

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

Par : ~~Madame~~ / Monsieur – Angelo BOZIC

**Titre du mémoire : Impacts prospectifs de la réforme Résiliation Infra
Annuelle en santé individuelle dans un environnement post-100% santé**

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

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

Membres présents du jury de la
filiale :

Signature :

Entreprise : SwissLife Santé &
Prévoyance

Nom : FONTANES Ariunaa

Signature : 

Directeur de mémoire en
entreprise

Membres présents du jury de
l'Institut des Actuaires :

Signature :

Nom : PONZA Laurent

Signature : 

Invité :

Nom :

Signature :

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

Signature du responsable
entreprise :



Signature du candidat :



Table des matières

Table des matières	3
Résumé	6
Abstract	7
Remerciements	9
Note de synthèse	10
Summary	14
1 Le système de santé en France	19
1.1 La Sécurité Sociale	19
1.1.1 Histoire de la Sécurité Sociale	19
1.1.2 Les branches de la Sécurité Sociale	20
1.2 Les complémentaires santé.....	20
1.2.1 Principe.....	20
1.2.2 La Complémentaire Santé Solidaire	21
1.3 Les contrats « responsables et solidaires »	22
2 Les différentes réformes réglementaires	24
2.1 L'Accord National Interprofessionnel.....	24
2.2 La réforme Résiliation Infra Annuelle (RIA).....	24
2.2.1 Description et objectifs de la réforme.....	24
2.2.2 Inquiétudes des assureurs.....	25
2.2.3 Possibles dispositions des assureurs	25
2.3 La réforme 100% santé.....	26
2.3.1 Description et objectifs de la réforme.....	26
2.3.2 Inquiétudes des assureurs.....	28
3 Présentation du portefeuille standard santé	30
3.1 Le portefeuille santé individuelle et collective.....	30
3.1.1 Structure du portefeuille santé individuelle	30
3.1.2 Taux d'extinction du portefeuille	31
3.1.3 Répartition des prestations par génération et par poste de soins.....	32
3.1.4 Evolution de la consommation dans le temps par génération.....	34
3.1.5 Evolution de la rentabilité par génération au cours de la durée de vie du contrat	36
3.1.6 Conclusion de l'étude	37
4 Modélisation des résiliations	40
4.1 Approche à priori.....	40

4.1.1	Impact déjà observé de la réforme RIA.....	40
4.1.2	Hypothèse considérée dans la modélisation des résiliations.....	41
4.2	Modélisation des taux bruts de résiliation.....	42
4.2.1	L'estimateur de Kaplan-Meier.....	42
4.3	Lissage uni-dimensionnel des taux bruts.....	46
4.3.1	Méthode de Whittaker-Henderson en dimension 1 :.....	46
4.3.2	Taux lissés obtenus.....	49
4.3.3	Validation des taux lissés.....	50
5	Modélisation des Affaires Nouvelles.....	52
5.1	Hypothèse considérée dans la modélisation des affaires nouvelles.....	52
5.2	Séries temporelles.....	52
5.2.1	Généralités sur les séries temporelles.....	52
5.2.2	Moyenne mobile.....	53
5.2.3	Décomposition de la série temporelle.....	53
5.3	Modélisation dans le cadre du scénario Hors RIA.....	55
5.3.1	Lissage de Holt Winters.....	55
5.3.2	Les processus ARIMA et SARIMA.....	59
5.4	Modélisation dans le cadre du scénario RIA.....	64
5.4.1	La méthode SES (Simple Exponential Smoothing).....	64
5.4.2	Utilisation de l'intervalle de confiance à 95% de la méthode Hors RIA.....	65
5.4.3	Cohérence avec la projection des taux de chutes.....	66
6	Modélisation des prestations.....	68
6.1	Statistiques descriptives.....	68
6.1.1	Historique de données utilisé.....	68
6.1.2	Considération de l'exposition.....	68
6.1.3	Statistiques descriptives des variables explicatives.....	69
6.1.4	Détermination du seuil de prestations extrêmes.....	73
6.2	Critère d'évaluation des modèles.....	75
6.3	Arbres de décision.....	76
6.3.1	Principe.....	76
6.3.2	Critère de division.....	76
6.3.3	Critère d'arrêt.....	77
6.3.4	Elagage de l'arbre.....	77
6.3.5	Arbre optimal.....	78
6.4	Modèles Linéaires Généralisés.....	79
6.4.1	Principe et théorie mathématique.....	79

6.4.2	Modèle optimal	82
6.4.3	Analyse des résidus de déviance	84
6.4.4	Analyse des valeurs aberrantes.....	84
7	Projection du portefeuille	87
7.1	Approche de projection.....	87
7.1.1	Méthode utilisée pour projeter les générations 2021 et antérieures :	87
7.1.2	Méthode utilisée pour les générations postérieures à 2021	88
7.2	Résultats Obtenus	89
7.2.1	Taux d'extinction du portefeuille	89
7.2.2	Rentabilité et profitabilité.....	90
	Conclusion générale	93
	Annexe.....	95
	Table des figures	100
	Bibliographie.....	102

Résumé

Le marché de l'assurance santé est au cœur d'un contexte réglementaire en forte évolution. En effet, au cours des dernières années, de nombreuses réformes réglementaires ont bouleversé le secteur de l'assurance santé, obligeant les assureurs à s'adapter en essayant d'anticiper les éventuels impacts de celles-ci. Ainsi, lors de ce mémoire, les différents effets des récentes réformes ont été étudiés. Cependant, deux réformes réglementaires préoccupent davantage les assureurs car celles-ci sont plus récentes et donc difficilement prévisible. Il est donc nécessaire de réadapter les modèles de projection.

La première est la réforme 100% Santé, mise en place progressivement depuis 2019. Celle-ci doit permettre de lutter contre le renoncement aux soins en donnant accès aux assurés à des paniers de soins totalement remboursés. Cette réforme, avantageuse donc pour les assurés, l'est moins pour les assureurs, devant dorénavant rembourser une part plus importante de prestations.

La seconde, la réforme Résiliation Infra Annuelle (RIA), entrée en vigueur le 1^{er} décembre 2020, facilite les résiliations des contrats de complémentaires santé. En effet, avant cette réforme, les assurés pouvaient résilier leur complémentaire santé seulement à la date d'anniversaire du contrat, soit une fois par an. Depuis cette réforme, les assurés ayant au moins un an d'ancienneté peuvent maintenant résilier à tout moment leur contrat, sans pénalité. Une nouvelle fois, cette réforme est avantageuse pour les assurés mais induit une plus forte incertitude pour les assureurs dans un marché déjà très concurrentiel.

Ainsi, l'objectif de ce mémoire est d'étudier les impacts déjà observés de ces réformes afin d'ensuite prédire les conséquences à moyen terme, notamment sur la rentabilité, de ces différentes réformes. Les répercussions étant très incertaines dues à la jeunesse de ces réformes, différentes hypothèses et plusieurs scénarios ont été utilisés.

Ces deux réformes, impactant respectivement les flux (primes et prestations) et les effectifs (résiliations et affaires nouvelles), différentes modélisations ont été utilisées. Pour la modélisation des prestations, les GLM ou encore les arbres de décisions ont été utilisés. Concernant les résiliations, l'estimateur de Kaplan-Meier a permis d'obtenir les taux bruts de résiliation, ensuite lissés grâce à Whittaker-Henderson. Enfin, les affaires nouvelles ont été modélisées et prédites grâce aux séries temporelles.

Mots clés : *complémentaire santé, réforme 100% Santé, réforme RIA, GLM, arbres de décisions, Kaplan-Meier, Whittaker-Henderson, Séries temporelles*

Abstract

The health insurance market is at the heart of a rapidly changing regulatory environment. Indeed, over the last few years, numerous regulatory reforms have disrupted the health insurance sector, forcing insurers to adapt by trying to anticipate their possible impacts. Thus, in this thesis, the various effects of recent reforms have been studied. However, two regulatory reforms are of greater concern to insurers as they are more recent and therefore difficult to predict. It is therefore necessary to readapt the projection models.

The first is the 100% Health reform, which has been gradually implemented since 2019. This should make it possible to fight against the renunciation of care by giving the insured access to fully reimbursed care baskets. This reform, which is therefore advantageous for policyholders, is less so for insurers, who must now reimburse a larger share of benefits.

The second reform, the Résiliation Infra Annuelle (RIA), which came into force on 1st December 2020, makes it easier to terminate supplementary health insurance contracts. Before this reform, policyholders could only cancel their supplementary health insurance on the anniversary date of the contract, i.e. once a year. Since this reform, policyholders with at least one year of seniority can now cancel their contract at any time, without penalty. Once again, this reform is advantageous for policyholders but creates greater uncertainty for insurers in an already highly competitive market.

Thus, the objective of this thesis is to study the already observed impacts of these reforms in order to then predict the medium-term consequences, notably on profitability, of these different reforms. As the repercussions are very uncertain due to the youth of these reforms, different hypotheses and several scenarios have been used.

As these two reforms impact respectively the flows (premiums and benefits) and the numbers (terminations and new business), different models were used. For the modelling of benefits, GLMs or decision trees were used. For terminations, the Kaplan-Meier estimator was used to obtain gross termination rates, which were then smoothed using Whittaker-Henderson. Finally, new business was modelled and predicted using time series.

Keywords: complementary health, 100% Health reform, RIA reform, GLM, decision trees, Kaplan-Meier, Whittaker-Henderson, Time series

Remerciements

Je souhaite remercier en premier lieu Ariunaa FONTANES, Pierre MARJOLLET, Pierre MONTAGNE et Laurent PONZA qui ont été d'une grande aide dans la construction de ce mémoire.

Je remercie également Maud THOMAS, maître de conférences à Sorbonne Université et co-responsable du Master Actuariat de l'ISUP pour ses précieux conseils.

De plus, je souhaite remercier l'ensemble des collaborateurs Swiss Life, notamment Anais LANGLOIS, Thibault NICOLAS, Adomo ALLO, Alexandre Dorot, Samir Mameri,... pour leur accueil et leur disponibilité tout au long de mon année d'apprentissage.

Enfin, je tiens à remercier Delphine Drouet D'Aubigny, directrice de la Direction Pilotage & Solvabilité SLPS et Sébastien Lebreton, responsable du Département Inventaire & MTP, pour la confiance qu'ils m'ont accordée en m'acceptant dans leur équipe.

Note de synthèse

Dans un contexte réglementaire très instable, les complémentaires santé doivent constamment pouvoir analyser et prévoir les impacts financiers des nombreuses réformes réglementaires.

Les deux réformes impactant le plus les assureurs sont les réformes 100% santé et RIA (résiliation infra annuelle). Le but de ce mémoire est donc d'analyser les impacts déjà observés de ces réformes et de prévoir les impacts à horizon 3 ans de ces réformes, notamment la RIA.

Pour ce faire, une projection complète du portefeuille sera réalisée. Cependant, face à la jeunesse de la réforme RIA, l'incertitude de son impact sur les effectifs est très forte. En effet, à l'instar de la loi Hamon entrée en vigueur en 2015 et offrant exactement les mêmes droits que la RIA mais pour des contrats d'assurance automobile, moto, multirisque habitation..., l'incidence sur les prochaines années est incertaine. L'impact de la loi Hamon a été significatif la première année mais les effets se sont ensuite estompés voire nullifiés quelques années après.

Ainsi, dans une optique de projection du portefeuille à 3 ans, il est nécessaire de prendre en compte cet aléa. Pour cela, nous allons construire deux scénarios qui impacteront les chutes puis les affaires nouvelles :

- Scénario dit « RIA » dans lequel on suppose que l'effet de la RIA ne s'estompe pas au cours du temps
- Scénario dit « Hors RIA » dans lequel l'impact sur les chutes devient nul après 2021

Certes le deuxième scénario est un peu brutal mais il servira surtout de comparaison avec le premier scénario afin d'observer les différents effets prospectifs de la réforme.

L'historique des chutes sera donc différent selon le scénario :

- Scénario RIA : données de 2021 permettant de capter la hausse des chutes liées à la réforme
- Scénario Hors RIA : données de 2018 à 2020 pour ne pas capter la hausse des chutes

Par ailleurs, les contrats collectifs étant très peu impactés par ces évolutions réglementaires, la projection se focalisera sur les contrats individuels.

Effectivement, concernant la réforme RIA, il est plus délicat de résilier un contrat d'assurance santé collectif impactant tous les salariés, qu'un contrat individuel. Mais surtout un tel changement doit obligatoirement se faire dans la concertation et selon un certain formalisme. Il est logique que les entreprises préfèrent donc conserver leur calendrier initial et inclure les questions relatives à la mutuelle obligatoire lors des réunions déjà programmées avec les représentants du personnel. Concernant la réforme 100% santé, les garanties en collectifs étant généralement plus fortes, le recours au 100% santé est moins important qu'en individuel. L'incertitude liée à ces réformes est donc moindre pour les contrats collectifs.

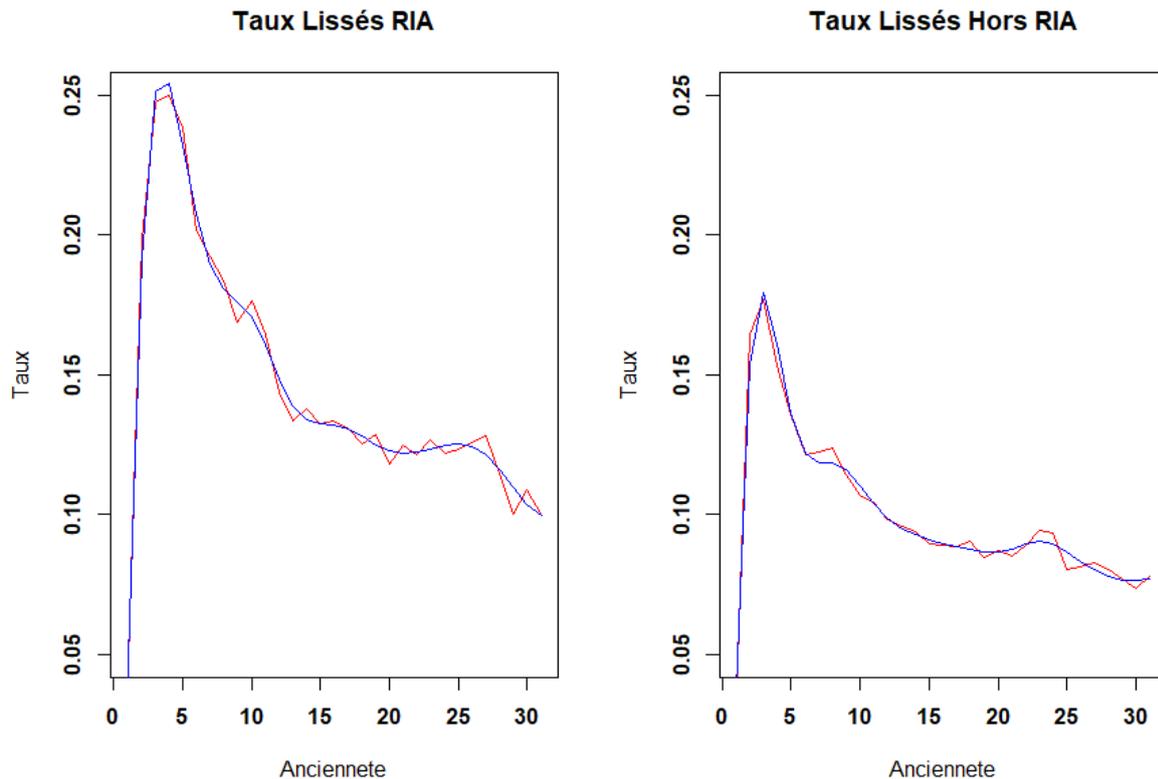
Ainsi, la projection du portefeuille santé individuelle se fera en considération des résiliations et des affaires nouvelles permettant d'obtenir les effectifs complets. Les prestations seront ensuite projetées permettant d'obtenir les flux de chaque assuré présent. Les montants de primes, considérées elles-mêmes comme plus stables, ne seront pas modélisées. Le montant de primes de chaque assuré sera utilisé

années après années tel quel, sauf pour les affaires nouvelles. En effet, les primes des nouveaux assurés seront annualisées en tenant compte de l'exposition.

Projection des résiliations

Les taux de résiliations, élément clef de la réforme RIA, ont été modélisée grâce à l'estimateur de Kaplan Meier et ensuite lissés par Whittaker-Henderson.

Les taux obtenus pour chaque scénario sont les suivants :



Les contrats ayant une ancienneté faible résiliaient initialement beaucoup plus que les anciens contrats. La réforme a ainsi accentué cet effet mais pas seulement. Le taux de résiliation est également plus élevé peu importe l'ancienneté. Ces taux seront ainsi utilisés pour la modélisation des assurés sortant.

Projection des affaires nouvelles

La prise en compte des affaires nouvelles est essentielle afin d'obtenir une assiette de prime et de sinistre complète.

L'utilisation des séries temporelles a permis d'obtenir une prédiction fiable.

Cependant, la modélisation des affaires nouvelles a été réalisée sous l'hypothèse que celles-ci soient corrélées aux résiliations. En effet, il est intuitif de penser que dans un environnement de fortes résiliations, les affaires nouvelles vont-elles aussi être impactées car les assurés ayant résilié l'auront fait afin d'obtenir de meilleures garanties ou un meilleur tarif. Ils seront susceptibles de chercher un nouvel assureur correspondant à leur besoin. Cette hypothèse selon laquelle les chutes et les affaires

nouvelles sont liées a été vérifiée de deux manières : graphiquement (tendance de la série temporelle) et statistiquement (test de corrélation).

Ainsi, pour la projection des nouveaux entrants, comme pour les chutes, on optera pour une modélisation différente avec et sans RIA :

- Scénario Hors RIA : modélisation des affaires nouvelles sans l'année 2021
- Scénario RIA : modélisation des affaires nouvelles avec l'année 2021

Dans le deuxième scénario, la prise en compte de l'année 2021 a été plus délicate du fait du manque d'observations. En effet, à la différence des résiliations, une seule année d'historique ne permet pas aux séries temporelles d'obtenir des prédictions robustes. Ainsi, plusieurs alternatives ont été proposées.

Projection des prestations

Afin de prendre en compte l'impact du 100% santé dans notre prévision, l'historique des prestations utilisé correspond uniquement à l'année 2021. En effet, la réforme fut mise en place pour la première fois en 2019, mais sa généralisation à tous les postes concernés n'est intervenue qu'en 2021. De plus, la crise sanitaire ayant entraîné des confinements successifs, l'utilisation de l'année 2020 viendrait biaiser nos prédictions.

De plus, dans le but d'obtenir des prédictions fiables, la théorie des valeurs extrêmes nous a permis de définir un seuil de prestations extrêmes et ainsi de supprimer de nos données d'apprentissage les montants extrêmes.

De ce fait, plusieurs méthodes de prédictions des prestations ont été testées. Les arbres de décision ont permis d'identifier les profils d'assurés les plus à risque de consommer. Cependant, les modèles linéaires généralisés ont obtenu les meilleurs résultats en termes de précision. Les prédictions issues des GLM Gamma seront donc utilisées pour prédire les consommations des assurés présents.

Impacts financiers

Le tableau ci-dessous représente l'évolution du profit par rapport à la dernière année d'observation et par rapport au scénario RIA par année de projection.

Le profit est défini comme la différence entre les primes encaissées et les sinistres de chaque année de projection. Afin de calculer la perte de profit, nous partons du montant de profit l'année antérieure et nous calculons un taux de croissance par rapport au montant de profit de l'année en cours pour chaque scénario.

	2022		2023		2024	
	RIA	Hors RIA	RIA	Hors RIA	RIA	Hors RIA
Ecart de S/P RIA	0,1%		0,9%		1,7%	
Perte de profit N-1	-7,8%	-5,1%	-11,8%	-6,7%	-15,2%	-7,3%
Ecart de profit RIA	-2,7%		-5,1%		-7,9%	

Interprétons 2022 de façon détaillée afin de mieux comprendre le tableau :

- 0,1% : le S/P RIA est supérieur au S/P Hors RIA de 0,1%
- -7,8% : dans le scénario RIA, la perte de profit par rapport à 2021 est de 7,8% contre 5,1% pour le scénario Hors RIA.
- Le scénario RIA fait perdre 2,7% de profit en plus par rapport au scénario Hors RIA.

Ainsi, l'impact de la réforme RIA est net en termes de rentabilité et de profitabilité. De plus, de manière générale, les affaires nouvelles ne permettent pas de garder un niveau de profit constant au cours du temps. En 2024, la perte de profit estimée est deux fois plus importante dans le scénario RIA (-7,3% contre -15,2%) et le S/P est dégradé de 1,7%.

Conclusion :

Bien que les impacts prospectifs de la réforme RIA semblent incertains, différentes hypothèses ont permis de projeter le portefeuille selon deux scénarios opposés afin d'analyser et comparer les résultats finaux.

Néanmoins, les taux de résiliations étant uni-dimensionnels, l'impact d'autres variables tel que l'âge n'a pas été pris en compte dans la modélisation. Ainsi, une modélisation multi-dimensionnelle permettrait d'obtenir des taux plus précis. Pour ce faire, les graphiques des probabilités de survies segmentés par variables présent en annexe permettent de se faire une idée des variables impactant significativement les résiliations. Ceci pourra alors être utilisé dans une possible modélisation des taux de chutes plus optimale. Les travaux réalisés dans la projection des prestations pourront aussi être utilisés dans le but d'une nouvelle approche de tarification du risque suite à la hausse des résiliations.

Summary

In a very unstable regulatory context, complementary health insurance companies must constantly be able to analyse and forecast the financial impact of the numerous regulatory reforms.

The two reforms that have the greatest impact on insurers are the 100% health and RIA reforms. The aim of this thesis is therefore to analyse the already observed impacts of these reforms and to forecast the impacts of these reforms, in particular the RIA, over a 3-year horizon.

In order to do this, a complete projection of the portfolio will be carried out. However, given the youth of the RIA reform, the uncertainty of its impact on the workforce is very high. Indeed, like the Hamon law which came into force in 2015 and offers exactly the same rights as the RIA but for car, motorbike, multi-risk home insurance contracts, etc., the impact on the coming years is uncertain. The impact of the Hamon law was significant in the first year but the effects faded or even disappeared a few years later.

Thus, from the point of view of a 3-year portfolio projection, it is necessary to take this hazard into account. To do this, we will construct 2 scenarios that will impact on the falls and then on new business:

- RIA scenario in which it is assumed that the effect of the RIA does not fade over time
- Non-RIA scenario in which the impact on falls becomes zero after 2021

Admittedly, the second scenario is a little brutal, but it will serve mainly as a comparison with the first scenario in order to observe the various prospective effects of the reform.

The history of falls will therefore differ according to the scenario:

- RIA scenario: data from 2021 to capture the increase in falls due to the reform
- Non-RIA scenario: data from 2018 to 2020 to not capture the increase in falls

Moreover, as group contracts are hardly affected by these regulatory changes, the projection will focus on individual contracts.

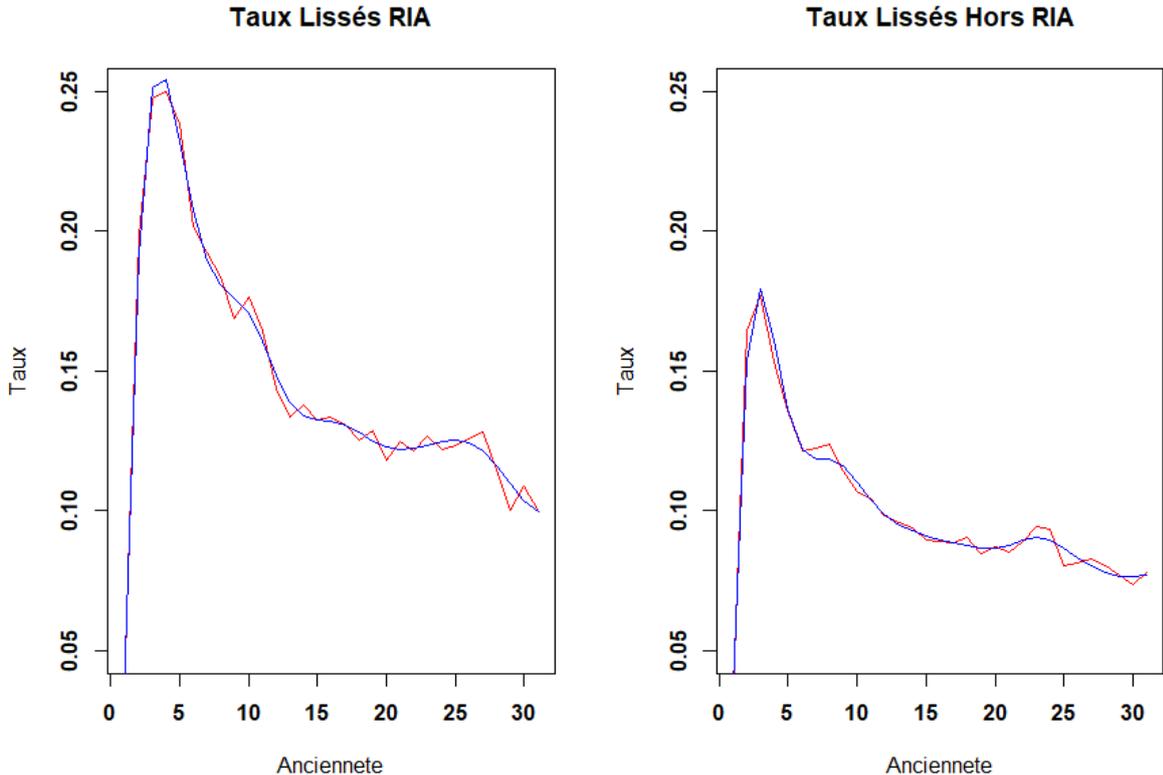
Indeed, with regard to the RIA reform, it is more difficult to terminate a group health insurance contract affecting all employees than an individual contract. But above all, such a change must be made in consultation and according to a certain formalism. It is logical that companies prefer to keep their original timetable and to include questions relating to the compulsory mutual insurance scheme in the meetings already scheduled with the staff representatives. With regard to the 100% health reform, as group coverage is generally stronger, the use of 100% health is less important than for individuals. The uncertainty linked to these reforms is therefore less for group contracts.

Thus, the projection of the individual health portfolio will take into account terminations and new business in order to obtain the full number of employees. The benefits will then be projected to obtain the flows of each insured present. Premiums, considered more stable, will not be modelled. The amount of premiums for each policyholder will be used year after year as is, except for new business. Indeed, the premiums of new insureds will be annualised by taking into account the exposure.

Projection of terminations

Termination rates, a key element of the RIA reform, were modelled using the Kaplan Meier estimator and then smoothed by Whittaker-Henderson.

The rates obtained for each scenario are as follows:



Contracts with low seniority initially terminated much more than older contracts. The reform has thus accentuated this effect, but not only. The cancellation rate is also higher regardless of seniority. These rates will thus be used for the modelling of outgoing policyholders.

New business projection

The inclusion of new business is essential to obtain a complete premium and loss base. The use of time series has provided a reliable prediction.

However, the modelling of new business was done under the assumption that it is correlated with terminations. Indeed, it is intuitive to think that in an environment of high cancellations, new business will also be impacted because policyholders who have cancelled will have done so in order to obtain better cover or a better rate. They will be likely to look for a new insurer to meet their needs. This hypothesis that churn and new business are linked has been verified in two ways: graphically (time series trend) and statistically (correlation test). Thus, for the projection of new entrants, as for the falls, we will opt for a different modelling with and without RIA:

- Non-RIA scenario: modelling new business without the year 2021
- RIA scenario: modelling new business with the year 2021

In the second scenario, the inclusion of the year 2021 was more difficult because, unlike the modelling of terminations, a single year of history does not allow the time series to make robust predictions. Thus, several alternatives were proposed.

Claims projection

In order to take into account the impact of 100% health in our forecast, the history of benefits used corresponds only to the year 2021. Indeed, the reform was first implemented in 2019, but its generalisation to all the posts concerned did not take place until 2021. Moreover, the health crisis having led to successive confinements, the use of the year 2020 would bias our predictions. Furthermore, in order to obtain reliable predictions, the extreme value theory allowed us to define a threshold of extreme benefits and thus to remove the extreme amounts from our training data.

As a result, several methods of benefit prediction were tested. Decision trees were used to identify the profiles of policyholders most likely to consume. However, the generalized linear models obtained the best results in terms of accuracy. The predictions from the Gamma GLMs will therefore be used to predict the consumption of the insureds present.

Financial impacts

The table below shows the evolution of profit compared to the last year of observation and to the RIA scenario by projection year.

Profit is defined as the difference between premiums collected and claims in each projection year.

	2022		2023		2024	
	RIA	Non-RIA	RIA	Non-RIA	RIA	Non-RIA
Loss ratio gap RIA	0,1%		0,9%		1,7%	
Loss of profit N-1	-7,8%	-5,1%	-11,8%	-6,7%	-15,2%	-7,3%
Profit gap RIA	-2,7%		-5,1%		-7,9%	

Let's interpret 2022 in detail to better understand the table:

- 0.1%: the S/P RIA is higher than the S/P Non-RIA by 0.1%.
- -7.8%: in the RIA scenario, the loss of profit compared to 2021 is 7.8% against 5.1% for the RIA scenario.

- The RIA scenario results in a loss of 2.7% more profit than the Non-RIA scenario.

Thus, the impact of the RIA reform is clear in terms of profitability. Moreover, in general, new business does not allow a constant level of profit to be maintained over time. In 2024, the estimated loss of profit is twice as important in the RIA scenario (-7.3% against -15.2%) and the S/P is degraded by 1.7%.

Conclusion

Although the prospective impacts of the RIA reform seem uncertain, different assumptions have allowed the portfolio to be projected under two opposite scenarios in order to analyse and compare the final results.

However, as termination rates are uni-dimensional, the impact of other variables such as age was not taken into account in the modelling. Thus, multi-dimensional modelling would allow for more accurate rates. To this end, the graphs of survival probabilities segmented by variables in the appendix give an idea of the variables that have a significant impact on terminations. This can then be used in a possible more optimal modelling of drop-out rates. The work done in the projection of benefits could also be used for a new approach to pricing risk following the increase in terminations.

1 Le système de santé en France

1.1 La Sécurité Sociale

1.1.1 Histoire de la Sécurité Sociale

Dès le début du XIXe siècle, l'influence de l'industrialisation et le développement du prolétariat révèlent le développement de dangers sociaux : accidents, chômage, licenciements. Face à l'insécurité au travail, à la pauvreté et au pouvoir patronal, le prolétariat crée les premières sociétés de secours mutuels et les mutuelles ouvrières s'organisant par grandes branches d'activité : mines, chemins de fer. Ces organismes prélèvent alors des cotisations sur les travailleurs et leur versent ainsi des indemnités en cas de besoin. Cependant, le droit d'association étant très restreint à cette époque, elles faisaient l'objet d'un contrôle très strict du pouvoir politique. Ce n'est qu'à partir de 1889 qu'elles pourront se constituer sans autorisation administrative.

Ainsi, au milieu du XIXe siècle, période de grande industrialisation, entraînant une main d'œuvre rurale et instable, le patronat s'est donc montré favorable à la création de régimes de retraites et de caisses de secours pour les salariés des industriels des mines et des chemins de fer. Par la suite, des hôpitaux, des crèches et même des écoles professionnelles ont vu le jour.

C'est alors qu'après plusieurs créations de régimes s'étant soldé par des échecs avant la première guerre mondiale, les lois de 1928 et 1930 sont votées, instituant les assurances sociales. La Loi du 5 avril 1928 crée une Assurance Vieillesse et une Assurance Maladie pour tous les salariés, basée sur le principe de la capitalisation. Cette loi fut complétée par la Loi du 30 avril 1930. Depuis, l'ensemble des salariés sont désormais couverts contre les risques maladie, invalidité et vieillesse.

Après la 2ème Guerre Mondiale, le préambule de la Constitution de la IVème République atteste pour la première fois le droit à la Sécurité Sociale, au côté des Droits de l'Homme et du Citoyen. A cette date, l'Ordonnance du 4 octobre 1945, acte la création du Régime Général.

