

Mémoire présenté devant l'Institut du Risk Management pour la validation du cursus à la Formation d'Actuaire de l'Institut du Risk Management et l'admission à l'Institut des actuaires

le

Par : **Alexandre COUILLAUX**

Titre : **Sensibilité au prix des contrats Habitation et optimisation de la stratégie de renouvellement**

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

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

Membres présents du jury de l'Institut des Actuaires :

Membres présents du jury de l'institut du Risk Management :

Secrétariat :
Bibliothèque :

Entreprise : *Societe Generale Insurance*

Nom : SOGECAP

Signature et cachet :



Siege : Tour D2, 17 bis Place des Reflets
92919 Paris la Défense 2

Adresse postale : TSA 61101 - 92919 Paris la Défense Cedex
086 380 730 RCS Nanterre

Directeur de mémoire en entreprise :

Nom : **Anthony DERIEN**

Signature :



Invité :

Nom :

Signature :

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

(après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise



Signature du candidat



Sommaire

Remerciements	5
Résumé	6
Abstract	7
Introduction	8
I. Présentation de l'étude	9
1. Contexte réglementaire	9
2. Périmètre de l'étude	10
3. Lien entre le comportement des assurés et leur sensibilité au prix	12
4. Politique tarifaire.....	13
II. Structure des données	17
1. Construction de la base de données	17
2. Analyse descriptive.....	21
a. Contexte et analyse macro.....	21
b. Analyse de la base et des variables explicatives	23
3. Analyse en correspondances multiples.....	34
a. Rappels théoriques.....	34
b. Application.....	35
III. Modélisation de la résiliation et sensibilité au prix	39
1. Régression logistique.....	39
a. Quelques rappels théoriques	39
b. Sélection des variables via une technique de <i>machine learning</i>	40
c. Matrice de corrélation.....	44
d. Modèle de résiliation.....	45
2. Modèle de survie.....	54
a. Théorie.....	54
b. Base d'étude.....	57
c. Résultats	58
d. Limites du modèle	62
3. Modèle de sensibilité au prix	62
a. Théorie.....	62
b. Résultats	63
IV. Modélisation de la prime commerciale	67
1. Objectif	67

2.	Produit Habitation	67
3.	Base de données.....	68
a.	Période d'analyse	68
b.	Données contrats	68
c.	Données sinistres	69
4.	Préambule technique	69
a.	Rappels théoriques.....	69
b.	Modèle de fréquence	70
c.	Modèle de coût moyen	71
d.	Mesure de qualité du modèle	71
5.	Modélisation.....	72
a.	Principe.....	72
b.	Significativité des variables	73
c.	Validation des modèles	74
d.	Modélisation de la fréquence	75
e.	Modélisation du coût moyen	77
f.	Modélisation de la prime pure	79
6.	Calcul de la prime commerciale	84
7.	Estimation de la marge.....	84
V.	Optimisation de la stratégie de renouvellement	87
1.	Définition des approches d'optimisation	87
a.	Approche non paramétrique	87
b.	Approche paramétrique	89
2.	Optimisation du chiffre d'affaires sous contrainte d'un nombre de contrats renouvelés	89
a.	Présentation de la base de simulation	89
b.	Fonction d'optimisation du chiffre d'affaires.....	89
c.	Matrice paramétrique actuelle	90
d.	Matrice paramétrique actuelle optimisée	91
e.	Matrice paramétrique optimisée sans contrainte sur les variables.....	92
f.	Approche non paramétrique	93
g.	Résultats des stratégies pour l'optimisation du chiffre d'affaires	95
3.	Optimisation de la marge sous contrainte d'un nombre de contrats renouvelés	96
a.	Présentation de la base de simulation	96
b.	Fonction d'optimisation	96

c. Matrice paramétrique actuelle optimisée	97
d. Matrice paramétrique optimisée sans contrainte sur les variables.....	98
e. Résultats des stratégies pour l’optimisation de la marge sous contrainte d’un nombre de contrats renouvelés.....	99
4. Optimisation finale	99
Conclusion	103
Liste des figures.....	105
Lexique	108
Bibliographie.....	109
Annexes	110
1. Triangle de résiliations cumulé par mois et année d’échéance	110
2. Mécanisme de construction de la valeur client assurantiel sur le périmètre dommages ...	111
3. Taux de résiliation croisé avec la majoration en euro au terme et la formule Investisseur ...	112
4. Taux de résiliation croisé avec la majoration au terme et la formule Etudiant	113
5. Analyse descriptive - taux de résiliation en fonction d’autres variables explicatives	114
6. Analyse en correspondances multiples : rappels théoriques et inertie	119
7. Arbre de décision : arbre complet après élagage.....	122
8. Modèle de Cox : graphiques d’hypothèse des risques proportionnels.....	123
9. Sensibilité au prix : graphiques sur d’autres variables significatives	124
10. Modélisation de la prime commerciale : formules de garanties	125
11. Modélisation de la prime commerciale : résidus de la modélisation de la fréquence de la garantie dégât des eaux	126
12. Modélisation de la prime commerciale : marge en fonction du statut-type croisée avec la formule	127
13. Optimisation de la marge : distribution des majorations de prime pour l’approche non paramétrique.....	128

Remerciements

Mes remerciements vont tout d'abord à *Societe Generale Insurance* et plus particulièrement Laurent Bruno et Alexandre You qui m'ont permis d'accéder à la formation de l'Institut du Risk Management (anciennement appelée Centre d'Etudes Actuarielles).

Je tiens à remercier Anthony Derien, responsable actuariat dommages et directeur de mémoire en entreprise, pour sa disponibilité, ses conseils avisés et ses nombreuses relectures.

Je souhaite également remercier l'équipe Dommages aux biens pour leur aide et plus particulièrement, Daisy Ménard et Manon Beaupoil. Merci aussi à Manon Pieren et Mohamed Gahbiche pour leur assistance sur le logiciel *R* et sur certains aspects techniques du mémoire.

Un grand merci à Christophe Dutang, conseiller scientifique, qui m'a apporté des conseils techniques et formulé des remarques pertinentes pour l'élaboration de ce mémoire.

Je n'oublie pas de remercier Steeve Monteiro et Sébastien Saubert pour avoir apporté un regard extérieur.

Enfin, je souhaite sincèrement remercier mes proches pour leur soutien quotidien et en particulier ma compagne Nathalie et ma petite Anna.

Résumé

Ce mémoire, réalisé au sein du service actuariat dommages de *Societe Generale Insurance*, s'inscrit dans le contexte d'un renforcement des droits pour les consommateurs avec l'arrivée récente de la loi Hamon. La faculté de résilier à tout moment, combinée à l'essor des comparateurs d'assurances et de nouveaux acteurs tels que les assureurs directs ou les GAFAs, induit une plus grande rotation des clients obligeant les compagnies à s'intéresser de près au sujet de la rétention des assurés en portefeuille.

L'objet du mémoire est l'étude de l'élasticité au prix des assurés multirisques habitation face à une majoration de prime à l'échéance. L'élasticité au prix traduit le rapport entre la variation de la demande (renouvellement ou résiliation de la police d'assurance) sur la variation du prix. En utilisant les caractéristiques individuelles des contrats MRH, plusieurs modèles vont être étudiés pour modéliser la résiliation : un modèle de régression logistique et un modèle de survie.

En parallèle, un modèle linéaire généralisé de fréquence et coût moyen sera développé pour établir un tarif technique et ainsi visualiser les segments rentables ou déficitaires dans notre portefeuille.

Ainsi, grâce aux modèles précités, une optimisation de la stratégie de renouvellement de la compagnie sera réalisée. Une première optimisation du chiffre d'affaires sous contrainte d'un nombre de contrats renouvelés sera recherchée. Une optimisation de la marge sera également étudiée pour conserver en portefeuille des contrats rentables. Enfin, il sera proposé une optimisation d'allocation des majorations au terme croisant le chiffre d'affaires, la marge et le volume de contrats.

Mots-clés : IARD, assurance habitation, Loi Hamon, résiliation, arbre de décision, régression logistique, sensibilité au prix, modèle de Cox, modèle linéaire généralisé, tarification, optimisation

Abstract

This thesis, produced within *Societe Generale Insurance's* actuarial department, is part of a strengthening of consumer rights with the recent arrival of the Hamon law. The ability to terminate the contracts at any time, combined with the growth of insurance prices comparison websites and new players such as direct insurers or GAFAs, leads to greater rotation of customers forcing companies to take a close interest in the retention of policyholders in the portfolio.

This thesis aims to determine the price elasticity of insured facing to a premium increase at renewal. Price elasticity reflects the link between the change in demand (renewal or termination of the insurance policy) divided by the price change. Several models, a logistic regression model and a survival model, will be studied to model the termination by using the individual characteristics of property insurance contracts.

In parallel, a generalized linear model of frequency and average cost will be developed to establish a technical tariff and thus visualize the profitable or deficit segments in our portfolio.

Thanks to the aforementioned models, an optimization of the company's renewal strategy will be achieved. A first optimization of the turnover under constraints of renewed contracts will be sought. An optimization of the margin will also be studied to keep profitable contracts in the portfolio. Finally, it will be proposed to optimize the allocation of premium increase at the renewal by combining the turnover, the margin and the volume of contracts.

Keywords : Non-life insurance, property insurance, Hamon law, terminations, decision tree, logistic model, price elasticity, Cox model, Generalized Linear Model, pure premium, optimization

Introduction

Le marché de l'assurance français est soumis à un contexte concurrentiel fort impliquant une pression sur les prix des contrats d'assurance de particuliers et plus particulièrement sur l'habitation.

L'arrivée de nouveaux acteurs tels que les assureurs directs et l'essor des GAFA (Amazon notamment qui commence à s'intéresser à l'assurance) rendent la compétition entre les compagnies d'assurance ardue. Les clients ont désormais, via Internet et les comparateurs d'assurances, la possibilité d'accéder à des informations tarifaires de la concurrence et ainsi comparer leur prime. Or, la rentabilité d'un portefeuille d'assurances est souvent liée à la fidélité des clients en portefeuille ; les affaires nouvelles étant généralement mieux traitées, d'un point de vue tarifaire, que le stock de contrats déjà en portefeuille. Ainsi, le ratio combiné de la MRH, indicateur de la rentabilité en IARD, s'établit à 97% en 2017, en hausse de 3 points par rapport à 2016 sous l'effet de l'augmentation des événements climatiques, laissant une marge de manœuvre limitée pour les assureurs. De plus, les objets connectés et la domotique qui commencent à se développer dans les habitations pourraient à terme faire diminuer les fréquences et les coûts moyens et ainsi faire baisser le chiffre d'affaires des assureurs sur ce marché.

En parallèle, le cadre réglementaire assurantiel a évolué sensiblement ces dernières années avec la loi Châtel instaurée en 2008 puis la loi Hamon entrée en vigueur en 2015. Ces évolutions sont une réelle avancée pour le client mais elles engendrent également des coûts supplémentaires pour les compagnies avec une rotation plus forte des clients (augmentation des coûts d'acquisition d'une part et des coûts de gestion des contrats avec une hausse des avenants d'autre part).

L'objet de ce mémoire est d'analyser le comportement de résiliation des assurés multirisque habitation (MRH) suite à une majoration tarifaire appliquée au terme. L'analyse de cette résiliation sera réalisée non seulement au terme mais également postérieurement au terme. Cette variation de la demande en fonction de la variation de prix correspond à la notion d'élasticité au prix ou sensibilité au prix. Par ailleurs, la construction d'un tarif technique permettra de définir les segments rentables et déficitaires en comparant la prime termée en portefeuille. La combinaison de ces modélisations, associée à une optimisation sous contraintes, servira à déterminer une stratégie optimale d'allocation des majorations au terme avec un objectif d'amélioration du chiffre d'affaires ou de la marge.

I. Présentation de l'étude

1. Contexte réglementaire

La loi Châtel (article L. 113-15-1 du code des assurances) prévoit que les assureurs doivent informer les assurés de la durée de préavis pour résilier à l'échéance leur contrat (délai de 15 à 20 jours pour résilier à compter de la réception de l'avis d'échéance). Depuis 2015 et la mise en vigueur de la loi du 17 mars 2014 relative à la consommation ou loi Hamon (article L. 113-15-2 du code des assurances), l'assuré a désormais la faculté de résilier à tout moment son contrat d'assurance, après une période d'un an de couverture. La loi Hamon a produit ses premiers effets avec des taux de résiliations Habitation en hausse de près d'un point d'après la Fédération Française de l'Assurance (de 12.7% en 2014 à 13.4% en 2017).

Dans cette partie, nous détaillons les deux types de résiliations auxquels nous nous intéresserons dans ce mémoire et qui sont liés à une majoration de la prime à l'échéance. Les contrats d'assurance étant à tacite reconduction, l'assuré doit se manifester auprès de son assureur pour voir son contrat résilié (excepté dans le cas des résiliations à l'initiative de l'assureur). Pour illustrer notre propos, nous prenons le cas d'une échéance principale au 1^{er} janvier N.

Résiliation loi Châtel (depuis 2008)

Il s'agit de la résiliation la plus usitée par les clients du portefeuille Sogessur jusqu'en 2015. A réception de son avis d'échéance, le client dispose d'un délai de quinze jours (si l'avis a été transmis par l'assureur au moins 2 mois et demi avant la date d'échéance, soit le 16 octobre au plus tard) ou de vingt jours (si l'avis est transmis après le 16 octobre) pour transmettre son courrier de résiliation à l'assureur. La prise d'effet de la résiliation a lieu à l'échéance, c'est-à-dire au 1^{er} janvier.

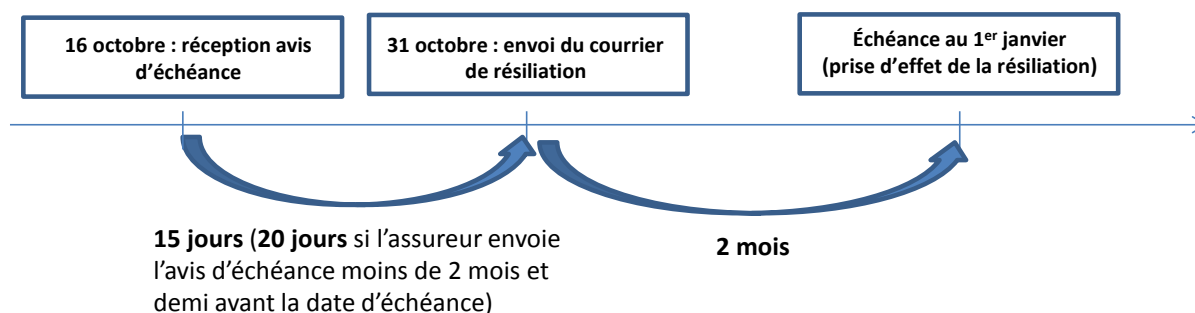


Figure 1 : Résiliation loi Châtel

Résiliation loi Hamon (depuis 2015)

La loi Hamon, en vigueur depuis 2015, permet à l'assuré de résilier à tout moment son contrat d'assurance dès lors qu'il a au moins un an d'ancienneté. Dans l'exemple ci-dessous, si l'assuré envoie son courrier de résiliation le 31 mars, la date d'effet de la résiliation sera un mois plus tard, soit le 30 avril. Nous pouvons remarquer que la date de résiliation est donc déconnectée de la date

d'échéance principale de l'assuré. Ce type de résiliation est majoritaire dans le portefeuille Habitation de Sogessur depuis 2016.

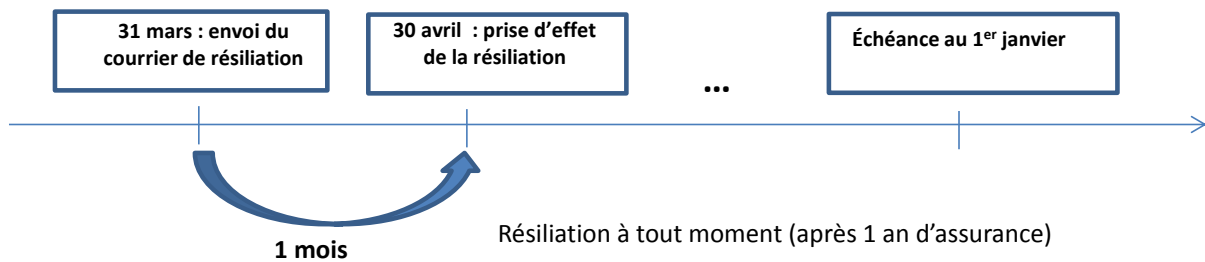


Figure 2 : Résiliation loi Hamon

2. Périmètre de l'étude

L'étude détaillée dans ce document exclut les résiliations à l'initiative de l'assureur (non paiement de la prime, absence de pièces justificatives ou surveillance du portefeuille suite à une sinistralité atypique) qui représente 16% des résiliations à fin 2017. Sont également exclues les résiliations hors échéances en cas de changement ou disparition du risque (vente du bien, déménagement, ...) qui pèsent pour 42% des résiliations.

Le périmètre du mémoire porte donc sur les deux types de résiliations détaillées dans le paragraphe précédent et pouvant être à l'initiative de l'assuré suite à une majoration de prix. Ces résiliations représentent 42% des résiliations enregistrées sur le produit Habitation à fin décembre 2017 et sont réparties comme suit : Hamon (LH) à hauteur de 29% et Châtel (LC) pour 13%.

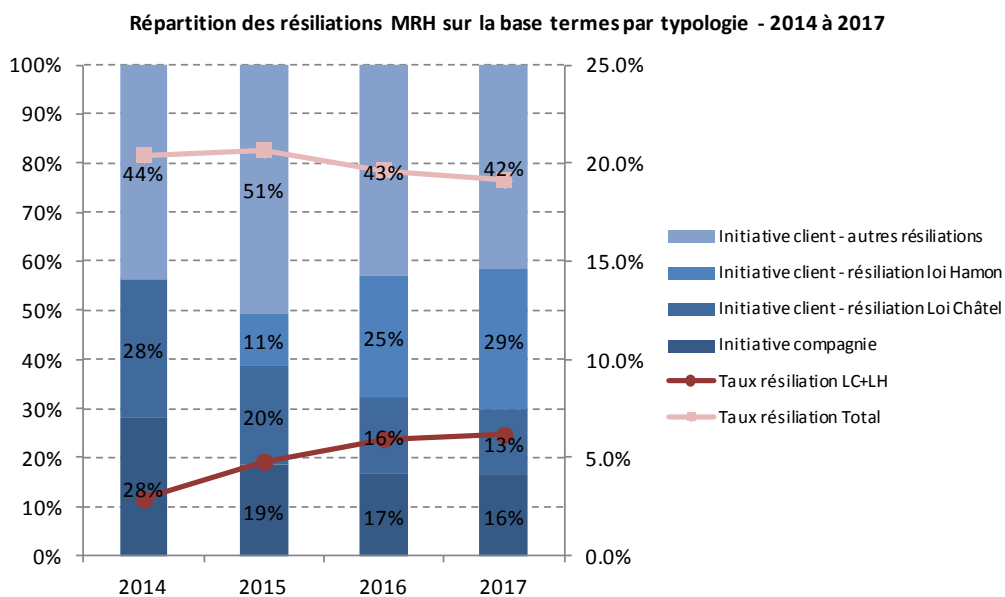


Figure 3 : Répartition des résiliations Habitation par typologie de 2014 à 2017

Le graphique ci-dessus présente la répartition des résiliations sur le produit Habitation entre 2014 et 2017. Il appelle les commentaires suivants :

- Malgré l’instauration de la loi Hamon, le taux de résiliation total du produit MRH diminue légèrement à 19% à fin 2017 grâce aux mesures de fidélisation entreprises depuis quelques années par la Compagnie (mise en place d’une valeur client, campagne d’appels téléphoniques sortants en cas de risque identifié de résiliation, etc.) ainsi qu’à la baisse des résiliations à l’initiative de la compagnie ;
- L’entrée en vigueur de la loi Hamon en 2015 est marquée par une augmentation du taux de résiliation sur le périmètre de l’étude (Châtel et Hamon) qui passe de 3% à plus de 6% en 2017. Par ailleurs, la part des résiliations au titre de la loi Châtel diminue nettement comme indiqué dans la figure 3 ;
- Enfin, les résiliations hors échéances à l’initiative de l’assuré (ou autres résiliations dans le graphique), correspondant à des changements ou des disparitions de risque (vente du bien, décès de l’assuré,...), restent stable à 42% à fin 2017.

Nous avons vu précédemment que l’arrivée de la loi Hamon sur le marché de l’assurance a engendré une augmentation du taux de résiliation à l’initiative des clients. Mais est-ce que cette augmentation de la résiliation est uniquement visible à l’échéance du contrat ou est-elle répartie tout au long de l’année ? Autrement dit, est-ce que les clients ont intégré la faculté de résilier à tout moment dès la mise en vigueur de la loi Hamon ?

Pour répondre à cette question, une étude a été menée sur la base des résiliations et des contrats terminés en MRH entre mars 2014 et décembre 2017. La constitution d’un triangle cumulé de taux de résiliations avec en ligne, l’année et le mois de l’échéance du contrat, et en colonne, l’écart entre le mois du terme et la résiliation, permet de visualiser une répartition homogène des résiliations dans l’année depuis 2015, indépendamment de l’échéance du contrat (cf. annexe 1). Les développements manquants, lorsque l’échéance n’est pas observée avec 12 mois de recul, ont été prolongés par une méthode de projection de type Chain-Ladder.

