i_{ans}}}{PP_{44_{ans}}} \tag{4.1}$$

Afin de déterminer la courbe d'âge, un lissage a été appliqué sur les courbes de prime pure en fonction de l'âge. En effet, les coefficients ont été lissés en appliquant les principes suivants :

— Les courbes croissantes ont été projetées : lorsque le coefficient de l'âge i+1 est supérieur à celui de l'âge i alors les coefficients sont correctement référencés et ils peuvent ainsi être intégrés à la courbe des âges finale ;

— Les courbes décroissantes ont été considérées comme stables : lorsque le coefficient de l'âge i+1 est inférieur à celui de l'âge i, alors il est considéré que le coefficient de l'âge i+1 est égal à celui de l'âge i. En effet, la valeur du coefficient lié à l'âge suivant ne peut être inférieure au coefficient de l'âge précédent.

En appliquant ce lissage, une courbe des âges croissante est finalement obtenue. Des courbes des âges par poste telles que celle du poste "Dentaire" peuvent alors être observées :

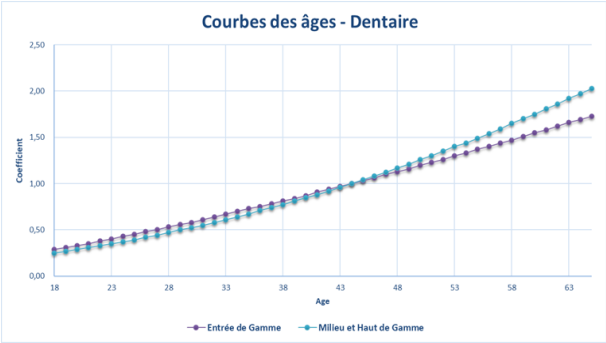


FIGURE 4.2 – Courbe des âges pour le poste Dentaire

Ces courbes des âges ainsi obtenues sont ensuite appliquées sur la base retenue en considérant pour chaque acte le coût du risque à 44 ans comme le point normalisé à « 1 ». Il y a également un tableau récapitulatif des

coefficients d'âge par poste et par niveau de garantie (EDG = Entrée de Gamme, MDG= Milieu de Gamme, HDG = Haut de Gamme) qui a été créé :

Age	Hospitalisation / Soins Courants		Optique	Dentaire	
	EDG & MDG	HDG	EDG & MDG & HDG	EDG	MDG & HDG
18	0,95	0,99	0,45	0,29	0,25
19	0,95	0,99	0,46	0,31	0,27
20	0,95	0,99	0,48	0,33	0,29
21	0,95	0,99	0,49	0,35	0,31
22	0,95	0,99	0,51	0,38	0,33
23	0,95	0,99	0,52	0,40	0,35
24	0,95	0,99	0,54	0,43	0,37
25	0,95	0,99	0,56	0,45	0,39
26	0,95	0,99	0,58	0,48	0,42
27	0,95	0,99	0,60	0,50	0,44
28	0,95	0,99	0,62	0,53	0,47
29	0,95	0,99	0,64	0,56	0,50
30	0,95	0,99	0,66	0,58	0,52
31	0,95	0,99	0,68	0,61	0,55
32	0,95	0,99	0,70	0,64	0,58
33	0,95	0,99	0,72	0,67	0,61
34	0,95	0,99	0,74	0,70	0,64
35	0,95	0,99	0,77	0,73	0,67
36	0,95	0,99	0,79	0,75	0,71
37	0,95	0,99	0,81	0,78	0,74
38	0,95	0,99	0,84	0,81	0,77
39	0,96	0,99	0,86	0,84	0,81
40	0,96	0,99	0,89	0,87	0,85
41	0,97	0,99	0,92	0,91	0,88
42	0,98	0,99	0,94	0,94	0,92
43	0,99	0,99	0,97	0,97	0,96
44	1,00	1,00	1,00	1,00	1,00
45	1,01	1,01	1,03	1,03	1,04
46	1,03	1,03	1,06	1,06	1,08
47	1,05	1,05	1,09	1,10	1,12
48	1,07	1,07	1,12	1,13	1,17
49	1,09	1,10	1,15	1,16	1,21
50	1,11	1,13	1,18	1,20	1,26
51	1,14	1,16	1,21	1,23	1,30
52	1,16	1,20	1,24	1,26	1,35
53	1,19	1,25	1,28	1,30	1,40
54	1,22	1,29	1,31	1,33	1,44
55	1,25	1,34	1,35	1,37	1,49
56	1,29	1,40	1,38	1,40	1,54
57	1,32	1,46	1,42	1,44	1,59
58	1,36	1,52	1,45	1,47	1,65
59	1,40	1,58	1,49	1,51	1,70
60	1,44	1,65	1,52	1,55	1,75
61	1,48	1,73	1,56	1,58	1,81
62	1,53	1,80	1,60	1,62	1,86
63	1,58	1,89	1,64	1,66	1,92
64	1,62	1,97	1,68	1,69	1,97
65	1,67	2,06	1,72	1,73	2,03

FIGURE 4.3 – Tableau récapitulatif des coefficients d'âge par poste et par niveau de garantie

4.1.4 Création du zonier

Il y a une consommation en frais de santé différente selon les régions où sont implantées les entreprises, expliquée essentiellement par une offre de soins disparate. Dans le cadre de ces travaux relatifs à la création des dernières normes tarifaires, un zonier a été conçu pour avoir une segmentation en fonction de la localisation géographique des assurés. Ce zonier a été réalisé en plusieurs étapes.

Première étape : Zonier Best Estimate

Dans un premier temps, une classification sur les données est réalisée avec un cluster en fonction des départements et de leurs primes pures. Cette classification permet de classer les départements en un nombre adéquat de classes en fonction de la prime pure. 7 Classes ont été retenues et la carte suivante est alors obtenue :

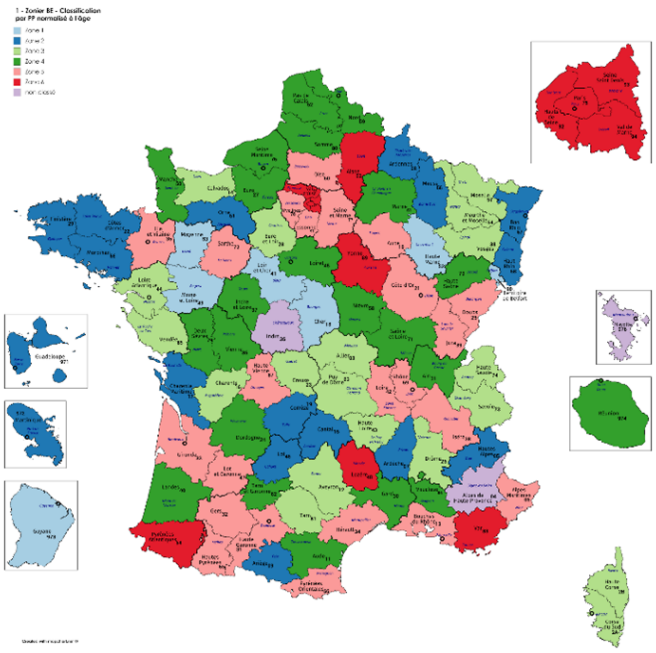


FIGURE 4.4 – Carte du zonier Best Estimate

Ensuite, avec un GLM, les coefficients de chaque zone sont obtenus :

Zone	pop	Coeff	Coeff Finaux
Zone 1	1 771	1,00	62%
Zone 2	15 345	1,20	74%
Zone 3	18 214	1,38	85%
Zone 4	22 489	1,51	93%
Zone 5	46 257	1,62	100%
Zone 6	30 603	1,73	106%

FIGURE 4.5 – Coefficients du zonier Best Estimate

Les coefficients finaux sont déterminés en divisant les coefficients par le coefficient de la zone d'étude (zones 5 et 6) qui est calculé de la manière suivante :

$$\text{Coeff de la zone d'étude} = \frac{(pop(zone5)\times coeff(zone5)+pop(zone6)\times coeff(zone6))}{pop(zone5)+pop(zone6)}$$

Deuxième étape : suppression des départements peu significatifs

Les départements avec très peu de population et donc peu significatifs sont écartés du zonier pour les reclasser ensuite lors de l'étape 3. Les départements non significatifs correspondent à des départements ayant moins de 1500 personnes assurées. Une carte de ce type est obtenue :

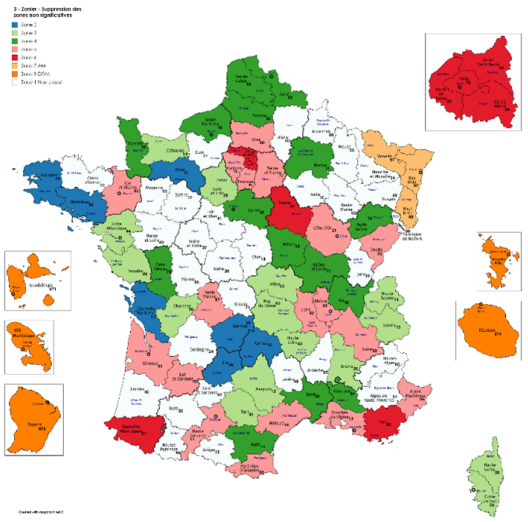


FIGURE 4.6 – Carte du zonier avec les zones à reclasser

Les zones blanches correspondent aux départements qui vont être reclas-

sés. Par ailleurs, deux nouvelles zones ont aussi été créées : La zone de l'Alsace Moselle et des dom tom. En effet, ces zones doivent être étudiées différemment car elles comportent certaines particularités qui doivent entraîner des différences dans les normes tarifaires.

Troisième étape : données externes, offre de soins

Pour faire le reclassement des zones blanches, un zonier est créé à partir de données externes composées de données démographiques et d'offres de soins.

Une ACP a été faite sur cette base pour mieux visualiser les données. L'ACP synthétise les données en seulement quelques nouvelles variables appelées **composantes principales**. Voici le résultat sur les corrélations entre les variables :

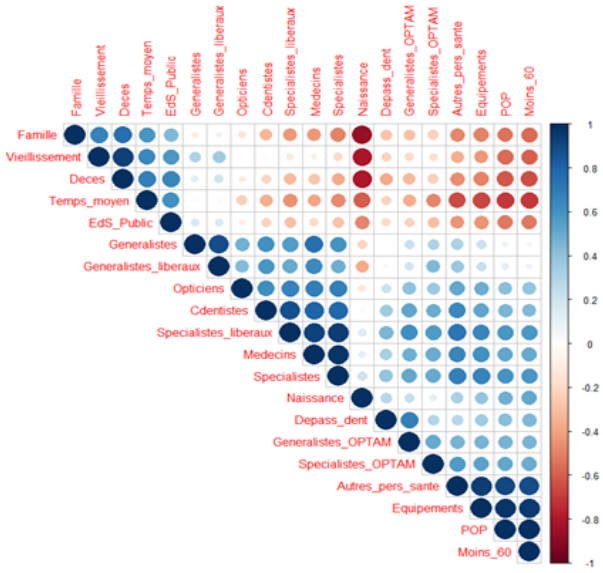


FIGURE 4.7 – Résultat de l'ACP

Ensuite, les valeurs propres sont observées pour mesurer la quantité de variance expliquée par chaque axe principal.

Quatrième étape : Classement des zones en fonction des données externes

Les départements en zone blanche sont reclassés selon la méthodologie suivante : la zone du département retenue est celle qui lui est le plus proche en termes de données externes, pour lequel le zonier BE et le zonier issu des précédents travaux de normes tarifaires (zonier AN) sont cohérents (zones quasi identiques).

Exemple : Reclassement du département 22 : le département qui lui ressemble est le 28 qui est en zone 4 sur le zonier BE et aussi en zone 4 sur le zonier AN et comme le département 22 est aussi en zone 4 sur le zonier AN, il est donc reclassé sur la zone 4.

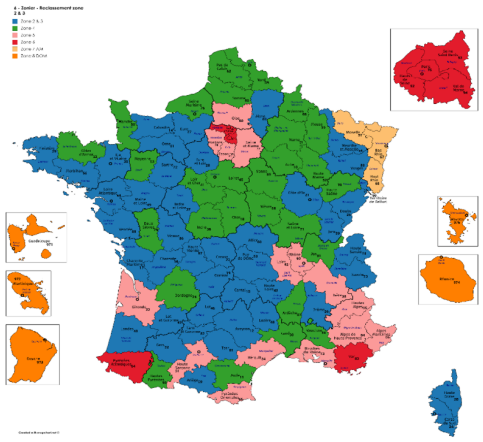
Cinquième étape : Reclassement des zones incohérentes

Plusieurs départements comme la Côte d'Or (21), le Doubs (25), l'Ille-et-Vilaine (35), le Lot-et-Garonne (47) , la Haute-Vienne (87), l'Yonne (89) ont été classés sur des zones élevées alors qu'il ne semble pas qu'ils devraient l'être. La même méthode que celle de l'étape 4 est appliquée et les résultats suivants sont obtenus avec un GLM :

Parameter	Level1	DF	Estimate	StdErr	LowerWald CL	UpperWald CL	ChiSq	ProbChiSq	coeff	borne_Inf	borne_sup
Intercept		1	6,25	0,03	6,19	6,31	38 610,68	-	518,15	486,83	551,48
Zone	Zone DOM	1	- 0,24	0,04	- 0,32	- 0,17	37,83	0,00	0,78	0,73	0,85
Zone	Zone AM	1	- 0,16	0,09	- 0,34	0,01	3,57	0,06	0,85	0,72	1,01
Zone	Zone 6	1	0,19	0,03	0,13	0,25	39,85	0,00	1,21	1,14	1,28
Zone	Zone 5	1	0,14	0,03	0,08	0,19	23,04	0,00	1,15	1,08	1,21
Zone	Zone 4	1	0,06	0,03	0,00	0,12	4,13	0,04	1,06	1,00	1,13
Zone	Zone 3	1	- 0,01	0,03	- 0,07	0,05	0,04	0,85	0,99	0,94	1,05
Zone	Zone 2	0	-	-	-	-			1,00	1,00	1,00
csp	NON CADRE	1	- 0,26	0,02	- 0,31	- 0,22	143,64	0,00	0,77	0,74	0,80
csp	ENSEMBLE PERSONNEL	1	- 0,14	0,02	- 0,18	- 0,10	43,27	0,00	0,87	0,84	0,91
csp	CADRE	0	-	-	-	-			1,00	1,00	1,00
COSEX	M	1	- 0,29	0,01	- 0,32	- 0,26	384,35	0,00	0,75	0,73	0,77
COSEX	F	0	-	-	-	-			1,00	1,00	1,00
Scale		1	2,70	0,29	2,18	3,34			14,89	8,85	28,36

FIGURE 4.8 – Résultat du GLM après reclassement des zones incohérentes

Au vu des résultats du GLM, la zone 3 n'est pas significative et ayant un coefficient proche de la zone 2, les zones 2 et 3 sont donc regroupées.
Le zonier final obtenu est représenté ci-dessous :



Carte du zonier

Zone	pop	Coeff	Coeff Finaux
Zone 23	36 414	1,00	87%
Zone 4	24 273	1,07	93%
Zone 5	37 037	1,15	100%
Zone 6	28 201	1,21	105%
Zone AM	1 049	0,85	74%
Zone DOM	7 809	0,79	68%

Coefficients du zonier

FIGURE 4.9 – Carte et coefficients du zonier final

4.1.5 Coefficients correcteurs des caractéristiques de l'assuré

La méthode utilisée pour déterminer les coefficients correspondant aux profils des assurés est un modèle linéaire généralisé sur la base de données utilisée.

Le GLM calcule les différents coefficients liés aux caractéristiques des assurés tels que la catégorie socio professionnelle et le sexe.

Ainsi, l'utilisation du modèle linéaire généralisé permet de récupérer les coefficients à appliquer aux primes pures afin de pouvoir les adapter en fonction des caractéristiques des assurés.

4.1.6 Calcul des primes pures

AG2R La Mondiale utilise une méthode interne afin de calculer les primes pures qui seront ensuite indiqués au sein de l'outil de tarification de l'entreprise pour le produit appelé « Simpleo ». Simpleo est l'offre sur-mesure lancée par le groupe en Juin 2019, avec les premiers contrats à effet 01/01/2020. Sa cible initiale étaient les entreprises entre 50 et 299 salariés.

Depuis 2022, notamment sous l'impulsion de la direction des grands comptes, Simpleo peut être utilisé jusqu'à 499 têtes. Pour chaque acte, la prime pure est calculée en fonction du niveau de garantie. Pour calculer la prime pure, plusieurs méthodes sont utilisées :

— Prime pure globale

— Modélisation de la prime pure par niveau de garantie en analysant la tendance et en maximisant la métrique R^2.