Les trois principes fondamentaux de cette Ordonnance sont :

- la généralisation progressive de la Sécurité Sociale pour toute la population, financée par une double cotisation payée par les employeurs et les salariés
- l'union des Institutions et l'universalité des risques : un régime unique doit être mis en place et des Caisses à compétence générale sont responsables d'appliquer les législations sociales
- la démocratie sociale : les Organismes de Sécurité Sociale étant des Organismes de droit privé dotés d'une autonomie de gestion, ils sont gérés par des Conseils d'Administration composés de représentants de salariés et d'employeurs

L'affiliation à la Sécurité Sociale est aujourd'hui obligatoire pour toutes les personnes résidentes en France. Le régime de rattachement dépend lui de la situation professionnelle de l'assuré. En France, le régime général et le régime agricole sont les deux principaux régimes de santé obligatoires. Le régime général concerne les salariés et les indépendants. Le régime agricole concerne lui les exploitants et salariés agricoles et les employés des industries agroalimentaires.

1.1.2 Les branches de la Sécurité Sociale

L'activité de la Sécurité Sociale est répartie en cinq branches :

- La maladie incluant la maladie, la maternité, l'invalidité, le décès
- Les accidents du travail et maladies professionnelles
- La famille incluant le handicap, le logement et le RSA
- La retraite incluant la vieillesse et le veuvage
- Le recouvrement

La branche maladie de la Sécurité Sociale garantit l'accès aux soins et assure le remboursement des dépenses de santé des assurés malades. Dans la majorité des régimes, la branche maladie gère également lorsqu'ils existent les risques maternité, invalidité et décès.

1.2 Les complémentaires santé

1.2.1 Principe

La complémentaire santé est une assurance dont le but est de venir en complément des remboursements de l'assurance maladie car celle-ci ne prend en charge qu'une partie des frais médicaux. Le reste à charge de l'assuré se verra ainsi diminuer ou même annuler dans le meilleur des cas. Cependant, le remboursement de l'assurance maladie puis de la complémentaire santé ne peut dépasser le montant initial dépensé par l'assuré. C'est le principe de non-enrichissement par l'assurance.

Trois types d'organisme proposent des contrats d'assurance complémentaires en France. Cependant, chacun d'eux est régi par une réglementation différente :

- Les mutuelles, régies par le code de la Mutualité
- Les Institutions de Prévoyance, régies par le code de la Sécurité Sociale
- Les Sociétés d'Assurances, régies par le code des Assurances.

Globalement les mutuelles restent majoritaires en termes de cotisations reçues. En effet, en 2020, les mutuelles collectent 49% des cotisations, les sociétés d'assurances 34% et les instituts de prévoyances 17%.

L'assurance complémentaire santé peut être souscrite de façon collective (en adhérant au contrat souscrit et proposé par un employeur ou par une association, professionnelle ou non), ou individuelle. Selon le type de contrat, la souscription est différente. En effet, pour un contrat d'assurance complémentaire individuelle, l'adhésion est facultative et concerne les particuliers. Cependant, pour les contrats collectifs, l'adhésion est, dans la plupart des cas, faite par l'employeur. L'adhésion de ces contrats collectifs s'effectue soit par un accord d'entreprise, par un référendum ou enfin par une décision unilatérale de l'employeur. Ainsi, depuis la réforme de l'ANI en 2016, décrite plus tard, l'employeur a l'obligation de mettre en place une complémentaire santé pour ses salariés.

La rentabilité des contrats individuels et collectifs diffère. Effectivement,

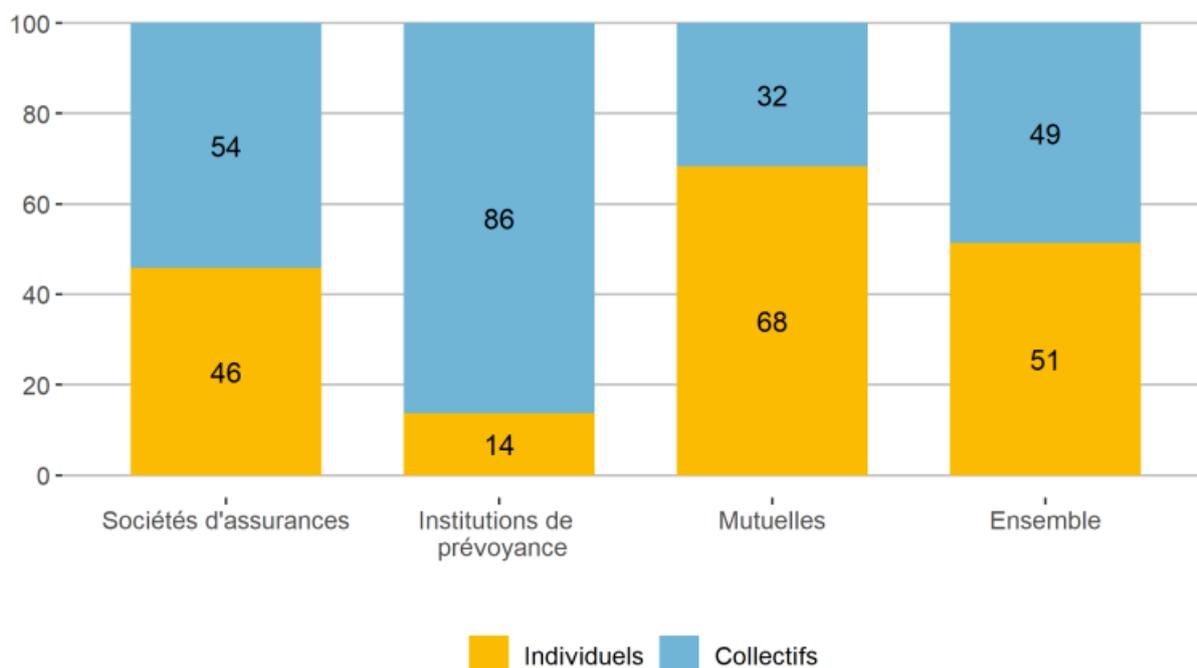


Figure 1 Source : ACPR, calculs DREES. Part des contrats collectifs et individuels dans l'ensemble des cotisations collectées en santé par les différents types d'organismes en 2020, en % de cotisations collectées.

1.2.2 La Complémentaire Santé Solidaire

La CMU-C, pour Couverture Maladie Universelle Complémentaire, a été créée en 1999 afin de réduire les disparités de couverture et donc de remboursement. En effet, le niveau de couverture dépend souvent du montant de ces cotisations. Néanmoins, compte tenu des différences de pouvoir d'achat entre les ménages, certaines personnes n'étaient pas en mesure de s'offrir une assurance complémentaire. La CMU-C prend alors en charge la totalité du ticket modérateur d'un acte de soins, annulant donc le reste à charge de l'assuré. Pour bénéficier de la CMU-C, plusieurs conditions doivent être vérifiées :

- Résider en France de manière régulière
- Résider sur le territoire français depuis plus de 3 mois
- Avoir des ressources inférieures à un certain plafond

L'ACS (Aide au paiement d'une Complémentaire Santé), donne droit à une aide financière pour payer un contrat de complémentaire santé pour les personnes dont les ressources sont légèrement supérieures au plafond d'attribution de la CMU-C. Le patient bénéficiant de l'ACS se voit, au même titre que la CMU-C, exonéré de reste à charge.

Ces 2 dispositifs d'aide à la prise en charge des frais de santé non remboursés par la Sécurité sociale, ont fusionné en un seul depuis le 1^{er} novembre 2019, nommé « **Complémentaire santé solidaire** » (CSS).

1.3 Les contrats « responsables et solidaires »

Les contrats « responsables et solidaires » ont pour objectif de responsabiliser l'assuré en encadrant ses dépenses de santé dans le but de limiter les dérives de sinistralité. Pour être considéré comme « responsable et solidaire », un contrat doit respecter un cahier des charges strict.

Ainsi, pour être considéré comme « responsable », si le parcours de soin est respecté, le contrat doit prendre en charge :

- Le ticket modérateur
- 30% minimum de la base de la Sécurité Sociale pour les médicaments à vignettes blanches
- 35% minimum de la base de la Sécurité Sociale pour les actes d'analyse et de laboratoires
- L'intégralité des dépassements d'honoraires des professionnels de chirurgie, d'obstétrique et d'anesthésie-réanimation, et également des soins dentaires et optiques
- La couverture de 2 prestations de prévention jugées prioritaires par la Haute Autorité de Santé : vaccination, détartrage annuel, ...

De plus, pour être considéré comme « solidaire », le contrat ne doit pas avoir de questionnaire médical à la souscription (pas de sélection médicale) et le tarif des cotisations ne doit pas dépendre de l'état de santé de l'assuré.

2 Les différentes réformes réglementaires

2.1 L'Accord National Interprofessionnel

Issu d'une négociation entre les partenaires sociaux (représentants des entreprises et des salariés), l'Accord National Interprofessionnel (ANI) du 11 janvier 2013 a été transcrit en droit par le pouvoir législatif.

Cet accord oblige tout employeur à mettre en place à titre obligatoire une complémentaire santé pour ses salariés.

Ainsi, depuis le 1er janvier 2016, toutes les entreprises ont eu l'obligation de mettre en place une complémentaire santé pour l'ensemble de leurs salariés. L'obligation concerne uniquement le salarié et non son conjoint ou ses ayants-droits. Le financement est pris en charge par l'employeur à hauteur de minimum 50 % de la cotisation.

En contrepartie, l'employeur profite d'avantages fiscaux et sociaux :

- Les cotisations versées par l'employeur pour la complémentaire santé sont déductibles du bénéfice imposable de l'entreprise
- Les cotisations versées par l'employeur sont exonérées de charges sociales, dans la limite de 6 % du Plafond annuel de la Sécurité sociale + 1,5 % du salaire annuel brut et dans la limite de 12 % du Plafond annuel de la Sécurité sociale

De ce fait, tous les salariés de l'entreprise ont l'obligation d'adhérer au contrat d'assurance complémentaire santé collectif responsable et solidaire proposé par leur employeur, mais certains cas de dispense existent. L'impact de cette réforme sur notre portefeuille sera illustré dans la partie 3.

2.2 La réforme Résiliation Infra Annuelle (RIA)

2.2.1 Description et objectifs de la réforme

A compter du 1er décembre 2020, les assurés pourront résilier leur complémentaire santé à tout moment, après une année de détention.

En effet, avant cette date, afin de résilier un contrat d'assurance santé, l'assuré ou l'entreprise souscriptrice devait adresser une lettre recommandée à l'assureur au moins 2 mois avant la date d'échéance. Passé ce délai, l'adhérent était réengagé automatiquement pour une durée d'un an, ne pouvant donc plus rompre son contrat avant la nouvelle date d'anniversaire.

De plus, la part de cotisation relative à la période non couverte (après la date d'effet de la résiliation) est remboursée à l'adhérent dans les 30 jours. Dans le cas contraire, des intérêts sont dus à l'adhérent. Cette réforme, dont seuls les contrats santé étant concernés par la réforme, a deux principaux objectifs :

- Améliorer le pouvoir d'achat des assurés en leur permettant de choisir à tout moment un contrat de complémentaire santé adaptée à leurs besoins
- Simplifier la vie quotidienne des assurés car l'assuré n'aura plus à rechercher le calendrier précis lui permettant de résilier le contrat

2.2.2 Inquiétudes des assureurs

Du point de vue des assureurs, cette réforme inquiète. En effet, plusieurs effets néfastes peuvent impacter les assureurs :

- **Risque de démutualisation** : la mutualisation étant l'un des principes fondamentaux de l'assurance, cette réforme pourrait l'empêcher et donc accentuer l'anti-sélection. Effectivement, l'assuré souscrit un contrat très protecteur compte tenu d'importantes dépenses à réaliser et pourrait résilier aussitôt que ces soins lui ont été remboursés.
- **Perte de primes** : les primes reçues (acquises) ne le sont pas vraiment car susceptibles d'être remboursées en partie si la résiliation a lieu pendant la période de couverture future.
- **Charge de gestion supplémentaire pour les assureurs** : si l'adhérent résilie pour changer d'organisme assureur, c'est le nouvel organisme assureur qui effectue toutes les démarches nécessaires à la résiliation, en garantissant la continuité de couverture.

2.2.3 Possibles dispositions des assureurs

Afin de se protéger contre ces possibles conséquences, les assureurs pourraient prendre quelques dispositions :

- **Remise en place des délais de carence** : le fait qu'il soit plus facile de résilier et donc de changer de complémentaire santé peut amener certains organismes à mettre en place des délais de carence plus ou moins longs. Le but étant d'éviter une hausse des comportements opportunistes (un assuré résiliant son contrat suite à des soins importants).
- **Tarifification selon les risques** : Le fait qu'il n'y ait pas de durée minimale d'engagement met à mal le principe de mutualisation des risques et peut amener à une prise en compte du profil des assurés lors de la tarification, c'est-à-dire proposer des tarifs plus ou moins élevés selon les risques que présente l'assuré. Cette réforme encouragerait les complémentaires à sélectionner les individus présentant de bons risques (personnes jeunes et bien portantes) et de délaisser ceux qui portent de mauvais risques (personnes âgées ou malades).
- **Hausse des frais de gestion** : Si l'assuré résilie pour changer d'organisme assureur, c'est le nouvel organisme assureur qui effectue toutes les démarches nécessaires à la résiliation, en garantissant la continuité de couverture. Dans ce sens, faire croître les frais de gestion permettra de compenser l'anti-sélection et les résiliations qui interviendront à tout moment de l'année.

2.3 La réforme 100% santé

2.3.1 Description et objectifs de la réforme

La réforme 100% Santé a pour objectif d'améliorer l'accès aux soins en réduisant les restes à charge pour les patients en audiologie, optique et dentaire.

En effet, d'après l'enquête Statistiques sur les ressources et conditions de vie de l'Insee, réalisée en 2017, 3,1 % des personnes de 16 ans ou plus vivant en France métropolitaine ont renoncé à des soins médicaux, soit environ 1,6 million de personnes.

De ce fait, la réforme oblige à terme les praticiens de santé à proposer des paniers ayant un reste à charge nul, appelés paniers 100% santé. Ainsi, en choisissant ces types de paniers, les soins et équipements seront intégralement remboursés par l'Assurance maladie obligatoire et la complémentaire santé.

Le cahier des charges des produits proposés ayant un reste à charge nul doit respecter un niveau minimum de qualité et doit couvrir tous les niveaux de soins pour que chaque client puisse trouver un panier lui correspondant. La seule contrainte pour accéder aux paniers de reste à charge nul est de posséder un contrat responsable. Si l'assuré n'est pas adhérent à une complémentaire santé ou qu'il a souscrit un contrat de mutuelle non-responsable, il ne pourra pas bénéficier de l'offre.

Voici concrètement la composition des paniers 100% santé et des paniers libres par poste de soin :

Optique : Pour les lunettes, deux classes (A et B) ont été définies :

Panier 100% santé	Panier libre
<p>Monture : prix limite de vente à 30 € ; norme CE</p> <p>Verres de classe A traitant tout trouble visuel avec amincissement, durcissement et antireflet</p> <p>Réévaluation de la base de remboursement des verres de classe A en moyenne de 30%</p>	<p>Monture : prix libres mais passage de 150 € à 100 € de remboursement maximum dans les contrats santé responsables</p> <p>Verres de la classe B dont la base de remboursement de la Sécurité sociale va fortement diminuer</p>

Source : *Qu'est-ce que le « 100% santé » en optique, audiologie et dentaire ? [2018] | <https://www.swisslife.fr>*

Audio : Les aides auditives sont divisées en deux groupes (1 et 2) selon leurs caractéristiques techniques et du nombre d'options choisies :

Panier 100% santé	Panier libre
<p>Aides auditive de Classe 1 :</p> <ul style="list-style-type: none"> - Tous les types d'appareils : contour d'oreille classique, contour à écouteur déporté ou intra-auriculaire - Avec au moins 12 canaux de réglage (ou dispositif de qualité équivalente) ; système permettant l'amplification des sons extérieurs restituée à hauteur d'au moins 30 dB ; - 4 ans de garantie ; - Et avoir au moins 3 options <p>Augmentation du niveau de remboursement de la Sécurité sociale</p> <p>Prix limite de vente qui diminueront progressivement sur 3 ans</p>	<p>Aides auditives de classe 2</p> <p>Avec des caractéristiques et des options supérieures aux aides auditives de classe 1.</p>

Source : Qu'est-ce que le « 100% santé » en optique, audiologie et dentaire ? [2018] | <https://www.swisslife.fr>

Dentaire : pour les prothèses dentaires, 3 types de paniers ont été définis. Les matériaux des prothèses en fonction de la localisation des dents conditionnent le niveau de remboursement

Panier 100% santé	Paniers libres	
<p>Inlays core et couronnes transitoires ;</p> <p>Prothèses fixes ou mobiles</p> <p>⇒ Couronnes</p> <ul style="list-style-type: none"> - Pour les dents visibles (les incisives, canines et 1e prémolaire) : céramique monolithique et céramo-métalliques ; - Pour les incisives et canines : céramique monolithique zircons ; - toute localisation : métalliques. <p>⇒ Bridges</p> <ul style="list-style-type: none"> - incisives et canines : céramo-métalliques ; - toute localisation : full zircon et métalliques ; - prothèses amovibles à base résine. <p>Augmentation du niveau de remboursement de la Sécurité sociale</p> <p>Prix limite de vente qui diminueront progressivement sur 3 ans</p>	<p>Tarifs maîtrisés</p> <p>⇒ Couronnes et bridges céramo-métalliques pour les dents intermédiaires et postérieures</p> <p>⇒ Appareils amovibles en stellite</p> <p>Augmentation du niveau de remboursement de la Sécurité sociale</p> <p>Prix limite de vente qui diminueront progressivement sur 3 ans</p>	<p>Tarifs libres</p> <p>⇒ Couronnes esthétiques 100% céramique</p> <p>⇒ Implantologie</p> <p>⇒ Parodontie</p> <p>⇒ Orthodontie</p> <p>⇒ Cosmétique dentaire</p>

Source : Qu'est-ce que le « 100% santé » en optique, audiologie et dentaire ? [2018] | <https://www.swisslife.fr>

Le déploiement de la réforme va différer selon les postes de soin. En effet, elle a été brutale pour le poste optique (mis en place générale en janvier 2020) mais plus étendue pour les 2 autres postes avec une mise en place progressive depuis 2019.

	2019	2020	2021
		Panier optique 100% santé garanti sans reste à charge	
	Plafonnement des prix pour les équipements du panier auditif « 100% santé » et augmentation du remboursement de la Sécurité sociale ⇒ Environ 200€ de moins par oreille par rapport à aujourd'hui	Montants plafonnés du panier auditif « 100% santé » encore réduits et nouvelle augmentation du remboursement de la Sécurité sociale ⇒ Environ 250€ de moins par oreille par rapport à aujourd'hui	Panier auditif 100% santé garanti sans reste à charge
	Plafonnement des tarifs des prothèses	Pas de reste à charge pour les couronnes et bridges du panier 100% santé	Panier dentaire 100% santé garanti sans reste à charge pour les autres équipements

Source : Qu'est-ce que le « 100% santé » en optique, audiologie et dentaire ? [2018] | <https://www.swisslife.fr>

2.3.2 Inquiétudes des assureurs

La réforme permet aux assurés de réduire fortement leur reste à charge. Néanmoins, la part de la sécurité sociale évoluant peu, c'est la part des complémentaires santé qui va potentiellement augmenter. Une hausse des coûts moyens est alors à prévoir. Une analyse plus détaillée montrant cet impact est située en partie 3.1.3.

3 Présentation du portefeuille standard santé

3.1 Le portefeuille santé individuelle et collective

Afin d'analyser la structure et la rentabilité du portefeuille santé, une étude est réalisée. L'objectif de celle-ci est de regarder l'évolution du portefeuille sur les 10 dernières années par génération. La génération désigne la date d'effet du contrat, c'est-à-dire la période durant laquelle les primes commencent à couvrir les garanties du contrat.

3.1.1 Structure du portefeuille santé individuelle

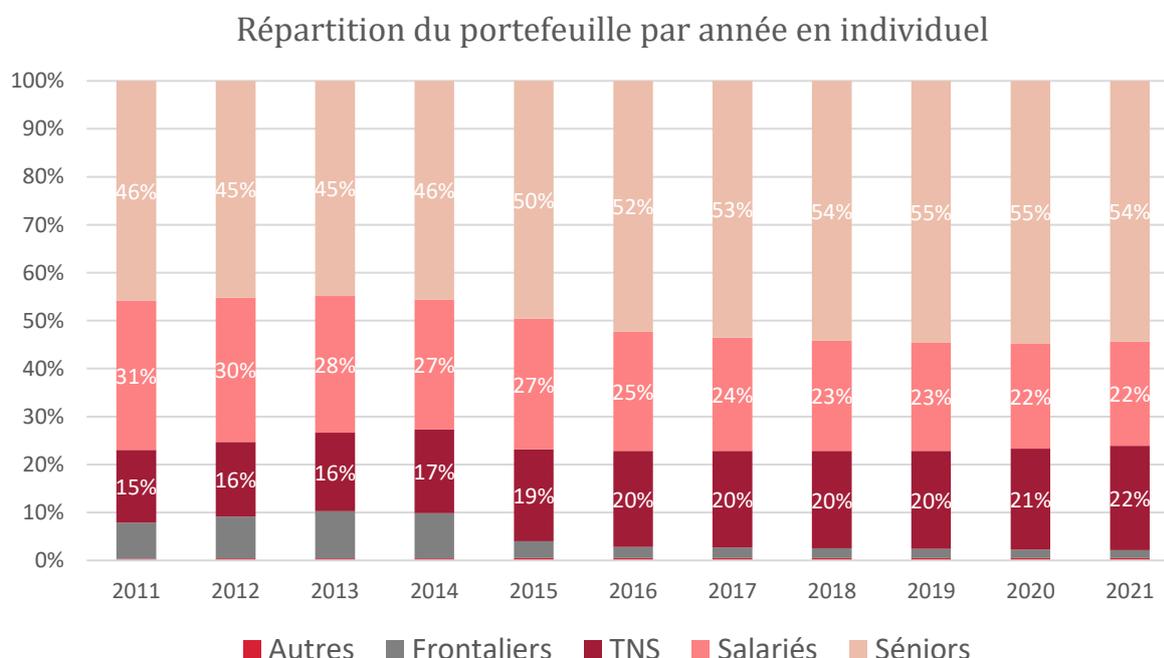


Figure 2 Répartition du portefeuille par année - individuel

La structure du portefeuille depuis 2011 a beaucoup évolué. En effet, on remarque par exemple que la part des séniors, population la plus à risque de consommer, est passée de 46% en 2011 à 56% en 2021 avec un premier changement de structure en 2015. En effet, cette année-là, la part des séniors a pour la première fois augmenté.

Une explication assez directe permet d'expliquer ce phénomène : la réforme des frontaliers. Précisément, on remarque que la part des frontaliers, la même année, a fortement diminué jusqu'à quasiment devenir nulle : environ 10% de frontaliers en 2014 contre 3.6% en 2015 et atteindre 1.7% en 2021. Cette réforme, qui a pris effet au 1er juin 2014, a bouleversé l'assurance maladie des frontaliers puisqu'elle a entraîné la disparition du régime privé d'assurance maladie, alors que 90%

d'entre eux y étaient affiliés. Les 1.7% de contrats frontaliers restants ne sont plus que des complémentaires frontaliers.

La perte donc des frontaliers a entraîné au global une plus forte part des séniors.

L'impact d'une seconde réforme est aussi visible : l'ANI.

En effet, la loi ANI, qui s'appuie sur l'Accord National Interprofessionnel signé en 2013, entre ainsi en vigueur le 1er janvier 2016. Elle oblige tous les employeurs du privé et les associations à fournir une complémentaire santé à leurs salariés, financée au moins à 50% par la société. Cette réforme a donc entraîné une perte importante de contrats individuels salariés : 31% de salariés en 2011 contre 22% en 2021. La part des TNS (Travailleurs Non-Salariés) a donc globalement augmenté pour atteindre une part égale à celle des salariés en 2021.

3.1.2 Taux d'extinction du portefeuille

Une génération correspond à une année d'effet des contrats. Ainsi, par exemple, la génération 2018 correspond à l'ensemble des contrats ayant pris effet en 2018. Ils correspondent donc à une ancienneté de 3 ans à fin 2021. Le taux d'extinction du portefeuille mesure alors le volume de primes restant en portefeuille par rapport au volume de primes initial.

Triangle de Primes		INDIVIDUEL										
Génération	Survenance											
	N	N+1	N+2	N+3	N+4	N+5	N+6	N+7	N+8	N+9	N+10	
2011		100%	81%	65%	49%	39%	34%	30%	27%	25%	23%	
2012		100%	80%	60%	47%	41%	36%	33%	29%	26%	6%	
2013		100%	79%	61%	53%	48%	43%	39%	35%	9%		
2014		100%	80%	69%	62%	56%	52%	46%	11%			
2015		100%	88%	78%	70%	65%	57%	14%				
2016		100%	88%	78%	71%	62%	15%					
2017		100%	86%	75%	63%	15%						
2018		100%	87%	73%	17%							
2019		100%	84%	20%								
2020		100%	23%									
2021		100%										
TOTAL		100%	82%	68%	56%	48%	41%	35%	30%	26%	23%	

année incomplète : 2022

Figure 3 Taux d'extinction du portefeuille par génération individuel

Remarque : La survenance N+1 est utilisée comme point de départ de la liquidation des primes dans le temps car lors de la survenance N, certains contrats peuvent prendre effet en fin d'année. Par conséquent, en prenant comme point de départ la survenance N, une grande partie des primes manquerait. Prenons un exemple concret : un contrat prend effet le 01/06/N. En ne considérant que les primes payées lors de l'année de survenance N, la moitié des primes manquerait car le contrat a pris effet en milieu d'année. En considérant comme point de départ les primes de l'année suivante, on obtient un volume de prime annuel et total, ce qui permet ensuite de voir le taux de primes qui disparaît dans le temps.

En moyenne, 5 ans après la date d'effet du contrat (survenance N+6), il ne reste que 41% du portefeuille individuel. De plus, on remarque que plus les générations sont récentes, plus le

portefeuille s'éteint lentement. En effet, pour la génération 2011, il ne restait que 34% des primes 5 ans après la date d'effet du contrat contre 57% pour la génération 2015.

Triangle de Primes		COLLECTIF										
Génération	Survénance											
	N	N+1	N+2	N+3	N+4	N+5	N+6	N+7	N+8	N+9	N+10	
2011		100%	90%	83%	73%	59%	45%	1%	1%	1%	0%	
2012		100%	93%	85%	56%	46%	1%	1%	1%	0%		
2013		100%	93%	75%	64%	1%	1%	1%	0%			
2014		100%	86%	79%	3%	3%	3%	2%	0%			
2015		100%	94%	85%	82%	74%	67%	4%				
2016		100%	91%	87%	78%	71%	4%					
2017		100%	93%	81%	77%	5%						
2018		100%	89%	81%	4%							
2019		100%	93%	8%								
2020		100%	9%									
2021		100%										
TOTAL	0	100%	91%	83%	66%	53%	25%	2%	1%	0%	0%	

année incomplète : 2022

Figure 4 Taux d'extinction du portefeuille collectif

Concernant le collectif, en moyenne, 5 ans après la date d'effet du contrat, il ne reste que 25% du portefeuille, soit 16% de moins qu'en collectif. De plus, on remarque que les taux d'extinctions sont beaucoup plus volatiles, en particulier pour les générations inférieures à 2015. L'impact de l'ANI est une nouvelle fois visible. En effet, en voyant les volumes de primes dans la partie droite du tableau, on voit clairement qu'à partir de 2016, l'année de la réforme, les montants de primes ont fortement augmenté (on peut observer l'effet inverse pour le portefeuille individuel car les salariés avaient des contrats individuels avant de passer en collectif suite à la réforme). La volatilité des montants est alors expliquée par le fait qu'avant la réforme, le portefeuille collectif n'était pas très développé (environ 3 000 contrats). Il suffit donc que quelques clients résilient pour que l'impact sur le taux soit immédiat, là où après 2016 l'impact serait atténué par le plus grand nombre de contrats.

3.1.3 Répartition des prestations par génération et par poste de soins

Evolution des prestations par poste - Individuel

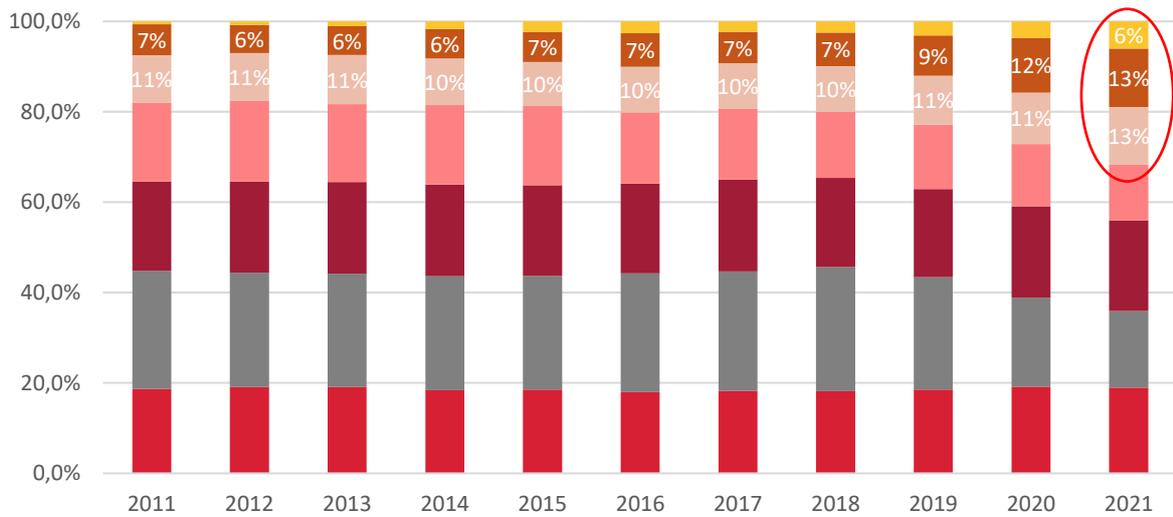


Figure 5 Evolution des prestations par poste - Individuel

Evolution du CM/assuré en montant - Individuel

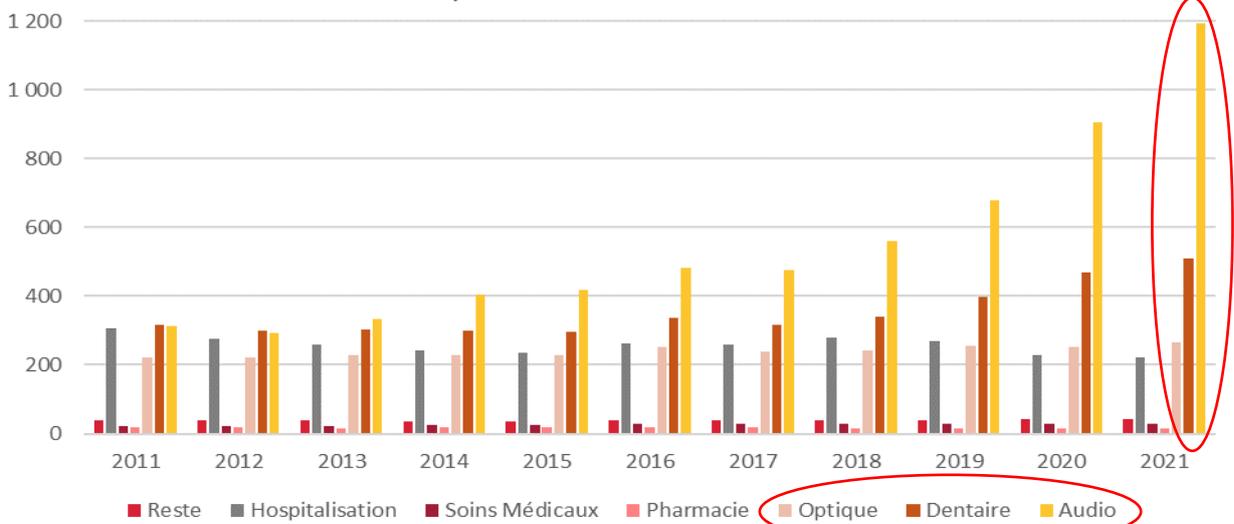


Figure 6 Evolution du CM/assuré en montant - Individuel

L'impact de la réforme 100% santé est ici clairement visible. En effet, en observant le 1er graphique, on voit que la part des prestations réalisées au sein des postes 100% santé (entourés en rouge) ne représentait que 18% des prestations totales pour la génération 2011, le poste audio étant quasiment nul. Cependant, pour la génération 2021, cette part atteint 32%, soit un tiers des prestations totales, avec une grosse évolution du poste audio. Cette évolution est nettement plus visible en termes de coût moyen/assuré grâce au second graphique. En 4 ans, le coût moyen par assuré du poste audio a plus que doublé, passant de 560€ en 2018 à 1200€ en 2021. Le coût moyen du poste dentaire a quant à lui également fortement augmenté, passant de 350€ en 2018 à 500€ en 2021, contrairement au poste optique qui reste stable malgré la réforme.