Illustrons notre propos sur une échéance principale (EP) en avril entre 2014 et 2017 :

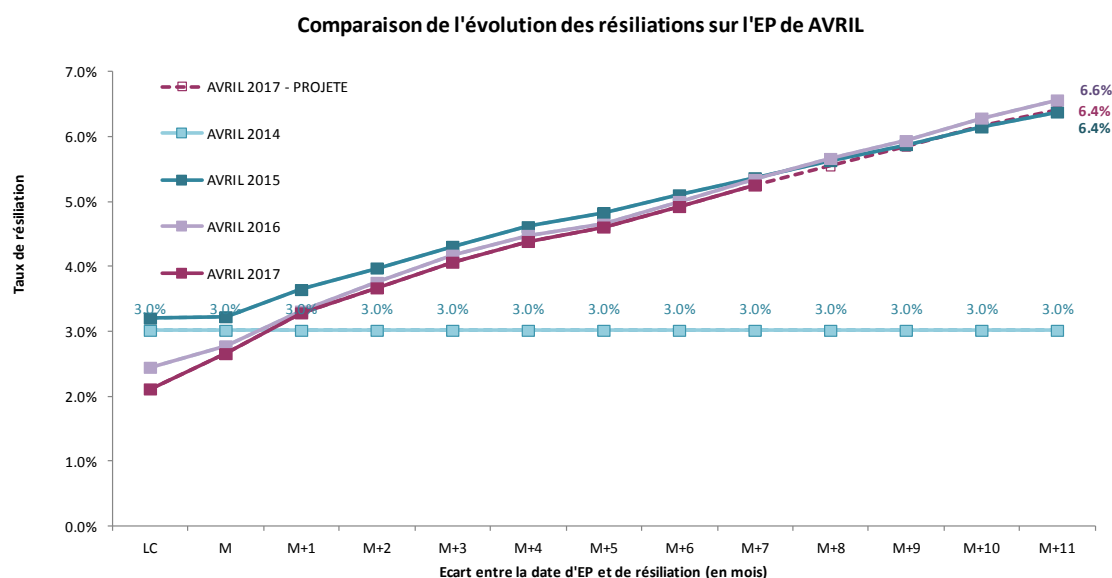


Figure 4 : Evolution des taux de résiliations depuis la loi Hamon - écart entre la date d'échéance d'avril et de résiliation

En 2014, le taux de résiliation Loi Châtel (LC) s'élève à 3% pour l'échéance d'avril et reste stable tout au long de l'année.

A partir de 2015 et encore plus nettement en 2016 et 2017, le taux de résiliation diminue à l'échéance (« effet Châtel ») et progresse de manière quasiment linéaire tout au long de l'année (« effet Hamon ») pour atteindre environ 6.5%. Le niveau de résiliations progresse donc significativement et se répartit, pour une échéance donnée, de manière plutôt homogène dans l'année.

Par ailleurs, comme l'étude est arrêtée à fin décembre 2017, les résiliations relatives à l'échéance d'avril 2017 ne sont pas exhaustives puisque le client a la faculté de résilier jusqu'en mars 2018. Ainsi, le terme d'avril 2017 a été prolongé en pointillé dans le graphique par la méthode Chain-Ladder. Nous introduisons donc dans l'étude deux variables liées la date du terme (mois de recul et mois de développement - cf. paragraphe II.1).

3. Lien entre le comportement des assurés et leur sensibilité au prix

Dans ce paragraphe, nous allons illustrer le comportement de l'assuré face à une majoration de sa prime habitation à l'échéance de son contrat.

Tout d'abord, nous pouvons assimiler l'assurance habitation à une assurance obligatoire. Elle est obligatoire pour les locataires de leurs appartements et pour les copropriétaires, au travers de la garantie responsabilité civile. Il reste donc le cas des propriétaires de maisons pour lesquels l'assurance MRH n'est pas obligatoire mais fortement recommandée au regard du préjudice subi en cas d'incendie ou d'événements climatiques sur un logement non assuré. Selon le taux de couverture communiqué par la FFA dans leur rapport sur l'assurance habitation (97.2% à fin 2016 dont 99.7% pour les résidences principales), la MRH peut être assimilée à un produit obligatoire.

Nous pouvons également intuitivement (et nous le démontrerons par la suite dans ce mémoire) que si l'augmentation à l'échéance est considérée comme acceptable par l'assuré, il aura une forte probabilité de renouveler sa police. Si l'augmentation est trop forte, il aura probablement comme réaction de rechercher un meilleur tarif auprès de la concurrence et s'il en trouve un, il résiliera sa police.

Ces constats nous amènent donc la question suivante : comment modéliser l'élasticité au prix ?

Arrêtons-nous tout d'abord sur la définition de l'élasticité au prix (notée $\varepsilon(p, D)$). Elle traduit le rapport entre la variation de la demande (renouvellement ou résiliation de la police d'assurance) sur la variation du prix (écart en euro ou en pourcentage entre la prime terminée en N et N-1) :

$$\varepsilon(p, D) = \frac{\frac{\Delta D}{D}}{\frac{\Delta p}{p}}$$

Ainsi, lors de l'analyse du taux de renouvellement de la police, l'élasticité sera négative puisque la demande diminuera avec la progression du prix. A l'inverse, l'élasticité au prix est positive au regard du taux de résiliation.

Deux approches sont souvent opposées pour réaliser une étude de sensibilité au prix.

La première consiste à isoler un échantillon du portefeuille et à leur appliquer une étendue de hausse tarifaire (par exemple, de +0% à +10%). En fonction de la réaction des assurés à ces majorations, nous pouvons en déduire des groupes sensibles au prix (pour lesquels il faudra limiter la majoration) et des groupes inélastiques au prix (ceux-ci pourront accepter des majorations plus élevées). Ce test permettrait ainsi de calibrer de manière optimale la politique tarifaire pour l'ensemble du portefeuille. Elle a néanmoins le désavantage d'être complexe à mettre en œuvre opérationnellement et également être difficile à faire accepter aux distributeurs, en l'occurrence la Société Générale et le Crédit du Nord, qui ne verraient pas d'un bon œil ce « test » sur leurs clients bancaires.

La seconde solution, plus courante dans les compagnies d'assurances traditionnelles, consiste à regarder le passé et donc à utiliser les données historiques de majorations de politique tarifaire à l'échéance. Elle a l'avantage d'être utilisable de suite mais peut également refléter une situation passée qui n'est plus en vigueur (exemple dans le cas d'un changement réglementaire des modalités de résiliation). Enfin, cette solution est rendue possible si les majorations des politiques tarifaires passées sont assez différentes d'une année à l'autre. En effet, l'étendue des majorations tarifaires est un préalable pour s'assurer que le modèle pourra « apprendre » en cas de faibles ou fortes majorations.

4. Politique tarifaire

Précisons maintenant dans quel cadre se déroule la politique de renouvellement chez *Societe Generale Insurance*. La politique tarifaire sert à tenir compte des différents facteurs qui peuvent s'appliquer aux polices d'assurance au cours du temps. Ils peuvent être regroupés en deux catégories :

- Effets exogènes : l'inflation des coûts de sinistres (frais d'experts, indice des prix d'entretien et d'amélioration, coût salarial et des matériaux, indice FFB), la multiplicité des événements climatiques et catastrophes naturelles, les évolutions réglementaires (convention IRSI), etc. A titre d'illustration, en 2019, une étude d'Actuaris prévoit une inflation de la prime pure Habitation de +2.2% expliquée par une progression de la sinistralité des garanties Vol et Incendie ;
- Effets endogènes : objectif de rentabilité du produit (ratio combiné technique hors produits financiers inférieur à 100%), objectif de rétention du portefeuille, récompenses des « bons clients » (notion de valeur client détaillée dans le paragraphe II.1), effet mix du portefeuille (descente en gamme vers des formules moins couvrantes ou progression de la nouvelle offre disposant de primes moyennes inférieures à l'ancienne offre), etc.

La politique tarifaire chez *Societe Generale Insurance* est donc un moment clé de la vie de l'entreprise. Elle se déroule sur près de 6 mois et fait intervenir de nombreux acteurs de la compagnie (Directions Marketing, Technique, Commerciale, Informatique, Contrôle de Gestion, ...) mais également les distributeurs Société Générale et Crédit du Nord. En effet, avant la décision d'évolution des primes d'assurances au 1^{er} janvier de chaque année, un système de gouvernance est mis en place pour faire valider les décisions de la politique tarifaire en interne puis auprès de nos distributeurs.

Les principales étapes de la politique tarifaire sont précisées dans le schéma ci-dessous :

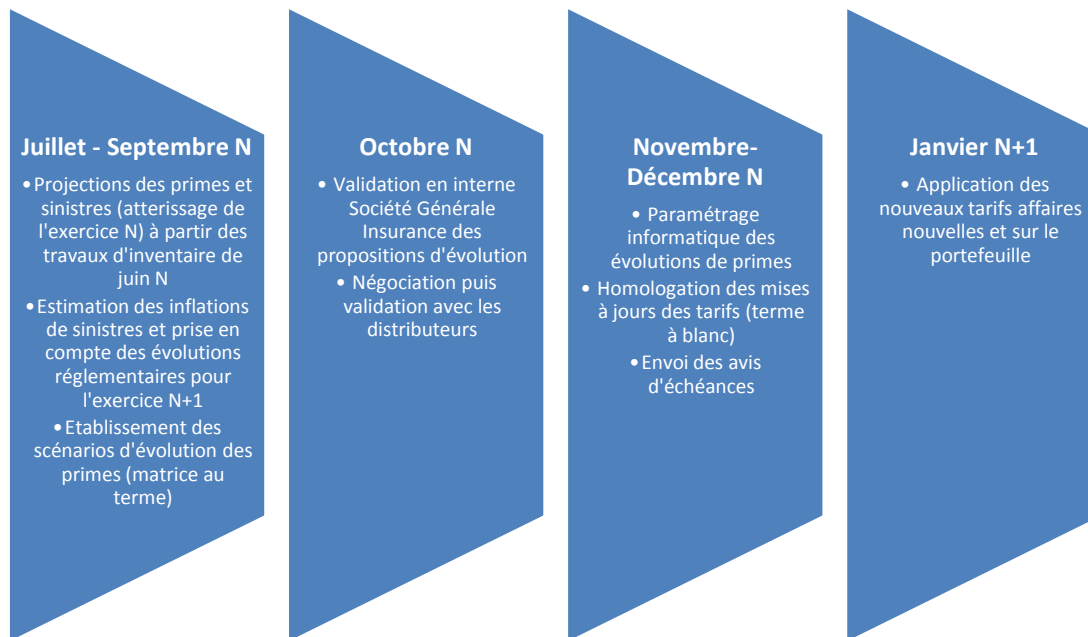


Figure 5 : Temps forts de la politique tarifaire chez Societe Generale Insurance

En général, pour un même produit d'assurance, deux politiques tarifaires distinctes sont décidées : une pour les affaires nouvelles et une pour le portefeuille. Le mémoire portera sur l'optimisation de la politique tarifaire du terme. La mise à jour du tarif à l'échéance fonctionne sur un principe de revalorisation et dépend de plusieurs critères :

- La formule (niveau de garanties) ;
- Le statut d'occupation (locataire, propriétaire occupant, propriétaire non occupant) ;
- La nature du bien (appartement / maison) ;
- La sinistralité (non sinistré / sinistré) ;
- L'ancienneté du contrat en portefeuille ;
- La valeur client.

Cette unique politique tarifaire à l'échéance reposant sur une pratique historique, on pourrait s'interroger sur la pertinence de ce choix. Existe-t-il une saisonnalité dans le taux de résiliation qui pourrait nous conduire à envisager plusieurs politiques tarifaires ?

Illustrons cette question par deux graphiques.

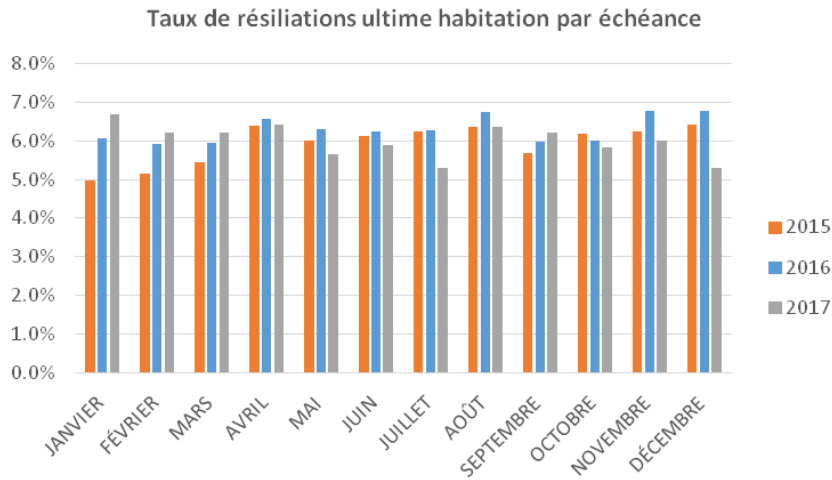


Figure 6 : Taux de résiliation vu à l'ultime par échéance principale entre 2015 et 2017

Le graphique ci-dessus ne présente pas de tendance particulière au niveau du taux de résiliation par échéance principale et pour les années 2015 à 2017. Les taux de résiliations sont projetés à l'ultime pour l'année 2017 car certaines termes ne présentent pas 12 mois de recul (cf. explications au paragraphe I.2).

En choisissant l'année 2016 et en regardant comment évolue le taux de résiliation sur chaque échéance et en fonction de la date d'enregistrement de la résiliation, on obtient une nappe relativement constante.

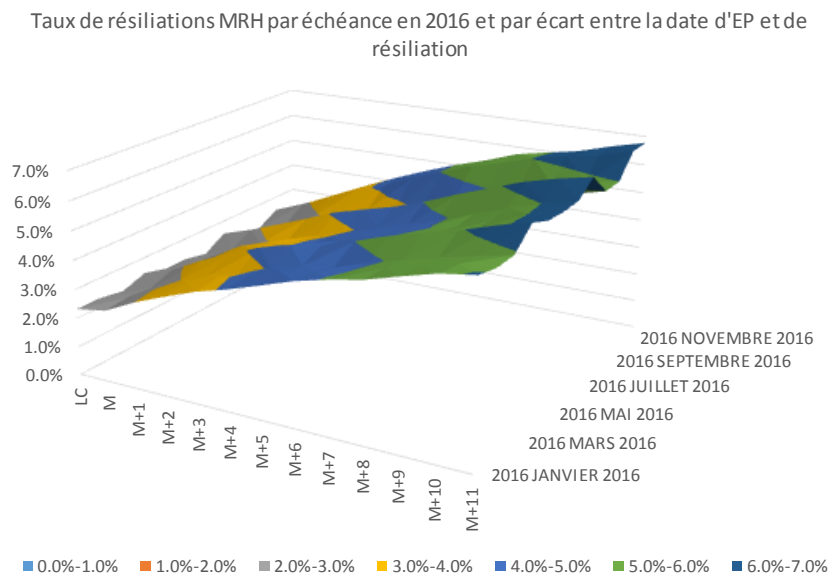


Figure 7 : Nappe de résiliation MRH pour les échéances 2016 et en fonction de l'écart entre la date d'échéance et la date de résiliation

Ainsi, le choix d'une politique tarifaire unique au terme ne semble pas devoir être remise en cause.

II. Structure des données

1. Construction de la base de données

L'élaboration d'une base de données est une étape essentielle à toute étude statistique et actuarielle. Nous allons détailler dans ce chapitre les variables utilisées pour construire notre base de modélisation et l'origine de ces données. L'ensemble des étapes lié à la construction de la base et au traitement des données a été réalisé sous le logiciel SAS.

Notre étude porte sur l'ensemble des contrats Habitation distribués par les réseaux Société Générale et Crédit du Nord et qui ont passé un terme entre janvier 2015 et décembre 2017, soit près de 1,3 millions d'occurrences. Le choix des années n'est pas laissé au hasard et correspond évidemment à l'entrée en vigueur de la loi Hamon. Pour un contrat, nous pouvons donc visualiser jusqu'à 3 avenants et donc 3 lignes dans la base de données lié à sa situation (renouvelé ou résilié) à chaque échéance entre 2015 et 2017 (cf. extrait ci-dessous). La variable à expliquer sera donc une indicatrice de renouvellement du contrat (1 : Oui / 0 : Non). En parallèle, la prime évoluera à chaque échéance en fonction de la politique tarifaire en vigueur.

	N° Contrat	N° Avenant	N° Personne	CO-Code Formule	CO-Code Fractionnement	CO-Code Situation Police	CO-Date Création Affaire Nouvelle	CO-Début Effet contrat	CO-Début Effet Avenant	CO-Fin Effet Avenant	CO-Montant Prime TTC	Var_perc	Var_euro
1		023		HB	S	Q	1997-05-09	1997-05-10	2015-05-01	2016-04-30	293.44	0.0275949012	7.88
2		024		HB	S	Q	1997-05-09	1997-05-10	2016-05-01	2017-04-30	305.26	0.040280807	11.82
3		025		HB	S	E	1997-05-09	1997-05-10	2017-05-01	2017-12-31	320.58	0.0501867261	15.32

Figure 8 : Extrait de quelques variables de la base de données

Notre base de données finale contient différentes variables réparties suivant les catégories détaillées ci-après et provenant de plusieurs sources de données :

- Les données **contrats et risques** correspondent aux variables traditionnelles relatives à la souscription (date de souscription, date de début d'effet du contrat, date de résiliation, date d'avenant, ...) et aux données liées aux caractéristiques du risque que nous précisons ici :
 - Statut d'occupation : locataire, propriétaire occupant et propriétaire non occupant ;
 - Nature du bien : appartement et maison ;
 - Type de résidence : résidence principale, secondaire et donné en location ;
 - Nombre de pièces de l'habitation : de 1 à 20 ;
 - Formule de garanties : niveau de couverture en terme de garanties choisi par le client. Deux gammes de produits coexistent dans les bases : une ancienne offre antérieure à novembre 2010 et une nouvelle offre distribuée depuis. Le nombre de formules s'élevant à 11, nous avons choisi de les regrouper en 5 formules (le détail des formules iii. à v. est précisé dans le chapitre IV relatif à la prime commerciale) :
 - i. Etudiant ;
 - ii. Propriétaire non occupant ;
 - iii. Entrée de gamme ;
 - iv. Moyen de gamme ;
 - v. Haut de gamme.
 - Surface des dépendances ;

- Canal de distribution : Société Générale, Crédit du Nord ou Sogessur (vente à distance) ;
- Type de vente (en agence, à distance) ;
- Adresse du souscripteur : le souscripteur habite t'il la même adresse que le risque assuré ?
- Année de construction du bien ;
- Niveau des capitaux objets usuels et objets de valeurs.
- Les données **sinistres** sont également étudiées et rapprochées des données contrats :
 - Les sinistres garantis et déclarés 12 mois avant le terme (indicatrice de la présence d'un sinistre et éventuellement le nombre de sinistres sur la période d'analyse) ;
 - Le nombre de signalements en surveillance du portefeuille en cas de sinistralité atypique.
- En tant que bancassureur, Sogessur dispose également de certaines données **bancaires** de la Société Générale dont celles étudiées dans le cadre de ce mémoire. Elles sont précisées ci-après :
 - Ventes liées : est-ce que le contrat d'assurance habitation est souscrit concomitamment avec un prêt immobilier (PPI) ? Le rapprochement avec l'assurance MRH est également réalisé pour la souscription d'un prêt consommation (utile par exemple dans le cadre de travaux de rénovation d'un bien immobilier) ou la détention d'un compte à vue ;
 - Segmentation bancaire du client : la Société Générale propose les modalités suivantes (où $y > x$) :
 - i. Grand Public : si les avoirs en banque sont inférieurs à x K€ ;
 - ii. Bonne gamme : si les avoirs en banque sont compris entre x et y K€ ;
 - iii. Patrimonial : si les avoirs en banque sont supérieurs à y K€.
 - Jeune actif : il s'agit d'une variable bancaire renseignant des clients âgés entre 18 et 29 ans, n'ayant plus le statut étudiant et exerçant une activité professionnelle ou étant en recherche d'emploi.
- Les variables de type **clients** sont également ajoutées à la base de données :
 - Age de l'assuré ;
 - Catégorie socio professionnelle : salarié, fonctionnaire, étudiant, agriculteur, profession libérale, etc. ;
 - Ancienneté du contrat ;
 - Multi équipement du client : présence d'autres contrats Automobile, Garantie des accidents de la vie, Protection Juridique (indicatrice et les nombres de contrats associés) ;
 - Le nombre d'impayés sur les 12 derniers mois ;
 - L'assuré est-il un salarié du groupe Société Générale ?
 - Assureur précédent : cette variable présentant plusieurs centaines de modalités, nous avons choisi de les classer suivant la provenance du précédent assureur :
 - i. 1^{ère} assurance ;
 - ii. Bancassureur ;
 - iii. Courtier ;
 - iv. Direct ;
 - v. Mutualiste ;
 - vi. Sogessur ;
 - vii. Traditionnel ;
 - viii. Autres ;

- ix. Ne sait pas.
- o Valeur client : cet indicateur calculé depuis 2015 permet d'identifier les « bons clients » du point de vue de plusieurs critères : leur sinistralité vis-à-vis du S/P d'équilibre de chaque produit, multi-équipement, incidents de paiement de la prime, ancienneté du client. Ce score évolue de 1A à 6 avec la description de chaque valeur dans le tableau ci-dessous (figure 9). Le détail de la classification du score est disponible en annexe 2.
- o Segment de rentabilité : classification du score de valeur client en trois segments :
 - i. Très rentables : score de 1A à 1C ;
 - ii. Rentables : scores de 2 à 4 ;
 - iii. Pas rentables : scores de 5 à 6.

Valeur	Rentabilité	Libellé	Description synthétique
1A	Clients rentables	Top multi-équipé	Clients extrêmement rentables et détenant plus de 3 contrats à Sogessur (parmi AUTO, MRH, GAV, PJ)
1B		Top à équiper	Clients extrêmement rentables et détenant 1 ou 2 contrats à Sogessur (parmi AUTO, MRH, GAV, PJ)
1C		Très bon	Clients très rentables, avec une ancienneté de plus de 10 ans chez Sogessur
2		Bon	Clients moyennement rentables (très rentables avec une ancienneté moyenne inférieure à 10 ans OU rentables avec une ancienneté moyenne supérieure à 10 ans)
3		Standard	Clients peu rentables, avec une ancienneté moyenne inférieure à 10 ans
4		Risqué	Clients rentables mais considérés risqués* pour Sogessur
5	Clients non rentables	S/C déséquilibré	Clients non rentables au regard du montant des primes payées et du montant des sinistres indemnisés par Sogessur
6		S/C très déséquilibré	
NA	Clients sans segment de valeur (Clients détenteurs AUTO, MRH, GAV/AV ou PJ depuis moins de 6 mois)		

Figure 9 : Détail du score de valeur client

- Des données **comportementales** sont historisées à fréquence hebdomadaire et mensuelle dans les bases techniques de la Compagnie et viennent enrichir notre étude :
 - o Nombre de contacts (e-mail, téléphone, courrier et fax) enregistré au niveau du client auprès de la plateforme de gestion de contrats deux mois avant et dix mois après la date de terme du contrat pour capter les effets « Châtel » et « Hamon » ;
 - o Degré de risque de ses contacts clients classés suivant les motifs d'appels en : faible, moyen et élevé.