La méthode de la prime pure globale est utilisée pour les actes dont les dépassements d'honoraires sont inexistants. Ainsi, la garantie restera toujours la même indépendamment de la formule souscrite par l'assuré (le niveau de couverture).

Cette méthode est utilisée pour calculer les primes pures des actes forfait hospitalier, transport et pharmacie. Pour chaque acte, les résultats sont retranscrits dans un tableau (par confidentialité, il sera observé des chiffres fictifs) :

(A)	(B)	(C)	(D)	(E)	(F)	(G)	(H)	(I)
Niveau	**Garantie**	**Population**	**Âge moyen**	**Coeff âge**	**Prestations**	**PP**	**PP normée**	**PP Simpleo**
1	100%	3902	49	1,09	8 350 €	2,14 €	1,96 €	1,40 €
2	100%	17010	44	1	25 570 €	1,50 €	1,50 €	
3	100%	14563	43	0,99	18 609 €	1,28 €	1,29 €	
4	100%	10608	45	1,01	12 416 €	1,17 €	1,16 €	

Tableau 4.1 : Tableau de calcul de la prime pure par acte et niveau de garantie

Avec :

— A et B = Niveau de couverture et pourcentage de remboursement de la garantie ;

— C = Correspond à la population bénéficiant de la garantie (Exposition)

— D = Age moyen de la population

— E = Coefficient d'âge venant de la courbe des âges

— F = Montant des remboursements de l'entreprise

— G = Prime pure calculée en appliquant le calcul PP = $\frac{Prestations}{Population}$

— H = Prime pure normalisée à l'âge (44 ans) calculée en appliquant le calcul PP normalisée = $\frac{Prime\ Pure}{Coeff_{age}}$

— I = Prime pure finale, moyenne pondérée des primes pures normées en fonction de la population

Pour les autres actes pouvant présenter des dépassements d'honoraires, c'est la méthode de la courbe de tendance de la prime pure qui est utilisée. Dans le cas de cette méthode, il est déterminé, pour chaque bénéficiaire, une fonction qui passe au plus proche des points connus. En effet, dans le cas où une garantie est présente sur plusieurs formules, il faut connaitre le coût moyen pour tous les niveaux de couverture. Plusieurs fonctions sont possibles :

- Linéaire : $y = Ax$
- Polynomiale 2nd degrés : $y = Ax^2 + Bx + C$
- Exponentiel : $y = A \times exp(Bx)$
- Logarithme : $y = log(Ax) + B$
- Puissance : $y = Ax^B$

Le choix de la fonction se fera en fonction de la métrique R^2 : la fonction dont le R^2 est le plus proche de 1 sera choisie tout en restant cohérent avec les résultats obtenus.

Le R^2, nommé aussi le coefficient de détermination, est compris entre 0 et 1, et croît avec l'adéquation de la régression au modèle. C'est un indicateur qui permet de mesurer la qualité d'une régression linéaire.

Cette méthode est illustrée ci-dessous (par confidentialité, les chiffres indiqués sont toujours fictifs) :

Niveau	Garantie	Population	Age moyen	Coeff âge	Prestations	PP	PP normée	PP SIMPLEO
1	150 €	3 098	42	0,94	9 502 €	3,07 €	3,26 €	3,19 €
2	400 €	14 565	45	1,03	132 455 €	9,09 €	8,83 €	9,21 €
3	600 €	20 263	44	1	303 884 €	15,00 €	15,00 €	15,64 €
4	900 €	11 784	45	1,03	271 231 €	23,02 €	22,35 €	21,88 €

Tableau 4.2 : Tableau du calcul de la prime pure pour un acte avec plusieurs niveaux de garanties

Dans un premier temps, il faut placer les points sur un graphique (Garantie x PP normé) :

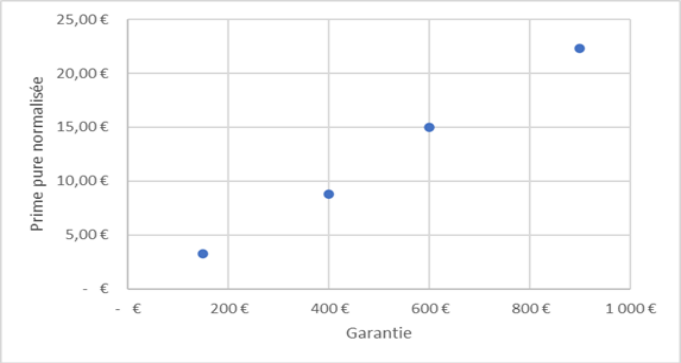


FIGURE 4.10 – Nuage de points des primes pures normalisées en fonction du niveau de garantie

Il faut tracer ensuite la courbe de tendance qui passe le plus proche des points connus :

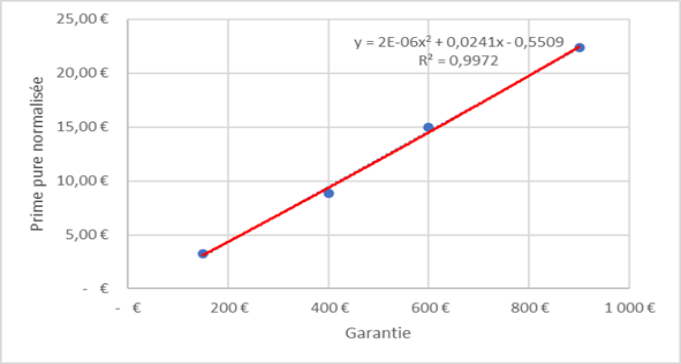


FIGURE 4.11 – Courbe de tendance des primes pures normalisées en fonction du niveau de garantie

La prime pure « Simpleo » est calculée en appliquant l'équation de la fonction trouvée. La méthode consiste à placer sur un graphique les quatre primes pures « Simpleo » calculées et données ci-dessus. Ensuite, différentes courbes de tendances dont les lois sont celles citées ci-dessus sont testées.

Le choix de conserver celle dont le R^2 est le plus proche de 1 est fait. Ainsi, les primes pures allouées aux garanties correspondantes sont obtenues. Par la suite, les coefficients correctifs de la CSP, du sexe et de la localisation sont

appliqués à ces primes pures afin d'obtenir les primes pures finales.

L'objectif du mémoire est de challenger ces normes tarifaires, construites sur des données 2016 – 2017 avec des données plus récentes et d'utiliser les modèles GAM en testant un nouvel outil de création de normes tarifaires nommé « Akur8 ».

4.2 Présentation d'un nouvel outil de tarification : Akur8

Dans le cadre de ce mémoire, une nouvelle méthode de tarification est mise en œuvre en utilisant les modèles additifs généralisés. L'outil de modélisation Akur8 permet d'exploiter les modèles GAM et le choix a donc été fait d'utiliser cet outil pour réaliser les normes tarifaires.

L'utilisation des modèles additifs généralisés (GAM) dans Akur8 permet de lutter contre la problématique de "boîte noire" des modèles d'intelligence artificielles qui les rendent inapplicables à la tarification des assurances du fait notamment des exigences réglementaires de transparence des régulateurs. Ainsi, le logiciel AKUR8 sera utilisé afin de mettre en place les différents modèles de tarification.

4.2.1 Description des modèles utilisés

Akur8 utilise les modèles additifs généralisés ayant la structure suivante :

$$\hat{y}(x) = g^{-1}(\sum_{j=1}^{N} \beta_j(X_j)) \tag{4.2}$$

où :

- $\hat{y}$ représente le vecteur de prédictions ;
- β_j représente des fonctions constantes par morceaux qui permettent de prendre en compte tout type de forme non paramétrique. β_j correspond au coefficient de la variable j ;
- X_j représente le vecteur prédictif de la variable j ;
- g^{-1} représente l'inverse de la fonction de lien utilisée.

Les modèles additifs généralisés permettent de décrire une relation non-linéaire entre des prédicteurs et une variable réponse. Tout comme les modèles linéaires généralisés, ils peuvent inclure des effets aléatoires afin de représenter des données groupées.

Dans le cadre de cette étude, le but recherché est la création des modèles de coût total avec la distribution Tweedie qui vont prédire directement la prime pure. Une fonction de lien de type logarithmique sera ainsi utilisée pour la création des modèles. Cette fonction de lien permet aux modèles d'acquérir une structure multiplicative car l'inverse d'une fonction logarithmique est une fonction exponentielle et la structure de modèle sera donc la suivante :

$$\hat{y}(x) = exp(\sum_{j=1}^{N} \beta_j(X_j)) = \prod_{j=1}^{N} \beta_j(X_j) \tag{4.3}$$

Ainsi, la déduction des différents coefficients β_j permettra d'ajuster les modèles. Le résultat final sera de la forme :

$\hat{y}(x) = Base \times Coefficient_1 \times Coefficient_2 \times \ldots \times Coefficient_n$ avec n étant le nombre de coefficients.

L'intercept « Base » est calculé à la fin du fit du modèle en résolvant un problème d'optimisation selon la loi choisie. La méthode de calcul de l'intercept sera détaillée dans la partie consacrée à la création du modèle de coût total avec Akur8.

Dans les modèles GAM usuels, les coefficients β_j sont calculés selon l'approche suivante :

$$\beta^* = arg\,max(p(y|\hat{y}\beta)) = arg\,max(LogLikelihood(x, y, \beta)) \tag{4.4}$$

Avec $p(y|\hat{y}\beta)$ qui suit la distribution choisie pour le modèle. Dans cette étude, il s'agit de la distribution de Tweedie. Les distributions Tweedie appartiennent à la classe des modèles de dispersion exponentielle, souvent utilisée dans les modèles linéaires généralisés. C'est une famille de distributions de probabilité qui comprend des distributions continues telles que la distribution Normale et Gamma, la distribution de Poisson exclusivement discrète,

et la classe de distributions composées mixtes Poisson-Gamma qui ont une quantité importante de zéros.

Toutefois, dans Akur8, cette métrique n'est jamais estimée de manière directe. Les β_j sont déterminés sur une base de fonctions en escaliers. Des contraintes sont ajoutées afin d'éviter le surajustement des coefficients β_j fittant un GLM dit « naïf ». L'approche du Maximum de Vraisemblance intègre ainsi directement l'hypothèse préalable :

$$\beta^* = arg\,max_\beta(p(y|\hat{y}\beta) \times p_{prior}(\beta)) \tag{4.5}$$

Parmi les contraintes ajoutées et intégrées au modèle, il est notamment supposé que les fonctions $\beta_j(...)$ sont constantes. Cette hypothèse implique que chaque palier des fonctions β_j sont égaux ainsi que l'hypothèse de nullité suivante : $\beta_{i,j} = \beta_{i+1,j}$

Cette hypothèse est confrontée aux données disponibles via le test statistique Chi-square. Ainsi, si l'hypothèse est rejetée, alors $\beta_{i,j} \neq \beta_{i+1,j}$ et les paliers correspondant aux deux coefficients $\beta_{i,j}$ et $\beta_{i+1,j}$ ne seront pas égaux. Par conséquent, la fonction β_j ne sera pas constante. Cependant, si l'hypothèse n'est pas rejetée, alors $\beta_{i,j} = \beta_{i+1,j}$. Le fonctionnement de cette hypothèse dans les deux cas est illustré avec le schéma suivant :

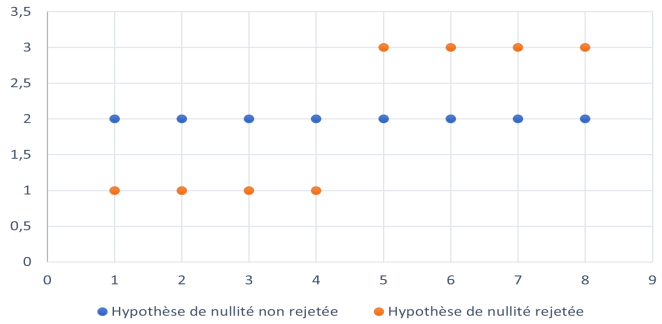


FIGURE 4.12 – Schéma du fonctionnement des coefficients beta j en fonction du résultat de l'hypothèse

Dans le cas où l'hypothèse de nullité est rejetée, ce rejet se fait entre $\beta_{4,j}$ et $\beta_{5,j}$ et la fonction se caractérise alors comme une fonction constante sur deux

morceaux. Par opposition, la fonction est constante dans le cas où l'hypothèse de nullité n'est pas refusée.

Les coefficients β_j peuvent alors être déterminés plus simplement. Néanmoins, l'hypothèse génère des groupements de valeurs au sein des variables pouvant être minimisés par le seuil de rejet de l'hypothèse de nullité. En conséquence, les fonctions β_j acquièrent une structure de fonction constante par morceaux.

En effet, un seuil faible de rejet produira de nombreux coefficients différents (et donc un grand « nombre de morceaux » de la fonction β_j) mais le modèle sera sujet au surapprentissage. Au contraire, un seuil élevé engendrera de larges groupes de coefficients et donc un plus petit « nombre de morceaux ». Le modèle sera alors beaucoup plus robuste mais moins performant vis-à-vis des données.

Par conséquent, Akur8 créera plusieurs modèles et l'Actuaire doit ensuite décider quel modèle est le plus adéquate à être utilisé en analysant les statistiques de performance.

4.2.2 Création de modèles sous Akur8

Akur8 permet de créer des modèles à partir de bases de données exportées préalablement construites. Avant de créer le modèle, il faut définir les objectifs du projet de modélisation en sélectionnant la variable à prédire, la variable d'exposition, la variable temporelle (l'année de survenance), la variable de partitionnement ainsi que les variables qui seront utilisées lors de la modélisation.

La variable de partitionnement va permettre d'entrainer le modèle sur les données d'apprentissage (80% de la base) et pouvoir tester sa performance sur les données de test (20% de la base). Ainsi, une base de données séparée en deux sera obtenue car la partie validation sera exécutée par cross-validation.

Dès lors, la création des modèles peut débuter. Pour chaque acte, il peut être créé un modèle de fréquence, un modèle pour le coût moyen ou un modèle du coût total.

Modèle « Fréquence X Coût Moyen »

Dans l'objectif de modéliser la prime pure, un modèle « Fréquence X Coût Moyen » peut être mis en place dans Akur8 en modélisant dans un premier temps la fréquence et le coût moyen séparément puis en agrégeant les deux modèles.

Pour modéliser la fréquence, un modèle GAM de loi de Poisson a été choisi en utilisant une fonction de lien de type logarithmique afin de permettre aux modèles d'avoir une structure multiplicative.

Pour modéliser le coût, un modèle GAM de loi Gamma a été choisi en utilisant également une fonction de lien de type logarithmique pour les mêmes raisons que celles du modèle de fréquence.

Les modèles de prime pure peuvent ensuite être créés après avoir validé les modèles de fréquence et de coût. En effet, les modèles de fréquence et de coût moyen peuvent être agrégés sous forme de produit afin d'obtenir un modèle de prime pure.

Il est ainsi possible de créer un modèle de prime pure pour chaque acte. Les modèles de fréquence et de coût moyen étant des modèles de type GAM et donc des modèles à coefficients multiplicatifs, le modèle de prime pure par acte sera également un modèle de type GAM à coefficients multiplicatifs.

Modèle de coût total

Il y a également la possibilité de créer directement un modèle du coût total. Afin de modéliser le coût total, il a été choisi de prendre un modèle GAM de loi Tweedie en utilisant également une fonction de lien de type logarithmique. Ce modèle a été choisi dans le cadre de cette étude pour des raisons de rapidité et d'efficacité.

Le paramètre de dispersion p de la loi Tweedie, compris entre 1 et 2, est fixé arbitrairement à 1,5 dans Akur8 et il peut être modifié selon le niveau de variance souhaité pour le modèle.

Lorsque le modèle a été créé, l'analyse des statistiques du modèle peut être

effectué. Après validation du modèle, l'Actuaire peut exporter un fichier regroupant l'intercept et les différents coefficients multiplicatifs liés aux différentes valeurs possibles que peuvent prendre les variables. Le calcul de la prime pure est le suivant :

$$Prime\ Pure_i = Base \times C_{1i} \times C_{2i} \times ... \times C_{ni}$$

Avec la Base représentant l'intercept et les C_{ji} qui correspondent aux coefficients prédits par le modèle pour la variable i.