Evolution des prestations par poste - collectif

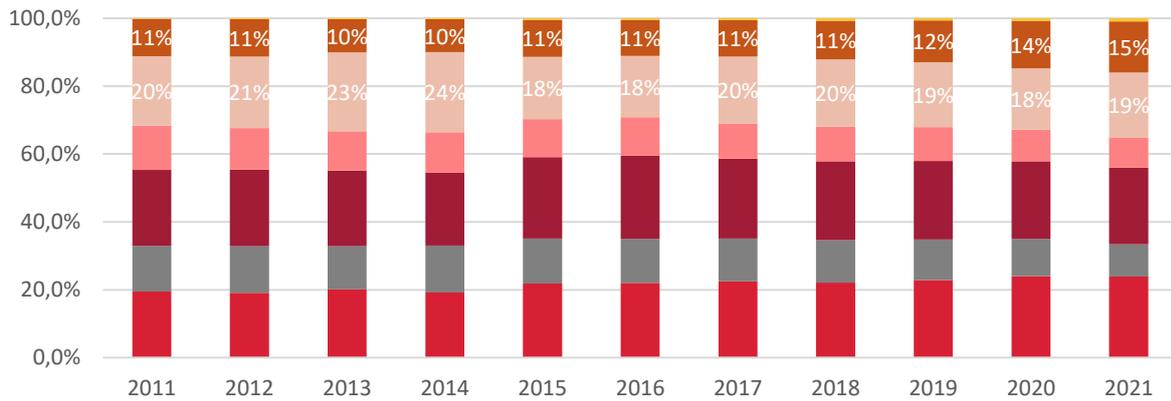


Figure 7 Evolution des prestations par poste - collectif

Evolution du CM/assuré en montant - collectif

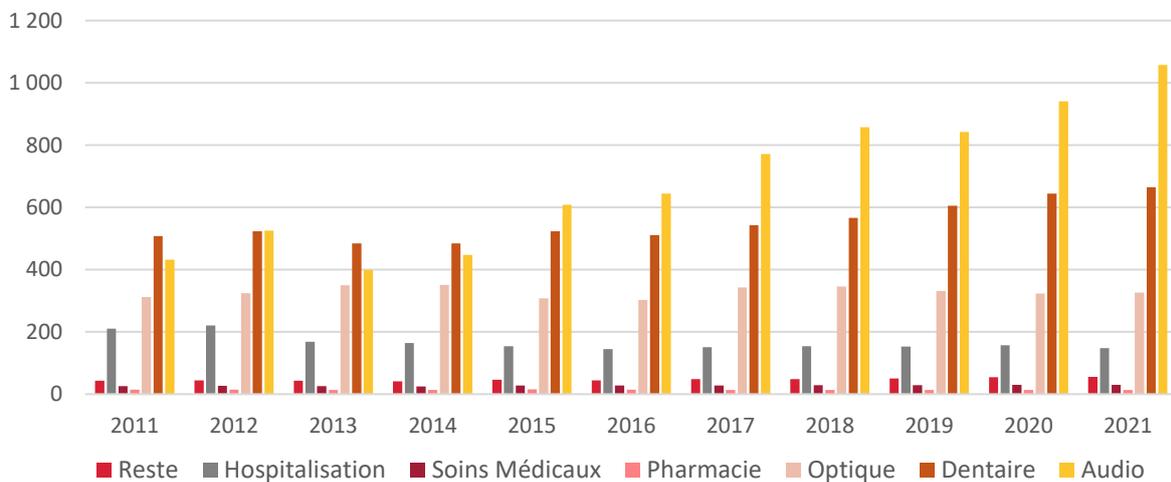


Figure 8 Evolution du CM/assuré - collectif

Concernant le collectif, l'impact de la réforme est nettement moins visible. Que ce soit en termes de coût moyen par assuré ou en part des prestations totales, les postes 100% santé sont restés assez stables après la réforme. En effet, les garanties en collectif étant plus élevées (produits plus haut de gamme), la recours aux offres 100% santé est ainsi limitée.

3.1.4 Evolution de la consommation dans le temps par génération

En prenant deux générations de référence, ayant à la fois des dates suffisamment différentes pour pouvoir les comparer, et suffisamment d'historique pour pouvoir analyser assez d'années de survéance, on peut alors comparer l'évolution de la consommation et de la rentabilité dans le temps.

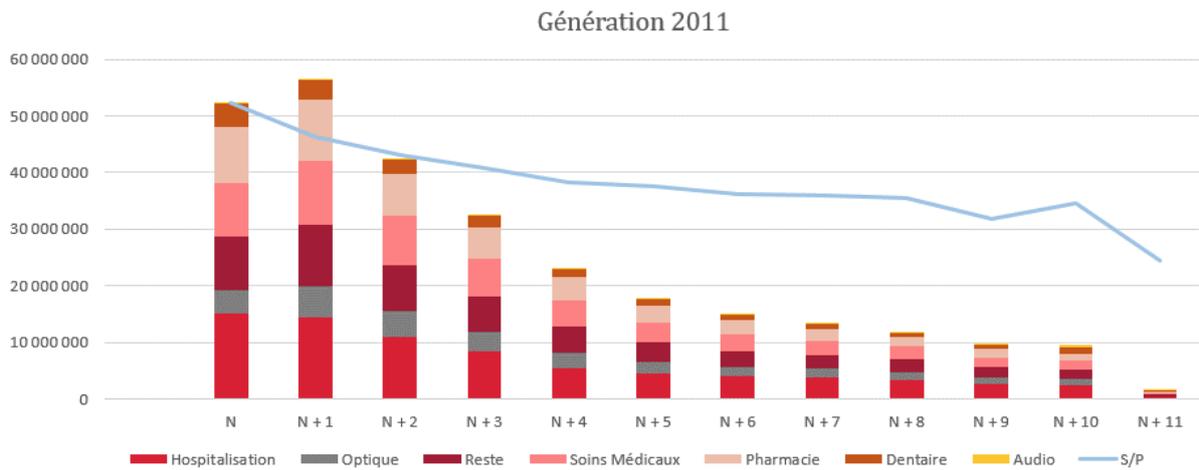


Figure 9 Evolution de la consommation dans le temps - génération 2011 *individuel*

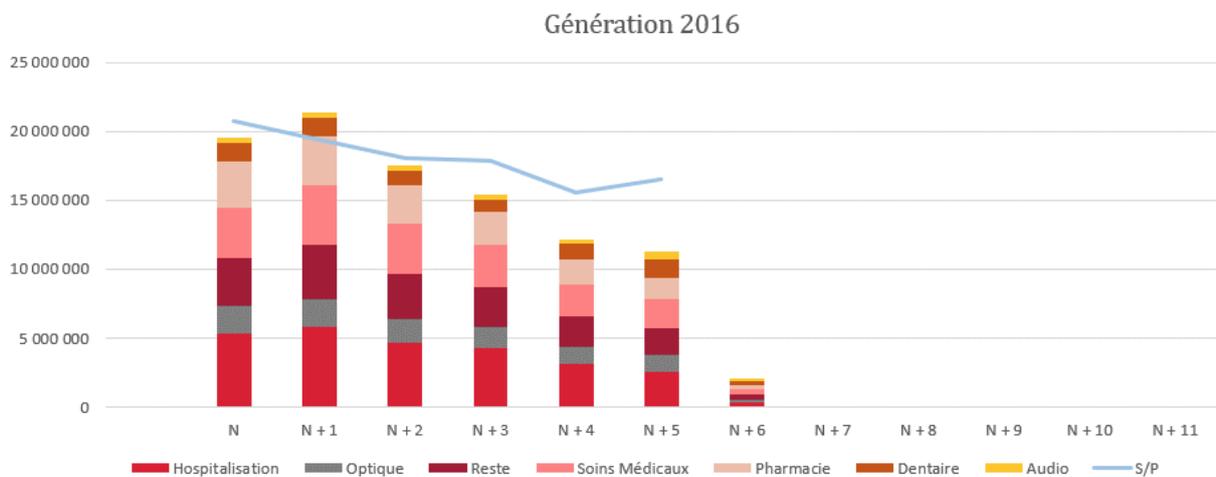


Figure 10 Evolution de la consommation dans le temps - génération 2016 *individuel*

Remarque : par soucis de confidentialité, l'échelle des S/P n'est ici pas indiquée

Contrairement aux contrats collectifs qui depuis l'ANI obligent les entreprises à proposer une assurance santé aux employés, l'assuré individuel a lui le choix de souscrire ou non à une assurance santé. Sa souscription vient donc d'un besoin de soins assez immédiat. On voit clairement ce phénomène en observant les S/P années après années. Le S/P, indicateur de rentabilité, mesurant le rapport entre les sinistres et les primes encaissées, est élevé les premières années du contrat et est surtout décroissant. Précisément, une fois les soins de l'assuré effectués assez rapidement après la souscription, une routine s'installe alors. L'assuré n'a plus besoin de soins et donc la rentabilité s'améliore années après années.

Cependant, on remarque une augmentation du S/P en N+10 pour la génération 2011 et en N+5 pour la génération 2016. Cela correspond donc à l'année de survenance 2021, année du 100% santé. L'impact de la réforme est une nouvelle fois visible. Alors que la rentabilité s'améliorait au fur et à mesure de la vie du contrat, la réforme vient dégrader celle-ci.

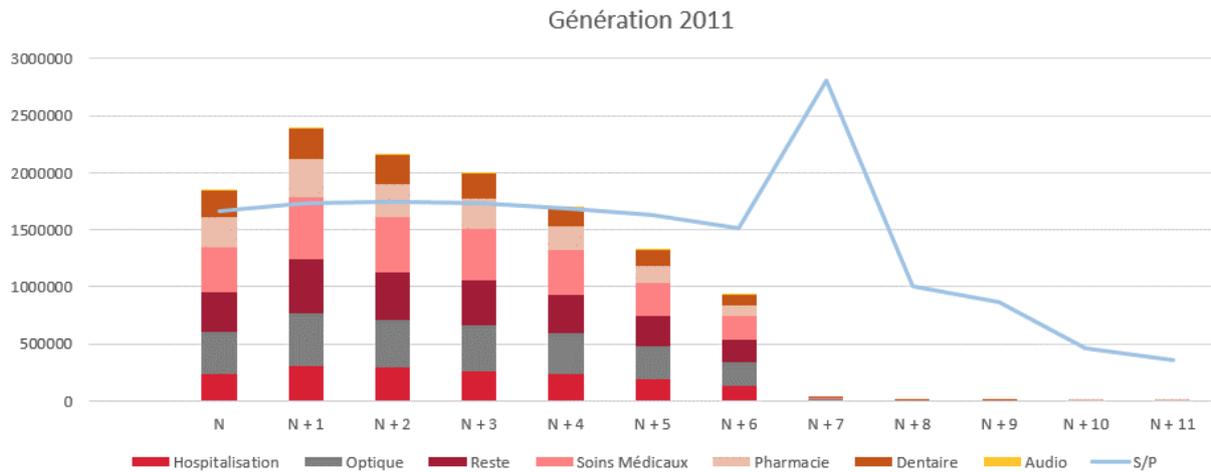


Figure 11 Evolution de la consommation dans le temps - génération 2011 collectif

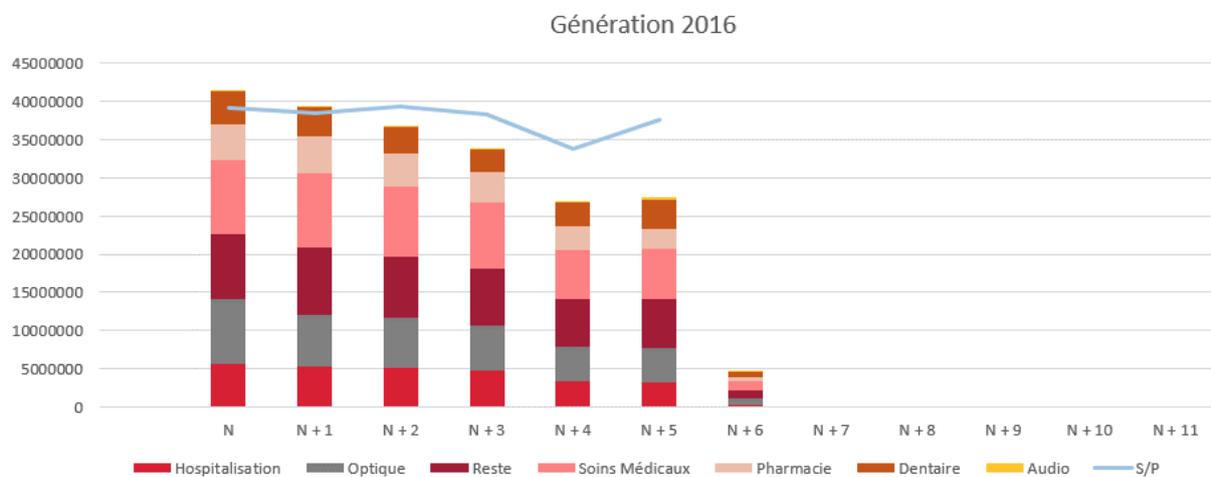


Figure 12 Evolution de la consommation dans le temps - génération 2016 collectif

La rentabilité en collectif se comporte différemment de l'individuel. Effectivement, on remarque que les S/P sont assez linéaires au fil des années, ce qui est cohérent avec le fait que les contrats collectifs santé sont obligatoires. La souscription du contrat ne vient pas d'un besoin immédiat de l'assuré, il n'y a donc pas de hausse des soins lors des premières années du contrat.

3.1.5 Evolution de la rentabilité par génération au cours de la durée de vie du contrat

Les deux graphiques suivants illustrent les écarts de S/P tout au long de la durée de vie des contrats par rapport au S/P initial.

Triangle de S/P	INDIVIDUEL										
Génération	Survenance										
	N	N + 1	N + 2	N + 3	N + 4	N + 5	N + 6	N + 7	N + 8	N + 9	N + 10
2011	100,0%	-10,1%	-15,4%	-19,2%	-23,6%	-24,7%	-26,9%	-27,2%	-28,2%	-34,1%	-29,6%
2012	100,0%	-6,3%	-11,0%	-15,1%	-17,6%	-19,9%	-21,6%	-22,6%	-28,4%	-22,6%	
2013	100,0%	-2,3%	-8,0%	-10,4%	-13,0%	-14,2%	-14,7%	-21,2%	-15,9%		
2014	100,0%	-1,1%	-5,1%	-10,3%	-10,4%	-11,9%	-18,4%	-13,0%			
2015	100,0%	-0,7%	-8,9%	-10,1%	-12,4%	-19,8%	-13,4%				
2016	100,0%	-5,5%	-11,0%	-11,8%	-20,9%	-16,9%					
2017	100,0%	-6,8%	-14,2%	-22,9%	-17,8%						
2018	100,0%	-4,6%	-17,3%	-13,2%							
2019	100,0%	-14,8%	-11,0%								
2020	100,0%	5,0%									
2021	100,0%										
TOTAL	100%	-5%	-11%	-14%	-17%	-18%	-19%	-21%	-24%	-28%	-30%

année incomplète : 2022

Figure 13 Evolution de la rentabilité par génération - individuel

Triangle de S/P	COLLECTIF										
Génération	Survenance										
	N	N + 1	N + 2	N + 3	N + 4	N + 5	N + 6	N + 7	N + 8	N + 9	N + 10
2011	100,0%	3,4%	3,7%	3,2%	1,2%	-1,3%	-7,1%	53,4%	-30,7%	-37,0%	-56,2%
2012	100,0%	7,7%	3,9%	1,5%	-2,0%	-10,1%	58,5%	-1,8%	-11,8%		
2013	100,0%	4,8%	1,7%	3,9%	-2,6%	-2,5%	-37,5%	-42,2%	101,2%		
2014	100,0%	5,6%	1,3%	-7,1%	13,3%	-19,1%	-19,8%	-15,8%			
2015	100,0%	6,2%	0,0%	-1,9%	-4,3%	-14,9%	-5,4%				
2016	100,0%	-1,2%	0,5%	-1,8%	-11,9%	-3,4%					
2017	100,0%	1,4%	3,2%	-3,8%	4,8%						
2018	100,0%	-0,1%	-6,1%	1,9%							
2019	100,0%	-4,7%	6,5%								
2020	100,0%	7,0%									
2021	100,0%										
TOTAL	100%	3%	2%	-1%	0%	-9%	-2%	-2%	20%	-37%	-56%

année incomplète : 2022

Figure 14 Evolution de la rentabilité par génération - collectif

L'interprétation effectuée précédemment est illustrée ici pour toutes les générations. Effectivement, pour les contrats individuels, la rentabilité s'améliore au fur et à mesure des années. On remarque qu'au global, le S/P baisse de 18% au bout de 5 années d'ancienneté. Les contrats collectifs eux ont une rentabilité beaucoup plus stable au cours du temps. En effet, celle-ci est pratiquement linéaire sur les 5 premières années avant d'être beaucoup plus volatile les années suivantes, lié au volume faible de contrats restant dans le portefeuille après 5 ans.

3.1.6 Conclusion de l'étude

Les différentes réformes réglementaires ont donc eu un impact sur notre portefeuille, que ce soit en termes de structure, de durée de vie des contrats ou de rentabilité. Cependant, ces réformes, notamment la RIA et le 100% santé, étant récentes, leurs impacts à terme est incertain. En effet, il est important de pouvoir projeter de la manière la plus précise possible la rentabilité ou la durée de vie des contrats afin d'anticiper les impacts, notamment négatifs de ces récentes réformes.

Ainsi, la projection du portefeuille se fera à 3 ans et sera concentrée sur les contrats individuels. Effectivement, les contrats collectifs n'ont pas été très impactés par la réforme RIA et par le 100% santé. Précisément, la résiliation d'un contrat collectif est beaucoup plus compliquée qu'un contrat individuel. La réforme RIA a donc beaucoup moins d'impact en collectif. Les quelques ayant déjà eu recours à la résiliation infra-annuelle l'ont fait principalement dans un contexte de mécontentement quant aux garanties, au suivi ou aux tarifs appliqués par l'organisme de complémentaire santé qu'elles avaient choisi initialement.

De plus, concernant le 100% santé, comme dit précédemment, les garanties en collectifs étant meilleures, le recours au 100% santé est beaucoup moindre en collectif qu'en individuel.

La projection de la rentabilité est donc incertaine pour les contrats individuels. La rentabilité repose sur les flux (primes et prestations) et les effectifs, la réforme RIA offre donc une vision très incertaine des effectifs car une hausse prolongée des résiliations entrainerait un environnement instable et donc plus compliqué à piloter. La réforme 100% santé perturbe les prestations. Ce mémoire va donc tenter d'éclaircir la vision à 3 ans de la rentabilité, en considérant notamment plusieurs scénarios.

4 Modélisation des résiliations

4.1 Approche à priori

4.1.1 Impact déjà observé de la réforme RIA

La réforme RIA offre à l'assuré la possibilité de résiliation à tout moment après un an d'ancienneté. Avant la réforme, les assurés devaient attendre la date d'anniversaire du contrat. Par conséquent, les résiliations sont beaucoup plus simples et rapides qu'auparavant. L'impact sur le nombre et le taux de chutes devraient être significatif :

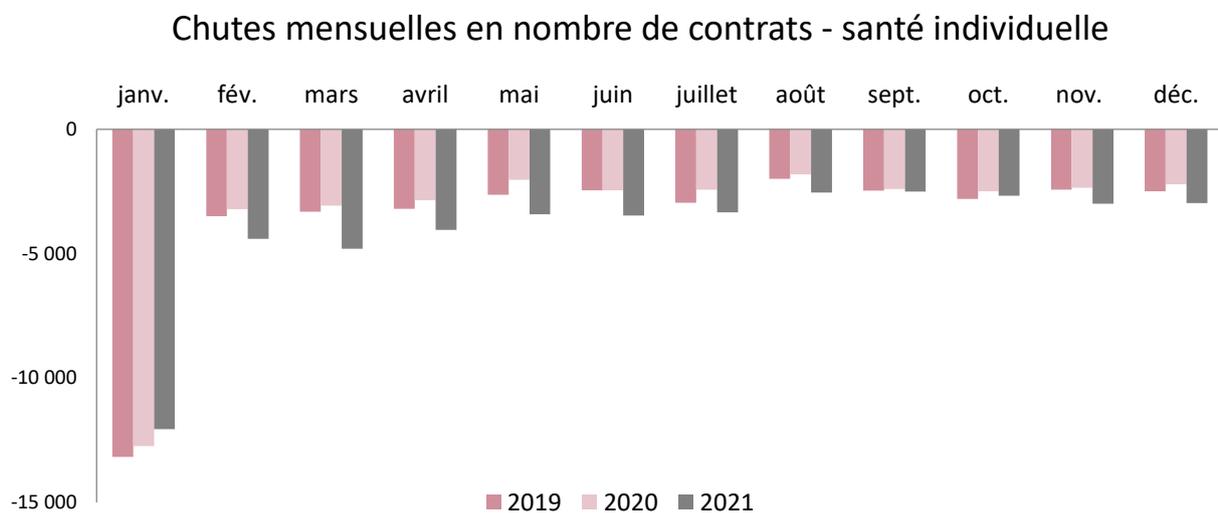


Figure 15 Chutes mensuelles en nombre de contrats - santé individuelle

On remarque que les chutes sont beaucoup plus mensuelles qu'avant la réforme.

Voici l'impact annuel sur le taux de chutes :

Taux de chutes - santé individuelle

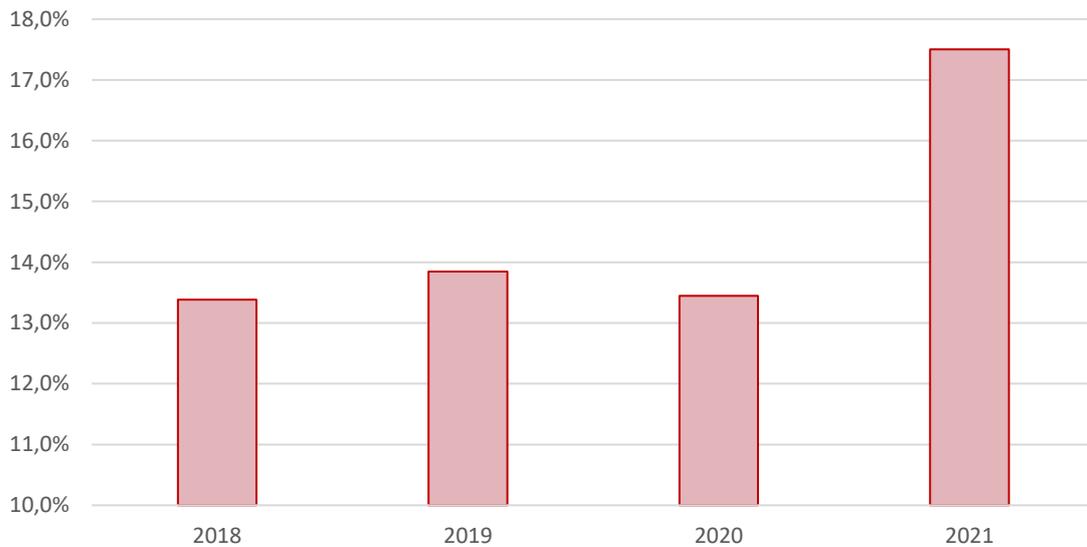


Figure 16 Taux de chutes annuels - santé individuelle

Avant la réforme, le taux de chutes était stable. Il était en moyenne de 13,6%. En 2021, il a fortement augmenté pour atteindre 17,5%, soit presque 4% de chutes en plus ce qui représente une hausse de 28,7% ($\frac{17,5-13,6}{13,6}$). L'impact est donc très significatif.

4.1.2 Hypothèse considérée dans la modélisation des résiliations

Cependant, à l'instar de la loi Hamon entrée en vigueur en 2015 et offrant exactement les mêmes droits que la RIA mais pour des contrats d'assurance automobile, moto, multirisque habitation..., l'impact sur les prochaines années est incertain. En effet, l'impact de la loi Hamon a été significatif la première année mais les effets se sont ensuite estompés voir nullifiés quelques années après.

Ainsi, dans une optique de projection du portefeuille à 3 ans, il est nécessaire de prendre en compte cet aléa. Pour cela, nous allons construire 2 scénarios qui impacteront les chutes puis les affaires nouvelles :

- Scénario dit « RIA » dans lequel on suppose que l'effet de la RIA ne s'estompe pas au cours du temps
- Scénario dit « Hors RIA » dans lequel l'impact sur les chutes devient nul après 2021

Certes le deuxième scénario est un peu brutal mais il servira surtout de comparaison avec le premier scénario afin d'observer les différents effets prospectifs de la réforme.

L'historique des chutes sera donc différent selon le scénario :

- Scénario RIA : données de 2021 permettant de capter la hausse des chutes de la réforme
- Scénario Hors RIA : données de 2018 à 2020 pour ne pas capter la hausse des chutes

Dans le second scénario, nous avons choisi cet historique de données pour 2 raisons.

La première raison est que la réforme de l'ANI entrée en vigueur en 2016 a provoqué une hausse des chutes en santé individuelle car les contrats individuels sont devenus des contrats obligatoires collectifs. Pour ne pas polluer nos données de cette hausse des chutes s'étalant jusqu'à 2017, nous décidons de ne pas prendre 2016 et 2017 dans notre historique de résiliations.

De plus, pour ne pas avoir des comportements de résiliations obsolètes, on décide de débiter notre historique à 2018. Nous aurons donc 3 années d'historique pour le Scénario Hors RIA, ce qui semble suffisant pour construire un modèle robuste représentant le comportement des résiliations avant la réforme.

4.2 Modélisation des taux bruts de résiliation

4.2.1 L'estimateur de Kaplan-Meier

4.2.1.1 Prise en compte de la censure

L'étude des résiliations est réalisée sur une période restreinte dans les 2 scénarios : 1 an pour le scénario RIA et 3 ans pour le scénario Hors RIA. Ainsi, la majorité des individus n'auront pas résilié à la fin de ces périodes d'observation. La variable d'intérêt (la résiliation) n'est alors pas observée. Ce phénomène est appelé « censure à droite ». Pour les individus censurés, on observe donc une durée plus petite que la variable d'intérêt.

De ce fait, si la censure n'est pas prise en compte, en faisant comme si une donnée censurée est égale à notre variable d'intérêt, les durées auront tendance à être sous évaluées. L'estimateur de Kaplan-Meier permet alors la prise en compte de ce phénomène.

On peut définir le terme "durée de survie" comme le temps restant jusqu'à la survenance d'un évènement particulier. Dans notre cas, la résiliation.

L'estimateur non-paramétrique de Kaplan Meier permet de construire une fonction de survie.

4.2.1.2 Théorie mathématique

- Principe de l'estimateur :

L'estimateur s'appuie sur l'idée suivante : dans l'intervalle d'ancienneté $[x; x + 1[$, la probabilité de ne pas résilier son contrat dans un temps t_i peut s'écrire :

$$S_x(t_m) = P(T > t_m)$$

D'après la formule de Bayes, $\forall t_m > t_{m-1} > \dots > t_2 > t_1$ la fonction de survie peut donc s'écrire :

$$\begin{aligned}
S_x(t_m) &= P(T > t_m | T > t_{m-1}) \cdot S(t_{m-1}) \\
&= P(T > t_m | T > t_{m-1}) \cdot P(T > t_{m-1} | T > t_{m-2}) \cdot S(t_{m-2}) \\
&= \dots \\
&= P(T > t_m | T > t_{m-1}) \dots P(T > t_2 | T > t_1) \cdot S(t_1)
\end{aligned}$$

avec T l'ancienneté à la survenance de l'évènement.

En choisissant comme instants de conditionnement les anciennetés $t_{(j)} \in [x; x + 1[$ où se produisent des évènements avec $t_{(0)} = x$, cela revient donc à estimer les probabilités :

$$p_j = P(T \geq t_{(j+1)} | T > t_{(j)}) \text{ et } i_j = P(T < t_{(j+1)} | T > t_{(j)})$$

Avec p_j la probabilité de ne pas résilier son contrat sur l'intervalle d'ancienneté $[t_{(j)}; t_{(j+1)}[$ sachant qu'on est présent à l'ancienneté de $t_{(j)}$.

Et $i_j = 1 - p_j$ la probabilité de résilier son contrat sur l'intervalle d'ancienneté $[t_{(j)}; t_{(j+1)}[$ sachant qu'on est présent à l'ancienneté de $t_{(j)}$.

La fonction de survie au point t pour une ancienneté x

$$S_x(t) = p_0 p_1 \dots p_{t-1} = \prod_{j|t_{(j)} < x} (1 - i_j)$$

- Expression de l'estimateur

Les $i_j = 1 - p_j$ sont alors estimés par la méthode du maximum de vraisemblance à partir de l'observation d'un n -échantillon.

Notons :

- d_j : le nombre de résiliations en t_j
- n_j : le nombre d'assurés exposés au risque juste avant t_j .

Le nombre de résiliation suit une loi Binomiale $\mathcal{B}(n_j, i_j)$. La fonction de vraisemblance s'écrit alors :

$$\mathcal{L}(i_1; i_2; \dots i_m) = \prod_{j=1}^m C_{n_j}^{d_j} i_j^{d_j} (1 - i_j)^{n_j - d_j}$$

Pour trouver le maximum de vraisemblance, on égalise à 0 la dérivée du premier ordre de la log-vraisemblance :

$$\begin{aligned} \frac{\partial \ln(L)}{\partial i_j} &= 0 \quad \forall j \in \{1, \dots, m\} \\ \Leftrightarrow \frac{\partial \sum_{i=1}^n \ln C_{n_j}^{d_j} + d_j \ln(i_j) + (n_j - d_j) \ln(1 - d_j)}{\partial i_j} &= 0 \quad \forall j \in \{1, \dots, m\} \\ \Leftrightarrow \frac{d_j}{i_j} - \frac{(n_j - d_j)}{1 - i_j} &= 0 \quad \forall j \in \{1, \dots, m\} \\ \Leftrightarrow \hat{i}_j &= \frac{d_j}{n_j} \quad \forall j \in \{1, \dots, m\} \end{aligned}$$

On en déduit alors l'estimateur de la fonction de survie :

$$\hat{S}_x(t) = \prod_{j|t_{(j)} < x} (1 - \hat{i}_j) = \prod_{j|t_{(j)} < x} \left(1 - \frac{d_j}{n_j}\right)$$

Et donc :

$$\frac{\hat{S}_{x+1}}{\hat{S}_x} = \prod_{j|x < t_{(j)} < x+1} \left(1 - \frac{d_{x,j}}{n_{x,j}}\right)$$

On peut alors en déduire l'estimation de la probabilité de résilier à l'ancienneté x (sur l'intervalle $[x, x + 1[$) :

$$\begin{aligned} \hat{i}_x &= 1 - \frac{\hat{S}_{x+1}}{\hat{S}_x} \\ &= 1 - \prod_{j|x < t_{(j)} < x+1} \left(1 - \frac{d_j}{n_j}\right) \end{aligned}$$

Nous remarquons que le nombre d'individus $n_{x,j}$ sous risque à l'ancienneté $t_{(j)}$ varie en fonction de ceux présents à l'instant $j - 1$ et de ceux censurés ou tronqués dans l'intervalle $[t_{(j-1)}; t_{(j)}[$.

