



**Mémoire présenté devant l'Institut des Actuares pour l'obtention du Diplôme
Universitaire des Actuares de Strasbourg**

**DÉFINITION D'UNE VALEUR CONTRAT À LA SOUSCRIPTION POUR LES CONTRATS
D'ASSURANCE HABITATION**

Par : Mona Armanteras
Tuteurs : François Luu et Romain Toesca
Entreprise : AXA France
Ecole : DUAS

Résumé

En France, le marché de l'assurance habitation est un marché fortement concurrentiel. Pour s'adapter à cette situation, les assureurs doivent innover et développer de nouvelles stratégies pour attirer de nouveaux clients. C'est dans ce contexte qu'AXA France a mis en place une stratégie qui consiste à proposer des tarifs à la souscription suffisamment attractifs pour satisfaire ses objectifs de volumétrie d'affaires nouvelles. Les tarifs proposés ne permettent alors pas nécessairement d'être rentable à la souscription des contrats. Il est donc nécessaire de s'assurer que les contrats deviendront rentables au fil de leur vieillissement.

Afin de déterminer la rentabilité à long terme des contrats, l'indicateur de valeur contrat à la souscription peut être utilisé. Cet indicateur correspond à la valeur actuelle probable des flux de résultat générés par un contrat sur l'ensemble de sa vie. Pour construire cet indicateur, il est notamment nécessaire de disposer d'une estimation des probabilités de survie du contrat dans le portefeuille.

Cette étude consiste à la construction de l'indicateur de valeur contrat et l'analyse des résultats qui s'y rapportent. Différents exemples d'utilisation de cet indicateur sont explorés, le principal étant la détermination des profils de contrats dont la rentabilité prédite est la plus élevée, afin d'optimiser la souscription des contrats d'assurance habitation.

Mots clés

Assurance multirisque habitation, Stratégie tarifaire, Rentabilité espérée, Valeur actuelle probable, Modèle de survie, Kaplan-Meier, Cox, Gradient Boosting

Abstract

In France, the home insurance market is highly competitive. Insurers must innovate and develop new pricing strategies in order to attract new customers and maintain their profitability. Within this environment, AXA France has adopted a strategy with attractive price so that it can fulfil its objectives in terms of volume. At the subscription of the contract, the offered price is set up at a low level in order to attract the customer but remains not profitable. It is therefore necessary to ensure that contracts bring positive cashflow in the future.

To determine the long-term profitability, the contract value at the subscription can be calculated. This indicator is defined as the present value of the future result cashflows generated by a contract over its entire life. To build this indicator, it is necessary to have an estimation of the contract's survival probabilities in the portfolio.

The aims of this study are to build the contract value at the subscription and to analyze the results brought by this indicator. Different examples of the use of this value are explored, the main one is the determination of the contract profiles which have the highest predicted profitability. Therefore, it is possible to optimize the underwriting strategy of the home insurance policies.

Keywords

Multi-risk home insurance, Pricing strategy, Expected profitability, Present value, Survival model, Kaplan-Meier, Cox, Gradient Boosting

Note de synthèse

Contexte de l'étude

L'objectif de ce mémoire est de construire un indicateur de prédiction de la rentabilité future des contrats d'assurance multirisque habitation.

La forte concurrence sur le marché de l'assurance habitation a mené AXA France à revoir à la baisse ses tarifs à la souscription afin de proposer des tarifs suffisamment attractifs pour satisfaire ses objectifs de volumétrie d'affaires nouvelles. Les contrats ne sont alors pas nécessairement rentables à leur souscription, mais leurs primes sont majorées annuellement afin d'améliorer leur rentabilité au fil de leur vieillissement.

Disposer d'un indicateur correspondant à la prédiction de la rentabilité future des contrats, sur l'ensemble de leur vie, vue à la souscription, permettrait d'évaluer la pertinence de cette stratégie. Il serait possible d'estimer si une nouvelle génération de contrats sera rentable sur l'ensemble de sa vie, ou si des contrats en particulier le seront.

Définition de l'indicateur de valeur contrat à la souscription

La valeur contrat à la souscription correspond à l'estimation, vue à la souscription, de la rentabilité future d'un contrat, c'est à dire à la valeur actuelle probable des flux de résultat futurs du contrat. La valeur contrat d'un contrat x s'exprime donc ainsi :

$$\text{Valeur contrat}(x) = \sum_{i=0}^{\infty} \mathbb{E}(\text{Résultat}_i(x)) \times \left(\frac{1}{1+\text{taux actualisation}_i}\right)^i$$

A chaque année de vie d'un contrat, le résultat va correspondre à la marge dégagée sur l'année, associée à ce contrat. Cette marge va dépendre de la valeur de la prime commerciale payée par l'assuré, de la prime pure associée au contrat pour l'année, des frais dépensés par la compagnie pour la mise en place et la gestion du contrat sur l'année, du montant des commissions associées au contrat sur l'année, et du coût de la réassurance associé au contrat pour l'année. Le résultat d'un contrat x sur une année i se calcule de la manière suivante :

$$\text{Résultat}_i(x) = \text{Prime commerciale}_i(x) - \text{Prime pure}_i(x) - \text{Frais}_i(x) - \text{Commissions}_i(x) - \text{Réassurance}_i(x).$$

Alors :

$$\begin{aligned} \text{Valeur contrat}(x) = & \sum_{i=0}^{\infty} (\mathbb{E}(\text{Prime commerciale}_i(x)) - \mathbb{E}(\text{Prime pure}_i(x)) - \mathbb{E}(\text{Frais}_i(x)) \\ & - \mathbb{E}(\text{Commissions}_i(x)) - \mathbb{E}(\text{Réassurance}_i(x))) \times \left(\frac{1}{1 + \text{taux actualisation}_i}\right)^i \end{aligned}$$

Pour déterminer les espérances des valeurs annuelles du résultat, et de ses composantes, il faut considérer et probabiliser les différentes situations dans lesquelles peut se retrouver chaque contrat à chaque instant de sa vie. Deux situations évidentes sont à considérer : la première est la situation dans laquelle le contrat a été résilié, et la seconde celle dans laquelle il est encore en cours. Dans la situation où le contrat serait encore en cours, deux nouvelles situations sont définies : celle dans laquelle le risque assuré n'a pas changé, et celle dans laquelle le risque a évolué. Pour estimer si le risque d'un contrat a évolué, un seul type d'évolution est

considéré, le changement de qualité de l'assuré (passage de locataire à propriétaire, ou inversement), qui est l'évolution ayant le plus d'impact sur la rentabilité d'un contrat. Il est supposé qu'un seul remplacement pour changement de qualité peut avoir lieu dans la vie d'un contrat. A chaque âge, trois situations possibles du contrat sont donc considérées :

- le contrat a été résilié depuis sa souscription ;
- le contrat a connu un remplacement pour changement de qualité depuis sa souscription ;
- le contrat n'a vécu aucun changement depuis sa souscription (n'a été ni résilié, ni remplacé).

La possible survenance des situations, au fil des moments de vie d'un contrat, peut être représentée visuellement :

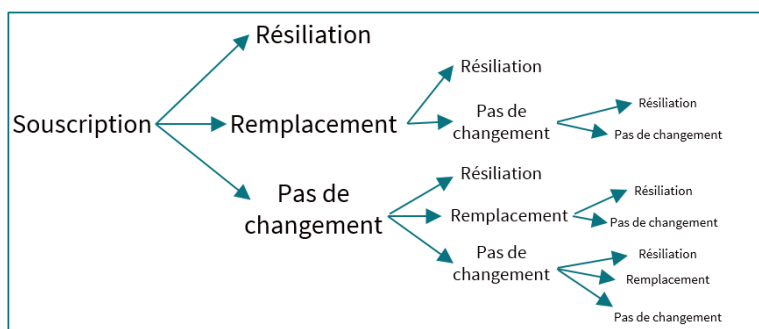


FIGURE - Représentation visuelle des situations dans lesquelles peuvent se retrouver les contrats

Les étapes de construction de l'indicateur

Détermination des probabilités de résiliation

Pour probabiliser les situations dans lesquelles peut se retrouver chaque contrat, il est d'abord nécessaire de disposer d'une estimation des probabilités de résiliation du contrat à chaque âge. Pour cela un modèle de survie des contrats est utilisé. Un modèle de Cox, qui est un modèle semi-paramétrique, a pu être construit sur 15 ans, pour prédire les probabilités de survie des contrats sur les 15 premières années de leur vie, en fonction de leurs caractéristiques à la souscription. Les contrats d'assurance habitation pouvant subsister au-delà de 15 ans, en particulier les contrats de propriétaires, il était nécessaire de prolonger l'estimation de la survie des contrats au-delà de 15 ans. Les courbes de survie des contrats, estimées par le modèle de Cox, ont été prolongées en utilisant des estimations de Kaplan-Meier des courbes de survie des contrats jusqu'à 50 ans, selon leur segment MRH (les segments MRH distinguent les contrats selon la qualité de l'assuré, selon le type de résidence, et selon le nombre de pièces du logement).

Pour chaque nouveau contrat d'assurance MRH, les probabilités de résiliation du contrat à chaque âge, jusqu'à 50 ans, sont estimées selon ses caractéristiques à la souscription.

Détermination des probabilités de remplacement

Pour évaluer les probabilités que chaque contrat se retrouve dans la situation où il aurait connu un remplacement pour changement de qualité, un modèle de survie des contrats, où le remplacement est considéré comme évènement de "décès", est utilisé. Un autre modèle de Cox a été construit sur 15 ans, pour prédire les probabilités de survie des contrats avant remplacement sur les 15 premières années de leur vie, selon leurs caractéristiques à la souscription. Au delà de 15 ans la survenance de remplacements a été négligée.

Pour chaque nouveau contrat d'assurance MRH, les probabilités de remplacement pour changement de qualité du contrat à chaque âge, jusqu'à 15 ans, sont estimées selon ses caractéristiques à la souscription.

Détermination des flux de résultat

Pour déterminer l'espérance des flux de résultat des contrats à chaque âge, il est nécessaire d'estimer la valeur du résultat dans chaque situation dans laquelle peuvent se retrouver les contrats, à chaque âge. Soit, pour un contrat, à chaque âge i :

- ${}_{i+1}q_0$, la probabilité de résiliation sur $[0; i + 1[$;

-
- ${}_{i+1}\hat{r}_0$, la probabilité de remplacement sur $[0; i + 1[$;
 - ${}_{i+1}\hat{p}_0 = 1 - {}_{i+1}\hat{q}_0 - {}_{i+1}\hat{r}_0$, la probabilité que la situation ne change pas sur $[0; i + 1[$.

L'espérance du résultat d'un contrat x à l'âge i se détermine ainsi :

$$\mathbb{E}(Résultat_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Résultat(\text{après remplacement})_i(x) + {}_{i+1}p_0(x) \times Résultat_i(x)$$

Chaque composante du résultat est donc déterminée à chaque âge dans le cas où le contrat n'a pas connu de changement, et dans le cas où il a connu un remplacement pour changement de qualité.

Pour l'évaluation de la prime commerciale au fil des années, l'application annuelle du taux de majoration au terme est prise en compte, en supposant que ce taux est fixe au fil du temps. Et un taux moyen d'évolution de la prime commerciale en cas de remplacement est estimé, selon le segment MRH du contrat.

Pour l'évaluation de la prime pure au fil des années, un taux annuel d'évolution naturelle de la prime pure au fil du temps est estimé et est pris en compte. Ce taux reflète la tendance d'évolution de la fréquence de sinistralité, ou la tendance d'évolution du coût de réparation des sinistres. Et un taux moyen d'évolution de la prime pure en cas de remplacement est estimé, selon la qualité de l'assuré du contrat.

Les frais sont divisés en frais fixes, qui sont des montants uniques associés à tous les contrats chaque année, et en frais variables, qui sont des taux applicables à la prime commerciale annuelle de chaque contrat. Le montant des frais dans chaque situation dépend donc du montant de la prime commerciale dans la situation. Il est supposé que les montants et taux de frais n'évolueront pas.

Les commissions associées au contrat correspondent à un taux applicable à la prime commerciale annuelle de chaque contrat. Le montant des commissions dans chaque situation dépend donc du montant de la prime commerciale dans la situation. Il est supposé que le taux de commissions n'évoluera pas.

Le coût de la réassurance associé au contrat correspond également à un taux applicable à la prime commerciale annuelle de chaque contrat. Le montant du coût de la réassurance dans chaque situation dépend donc du montant de la prime commerciale dans la situation. Il est supposé que le taux de coût de la réassurance n'évoluera pas.

Dans la définition de l'indicateur de valeur contrat à la souscription, les flux de résultat futurs des contrats sont actualisés pour être vus à la souscription des contrats. Les taux d'actualisation utilisés sont définis par les hypothèses financières fixées par AXA France.

Analyse des résultats

La valeur contrat des affaires nouvelles observées entre janvier et juin 2021 a été estimée, permettant ainsi un aperçu de la rentabilité prédite de cette nouvelle génération de contrats.

L'analyse globale des résultats obtenus permet de voir les différences de rentabilité des contrats selon leur segment MRH :

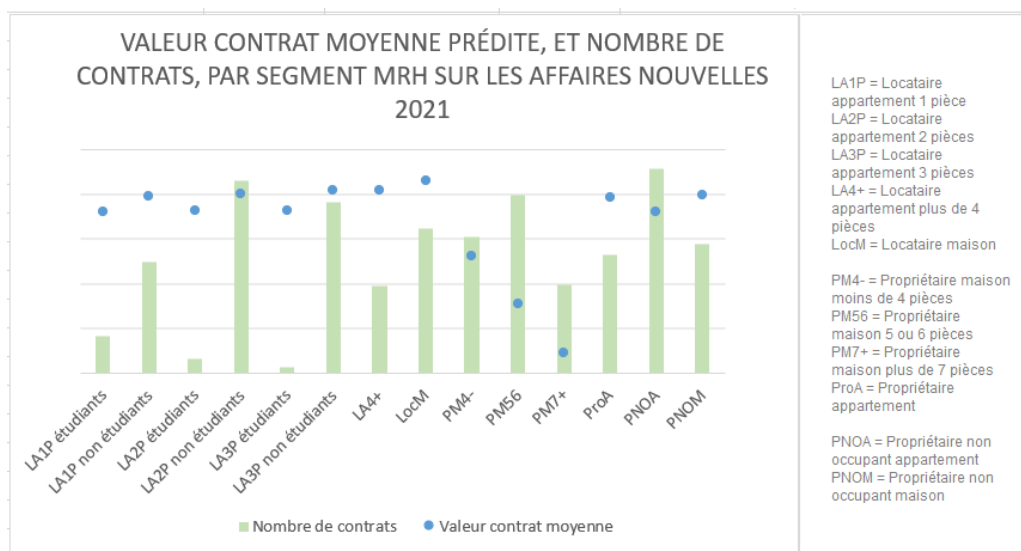


FIGURE - Valeur contrat moyenne prédite par segment MRH sur les affaires nouvelles de 2021

Cela permet d'observer, notamment, que les contrats de propriétaires de maisons sont les moins rentables et que les contrats de locataires de maisons sont les plus rentables. Il apparaît aussi que les contrats étudiants de locataires sont moins rentables que les contrats non-étudiants de locataires, ce qui est logique, la formule étudiante étant une offre d'assurance avantageuse pour les étudiants.

Il est également possible d'observer l'évolution prédite au fil des années de la valeur contrat globale cumulée, sur un contrat ou un ensemble de contrats. Il est possible de voir ainsi si la rentabilité des contrats a bien tendance à s'améliorer au fil des années, comme le veut la stratégie tarifaire consistant à majorer la prime des contrats chaque année. Il est visible, notamment, que la stratégie fonctionne bien pour les contrats de locataires non-étudiants :

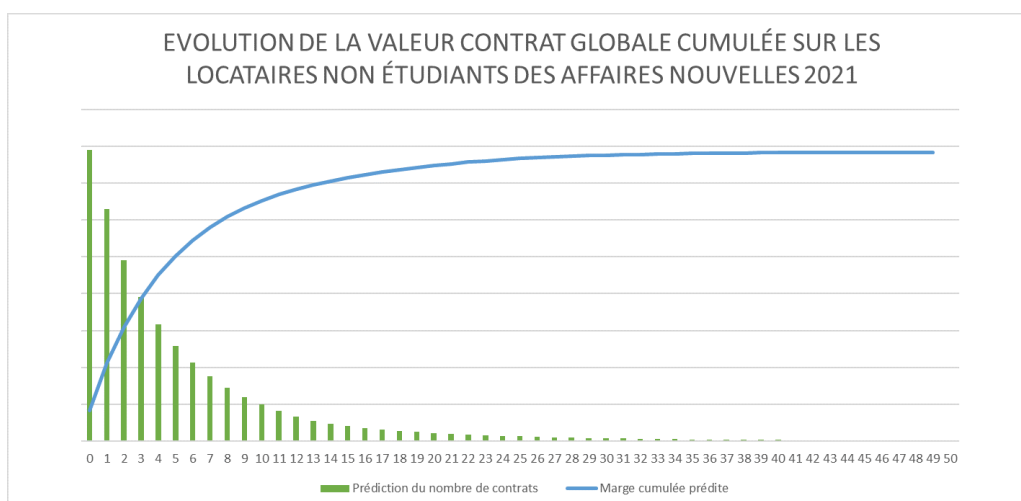


FIGURE - Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des locataires non-étudiants

La première année, les contrats de locataires non-étudiants ne sont pas rentables, et ils ont tendance à devenir rentables quelques années plus tard.

Exploitation des résultats

Les résultats obtenus par la prédiction de la valeur contrat à la souscription peuvent être utilisés dans différentes applications.

L'étude de la valeur contrat peut permettre de définir des stratégies de souscription qui seraient bénéfiques pour le résultat de la compagnie. Cela consisterait à définir des profils de contrats cibles, dont la rentabilité attendue est meilleure que celle d'autres profils, à souscrire en priorité. Il pourrait par exemple être décidé de faire des opérations commerciales, ou des campagnes marketing, ayant pour objectif d'augmenter la souscription de contrats de locataires de maisons, qui est le segment des contrats les plus rentables, d'après l'étude de la valeur contrat prédite des affaires nouvelles de 2021.

L'observation de la valeur contrat peut également permettre d'établir de nouvelles stratégies tarifaires. Il est par exemple possible de calculer un taux à appliquer à la prime commerciale à l'affaire nouvelle des contrats, selon le segment MRH par exemple, qui permettrait d'obtenir une valeur cible de rentabilité sur l'ensemble du segment.

Conclusion

L'indicateur de valeur contrat à la souscription permet de prédire la rentabilité future des contrats dès leur souscription. Il peut permettre de savoir à l'avance quels contrats devraient être les plus rentables et lesquels seraient les moins bénéfiques au résultat de la compagnie. Il peut ainsi permettre d'élaborer des stratégies de souscription. Il peut également guider des stratégies tarifaires. Voir quels profils de contrats sont les moins rentables permettrait de savoir quels types de contrats en priorité devraient voir leurs tarifs augmenter.

Summary

Context of the study

The objective of this dissertation is to construct a predictive indicator of the future profitability of multi-risk home insurance contracts.

Strong competition in the home insurance market has led AXA France to revise downwards its underwriting rates in order to offer underwriting rates enough attractive to meet its new contracts volume target. The contracts are not necessarily profitable at the time of subscription, but their premiums are annually increased to improve their profitability as they age.

Having an indicator corresponding to the prediction of the future profitability of the contracts, over their entire life, seen at the time of subscription, would allow to evaluate the relevance of this strategy. It would be possible to estimate whether a new generation of contracts will be profitable over its lifetime, or whether specific contracts will be profitable.

Definition of the contract value at subscription indicator

The contract value at subscription indicator is the estimate, seen at subscription, of the future profitability of a contract, which is the present value of the contract's future result flows. The contract value of a contract x is expressed as :

$$\text{Contract value}(x) = \sum_{i=0}^{\infty} \mathbb{E}(\text{Result}_i(x)) \times \left(\frac{1}{1+\text{discount rate}_i}\right)^i$$

Every year of the contract's life, the result will be the margin generated over the year, associated with this contract. This margin's value will depend on the value of the commercial premium paid by the policyholder, the pure premium associated with the policy for the year, the costs incurred by the company in setting up and managing the policy for the year, the amount of commission associated with the policy for the year, and the cost of reinsurance associated with the policy for the year. The result of a contract x for the year i is calculated as follows :

$$\text{Result}_i(x) = \text{Commercial premium}_i(x) - \text{Pure premium}_i(x) - \text{Costs}_i(x) - \text{Commission}_i(x) - \text{Reinsurance}_i(x).$$

Then :

$$\begin{aligned} \text{Contract value}(x) = \sum_{i=0}^{\infty} (\mathbb{E}(\text{Commercial premium}_i(x)) - \mathbb{E}(\text{Pure premium}_i(x)) - \mathbb{E}(\text{Costs}_i(x)) \\ - \mathbb{E}(\text{Commission}_i(x)) - \mathbb{E}(\text{Reinsurance}_i(x))) \times \left(\frac{1}{1 + \text{discount rate}_i}\right)^i \end{aligned}$$

In order to determine the expected annual values of the result and its components, the different situations in which each contract may be at each moment of its life must be considered and probabilsed. There are two obvious situations to consider : the first is the one where the contract has been terminated, and the second is the one where it is still active. In the situation where the contract is still active, two new situations are defined : one where the insured risk has not changed, and one where the risk has changed. To estimate whether the risk of a contract has changed, only one type of change is considered, the change in

the status of the insured (from tenant to owner, or conversely), which is the change that has the greatest impact on the profitability of a contract. It is assumed that only one change in status can take place in a contract's life. At each age, three possible contract situations are therefore considered :

- the contract has been terminated since its inception ;
- the contract has been replaced due to a change in status since its inception ;
- the contract has not undergone any change since its inception (has not been terminated or replaced).

It is possible to visually represent the possible occurrence of situations throughout a contract's life :

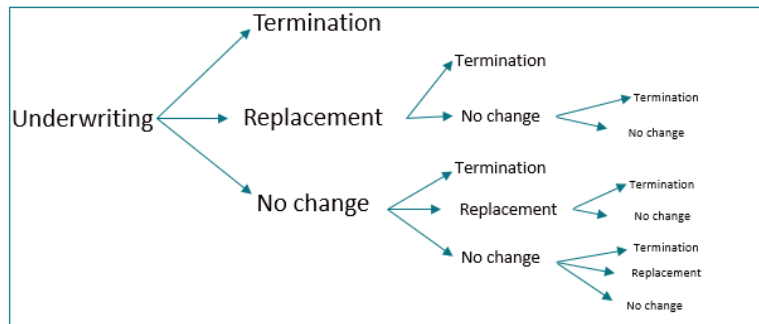


FIGURE - Visual mapping of the situations in which the contracts may be

Stages in the construction of the indicator

Determining the probability of termination

In order to probabilise the situations in which each contract may be, it is first necessary to have an estimate of the contract's probabilities of termination at each age. For this purpose a contract survival model is used. A Cox model, which is a semi-parametric model, could be constructed over 15 years to predict the survival probabilities of contracts over the first 15 years of their life, depending on their characteristics at the time of the underwriting. As home insurance policies can survive beyond 15 years, especially homeowners' policies, it was necessary to extend the estimate of the contracts' survival beyond 15 years. The contracts' survival curves, estimated by the Cox model, were extended using Kaplan-Meier estimates of 50-year survival curves of the contracts according to their MRH segment (the MRH segments distinguish contracts according to the policyholder's status, according to the type of residence, and according to the number of rooms in the dwelling).

For each new home insurance contract, the likelihood of the policy being terminated at each age up to 50 is estimated according to its characteristics at the time of the underwriting.

Determining the replacement probabilities

To assess the probabilities that each contract would be in the situation where it would have experienced change in status, a pre-replacement contract survival model is used. Another 15-year Cox model was constructed to predict the contracts' survival probabilities before replacement over the first 15 years of their life, depending on their characteristics at the time of the underwriting. Beyond 15 years, the occurrence of replacements has been overlooked.

For each new home insurance contract, the likelihood of the policy being replaced at each age up to 15 years is estimated according to its characteristics at the time of underwriting.

Determining the result flows

In order to determine the contracts' expected flows of result at each age, it is necessary to estimate the value of the result in each situation in which the contracts may be, at each age. Let, for a contract, at each age i :

- ${}_{i+1}\hat{q}_0$, the termination probability over $[0; i + 1[$;
- ${}_{i+1}\hat{r}_0$, the replacement probability over $[0; i + 1[$;
- ${}_{i+1}\hat{p}_0 = 1 - {}_{i+1}\hat{q}_0 - {}_{i+1}\hat{r}_0$, the probability that the situation does not change over $[0; i + 1[$.

The result's expectation of a contract x at age i is determined as follows :

$$\mathbb{E}(Result_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Result(after\ replacement)_i(x) + {}_{i+1}p_0(x) \times Result_i(x)$$

Each component of the result is therefore determined at each age in the case where the contract has not undergone a change, and in the case where it has undergone a replacement for change in status.

For the assessment of the commercial premium over the years, the annual application of the gross-up rate at term is taken into account, assuming that this rate is fixed over time. And an average rate of change in the commercial premium in case of replacement is estimated, depending on the MRH segment of the contract.

For the assessment of the pure premium over time, an annual rate of natural change in the pure premium over time is estimated and taken into account. This rate reflects the trend in the frequency of claims, or the trend in the cost of repairing claims. And an average rate of change in the pure premium in case of replacement is estimated, depending on the policyholder's status.

The costs are divided into fixed costs, which are single amounts associated with all contracts each year, and variable costs, which are rates applicable to the annual commercial premium of each contract. The amount of the costs in each situation therefore depends on the amount of the commercial premium in the situation. It is assumed that the amounts and rates of costs will not change.

The commissions associated with the contract correspond to a rate applicable to the annual commercial premium of each contract. The amount of commission in each situation therefore depends on the amount of the commercial premium in the situation. It is assumed that the commission rate will not change.

The cost of reinsurance associated with the contract is also a rate applicable to the annual commercial premium of each contract. The amount of the reinsurance cost in each situation therefore depends on the amount of the commercial premium in the situation. It is assumed that the reinsurance cost rate will not change.

In the definition of the contract value at subscription indicator, the future result flows of the contracts are discounted to be seen at the underwriting of the contracts. The discount rates used are defined by the financial assumptions set by AXA France.

Analysis of the results

The contract value of the new contracts observed between January and June 2021 has been estimated, allowing an overview of the predicted profitability of this new generation of contracts.

The overall analysis of the results obtained shows the differences in the profitability of contracts according to their MRH segment :



FIGURE - Average contract value predicted, by MRH segment, on the new contracts of 2021

In particular, it can be seen that homeowners' policies are the least profitable and home renters' policies are the most profitable. It also appears that student tenant contracts are less profitable than non-student tenant contracts, which is logical, as the student policy is an advantageous insurance offer for students.

It is also possible to observe the predicted evolution over the years of the overall cumulative contract value, on a contract or a group of contracts. It is possible to see whether the profitability of the contracts tends to improve over the years, as the pricing strategy of increasing the premium of the contracts each year requires. In particular, it can be seen that the strategy works well for the non-students tenant contracts :

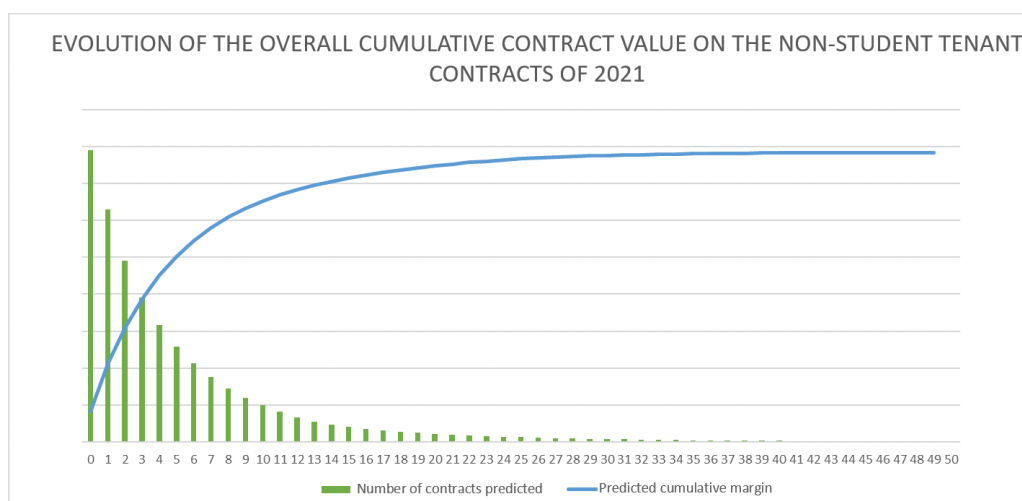


FIGURE - Predicted evolution of the overall cumulative contract value on the non-student tenant new contracts of 2021

The non-student tenant contracts are not profitable in the first year, and they tend to become profitable a few years later.

Exploitation of the results

The results obtained from the prediction of the contract value at subscription can be used in different applications.

The study of the contract value can help to define underwriting strategies that would be beneficial to the company's result. This would consist of defining target contract profiles, whose expected profitability

is better than the other profiles' expected profitability, to be underwritten as a priority. For example, it could be decided to run commercial operations or marketing campaigns aimed at increasing the number of subscriptions of house tenant's contracts, which is the most profitable contract segment, according to the study of the predicted contract value of the new contracts of 2021.

The observation of the contract value can also make it possible to establish new pricing strategies. For example, it is possible to calculate a rate to be applied to the commercial premium at subscription of the contracts, depending on the MRH segment for example, which would allow to obtain a target profitability value for the entire segment.

Conclusion

The contract value at subscription indicator allows the future profitability of contracts to be predicted as soon as they are subscribed. It can help to identify in advance which contracts are likely to be the most profitable and which would be the least beneficial to the company's result. It can thus help develop underwriting strategies. It can also guide pricing strategies. Seeing which contract profiles are the least profitable can help to identify which types of contracts should be prioritised for tariff increases.

Remerciements

Je tiens à remercier en premier lieu mon tuteur en entreprise, Romain Toesca. Je le remercie pour l'encadrement très complet dont j'ai bénéficié, pour sa disponibilité et l'attention accordée pour la réalisation de ce mémoire. Je le remercie également de m'avoir conseillée tout au long de cette étude.

Je tiens également à remercier François Luu, responsable de l'équipe de tarification MRH, pour m'avoir suivie pendant toute la durée de cette étude, et m'avoir apporté des recommandations qui ont été très utiles à l'avancée de mon travail.

Je remercie également Hugo Hammerer, *data scientist* dans l'équipe de tarification MRH, pour l'aide qu'il m'a apportée et les conseils qu'il m'a donnés dans la réalisation de cette étude.

Je remercie aussi les autres membres de l'équipe de tarification MRH pour leur disponibilité et leur bienveillance.

Je souhaite également remercier Julien Durand, responsable du service Multirisque habitation, et tous les autres membres du service, au sein duquel j'ai pu évoluer dans d'excellentes conditions pour réaliser ce mémoire.

Enfin, je remercie mon tuteur académique, monsieur Jean Berard, pour le suivi lors de cette étude, et les conseils donnés.

Table des matières

Résumé	1
Abstract	2
Note de synthèse	3
Summary	8
Table des figures	17
Liste des tableaux	20
Introduction	22
1 Contexte de l'étude	24
1.1 Le marché de l'assurance Multirisque Habitation	24
1.1.1 Le marché de l'assurance IARD	24
1.1.2 L'assurance multirisque habitation	25
1.1.3 AXA France sur le marché de l'assurance multirisque habitation	26
1.2 Les problématiques de l'étude	29
1.2.1 La stratégie d'AXA France	29
1.2.2 Les indicateurs de suivi de la stratégie	29
1.2.3 Nécessité d'un indicateur de la valeur contrat	30
1.2.4 La construction de l'indicateur de la valeur contrat	31
1.3 Présentation de la base Image MRH	33
1.3.1 La construction de la base	33
1.3.2 Présentation des variables principales de la base	33
1.4 Synthèse du chapitre	35
2 Modélisation non-paramétrique de la probabilité de résiliation	36
2.1 Études statistiques descriptives	36
2.1.1 Description du portefeuille	36
2.1.2 Caractéristiques des résiliations	39
2.1.3 Synthèse de l'analyse descriptive	44
2.2 Modélisation préliminaire : Estimation des courbes de survie par Kaplan Meier	45
2.2.1 Observation des données de survie	45
2.2.2 Estimation des courbes de survie avec Kaplan-Meier	47
2.2.3 Lissages non-paramétriques des taux annuels de mortalité	50
2.2.4 Analyse des résultats	51
2.3 Synthèse du chapitre	55

3	Modélisation semi-paramétrique de la probabilité de résiliation	56
3.1	Construction d'une nouvelle base de données	56
3.1.1	Définition de la base	56
3.1.2	Description des variables	57
3.1.3	Traitement de la base de données	58
3.1.3.1	Traitement des valeurs manquantes	58
3.1.3.2	Traitement des valeurs aberrantes	59
3.1.3.3	Étude des corrélations entre variables	60
3.2	Modélisation semi-paramétrique : Modèle de Cox	63
3.2.1	Définition du modèle	63
3.2.2	Vérification de l'hypothèse de proportionnalité des risques instantanés	64
3.2.3	Modélisation classique	66
3.2.3.1	Présentation de la méthodologie	66
3.2.3.2	Implémentation informatique du modèle	66
3.2.3.3	Sélection des variables	66
3.2.3.4	Étude de la signification des coefficients	67
3.2.3.5	Construction d'un modèle sur les âges entiers	67
3.2.3.6	Construction d'un modèle sur les âges exacts	69
3.2.3.7	Prolongation des courbes de survies estimées par le modèle	74
3.2.4	Modélisation avec utilisation d'un algorithme de <i>Gradient Boosting</i>	76
3.2.4.1	Présentation de l'algorithme de <i>Gradient Boosting</i>	76
3.2.4.2	Implémentation informatique du modèle	77
3.2.4.3	Construction d'un modèle sur les âges exacts	77
3.2.4.4	Prolongation des courbes de survies estimées par le modèle	80
3.2.5	Choix du modèle	82
3.3	Synthèse du chapitre	83
4	Modélisation de la probabilité de remplacement	84
4.1	Étude statistique descriptive préalable	84
4.2	Construction d'une nouvelle base de données	88
4.3	Modélisation semi-paramétrique : Modèle de Cox	90
4.3.1	Vérification de l'hypothèse de proportionnalité des risques instantanés	90
4.3.2	Modélisation avec utilisation d'un algorithme de <i>Gradient Boosting</i>	90
4.3.2.1	Préférence pour la modélisation avec utilisation du <i>Gradient Boosting</i>	90
4.3.2.2	Construction d'un modèle sur les âges exacts	91
4.4	Synthèse du chapitre	99
5	Construction de la valeur contrat à la souscription	100
5.1	Définition de l'indicateur de valeur contrat	100
5.1.1	Probabilisation des situations dans lesquelles peuvent se trouver les contrats	100
5.1.2	Définition de l'espérance de la prime commerciale	101
5.1.3	Définition de l'espérance de la prime pure	102
5.1.4	Définition de l'espérance des frais	104
5.1.5	Définition de l'espérance des commissions	104
5.1.6	Définition de l'espérance du coût de la réassurance	105
5.1.7	Définition du taux d'actualisation	105
5.2	Étude des valeurs contrat obtenues sur les affaires nouvelles de 2021	106
5.3	Tests de sensibilité	112
5.3.1	Suppression du risque de remplacement	112
5.3.2	Modification du taux d'évolution de la prime pure	112
5.3.3	Modification des probabilités de résiliation	113
5.3.4	Définition d'une contrainte de résiliation des contrats à un âge maximal de 15 ans	115
5.4	Possibles évolutions de la prédiction de l'indicateur de valeur contrat	116
5.5	Synthèse du chapitre	117

Conclusion	118
Bibliographie	121
Annexes	122

Table des figures

1.1	Répartition des cotisations perçues sur le marché IARD en 2019 en France d'après la Fédération Française de l'Assurance [11]	24
1.2	Cotisations perçues sur le marché IARD et MRH par année en France d'après la Fédération Française de l'Assurance	26
1.3	Nombre de contrats à chaque fin d'année, au sein du portefeuille MRH d'AXA France	27
1.4	Nombre d'affaires nouvelles et de résiliations par an, au sein du portefeuille MRH d'AXA France	27
1.5	Chiffre d'affaire TTC annuel du portefeuille MRH d'AXA France	28
2.1	Nombre de contrats dans le portefeuille MRH d'AXA France	36
2.2	Part des différents produits MRH en nombre de contrats	37
2.3	Répartition des contrats selon leur segmentation MRH	37
2.4	Age moyen des contrats en portefeuille par segment MRH (vue à fin août 2020)	38
2.5	Evolution de l'âge moyen des contrats en portefeuille de 2016 à 2020 (vue à fin août)	38
2.6	Distribution des âges des contrats en portefeuille (vue à fin août 2020)	39
2.7	Prime moyenne des contrats en portefeuille par segment MRH (vue à fin août 2020)	39
2.8	Age moyen des contrats résiliés au moment de leur résiliation par segment MRH (vue à fin août 2020)	40
2.9	Age moyen des contrats résiliés au moment de leur résiliation, selon l'âge des assurés à la souscription (vue à fin août 2020)	40
2.10	Age moyen des contrats résiliés au moment de leur résiliation, selon la région AXA (vue à fin août 2020)	41
2.11	Age moyen des contrats résiliés au moment de leur résiliation, selon la catégorie socio-professionnelle des assurés (vue à fin août 2020)	41
2.12	Distribution des âges des contrats résiliés au moment de leur résiliation (vue à fin août 2020)	42
2.13	Prime moyenne à la résiliation des contrats résiliés par segment MRH (vue à fin août 2020)	42
2.14	Part de contrats résiliés à leur terme parmi les contrats résiliés (vue à fin décembre)	43
2.15	Part de contrats résiliés à leur premier terme parmi ceux résiliés à leur terme (vue à fin décembre)	44
2.16	Présentation d'un extrait de la table regroupant les données observées	46
2.17	Présentation de la visualisation des différents types de données	47
2.18	Tracé des probabilités de survie estimées à chaque âge et par année d'observation	48
2.19	Tracé des fonctions de survie estimées par année d'observation	49
2.20	Tracé de la fonction de survie estimée globale	49
2.21	Tracé de la courbe des \hat{q}_t bruts et de la courbe des \hat{q}_t lissée	51
2.22	Exemple de l'observation de la vie d'un contrat, par rapport au déroulement réel de sa vie	52
2.23	Distribution des résiliations au sein des 12 mois suivants l'anniversaire des contrats	52
2.24	Tracé final de la fonction de survie globale des contrats estimée	53
2.25	Tracé final des fonctions de survie des contrats estimées selon le segment MRH	54

3.1	Histogramme de la variable indiquant le dernier âge observé des contrats non-censurés . . .	57
3.2	Étude des corrélations entre les variables quantitatives de la base de données	61
3.3	Test de proportionnalité pour la variable indicatrice d'un contrat étudiant	65
3.4	Courbes de survie prédites, par le modèle de Cox sur les âges entiers, moyennes par segment MRH	68
3.5	Comparaison entre la courbe de survie moyenne prédite par le modèle de Cox sur 15 ans, la courbe de survie de Kaplan-Meier estimée sur 15 ans, et la courbe de survie de Kaplan-Meier estimée sur la base Image sur 50 ans, sur l'ensemble des contrats	69
3.6	Courbes de survie prédites, par le modèle de Cox sur les durées de vie exactes, moyennes par segment (à gauche) et selon la présence de formule étudiante et par segment MRH (à droite)	70
3.7	Comparaison entre la courbe de survie moyenne prédite par le modèle de Cox sur les âges exacts sur 15 ans, celle prédite par le modèle sur les âges entiers, la courbe de survie de Kaplan-Meier estimée sur 15 ans, et la courbe de survie de Kaplan-Meier estimée sur la base Image sur 50 ans, sur l'ensemble des contrats	70
3.8	Mise en place de la méthode de validation croisée	73
3.9	Valeurs de la concordance des prédictions des modèles de Cox classique construits par validation croisée, sur dix sous-ensembles de validation de la base de données	73
3.10	Comparaison, pour chaque segment MRH, entre la courbe de survie moyenne prédite par Cox classique et prolongée, et la courbe de survie estimée par Kaplan-Meier sur la base Image	75
3.11	Courbes de survie prédites, par le modèle de Cox <i>Gradient Boosting</i> sur les durées de vie exactes, moyennes par segment MRH (à gauche) et selon la présence de la formule étudiante et par segment MRH (à droite)	78
3.12	Impact de la valeur des variables sur le risque instantané relatif de résiliation des contrats d'après le modèle de Cox <i>Gradient Boosting</i>	79
3.13	Impact de la valeur des variables sur le risque instantané relatif de résiliation d'un premier contrat d'après le modèle de Cox <i>Gradient Boosting</i>	79
3.14	Impact de la valeur des variables sur le risque instantané relatif de résiliation d'un second contrat d'après le modèle de Cox <i>Gradient Boosting</i>	80
3.15	Valeurs de la concordance des prédictions des modèles de Cox <i>Gradient Boosting</i> construits par validation croisée, sur dix sous-ensembles de validation de la base de données	80
3.16	Comparaison, pour chaque segment MRH, entre la courbe de survie moyenne prédite par Cox <i>Gradient Boosting</i> et prolongée, et la courbe de survie estimée par Kaplan-Meier sur la base Image	81
4.1	Marge moyenne des contrats, par segment, sur leur deuxième année de vie	84
4.2	Répartition des remplacements pour changement de qualité au sein des segments MRH, par rapport à la répartition des contrats en portefeuille au sein des segments MRH	85
4.3	Répartition des remplacements pour changement de qualité au sein des régions AXA, par rapport à la répartition des contrats en portefeuille au sein des régions AXA	86
4.4	Répartition des remplacements pour changement de qualité au sein des tranches d'âge des clients, par rapport à la répartition des contrats en portefeuille au sein des tranches d'âge des clients	86
4.5	Répartition des remplacements pour changement de qualité au sein des catégories socio-professionnelles des clients, par rapport à la répartition des contrats en portefeuille au sein des catégories socio-professionnelles des clients	87
4.6	Histogramme de la variable indiquant le dernier âge observé des contrats non-censurés	88
4.7	Test de proportionnalité pour la variable indicatrice d'un contrat étudiant	90
4.8	Courbes de survie prédites, par le modèle de Cox <i>Gradient Boosting</i> sur les durées de vie exactes avant remplacement, moyennes par segment MRH (à gauche) et selon la présence de la formule étudiante et par segment MRH (à droite)	92
4.9	Comparaison entre la courbe de survie avant remplacement moyenne prédite par le modèle de Cox, et la courbe de survie avant remplacement de Kaplan-Meier, par segment sur les contrats étudiants ou non étudiants	93

4.10	Comparaison entre la courbe de survie avant remplacement moyenne prédite par le modèle de Cox construit sur l'ensemble des données, la courbe de survie avant remplacement moyenne prédite par le modèle de Cox construit spécifiquement pour les contrats étudiants, et la courbe de survie avant remplacement de Kaplan-Meier, par segment sur les contrats étudiants	94
4.11	Courbe de survie avant remplacement moyenne prédite par le modèle de Cox construit spécifiquement pour les contrats étudiants, par segment sur les contrats étudiants, et évolution du nombre de contrats étudiants soumis au risque de remplacement au fil des âges possibles des contrats	94
4.12	Courbes de survie prédites, par les modèles de Cox <i>Gradient Boosting</i> sur les durées de vie exactes avant remplacement, construits en séparant les étudiants, des locataires non-étudiants, des propriétaires, moyennes selon la présence de la formule étudiante et par segment MRH	95
4.13	Impact de la valeur des variables sur le risque instantané relatif de remplacement des contrats étudiants d'après le modèle de Cox <i>Gradient Boosting</i>	96
4.14	Impact de la valeur des variables sur le risque instantané relatif de remplacement des contrats locataires non-étudiants d'après le modèle de Cox <i>Gradient Boosting</i>	97
4.15	Impact de la valeur des variables sur le risque instantané relatif de remplacement des contrats propriétaires d'après le modèle de Cox <i>Gradient Boosting</i>	97
4.16	Valeurs de la concordance des prédictions des modèles de Cox <i>Gradient Boosting</i> segmentés construits par validation croisée, sur dix sous-ensembles de validation de la base de données	98
5.1	Représentation visuelle des situations dans lesquelles peuvent se retrouver les contrats	101
5.2	Evolution de la prime pure observée sur le portefeuille MRH entre 2009 et 2017, avec deux ans de vieillissement de la charge sinistre	103
5.3	Valeur contrat moyenne prédite par segment MRH sur les affaires nouvelles de 2021	106
5.4	Marge globale prédite par segment MRH sur les affaires nouvelles de 2021	107
5.5	Proportion de contrats au sein des différentes tranches de marge prédites, par segment sur les affaires nouvelles de 2021	107
5.6	Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des locataires non-étudiants	108
5.7	Valeur contrat moyenne prédite des contrats PM7+ selon le nombre de pièces du logement, sur les affaires nouvelles de 2021	109
5.8	Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, selon la prise en compte du risque de remplacement	112
5.9	Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, selon différentes valeurs du taux d'évolution de la prime pure	113
5.10	Étude par Kaplan-Meier des éventuelles différences de tendance de résiliation des contrats selon leur année de souscription, entre 2006 et 2020	114
5.11	Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, selon différents scénarios d'évolution des probabilités de résiliation	114
5.12	Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, avec l'hypothèse d'un âge maximal des contrats de 15 ans	115

Liste des tableaux

2.1	Estimation de l'espérance de vie à la souscription, au global, et selon les segments MRH	53
3.1	Correction des valeurs aberrantes sur les variables concernées	59
3.2	Valeurs de la concordance de l'application du modèle construit sur les âges entiers	69
3.3	Valeurs de la concordance des prédictions des modèles construits sur les âges entiers ou sur les âges exacts	71
3.4	Valeurs de la concordance des prédictions des modèles construits sur les âges exacts avec ou sans séparation locataires/propriétaires/PNOs	72
3.5	Valeurs de la concordance de l'application du modèle <i>Gradient Boosting</i>	78
3.6	Erreurs sur les estimations de l'espérance de vie à la souscription selon les segments MRH, selon les prédictions moyennes de la fonction de survie, des modèles de cox, et des estimations de Kaplan-Meier	82
3.7	Valeur de la concordance des prédictions du modèle de Cox classique et du modèle de Cox <i>Gradient Boosting</i>	82
3.8	Erreur de prédiction moyenne, par segment, avec le modèle de Cox classique et le modèle de Cox <i>Gradient Boosting</i> , de l'espérance de vie à la souscription des contrats non censurés, par rapport aux durées de vie réellement observées	82
4.1	Évolution de la prime hors taxe lors d'un remplacement pour changement de qualité, selon le type de remplacement	85
4.2	Valeur de la concordance des prédictions du modèle de Cox classique et du modèle de Cox <i>Gradient Boosting</i>	91
4.3	Valeur de la concordance des prédictions du modèle de Cox <i>Gradient Boosting</i> initial et du modèle segmenté	95
5.1	Taux d'évolution moyen de la prime commerciale lors d'un remplacement, selon le segment MRH du contrat à la souscription	102
5.2	Taux d'évolution par segment MRH à appliquer à la prime commerciale des contrats à l'affaire nouvelle, pour prédire une marge globale par segment nulle	110
5.3	Prime moyenne à l'affaire nouvelle des contrats, par segment, avec et sans application du taux d'évolution à appliquer pour prédire une marge globale par segment nulle	110

Information

Par souci de confidentialité, les données relatives à AXA France contenues dans ce mémoire ont été modifiées volontairement. Les données numériques présentées ici ont une valeur purement indicative. Toutefois, ces modifications ont été réalisées de manière à ce que les valeurs restent comparables entre elles.

Introduction

En France, le marché de l'assurance IARD, et en particulier celui de l'assurance habitation, est fortement concurrentiel. D'une part, la loi Hamon, entrée en vigueur en 2015, a offert aux assurés une plus grande facilité de résiliation. D'autre part, aujourd'hui les clients sont beaucoup plus présents sur internet, et comparent davantage les offres d'assurances sur le marché. Ils sont donc désormais très sensibles à la dimension prix, et les assureurs doivent prendre cela en compte dans leur stratégie tarifaire.

Face à cette problématique, AXA France a développé une stratégie qui consiste à proposer des tarifs à la souscription suffisamment attractifs pour satisfaire ses objectifs de volumétrie d'affaires nouvelles. Alors les contrats ne seront pas nécessairement rentables à la souscription, mais chaque année, une majoration des primes commerciales sera appliquée avec pour volonté de rendre les contrats rentables au fil de leur vieillissement.

Aujourd'hui le portefeuille d'assurance multirisque habitation d'AXA France est rentable en globalité, car la marge dégagée par les anciens contrats compense la perte générée par les nouveaux contrats. Cependant pour savoir si cette stratégie est viable à long terme, avec les politiques tarifaires actuelles, il serait nécessaire de vérifier que chaque génération de nouvelles affaires souscrites sera rentable sur la durée de vie totale de ses contrats.

L'objectif de cette étude est donc de construire un indicateur de valeur contrat à la souscription. Cet indicateur correspondra, pour chaque nouveau contrat, à la valeur actuelle probable des flux de résultat futurs qu'il devrait générer sur l'ensemble de sa vie.

Une première partie présentera plus en détail le contexte qui a mené à la nécessité de disposer d'un indicateur de valeur contrat à la souscription.

La première notion clé dans la définition de la valeur contrat, est la prédiction de l'espérance de vie des contrats. Pour construire l'indicateur de valeur contrat à la souscription, il est donc nécessaire de développer un modèle de survie des contrats, à la souscription, face au risque de résiliation, afin de disposer d'une estimation des probabilités de résiliation, à chaque âge, des contrats. Notamment, le cours *Modèle de survie en actuariat* de COUSIN A. [5] a été un guide dans l'étude des différents types de modèles de survie pouvant être développés dans le cadre de cette étude. Dans la deuxième partie de ce mémoire, une modélisation non-paramétrique de la survie des contrats sera d'abord étudiée. Puis, dans la troisième partie, il sera choisi de développer un modèle de survie semi-paramétrique, permettant de prendre en compte l'impact des caractéristiques à la souscription des contrats, sur leur survie. En particulier, le modèle de Cox dit à risques proportionnels, présenté par COX D. dans un article paru en 1972, "Regression Models and Life-Tables" dans *Journal of the Royal Statistical Society : Series B (Methodological)* [6], sera adapté à cette problématique.

Aussi, il sera décidé d'intégrer à la définition de l'indicateur de valeur contrat, la notion d'évolution du risque au fil de la vie des contrats, pouvant avoir un impact sur leur rentabilité. Un modèle de survie semi-paramétrique des contrats, à la souscription, face au risque de remplacement pour changement de qualité, sera donc développé, dans une quatrième partie, pour disposer, pour chaque contrat, d'une estimation de la probabilité de remplacement, à chaque âge. Le remplacement pour changement de qualité est le type de remplacement ayant l'impact le plus marqué sur la rentabilité des contrats, c'est pourquoi c'est

la survenance de ce type de remplacement qui a été étudié. Et l'impact d'un éventuel remplacement pour changement de qualité sur le résultat des contrats sera ainsi intégré dans la prédiction de la valeur contrat.

Toutes les autres notions nécessaires à la construction de l'indicateur de valeur contrat à la souscription seront précisément définies dans la cinquième partie de ce mémoire. La prédiction de la valeur contrat à la souscription sera appliquée aux affaires nouvelles observées de janvier à juin 2021. L'étude des résultats obtenus permettra un premier aperçu global de la rentabilité attendue sur cette nouvelle génération de contrat, selon les hypothèses retenues dans la définition de l'indicateur. Une analyse plus précise des valeurs contrats obtenues, permettra d'apprécier la rentabilité attendue de différents profils prédéfinis de contrats. Différents exemples d'utilité de cet indicateur seront alors également présentés. Le principal étant qu'il permettrait à AXA d'optimiser la souscription des contrats d'assurance habitation, en souscrivant en priorité des profils de contrats ayant une valeur contrat prédite plus élevée. Enfin, les hypothèses prises pour la construction de l'indicateur seront remises en question et l'impact de certaines éventuelles évolutions sera étudié.