Dans le cas du modèle Tweedie qui sera utilisé dans cette étude, l'intercept est obtenu par cette formule :

$$Intercept\ =\ \log\left(\frac{\sum_i wi \times yi \times \widehat{yi}^{\,1-p}}{\sum_i wi \times \widehat{yi}^{\,2-p}}\right)$$

Où :

- w_i = poids de chaque observation ;
- y_i = observations ;
- $\hat{y}_i$ = prédictions ;
- p = paramètre de dispersion de la loi Tweedie.

Akur8 permet également de sommer chaque modèle de coût total par actes afin d'avoir le modèle de coût total par poste. Cependant, la somme de modèles GAM ne donne pas un modèle GAM. Akur8 ne peut en effet pas toujours utiliser le cadre GAM avec des coefficients explicites pour afficher les résultats d'un modèle agrégé.

Dans ce cas, il est toujours possible d'inspecter les variables mais les dépendances partielles seront affichées à la place des coefficients. Effectivement, lorsqu'un modèle ne peut pas être exprimé dans un format GAM, aucun coefficient ne peut être utilisé pour comprendre le comportement du modèle par rapport aux différentes variables. Pour résoudre ce problème, Akur8 utilise donc une méthode couramment utilisée dans le domaine de l'apprentissage automatique appelée « Partial Dependance Plot ».

Dans le cadre de cette étude, afin de déterminer la prime pure globale par acte, il a été décidé de ne pas agréger les modèles par acte dans l'outil Akur8 car cette solution ne correspondrait plus à des modèles GAM et ne donnerait pas la possibilité d'obtenir les coefficients de chaque variable en fonction d'un acte. Or, l'objectif recherché est d'obtenir des normes tarifaires modulaires par acte et l'agrégation des modèles par acte se fera donc manuellement en utilisant Excel.

Ensuite, lors de l'étape de la génération du modèle, il est nécessaire de définir certains paramètres :

- **Le nombre de variables sélectionnées dans les modèles** : les modèles créés comprendront un certain nombre de variables appartenant à l'intervalle choisi. La complexité augmentera avec le nombre de variables comprises dans le modèle.
- **Le nombre de pas de parcimonie** : chaque valeur de parcimonie correspond à un groupe de modèles ayant le même nombre de variables. Par exemple, un nombre de pas de parcimonie égal à 5 va générer cinq groupes de modèles, chaque groupe ayant un nombre spécifique de variables dans l'intervalle choisi.
- **Le nombre de pas de lissage** : il indique le nombre de niveaux de lissage. Par exemple, avec un lissage de 6, l'outil générera, pour chaque niveau de parcimonie, 6 modèles différents. Moins le modèle est lisse et plus il sera fidèle aux données tandis qu'un modèle plus lisse sera moins fidèle aux données mais beaucoup plus robuste.

Les paramètres de parcimonie et de lissage définissent le nombre de modèles créés : des valeurs plus élevées signifient des choix plus précis dans la création des modèles, mais aussi un temps de calcul plus long. Pour chaque combinaison de lissage et de parcimonie, 5 modèles différents sont créés :

— 4 modèles de validation croisée sur le sous-ensemble de modélisation de la base de données

— 1 modèle sur l'ensemble du sous-ensemble de modélisation.

L'utilisation de la validation croisée permet d'avoir une estimation fiable des performances des modèles créés. La validation croisée est une technique très polyvalente qui permet de sélectionner des modèles et d'estimer l'erreur de

généralisation. Dans Akur8, les modèles sont automatiquement validés de manière croisée sur l'ensemble de modélisation.
Par la suite, les scores de performance des modèles seront les scores moyens sur les 4 échantillons. Akur8 peut alors construire un ensemble de modèles basés sur des combinaisons distinctes de paramètres avec la méthode appelée « Grid Search ».

4.2.3 Visualisation des résultats et choix du modèle

Akur8 a ainsi recours à une méthode appelée « Grid Search » qui est une méthode d'optimisation permettant de tester une série de paramètres et de comparer les performances pour déterminer le meilleur paramétrage.

L'illustration ci-dessous est une extraction de l'outil à la suite de la mise en place de la méthode du GridSearch. Les modèles créés dans un graphique peuvent être observés, où chaque point représente un modèle différent :

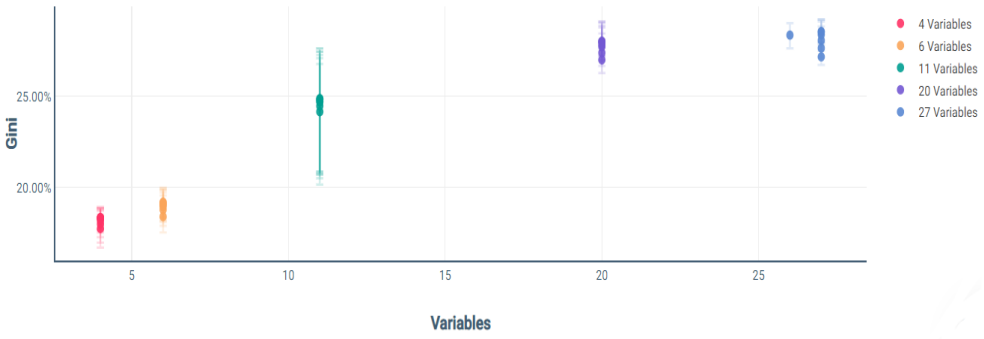


FIGURE 4.13 – Exemple de Grid Search de l'outil Akur8

Dans ce cas-là, il a été demandé 5 niveaux de parcimonie et 7 niveaux de lissage donc 35 solutions au problème de modélisation ont été élaborées avec différents nombres de variables et différents lissages. Le nombre de variables dans le modèle est indiqué sur l'axe horizontale et les performances hors échantillon du modèle sont indiqués sur l'axe vertical. La métrique par défaut est le Gini.

Ce graphique permet de visualiser que plus le nombre de variables augmente,

plus le modèle est performant. Il faut donc choisir le modèle afin d'obtenir le meilleur équilibre possible entre la complexité du modèle et ses performances. En effet, un modèle avec plus de variables sera plus difficile à interpréter et augmentera le coût de calcul.

4.2.4 Inspection du modèle selectionné

Lorsque le modèle est choisi, il y a la possibilité d'avoir un aperçu du modèle et des informations sur les performances du modèle.

Graphique de spread

Les variables sélectionnées par le modèle ainsi que leurs importances peuvent être observés. Les spreads reposent sur les coefficients calculés par le modèle. Il existe deux types de spreads proposés par l'outil :

— **Les spreads 100/0** : Ces types de spread permettent de récupérer l'influence de la variable observée sur la variable à prédire. Il est calculé de la manière suivante :

$$Spread = \frac{Max(Coefficient)}{Min(Coefficient)} - 1 \tag{4.6}$$

Voici ci-dessous une illustration de ce spread 100/0 :

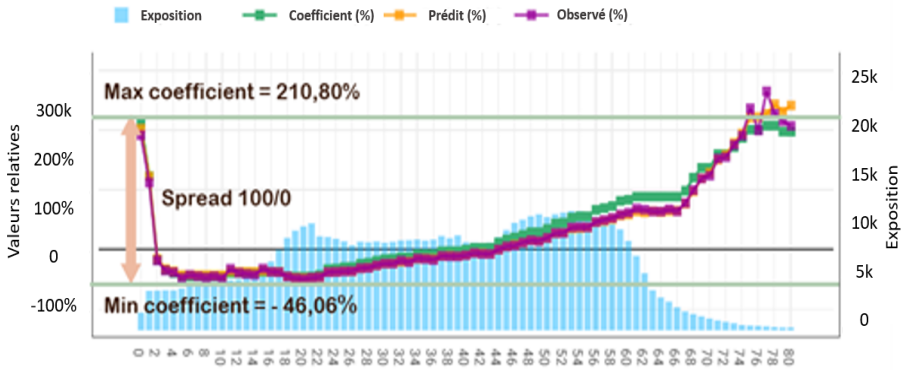


FIGURE 4.14 – Exemple de spread 100/0 de l'outil Akur8

Les coefficients $\beta_{i,j}$ calculés par le modèle sont représentés par la courbe verte.

— **Les spreads 95/5** : Ils correspondent à la même mesure que les spreads 100/0 après suppression des coefficients correspondant aux 5% les plus élevés et les plus bas du jeu de données. L'outil supprime les coefficients les plus élevés jusqu'à ce que l'exposition totale tombe à 95% et supprime également les coefficients les plus faibles jusqu'à ce que l'exposition totale tombe à 95% de l'exposition de départ. Le spread 95/5 permet ainsi de ne pas tenir compte des valeurs extrêmes.

De plus, les spreads 100/0 et 95/5 pour chacune des variables sélectionnées dans le modèle sont affichés :

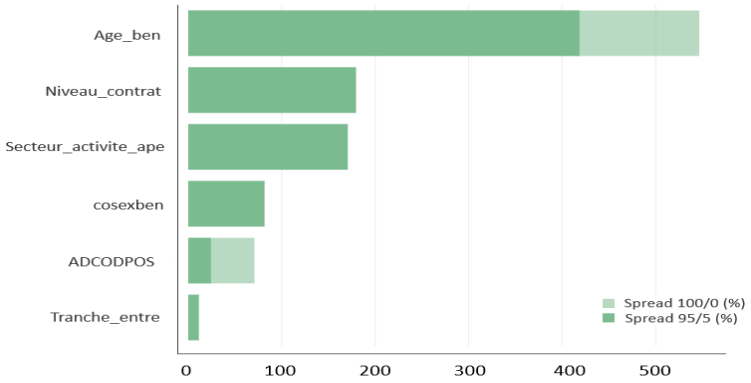


FIGURE 4.15 – Exemple de spread 100/0 et 95/5 de l'outil Akur8

Courbe de Lorenz

La courbe de Lorenz décrit la qualité des prédictions du modèle. Elle est construite en ordonnant toutes les observations de la fréquence estimée la plus basse à la plus élevée et en calculant le nombre cumulé de sinistres observés pour chaque observation.

La courbe représente ainsi une mesure de segmentation du portefeuille analysé. Une prédiction aléatoire du risque suivrait la diagonale tandis que des prédictions parfaites conduiraient à une courbe de Lorenz augmentant très fortement pour atteindre 100% et y rester. Dans le cas de la courbe de Lorenz ci-dessous, 40% des contrats prédisant les coûts les plus élevés représentent 61% des coûts observées sur le portefeuille :

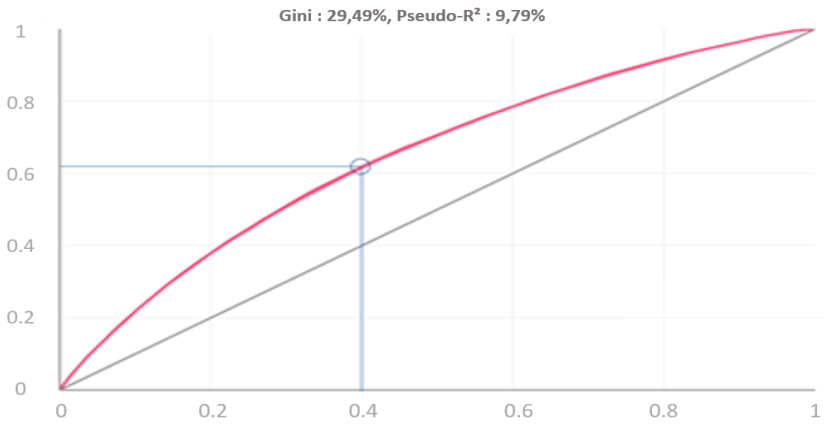


FIGURE 4.16 – Exemple de courbe de Lorenz de l'outil Akur8

Courbe de Lift

La courbe de Lift est construite en triant les prédictions de la plus basse à la plus haute et en les regroupant en 20 groupes qui représentent chacun 5% des prédictions. Pour chaque groupe, la prédiction moyenne du modèle et l'observation moyenne sont affichés afin d'évaluer la qualité des prédictions du modèle.

Les courbes de Lift sont construites sur les données de test de la validation croisée afin de refléter la performance hors échantillon du modèle :

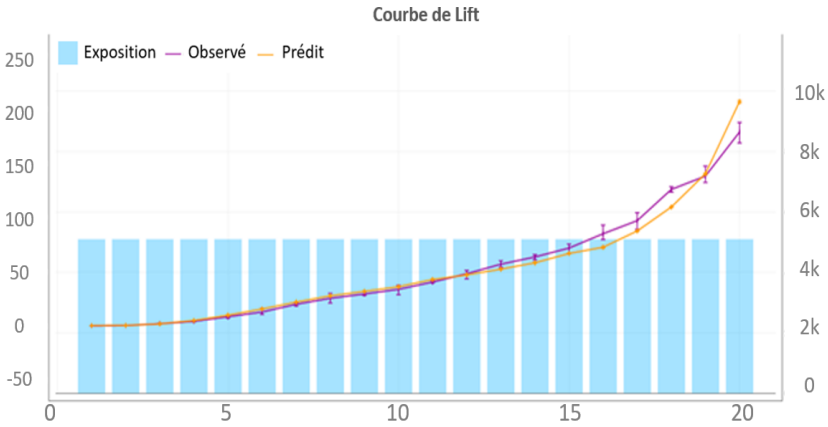


FIGURE 4.17 – Exemple de courbe de Lift de l'outil Akur8

Résidus quantiles randomisés

Dans un modèle GAM valide, les résidus sont censés suivre une loi normale. Donc, les résidus quantiles normalisés randomisés doivent suivre en théorie une loi normale centrée réduite.

Voici leur expression :

$$R = F^{-1}_{N(0,1)}(\hat{F}_i(Y_i))$$

Où $F_{\hat{i}}$ est la fonction de répartition estimée pour la réponse Y_i.

Les résidus quantiles randomisés sont utiles pour voir l'adéquation ou non de la loi utilisée par le modèle à la distribution des données. Afin de valider cette adéquation ou non, il faut regarder la centralité des résidus par rapport à l'axe des ordonnés et l'existence d'une « tache » uniforme des résidus :

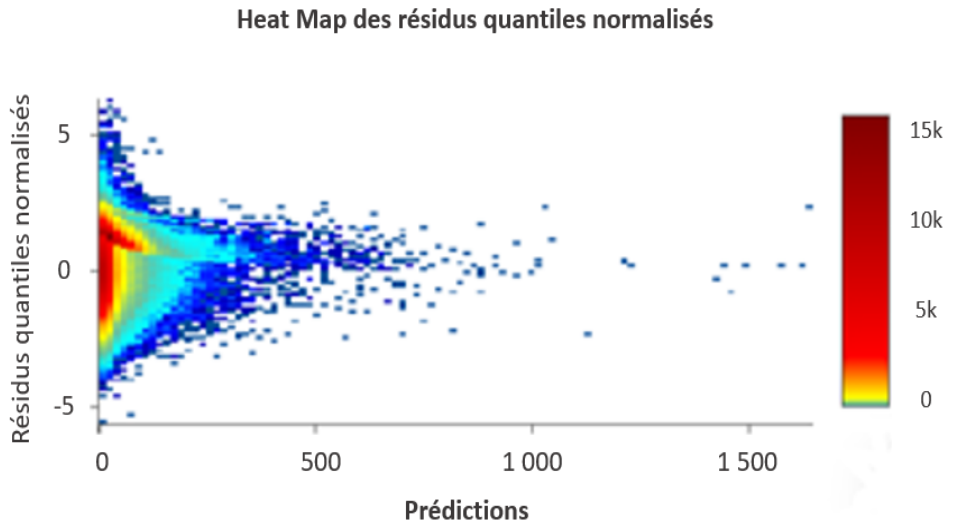


FIGURE 4.18 – Exemple de courbe de résidus quantiles normalisés dans l'outil Akur8

Dans le cas où il y aurait deux taches bien distinctes sur le graphique, il serait intéressant de procéder à un découpage de la variable observée. En effet, cette situation traduirait deux comportements distincts au sein de la variable et chacun devrait donc être analysé séparément.

Toutefois, n'ayant pas d'explications précises sur le fonctionnement de ces résidus et du mode de calcul des quantiles, il a été décidé de se concentrer sur d'autres métriques de performances en priorité avant d'étudier les résidus afin de confirmer le choix d'un modèle.

Statistiques

Il y a également un récapitulatif des statistiques du modèle telles que le coefficient de Gini, le pseudo R^2, la RMSE, la déviance, la MAE ou encore le nombre de variables utilisés.