Près de 3 millions de contacts client sont ainsi répertoriés entre 2014 et 2017. Les motifs d'appels liés à la résiliation sont évidemment retirés de la base pour éviter une prophétie « auto réalisatrice » du modèle. La table ci-dessous détaille la classification « à dire d'expert » des motifs d'appels.

Contacts à risque élevé	Explication et modif comptable Demande de Relevé d'Information (RI) Divers complexes Devis d'avenant Réclamation Informations terme Dde Geste Commercial Comptabilite complexe Demande sur les antécédents
Contacts à risque moyen	Info garanties et procédures Comptabilite simple Autres documents Annulation MRH/PJ/SCO/Auto/CCR/GAV Notification opposition Informations sur contrats Comission de surendettement Complément d'informations
Contacts à risque faible	Demande attest, CV, sit cpte.. Mandat de gestion VA Demande de modifications Demande de souscription Déclaration SIN ou suivi Relances justificatifs Divers simples Autre renseignement Vos besoins, nos conseils Aide à la souscription Assistance outil IARD Information produit non détenu Dde changement date effet Attest détecteur de fumée Renseignement adhésion Erreur d'appel

Figure 10 : Classification des contacts clients auprès de la plateforme de gestion

- Les variables de **prix** sont également ajoutées à la base de données :
 - Niveau de la prime d'assurance au moment du terme ;
 - Variation de la prime en euro et en pourcentage par rapport à l'année précédente ;
 - Variation de la prime en euro et en pourcentage sur les deux et trois dernières années;
 - Ecart entre le tarif affaires nouvelles et portefeuille ;
 - Coefficient de défense de portefeuille (l'assuré appelle à l'échéance pour négocier ou annuler une majoration au terme). Ce coefficient est donc inférieur à 1.
- Les variables liées à la **résiliation** sont prises en compte :
 - Type de résiliation à l'initiative du client : résiliation en « Loi Châtel » et « Loi Hamon » ;
 - Mois de développement : écart entre la date de résiliation et la date du terme ;
 - Mois de recul : recul entre la date du terme et la date d'observation des données (31 décembre 2017), plafonné à 12 mois. Par exemple, pour l'échéance d'avril 2017, le mois de recul est de 9.

Par ailleurs, nous avons choisi d'exclure certaines typologies de données :

- Les formules Etudiant et Investisseur (Propriétaire non occupant ou PNO) sont exclues du périmètre car ces assurés ont des caractéristiques particulières (faible durée de vie dans le portefeuille, petites primes) qui engendrent un comportement inélastique au prix (cf.

graphiques en vision univariée en annexes 3 et 4). Par ailleurs, les investisseurs peuvent inclure leur prime d'assurance dans leur charges déductibles lors de l'imposition de leurs revenus fonciers ;

- Les résiliations à l'initiative de la compagnie ou hors échéances (déménagement, vente du bien, décès de l'assuré, ...) sont conservées dans la base. Néanmoins, ces occurrences sont retraitées pour être considérées comme toujours en cours dans la mesure où la majoration de prime n'est pas à l'origine du départ des assurés ;
- Les contrats du distributeur Crédit du Nord pour lesquels nous ne disposons pas de données bancaires (5.3% du volume de la base totale) sont également retirés de la base ;
- Enfin, les affaires nouvelles souscrites en 2017 sont exclues du scope de l'analyse car elles n'ont pas passé de terme durant l'année et il est donc impossible de visualiser une évolution de la prime entre deux échéances.

2. Analyse descriptive

a. Contexte et analyse macro

Le portefeuille habitation chez Sogessur fait partie des principaux contributeurs au niveau du chiffre d'affaires de la Compagnie. La croissance est régulière depuis plusieurs années avec une augmentation d'environ 10% du chiffre d'affaires chaque année portée principalement par un effet volume. A fin 2017, le portefeuille compte environ 680 000 contrats pour 148 M€ de primes acquises, représentant une prime moyenne HT de 225€ (contre 242€ pour le marché à fin 2016, différence expliquée par l'écart de propriétaires et locataires entre Sogessur et le marché).

Malgré le fort développement de Sogessur, la Compagnie reste un acteur modeste sur le marché français de l'Habitation avec une part de marché estimée à 1.4% en 2017.

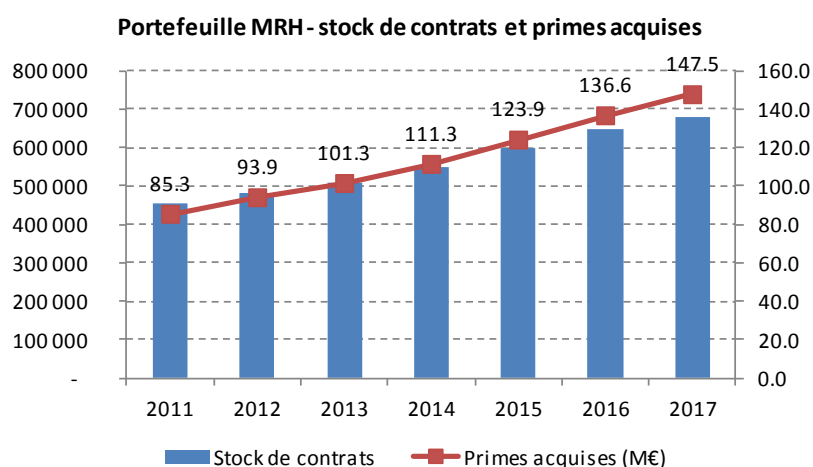


Figure 11 : Stock de contrats et primes acquises MRH entre 2011 et 2017

La forte croissance s'explique par un solde positif de contrats (écart entre les affaires nouvelles et les résiliations). Cet effet s'est accru en 2015, année de mise en vigueur de la « loi Hamon » qui a pour objectif d'augmenter la rotation des contrats au sein des portefeuilles des assureurs. Sur le graphique ci-dessous, le niveau d'affaires nouvelles n'est pas comparable entre Sogessur et le marché dans la mesure où le portefeuille du bancassureur est encore modéré et donc le poids des affaires nouvelles est relativement important. Néanmoins, l'effet « Loi Hamon » est bien visible à partir de 2015 avec une progression comparable du ratio affaires nouvelles divisé par le stock de contrats en N-1 entre Sogessur (+8%) et le marché (+10%). A partir de 2016, la diminution du taux d'affaires nouvelles s'explique par l'accroissement relatif du portefeuille et à une stabilité du niveau d'affaires nouvelles.

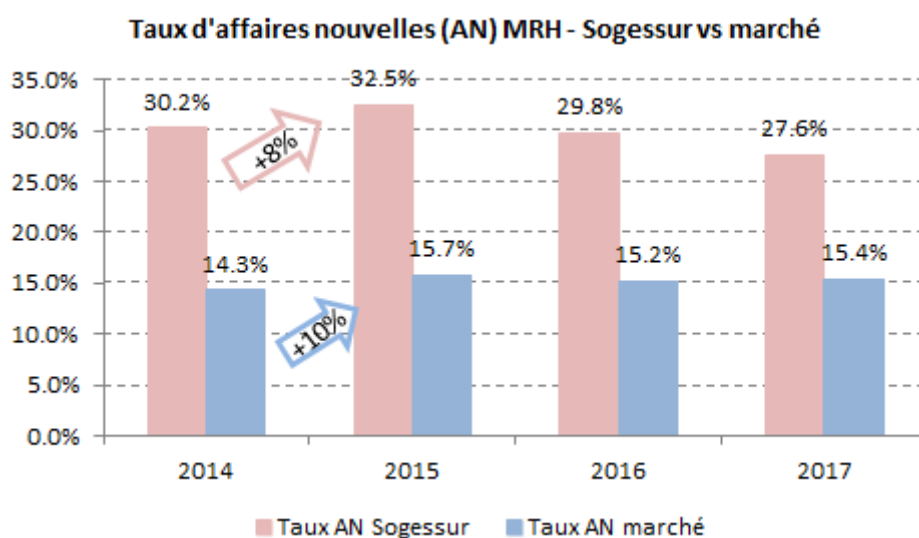


Figure 12 : Comparaison du taux d'affaires nouvelles entre Sogessur et le marché

La croissance du portefeuille s'explique par la montée en puissance de l'assurance au sein du réseau bancaire Société Générale et plus récemment du Crédit du Nord. Les contacts entre le prospect et le banquier sont plus importants en comparaison avec un assureur traditionnel et sont désormais l'occasion d'échanger sur l'environnement assurantiel du client. Par ailleurs, suivant le niveau des taux d'intérêts, le nombre de souscriptions de prêts immobiliers peut varier favorablement et ainsi avoir une répercussion positive sur le volume d'assurances habitation.

La figure 13 présente les volumes de prêts immobiliers distribués par le réseau Société Générale et les taux d'emprunt d'Etat à 10 ans, référence courante pour déterminer les taux des crédits immobiliers, entre 2013 et 2017.

Sur cette période, nous pouvons constater 3 phases :

1. Stabilité des volumes de prêts immobiliers entre 2013 et 2014 expliquée par une stabilité des taux d'emprunt jusqu'en 2014 et un décalage dans la répercussion de la baisse des taux dans les taux des prêts immobiliers ;

2. Croissance des prêts immobiliers en 2015 expliqué par le niveau bas des taux d'emprunt et à un effet communication des médias. Par ailleurs, la renégociation des prêts souscrits antérieurement à 2015 explique un afflux de volumes ;
3. Niveau des prêts immobiliers en 2016 et 2017 supérieur à 2013-2014 expliqué par le maintien des taux d'emprunt d'Etat à environ 0.5% et par la diminution du spread entre les taux OAT et les taux des prêts immobiliers.

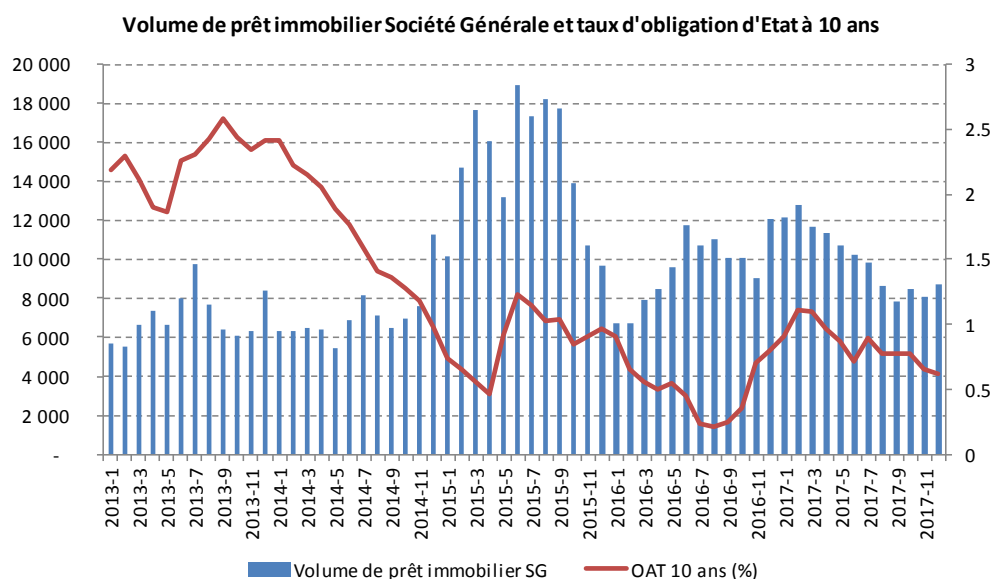


Figure 13 : Comparaison du volume de prêts immobiliers et des taux d'obligation d'Etat à 10 ans entre 2013 et 2017

b. Analyse de la base et des variables explicatives

Analysons désormais la base d'étude construite au paragraphe II.1. Dans cette base, nous dénombrons plus d'un million d'occurrences pour un taux de résiliation moyen de 6.22%.

L'étude détaillée des variables explicatives permet de mettre en lumière des données qui pourraient être pertinentes à utiliser dans le modèle de résiliation. Par ailleurs, cette analyse permet de repérer des valeurs extrêmes, de faire des regroupements de modalités si les volumes ne sont pas assez significatifs ou de mettre en exergue des tendances.

Les variables principales sont détaillées dans cette partie tandis quelques variables secondaires sont présentées en annexe.

Les graphiques s'interprètent de la façon suivante :

- L'axe des abscisses détaille les modalités de la variable considérée ;
- L'axe des ordonnées représente, à gauche, le taux de résiliation en bleu et à droite, le volume de contrats terminés avec des histogrammes de couleur rouge.

- L'année du terme

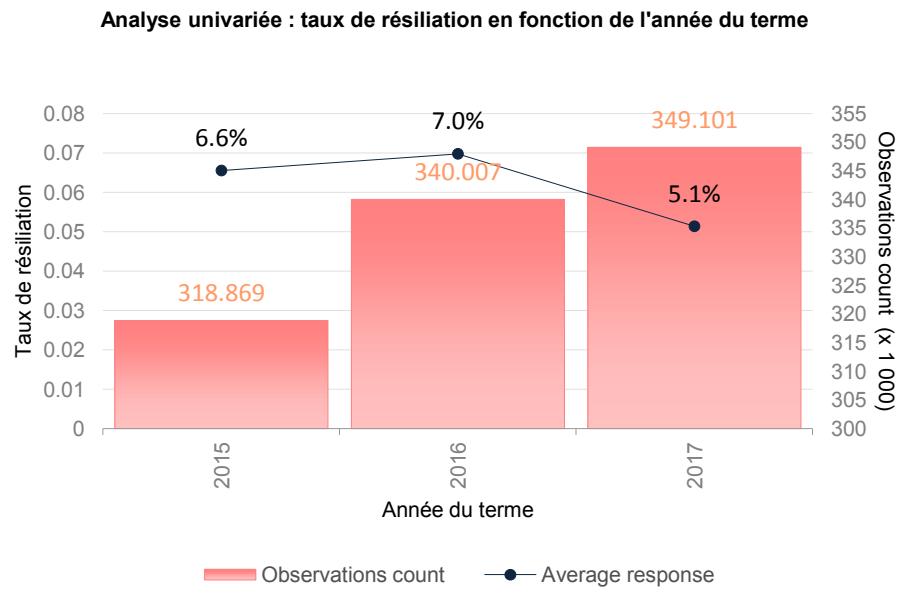


Figure 14 : Taux de résiliation en fonction de l'année de l'échéance

Comme indiqué précédemment, l'étude porte sur les contrats terminés entre janvier 2015 et décembre 2017. Ce graphique est intéressant car il justifie l'emploi de la variable présentée ci-dessous et qui s'intitule « mois de recul ». En effet, une lecture rapide du graphique pourrait indiquer que le taux de résiliation diminue en 2017 par rapport à 2015 et 2016. En réalité, cette hausse s'explique seulement par le fait que la vue de 2017 est incomplète dans la mesure où les assurés d'échéance novembre 2017 ont la possibilité de résilier au titre de la loi Hamon jusqu'en octobre 2018. Ainsi, l'exercice 2017 contient la totalité des résiliations « loi Châtel » mais une partie seulement des résiliations « Loi Hamon ».

- Le mois de recul

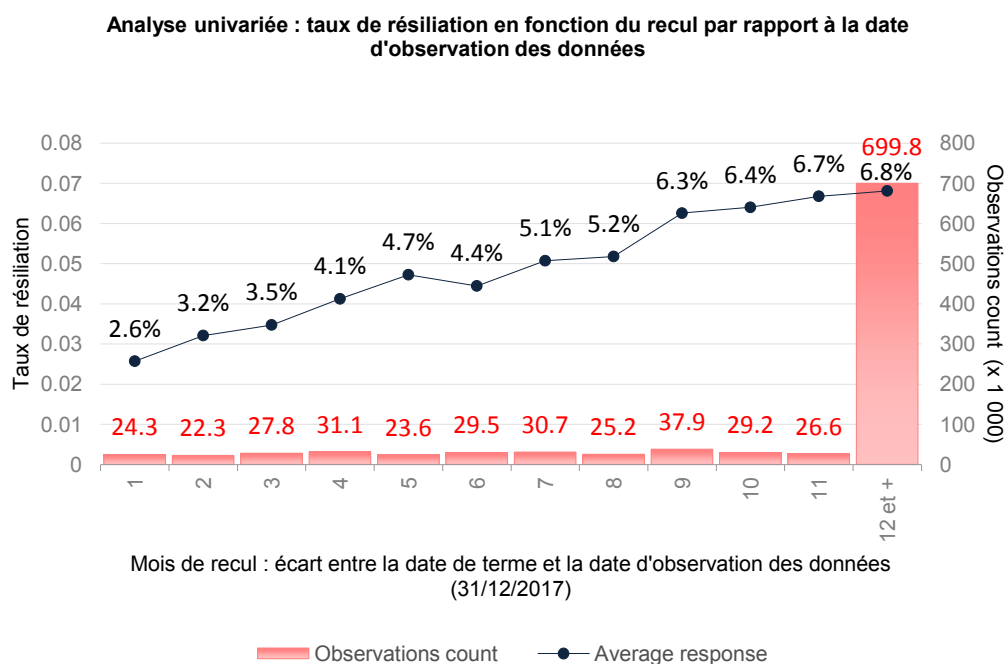


Figure 15 : Taux de résiliation en fonction de l'écart entre le terme et l'observation des données

Cette variable quantitative à valeurs entières vient conforter l'analyse supra. Le taux de résiliation progresse en fonction du mois de recul. Plus le recul est faible (i.e. la date de terme est proche de la date d'observation des données), plus le taux de résiliation est faible puisque les résiliations au motif « loi Châtel » sont intégrées mais celles au motif « loi Hamon » le sont de manière incomplète. Ainsi, les valeurs 1 à 11 de la variable mois de recul correspondent aux termes de l'année 2017. Cette variable devra donc être intégrée dans le modèle de résiliation pour tenir compte de l'absence de recul sur les termes de 2017 à la différence des termes 2015 et 2016.

- Données prix : La variation de la prime en euro au terme

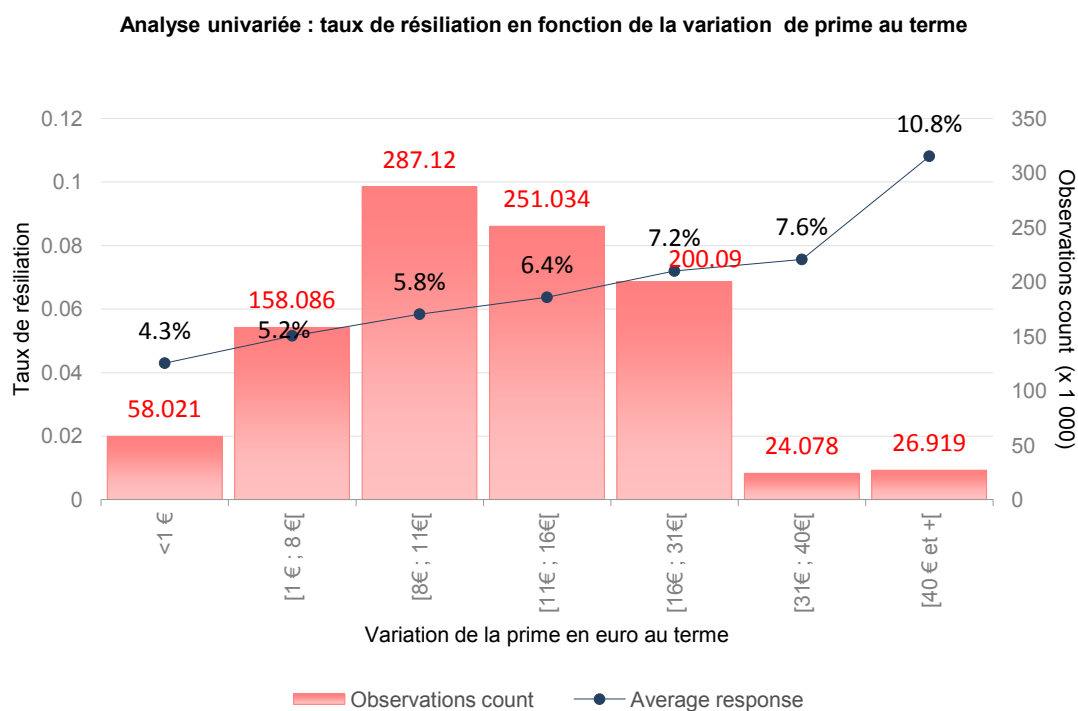


Figure 16 : Taux de résiliation en fonction de la variation de prime au terme

La variation de prime en euro au terme est une variable quantitative continue qui correspond à l'écart de prime payée par le client entre le terme N et N-1 suite aux mesures tarifaires définies dans le cadre de la politique tarifaire. Cette variable a été discrétisée en calculant les percentiles à 5%, 25%, 50%, 75% et 95%. Par ailleurs, le plafond de majoration sur le produit Habitation étant paramétré à 40€, ce seuil a été ajouté à la catégorisation de la variable.

Plus la majoration tarifaire est élevée, plus le taux de résiliation augmente. La hausse importante du taux de résiliation observée pour une variation de prime supérieure à 40€ s'explique par la surveillance de portefeuille mise en place par la Compagnie pour les clients sur-sinistrés et qui peut générer des majorations plus élevées que le plafond de la politique tarifaire.

Par ailleurs, l'analyse du taux de résiliation en fonction des variations cumulées sur les deux derniers termes est également disponible en annexe 5. Les mêmes conclusions s'appliquent, à savoir une progression du taux de résiliation des contrats en fonction du niveau de majoration de la prime au terme.