— Chapitre 1 —

Contexte de l'étude

1.1 Le marché de l'assurance Multirisque Habitation

1.1.1 Le marché de l'assurance IARD

L'assurance IARD, signifiant « Incendie, Accidents et Risques Divers », est une famille d'assurances qui offre aux particuliers, aux entreprises et autres entités un moyen de protection et de prévoyance contre les aléas, les risques et leurs conséquences ne relevant pas de la durée de la vie humaine. Plus précisément l'assurance IARD, également appelée assurance dommages ou encore assurance non-vie, a pour objet la couverture du patrimoine de l'assuré au quotidien en cas de dommages, en compensant alors les pertes financières dues à la dégradation ou à la destruction. Les biens de l'assuré peuvent être endommagés par des évènements tels que l'incendie, le vol, le dégât des eaux, le bris de glace ou encore les catastrophes naturelles. L'assurance IARD intègre également des garanties complémentaires, telles que la garantie responsabilité civile afin de couvrir les dommages causés à autrui.

L'assurance IARD regroupe plusieurs catégories d'assurance. Les plus connues sont l'automobile, l'habitation et les multirisques professionnelles.

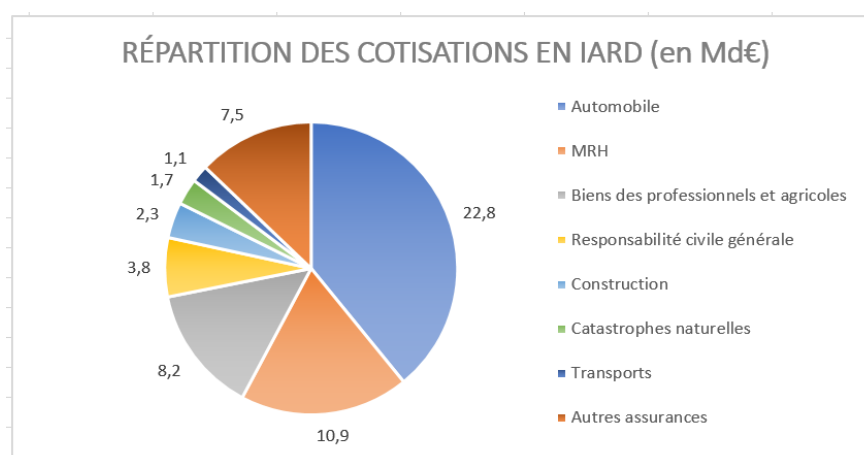


FIGURE 1.1 – Répartition des cotisations perçues sur le marché IARD en 2019 en France d'après la Fédération Française de l'Assurance [11]

Il apparaît sur ce graphique circulaire que l'assurance multirisque habitation (MRH) est la deuxième assurance IARD la plus importante en termes de cotisations en France, c'est sur ce type d'assurance que ce mémoire se concentrera.

En assurance IARD les produits peuvent être destinés à des particuliers ou à des entreprises. Ce mémoire se concentrera sur des produits d'assurance MRH destinés à des particuliers.

1.1.2 L'assurance multirisque habitation

Présentation générale de l'assurance multirisque habitation

Initialement l'assurance habitation ne couvrait que l'incendie. Aujourd'hui il s'agit finalement d'un contrat familial global qui a pour vocation de couvrir l'habitation d'un particulier contre les dommages matériels, mais également la responsabilité civile de ses occupants, ainsi que leur protection juridique.

L'assurance habitation est obligatoire pour les locataires et les copropriétaires occupants ou non-occupants, et elle est facultative pour les propriétaires dont le logement n'est pas en copropriété.

Les différentes garanties généralement proposées par un contrat multirisque habitation pour un particulier sont les suivantes :

- Incendie : elle couvre les dégâts causés par le feu et la fumée, et par les pompiers, à condition que l'incendie soit d'origine accidentelle ;
- Vol et vandalisme : des montants maximaux de remboursement sont fixés dans les contrats, ils distinguent les meubles, les objets de valeur et les objets précieux ;
- Dégât des eaux : la garantie prend en charge les dégradations survenues chez l'assuré et chez ses voisins si leur habitation a été touchée par le dégât ;
- Bris de glace : les éléments de séparation avec l'extérieur ou qui délimitent une pièce sont principalement pris en charge par cette couverture ;
- Catastrophes naturelles : l'événement doit avoir fait l'objet d'une déclaration par arrêté interministériel pour entraîner un dédommagement ;
- Responsabilité civile : la garantie couvre les dommages causés par l'assuré à autrui, l'assureur indemnise la victime à la place de l'assuré ;
- Protection juridique : cela correspond à la défense de l'assuré en cas de contentieux dirigés contre lui, le plus souvent liés à la responsabilité civile, et cela se traduit par une assistance judiciaire et la prise en charge des frais de procès ;
- Accidents liés aux attentats : couvrant les dommages causés par un attentat ou un acte de terrorisme, tel que défini aux articles 421-1 et 421-2 du Code pénal ;
- Assistance à domicile.

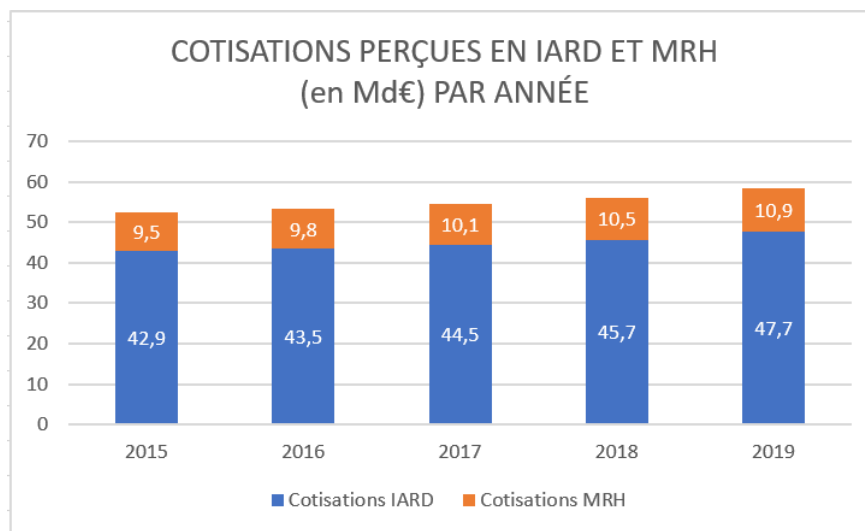


FIGURE 1.2 – Cotisations perçues sur le marché IARD et MRH par année en France d’après la Fédération Française de l’Assurance

De manière visible sur ces diagrammes empilés, les montants des cotisations perçues sur le marché IARD et MRH ont tendance à augmenter au fil des années. Concernant le marché MRH, les cotisations sont passées de 9,5 milliards d’euros en 2015 à 10,9 milliards d’euros en 2019 (+14,7%). Cette augmentation peut s’expliquer d’abord par le fait que l’assurance des habitations est de plus en plus risquée, en raison de l’augmentation du risque de catastrophes naturelles. De plus, d’après l’INSEE, il y a une hausse du marché immobilier en France, et donc plus de logements à assurer, en 2014 la France métropolitaine comptait 33,9 millions de logements [12], et en 2020 36,1 millions [13]. Et en 2019 l’assurance MRH représentait 18,6% des cotisations perçues en assurance IARD.

L’environnement concurrentiel du marché de l’assurance multirisque habitation

En France, le marché de l’assurance MRH est particulièrement concurrentiel. En 2015, la loi relative à la consommation, dite loi « Hamon », est entrée en vigueur. Cette loi fixe de nouveaux droits pour les assurés, et de nouvelles obligations pour les assureurs. De nombreux domaines sont concernés par cette loi dont celui de l’assurance habitation. Alors qu’auparavant les assurés ne pouvaient résilier leur assurance que dans les deux mois qui précédaient l’échéance de leur contrat (autrement ce dernier était renouvelé tacitement), la loi Hamon permet aux clients de pouvoir résilier à n’importe quel moment leur contrat d’assurance automobile ou habitation à partir d’un an. Ainsi, cette loi permet aux assurés de faire jouer la concurrence, et est donc vecteur de compétitivité. Les assureurs doivent constamment étudier leur tarification afin de conserver leurs parts de marché et stimuler leurs résultats.

De plus, les assureurs doivent s’adapter au comportement des clients d’aujourd’hui qui sont de plus en plus présents sur internet, les réseaux sociaux et qui comparent davantage les offres. Les particuliers sont désormais très sensibles à la dimension prix et à la personnalisation des offres.

1.1.3 AXA France sur le marché de l’assurance multirisque habitation

Le service MRH d’AXA France encadre principalement deux produits d’assurance multirisque habitation pour les particuliers.

Le produit "Confort" est un ancien produit qui a vocation à disparaître. Il ne peut donc normalement plus être souscrit, cependant les assurés ayant souscrit au produit "Confort" auparavant ne sont pas dans l’obligation de changer de produit tant qu’ils ne le désirent pas. Ainsi il va subsister pendant des années encore des contrats d’assurance « Confort » dans les bases de données de suivi.

Le produit "Ma Maison" est un nouveau produit qui a été lancé en mai 2017. Il se veut plus moderne en étant un produit totalement modulable selon les besoins de chaque client. Le produit propose un socle de garanties, et les assurés ont ensuite la possibilité de souscrire à des options à la carte, afin de définir le

contrat d'assurance habitation qui leur convient le mieux (il y a par exemple une option piscine, une option casse des appareils nomades, une option énergies vertes, etc.). Étant très personnalisable et proposant des primes mieux ajustées au risque associé à chaque assuré, le produit a permis à AXA de continuer de s'imposer dans le marché de l'assurance habitation de plus en plus concurrentiel.

Il est possible d'observer quelques chiffres pour se rendre compte de la taille du portefeuille MRH d'AXA France et de sa place sur le marché.

Fin 2020 le portefeuille MRH d'AXA France était constitué de 3,5 millions contrats. Au fil des années le nombre de contrats en portefeuille a eu tendance à diminuer d'environ 2% par an.

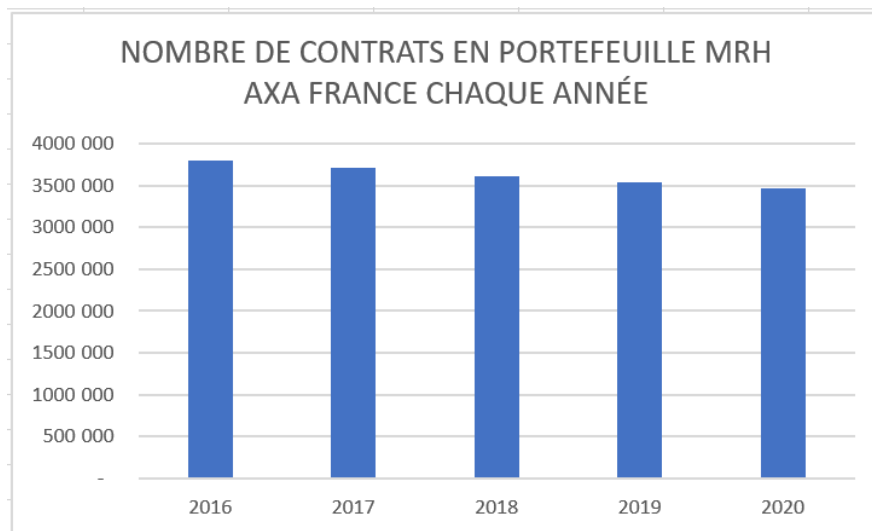


FIGURE 1.3 – Nombre de contrats à chaque fin d'année, au sein du portefeuille MRH d'AXA France

En effet l'apport net est négatif tous les ans, c'est à dire qu'il y a plus de résiliations que d'affaires nouvelles chaque année.

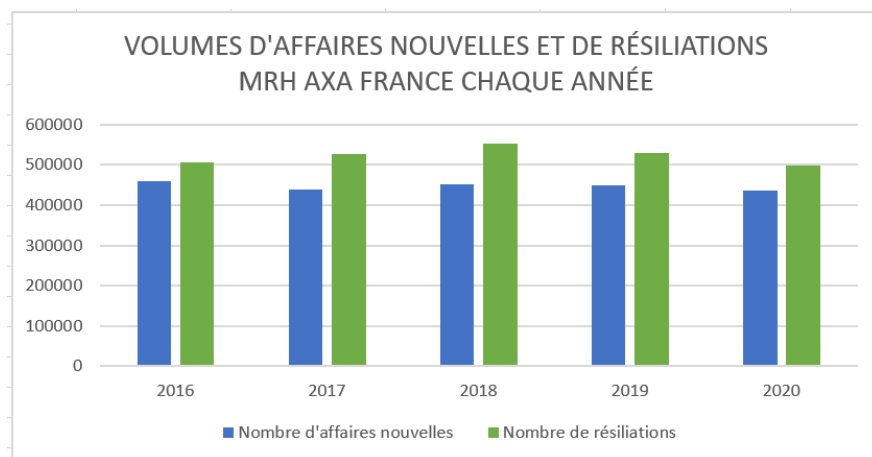


FIGURE 1.4 – Nombre d'affaires nouvelles et de résiliations par an, au sein du portefeuille MRH d'AXA France

Malgré une diminution du nombre de contrats en portefeuille chaque année, le chiffre d'affaire MRH d'AXA France est resté relativement stable et est d'environ 1,2 milliards d'euros en 2020.

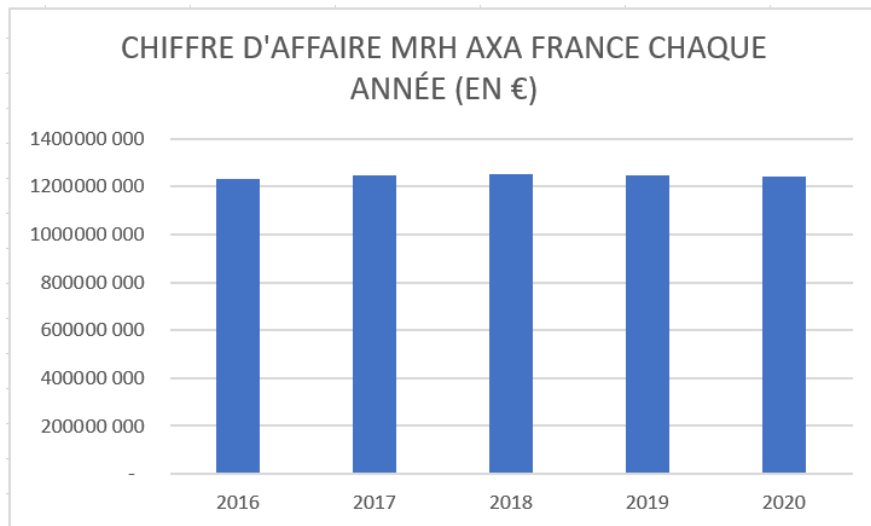


FIGURE 1.5 – Chiffre d'affaire TTC annuel du portefeuille MRH d'AXA France

Aujourd'hui AXA est le deuxième assureur Français en chiffre d'affaire sur le marché d'assurance auto et MRH, d'après l'Argus de l'assurance [14].

1.2 Les problématiques de l'étude

1.2.1 La stratégie d'AXA France

L'objectif d'AXA France en tant qu'assureur est de protéger ses assurés. Pour cela, une entreprise se doit de générer suffisamment de résultat pour pouvoir respecter ses engagements. Pour cela, AXA se doit alors notamment d'être prudent dans sa prédiction de la sinistralité de ses assurés, et également d'assurer un grand nombre de risques afin d'avoir une mutualisation efficace.

Il est important de se rendre compte que ces deux sous-objectifs sont difficiles à concilier. En effet afin d'avoir un grand nombre d'assurés, AXA doit essayer de maximiser son nombre d'affaires nouvelles. Cependant comme le marché de l'assurance habitation est très concurrentiel, pour attirer de nouveaux assurés il faut avoir un tarif attractif, et cela ne semble au premier abord pas forcément compatible avec le principe de prudence qui veut que les primes des contrats soient assez élevées pour que les contrats soient rentables. Une stratégie particulière a donc été mise en place pour remédier à ce problème.

Les primes commerciales des contrats à l'affaire nouvelle sont proposées comme étant les plus faibles possibles pour pouvoir attirer le plus grand nombre de nouveaux clients, mais les contrats ne sont alors pas nécessairement rentables dans les premières années de leur vie.

La prime commerciale d'un contrat est composée d'abord de la prime pure modélisée, qui correspond à l'espérance de la charge de sinistralité qui va toucher les assurés. Ensuite la prime commerciale prend en compte les frais auxquels la compagnie va être soumise pour la prise en charge du contrat (les frais d'acquisition, de souscription, de gestion, les commissions, le coût de la réassurance...). Le montant des taxes fiscales auxquelles sont assujettis les produits d'assurance est aussi pris en compte si la prime commerciale est vue "toutes taxes comprises" (et non "hors taxes"). Et enfin la prime commerciale contient une marge de prudence, qui permet d'anticiper l'éventualité d'avoir une sinistralité réelle supérieure à la sinistralité estimée, et une marge technique qui permet de rémunérer les investisseurs. La prime commerciale peut aussi prendre en compte des mesures commerciales en fonction des orientations stratégiques de la direction. Cela consisterait par exemple à diminuer les primes de certains segments pour essayer de souscrire plus de contrats de ce segment.

Sur la prime commerciale de ses affaires nouvelles AXA prend le risque de ne pas couvrir la totalité de ses frais. Pour compenser cette perte, c'est sur l'ensemble des contrats, toutes générations confondues, que les frais, la prime pure, et les commissions, doivent être couverts par les primes. Il faut donc s'assurer qu'un nouveau contrat qui n'est pas rentable à l'affaire nouvelle le deviendra au fil du temps. Il a donc été défini qu'un terme intervient à chaque date d'anniversaire d'un contrat. Cela consiste à définir des majorations qui seront appliquées annuellement à la prime commerciale du contrat, dépendant de différents critères comme par exemple la sinistralité observée sur le contrat. Alors tout contrat qui reste en portefeuille est censé devenir rentable à partir d'une certaine date, et permettre de compenser les pertes de résultat provoquées par les jeunes contrats.

Avec cette stratégie, le nombre d'affaires nouvelles est maximisé chaque année et les contrats vus en globalité dégagent suffisamment de résultat pour qu'AXA respecte ses engagements.

1.2.2 Les indicateurs de suivi de la stratégie

Différents indicateurs sont utiles au suivi de la stratégie d'AXA, présentée précédemment :

- Le taux de transformation des devis : Cet indicateur correspond au rapport entre le nombre de devis convertis en contrats, et le nombre total de devis effectués, sur une période. L'étude de ce taux permet de se rendre compte de la satisfaction des prospects vis à vis du tarif proposé par AXA.

-L'élasticité à l'affaire nouvelle : Pour définir cet indicateur il faut comparer les taux de transformation des devis, sur une catégorie de contrat aux profils homogènes, dans deux situations tarifaires différentes, toutes choses égales par ailleurs. L'observation de cet indicateur permet de se rendre compte de l'impact des évolutions tarifaires à l'affaire nouvelle sur la concrétisation des devis. Cet indicateur est donc utile pour définir le tarif maximisant le nombre d'affaires nouvelles.

- L'élasticité au terme : Cet indicateur consiste à comparer, sur un ensemble de contrats aux profils homogènes, la proportion de contrats qui ont été résiliés à l'application d'un certain taux de terme par rapport à la proportion de résiliés à l'application d'un autre taux. Par exemple si en 2015 un taux de terme de x% a été appliqué aux contrats en portefeuille cette année, la proportion de résiliés face à ce taux a été mesurée notamment sur les locataires. En supposant que l'année suivante un taux de terme de x+1% a été appliqué, la proportion de résiliés face à ce nouveau taux a été mesurée à nouveau sur les locataires. Alors il est possible de voir si une augmentation du taux de terme a entraîné une augmentation du taux de résiliation chez les locataires. La valeur d'élasticité au terme va correspondre au taux de résiliation en 2016, moins le taux de résiliation en 2015, le tout divisé par la différence entre le taux de terme en 2016 et le taux de terme en 2015. La mesure de cette élasticité au terme selon les profils des contrats permet de définir des valeurs de terme optimales pour rendre les contrats rentables au fil des années sans non plus pousser les clients vers une insatisfaction de leur tarif qui les inciterait à résilier.

- Le taux de sinistralité (ou *loss ratio*) : Cet indicateur se définit, sur un ensemble de contrats ou un seul contrat, comme le ratio entre la somme des charges sinistres observées et la somme des primes commerciales acquises, sur une période :

$$\text{Loss ratio} = \frac{S}{C} = \frac{\text{Charges de sinistres}}{\text{Primes acquises}}$$

Il permet de se rendre compte de l'importance de la sinistralité observée sur différentes périodes. En particulier, il permet de vérifier que la tarification est bien adaptée aux risques.

- Le ratio de perte modélisée (ou *expected loss ratio*, ELR) : Cet indicateur correspond, pour un ensemble de contrats ou un seul contrat, au ratio entre la somme des primes pures prédites et la somme des primes commerciales :

$$\text{Expected Loss Ratio} = \text{ELR} = \frac{\text{Primes pures prédites}}{\text{Primes commerciales}}$$

Cet indicateur permet de voir si la tarification sera adaptée aux risques dans le futur. Il permet aussi de voir quels contrats, ou profils de contrats, seront les plus rentables, et donc d'avoir une idée des contrats que l'assureur souhaite en particulier conserver dans le portefeuille.

- Les volumes, de contrats en portefeuille, d'affaires nouvelles, et de résiliations : Ces indicateurs permettent de constater la croissance passée du portefeuille de contrats.

-Le chiffre d'affaire : Cet indicateur permet de suivre l'ampleur de la production passée.

1.2.3 Nécessité d'un indicateur de la valeur contrat

Tous les indicateurs présentés précédemment sont utiles pour vérifier que la stratégie d'AXA fonctionne bien, mais ils permettent difficilement de prédire l'efficacité de cette stratégie. Pour cela un autre indicateur pourrait être utilisé, un indicateur de prédiction de la durée des contrats.

Construire un indicateur de durée des contrats, permettrait d'avoir une vision de l'espérance de vie des contrats. Cette notion serait utile de manière générale pour mieux connaître le portefeuille MRH d'AXA France, mais elle serait aussi particulièrement utile pour l'évaluation de l'efficacité de notre stratégie dans le futur. En effet, notre stratégie repose sur l'hypothèse que les contrats souscrits restent en portefeuille suffisamment longtemps pour être rentables, et pour qu'une certaine rentabilité soit obtenue sur l'ensemble des contrats, toutes générations confondues.

Pour évaluer l'efficacité de la stratégie dans le futur, la notion de durée seule ne peut pas suffire, mais elle permettra de définir la notion de valeur contrat à la souscription. Ce nouvel indicateur permettra de centraliser les divers indicateurs mentionnés auparavant en un indicateur unique. Cet indicateur sera donc utile pour éclairer les décideurs de la compagnie.

Aujourd'hui il est visible que la stratégie fonctionne car il y a un bon équilibre entre le résultat des anciens contrats et celui des nouveaux. Mais nous ne savons pas si à l'avenir cela fonctionnera toujours. Il est possible que notre rentabilité au global soit aujourd'hui portée par nos anciens contrats qui ont connu les politiques d'évolution tarifaire passées. Mais qu'à l'avenir lorsque les anciens contrats seront des contrats ayant connu dès leur souscription la stratégie d'évolution tarifaire actuelle, ceux-ci ne seront plus capables de compenser les pertes de résultat provoquées par les jeunes contrats.

Pour répondre à ce questionnement, il sera nécessaire de calculer une valeur contrat à la souscription,

cela va correspondre à l'expression du résultat futur, en euros, d'un contrat sur toute sa vie, vu à la souscription. Une fois que cette valeur sera disponible pour chacune de nos affaires nouvelles, il sera possible de voir, sur l'ensemble des affaires nouvelles d'une année, par exemple, la valeur du résultat total qui sera dégagé sur toute la vie de ces contrats. Si cette valeur de résultat total prédit est positive, cela signifiera que cette seule génération est capable de porter la stratégie tarifaire actuelle. Un tel résultat serait rassurant sur le fait que la stratégie actuelle soit viable à long terme.

1.2.4 La construction de l'indicateur de la valeur contrat

La valeur contrat correspond à l'estimation, à la souscription, de la rentabilité future d'un contrat. Cela correspond donc à la valeur actuelle probable des flux de résultat futurs du contrat.

À chaque année de vie d'un contrat, le résultat va correspondre à la marge dégagée, grâce à l'encaissement de la prime commerciale, après paiement des frais engendrés par la souscription, la gestion du contrat, etc., des commissions, du coût de la réassurance, et après mise en provision du montant de la prime pure.

Pour un contrat x , le résultat l'année i vaut :

$$\begin{aligned} \text{Résultat}_i(x) = & \text{Prime commerciale}_i(x) + \text{Produits financiers}_i(x) - \text{Prime pure}_i(x) - \text{Frais}_i(x) \\ & - \text{Commissions}_i(x) - \text{Réassurance}_i(x) - \text{Impots}_i(x) - \text{Taxes}_i(x) \end{aligned}$$

- $\text{Prime commerciale}_i(x)$ correspond à la prime payée par l'assuré du contrat x pour l'année i .
- $\text{Produits financiers}_i(x)$ correspond aux produits financiers obtenus par le placement de la prime jusqu'à utilisation de celle-ci pour dédommagement de sinistres. Dans notre étude, les produits financiers seront négligés.
- $\text{Prime pure}_i(x)$ correspond à la prime pure associée au contrat x pour l'année i , elle sera provisionnée jusqu'à utilisation.
- $\text{Frais}_i(x)$ correspond aux frais dépensés par la compagnie pour la mise en place et la gestion du contrat x sur l'année i . Les frais sont composés de frais d'acquisition (appliqués uniquement la première année du contrat), de frais d'administration des contrats, de frais de gestion de sinistres, d'autres charges techniques et de charges financières. Et ces différents types de frais sont divisés en frais fixes et en frais variables dépendant de la valeur de la prime des contrats.
- $\text{Commissions}_i(x)$ correspond aux commissions associées au contrat x pour l'année i , reversées à l'agent d'assurance AXA qui est en charge du contrat, c'est un taux appliqué au montant de la prime commerciale.
- $\text{Réassurance}_i(x)$ correspond au coût de la réassurance associé au contrat x pour l'année i , c'est aussi un taux appliqué au montant de la prime commerciale.
- $\text{Impots}_i(x)$ correspond normalement à l'imposition du résultat, mais elle ne sera pas prise en compte car les calculs seront faits dans le but d'obtenir un résultat brut.
- $\text{Taxes}_i(x)$ correspond au montant des taxes fiscales auxquelles est assujéti le produit d'assurance, associé au contrat x pour l'année i . Cependant les primes commerciales seront vues "hors taxes", donc les taxes seront déjà soustraites.

La formule retenue est donc la suivante :

$$\text{Résultat}_i(x) = \text{Prime commerciale HT}_i(x) - \text{Prime pure}_i(x) - \text{Frais}_i(x) - \text{Commissions}_i(x) - \text{Réassurance}_i(x).$$

La valeur actuelle probable, à la souscription, des flux de résultat futurs peut donc s'exprimer ainsi :

$$\text{Valeur contrat}(x) = \sum_{i=0}^{\infty} \mathbb{E}(\text{Résultat}_i(x)) \times \left(\frac{1}{1+\text{taux actualisation}_i} \right)^i$$

$$\begin{aligned} \mathbb{E}(\text{Résultat}_i(x)) = & \mathbb{E}(\text{Prime commerciale HT}_i(x)) - \mathbb{E}(\text{Prime pure}_i(x)) - \mathbb{E}(\text{Frais}_i(x)) - \\ & \mathbb{E}(\text{Commissions}_i(x)) - \mathbb{E}(\text{Réassurance}_i(x)) \end{aligned}$$

Pour déterminer les espérances des valeurs annuelles du résultat, et de ses composantes (prime commerciale, prime pure, frais, etc.), il faut considérer les valeurs dans les différentes situations dans lesquelles

peut se retrouver chaque contrat à chaque âge, et leur probabilité de survenance. Trois situations ont été considérées à chaque âge i :

- Le contrat a été résilié entre la souscription et l'âge i compris : dans ce cas, toutes les composantes du résultat sont nulles ;
- Le contrat a vécu un remplacement pour changement de qualité (passage de locataire à propriétaire, ou inversement) entre la souscription et l'âge i compris : dans ce cas, le remplacement a eu un impact sur toutes les composantes du résultat car le risque assuré n'est plus le même en i qu'à la souscription, et les valeurs initiales des composantes ont également évolué avec le temps et ne sont pas les mêmes qu'à la souscription (en raison de l'application annuelle du terme à la prime commerciale, etc.) ;
- Le contrat n'a vécu aucun changement (n'a été ni résilié, ni remplacé pour changement de qualité) entre la souscription et l'âge i compris : dans ce cas, le risque n'a pas changé, seules les évolutions "naturelles" des composantes du résultat au fil du temps sont prises en compte.

Généralement, seules deux situations sont prises en compte dans la définition d'une valeur contrat, la notion de remplacement à la maille contrat est souvent négligée. Il a été décidé dans cette étude de prendre en compte individuellement la possibilité que le risque assuré change durant la vie d'un contrat. Mais prédire la survenance de tous les motifs de remplacement possibles aurait été assez complexe. Par besoin de simplification, il a été décidé de ne prendre en compte que le motif de remplacement pour changement de qualité, qui est le motif ayant l'impact le plus important sur le taux de sinistralité, et sur le ratio de perte modélisé des contrats.

L'espérance de chaque composante peut donc être exprimée, en utilisant pour chaque contrat x et pour chaque âge i la probabilité que le contrat ait été résilié entre sa souscription et l'âge i compris, la probabilité que le contrat ait eu un remplacement pour changement de qualité entre sa souscription et l'âge i compris, et la probabilité que le contrat n'ait pas vécu de changement entre sa souscription et l'âge i compris. Ces probabilités pourront être estimées en développant des modèles de durée, pour la prédiction de durée de vie des contrats avant résiliation et la prédiction de la durée de vie des contrats avant un remplacement pour changement de qualité.

Après détermination de tous les paramètres nécessaires aux calculs d'espérances (le taux de majoration au terme, les valeurs des frais fixes et taux de frais variables, le taux de commission, le taux de réassurance, etc.), détermination des probabilités de survenance de chaque situation à chaque âge et pour chaque contrat, et détermination des taux d'actualisation des flux à utiliser, il sera possible de calculer la valeur contrat vue à la souscription de chaque affaire nouvelle.

1.3 Présentation de la base Image MRH

La principale base de données utilisée, qui repertorie les informations des contrats en portefeuille, est la base Image MRH. Elle permet d'avoir chaque mois la vision de tous les contrats d'assurance multirisque habitation du portefeuille, qu'ils appartiennent au produit « Confort » ou au produit « Ma Maison ».

1.3.1 La construction de la base

La base Image a été construite à partir de 2016, ainsi elle regroupe les informations à partir de 2016 des contrats qui ont été ou sont encore en portefeuille.

Pour se construire, la base Image MRH utilise des bases de données regroupant les informations de tous les contrats d'assurance IARD d'AXA France. Il y a une base de données par région AXA. Il faut savoir qu'au sein d'AXA France les contrats sont répartis selon des régions qui correspondent d'abord à un découpage de la France métropolitaine, avec une région Ile-de-France, une Nord-Est, une Ouest, une Sud-Est et une Sud-Ouest, à une région Dom-Tom et enfin à une région "AXA Partenaires" qui, elle, n'a pas de signification géographique mais représente les clients AXA qui travaillent au sein de l'entreprise. Il y a donc une extraction qui est faite dans chacune de ces bases pour récupérer, et regrouper, les informations des contrats MRH uniquement.

Nous retrouvons dans cette base les caractéristiques principales de chaque contrat, à savoir, son produit d'appartenance, son canal de distribution, la nature de son risque, sa prime payée, sa date d'affaire nouvelle, sa date de résiliation...

Pour rappel, le produit Ma Maison a été lancé à partir de mai 2017, alors il y a dans la base Image MRH des contrats du produit Confort et des contrats du produit Ma Maison. Selon le produit les garanties souscrites au contrat sont indiquées par des variables spécifiques au produit. Il y a alors eu un travail fait pour traduire les options de garanties Confort en options de garanties Ma Maison, pour que les mêmes variables soient utilisées pour tous les contrats.

La base Image MRH est mise à jour tous les mois. Chaque mois, les nouveaux contrats souscrits sont rajoutés. Et lorsqu'un contrat a vu sa prime, sa date de remplacement, ou sa date de résiliation, évoluer au cours du mois, alors une nouvelle ligne, comportant les nouvelles informations du contrat, est rajoutée.

A fin décembre 2020, la base Image MRH regroupait les informations de 6,9 millions de contrats, dont 49% qui sont des contrats résiliés depuis 2016.

1.3.2 Présentation des variables principales de la base

Dans cette section, les principales variables utilisées pour décrire les contrats dans la base seront énumérées.

Une variable permet de savoir si le contrat est associé au produit Confort, au produit Ma Maison, ou bien à un autre produit (un produit d'assurance habitation plus ancien par exemple).

Il est également possible de connaître le segment MRH du contrat, par le biais d'une variable. Les segments possibles sont les suivants :

- LA1P est associé aux contrats dont l'assuré est locataire, et dont le bien est un appartement une pièce ;
- LA2P est associé aux contrats dont l'assuré est locataire, et dont le bien est un appartement deux pièces ;
- LA3P est associé aux contrats dont l'assuré est locataire, et dont le bien est un appartement trois pièces ;
- LA4+ est associé aux contrats dont l'assuré est locataire, et dont le bien est un appartement de quatre pièces et plus ;
- ProA est associé aux contrats dont l'assuré est propriétaire, et dont le bien est un appartement ;
- PNOA est associé aux contrats dont l'assuré est propriétaire, et dont le bien est un appartement non occupé ;

- LocM est associé aux contrats dont l'assuré est locataire, et dont le bien est une maison ;
- PM4- est associé aux contrats dont l'assuré est propriétaire, et dont le bien est une maison de quatre pièces et moins ;
- PM56 est associé aux contrats dont l'assuré est propriétaire, et dont le bien est une maison de cinq ou six pièces ;
- PM7+ est associé aux contrats dont l'assuré est propriétaire, et dont le bien est une maison de sept pièces et plus ;
- PNOM est associé aux contrats dont l'assuré est propriétaire, et dont le bien est une maison non occupé.

Il est également possible de reconstruire ces segments à l'aide des caractéristiques de contrats brutes, représentées par d'autres variables :

- Une variable indique si l'assuré est un locataire, un propriétaire ou un propriétaire non-occupant ;
- Une indique si le logement assuré est une maison, un appartement au rez-de-chaussée ou un appartement à un autre niveau ;
- Une indique si le bien est occupé ou non ;
- Une autre indique le nombre de pièces du logement.

Aussi une indique si le logement assuré est une résidence principale, secondaire, louée ou non occupée.

Une variable donne la date d'émission de l'affaire nouvelle, c'est à dire la date où la demande de souscription a été faite. Et une autre donne la date d'effet de l'affaire nouvelle, c'est à dire la date à partir de laquelle le risque est couvert et le client paye sa prime.

Une variable donne la date d'émission de la résiliation, c'est à dire la date où la demande de résiliation a été faite. Et une autre donne la date d'effet de la résiliation, c'est à dire la date ou le contrat a pris fin.

Des variables présentent les montants de la prime hors taxe et de la prime toutes taxes comprises.

D'autres variables permettent de connaître d'autres caractéristiques du contrat, telles que la région AXA associée, ou le canal de distribution utilisé, si le contrat est un contrat étudiant, etc.

Une autre variable indique le numéro de client de l'assuré du contrat. Cette variable peut permettre de lier la base Image avec une base clients, regroupant les informations de tous les clients AXA. Alors il est possible de disposer d'autres variables comme celles indiquant l'âge de l'assuré, l'ancienneté de l'assuré, la catégorie socio-professionnelle de l'assuré, etc.

1.4 Synthèse du chapitre

Le marché de l'assurance MRH est très concurrentiel, et il est important d'avoir un tarif attractif à l'affaire nouvelle, pour attirer de nouveaux clients.

AXA a donc développé une stratégie pour garantir un résultat satisfaisant malgré cette situation. Le tarif à l'affaire nouvelle est proposé comme suffisamment faible pour permettre de satisfaire les objectifs de volumétrie d'affaires nouvelles. Les contrats ne sont donc pas toujours rentables dans les premières années de leur vie, mais une majoration des primes commerciales est appliquée annuellement afin de rendre les contrats rentables au fil du temps. Cette majoration doit cependant être définie de telle sorte qu'elle n'incite pas les assurés à résilier leur contrat. Un résultat satisfaisant est ainsi obtenu sur l'ensemble des contrats, toutes générations confondues. La marge dégagée par les anciens contrats permet de compenser la perte générée par les nouveaux contrats.

Cette stratégie fonctionne actuellement, mais il n'est aujourd'hui pas possible de savoir si elle continuera d'être efficace dans le futur. Un indicateur de valeur contrat à la souscription doit donc être développé, prenant en compte l'espérance de vie des contrats et des probabilités d'évolution du risque, pour avoir une vision du résultat total attendu sur toute la vie de chaque contrat. Il sera alors possible de voir si la stratégie actuelle sera viable sur les générations futures, ou si elle fonctionne actuellement surtout grâce au poids dans le résultat des anciens contrats présents en portefeuille actuellement, caractérisé par les anciennes stratégies tarifaires qu'ils ont connu.

— Chapitre 2 —

Modélisation non-paramétrique de la probabilité de résiliation

2.1 Études statistiques descriptives

Cette étude du risque de résiliation commence par une analyse descriptive afin d'avoir une meilleure connaissance du portefeuille, et des caractéristiques des résiliations. Cette analyse a été faite sur la base Image MRH avec le logiciel WPS (version 3.3.1) en utilisant le langage SAS.

2.1.1 Description du portefeuille

Le nombre de contrats MRH en portefeuille a eu tendance à diminuer au fil des dernières années en raison d'un plus grand nombre de résiliations que d'affaires nouvelles, et fin 2019 environ 3,5 Millions de contrats sont encore en portefeuille :

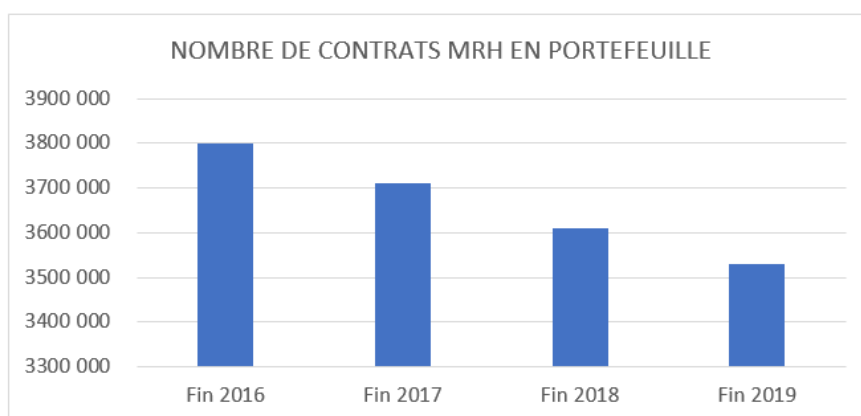


FIGURE 2.1 – Nombre de contrats dans le portefeuille MRH d'AXA France

Les contrats sont majoritairement répartis entre le produit Confort et le produit Ma Maison, qui prend de plus en plus de place au sein du portefeuille :

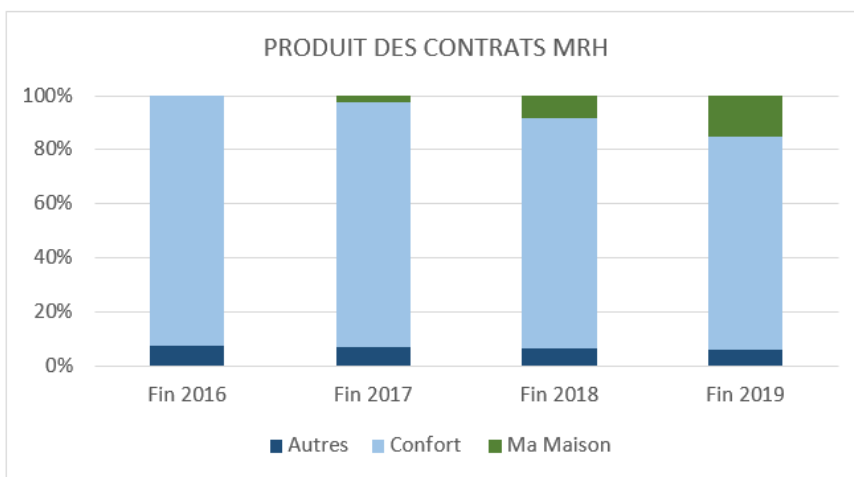


FIGURE 2.2 – Part des différents produits MRH en nombre de contrats

Les contrats peuvent également se répartir selon leur segmentation MRH :

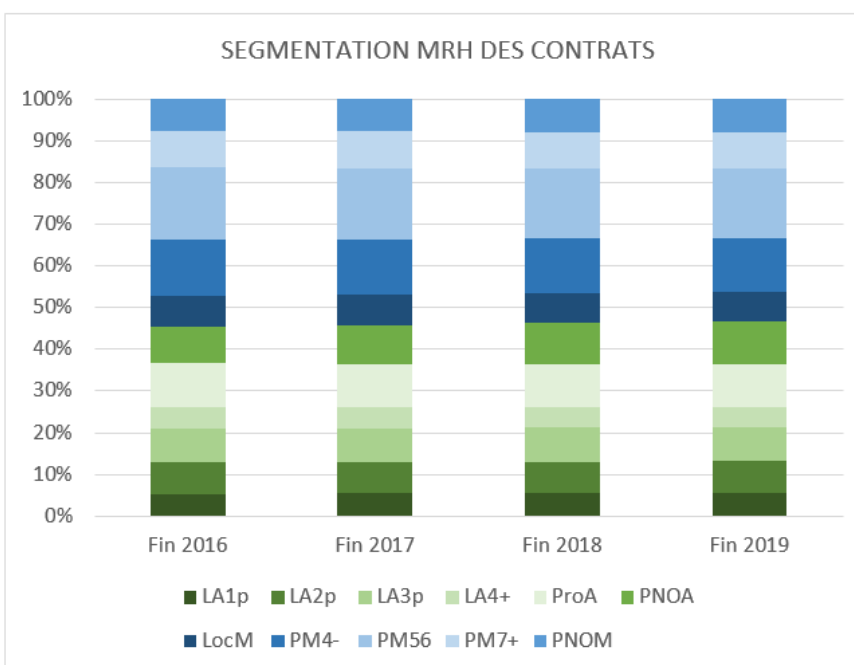


FIGURE 2.3 – Répartition des contrats selon leur segmentation MRH

Cela permet de se rendre compte que les segments les plus représentés dans le portefeuille sont ceux des propriétaires de maisons de cinq ou six pièces, des propriétaires de maisons de moins de quatre pièces, et des propriétaires d'appartements.

En utilisant la variable indiquant le date d'effet des contrats, il est possible de déduire leur âge d'effet. En observant le portefeuille à fin août 2020, il est possible de voir que l'âge moyen des contrats en portefeuille est de 10 ans et sept mois. L'âge moyen des contrats en portefeuille a pu être calculé selon le segment MRH des contrats, ce qui permet de voir que les contrats de propriétaires ont une plus grande ancienneté en portefeuille que les contrats de locataires :

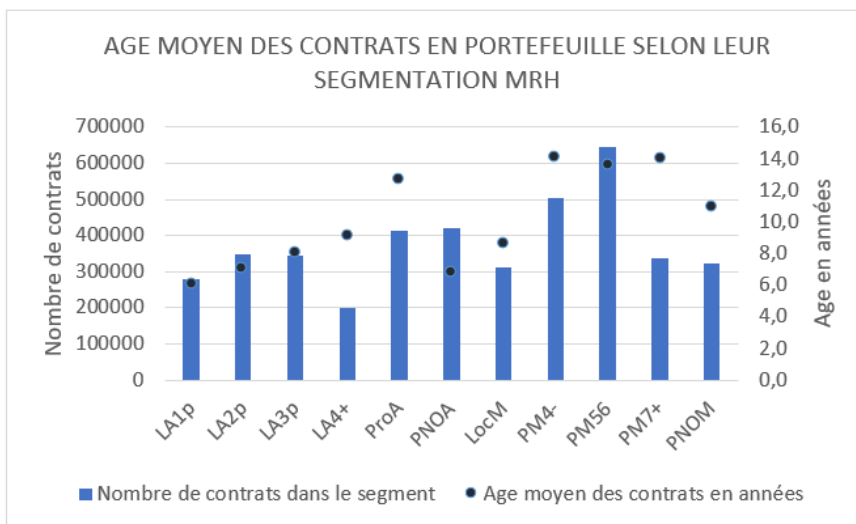


FIGURE 2.4 – Age moyen des contrats en portefeuille par segment MRH (vue à fin août 2020)

Le fait que les contrats de locataires ont des âges moyens plus faibles, peut s'expliquer par le fait que les locataires sont souvent plus jeunes et vont être plus attentifs aux tarifs de la concurrence, via internet, et ainsi auront tendance à résilier plus facilement. Aussi les locataires déménagent plus et il est courant de changer d'assurance habitation lors d'un déménagement.

L'âge moyen des contrats en portefeuille au 31 août de chaque année depuis 2016 a été observé :

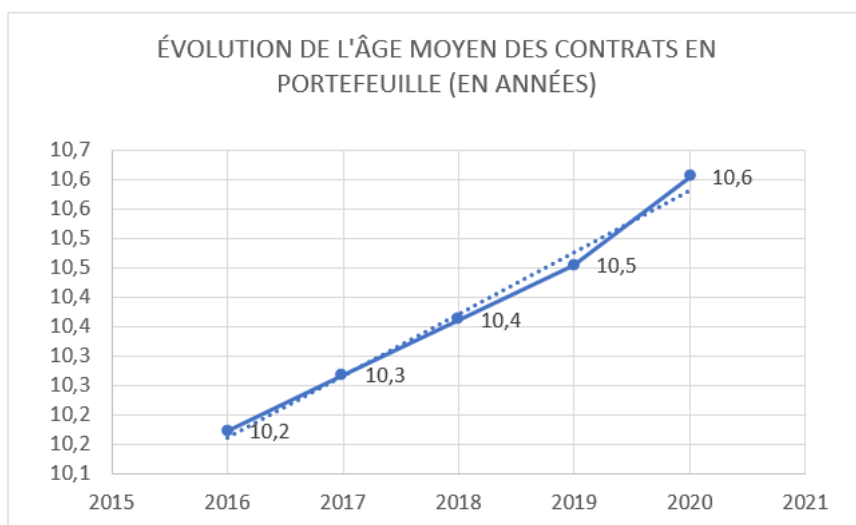


FIGURE 2.5 – Evolution de l'âge moyen des contrats en portefeuille de 2016 à 2020 (vue à fin août)

En août 2016 les contrats en portefeuille avaient en moyenne 10 ans et deux mois, et cet âge moyen a augmenté d'environ 1 mois et 8 jours par an jusqu'à 2020. Tous les segments MRH ont connu une augmentation annuelle semblable de l'âge moyen des contrats en portefeuille, cela peut être dû à la politique AXA qui consiste à faire en sorte de fidéliser les clients.

La distribution des âges des contrats en portefeuille a pu être observée :

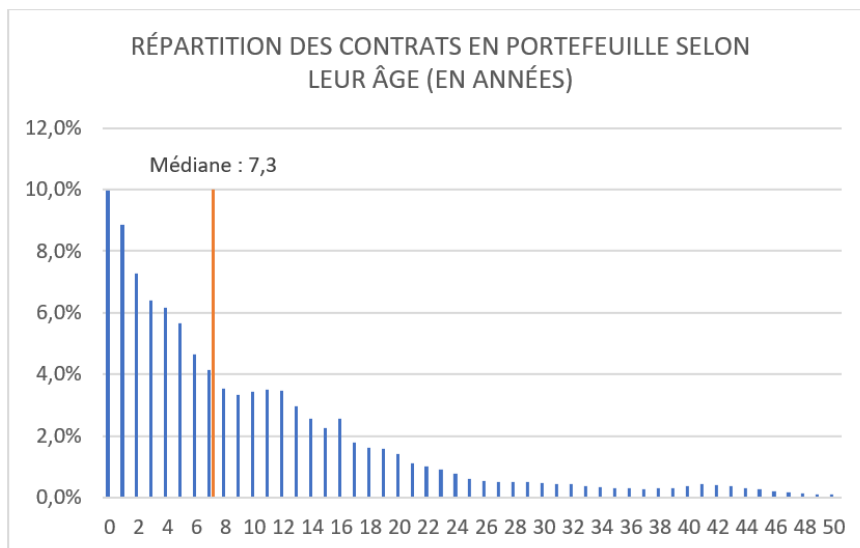


FIGURE 2.6 – Distribution des âges des contrats en portefeuille (vue à fin août 2020)

La médiane des âges des contrats en portefeuille est alors de sept ans et quatre mois.

Il est possible de se demander si le niveau de prime d'un contrat d'assurance peut avoir un impact sur la volonté de résiliation de l'assuré. Alors la prime moyenne des contrat en portefeuille est observée, et sera mise en comparaison, dans la sous-partie suivante, avec la prime moyenne des contrats résiliés. La prime moyenne des contrats en portefeuille vaut 330,10€. En observant la prime moyenne par segment MRH il est visible que plus le bien assuré est grand plus la prime est élevée, et que la prime d'une maison est plus élevée que celle d'un appartement, et également que la prime d'un propriétaire est plus élevée que la prime d'un locataire, traduisant un niveau de risque plus élevé :

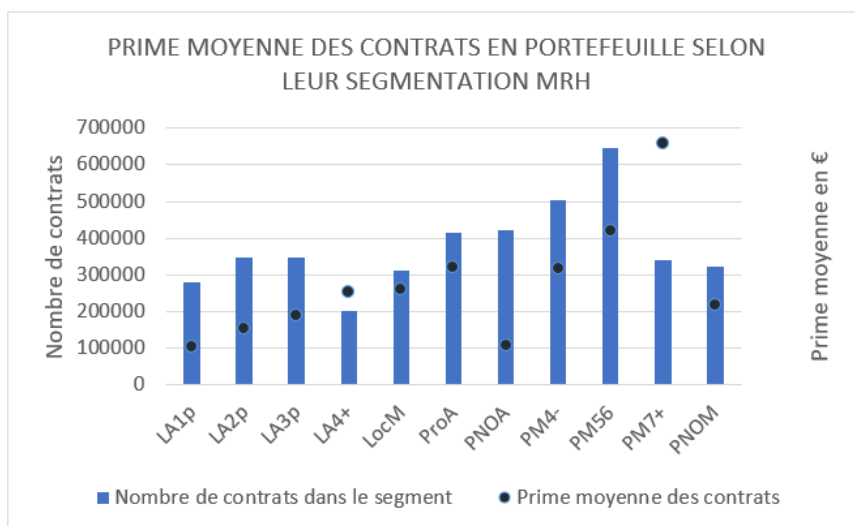


FIGURE 2.7 – Prime moyenne des contrats en portefeuille par segment MRH (vue à fin août 2020)

2.1.2 Caractéristiques des résiliations

Lorsqu'un contrat est résilié ses informations sont conservées dans la base Image. Comme la première image de la base Image date de janvier 2016, les caractéristiques des contrats résiliés depuis 2016 ont donc pu être étudiées à l'aide du logiciel WPS.