Pour évaluer un modèle, il a été choisi de prendre en compte les graphiques de spread, le coefficient de Gini avec la courbe de Lorentz, la courbe de Lift, le Pseudo-R^2 ainsi que le nombre de variables :

METRIC	TRAIN FULL	TRAIN K-FOLD	TEST K-FOLD
GINI	29.55 %	29.57 %	29.49 %
NORM. GINI	38.66 %	38.68 %	38.58 %
PSEUDO-R²	9.83 %	9.84 %	9.79 %
RMSE	122	122	121.9
DEVIANCE	7667000	5750000	1918000
AVG. DEVIANCE	17.15	17.15	17.16
MAE	50.49	50.49	50.5
NB. VARIABLES	5	5	5

FIGURE 4.19 – Exemple de statistiques dans Akur8

4.2.5 Optimisation du modèle

Lorsque l'analyse du modèle choisi est terminée, des améliorations peuvent encore y être apportées. Premièrement, un meilleur niveau de lissage des coefficients peut être choisi. Lors du choix d'un modèle, si les coefficients du

modèle sont trop robustes ou bien que les coefficients du modèle semblent capturer du bruit, il est possible de modifier le niveau de lissage afin de sélectionner le meilleur compromis entre ces deux situations. Par exemple, si des coefficients sont susceptibles d'être en sur-ajustement, il faut alors sélectionner un modèle plus lisse.

Deuxièmement, des interactions entre les variables du modèle peuvent être ajoutées au modèle. En effet, l'Actuaire peut choisir d'ajouter des interactions proposées par l'outil ou de créer manuellement des interactions. Les interactions sont ajustées de manière à ce que les effets sur la variable unique ne soient pas affectés par l'ajout des interactions entre les deux variables : les interactions capturent l'effet résiduel.

Troisièmement, il y a également la possibilité de rajouter une composante géographique au modèle en créant un zonier. L'Actuaire a encore le choix du nombre de pas de lissage à utiliser pour le Grid Search géographique. Le nombre de pas définit le nombre de modèles géographiques à générer, chacun présentant un différent niveau de lissage de leurs coefficients.

Enfin, la dernière possibilité d'amélioration du modèle consiste à l'enrichir avec l'ajout de variables externes. En effet, lors de la création du modèle, toutes les variables régionales et externes ont été exclues afin de décorréler l'effet régional pur déterminé uniquement par la localisation du code postal de celui reflété par les variables externes qui sont reliées à un code postal également. Après avoir ajouté la composante géographique au modèle, les variables externes peuvent être ajoutées. Ces variables externes permettent de compléter le zonier en créant un effet dans les zones où il y a peu de données.

4.2.6 Cas d'usage sur l'acte Pharmacie

Cette partie est consacrée à la description d'un cas d'usage sur la création d'un modèle pour l'acte « Pharmacie » pour l'année de survenance 2021. Ainsi, les résultats des modélisations du coût total du modèle GAM réalisé avec l'outil Akur8 pour l'acte Pharmacie seront présentés.

Modèle de coût total

Le modèle de coût total est construit avec une loi Tweedie avec le paramètre de dispersion p égal à 1,5. Plusieurs paramètres de dispersion ont été testés dans le cadre de cette étude mais la valeur 1,5 fixée par Akur8 s'est avérée être le meilleur compromis entre performance et variance du modèle.

Dans cette étude, les variables importantes telles que l'âge du bénéficiaire, le sexe, le type de bénéficiaire, la tranche d'effectif de son entreprise, son secteur d'activité, le niveau de gamme du contrat ou encore la garantie ont été conservées.

Le Grid search du modèle sur la Pharmacie est représenté ci-dessous :

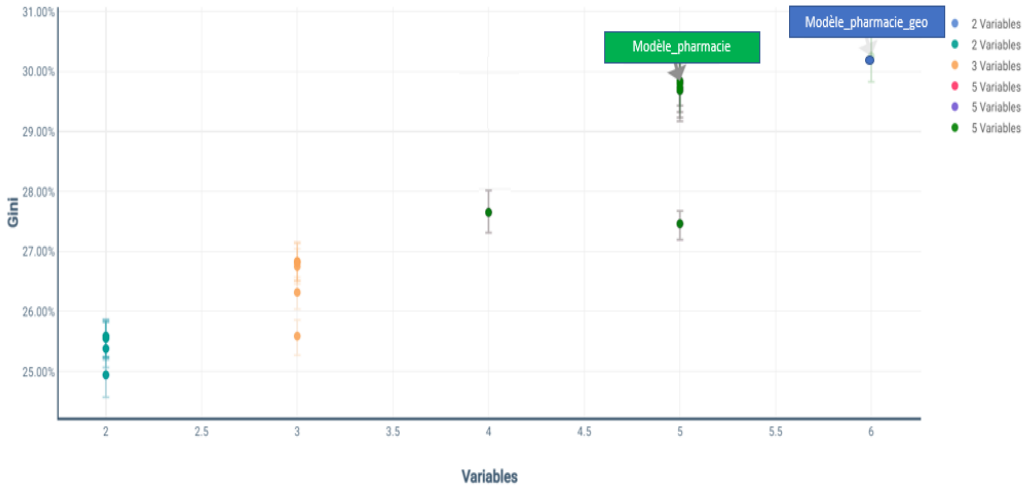


FIGURE 4.20 – Grid search de l'acte Pharmacie en 2021

Dans la figure ci-dessus, chaque couleur représente une série de modèles construits avec un nombre de variables indiqué en abscisse. Pour chaque série représentée par une même couleur, les modèles sont différents selon leurs niveaux de lissage.

Pour cet acte, le modèle avec le meilleur Gini a été choisi. Le modèle « *modele_pharmacie* » encadré en vert, construit avec 5 variables, a un bon Gini de 29,6% qui est bien supérieur au modèle à 4 variables qui a un Gini

de 27,7%.
De plus, en ajoutant la composante géographique avec la création d'un zonier, le modèle « *modele_pharmacie_geo*» encadré en bleu et légèrement meilleur avec 6 variables et un Gini d'environ 30% est obtenu.

L'ajout d'interactions et de variables externes n'a pas apporté de gain assez significatif pour les conserver dans le modèle. Afin de ne pas complexifier le modèle, il a ainsi été décidé de ne pas ajouter d'interactions et de variables externes pour ce modèle sur l'acte Pharmacie.

Grâce au modèle comportant les informations sur la géographie, l'influence du lieu de travail sur le coût total pour le poste Pharmacie peut être observé.

Voici ci-dessous la carte géographique permettant de voir les régions qui ont le plus d'influence sur le coût total pour le poste Pharmacie :

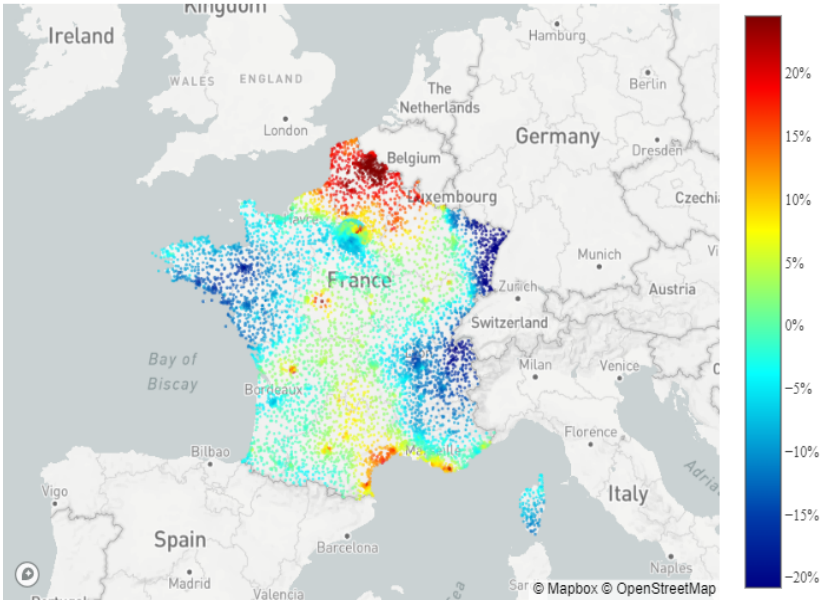


FIGURE 4.21 – Cartographie des coefficients par code postal en Pharmacie

Il est donc constaté qu'il y a des coefficients plus élevés en Pharmacie dans les régions du nord ainsi qu'aux alentours de Bordeaux. La région de l'Ile-de-France n'est pas particulièrement influente tandis que l'Alsace et les régions

du Sud-Est et du Nord-Ouest influencent négativement le coût total en Pharmacie.

Pour l'acte Pharmacie, le modèle à 5 variables sélectionné auquel s'ajoute la variable géographique relatif au code postal de l'entreprise de l'assuré « ADCODPOS » est retenue.
Les variables vont désormais être présentées selon leurs ordres d'importance :

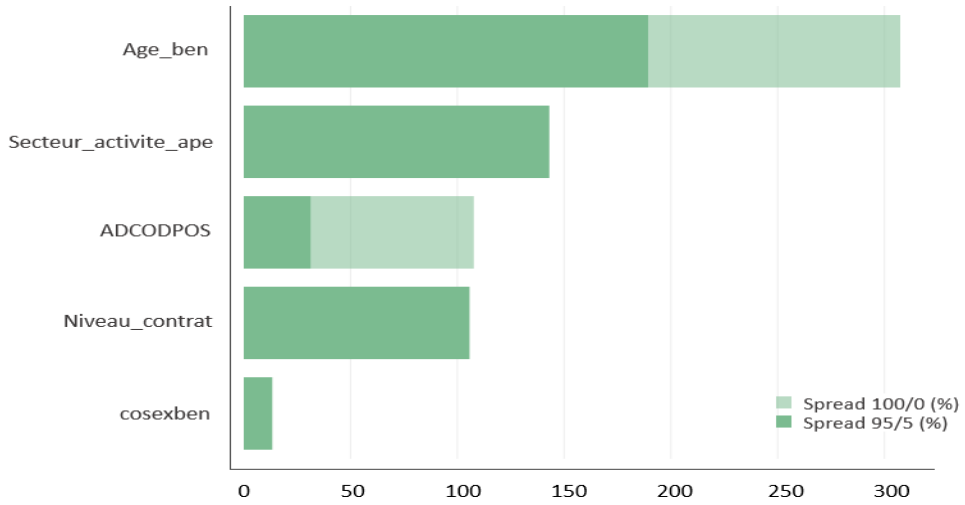


FIGURE 4.22 – Classification des variables selon leur importance pour le modèle de coût total en Pharmacie

Dans ce modèle, l'âge du bénéficiaire a un spread 95/5 de 189% et un spread 100/0 de 307%. C'est la variable la plus discriminante du modèle et il existe donc de fortes disparités entre les coefficients du modèle même en écartant les 5% plus faibles et les 5% plus forts.

Un niveau de lissage plus fort a été choisi pour ce modèle. En effet, certaines variables telles que l'âge nécessitaient d'appliquer un niveau de lissage supérieur à celui proposé initialement car les coefficients du modèle semblaient capturer du bruit.

Désormais, il va être détaillé les effets des variables sur le coût total en Pharmacie avec le tracé des coefficients accordés à chacune des modalités de la variable concernée. Voici les effets de l'âge sur le coût total en Pharmacie :

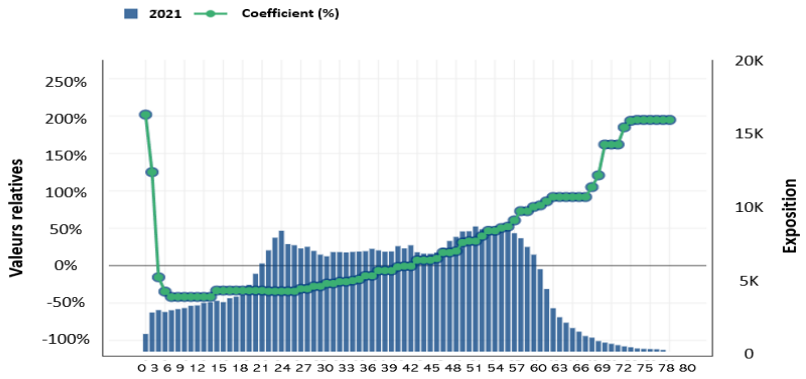


FIGURE 4.23 – Effets de l'âge du bénéficiaire sur le coût total en Pharmacie

Le graphique affiche les valeurs relatives sur l'axe vertical en fonction de l'âge sur l'axe horizontal. Les valeurs relatives représentent les valeurs brutes normalisées par la moyenne des valeurs brutes. Les bénéficiaires entre 0 et 2 ans ont un effet fort sur le coût total et cet effet est en augmentation légère de 2 à 66 ans puis en augmentation plus importante à partir de 66 ans. Ces constations sont cohérentes avec les statistiques observées.

Dans un deuxième temps, les effets du secteur d'activité de l'entreprise du salarié sur le coût total en Pharmacie vont être étudiés :

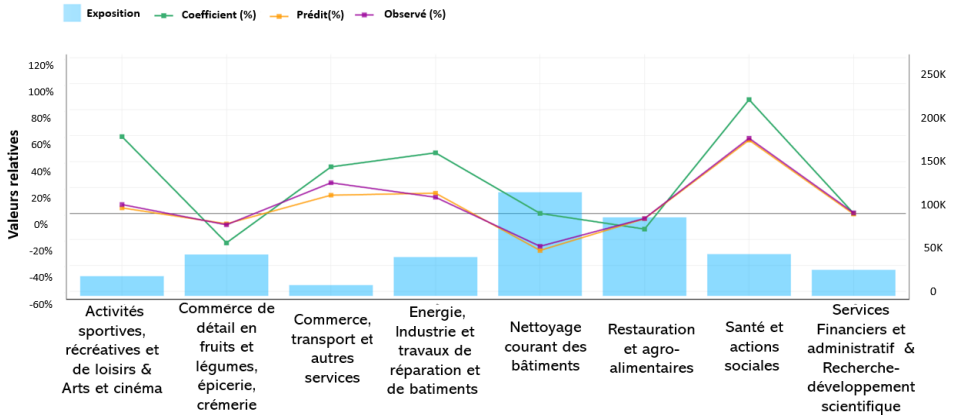


FIGURE 4.24 – Effets du secteur d'activité de l'entreprise sur le coût total en Pharmacie

Le coût total est influencé positivement par les secteurs d'activité tels que « Activités sportives, récréatives et de loisirs & Arts et cinéma », « Energie, Industrie et travaux de réparation et de batiments » et « Santé et actions sociales » notamment. Au contraire, le coût total est influencé à la baisse par les secteurs d'activité tels que « Commerce de détail en fruits et légumes, épicerie, crémerie » et « Restauration et agro-alimentaires ».

De plus, les effets du sexe du bénéficiaire sur le coût total en Pharmacie sont présentés ci-dessous :

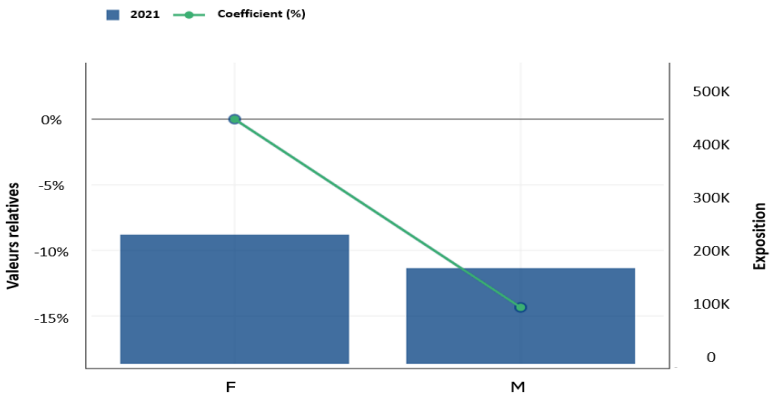


FIGURE 4.25 – Effets du sexe du bénéficiaire sur le coût total en Pharmacie

Le graphique indique que ce sont les femmes qui influencent à la hausse le coût total en Pharmacie.