En effet, pour une partie des assurés, toute l'information n'est pas observable, on parle alors de censure.

Pour estimer une loi de résiliation par l'ancienneté, il est nécessaire d'observer les individus d'un échantillon pendant l'année entière où leur ancienneté est de x années. Nous observons seulement les assurés sur la durée de présence qu'ils ont dans le portefeuille ce qui n'inclut pas toujours l'année entière où l'assuré à l'ancienneté x . Lorsqu'il manque l'observation en début d'ancienneté, on parle de troncature (ou censure gauche) et en fin d'ancienneté, on parle de censure droite.

Ici, étant donné que nous avons une visibilité du contrat depuis son origine, les cas de troncature n'existent pas. Seules les censures à droite peuvent être présentes notamment lorsque :

- L'assuré résilie son contrat avant d'atteindre l'ancienneté $x + 1$,
- L'assuré décède avant d'atteindre l'ancienneté $x + 1$,
- L'assuré n'a pas encore atteint l'ancienneté $x + 1$ à la fin de la période d'observation

Soient :

- c_j l'effectif des individus censurés entre T_{j-1} et T_j ,

Nous avons alors : $n_j = n_{j-1} - d_{j-1} - c_{j-1}$

- Propriétés de l'estimateur

L'estimateur des i_x utilisé par Kaplan-Meier est l'estimateur du maximum de vraisemblance.

D'après le Lemme de Fleming et Harrington, l'estimateur de Kaplan-Meier est un estimateur non paramétrique sans biais : $E(\hat{S}(t)) = S(t)$.

Un estimateur de la variance de Kaplan-Meier est donné par la formule de Greenwood :

$$Var(\hat{S}) = \hat{S}(t)^2 \cdot \sum_{j|t_j \leq t} \frac{d_j}{(n_j - d_j)n_j}$$

4.2.1.3 Taux bruts estimés par Kaplan-Meier

Afin d'être cohérent dans l'approche générationnelle, l'étude a été regroupée par ancienneté. De cette manière, les taux de chutes sont calculés ancienneté par ancienneté. Les affaires nouvelles (ancienneté inférieure à 1 ans) ne viennent donc pas biaiser les taux. De plus, pour garder une volumétrie suffisante de données par segments, les taux des chutes sont unidimensionnels (seulement segmentés par ancienneté).

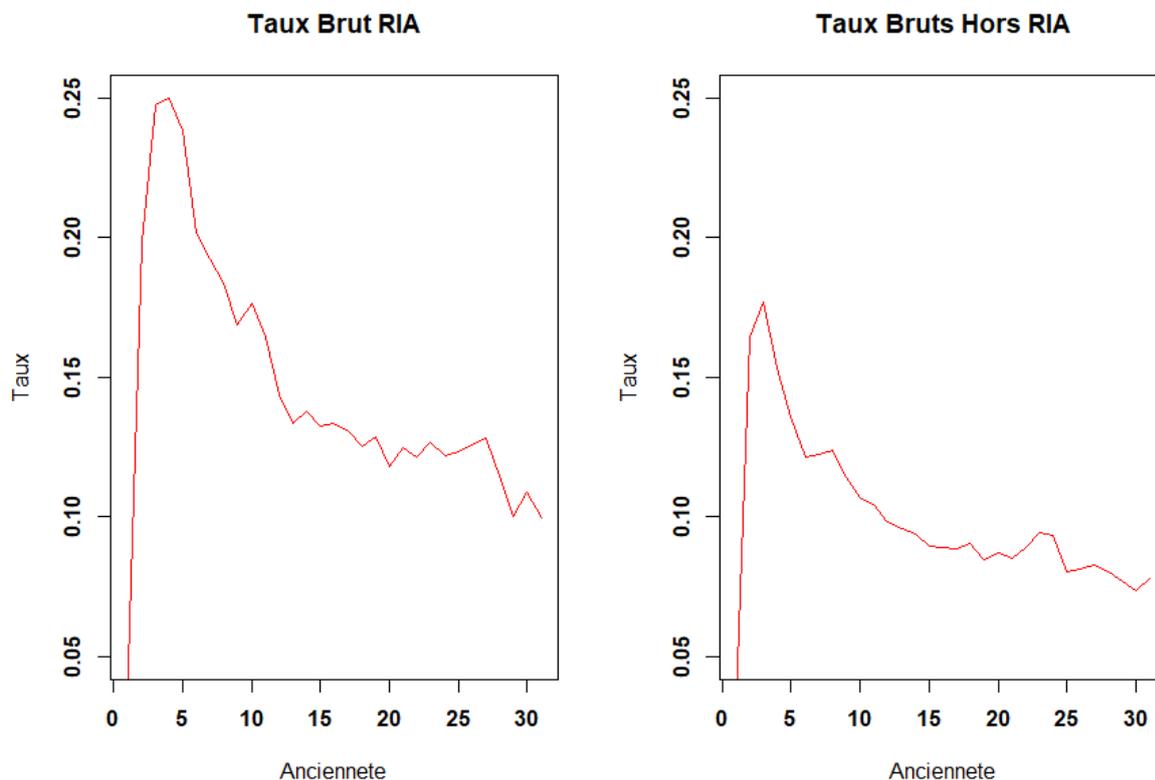


Figure 17 Taux bruts estimés par Kaplan-Meier

L'impact de la réforme est clair : la tendance des chutes par ancienneté reste similaire mais le seuil de chutes atteint est beaucoup plus élevé, surtout pour les anciennetés les plus récentes. Cependant, les contrats jeunes ont un S/P supérieurs aux anciens contrats. Cette hausse des résiliations des jeunes anciennetés pourrait donc avoir un impact positif sur le S/P.

4.3 Lissage uni-dimensionnel des taux bruts

Dans le but de corriger les irrégularités des taux bruts, un lissage de ces taux par Whittaker-Henderson a été réalisé. Les taux bruts étant assez « lisses » et afin de garder des taux lissés les plus proches possibles de ceux estimés par Kaplan-Meier, le critère de fidélité est maximisé et le critère de régularité est minimisé.

4.3.1 Méthode de Whittaker-Henderson en dimension 1 :

4.3.1.1 Théorie mathématique

La méthode de lissage de Whittaker-Henderson est basé sur deux critères :

- un critère de fidélité entre la courbe de référence obtenue par Kaplan-Meier et la courbe lissée
- un critère de régularité qui correspond au lissage de la courbe de référence

Critère de fidélité :

Le critère de fidélité mesure la distance quadratique entre la mortalité lissée \tilde{q}_x et la mortalité estimée \tilde{Q}_x . Chaque distance est pondérée par le poids W_x :

$$F = \sum_{x=1}^{\omega} W_x (\tilde{q}_x - \tilde{Q}_x)^2,$$

ω est l'âge de vie maximal.

La distance représente l'écart aux valeurs estimées. Les poids sont ajustables selon différentes modalités.

Critère de régularité :

Le critère de régularité permet d'évaluer la régularité de la courbe lissée :

$$K = \sum_{x=1}^{n-z} (\Delta^z \tilde{q}_x)^2$$

Où z est un premier paramètre du modèle, en entier positif, et où $\Delta^z \tilde{q}_x := \sum_{k=0}^z C_z^k (-1)^{z-k} \tilde{q}_{x+k}$.

Plus la forme de la courbe sera lissée, plus la valeur de K diminuera.

La méthode de lissage de Whittaker-Henderson consiste à trouver les valeurs lissées \tilde{q}_x qui minimisent M :

$$M := F + hK,$$

où h est un second paramètre du modèle.

On définit alors les matrices suivantes :

$$Q = \begin{bmatrix} \tilde{Q}_1 \\ \tilde{Q}_2 \\ \dots \\ \tilde{Q}_n \end{bmatrix}; \quad V = \begin{bmatrix} \tilde{q}_1 \\ \tilde{q}_2 \\ \dots \\ \tilde{q}_n \end{bmatrix}; \quad W = \begin{pmatrix} w_1 & 0 & \dots & \dots & 0 \\ 0 & w_2 & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & w_{n-1} & 0 \\ 0 & \dots & \dots & 0 & w_n \end{pmatrix}$$

Où W est la matrice diagonale correspondant aux poids donnés à chaque valeur estimée.

Voici la notation de F en notation matricielle :

$$F = (V - Q)^T \times W \times (V - Q)$$

Pour le critère de régularité, on définit :

$$\Delta^z V = \begin{pmatrix} \Delta^z \tilde{q}_1 \\ \Delta^z \tilde{q}_2 \\ \dots \\ \dots \\ \Delta^z \tilde{q}_{n-z} \end{pmatrix}$$

La notation matricielle de R est donc :

$$R = (\Delta^z V)^T \cdot (\Delta^z V).$$

On introduit alors une matrice K_z de taille $(n - z, n)$, contenant les coefficients binomiaux d'ordre z dont le signe alterne. Pour z pair, le signe est positif.

Par exemple pour $z = 2$ et $p = 4$, on aura :

$$K_2 = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{pmatrix}$$

On a ainsi : $\Delta^z V = K_z V$.

Ce qui nous permet d'écrire le critère M sous la forme :

$$\begin{aligned} M &= F + hR \\ &= (V - Q)^T \cdot W \cdot (V - Q) + h(\Delta^z V)^T \cdot (\Delta^z V) \\ &= (V - Q)^T \cdot W \cdot (V - Q) + hV^T \cdot K_z^T \cdot K_z \cdot V. \end{aligned}$$

Après développement puis regroupement des termes, on obtient :

$$M = V^T \cdot W \cdot V - 2V^T \cdot W \cdot Q + Q^T \cdot W \cdot Q + hV^T \cdot K_z^T \cdot K_z \cdot V.$$

Les \tilde{q}_x lissés sont alors obtenus en annulant la dérivée partielle de M par rapport à chacun des \tilde{q}_x .

On dérive alors M vectoriellement :

$$\begin{aligned} \frac{\Delta M}{\Delta V} &= 2W \cdot V - 2W \cdot Q + 2hK_z^T \cdot K_z \cdot V = 0 \\ \Leftrightarrow W \cdot V + 2hK_z^T \cdot K_z \cdot V &= 2W \cdot Q. \end{aligned}$$

On pose $C = W + hK_z^T \cdot K_z$ et on a donc :

$$C \cdot V = W \cdot Q.$$

Enfin, si C est inversible, nous obtenons l'expression suivante des taux lissés :

$$V = c^{-1}.W.Q.$$

4.3.1.2 Choix des poids

Pour limiter les poids donnés aux points aberrants, une pondération par effectif à l'ancienneté x rapporté à l'effectif moyen pour toutes les anciennetés est utilisée :

$$w_x = \frac{n_x}{\bar{n}}$$

où w_x correspondent aux poids, n_x est l'effectif à l'ancienneté x et $\bar{n} = \frac{\sum_{x=x_{inf}}^{x_{sup}} n_x}{x_{sup} - x_{inf} + 1}$

4.3.2 Taux lissés obtenus

Comme dit précédemment, les taux bruts étant assez réguliers, on décide ne pas donner une grande valeur au paramètre de lissage.

Les taux obtenus ci-dessous sont ont donc un critère de lissage égal à 1 et un critère de fidélité égal à 3.

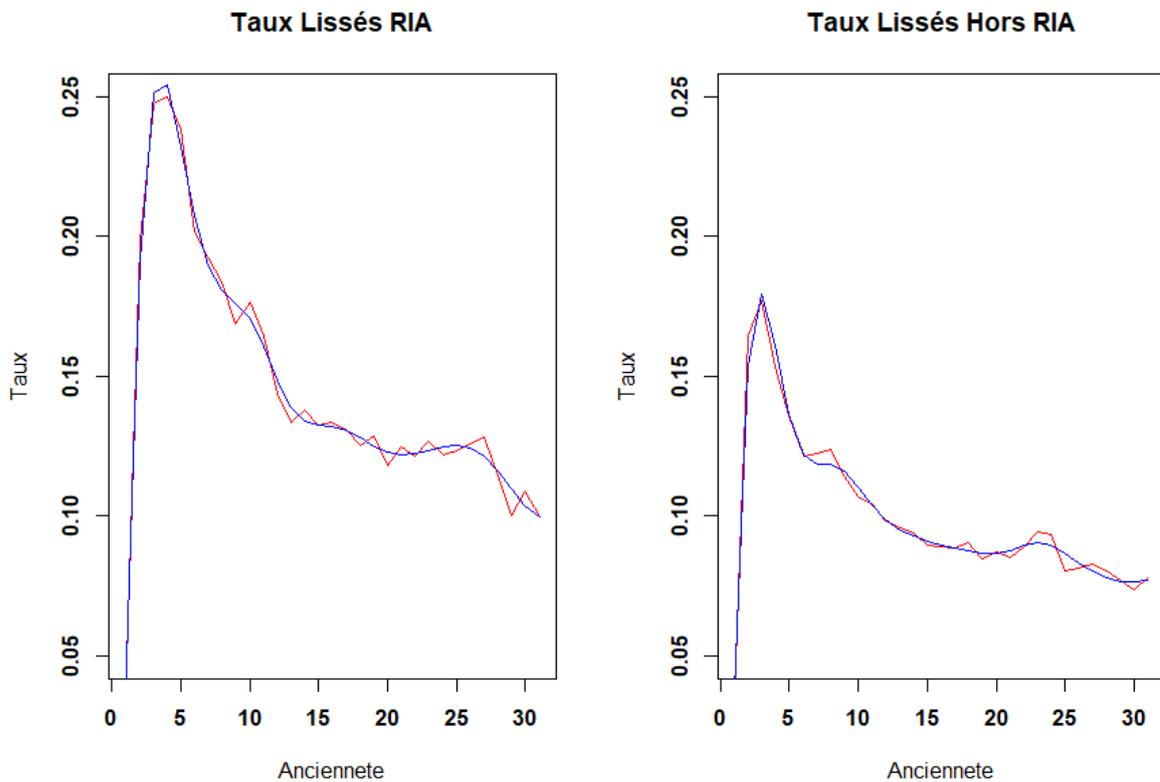


Figure 18 Taux lissés par Whittaker-Henderson

Les taux ainsi lissés ne sont pas strictement décroissants en fonction de l'ancienneté. On remarque de légères stagnations et même une légère croissance des taux pour les anciennetés les plus grande. Cela peut s'expliquer par le fait que dans les grandes anciennetés, les effectifs sont moindres donc les fluctuations sont plus fréquentes.

Cependant, contrairement à des taux de mortalités calculés en fonction de l'âge, ce n'est pas problématique. En effet, les taux de mortalités doivent être monotones au cours des âges car les taux de mortalités impactent la tarification. Ainsi, des taux croissants puis décroissants ne seraient pas explicables commercialement car cela signifierait qu'un assuré plus jeune paierait plus cher qu'un assuré plus âgé.

4.3.3 Validation des taux lissés

Le lissage des taux bruts modifie la structure de ces derniers. C'est pourquoi, il est nécessaire de s'assurer que ce lissage reste valide.

Un test non-paramétrique présenté par Planchet [Planchet, 2020] permet de valider cet ajustement : le test du changement de signe.

Il s'agit là d'un test statistique qui vient compter le nombre de changements de signe entre les différences des taux bruts avec les taux lissés.

Formellement, et dans le cas d'un lissage bi-dimensionnel, $\forall i \in \llbracket 1, x \rrbracket$ la grille Ancienneté des taux bruts et lissés, la différence est définie comme $\Delta_k = \Delta_j = q_j - \hat{q}_j$ où $(q_i)_{i \in \llbracket 1, x \rrbracket}$ sont les taux de sortie bruts et $(\hat{q}_i)_{i \in \llbracket 1, x \rrbracket}$ les taux de sortie lissés. Compter le nombre de changement de signe revient à compter de nombre de différence strictement positive ou négative. Pour les différences positives, le nombre de changement de signe s'écrit donc :

$$n = \sum_{k=1}^{t-1} \mathbb{1}_{\mathbb{R}_+^*} (-\Delta_k \Delta_{k+1})$$

Où il est attendu que la probabilité que Δ_k et Δ_{k+1} soient de signes différents est de $\frac{1}{2}$.

L'approximation gaussienne doit vérifier l'hypothèse statistique :

$$\frac{2n - (t - 1)}{\sqrt{t - 1}} \sim \mathcal{N}(0,1)$$

La validation du lissage à l'aide du test de changement de signes affichait une p-valeur de 0,824.

Au seuil de 5%, le lissage est donc conforme.

Au vu du choix fait a priori de privilégier la fidélité à la régularité, ce résultat est cohérent.

5 Modélisation des Affaires Nouvelles

5.1 Hypothèse considérée dans la modélisation des affaires nouvelles

Une fois les résiliations projetées, il est primordial de considérer les nouveaux contrats afin d'avoir les effectifs au complet. Effectivement, ne considérer que les chutes sous estimerait le nombre d'assurés couverts et donc le chiffre d'affaires.

Pour cela, la méthode de projection des affaires nouvelles doit se faire de manière cohérente avec les chutes. Ainsi, l'idée à priori est que les chutes et les affaires nouvelles sont corrélées. Précisément, dans un environnement de fortes résiliations, il est logique de se dire que les affaires nouvelles vont aussi augmenter. Les assurés ayant résilié leur contrat dans une entreprise vont en partie chercher un nouvel assureur.

Pour vérifier cette hypothèse, un test de corrélation est effectué. L'historique des résiliations et des affaires nouvelles de 2015 à 2021 a été utilisé. Evidemment, une si faible volumétrie de données ne donnera pas un résultat très pertinent mais cela permet de se faire une première idée. Les coefficients de corrélation obtenus sont les suivants :

Pearson	74%
Sperman	86%
Kendall	72%

La corrélation entre les chutes et les affaires nouvelles semble bien réelle.

Nous verrons par la suite que cette hypothèse sera d'autant plus vérifiée.

5.2 Séries temporelles

Nous disposons d'un historique mensuel des affaires nouvelles de 2013 à 2021. Le but est de déterminer un modèle de prévision adéquat afin d'obtenir des prédictions fiables.

5.2.1 Généralités sur les séries temporelles

Une série temporelle est une suite réelle finie $(X_t)_{1 \leq t \leq n}$, où t représente le temps. Elle se décompose principalement en trois éléments :

- La tendance (ft) d'une série représente son évolution générale à long terme. Elle traduit son comportement « moyen ».
- La saisonnalité (st) correspond à un phénomène périodique, qui se répète dans le temps.
- Les résidus ou bruit (st) correspondent à des fluctuations irrégulières, de nature aléatoire.

Dans le but d'identifier ces composantes, nous allons utiliser la méthode des moyennes mobiles afin de décomposer la série des affaires nouvelles.

5.2.2 Moyenne mobile

On appelle moyenne mobile, une transformation de X_t s'écrivant comme combinaison linéaire finie des valeurs de la série correspondant à des dates proches de t . La série transformée s'écrit :

$$M_{m_1+m_2+1}X_t = \sum_{i=m_1}^{m_2} \theta_i X_{t+i}$$

où $\theta_{-m_1}, \dots, \theta_{m_2}$ sont des réels et m_1 et m_2 des entiers. L'ordre de la moyenne mobile est $m_1 + m_2 + 1$.

Une moyenne mobile est centrée si $m_1 = m_2 = m$ et est symétrique si et seulement si $\theta_{-i} = \theta_i$
Une moyenne mobile arithmétique est une moyenne mobile centrée d'ordre impaire et telle que $\theta_i = \frac{1}{2m+1}$

5.2.3 Décomposition de la série temporelle

Voici la représentation de la série temporelles issue de notre historique des affaires nouvelles depuis 2013 :

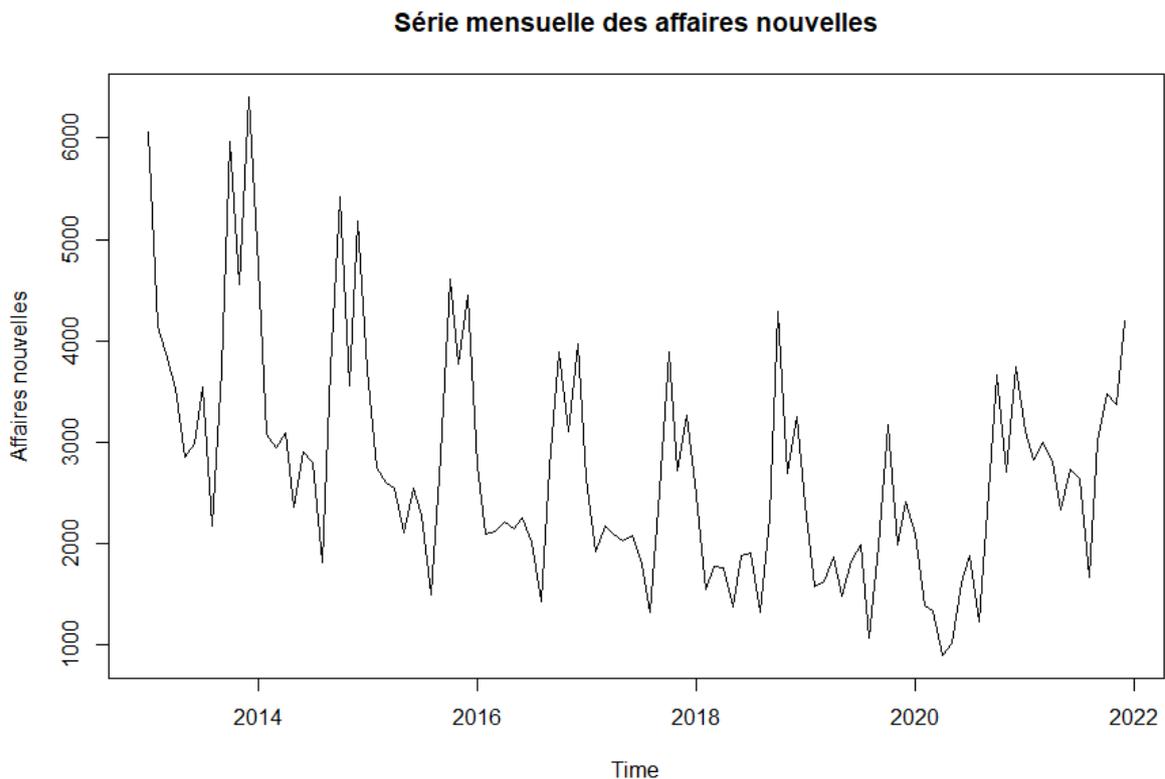


Figure 19 Série temporelle des affaires nouvelles

Au vu de la série, une décomposition additive semble adaptée. Nous la considérons de la forme suivante :

$$X_t = ft + st + st$$

Avec ft la saisonnalité et st les résidus.

Une première analyse peut être réalisée : on voit clairement que le nombre d'affaires nouvelles a augmenté depuis 2021, date de la réforme RIA et donc de la hausse des chutes. Affichons la décomposition de la série afin d'avoir plus d'éléments d'analyse :

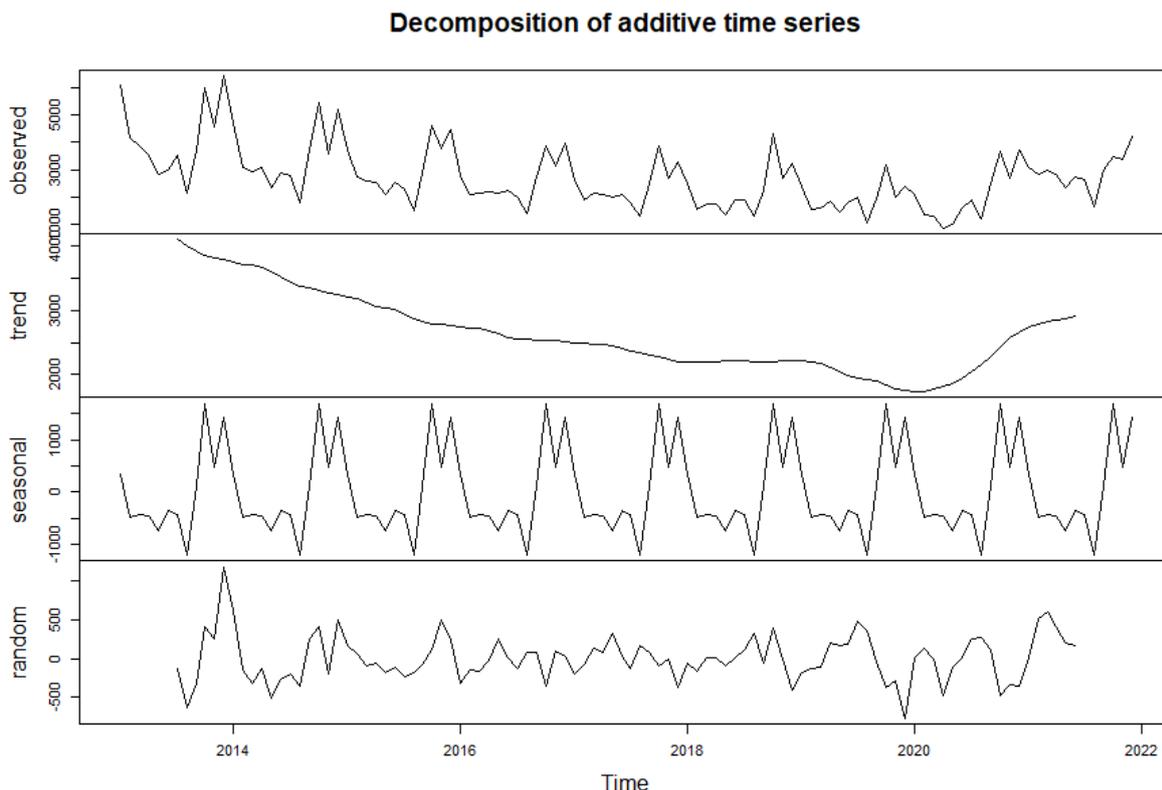


Figure 20 Décomposition de la série des affaires nouvelles

La tendance décrite juste avant est encore plus visible : entre 2013 et fin 2020, la tendance était linéairement décroissante. Depuis 2021, date d'impact de la réforme RIA et donc de hausse des chutes, la tendance a complètement changé et est strictement croissante. Cette analyse montre une fois de plus le lien entre les chutes et les affaires nouvelles. Cette corrélation doit donc être prise en compte dans la modélisation. Comme pour les chutes, on optera pour une modélisation différente avec et sans RIA :

- Scénario Hors RIA : modélisation des affaires nouvelles sans l'année 2021
- Scénario RIA : modélisation des affaires nouvelles avec l'année 2021

Cependant, dans le cadre du scénario RIA, l'application directe des séries temporelles sera inadaptée. En effet, utiliser une seule année d'historique n'est clairement pas suffisant. On proposera donc plusieurs alternatives.

5.3 Modélisation dans le cadre du scénario Hors RIA

Avant toute chose, Idéalement, nous voulons gérer des séries qui sont stationnaires (c'est-à-dire qui ont une moyenne et une variance constantes dans le temps). Pour vérifier cette hypothèse, un test de Dickey-Fuller a été réalisé. La p valeur obtenue étant de 0,01, on rejette l'hypothèse H_0 : la série n'est pas stationnaire, la série est donc stationnaire.

```
> adf.test(dfts, alternative="stationary")  
  
      Augmented Dickey-Fuller Test  
  
data: dfts  
Dickey-Fuller = -5.7425, Lag order = 4, p-value = 0.01  
alternative hypothesis: stationary
```

5.3.1 Lissage de Holt Winters

Les méthodes de lissage exponentiel sont des techniques empiriques de prévision qui accordent une certaine importance aux observations passées d'une série temporelle. Parmi ces méthodes, on retrouve le lissage de Holt Winters, et particulièrement le modèle additif de Holt Winters.

Ce modèle est une extension du lissage exponentiel de Holt qui prend en compte la saisonnalité. Cette méthode génère des valeurs lissées de façon exponentielle pour le niveau, la tendance et l'ajustement saisonnier de la prévision. Effectivement, d'après la décomposition de notre série, on observe une saisonnalité de 12 mois, il faut donc prendre en compte ce phénomène dans la modélisation.

5.3.1.1 Théorie mathématique

Nous supposons qu'au voisinage de T , la série peut être approchée par $X_t = (a(t - T) + b) + S_t + \epsilon_t$ avec $t = 1, \dots, T$

S_t un facteur de saisonnalité prenant s valeurs différentes : S_1, \dots, S_s

$\epsilon_t \sim BB(0, \sigma^2)$

La méthode de Holt Winters propose pour l'estimation de a, b et S_t les formules de mise à jour suivantes :

$$\begin{cases} \hat{a}_T = (1 - \beta)\hat{a}_{T-1} + \beta(\hat{b}_T - \hat{b}_{T-1}) \\ \hat{b}_T = \alpha(X_T - \hat{S}_{T-p}) + (1 - \alpha)(\hat{b}_{T-1} + \hat{a}_{T-1}) \\ \hat{S}_T = \gamma(X_T - \hat{b}_T) + (1 - \gamma)\hat{S}_{T-p} \end{cases}$$

où α, β et $\gamma \in]0, 1[$

On en déduit donc une prévision à l'horizon h :

$$\hat{X}_T(h) = \begin{cases} \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-s}, \text{ pour } 1 \leq h \leq s \\ \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-2s}, \text{ pour } 1 \leq h \leq 2s \\ \dots \\ \hat{a}_1(T) + h\hat{a}_2(T) + \hat{S}_{T+h-ns}, \text{ pour } 1 \leq h \leq ns, \text{ avec } n \rightarrow +\infty \end{cases}$$

5.3.1.2 Séparation série d'entraînement – série de test

Afin de vérifier une prévision temporelle, il est important de séparer les données en deux séries : une série d'entraînement et une série de test :



Ayant des données disponibles de 2013 à 2020, on décide de séparer les données de sorte à avoir environ 75% des données pour la série d'entraînement et 25% pour la série de test. Ainsi, les données de 2013 à 2018 servent à l'entraînement et les données de 2019 à 2020 servent au test.

5.3.1.3 Application du modèle

A l'aide du package « forecast », nous réalisons cette modélisation sous R. Les coefficients de lissage Holt Winters estimés sont présentés ci-dessous :

```
ETS(M,Ad,M)
```

```
Call:
```

```
ets(y = dfts_train)
```

```
Smoothing parameters:
```

```
alpha = 0.114  
beta = 0.0041  
gamma = 3e-04  
phi = 0.9759
```

```
Initial states:
```

```
l = 4717.3958  
b = -70.4637  
s = 1.5855 1.2108 1.6346 1.0326 0.5446 0.8139  
0.8521 0.7282 0.8255 0.8324 0.8017 1.1383
```

```
sigma: 0.09
```

Le modèle retenu est un ETS(M,Ad,M), définissant dans l'ordre le type d'erreur, de tendance et de saison. Nous avons donc une erreur de type M (multiplicative), une tendance de type A (additive) et une saison de type M (multiplicative). La constante alpha étant proche de 0, le modèle n'accorde pas énormément d'importance sur les événements les plus récents. Le coefficient beta est positif mais presque nul, la tendance sera donc pratiquement constante. La constante de saisonnalité gamma est très faible, la saisonnalité n'aura donc pas beaucoup d'impact sur notre prévision.