En moyenne, les contrats résiliés avaient été souscrits 8 ans et deux mois auparavant. Il est possible d'observer la durée de vie moyenne des contrats résiliés selon le segment MRH qui les caractérisait au moment de la résiliation. Il est vu, en particulier, que les contrats des propriétaires d'appartement et de

maisons ont, en moyenne, été résilié après une période plus longue que ceux des locataires d'appartements ou de maisons :

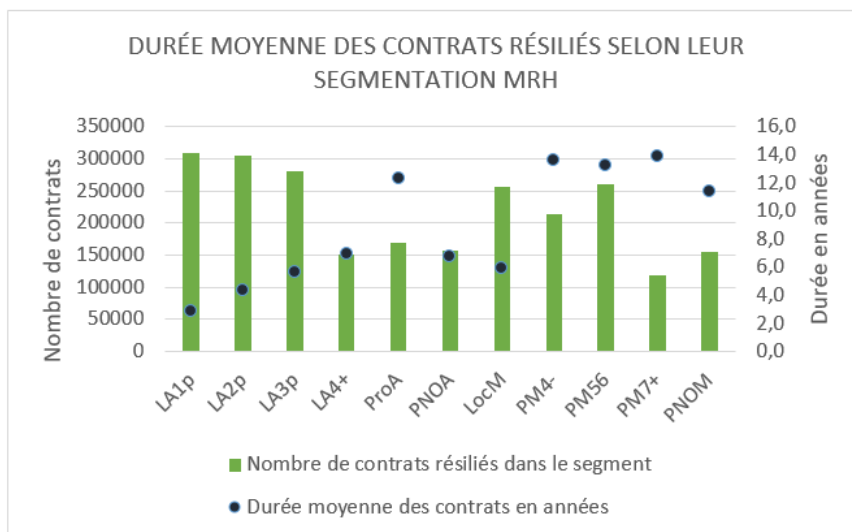


FIGURE 2.8 – Age moyen des contrats résiliés au moment de leur résiliation par segment MRH (vue à fin août 2020)

En observant la durée moyenne des contrats résiliés selon l'âge des assurés à la souscription, nous voyons que les assurés souscrivant un contrat alors qu'ils ont entre 30 et 40 ans auront tendance à le résilier légèrement plus tard que d'autres assurés :

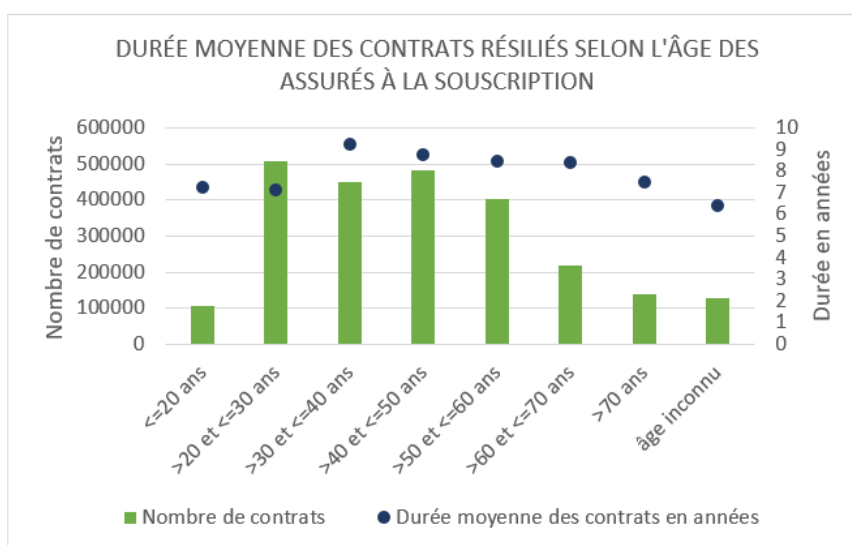


FIGURE 2.9 – Age moyen des contrats résiliés au moment de leur résiliation, selon l'âge des assurés à la souscription (vue à fin août 2020)

Cela peut être dû au fait qu'avant 30 ans, les assurés ont tendance à plus déménager, et, comme dit précédemment, il est courant de changer d'assurance habitation lors d'un déménagement.

La durée moyenne des contrats résiliés selon la région AXA a également été étudiée :

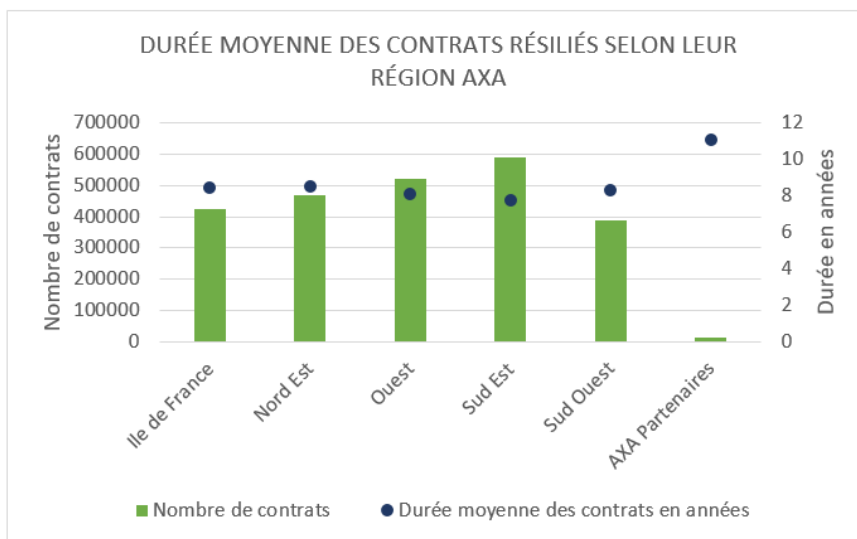


FIGURE 2.10 – Age moyen des contrats résiliés au moment de leur résiliation, selon la région AXA (vue à fin août 2020)

Il apparaît que les contrats de la région du Sud-Est ont une durée de vie moyenne légèrement plus faible que celle des contrats des régions restantes. Aussi les contrats "AXA Partenaires", qui n'est pas une région géographique mais représente les assurés travaillant au sein d'AXA, ont eux une durée de vie moyenne plus élevée d'environ deux ans que celle des contrats des autres régions. Il est possible de supposer que ces assurés seront éventuellement plus fidèles que des assurés lambda, car ils auront un attachement particulier à la marque AXA.

De plus, en observant désormais la durée moyenne des contrats résiliés selon la catégorie socio-professionnelle des assurés à la résiliation, nous pouvons relever le fait que les contrats des retraités ont tendance à avoir été résiliés plus tard que ceux des autres assurés :

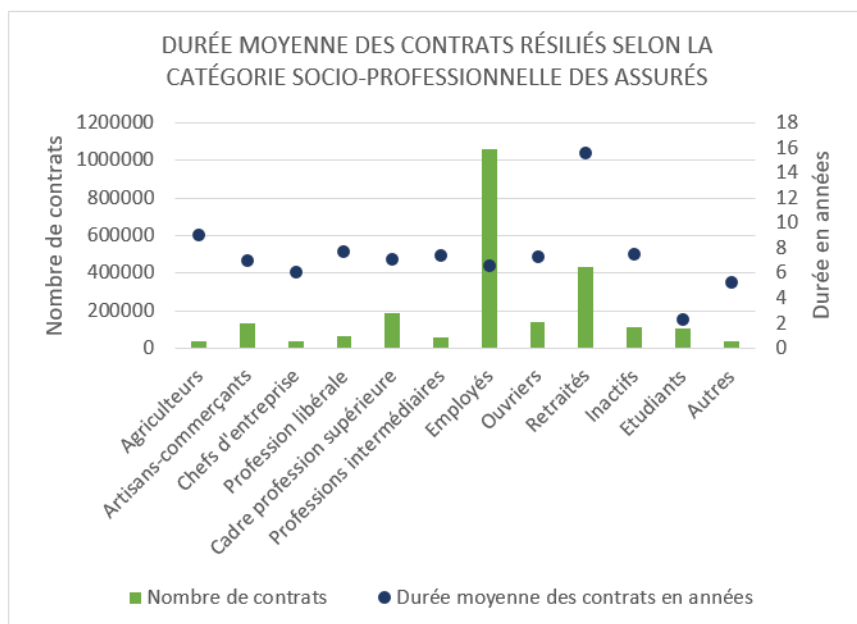


FIGURE 2.11 – Age moyen des contrats résiliés au moment de leur résiliation, selon la catégorie socio-professionnelle des assurés (vue à fin août 2020)

Cela peut indiquer que la majorité des contrats des retraités ont été souscrits pendant la vie active, et qu'il est plus rare de souscrire un contrat d'assurance MRH durant la retraite. Il y a peu de jeunes contrats au sein des contrats des retraités.

Il est possible d'analyser la répartition des contrats résiliés selon leur âge et de la comparer à la répartition des contrats en portefeuille selon leur âge, vue précédemment :

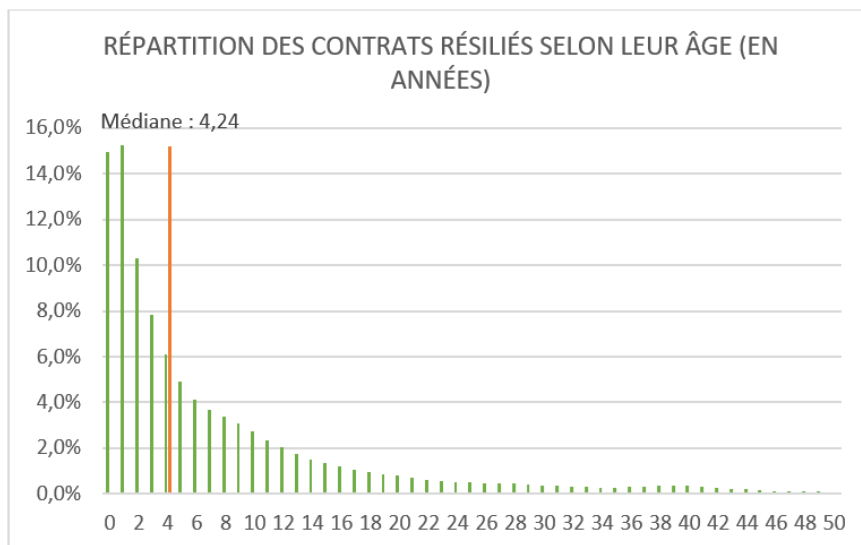


FIGURE 2.12 – Distribution des âges des contrats résiliés au moment de leur résiliation (vue à fin août 2020)

Les contrats en portefeuille apparaissent ainsi plus âgés que les contrats résiliés. Et finalement environ 54% des contrats résiliés ont été résiliés avant leur cinq ans, et la médiane de l'âge des contrats à la résiliation vaut quatre ans et trois mois. Il est donc important de souligner le fait que la majorité des résiliations ont lieu dans les premières années de la vie des contrats.

Les contrats résiliés avaient en moyenne une prime de 294,45€ au moment de leur résiliation. Les primes moyennes des contrats de propriétaires de maisons au moment de leur résiliation sont globalement élevées. En effet, en comparaison avec l'étude des primes moyennes des contrats en portefeuille présentée précédemment, les primes moyennes des contrats de propriétaires de maisons encore en portefeuille sont plus faibles :

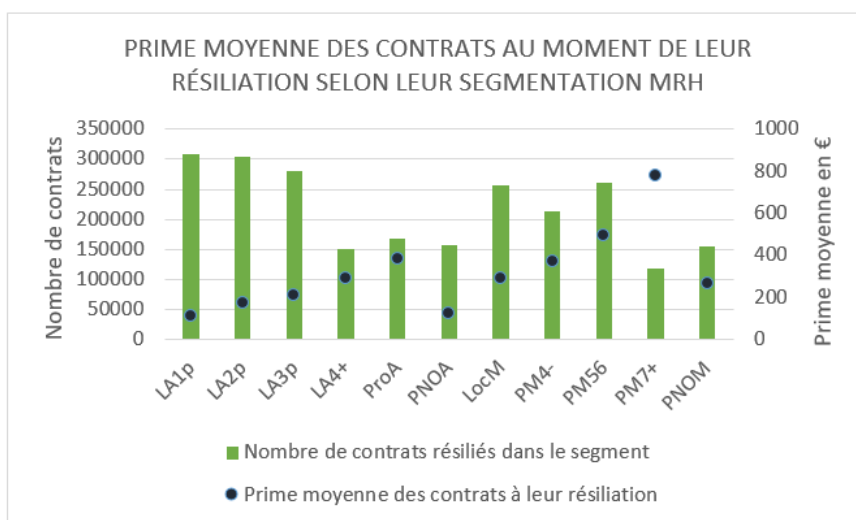


FIGURE 2.13 – Prime moyenne à la résiliation des contrats résiliés par segment MRH (vue à fin août 2020)

Par exemple, les contrats du segment des propriétaires de maisons de cinq ou six pièces résiliés avaient une prime moyenne à la résiliation de 3% plus élevée que la prime moyenne des contrats en portefeuille de ce même segment.

L'une des raisons possibles d'une résiliation serait que l'assuré considère que sa prime d'assurance est élevée. La prime d'un contrat d'assurance d'un assuré, qui reste en portefeuille, évolue et a tendance à augmenter au fil du temps, au fil des termes. Il est alors possible que l'assuré considère que son dernier terme était particulièrement élevé, et décide de résilier son contrat.

Il existe un certain nombre de motifs de résiliation. Un contrat peut être résilié, par exemple, en raison de la vente du bien assuré, ou du décès de l'assuré, ou bien dans le cas de non-paiement, et en particulier un assuré peut résilier son contrat pour souscrire chez un assureur concurrent en invoquant alors la loi Hamon, ou tout simplement en se disant insatisfait vis-à-vis de son terme. Il a donc été étudié la proportion des contrats qui ont été résiliés entre 1 mois avant leur terme et 3 mois après leur terme, et dont l'assuré a invoqué l'application de la loi Hamon comme motif de résiliation, ainsi que ceux dont l'assuré a clairement indiqué qu'il résiliait au terme car il était insatisfait de sa majoration, par rapport à l'ensemble des contrats résiliés. Ainsi il est possible de voir que ces dernières années environ 21% des contrats résiliés ont probablement été résiliés en raison d'une insatisfaction vis-à-vis du terme :

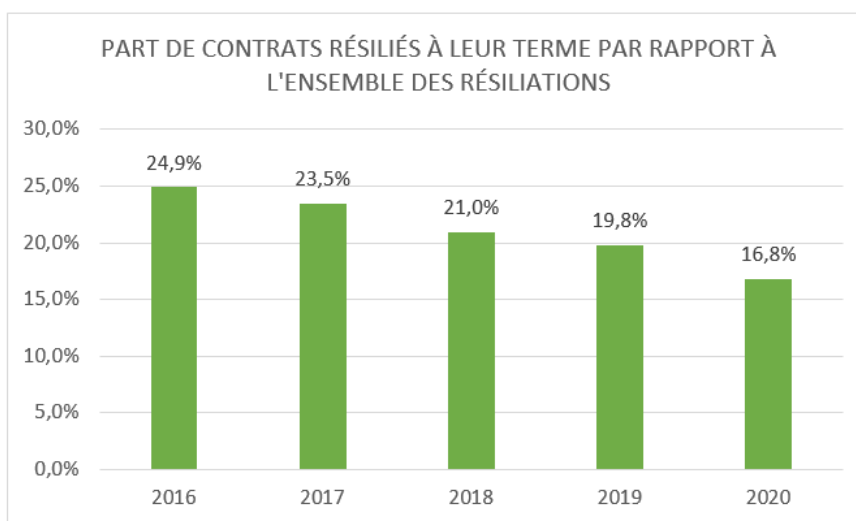


FIGURE 2.14 – Part de contrats résiliés à leur terme parmi les contrats résiliés (vue à fin décembre)

Il apparaît également que la part de contrats résiliés en raison d'une insatisfaction du tarif, a diminué au fil des années. Or, au fil du temps, AXA a souhaité développer sa fidélisation client. Pour cela les taux de majoration au terme ont été diminués ces dernières années. Cette stratégie a visiblement fonctionné, et a amélioré la satisfaction des clients vis à vis de leur tarif au fil des années.

Par la suite, parmi les contrats ayant résilié à leur terme, la part des contrats pour qui il s'agissait du premier terme a été étudiée :

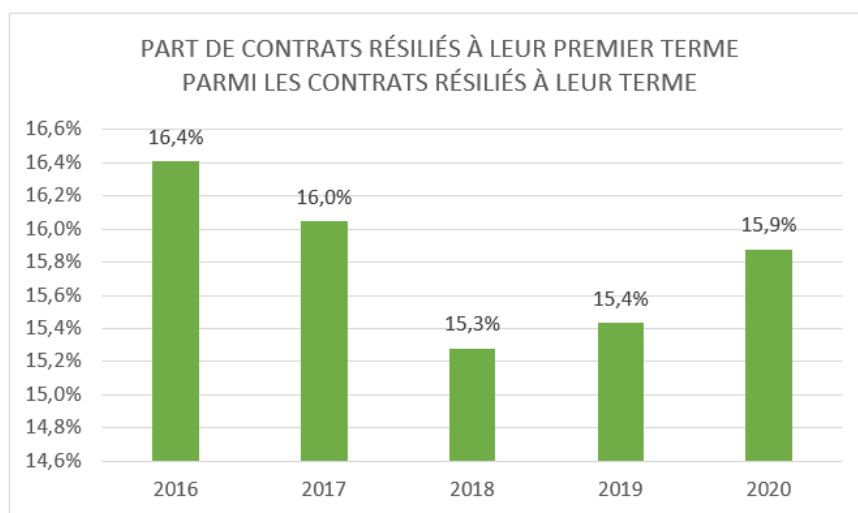


FIGURE 2.15 – Part de contrats résiliés à leur premier terme parmi ceux résiliés à leur terme (vue à fin décembre)

Parmi les contrats résiliés à leur terme avec l'application de la loi Hamon, ou une insatisfaction vis-à-vis du terme, comme motif de résiliation, environ 16% ont été résiliés à leur premier terme sur ces cinq dernières années. Le premier terme est un passage sensible pour le client.

2.1.3 Synthèse de l'analyse descriptive

L'analyse descriptive qui a été réalisée précédemment, sur les contrats en portefeuille puis sur les contrats résiliés au cours de ces dernières années, a permis de faire ressortir des variables qui pourront éventuellement être utilisées dans les modélisations qui vont suivre.

La variable qui semble avoir le plus d'impact sur la durée de vie des contrats, parmi les variables étudiées, est la variable indiquant le segment MRH du contrat.

L'âge de l'assuré à la souscription, et sa catégorie professionnelle, la région AXA du contrat, ainsi que la prime du contrat, semblent également, au premier abord, avoir un léger impact sur la durée de vie des contrats MRH.

Nous allons pouvoir nous représenter dans la suite de l'étude si ces variables feront partie des variables principales dans la modélisation.

2.2 Modélisation préliminaire : Estimation des courbes de survie par Kaplan Meier

Afin d'avoir une première vision de la survie des contrats il a été choisi d'appliquer la méthode d'estimation des courbes de survie par Kaplan Meier qui est non paramétrique. Cette méthode suppose alors que la survie des contrats ne dépend que du temps. Une courbe de survie sera estimée sur l'ensemble des contrats sans distinction. Puis pour voir l'impact de la variable de la segmentation MRH, des courbes de survie seront estimées sur les contrats séparés en différentes bases selon leur segment MRH. Ainsi, il sera possible d'avoir un premier aperçu de l'influence du segment MRH sur la survie d'un contrat.

Il y a eu la volonté d'estimer la fonction de survie des contrats sur une période étendue, sur plusieurs décennies, car l'assurance multirisque habitation présente la spécificité d'avoir des contrats pouvant rester en vigueur très longtemps, en particulier lorsque les assurés ne vivent pas de déménagements. Il a finalement été décidé de ne faire cette étude que sur les contrats souscrits à partir de 1970, cela permettant d'avoir une estimation de la fonction de survie jusqu'à l'âge 50, ce qui sembla suffisant au vu du peu de contrats restant en portefeuille souscrits avant 1970. Cependant il faut noter que comme il a été décidé d'estimer la fonction de survie jusqu'à l'âge 50 ans, les bases de données disponibles, et en particulier la base Image MRH que nous utilisons dans cette section, n'existaient pas au moment de la souscription de la majorité des contrats étudiés. Alors nous ne disposons pas des caractéristiques des contrats à la souscription mais de leurs caractéristiques à partir de 2016 uniquement. En raison de cette contrainte de données, une hypothèse a donc été prise pour la construction des fonctions de survie selon le segment MRH : il est supposé que les caractéristiques des contrats étudiés, et par conséquent leur segment MRH, n'ont pas changé au cours de leur vie.

2.2.1 Observation des données de survie

La manière idéale pour observer les données de survie des contrats du portefeuille serait de pouvoir observer chaque contrat sur la période allant de sa date de souscription à sa date de résiliation. Cela n'est en réalité pas possible car le portefeuille n'est pas éteint.

L'étude a donc été réalisée avec les données de la base Image MRH sur la période d'observation de 2016 jusqu'à fin 2020, qui correspond à la période sur laquelle la base Image a été construite. Sur chaque année il a pu être observé le nombre de contrats en portefeuille, à risque, selon l'année de souscription et l'âge du contrat. Ainsi à chaque âge, observable entre 2016 et 2020, il est possible de voir la part de contrats qui ont été résiliés à cet âge, et finalement la part qui est restée en portefeuille jusqu'à l'âge suivant (la part de contrats qui a survécu un an de plus). Cela permet d'estimer des probabilités de résiliation sur les données observées, pour chaque âge observé selon l'année de souscription, avec la possibilité de calculer à chaque âge la probabilité de résiliation moyenne sur l'ensemble des années de souscription dont les contrats ont pu être observés à cet âge entre 2016 et 2020. Les données ont été extraites de la base Image avec le logiciel SAS® puis regroupées et traitées sous forme de tables avec le logiciel Office Excel®.

Les données ont été regroupées dans une table sous la forme suivante :

Age des contrats à l'observation	Année de souscription des contrats										
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
0							516623	487921	488215	476863	453379
1						509189	437885	396947	392651	393560	
2					403693	430880	358483	318284	316609		
3				332031	350636	375561	303135	269214			
4			262494	292522	309465	331824	263546				
5		230988	234225	260535	276208	299011					
6	233984	208325	210118	233323	250369						
7	212453	188481	188982	211445							
8	192974	170742	172315								
9	175433	156343									
10	161054										

FIGURE 2.16 – Présentation d'un extrait de la table regroupant les données observées

En particulier à la case d'année de souscription des contrats 2016 et d'âge 0 nous avons le nombre de contrats souscrits en 2016. Par exemple à la case d'année de souscription 2014 et d'âge 2 nous avons le nombre de contrats souscrits en 2014 qui sont restés en portefeuille au moins 2 ans, cela correspond au nombre de contrats souscrits en 2014 qui sont restés en portefeuille au moins 1 an moins le nombre de contrats souscrits en 2014 résiliés à l'âge de 1 an . . .

Il faut noter qu'en observant les données de cette manière, seules les probabilités de résiliation annuelles seront estimées.

Les données de survie de l'ensemble des contrats ont donc été observées de cette manière, en se limitant aux contrats dont l'année de souscription était postérieure à 1970. Cette limitation nous prive de 0,4% des contrats, souscrits avant 1970 et encore en cours. Dans la suite de l'étude, il sera supposé que tous les contrats résilient au plus tard à 50 ans d'ancienneté.

Ces données ont également été observées en faisant une distinction entre les contrats selon les segments MRH auquel ils sont associés.

Cela permettra d'étudier la durée de survie des contrats selon certaines de leurs caractéristiques.

Il faut noter que, de cette manière, différents types de durées de vie des contrats seront observés :

- Certaines durées pourront être tronquées : cela signifie que la souscription du contrat a eu lieu à une date antérieure à la date de début d'observation de l'étude ;
- Certaines durées pourront être censurées : cela signifie que la date de fin d'observation du contrat est postérieure à la date de fin d'observation de l'étude ;
- Certaines durées pourront être à la fois tronquées et censurées ;
- Certaines durées seront entièrement observables, sans censure ni troncature ;
- Certaines durées ne pourront tout simplement pas être observées sur la période d'observation.

Le schéma ci-dessous permet la visualisation de ces différents types de données :

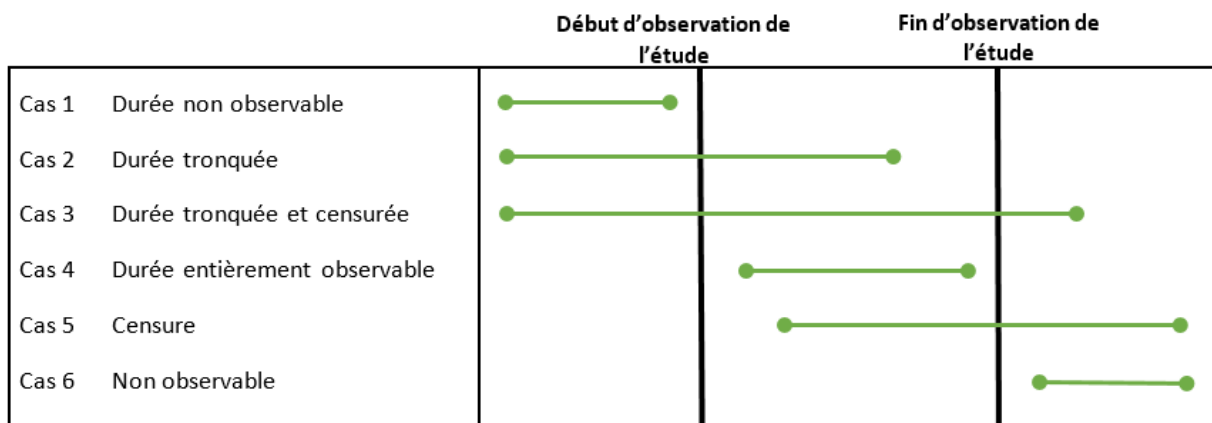


FIGURE 2.17 – Présentation de la visualisation des différents types de données

2.2.2 Estimation des courbes de survie avec Kaplan-Meier

Kaplan et Meier ont développé une méthode de calcul des lois de survie des individus de manière discrète. Cette méthode est notamment décrite dans le mémoire d'actuariat de BALTESAR B. *Construction d'une table de mortalité sur un portefeuille de temporaires décès* (2013) [3]. Afin d'appliquer cette méthode, il faut noter que pour chaque année de souscription des contrats nous disposons de :

- $t \in \{0, \dots, m\}$, avec m l'âge maximal observable, les âges possibles des contrats ;
- n_t le nombre de contrats en portefeuille à l'âge t , ou le nombre de sujets dits « à risque » sur l'intervalle $]t; t + 1]$;
- d_t le nombre de contrats résiliés pendant l'intervalle $]t; t + 1]$;
- c_t le nombre de contrats censurés pendant l'intervalle $]t; t + 1]$;
- tr_t le nombre de contrats tronqués pendant l'intervalle $]t; t + 1]$.

Alors $\forall t \in \{1, \dots, m\}$ la formule suivante est valable : $n_t = n_{t-1} - d_{t-1} - c_{t-1} + tr_{t-1}$.

De manière plus schématiques cette formule peut s'exprimer ainsi : le nombre de contrats exposés au risque de résiliation à l'âge t correspond au nombre de contrats exposés au risque à l'âge $t - 1$, moins le nombre de contrats sortis à l'âge $t - 1$, plus le nombre de contrats entrés à l'âge $t - 1$.

Les résiliations et les censures à l'âge $t - 1$ représentent le nombre de contrats sortants à cet âge et les troncatures à l'âge $t - 1$ représentent le nombre de contrats entrants à cet âge.

Kaplan-Meier repose sur la notion de fonction de survie.

Pour développer ce sujet il est nécessaire de rappeler les principales notions utilisées en mathématiques de la mortalité, qui sont notamment présentées dans le mémoire de DAMASSE A. *Mortalité prospective et construction de tables de mortalité d'expérience* :

- Une variable aléatoire T symbolise le moment de décès, \mathbb{P} est la fonction de probabilité et t est la variable temps ;
- La probabilité pour un individu de survivre k années de plus sachant qu'il a vécu au moins t années s'exprime ainsi : ${}_k p_t = \mathbb{P}(T > t + k | T > t)$, quand $k = 1$ la notation est p_t ;
- La probabilité pour un individu de décéder dans les k années de suivantes sachant qu'il a vécu au moins t années s'exprime ainsi : ${}_k q_t = 1 - {}_k p_t$ car ${}_k q_t = \mathbb{P}(T \leq t + k | T > t) = 1 - \mathbb{P}(T > t + k | T > t)$, quand $k = 1$ la notation est q_t ;
- La fonction de survie de T se défini ainsi : $S(t) = \mathbb{P}(T > t)$.

Cette formule de la fonction de survie peut se développer de la manière suivante :

$$S(t) = \mathbb{P}(T > t) = \mathbb{P}(T > t | T > x) \times S(x) = S_x(t) \times S(x) \text{ avec } t > x.$$

Alors avec la notation $p_t = \mathbb{P}(T > t + 1 | T > t)$ la fonction de survie à chaque âge peut s'exprimer ainsi :

$$S(1) = \mathbb{P}(T > 1 | T > 0) \times S(0) = p_0$$

$$S(2) = \mathbb{P}(T > 2 | T > 1) \times S(1) = p_1 \times p_0$$

...

$$S(t) = \mathbb{P}(T > t | T > t - 1) \times S(t - 1) = \prod_{k=0}^{t-1} p_k$$

La probabilité de survie entre l'âge t et l'âge $t+1$, p_t , peut être estimée comme correspondant au nombre de contrats encore en portefeuille à l'âge $t+1$ divisé par le nombre de contrats encore en portefeuille à l'âge t , dans le cas où il n'y a pas de censures ou de troncatures. La formule suivante se dessine donc : $\hat{p}_t = \frac{n_t - d_t}{n_t}$.

Finalement l'estimateur de Kaplan-Meier de la fonction de survie se définit de la manière suivante :

$$\widehat{S}_{KM}(t) = \prod_{k=0}^{t-1} \hat{p}_k = \prod_{k=0}^{t-1} \frac{n_k - d_k}{n_k}$$

Il est possible d'introduire la notion de probabilité de résiliation entre l'âge t et l'âge $t+1$, q_t , qui peut être estimée comme le nombre de contrats résiliés durant l'âge t divisé par le nombre de contrats en portefeuille à l'âge t Alors : $\hat{q}_t = \frac{d_t}{n_t} = 1 - \hat{p}_t$.

L'estimateur de Kaplan-Meier de la fonction de survie s'exprime donc aussi de la manière suivante :

$$\widehat{S}_{KM}(t) = \prod_{k=0}^{t-1} \frac{n_k - d_k}{n_k} = \prod_{k=0}^{t-1} 1 - \frac{d_k}{n_k} = \prod_{k=0}^{t-1} 1 - \hat{q}_k$$

Les probabilités de survie, de résiliation et finalement la fonction de survie ont pu être estimées sur les données qui avaient été observées de 2016 à 2020.

Pour chaque année d'observation de 2016 à 2019, les contrats à risque ont été classés selon leurs âges, et les nombres de contrats à chaque âge ont été relevés. Puis ces quantités ont été comparées aux nombres de contrats aux âges directement supérieurs pour les mêmes années de souscription (le nombre de contrats restants à risque à l'année d'observation suivante), pour déterminer des probabilités de survie ou de résiliation à chaque âge pour chaque année de 2016 à 2019.

Voici par exemple le tracé des valeurs des estimations des probabilités de survie à chaque âge selon l'année d'observation, pour l'étude sur l'ensemble des contrats :

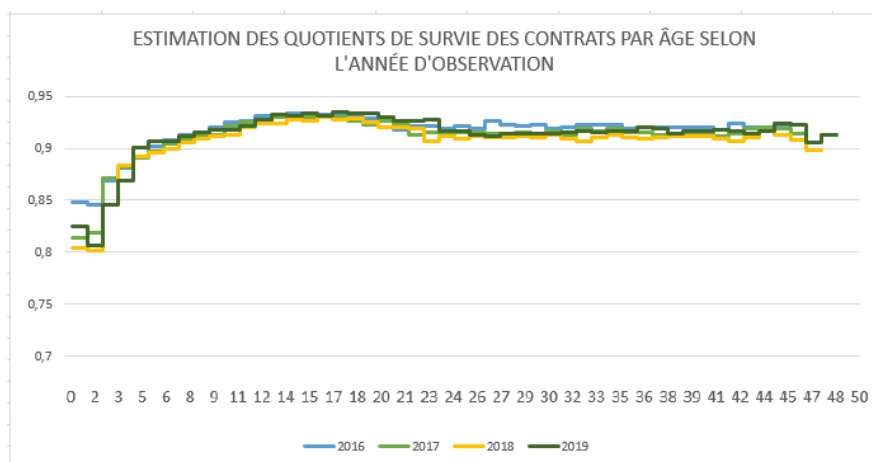


FIGURE 2.18 – Tracé des probabilités de survie estimées à chaque âge et par année d'observation

Il est visible, par exemple, qu'en 2018 les contrats de 2 ans avaient en moyenne 80% de chance de survivre au moins un an de plus.

Ce graphique permet de voir que les clients ont tendance à avoir une forte propension à résilier au

début de la vie de leur contrat. Puis à partir de 5 ans d'ancienneté du contrat, la probabilité de survie du contrat est plus élevée et se stabilise.

Après l'obtention de ces estimations de probabilités de survie, les valeurs de la fonction de survie à chaque âge ont pu être estimées :

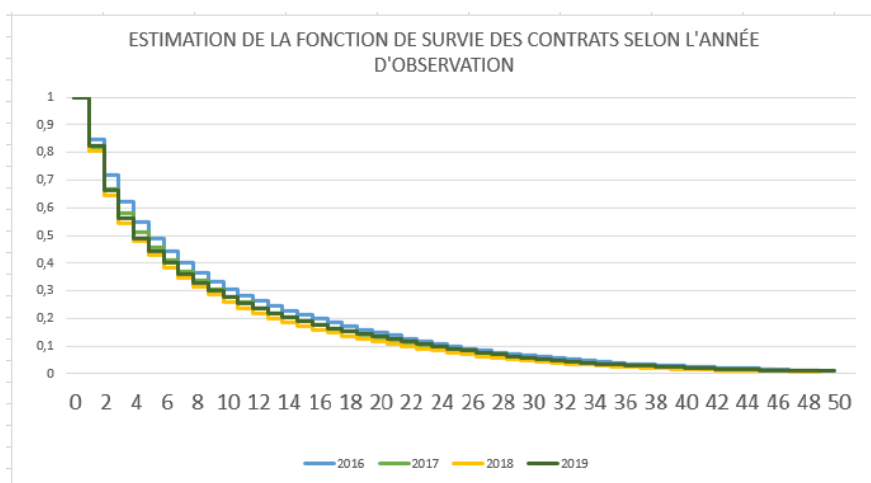


FIGURE 2.19 – Tracé des fonctions de survie estimées par année d'observation

Il est visible sur ces courbes que les contrats à risques sur les années de 2016 à 2019 ont eu la même tendance de résiliation, il est donc possible dans un premier temps de faire des estimations moyennes sur l'ensemble de ces années. Cette hypothèse sera à nouveau étudiée dans la suite de l'étude.

Finalement des estimations ne dépendant pas de l'année d'observation ont été obtenues en calculant les estimations à chaque âge sur l'ensemble des années d'observation :

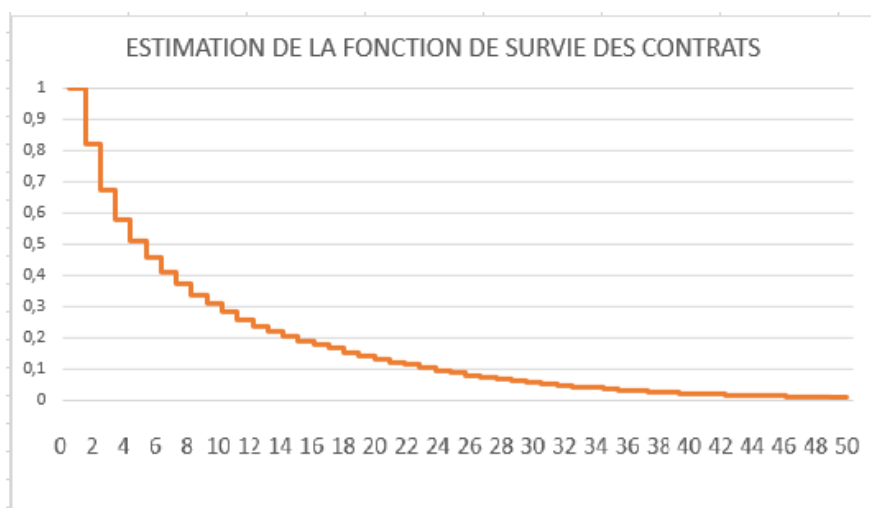


FIGURE 2.20 – Tracé de la fonction de survie estimée globale

Il est nécessaire de noter que les fonctions de survie estimées (que ce soit sur l'ensemble des contrats, où sur les contrats différenciés par leur segmentation MRH) n'atteignent pas la valeur 0 car nous avons des données censurées. En effet nous n'avons pas la possibilité de savoir à quel âge seront résiliés tous les contrats restants en portefeuille à la fin de notre durée d'observation. C'est également la raison pour laquelle l'estimation de la probabilité de résiliation des contrats n'atteint pas la valeur 1.

2.2.3 Lissages non-paramétriques des taux annuels de mortalité

Les estimations des probabilités de résiliations à chaque âge ont pu être calculées sur les données observées, sur l'étude au global et les études selon les segments MRH. Pour rappel, ces estimations peuvent se calculer ainsi : $\hat{q}_t = \frac{d_t}{n_t} = 1 - \hat{p}_t$ ou comme $\hat{q}_t = 1 - \frac{\widehat{S_{KM}(t+1)}}{S_{KM}(t)}$. Les taux \hat{q}_t doivent faire l'objet d'un lissage car la vraie courbe des q_t doit être proche de la courbe empirique \hat{q}_t mais également suffisamment régulière, pour supprimer la volatilité présente à cause du manque de données, notamment sur les grands âges. Deux méthodes de lissage non-paramétrique des taux de mortalité ont alors été considérées, dans un premier temps la méthode des moyennes mobiles pondérées puis la méthode de Whittaker-Henderson. Ces méthodes sont présentées par COUSIN A. dans son cours *Modèle de survie en actuariat* [5].

Méthode des moyennes mobiles pondérées

Pour déterminer des estimateurs des taux de mortalité lissés avec la méthode des moyennes mobiles il faut d'abord choisir des valeurs h' et h qui vont permettre de définir pour un âge t la fenêtre $[t - h'; t + h]$ sur laquelle la moyenne sera calculée. Il faut alors t tel que $t - h' \geq 0$ (l'âge minimal) et $t + h \leq m$ (l'âge maximal). Il est possible de choisir une fenêtre non centrée, avec $h' \neq h$.

Alors pour un âge t , satisfaisant les contraintes énoncées précédemment, l'estimateur du taux lissé s'exprime ainsi : $\hat{q}_t^l = \frac{1}{\sum_{i=t-h'}^{t+h} \alpha_i} \times \sum_{i=t-h'}^{t+h} \alpha_i \hat{q}_i$, avec α_i le poids attribué à la position i dans la fenêtre $[t - h'; t + h]$.

Il est alors visible que cette méthode ne prend pas en compte la fiabilité des valeurs. En effet dans les premières années de la vie des contrats beaucoup de volume est disponible pour estimer les taux de résiliation, alors il est moins nécessaire de lisser les valeurs que lorsque peu de données de contrats sont disponibles. Si la valeur de h pouvait évoluer au fil du temps cet aspect serait pris en compte, mais ce n'est pas le cas. Cette méthode n'a donc pas été retenue pour le lissage des taux de mortalité des contrats.

Méthode de Whittaker-Henderson

Finalement la méthode de lissage de Whittaker-Henderson a été appliquée, avec l'utilisation du logiciel R. Cela a permis d'obtenir pour chaque étude des données de survie une courbe des \hat{q}_t lissée. Cette méthode consiste à rechercher le meilleur compromis entre l'adéquation aux données brutes et la régularité de la courbe de mortalité.

L'estimateur du taux de mortalité lissé \hat{q}_t^l , pour chaque âge t , correspond au taux minimisant $WH_h(c) = F(c) + hS(c)$, avec pour toute courbe $c : t \rightarrow c_t$, $F(c) = \sum_{t=0}^m w_t (c_t - \hat{q}_t)^2$ (avec (w_t) une suite de poids positifs) le critère de précision et $S(c) = \sum_{t=0}^{m-z} [(\Delta^z c)_t]^2$ (avec un ordre z qui a été choisi ici comme valant 2) le critère de régularité.

h correspond à un paramètre donnant l'importance de la contrainte de régularité. Pour vérifier que le choix de h était cohérent un test d'adéquation a été effectué à chaque lissage. Ce test d'adéquation opposait l'hypothèse $H_0 : q_t = \hat{q}_t^l$ à l'hypothèse $H_1 : q_t \neq \hat{q}_t^l$. Cela consiste d'abord à définir un $D_{réalisé} = \sum_{t=0}^m (\hat{q}_t - \hat{q}_t^l)^2$. Puis à se placer lorsque H_0 est vrai, ainsi chaque \hat{q}_t doit suivre la loi normale de moyenne \hat{q}_t^l et de variance $\frac{\hat{q}_t^l(1-\hat{q}_t^l)}{n_t}$. Un certain nombre de réalisations, disons i réalisations, de cette loi sont produites pour chaque t . Ensuite va être calculée pour chaque réalisation (ensemble de nouveaux taux) la valeur $D = \sum_{t=0}^m (\hat{q}_t - \hat{q}_t^l)^2$. Ont donc été obtenues i valeurs de D . Il faut calculer le nombre de D supérieurs ou égaux à $D_{réalisé}$ divisé par le nombre total de D . Une valeur entre 0 et 1 est obtenue, transformée en pourcentage elle devient la p-valeur du test d'adéquation. Lorsque la p-valeur de ce test était inférieure ou égale à 5% l'hypothèse H_0 et donc le choix de h était rejeté. Le h a finalement été choisi selon l'appréciation visuelle des courbes obtenues et dans la mesure où le test d'adéquation donnait une p-valeur supérieure à 5%.

La suite de poids positifs w_t associée aux âges t peut être choisie de différentes manières. Cependant, avec l'approximation normale de l'estimateur des taux bruts \hat{q}_t , pour un âge t , pour n_t grand : $\hat{q}_t - q_t \rightarrow \mathcal{N}(0, \frac{q_t(1-q_t)}{n_t})$. Ainsi il peut être choisi de donner plus d'importance à la précision sur les âges où $\frac{q_t(1-q_t)}{n_t}$ est faible (où il y a une incertitude faible sur q_t). Alors plus n_t est grand, moins il y a une incertitude sur

q_t . Les w_t ont donc été choisis comme valant $w_t = \frac{n_t}{\bar{n}}$ avec $\bar{n} = \frac{\sum_{t=0}^m n_t}{m+1}$ l'effectif moyen sur la plage d'âges étudiée.

La méthode de lissage de Whittaker-Henderson a permis d'obtenir une courbe des \hat{q}_t lissée pouvant être comparée à la courbe des \hat{q}_t bruts. Voici en particulier le résultat obtenu pour l'étude sur l'ensemble des contrats :

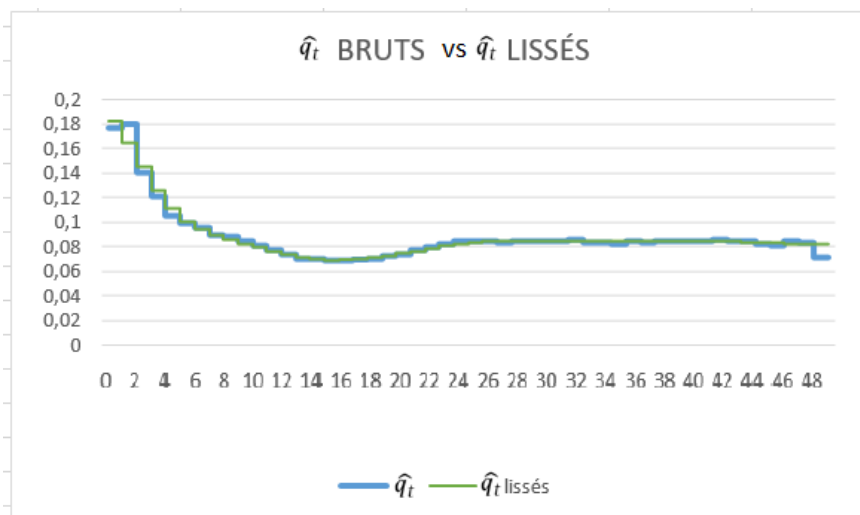


FIGURE 2.21 – Tracé de la courbe des \hat{q}_t bruts et de la courbe des \hat{q}_t lissée

2.2.4 Analyse des résultats

Finalement il a été obtenu des probabilités de résiliation lissées et des probabilités de survie à chaque âge, et une estimation de la fonction de survie :

- Pour l'ensemble des contrats ;
- Pour chaque segment MRH (Locataires d'appartement une pièce, deux pièces...).

Pour chacune des catégories de contrats étudiées, et listées ci-dessus, il est alors possible de calculer l'espérance de vie des contrats à la souscription. Pour cela, la probabilité, pour un contrat, d'être résilié au cours de chaque âge va être considérée. Nous devons donc d'abord nous rendre compte de la manière selon laquelle sont réparties les résiliations au sein d'une année. En effet, comme définie notamment par la société de conseil Spac actuair [17], l'espérance de vie à l'âge t se calcule ainsi : $E(t) = \sum_{k=0}^{\infty} k p_t \times q_{k+t} \times (k + \frac{1}{2})$. On considère un âge m qui correspond à l'âge auquel il n'y a plus de contrats en portefeuille, et alors finalement pour tout $i > m$, $S(i) = 0$ et pour $i = m$, $q_i = 1$. Et donc l'espérance de vie à la souscription est définie ainsi : $E(0) = \sum_{k=0}^m S(k) \times q_k \times (k + \frac{1}{2})$. Dans cette définition de l'espérance de vie il a été supposé que la distribution des résiliations au sein d'un âge fût uniforme. Alors lorsqu'un contrat est officiellement résilié à 1 an, par exemple, il est supposé qu'il vive en réalité en moyenne 1,5 an, car il ne sera pas possible de savoir à quel moment de l'année exactement il a été résilié. La valeur $\frac{1}{2}$ dans la formule représente donc la sous-estimation moyenne de l'âge à la résiliation.

Pour mieux comprendre comment est apparue cette sous-estimation à la résiliation il est possible de prendre un exemple :

Un contrat est souscrit le 31 avril 2016 et reste en portefeuille au moins un an, il est donc encore visible le 31 avril 2017. Ensuite si ce contrat est résilié le 31 mars 2018, il ne sera plus observable à son deuxième anniversaire en avril 2018. Dans la méthode d'observation des contrats utilisée, le contrat apparaîtra donc comme résilié après la première année car la méthode ne permet pas de déceler la date exacte de la résiliation. Ici, l'âge est sous-estimé de 11 mois.

En effet son âge réel à sa résiliation aura en réalité été :

dernier âge entier observé + sous estimation à la résiliation = 1 an + 11 mois = 1 an et 11 mois

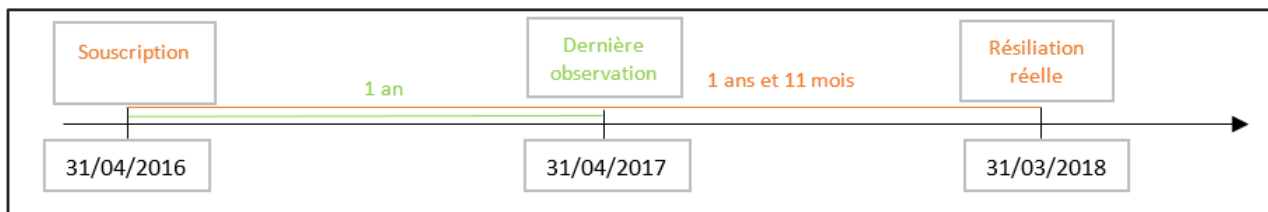


FIGURE 2.22 – Exemple de l’observation de la vie d’un contrat, par rapport au déroulement réel de sa vie

Cette formule de l’espérance ne peut s’appliquer directement à nos données car la distribution des résiliations au sein d’une année n’est pas nécessairement uniforme.

Pour pouvoir adapter cette formule nous allons donc observer la distribution des résiliations au sein des 12 mois suivants le dernier anniversaire des contrats avant leur résiliation, et calculer la sous-estimation moyenne qui a lieu.

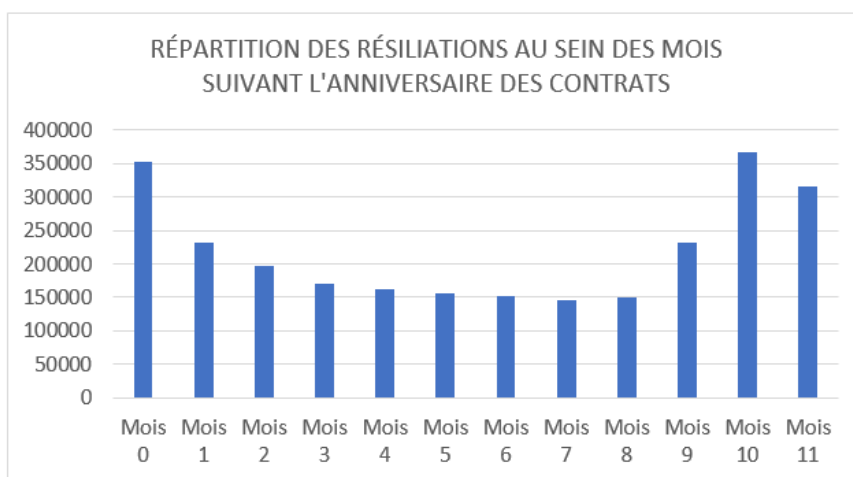


FIGURE 2.23 – Distribution des résiliations au sein des 12 mois suivants l’anniversaire des contrats

Nous pouvons en déduire que lorsque nous considérons qu’un contrat a été résilié à un âge $t \in \{1, \dots, m\}$, nous sous-estimons son âge à la résiliation de 5,67 mois en moyenne.

L’espérance de vie des contrats à la souscription se calcule donc ainsi :

$E(0) = \sum_{k=0}^m S(k) \times q_k \times (k + \frac{5,67}{12})$, m correspond à l’âge auquel il n’y a plus de contrats en portefeuille. Il n’a pas été possible de déterminer cet âge comme le portefeuille n’est pas éteint, et qu’à la fin de la période d’observation des contrats sont encore en portefeuille à tout âge. Cependant il a été choisi de supposer qu’au dernier âge observé (50 ans pour l’étude de l’ensemble des contrats et les études selon la segmentation MRH) l’estimation de la fonction de survie, dont la valeur est assez proche de zéro, pouvait être arrondie à zéro, et alors que l’estimation de la probabilité de résiliation à cet âge pouvait être mise égale à un. L’espérance de vie des contrats peut donc être estimée.