Enfin, voici les performances obtenues pour ce modèle :

Gini	**MAE**	**Pseudo-R²**	**Observed target average**	**Predicted target average**
30,01%	47,5	9,84%	42,96	42,45

Les performances sont relativement bonnes avec un Gini de 30,01%, une MAE de 47,5, un pseudo-R² de 9,84% et une valeur cible moyenne observé proche de celle prédite. Néanmoins, l'erreur absolue moyenne (MAE) de prédiction est assez importante et s'explique probablement par la présence de prestations en Pharmacie extrêmes dans la base.
Ce modèle GAM peut être sélectionné et analysé plus en détail :

— **Courbe de lift :**

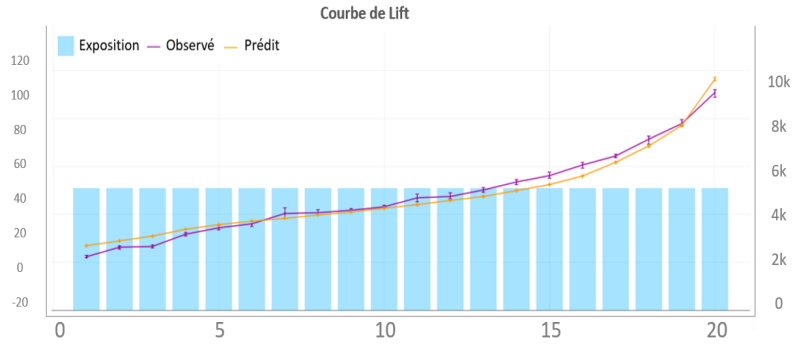


FIGURE 4.26 – Courbe Lift pour le modèle GAM de coût total en Pharmacie

La courbe Lift montre que les prédictions sont moins précises pour les montants de prestations extrêmes en Pharmacie car pour les groupes de 1 à 4 ainsi que le 20ème groupe, les observations sont légèrement au dessus des prédictions. A l'inverse, les prédictions sont légèrement en dessous des observations du 11ème au 18ème groupe. Néanmoins, le modèle prédit parfaitement les classes moins risquées de 6 à 10.

— **Courbe de Lorenz :**

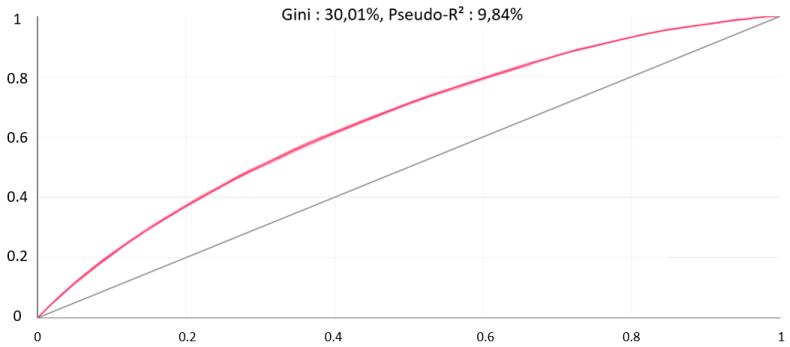


FIGURE 4.27 – Courbe de Lorenz pour le modèle GAM de coût total en Pharmacie

La courbe de Lorenz montre ici que 20% des contrats pour lesquels sont

prédit les coûts totaux les plus élevés détiennent 38% des coûts totaux observés sur le portefeuille.

— **Les résidus quantiles normalisés :**

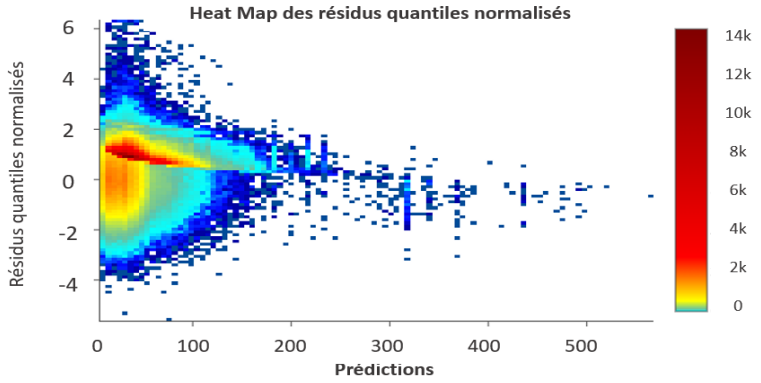


FIGURE 4.28 – Résidus quantiles normalisés pour la loi de coût total en Pharmacie

Après l'analyse des performances du modèle, l'étude des résidus quantiles normalisés randomisés permet de vérifier la validité du modèle car dans un modèle GAM valide, les résidus sont censés suivre une loi normalee centrée réduite.

Dans le cadre de ce modèle sur l'acte "Pharmacie", les résidus sont centrés autour de 0 et sont symétriques par rapport à l'axe horizontale d'ordonnée 0. L'échelle de couleurs représentant l'exposition, l'essentiel des résidus se situe dans la zone rouge/orange/jaune dans l'intervalle [-2 : 2]. Le modèle se base bien sur l'hypothèse de la loi normale centrée réduite pour ces résidus et peut donc être validé.

Chapitre 5

Analyse des résultats

Dans cette partie, les résultats obtenus avec la nouvelle méthode pour la création des normes tarifaires vont être analysés. Ils seront également comparés avec les résultats de la méthode appliquée lors des dernières normes tarifaires réalisées pour le produit « Simpleo ».

5.1 Comparaison et analyse des résultats par poste

Il a été choisi de traiter prioritairement un acte dans chacun des 6 grands postes de soins afin d'avoir une vision globale des résultats pour chaque poste. Ainsi, les résultats d'un acte par poste sur l'année de survenance 2021 vont être analysés et comparés dans cette partie.

Par confidentialité, les chiffres présentés seront fictifs mais ils ont été affectés d'un coefficient unique pour garder le même ordre de grandeur entre l'observé, les primes pures d'Akur8 et les primes pures de Simpleo.

Dans un premier temps, l'analyse des modèles a été effectuée avec les courbes de Lorenz et Lift afin d'étudier la performance.

Par la suite, une comparaison des primes pures et des coûts moyens observés a été effectuée afin de mesurer la précision de la tarification des modèles.

A noter, la cible du produit Simpleo sont les entreprises de 50 à 499 têtes et

donc les normes tarifaires associées ont été travaillées pour cette cible spécifique. Les données de cette étude concernent les entreprises de 0 à plus de 100 000 salariés avec environ 40% des assurés qui font partie de grandes entreprises (plus de 1 000 salariés) et environ 40% appartenant à des petites entreprises de 0 à 50 salariés.

Il y a donc un biais dans la comparaison des primes pures utilisées pour le produit Simpleo et les primes pures définies sous Akur8, du fait d'une cible différente. Cependant, la comparaison permettra d'observer si l'évolution du tarif en fonction de l'âge est similaire et de donner un ordre de grandeur.

Pour la deuxième partie de la comparaison, le choix d'établir un cadre de comparaison avec des hypothèses communes à chaque acte a été fait :

- Les bénéficiaires âgés entre 0 et 18 ans correspondent aux enfants ;
- A partir de 70 ans, l'assuré est considéré comme étant un retraité donc uniquement les âges de 0 à 70 ans seront étudiés.
  De plus, il n'y a pas suffisamment d'exposition pour les âges supérieurs à 70 ans étant donné que les lois Evin ont été écartés de l'étude.

5.1.1 Le poste "Pharmacie"

La Pharmacie est à la fois un poste et un acte. C'est donc l'acte Pharmacie qui va être étudié afin de comparer les deux méthodes de tarification. Cet acte a été présenté précédemment lors du cas d'usage dans la partie présentant l'outil Akur8.

Pour rappel, la courbe de Lift est construite en triant les prédictions de la plus basse à la plus haute et en les regroupant en 20 groupes qui représentent chacun 5% des prédictions.

Comme il a été vu dans le cas d'usage sur la Pharmacie, le modèle obtenu avec Akur8 est très précis pour cet acte pour chaque classe de risque. Il est toutefois légèrement en sur-prédiction pour les classes de risques faibles et en sous-prédiction pour les classes de risques 11 à 18.

Les courbes de Lorenz des deux modèles sont représentées ci-dessous :

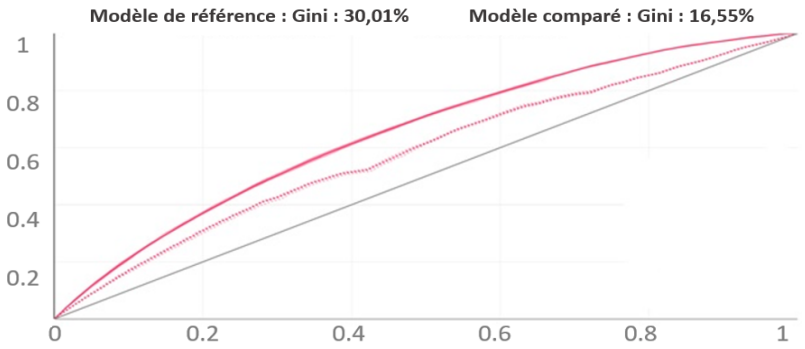


FIGURE 5.1 – Comparaison des Gini avec la courbe de Lorenz entre Akur8 et Simpleo pour l'acte "Pharmacie"

Le modèle de référence avec la courbe en trait continu est le modèle obtenu avec Akur8 tandis que le modèle comparé avec la courbe en pointillé est le modèle Simpleo. Le modèle obtenu avec Akur8 est ainsi plus performant avec environ 13 points de Gini d'écart.

Désormais, les graphiques des primes pures et des coûts moyens par bénéficiaire observés de l'acte "Pharmacie" par âge et par sexe vont être analysés :

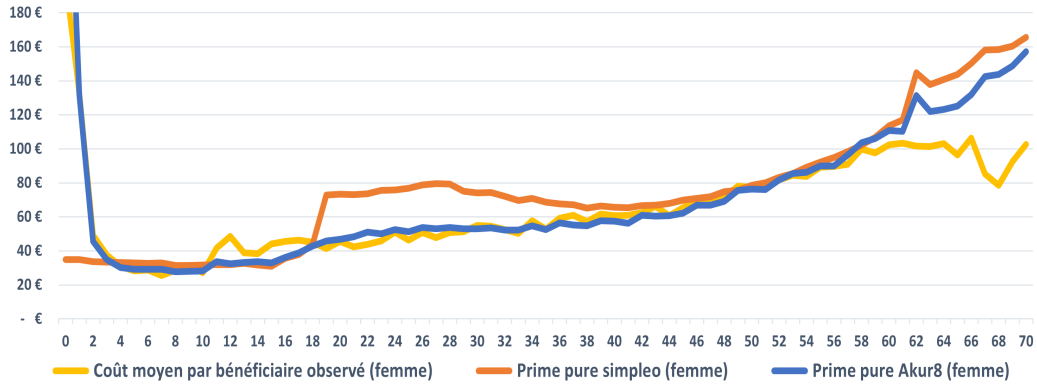


FIGURE 5.2 – Comparaison des primes pures et de l'observé pour les femmes en fonction de l'âge pour l'acte « Pharmacie »

En pharmacie, les primes pures obtenues pour les femmes avec Akur8 sont très proches des coûts moyens observés. Comparé au modèle « Simpleo », le

modèle Akur8 est plus précis, notamment pour les jeunes adultes ainsi que les nouveau-nés.

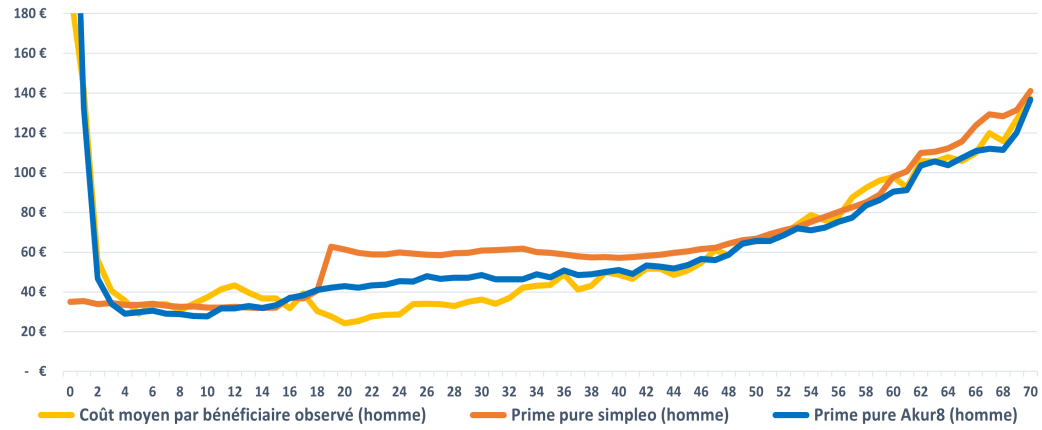


FIGURE 5.3 – Comparaison des primes pures et de l'observé pour les hommes en fonction de l'âge pour l'acte « Pharmacie »

Concernant la comparaison des primes pures pour les hommes, les mêmes constations que celles chez les femmes peuvent être faites.

Ainsi, pour l'acte Pharmacie, les primes pures obtenues avec Akur8 sont plus précises et adaptées au portefeuille que celles obtenues avec Simpleo. En revanche, elles sont un peu sur-tarifées à partir de 60 ans chez les femmes.

5.1.2 Le poste "Actes médicaux"

Pour le poste « Actes médicaux », l'acte qui va être étudié est « Consultations et visites spécialistes non-CAS ». Cet acte représente environ 44% des prestations du poste et 38% des bénéficiaires consommant en 2021.

Pour cet acte, le modèle réalisé avec Akur8 a retenu 7 variables dont l'âge du bénéficiaire, son sexe, le niveau de gamme du contrat, le secteur d'activité de l'entreprise du salarié, la tranche d'effectif de l'entreprise du salarié, le niveau de gamme du poste « soins courants » ainsi que le code postal.

La courbe de Lift du modèle est affichée ci-dessous :

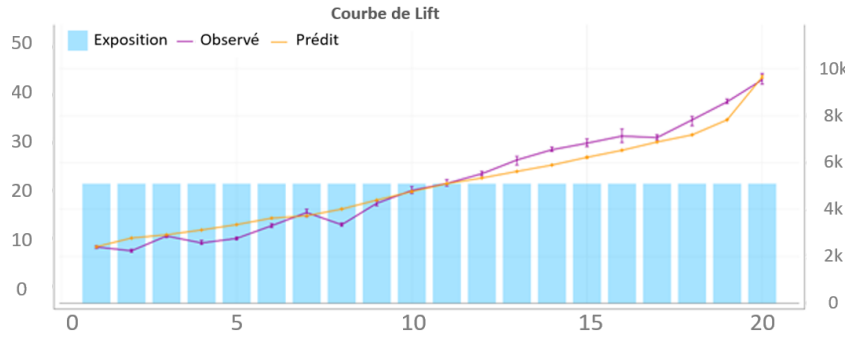


FIGURE 5.4 – Courbe Lift pour le modèle GAM obtenu avec Akur8 en consultations et visites spécialistes non-CAS

La courbe orange représente les prédictions et la courbe violette représente les observations. Le modèle obtenu avec Akur8 est assez précis pour chaque classe de risque mais il est légèrement en sur-prédiction pour les classes de risques faibles et en sous-prédiction pour les classes de risques 12 à 20. Il s'agit ici d'un modèle ayant un niveau de lissage assez fort.
Les courbes de Lorenz des deux modèles vont maintenant être étudiées :

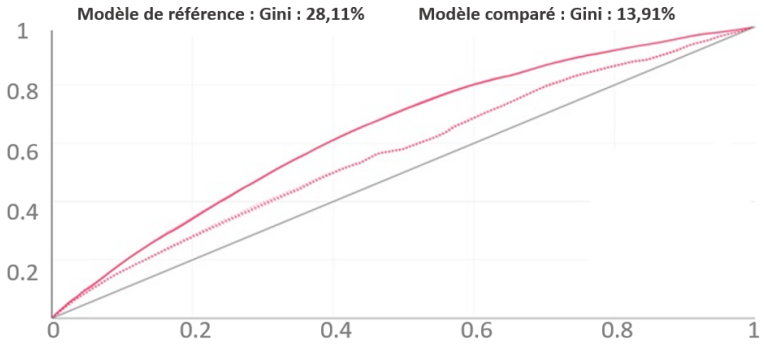


FIGURE 5.5 – Comparaison des Gini avec la courbe de Lorenz entre Akur8 et Simpleo pour l'acte "consultations et visites spécialistes non-CAS"

Le modèle obtenu avec Akur8 est ainsi plus performant en Gini avec environ 14 points d'écart.