- Données prix : le niveau de prime client au terme

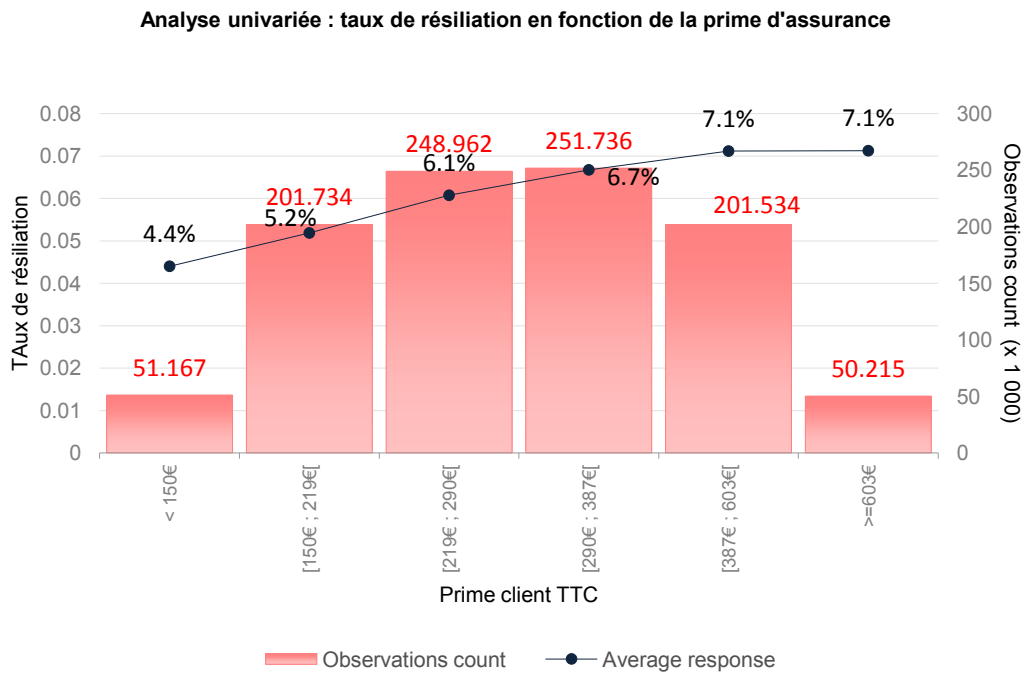


Figure 17 : Taux de résiliation en fonction de la prime client

La variable quantitative continue analysée ici correspond à la prime d'assurance TTC payée par le client au terme. Cette variable a été catégorisée en calculant les percentiles définis précédemment. Plus la prime commerciale augmente, plus le taux de résiliation de l'assuré s'accroît ce qui peut s'expliquer par une mise en concurrence plus forte des clients pour des primes élevées. La prime moyenne de la base d'analyse s'élève à 322€ TTC sur la période. Ce niveau est supérieur à celle présentée au paragraphe précédent dans la mesure où les formules Etudiant et Investisseur (présentant des primes faibles) sont exclues.

- Données client : l'âge de l'assuré

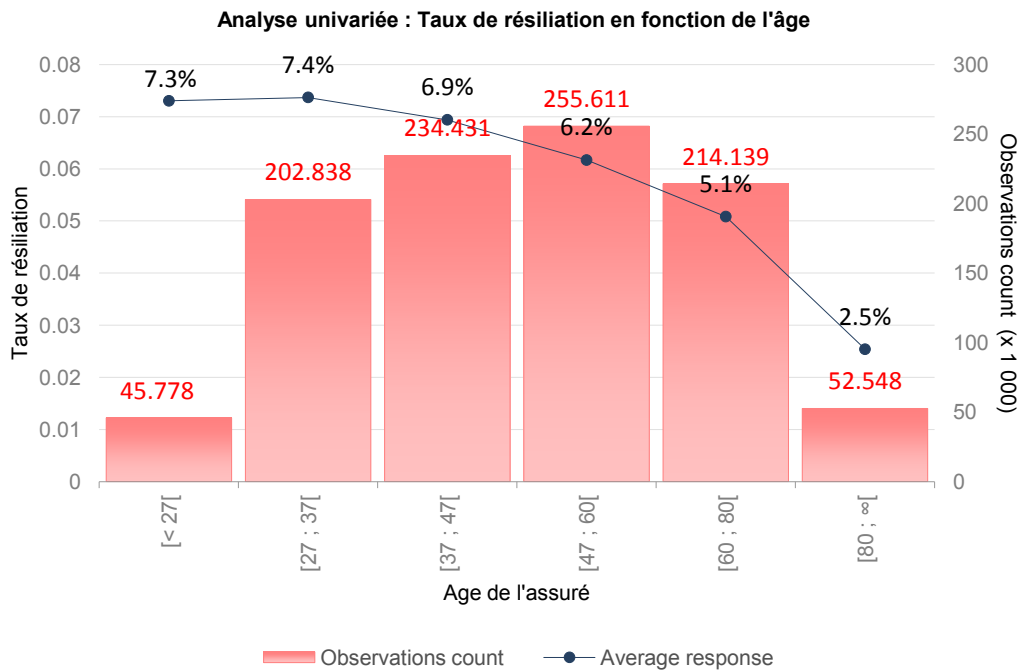


Figure 18 : Taux de résiliation en fonction de l'âge de l'assuré

L'âge de l'assuré est une variable quantitative à valeurs entières observée à la date du terme. La variable a été discrétisée en calculant les percentiles à 5%, 25%, 50%, 75% et 95%. Le taux de résiliation des polices diminue avec l'âge de l'assuré, ce qui peut s'expliquer par la plus forte mobilité des « jeunes souscripteurs » et leur appétence pour Internet et les comparateurs de prix d'assurance. Ces propos sont néanmoins à nuancer dans la mesure où les jeunes sont souvent des locataires qui ont un taux de résiliation plus faible donc il faudra confirmer cette analyse avec la modélisation toutes choses égales par ailleurs.

La variable âge peut néanmoins poser des difficultés pour l'assureur d'un point de vue éthique dans la mesure où ce type de graphique conduirait à majorer davantage les personnes plus âgées. Une alternative à l'utilisation d'une telle variable pourrait être l'étude de l'ancienneté du contrat (cf. annexe 5).

- Données client : segment de rentabilité

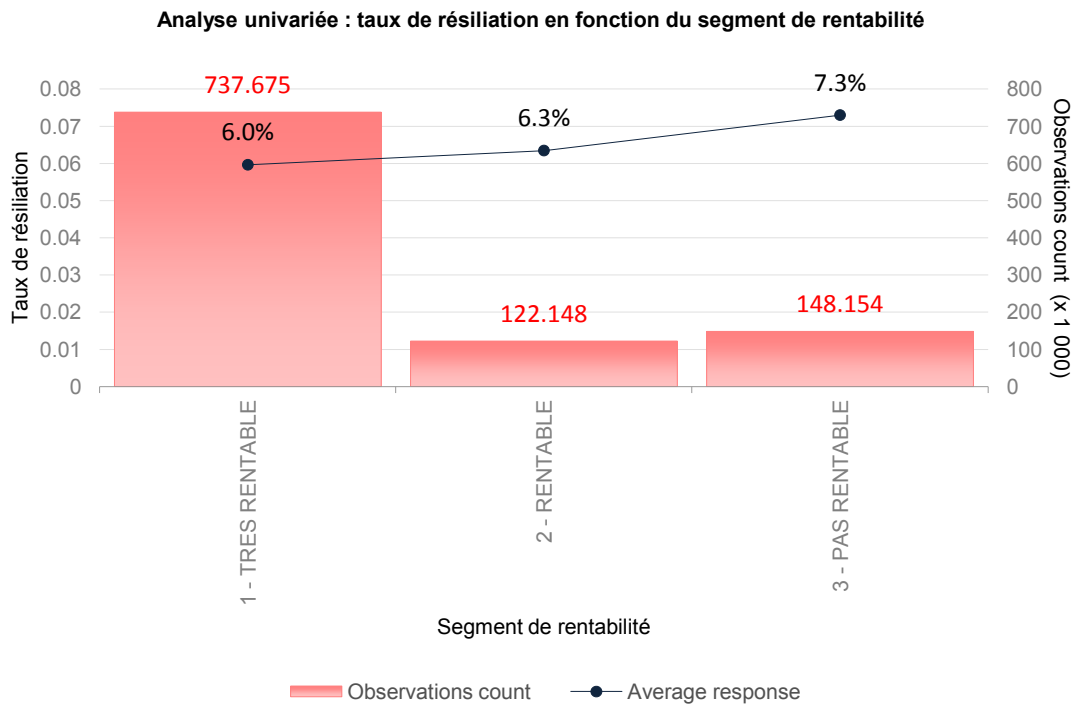


Figure 19 : Taux de résiliation en fonction du segment de rentabilité

La variable segment de rentabilité résulte du score de la valeur client assurantielle (note de 1A à 6) qui est déterminé à partir de plusieurs critères (sinistralité, multi-équipement, impayés et ancienneté du client). Le taux de résiliation en fonction de ces différents critères est présenté en annexe 5.

Le graphique 19 indique qu'un « bon » client présente une probabilité de résilier inférieure à un client non rentable.

- Données client : typologie de contacts clients auprès de la plateforme de gestion de contrats

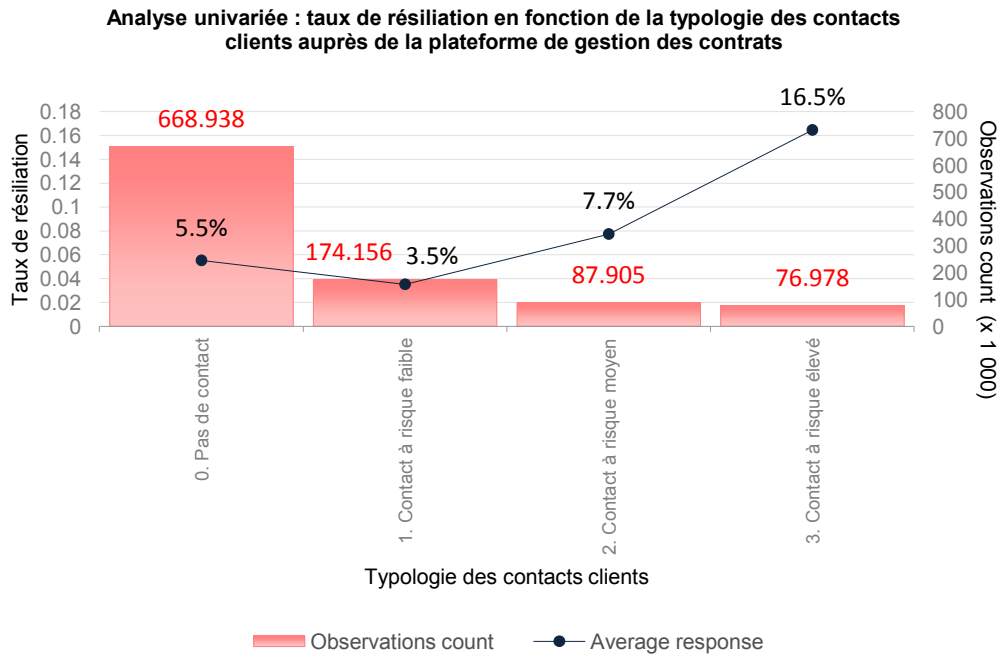


Figure 20 : Taux de résiliation en fonction de la typologie des contacts clients auprès de RCC

La variable typologie des contacts clients auprès de la plateforme de gestion de contrats est une variable catégorielle prenant quatre modalités (pas de contact, au moins 1 contact à risque faible, modéré et élevé). Cette variable indique si le client a contacté la plateforme RCC par téléphone, e-mail ou courrier dans une période encadrant l'échéance de son contrat et si ce contact est potentiellement risqué pour la suite de la relation avec l'assureur.

La typologie de contacts clients est une variable discriminante dans l'analyse du taux de résiliation de l'assuré. Ainsi plus un client contacte la plateforme de gestion pour un motif à risque élevé (informations sur le terme, demande de relevé d'information, explication comptable,..), plus son taux de résiliation progresse. Dans la base d'étude, environ un tiers des clients ont contactés la plateforme de gestion pour des questions liées à leurs contrats d'assurance.

- Données risque : le statut d'occupation

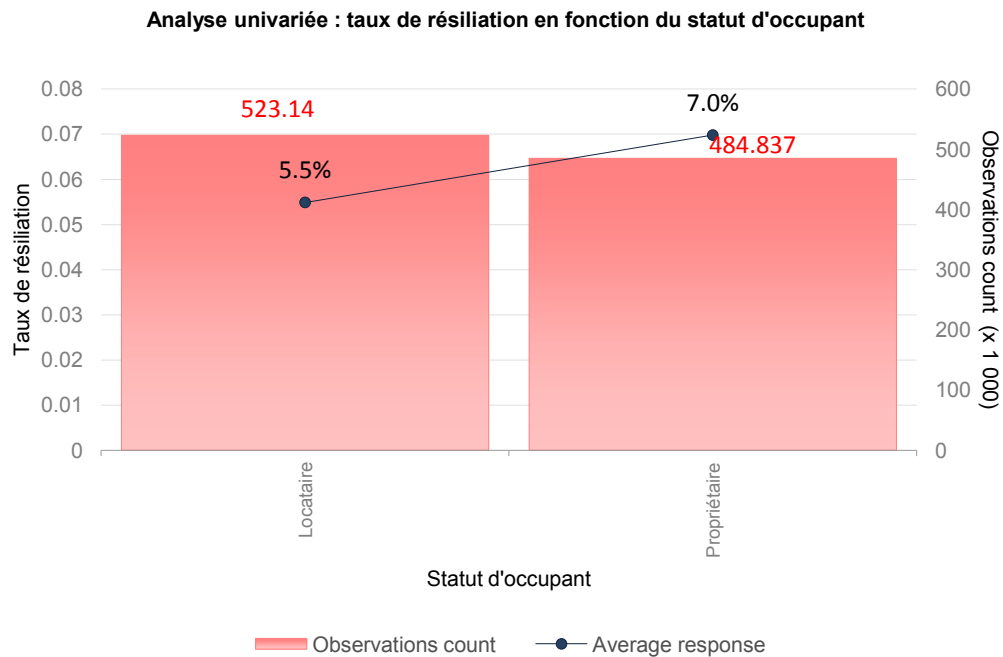


Figure 21 : Taux de résiliation en fonction du statut d'occupation

La variable étudiée correspond au statut de l'occupant, locataire ou propriétaire. La proportion des deux modalités est quasiment équivalente dans la base d'étude (51.9% de locataire et 48.1% de propriétaire).

Les propriétaires résilient davantage suite à une majoration que les locataires du probablement à leur niveau de prime supérieur (381€ contre 267€ pour les locataires). D'ailleurs l'analyse du V de Cramer (0.51) entre les variables statut d'occupant et prime payée par le client nous conforte dans cette hypothèse.

Test V de Cramer

Ce test mesure l'intensité de la liaison entre deux variables qualitatives à $K \geq 2$ modalités.

Le V de Cramer est le rapport entre la racine carrée du test du Khi-deux divisée par l'effectif multiplié par le plus petit côté du tableau (nombre de lignes L ou de colonnes C) moins 1.

$$V = \sqrt{\frac{\chi^2}{N \times \min(C, L) - 1}}$$

Une valeur 0 indique une absence de relation statistique entre les deux variables tandis qu'une valeur approchant 1 montre une colinéarité. Nous pouvons considérer qu'à partir d'une valeur de 0.30, la liaison entre les deux variables est forte.

Il est à noter que l'analyse du taux de résiliation total (incluant les avenants de changement ou disparation de risque) indique l'inverse, à savoir que les locataires résilient plus que les propriétaires.

- Données risque : présence d'un coefficient de surveillance du portefeuille

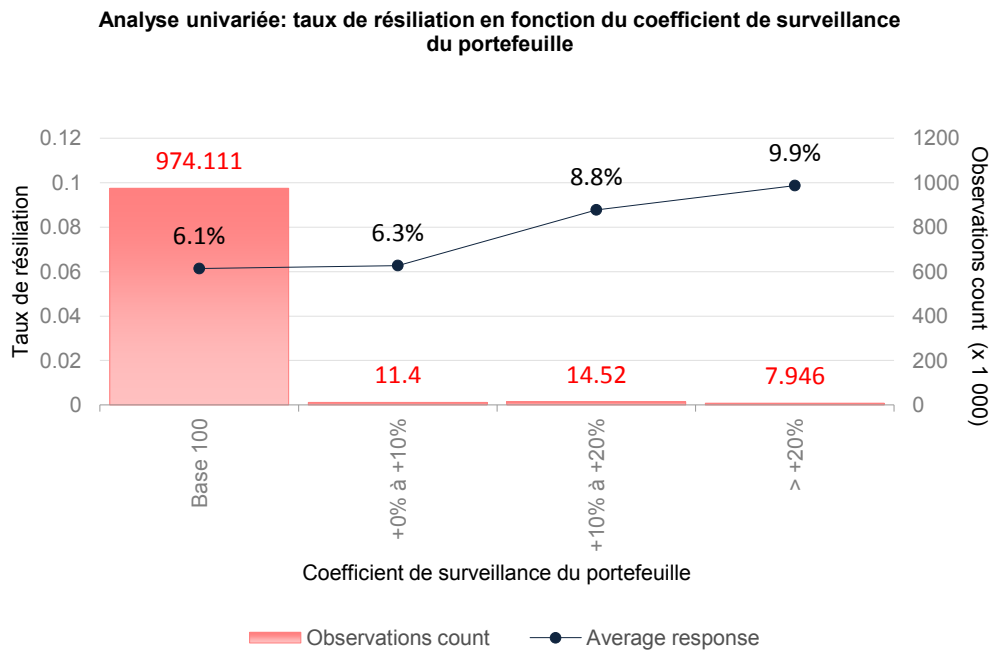


Figure 22 : Taux de résiliation en fonction du niveau du coefficient de surveillance du portefeuille

La présence d'un coefficient de surveillance du portefeuille est également un facteur discriminant pour le taux de résiliation. Une amplitude de près de 4 points est constatée sur le taux de résiliation suivant le niveau du coefficient de SDP. L'explication est la suivante : un client ayant eu plusieurs sinistres dans le passé verra sa prime augmentée fortement à l'échéance et aura donc tendance à résilier son contrat.

L'analyse du nombre de sinistres de l'assuré sur les 12 derniers mois présentée en annexe 5 conduit à la même conclusion.

- Données bancaires : la segmentation bancaire de la Société Générale

Analyse univariée : taux de résiliation en fonction de la segmentation bancaire

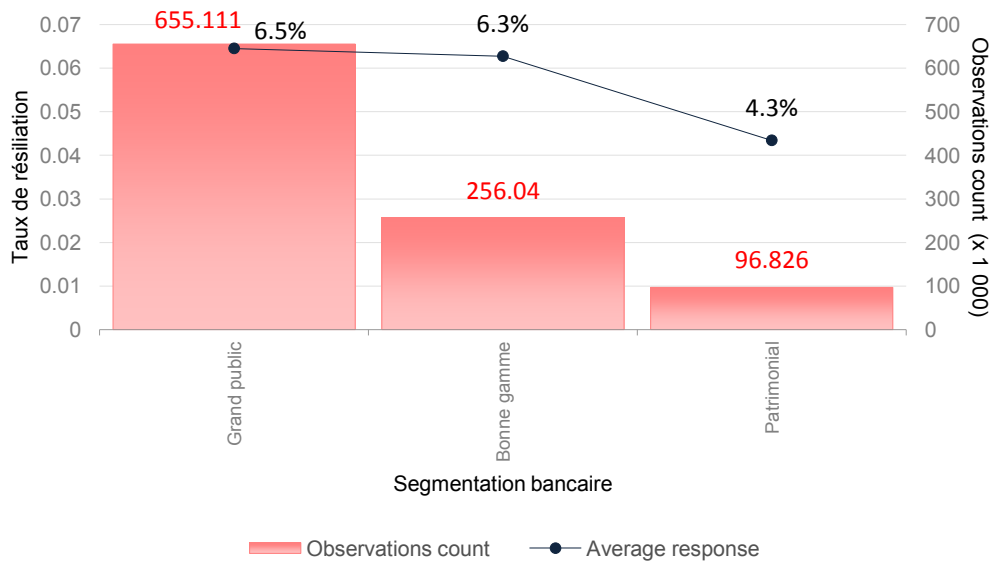


Figure 23 : Taux de résiliation en fonction de la segmentation bancaire

La première variable bancaire analysée est la segmentation client suivant les avoirs détenus au sein de la Société Générale. Le taux de résiliation diminue avec le niveau de ressources des clients : un client ayant des revenus contraints sera probablement plus regardant sur le montant de la prime d'assurance qu'une personne aisée.

- Données bancaires : la présence d'un prêt immobilier

Analyse univariée: taux de résiliation en fonction de la présence ou non d'un prêt immobilier

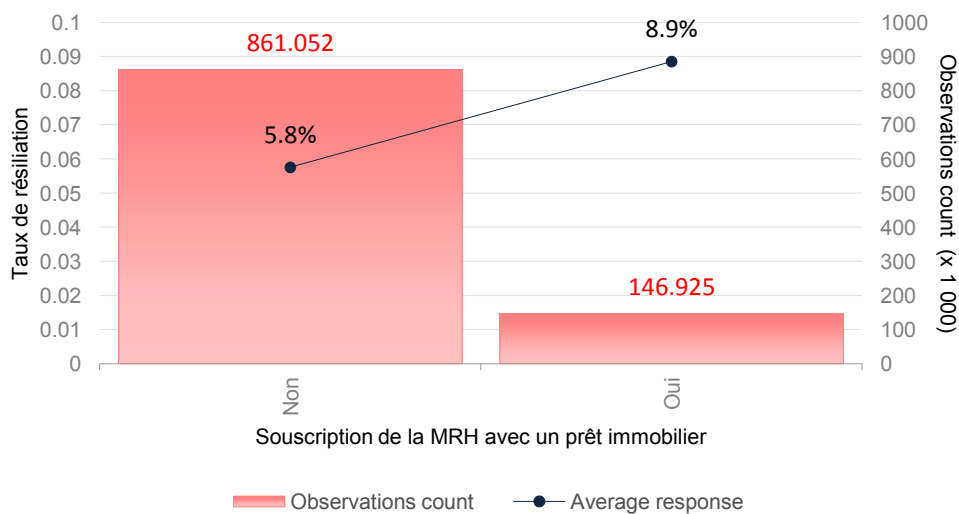


Figure 24 : Taux de résiliation en fonction de la souscription de l'assurance avec un prêt immobilier

La souscription conjointe d'un prêt immobilier (PPI) avec l'assurance habitation est également un indicateur intéressant à visualiser. En effet, un client ayant souscrit un prêt pour l'achat d'un bien immobilier se voit régulièrement proposer l'assurance habitation. Le client n'a pas forcément le temps de comparer les offres des assureurs, choisit l'offre proposée par le conseiller et réfléchit postérieurement sur les niveaux de couvertures et de prix associés à sa police. Dans ce cas, il aura une propension à résilier plus importante qu'un client n'ayant pas souscrit de PPI.