Le tableau suivant présente les espérances de vie de contrats estimées pour chaque étude :

Catégorie de contrats étudiée	Espérance de vie à la souscription
Ensemble des contrats	7 ans et 2 mois
Contrats LA1P	3 ans et 9 mois
Contrats LA2P	4 ans et 8 mois
Contrats LA3P	5 ans et 2 mois
Contrats LA4+	5 ans et 7 mois
Contrats LocM	5 ans
Contrats PM4-	10 ans
Contrats PM56	10 ans et 8 mois
Contrats PM7+	12 ans et 4 mois
Contrats ProA	10 ans et 3 mois
Contrats PNOA	10 ans et 5 mois
Contrats PNOM	9 ans et 3 mois

TABLE 2.1 – Estimation de l’espérance de vie à la souscription, au global, et selon les segments MRH

La courbe finale des estimations de la fonction de survie générale (avec les taux estimés et lissés au cours des étapes précédentes) pour un contrat est la suivante :

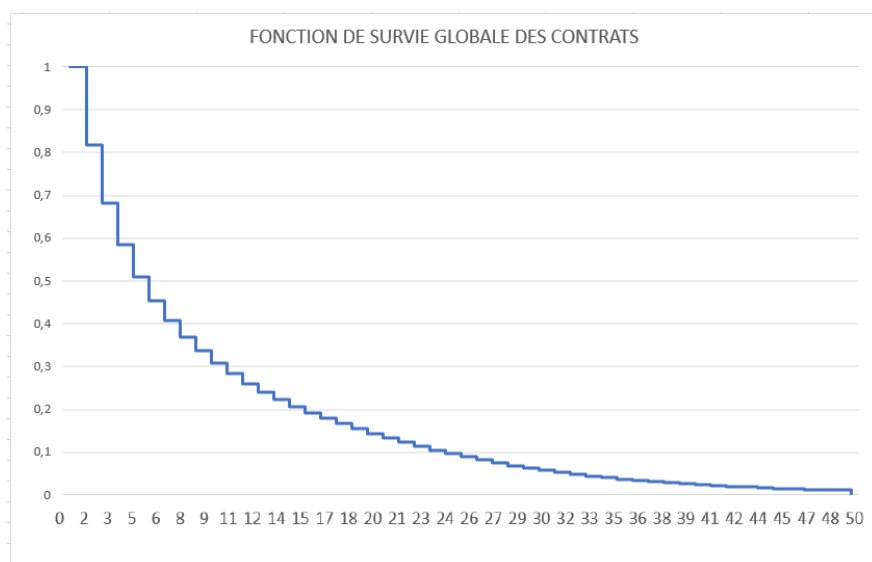


FIGURE 2.24 – Tracé final de la fonction de survie globale des contrats estimée

La courbe de la fonction de survie estimée pour un contrat, selon son segment MRH, est donnée par le graphique suivant :

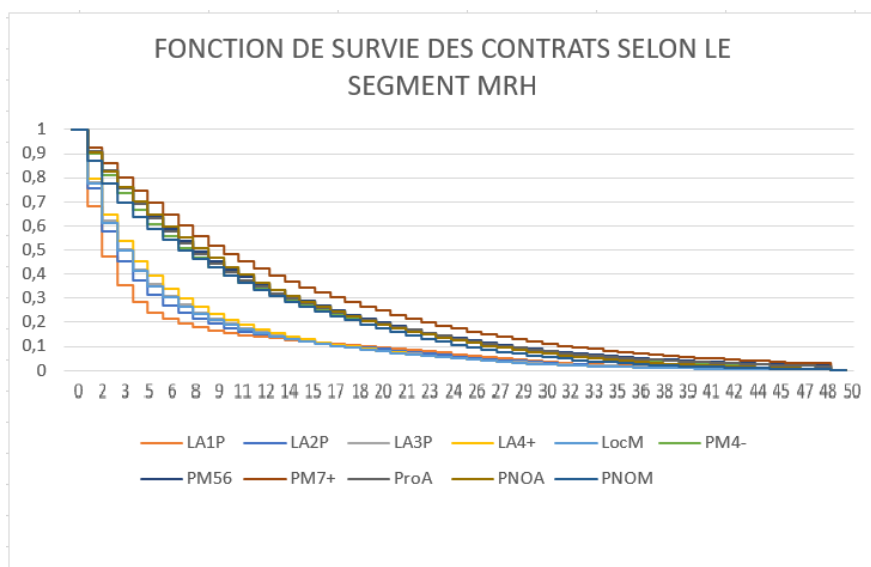


FIGURE 2.25 – Tracé final des fonctions de survie des contrats estimées selon le segment MRH

Cette représentation de la fonction de survie selon le segment MRH du contrat permet de voir graphiquement que le contrat d'assurance habitation d'un locataire aura une espérance de survie plus faible que le contrat d'assurance d'un propriétaire. Cela semble logique car 48% des résiliations observées sur le portefeuille sont liés à des déménagements, et par définition les locataires déménagent plus souvent que les propriétaires.

Ces résultats nous permettent d'avoir une vision sur la durée des contrats, mais il faut noter que les indicateurs de mortalité (résiliation) des contrats calculés ne correspondent pas nécessairement à la tendance de marché actuelle. En effet, en particulier pour les indicateurs des grands âges, nous avons utilisé les données de contrats souscrits plusieurs années auparavant, et donc souscrits dans un contexte différent de celui actuel. Dans le contexte actuel, nous pouvons supposer qu'il est plus facile pour les assurés de résilier leur contrat grâce à une communication simplifiée par l'accès à internet. Ainsi la tendance de marché d'aujourd'hui n'est pas forcément la même qu'hier.

2.3 Synthèse du chapitre

L'objectif de cette partie était d'avoir un premier aperçu des tendances de résiliation des contrats d'assurance MRH.

Un modèle de survie non-paramétrique, a été construit sur les données de la base Image MRH, par la méthode de Kaplan-Meier. Il a permis d'obtenir, notamment, pour chaque segment MRH, une estimation de la probabilité de résiliation à chaque âge sur 50 ans. Cependant les données des contrats à la souscription n'étaient pas disponibles dans la base Image, c'est donc le segment des contrats vu en 2016 qui a été utilisé pour segmenter les contrats et obtenir plusieurs modélisations de Kaplan-Meier. L'hypothèse forte selon laquelle les contrats ne changeaient pas de segment au cours de leur vie avait donc été prise.

Ce premier modèle a permis de voir que les contrats de locataires ont en moyenne une espérance de vie plus faible que les contrats de propriétaires.

— Chapitre 3 —

Modélisation semi-paramétrique de la probabilité de résiliation

3.1 Construction d'une nouvelle base de données

Précédemment, une première modélisation de courbes de survie par Kaplan-Meier a été effectuée sur 50 ans, en distinguant les contrats selon leur segment MRH. Cependant pour cette modélisation une hypothèse forte avait été prise : celle que les contrats ne changent pas de segment MRH au cours de leur vie, et ainsi que l'information dont nous disposons sur leur segment à partir de 2016 était la même qu'à leur souscription.

L'objectif de la suite de l'étude est de construire un modèle semi-paramétrique de prédiction de la courbe de survie associée à chaque contrat, selon ses caractéristiques à la souscription, afin de se rendre compte de l'espérance de vie à la souscription de chaque contrat. Ainsi il est préférable de disposer d'une base de données dans laquelle les informations contrats disponibles sont celles à la souscription. Et il est nécessaire de pouvoir observer le vieillissement des contrats sur une durée assez élevée (d'au moins 12 ans), étant donnée que la première étude a permis une estimation de l'espérance de vie des contrats MRH qui est de 7 ans et 2 mois en moyenne, et qui peut aller jusqu'à 12 ans et 4 mois pour la catégorie des contrats PM7+. La base Image disponible, répertoriant les informations des contrats à partir de 2016 n'est donc pas satisfaisante, c'est pourquoi une nouvelle base de données plus étendue a du être construite.

3.1.1 Définition de la base

Une base de données regroupant les informations des contrats souscrits à partir de 2006 a été construite.

La méthodologie utilisée pour retrouver les informations contrats a été la même que celle utilisée pour la construction de la base Image. Les informations ont donc été récupérées dans des bases de données regroupant les informations de tous les contrats d'assurance IARD d'AXA France, et séparant les contrats selon leur région AXA. Une extraction a donc été faite dans chacune de ces bases pour récupérer, et regrouper, les informations des contrats MRH. Ensuite il y a eu une traduction des options de garantie Confort en options de garanties Ma Maison. Et enfin une liaison avec une base client a pu être faite pour avoir les informations clients à la souscription, associées à chaque contrat.

Le périmètre des contrats de la nouvelle base de données est donc tel que la plage temporelle étudiée est composée de 15 années d'observations, du 01/01/2006 au 31/12/2020. Le portefeuille étudié est le portefeuille MRH classique.

La structure est telle qu'il y a une image par contrat souscrit entre 2006 et 2020. A chaque contrat sont associés ses caractéristiques à la fin de l'année de son affaire nouvelle, une variable qui indique le dernier

âge observé pour le contrat (par exemple 5,5 ans), et une variable qui indique si l'évènement de résiliation a été observé pour le contrat ou s'il a été censuré. Cette méthode de construction n'est pas parfaite, car il n'est pas possible d'avoir une vision du contrat à l'instant même de sa souscription mais seulement en décembre de la même année. Ceci est dû au fait que les bases de données MRH mensuelles ne sont plus disponibles avant 2016, seule une image par année l'est. C'est une limite qui aurait un impact pour environ 4,6% des contrats (d'après une étude sur les contrats souscrits entre 2016 et 2020), qui connaîtraient un remplacement durant l'année civile de leur souscription, susceptible de modifier au moins une de leurs caractéristiques.

Le dernier âge observé des contrats est exprimé en nombre de mois (les valeurs possibles sont des multiples de 1/12). Il est possible de ramener les valeurs à une vision annuelle en prenant la partie entière des valeurs indiquées.

Il est possible de représenter l'histogramme de la variable indiquant le dernier âge observé, qui sera la variable d'intérêt, pour les contrats non-censurés, l'objectif de ce chapitre étant d'estimer la probabilité de résiliation des contrats à chaque année de vie :

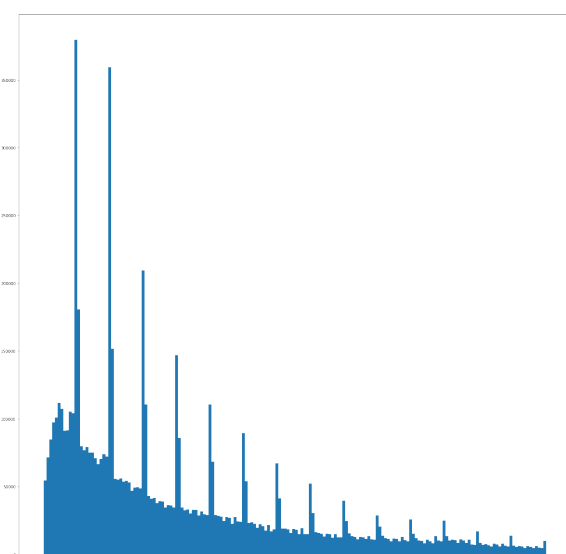


FIGURE 3.1 – Histogramme de la variable indiquant le dernier âge observé des contrats non-censurés

Il apparaît sur cet histogramme représentant le nombre de résiliations de contrats à chaque mois de vie possible, que les contrats ont tendance à être résiliés aux années complètes, c'est-à-dire à leurs anniversaires. Cela peut traduire un impact de la majoration au terme sur la volonté de résiliation des assurés de leurs contrats.

La base de données obtenue regroupe les informations de 6 703 880 contrats.

3.1.2 Description des variables

Les variables conservées dans la base de données apportent les informations associées aux contrats à l'affaire nouvelle, et elles peuvent être divisées en trois catégories.

Des variables présentent des informations sur le risque :

- le type de résidence du bien immobilier assuré (résidence principale, secondaire, louée, non occupée) ;
- la qualité de l'assuré (locataire, propriétaire, propriétaire non-occupant) ;
- le type d'habitation (maison, appartement au rez-de-chaussée, appartement à un autre niveau) ;
- la qualité d'occupation du bien ;
- l'étage du logement ;
- le taux d'objets de valeur dans le logement ;
- l'ancienneté du logement selon trois plages d'âges possibles ;
- le nombre de pièces du logement ;

- le segment MRH du contrat, qui correspond à une combinaison d'autres variables ;
- le capital total déclaré ;
- la surface des dépendances rattachées au logement ;
- le nombre de garanties optionnelles souscrites en plus du socle de garanties.

Des variables présentent des informations sur le type de contrat :

- s'il s'agit d'un contrat étudiant ;
- si le contrat est rattaché à une personne morale ;
- la région AXA auquel est rattaché le contrat ;
- le canal de distribution utilisé pour souscrire le contrat ;
- le fractionnement des primes du contrat ;
- la prime commerciale annuelle hors taxes du contrat (qui est donc la prime à l'affaire nouvelle) ;
- le type de franchise choisi pour le contrat ;
- l'année de souscription du contrat ;
- le produit associé au contrat (Confort, Ma Maison, ...).

Des variables présentent des informations sur l'assuré du contrat :

- si l'assuré est un enfant d'assuré ;
- le premier type de produits que l'assuré a souscrit chez AXA (par exemple un produit d'assurance santé, un produit d'épargne, etc.) ;
- l'âge de l'assuré ;
- l'ancienneté de l'assuré au sein d'AXA ;
- la catégorie socio-professionnelle de l'assuré ;
- le nombre de contrats AXA en cours déjà souscrits par le client.

Il est visible que certaines de ces variables sont qualitatives, comme par exemple le type de résidence du bien, la région AXA associée au contrat ou la catégorie socio-professionnelle de l'assuré.

Et d'autres variables sont quantitatives comme le nombre de pièces, ou l'âge du client.

Il est important de noter que l'indication de l'année de souscription des contrats permettra de se rendre compte de l'impact de l'année de souscription sur la tendance de résiliation des contrats.

3.1.3 Traitement de la base de données

3.1.3.1 Traitement des valeurs manquantes

Pour certains contrats, toutes les informations demandées ne sont pas disponibles, il y a alors des valeurs manquantes dans les données. Il a été déterminé le taux de valeurs manquantes pour chaque variable.

Le tableau regroupant ces taux est présenté en annexe 1.

Pour la majorité des variables, la part de valeurs manquantes est faible voire nulle. Ce taux est élevé seulement pour la variable qui indique la catégorie socio-professionnelle du client, pour laquelle il est de 29,69%, et la variable qui indique le nombre de contrats AXA déjà souscrits au nom du client, pour laquelle il est de 41,38%.

Plusieurs méthodes de traitement des valeurs manquantes ont été envisagées. La première méthode consistait à supprimer les individus qui présentent des valeurs manquantes. Mais cette méthode ne peut être utilisée que lorsqu'il y a peu d'individus concernés, sinon il y a un risque de biaiser les résultats. Dans nos données, avec cette méthode, au moins 41,38% des contrats auraient été supprimés de la base. C'est une proportion très importante, il était donc finalement inenvisageable d'utiliser cette méthode.

La deuxième méthode consistait à remplacer les valeurs manquantes par une valeur fixe, faisant partie des valeurs possibles déjà attribuées par les variables. Pour chaque variable quantitative il aurait pu, par exemple, être mesurée la valeur médiane, et cette valeur médiane aurait pu être choisie comme valeur par défaut pour remplacer toute valeur manquante de la variable. Pour chaque variable qualitative, c'est la modalité la plus représentée qui aurait pu être choisie comme modalité de remplacement des valeurs

manquantes. Cependant cette méthode est aussi à privilégier lorsqu'il y a peu de valeurs manquantes, car cela peu réduire la variance des variables. Cette méthode n'a donc pas été retenue pour cette étude.

Une autre méthode étudiée était une méthode de prédiction des "vraies valeurs" associées aux contrats présentant des valeurs manquantes, avec des arbres de décision. Pour chaque variable, un arbre de décision prédisant la valeur ou la modalité de la variable aurait été construit sur les données pour lesquelles la valeur n'est pas manquante. Cet arbre aurait ensuite prédit une valeur de remplacement de la valeur manquante, pour chaque contrat, en fonction de ses autres caractéristiques. Cette méthode semblait pertinente et plus adaptée à nos données. Cependant la question s'est posée de savoir pourquoi certaines des données sont manquantes. Cela serait dû à des anomalies qui ne semblent pas pouvoir être expliquées et continueront d'affecter nos données futures. Mais il est possible qu'il y ait en réalité une explication à ces anomalies, qui ne peut être perçue aujourd'hui. Il est possible que ces anomalies n'affectent qu'une certaine catégorie de contrats. Le fait d'avoir une ou plusieurs valeurs manquantes sur un individu pourrait finalement implicitement apporter une certaine information sur l'individu en le distinguant des autres contrats, pour une raison que nous ne pouvons expliquer aujourd'hui mais qui n'est éventuellement pas négligeable. Remplacer les valeurs manquantes en prédisant des "vraies valeurs" supposées pourrait finalement provoquer une perte d'information. Cette méthode de traitement des valeurs manquantes n'a alors pas été retenue non-plus.

Finalement la méthode de traitement des valeurs manquantes qui a été retenue, consistait à construire une valeur ou modalité de variable, indiquant qu'il s'agissait d'une valeur manquante. Les variables quantitatives, telles que le nombre de pièces du logement ou l'âge du client, ne peuvent prendre que des valeurs positives. Alors pour distinguer les contrats présentant une valeur manquante, la valeur "-1" a été choisie comme valeur de remplacement des valeurs manquantes de ces variables. Pour les variables qualitatives, telles que le type de logement ou la qualité de l'assuré, une modalité "nan" a tout simplement été construite pour les contrats ayant une valeur manquante. Ainsi le cas de présence d'une valeur manquante pour une variable est considéré comme une caractéristique à proprement parler et il n'y a pas d'éventuelle perte d'information.

3.1.3.2 Traitement des valeurs aberrantes

Des valeurs aberrantes, c'est à dire des valeurs qui sont, de manière certaine ou presque certaine, fausses ou illogiques, ont été observées au niveau de variables quantitatives. Elles correspondaient à des valeurs étonnement élevées ou faibles, ou même négatives alors que c'était normalement impossible. Un mécanisme de correction des valeurs aberrantes a donc été mis en place pour chaque variable concernée. Il consistait à définir une valeur minimale et une valeur maximale possible, et lorsqu'un contrat présentait une valeur inférieure au minimum choisi alors cette valeur était remplacée par le minimum, et de la même manière si la valeur était supérieure au maximum choisi elle était remplacée par le maximum. Le tableau suivant récapitule les corrections effectuées sur les variables concernées :

Variable concernée	Étendue des valeurs	Corrections
Nombre de pièces	De 1 à 99	Maximum choisi à 24
Âge du client	De -30 ans à 126 ans	Minimum choisi à 18, maximum à 117
Ancienneté du client	De -5 ans à 120 ans	Minimum choisi à 0, Maximum à 41
Surface des dépendances	De 0 m ² à 2 019 061 m ²	Maximum choisi à 500

TABLE 3.1 – Correction des valeurs aberrantes sur les variables concernées

Pour la variable indicatrice du nombre de pièces du logement assuré, un maximum a été choisi à 24 car c'est normalement le nombre maximal de pièces autorisé d'après les guides de souscription des produits d'assurance MRH d'AXA France. Pour cette variable il y avait alors 0,001% de valeurs aberrantes.

Pour la variable qui indique la surface des dépendances, de la même manière un maximum a été choisi à 500 m² car c'est la surface maximale autorisée par les guides de souscription. 0,130% des contrats ont donc eu une surface des dépendances réaffectée à 500 m².

Pour l'âge du client, un minimum à été choisi à 18 ans, car cela correspond à la majorité en France, et un maximum a été choisi à 117 ans car c'est l'âge maximal atteint en France. Il y a eu une réaffectation de valeur pour 0,011% des contrats.

Pour l'ancienneté du client, un minimum a été choisi à 0 an pour qu'il n'y ait pas d'anciennetés négatives, cela concernait 0,008% des contrats. Et un maximum a été choisi à 41 ans car cela ne semblait pas très réaliste qu'un client ait une ancienneté plus élevée, et cela ne concernait que 1,079% qui ont donc vu leur valeur d'ancienneté être réaffectée à 41 ans.

3.1.3.3 Étude des corrélations entre variables

La base de données obtenue a été construite dans le but de construire un modèle de survie semi-paramétrique sur la durée de vie des contrats. Le modèle utilisé sera un modèle de Cox. Pour la construction d'un modèle de ce type il est préférable d'avoir une éventuelle corrélation entre variables limitée. En effet en cas de présence de corrélations importantes il serait difficile de séparer les effets individuels de chaque variable explicative sur la variable réponse. Une sélection de variables à conserver dans la base a donc été faite pour limiter les niveaux de corrélation, et limiter un risque de sur-information.

Pour calculer les niveaux de corrélation entre les variables quantitatives de la base, le coefficient de corrélation de Bravais-Pearson a été utilisé. Ce coefficient est notamment présenté par MAUMY-BERTRAND M. et BERTRAND F. dans leur livre *Initiation à la statistique avec R* [4].

Pour deux variables aléatoires X et Y, le coefficient de corrélation de Bravais-Pearson théorique vaut :

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Avec $Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$ la covariance entre X et Y et $Var(X)$, $Var(Y)$ les variances de X et Y.

Le coefficient de Bravais-Pearson observé vaut :

$$r(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x}_n)(y_i - \bar{y}_n)}{(\sum_{i=1}^n (x_i - \bar{x}_n)^2)(\sum_{i=1}^n (y_i - \bar{y}_n)^2)}$$

$r(x, y)$ correspond à une réalisation de $\rho(X, Y)$, n est le nombre d'observations, pour tout $i \leq n$ x_i et y_i sont les valeurs observées des variables X et Y pour l'individu i , et \bar{x}_n et \bar{y}_n sont les moyennes des valeurs observées des variables X et Y sur l'ensemble des individus.

Les corrélations de Bravais-Pearson entre toutes les variables quantitatives de la base ont pu être déterminées :

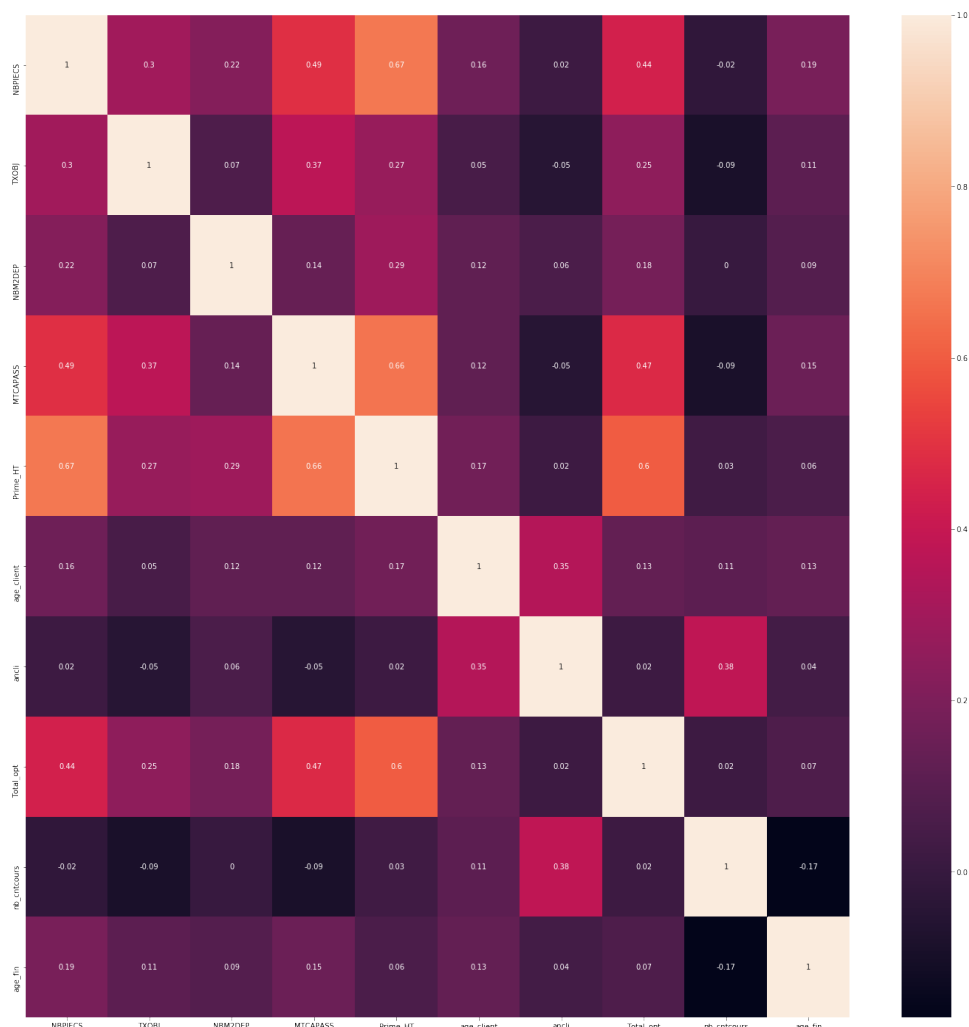


FIGURE 3.2 – Étude des corrélations entre les variables quantitatives de la base de données

Il a été décidé de considérer les corrélations au delà du seuil de 0,4/-0,4, comme trop importantes pour être conservées.

Alors il est apparu que la variable indiquant la prime hors taxes du contrat était très corrélée avec les variables indiquant le nombre de pièces du logement, le montant de capital déclaré, et le nombre d'options du contrat. Ces corrélations sont logiques étant donné que la prime d'un contrat dépend des caractéristiques du risque assuré. Il a donc été choisi de ne pas conserver la variable indiquant la prime hors taxes.

Les variables indiquant le nombre de pièces, le montant de capital déclaré, et le nombre d'options, sont également fortement corrélées entre elles. Entre ces trois variables, il a été choisi de conserver la variable indiquant le nombre de pièces. Car c'est une des variables les plus importantes pour distinguer les différents niveaux de risques des contrats.

Pour faire une sélection au niveau des variables qualitatives, les significations des variables ont été comparées et des variables ont été retirées de manière à éviter un sur-apprentissage.

La variable indiquant l'étage du logement a été retirée car cette information est déjà partiellement apportée par la variable indiquant le type d'habitation, qui distingue les appartements au rez-de-chaussée des appartements à un autre niveau).

La variable indiquant la catégorie socio-professionnelle du client a été retirée. Non-seulement il a été vu précédemment qu'elle présentait une part importante des valeurs manquantes, mais également il est surtout important de faire une distinction entre les clients étudiants et les autres, donc il était préférable de conserver uniquement la variable indiquant si le client est un étudiant.

La variable indiquant le segment MRH du contrat a été retirée car cette variable est en réalité une

combinaison d'autres variables sur le risque (la qualité de l'assuré, le type de résidence, le nombre de pièces et la qualité d'occupation du bien).

3.2 Modélisation semi-paramétrique : Modèle de Cox

Le modèle de Cox offre la possibilité d'intégrer des variables explicatives au modèle. En effet c'est un modèle de durée semi-paramétrique, c'est à dire qu'il prend en compte des paramètres et que la loi de la fonction de survie n'est ni prédéterminée ni connue. La construction d'un modèle de Cox, pour prédire la durée de vie des contrats d'assurance MRH, permettra d'obtenir pour chaque contrat une estimation de sa fonction de survie selon ses caractéristiques à la souscription.

3.2.1 Définition du modèle

Le modèle de Cox dit à risques proportionnels a été présenté par COX D. dans un article paru en 1972, "Regression Models and Life-Tables" dans *Journal of the Royal Statistical Society : Series B (Methodological)* [6].

Le modèle de Cox est un modèle multivarié qui permet de mesurer l'effet des facteurs qualitatifs ou quantitatifs sur la survie, en exprimant le risque instantané de survenance de l'événement étudié, aussi appelé fonction de hasard.

La fonction de hasard se définit de la manière suivante :

$$h(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln(S(t)) \text{ avec } f \text{ la fonction densité de la durée de vie des contrats.}$$

$$\text{Dans le cas discret elle se définit ainsi : } h(t) = \mathbb{P}(T = t | T \geq t) = \frac{\mathbb{P}(T=t)}{S(t)}.$$

Il en résulte qu'à partir de la fonction de hasard, la loi de la durée de vie peut être déterminée avec la relation suivante :

$$S(t) = \exp\left(-\int_0^t h(s) ds\right).$$

Il est possible de noter $H(t) = \int_0^t h(s) ds$ la fonction de hasard cumulée, ainsi $S(t) = \exp(-H(t))$.

Dans le cas discret, la fonction de survie de T se définit donc ainsi : $S(t) = \exp(-\sum_{k=0}^t h(k))$. Et la fonction de hasard cumulée de la manière suivante : $H(t) = \sum_{k=0}^t h(k)$. Dans la suite de l'explication du modèle, et dans son application, le cas discret sera privilégié.

Le modèle de Cox repose sur l'idée selon laquelle la fonction de hasard d'un individu (ici d'un contrat) se décompose comme un risque de base commun à tous les individus, dépendant du temps, et une fonction linéaire des facteurs qualitatifs et quantitatifs caractérisant chaque individu.

Le modèle de Cox s'écrit donc ainsi :

$$h(t|Z_i) = \underbrace{h_0(t)}_{\text{risque de base}} \underbrace{\exp(Z_i \beta)}_{\text{Risque relatif}}$$

Z_i correspond au vecteur des covariables pour un individu i (Z_{i1}, \dots, Z_{in}). β correspond au vecteur des coefficients inconnus associés aux caractéristiques des individus, les paramètres du modèle, $(\beta_1, \dots, \beta_n)$.

L'objectif du modèle est alors d'estimer le risque de base $h_0(t)$ et les coefficients associés aux caractéristiques des individus (des contrats) β_1, \dots, β_n .

Le modèle de Cox repose sur deux hypothèses de base :

- Hypothèse de proportionnalité, au cours du temps, entre le risque instantané et le risque de base :

$$\ln\left(\frac{h(t|Z_i)}{h_0(t)}\right) = Z_i \beta$$

Cette hypothèse peut aussi s'appeler l'hypothèse de log-linéarité car il en découle :

$$\ln(h(t|Z_i)) = \ln(h_0(t)) + Z_i \beta$$

- Hypothèse de linéarité (ou des risques proportionnels), le rapport des fonctions de risque instantané pour deux individus i et j ne dépend pas du temps :

$$\frac{h(t|Z_i)}{h(t|Z_j)} = \frac{\exp(Z_i \beta)}{\exp(Z_j \beta)} = \text{constante}$$

Donc, pour tout individu i , les β_k représentent l'effet de la caractéristique Z_{ik} sur la survenance de l'évènement étudié :

- Si β_k est nul ($\exp(\beta_k) = 1$), alors la k ième caractéristique n'a pas d'influence sur la survenance de l'évènement.

- Si β_k est positif ($\exp(\beta_k) > 1$), toutes choses égales par ailleurs, la présence de cette particularité, pour une variable qualitative, ou des valeurs élevées de la k ième caractéristique, pour une variable quantitative, sont associées à un risque instantané plus important.

- Et ainsi si β_k est négatif ($\exp(\beta_k) < 1$), alors la présence de cette particularité, pour une variable qualitative, ou des valeurs élevées de la k ième caractéristique, pour une variable quantitative, sont associées à un risque instantané plus faible.

Pour estimer les paramètres β_k , la méthode du maximum de vraisemblance est utilisée.

En imaginant que les données soient constituées de s durées de survie distinctes, triées, telles que $t_1 < \dots < t_s$. Et en appelant $R_i(t)$ l'indicatrice que l'individu i soit encore à risque au temps $t \in (t_l)_{s \geq l \geq 1}$. Il est possible d'exprimer la probabilité conditionnelle que cet i ième individu meurt (que le contrat soit résilié) en t , sachant qu'un individu parmi les y individus meurt en t :

$$\frac{h(t|Z_i)}{\sum_{j=1}^y R_j(t)h(t|Z_j)} = \frac{h_0(t) \times \exp(Z_i \beta)}{\sum_{j=1}^y R_j(t) \times h_0(t) \times \exp(Z_j \beta)} = \frac{\exp(Z_i \beta)}{\sum_{j=1}^y R_j(t) \times \exp(Z_j \beta)}$$

Il est visible que la formule obtenue correspond au risque instantané de l'individu i au temps t divisé par la somme des risques instantanés des individus encore à risque en t .

La vraisemblance partielle proposée par Cox correspond alors au produit de ces quantités pour l'ensemble des individus, en prenant en compte la possibilité que des individus soient censurés :

$$\mathcal{L}(\beta) = \prod_{i=1}^y \left(\frac{\exp(Z_i \beta)}{\sum_{j=1}^y R_j(t) \times \exp(Z_j \beta)} \right)^{d_i} \text{ avec } d_i \text{ qui vaut 0 si l'individu } i \text{ est censuré et 1 sinon.}$$

En maximisant la vraisemblance de Cox il est alors possible d'estimer le vecteur de paramètres :

$$\hat{\beta} = \operatorname{argmax}_{\beta} (\mathcal{L}(\beta)).$$

Il est possible de calculer la log vraisemblance : $\ln(\mathcal{L}(\beta)) = \sum_{i=1}^y d_i [Z_i \beta - \ln(\sum_{j=1}^y R_j(t) \exp(Z_j \beta))]$.

Ensuite la dérivée de la log vraisemblance par rapport à β vaut :

$$\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta} = \sum_{i=1}^y d_i \left[Z_i - \frac{\sum_{j=1}^y R_j(t) \times Z_j \times \exp(Z_j \beta)}{\sum_{j=1}^y R_j(t) \times \exp(Z_j \beta)} \right].$$

La dérivée de la log vraisemblance vaut 0 lorsque $\beta = \hat{\beta}$. Alors la détermination de l'estimation de β se fait en en résolvant l'équation $\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta} = 0$.

Plusieurs méthodes de convergence peuvent être utilisées pour résoudre cette équation. Deux méthodes seront étudiées et présentées dans les sous-parties suivantes.

Le risque de base cumulé peut alors être estimé ainsi :

$$\hat{H}_0(t) = \sum_{t_l \leq t} \frac{\sum_{j=1}^y R_j(t_l) - \sum_{j=1}^y R_j(t_{l+1})}{\sum_{j=1}^y R_j(t_l) \exp(Z_j \hat{\beta})}$$

Et la fonction de survie peut donc également être estimée pour tout individu i :

$$\hat{S}(t|Z_i) = \exp(-\hat{H}_0(t) \times \exp(Z_i \hat{\beta}))$$

Implémenter un modèle de Cox peut donc permettre d'estimer la courbe de survie d'un contrat en portefeuille, en prenant en compte l'impact de ses caractéristiques à la souscription.

3.2.2 Vérification de l'hypothèse de proportionnalité des risques instantanés

Pour implémenter le modèle il faut d'abord vérifier l'hypothèse de proportionnalité des risques instantanés, qui est l'hypothèse principale du modèle de Cox.

Cette hypothèse peut être vérifiée graphiquement pour chaque variable. Pour cela la fonction $\ln(-\ln(\hat{S}(t)))$ peut être tracée, en utilisant l'estimation de la fonction de survie de Kaplan-Meier, pour les ensembles d'individus représentant chaque classe de la variable étudiée. Si les courbes sont parallèles pour les différentes caractéristiques, alors l'effet de la covariable est considéré comme indépendant du temps. Mais si les courbes se croisent alors l'effet de la covariable n'est pas indépendant du temps.

Par exemple il est visible que pour la variable indiquant si le contrat est un contrat étudiant, l'hypothèse est vérifiée :

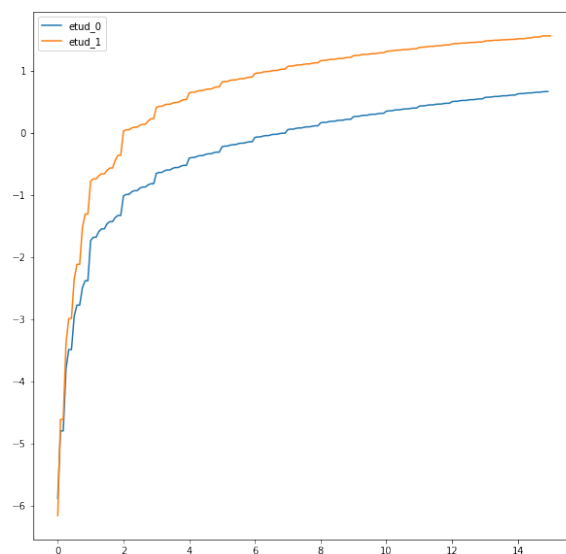


FIGURE 3.3 – Test de proportionnalité pour la variable indicatrice d'un contrat étudiant

Cette vérification a été effectuée sur l'ensemble des variables et a permis une première sélection de variables pour le modèle. Il a alors été décidé de ne pas prendre en compte la variable indiquant la classe d'âge du logement, pour qui l'hypothèse n'était pas vérifiée.

3.2.3 Modélisation classique

Dans un premier temps un modèle de Cox a été implémenté de manière classique. C'est-à-dire que la maximisation de la vraisemblance, pour déterminer les paramètres du modèle, se fait avec l'algorithme de Newton-Raphson. Dans la suite de l'étude, un autre algorithme sera utilisé, celui du *Gradient Boosting*.

3.2.3.1 Présentation de la méthodologie

L'utilisation de l'algorithme de Newton-Raphson pour résoudre des équations numériques, est une méthode itérative. Elle a été initiée par NEWTON I. vers 1669 sur des exemples numériques. Et dix ans plus tard, RAPHSON J. a mis en évidence une formule de récurrence.

Selon cet algorithme, pour déterminer l'estimateur $\hat{\beta}$ de β , dans l'équation $\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta} = 0$, il faut partir d'une valeur initiale $\hat{\beta}_0$. Puis à chaque étape $k+1$, $\hat{\beta}_{k+1} = \hat{\beta}_k - H^{-1}(\hat{\beta}_k)U(\hat{\beta}_k)$ avec H la matrice Hessienne de la log vraisemblance et U le vecteur gradient de la log-vraisemblance, sera de plus en plus proche de la valeur permettant de résoudre l'équation.

Il existe une variante qui consiste à remplacer la matrice Hessienne H par son espérance, la formule devient alors : $\hat{\beta}_{k+1} = \hat{\beta}_k - I^{-1}(\hat{\beta}_k)U(\hat{\beta}_k)$ avec I l'information de Fisher définie par $I = -\mathbb{E}\left[\frac{\partial^2 \ln(\mathcal{L}(\beta))}{\partial \beta \partial \beta^t}\right]$.

3.2.3.2 Implémentation informatique du modèle

La construction du modèle de Cox pour la prédiction de la survie des contrats a été faite sur la base de données construite avec les informations des contrats souscrits à partir de 2006. La base de données a été divisée en une base d'entraînement, représentant 80% de la base, et une base test, représentant les 20% restants. Le modèle a été implémenté avec les données de la base d'entraînement et les prédictions sur la base test ont été étudiées.

Pour cela le langage *Python* (version 3.7.3) a été utilisé, et en particulier le package *lifelines* (version 0.25.10). Le package *lifelines* présente des fonctions utiles à l'analyse de survie, et en particulier la fonction *CoxPHFitter* qui permet de construire un modèle de Cox à partir de données de survie. Le modèle obtenu prédit directement la valeur de risque instantané de chaque contrat, et il est possible d'obtenir la fonction de survie pour chaque contrat.

Pour l'utilisation de cette fonction, plusieurs paramètres ont été choisis :

- Le seuil de significativité, *alpha*, à 0,05.
- La méthode de traitement des égalités, *tie_method*, Efron. Cette méthode permet de traiter les cas où plusieurs observations ont lieu aux mêmes âges, ce qui n'est pas pris en compte dans la définition initiale de la vraisemblance partielle de Cox. Cela consiste à utiliser une expression de la vraisemblance partielle telle que s'il n'y a pas d'égalités, elle correspond à la vraisemblance partielle de Cox initiale.
- La pénalisation dans la régression, *penalizer*, à 0,01. Cela permet d'associer une pénalité à la taille des coefficients, durant la régression. Et cela permet d'améliorer la stabilité des estimations et de mieux contrôler les cas de corrélation entre des covariables.

La fonction prend en entrée les informations des contrats, le dernier âge observé des contrats, et l'indicateur de censure. Les variables caractéristiques des contrats ont préalablement été transformées en indicatrices, et les caractéristiques les plus représentées ont été définies comme valeurs de référence, afin de représenter le risque de base. Le modèle de Cox va alors se construire en estimant l'effet des caractéristiques sur le risque de base.

3.2.3.3 Sélection des variables

Pour sélectionner des variables dans la construction du modèle la méthode de sélection *backward* a été employée, en se basant sur la concordance comme indicateur de qualité.

L'indicateur de concordance statistique est notamment présenté dans une étude de THERNEAU T. et ATKINSON E. *Concordance* [2]. C'est un indicateur qui, contrairement aux métriques d'erreurs classiques,

peut être utilisé en présence de données censurées, c'est pourquoi il est utilisé dans l'analyse de modèles de Cox. La valeur de la concordance se situe entre 0 et 1. Elle représente la part des paires de contrats dont les durées de survie prédites sont correctement ordonnées, l'ordre est le même que celui des durées réellement observées, par rapport à l'ensemble des paires de contrats qui peuvent être ordonnées selon leur dernier âge observé. Alors il peut être déduit que plus la valeur de la concordance est proche de 1, meilleure est la prédiction de la survie.

La méthode *backward* consiste à partir d'un modèle complet, où toutes les variables dont nous disposons sont explicatives, et retirer une à une les variables afin de voir quel retrait a le moins d'impact sur la performance de la prédiction du modèle, et donc voir au fur et à mesure quelles variables ne seront pas considérées comme explicatives. L'impact des retraites de variables a donc été évalué en observant les valeurs de la concordance. Le retrait d'une variable ne faisant pas diminuer la valeur de la concordance pouvait donc être considéré comme légitime. Le retrait des variables indiquant, le nombre de mètres carrés des dépendances, l'ancienneté de l'assuré, le sexe de l'assuré, ou le nombre de contrats déjà souscrits par l'assuré au sein d'AXA, a alors été envisagé.

3.2.3.4 Étude de la signification des coefficients

Lors de la construction du modèle il est possible de tester l'influence significative des coefficients à l'aide du test de Wald. Le test de Wald est notamment présenté dans le cours *Statistique inférentielle : tests* de MONBET V [10].

L'objectif est de se rendre compte, pour chaque coefficient, s'il a réellement une influence significative sur la survie. L'hypothèse de référence du test est donc la suivante :

$$\begin{cases} H_0 & : \beta_j = 0 \\ & \Leftrightarrow \text{Pas d'influence significative de la covariable} \\ H_1 & : \beta_j \neq 0 \end{cases}$$

La statistique du test est $\frac{\sqrt{n}\hat{\beta}}{se(\hat{\beta})} \rightarrow \mathcal{N}(0, 1)$ où $se(\hat{\beta})$ est l'écart-type de l'estimateur de β .

L'hypothèse nulle H_0 , de non significativité de l'influence de la covariable, est rejetée si la p-valeur est inférieure à un seuil fixé $\alpha = 0,05$. Ce seuil correspond à l'erreur de première espèce, c'est-à-dire l'erreur éventuelle que l'hypothèse nulle H_0 soit vue comme fausse, alors qu'elle est vraie.

L'ensemble des p-valeurs obtenues ont été étudiées, et les variables indicatrices qui présentaient une p-valeur supérieure à 0,05 n'ont pas été intégrées à la modélisation.

3.2.3.5 Construction d'un modèle sur les âges entiers

Une première modélisation de la survie avec un modèle de Cox a été effectuée. Elle a été faite sur les données des contrats souscrits à partir de 2006, pour lesquels au lieu de prendre le dernier âge observé exact, il a été pris l'entier supérieur du dernier âge observé. Ainsi si, par exemple, un contrat était résilié à l'âge 0,5, alors étant donné qu'il n'était plus observable qu'à partir de l'âge entier 1, son âge de fin d'observation était défini à 1.

L'objectif était de prédire des courbes de survie moyennes par segment MRH, par modèle de Cox, directement comparables avec les courbes de survie estimées précédemment par Kaplan-Meier, par segment, entre 1970 et 2020. Même si lors de la construction de ces courbes de Kaplan-Meier, le segment MRH observé de chaque contrat n'était pas nécessairement le segment MRH à la souscription, il y a eu une volonté de se rendre compte de la proximité de ces estimations avec les nouvelles estimations faites avec un modèle de Cox.

Avec ce modèle une courbe de survie par contrat a été prédite. Ensuite il a été décidé d'observer la courbe de survie prédite moyenne sur tous les contrats associés à chaque segment MRH. Il est visible que la survie des contrats de propriétaires est prédite comme globalement meilleure que celle des locataires (pour rappel l'estimation de Kaplan-Meier sur la base Image de 1970 à 2020 avait apporté la même observation) :

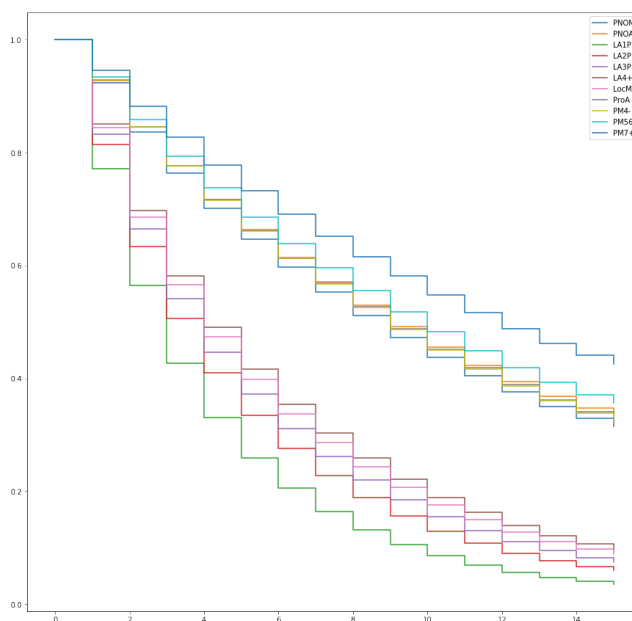


FIGURE 3.4 – Courbes de survie prédites, par le modèle de Cox sur les âges entiers, moyennes par segment MRH

Pour analyser la performance de ce modèle et des suivants qui seront mis en place, et éventuellement les comparer, les indicateurs pouvant être utilisés seront :

- La concordance, de l'application du modèle sur la base d'entraînement ou sur la base de test ;
- L'écart entre les valeurs estimées de la courbe de survie moyenne, à chaque âge entier, sur différents sous-ensembles de données, et les valeurs de la courbe de survie estimée par Kaplan-Meier sur les mêmes données ;
- L'écart entre l'estimation d'une espérance de vie moyenne, en utilisant les valeurs estimées de la courbe de survie moyenne aux âges entiers, sur différents sous-ensembles de données (sur les différents segment MRH), et l'estimation d'une espérance de vie en utilisant les valeurs de la courbe de survie estimée par Kaplan-Meier sur les mêmes données ;
- L'écart entre la moyenne des espérances de vie prédites sur les contrats non-censurés, sur différents sous-ensembles de données, et la moyenne des durées de vie réellement observées sur ces mêmes contrats.

Il est donc possible de comparer ces courbes prédites moyennes par segment, avec les courbes de survie de Kaplan-Meier estimées sur la base Image sur 50 ans, mais également avec des nouvelles courbes de survie de Kaplan-Meier estimées par segment sur la nouvelle base de données sur les contrats souscrits à partir de 2006. En particulier il sera préférable de comparer les courbes moyennes par segment prédites avec le modèle de Cox, avec ces nouvelles courbes de Kaplan-Meier pour lesquelles le segment MRH associé aux contrats est bien le segment à la souscription.

Sur l'ensemble des contrats, le modèle de Cox a tendance à sur-estimer les valeurs de la courbe de survie moyenne de 7,16% , par rapport aux estimations de Kaplan-Meier :

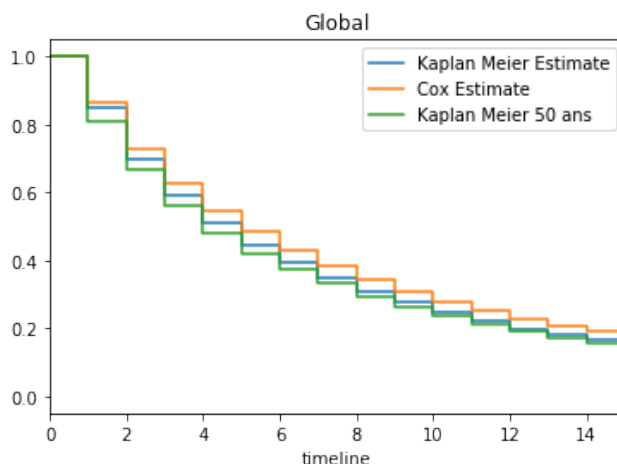


FIGURE 3.5 – Comparaison entre la courbe de survie moyenne prédite par le modèle de Cox sur 15 ans, la courbe de survie de Kaplan-Meier estimée sur 15 ans, et la courbe de survie de Kaplan-Meier estimée sur la base Image sur 50 ans, sur l'ensemble des contrats

Les segments sur lesquelles le modèle de Cox sur-estime le moins les valeurs de la courbe de survie moyenne, par rapport aux estimations de Kaplan-Meier, sont le LA3P (3,10%) et LA4+ (3,90%).

Et les segments sur lesquelles il surestime le plus sont le LA1P (16,90%), puis PNOM (8,50%).

Les graphiques de comparaison pour chaque segment sont présentés en annexe 2.

En appliquant ce modèle à la base d'entraînement et à la base test les valeurs de concordance suivantes ont été obtenues :

	Base d'entraînement	Base test
Concordance	0,6794	0,6793

TABLE 3.2 – Valeurs de la concordance de l'application du modèle construit sur les âges entiers

La valeur de l'indicateur de concordance d'un modèle de survie est généralement compris entre 0,55 et 0,75. Le modèle de Cox construit au cours de cette étude présente une concordance d'environ 0,68 ce qui est donc satisfaisant. De plus, l'indicateur est stable entre l'application du modèle à la base d'entraînement et à la base test.

3.2.3.6 Construction d'un modèle sur les âges exacts

Une nouvelle modélisation de la durée de vie des contrats par Cox a été mise en oeuvre, mais cette fois-ci sur les durées de vie exactes des contrats (c'est-à-dire exprimées en mois). Cela a permis d'obtenir une prédiction de la fonction de survie plus précise, pour chaque contrat. Cela permet de voir la tendance de résiliation au cours d'une année, et en particulier de visualiser l'effet du terme, qui augmenterait le risque de résiliation, d'après l'étude descriptive effectuée précédemment.

A nouveau il a été possible d'observer la courbe de survie prédite moyenne sur tous les contrats associés à chaque segment MRH, mais aussi sur les contrats associés aux segments LA1P, LA2P et LA3P selon la présence de la formule étudiante :

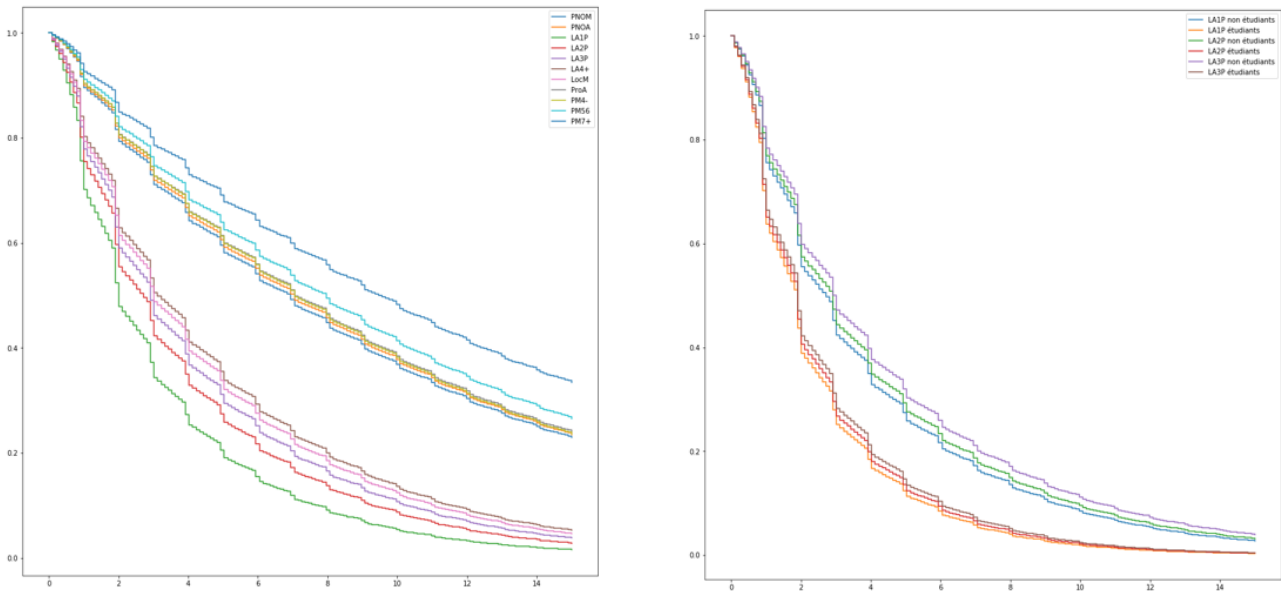


FIGURE 3.6 – Courbes de survie prédites, par le modèle de Cox sur les durées de vie exactes, moyennes par segment (à gauche) et selon la présence de formule étudiante et par segment MRH (à droite)

Il apparaît que les contrats avec la formule étudiante ont une survie plus faible que celle des contrats sans la formule. Cela semble logique car les étudiants ont tendance à avoir une mobilité plus importante, et alors restent moins longtemps dans un même logement et auront alors plus tendance à résilier leur contrat.