Désormais, les graphiques des primes pures et des coûts moyens par bénéficiaire observés de l'acte « Consultations et visites spécialistes non-CAS » par âge et par sexe vont être analysés :

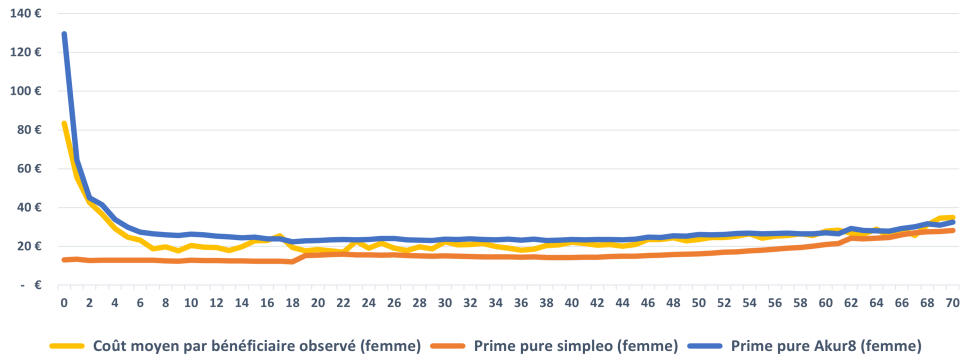


FIGURE 5.6 – Comparaison des primes pures et de l'observé pour les femmes par âge pour l'acte « Consultations et visites spécialistes non-CAS »

Pour l'acte « Consultations et visites spécialistes non-CAS », à l'exception des nouveau-nés, les primes pures obtenues avec Akur8 sont proches des coûts moyens observés, avec une légère sur-tarification notamment chez les moins de 40 ans et particulièrement chez les nouveau-nés.

Comme pour l'acte Pharmacie, le modèle Akur8 est plus précis que celui de « Simpleo ». Ce dernier a tendance à proposer des primes pures plus faibles.

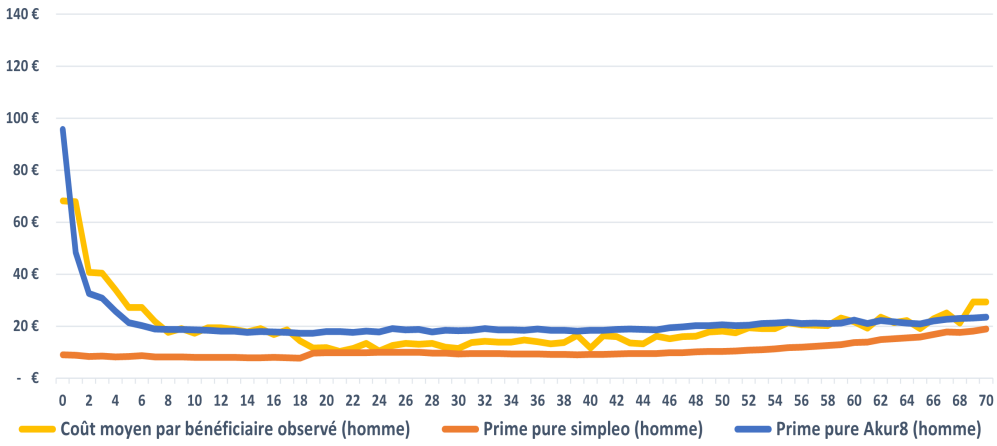


FIGURE 5.7 – Comparaison des primes pures et de l'observé pour les hommes par âge pour l'acte « Consultations et visites spécialistes non-CAS »

En général, sur le graphique de la comparaison des primes pures pour les hommes, les mêmes constations que celles chez les femmes peuvent être faites avec un peu plus d'écart entre les primes prévisionnelles Akur8 et les coûts moyens observés, ainsi qu'une sous-tarification chez les enfants entre 2 et 8 ans.

Les primes pures obtenues avec Akur8 sont, encore une fois, plus précises et adaptées au portefeuille que celles obtenues avec Simpleo pour l'acte « Consultations et visites spécialistes non-CAS ».

5.1.3 Le poste "Optique"

Pour le poste « Optique », l'acte qui va être étudié est « verres simples adulte ». Cet acte représente environ 70% des prestations du poste et 49% des bénéficiaires consommant en 2021.

Pour cet acte, le modèle réalisé avec Akur8 a retenu 6 variables dont l'âge du bénéficiaire, son sexe, le niveau de gamme du contrat, le secteur d'activité de l'entreprise du salarié, la tranche d'effectif de l'entreprise du salarié ainsi que le code postal de l'entreprise. La courbe de Lift du modèle est représentée ci-dessous :

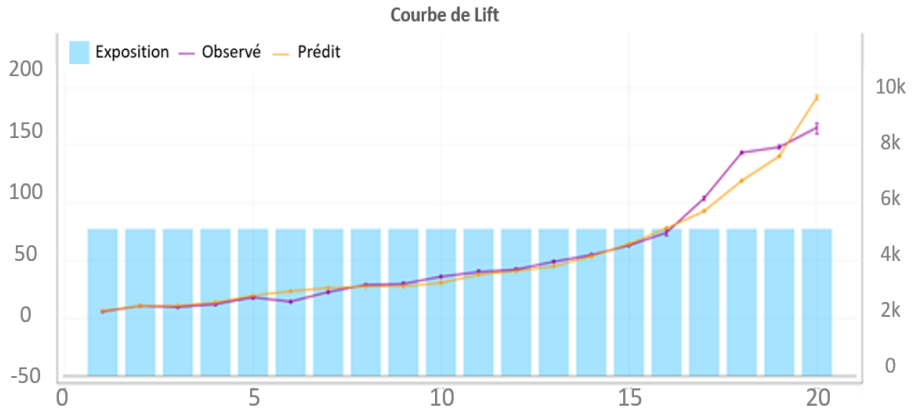


FIGURE 5.8 – Courbe Lift pour le modèle Akur8 en verres simples adulte

Le modèle obtenu avec Akur8 donne des prédictions très précises pour chaque classe de risque à l'exception des classes de risques 6 et 20 qui sont en sur-prédiction et les classes 16 et 17 qui sont en sous-prédiction.

Les courbes de Lorenz des deux modèles sont représentées ci-dessous :

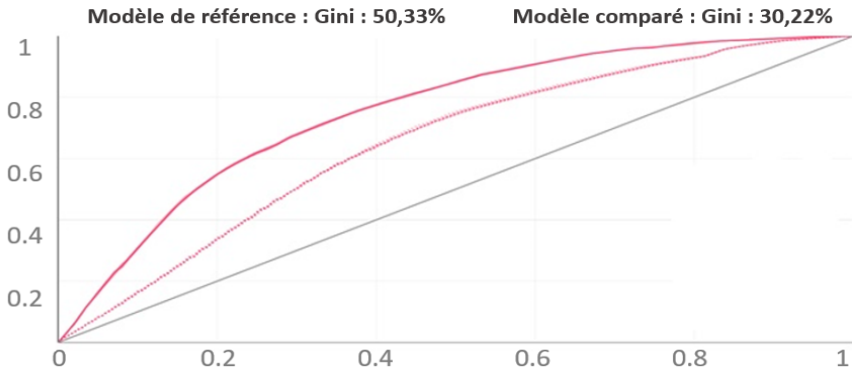


FIGURE 5.9 – Comparaison des Gini avec la courbe de Lorenz entre Akur8 et Simpleo pour l'acte "verres simples adulte"

Le modèle obtenu avec Akur8 est ainsi clairement plus performant avec environ 20 points de Gini d'écart. Le Gini obtenu pour cet acte est notablement élevé, ce qui signifie que le modèle prédit très bien les prestations observées.

Désormais, les graphiques des primes pures et des coûts moyens par bénéficiaire observés de l'acte « verres simples adulte » par âge et par sexe vont être analysés :

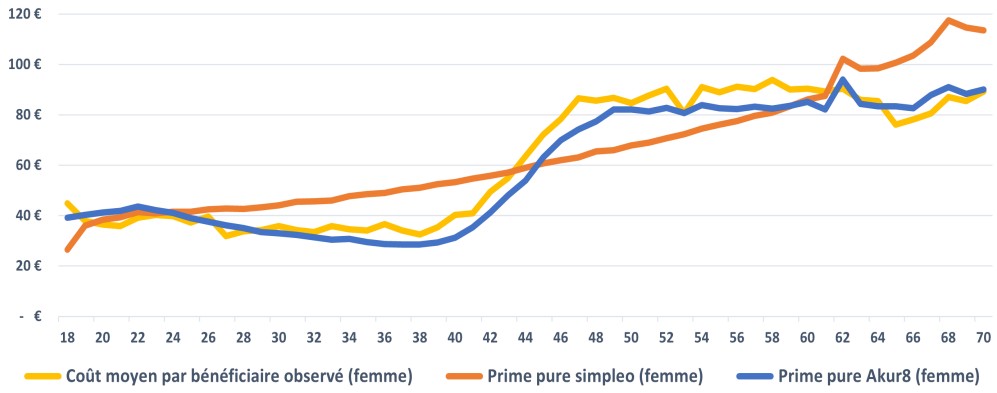


FIGURE 5.10 – Comparaison des primes pures et de l'observé pour les femmes en fonction de l'âge pour l'acte « verres simples adulte »

Pour l'acte « verres simples adulte », les primes pures obtenues avec Akur8 sont proches des coûts moyens observés, avec une légère sous-tarification pour les adultes de 32 à 60 ans. De plus, le modèle obtenu avec Akur8 sur-tarifie légèrement pour les adultes âgés de plus de 64 ans.

Comparé au modèle « Simpleo », le modèle Akur8 est plus précis. En effet, les primes pures Simpleo ne sont pas adaptées aux données en étant plus fortes que la réalité pour les jeunes adultes et ceux âgés de plus de 62 ans. En revanche, elles sont plus faibles que la réalité entre 40 et 62 ans.

Le modèle « Simpleo » ne permet donc pas d'obtenir une tarification suffisamment fine par rapport à la population visée.

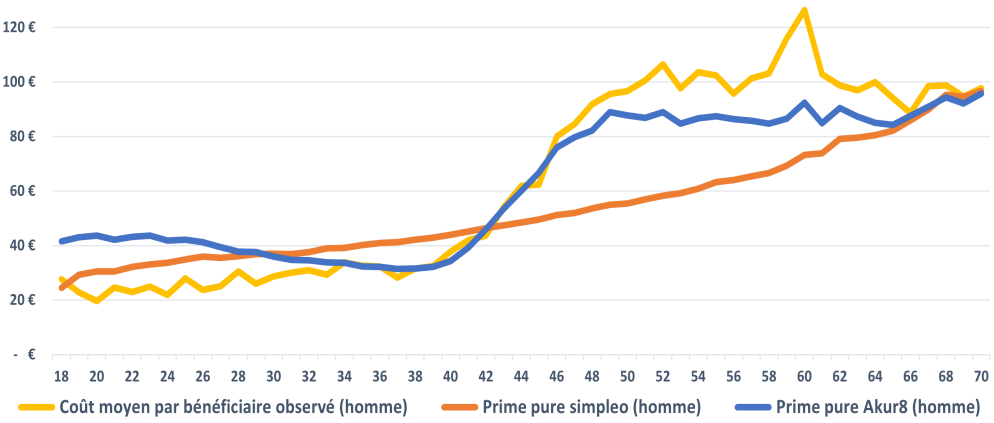


FIGURE 5.11 – Comparaison des primes pures et de l'observé pour les hommes en fonction de l'âge pour l'acte « verres simples adulte »

Les primes pures Akur8 pour les hommes présentent plus d'écart entre les primes prévisionnelles Akur8 et les coûts moyens observés que chez les femmes. Elles sont notamment en sur-tarification pour les jeunes adultes avant de sous-prédire les adultes âgés de plus de 50 ans.

Les primes pures obtenues avec Akur8 sont donc plus précises et adaptées au portefeuille que celles obtenues avec Simpleo pour l'acte « verres simples adulte ».

5.1.4 Le poste "Dentaire"

Pour le poste « Dentaire », l'acte qui va être étudié est « Prothèses dentaires remboursées ». Cet acte représente environ 53% des prestations du poste et 13% des bénéficiaires consommant. Le choix de cet acte a été fait car c'est un acte assez intéressant à étudier et également du fait de sa grande proportion de prestations au sein du poste « Dentaire ». Pour cet acte, le modèle réalisé avec Akur8 a retenu 4 variables dont l'âge du bénéficiaire, le niveau de gamme du contrat, le niveau de gamme du poste « Dentaire » ainsi que le code postal de l'entreprise. La courbe de Lift est affichée ci-dessous :

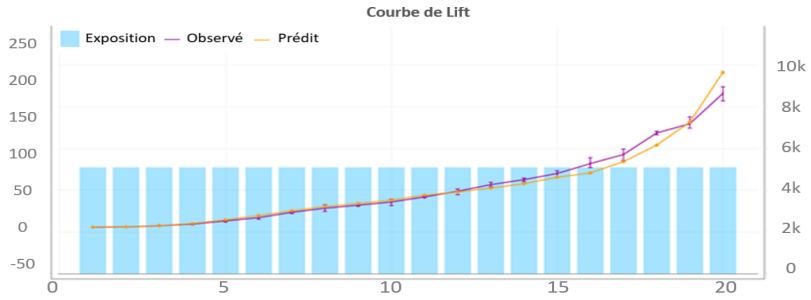


FIGURE 5.12 – Courbe Lift pour le modèle GAM obtenu avec Akur8 en prothèses dentaires remboursées

Le modèle obtenu avec Akur8 est très précis sur cet acte pour chaque classe de risque mais il est légèrement en sous-prédiction pour les classes de risques 13 à 18 et en sur-prédiction pour la classe de risque la plus élevée. De plus, les classes de risques faibles sont parfaitement prédites par le modèle. Les courbes de Lorenz des deux modèles sont représentées ci-dessous :

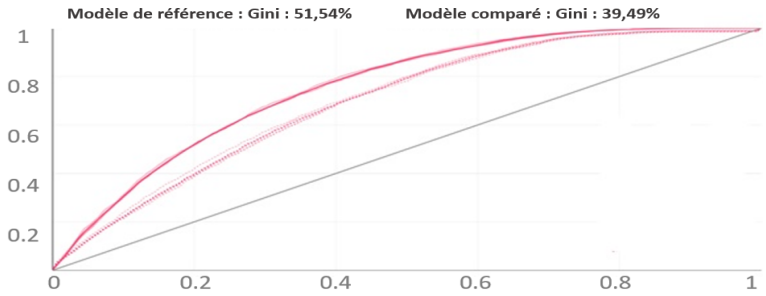


FIGURE 5.13 – Comparaison des Gini avec la courbe de Lorenz entre Akur8 et Simpleo pour l'acte "Prothèses dentaires remboursées"

Le modèle obtenu avec Akur8 est ainsi plus performant avec environ 11 points de Gini d'écart.

Désormais, les graphiques des primes pures et des coûts moyens par bénéficiaire observés de l'acte « Prothèses dentaires remboursées » par âge et par sexe vont être analysés :

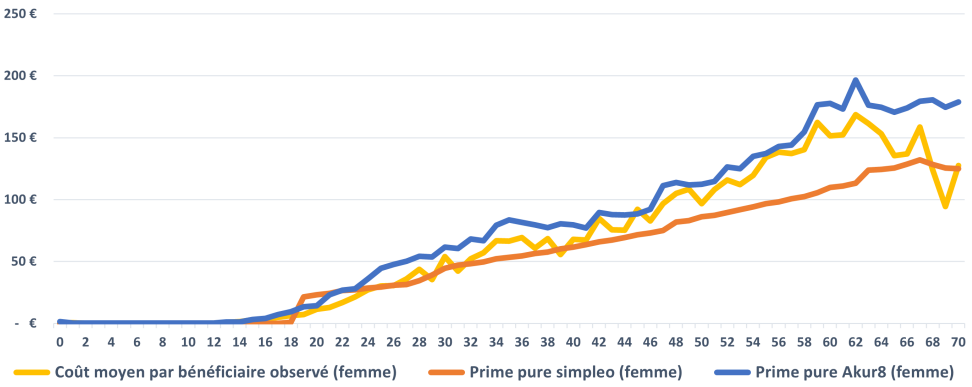


FIGURE 5.14 – Comparaison des primes pures et de l'observé pour les femmes en fonction de l'âge pour l'acte « Prothèses dentaires remboursées »

FIGURE 5.15 – Comparaison des primes pures et de l'observé pour les hommes en fonction de l'âge pour l'acte « Prothèses dentaires remboursées»

Cet acte présente une forte volatilité entre les différents âges. Il peut aussi

être observé que les enfants ne consomment quasiment pas de prothèses dentaires remboursées. Les primes pures obtenues avec Akur8 pour cet acte sont proches des coûts moyens observés mais elles sont légèrement en surprédiction pour les adultes et plus particulièrement à partir de 60 ans. Le modèle Akur8 est plus précis que celui de « Simpleo » qui a tendance à proposer des primes pures plus élevées que les coût moyens observés pour les jeunes adultes. De plus, le modèle Simpleo est plus lissé que celui d'Akur8.