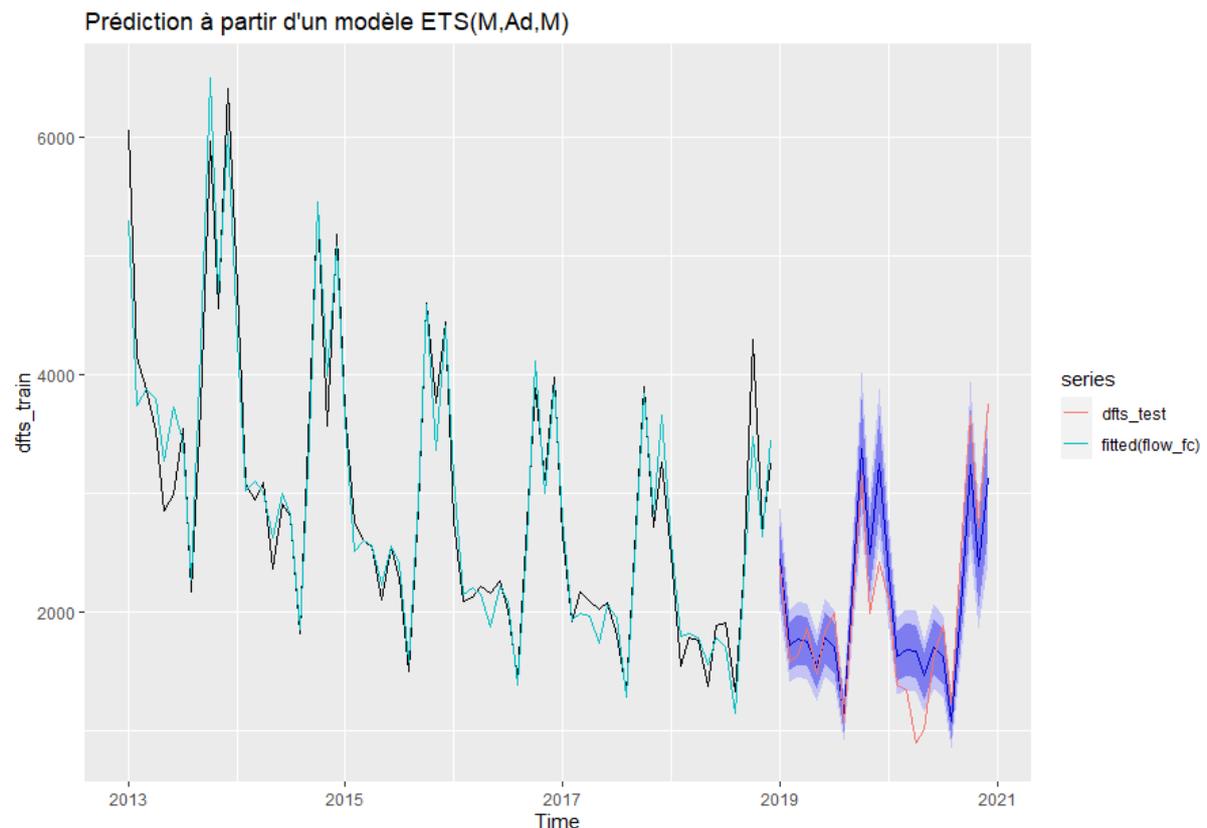


Figure 21 Prédiction des affaires nouvelles à partir d'un modèle ETS

En vert, les valeurs obtenues par le lissage sont plutôt bien ajustées. On ne remarque pas trop d'écart entre la série réelle et ajustée. Concernant la prédiction en bleu, le constat est le même : la prédiction est plutôt satisfaisante.

En effet, pour l'année 2019 le nombre d'affaire nouvelles prédit est de 25 095 contre 23 395 réellement observé, soit une erreur de 7,2%. En revanche, concernant l'année 2020, on remarque que les écarts mensuels sont plus volatils mais se compensent au global : 24 000 affaires nouvelles prédites contre 24 101 réelles, soit une erreur de 0,4%. L'erreur moyenne est donc de 3,8%, ce qui est satisfaisant. La somme d'affaire nouvelle globale sur les 2 années est donc de 49 095. L'intervalle de confiance à 95%, représenté ici en bleu foncé donne comme intervalle [39 342 ;58 848], soit un intervalle plutôt bien ajusté.

5.3.1.4 Analyse des résidus

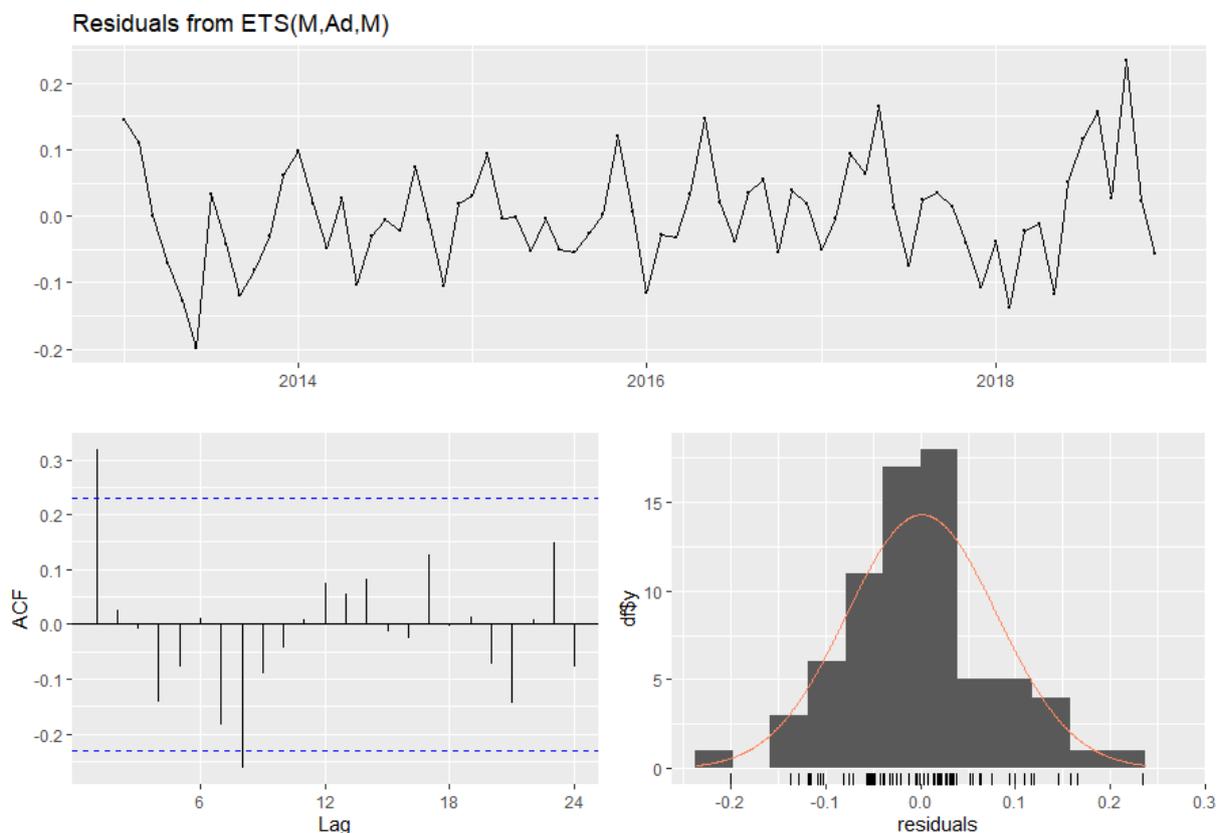


Figure 22 Résidus modèle ETS

Les résidus semblent normaux et centrés en 0. Cette hypothèse est vérifiée à l'aide d'un test de Student et de Shapiro-Wilk.

- Test de Student : H_0 : La moyenne de l'échantillon est nulle
 H_1 : La moyenne de l'échantillon n'est pas nulle
- Test de Shapiro-Wilk : H_0 : L'échantillon suit une loi normale
 H_1 : L'échantillon ne suit pas une loi normale

La p-valeur du test de Student étant de 0,88 et celle du test de Shapiro-Wilk de 0,52, on confirme que les résidus suivent une loi normale de moyenne nulle. Cependant, ils ne semblent pas former un bruit blanc. Afin de vérifier cette hypothèse, nous réalisons un test de Ljung-Box, dont les hypothèses sont :

H_0 : Les données sont distribuées indépendamment

H_1 : Les données ne sont pas distribuées indépendamment

Ljung-Box test

```
data: Residuals from ETS(M,Ad,M)
Q* = 22.771, df = 3, p-value = 4.508e-05
Model df: 17. Total lags used: 20
```

La p-valeur étant en dessous du seuil de 0.05, on rejette l'hypothèse H_0 , les résidus ne sont donc pas des bruits blancs.

5.3.2 Les processus ARIMA et SARIMA

5.3.2.1 Théorie mathématique

Un processus $(X_t)_{t \in \mathbb{N}}$ est un processus ARIMA (AutoRegressive Integrated Moving Average) d'ordre (p, d, q) , noté $ARIMA(p, d, q)$, si :

$$\Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t$$

où :

- $\nabla^d = (I - B)^d$,
- $\Phi(B) = I - \varphi_1 B - \dots - \varphi_p B^p$ où $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$ et $\varphi_p \neq 0$,
- $\Theta(B) = I + \theta_1 B + \dots + \theta_q B^q$ où et $\theta_q \neq 0$.

Remarque :

- Les modèles ARIMA permettent de modéliser des séries temporelles présentant une tendance polynomiale.
- $(I - B)^d X_t$ est équivalent asymptotiquement à un processus ARMA(p, q).

On dit qu'un processus $(X_t)_{t \in \mathbb{N}}$ est un processus SARIMA (Seasonnal AutoRegressive Integrated Moving Average) d'ordre $(p, d, q)(P, D, Q)_s$, noté $SARIMA(p, d, q)(P, D, Q)_s$ si :

$$\Phi(B)\Phi'(B^s)\nabla^d\nabla_s^D X_t = \Theta(B)\Theta'(B^s)\varepsilon_t$$

où :

- $\nabla^d = (I - B)^d$
- $\nabla_s^D = (I - B^s)^D$,
- $\Phi(B) = I - \varphi_1 B - \dots - \varphi_p B^p$ où
 - $(\varphi_1, \dots, \varphi_p) \in \mathbb{R}^p$ et $\varphi_p \neq 0$,
- $\Phi'(B) = I - \varphi'_1 B - \dots - \varphi'_p B^p$ où
 - $(\varphi'_1, \dots, \varphi'_p) \in \mathbb{R}^p$ et $\varphi'_p \neq 0$,
- $\Theta(B) = I + \theta_1 B + \dots + \theta_q B^q$ où
 - $(\theta_1, \dots, \theta_q) \in \mathbb{R}^q$ et $\theta_q \neq 0$,
- $\Theta'(B) = I + \theta'_1 B + \dots + \theta'_q B^q$ où
 - $(\theta'_1, \dots, \theta'_q) \in \mathbb{R}^q$ et $\theta'_q \neq 0$.

Remarque :

- *Les modèles SARIMA permettent de modéliser des séries présentant une saisonnalité.*
- *Estimer un modèle SARIMA se ramène à l'estimation d'un modèle ARMA sur la série différenciée.*

5.3.2.2 Critères de validation du modèle

Afin de choisir le modèle le plus adapté, nous allons utiliser la fonction R `auto.arima` du package « `forecast` ».

La fonction `auto.arima` utilise une variante de l'algorithme de Hyndman-Khandakar. Elle va permettre de choisir le meilleur modèle selon les critères *AIC*, *AIC_C* et *BIC*.

Le critère d'information d'Akaike (AIC) est un estimateur de l'erreur de prédiction. L'AIC utilise le maximum de vraisemblance en pénalisant les modèles comportant trop de variables, qui « surapprennent » les données et ne parviennent pas à généraliser.

Sa formulation est la suivante :

$$AIC = -2\log(L) + 2k$$

où L est la vraisemblance maximisée et k le nombre de paramètres à estimer dans le modèle. Ce critère permet donc de pénaliser 2 fois la déviance du modèle ($-2\log(L)$).

Lorsque le nombre de paramètre est grand comparé au nombre d'observations n (si $\frac{n}{k} < 40$), il est recommandé d'utiliser l'AIC « corrigé » :

$$AIC_C = AIC + \frac{2k(k+1)}{n-k-1}$$

Enfin le critère BIC pénalise encore d'avantage le sur-paramétrage que l'AIC :

$$BIC = -2\log(L) + k\log(N)$$

5.3.2.3 Application

```
> auto.arima(dfts_train, trace=TRUE)

ARIMA(2,1,2)(1,1,1)[12]      : Inf
ARIMA(0,1,0)(0,1,0)[12]    : 862.9665
ARIMA(1,1,0)(1,1,0)[12]    : 863.5397
ARIMA(0,1,1)(0,1,1)[12]    : 858.204
ARIMA(0,1,1)(0,1,0)[12]    : 856.0402
ARIMA(0,1,1)(1,1,0)[12]    : 858.1899
ARIMA(0,1,1)(1,1,1)[12]    : 860.11
ARIMA(1,1,1)(0,1,0)[12]    : 852.6404
ARIMA(1,1,1)(1,1,0)[12]    : 854.5708
ARIMA(1,1,1)(0,1,1)[12]    : 854.6148
ARIMA(1,1,1)(1,1,1)[12]    : 856.8821
ARIMA(1,1,0)(0,1,0)[12]    : 861.333
ARIMA(2,1,1)(0,1,0)[12]    : 854.3411
ARIMA(1,1,2)(0,1,0)[12]    : 853.8774
ARIMA(0,1,2)(0,1,0)[12]    : 851.5896
ARIMA(0,1,2)(1,1,0)[12]    : 852.9725
ARIMA(0,1,2)(0,1,1)[12]    : 853.096
ARIMA(0,1,2)(1,1,1)[12]    : 855.2996
ARIMA(0,1,3)(0,1,0)[12]    : 853.881
ARIMA(1,1,3)(0,1,0)[12]    : Inf

Best model: ARIMA(0,1,2)(0,1,0)[12]

Series: dfts_train
ARIMA(0,1,2)(0,1,0)[12]

Coefficients:
      ma1      ma2
-0.4776 -0.3531
s.e.    0.1299  0.1230

sigma^2 = 99058: log likelihood = -422.58
AIC=851.15  AICc=851.59  BIC=857.39
```

Le modèle retenu par la fonction est un $ARIMA(0,1,2)(0,1,0)_{12}$ donc un SARIMA, prenant en compte la saisonnalité sur 12 mois. Nous sommes dans le cas où $\frac{n}{k} < 40$, c'est-à-dire que le nombre de paramètres à estimer est grand comparé au nombre d'observations. Le critère AIC_C est alors utilisé. On voit bien que ce modèle minimise le AIC_C parmi tous les autres modèles initialement retenus par l'algorithme avec un AIC_C de 851.59 et un AIC de 851.15. On remarque donc que ce modèle possède un meilleur AIC que le lissage de Holt-Winters.

5.3.2.4 Analyse des résidus

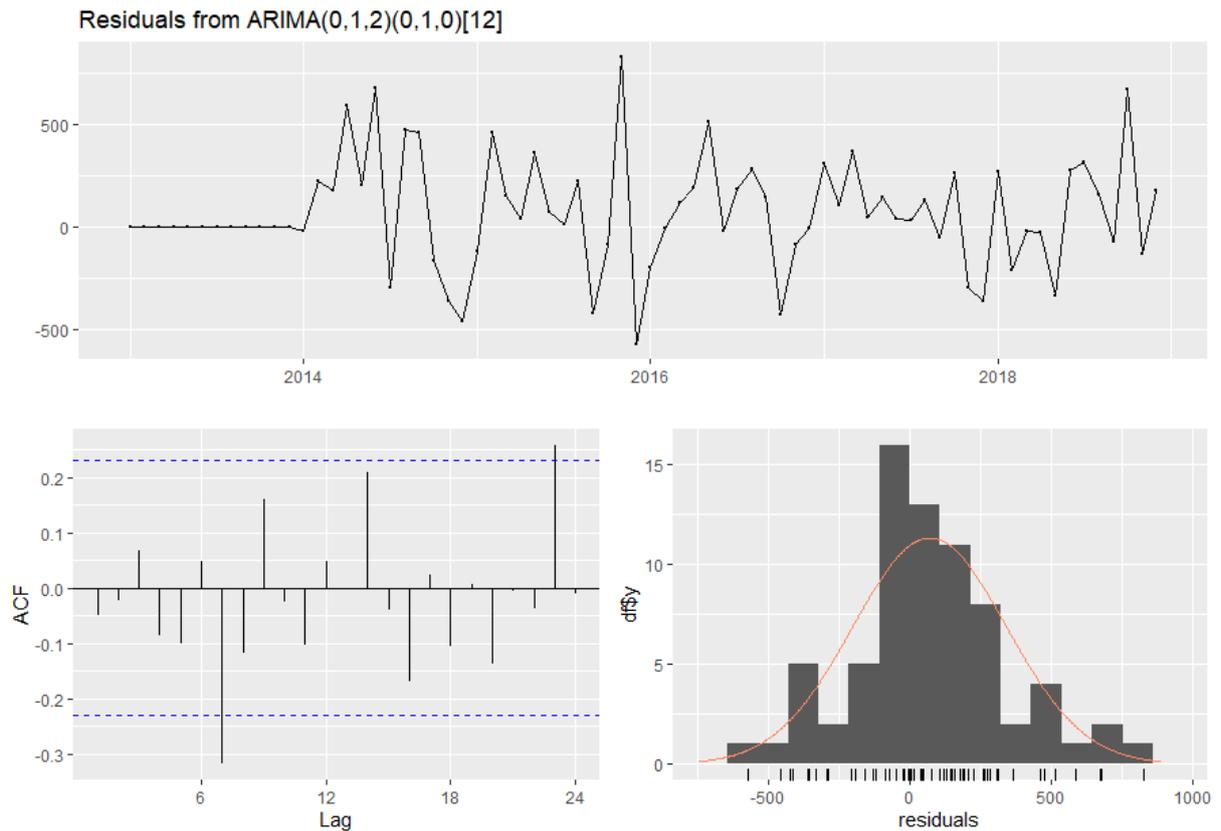


Figure 23 Résidus modèle SARIMA

Il est primordial de vérifier que les résidus sont de moyenne nulle et indépendants et identiquement distribués.

Le test de Student donne une p-valeur de 0,06 tandis que la p-valeur du test de Ljung-Box donne une p-valeur de 0,34. Par conséquent, malgré le fait que la p-valeur du test de Student ne soit pas significativement supérieure à 0,05, on en déduit que les hypothèses sur les résidus sont vérifiées. Nos résidus sont donc un bruit blanc, nous pouvons accepter notre modélisation.

5.3.2.5 Modélisation

De la même manière que pour le lissage de Holt-Winters, nous avons séparé notre série en deux : une série test et une série d'entraînement.

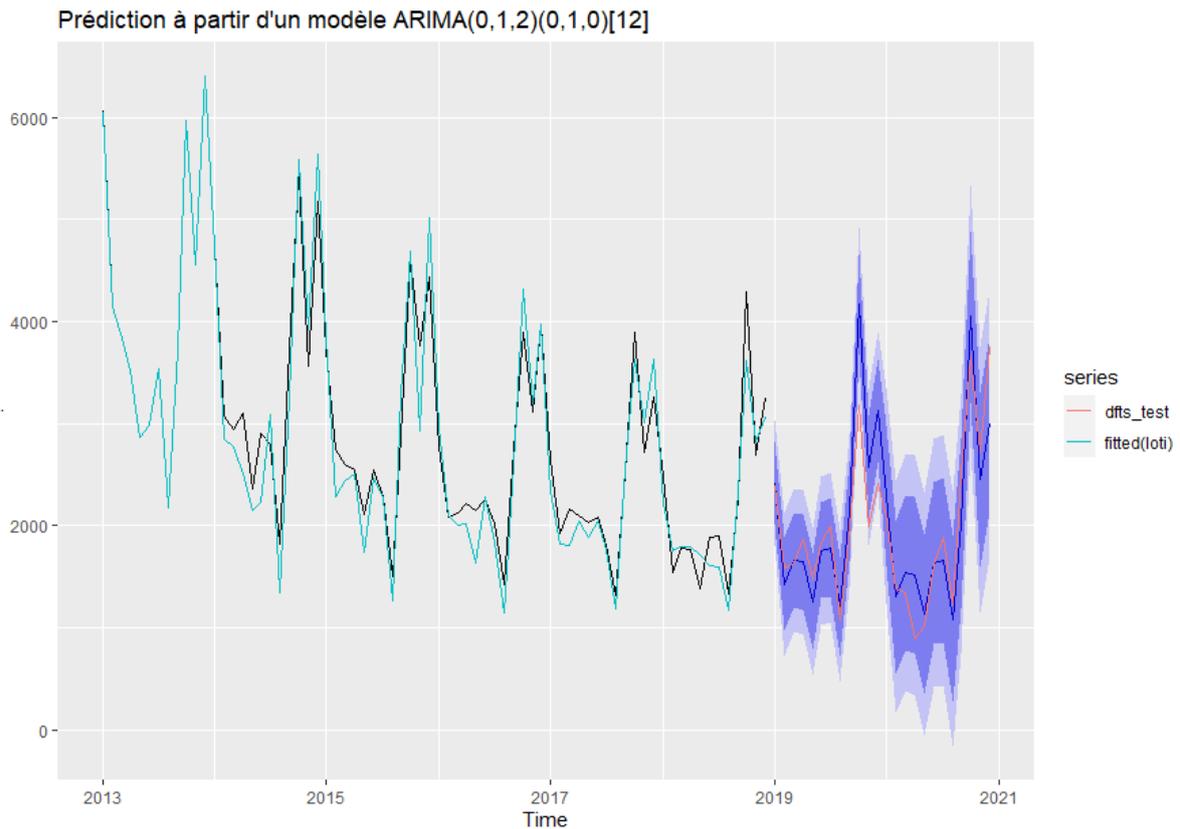


Figure 24 Prédiction à partir du modèle SARIMA

En vert, les valeurs obtenues par le lissage sont plutôt bien ajustées. On ne remarque pas trop d'écart entre la série réelle et ajustée.

Concrètement les prédictions en bleu, on remarque que les écarts mensuels sont plus volatils mais se compensent au global : pour l'année 2019 le nombre d'affaires nouvelles prédit est de 25 165 contre 23 395 réellement observé, soit une erreur de 7,6%. Concernant l'année 2020, 23 690 affaires nouvelles sont prédites contre 24 101 réelles, soit une erreur de 1,7%. L'erreur moyenne est donc de 4,65%, ce qui est satisfaisant, mais légèrement supérieur au lissage de Holt-Winters. La somme d'affaires nouvelles globale sur les 2 années est donc de 49 095. L'intervalle de confiance à 95% est aussi plus large que le modèle précédent : $IC = [25\ 592 ; 72\ 119]$.

Les prédictions du modèle SARIMA sont donc légèrement moins précises que celle de Holt-Winters, cependant, à cause des hypothèses des résidus, on décide de garder le modèle SARIMA.

Les prévisions d'affaires nouvelles pour les 3 prochaines années sont obtenues ci-dessous :

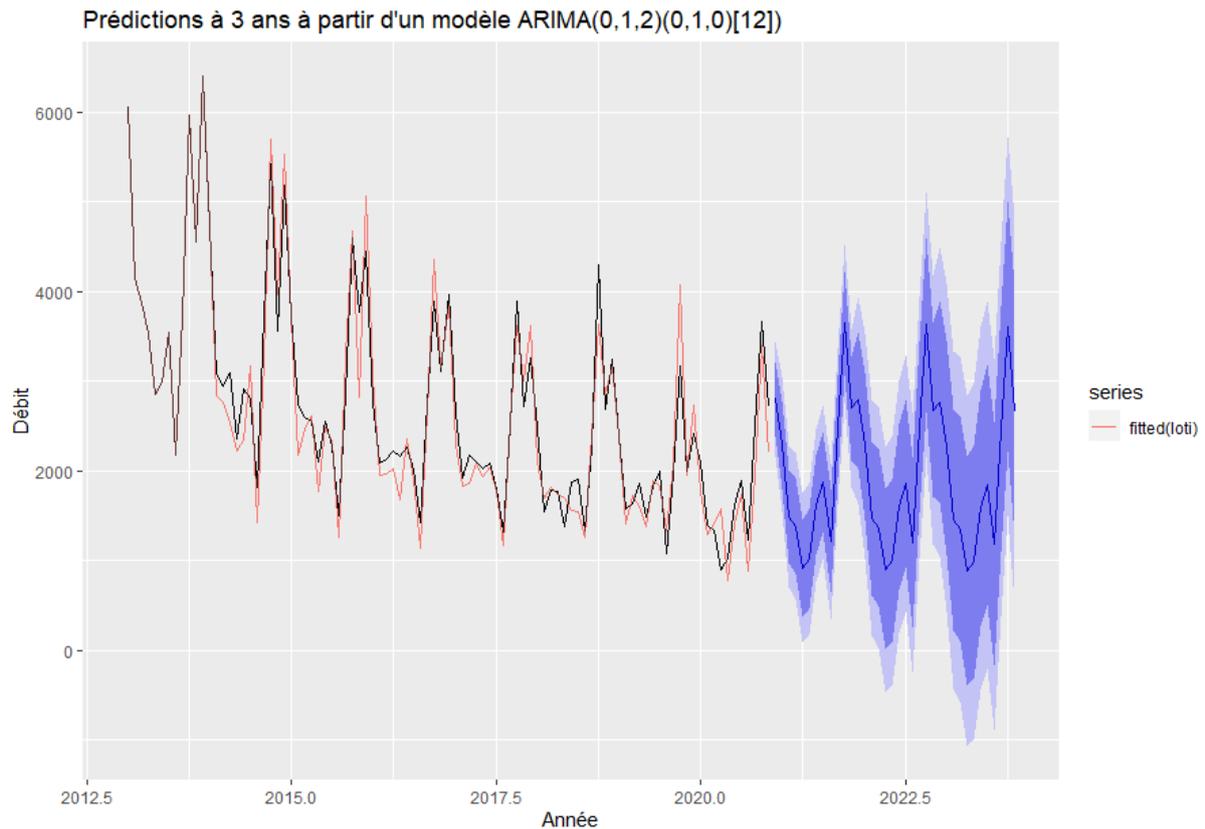


Figure 25 Prédictions finales à partir du modèle SARIMA optimal

Les prédictions à 3 ans du modèle sont les suivantes : 23 547 affaires nouvelles pour 2022, 23 052 pour 2023 et 22 614 pour l'année 2024.

5.4 Modélisation dans le cadre du scénario RIA

Comme dit précédemment, la modélisation des affaires nouvelles de ce scénario est plus délicate. En effet, utiliser les séries temporelles avec une seule année d'historique n'a pas vraiment d'intérêt et est beaucoup moins fiable. Il existe cependant plusieurs alternatives afin d'obtenir des résultats les plus cohérents possible.

5.4.1 La méthode SES (Simple Exponential Smoothing)

Cette méthode applique aux valeurs précédentes des poids exponentiellement décroissants selon leur ancienneté. La prévision par SES sera alors une moyenne pondérée des dernières observations, en appliquant plus de poids aux observations plus récentes.

Voici la formule mathématique de cette méthode :

$$\hat{y}_{t+h|t} = \alpha y_t + \alpha(1 - \alpha)y_{t-1} + \alpha(1 - \alpha)^2 y_{t-2} + \dots$$

avec $\hat{y}_{t+h|t}$ la prévision de y au temps $t + h|t$.

Le paramètre $\alpha \in]0 ; 1[$ et décrit la distribution des poids. Le paramètre ne peut pas prendre les valeurs 0 et 1 car si $\alpha = 0, \hat{y}_{t+h|t} = 0$ et si $\alpha = 1, \hat{y}_{t+h|t} = y_t$.

Plus la valeur de α est élevée, plus les événements récents auront un poids élevé dans la prévision. Lorsque les pas de temps antérieurs tendent vers l'infini, la somme de tous les poids α tend vers 1.

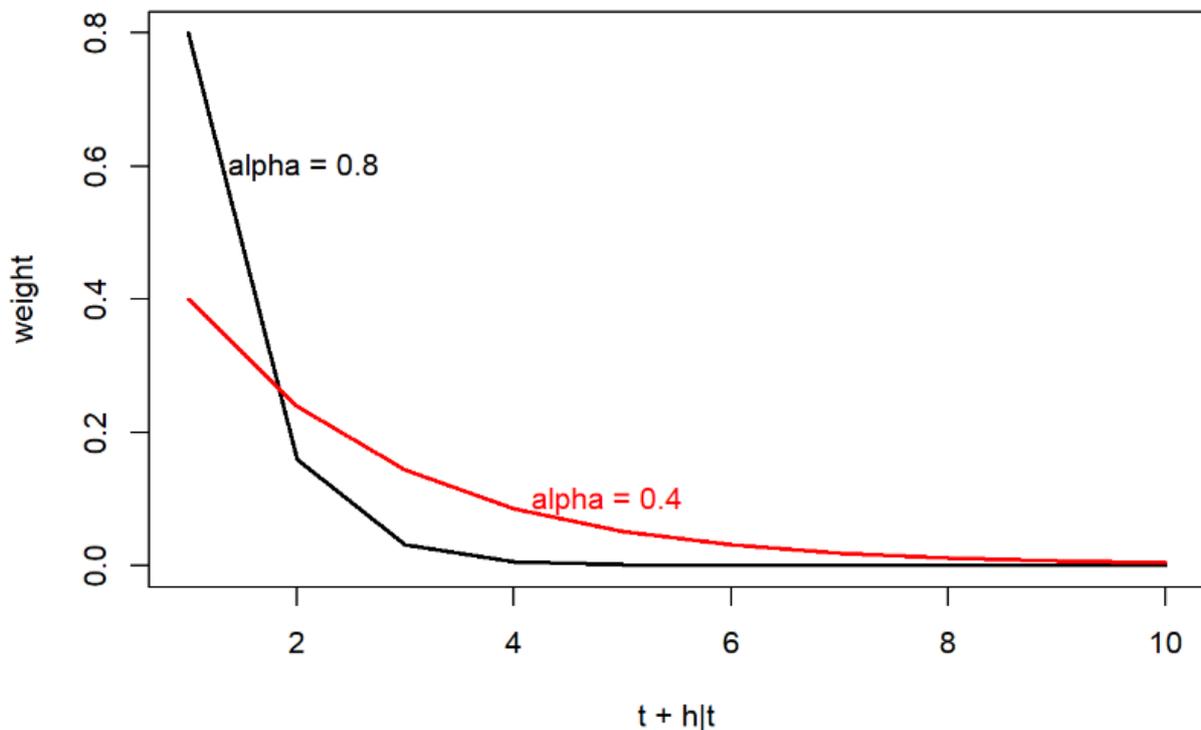


Figure 26 Modèle SES - impact du choix de la variable alpha

Avantages de cette méthode : permet de garder notre profondeur d'historique en accordant d'avantage de poids aux événements récents, donc aux affaires nouvelles de 2021.

Inconvénients de cette méthode : en ayant la liberté de choisir la valeur du paramètre α , la prédiction ne dépend plus de la fiabilité du modèle mais plutôt de l'importance que l'on veut accorder aux événements récents. En ayant choisi un paramètre $\alpha = 0,86$, le montant d'affaires nouvelles prédit vaut le montant des affaires nouvelles de 2021. Cette valeur est donc une valeur pivot : plus la valeur dépasse 0,86, plus le montant d'affaires nouvelles va être grand et inversement lorsque la valeur du paramètre diminue par rapport à 0,86. Cette méthode est donc beaucoup trop paramétrique car on peut ajuster nous même les valeurs prédites.

5.4.2 Utilisation de l'intervalle de confiance à 95% de la méthode Hors RIA

En observant la borne supérieure de l'intervalle de confiance à 95% de la méthode Hors RIA, on trouve 33 335 affaires nouvelles prédites pour l'année suivante. Précisément, le montant d'affaires nouvelles réel étant de 35 179 en 2021, l'erreur de prédiction est donc de 5,5%. Cette borne supérieure de l'intervalle de confiance à 95% peut alors s'apparenter à un scénario aggravé des affaires nouvelles, précisément le scénario que l'on souhaite modéliser.

Avantages de cette méthode : moins paramétrique que la première méthode et bonne précision.

Inconvénients de cette méthode : la précision de la prédiction de l'intervalle de confiance est peut-être une coïncidence. Vu que la prédiction se base sur un historique et une tendance différente de celle de 2021, les montants des années suivantes seront possiblement biaisés. De plus, l'intervalle de confiance s'élargit avec le temps donc les biais risquent de s'aggraver.

5.4.3 Cohérence avec la projection des taux de chutes

L'hypothèse principale de la modélisation des effectifs est que les chutes et les affaires nouvelles sont liées. L'idée d'utiliser cette hypothèse lors de prévision des affaires nouvelles dans le scénario Hors RIA est donc possiblement correcte.