Désormais il est possible de comparer ces courbes prédites moyennes par segment, avec les courbes prédites par segment avec le modèle sur les âges entiers, et les courbes de survie de Kaplan-Meier estimées sur la base Image sur 50 ans, ainsi que les courbes de survie de Kaplan-Meier estimées par segment sur la nouvelle base de données sur les contrats souscrits à partir de 2006. La comparaison a d'abord été faite sur l'ensemble des contrats :

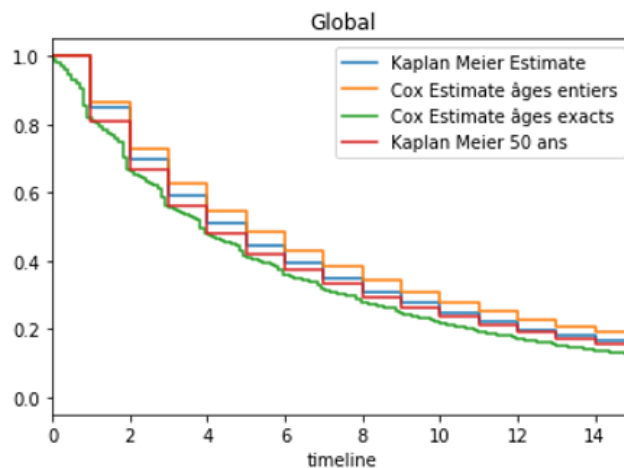


FIGURE 3.7 – Comparaison entre la courbe de survie moyenne prédite par le modèle de Cox sur les âges exacts sur 15 ans, celle prédite par le modèle sur les âges entiers, la courbe de survie de Kaplan-Meier estimée sur 15 ans, et la courbe de survie de Kaplan-Meier estimée sur la base Image sur 50 ans, sur l'ensemble des contrats

Contrairement au modèle sur les âges entiers, le modèle sur les âges exacts a tendance à légèrement sous-estimer la survie des contrats, de 6,9% (contre une sur-estimation de 7,16% avec le modèle sur les âges entiers), par rapport à l'estimation de Kaplan-Meier sur les contrats souscrits à partir de 2006. Et il

apparaît que les segments sur lesquels le modèle sous-estime le plus, sont le LA2P, le LA3P et le LA4+.

Les graphiques de comparaison pour chaque segment sont présentés en annexe 3.

Il est possible de relever les valeurs moyennes par segment des fonctions de survie, à chaque âge entier, prédites par le modèle de Cox construit sur les âges exacts, ou celui construit sur les âges entiers, ou estimées par Kaplan-Meier sur la nouvelle base de données. Ainsi il est possible d'estimer une espérance de vie (définie pour une représentation discrétisée de la durée de survie) par segment selon les différentes prédictions. En effet en partant de la définition suivante de l'espérance de vie à la souscription (simplifiée, sans prise en compte de la répartition des résiliations au sein d'une année) :

$$E(0) = \sum_{k=0}^{\infty} S(k) \times q_k \times k.$$

Il sera supposé que le dernier âge possible que peut atteindre un contrat est 15 ans et qu'au delà de cet âge la fonction de survie est nulle (ce qui ne peut être vrai en pratique, mais que nous allons supposer vrai pour pouvoir comparer les différentes prédictions). Alors l'espérance de vie peut être estimée :

$$\hat{E}(0) = \sum_{k=0}^{\infty} \hat{S}(k) \times \hat{q}_k \times k = \sum_{k=0}^{\infty} \hat{S}(k) \times \left(1 - \frac{\hat{S}(k+1)}{\hat{S}(k)}\right) \times k = \sum_{k=1}^{15} \hat{S}(k)$$

Cette formule a donc permis d'estimer une espérance de vie à la souscription avec les prédictions moyennes de chaque modèle de Cox, et des estimations de Kaplan-Meier, par segment.

Les résultats obtenus sont précisément présentés dans un tableau en annexe 4.

Il est visible que les trois estimations de l'espérance de vie moyenne par segment sont cohérentes entre elles, en effet si les segments étaient classés selon leur valeur d'espérance moyenne prédite avec les trois modèles, les classements seraient très similaires. Sur l'ensemble des segments, le modèle de Cox sur les âges exacts a tendance à sous-estimer l'espérance de vie des contrats de 7,9% par rapport aux estimations par Kaplan-Meier. Le modèle de Cox sur les âges entiers a tendance à surestimer de 8,7%. L'erreur du modèle sur les âges exacts est donc légèrement plus faible, et de plus il est préférable, par prudence, d'avoir un modèle qui sous-estime légèrement la survie des contrats, plutôt qu'un qui la surestime.

Il est également possible d'observer les valeurs de concordance des prédictions du modèle lors de son application, et de les comparer à celles de l'application du modèle sur les âges entiers :

	Base d'entraînement	Base test
Concordance modèle sur les âges entiers	0,6794	0,6792
Concordance modèle sur les âges exacts	0,6827	0,6825

TABLE 3.3 – Valeurs de la concordance des prédictions des modèles construits sur les âges entiers ou sur les âges exacts

Il apparaît que les prédictions du modèle de Cox construit sur les durées de vie exactes des contrats sont de meilleure qualité, car leur ordre correspond plus précisément à l'ordre des durées de vie réelles.

De plus il peut être envisagé d'observer la qualité de la prédiction de l'espérance de vie à la souscription par rapport à l'âge de résiliation observé pour des contrats non-censurés. L'espérance de vie à la souscription a donc été prédite pour chaque contrat selon la définition mathématique de l'espérance pour une représentation continue de la durée de survie T : $\mathbf{E}(T) = \int_0^{\infty} S(t)dt$.

Il a été observé qu'en faisant la moyenne des espérance prédites sur les contrats non-censurés de chaque segment MRH, les valeurs obtenues étaient visiblement plus élevées que la moyenne des durées de vie réellement observées. Une autre manière de prédire l'espérance de vie des contrats a donc été envisagée. Il est commun d'utiliser la médiane de la courbe de survie prédite d'un contrat pour estimer son espérance de vie. Seulement sur nos données les courbes de survie ne peuvent être estimées que jusqu'à 15 ans, certains contrats (plus généralement ceux des propriétaires) peuvent avoir une courbe de survie estimée qui reste supérieure à 0,5 jusqu'à 15 ans, pour ces contrats la médiane est donc infinie. La nouvelle définition

de l'estimation de l'espérance de vie d'un contrat a donc été choisie comme correspondant au minimum entre la médiane de la courbe de survie et l'espérance de la durée de survie du contrat. Les espérances de vie moyennes par segment obtenues ainsi étaient beaucoup plus proches des durées de vie moyennes observées. Il est possible d'observer l'erreur de prédiction moyenne en années, par segment.

L'ensemble des résultats obtenus sont présentés dans un tableau en annexe 5.

Sur l'ensemble des contrats non-censurés l'erreur de prédiction de l'espérance de vie est d'environ 1,1 ans. En particulier, sur les contrats de locataires elle est en moyenne de 0,2 an, sur ceux des propriétaires elle est de 3 ans, et sur ceux des PNOs de 3,3 ans. Cela semble dû au fait que les contrats de locataires ont une espérance de vie plus faible, et ainsi la majorité (74,7%) des résiliations des contrats des locataires ont pu être observées, ce qui n'est pas le cas pour les contrats des propriétaires (47,2%), ni pour ceux des PNOs (45,8%).

Le modèle de Cox utilisé pour ces prédictions est donc celui construit sur les âges de fin exacts, mais il est possible de calculer les erreurs de prédictions moyennes avec le modèle construit sur les âges entiers. Alors il apparaît que l'erreur de prédiction moyenne par segment est toujours plus importante, avec une erreur de prédiction sur l'ensemble des contrats non-censurés de 1,3 ans.

Donc, au vu des résultats obtenus lors des comparaisons effectuées entre le modèle de Cox construit sur les âges exacts et le modèle construit sur les âges entiers, il est possible d'affirmer que le modèle sur les âges exacts, en plus d'apporter une visibilité sur la répartition des résiliations au sein des années, semble également plus adéquat au niveau de ses prédictions.

Une approche pour essayer d'améliorer les prédictions construites avec un modèle de Cox sur les âges exacts, c'est à dire de les rapprocher, en moyenne, des estimations de Kaplan-Meier des courbes de survie par segment, a été mise en oeuvre. Les données ont été séparées avec dans une base de données uniquement les contrats de locataires, dans une autre uniquement les contrats de propriétaires, et dans une autres ceux des propriétaires non-occupant (PNO). Ensuite sur chaque base un modèle de Cox a été construit pour prédire la survie des locataires, la survie des propriétaires et la survie des PNOs. En utilisant chacun des trois modèles la courbe de survie prédite moyenne pour chacun des segments MRH a été représentée, et comparée avec la courbe de survie estimée par Kaplan-Meier, ainsi que la courbe de survie moyenne prédite par le premier modèle de Cox construit sur l'ensemble des données, sur les âges exacts. Finalement, cette méthode n'améliorait pas les prédictions, ni les valeurs de la concordance. Cette méthode n'a donc pas été retenue.

	Base d'entraînement	Base test
Concordance modèle sans séparation	0,6827	0,6825
Concordance modèle avec séparation	0,6820	0,6818

TABLE 3.4 – Valeurs de la concordance des prédictions des modèles construits sur les âges exacts avec ou sans séparation locataires/propriétaires/PNOs

Le modèle de Cox construit sur l'ensemble des contrats, à partir des durées de vie des contrats exactes, sera donc retenu comme étant le modèle de Cox classique le plus performant.

Suivant ce modèle, les variables ayant le plus d'impact sur le risque de résiliation des contrats sont notamment :

- la variable indiquant si le contrat est rattaché à une personne morale ;
- celle indiquant le produit d'entrée au sein d'AXA de l'assuré ;
- celle indiquant la qualité de l'assuré (si l'assuré est propriétaire le risque de résiliation du contrat sera plus faible, toutes choses égales par ailleurs) ;

- celle indiquant le type de résidence du bien assuré (si c'est une résidence non-occupée et non-louée le risque de résiliation du contrat sera plus élevé, toutes choses égales par ailleurs) ;
- celle indiquant si le contrat est un contrat étudiant (si c'est un contrat étudiant son risque de résiliation sera plus élevé, toutes choses égales par ailleurs).

L'impact des différentes caractéristiques possibles sur le risque de résiliation des contrats est présenté en annexe 6.

Pour apprécier la performance du modèle il est possible de vérifier qu'il n'y a pas de sur-apprentissage, pour cela la méthode de validation croisée peut être utilisée. Cette méthode consiste à diviser la base de données en k échantillons. Le modèle va ensuite être reconstruit k fois avec les mêmes paramètres que le modèle sélectionné. A chaque construction, $k - 1$ des échantillons vont être utilisés comme ensemble d'apprentissage, et l'échantillon restant sera utilisé comme ensemble de validation. A chaque fois l'échantillon de validation changera, et ainsi chaque échantillon aura été utilisé comme ensemble de validation. Pour chaque modèle construit, la concordance obtenue par son application sur son ensemble de validation sera mesurée.

Il est possible de représenter schématiquement la mise en place de la méthode de validation croisée. En choisissant de diviser la base en $k = 10$ échantillons, il est possible de se représenter quels échantillons sont utilisés pour l'apprentissage et pour la validation à chaque $i - ième$ reconstruction du modèle :

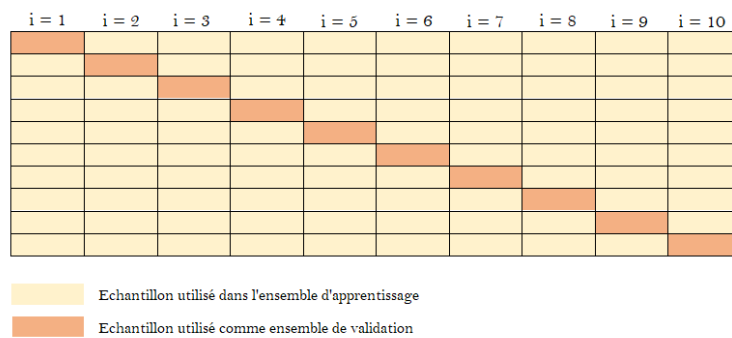


FIGURE 3.8 – Mise en place de la méthode de validation croisée

Les valeurs de concordance mesurées sur l'application de chaque modèle à la base de validation associée peuvent être représentées graphiquement :

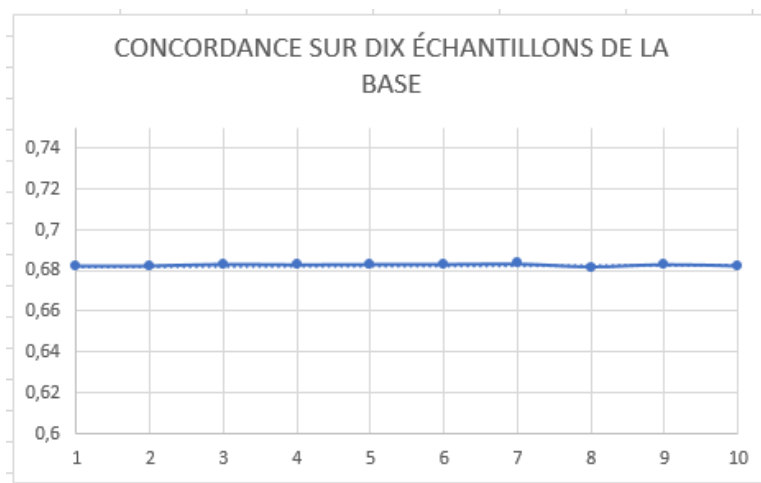


FIGURE 3.9 – Valeurs de la concordance des prédictions des modèles de Cox classique construits par validation croisée, sur dix sous-ensembles de validation de la base de données

L'indice de concordance est stable sur les dix sous-ensembles de validation ce qui est satisfaisant. Les résultats de cette validation croisée indiquent qu'il n'y a pas de sur-apprentissage du modèle.

3.2.3.7 Prolongation des courbes de survies estimées par le modèle

Le modèle de Cox implémenté permet de prédire pour chaque contrat sa courbe de survie, selon ses informations à la souscription, jusqu'à 15 ans. Mais il est visible qu'à 15 ans, en moyenne, la valeur de la courbe de survie prédite reste élevée, en particulier pour les contrats de propriétaires. Conserver uniquement le modèle de Cox sur 15 ans n'est donc pas acceptable. Il est donc nécessaire de prolonger les courbes de survie prédites au delà de 15 ans, avec une approximation utilisant les taux de survie qui avaient été estimés entre 0 et 50 ans sur la base Image lors des études préalables à la modélisation semi-paramétrique. Les taux de survie avaient été estimés par segment MRH, en supposant que le segment des contrats ne changeait pas au cours de leur vie. Cette hypothèse va être conservée pour que les taux de survie entre 15 et 50 ans puissent être utilisés pour prolonger les courbes de survie prédites par le modèle de Cox pour chaque contrat, selon son segment MRH à la souscription.

Les étapes de la construction de la prédiction d'une courbe de survie prolongée jusqu'à 50 ans, pour chaque contrat, sont les suivantes :

- Application du modèle de Cox pour construire une prédiction de la courbe de survie entre 0 et 15 ans. Cette courbe est nommée $\hat{S}_{début}$.

- Construction d'une courbe de survie estimée, appelée \hat{S}_{fin} . Pour construire cette courbe, les valeurs de la courbe de Kaplan-Meier construite sur la nouvelle base entre 0 et 15 ans, associée au segment MRH du contrat, appelée $\hat{S}_{KMbaseCox}$, sont récupérées. Et à partir de 16 ans, chaque nouvelle valeur de la courbe de survie est calculée à partir de la valeur précédente de la courbe de survie et du taux de survie $\hat{p}_x^{baseImage}$ précédent qui avait été obtenu sur la base Image jusqu'à 50 ans, associé au segment MRH du contrat. La

définition de la courbe est donc
$$\hat{S}_{fin}(x) = \begin{cases} \hat{S}_{KMbaseCox}(x) , & \text{si } x \leq 15 \\ \hat{S}_{fin}(x-1) \times \hat{p}_{x-1}^{baseImage} , & \text{si } x > 15 \end{cases}$$

- Modification de l'ordre de grandeur des valeurs de la courbe obtenue, de telle sorte qu'elle puisse prolonger la courbe prédite par le modèle de Cox entre 0 et 15 ans pour le contrat. La courbe $\hat{S}_{prolongement}$ qui sera utilisée entre 15 et 50 ans pour le prolongement se définit comme $\hat{S}_{prolongement}(x) = \hat{S}_{fin}(x) \times \frac{\hat{S}_{début}(15)}{\hat{S}_{fin}(15)}$.

Une fois la courbe de survie prédite prolongée pour chaque contrat selon son segment MRH, il est possible d'observer la courbe de survie moyenne prédite par segment sur 50 ans et de la comparer avec la courbe de survie estimée par Kaplan-Meier sur 50 ans sur la base Image. Des écarts sont visibles, ces écarts semblent dûs au fait que, pour la construction de la courbe de survie par Kaplan-Meier par segment, il ne s'agissait pas nécessairement du segment MRH à la souscription.

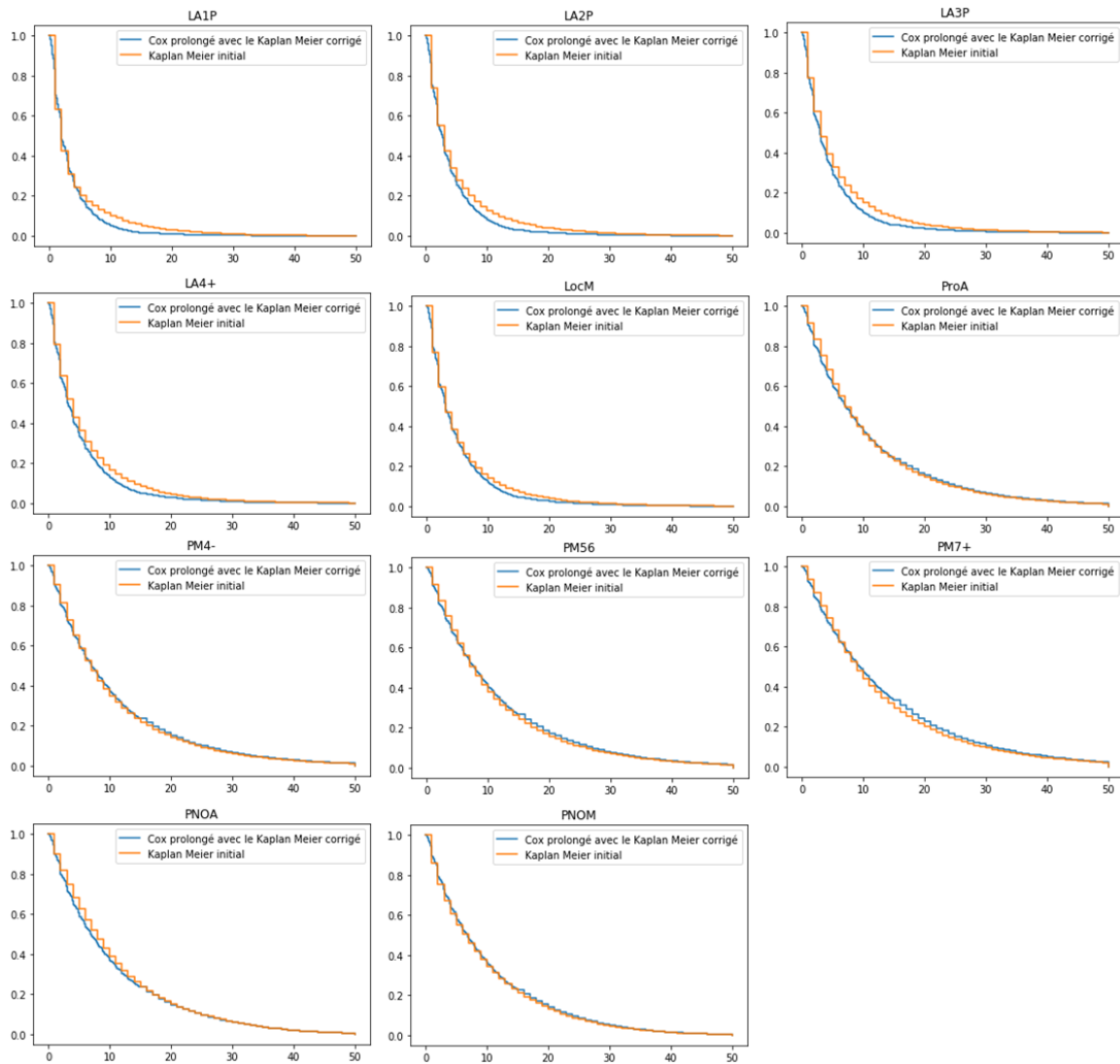


FIGURE 3.10 – Comparaison, pour chaque segment MRH, entre la courbe de survie moyenne prédite par Cox classique et prolongée, et la courbe de survie estimée par Kaplan-Meier sur la base Image

En comparant ces courbes de survie moyennes aux courbes de survie construites \hat{S}_{fin} les écarts sont réduits. Ces courbes reprennent sur les 15 premières années la dynamique de résiliations des contrats selon leur segment, en disposant bien du segment MRH à la souscription. Au delà de 16 ans c'est la dynamique de résiliation des contrats selon leur segment, mais en ne disposant pas nécessairement du segment à la souscription, qui a été reprise. Le fait qu'il ne s'agisse pas nécessairement du segment à la souscription n'a un impact qu'à partir de 16 ans, contrairement à la courbe de survie initiale construite par Kaplan-Meier sur la base Image, pour laquelle cela avait un impact sur tous les âges. C'est pourquoi les écarts sont réduits avec les courbes de survie construites \hat{S}_{fin} .

Les graphiques représentant la comparaison pour chaque segment sont présentés en annexe 7.

3.2.4 Modélisation avec utilisation d'un algorithme de *Gradient Boosting*

Un premier modèle de Cox a été implémenté de manière classique. Désormais il va être étudié dans la suite de l'étude, la possibilité d'implémenter un modèle pour lequel la maximisation de la vraisemblance, pour estimer les paramètres β du modèle, se ferait par un algorithme de *Gradient Boosting*.

3.2.4.1 Présentation de l'algorithme de *Gradient Boosting*

Le *Gradient Boosting* est un algorithme de *Machine learning*.

Le *Machine learning* ou apprentissage automatique est une technologie d'intelligence artificielle dont le principe est d'apprendre aux machines à prédire à partir de données, sans les avoir explicitement programmées pour cela. La mise en place d'un algorithme de *Machine learning* nécessite donc une quantité importante de données, mais également des données pertinentes pour la prédiction voulue. Généralement les méthodes de *Machine learning* sont particulièrement efficaces.

Au sein du *Machine learning* existent deux types d'apprentissage. Ces deux notions sont notamment présentés dans le mémoire de KOYE G. K. *Comparaison des méthodes classiques et alternatives avec le machine learning pour la construction d'une table de mortalité d'expérience best estimate* [8]. Le premier type d'apprentissage est l'apprentissage non-supervisé. L'apprentissage non-supervisé consiste à prédire sur des données non-étiquetées. La machine va devoir déterminer les structures sous-jacentes des données, car celles-ci ne seront pas connues à l'avance. Le second est l'apprentissage supervisé. Dans le cas de l'apprentissage supervisé, des informations sur les résultats attendus sont présentées à la machine pour qu'elle apprenne de ceux-ci. En particulier, la structure des données peut être préalablement explicitée, et la machine va devoir apprendre à classer de nouvelles données au sein de cette structure. Pour la construction d'un modèle de Cox, c'est un apprentissage supervisé qui est utilisé, car les dates de mort (ou résiliation) connues des individus sont communiquées à la machine pour qu'elle apprenne à prédire la mort (ou résiliation) d'autres individus.

Le *Gradient Boosting* est un algorithme d'apprentissage supervisé qui peut être utilisé dans la construction d'un modèle de Cox. Le fonctionnement de cet algorithme est notamment décrit dans le cours de LASGOUTTES J.M.. *Introduction au boosting* [9], il sera résumé dans les paragraphes suivants.

Le *Gradient Boosting* repose sur l'utilisation d'arbres de décision. Un arbre de décision est une méthode d'apprentissage non-supervisé, utile pour effectuer une régression ou une classification. Elle consiste à classer les individus d'une base d'apprentissage en différents groupes les plus homogènes possible, au niveau de leur valeur de la variable cible, en fonction des valeurs des différentes variables explicatives disponibles. Lorsque les règles de décisions vis-à-vis des variables explicatives sont déterminées, la prédiction associée à chaque groupe d'individus obtenu est définie. Pour une régression, il s'agira de la moyenne des valeurs observées, de la variables cible, des individus de la base d'apprentissage associés au groupe. Pour une classification, il s'agira de la modalité de la variable cible la plus représentée, au sein des individus de la base d'apprentissage associés au groupe.

De manière classique, le *Gradient Boosting*, pour effectuer une régression, fonctionne ainsi :

- Le modèle veut prédire la valeur de la variable cible de chaque individu d'une base de données.
- Un premier arbre de décision est construit sur les données d'apprentissage disponibles, pour construire une première prédiction.
- L'arbre de décision obtenu est appliqué aux données d'apprentissage et l'écart entre la valeur réelle et la valeur prédite de la variable cible est déterminée pour chaque individu. Cet écart correspond à un résidu représentatif de l'insuffisance du modèle.
- Un nouvel arbre de décision est construit mais cette fois-ci sur les résidus obtenus précédemment.
- Ce nouvel arbre est appliqué aux données d'apprentissage et une prédiction de l'erreur de prédiction du premier arbre est obtenue pour chaque individu. Cette erreur prédite va être sommée à la valeur de la variable cible prédite par le premier arbre, proportionnellement à un certain poids, prenant une valeur entre 0 et 1, appelé le *learning rate*, correspondant à la vitesse d'apprentissage du *Gradient Boosting*. Cela permet d'obtenir pour chaque individu une nouvelle prédiction de la valeur de la variable cible, normalement plus proche de la valeur réelle.

- Le nouveau résidu obtenu pour chaque individu peut être calculé et un nouvel arbre de décision peut à nouveau être construit pour prédire ces résidus. La prédiction obtenue par ce nouvel arbre peut être agrégée à la prédiction obtenue précédemment, toujours proportionnellement au *learning rate*. Et ces étapes peuvent être appliquées à nouveau successivement un certain nombre de fois prédéterminé, correspondant au nombre d'arbres choisi.

Le *Gradient Boosting* est donc un algorithme utilisant une succession d'arbres de décision, apprenant des erreurs de leurs prédécesseurs.

Le *Gradient Boosting* peut s'adapter à différents contextes, pour cela l'évaluation des erreurs des arbres de décision précédents, à chaque itération, peut être faite de manières différentes. En effet de manière classique, comme présenté précédemment, les valeurs des résidus peuvent être utilisées, mais les valeurs d'autres indicateurs peuvent aussi être utilisées. Pour construire un modèle de Cox par *Gradient Boosting*, l'objectif étant de prédire une estimation de β , c'est la valeur de la log-vraisemblance partielle obtenue qui est étudiée à chaque itération : $\ln(\mathcal{L}(\beta)) = \sum_{i=1}^y d_i [Z_i \beta - \ln(\sum_{j=1}^y R_j(t) \exp(Z_j \beta))]$. En particulier l'écart entre $\frac{\partial \ln(\mathcal{L}(\beta))}{\partial \beta}$ et 0 sera étudié dans le but de le minimiser au fil des itérations.

Cette utilisation du *Gradient Boosting* est une manière efficace de déterminer les paramètres d'un modèle de Cox.

3.2.4.2 Implémentation informatique du modèle

Pour construire un modèle de Cox avec utilisation du *Gradient Boosting* il est possible d'utiliser, dans le langage *Python*, le package *XgBoost* (version 1.4.2) et en particulier la fonction *XGBRegressor* avec le paramètre " objective='survival : cox' " qui permet de construire un modèle de Cox. Le modèle obtenu prédit la valeur de risque instantané relatif de chaque contrat i , $\exp(Z_i \hat{\beta})$. Il est ensuite possible de déduire le risque de base, $\hat{H}_0(t) = \sum_{t_i \leq t} \frac{\sum_{j=1}^y R_j(t_i) - \sum_{j=1}^y R_j(t_{i+1})}{\sum_{j=1}^y R_j(t_i) \exp(Z_j \hat{\beta})}$. Et finalement la prédiction de la fonction de survie pour chaque contrat i , $\hat{S}(t|Z_i) = \exp(-\hat{H}_0(t) \times \exp(Z_i \hat{\beta}))$.

Pour déterminer les paramètres les plus adaptés à la construction d'un *Gradient Boosting* sur les données disponibles, l'indicateur de concordance a été utilisé. Un grand nombre de combinaisons de valeurs de paramètres différentes ont été utilisées pour la construction de modèles, et, à chaque construction, le modèle obtenu était appliqué à la base test, et la valeur de concordance des prédictions était étudiée. Les valeurs de paramètres du modèle dont les prédictions présentaient la concordance la plus élevée étaient alors retenues comme valeurs de paramètres optimales.

3.2.4.3 Construction d'un modèle sur les âges exacts

Un modèle de Cox avec utilisation du *Gradient boosting* a été construit directement sur les données présentant les durées de vie exactes des contrats. Il a été vu précédemment, lors de la construction d'un modèle classique, que la construction à partir des durées de vies en années entières était moins pertinente, donc cette méthode n'a pas été envisagée dans la construction du modèle de *Gradient Boosting*.

Lors de la détermination des paramètres du modèle, plusieurs combinaisons de valeurs de paramètres ont été étudiées, et il a été retenu que le modèle de Cox *Gradient Boosting* devait être construit avec, notamment, 100 arbres, une profondeur maximale des arbres de 21 et un taux d'apprentissage de 0,02.

Après implémentation du modèle avec ces paramètres optimaux, il a été possible d'observer la courbe de survie prédite moyenne sur tous les contrats associés à chaque segment MRH, et aussi sur les contrats associés aux segments LA1P, LA2P et LA3P selon la présence de la formule étudiante :

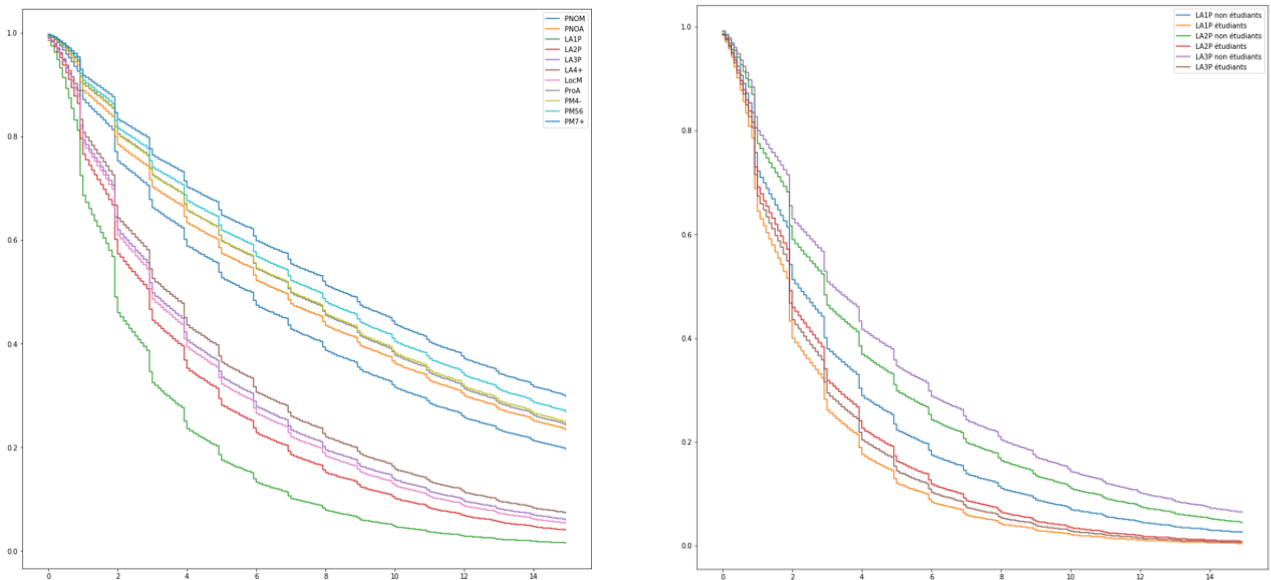


FIGURE 3.11 – Courbes de survie prédites, par le modèle de Cox *Gradient Boosting* sur les durées de vie exactes, moyennes par segment MRH (à gauche) et selon la présence de la formule étudiante et par segment MRH (à droite)

Comme lors de la construction classique d'un modèle de Cox, il apparaît que les contrats de propriétaires ont une meilleure survie que les contrats des locataires. Les modélisations sont donc cohérentes entre elles.

Aussi, à nouveau, il apparaît que les contrats avec la formule étudiante ont une survie plus faible que celle des contrats sans la formule. Par rapport aux modèle classique, le modèle *Gradient Boosting* prédit une survie plus faible pour les contrats de locataires d'appartement une pièce et non-étudiants.

Il est possible de se rendre compte de l'écart entre les estimations des valeurs de la courbe de survie des contrats du modèle de Cox *Gradient Boosting* et du modèle de Kaplan Meier. Une estimation de l'espérance de vie à la souscription par segment MRH, selon les prédictions moyennes de la fonction de survie, du modèle de Cox *Gradient Boosting* et des estimations de Kaplan-Meier, a donc été calculé dans ce but.

Les résultats obtenus sont présentés dans un tableau en annexe 8.

Sur l'ensemble des segments, le modèle de Cox *Gradient Boosting* sur les âges exacts a tendance à sous-estimer l'espérance de vie des contrats de 7,7% par rapport aux estimations par Kaplan-Meier. Comme dit précédemment, il est plus prudent d'avoir un modèle qui sous estime légèrement l'espérance de vie des contrats que l'inverse donc ces résultats sont satisfaisants.

Les valeurs de concordance des prédictions du modèle lors de son application à la base d'entraînement et à la base test ont été observées :

	Base d'entraînement	Base test
Concordance	0,7029	0,7114

TABLE 3.5 – Valeurs de la concordance de l'application du modèle *Gradient Boosting*

La valeur de l'indicateur de concordance de ce modèle est d'environ 0,70 ce qui est satisfaisant.

Enfin, la qualité de la prédiction de l'espérance de vie à la souscription par rapport à l'âge de résiliation observé, des contrats non-censurés, par segment, a également été observée.

L'ensemble des résultats obtenus sont présentés dans un tableau en annexe 9.

Sur l'ensemble des contrats non-censurés, l'erreur de prédiction de l'espérance de vie est d'environ 1 an. En particulier, sur les contrats de locataires elle est en moyenne de 0,3 an, sur ceux des propriétaires elle est de 2,5 ans, et sur ceux des PNOs de 2,5 ans.

Afin de comprendre la manière dont ce modèle construit ses prédictions, il est possible d'observer la représentation suivante de l'impact des variables sur le risque de résiliation, nommée "SHAP values" :

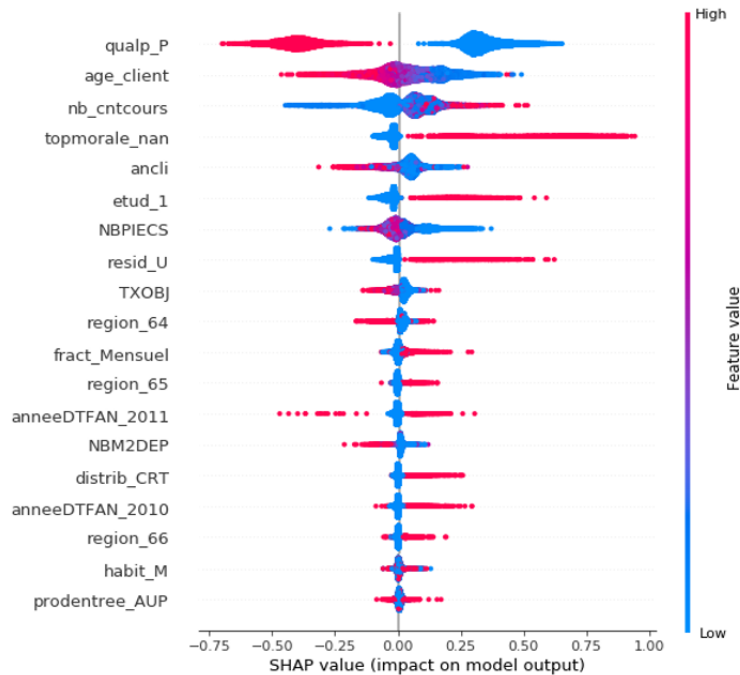


FIGURE 3.12 – Impact de la valeur des variables sur le risque instantané relatif de résiliation des contrats d'après le modèle de Cox *Gradient Boosting*

En observant cette représentation visuelle il est possible de dire, par exemple, que d'après ce modèle de Cox, si un contrat est associé à un propriétaire son risque de résiliation sera plus faible, ou si un contrat est associé à un client plus jeune son risque de résiliation sera plus élevé, toutes choses égales par ailleurs.

En prenant deux contrats dans les données, il est possible de représenter visuellement la manière selon laquelle le risque instantané relatif de résiliation du contrat a été calculé. Pour le premier contrat, il apparaît notamment que la région associée au contrat, et l'âge du client, qui est relativement jeune, augmentent le risque de résiliation, tandis que le fait que le client soit propriétaire, ou que son logement possède six pièces, diminuent le risque de résiliation :

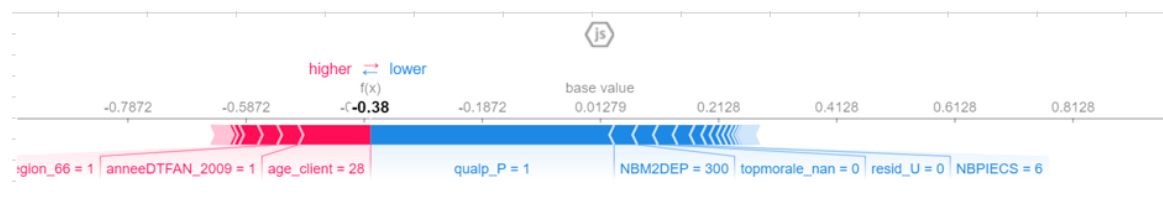


FIGURE 3.13 – Impact de la valeur des variables sur le risque instantané relatif de résiliation d'un premier contrat d'après le modèle de Cox *Gradient Boosting*

Pour le second contrat, il apparaît que l'ancienneté du client, qui est assez élevée, diminue le risque de résiliation, mais que, notamment, le fait que le client ne soit pas propriétaire, qu'il soit relativement jeune,

ou que son logement ne possède que deux pièces, augmentent le risque de résiliation :



FIGURE 3.14 – Impact de la valeur des variables sur le risque instantané relatif de résiliation d'un second contrat d'après le modèle de Cox *Gradient Boosting*

Pour apprécier la qualité de la construction du modèle de Cox *Gradient Boosting*, la méthode de validation croisée a à nouveau été utilisée.

Les valeurs de concordance mesurées sur l'application de chaque modèle de *Gradient Boosting* à la base de validation associée ont pu être représentées :

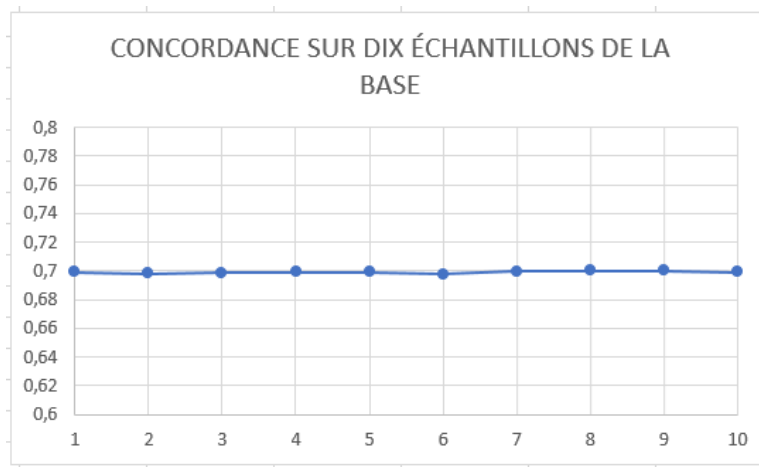


FIGURE 3.15 – Valeurs de la concordance des prédictions des modèles de Cox *Gradient Boosting* construits par validation croisée, sur dix sous-ensembles de validation de la base de données

L'indice de concordance est stable sur les dix sous-ensembles de validation. Les résultats de cette validation croisée indiquent donc qu'il n'y a pas de sur-apprentissage du modèle.

3.2.4.4 Prolongation des courbes de survies estimées par le modèle

De la même manière que lors de la construction d'un modèle de Cox classique, le modèle obtenu par *Gradient Boosting* permet de prédire, pour chaque contrat, sa courbe de survie jusqu'à ses 15 ans. Cependant à 15 ans les valeurs des courbes de survie restent relativement élevées. Il est donc important de les prolonger. La même méthode que celle qui a été mise en oeuvre pour le prolongement des courbes obtenues avec le modèle de Cox classique jusqu'à 50 ans, a été utilisée. Les taux de survie qui avaient été obtenus par Kaplan-Meier, par segment, sur 50 ans ont permis une prolongation des courbes de survie des contrats selon leur segment MRH.

Il est possible de représenter la courbe de survie moyenne prédite par segment sur 50 ans, obtenue grâce à la prolongation, et de la comparer à la courbe de survie qui avait été estimée par Kaplan-Meier sur 50 ans sur la base Image MRH :

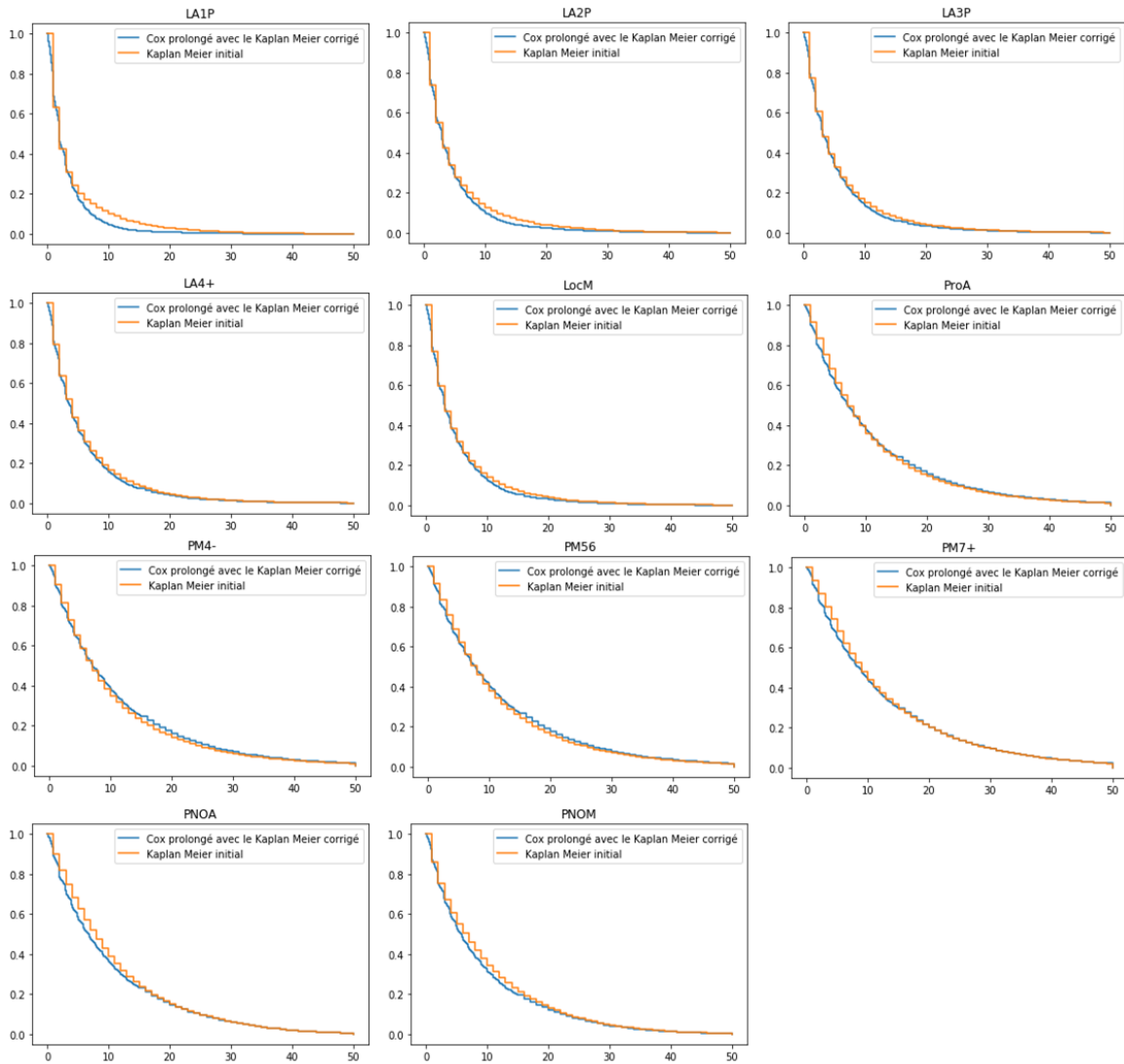


FIGURE 3.16 – Comparaison, pour chaque segment MRH, entre la courbe de survie moyenne prédite par *Cox Gradient Boosting* et prolongée, et la courbe de survie estimée par Kaplan-Meier sur la base Image

Comme lors de la construction du modèle de Cox classique des écarts sont visibles entre les courbes prédites par *Cox Gradient Boosting* et celles prédites par Kaplan-Meier.

Il est aussi possible de comparer ces courbes de survie moyennes aux courbes de survie construites \hat{S}_{fin} et alors les écarts sont également réduits.

Les graphiques représentant la comparaison pour chaque segment sont présentés en annexe 10.

3.2.5 Choix du modèle

Pour prédire la survie des contrats d'assurance MRH, face au risque de résiliation, deux modèles ont été développés : un premier modèle de Cox classique, et un second modèle de Cox utilisant l'algorithme de *Gradient Boosting*. Le modèle le plus satisfaisant entre les deux va alors être retenu comme modèle à utiliser pour la construction de la valeur contrat à la souscription.

En reprenant l'étude des valeurs des différents indicateurs calculées sur les résultats obtenus avec les deux modèles, le modèle de Cox par *Gradient Boosting* est plus performant à plusieurs niveaux, c'est donc lui qui a été retenu pour la construction de la valeur contrat.

En regardant, pour chaque segment et pour chaque modèle de Cox, l'espérance des contrats estimée avec les valeurs moyennes des fonctions de survie, à chaque âge entier, prédites par le modèle sur les 15 premières années, et en la comparant avec l'espérance estimée avec le modèle de Kaplan-Meier sur 15 ans, le modèle de Cox par *Gradient Boosting* est meilleur. Il présente des erreurs plus faibles et surtout plus stables entre les différents segments MRH :

Segment	Erreur sur les espérances Modèle de Cox classique par rapport à Kaplan-Meier	Erreur sur les espérances Modèle de Cox Gradient Boosting par rapport à Kaplan-Meier
Locataires	-12,8%	-7,7%
Propriétaires et PNOs	-5,1%	-7,8%
Global	-7,9%	-7,7%

TABLE 3.6 – Erreurs sur les estimations de l'espérance de vie à la souscription selon les segments MRH, selon les prédictions moyennes de la fonction de survie, des modèles de cox, et des estimations de Kaplan-Meier

En observant les valeurs de l'indicateur de concordance des prédictions des modèles lors de leur application à la base d'entraînement et à la base test, le modèle par *Gradient Boosting* présente une meilleure concordance :

	Base d'entraînement	Base test
Concordance modèle classique	0,6827	0,6825
Concordance modèle <i>Gradient Boosting</i>	0,7029	0,7114

TABLE 3.7 – Valeur de la concordance des prédictions du modèle de Cox classique et du modèle de Cox *Gradient Boosting*

Il apparaît, en observant l'erreur de prédiction de l'espérance de vie à la souscription par rapport à l'âge de résiliation observé, des contrats non-censurés, par segment, que le modèle par *Gradient Boosting* présente des erreurs plus faibles, en particulier au niveau des propriétaires et PNOs :

Segment	Erreur moyenne de prédiction (en années) Modèle classique	Erreur moyenne de prédiction (en années) Modèle Gradient Boosting
Locataires	0,2	0,3
Propriétaires et PNOs	3,1	2,5
Global	1,1	1

TABLE 3.8 – Erreur de prédiction moyenne, par segment, avec le modèle de Cox classique et le modèle de Cox *Gradient Boosting*, de l'espérance de vie à la souscription des contrats non censurés, par rapport aux durées de vie réellement observées

3.3 Synthèse du chapitre

L'objectif de ce chapitre était de construire un modèle de prédiction de la survie des contrats d'assurance MRH face au risque de résiliation, prenant en compte les caractéristiques des contrats à la souscription. Cela consistait donc à construire un modèle de survie semi-paramétrique.

Pour construire un modèle de survie semi-paramétrique, il était nécessaire de bien disposer des informations des contrats à la souscription. Une nouvelle base de données a donc été construite, regroupant les informations contrats à la souscription et les informations sur la survie des contrats, des contrats d'assurance MRH souscrits entre 2006 et 2020. Avant 2006, les données n'étaient pas disponibles.

Des modèles de Cox ont été construits sur cette base de données. Un premier modèle de Cox a été construit de manière classique, et un second a été construit en utilisant l'algorithme du *Gradient Boosting*. Ces modèles permettaient d'avoir, pour chaque contrat, une prédiction de sa survie sur les 15 premières années de sa vie. Pour comparer la qualité de prédiction des deux modèles, les valeurs de l'indicateur de concordance sur leur application ont été étudiées. Aussi, l'écart entre leurs prédictions et les prédictions faites par un modèle de Kaplan-Meier sur les mêmes données, et la qualité de la prédiction de l'espérance de vie sur les contrats dont la résiliation a été observée, ont été étudiés. Finalement, le modèle de Cox construit par *Gradient Boosting* s'est avéré être le meilleur modèle de survie obtenu pour prédire la survie des contrats. Ce modèle a donc été retenu pour la construction de la valeur contrat.

Avoir un modèle de survie sur 15 ans ne suffisait pas. Donc, pour chaque contrat, la courbe de survie obtenue sur 15 ans, a été prolongée jusqu'à 50 ans, en utilisant les valeurs de probabilités de résiliation à chaque âge obtenues pour le segment du contrat, lors de la modélisation de Kaplan-Meier effectuée dans le chapitre précédent, sur la base Image.

— Chapitre 4 —

Modélisation de la probabilité de remplacement

4.1 Étude statistique descriptive préalable

Cette étude du risque de remplacement commence par une analyse descriptive afin d'avoir une meilleure vision des caractéristiques des remplacements qui peuvent avoir lieu au cours de la vie des contrats. Cette analyse a été faite sur la base Image MRH.

Au cours de la vie d'un contrat des remplacements peuvent avoir lieu, ils correspondent à une modification du contrat. Il peut notamment y avoir un remplacement pour augmenter ou diminuer le montant des capitaux assurés, pour souscrire de nouvelles options d'assurance ou au contraire en enlever, ou lorsque l'assuré déménage et que le logement à assurer change, ou lorsque la qualité de l'assuré change.