3. Analyse en correspondances multiples

Les techniques d'analyse de données sont des méthodes descriptives car elles ne supposent pas l'existence d'un modèle mathématique décrivant des liens entre les variables. La visualisation d'une base de données constituée de milliers de lignes et de dizaines de variables n'étant pas aisée, l'objectif de ces méthodes est de représenter graphiquement un nuage de n individus décrit par p variables qualitatives ou quantitatives. Il existe plusieurs techniques d'analyse de données (pour en savoir plus, se référer au livre de Gilbert Sapota mentionné dans le chapitre bibliographie) mais nous allons étudier ici l'analyse en correspondances multiples (ou ACM) qui est une extension de l'analyse factorielle des correspondances.

a. Rappels théoriques

Les rappels théoriques au sujet de l'analyse factorielle des correspondances (AFC) sont détaillés dans l'annexe 6.

L'analyse factorielle des correspondances ne permettant pas l'analyse de plus de deux variables, une extension de l'AFC est proposée dans le paragraphe suivant.

Analyse en correspondance multiples (ACM)

L'analyse en correspondances multiples peut être considérée comme une généralisation de l'AFC et permet d'analyser un nombre quelconque de variables. A partir d'un tableau qui croise les individus et les variables, nous allons substituer les variables par les modalités prises par lesdites variables.

En considérant n individus et M variables qualitatives et les notations suivantes : les variables $x^1, \dots, x^m, \dots, x^M$ et les modalités $X_1^m, \dots, X_j^m, \dots, X_{p_m}^m$ associées à la variable x^m , nous allons construire le tableau disjonctif complet Z :

$$Z_{(n,p)} = \left(\begin{array}{ccc|ccc|ccc} X_1^1, \dots, X_{p_1}^1 & X_1^2, \dots, X_{p_2}^2 & X_1^m, \dots, X_{p_m}^m \\ 0 \dots 1 \dots 0 & 0 \dots 1 \dots 0 & 0 \dots 0 \dots 1 \end{array} \right)$$

Sur chaque ligne, il y aura M valeurs 1 et $p - M$ valeurs 0 avec p le nombre de modalités toutes variables confondues, i.e. : $p = \sum_{m=1}^M p_m$. Il y aura donc autant de valeurs 1 qu'il y a de variables.

L'ACM consiste à effectuer une analyse factorielle des correspondances du tableau Z.

Nuage des individus

Chaque individu i est représenté par un vecteur de \mathbb{R}^p de composantes :

$$\frac{1}{\sqrt{\frac{f_j^m}{M}}} \frac{z_j^m(i)}{M}$$

Avec $z_j^m(i)$ un point du tableau Z égal à : $\left\{ \begin{array}{l} 1 \text{ si l'individu } i \text{ prend la modalité } X_j^m \text{ de } x^m \\ 0 \text{ sinon} \end{array} \right\}$ et

f_j^m la fréquence de la modalité X_j^m dans la population

Nous pouvons également écrire la distance entre deux individus :

$$d_{\chi^2}^2(i, i') = \sum_{n=1}^M \sum_{j=1}^{p_m} \frac{1}{M \cdot f_j^m} (z_j^m(i) - z_j^m(i'))^2$$

Le terme entre parenthèses est égal à 1 si X_j^m est prise par un et un seul des individus. Ainsi, un individu présentant une modalité rare est éloigné des autres individus.

Nuage des modalités

Chaque modalité X_j^m est représentée par un vecteur de \mathbb{R}^n :

$$\frac{1}{\sqrt{\frac{1}{n}}} \frac{z_j^m(i)}{n_j^m}$$

Avec n_j^m le nombre d'individus ayant la modalité X_j^m

Comme pour une analyse factorielle des correspondances, on obtient dans les plans factoriels l'ensemble des modalités des individus et des modalités des variables.

b. Application

L'ACM a été réalisée sous le logiciel R (via la fonction *ade4*), à partir de la base constituée dans le paragraphe II.1 et limitée à onze variables (cf. annexe 6 pour le détail des variables) pour conserver une certaine lisibilité dans la restitution graphique.

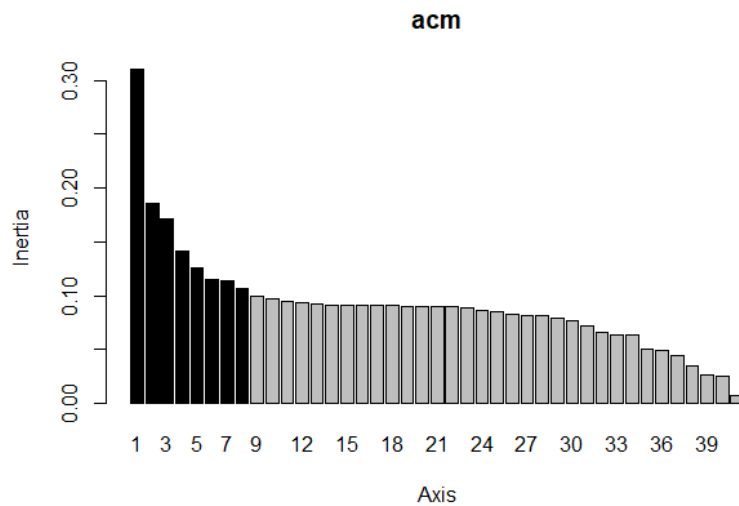


Figure 25 : Histogramme des valeurs propres de l'ACM

Le premier graphique présenté pour l'ACM correspond à l'histogramme des valeurs propres. Le nombre d'axes (abscisse de la figure 26) est égal à 41, correspondant à la somme des $p_m - 1$ modalités des onze variables sélectionnées.

La valeur propre de l'axe 1 représente 0.31 soit une partie d'inertie expliquée par cet axe de 8.3%. Pour l'axe 2, la valeur propre est égal à 0.19, soit une partie d'inertie expliquée par cet axe de 5.0%. Dans une ACM, la part d'inertie expliquée est en général faible (cf. annexe 6 pour visualiser l'inertie cumulée).

La figure 26 présente les modalités des variables dans le plan factoriel des axes 1 et 2.

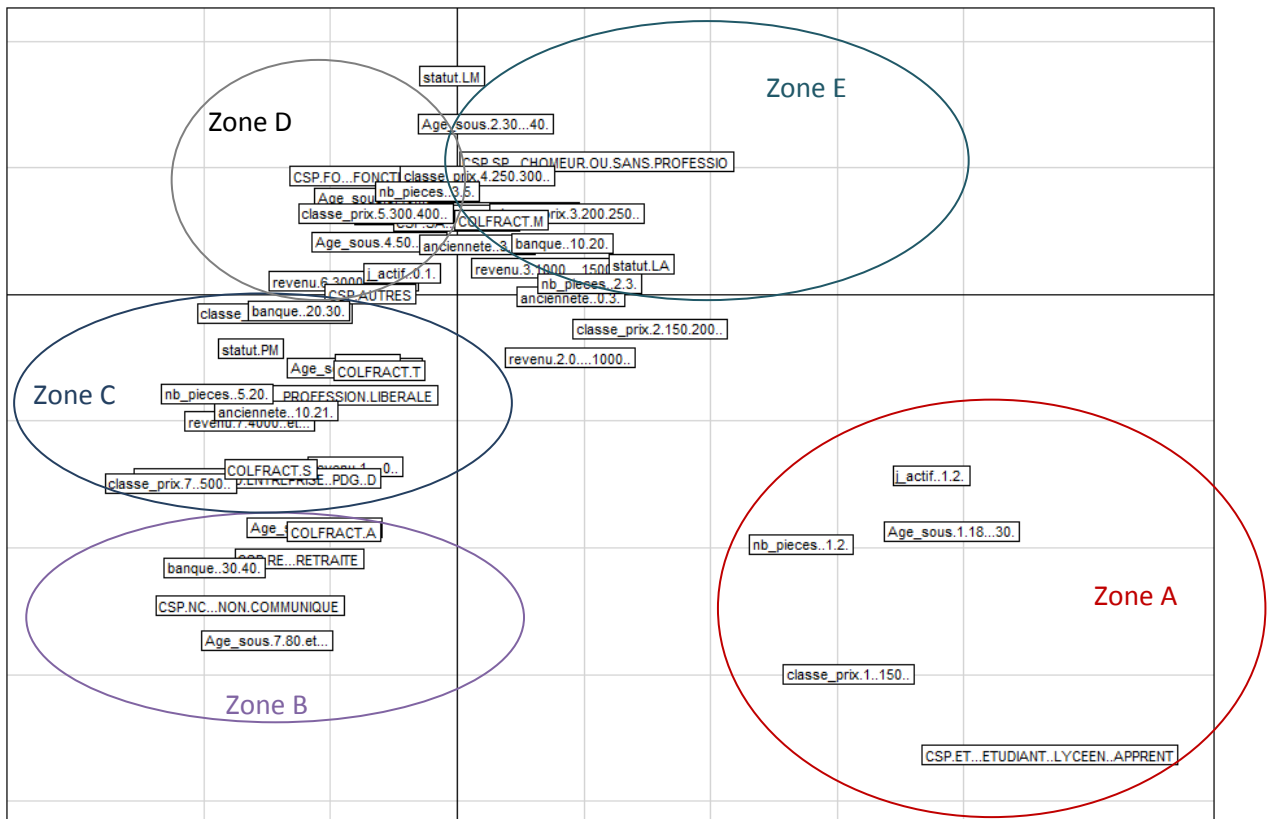


Figure 26 : Représentation des modalités des variables dans le plan factoriel des axes 1 et 2

Dans la zone A, les populations jeunes (Age_sous_18_30ans) ont une catégorie socio professionnelle étudiant ou apprenti (CSP_etudiant_lyceen_apprenti) ou sont qualifiés de jeunes actifs (j_actif_1_2) au sens de la banque (clients âgés de 18 à 29 ans, en activité professionnelle ou en recherche d'emploi mais n'ayant plus le statut d'étudiant). Ils habitent généralement dans des logements petits (nb_pieces_1_2) et présentent une prime d'assurance faible (classe_prix_inférieur_150€).

Dans la zone B, les clients âgés (Age_sous_80ans_et_+) et retraités (CSP_Retraité), dispose d'un patrimoine conséquent (banque_30_40 signifiant un client de type patrimonial) et paye leur prime en fractionnement annuel (COLFRACT_A).

Dans la zone C, les populations d'un certain âge (entre 60 et 70 ans) ont une catégorie socio professionnelle favorisée (CSP_Chef_d'entreprise_PDG et CSP_Profession_Libérale). Leur patrimoine relativement élevé (banque_20_30 signifiant un segment bonne gamme) leur permet d'être propriétaire de maison (statut_PM) de plus de 5 pièces (nb_pieces_5_et_+). En lien avec leur âge, leur ancienneté est plutôt importante (ancienneté_10ans_et_+) et ils payent en fractionnement semestriel ou trimestriel (COLFRACT_S_T) leur prime d'assurance de plus de 500€ (classe_prix_500_et_+).

Dans la zone D, des locataires d'appartement (statut_LA) âgés de 30 à 40 ans (Age_sous_30_40) vivent dans des logements de taille modeste (nb_pieces_2_3), n'ont pas de travail (CSP_Chômeur_sans_profession) et sont catégorisés grand public selon la terminologie de la banque

(banque_10_20). Ils payent leur prime d'assurance que mensuellement (COLFRACT_M) pour un montant compris entre 200 et 250€ (classe_prix_200_250€).

Dans la dernière zone, des populations d'âge compris entre 40 et 50 ans sont fonctionnaires (CSP_Fonctionnaire) et ne sont donc pas considérés comme jeunes actifs (j_actif_0_1). Ils habitent des logements de 3 à 4 pièces (nb_pieces_3_5) moyennant une cotisation d'assurance comprise entre 300 et 400€ (classe_prix_300_400€).

Le dernier graphique correspond au cercle de corrélation des axes 1 et 3.

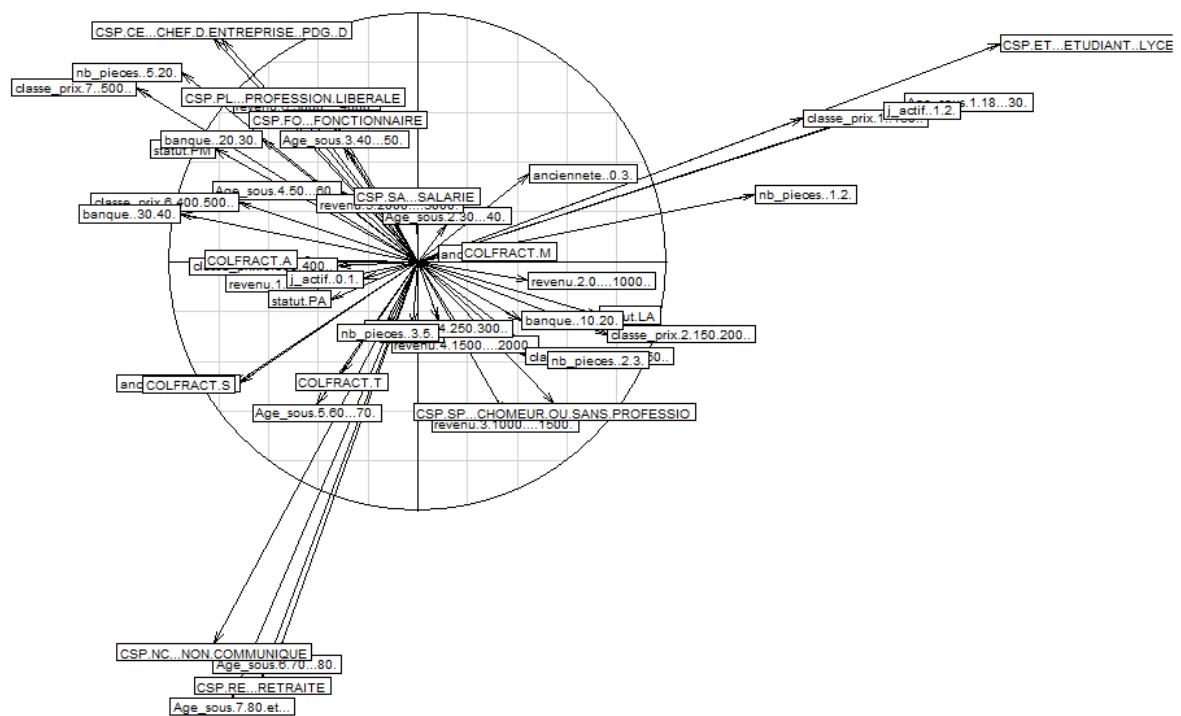


Figure 27 : Représentation du cercle des corrélations des axes 1 et 3

Le cercle de corrélations entre les axes 1 et 3 de l'ACM aboutit aux mêmes conclusions que lors de l'analyse des axes 1 et 2. Nous observons une confrontation des populations jeunes et âgées, des travailleurs et des personnes sans emploi ainsi que des personnes aisées avec une population aux ressources plus modestes. Tous ces effets tendent à mettre en exergue dans nos données un effet de classe sociale.

III. Modélisation de la résiliation et sensibilité au prix

1. Régression logistique

a. Quelques rappels théoriques

La variable à expliquer, la résiliation du contrat, est une variable binaire qui ne peut prendre que des valeurs 1 (le contrat est résilié) ou 0 (le contrat est renouvelé). Cette variable est modélisée à l'aide de prédicteurs qualitatifs ou quantitatifs. Il s'agit donc d'un apprentissage supervisé qui sera construit à partir d'une régression logistique.

Plusieurs types de modèles existent pour cette régression : les modèles logit, probit, tobit et complementary log-log model.

Nous utiliserons le modèle logit dans la mesure où sa capacité d'interprétation des paramètres β est simple et du fait d'une tendance à attribuer aux événements « extrêmes » une probabilité plus forte que la distribution normale.

On souhaite déterminer si y_i vaut 0 ou 1 à partir de variables explicatives. Pour ce faire, on écrit le modèle à variables latentes (inobservables) π_i^* suivant :

$$\pi_i^* = X_i * \beta + \varepsilon_i \text{ où } i \in [1, \dots, N] \text{ et } \varepsilon_i \text{ un résidu d'espérance nulle}$$

On passe ensuite de la variable latente π_i^* à la variable y_i par un seuillage :

$$y_i = \begin{cases} 1 & \text{si } \pi_i^* \geq \pi_0 \\ 0 & \text{sinon} \end{cases}$$

$\pi_i^* = P(y_i = 1 | X_i) = F(X_i * \beta)$ où F est la fonction de répartition des X_i :

$$F(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}$$

La fonction de répartition de la loi logit est représentée ci-dessous :

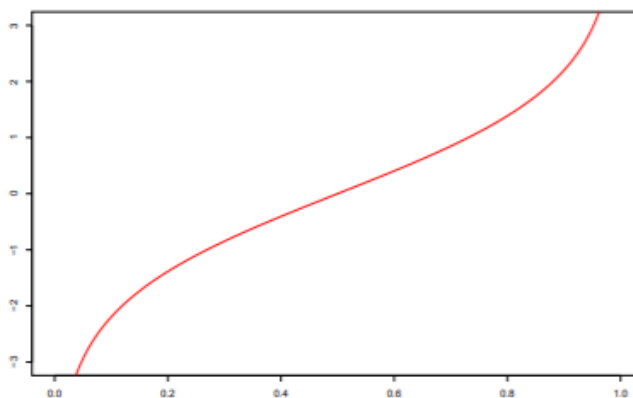


Figure 28 : Fonction de répartition de la loi logit

Ainsi, la variable latente peut s'écrire :

$$\pi_i^* = P(y_i = 1 | X_i) = F(X_i * \beta) = \frac{1}{1 + \exp(-X_i * \beta)}$$

$$\text{Et : } \frac{\pi_i^*}{1 - \pi_i^*} = \frac{\frac{1}{1 + \exp(-X_i * \beta)}}{1 - \frac{1}{1 + \exp(-X_i * \beta)}} = \frac{\frac{1}{1 + \exp(-X_i * \beta)}}{\frac{1 + \exp(-X_i * \beta) - 1}{1 + \exp(-X_i * \beta)}} = \frac{1}{\exp(-X_i * \beta)} = \exp(X_i * \beta)$$

En prenant le logarithme et en posant $\frac{\pi_i^*}{1 - \pi_i^*} = c_i$, on obtient :

$$\ln(c_i) = \ln(\exp(X_i * \beta)) = X_i * \beta \cong y_i$$

On peut interpréter cette égalité en indiquant que pour un individu i , il y a c_i fois plus de chances que y_i soit égal à 1 plutôt que 0. Ainsi, $\exp(X_i * \beta)$ exprime l'odds ratio ou rapport de cote.

b. Sélection des variables via une technique de *machine learning*

Théorie :

La base de données contenant plus de 100 variables, il est nécessaire de faire une première sélection des variables discriminantes à partir d'une technique de *machine learning*.

Dans le paragraphe ci-dessous, s'inspirant de l'ouvrage de Gilbert Saporta (cf. bibliographie), nous introduirons brièvement la théorie des arbres de décision.

Développées dans les années 1960, les arbres de régression et de décisions ont connu un regain d'intérêt dans les années 2000 avec les travaux de Breinman & al. (1984) et l'augmentation de la puissance de calcul. Par ailleurs, la lisibilité de leurs résultats leur confère un atout supplémentaire.

Lorsque la variable à expliquer y est une variable quantitative, on parlera d'arbres de régression et lorsque y est une variable qualitative, on utilisera la terminologie arbres de décision ou de classification. Dans la suite du paragraphe, on considèrera que y est une variable qualitative pour être cohérent avec l'application de cette technique de *machine learning* sur la prédiction de la résiliation du client.

Le principe de construction de l'arbre est de segmenter l'échantillon en deux sous-ensembles à l'aide des variables x^1, x^2, \dots, x^p . Le procédé est itéré à nouveau dans chaque sous-ensemble. L'objectif est de déterminer la meilleure partition des modalités des variables selon un critère d'explication de y .

Deux mesures d'impuretés ou fonctions d'hétérogénéité existent dans le cas où la variable à expliquer y est qualitative avec k modalités :

- L'entropie définie par : $\sum_{i=1}^k p_i \ln(p_i)$
- L'indice de diversité de Gini qui s'écrit : $\sum_{i=1}^k p_i(1 - p_i)$

L'objectif est d'obtenir la segmentation en deux sous-ensembles conduisant à l'impureté la plus faible.

Pour une variable à expliquer y à deux modalités et avec les proportions respectives p et $1 - p$, l'indice de Gini vaut $2p(1 - p)$ et a un comportement similaire à l'entropie.

y étant une variable qualitative, l'observation x sera classée dans le groupe ayant la plus forte fréquence en construisant un tableau de classement comme en discrimination.

Lors de la construction de l'arbre, il faut faire attention au phénomène de sur-apprentissage. Plus l'arbre est profond, plus le nombre de nœuds augmente. Si aucune limite n'est définie, les nœuds terminaux pourraient être réduits à une seule observation aboutissant à un taux d'erreur nul sur l'échantillon, l'arbre étant fidèle aux données d'apprentissage. Cependant, ce type d'arbre ne pourrait pas être réutilisé sur une autre base de données.