Lors de la projection du portefeuille, on suppose que les taux de chutes par génération RIA resteront constants au cours des 3 prochaines années. Nous pouvons donc, par soucis de cohérence et de prudence, faire de même pour les affaires nouvelles : projeter 35 179 affaires nouvelles pour les années 2022, 2023 et 2024.

Avantages de cette méthode : facilité de mise en place

Inconvénients de cette méthode : incohérence avec la modélisation des affaires nouvelles Hors RIA : effectivement, on projette des montants différents pour les 3 prochaines années dans le scénario Hors RIA.

C'est cette dernière alternative, plus simple et plus prudente qui sera ainsi considérée.

6 Modélisation des prestations

6.1 Statistiques descriptives

6.1.1 Historique de données utilisé

Afin de prendre en compte la réforme 100% santé, l'historique des prestations utilisé est uniquement celui de l'année 2021. En effet, les postes audio et dentaire n'ayant été impactés que progressivement par la réforme, tant dit que le poste optique a été impacté brutalement en 2020, prendre un historique de consommation de plusieurs années biaiserait nos prédictions. Les coûts moyens en dentaire, par exemple, sont sensiblement différents en 2019 qu'en 2021, une fois la réforme déployée intégralement.

De plus, l'année 2020 étant marquée par le Covid-19 et ainsi les nombreux confinements, utiliser cette année dans notre historique de données viendrait également biaiser nos prédictions. Effectivement, le montant des prestations cette année est nettement plus faible à cause des confinements.

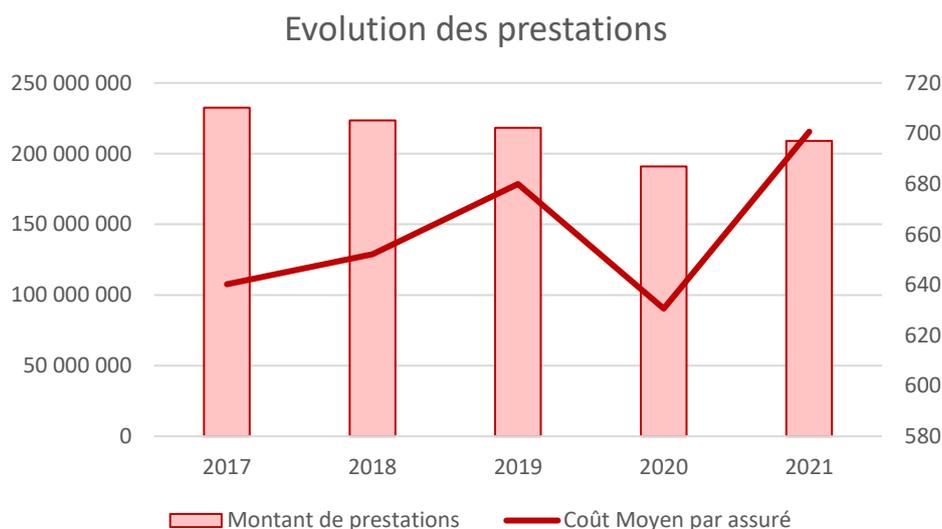


Figure 27 Evolution des prestations et du CM/assuré depuis 2017

Les coûts moyens avant la réforme étaient stables (2017 et 2018) et ont subitement augmenté en 2019 (début réforme 100% santé) avant de chuter en 2020 (confinements) et enfin fortement réaugmenté en 2021 (fin réforme 100% santé).

6.1.2 Considération de l'exposition

L'exposition, définie comme le taux de présence de chaque assuré au cours de l'année est une variable primordiale dans la prédiction des prestations. En effet, un individu présent la moitié de l'année aura probablement un montant de prestations plus faible qu'un assuré présent toute l'année.

Il convient donc de définir cette métrique de la manière la plus précise possible et d'ensuite retirer ce biais de nos données.

On définit l'exposition en 2021 de la façon suivante :

$$E_{2021} = \frac{\text{MIN}(\text{DAT_EFFT_CHUT}; 31/12/2021) - \text{MAX}(\text{DAT_EFFT_CT}; 01/01/2021)}{365}$$

- DAT_EFFT_CHUT représente la date d'effet de la résiliation du contrat
- DAT_EFFT_CT représente la date d'effet du contrat
- $E_{2021} \in]0 ; 100\%]$

Une fois l'exposition définie, il faut donc corriger ce biais de nos données. Pour ce faire, 2 approches peuvent être considérées :

- Ne garder que les assurés ayant une exposition de 100%
- Proratiser les prestations avec le taux d'exposition

La première approche ne peut être considérée que si la suppression des expositions différentes de 1 n'entraîne pas une trop grande perte de volumétrie de données. De ce fait, sur 298 477 assurés principaux, 225 268 ont une exposition de 100%. Ainsi, Supprimer les assurés principaux n'ayant pas 100% d'exposition supprimerait un peu moins de 25% de notre jeu de données. La volumétrie de données finale reste suffisante, cette approche peut être considérée.

La seconde approche présente un inconvénient important. En effet, voici une formalisation de cette approche :

Soit un individu X ayant un montant de prestations P_X et une exposition E_{2021}^X . Son montant de prestation corrigé est alors $P'_X = \frac{P_X}{E_{2021}^X}$.

Cependant, cette approche engendrerait trop de biais dans l'estimation des prestations manquantes, notamment pour les individus n'ayant pas ou très peu consommé et ayant une faible exposition. De même pour les individus ayant subi une grosse dépense de soins en début de période d'exposition, cette méthode supposerait alors que ces individus consommeront de la même manière le reste de la période et cela augmenterait alors le phénomène de valeurs aberrantes.

La première approche est donc considérée.

6.1.3 Statistiques descriptives des variables explicatives

La base de données utilisée pour la modélisation des prestations est issue de la jointure de plusieurs sources de données issues du système de gestion interne. En effet, Swiss Life dispose d'un outil nommé le « Datamart » regroupant des dizaines de tables de données elles-mêmes rassemblant différentes informations sur les assurés.

Plusieurs retraitements ont été préalablement réalisés afin de « nettoyer » la base de données.

Les informations ainsi utilisées pour notre étude sont les suivantes :

- Clé d'identification de l'assuré principal du contrat

- Clé d'identification des bénéficiaires du contrat permettant de créer la variable nombre de bénéficiaires (NB_BENEF)
- Date de naissance permettant de créer la variable âge
- Date d'effet du contrat permettant de créer la variable ancienneté
- Date de résiliation du contrat
- CSP
- Produit cible
- Sexe de l'assuré
- Montant de primes
- Montant de prestation

6.1.3.1 Age atteint lors de la période d'étude

La moyenne d'âge des assurés principaux est de 59 ans. L'Age minimal est de 18 ans et l'âge maximal est de 112 ans. Effectivement, pour souscrire un contrat d'assurance santé, il faut être majeur.

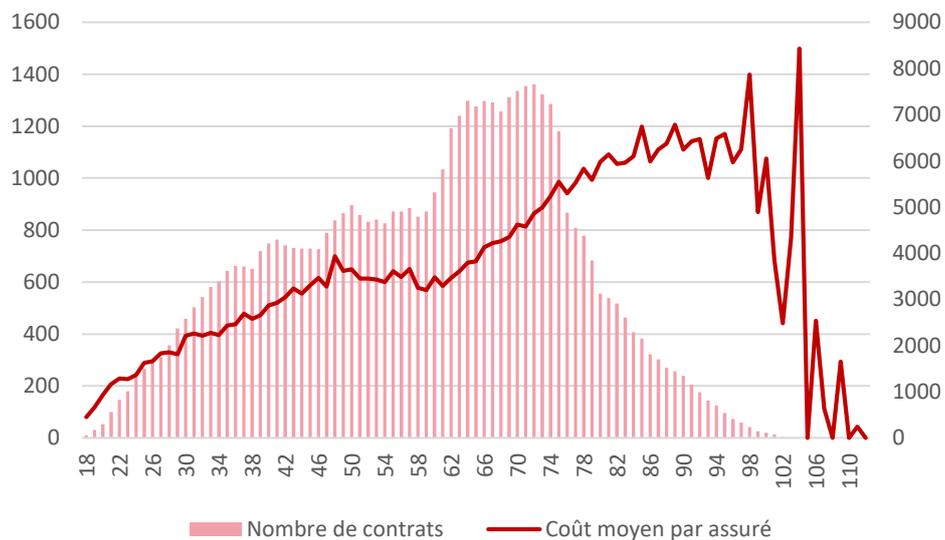


Figure 28 Evolution du CM/assuré et nombre de contrats en fonction de l'âge

On observe que le coût moyen par assuré augmente avec l'âge. Cependant, cela ne veut pas forcément dire que la rentabilité de ces contrats est impactée. En effet, la tarification doit permettre de compenser cette hausse de soins.

Evolution du S/P par âge

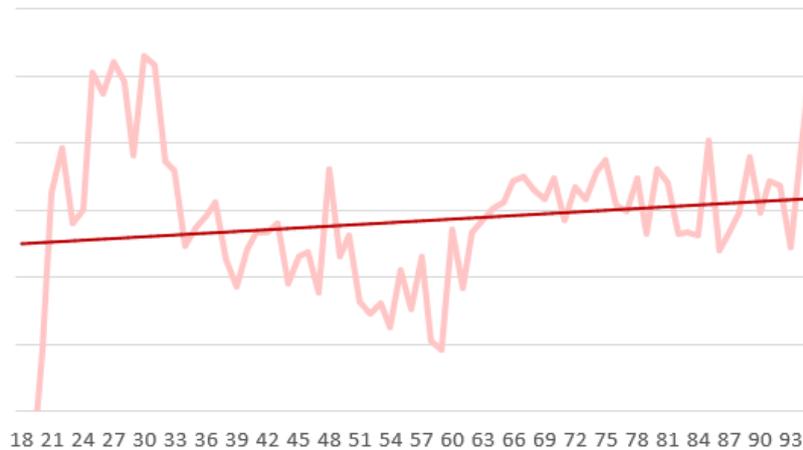


Figure 29 Evolution du S/P par âge

Remarque : par soucis de confidentialité, l'axe des S/P n'est pas représenté

La droite rouge représente la courbe de tendance linéaire.

La remarque précédente est ici clairement illustrée : les contrats âgés réalisant le plus de sinistres ne sont pas ceux où la rentabilité est la moins bonne. Au contraire, les contrats jeunes, qui à priori ne devraient pas beaucoup consommer et ont donc une prime plus faible ont la pire rentabilité. Cependant, au vu du peu de volumétrie de contrats dans les premiers âges, cette remarque est à relativiser. Le principe de mutualisation est ainsi clairement visible ici.

Une seconde information est visible ici : la rentabilité entre 50 et 60 ans est clairement meilleure que pour des âges supérieurs à 60 ans. La tarification a donc permis de réduire les écarts de consommation mais pas les annulés.

6.1.3.2 Répartition Homme Femme

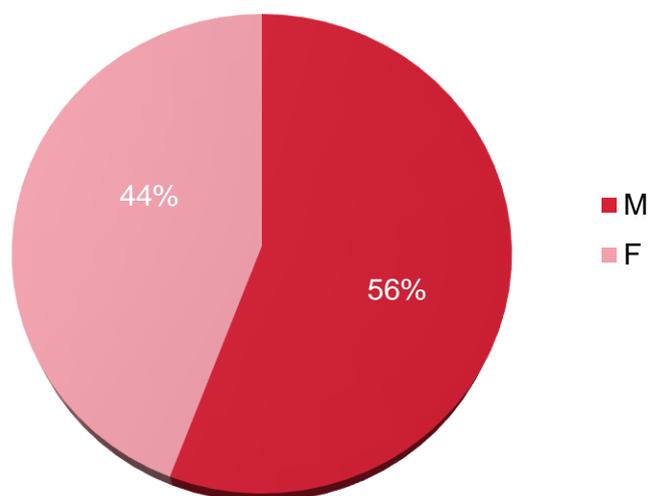


Figure 30 Répartition Homme/Femme

Le portefeuille est composé majoritairement d'homme

6.1.3.3 Répartition CSP

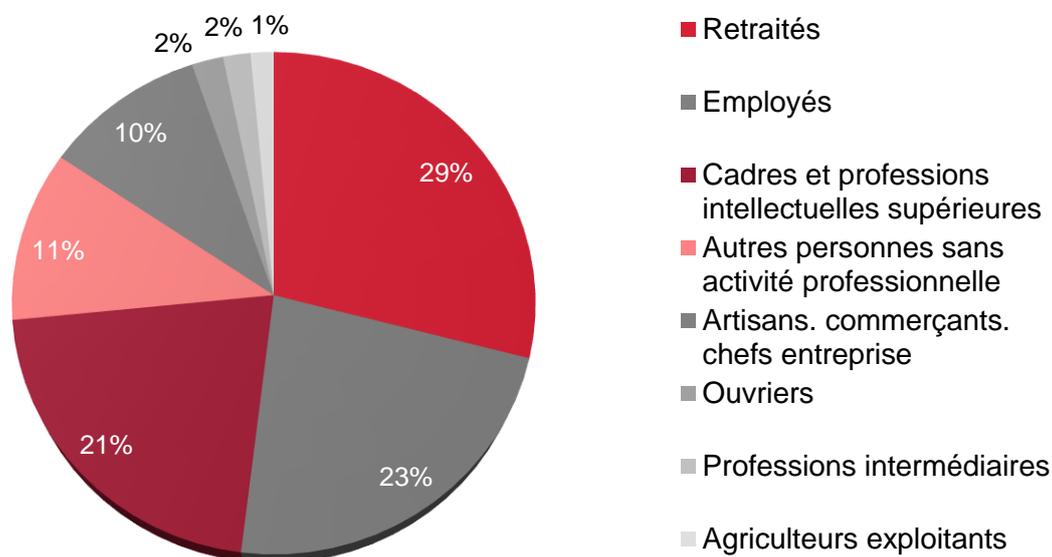


Figure 31 Répartition des CSP

6.1.3.4 Nombre de bénéficiaires du contrat

Dans notre portefeuille d'étude, le nombre maximal de bénéficiaires par contrat est de 15. Cependant, le nombre de contrats dépassant 7 bénéficiaires étant très faible (moins de 250 contrats), on décide d'isoler l'illustration graphique à 7 bénéficiaires maximum.

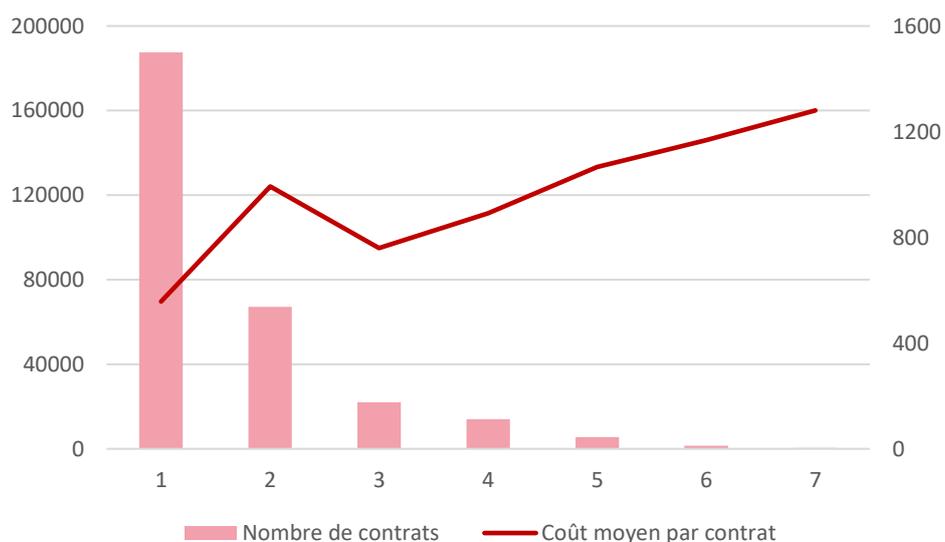


Figure 32 Evolution du CM/assuré et nombre de contrats en fonction du nombre de bénéficiaires du contrat

Le coût moyen augmente avec le nombre de bénéficiaires (1200 € pour 7 bénéficiaires contre 557€ pour 1 seul bénéficiaire, soit près du double). Ce résultat est assez intuitif car plus il y a de bénéficiaires dans le contrat, plus le nombre de recours aux soins risque d'être élevé.

6.1.3.5 Ancienneté des contrats

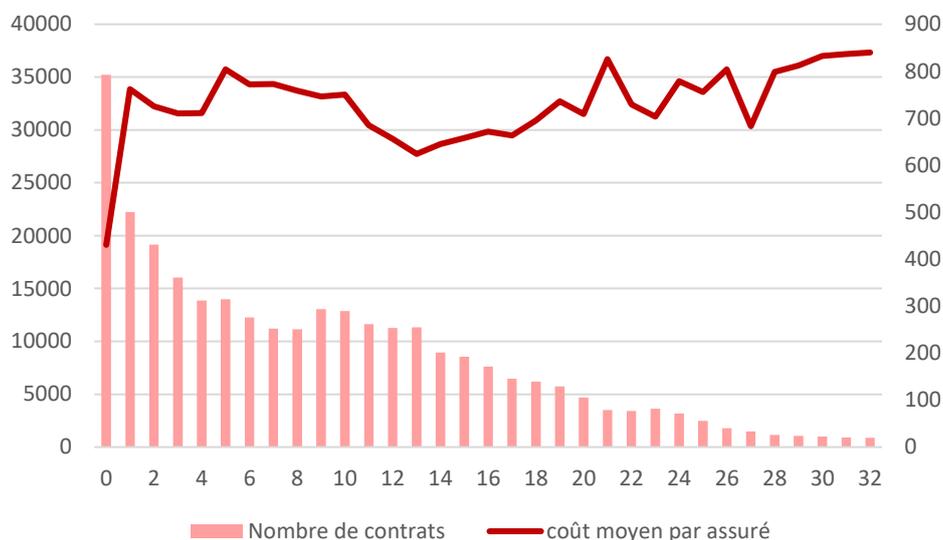


Figure 33 Evolution du CM/assuré et nombre de contrats en fonction de l'ancienneté

L'ancienneté inférieure à 1 an est la plus représentée. En effet, la réforme RIA ayant accentuée les chutes et les affaires nouvelles, le constat est cohérent. Le coût moyen de ces contrats est plus faible que le reste des anciennetés car certains contrats ont pris effet en fin d'années et donc les prestations ou leurs remboursements n'ont pas encore eu lieu. Ainsi, le coût moyen par assuré est assez stable par ancienneté pour des nombres de contrats significatifs.

6.1.4 Détermination du seuil de prestations extrêmes

6.1.4.1 Définition des valeurs aberrantes

Les valeurs aberrantes sont des valeurs extrêmes qui sont inhabituellement différentes de la distribution d'une variable. En pratique, une fois ces valeurs aberrantes, dites « extrêmes », détectées, il n'est pas toujours nécessaire de les supprimer. Cependant, dans notre cas, la suppression de ces valeurs est nécessaire. En effet, après avoir une première fois entraîné nos modèles sur nos données non corrigées, les résultats de prédictions n'étaient pas du tout satisfaisants. Ainsi, pour ne plus baisser nos modèles, nous devons d'abord déterminer à partir de quel seuil une valeur est considérée comme extrême. Ce seuil sera déterminé grâce à un outils issus de la théorie des valeurs extrêmes : le Mean Excess Plot.

6.1.4.2 La fonction de dépassement moyen (mean excess function)

Un outil graphique important pour étudier les valeurs extrêmes est la fonction de dépassement moyen définie par :

$$e(u) = \mathbb{E}(X - u | X > u).$$

Soit $\Delta_n(u) = \{i : i = 1, \dots, n, X_i > u\}$. La fonction de dépassement moyen empirique est définie par :

$$e_n(u) = \frac{\sum_{i \in \Delta_n(u)} (X_i - u)}{\text{card} \Delta_n(u)}.$$

A partir de cette fonction, on définit le graphique :

$$\{X_{(i)}, e_n(X_{(i)}) : i = 1, \dots, n\}$$

Ce graphique est appelé graphique de dépassement moyen (mean excess plot). C'est à partir de ce graphique qu'est déterminé le seuil u de valeur extrême.

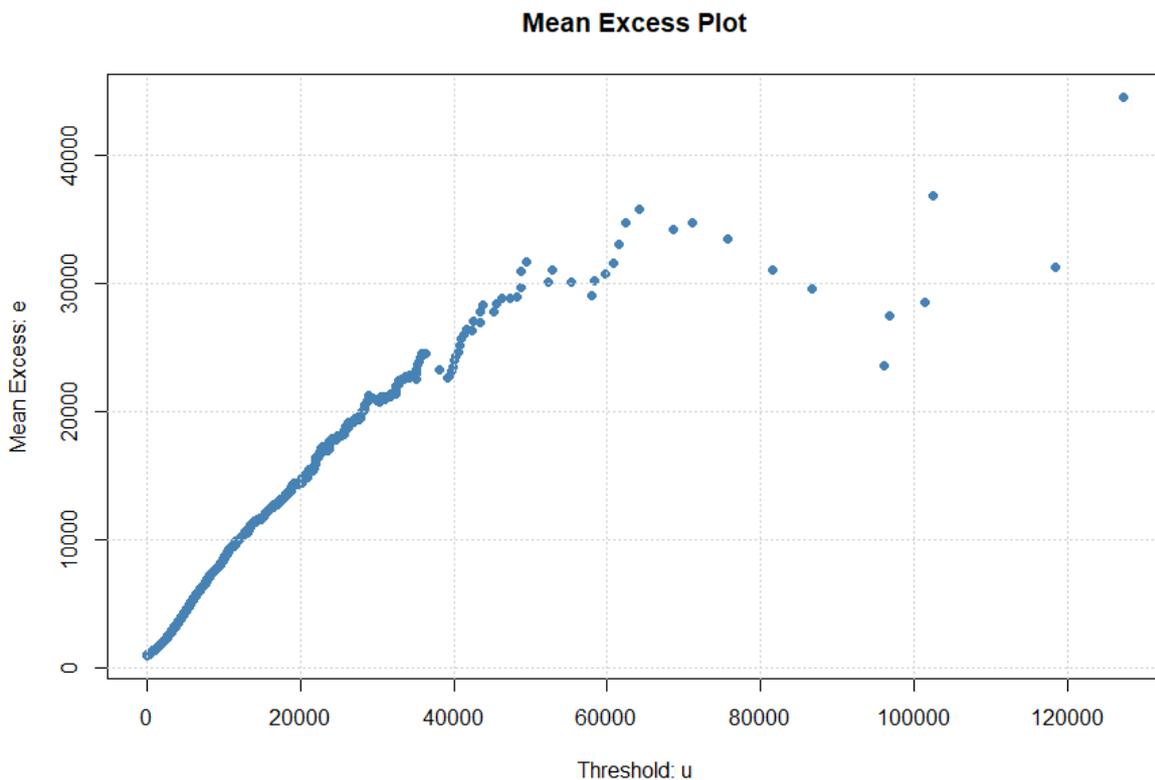


Figure 34 Mean Excess Plot

Graphiquement, le seuil est défini lors de la première rupture de linéarité des points. Cependant, il est difficile de déterminer précisément ce seuil immédiatement. Pour avoir une vision plus précise, faisons un zoom sur la première partie linéaire du graphique :

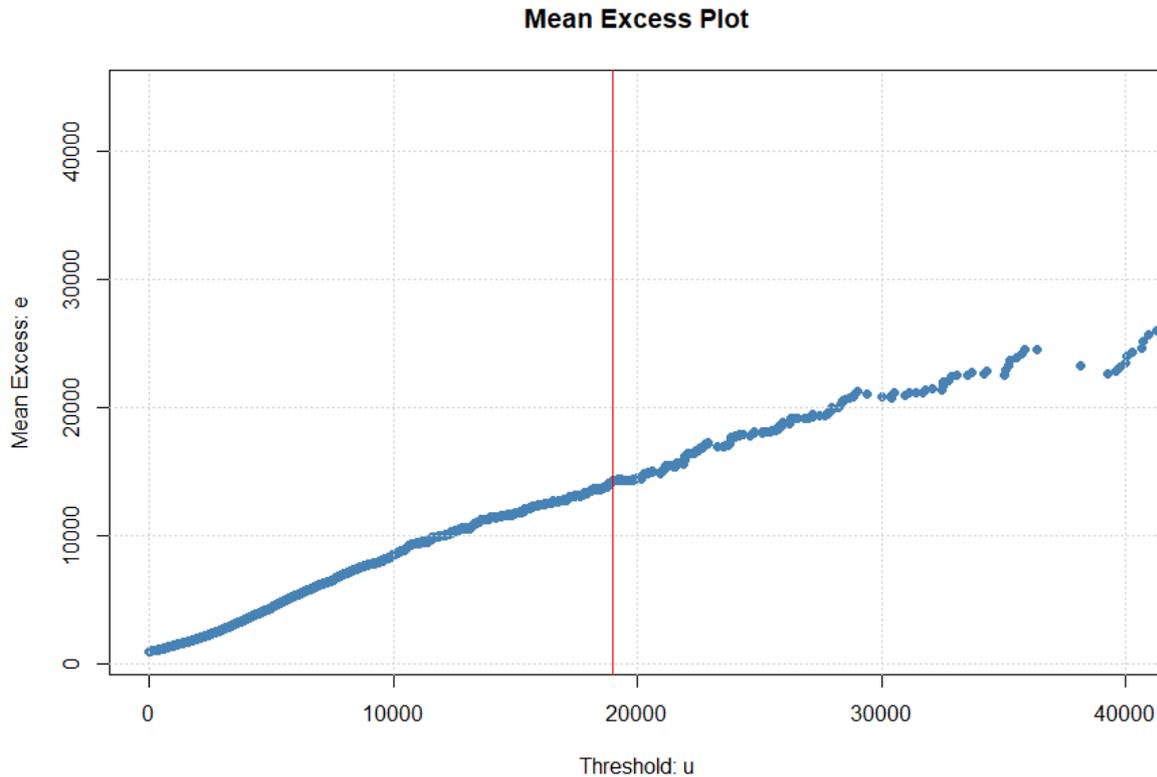


Figure 35 Mean Excess Plot zoomé

La détermination du seuil est maintenant plus optimale. On retient donc le seuil 19 000. Ainsi, les prestations au-delà de 19 000 € sont considérées comme extrêmes. Ces prestations considérées comme extrêmes seront alors supprimées de nos données d'entraînement pour ne pas biaiser nos prédictions.

6.2 Critère d'évaluation des modèles

Afin de déterminer le meilleur modèle, il est primordial de définir un critère d'évaluation. Dans notre étude, le RMSE (Root Mean Squared Error) est choisi comme critère de performance de la prédiction. L'erreur quadratique moyenne est une mesure de l'erreur absolue qui met au carré l'écart pour empêcher les écarts positifs et négatifs de s'annuler. Voici sa formalisation mathématique :

$$RMSE(M) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

où M représente un modèle prédictif

y_i représente l'observation i variant de 1 à n

\hat{y}_i représente l'estimation de y_i issue de M

6.3 Arbres de décision

6.3.1 Principe

Un arbre de décision est un algorithme qui répond successivement à une question par « oui » ou « non » pour classifier les observations selon les variables explicatives utilisées pour le modèle. Deux catégories existent :

- Les arbres de classification qui ont pour objectif de prédire une variable discrète (par exemple 0 ou 1) ;
- Les arbres de régression qui ont pour objectif de prédire une variable continue (dans notre cas : le montant d'un sinistre).

Ces types d'arbres ont été présentés par Breiman [Breiman, 1984] sous le nom de Classification And Regression Tree (CART).

Ce sont les arbres de régression qui seront utilisés ici pour modéliser les montants de sinistres.

Un arbre CART est constitué de plusieurs éléments : une racine qui est le critère de division le plus important selon l'algorithme, de nœuds qui n'est ni plus ni moins qu'un nouveau critère de division, et de feuilles qui sont les réponses aux différents parcours de nœuds (nommé aussi le nœud terminal).

La méthode de construction d'un arbre quelconque est récursive et se base sur 2 éléments :

- La détermination d'un critère de division à chacun des nœuds (choix de la variable explicative et condition de division binaire)
- La détermination d'un critère d'arrêt (détermine si un nœud est terminal ou non).

La méthode de construction de l'arbre optimal est la suivante :

- Créer un arbre maximal créant autant de nœuds que nécessaire afin d'avoir 0% d'erreur sur la base d'apprentissage. Cet arbre maximal sera parfait pour la base d'apprentissage, mais ne plus du tout robuste pour tout nouvel échantillonnage puisqu'il sera basé sur les spécificités de la base sur laquelle il a appris. C'est ce que l'on appelle le sur-apprentissage.
- Elaguer cet arbre maximal dans le but d'arriver à un arbre suffisamment précis et robuste quel que soit l'échantillon ou la base de données sur laquelle il sera appliqué.

Quel que soit l'arbre, chacune des observations de la base appartient à une seule et unique feuille.

6.3.2 Critère de division

A chaque nœud, le choix de la covariable à diviser et son critère de division binaire est obtenu selon la méthode suivante :

Supposons que le nœud soit divisé en deux branches : B_L et B_R ayant respectivement pour moyenne \bar{y}_L et \bar{y}_R , alors l'algorithme va chercher à minimiser le *Sum of Squares Error (SSE)* :

$$SSE = \sum_{i \in B_L} (y_i - \bar{y}_L)^2 + \sum_{j \in B_R} (y_j - \bar{y}_R)^2$$

où y_i représente la prédiction.

6.3.3 Critère d'arrêt

Ces étapes récursives de division à chaque nœud s'arrêtent lorsque qu'une règle définie au préalable est vérifiée. Il s'agit très souvent d'un nombre de feuilles minimal à avoir pour éviter le surapprentissage, ou encore d'un seuil minimal de gain en hétérogénéité.

6.3.4 Elagage de l'arbre

Dans la librairie *rpart*, le paramètre de complexité « cp » (*complexity parameter*) représente l'amélioration minimale du modèle nécessaire à chaque nœud. Ce paramètre se base sur le coût de la complexité du modèle.

Il représente un critère d'arrêt. Cela aide ainsi à accélérer la recherche de fractionnements en déterminant les divisions qui ne répondent pas au critère et les élaguer avant d'aller trop loin.

A l'aide la fonction R « plotcp », nous pouvons déterminer graphiquement le paramètre de complexité optimal :

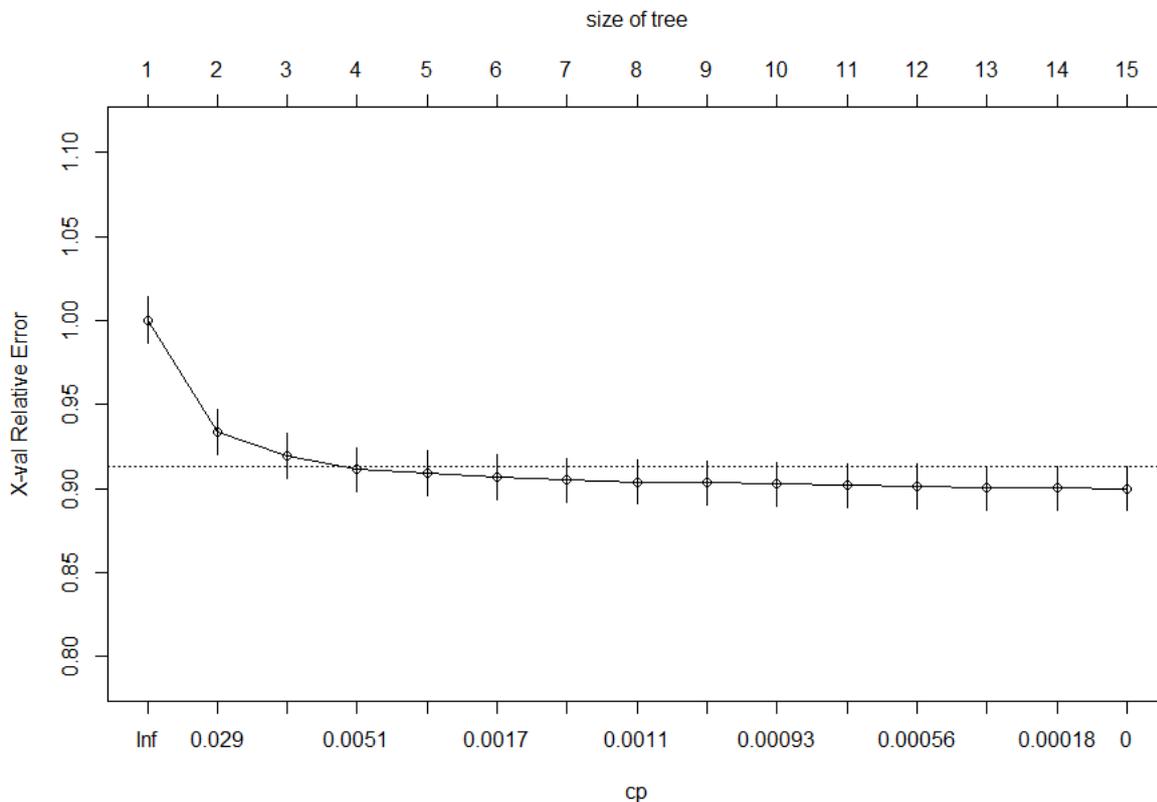


Figure 36 Evolution du paramètre de complexité en fonction de la taille de l'arbre

On remarque que l'erreur relative diminue avec le paramètre de complexité. Ainsi, l'erreur relative optimale est obtenue pour un paramètre de complexité nul. L'arbre optimal sera donc composé de 15 feuilles.