Il a été décidé de considérer en particulier le remplacement pour changement de qualité, car il apparaît que la rentabilité d'un contrat est particulièrement différente entre le cas où l'assuré est locataire et le cas où il est propriétaire. Cela peut se voir par exemple au niveau de la rentabilité des contrats, observée à leur deuxième année de vie. La marge moyenne de tous les segments de locataires est supérieure à 10%, alors que celle des segments de propriétaires et PNOs est inférieure à 10% :

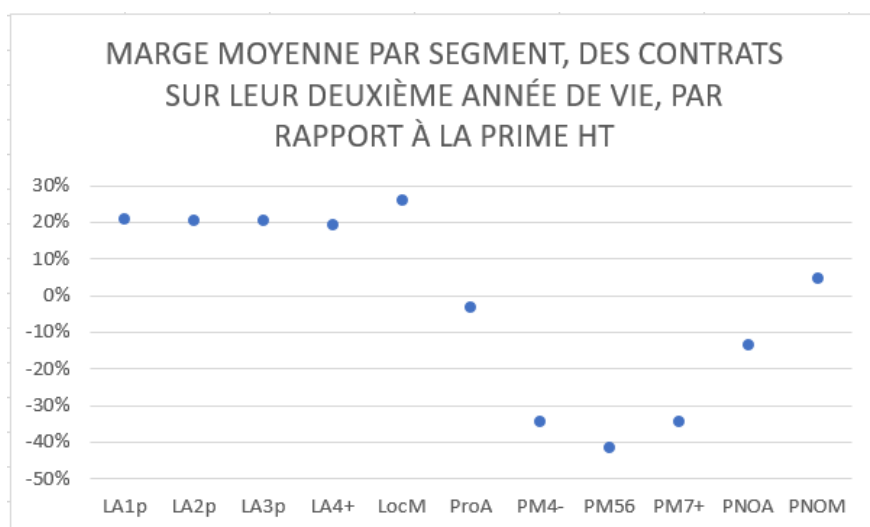


FIGURE 4.1 – Marge moyenne des contrats, par segment, sur leur deuxième année de vie

Pour la construction de la valeur contrat à la souscription, il serait donc intéressant de considérer la possibilité que ce type de remplacement ait lieu durant la vie de chaque contrat.

L'impact d'un remplacement de ce type sur la prime commerciale hors taxes peut être observé pour le passage de locataire à propriétaire, et le passage de propriétaire à locataire :

Type de remplacement	Évolution moyenne de la prime HT en €	Évolution moyenne de la prime HT en %
Passage de propriétaire à locataire	-76,87 €	-13%
Passage de locataire à propriétaire	+44,96 €	+32%

TABLE 4.1 – Évolution de la prime hors taxe lors d'un remplacement pour changement de qualité, selon le type de remplacement

Il apparaît que la prime commerciale a tendance à augmenter lorsqu'un assuré locataire devient propriétaire, et a tendance à diminuer lorsqu'un propriétaire devient locataire. Cela est logique car le risque assuré est plus important pour un propriétaire que pour un locataire.

Les remplacements pour changement de qualité représentent 4,58% des remplacements. La majorité, 67%, des remplacements pour changement de qualité sont des remplacements pour passer de locataire à propriétaire. Or il était visible que la rentabilité des contrats de propriétaires est moins bonne que celle des locataires, donc prendre en compte les remplacements pour changement de qualité dans la construction de la valeur contrat serait plus prudent, pour ne pas surestimer la rentabilité des contrats de locataires.

Il est possible de regarder si des segments MRH sont particulièrement concernés par les remplacements pour changement de qualité :

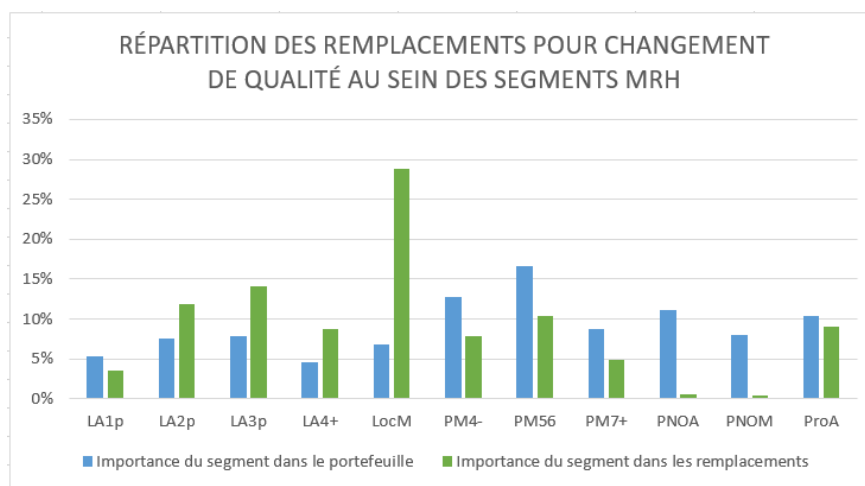


FIGURE 4.2 – Répartition des remplacements pour changement de qualité au sein des segments MRH, par rapport à la répartition des contrats en portefeuille au sein des segments MRH

Par rapport à l'importance des segments dans le portefeuille MRH, les remplacements pour changement de qualité concernent particulièrement les locataires de maisons, et les locataires d'appartements de plus de deux pièces, et, dans une mesure un peu moins importante, les propriétaires d'appartements.

De la même manière il est possible de regarder si les remplacements pour changement de qualité concernent les contrats d'une région AXA en particulier :

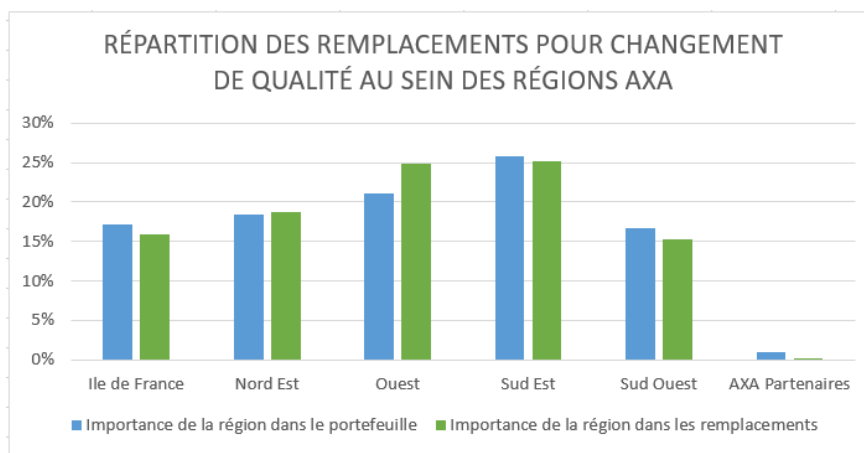


FIGURE 4.3 – Répartition des remplacements pour changement de qualité au sein des régions AXA, par rapport à la répartition des contrats en portefeuille au sein des régions AXA

La région Ouest présente une plus forte proportion de remplacements par rapport à sa place dans le portefeuille MRH, mais la différence par rapport aux autres régions est très peu marquée.

Il est aussi possible de s'intéresser à l'impact des caractéristiques des assurés sur la survenance de remplacements pour changement de qualité. Il a donc été observé si des tranches d'âge de clients étaient particulièrement concernées par ces remplacements :

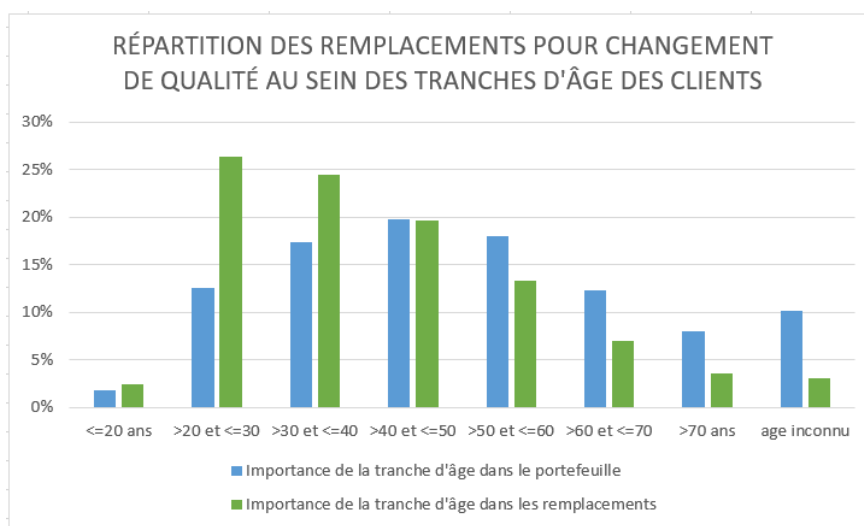


FIGURE 4.4 – Répartition des remplacements pour changement de qualité au sein des tranches d'âge des clients, par rapport à la répartition des contrats en portefeuille au sein des tranches d'âge des clients

Les assurés âgés entre 20 et 40 ans sont particulièrement concernés par les remplacements pour changement de qualité, et au contraire les assurés de plus de 50 ans sont moins concernés.

La répartition des remplacements au sein des différentes catégories socio-professionnelles des assurés a également été représentée :

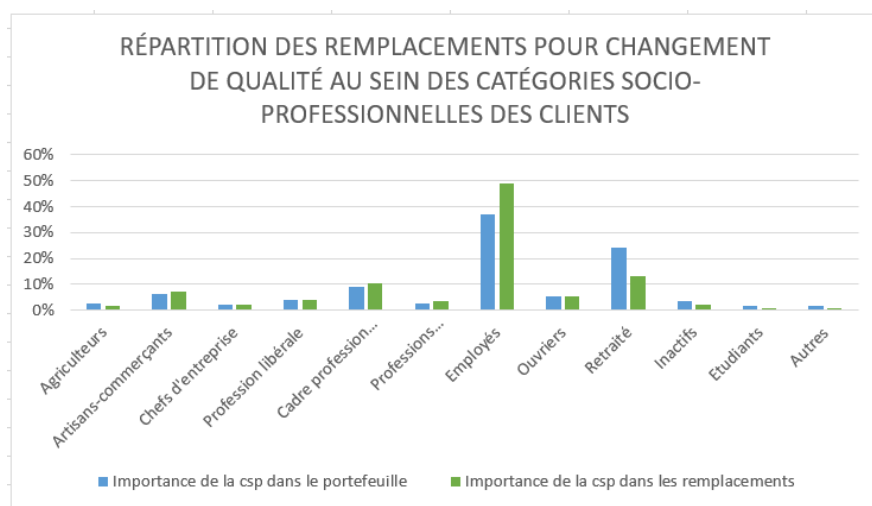


FIGURE 4.5 – Répartition des remplacements pour changement de qualité au sein des catégories socio-professionnelles des clients, par rapport à la répartition des contrats en portefeuille au sein des catégories socio-professionnelles des clients

Les employés sont plus concernés par les remplacements pour changement de qualité que les autres catégories socio-professionnelles, et au contraire les retraités et les étudiants sont moins concernés.

Cette étude statistique descriptive sur les remplacements pour changement de qualité a permis de se rendre compte que certains profils de contrats sont particulièrement concernés par ce type de remplacements. Le segment MRH du contrat et l'âge de l'assuré, par exemple, ont un impact important sur la survenance de ce type de remplacements.

4.2 Construction d'une nouvelle base de données

Pour modéliser la survie des contrats MRH avant remplacement pour changement de qualité, et avoir la probabilité de remplacement à chaque âge, pour chaque contrat, il a été envisagé de faire directement une modélisation semi-paramétrique. Pour cela, les informations des contrats à l'affaire nouvelle étaient donc nécessaires, comme pour la modélisation de la survie des contrats avant résiliation.

Pour modéliser la survie avant remplacement pour changement de qualité, il a été supposé que seul un remplacement de ce type pouvait survenir au cours de la vie d'un contrat. Il a donc été choisi de retenir uniquement le premier remplacement pour changement de qualité des contrats.

La nouvelle base de données a été construite en récupérant la base de données qui avait été construite dans le chapitre 3 pour la modélisation de la survie des contrats face au risque de résiliation, présentant une image par contrat. La base regroupe donc les informations contrats et clients, à la souscription, des contrats MRH souscrits entre le 01/01/2006 et le 31/12/2020. La variable indiquant le dernier âge observé pour le contrat a été modifiée. Elle correspond désormais au dernier âge observé avant remplacement si un remplacement pour changement de qualité a eu lieu, dans ce cas il n'y a pas de censure, ou au dernier âge observé du contrat s'il a été résilié sans avoir connu de remplacement, alors il y a eu une censure, ou au dernier âge observé du contrat fin 2020 si le contrat n'a pas été résilié ni remplacé à la fin de la période d'observation, dans ce cas il y a également une censure. La variable indiquant si l'évènement a été observé pour le contrat, ou s'il a été censuré, a été mis à jour par rapport à l'évènement de remplacement pour changement de qualité à la place de l'évènement de résiliation.

Les variables apportant des informations sur le contrat et l'assuré sont donc les mêmes que dans la base de données construite pour modéliser le risque de résiliation, celles-ci ne seront donc pas présentées à nouveau. Les traitements de la base de données qui avait été faits ont aussi été conservés (traitement des valeurs manquantes et aberrantes, et sélection de variables en fonction des corrélations).

Il est possible de représenter l'histogramme de la variable indiquant le dernier âge observé, qui sera la variable d'intérêt, pour les contrats non-censurés, l'objectif de ce chapitre étant d'estimer la probabilité de remplacement des contrats à chaque année de vie :

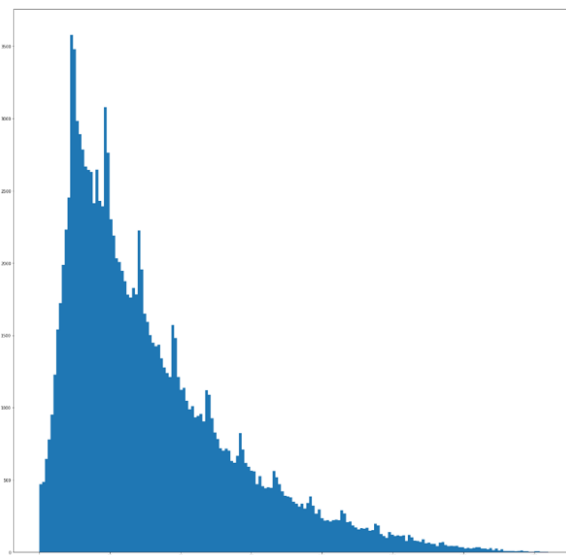


FIGURE 4.6 – Histogramme de la variable indiquant le dernier âge observé des contrats non-censurés

Il est visible sur cet histogramme représentant le nombre de remplacements pour changement de qualité de contrats à chaque mois de vie possible, qu'il y a une tendance de remplacement au terme mais qui est beaucoup moins marquée que la tendance de résiliation au terme qui avait été observée dans le chapitre 3.

La base de données obtenue regroupe les informations des mêmes 6 703 880 contrats que dans la base

de données présentée dans le chapitre 3.

4.3 Modélisation semi-paramétrique : Modèle de Cox

La construction d'un modèle de Cox permettra de prédire la probabilité de remplacement pour changement de qualité à chaque âge pour chaque contrat, selon ses caractéristiques à la souscription.

4.3.1 Vérification de l'hypothèse de proportionnalité des risques instantanés

Pour implémenter le modèle, il faut d'abord vérifier l'hypothèse de proportionnalité des risques instantanés, qui est l'hypothèse principale du modèle de Cox, présentée dans le chapitre 3.

Cette hypothèse a été vérifiée graphiquement pour chaque variable, de la même manière que lors de la modélisation de la survie avant résiliation. La fonction $\ln(-\ln(\hat{S}(t)))$ a été tracée, en utilisant l'estimation de la fonction de survie de Kaplan-Meier, pour les ensembles d'individus représentant chaque classe de la variable étudiée.

Pour les variables indiquant le type de franchise, si l'assuré est un enfant d'assuré, ou si c'est un contrat étudiant, les courbes construites se croisaient, cela signifiait donc que l'effet de ces covariables n'était pas indépendant du temps.

Par exemple cela est bien visible pour la variable indiquant si le contrat est un contrat étudiant :

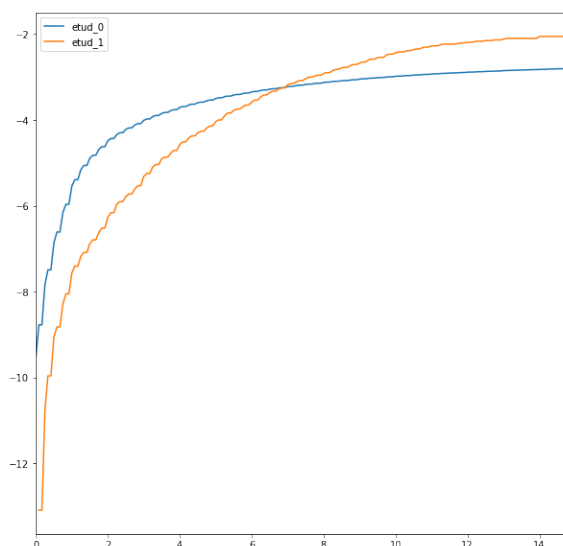


FIGURE 4.7 – Test de proportionnalité pour la variable indicatrice d'un contrat étudiant

Les variables citées précédemment, ne vérifiant pas l'hypothèse de proportionnalité des risques instantanés, ne peuvent donc pas être utilisées dans la construction d'un modèle de Cox pour prédire la survie des contrats avant remplacement pour changement de qualité. Il sera vu, dans la suite de l'étude, une solution pour prendre en compte ses informations si nécessaires.

4.3.2 Modélisation avec utilisation d'un algorithme de *Gradient Boosting*

Dans ce chapitre, il a directement été envisagé d'implémenter un modèle de Cox utilisant un algorithme de *Gradient Boosting* pour maximiser la vraisemblance partielle de Cox et déterminer les paramètres du modèle.

4.3.2.1 Préférence pour la modélisation avec utilisation du *Gradient Boosting*

Dans le chapitre 3, où l'objectif était de construire un modèle de Cox pour prédire la survie des contrats avant résiliation, il avait été vu que l'implémentation d'un modèle de Cox par *Gradient Boosting* était plus performante que l'implémentation d'un modèle de Cox classique.

Il a donc été envisagé dans ce chapitre, de directement construire un modèle par *Gradient Boosting*, mais il a tout de même été vérifié que cette méthode était aussi plus performante pour la prédiction de la survie des contrats avant remplacement.

Un modèle de Cox classique a donc été construit sur les données, sur les âges observés exacts (en durées mensuelles), avec la même méthodologie que dans le chapitre 3 (avec une sélection *Backward* des données,...). Il a été implémenté en *Python* avec utilisation du package *lifelines*. Un premier modèle de Cox *Gradient Boosting* a aussi été construit sur les données, avec la même méthodologie que dans le chapitre 3 (avec détermination des paramètres optimaux en fonction de la valeur de la concordance,...). Il a été implémenté en *Python* avec utilisation du package *XgBoost*.

Les valeurs de concordance obtenues lors de l'application des modèles à la base d'entraînement et à la base test ont été comparées :

	Base d'entraînement	Base test
Concordance modèle classique	0,7238	0,7246
Concordance modèle <i>Gradient Boosting</i>	0,8060	0,7728

TABLE 4.2 – Valeur de la concordance des prédictions du modèle de Cox classique et du modèle de Cox *Gradient Boosting*

Il apparaît que le modèle de Cox *Gradient Boosting* présente une concordance bien meilleure que le modèle de Cox classique. Cependant les valeurs obtenues ne sont pas très stables entre la base d'entraînement et la base test pour ce modèle. Dans la suite de l'étude, la construction d'un modèle de Cox *Gradient Boosting* sera donc approfondie de manière à construire un modèle optimal, si possible plus stable.

4.3.2.2 Construction d'un modèle sur les âges exacts

Un modèle de Cox *Gradient Boosting* a été construit sur les données présentant les durées de vie exacts avant remplacement des contrats.

Lors de la détermination des paramètres du modèle, plusieurs combinaisons de différentes valeurs de paramètres ont été étudiées, et il a été retenu que le modèle de Cox *Gradient Boosting* devait être construit avec, notamment, 180 arbres, une profondeur maximale des arbres de 25 et un taux d'apprentissage de 0,01.

Ce modèle permet d'obtenir, notamment, la courbe de survie avant remplacement de chaque contrat sur les 15 premières années de sa vie.

Il est possible d'observer la courbe de survie prédite moyenne sur tous les contrats associés à chaque segment MRH, et aussi sur les contrats associés aux segments LA1P, LA2P et LA3P selon la présence de la formule étudiante :

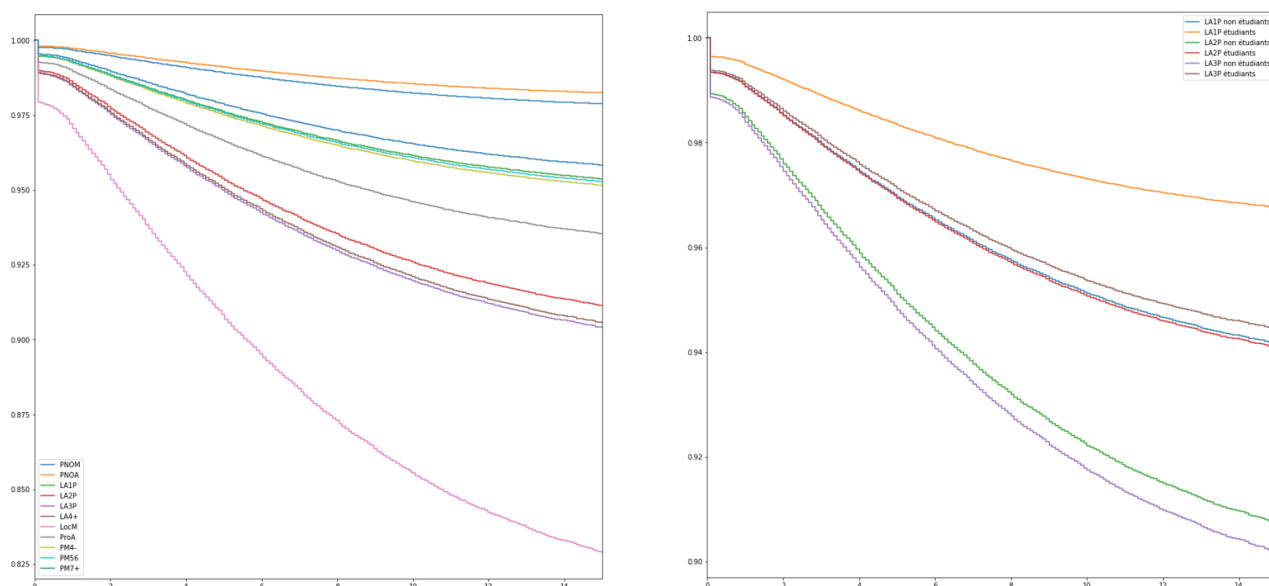


FIGURE 4.8 – Courbes de survie prédites, par le modèle de Cox *Gradient Boosting* sur les durées de vie exactes avant remplacement, moyennes par segment MRH (à gauche) et selon la présence de la formule étudiante et par segment MRH (à droite)

Il est à noter que les courbes de survie avant remplacement ont des valeurs beaucoup plus élevées que les courbes de survie avant résiliation. En effet, même à 15 ans, les valeurs moyennes de courbes de survie de tous les segments restent supérieures à 0,80. Ceci est dû au fait que tous les contrats ne vont pas être remplacés, donc contrairement à l'étude de la survie avant résiliation, une courbe de survie atteignant la valeur nulle n'est pas attendue. Il apparaît que les contrats de locataires de maison sont les plus exposés au risque de remplacement pour changement de qualité, cela est donc cohérent avec l'étude statistique descriptive effectuée précédemment. Aussi il est visible que les propriétaires non occupants sont très peu susceptibles de faire un remplacement de ce type. Cela paraît logique car, un locataire peut vouloir acheter un logement avec le temps, cependant un propriétaire non-occupant, qui utilise son logement pour le louer, par exemple, semble peu susceptible de vouloir un jour y habiter.

Aussi, la variable indiquant si le contrat est un contrat étudiant n'a pas pu être prise en compte par le modèle car elle ne vérifiait pas l'hypothèse de proportionnalité des risques instantanés. Il n'apparaît donc pas une différence évidente entre la survie avant remplacement des contrats étudiants et celle des contrats non étudiants.

Les courbes de survie avant remplacement moyennes par segment ont pu être comparées à des courbes de survie prédites par Kaplan-Meier par segment, sur les mêmes données. En particulier, au niveau des segments concernés par la formule étudiante il apparaît que les courbes de survie moyennes obtenues sur les contrats étudiants ne suivent pas la même tendance de remplacement que celles prédites par Kaplan-Meier :

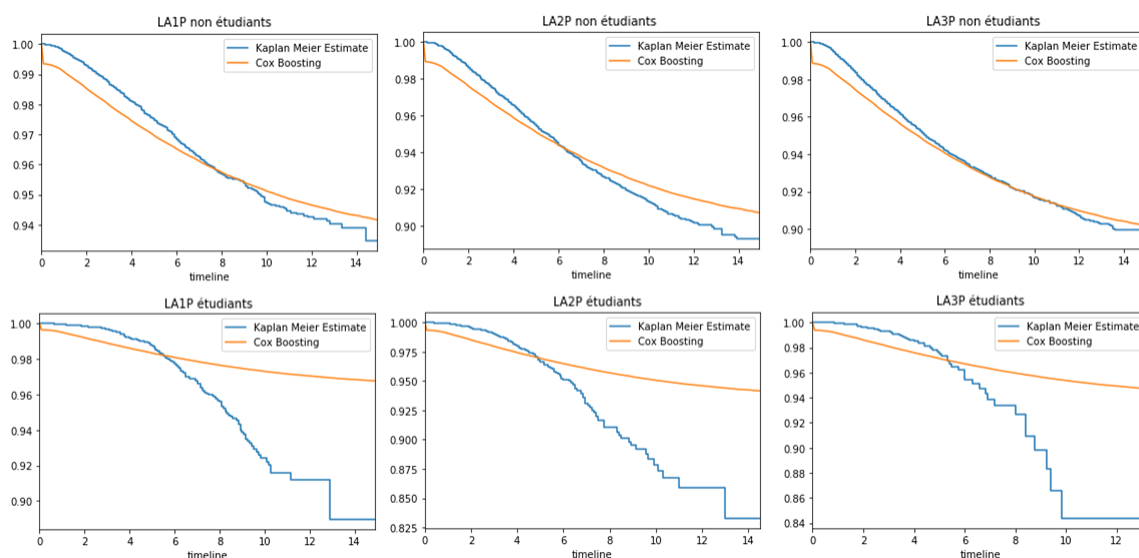


FIGURE 4.9 – Comparaison entre la courbe de survie avant remplacement moyenne prédite par le modèle de Cox, et la courbe de survie avant remplacement de Kaplan-Meier, par segment sur les contrats étudiants ou non étudiants

Il semble donc important de prendre en compte la situation d'étudiant dans la modélisation de la survie avant remplacement. La variable indiquant si c'est un contrat étudiant ne pouvant être prise en compte dans le modèle global, il faut construire un modèle spécifique sur les données des contrats étudiants.

Un modèle de Cox *Gradient Boosting* spécifique à été construit pour les contrats étudiants. Les paramètres optimaux déterminés pour ce modèle étaient, notamment, un nombre d'arbres de 170, une profondeur maximale des arbres de 16 et un taux d'apprentissage de 0,007.

Un modèle de Cox *Gradient Boosting* spécifique à été construit pour les contrats de locataires non-étudiants. Les paramètres optimaux déterminés pour ce modèle étaient, notamment, un nombre d'arbres de 150, une profondeur maximale des arbres de 22 et un taux d'apprentissage de 0,015.

Un modèle de Cox *Gradient Boosting* spécifique à également été construit pour les contrats de propriétaires. Les paramètres optimaux déterminés pour ce modèle étaient, notamment, un nombre d'arbres de 100, une profondeur maximale des arbres de 25 et un taux d'apprentissage de 0,004.

Aussi il a été décidé de négliger les remplacements pour changement de qualité pour les PNOs car moins de 1% des contrats PNOs observés étaient concernés.

Finalement cette nouvelle méthode de modélisation permet d'obtenir des courbes de survie avant remplacement moyennes par segment pour les contrats étudiants, plus proches des courbes de survie obtenues par Kaplan-Meier sur les mêmes catégories de contrats :

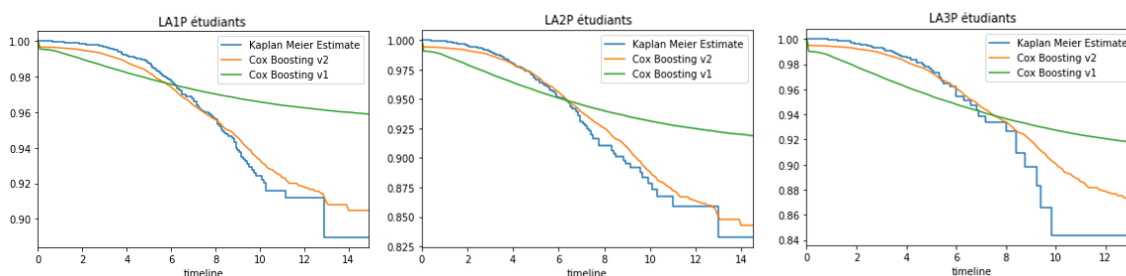


FIGURE 4.10 – Comparaison entre la courbe de survie avant remplacement moyenne prédite par le modèle de Cox construit sur l’ensemble des données, la courbe de survie avant remplacement moyenne prédite par le modèle de Cox construit spécifiquement pour les contrats étudiants, et la courbe de survie avant remplacement de Kaplan-Meier, par segment sur les contrats étudiants

Il est visible que cette nouvelle modélisation présente tout de même un inconvénient : comme le modèle a été construit uniquement sur les contrats étudiants, l’estimation des valeurs du risque instantané de base a été faite sur un nombre de données limité, et en particulier après 10 ans il y a des âges auxquels aucun remplacement n’a été observé pour ces données, et la courbe du risque instantané de base estimé présente donc des irrégularités. Cela peut se voir au niveau des courbes de survie moyennes par segment :

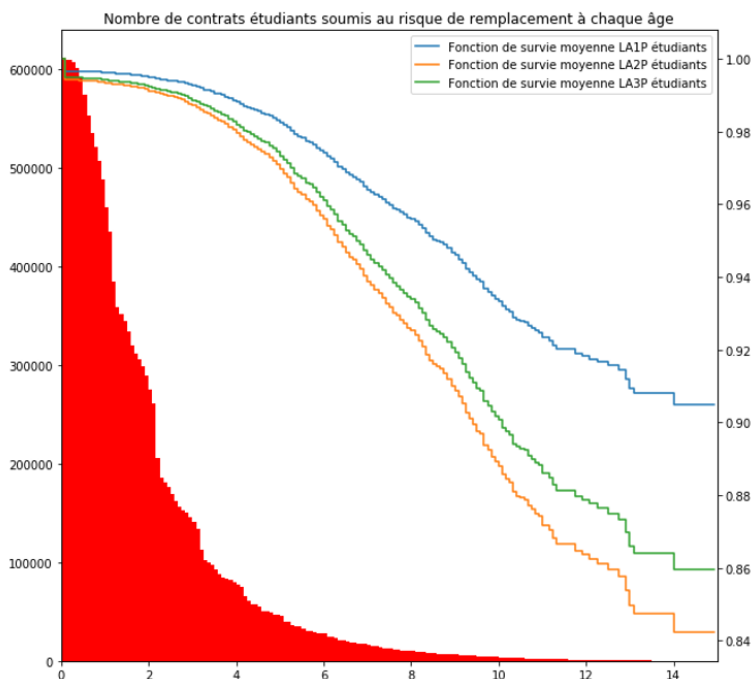


FIGURE 4.11 – Courbe de survie avant remplacement moyenne prédite par le modèle de Cox construit spécifiquement pour les contrats étudiants, par segment sur les contrats étudiants, et évolution du nombre de contrats étudiants soumis au risque de remplacement au fil des âges possibles des contrats

Pour pallier ce problème, il a été décidé de lisser les valeurs estimées de la courbe du risque instantané de base avec la méthode de lissage de Whittaker-Henderson, présentée dans la section 2.2.3.

Finalement, en représentant les courbes de survie avant remplacement moyennes par segment, pour les contrats étudiants et les non étudiants, obtenues avec la nouvelle méthode de modélisation, il apparaît qu’il y a une tendance de remplacement bien différente selon si l’assuré est étudiant ou non, et avec la mise en oeuvre du lissage des valeurs du risque instantané de base pour les étudiants, les courbes obtenues sont régulières :

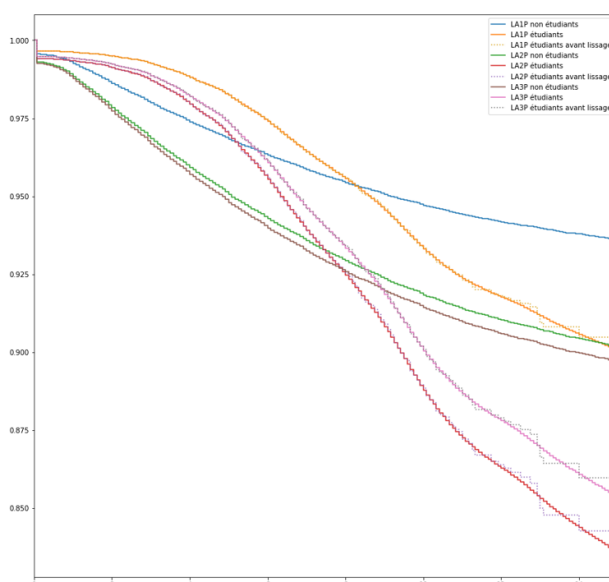


FIGURE 4.12 – Courbes de survie prédites, par les modèles de Cox *Gradient Boosting* sur les durées de vie exactes avant remplacement, construits en séparant les étudiants, des locataires non-étudiants, des propriétaires, moyennes selon la présence de la formule étudiante et par segment MRH

Finalement, pour s'assurer du fait que la modélisation de la survie avant remplacement des contrats faite en séparant les contrats étudiants, des locataires non-étudiants, des propriétaires, est plus performante que la modélisation faite sur l'ensemble des données, ces deux méthodes vont être comparées à d'autres niveaux.

Il est possible de se rendre compte de l'écart entre les estimations des valeurs de la courbe de survie des contrats du modèle de Cox *Gradient Boosting* initial, et les estimations des valeurs de la courbe de survie des contrats du modèle de Cox *Gradient Boosting* segmenté, avec celles du modèle de Kaplan Meier. Une estimation de l'espérance de vie, avant remplacement, à la souscription par segment MRH, selon les prédictions moyennes de la fonction de survie avant remplacement, de chaque modèle de Cox *Gradient Boosting* et des estimations de Kaplan-Meier, a donc été calculée dans ce but.

Les résultats obtenus sont présentés dans un tableau en annexe 11.

La modélisation segmentant les données présente des erreurs de prédiction plus faibles par rapport à Kaplan-Meier que la modélisation initiale. Au global, l'erreur du modèle initial est de 0,24% et celle du modèle segmenté est de -0,09%.

Enfin les valeurs de concordance des prédictions des deux modélisations, lors de leur application à la base d'entraînement et à la base test, ont été observées :

	Base d'entraînement	Base test
Concordance modèle <i>Gradient Boosting</i> initial	0,8060	0,7728
Concordance modèle <i>Gradient Boosting</i> segmenté	0,7857	0,7836

TABLE 4.3 – Valeur de la concordance des prédictions du modèle de Cox *Gradient Boosting* initial et du modèle segmenté

Le modèle de Cox *Gradient Boosting* construit sur les données segmentées, présente une valeur de concordance plus stable entre son application à la base d'entraînement et à la base de test, ce qui est

préférable.

Au vu des résultats obtenus lors des comparaisons effectuées entre le modèle de *Cox Gradient Boosting* construit sur l'ensemble des données, et le modèle de *Cox Gradient Boosting* segmenté en un modèle pour les contrats étudiants, un pour les locataires non-étudiants, et un pour les propriétaires, il apparaît que le modèle segmenté est plus performant et plus stable. C'est donc ce modèle qui sera utilisé pour la prédiction des probabilités de remplacement pour changement de qualité, dans la construction de la valeur contrat à la souscription.

Il est possible d'observer l'impact des variables sur le risque de remplacement, selon les trois modèles constituant le modèle segmenté :

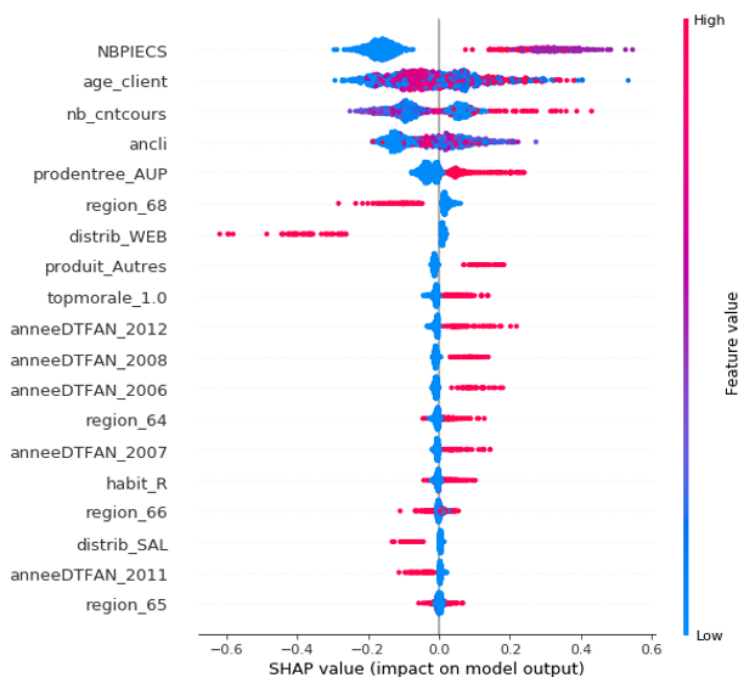


FIGURE 4.13 – Impact de la valeur des variables sur le risque instantané relatif de remplacement des contrats étudiants d'après le modèle de *Cox Gradient Boosting*

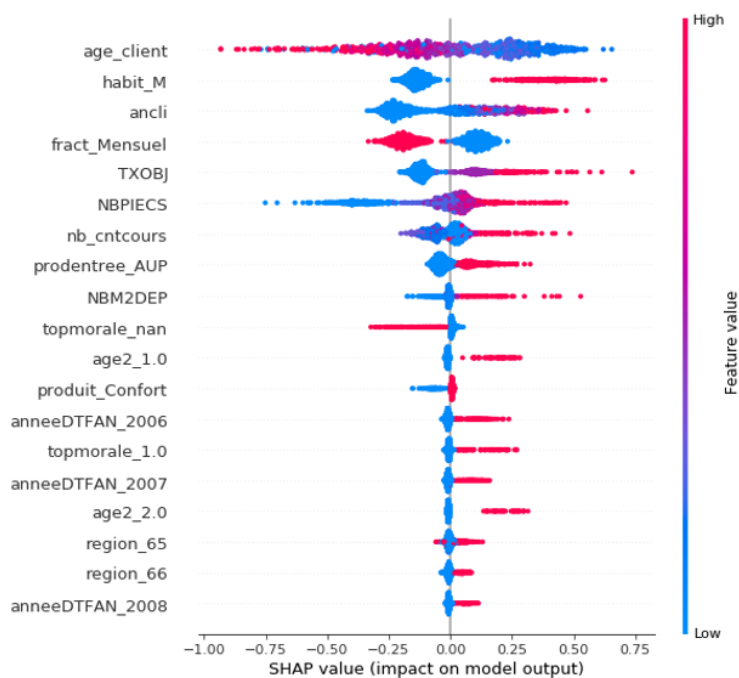


FIGURE 4.14 – Impact de la valeur des variables sur le risque instantané relatif de remplacement des contrats locataires non-étudiants d’après le modèle de Cox *Gradient Boosting*

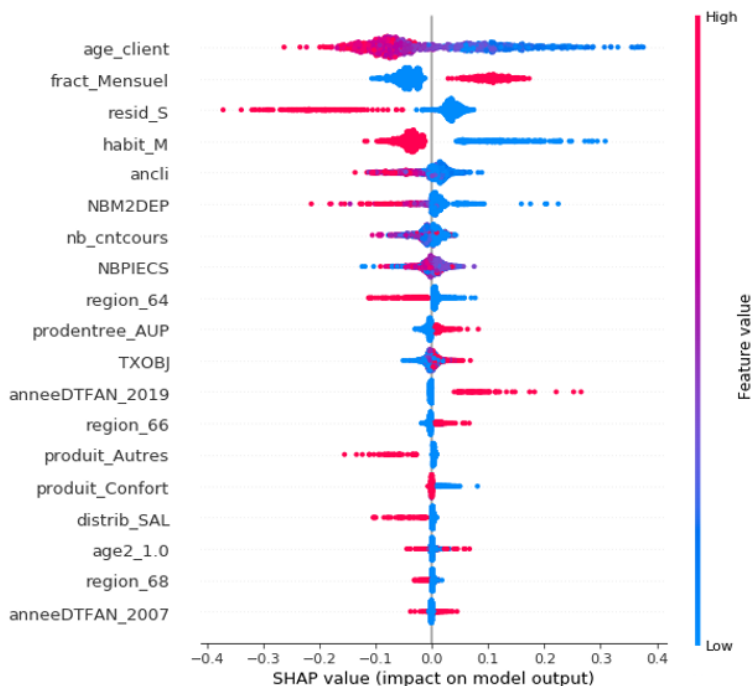


FIGURE 4.15 – Impact de la valeur des variables sur le risque instantané relatif de remplacement des contrats propriétaires d’après le modèle de Cox *Gradient Boosting*

En observant ces représentations visuelles il est possible de dire, par exemple, que d’après ces modèles de Cox, pour un contrat étudiant, la variable qui a le plus d’impact sur son risque de remplacement est la variable indiquant le nombre de pièces du logements. Plus le logement comporte de pièces, plus le risque instantané de remplacement du contrat pour changement de qualité est élevé, toutes choses égales par ailleurs. Cela semble logique car un étudiant souscrivant un contrat pour un logement de trois pièces serait

plus susceptible de rester à long terme dans son logement, que dans le cas d'un studio. Et finalement, avec le temps, il aurait plus de chance de désirer devenir propriétaire. Pour un contrat locataire non-étudiant ou propriétaire, la variable la plus importante, d'après les modèles, est la variable indiquant l'âge du client. Plus le client est jeune, plus le risque instantané de remplacement du contrat est élevé, toutes choses égales par ailleurs. Cela semble logique car plus des clients vieillissent, plus ils sont susceptibles d'avoir des enfants, et moins ils sont mobiles.

De la même manière que lors de la construction des modèles de Cox pour prédire le risque de résiliation, la qualité de la construction du modèle de Cox *Gradient Boosting* pour prédire le risque de remplacement pour changement de qualité peut être appréciée en utilisant la méthode de validation croisée.

Les valeurs de concordance mesurées sur l'application de chaque modèle de *Gradient Boosting* segmenté à la base de validation associée ont pu être représentées :

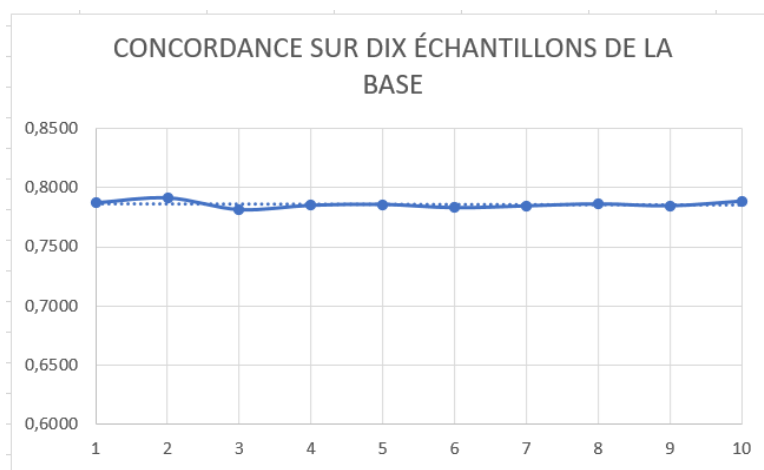


FIGURE 4.16 – Valeurs de la concordance des prédictions des modèles de Cox *Gradient Boosting* segmentés construits par validation croisée, sur dix sous-ensembles de validation de la base de données

L'indice de concordance est stable sur les dix sous-ensembles de validation. D'après les résultats de cette validation croisée, il n'y a pas de sur-apprentissage du modèle.

4.4 Synthèse du chapitre

L'objectif de ce chapitre était de construire un modèle de prédiction de la survie des contrats d'assurance MRH face au risque de remplacement pour changement de qualité, qui est un motif de remplacement ayant un fort impact sur la rentabilité des contrats.

Pour construire un modèle de survie semi-paramétrique il était nécessaire de disposer des informations des contrats à la souscription. La base de données qui avait été construite dans le chapitre 3, regroupant les informations contrats à la souscription des contrats d'assurance MRH souscrits entre 2006 et 2020, a été récupérée. Les informations sur la survie des contrats avant résiliation y ont été remplacées par les informations sur la survie de contrats avant remplacement

Des modèles de Cox ont été construits sur cette base de données. La construction de modèles utilisant l'algorithme du *Gradient Boosting* a été considérée comme plus pertinente que la construction de modèles de Cox classiques. Un premier modèle par *Gradient Boosting* a été construit sur l'ensemble des données (dit modèle initial), puis un second modèle par *Gradient Boosting* a été construit en séparant les contrats étudiants, des locataires non-étudiants, des propriétaires, et en négligeant le remplacement des contrats PNOs (dit modèle segmenté). Ces modèles permettaient d'avoir, pour chaque contrat, une prédiction de sa survie avant remplacement sur les 15 premières années de sa vie. Pour comparer la qualité de prédiction des deux modèles, les valeurs de l'indicateur de concordance sur leur application ont été étudiées. L'écart entre leurs prédictions et les prédictions faites par un modèle de Kaplan-Meier sur les mêmes données a été étudié. Finalement, le modèle de Cox *Gradient Boosting* segmenté est le meilleur modèle de survie testé pour prédire les probabilités de remplacement des contrats et a été retenu pour la construction de la valeur contrat.

— Chapitre 5 —

Construction de la valeur contrat à la souscription

5.1 Définition de l'indicateur de valeur contrat

La dernière étape de cette étude consiste à construire l'indicateur de valeur contrat à la souscription. Dans le chapitre 1, de contextualisation, la valeur contrat d'un contrat d'assurance a été définie comme la valeur actuelle probable, à la souscription, des flux de résultat futurs du contrat. Le résultat d'un contrat sur une année dépend des flux entrants, la prime commerciale, et des flux sortants, la prime pure du contrat sur cette année (l'espérance de la sinistralité), les montants des frais, les commissions et le coût de la réassurance associés à ce contrat sur cette année. La valeur contrat, d'un contrat x , s'exprime donc ainsi :

$$\text{Valeur contrat}(x) = \sum_{i=0}^{\infty} \mathbb{E}(\text{Résultat}_i(x)) \times \left(\frac{1}{1+\text{taux actualisation}_i}\right)^i$$

Avec pour toute année i : $\mathbb{E}(\text{Résultat}_i(x)) = \mathbb{E}(\text{Prime commerciale}_i(x)) - \mathbb{E}(\text{Prime pure}_i(x)) - \mathbb{E}(\text{Frais}_i(x)) - \mathbb{E}(\text{Commissions}_i(x)) - \mathbb{E}(\text{Réassurance}_i(x))$

L'estimation de chacun des éléments composants la valeur contrat sera présentée dans les sous-parties suivantes.

5.1.1 Probabilisation des situations dans lesquelles peuvent se trouver les contrats

Il a notamment été exposé dans le chapitre 1, qu'à chaque âge i , un contrat d'assurance habitation pouvait se trouver dans trois situations différentes. Pour calculer l'espérance des éléments constituant le résultat d'un contrat x en i , il faut donc connaître la probabilité que le contrat se trouve dans chacune des situations possibles.

Les trois situations possibles à l'âge i sont les suivantes :

- Le contrat a été résilié entre la souscription et l'âge i compris, et une fois résilié, le contrat n'enregistrera plus de flux pour tout âge postérieur à la résiliation ;
- Le contrat a vécu un remplacement pour changement de qualité (passage de locataire à propriétaire, ou inversement) entre la souscription et l'âge i compris, et une fois remplacé, les flux enregistrés les âges suivants le remplacement dépendront de l'impact du remplacement ;
- Le contrat n'a vécu aucun changement (n'a été ni résilié, ni remplacé pour changement de qualité) entre la souscription et l'âge i compris.

Pour rappel, il est supposé qu'un seul remplacement pour changement de qualité peut survenir au cours de la vie d'un contrat.

Il est possible de représenter visuellement la possible survenance des situations, au fil des moments de vie d'un contrat :

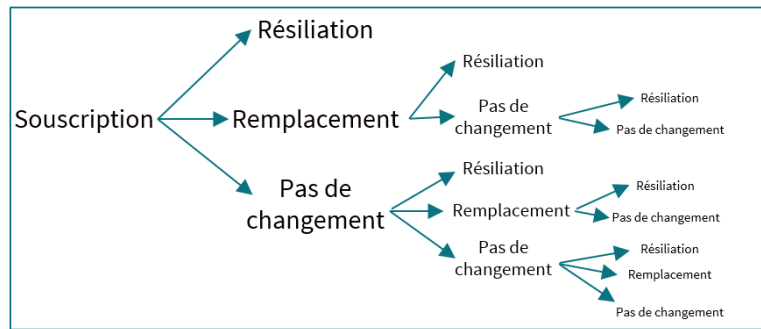


FIGURE 5.1 – Représentation visuelle des situations dans lesquelles peuvent se retrouver les contrats

Dans le chapitre 3, un modèle de durée semi-paramétrique a été construit pour prédire la probabilité de résiliation à chaque âge, jusqu'à 49 ans, des contrats d'assurance MRH, selon leurs caractéristiques à la souscription. Dans le chapitre 4, un autre modèle de durée a été construit pour prédire la probabilité de remplacement pour changement de qualité à chaque âge, jusqu'à 14 ans, des contrats, selon leurs caractéristiques à la souscription. Au delà de 14 ans, le risque de remplacement a été négligé.

Les modèles ont en fait permis de disposer, pour chaque contrat, d'une estimation de la fonction de survie face au risque de résiliation à chaque âge i , $\hat{S}^{résil}(i)$, et d'une estimation de la fonction de survie face au risque de remplacement à chaque âge i , $\hat{S}^{remp}(i)$. Les estimations des probabilités de résiliation, ${}_1\hat{q}_i$, et de remplacement, ${}_1\hat{r}_i$, à chaque âge i sont donc automatiquement déduites, ainsi que celle de la probabilité qu'il n'y ait aucun changement, ${}_1\hat{p}_i$:

$${}_1\hat{q}_i = 1 - \frac{\hat{S}^{résil}(i+1)}{\hat{S}^{résil}(i)}$$

$${}_1\hat{r}_i = 1 - \frac{\hat{S}^{remp}(i+1)}{\hat{S}^{remp}(i)}$$

$${}_1\hat{p}_i = 1 - {}_1\hat{q}_i - {}_1\hat{r}_i$$

Il est alors possible de disposer, pour chaque contrat et chaque âge i , des estimations des probabilités nécessaires pour la construction de la valeur contrat :

— La probabilité de résiliation sur $[0; i + 1[$: ${}_i\hat{q}_0 = 1 - \hat{S}^{résil}(i + 1)$

— La probabilité de remplacement sur $[0; i + 1[$:

$${}_{i+1}\hat{r}_0 = {}_1\hat{r}_0 \text{ pour } i = 0$$

$${}_{i+1}\hat{r}_0 = {}_1\hat{r}_0 {}_1\hat{p}_1 + {}_1\hat{p}_0 {}_1\hat{r}_1 \text{ pour } i = 1$$

$${}_{i+1}\hat{r}_0 = {}_1\hat{r}_0 \prod_{k=1}^i {}_1\hat{p}_k + \sum_{l=0}^{i-2} [(\prod_{j=0}^l {}_1\hat{p}_j) {}_1\hat{r}_{l+1} (\prod_{j=l+2}^i {}_1\hat{p}_j)] + \prod_{k=0}^{i-1} {}_1\hat{p}_k {}_1\hat{r}_i \text{ pour } i \geq 2$$

— La probabilité que la situation ne change pas sur $[0; i + 1[$: ${}_i\hat{p}_0 = 1 - {}_i\hat{q}_0 - {}_i\hat{r}_0$

Il est à noter que lorsqu'une résiliation ou un remplacement a lieu au cours de l'année i , il est considéré que la résiliation ou le remplacement a eu lieu en début d'année. Ainsi au moment de l'estimation de l'espérance du résultat en i , ce sont les probabilités de résiliation, et de remplacement sur $[0; i + 1[$ qui sont considérées.