La méthode « CART », pour *Classification and Regression Trees*, consiste à construire l'arbre maximal en fixant un nombre minimal d'observations par nœud, puis à l'élaguer en utilisant une procédure de validation croisée ou un échantillon test.

L'objectif est de déterminer un sous-arbre T de l'arbre maximal T_{max} réalisant un bon compromis entre performance mesurée par le coût d'erreur $R(T)$ et sa complexité calculée par le nombre de segments terminaux $|T|$. Le critère d'élagage est défini par : $crit_{\alpha}(T) = R(T) + \alpha |T|$ où α est un paramètre de pénalisation.

A la fin de l'étape d'élagage, il faut choisir le sous-arbre minimisant l'erreur. Dans le cas de la validation croisée permettant d'évaluer la valeur de α , les données disponibles sont divisées en V morceaux. A l'étape i , on utilise tous les morceaux sauf le $i^{ème}$ pour construire un sous-ensemble. On réitère pour i variant de 1 à V et on choisit pour α la valeur i qui minimise la moyenne des coûts d'erreurs.

Mise en application :

La mise en pratique de l'arbre de décision a été construite à partir du logiciel R et en utilisant la fonction *rpart* du package portant le même nom sur la base d'apprentissage. La variable à expliquer est la résiliation client et nous limitons la profondeur de l'arbre à 8 avec un minimum de feuilles paramétré à 3 000 individus.

L'arbre de décision va ainsi nous permettre de déterminer les variables les plus discriminantes dans l'explication de la résiliation client.

Illustrons graphiquement les résultats obtenus sur l'arbre avec un modèle simplifié :

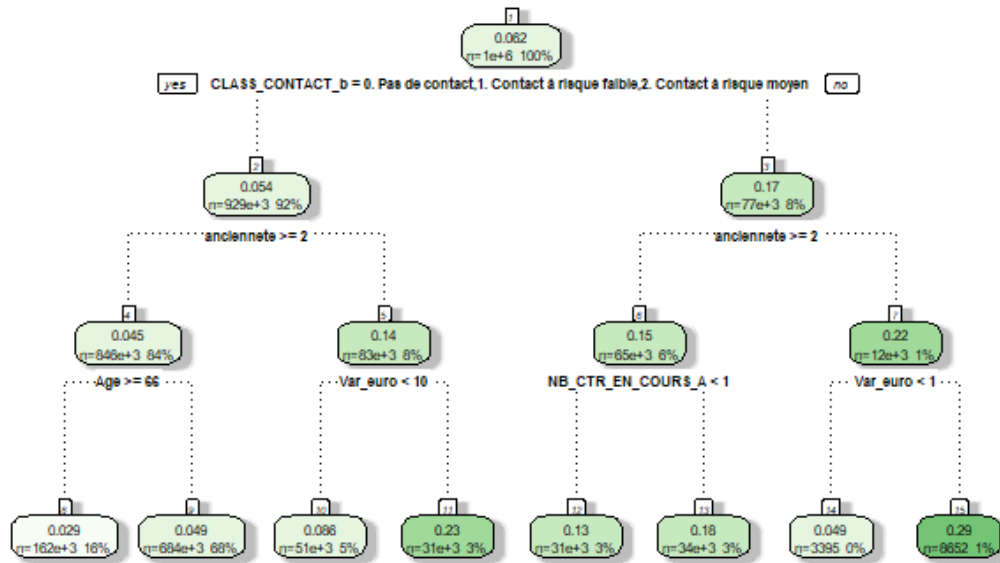


Figure 29 : Arbre de décision modèle simplifié

Le graphique ci-dessus présente un arbre de décision pour le modèle simplifié (limité à une profondeur maximal de 3).

Chaque nœud ou feuille présente les informations suivantes :

- Taux de résiliation (pour le nœud principal : 6.2%) ;
- Volume de contrats (soit 1 million d'occurrences) ;
- Pourcentage de contrats classé (100% pour le 1^{er} nœud).

Dans le modèle simplifié, les variables suivantes apparaissent comme discriminantes : typologie de contacts auprès de RCC, l'ancienneté du contrat, l'âge de l'assuré, le nombre de contrats en cours et la majoration en euro au terme.

Ainsi, le taux de résiliation le plus faible, 2.9% pour le nœud en bas à gauche avec 16% des contrats, est obtenu avec les conditions suivantes :

- Si la typologie de contacts avec la plateforme de gestion ne présente pas de risque élevé et
- Si l'ancienneté du contrat est supérieur ou égal à 2 ans et
- Si l'âge de l'assuré est supérieur ou égal à 66 ans.

A l'inverse, un taux de résiliation élevé, 29% pour le nœud en bas à droite et 1% du volume, est obtenu avec les modalités suivantes :

- Si la typologie de contacts avec la plateforme de gestion présente un risque élevé et
- Si l'ancienneté du contrat est inférieure à 2 ans et
- Si la majoration de prime à l'échéance a été supérieure ou égale à 1€.

Pour construire le modèle complet, nous appliquons la fonction *printcp* à l'arbre en sortie de la fonction *rpart*. Pour déterminer le bon niveau de simplification, nous allons procéder par validation croisée.

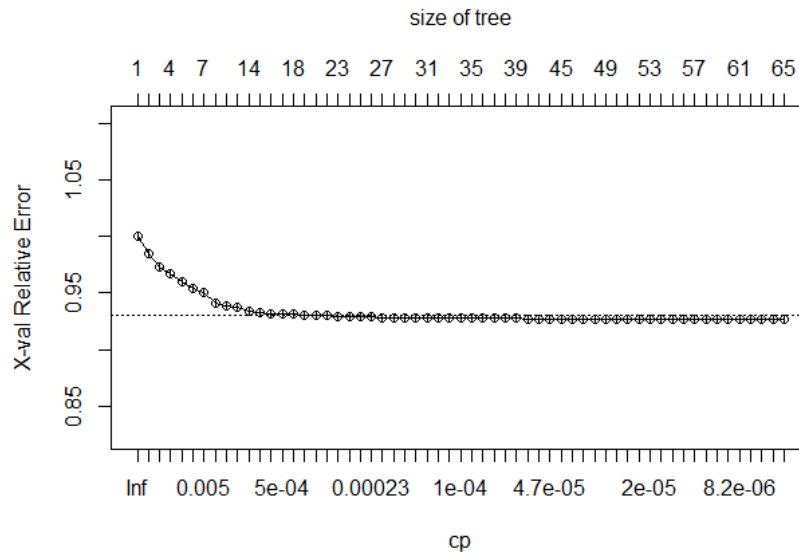


Figure 30 : Erreur commise en validation croisée en fonction de la taille de l'arbre

Le graphique 30 indique l'erreur commise en validation croisée suivant le nombre de feuilles. Plus la taille de l'arbre de l'arbre augmente, plus les performances s'améliorent. Lorsque le nombre de feuilles augmente trop, l'erreur augmente en raison du sur-apprentissage. Ici, l'erreur minimale est obtenue pour la valeur de cp égal à 3.44×10^{-6} correspondant à 61 feuilles. Afin d'élaguer l'arbre saturé, le paramètre cp , correspondant au paramètre de pénalisation α présenté précédemment, est ainsi modifié dans la fonction *rpart* avec cette nouvelle valeur.

Le modèle complet après élagage est disponible en annexe 7. En parallèle, un diagramme en barre présentant les variables les plus significatives est détaillé infra :

Classement des variables les plus discriminantes - Arbre de décision - modèle complet

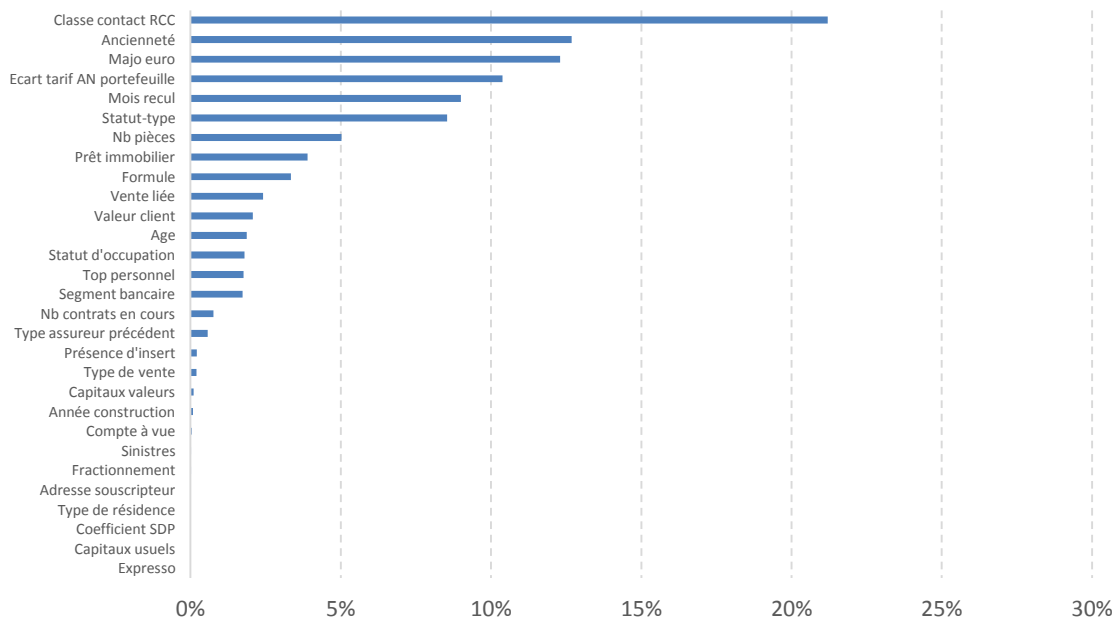


Figure 31 : Arbre de décision complet - classement des variables les plus significatives

A partir de l'arbre de décision complet, on aboutit aux variables précitées dans le modèle simplifié ainsi qu'aux données principales suivantes : écart de tarif entre affaires nouvelles et portefeuille, le mois de recul, le statut-type, le nombre de pièces, la valeur client et la présence d'un prêt immobilier.

Par ailleurs, certaines données bancaires apparaissent comme peu significatives comme l'équipement du client en compte à vue ou prêt consommation. La notion de vente combinée comprenant une information plus générique sera donc préférée.

c. Matrice de corrélation

Grâce au logiciel *R* et avec la fonction *corrplot* du package susnommé, nous avons calculé la matrice de corrélation.

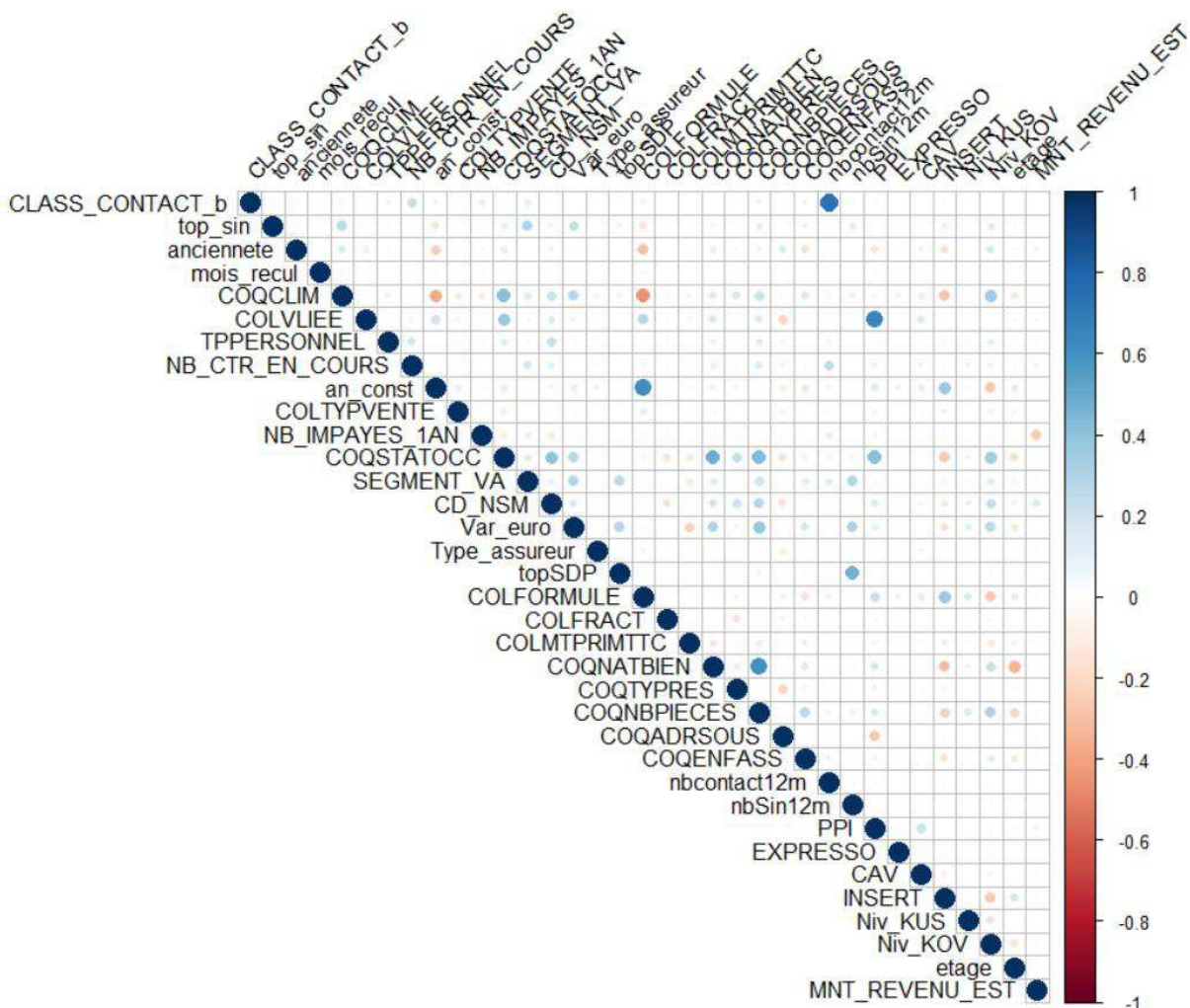


Figure 32 : Matrice de corrélation de la base d'étude

Les variables les plus corrélées sont mises en exergues ci-après :

- La typologie de contact auprès de la plateforme et le nombre de contacts auprès de cette même plateforme ;
- L'indicateur vente combinée d'une assurance et d'un prêt immobilier (variable PPI) ;
- L'année de construction du bien et la formule de garanties : on peut supposer qu'un occupant d'un logement récent sera davantage enclin à prendre une formule haut de gamme que s'il possède un bien ancien (sur le même principe qu'en automobile avec l'année de commercialisation du véhicule) ;
- La nature du bien (maison ou appartement) et le nombre de pièces : une maison est, en moyenne, plus grande en terme de superficie comparée à un appartement.

d. Modèle de résiliation

Le modèle de résiliation est développé à partir de la base d'étude construite dans le paragraphe II.1. Pour vérifier que notre modèle ne sur-apprend pas, nous découpons notre table en deux bases : une base d'apprentissage contenant 80% de l'échantillon sélectionné aléatoirement et une base de validation comprenant les 20% de lignes restantes.

Significativité et sélection des variables

Le modèle le plus performant n'est pas le modèle contenant le plus de variables explicatives mais celui qui est un bon compromis entre la qualité de l'ajustement et sa complexité. Le critère retenu doit diminuer lorsque des variables améliorant la qualité du modèle sont introduites et il doit diminuer lorsque des variables n'apportant aucunes informations supplémentaires sont ajoutées. Le paragraphe suivant définit deux critères de vraisemblance pénalisée :

- Le critère d'AIC d'Akaïké s'écrit : $-2 \ln L(\hat{\theta}) + 2p$ où $\ln L(\hat{\theta})$ est la log-vraisemblance et p le nombre de paramètres du modèle ($p - 1$ étant le nombre de variables explicatives). Le meilleur modèle est celui qui minimise le critère AIC ;
- Le critère BIC de Schwartz vaut : $-2 \ln L(\hat{\theta}) + \ln(n) \times p$ où $\ln L(\hat{\theta})$ est la log-vraisemblance, n le nombre d'observations et p le nombre de paramètres du modèle. Le meilleur modèle est celui qui minimise le critère BIC. Il est à noter que le critère BIC dépend du nombre d'observations donc il privilégiera des modèles avec moins de variables explicatives pour des grands échantillons par rapport au critère AIC.

Concernant le procédé de sélection des variables, plusieurs méthodes dites « pas à pas » existent. Elles sont détaillées succinctement dans le paragraphe suivant :

- Sélection *Forward* ou méthode ascendante : le modèle ne contient aucune variable et le test consiste à ajouter les variables une à une en choisissant celle qui minimise le critère retenu

(AIC ou BIC). La procédure est itérée avec une nouvelle variable et la méthode s'arrête lorsque plus aucune variable ne permet d'améliorer la régression ;

- Sélection *Backward* ou méthode descendante : il s'agit du procédé inverse. L'algorithme commence avec l'ensemble des variables disponibles et élimine la variable la moins significative au regard du critère mentionné plus haut. La procédure est répétée tant que le modèle peut être amélioré. Il est à noter que l'implémentation des méthodes Forward et Backward peut également être choisie avec le critère du test de Wald. Dans ce cas, l'algorithme se termine lorsque les variables sont supérieures à un seuil de significativité (par exemple 5%) ;
- Sélection *Stepwise* : cet algorithme est un perfectionnement de la première méthode dite ascendante. En plus des itérations, elle consiste à effectuer des test de significativité pour ne pas introduire des variables non significatives et éventuellement éliminer des variables introduites précédemment n'apportant plus d'informations compte tenu de la dernière variable sélectionné. La méthode se termine lorsque le modèle ne peut plus ajouter ni supprimer des variables. C'est cette dernière méthode qui sera utilisée dans ce mémoire.

Classement des variables les plus discriminantes - Régression logistique

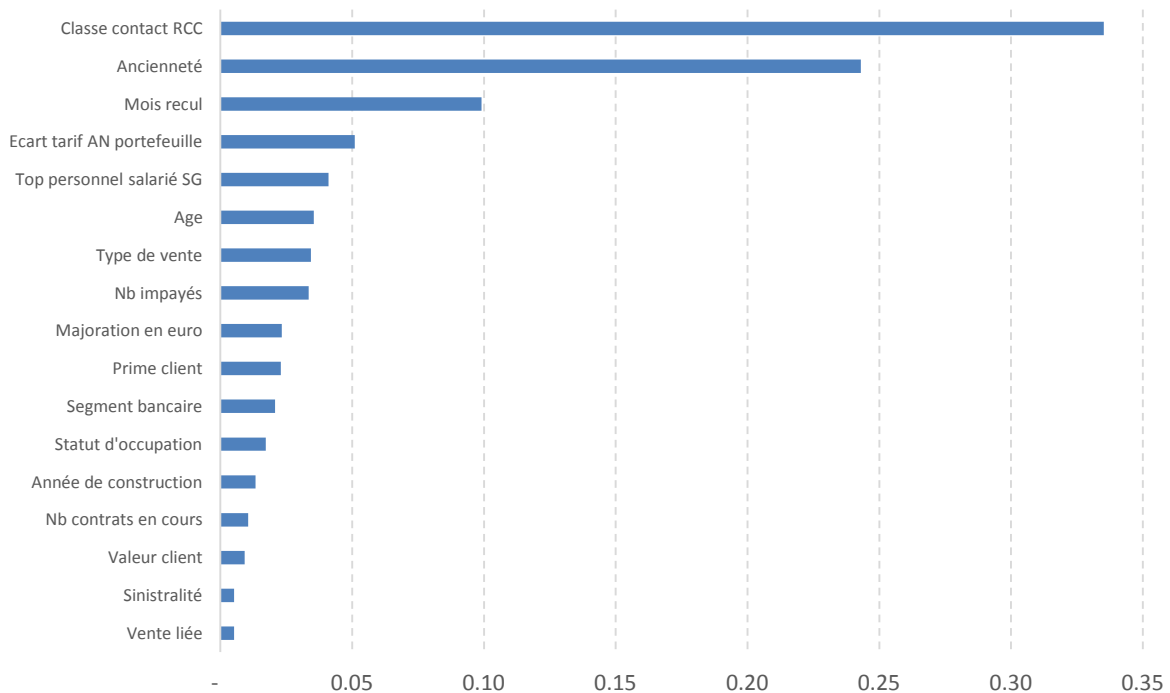


Figure 33 : Régression logistique - classement des variables les plus significatives

Les variables significatives obtenues par le modèle de régression logistique sont similaires à celles obtenues pour l'arbre de décision. Leur classement en figure 34 est déterminé par le rapport du khi-deux de chaque variable sur la somme totale des khi-deux.

La significativité du modèle est également assurée puisque les p_value de chaque variable sont significatives au seuil de 5%.

Criteria	Degrees of freedom	Chi-2	Pr > Chi-2
Classe contact RCC	3	11 652	-
Ancienneté	4	8 445	-
Mois recul	10	3 444	-
Ecart tarif AN portefeuille	9	1 773	-
Top personnel salarié SG	1	1 426	-
Age	5	1 231	0,00
Type de vente	1	1 192	0,00
Nb impayés	2	1 163	0,00
Majoration en euro	3	810	0,00
Prime client	8	795	0,00
Segment bancaire	2	721	0,00
Statut d'occupation	1	596	0,00
Année de construction	3	464	0,00
Nb contrats en cours	2	365	0,00
Valeur client	5	318	0,00
Sinistralité	5	182	0,00
Vente liée	1	180	0,00

Figure 34 : Significativité des variables du modèle logistique

Résultats du modèle de régression logistique

Le modèle de régression logistique a été réalisé sous *Addactis Pricing* module *behavior*.

L'histogramme des prédictions des taux de résiliations sur les bases d'apprentissage et de validation est présenté dans la figure 35.