6.3.5 Arbre optimal

L'arbre optimal obtenu est alors obtenu en figure 37, implémenter sous R grâce à la fonction « rpart » puis rpart.plot pour afficher cet arbre. La fonction « rpart » à l'avantage de permettre une validation croisée.

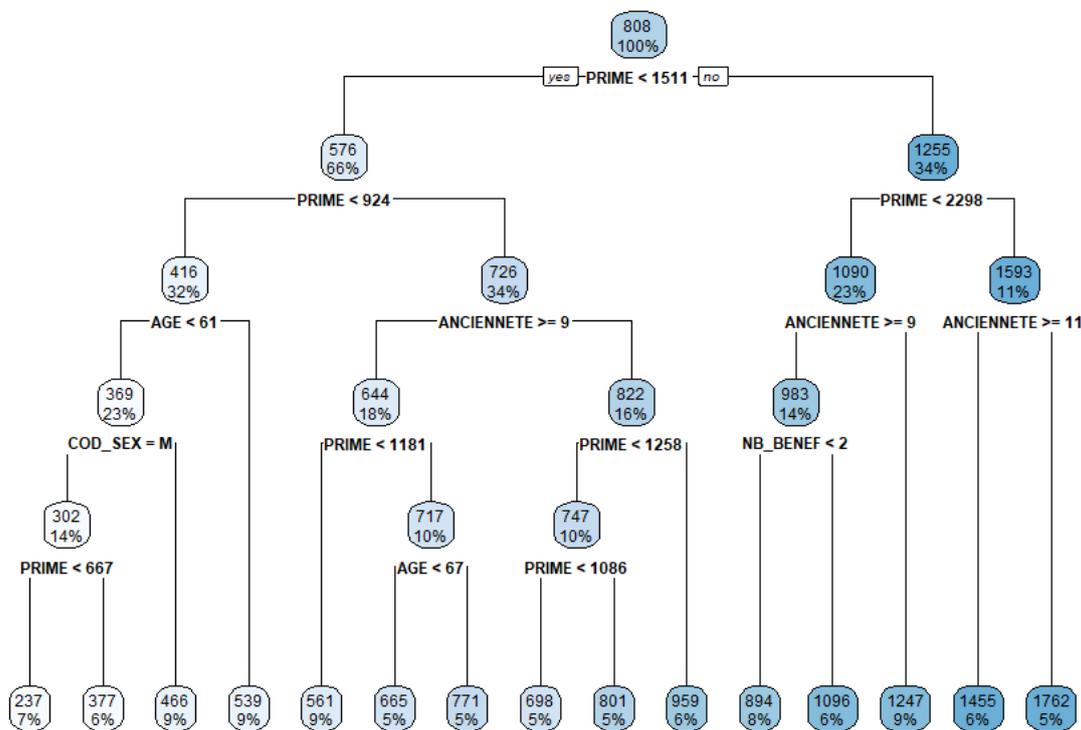


Figure 37 Arbre de régression optimal

Graphiquement, plus la couleur du nœud est foncée, plus l'individu est risqué.

On remarque que la variable la plus utilisée pour la décision est la prime. Effectivement, la prime pure correspondant à l'espérance de la sinistralité, c'est donc assez logique.

Ainsi, la distribution des sinistres prédits varie de 237€ à 1762€. Le profil le moins risqué correspond à un homme âgé de moins de 61 ans et ayant une prime de moins de 667 €. Le profil le plus risqué correspond à un assuré ayant une prime supérieure à 2298€ et une ancienneté inférieure à 11 ans.

Le RMSE de l'arbre de décision est de 378. Sachant que la prestation moyenne des assurés présent est d'environ 1000€, le modèle a une erreur moyenne d'environ 37,8%. Nous allons comparer ce résultat avec un autre modèle afin de voir s'il est possible de trouver un modèle plus précis.

6.4 Modèles Linéaires Généralisés

6.4.1 Principe et théorie mathématique

Le modèle linéaire classique est souvent inadapté quant à la loi qu'il associe aux variables d'intérêt. Grâce aux GLM il est possible de leur associer d'autres lois que la loi normale. Ces lois font partie de la famille exponentielle qui offre un cadre commun d'estimation et de modélisation.

6.4.1.1 La famille exponentielle

Une variable Y a une loi faisant partie de la famille exponentielle si sa densité peut se mettre sous la forme :

$$f(y|\theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{\phi} + c(y, \phi)\right), \quad y \in S$$

Avec :

- θ : Paramètre naturel, aussi appelé paramètre canonique ou encore paramètre de la moyenne.
- ϕ : Paramètre de dispersion. Il n'existe pas pour certaines lois de la famille exponentielle, notamment lorsque la loi de Y ne dépend que d'un seul paramètre (on pose dans ces cas $\phi = 1$). Sinon il s'agit d'un paramètre de nuisance qu'il faut estimer. Comme son nom l'indique, ce paramètre est lié à la variance de la loi. C'est aussi un paramètre très important dans la mesure où il contrôle la variance et donc le risque. Dans certains cas pondération est nécessaire pour accorder des importances relatives aux différentes observations et le paramètre ϕ est remplacé par $\frac{\phi}{\omega}$, ω étant un poids connu à priori.
- S : Support de la loi, sous-ensemble de \mathbb{R} ou \mathbb{N} .
- La fonction $b(\cdot)$ (resp. $c(\cdot)$) est une fonction de θ

6.4.1.2 Composante déterministe ou prédicteur linéaire

Supposons que nous disposions de $p \in \mathbb{N}$ variables explicatives pour expliquer Y . A chaque réponse Y_i pour $i \in \llbracket 1, n \rrbracket$, est associé un vecteur de variable explicatives (variables observées) $X_i = (X_{ij})_{j \in \llbracket 1, p \rrbracket}$. On note sous forme matricielle $X = (X_i)_{i \in \llbracket 1, n \rrbracket}$. Soit $\beta = (\beta_j)_{j \in \llbracket 1, p \rrbracket}$ un vecteur de p paramètres. Le prédicteur linéaire à n composante associée aux observations X et aux réponses Y s'écrit sous forme matricielle $\eta = X\beta$.

6.4.1.3 Fonction lien

C'est une fonction supposée monotone et différentiable faisant le « lien » entre la composante aléatoire et le d-prédicteur linéaire. En notant g , elle vérifie :

$$\forall_i \in \llbracket 1, n \rrbracket, g(\mathbb{E}[Y_i|X_i]) = \eta_i = \beta' X_i$$

On note $\mu_i = \mathbb{E}[Y_i|X_i]$ Les GLM sont une généralisation du modèle linéaire gaussien.

Il est possible de réécrire l'expression de la densité de la structure exponentielle en posant :

$$Q(\theta) = \frac{\theta}{\phi}, v(\theta) = \exp\left(-\frac{a(\theta)}{\phi}\right) w(y) = \exp\{b(y, \phi)\}$$

Les variables aléatoires de la famille exponentielles possèdent des propriétés intéressantes pour leur espérance et leur variance qui peuvent s'écrire explicitement à partir de l'écriture de leur densité.

Proposition :

Soit Y une variable aléatoire de la famille exponentielle admettant une densité $f(y|\theta, \phi)$.

$$\text{On a } \mathbb{E}[Y] = b'(\theta) \text{ et } \mathbb{V}[Y] = b''(\theta)\phi$$

g est appelée fonction lien. On distingue cependant pour chaque loi de la famille exponentielle une fonction lien particulière, la fonction canonique.

Définition : Fonction lien canonique

Avec la notation précédente, la fonction de lien canonique est la fonction réelle telle que $g(\mu) = \theta$.

Cette fonction permet de faire le lien entre l'espérance μ et le paramètre θ , dit paramètre canonique. Le tableau ci-dessous présente les fonctions liens canoniques pour quelques lois usuelles appartenant à la famille exponentielle :

Loi	Fonction lien canonique
-----	-------------------------

Normale	μ
Poisson	$\ln(\mu)$
Gamma	$\frac{1}{\mu}$
Binomiale	$\ln(\mu) - \ln(1 - \mu)$

Avec ces premières définitions, nous constatons que les modèles linéaires généraux sont des cas particuliers des modèles linéaires généralisés.

6.4.1.4 Estimation des paramètres du modèle

L'estimation des paramètres utilise la maximisation de la vraisemblance ou de la log-vraisemblance du modèle. On cherche un estimateur $\hat{\beta}$ tel que

$$\hat{\beta} = \operatorname{argmax}_{\beta \in \mathbb{R}} l_Y(\theta, \phi)$$

Dans le cas général avec les notations introduites précédemment, la vraisemblance du modèle s'écrit

$$\mathcal{L}_Y(\theta, \phi) = \prod_{i=1}^n f(y_i | \theta_i, \phi) = \exp\left(\sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi} + \sum_{i=1}^n c(y_i, \phi)\right)$$

D'où on déduit aisément la log-vraisemblance

$$l_Y(\theta, \phi) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{\phi} + \sum_{i=1}^n c(y_i, \phi)$$

Les équations du 1^{er} ordre à vérifier pour la résolution de ce programme de maximisation sont

$$\forall j \in \llbracket 1, p \rrbracket, \sum_{i=1}^n \frac{(y_i - \mu_i) x_{ij}}{\mathbb{V}[Y_i]} \frac{\partial \mu_i}{\partial \eta_j} = 0$$

Si le choix de la fonction lien canonique comme fonction lien est fait, il est possible de simplifier les équations obtenues ci-dessus. Si g est la fonction lien canonique associée à Y , alors par définition $g(\mu_i) = \eta_i = \theta_i$. Alors

$$\begin{aligned}\frac{\partial \mu_i}{\partial \eta_i} &= \frac{\partial \mu_i}{\partial \theta_i} \\ &= \frac{\partial b'}{\partial \theta_i}(\theta_i) \\ &= b''(\theta_i)\end{aligned}$$

Ce qui donne :

$$\forall j \in \llbracket 1, p \rrbracket, \sum_{i=1}^n (Y_i - \mu_i) x_{ij}$$

Les GLM permettent d'étudier les données qui ne satisferaient pas les conditions restrictives des modèles linéaires généraux (normalité et homoscedasticité des résidus). Ils permettent de plus d'étudier comme variables explicatives à la fois des variables quantitatives et qualitatives.

Effectivement, peu de problèmes de modélisation en assurance non-vie ne présentent que des variables explicatives quantitatives.

6.4.2 Modèle optimal

Le modèle retenu est un GLM Gamma avec fonction de lien log. Cependant, les GLM Gamma ne pouvant prédire des montants nuls, les assurés n'ayant pas consommé se sont vus attribuer un montant de prestation symbolique de 0,01 euros.

Afin d'obtenir un modèle le plus optimal possible, un autre critère utilisé précédemment va être utilisé pour sélectionner la combinaison interne de variables explicatives à utiliser. Pour ce faire, nous avons utilisé la fonction « stepAIC » du package R « MASS ». Cette fonction procède par étape. La première étape consiste à calculer l'AIC du modèle de référence ainsi que l'impact du retrait de chaque variable sur l'AIC. La variable dont le retrait permet une baisse maximale de l'AIC est alors supprimé jusqu'à ce que la suppression d'une des variables restantes augmente l'AIC.