5.1.2 Définition de l'espérance de la prime commerciale

Après détermination des probabilités nécessaires, il est possible d'estimer l'espérance de la première composante du résultat, la prime commerciale.

L'espérance de la prime commerciale du contrat x , l'année i , peut s'exprimer ainsi :

$$\mathbb{E}(Prime\ commerciale_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times PrimeCo(après\ remp)_i(x) + {}_{i+1}p_0(x) \times PrimeCo_i(x)$$

$PrimeCo_i(x)$ représente la prime commerciale en i du contrat x lorsqu'il n'a pas connu de changement, sa valeur va dépendre de la valeur de la prime commerciale à l'affaire nouvelle, $PrimeCo_0(x)$, et du taux de majoration au terme, $terme$, appliqué les années précédentes.

$PrimeCo(après\ remp)_i(x)$ correspond à la prime commerciale en i du contrat x dans le cas où le contrat a connu un remplacement, sa valeur va dépendre de la valeur de la prime commerciale à l'affaire nouvelle, $PrimeCo_0(x)$, du taux de majoration appliqué, noté $terme$, appliqué, et d'une estimation d'un taux d'évolution de la prime en raison du remplacement (il a été choisi de faire une estimation de ce taux selon les caractéristiques du risque assuré à l'affaire nouvelle), $TauxRempPrimeCo(x)$.

$$\mathbb{E}(Prime\ commerciale_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times PrimeCo_0(x) \times (1 + terme)^i \\ \times TauxRempPrimeCo(x) + {}_{i+1}p_0(x) \times PrimeCo_0(x) \times (1 + terme)^i$$

Le taux de terme est fixé par la politique de tarification d'AXA France et vaut actuellement 3%. Ne pouvant connaître les politiques tarifaires futures, il sera supposé, pour la construction de la valeur contrat, que ce taux sera constant au fil des années.

Le taux d'évolution de la prime en raison du remplacement a été estimé en moyenne sur chaque segment MRH. Pour chaque segment MRH des contrats à la souscription, a été calculé le rapport moyen entre la prime commerciale après remplacement et celle avant remplacement, pour les contrats ayant vécu un remplacement pour changement de qualité. Les taux d'évolution moyens de la prime commerciale lors d'un remplacement obtenus étaient les suivants :

Segment MRH	Taux d'évolution de la prime commerciale lors d'un remplacement
LA1p	1,56
LA2p	1,47
LA3p	1,44
LA4+	1,29
LocM	1,29
ProA	1,08
PM4-	1,00
PM56	0,92
PM7+	0,84

TABLE 5.1 – Taux d'évolution moyen de la prime commerciale lors d'un remplacement, selon le segment MRH du contrat à la souscription

Pour rappel le risque de remplacement pour changement de qualité est négligé pour les contrats de propriétaires non-occupants.

Tous les éléments présentés dans cette sous-partie permettent donc l'estimation de l'espérance de la prime commerciale à chaque âge pour chaque contrat d'assurance MRH.

5.1.3 Définition de l'espérance de la prime pure

L'espérance de la deuxième composante du résultat, la prime pure, peut également être estimée.

L'espérance de la prime pure du contrat x , l'année i , peut s'exprimer ainsi :

$$\mathbb{E}(Prime\ pure_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times PrimePure(après\ remp)_i(x) + {}_{i+1}p_0(x) \times PrimePure_i(x)$$

$PrimePure_i(x)$ représente la prime pure en i du contrat x lorsqu'il n'a pas connu de changement, sa valeur va dépendre de la valeur de la prime pure à l'affaire nouvelle, $PrimePure_0(x)$, et d'un taux estimé, représentatif de la tendance d'évolution de la prime pure au fil des années, $TauxPP$, qui reflète la

tendance d'évolution de la fréquence de sinistralité, ou la tendance d'évolution du coût de réparation des sinistres.

$PrimePure(après\ remp)_i(x)$ correspond à la prime pure en i du contrat x dans le cas où le contrat a connu un remplacement, sa valeur va dépendre de la valeur de la prime pure à l'affaire nouvelle, $PrimePure_0(x)$, du taux d'évolution de la prime pure au fil des années, $TauxPP$, et d'une estimation d'un taux d'évolution de la prime pure en raison du remplacement (l'estimation de ce taux a été faite selon si l'assuré est un locataire ou un propriétaire à la souscription), $TauxRempPrimePure(x)$.

$$\mathbb{E}(Prime\ pure_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times PrimePure_0(x) \times (1 + TauxPP)^i \\ \times TauxRempPrimePure(x) + {}_{i+1}p_0(x) \times PrimePure_0(x) \times (1 + TauxPP)^i$$

Le taux d'évolution de la prime pure au fil des années a été estimé en calculant le taux d'évolution de la prime pure du portefeuille MRH observé ces dernières années. Pour cela, la valeur de la charge sinistre du portefeuille, vieillie sur 2 ans, a été observée entre 2009 et 2017 et divisée par la valeur de l'exposition du portefeuille l'année correspondante. Il a alors été observé que la prime pure a eu tendance à augmenter de 0,62%, en moyenne, d'une année à la suivante :

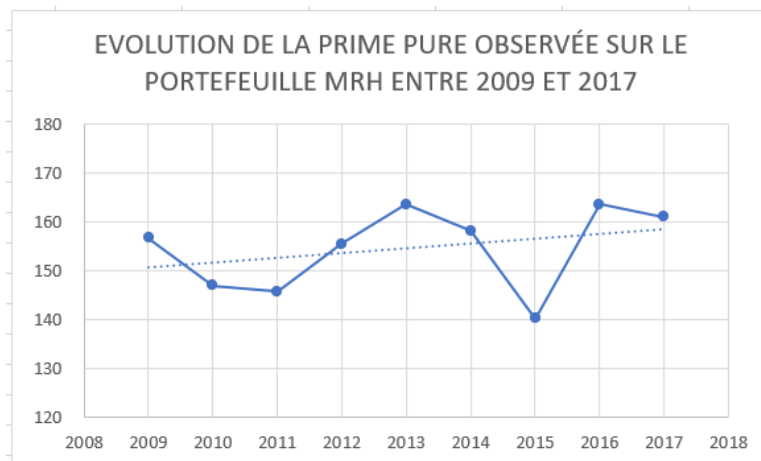


FIGURE 5.2 – Evolution de la prime pure observée sur le portefeuille MRH entre 2009 et 2017, avec deux ans de vieillissement de la charge sinistre

Il a été supposé que la prime pure conserverait cette tendance d'évolution moyenne à l'avenir. Le taux d'évolution de la prime pure d'une année à la suivante a donc été fixé à 0,62% pour la construction de la valeur contrat. Ce taux permet donc de capter une tendance d'augmentation de la gravité des sinistres, ou l'inflation des coûts de réparation.

Pour estimer le taux d'évolution de la prime pure en raison du remplacement, il n'a pas été possible d'utiliser la même méthode que pour l'estimation du taux d'évolution de la prime commerciale en raison d'un remplacement. Les valeurs de la prime pure des contrats avant et après remplacement pour changement de qualité n'étaient pas disponibles. Alors un taux moyen d'évolution de la prime pure lors d'un remplacement pour changement de qualité a été estimé, pour l'ensemble des contrats de locataires comme étant le rapport entre la prime pure moyenne à l'affaire nouvelle des propriétaires et celle des locataires, et pour les contrats de propriétaires comme étant le rapport entre la prime pure moyenne à l'affaire nouvelle des locataires et celle des propriétaires. Pour les contrats de locataires le taux d'évolution de la prime pure moyen obtenu était de 5,03 et pour les contrats de propriétaires de 0,20.

Il est désormais possible d'estimer l'espérance de la prime pure à chaque âge des contrats d'assurance MRH.

5.1.4 Définition de l'espérance des frais

Les frais associés à un contrat constituent la troisième composante du résultat du contrat. Leur espérance peut être estimée.

Les frais sont divisés en frais fixes, qui sont des montants uniques associés à tous les contrats, et en frais variables, qui sont des taux applicables à la prime commerciale de chaque contrat.

L'espérance des frais du contrat x , l'année i , peut s'exprimer ainsi :

$$\mathbb{E}(Frais_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Frais(après remp)_i(x) + {}_{i+1}p_0(x) \times Frais_i(x)$$

$Frais_i(x)$ représente les frais en i du contrat x lorsqu'il n'a pas connu de changement, sa valeur va dépendre de la valeur des frais fixes en i , $FraisFixes_i$, et des taux de frais variables en i , $FraisVar_i$, appliqués à la prime commerciale en i , $PrimeCo_i(x)$.

$Frais(après remp)_i(x)$ représente les frais en i du contrat x dans le cas où le contrat a connu un remplacement, sa valeur va dépendre de la valeur des frais fixes en i , $FraisFixes_i$, et des taux de frais variables en i , $FraisVar_i$, appliqués à la prime commerciale après remplacement en i , $PrimeCo(après remp)_i(x)$.

$$\mathbb{E}(Frais_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times [FraisFixes_i + FraisVar_i \times PrimeCo(après remp)_i(x)] + {}_{i+1}p_0(x) \times [FraisFixes_i + FraisVar_i \times PrimeCo_i(x)]$$

Les frais sont constitués :

- des frais d'acquisition, qui sont appliqués uniquement la première année des contrats, et représentent par exemple le coût des commissions d'acquisition, et les frais d'ouverture de dossier, ou de publicité ;
- des frais d'administration, qui représentent par exemple les frais de personnel, d'encaissement des primes, d'administration du portefeuille, ou de gestion des participations aux bénéficiaires ;
- d'autres charges techniques, qui correspondent à des frais internes et externes qui exceptionnellement ne peuvent être définis par la norme ;
- des charges financières, qui correspondent par exemple aux intérêts bancaires, et intérêts sur opérations de financement ;
- des frais de gestion de sinistres.

La connaissance des montants de frais fixes et des taux de frais variables permet d'estimer l'espérance des frais, à chaque âge, des contrats.

5.1.5 Définition de l'espérance des commissions

La quatrième composante du résultat du contrat correspond aux commissions. Leur espérance peut être estimée.

L'espérance des commissions associées au contrat x , l'année i , peut s'exprimer ainsi :

$$\mathbb{E}(Commission_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Com(après remp)_i(x) + {}_{i+1}p_0(x) \times Com_i(x)$$

$Com_i(x)$ représente la valeur des commissions en i du contrat x lorsqu'il n'a pas connu de changement, sa valeur va dépendre d'un taux de commissions, Com , appliqué à la prime commerciale en i , $PrimeCo_i(x)$.

$Com(après remp)_i(x)$ représente la valeur des commissions en i du contrat x dans le cas où le contrat a connu un remplacement, sa valeur va dépendre d'un taux de commissions, Com , appliqué à la prime commerciale après remplacement en i , $PrimeCo(après remp)_i(x)$.

$$\mathbb{E}(Commissions_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Com \times PrimeCo(après remp)_i(x) + {}_{i+1}p_0(x) \times Com \times PrimeCo_i(x)$$

Le taux de commissions est un taux fixé à 20% qui correspond au taux de rémunération des agents, sur la prime de chaque contrat de leur portefeuille.

L'espérance des commissions, à chaque âge, des contrats peut donc également être estimée.

5.1.6 Définition de l'espérance du coût de la réassurance

Le coût de la réassurance associé à un contrat correspond à la dernière composante du résultat du contrat. Son espérance peut être estimée.

L'espérance du coût de la réassurance du contrat x , l'année i , peut s'exprimer ainsi :

$$\mathbb{E}(Réassurance_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Réass(après remp)_i(x) + {}_{i+1}p_0(x) \times Réass_i(x)$$

$Réass_i(x)$ représente la valeur du coût de réassurance en i du contrat x lorsqu'il n'a pas connu de changement, sa valeur va dépendre d'un taux de coût de réassurance, $Réass$, appliqué à la prime commerciale en i , $PrimeCo_i(x)$.

$Réass(après remp)_i(x)$ représente la valeur du coût de réassurance en i du contrat x dans le cas où le contrat a connu un remplacement, sa valeur va dépendre d'un taux de coût de réassurance, $Réass$, appliqué à la prime commerciale après remplacement en i , $PrimeCo(après remp)_i(x)$.

$$\mathbb{E}(Réassurance_i(x)) = {}_{i+1}q_0(x) \times 0 + {}_{i+1}r_0(x) \times Réass \times PrimeCo(après remp)_i(x) + {}_{i+1}p_0(x) \times Réass \times PrimeCo_i(x)$$

Le taux de coût de réassurance est fixé à 3,6% et est déterminé par les traités de réassurance en vigueur. Il est supposé dans la construction de la valeur contrat, que ce taux n'évoluera pas dans le futur.

5.1.7 Définition du taux d'actualisation

La valeur contrat à la souscription correspond à la valeur actuelle probable, à la souscription, des flux de résultat futurs du contrat. Une actualisation des estimations annuelles de l'espérance du résultat est donc nécessaire.

Le taux d'actualisation peut s'exprimer de la manière suivante :

$$taux\ actualisation_i = rf_i + (\mathbb{E}(r_M) - rf_i) \times \beta_{AXA}$$

Avec rf_i le taux sans risque de l'année i , $\mathbb{E}(r_M)$ l'espérance de rentabilité du marché, et β_{AXA} la sensibilité de l'action AXA par rapport au marché.

Ces éléments sont communiqués par la direction *Risk Management* IARD d'AXA France, et permettent la définition du taux d'actualisation à utiliser pour la construction de la valeur contrat à la souscription des contrats d'assurance MRH.

Après définition du taux d'actualisation, tous les éléments nécessaire à la construction de la valeur contrat à la souscription sont disponibles.

5.2 Étude des valeurs contrat obtenues sur les affaires nouvelles de 2021

Après détermination de tous les éléments nécessaires à la construction de la prédiction de la valeur contrat à la souscription, présentés précédemment, une valeur contrat peut être estimée pour chaque nouvelle affaire. Un code a été développé en *Python* dans ce but, et a été appliqué aux affaires nouvelles MRH souscrites entre janvier et juin 2021.

Il est possible de se rendre compte de la valeur contrat moyenne par segment MRH, obtenue sur les affaires nouvelles de 2021 :

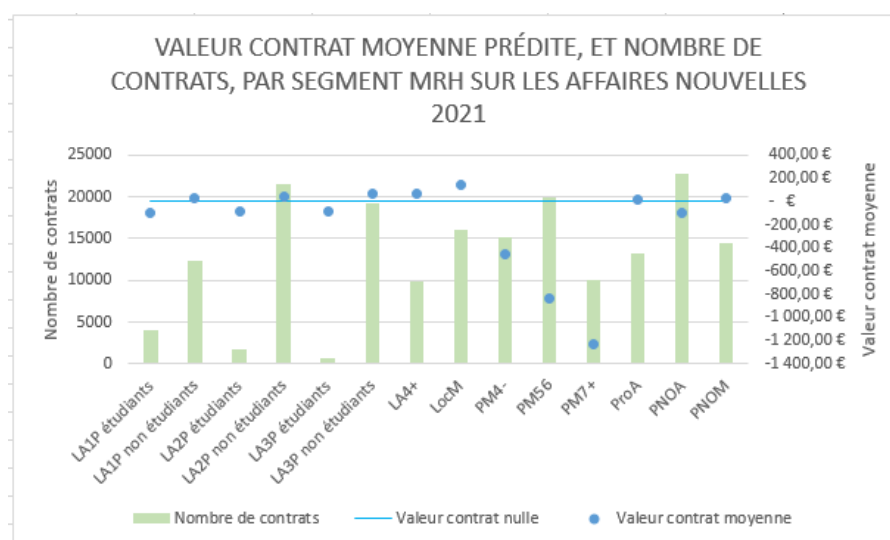


FIGURE 5.3 – Valeur contrat moyenne prédite par segment MRH sur les affaires nouvelles de 2021

Il apparaît que les contrats étudiants, les contrats de propriétaires de maisons, et ceux de propriétaires non-occupants d'appartement, ont une valeur contrat estimée négative, et particulièrement faible pour les propriétaires de maisons.

Pour chaque contrat, la marge attendue, correspondant au rapport entre la valeur contrat prédite et la valeur actuelle probable des flux d'espérance de la prime commerciale, a été calculée. La marge attendue globale sur l'ensemble des contrats est de -13,55%. La marge attendue globale par segment est présentée ci-dessous :

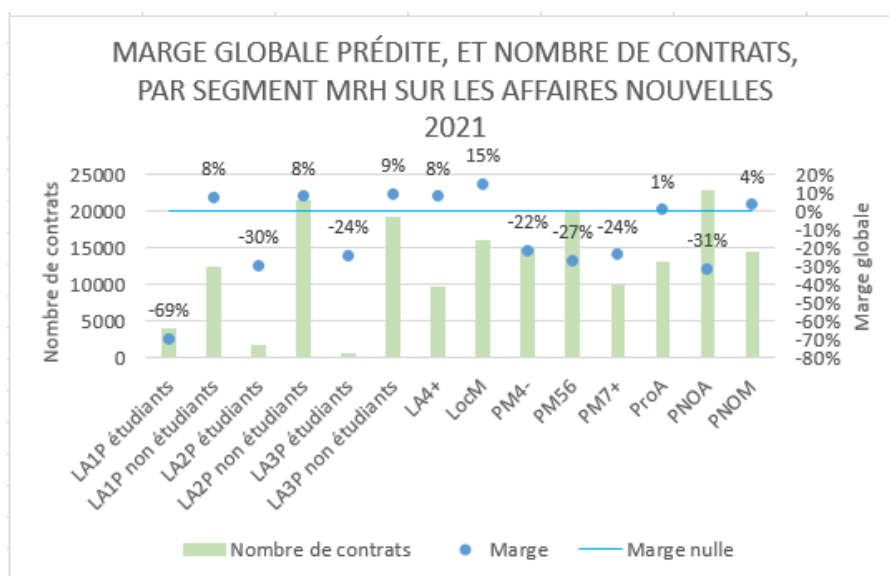


FIGURE 5.4 – Marge globale prédite par segment MRH sur les affaires nouvelles de 2021

En observant la marge attendue plutôt que la valeur contrat attendue, il apparaît que ce sont en fait les contrats étudiants de locataires d'appartements une pièce qui sont les moins rentables. Cependant ce type de contrats ne représentent que 2,25% des affaires nouvelles observées, et leur prime moyenne est relativement faible, autour de 130€ annuellement, ainsi la perte de résultat causée par ce segment ne représentera qu'une infime partie de la perte de résultat totale. C'est en effet ce qui a été observé lors du calcul de la valeur contrat moyenne par segment.

La volatilité de la marge des contrats au sein des segments MRH a été observée en représentant la proportion de contrats au sein des différentes tranches de marge possibles, par segment :

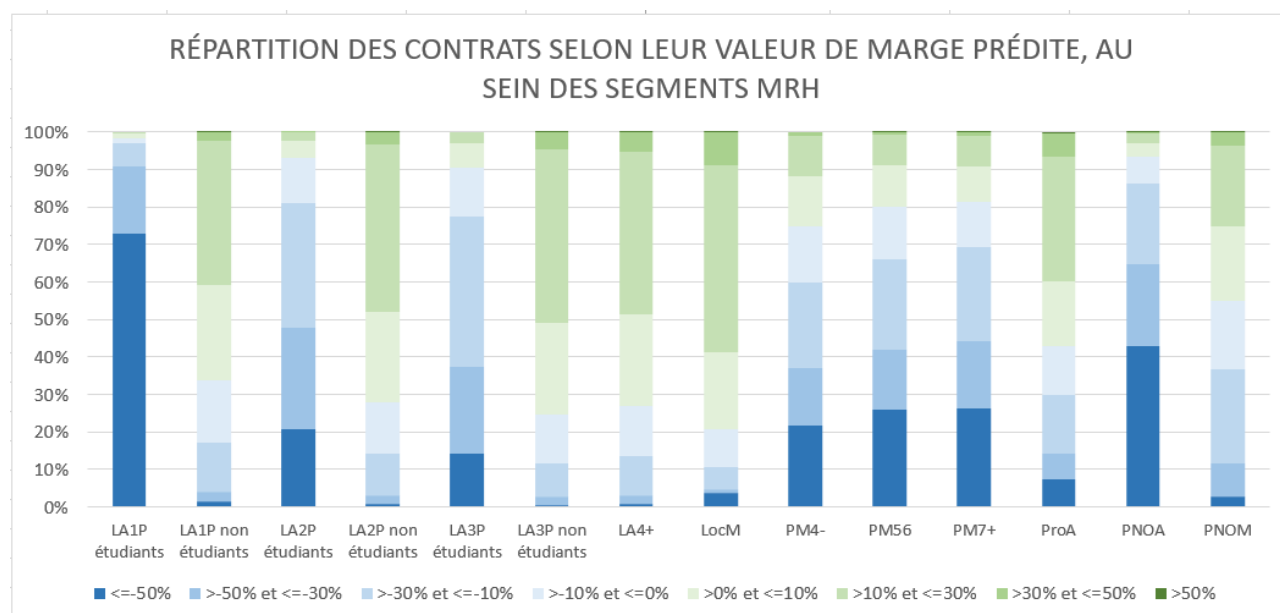


FIGURE 5.5 – Proportion de contrats au sein des différentes tranches de marge prédites, par segment sur les affaires nouvelles de 2021

Cette représentation permet de se rendre compte de la proportion de contrats dont la marge prédite est positive, au sein des segments. Les segments présentant le plus de contrats dont la marge prédite est positive sont les segments des locataires d'appartements de trois pièces non-étudiants (74%), des locataires d'appartements de plus de quatre pièces (72%) et des locataires de maisons (79%).

Il est possible d'observer l'évolution prédite de la valeur contrat globale cumulée sur les affaires nouvelles de 2021, selon le segment. Cela permet de se rendre compte de l'âge à partir duquel les contrats restants ne font plus augmenter ou diminuer la valeur contrat globale, en raison de leur faible nombre.

Cette évolution a, par exemple, été représentée pour les locataires non-étudiants dans le graphique suivant :

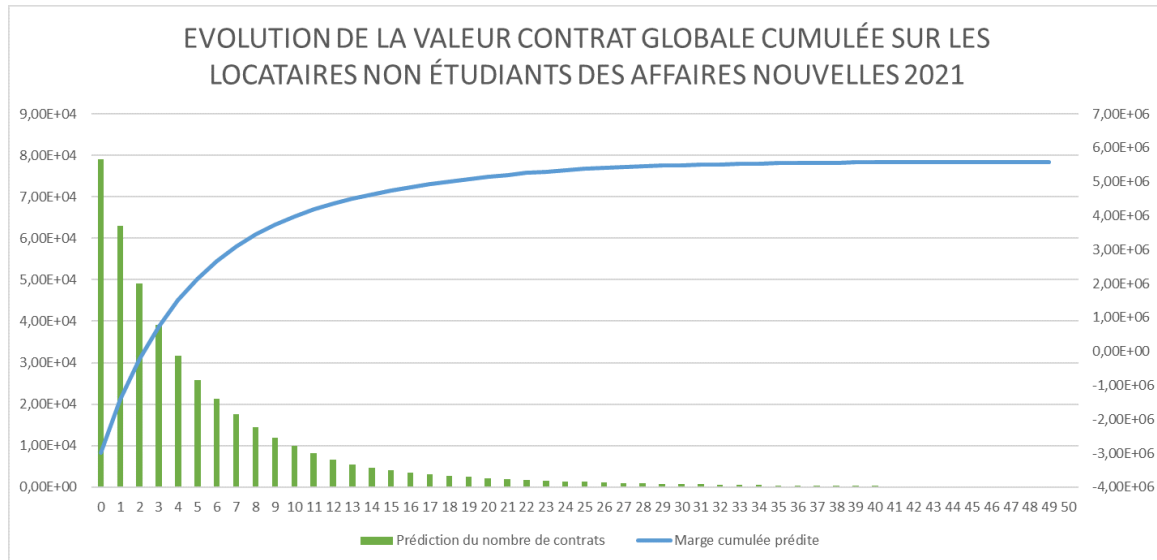


FIGURE 5.6 – Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des locataires non-étudiants

Les graphiques représentatifs des évolutions des autres segments sont présentés en annexe 12.

Pour les contrats de locataires étudiants, la valeur contrat globale cumulée se stabilise à une valeur négative à partir d'environ 12 ans, après avoir diminué. Pour les contrats de locataires non-étudiants, elle augmente puis se stabilise à une valeur positive à partir d'environ 28 ans. Pour les contrats de propriétaires elle diminue, puis augmente légèrement, et se stabilise à une valeur négative à partir d'environ 27 ans. Pour les contrats de propriétaires non-occupants d'appartements, elle diminue puis se stabilise à une valeur négative à partir d'environ 10 ans. Et enfin, pour les contrats de propriétaires non-occupants de maisons, elle augmente puis se stabilise à une valeur positive à partir d'environ 37 ans.

Au fil des années, le nombre de contrats de locataires non-étudiants diminue plus vite que celui des propriétaires. Mais les résultats annuels prédits d'un contrat de locataire non-étudiant sont bien meilleurs que les résultats prédits d'un contrat de propriétaire. Donc finalement la stabilisation de la marge globale des contrats de locataires non-étudiants est attendue environ au même âge que la stabilisation de la marge globale des contrats de propriétaires.

Pour aller plus loin sur l'étude des valeurs contrats obtenues sur les affaires nouvelles de 2021, une première application d'optimisation a été faite. L'objectif était de définir un scénario améliorant le plus possible la valeur contrat globale, tel que moins de 3% des contrats auraient été souscrits dans un autre segment MRH que celui réellement souscrit. C'est donc un scénario d'optimisation sous contrainte de volumétrie globale invariante. La finalité de ce travail est de définir quels segments de contrats sont les moins intéressants à souscrire, et lesquels sont les plus intéressants à souscrire pour AXA.

Selon l'étude précédente de la valeur contrat moyenne prédite par segment, les segments dont les contrats ont les valeurs contrat les plus faibles sont les contrats de propriétaires de maison. Plusieurs scénarios de diminution du nombre de contrat au sein de ces segments ont été étudiés. Le scénario qui améliorerait le plus la valeur contrat globale était une diminution du nombre de contrats de propriétaires de maisons de plus de sept pièces de 40%. Cela correspondait à une diminution du nombre de contrats de 2,2% au global, et à une amélioration de la valeur contrat globale de 15,1%.

Au sein du segment des PM7+, il a été cherché de définir les contrats les moins rentables en particulier. En observant la valeur contrat moyenne prédite des contrats PM7+ selon le nombre de pièces du logement, il est apparu que les contrats d'assurance de logements d'exactly sept pièces sont les moins rentables :

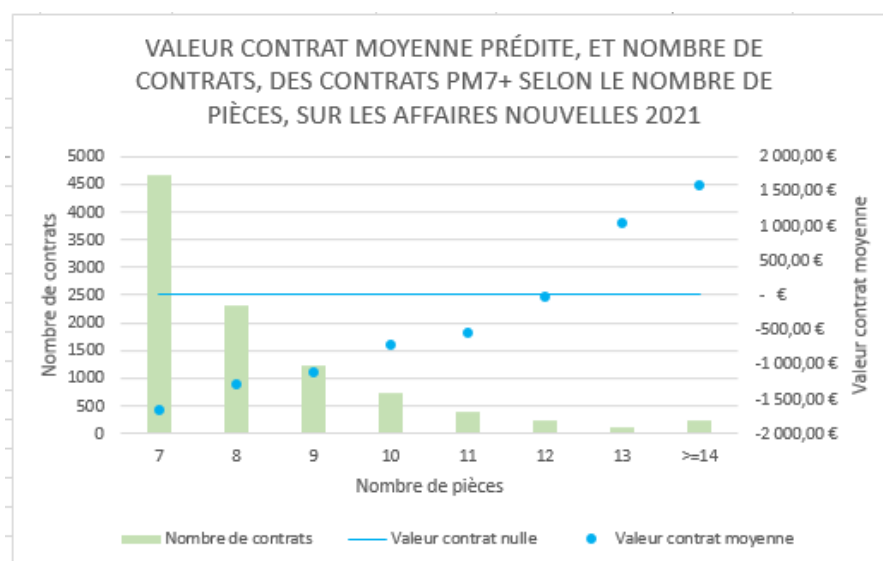


FIGURE 5.7 – Valeur contrat moyenne prédite des contrats PM7+ selon le nombre de pièces du logement, sur les affaires nouvelles de 2021

En supprimant autant de contrats mais uniquement des contrats de propriétaires de maisons de sept pièces, la valeur contrat globale était finalement améliorée de 20,2%.

Selon l'étude de la valeur contrat moyenne prédite par segment, le segment dont les contrats ont les valeurs contrat les plus élevées sont les contrats de locataires de maisons. Il a donc été envisagé que les contrats de propriétaires de maisons de sept pièces supprimés soient en fait de nouveaux contrats souscrits par des locataires de maisons. Selon ce scénario la valeur contrat globale était alors améliorée de 21,9%.

Cette étude de la valeur contrat des affaires nouvelles de 2021 a donc permis de simuler une stratégie de souscription qui serait bénéfique pour le résultat de la compagnie, qui consisterait à moins souscrire de contrats de propriétaires de maison de sept pièces et de plutôt souscrire des contrats de locataires de maisons. Cependant le changement doit être mineur, concerner, par exemple, moins de 3% des contrats au global, car AXA s'est fixé des objectifs de chiffre d'affaires à vérifier. Il faut donc continuer à souscrire un minimum de propriétaires de maisons, qui présentent des primes importantes. Ceci est un exemple d'utilité de la valeur contrat à la souscription.

Une seconde application a été faite. L'objectif était de déterminer un taux par segment, à appliquer à la prime commerciale des contrats à l'affaire nouvelle, sans modifier les autres paramètres, qui permettrait d'obtenir une marge globale prédite nulle sur l'ensemble des contrats de chaque segment MRH. Les taux déterminés sont les suivants :

Segment MRH	Taux d'évolution à appliquer à la prime commerciale à l'affaire nouvelle
LA1p étudiants	1,99
LA1p non-étudiants	0,89
LA2p étudiants	1,42
LA2p non-étudiants	0,88
LA3p étudiants	1,34
LA3p non-étudiants	0,87
LA4+	0,89
LocM	0,79
ProA	0,99
PM4-	1,31
PM56	1,38
PM7+	1,33
PNOA	1,45
PNOM	1,33

TABLE 5.2 – Taux d'évolution par segment MRH à appliquer à la prime commerciale des contrats à l'affaire nouvelle, pour prédire une marge globale par segment nulle

La prime commerciale à l'affaire nouvelle des contrats des propriétaires de maisons, des locataires d'appartements étudiants, et des propriétaires non-occupants d'appartements devrait être augmentées pour obtenir une marge globale nulle par segment. À l'inverse, la prime commerciale à l'affaire nouvelle des contrats des locataires d'appartements non-étudiants, des locataires de maisons, des propriétaires d'appartements, et des propriétaires non-occupants de maisons devrait être légèrement augmentée.

La prime moyenne à l'affaire nouvelles des contrats, par segment, avec et sans application du taux d'évolution déterminé, a été calculée :

Segment MRH	Prime moyenne à l'affaire nouvelle sans évolution (en €)	Prime moyenne à l'affaire nouvelle avec évolution (en €)
LA1p étudiants	54,9	109,26
LA1p non-étudiants	119,85	106,66
LA2p étudiants	86,38	122,65
LA2p non-étudiants	142,69	125,57
LA3p étudiants	98,45	131,92
LA3p non-étudiants	163,57	142,31
LA4+	206,13	183,46
LocM	237,94	187,97
ProA	220,95	218,75
PM4-	258,12	338,14
PM56	370,47	511,26
PM7+	586,65	780,25
PNOA	73,21	106,15
PNOM	201,69	191,61

TABLE 5.3 – Prime moyenne à l'affaire nouvelle des contrats, par segment, avec et sans application du taux d'évolution à appliquer pour prédire une marge globale par segment nulle

La nouvelle prime moyenne à appliquer à l'affaire nouvelle prend en compte l'espérance de vie des contrats au sein des segments. Ainsi, les contrats LA1P étudiants, ayant une espérance de vie plus faible que les contrats LA1P non-étudiants, ont une prime commerciale moyenne à l'affaire nouvelle, permettant de prédire une marge globale nulle sur le segment, plus élevée que celle des contrats LA1P non-étudiants.

L'étude de la prédiction de la valeur contrat des affaires nouvelles de 2021 peut donc aider à définir de nouvelles stratégies tarifaires à la souscription qui seraient bénéfiques pour le résultat de la compagnie.

5.3 Tests de sensibilité

Pour construire l'indicateur de valeur contrat à la souscription, beaucoup d'hypothèses ont été prises. Ces hypothèses peuvent être remises en question. En effet elles ont été prises selon la situation actuelle, mais dans le futur le contexte pourrait évoluer. Il serait pertinent de voir l'impact qu'aurait la modification de certaines hypothèses sur les valeurs contrat des affaires nouvelles de 2021 par exemple. Les hypothèses seront modifiées une à une, sans impacter les autres.

5.3.1 Suppression du risque de remplacement

Lors de la construction de la valeur contrat, le risque qu'un remplacement pour changement de qualité survienne au cours de la vie des contrats a été pris en compte. L'impact de la prise en compte de cette hypothèse peut être observé en recalculant la valeur contrat des affaires nouvelles sans prendre en compte les probabilités de remplacement.

Il est possible d'observer la marge globale prédite par segment dans le cas où le risque de remplacement est pris en compte et dans le cas où il est négligé :

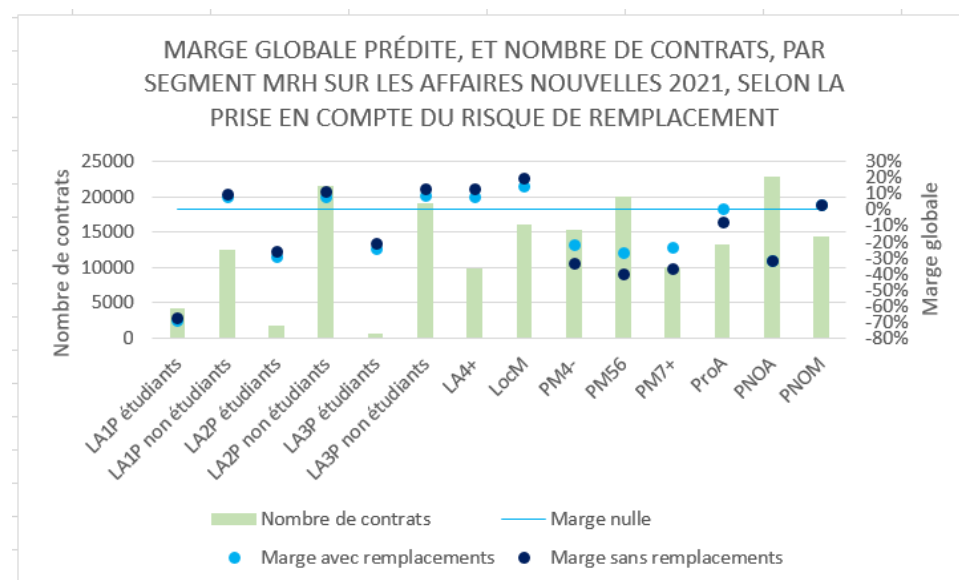


FIGURE 5.8 – Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, selon la prise en compte du risque de remplacement

Finalement prendre en compte le risque de remplacement est plus prudent pour la prédiction de la valeur contrat des contrats de locataires, un risque de sur-estimation de la marge de 3,38 points en moyenne est évité. Pour les contrats de propriétaires la marge estimée est plus élevée avec la prise en compte du risque de remplacement, de 11,53 points en moyenne, cela peut sembler moins prudent mais finalement c'est une vision plus représentative de la réalité.

5.3.2 Modification du taux d'évolution de la prime pure

D'après l'étude de l'évolution de la prime pure du portefeuille MRH ces dernières années, un taux d'évolution de la prime pure des contrats a été fixé à 0,62% pour la construction de la valeur contrat. En réalité ce taux est un taux d'évolution moyen qui continuera certainement d'évoluer avec la prise en compte des années suivantes. Il peut donc être envisagé d'observer l'impact d'une modification de ce taux fixe pour la construction de la valeur contrat. La prédiction de la valeur contrat des affaires nouvelles de 2021 a donc été recalculée dans le cas où le taux d'évolution de la prime pure valait 1,62% et dans le cas où il valait -0,38%, au lieu de 0,62%.

La marge globale prédite par segment dans ces deux cas a été calculée :

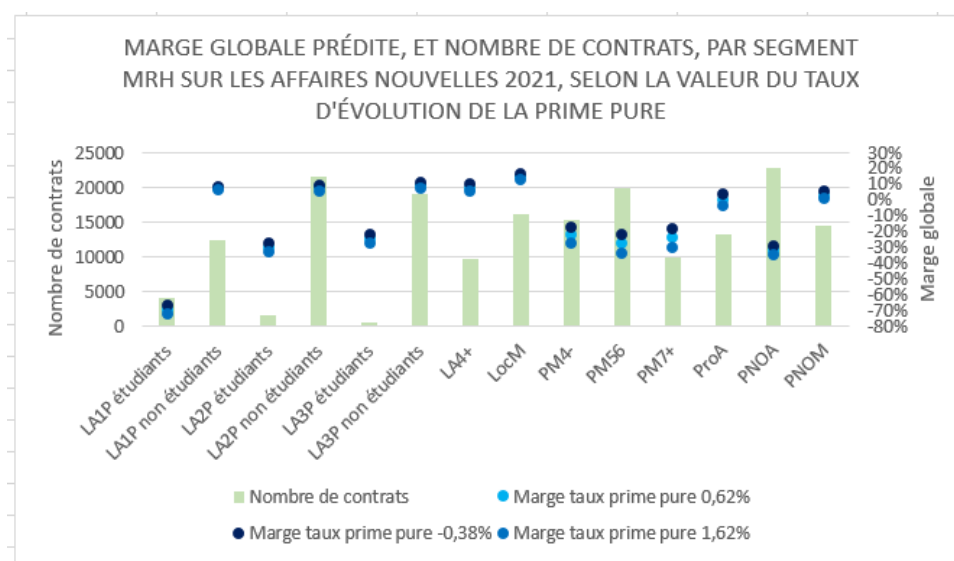


FIGURE 5.9 – Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, selon différentes valeurs du taux d'évolution de la prime pure

Sur l'ensemble des segments, il apparaît que diminuer le taux d'évolution de la prime pure de 1 point a tendance à augmenter la marge globale prédite de 4,17 points. Et, à l'inverse, augmenter le taux d'évolution de la prime pure de 1 point a tendance à diminuer la marge globale prédite de 4,70 points. Par principe de prudence il serait donc important de mettre à jour le taux d'évolution de la prime pure si celui-ci augmentait.

5.3.3 Modification des probabilités de résiliation

Pour construire la prédiction de l'indicateur de valeur contrat à la souscription, un modèle de prédiction du risque de résiliation des contrats a été développé pour obtenir les probabilités de résiliation des contrats au cours de leur vie. Pour rappel le modèle a été développé en observant les tendances de résiliation des contrats souscrits entre 2006 et 2020. Il est possible de se demander si les informations passées utilisées seront réellement représentatives du futur, ou si les tendances de résiliation des contrats évolueront. Il est difficile de prévoir quel type d'évolution peut être attendu. En effet en observant par exemple, à l'aide d'une modélisation de Kaplan-Meier, les tendances de résiliation des contrats selon leur année de souscription, sur les contrats utilisés pour la construction du modèle, il apparaît que les courbes de survie prédites d'une année à l'autre sont relativement proches et qu'il n'y a pas d'évolution continue :

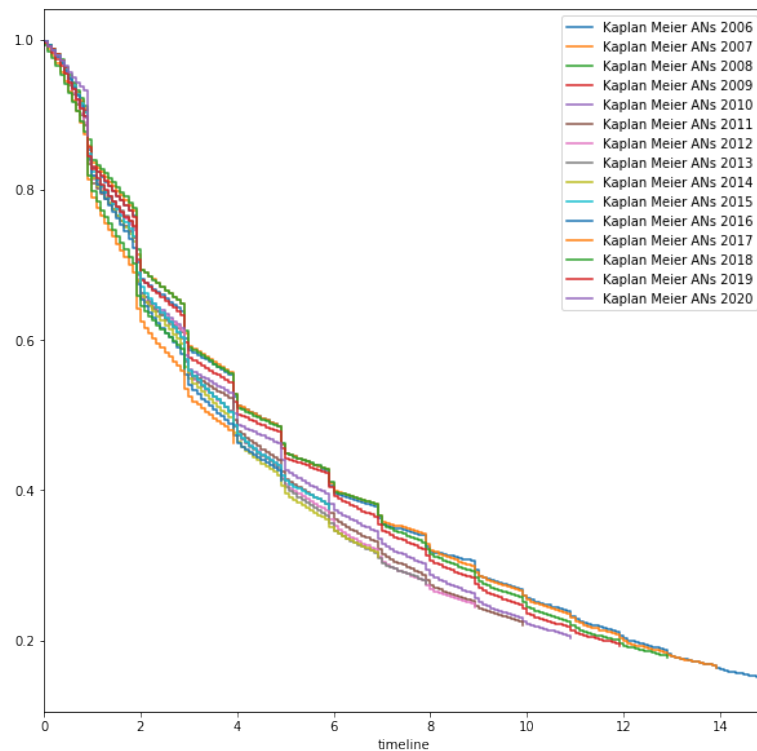


FIGURE 5.10 – Étude par Kaplan-Meier des éventuelles différences de tendance de résiliation des contrats selon leur année de souscription, entre 2006 et 2020

Pour se représenter deux cas d'évolution possibles des tendances de résiliation à l'avenir, deux scénarios peuvent être envisagés, et le calcul de la valeur contrat prédite des affaires nouvelles de 2021 peut être refait dans ces deux scénarios. Le premier scénario consistait à envisager le cas où la résiliation des contrats était plus forte à l'avenir, dans ce cas les probabilités de résiliation prédites à chaque âge ont été augmentées de 10%. Le second scénario consistait à envisager le cas où la résiliation des contrats diminuait à l'avenir, dans ce cas les probabilités de résiliation prédites à chaque âge ont été diminuées de 10%.

La marge globale prédite par segment sur les affaires nouvelles de 2021 dans ces deux scénarios a été observée :

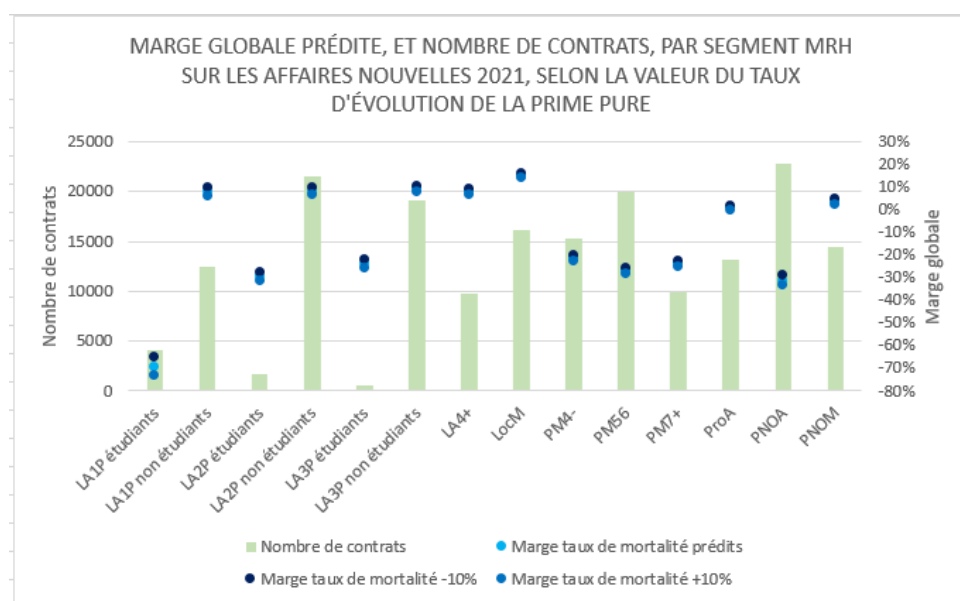


FIGURE 5.11 – Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, selon différents scénarios d'évolution des probabilités de résiliation

Une augmentation ou une diminution de 10% des taux de résiliation auraient un léger impact sur la valeur contrat. En effet sur l'ensemble des contrats, une diminution des taux de résiliation de 10% entraînerait une augmentation de la marge globale prédite de 1,50 points. Et une augmentation des taux de résiliation de 10% entraînerait une diminution de la marge globale prédite de 1,41 points. Le segment le plus affecté serait celui des locataires d'appartements une pièce étudiants, qui subirait une augmentation ou diminution de la marge globale prédite de 3,9 points.

5.3.4 Définition d'une contrainte de résiliation des contrats à un âge maximal de 15 ans

Selon le même raisonnement que celui utilisé dans la section précédente, il est possible d'envisager le fait que les nouveaux contrats d'assurance MRH d'AXA ne restent pas en portefeuille jusqu'à atteindre des âges très élevés comme ça a pu être le cas pour les générations précédentes. En effet, avec, par exemple, l'avancée constante de la digitalisation du marché de l'assurance il est possible que l'acte de résiliation soit plus facilement accessible aux assurés à l'avenir. Il serait possible d'envisager un scénario selon lequel, à l'avenir, les contrats d'assurance MRH ne dépasseraient pas un certain âge sans être résiliés, par exemple 15 ans. L'impact de cette potentielle hypothèse sur la prédiction de la valeur contrat à la souscription peut être étudié.

La marge globale prédite par segment des affaires nouvelles de 2021, avec l'hypothèse d'âge maximal des contrats de 15 ans, a été observée :

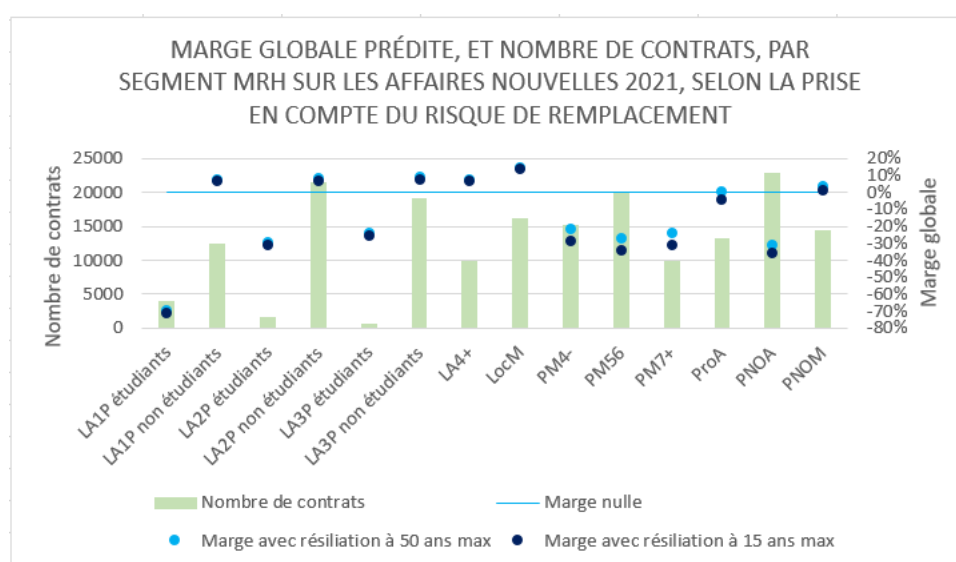


FIGURE 5.12 – Marge globale prédite par segment MRH sur les affaires nouvelles de 2021, avec l'hypothèse d'un âge maximal des contrats de 15 ans

Il apparaît que cette hypothèse a en particulier un impact sur les contrats de propriétaires. En effet sur les générations précédentes, il a été observé que ces contrats avaient tendance à rester en portefeuille plus longtemps que les autres, et donc leur stratégie de tarification prend cela en compte. D'après les résultats obtenus, si la situation évoluait, la stratégie actuellement en vigueur serait moins adaptée. En effet dans le cas où tous les nouveaux contrats résiliaient au maximum à l'âge de 15 ans, la marge globale prédite sur les contrats de propriétaires diminuerait de 5,6 points, alors que sur les contrats de locataires elle ne diminuerait que de 1,4 points.

5.4 Possibles évolutions de la prédiction de l'indicateur de valeur contrat

Il a été vu précédemment que la construction de l'indicateur de valeur contrat à la souscription, mise en oeuvre, repose sur un certain nombre d'hypothèses.

La principale hypothèse est celle selon laquelle les tendances de résiliation passées reflètent les tendances de résiliation futures. Dans le cas où cette hypothèse ne serait pas vérifiée à l'avenir, il a été vu lors des tests de sensibilité précédents, que la prédiction de la valeur contrat serait sensible à une évolution des taux de résiliation, ou à une limitation de l'âge des contrats, donc conserver cette hypothèse ne serait pas adapté. À court terme il serait difficile d'observer une évolution marquée des tendances de résiliation, mais dans quelques années les tendances de résiliations observées sur les dernières années pourraient être comparées aux tendances qui avaient été observées entre 2006 et 2020. Il serait alors possible de voir s'il est sensé de garder la prédiction actuelle du risque de résiliation, ou si celle-ci devrait être adaptée aux nouvelles tendances de résiliation observées.

Au niveau de la politique de tarification, un taux de majoration des contrats au terme fixe a été utilisé pour la construction de la valeur contrat, ce taux est le taux moyen actuellement en vigueur. À l'avenir une autre valeur du taux de majoration au terme pourrait être choisie, dans ce cas il faudrait mettre à jour la construction de la prédiction de la valeur contrat avec ce nouveau taux. Cependant une seule modification de ce taux ne pourrait suffire. En effet lorsqu'une évolution du taux de majoration au terme a lieu, il y a un risque que cela ait un impact sur la résiliation des contrats au moment de leur terme. Par exemple une augmentation du taux de majoration au terme pourrait dégrader le niveau de satisfaction des clients et augmenter le risque de résiliation. Il serait donc nécessaire, dès que cela serait possible, d'observer l'impact de l'évolution du taux de majoration sur la résiliation des contrats, et de prendre en compte ce potentiel impact lors de la prédiction des probabilités de résiliation des contrats, pour la prédiction de la valeur contrat.

Au niveau des coûts engendrés par les contrats, un taux d'évolution de la prime pure au fil des années a été estimé en calculant le taux d'évolution de la prime pure du portefeuille MRH observé ces dernières années. Ce taux pourrait être réévalué et mis à jour dans le calcul de la prédiction de la valeur contrat, en prenant en compte l'évolution de la prime pure observée sur les nouvelles années pouvant être prises en compte. Aussi, si les montants et taux de frais, de commissions et de coût de réassurance, étaient amenés à changer il serait nécessaire de mettre à jour ces informations dans le code de construction de la prédiction de la valeur contrat.

En ce qui concerne la prédiction de l'évolution du risque assuré, et donc des primes commerciales et primes pures associées, seule la prédiction des remplacements pour changement de qualité et de leur impact a été prise en compte. Mais cette prédiction d'évolution pourrait être enrichie par la prédiction de la survenance d'autres types de remplacements et de leur impact. En effet il avait été observé que les remplacements pour changement de qualité ne représentaient en fait que 4,58% des remplacements. Mais il a été vu que la prise en compte du risque de ce type de remplacements dans la prédiction de la valeur contrat était pertinente. C'est pourquoi il pourrait être intéressant, à l'avenir, d'essayer d'intégrer une prédiction du risque de survenance d'autres types de remplacement et de leur impact, dans la prédiction de la valeur contrat à la souscription.

La prédiction de la valeur contrat à la souscription dépend donc fortement d'hypothèses de construction qui pourront être enrichies, ou être mises à jours si des évolutions ont lieu.

5.5 Synthèse du chapitre

L'objectif de ce dernier chapitre était de finaliser la construction d'une prédiction de la valeur contrat à la souscription des contrats d'assurance MRH.