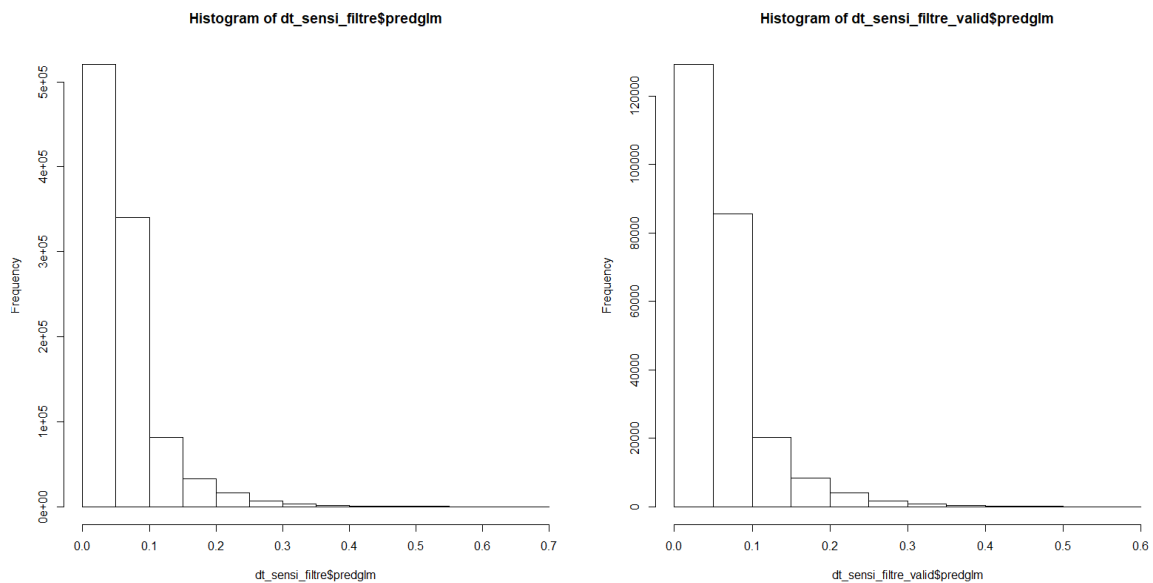


Figure 35 : Histogramme des prédictions des taux de résiliations sur la base d'apprentissage et de validation

La distribution des taux de résiliations est tout à fait comparable entre les deux bases d'apprentissage et de test. Ainsi, la moyenne des taux de résiliations modélisés obtenus sur les bases d'apprentissage et de validation est de respectivement 6.24% et 6.23%, cohérent avec le taux observé sur la base à 6.22% (cf. chapitre II.2.b).

Les graphiques suivants présentent quelques résultats obtenus avec la régression logistique.

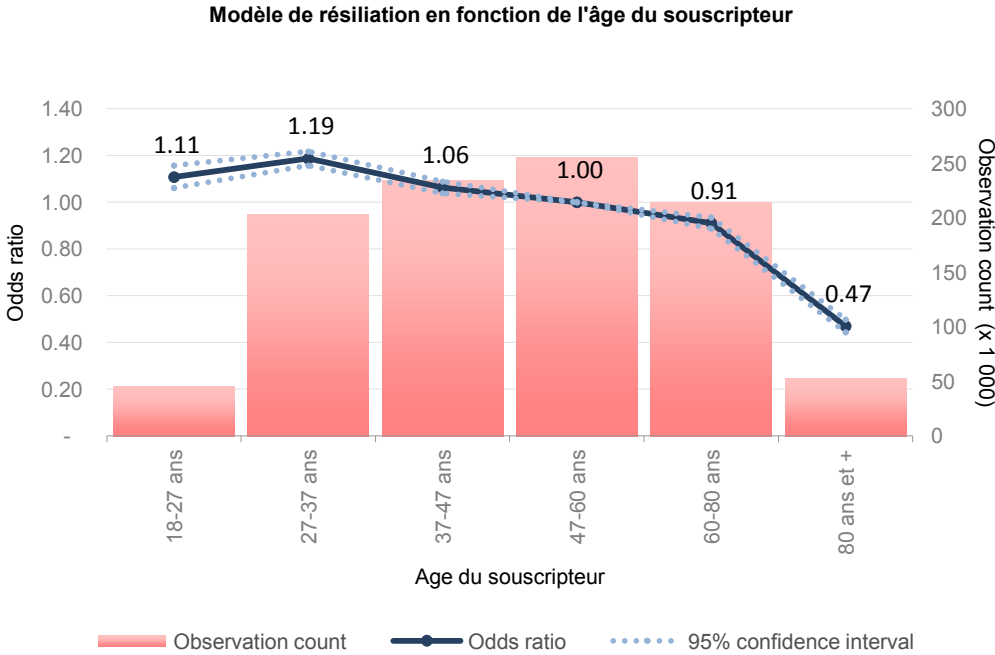


Figure 36 : Modèle logistique de résiliation suivant l'âge du souscripteur

La résiliation, toutes choses égales par ailleurs, diminue avec l'augmentation de l'âge du souscripteur. Un assuré d'âge 80 ans aura 53% de chances de moins de résilier qu'un client âgé entre 47 et 60 ans. Les courbes en pointillé représentent l'intervalle de confiance à 95% de l'odds ratio.

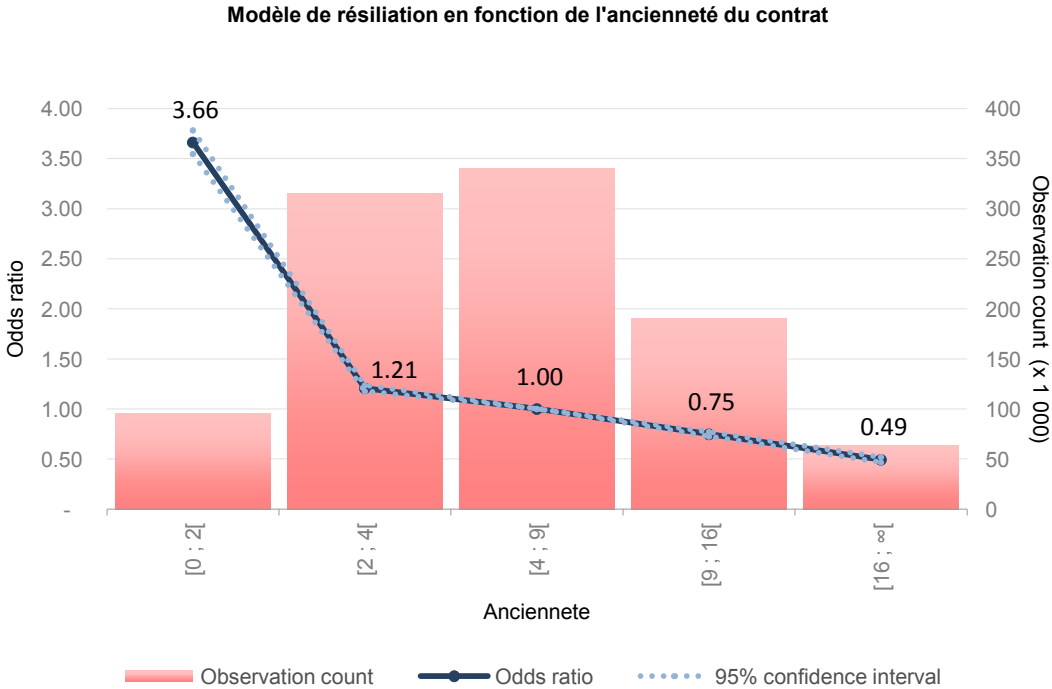


Figure 37 : Modèle logistique de résiliation suivant l'ancienneté

La résiliation, toutes choses égales par ailleurs, diminue plus le contrat est ancien en portefeuille. Un contrat en portefeuille depuis 10 ans aura ainsi 25% de chances de moins de résilier qu'un contrat plus récent de 5 ans.

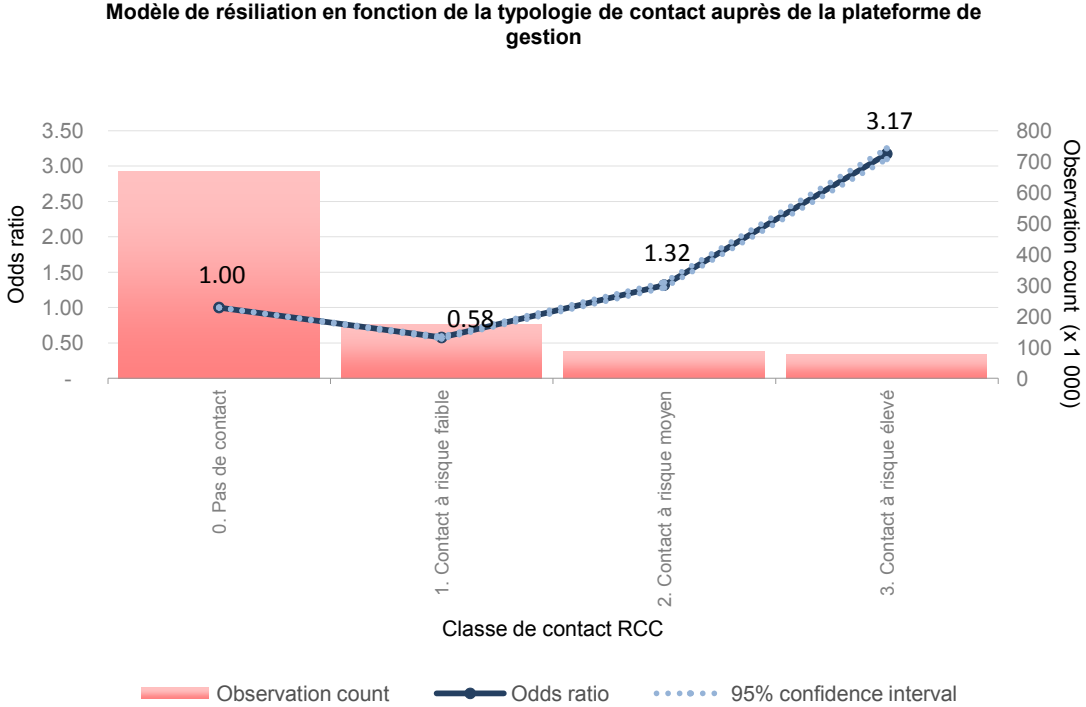


Figure 38 : Modèle logistique de résiliation suivant la typologie de contact RCC

La résiliation, toutes choses égales par ailleurs, augmente si le client a contacté pour un motif de risque élevé la plateforme de gestion de contrats appelée RCC. Ainsi, un contrat avec des contacts à risque élevé (informations sur le terme, demande de relevé d'information, explication comptable,..) aura 3 fois plus de chances de résilier qu'un client n'ayant pas eu de contact avec le service client.

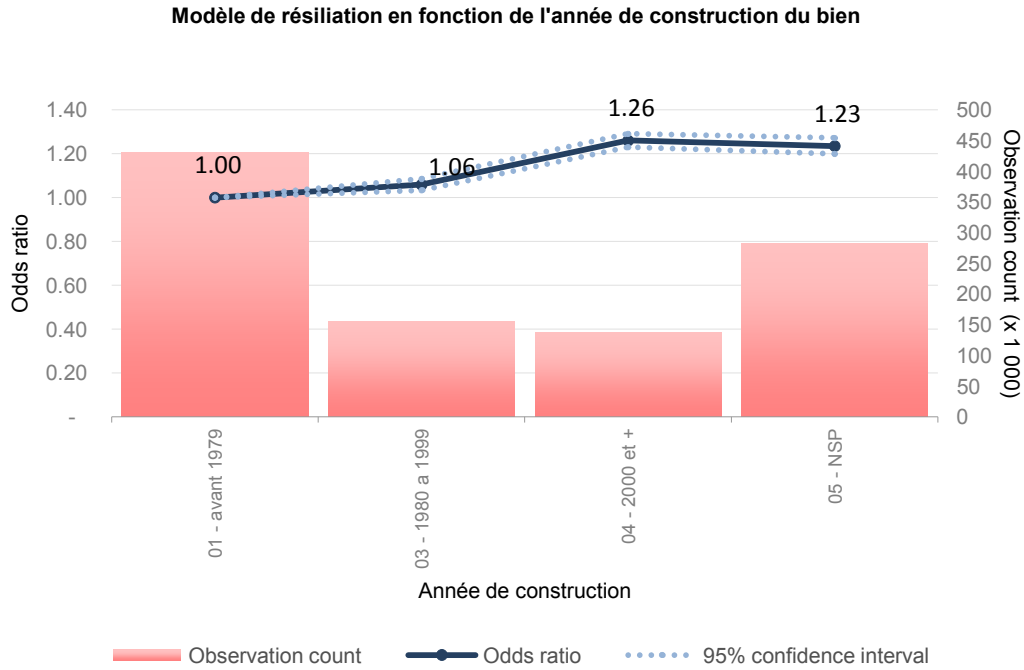


Figure 39 : Modèle logistique de résiliation suivant l'année de construction du bien

La résiliation, toutes choses égales par ailleurs, diminue avec la vétusté du bien. Un contrat correspondant à une habitation récente construite dans les années 2000 aura ainsi 26% de chances de résilier en plus qu'un contrat portant sur un logement des années 1970.

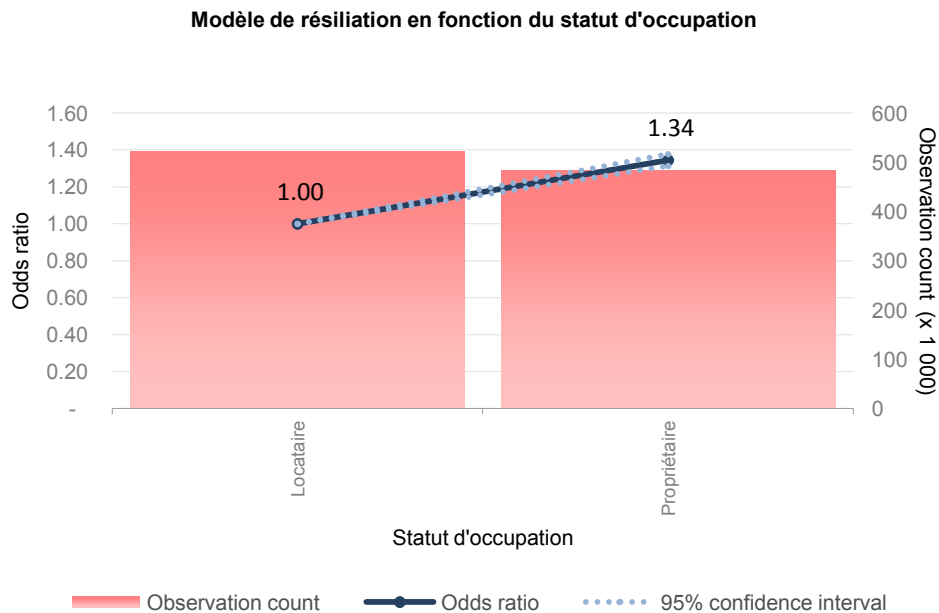


Figure 40 : Modèle logistique de résiliation suivant le statut d'occupation

La résiliation, toutes choses égales par ailleurs, augmente si l'occupant est propriétaire. Ainsi, un propriétaire aura 34% de résilier en plus qu'un locataire.

Validation du modèle

La représentation graphique du test de *Hosmer et Lemeshow* permet de mesurer l'adéquation entre probabilités estimées et probabilités théoriques. La statistique H de ce test est comparée à une distribution du khi-deux à $g - 2$ degrés de libertés où g représente le nombre de groupes :

$$H = \sum_{j=1}^g \frac{(O_j - N_j\pi_j)^2}{N_j\pi_j(1 - \pi_j)}$$

Avec :

- N_j le nombre de d'observations du $j^{\text{ème}}$ groupe ;
- O_j le nombre de résiliations observées du $j^{\text{ème}}$ groupe ;
- π_j la moyenne des probabilités estimées dans le $j^{\text{ème}}$ groupe.

Le graphe de la figure 41 présente donc :

- En abscisses : les moyennes des probabilités estimées par la régression logistique et triées par ordre croissant puis découpées en 50 classes d'amplitudes égales ;
- En ordonnées : la moyenne du taux de résiliation calculée sur les données de chaque classe.

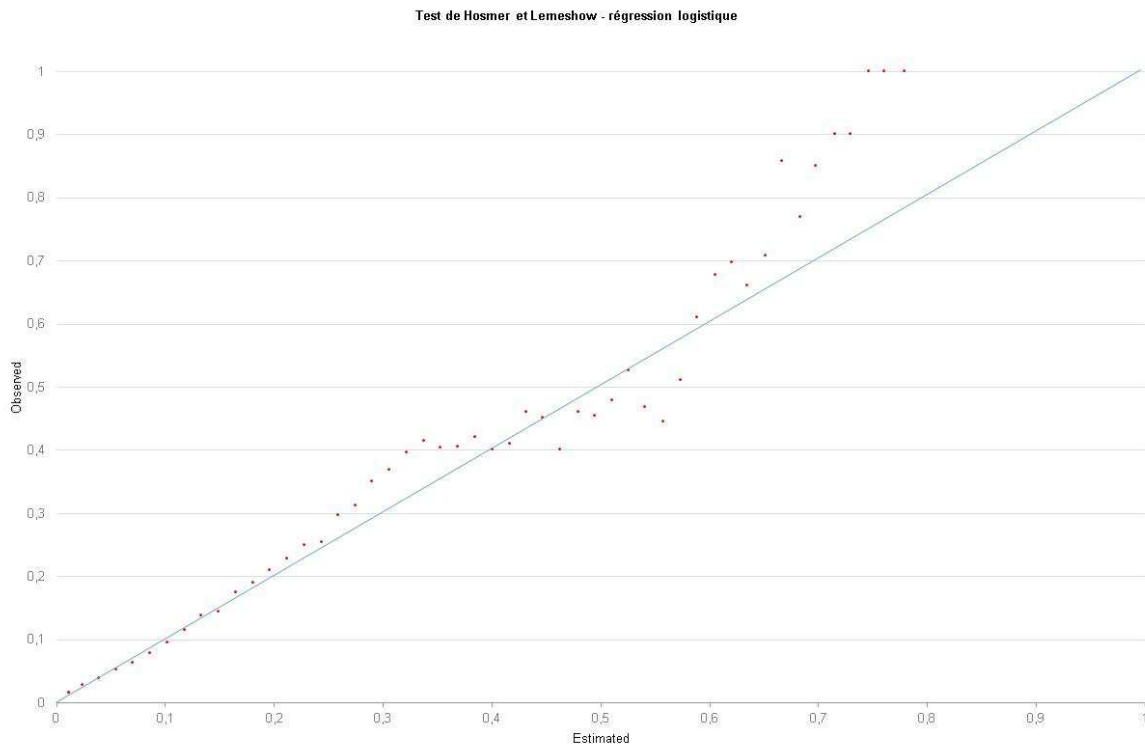


Figure 41 : Test de Hosmer et Lemeshow

Les points étant quasiment situés sur la bissectrice, le test de Hosmer et Lemeshow indique que notre modélisation est adéquate.

Par ailleurs, le pouvoir discriminant du modèle a été étudié en calculant l'aire sous la courbe ROC (*Receiver Operating Characteristic*) de la régression logistique.

La courbe ROC est un outil graphique qui permet de comparer plusieurs modèles de résiliations entre eux.

La démarche de construction de la courbe ROC est détaillée ci-dessous :

1. Pour chaque observation de la base, le modèle logistique donne une probabilité de résiliation estimée ;
2. Ligne à ligne, comparaison de la réponse estimée avec celle observée dans la base. Définition d'un seuil compris entre]0 ; 1[. Si la réponse estimée est supérieure au seuil, on note 1 sinon on inscrit 0 ;
3. La matrice de confusion est définie ci-dessous :

		Valeur prédite	
		0 (non résiliation)	1 (résiliation)
Valeur observée	0 (non résiliation)	Vrai négatif (VN)	Faux positif (FP)
	1 (résiliation)	Faux négatif (FN)	Vrai positif (VP)

Figure 42 : Matrice de confusion

Les valeurs définies au 2) sont reportées dans le tableau ci-dessus et plusieurs mesures y sont associées (la présentation n'est pas exhaustive et d'autres indicateurs existent) :

- La sensibilité ou le rappel est la fréquence des vrais positifs et se calcule comme suit : $\frac{VP}{VP+FN}$
 - La spécificité est la fréquence des vrais négatifs et se définit comme : $\frac{VN}{VN+FP}$
 - La précision ou valeur prédictive positive se détermine de la manière suivante : $\frac{VP}{VP+FP}$
 - Enfin, le taux d'erreur correspond aux valeurs mal prédites, i.e. à la formule : $\frac{FN+FP}{VN+FN+VP+FP}$
4. Sur un graphe est représenté, en abscisse, la valeur associée à 1 – spécificité et, en ordonnée, la sensibilité ;
 5. En faisant varier le seuil dans l'intervalle]0 ;1[, pour chaque valeur de seuil, on réitère les décisions prises aux étapes 2 à 4.