Le modèle retenu est présenté ici :

```
Call:
glm(formula = PRESTA ~ PRIME + AGE + ANCIENNETE + COD_SEX + produit_cible +
     NB_BENEF + CSP, family = Gamma(link = "log"), data = DATA_TRAIN)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.5350 -1.1107 -0.4320  0.2261  6.6160
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5.260e+00	4.926e-02	106.790	< 2e-16	***
PRIME	3.360e-04	5.871e-06	57.237	< 2e-16	***
AGE	1.005e-02	4.105e-04	24.477	< 2e-16	***
ANCIENNETE	-8.363e-03	4.199e-04	-19.918	< 2e-16	***
COD_SEXM	-1.050e-01	7.264e-03	-14.457	< 2e-16	***
produit_cibleFrontaliers Comp	7.009e-03	5.511e-02	0.127	0.8988	
produit_ciblesalariés	4.851e-02	4.866e-02	0.997	0.3187	
produit_cibleSenior	9.360e-02	4.900e-02	1.910	0.0561	.
produit_cibleTNS	1.177e-01	4.886e-02	2.410	0.0160	*
NB_BENEF	1.592e-01	4.026e-03	39.544	< 2e-16	***
CSPArtisans. commerçants. chefs entreprise	6.519e-02	3.584e-02	1.819	0.0690	.
CSPAutres personnes sans activité professionnelle	8.578e-02	3.486e-02	2.461	0.0139	*
CSPCadres et professions intellectuelles supérieures	8.544e-02	3.568e-02	2.395	0.0166	*
CSPEmployés	7.618e-02	3.464e-02	2.199	0.0279	*
CSPOuvriers	8.073e-02	4.092e-02	1.973	0.0485	*
CSPProfessions intermédiaires	7.580e-02	4.214e-02	1.799	0.0720	.
CSPRetraités	1.755e-01	3.459e-02	5.074	3.91e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Gamma family taken to be 1.681169)

Null deviance: 204399 on 136575 degrees of freedom
Residual deviance: 179881 on 136559 degrees of freedom
(388 observations effacées parce que manquantes)
AIC: 2090044

Number of Fisher Scoring iterations: 7

Les coefficients sont globalement tous significatifs, hormis les différents produits cibles où un seul d'entre eux a une p-valeur significative.

L'analyse de déviance ANOVA, partant du modèle constitué uniquement de l'intercept procède par ajout successif des variables explicatives en effectuant lors de chaque ajout un test du rapport de vraisemblance entre le modèle sans ajout et le modèle avec ajout de la variable explicative. L'idée est de tester si l'ajout d'une variable explicative diminue ou non la déviance du modèle.

Analysis of Deviance Table

Model: Gamma, link: log

Response: PRESTA

Terms added sequentially (first to last)

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			136575	204399	
PRIME	1	19265.8	136574	185133	< 2.2e-16 ***
AGE	1	944.8	136573	184188	< 2.2e-16 ***
ANCIENNETE	1	1223.5	136572	182965	< 2.2e-16 ***
COD_SEX	1	220.9	136571	182744	< 2.2e-16 ***
produit_cible	4	85.3	136567	182658	2.533e-10 ***
NB_BENEF	1	2594.3	136566	180064	< 2.2e-16 ***
CSP	7	182.7	136559	179881	< 2.2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Les variables gardées par la fonction « stepAIC » sont confirmée par le test ANOVA, au vu des p-valeur significatives au seuil de 5%.

6.4.3 Analyse des résidus de déviance

La déviance permet d'évaluer l'adéquation d'un modèle en le comparant au modèle plus général ayant le nombre maximum de paramètres à estimer. Ce modèle est appelé « le modèle saturé » et a un paramètre par observation.

Afin d'évaluer cet ajustement, la déviance examine la différence entre logarithmiques-vraisemblance du modèle saturé et du modèle étudié, en d'autres termes :

$$l(b_{sunt}, y) - l(b, y)$$

Où b_{sunt} désigne l'estimateur du maximum de vraisemblance du vecteur paramètre du modèle saturé
 b désigne l'estimateur du maximum de vraisemblance du vecteur paramètre du modèle étudié

Ainsi, la déviance est définie comme : $D = 2[l(b_{sunt}, y) - l(b, y)]$. Un « bon » modèle aura alors une déviance proche de 0.

Le résidu de déviance individuel est alors défini comme la racine carrée de la contribution individuelle de l'observation, multipliée par le signe du résidu brut.

Pour valider notre modèle, il est donc important de vérifier le comportement de ces résidus soient proche de 0.

Pour ce faire, affichons la sortie R des résidus de déviance :

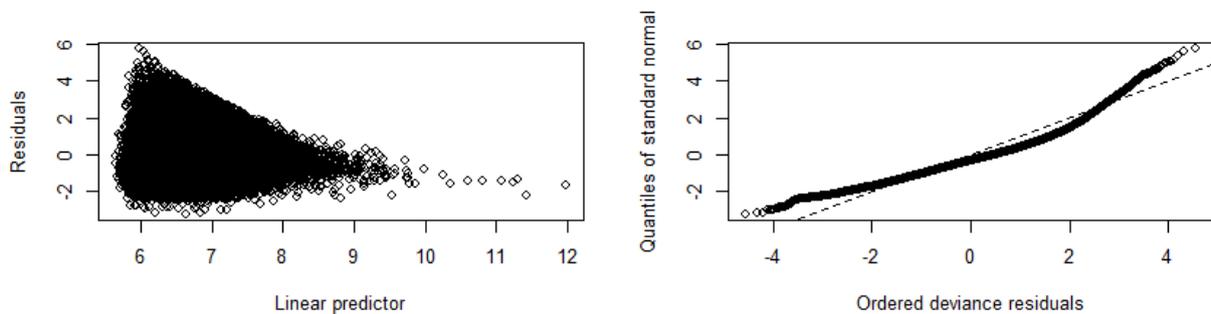


Figure 38 Représentation des résidus de déviance

Les résidus de déviance semblent centrés en 0 et relativement bien alignés sur la bissectrice du QQ-plot, hormis de légers décrochages pour les queues de distributions. Ainsi, l'analyse de ces résidus confirme la pertinence du modèle.

6.4.4 Analyse des valeurs aberrantes

Avant d'entraîner nos modèles, nous avons préalablement identifié et supprimé les valeurs aberrantes de nos données afin de ne pas biaiser nos prédictions. Dans le but d'être certain qu'aucune donnée extrême ne soit encore présente, nous allons utiliser un outil statistique : la distance de Cook. Cette métrique mesure l'effet de la suppression d'une donnée sur l'estimation des paramètres du modèle de régression. Ainsi, la distance de Cook est définie comme :

$$D_i = \frac{e_i^2}{k \text{ MSE}} \times \frac{h_{ii}}{(1 - h_{ii})^2}$$

Où e_i représente le i ème résidu

k représente le nombre de paramètre

MSE représente l'erreur quadratique moyenne du modèle de régression

h_{ii} représente la valeur de la i ème colonne et i ème ligne de la matrice de projection de projection des variables explicatives

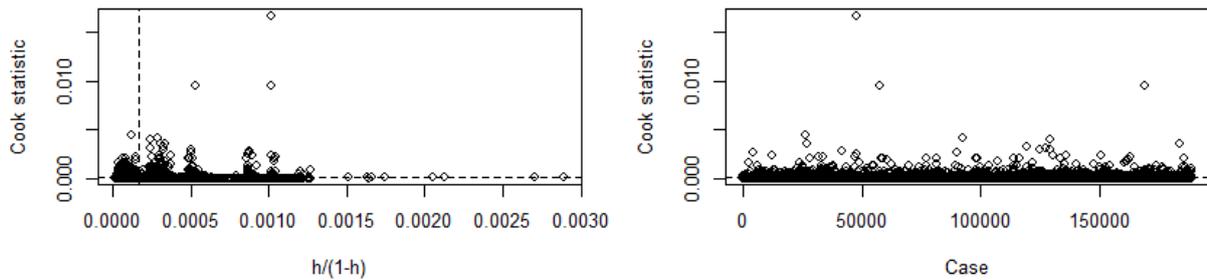


Figure 39 Représentation des distances de Cook

Le graphique de gauche illustre la statistique de Cook vs l'effet de levier standardisé :

- n = nombre d'observations, p = nombre de paramètres estimés
- l'équation de la droite horizontale est $\frac{8}{n-2p}$. Les points au-dessus de la droite exercent une grande influence sur le modèle
- l'équation de la droite verticale est $\frac{2p}{n-2p}$. Les points à droite de la droite ont un fort effet de levier.

Un point levier est une observation qui a une prédiction inhabituellement différente de la distribution des prédictions. Un point influent est une valeur dont la suppression entrainerait une modification importante des coefficients estimés du modèle.

Ainsi, on remarque qu'il n'y aucun point dépassant le seuil représenté à l'aide de la droite horizontale dans le graphique de droite (droite qui n'est pas présente car pas de points influents).

Sur le graphique de gauche, une majorité de points se trouve en dessous de la ligne horizontale et à gauche de la ligne verticale.

La détection des valeurs aberrantes grâce aux valeurs extrêmes ont donc permis de ne plus avoir de valeurs influentes dans nos données.

Le RMSE du GLM est de 289, donc meilleur que l'arbre de décision. En effet, son erreur est donc d'environ 29%. Les prédictions faites par le GLM seront donc utilisées pour la projection à 3 ans du portefeuille.

7 Projection du portefeuille

7.1 Approche de projection

Une fois les effectifs et les flux projetés, nous pouvons alors estimer différents indicateurs. En effet, les taux de chutes estimés par Kaplan-Meier et le nombre d'affaires nouvelles prédis grâce aux séries temporelles permettent de prédire le taux d'extinction du portefeuille individuel selon les 2 scénarios. Ainsi, les taux de chutes vont permettre de projeter les taux d'extinction futurs pour les générations actuelles. Le nombre d'affaires nouvelles permettra de projeter les taux d'extinction des générations futures, c'est-à-dire à partir de 2022. Les prestations estimées permettent alors de projeter le ratio de rentabilité S/P.

Pour ce faire, nous avons donc projeté le portefeuille en différenciant les générations 2021 et antérieures aux générations futures, à partir de 2022.

7.1.1 Méthode utilisée pour projeter les générations 2021 et antérieures :

Nous sommes partis de la base de données regroupant toutes les variables discriminantes des assurés en 2021, noté ici année « N ». A partir de ces infos, nous les avons regroupés en faisant un TCD afin d'obtenir une table simplifiée de nos données regroupant les informations suivantes :

- Ancienneté
- Age
- Nombre de bénéficiaires
- Sexe
- CSP
- Produit cible
- Somme de Primes
- Somme de Prestations
- Nombre de contrats

Par exemple en 2021, soit en N, si l'âge et l'ancienneté d'une tranche de données valent respectivement 54 et 5 ans, l'âge et l'ancienneté en 2022 vont valoir 55 et 6 ans, ceci ayant un impact sur les prestations et les taux de chutes.

Le terme « tranche » désigne une ligne de la base de données regroupant une combinaison de variable unique : par exemple : Ancienneté = 7, Age = 68, CSP=Retraité, Nombre de bénéficiaires=2, Sexe=H, Nombre de contrats de la tranche= 10.

Ainsi, la variable Prestation est la somme des prestations des 10 contrats de la tranche, de même pour les primes.

Pour projeter les prestations et les primes dans le temps, il suffira donc de corriger ce montant des chutes. En effet, si le taux de chutes est de 20% pour cette tranche, il ne restera alors que $10 \times (1 - 0,20) = 8$ contrats à la fin de l'année. Le montant des prestations sera alors égal à la somme des prestations

des 8 contrats restants + la somme des prestations des contrats chutés. Effectivement, les contrats chutés n'ayant plus une exposition de 1, il faut déterminer leur exposition afin d'estimer leur consommation car l'exposition influe énormément sur le montant de prestations.

Pour ce faire, nous avons simplement ventilé les expositions par ancienneté et calculé une moyenne. On obtient donc une moyenne d'exposition par ancienneté.

Ceci va permettre d'estimer l'exposition des contrats chuté et donc leur consommation.

En revenant sur l'exemple précédent, la somme des consommations des 2 contrats chutés est alors égal à la valeur donnée par le GLM multiplié par 2 et corrigé de l'exposition, en d'autres termes :

Prestation Contrats Chutés = Valeur GLM x 2 x Exposition

Prestation Stock = Valeur GLM x 8

Prestation Totale tranche = Prestation Stock + Prestation Contrat Chutés

Concernant les primes, on doit préalablement calculer la moyenne des primes des contrats de chaque tranche afin d'avoir un montant de primes par profil comme pour les prestations. En effet, contrairement aux prestations qui sont obtenues par GLM et donc directement par profil, les primes sont sommées par tranche.

Une fois les primes moyennes projetées de la même manière que les prestations, on obtient donc une projection fiable par année du portefeuille.

Voici donc une synthèse de la méthode de projection :

- Ancienneté N+1 = 1+ Ancienneté N
- Age N+1 = 1 + Age N
- Taux de chute N+1 = Taux de chute estimé par Kaplan-Meier par Ancienneté N+1
- Nombre de contrats N+1 = Nombre de contrats N x (1-Taux de chute N+1)

A partir de ces informations, on détermine les prestations :

- Prestations Stock N+1 = Prestations estimées par GLM selon les variables de la tranche x Nombre de contrats N+1
- Nombre de chutes N+1 = Taux de chutes x Nombre de contrat
- Exposition des chutes = Moyenne des expositions ventilées par ancienneté
- Equivalent Nombre de chutes = Exposition des chutes x Nombre de chutes N+1
- Prestations Chutes N+1 = Prestations estimées par GLM selon les variables de la tranche x Nombre de chutes N+1 x Exposition
- Primes Stock N+1 = Primes Moyennes N x Nombre de contrats N+1
- Primes Chutes N+1 = Primes Moyennes N x Nombre de chutes N+1 x Exposition

On obtient donc les montants de primes et de prestations pour les générations antérieures à 2021 en sommant les flux Stock et Chutes.

7.1.2 Méthode utilisée pour les générations postérieures à 2021

Cette méthode se différencie de la précédente car les affaires nouvelles vont ici impacter la projection. En effet, le profil de risque est différent entre le stock et les affaires nouvelles. Ainsi, afin d'avoir une base d'assuré ayant des profils les plus similaires aux « nouveaux clients », nous nous sommes basés

sur la table de données des affaires nouvelles de 2020 et 2021. La table des affaires nouvelles de 2020 sera utilisée pour la projection du scénario hors RIA et celle de 2021 pour le scénario RIA.

Nous avons donc récupéré les mêmes informations pour les assurés nouveaux que pour les anciens assurés : âge, ancienneté, CSP, Nombre de bénéficiaires, Sexe, Prestations, Primes. Ces informations clients représenteront donc les informations types des affaires nouvelles et serviront de base de prédiction des prestations. Ainsi, nous aurons des montants de prestations cohérent avec le profil de risque différent des affaires nouvelles. La projection sera ensuite réalisée de la même manière que lors de la méthode précédente.

Cependant, en 2020, le nombre d'affaires nouvelles, donc de lignes, était de 24 101. Le montant total de prestations et de primes sera donc proportionnel au nombre de contrat.

Néanmoins, lors de la projection du scénario Hors RIA, les montants d'affaires nouvelles seront différents. Par exemple, en 2022 dans le scénario Hors RIA, le montant d'affaires nouvelles prédit est de 23 447. Le montant de prestations sera donc plus faible. Pour corriger cet effet de nombre, nous appliqueront simplement un coefficient correcteur égal à $\frac{23\,447}{24\,101}$. Ainsi, si le montant de prestations prédit d'après notre base pour l'année 2022 et la génération 2022 est de 1 000 000€, le montant réel utilisé dans la projection sera de $1\,000\,000 \times \frac{23\,447}{24\,101} = 972\,864$ €. En faisant de même pour toutes les générations en fonction des 2 scénarios, nous obtenons notre projection des prestations à 3 ans de la rentabilité et du taux d'extinction.

Concernant les primes l'exposition est à prendre en compte. En effet, les expositions des affaires nouvelles sont en moyenne de 62% pour les affaires nouvelles de 2021 et 67% en 2020. Les montants de primes seront donc globalement inférieurs au montant réel annuel à utiliser pour les années suivantes. Pour corriger cela, on propose de proratiser les primes par le taux d'exposition. Par exemple, si le montant de primes pour une tranche ayant une exposition moyenne de 40% est de 300€, le montant utilisé pour l'année suivante de cette même tranche sera de $\frac{300}{0,4} = 750$ €. Ceci permet d'annualiser la prime initiale.

7.2 Résultats Obtenus

7.2.1 Taux d'extinction du portefeuille

Volume de primes	Taux d'extinction du portefeuille											
	N	N + 1	N + 2	N + 3	N + 4	N + 5	N + 6	N + 7	N + 8	N + 9	N + 10	
Scénario Hors RIA		100%	84%	69%	61%	57%	44%	42%	36%	29%	27%	
Scénario RIA		100%	75%	57%	48%	39%	32%	26%	19%	17%	13%	

Figure 40 Taux d'extinction du portefeuille projeté

Le volume de primes diminue fortement avec la RIA. On remarque qu'au bout de 5 ans, il y a 12% de primes en moins. Au bout de 9 ans, l'écart de primes est de 14%. Il ne reste ainsi que 13% du volume de primes initial, contre 27% dans le scénario sans prise en compte de la réforme.

La durée de vie des contrats est ainsi beaucoup plus courte en cas de RIA prolongée. La rentabilité des contrats santé individuelle s’améliorant au fur et à mesure de l’ancienneté, ce raccourcissement de la durée viendrait alors engendrée une perte de rentabilité.

7.2.2 Rentabilité et profitabilité

On définit le profit comme la différence entre les primes perçues et les sinistres.

	2022		2022	
	RIA	Hors RIA	RIA	Hors RIA
	Avant intégration des AN		Après intégration des AN	
Ecart de S/P RIA	-0,3%		0,1%	
Perte de profit N-1	-14,2%	-9,4%	-7,8%	-5,1%
Ecart de profit RIA	-4,8%		-2,7%	

Figure 41 Rentabilité et profitabilité projetées

- Ecart de S/P RIA : le S/P RIA est inférieur au S/P Hors RIA de 0,3%.
- Perte de profit N-1 : On part du profit réalisé l’année précédente, ici 2021 et on compare l’évolution en taux de croissance. Ainsi, avant intégration des AN, le scénario RIA a entraîné une perte de profit de -14,2% par rapport à 2021.
- Ecart de profit RIA : Avant intégration des AN, le scénario RIA engendre 4,8% de profit en moins par rapport au profit Hors RIA.

Avant intégration des affaires nouvelles, le S/P RIA est meilleur que le S/P Hors RIA. En effet, dans le scénario RIA, les taux de résiliations des anciennetés les plus faibles est plus élevé. Or, ces générations récentes ont une rentabilité mauvaise car la consommation est élevée au début du contrat. Ainsi, une plus grande part de mauvais risque est sorti du portefeuille avec la RIA, le S/P s’améliore donc. Cependant, en terme de profit, la hausse des résiliations a provoqué une perte de profit de 4,8%.

Après intégration des affaires nouvelles, le S/P RIA est supérieur à celui Hors RIA et l’écart de profit diminue. En effet, les affaires nouvelles étant plus nombreuses dans le scénario RIA, le profit augmente donc d’avantage. De plus, les affaires nouvelles ayant une ancienneté de moins d’un an, donc un S/P élevé, le S/P s’est donc dégradé.

De la même manière, on obtient les résultats pour les 2 années suivantes en ayant intégrer les affaires nouvelles :

	2023		2024	
	RIA	Hors RIA	RIA	Hors RIA
Ecart de S/P RIA	0,9%		1,7%	
Perte de profit N-1	-11,8%	-6,7%	-15,2%	-7,3%
Ecart de profit RIA	-5,1%		-7,9%	

Figure 41 Rentabilité et profitabilité projetées

L'écart de S/P entre les 2 scénarios s'est creusé. En effet, la rentabilité des contrats est la moins bonne entre la 1^{ère} et la 3^{ème} année du contrat. Les affaires nouvelles intégrées en 2022 ont donc une ancienneté de 1 an et donc une très mauvaise rentabilité. Le scénario RIA ayant un nombre beaucoup plus important d'affaires nouvelles, l'écart s'est ainsi creusé et accentué tous les ans.

La perte de profit peut donc s'expliquer par deux phénomènes :

- Un effet structure : baisse de la rentabilité du portefeuille
- Un effet volume : perte de contrats

Ces contrats moins rentables et moins nombreux ont ainsi diminué le profit chaque années, malgré les affaires nouvelles. La perte de profit cumulée sur les 3 ans est de -34,8% pour le scénario RIA et -19,1% dans le scénario sans réforme.

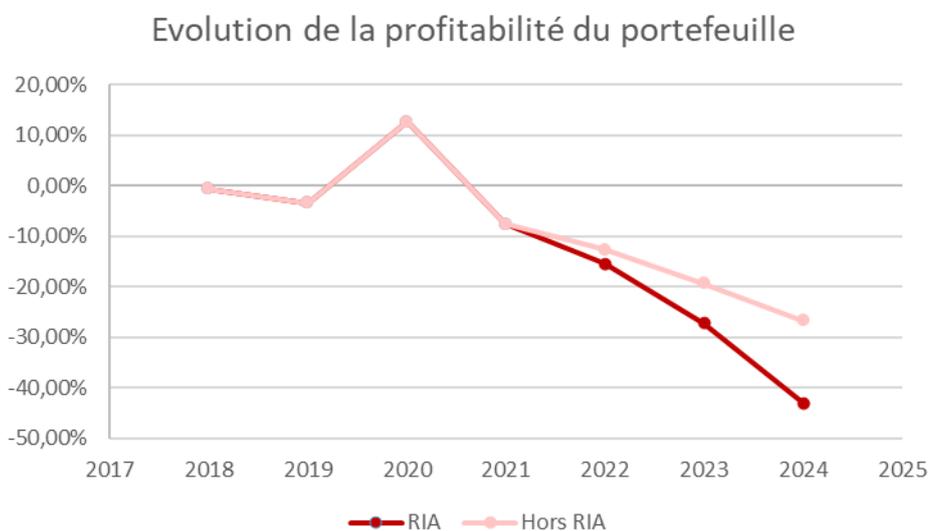


Figure 42 Evolution projetée de la profitabilité du portefeuille

De manière générale, les affaires nouvelles ne permettent pas de garder un niveau de profit équivalent chaque année. Hormis en 2020, année du confinement ayant provoqué une baisse des recours aux soins, le profit diminue tous les ans. La réforme RIA ayant accentué les résiliations dégrade d'autant plus le profit. De plus, la perte de profit depuis 2018 s'explique par l'ANI car le portefeuille individuel

a diminué au profit du collectif. La RIA accélère donc la perte de contrat subit depuis 2016 (date de la réforme) par l'ANI.

Conclusion générale

Après avoir analysé la structure du portefeuille et les effets déjà observés des évolutions réglementaires, l'objectif était d'estimer les impacts prospectifs de la réforme RIA suite à la mise en place complète du 100% Santé.

Ce mémoire a ainsi permis d'obtenir une projection complète du portefeuille. De ce fait, les effectifs au complet ainsi que leurs flux ont été projetés, selon deux scénarios avec et sans réforme. L'estimateur de Kaplan-Meier a permis d'obtenir des taux bruts de résiliation en tenant compte de la censure. La méthode de Whittaker-Henderson a permis de lisser ces taux bruts.

Néanmoins, les taux de chutes n'étant segmentés que par ancienneté, l'impact des autres variables n'est donc pas pris en compte, entraînant un biais. Les courbes de survies présentes en annexe offrent cependant une visualisation de l'impact de chaque variable sur la résiliation. Ceci pourrait alors servir de base dans la construction d'un taux de chutes plus précis ou également dans le but d'une meilleure appréhension du risque.

Les travaux réalisés dans la modélisation des prestations pourront également servir dans une optique de nouvelle approche de tarification du risque suite à ces réformes. En effet, les modèles linéaires généralisés ont obtenu les meilleurs résultats en termes de précision (critère RMSE) mais les arbres de régression ont servi dans la détection des profils d'assurés les plus à risque de consommer.

Enfin, les affaires nouvelles ont été modélisées grâce aux processus SARIMA tenant compte de la saisonnalité et d'une hypothèse de corrélation entre le nombre de résiliations et le nombre d'affaires nouvelles. En conséquence de cette hypothèse, la modélisation des affaires nouvelles a été adaptée selon le scénario avec et sans RIA.

Le volume de primes étant considéré comme connu et donc constant au cours du temps, des phénomènes tels que l'inflation ou l'indexation tarifaire n'ont pas été pris en compte. Cependant, le but principal étant d'analyser des ratios de S/P ou de profitabilité, en comparant ainsi les primes aux sinistres, ces phénomènes auraient donc été compensés par l'inflation médicale. L'effet réel de ce choix est donc à atténuer.

Ainsi, l'impact de la réforme RIA est net en termes de S/P et surtout de profitabilité du portefeuille. Cependant, d'autres impacts sont aussi à prévoir.

En effet, les résiliations et les affaires nouvelles étant plus nombreuses, une hausse de frais de gestion et d'acquisition ainsi qu'une baisse des taux de renouvellements est donc à soupçonner, augmentant la prime des assurés. La hausse des frais de gestion viendrait alors diminuer le ratio S/P combiné et donc la rentabilité technique.

De plus, la recherche perpétuelle du meilleur tarif est bénéfique pour l'assuré qui voit sa cotisation diminuer. Cependant, cela met à mal la mutualisation du risque. Effectivement, d'après les figures 43 et 47 présentes en annexe, le risque de résiliation diminue avec l'âge, les jeunes actifs résilient beaucoup plus que les seniors. La figure 51, illustrant la moyenne d'âge par ancienneté, confirme également cette hypothèse.

De ce fait, la prime pure étant directement indexée sur le risque, cela va entraîner un gain important de primes pour les jeunes actifs. A contrario, les primes d'assurance santé vont devenir inaccessibles pour les retraités qui ne pourront plus s'assurer.

L'équilibre global des contrats individuels basés sur le fait que les bons risques, payant une cotisation très élevée par rapport à leur risque, financent les très mauvais risques, est menacé. Les jeunes actifs

résiliant afin d'avoir un tarif plus attractif, seuls restent les retraités n'étant plus assurables ailleurs et qui devront à terme payer le prix de leur risque sans pouvoir résilier leurs contrats.

Annexe

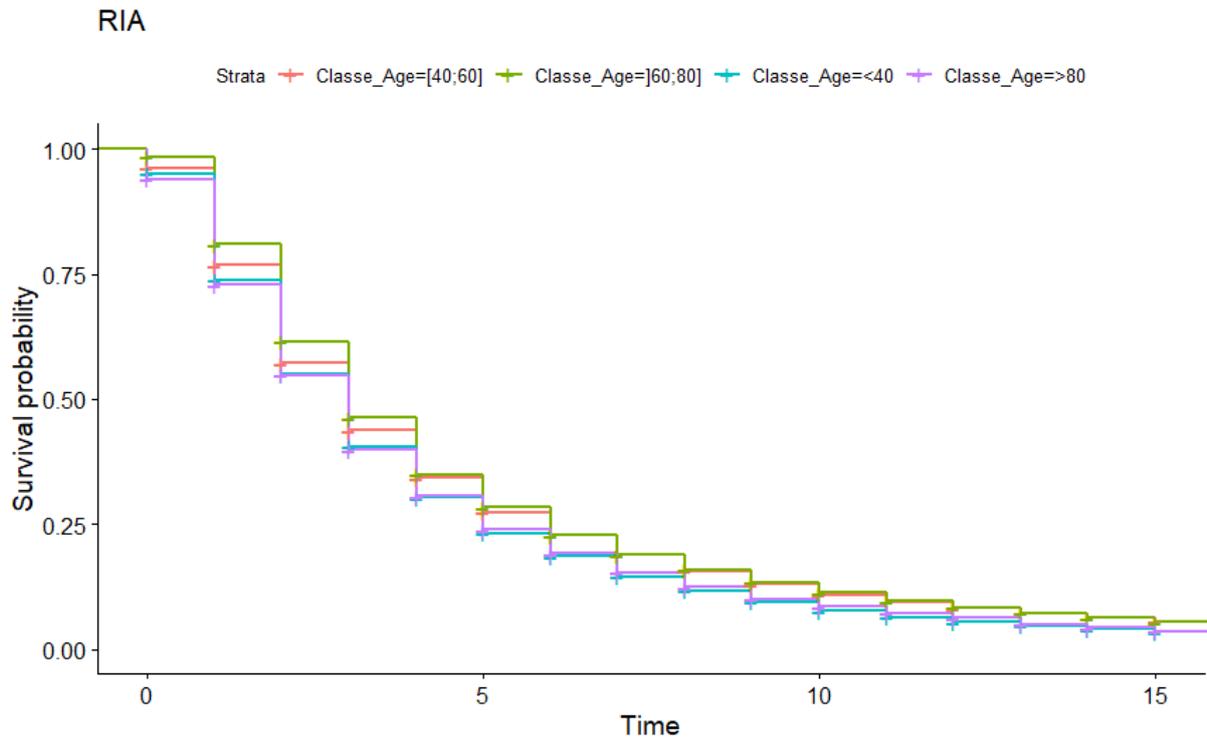


Figure 43 Courbe de survie RIA en fonction de l'âge

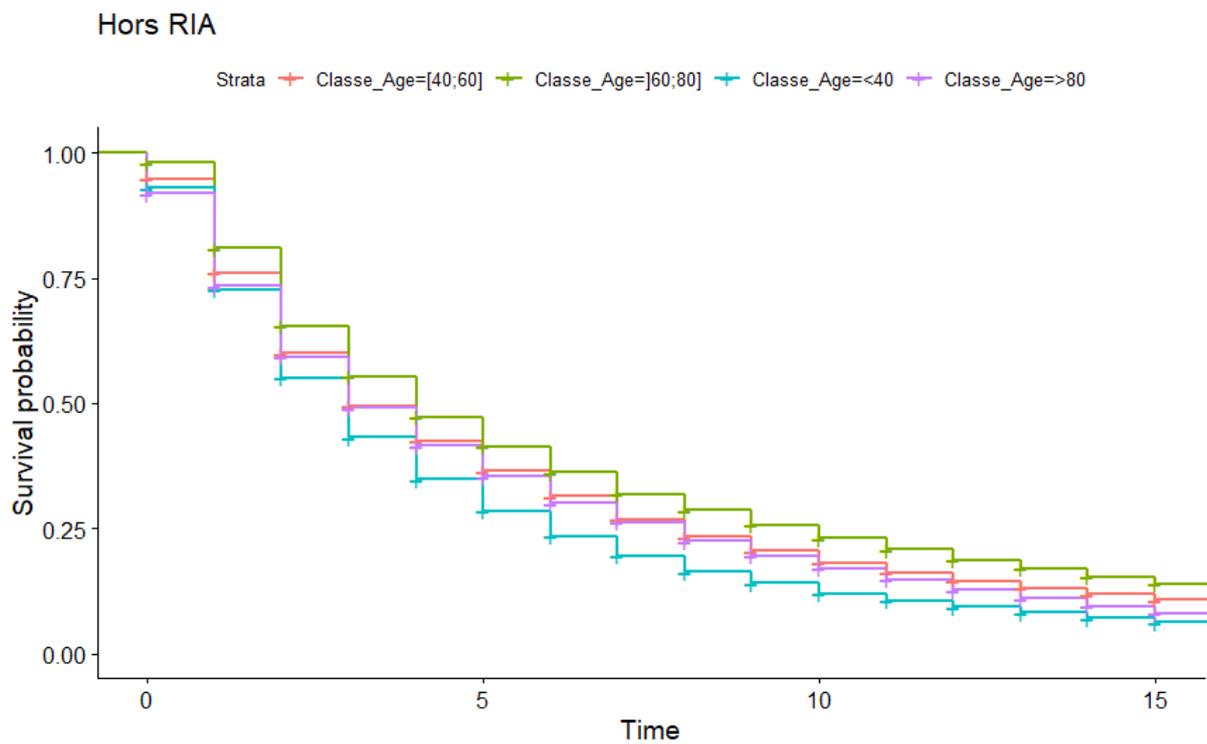


Figure 44 Courbe de survie Hors RIA en fonction de l'âge

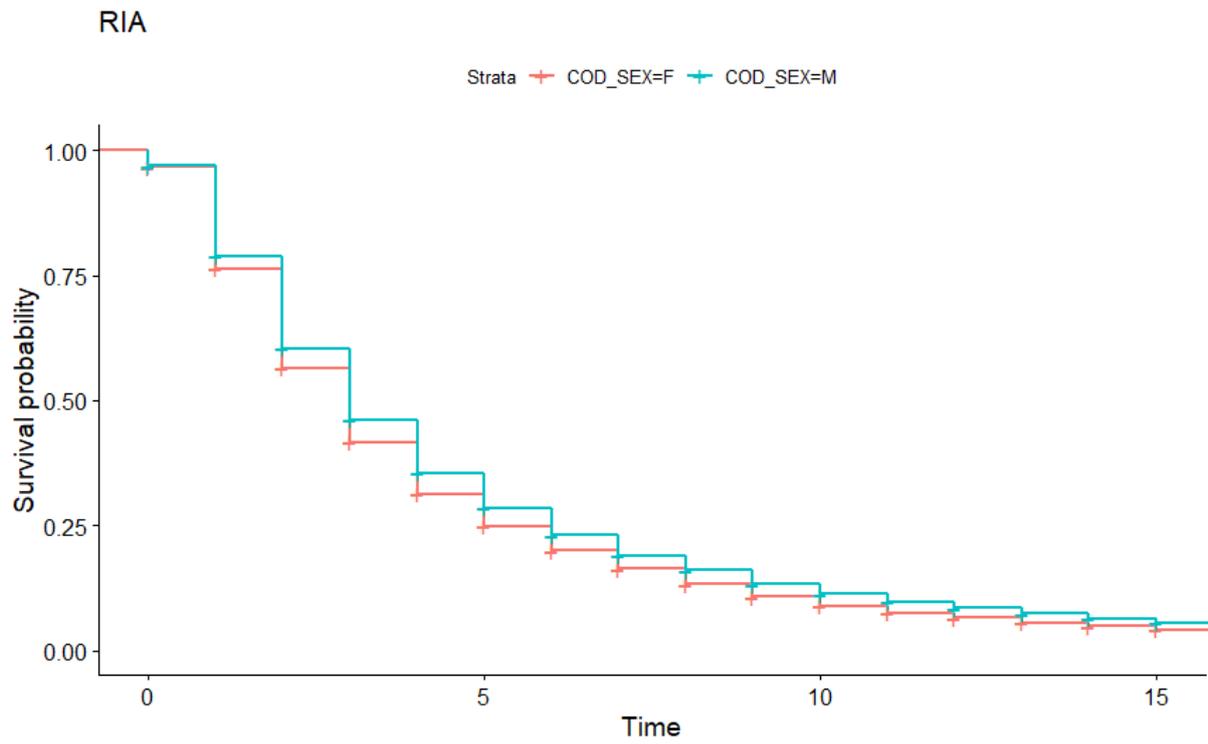


Figure 45 Courbe de survie RIA en fonction du sexe

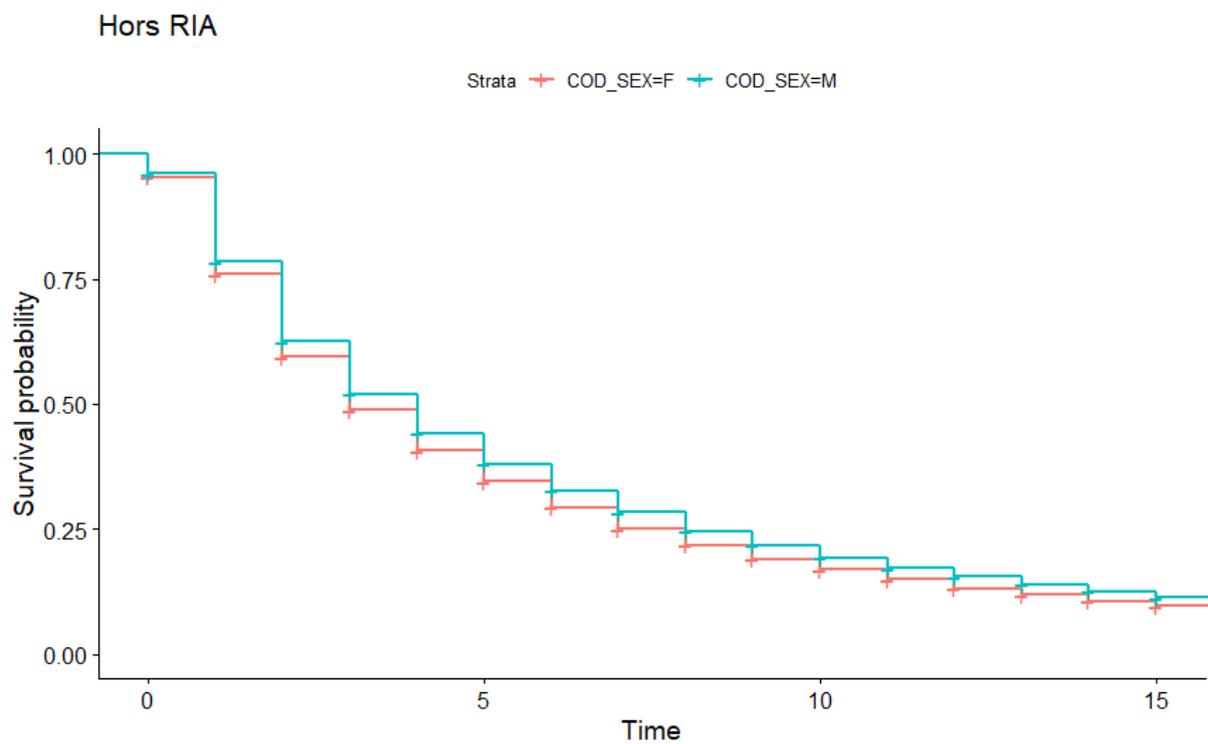


Figure 46 Courbe de survie Hors RIA en fonction du sexe

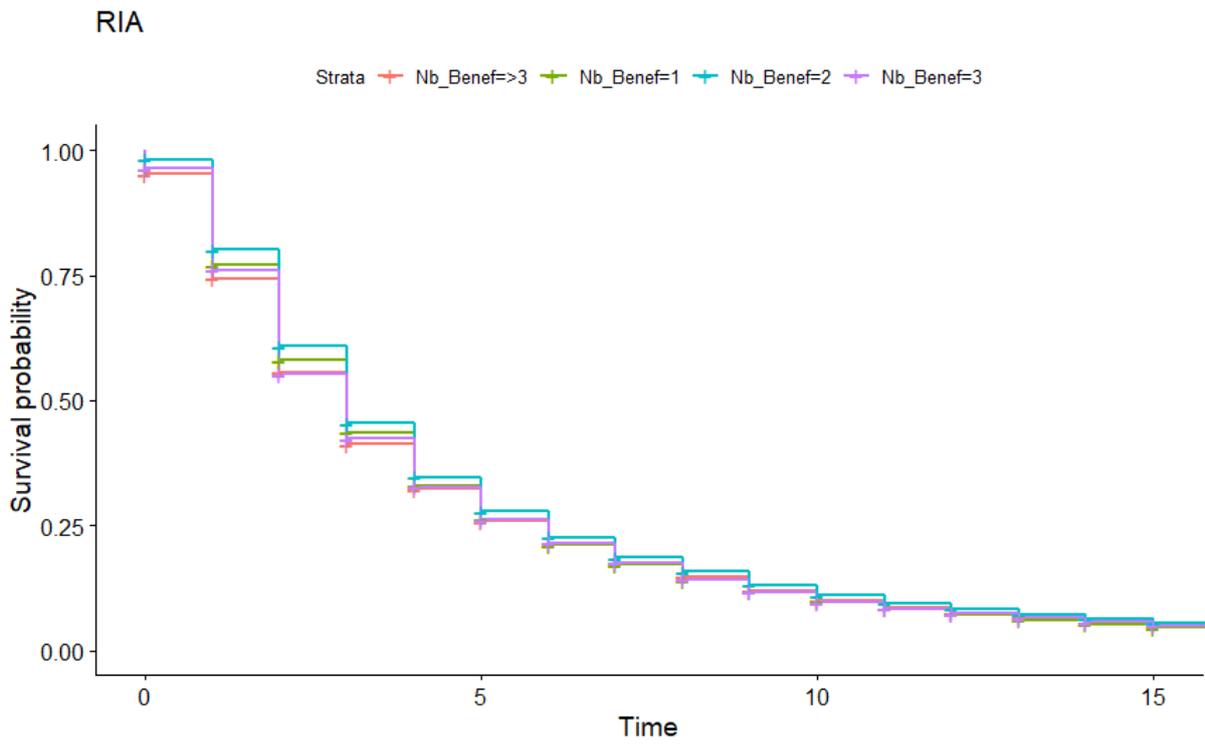


Figure 49 Courbe de survie RIA en fonction du nombre de bénéficiaires

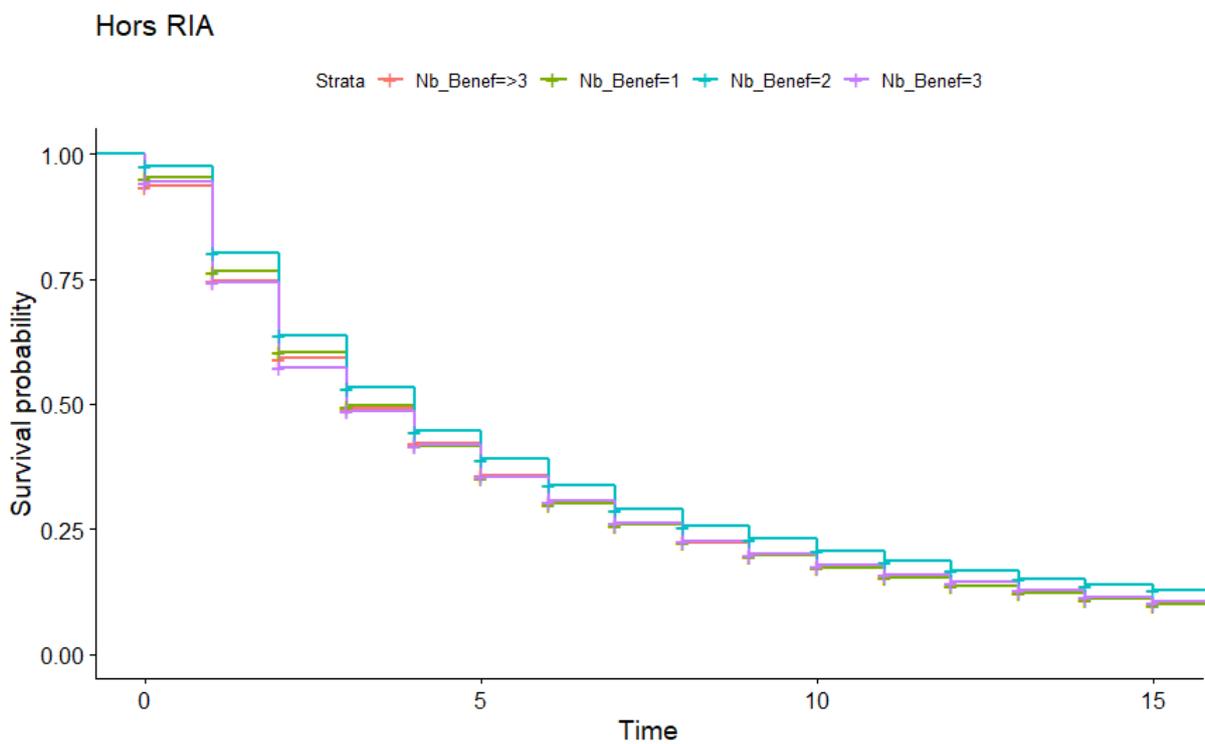


Figure 50 Courbe de survie Hors RIA en fonction du nombre de bénéficiaires

Moyenne d'âge par ancienneté

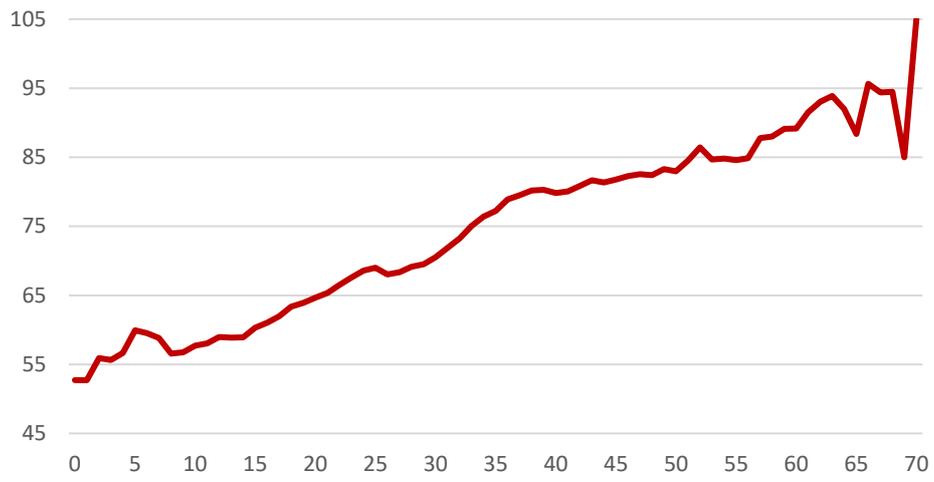


Figure 51 Moyenne d'âge par ancienneté

Table des figures

Figure 1 Source : ACPR, calculs DREES. Part des contrats collectifs et individuels dans l'ensemble des cotisations collectées en santé par les différents types d'organismes en 2020, en % de cotisations collectées. _____	21
Figure 2 Répartition du portefeuille par année - individuel _____	30
Figure 3 Taux d'extinction du portefeuille par génération individuel _____	31
Figure 4 Taux d'extinction du portefeuille collectif _____	32
Figure 5 Evolution des prestations par poste – Individuel _____	33
Figure 6 Evolution du CM/assuré en montant - Individuel _____	33
Figure 7 Evolution des prestations par poste - collectif _____	34
Figure 8 Evolution du CM/assuré - collectif _____	34
Figure 9 Evolution de la consommation dans le temps - génération 2011 individuel _____	35
Figure 10 Evolution de la consommation dans le temps - génération 2016 individuel _____	35
Figure 11 Evolution de la consommation dans le temps - génération 2011 collectif _____	36
Figure 12 Evolution de la consommation dans le temps - génération 2016 collectif _____	36
Figure 13 Evolution de la rentabilité par génération - individuel _____	37
Figure 14 Evolution de la rentabilité par génération - collectif _____	37
Figure 15 Chutes mensuelles en nombre de contrats - santé individuelle _____	40
Figure 16 Taux de chutes annuels - santé individuelle _____	41
Figure 17 Taux bruts estimés par Kaplan-Meier _____	46
Figure 18 Taux lissés par Whittaker-Henderson _____	49
Figure 19 Série temporelle des affaires nouvelles _____	53
Figure 20 Décomposition de la série des affaires nouvelles _____	54
Figure 21 Prédiction des affaires nouvelles à partir d'un modèle ETS _____	57
Figure 22 Résidus modèle ETS _____	58
Figure 23 Résidus modèle SARIMA _____	62
Figure 24 Prédiction à partir du modèle SARIMA _____	63
Figure 25 Prédictions finales à partir du modèle SARIMA optimal _____	64
Figure 26 Modèle SES - impact du choix de la variable alpha _____	65
Figure 27 Evolution des prestations et du CM/assuré depuis 2017 _____	68
Figure 28 Evolution du CM/assuré et nombre de contrats en fonction de l'âge _____	70
Figure 29 Evolution du S/P par âge _____	71
Figure 30 Répartition Homme/Femme _____	71
Figure 31 Répartition des CSP _____	72
Figure 32 Evolution du CM/assuré et nombre de contrats en fonction du nombre de bénéficiaires du contrat _____	72
Figure 33 Evolution du CM/assuré et nombre de contrats en fonction de l'ancienneté _____	73
Figure 34 Mean Excess Plot _____	74
Figure 35 Mean Excess Plot zoomé _____	75
Figure 36 Evolution du paramètre de complexité en fonction de la taille de l'arbre _____	77
Figure 37 Arbre de régression optimal _____	78
Figure 38 Représentation des résidus de déviance _____	84
Figure 39 Représentation des distances de Cook _____	85
Figure 40 Taux d'extinction du portefeuille projeté _____	89
Figure 41 Rentabilité et profitabilité projetées _____	91
Figure 42 Evolution projetée de la profitabilité du portefeuille _____	91
Figure 43 Courbe de survie RIA en fonction de l'âge _____	95
Figure 44 Courbe de survie Hors RIA en fonction de l'âge _____	95
Figure 45 Courbe de survie RIA en fonction du sexe _____	96
Figure 46 Courbe de survie Hors RIA en fonction du sexe _____	96
Figure 47 Courbe de survie RIA en fonction du type de produit cible _____	97

<i>Figure 48 Courbe de survie Hors RIA en fonction du type de produit cible</i>	<u>97</u>
<i>Figure 49 Courbe de survie RIA en fonction du nombre de bénéficiaires</i>	<u>98</u>
<i>Figure 50 Courbe de survie Hors RIA en fonction du nombre de bénéficiaires</i>	<u>98</u>
<i>Figure 51 Moyenne d'âge par ancienneté</i>	<u>99</u>

Bibliographie

Analyse de survie (larmarange.github.io) : <https://larmarange.github.io/analyse-R/analyse-de-survie.html>

FREDERIC PLANCHET [2022] : Modèles de durées

Ressources R pour les modèles de durée : ressources-actuarielles.net

THERNEAU T. [2022] : Package 'survival', Package 'rpart' | Logiciel R

Mathieu Ribatet : Modélisation paramétrique de la survie

12 Les séries temporelles | Analyse et modélisation d'agroécosystèmes (essicolo.github.io)

BRIAN MATTIS [2021] : Time Series Forecasting in R with Holt-Winters

Analysez et modélisez des séries temporelles [2020] | openclassrooms.com

PHUONG LINH [2020] : Forecast: Holt-Winters Exponential Smoothing

JAMES LONG, PAUL TEETOR [2019], R Cookbook, 2nd Edition, Time series analysis

MARIE-JEANNE VIEILLE [2016] : Un arbre de décision avec R : www.lovelyanalytics.com

LAURENT ROUVIERE [2015] : Sélection-"validation" de modèles

Dr. GUANGLIANG CHEN : Diagnostics for Leverage and Influence | Mathematics & Statistics, San José State University

ARTHUR CHARPENTIER [2013] : Actuariat IARD - ACT2040 | Partie 4 - modèles linéaires généralisés | Université du Québec à Montréal

PIERRE MARJOLLET [2020] : Nouvelles approches en tarification de produits santé pour expatriés, Mémoire d'actuariat

TIMOTHE KUEHM [2020] : Détection de la résiliation de contrats de complémentaires santé individuelles et impacts de la loi résiliation infra-annuelle, Mémoire d'actuariat

DALIA KOUKOU, Senior Consultante - GERALDINE GUITARD, Senior Manager [2021] : La résiliation infra-annuelle en santé : menace ou opportunité pour les organismes complémentaires d'assurance ? Qui seront les grands gagnants ?

GUERRIC BRAS [2020] : Quels impacts de la résiliation infra-annuelle en assurance santé ?

L'historique depuis 1945 - Ministère de la santé, 2019 | <https://solidarites-sante.gouv.fr>

Notre histoire | L'Assurance Maladie | <https://assurance-maladie.ameli.fr>

ACPR, calculs DREES. Part des contrats collectifs et individuels dans l'ensemble des cotisations collectées en santé par les différents types d'organismes en 2020, en % de cotisations collectées.

Qu'est-ce que le « 100% santé » en optique, audiologie et dentaire ? [2018] | <https://www.swisslife.fr>