Concernant les hommes, les primes pures obtenues avec Akur8 sont plus précises et variables que celles obtenues avec Simpleo. Les prédictions sont également plus proches des observations pour les hommes que pour les femmes. Pour le modèle Simpleo, les prédictions des hommes sont généralement plus faibles par rapport aux coûts moyens par bénéficiaire observés.

Ainsi, le modèle réalisé sur Akur8 est plus fin et adapté au portefeuille que celui de Simpleo en « Prothèses dentaires remboursées ».

5.1.5 Le poste "Hospitalisation"

Pour le poste « Hospitalisation », l'acte qui va être étudié est « Honoraires non-CAS ». Cet acte représente environ 10% des prestations du poste et 27% des bénéficiaires consommant.

Pour cet acte, le modèle réalisé avec Akur8 a retenu 5 variables dont l'âge du bénéficiaire, le niveau de gamme du contrat, la garantie, la tranche d'effectif de l'entreprise du salarié ainsi que le code postal de l'entreprise. La courbe de Lift du modèle est affichée ci-dessous :

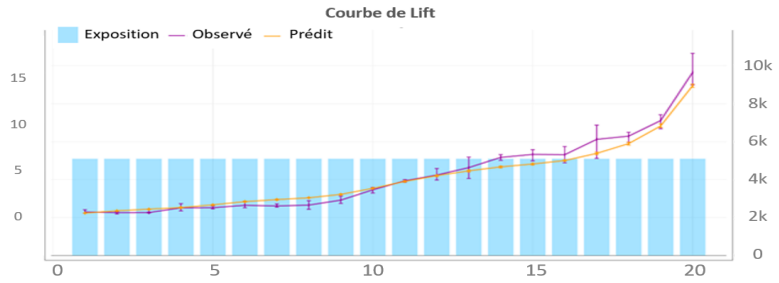


FIGURE 5.16 – Courbe Lift pour le modèle GAM obtenu avec Akur8 en Honoraires non-CAS

Le modèle obtenu avec Akur8 est assez précis pour chaque classe de risque mais il est légèrement en sur-prédiction pour les classes de risques faibles entre 6 et 9 et en sous-prédiction pour les classes de risques 14 à 18.
Les courbes de Lorenz des deux modèles sont représentées ci-dessous :

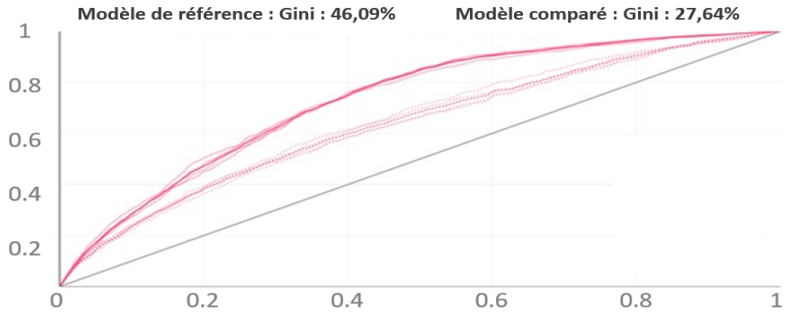


FIGURE 5.17 – Comparaison des Gini avec la courbe de Lorenz entre Akur8 et Simpleo pour l'acte « Honoraires non-CAS »

Le modèle obtenu avec Akur8 est ainsi bien plus performant avec environ 19 points de Gini d'écart. Avec un indice de Gini de 46.09, le modèle prédit très bien l'observé.
Désormais, les graphiques des primes pures et des coûts moyens par bénéficiaire observés de l'acte « Honoraires non-CAS » par âge et par sexe vont être analysés :

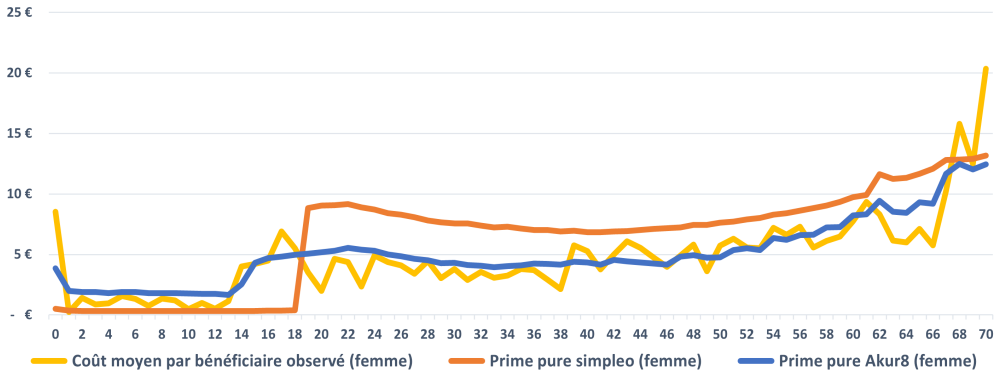


FIGURE 5.18 – Comparaison des primes pures et de l'observé pour les femmes en fonction de l'âge pour l'acte « Honoraires non-CAS »

Pour l'acte « Honoraires non-CAS », les primes pures obtenues avec Akur8 sont plus précises que celles obtenues avec Simpleo. Le modèle Simpleo a tendance a donner des primes pures légèrement supérieures à la réalité chez les femmes à partir de 18 ans mais les primes pures obtenues restent satisfaisantes.

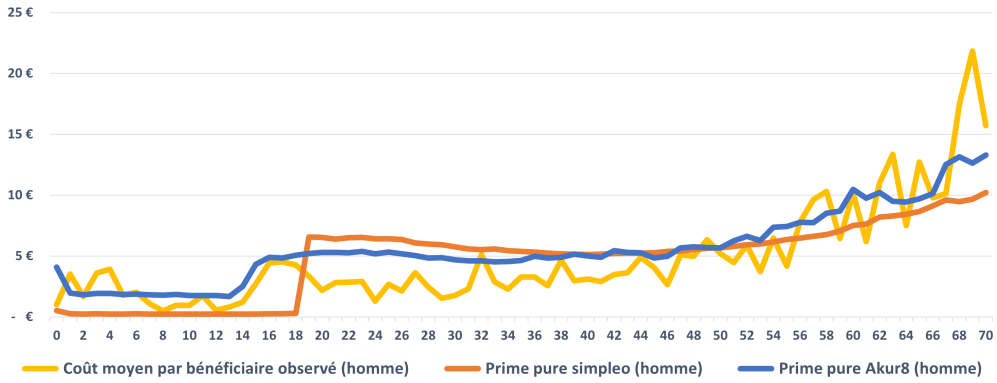


FIGURE 5.19 – Comparaison des primes pures et de l'observé pour les hommes en fonction de l'âge pour l'acte « Honoraires non-CAS »

Les primes pures obtenues pour les hommes avec Akur8 sont précises mais légèrement moins que pour les femmes. Elles sont effectivement en sous-prédiction par rapport à la réalité pour les adultes jusqu'à 58 ans.

Les primes pures du modèle « Simpleo » sont plus proches des coûts moyens par bénéficiaire observés à partir de 18 ans chez les hommes que chez les femmes. Le modèle « Simpleo » donne aussi des prédictions proches à celles obtenues avec Akur8 chez les hommes à partir de 18 ans.

Généralement, le modèle obtenu avec Akur8 est donc plus précis et performant que le modèle de Simpleo pour l'acte « Honoraires non-CAS ».

5.1.6 Le poste "Autres prestations"

Pour le poste « Autres prestations », l'acte qui va être étudié est « Analyses ». Cet acte représente environ 17% des prestations du poste et 32% des bénéficiaires consommant.

Pour cet acte, le modèle réalisé avec Akur8 a retenu 6 variables dont l'âge du bénéficiaire, son sexe, le niveau de gamme du contrat, le secteur d'activité de l'entreprise du salarié, la tranche d'effectif de l'entreprise du salarié ainsi que le code postal. La courbe de Lift du modèle est affichée ci-dessous :

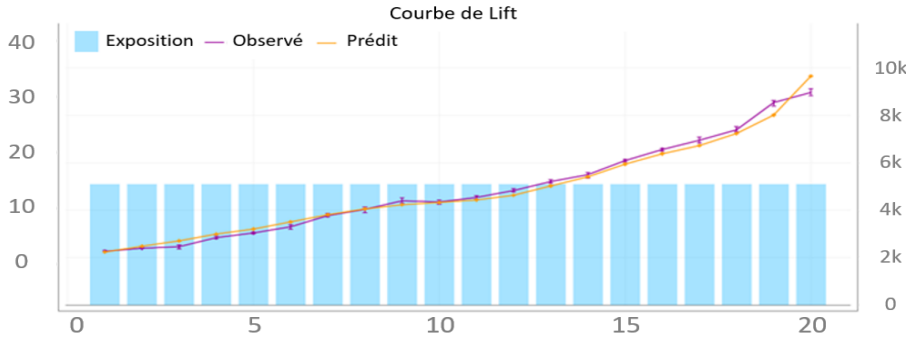


FIGURE 5.20 – Courbe Lift pour le modèle GAM obtenu avec Akur8 en Analyses

Le modèle pour l'acte « Analyses » prédit plutôt bien chaque classe de risque mises à part les classes de risque 19 et 20 qui sont respectivement en légère sous-prédiction et sur-prédiction.
Les courbes de Lorenz des deux modèles sont représentées ci-dessous :

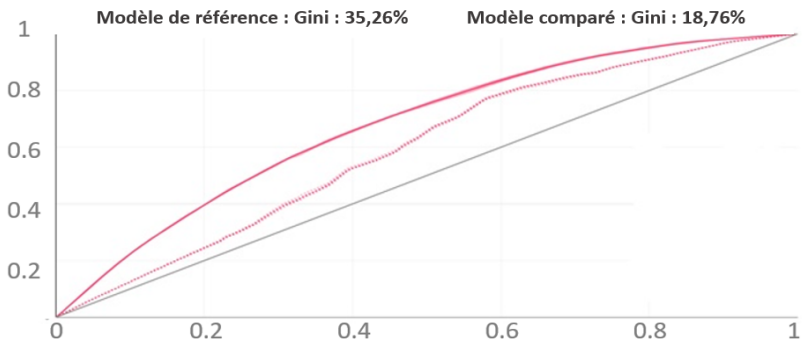


FIGURE 5.21 – Comparaison des Gini avec la courbe de Lorenz entre Akur8 et Simpleo pour l'acte « Analyses »

Le modèle obtenu avec Akur8 est ainsi presque deux fois plus performant en Gini.

Désormais, les graphiques des primes pures et des coûts moyens par bénéficiaire observés de l'acte « Analyses » par âge et par sexe vont être étudiés :

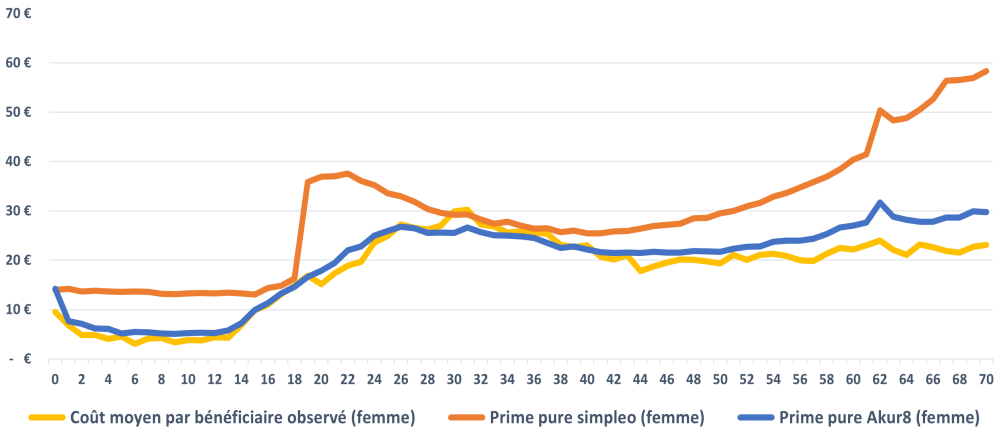


FIGURE 5.22 – Comparaison des primes pures et de l'observé pour les femmes en fonction de l'âge pour l'acte « Analyses »

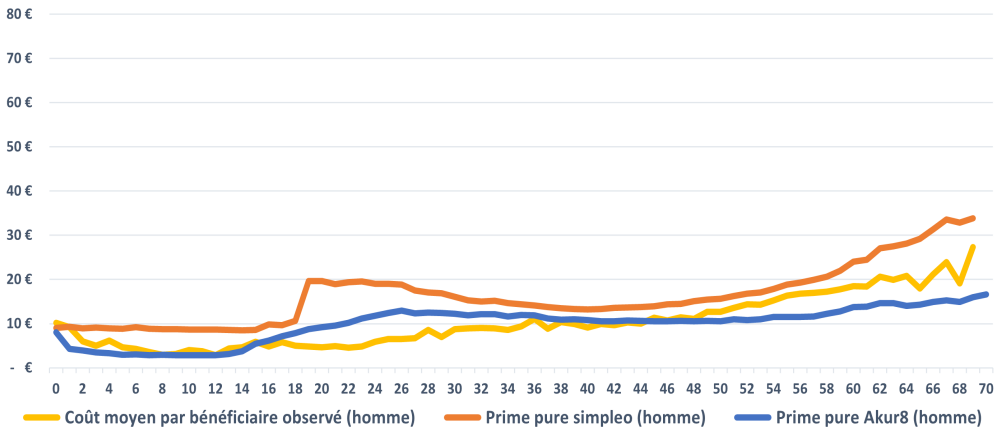


FIGURE 5.23 – Comparaison des primes pures et de l'observé pour les hommes en fonction de l'âge pour l'acte « Analyses »

Pour l'acte « Analyses », les primes pures obtenues avec Akur8 prédisent bien le coût moyen par bénéficiaire observé pour tous les âges à l'exception des

adultes âgés de plus 55 ans pour les femmes où les primes sont légèrement supérieures aux coûts moyens par bénéficiaire observés.

Le modèle Akur8 est plus précis que celui de « Simpleo ». En effet, le modèle « Simpleo » prédit convenablement pour les enfants mais ses prédictions sont supérieures aux coûts moyens par bénéficiaire observés pour les jeunes adultes ainsi que ceux âgés de plus de 40 ans. Les primes pures « Simpleo » ne sont donc pas adéquates au portefeuille pour cet acte.

Généralement, sur le graphique de la comparaison des primes pures pour les hommes, les mêmes constations que celles chez les femmes peuvent être faites avec un peu plus d'écart entre les primes pures et les coûts moyens observés. Les prédictions sont plus précises chez les femmes et les coûts moyens par bénéficiaire observés sont moins élevés chez les hommes pour les jeunes adultes.

Ainsi, les primes pures obtenues avec Akur8 sont également plus précises et adaptées au portefeuille que celles obtenues avec Simpleo pour l'acte « Analyses ».

5.2 Bilan de cette nouvelle méthode

Pour faire un bilan de cette nouvelle méthode mise en place dans le cadre de cette étude, il peut être affirmé qu'elle a permis de prédire des primes pures plus proches des observations que celles obtenues avec les dernières normes tarifaires. En effet, les modèles réalisés avec Akur8 sont plus précis par rapport à la population visée et plus performants en matière de Gini.

Akur8 permet également de modéliser assez rapidement le coût total pour chaque acte tout en contrôlant chaque étape de la modélisation. Les modèles obtenus sont également transparents en permettant de voir l'impact de chacune des variables sur le modèle. Par ailleurs, il est possible de créer un zonier assez rapidement et de l'améliorer en ajoutant des variables externes afin de pallier le manque de données sur certains code postaux.

Néanmoins, les prédictions du modèle de Simpleo, pour chaque acte étudié, peuvent être légèrement biaisées, du fait de cibles différentes. Ces comparaisons ont été réalisées afin d'avoir une idée générale sur l'amélioration ou non

des normes tarifaires en utilisant les modèles GAM avec Akur8. Finalement, elles ont permis de constater que cette nouvelle méthode apporte bien une amélioration des normes tarifaires dans plusieurs aspects.