Pour cela des hypothèses ont été définies pour l'estimation des flux probables de résultat futurs. Tous les paramètres nécessaires à l'estimation des primes commerciales, des primes pures, des frais, des commissions, des coûts de réassurance ont été présentés.

L'utilisation des modèles de prédiction du risque de résiliation et de remplacement pour changement de qualité des contrats pour la prédiction de la valeur contrat a été explicitée.

Une fois le calcul de la prédiction de l'indicateur de valeur contrat élaboré, celui-ci a été appliqué aux affaires nouvelles de 2021. Les résultats obtenus ont été observés. Ils permettent notamment de classer les contrats selon leur rentabilité prédite. Ainsi, il est apparu que les contrats dont la marge prédite était la plus faible étaient majoritairement des contrats de propriétaires de maisons. Les contrats présentant une marge attendue élevée étaient majoritairement des contrats de locataires de maisons. Ils permettent aussi de déterminer quels montants de prime commerciale à l'affaire nouvelle, les contrats devraient avoir pour être rentables, selon leur tendance de résiliation prédite, leur tendance de remplacement prédite, et la politique d'évolution tarifaire en vigueur.

La sensibilité de la prédiction de la valeur contrat à l'évolution de certaines hypothèses a été étudiée. Notamment il a été vu que ne pas prendre en compte le risque de remplacement pour changement de qualité aurait eu un impact sur les résultats obtenus pour les affaires nouvelles de 2021. En particulier cela aurait été moins prudent pour la prédiction de la valeur contrat des contrats de locataires. Aussi, il a été vu qu'une modification du taux d'évolution de la prime pure aurait eu un certain impact sur les résultats, ainsi qu'une évolution des taux de résiliation, ou une limitation de la vieillesse des contrats.

Les tests de sensibilité ont permis de voir qu'il serait utile de mettre à jour les hypothèses utilisées dans la prédiction de la valeur contrat si des évolutions de la tarification avaient lieu, ou si des tendances de résiliation étaient observées.

Conclusion

L'objectif de ce mémoire est de construire un indicateur de valeur contrat à la souscription des contrats d'assurance multirisque habitation afin d'avoir une vision à long terme de la performance de la stratégie tarifaire d'AXA France.

Pour construire cet indicateur il est nécessaire de disposer d'une prédiction des tendances de résiliation des contrats au fil de leur vieillissement. Une première étude de la survie des contrats, selon le segment MRH des contrats, permet de se rendre compte que les contrats de locataires ont tendance à être résiliés plus rapidement que les contrats de propriétaires. En effet, d'après des modèles de durée de Kaplan-Meier construits sur 50 ans sur l'ensemble des contrats, observés entre 2016 et 2020, de chaque segment MRH, vus en 2016, les espérances de vie moyennes des contrats selon leur segment sont les suivantes :

Catégorie de contrats étudiée	Espérance de vie à la souscription
Ensemble des contrats	7 ans et 2 mois
Contrats LA1P	3 ans et 9 mois
Contrats LA2P	4 ans et 8 mois
Contrats LA3P	5 ans et 2 mois
Contrats LA4+	5 ans et 7 mois
Contrats LocM	5 ans
Contrats PM4-	10 ans
Contrats PM56	10 ans et 8 mois
Contrats PM7+	12 ans et 4 mois
Contrats ProA	10 ans et 3 mois
Contrats PNOA	10 ans et 5 mois
Contrats PNOM	9 ans et 3 mois

TABLE - Estimation de l'espérance de vie à la souscription, au global, et selon les segments MRH

Malgré le fait que, dans la construction de ces modèles, le segment MRH observé des contrats ne soit pas nécessairement le segment des contrats à leur souscription, mais celui observé en 2016, il est visible que le segment MRH des contrats a un impact important sur leur survie.

D'autres caractéristiques pouvant également avoir un impact, un modèle de durée semi-paramétrique, modèle de Cox, a été développé, sur les contrats souscrits entre 2006 et 2020, pour prédire les probabilités de résiliation des contrats, à chaque âge. Ce modèle prend en compte les caractéristiques des contrats, ce sont alors leurs caractéristiques à la souscription qui ont été utilisées, pour disposer d'une prédiction la plus exacte possible de la survie des contrats à la souscription. Il n'a été possible de construire ce modèle que sur 15 ans car les informations des contrats à la souscription n'étaient plus disponibles avant 2006. Cependant en assurance habitation des contrats ont tendance à vivre plus longtemps que 15 ans, en particulier des contrats de propriétaires. Il a donc été nécessaire de prolonger les courbes de survie prédites par ce modèle au delà de 15 ans. Pour cela, les courbes de survie qui avaient été prédites par Kaplan-Meier sur 50 ans selon le segment MRH observé en 2016, ont été utilisées. D'après ce modèle de Cox, construit à l'aide de

l'algorithme du *Gradient Boosting*, les caractéristiques ayant le plus d'impact sur la tendance de résiliation d'un contrat sont :

- la qualité de l'assuré (être un propriétaire diminue le risque de résiliation) ;
- l'âge de l'assuré (être âgé diminue le risque) ;
- le nombre de contrats que l'assuré possède déjà au sein d'AXA (avoir beaucoup de contrats augmente le risque) ;
- l'ancienneté de l'assuré au sein d'AXA (avoir une grande ancienneté diminue le risque) ;
- si le contrat est un contrat étudiant (être étudiant augmente le risque) ;
- le nombre de pièces du logement (un faible nombre de pièces augmente le risque).

Cet indicateur de valeur contrat prend aussi en compte la probabilité individualisée que les contrats vivent un remplacement pour changement de qualité au cours de leur vie, et l'impact que cela aurait sur leur rentabilité. Une prédiction des tendances de remplacement pour changement de qualité des contrats au fil de leur vieillissement est donc utilisée. Plusieurs modèles de Cox ont été développés, sur les contrats souscrits entre 2006 et 2020, avec utilisation du *Gradient Boosting*, pour prédire les probabilités de remplacement des contrats sur 15 ans, à chaque âge, selon leurs caractéristiques à la souscription. Au-delà de 15 ans, le risque de remplacement a été négligé. Un modèle a été construit pour les contrats étudiants. D'après lui la caractéristique ayant le plus d'impact sur la tendance de remplacement d'un contrat étudiant est le nombre de pièces du logement. Plus le logement comporte de pièces, plus le risque instantané de remplacement du contrat pour changement de qualité est élevé, toutes choses égales par ailleurs. Un modèle a été construit pour les contrats de locataires non-étudiants, et un autre pour les contrats de propriétaires. D'après eux la caractéristique ayant le plus d'impact sur la tendance de remplacement d'un contrat de locataire non-étudiant ou d'un contrat de propriétaire est l'âge de l'assuré à la souscription. Plus le client est jeune, plus le risque instantané de remplacement du contrat est élevé, toutes choses égales par ailleurs. Pour les contrats de propriétaires non-occupants le risque de remplacement pour changement de qualité est négligé car il est très faible.

La prédiction de l'indicateur de valeur contrat à la souscription est construite en prenant en compte les prédictions des probabilités de résiliation et de remplacement pour changement de qualité des contrats, obtenues selon la méthode énoncée précédemment. Elle est aussi construite en supposant que les conditions actuelles d'évolution des primes commerciales et des primes pures, sont figées dans le temps, ainsi que la définition actuelle des frais, des commissions et du coût de la réassurance. La construction de cette prédiction dépend donc de beaucoup d'hypothèses. Il est alors nécessaire de prendre du recul vis-à-vis des montants de valeurs contrat prédits. Il serait difficile d'affirmer que ces montants seront précisément proches des vraies valeurs. Cependant, il est possible d'avoir confiance en l'ordre des contrats donné selon leur valeur contrat prédite, qui, d'après des tests de sensibilité, serait peu impacté par d'éventuelles évolutions des hypothèses.

La construction de cet indicateur peut être utilisée pour prédire la valeur contrat à la souscription des affaires nouvelles observées entre janvier et juin 2021. L'analyse des résultats obtenus indique que les contrats les plus rentables seraient les contrats de locataires de maison, et de locataires d'appartements non-étudiants, et les contrats les moins rentables seraient les contrats de propriétaires de maisons. A présent, cet indicateur pourra être calculé sur toutes les affaires nouvelles futures et permettra ainsi d'estimer et de suivre la rentabilité à moyen-long terme du portefeuille.

Il est possible d'aller au-delà de l'analyse descriptive des prédictions de valeurs contrats obtenus, et d'utiliser ces prédictions pour notamment optimiser la stratégie de souscription d'AXA France. En effet l'exploitation de l'indicateur de valeur contrat qui a été construit permet de définir des profils de contrats à souscrire en priorité pour améliorer la rentabilité future du portefeuille MRH d'AXA. Les conclusions obtenues pourraient guider les campagnes marketing et publicitaires futures. Aussi l'exploitation de cet indicateur pourrait guider les politiques d'évolution des tarifs à l'affaire nouvelle. En effet, il apparaît, notamment, que pour améliorer la marge globale sur l'ensemble des contrats du portefeuille MRH, les primes commerciales à l'affaire nouvelle des contrats des propriétaires de maisons, des locataires d'appartements étudiants, et des propriétaires non-occupants d'appartements pourraient être augmentées en priorité.

L'indicateur de valeur contrat qui a été construit est adapté au contexte de tarification actuel, et aux

tendances de résiliation et de remplacement actuelles. Mais le contexte de tarification, ou les tendances de résiliation ou de remplacement pourraient évoluer avec le temps. Alors pour pouvoir continuer d'utiliser l'indicateur de valeur contrat, les hypothèses prises dans sa construction devraient être mises à jour. Par exemple, si à l'avenir la politique de majoration au terme évoluait, le nouveau taux en vigueur pourrait remplacer le taux utilisé préalablement dans la construction de la valeur contrat. Aussi, si les tendances de résiliation évoluaient, le modèle de prédiction du risque de résiliation, utilisé dans la construction de la valeur contrat, pourrait être mis à jour en prenant en compte les tendances de résiliation observées sur les nouvelles années disponibles depuis 2020. L'indicateur de valeur contrat est donc modulable pour s'adapter au mieux au contexte dans lequel les contrats sur lequel il s'applique sont souscrits.

Les hypothèses de construction de la valeur contrat pourront être mises à jour, si nécessaire, mais elles pourraient aussi être enrichies. En effet, cette étude s'est limitée à la prise en compte du risque de remplacement pour changement de qualité au cours de la vie des contrats. Mais la construction de l'indicateur de valeur contrat pourrait être enrichie par la prédiction de la survenance d'autres types de remplacements et de leur impact.

Bibliographie

- [1] ADICEOM C. (2019) *Généralisation de l'optimisation de la majoration de la prime au terme des contrats d'assurance habitation*. Mémoire d'actuariat, ESSEC.
- [2] ATKINSON E. et THERNEAU T. (2021) *Concordance*. Article.
- [3] BALTESAR B. (2013) *Construction d'une table de mortalité sur un portefeuille de temporaires décès*. Mémoire d'actuariat, ISFA.
- [4] BERTRAND F. et MAUMY-BERTRAND M. (2018) *Initiation à la statistique avec R*. 3e édition. DUNOD. Ouvrage.
- [5] COUSIN A. *Modèle de survie en actuariat*. Cours, Université de Strasbourg.
- [6] COX D. (1972) *Journal of the Royal Statistical Society : Series B (Methodological)*. "Regression Models and Life-Tables". 187-202. Article.
- [7] HERLING A.S. (2019) *Évaluation de la qualité de la souscription en assurance automobile grâce à la construction d'un indicateur de valeur*. Mémoire d'actuariat, DUAS.
- [8] KOYE G.K. (2019) *Comparaison des méthodes classiques et alternatives avec le machine learning pour la construction d'une table de mortalité d'expérience best estimate*. Mémoire d'actuariat, ENSAE.
- [9] LASGOUTTES J.M. (2020) *Introduction au boosting*. Cours, INSA.
- [10] MONBET V. (2013) *Statistique inférentielle : tests*. Cours, Université de Rennes.

Webographie

- [11] Fédération Française de l'assurance. <https://www.ffa-assurance.fr/etudes-et-chiffres-cles/assurances-de-biens-et-de-responsabilite-donnees-cles-par-annee> (étude du marché de l'assurance IARD), site consulté le 30 janvier 2021
- [12] INSEE. <https://www.insee.fr/fr/statistiques/1379714> : :text=Au%201er%20janvier%202014%2C%20la%20France%20m%C3%A9ropolitaine%20compte%2033,dont%2083%20%25%20de%20r%C3%A9sidences%20principales (le nombre de logements en France au 1er janvier 2014), site consulté le 20 mai 2021
- [13] INSEE. <https://www.insee.fr/fr/statistiques/4985385> : :text=qu'en%20m%C3%A9tropole-,Le%20nombre%20de%20r%C3%A9sidences%20principales%20augmente%20moins%20vite%20sur%20la,de%20logements%20suppl%C3%A9mentaires%20depuis%201982 (le nombre de logements en France au 1er janvier 2020), site consulté le 20 mai 2021
- [14] L'Argus de l'assurance. <https://www.argusdelassurance.com/classements/classement-auto-et-mrh-2020.165481> (classement des assureurs en assurance auto et MRH), site consulté le 23 juillet 2021
- [15] Spac Actuaire. <https://www.spac-actuaire.fr/lexique/table-de-mortalite/> (définition de notions de mortalité), site consulté le 28 décembre 2020

Annexes

Plan des annexes

Annexe 1 : Traitement des valeurs manquantes : Taux de valeurs manquantes pour chaque variable de la base

Annexe 2 : Construction d'un modèle sur les âges entiers : Comparaison entre les courbes de survie moyennes prédites par le modèle de Cox, et celles prédites par Kaplan-Meier

Annexe 3 : Construction d'un modèle sur les âges exacts : Comparaison entre les courbes de survie moyennes prédites par les modèles de Cox, et celles prédites par Kaplan-Meier

Annexe 4 : Construction d'un modèle sur les âges exacts : Espérance de vie à la souscription estimée avec les prédictions moyennes de chaque modèle de Cox, et des estimations de Kaplan-Meier, par segment

Annexe 5 : Construction d'un modèle sur les âges exacts : Erreur de prédiction moyenne de l'espérance de vie à la souscription, des contrats non-censurés

Annexe 6 : Construction d'un modèle sur les âges exacts : Impact des valeurs des variables explicatives d'après le modèle de Cox classique retenu

Annexe 7 : Prolongation des courbes de survies estimées par le modèle : Comparaison des courbes de survie prolongées moyennes aux courbes de survie construites \hat{S}_{fin}

Annexe 8 : Construction d'un modèle sur les âges exacts : Espérance de vie à la souscription estimée avec les prédictions moyennes du modèle de Cox *Gradient Boosting*, et des estimations de Kaplan-Meier, par segment

Annexe 9 : Construction d'un modèle sur les âges exacts : Erreur de prédiction moyenne de l'espérance de vie à la souscription, des contrats non-censurés

Annexe 10 : Prolongation des courbes de survies estimées par le modèle : Comparaison des courbes de survie prolongées moyennes aux courbes de survie construites \hat{S}_{fin}

Annexe 11 : Construction d'un modèle sur les âges exacts : Espérance de vie à la souscription estimée avec les prédictions moyennes du modèle de Cox *Gradient Boosting* initial et du modèle segmenté, et des estimations de Kaplan-Meier, par segment

Annexe 12 : Étude des valeurs contrat obtenues sur les affaires nouvelles de 2021 : Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles de 2021

Annexe 1 : Traitement des valeurs manquantes : Taux de valeurs manquantes pour chaque variable de la base

Variable	Part de valeurs manquantes
Type de résidence	0,08%
Qualité de l'assuré	0,70%
Type de logement	0,60%
Étage du logement	0,00%
Taux d'objets de valeur	0,90%
Ancienneté du logement	0,80%
Nombre d'options	0,00%
Indicateur enfant d'assuré	0,00%
Région AXA	0,00%
Canal de distribution	0,00%
Fractionnement des primes	0,00%
Qualité d'occupation du bien	0,70%
Indicateur étudiant	0,00%
Indicateur personne morale	9,00%
Sexe de l'assuré	4,90%
Produit d'entrée chez AXA	0,60%
Nombre de pièces du logement	0,60%
Capital total déclaré	0,60%
Prime hors taxes	0,00%
Âge du client	4,90%
Ancienneté du client au sein d'AXA	1,90%
Type de franchise	0,00%
Surface des dépendances	10,00%
Année de souscription	0,00%
Catégorie socio-professionnelle du client	29,69%
Produit MRH	0,00%
Nombre de contrats AXA en cours	41,38%
Segment MRH	0,00%

TABLE - Proportion de valeurs manquantes par variable dans la nouvelle base de données

Annexe 2 : Construction d'un modèle sur les âges entiers : Comparaison entre les courbes de survie moyennes prédites par le modèle de Cox, et celles prédites par Kaplan-Meier

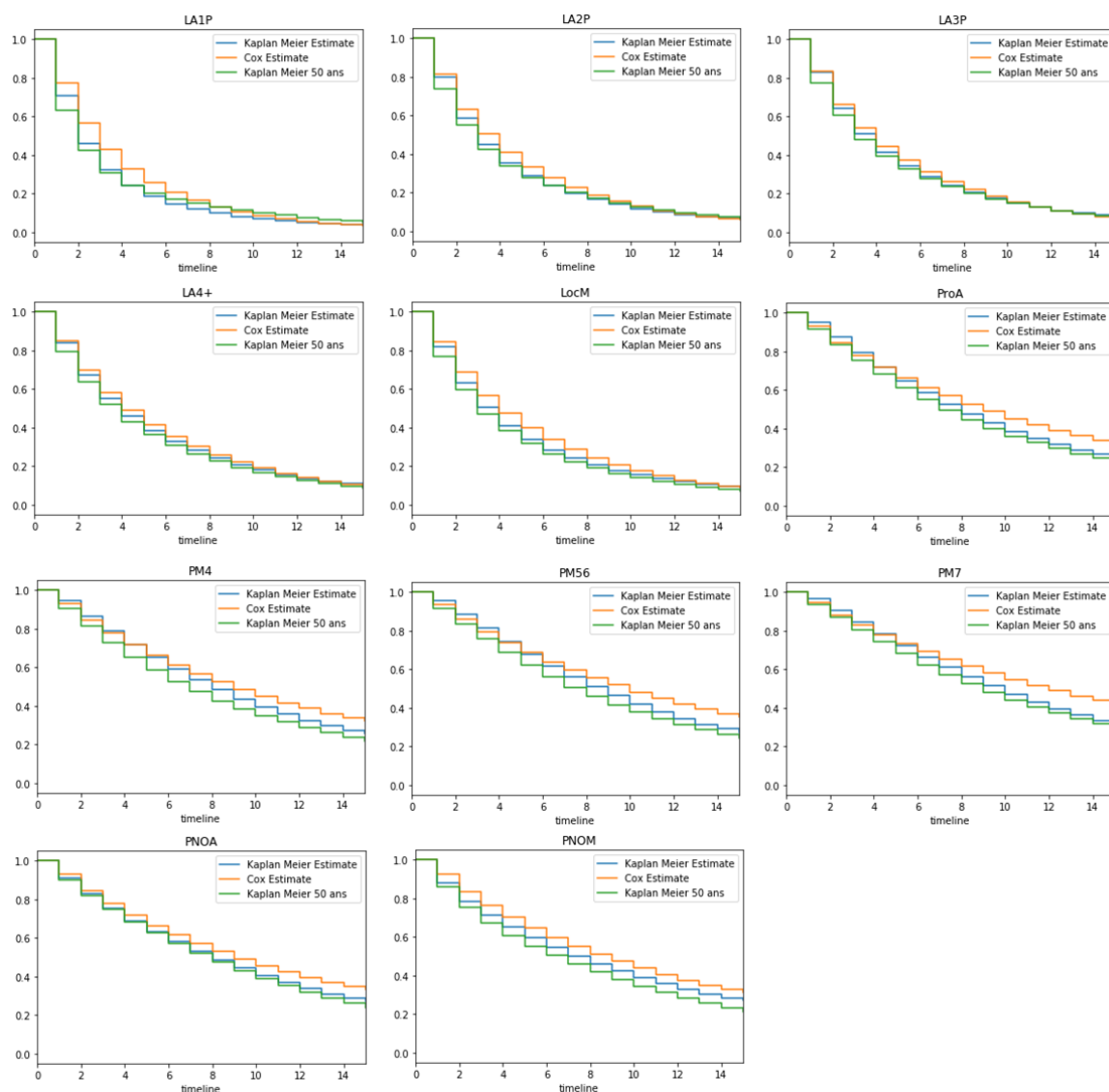


FIGURE - Comparaison entre la courbe de survie moyenne prédite par le modèle de Cox classé sur 15 ans, la courbe de survie de Kaplan-Meier estimée sur 15 ans, et la courbe de survie de Kaplan-Meier estimée sur la base Image sur 50 ans, sur l'ensemble des contrats de chaque segment MRH

Annexe 3 : Construction d'un modèle sur les âges exacts : Comparaison entre les courbes de survie moyennes prédites par les modèles de Cox, et celles prédites par Kaplan-Meier

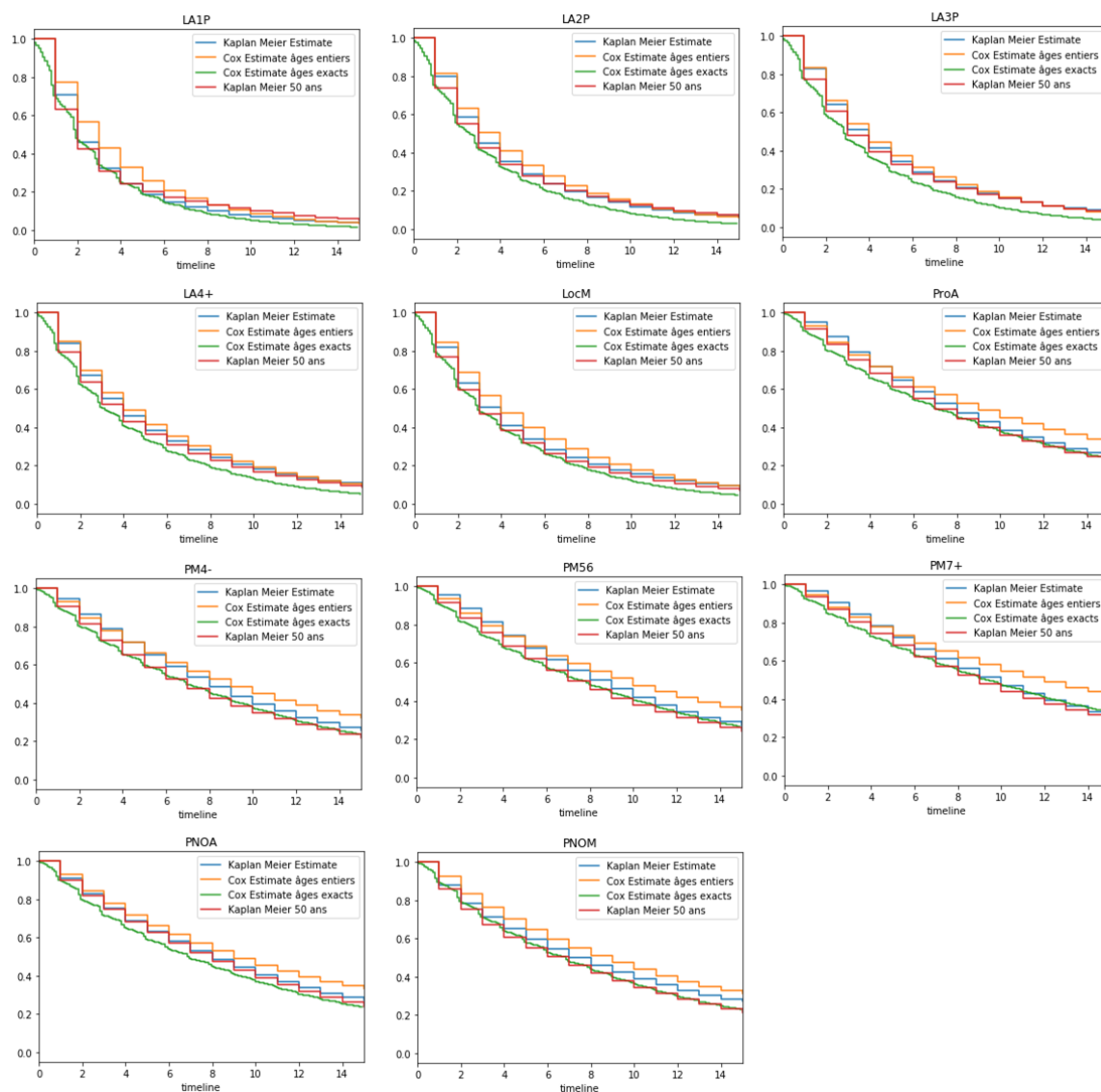


FIGURE - Comparaison entre la courbe de survie moyenne prédite par le modèle de Cox classique sur les âges exacts sur 15 ans, celle prédite par le modèle sur les âges entiers, la courbe de survie de Kaplan-Meier estimée sur 15 ans, et la courbe de survie de Kaplan-Meier estimée sur la base Image sur 50 ans, sur l'ensemble des contrats de chaque segment MRH

Annexe 4 : Construction d'un modèle sur les âges exacts : Espérance de vie à la souscription estimée avec les prédictions moyennes de chaque modèle de Cox, et des estimations de Kaplan-Meier, par segment

Segment MRH	Espérance de vie Modèle de Cox sur les âges entiers	Espérance de vie Modèle de Cox sur les âges exacts	Espérance de vie Kaplan-Meier sur les âges entiers
LA1P non étudiants	4,02	3,21	3,21
LA1P étudiants	2,46	1,80	1,97
LA2P non étudiants	4,27	3,40	3,87
LA2P étudiants	2,62	1,90	2,54
LA3P non étudiants	4,57	3,68	4,41
LA3P étudiants	2,76	2	2,37
LA4+	4,99	4,06	4,77
LocM	4,79	3,87	4,32
ProA	8,42	7,43	7,86
PM4-	8,4	7,39	7,92
PM56	8,79	7,8	8,26
PM7+	9,59	8,64	8,89
PNOA	8,46	7,33	7,84
PNOM	8,22	7,2	7,49
Erreur / à Kaplan Meier	8,7%	-7,9%	

TABLE - Estimations de l'espérance de vie à la souscription selon les segments MRH, selon les prédictions moyennes de la fonction de survie, du modèle de cox sur les âges entiers, de celui sur les âges exacts, et des estimations de Kaplan-Meier

Annexe 5 : Construction d'un modèle sur les âges exacts : Erreur de prédiction moyenne de l'espérance de vie à la souscription, des contrats non-censurés

Segment MRH	Nombre de contrats	Erreur moyenne de prédiction (en années)
LA1P non étudiants	85183	0,37
LA1P étudiants	75478	-0,02
LA2P non étudiants	123399	0,16
LA2P étudiants	17241	-0,21
LA3P non étudiants	108264	0,1
LA3P étudiants	6133	0,02
LA4+	54658	0,28
LocM	97661	0,39
ProA	42694	2,78
PM4-	50901	2,7
PM56	57236	3,1
PM7+	22696	3,99
PNOA	46191	3,49
PNOM	40238	3,06

TABLE - Erreur de prédiction moyenne par segment avec le modèle de Cox, de l'espérance de vie à la souscription des contrats non censurés, par rapport aux durées de vie réellement observée

Annexe 6 : Construction d'un modèle sur les âges exacts : Impact des valeurs des variables explicatives d'après le modèle de Cox classique retenu

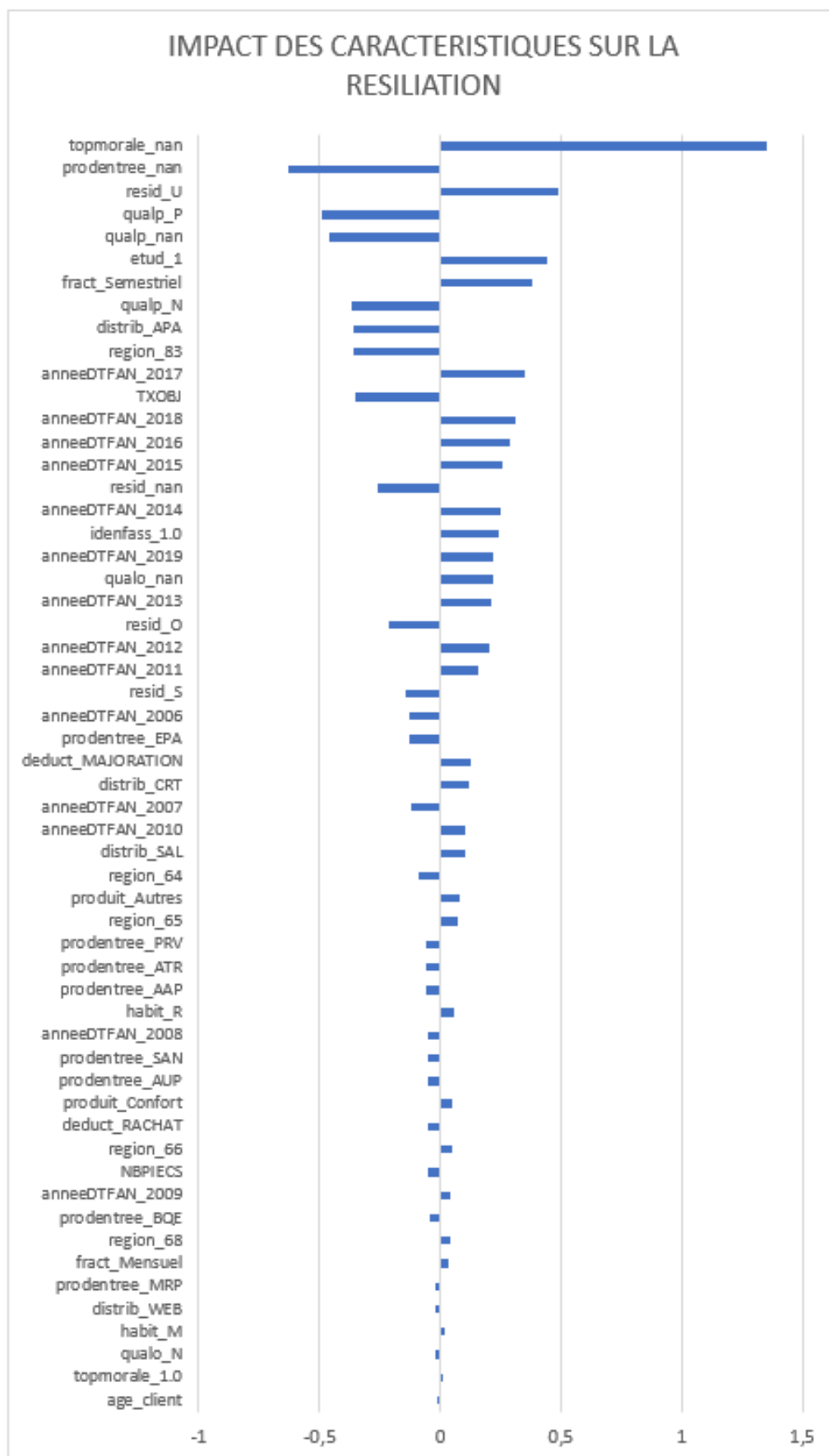


FIGURE - Représentation de l'impact de chaque caractéristique possible sur le risque de résiliation des contrats, d'après le modèle de Cox classique retenu

Annexe 7 : Prolongation des courbes de survies estimées par le modèle : Comparaison des courbes de survie prolongées moyennes aux courbes de survie construites \hat{S}_{fin}

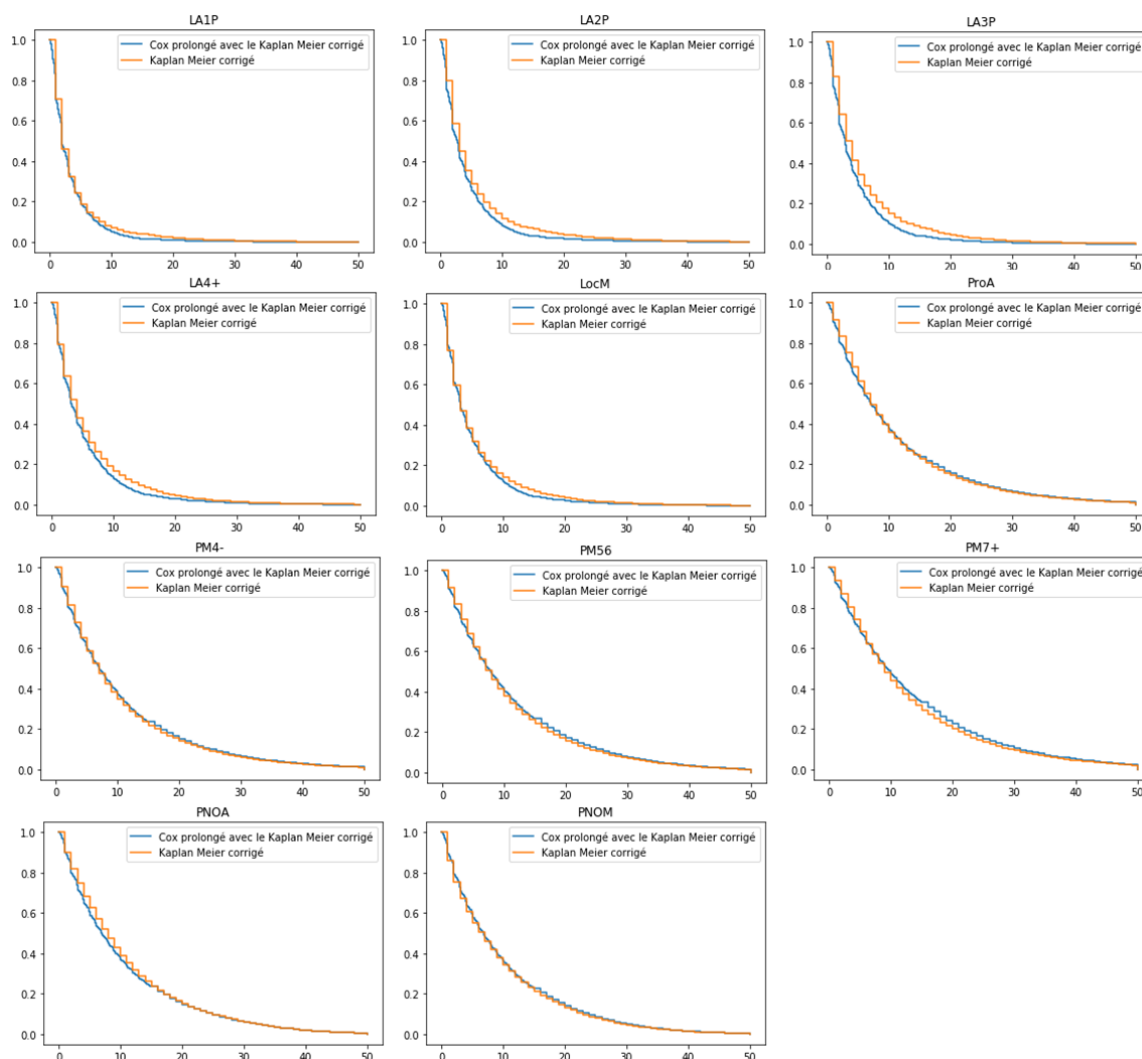


FIGURE - Comparaison, pour chaque segment MRH, entre la courbe de survie prédite, par le modèle de Cox classique, et prolongée moyenne, et la courbe de survie estimée \hat{S}_{fin}

Annexe 8 : Construction d'un modèle sur les âges exacts : Espérance de vie à la souscription estimée avec les prédictions moyennes du modèle de Cox Gradient Boosting, et des estimations de Kaplan-Meier, par segment

Segment MRH	Espérance de vie Modèle de Cox Gradient Boosting	Espérance de vie Kaplan-Meier sur les âges entiers
LA1P non étudiants	2,90	3,21
LA1P étudiants	1,88	1,97
LA2P non étudiants	3,66	3,87
LA2P étudiants	2,28	2,54
LA3P non étudiants	4,19	4,41
LA3P étudiants	2,1	2,37
LA4+	4,41	4,77
LocM	3,93	4,32
ProA	7,42	7,86
PM4-	7,46	7,92
PM56	7,75	8,26
PM7+	8,18	8,89
PNOA	7,16	7,84
PNOM	6,54	7,49
Erreur / à Kaplan Meier	-7,7%	

TABLE - Estimations de l'espérance de vie à la souscription selon les segments MRH, selon les prédictions moyennes de la fonction de survie, du modèle de cox *Gradient Boosting*, et des estimations de Kaplan-Meier

Annexe 9 : Construction d'un modèle sur les âges exacts : Erreur de prédiction moyenne de l'espérance de vie à la souscription, des contrats non-censurés

Segment MRH	Nombre de contrats	Erreur moyenne de prédiction (en années)
LA1P non étudiants	85183	0,09
LA1P étudiants	75478	-0,05
LA2P non étudiants	123399	0,28
LA2P étudiants	17241	-0,06
LA3P non étudiants	108264	0,43
LA3P étudiants	6133	0,02
LA4+	54658	0,48
LocM	97661	0,35
ProA	42694	2,26
PM4-	50901	2,32
PM56	57236	2,58
PM7+	22696	3
PNOA	46191	2,88
PNOM	40238	2,08

TABLE - Erreur de prédiction moyenne par segment avec le modèle de *Cox Gradient Boosting*, de l'espérance de vie à la souscription des contrats non censurés, par rapport aux durées de vie réellement observée

Annexe 10 : Prolongation des courbes de survies estimées par le modèle : Comparaison des courbes de survie prolongées moyennes aux courbes de survie construites \hat{S}_{fin}

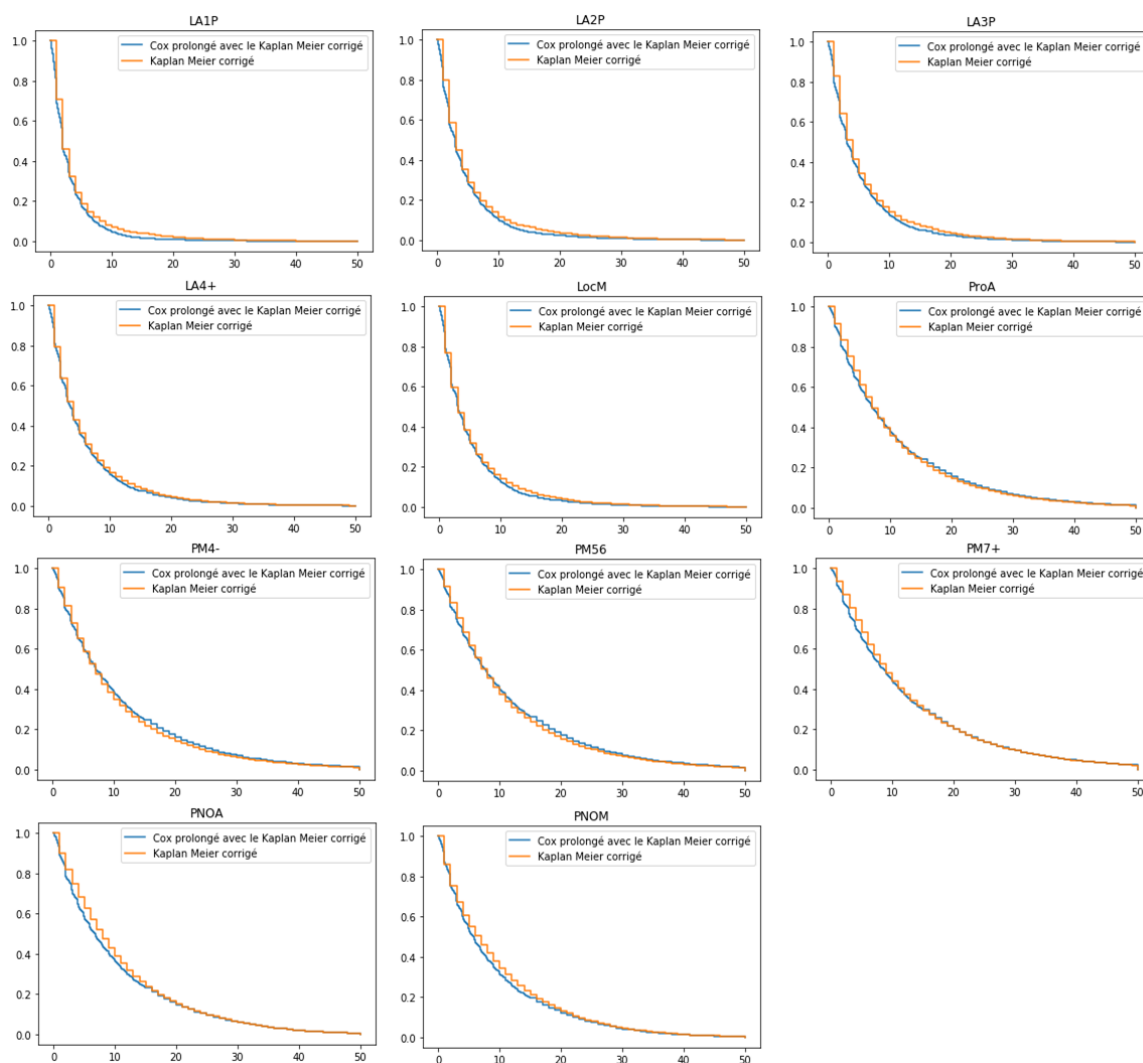


FIGURE - Comparaison, pour chaque segment MRH, entre la courbe de survie prédite, par le modèle de *Cox Gradient Boosting*, et prolongée moyenne, et la courbe de survie estimée \hat{S}_{fin}

Annexe 11 : Construction d'un modèle sur les âges exacts : Espérance de vie à la souscription estimée avec les prédictions moyennes du modèle de Cox Gradient Boosting initial et du modèle segmenté, et des estimations de Kaplan-Meier, par segment

Segment MRH	Espérance de vie Modèle de Cox Gradient Boosting initial	Espérance de vie Modèle de Cox Gradient Boosting segmenté	Espérance de vie Kaplan-Meier
LA1P non étudiants	14,48	14,38	14,42
LA1P étudiants	14,59	14,29	14,23
LA2P non étudiants	14,08	14,03	14,01
LA2P étudiants	14,17	13,80	13,71
LA3P non étudiants	14,02	13,98	14,02
LA3P étudiants	14,13	13,94	13,73
LA4+	14,05	14,02	14,08
LocM	13,24	13,22	13,26
ProA	14,35	14,45	14,51
PM4-	14,51	14,52	14,65
PM56	14,53	14,52	14,65
PM7+	14,58	14,54	14,69
PNOA	14,83	15	14,96
PNOM	14,79	15	14,93
Erreur globale / à Kaplan-Meier	0,24%	-0,09%	
Erreur Locataires	1,16%	0,17%	
Erreur Propriétaires et PNOs	-0,92%	-0,41%	

TABLE - Estimations de l'espérance de vie, avant remplacement, à la souscription selon les segments MRH, selon les prédictions moyennes de la fonction de survie, du modèle de cox *Gradient Boosting* initial et du modèle segmenté, et des estimations de Kaplan-Meier

Annexe 12 : Étude des valeurs contrat obtenues sur les affaires nouvelles de 2021 : Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles de 2021

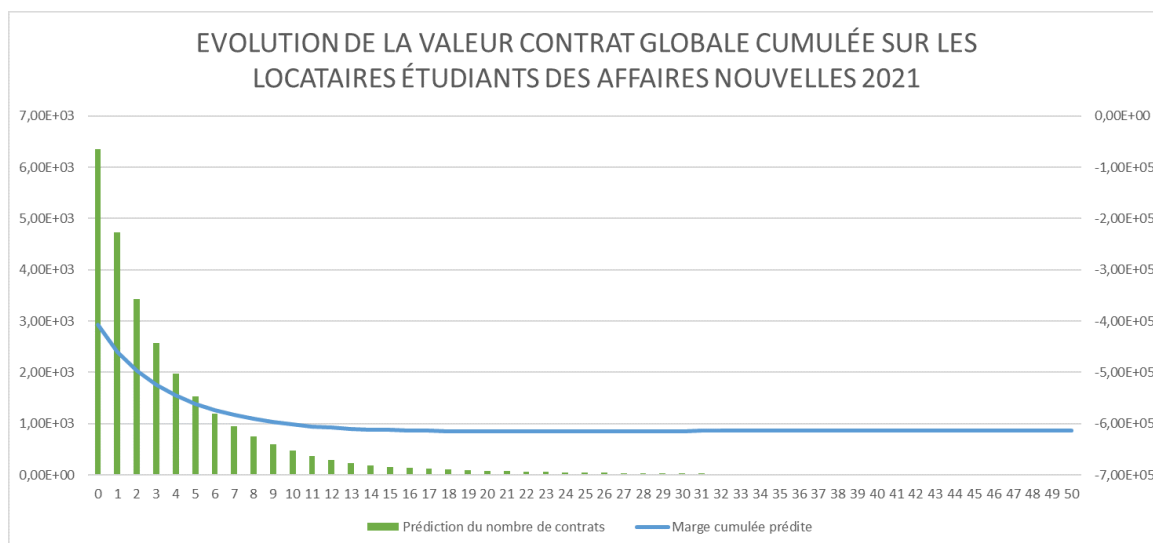


FIGURE - Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des locataires étudiants

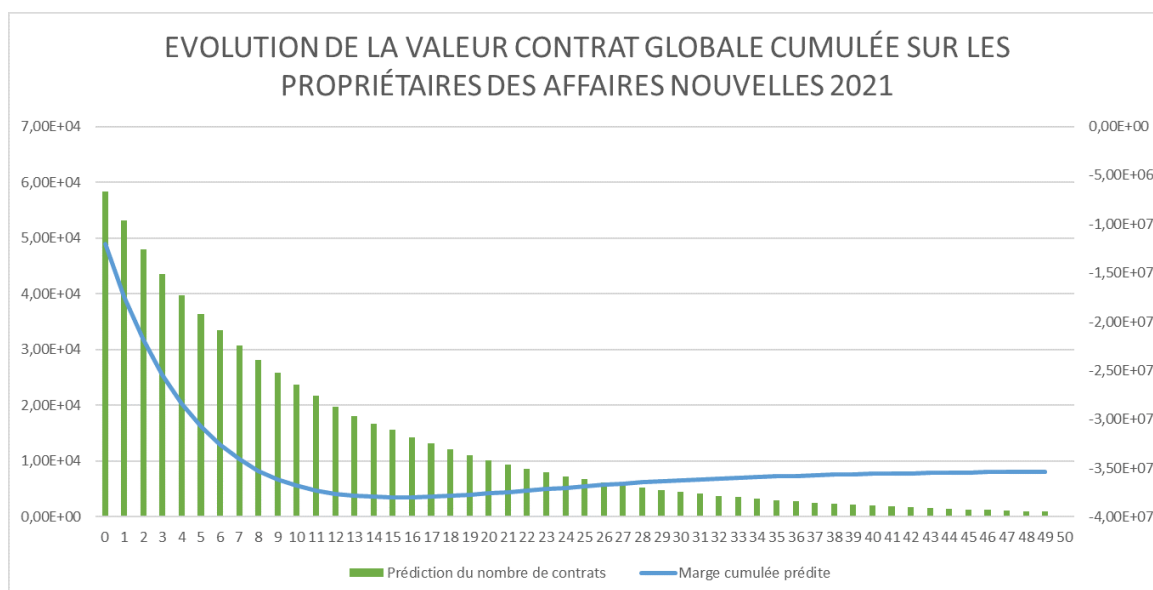


FIGURE - Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des propriétaires

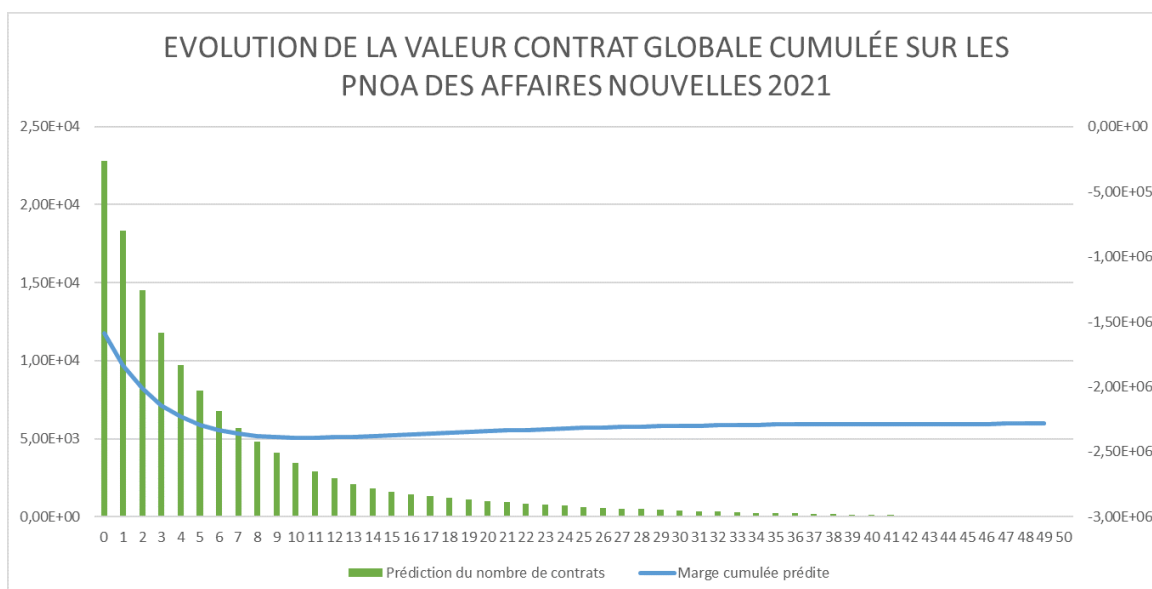


FIGURE - Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des propriétaires non-occupants d'appartements

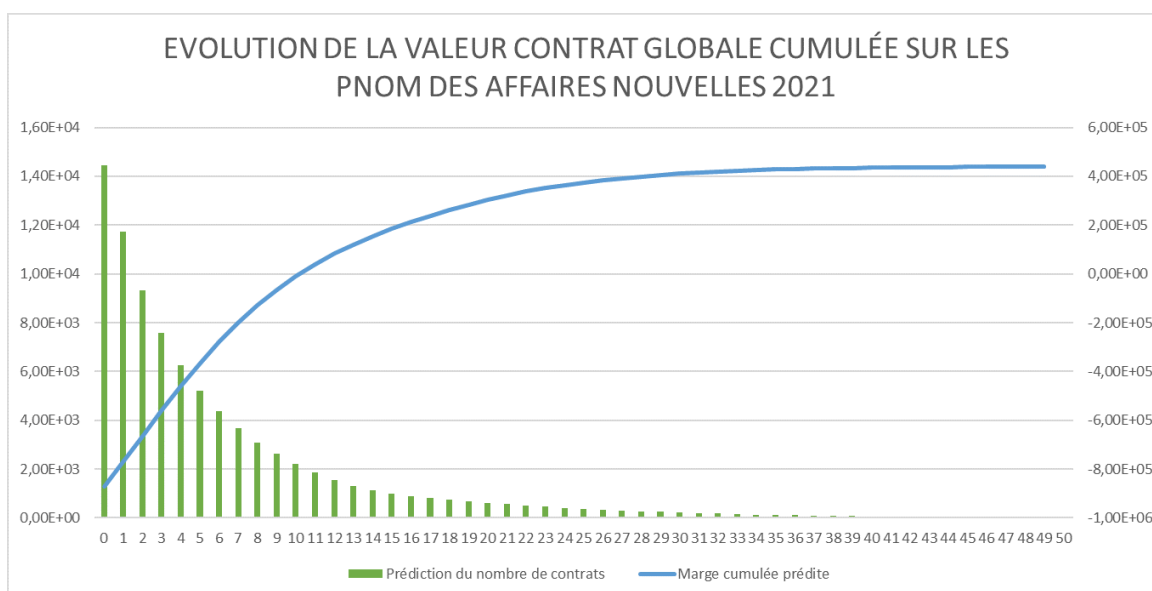


FIGURE - Évolution prédite de la valeur contrat globale cumulée des affaires nouvelles 2021 des propriétaires non-occupants de maisons