5.3 Limites et ouverture

Cette étude a mis en avant tous les aspects positifs apportés par cette nouvelle méthode de construction des normes tarifaires mais il existe de nombreux points d'améliorations à exploiter dans le futur :

- L'ajout de nouvelles variables telles que la CSP des assurés ou encore la situation familiale avec la DSN pourrait aboutir à une segmentation plus fine et obtenir ainsi de probables meilleurs résultats. En effet, la déclaration sociale nominative (DSN) permet de récupérer les informations issues de la paie au sens large ainsi que les signalements d'évènements pour l'Assurance Maladie et pour Pôle emploi.
- Les données de cette étude pourraient également être complétées par l'utilisation de données externes avec l'Open Damir.
  L'Open Damir est une base de données externe contenant les prestations remboursées par la Sécurité sociale et cette base est gérée par la Sécurité sociale. Cette base de données peut permettre de considérer des dépenses à une échelle plus grande et de comparer son portefeuille avec la population totale francaise. Elle pourrait donc permettre d'obtenir de meilleurs résultats en améliorant par exemple le zonier.

  Néanmoins, cette base ainsi que les données de la DSN requièrent un travail d'adaptation et de retraitement des données. C'est un processus qui prend du temps et qui est actuellement en cours au sein de la direction de l'Actuariat d'AG2R La Mondiale. De plus, étant donné qu'AG2R La Mondiale est un des leader de l'assurance santé en France, la quantité des données à disposition pour cette étude a été jugé plus que suffisante et l'utilisation de bases externes n'a donc pas été nécessaire.
- Le modèle des normes tarifaires « Simpleo » a été réalisé pour une cible différente de celle de cette étude. Comme expliqué précedemment, les résultats du modèle de Simpleo pour chaque comparaison réalisée peuvent ainsi être légèrement biaisés.

- Des modèles de fréquence et de coût moyen par acte auraient pu être créés avec Akur8 afin de calculer la prime pure selon la méthode « Fréquence X Coût Moyen » et comparer les résultats avec ceux obtenus en modélisant directement le coût total. Il s'agit d'un axe d'amélioration de cette étude qui n'a pas pu être effectué par manque de temps mais qui est prévu d'être réalisé dans une étude ultérieure.
- Des modèles GAM par acte auraient pu être modélisés hors Akur8 afin de comparer les résultats avec ceux obtenus en utilisant Akur8. De même, c'est un axe d'amélioration qui n'a pas pu être effectué par manque de temps mais qui est prévu d'être réalisé dans une étude ultérieure.
- Dans cette étude, il n'a pas été retenu les bénéficiaires appartenant au régime Alsace-Moselle ou encore les lois Evin et les TNS et il serait intéressant de réaliser une étude spécifique sur ces types de bénéficiaires afin d'obtenir un jeux de normes tarifaires complet.

Conclusion

Cette étude a ainsi permis d'entamer la construction des nouvelles normes tarifaires pour des contrats collectifs en assurance santé et de challenger la méthode utilisée lors de la construction des dernières normes tarifaires en testant l'utilisation de modèles GAM avec l'outil Akur8. La méthode de tarification retenue est une tarification acte par acte et 6 actes sur 38 ont été modélisés dans le cadre de cette étude.

L'utilisation des modèles GAM avec Akur8 a permis d'obtenir un gain significatif dans la performance des modèles. De plus, Akur8 apporte un gain de temps non négligeable lors de la modélisation des actes. La création du zonier, en particulier, est également simplifiée et apporte un réel gain de performance en Gini notamment. La transparence des modèles créés permet aussi de connaître l'impact de chaque variable dans la modélisation du coût total de chaque acte.

Les résultats obtenus indiquent que l'utilisation d'Akur8 permet également d'être plus performant en matière de précision par rapport à la population visée. Cette nouvelle méthode s'avère ainsi pertinente et plus performante, en matière de Gini et de précision, que l'approche utilisée lors de la construction des dernières normes tarifaires.

Cependant, cette étude présente certaines limites qui ont été indiquées précédemment comme le manque de variables telles que la CSP ou la situation familiale ainsi que de données externes avec des bases de données telles que l'Open Damir. D'autre part, il existe un léger biais dans la comparaison avec Simpleo. Enfin, une comparaison entre les modèles « coût total» créés dans cette étude et des modèles « Fréquence X Coût Moyen » avec Akur8 n'a pas été effectuée dans cette étude. Par conséquent, il serait intéressant de

poursuivre cette étude avec les axes d'amélioration qui ont été identifiés.

Par ailleurs, tous les travaux de cette étude ont été réalisés afin de pouvoir être facilement réutilisables dans le cadre de futures études, ainsi que finaliser le jeu de normes tarifaires complet, soit en modélisant tous les actes et en les agrégeant pour obtenir la prime pure totale.

Table des figures

Biblographie

Mémoires d'actuariat

— DEBATISTA L. (2022), *Conception d'un outil de tarification en santé*

— BOUCHETAT M-K. (2021), *Tarification santé du régime Alsace-Moselle*

— BONNIFAIT C. (2019), *Optimisation d'un outil de tarification santé destiné au pilotage des grands comptes et Branches professionnelles*

— TOUTAIN FH. (2018), *Création d'un outil de tarification santé*

— ABDOLLAHI F. (2017), *Tarification d'une complémentaire santé à destination des séniors, modulaire par poste de garanties et l'impact sur la solvabilité*

— MORIN J-B. (2012), *La tarification en santé*

— GRARI V. (2009), *Impact des données exogènes sur la tarification en santé*

— LAGADEC F. (2009), *Tarification d'un contrat de complémentaire santé par un Modèle Linéaire Généralisé*

Sites internet

— www.ameli.fr

— www.securite-sociale.fr

— www.argusdelassurance.com

— www.interieur.gouv.fr

— www.legifrance.gouv.fr

— www.ressources-actuarielles.net

Documentations et formations internes AG2R La Mondiale

— Approche de la tarification santé
— Modèle collectif
— Economie de la santé
— Formation sur Akur8

Annexe

Annexe A

Extrait de la grille de garantie de la gamme "Afflec"

AG2R LA MONDIALE
Santé

Résumé de garanties
Commerce de détail de fruits et légumes, épicerie et produits laitiers (Brochure 3244)

EnVisages

Les niveaux d'indemnisation définis ci-dessous s'entendent **y compris** les prestations versées par la Sécurité sociale, dans la limite des frais réellement engagés par les bénéficiaires.

Abréviations :
BR : Base de remboursement retenue par l'assurance maladie obligatoire pour déterminer le montant de son remboursement
DPTM (Dispositif de Pratique Tarifaire Maîtrisée) : OPTAM / OPTAM-CO : OPTAM : Option Pratique Tarifaire Maîtrisée. OPTAM-CO : Option Pratique Tarifaire Maîtrisée – Chirurgie Obstétrique.
€ : Euro
FR : Frais réels engagés par le bénéficiaire
HLF : Honoraires limites de facturation fixés selon la réglementation en vigueur à la date des soins effectués par le bénéficiaire
PLV : Prix limite de vente fixés selon la réglementation en vigueur à la date des soins effectués par le bénéficiaire
PMSS : Plafond Mensuel de la Sécurité sociale
RSS : Remboursement Sécurité sociale = montant remboursé par l'assurance maladie obligatoire et calculé par l'application du taux de remboursement légal en vigueur à la base de remboursement
TM : Ticket Modérateur soit partie de la base de remboursement non prise en charge par l'assurance maladie obligatoire (TM = BR – RSS)

Hospitalisation

Nature des frais		Base	Niveaux d'indemnisation		
			Option 1	Option 2	Option 3
En cas d'hospitalisation médicale, chirurgicale et de maternité					
Frais de séjour		220 % BR	+ 50 % BR	+ 100 % BR	+ 300 % BR
Forfait journalier hospitalier		100 % des FR limité au forfait réglementaire en vigueur	–	–	–
Honoraires : actes de chirurgie (ADC), actes d'anesthésie (ADA), actes techniques médicaux (ATM), autres honoraires	Pour les médecins adhérents à un DPTM	220 % BR	+ 50 % BR	+ 100 % BR	+ 300 % BR
	Pour les médecins non adhérents à un DPTM	200 % BR	–	–	–
Chambre particulière (Dans la limite de 60 jours en hospitalisation médicale et chirurgicale, de 90 jours par année civile en maison de repos, de convalescence ou d'accueil spécialisé pour handicapés en secteur psychiatrique)		70 € par jour	+ 10 € par jour	+ 20 € par jour	+ 30 € par jour
Frais d'accompagnement d'un enfant à charge de -16 ans (sur présentation d'un justificatif)		35 € par jour	–	+ 5 € par jour	+ 25 € par jour

Extrait de la grille de garantie de la gamme "Afflec" à partir de 2020

Annexe B

Règle de décision du niveau d'acte

Règle du niveau d'acte				
	Niveau d'acte			
Acte	**1**	**2**	**3**	**4**
Actes de chirurgie et technique	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 400% BR[	$\geq 400\%BR; 100\%FR$
Actes de chirurgie et technique CAS	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 400% BR[	$\geq 400\%BR; 100\%FR$
Consultations et visites généralistes	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 400% BR[	$\geq 400\%BR; 100\%FR$
Consultations et visites généralistes CAS	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 400% BR[	$\geq 400\%BR; 100\%FR$
Consultations et visites spécialistes	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 400% BR[	$\geq 400\%BR; 100\%FR$
Consultations et visites spécialistes CAS	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 400% BR[	$\geq 400\%BR; 100\%FR$
Analyses	[0% BR : 100% BR]	]100% BR : 150% BR]	]150% BR : 250% BR]	>250% BR ; 100% FR
Actes imagerie	[100% BR : 190% BR]	]190% BR : 240% BR]	]240% BR : 300% BR[	$\geq 300\%BR; 100\%FR$
Transport	[100% BR : 150% BR]	]150% BR : 250% BR]	]250% BR : 350% BR]	>350% BR ; 100% FR
Maternité	[0€ :100€ ; 10€/jour]	]100€ : 600€ ; 10€/jour : 100€/jour[	[600€ : 1000€ ; 100€/jour : 200€/jour[	>1000€ ; >200€/jour

Actes d'imagerie médicale, radiologie, et échographie	[100% BR : 150% BR]	]150% BR : 250% BR]	]250% BR : 350% BR]	>350% BR ; 100% FR
Auxiliaires médicaux	[0% BR : 100% BR]	]100% BR : 250% BR]	]250% BR : 350% BR]	>350% BR ; 100% FR
Prothèses auditives	[0€ :1000€]	]1000€ : 2000€[	]2000€ : 3000€[	>3000 € ; 100% FR
Soins dentaires	100% BR	]100% BR : 150% BR]	]150% BR : 250% BR]	>250% BR ; 100% FR
Prothèses dentaires	[0% BR : 200% BR[	]200% BR : 299% BR]	]300% BR : 460% BR]	$\geq 460\%BR; 100\%FR$
Implants dentaires	[0€ :100€]	]100€ :1350€ ; 400 €/implant : 800 €/implant]	]1350€ :1400€ ; 800 €/implant : 1200 €/implant[	>1400€ ;≥ $1200/implant$
Orthodontie	[0%BR : 95%BR]	[100% BR : 200% BR]	]200% BR : 400% BR]	>400% BR ; 100% FR
Parodontologie	[0€ :100€]	[100€ :400€]	[400€ :450€]	>450€
Inlay et onlay	[0% BR : 200% BR[	]200% BR : 375% BR]	]375% BR : 400% BR]	$\geq 400\%BR; 100\%FR$
Prothèses dentaires remboursées par le RO	[0% BR : 125% BR]	]125% BR : 390% BR]	]390% BR : 450% BR]	>450% BR ; 100% FR
Frais de séjour	]100% BR : 200% BR]	]200% BR : 300% BR]	]300% BR : 399% BR]	$\geq 400\%BR; 100\%FR$
Forfait hospitalier	[0 :100€/jour]	[100€/jour : 200€/jour]	[200€/jour : 400€/jour]	100%FR ; >400€/jour
Chambre particulière	[0€ :100€]	[10€/jour : 90€/jour[	[90€/jour : 100€/jour]	>100€/jour ; 100% FR
honoraires	[0% BR : 100% BR]	]100% BR : 19% BR]	]190% BR : 300% BR]	>300% BR ; 100% FR
honoraires CAS	[0% BR : 175% BR]	]175% BR : 300% BR[	[300% BR : 390% BR]	>390% BR ; 100% FR

Chirurgie de la myopie	[0€ :100€]	]100€ :1500€ ; 100 €/œil : 350 €/œil ; 200 €/an ; 400 €/an]	]350€/œil : 950 €/œil ; 400 €/an : 600 €/an]	>600 €/an ; 100% FR ; >950 €/oeil
Lentilles refusées par le RO	[0€ :100€ ;105 €/an]	[100€ :300€ ; 105 €/an : 330 €/an]	[300€ :450€ ; 330 €/an : 400 €/an]	>400 €/an ;>450€
verres multifocaux ou progressifs adulte	[0€ :170€]	[170€ :200€]	]200€ :300€]	>300€
verres simples adulte	[0€ :100€[	[100€ :200€]	]200€ :400€]	>400€
verres simples enfant	[0€ :100€[	[100€ :200€]	]200€ :300€]	>300€

Annexe C

Distribution de Tweedie

En assurance santé, la modélisation des sinistres peut être effectuée en ne considérant que le coût des sinistres (Modèle « coût total ») ou bien en considérant le nombre de sinistres et leur coût moyen (Modèle « Fréquence X Coût Moyen »).
Dans le cadre de cette étude, les modèles de coût total par acte ont été réalisés avec un modèle additif généralisé de loi Tweedie dans Akur8. Dans cette partie, l'aspect théorique de la distribution de Tweedie sera ainsi présenté et détaillé.

Soit Y une variable suivant une distribution de Tweedie et y une réalisation de Y . Soit $\mu \in \mathbb{R}$, l'espérance de Y , soit p le paramètre de forme de la distribution de Tweedie tel que $1 < p < 2$ et ϕ le paramètre de dispersion. Y suit une loi de la famille des distributions exponentielles si et seulement si

sa densité peut être définie sous la forme suivante :

$$f(y,\theta,\phi) = \frac{exp(y\theta - b(\theta))}{a(\phi)} + c(y,\phi) \tag{5.1}$$

avec :

- $y \in \mathbb{R}$
- θ le paramètre de la moyenne,
- ϕ le paramètre de dispersion lié à la variance,
- a une fonction définie sur $\mathbb{R}$ et non nulle,
- b une fonction définie sur $\mathbb{R}$, au moins deux fois dérivable et avec une dérivée seconde positive,
- c une fonction définie sur $\mathbb{R}$,
- $\beta = \frac{1}{\phi(1-p)}$ et $\lambda = \frac{1}{\phi(2-p)}$,
- $\theta = \frac{\mu^{1-p}}{1-p}$,
- $a(\phi) = \phi$,
- $b(\theta) = \frac{\mu^{2-p}}{2-p}$,
- $c(y,\theta) = \ln(\sum_{i=1}^{\infty} \frac{\beta^n \alpha}{\Gamma(n\alpha)}) \times \frac{\lambda^n}{n!} \times y^{n\alpha-1}$

La distribution de Tweedie est un cas particulier des modèles de dispersion exponentielle et elle est régulièrement utilisée comme distribution pour les modèles linéaires généralisés ou les modèles additifs généralisés.

C'est une famille de distribution de probabilité comprenant des distributions continues telles que la distribution Normale et Gamma, la distribution de Poisson et la classe de distributions mixtes Poisson-Gamma. Les distributions Normale (p=0), Poisson (p=1 avec ϕ=1), Gamma (p=2) et Gaussienne inverse (p=3) sont donc des cas particuliers :

Modèles de dispersion exponentielle	**p**	$V(\mu)$	ϕ
Normale	0	1	θ
Poisson	1	μ	1
Poisson-Gamma	1<p<2	μ^p	ϕ
Gamma	2	μ^2	ϕ

Avec la classe de distribution Poisson-Gamma ($1 < p < 2$), il peut y avoir un groupe d'éléments de données à zéro et cette propriété permet de modéliser les sinistres dans le secteur de l'assurance. En effet, afin de modéliser le coût par assuré d'un contrat en assurance, il est nécessaire de tenir compte des contrats qui ne présente pas de sinistres sur la période considérée par l'étude.

Or, les distributions Normale, Gamma et Poisson ne possèdent pas de masse de probabilité en zéro. Le paramètre de la distribution Tweedie doit ainsi être une valeur fixe comprise entre]1 : 2[.

Par ailleurs, il faut que la variable dépendante soit numérique avec des données supérieures ou égales à zéro.

Annexe D

Ratio S/P des modèles Akur8

Acte	S/P
Pharmacie	96%
Verres simples adulte	93%
Prothèses dentaires remboursées	87%
Consultations et visites spécialistes non-CAS	89%
Analyses	89%
Honoraires non-CAS	84%