Le graphique de la figure 43 est ainsi obtenu :

Modèle de résiliation - base d'apprentissage - courbe ROC

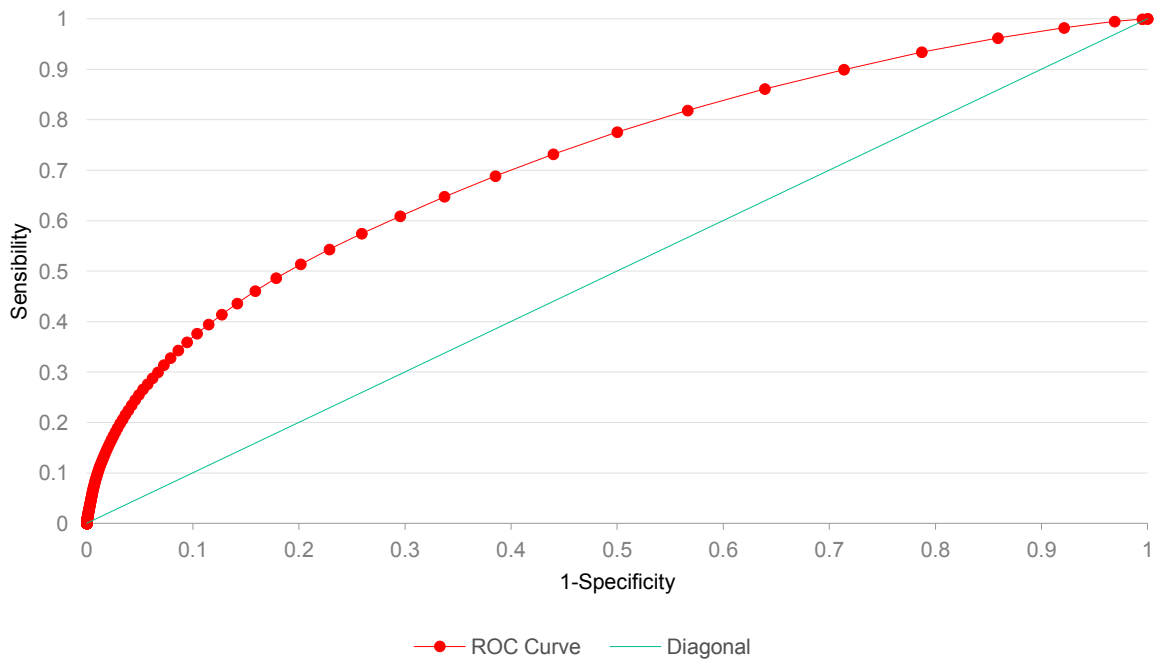


Figure 43 : Courbe ROC du modèle de régression logistique

La courbe rouge du modèle se situe au-dessus de la bissectrice verte représentant $y=x$. Pour comparer deux prédictions, l'aire sous la courbe (indicateur *AUC* pour *Area Under Curve*) est calculée. Sur la base d'apprentissage, une aire de 0.718 est ainsi déterminée et nous conduit à qualifier la discrimination du modèle de bonne

Le modèle de régression logistique a été testé sur la base de validation et une aire sous la courbe ROC de 0.715 est calculée, soit un niveau similaire à celle obtenue pour la construction du modèle. La représentation graphique des courbes ROC entre les bases d'apprentissage et de validation est disponible en figure 44.

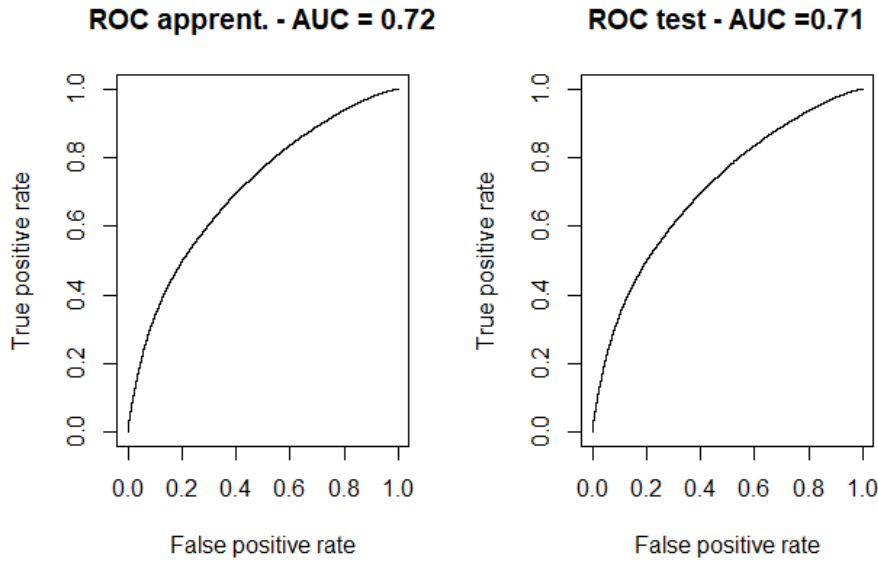


Figure 44 : Comparaison des courbes ROC de la régression logistique entre apprentissage et validation

2. Modèle de survie

Dans cette partie, une approche basée sur l'estimation de la durée de vie des contrats Habitation en portefeuille est utilisée pour modéliser le taux de résiliation. Les modèles de survie permettent d'estimer la probabilité de survie d'un contrat en portefeuille à chaque date t .

a. Théorie

Introduction :

Soit la variable aléatoire de durée de vie T à valeurs dans $[0, +\infty[$, F sa fonction de répartition définie par $F(t) = P(T \leq t)$ et f la densité. Définissons les fonctions caractérisant la distribution de la variable T :

- La fonction de survie est le complémentaire de la fonction de répartition et se définit par :

$$S(t) = 1 - F(t) = P(T > t)$$

S est une fonction continue monotone décroissante telle que : $S(0) = 1$ et $\lim_{t \rightarrow \infty} S(t) = 0$

- La survie conditionnelle S_u correspond à la durée de vie d'un individu après une date $u + t$ sachant qu'il a été en vie jusqu'en u . En posant $S_u(t) = P(T > u + t | T > u)$ et en utilisant la définition de la probabilité conditionnelle, on obtient :

$$S_u(t) = \frac{P(T > u + t)}{P(T > u)} = \frac{S(u + t)}{S(u)}$$

- La fonction de hasard λ ou fonction de risque correspond à un taux de panne ou à un taux de risque instantané. Elle s'écrit comme le rapport entre la densité et la fonction de survie :

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln(S(t))$$

car $F(t) = \int_0^t f(u)du$ et en prenant sa dérivée : $F'(t) = f(t) = -S'(t)$

La fonction de hasard λ indique la probabilité que la durée de vie soit comprise entre t et $t + u$ sachant qu'elle est plus grande que t .

Modèles à hasards proportionnels :

La distribution de survie n'étant pas connue a priori, le choix de modélisation a été porté sur des modèles semi-paramétriques. Les modèles à hasards proportionnels sont des modèles semi-paramétriques à structure multiplicative, ce qui se traduit par un effet multiplicatif des différentes covariables sur la durée de vie. Ils se caractérisent de la manière suivante :

$$\forall t > 0, \lambda(t|Z) = \lambda_0(t) \times h(\beta, Z) \text{ où}$$

- λ_0 est la fonction de hasard de base ;
- h une fonction positive ;
- Z le vecteur des covariables de longueur p ;
- β le vecteur des paramètres.

Notons que le deuxième facteur du produit ne dépend pas du temps contrairement à λ_0 . On peut considérer que l'effet des covariables se simplifie en une quantité $\beta'Z$.

Par ailleurs, en considérant deux individus i et j pour covariables respectives Z_i et Z_j , les fonctions de hasard sont proportionnelles puisque le rapport des fonctions λ ne dépend pas du temps :

$$\frac{\lambda(t|Z_i)}{\lambda(t|Z_j)} = \frac{h(\beta'Z_i)}{h(\beta'Z_j)}$$

Modèle de Cox :

Le modèle de Cox est donc un cas particulier des modèles à hasards proportionnels en considérant la fonction h comme une fonction exponentielle. Ainsi, le modèle de Cox s'écrit :

$$\lambda(t|Z) = \lambda_0(t) \times \exp(\beta'Z)$$

Le choix de la fonction exponentielle n'est pas un hasard dans la mesure où il s'agit d'une fonction à valeurs dans R_+^* et sachant que $\exp(0) = 1$. λ_0 ayant une forme inconnue, le modèle est appelé semi-paramétrique.

Estimation des paramètres :

Soit les notations suivantes :

- R le nombre de résiliations observées parmi les n clients ;
- $T_1 < T_2 < \dots < T_R$ les temps de résiliations distincts ;
- $(1), (2), \dots, (R)$ les indices des clients ayant résiliés respectivement en T_1, \dots, T_R ;
- Z_i la valeur des covariables de l'individu i ;
- $R(T_i)$ l'ensemble des individus encore à risque à T_i^- (juste avant T_i)

Pour le modèle de Cox, la méthode d'estimation des paramètres est la vraisemblance partielle. Il s'agit d'estimer uniquement le coefficient de régression β en supposant que la fonction de hasard de base λ_0 est un paramètre de nuisance. Ainsi, la fonction λ_0 ne sera pas estimée dans le modèle de Cox.

En prenant l'hypothèse qu'il n'y a qu'une seule résiliation à chaque intervalle, la probabilité qu'il y ait une résiliation à la date T_i s'écrit :

$$\sum_{j \in R(T_i)} \lambda_0(T_i) \times \exp(\beta' Z_j)$$

La probabilité que l'individu i connaisse une résiliation en T_i sachant qu'une résiliation a eu lieu à cette même date se traduit par :

$$\frac{\lambda_0(T_i) \times \exp(\beta' Z_{(i)})}{\sum_{j \in R(T_i)} \lambda_0(T_i) \times \exp(\beta' Z_j)} = \frac{\exp(\beta' Z_{(i)})}{\sum_{j \in R(T_i)} \exp(\beta' Z_j)}$$

Compte-tenu des contributions à la vraisemblance à chaque temps de résiliation, la vraisemblance partielle est définie comme le produit sur les temps de résiliations :

$$L_{Cox}(\beta) = \prod_{i=1}^R \left[\frac{\exp(\beta' Z_{(i)})}{\sum_{j \in R(T_i)} \exp(\beta' Z_j)} \right]$$

Le terme L_{Cox} ne dépendant pas de λ_0 , le coefficient β peut être évalué sans connaître la fonction de hasard de base, par maximisation de la vraisemblance partielle.

Tests de validité du modèle de Cox :

Deux hypothèses principales sont à vérifier pour le modèle de Cox :

- L'hypothèse de proportionnalité des risques :

Une méthode graphique est utilisée pour vérifier cette hypothèse de proportionnalité des risques. Pour chaque variable, elle consiste à tracer la fonction $\ln(-\ln(\hat{S}_t))$ en fonction de $\ln(t)$ avec t l'ancienneté du contrat en portefeuille :

$$\ln[-\ln(\hat{S}(t))] = \ln \left[\int_0^t \hat{\lambda}(u|Z) du \right] = \ln[\hat{\Lambda}_0(t) \times \exp(\hat{\beta}Z)] = \hat{\beta}Z + \ln(\hat{\Lambda}_0(t))$$

Avec $\hat{\Lambda}_0$ l'estimateur de Nelson-Aalen

Le résultat de cette hypothèse de proportionnalité des risques est vérifié pour certaines variables telles que le statut d'occupation mais par pour d'autres comme la majoration en euro au terme (cf. graphiques en annexe 8 et limites au paragraphe III.2.d). Dans le cas de risques proportionnels, on observera une translation des courbes de la fonction $\ln(-\ln(\hat{S}_t))$.

- Test de nullité simultanée des coefficients :

Le test de significativité des coefficients porte sur l'hypothèse : $\begin{cases} H_0 : \beta = 0 \\ H_1 : \beta \neq 0 \end{cases}$

Trois statistiques sont utilisées :

- Statistique du rapport de vraisemblance
- Test de Wald
- Statistique du Score (ou log-rank)

Le résultat de la statistique du score est affiché dans le bas de la figure 45 et indique avec une p_value de 0 que les coefficients sont bien significatifs.

b. Base d'étude

A partir de la base construite au chapitre II.1, les données ont été agrégées au niveau de chaque contrat (tandis qu'auparavant les données étaient triées par numéro de contrat et numéro d'avenant). Ainsi, les données de risques (statut d'occupation, nombre de pièces du logement, surface des dépendances, ...) et comportementales (nombre et typologie de contacts auprès de la plateforme de gestion) sont conservées. Pour les données clients (âge, ancienneté,...) et de primes (niveau de prime payée par le client au terme), la dernière situation est visible. Enfin, les évolutions de majoration de primes sont additionnées sur le nombre d'années exposées.

A ces données s'ajoute la notion de censure : cette variable prend la modalité 1 lorsque la résiliation n'est pas observée sur la période d'analyse (2015 à 2017) et prend la valeur 0 lorsque le client a résilié sur la période. Le modèle est dit censuré lorsque le résultat de l'expérience n'est visible que partiellement, c'est-à-dire qu'on n'a pas d'informations sur la durée de vie. On sait uniquement que celle-ci est supérieure à t , où t est la durée de l'étude. Ici, on traitera une censure à droite dans la mesure où la date de résiliation n'est pas connue pendant la période d'observation pour certains individus.

Ainsi, la base de données pour le modèle de survie contient environ 450 000 lignes.

c. Résultats

Le modèle de Cox a été développé sous *R* avec le package *survival* et la fonction *coxph* appliquée à la variable de censure et à l'exposition du contrat, c'est-à-dire son ancienneté en portefeuille.

Le graphique 45 présente les exponentiels des coefficients, communément appelés *hazard ratio* (ou rapport des risques instantanés) qui indiquent le risque relatif estimé sur la durée de l'étude. D'autres informations sont disponibles telles que l'intervalle de confiance des coefficients, le nombre d'occurrences ainsi que les *p-value*.

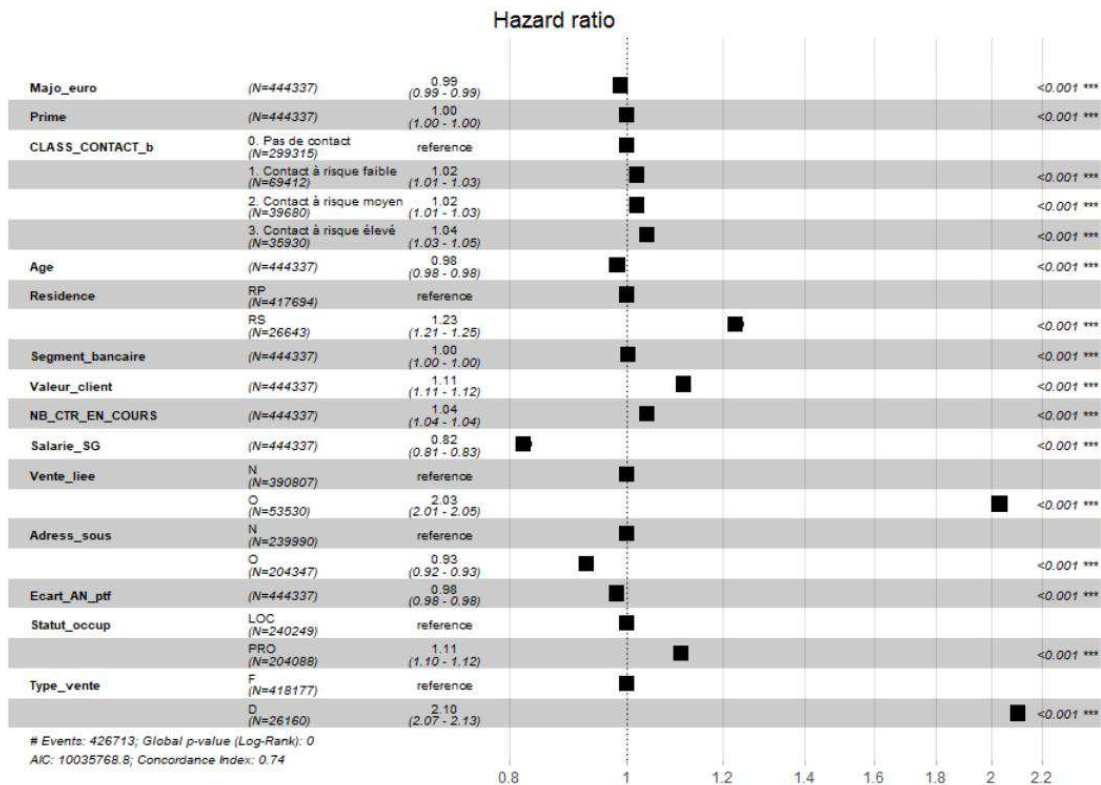


Figure 45 : Rapport des risques instantanés du modèle de Cox

Les p-value sont inférieures à 0.001 indiquant que toutes les variables et leurs modalités sont significatives au seuil de 5%.

Le taux de concordance du modèle s'établit à 0.74 ce qui indique un bon pouvoir discriminant du modèle. Le taux de concordance est équivalent à l'aire sous la courbe ROC défini dans le chapitre III.1.c

Par ailleurs, les variables catégorielles les plus discriminantes (qu'on peut analyser comme celles étant les plus éloignées de la ligne verticale du coefficient 1) sont les suivantes : type de résidence (principale/secondaire), salarier du groupe SG, vente combinée et type de vente (en agence ou à distance).

Intéressons-nous maintenant à la variable continue décrivant l'âge de l'assuré. L'effet marginal d'un an supplémentaire sur l'âge du client s'écrit :

$$\frac{\exp(MajoEuro_i \times \beta_{MajoEuro} + \dots + (Age_i + 1) \times \beta_{Age} + \dots + TypeVente_i \times \beta_{TypeVente})}{\exp(MajoEuro_i \times \beta_{MajoEuro} + \dots + Age_i \times \beta_{Age} + \dots + TypeVente_i \times \beta_{TypeVente})} = \frac{\exp((Age_i + 1) \times \beta_{Age})}{\exp(Age_i \times \beta_{Age})} = \frac{\exp(Age_i \times \beta_{Age}) \times \exp(\beta_{Age})}{\exp(Age_i \times \beta_{Age})} = \exp(\beta_{Age}) = 0.98$$

Quel que soit t , chaque augmentation de l'âge d'un an multiplie le risque instantané par un facteur de 0.98, soit une diminution de 2% de la probabilité de résilier.

Les graphiques suivants présentent les fonctions de survie du modèle sur quelques variables principales en utilisant la fonction *survfit* sous R. Cette fonction est appliquée pour les graphiques 47 à 50, au tableau contenant la moyenne des variables continues et en prenant la modalité la plus représentée pour les variables ordinales. La variable analysée, le score de valeur client dans le graphique 48 par exemple, est introduite comme un vecteur avec les modalités variant de 1 à 6. L'idée est ainsi de faire varier la fonction de survie des modalités de la valeur client pour un individu de référence (toutes les autres variables étant vues toutes choses égales par ailleurs).

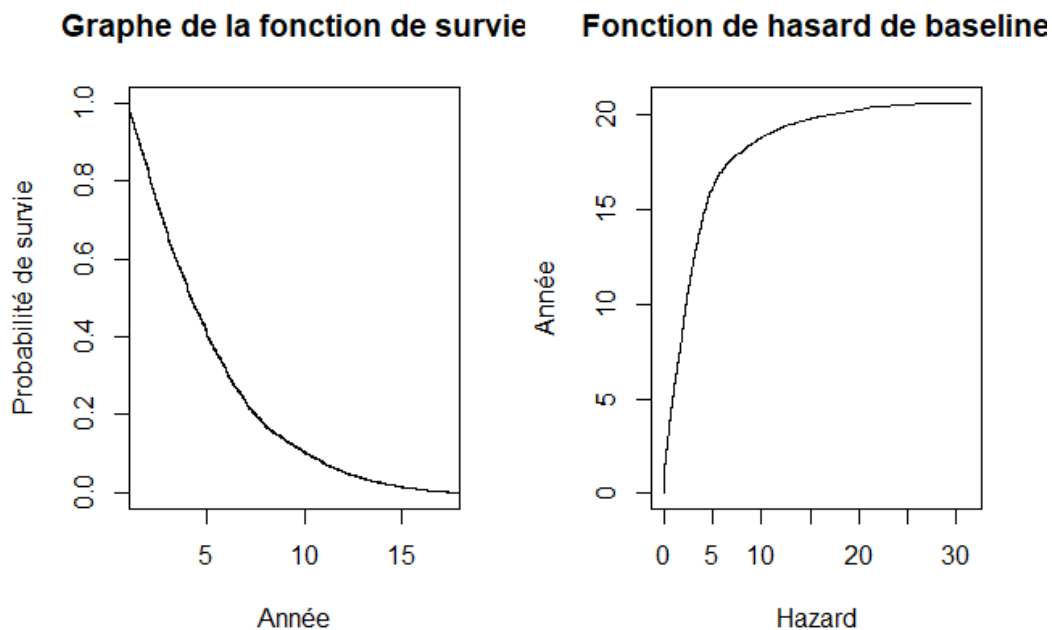


Figure 46 : Fonction de survie et de hasard du modèle de Cox

Le graphique supra présente la fonction de survie et de hasard du modèle de Cox. La fonction de survie permet de représenter la durée de survie des contrats dans le portefeuille. Ainsi, la probabilité qu'un contrat habitation reste en portefeuille après 10 ans d'ancienneté est d'environ 10%.

Graphe de la fonction de survie suivant l'âge

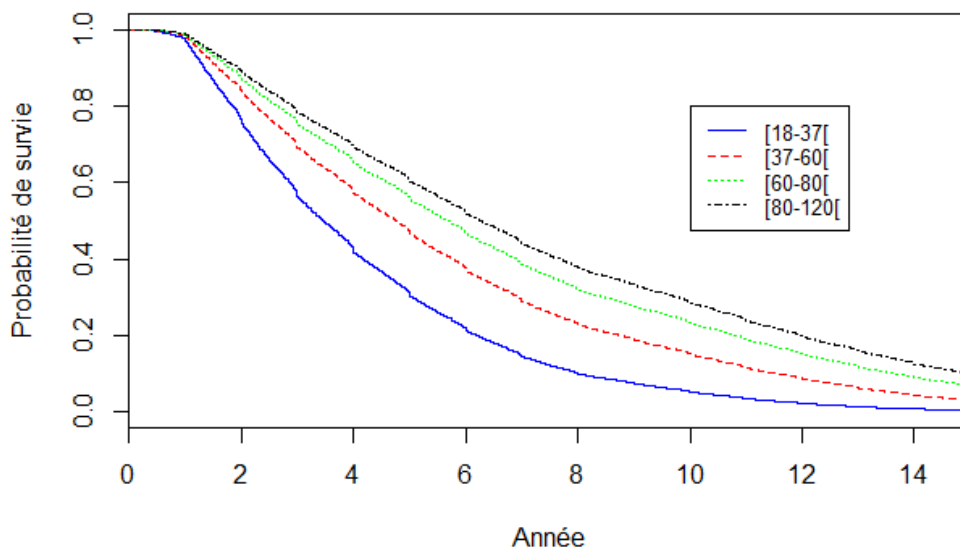


Figure 47 : Fonction de survie du modèle de Cox suivant l'âge du client

La figure 47 expose la fonction de survie suivant la variable âge du client. Le taux de chute décroît plus vite pour les jeunes clients. Après 5 ans, la probabilité de survie en portefeuille est de 30% pour un client d'âge 20 ans contre plus de 60% pour un assuré de 81 ans.

Graphe de la fonction de survie suivant la valeur client

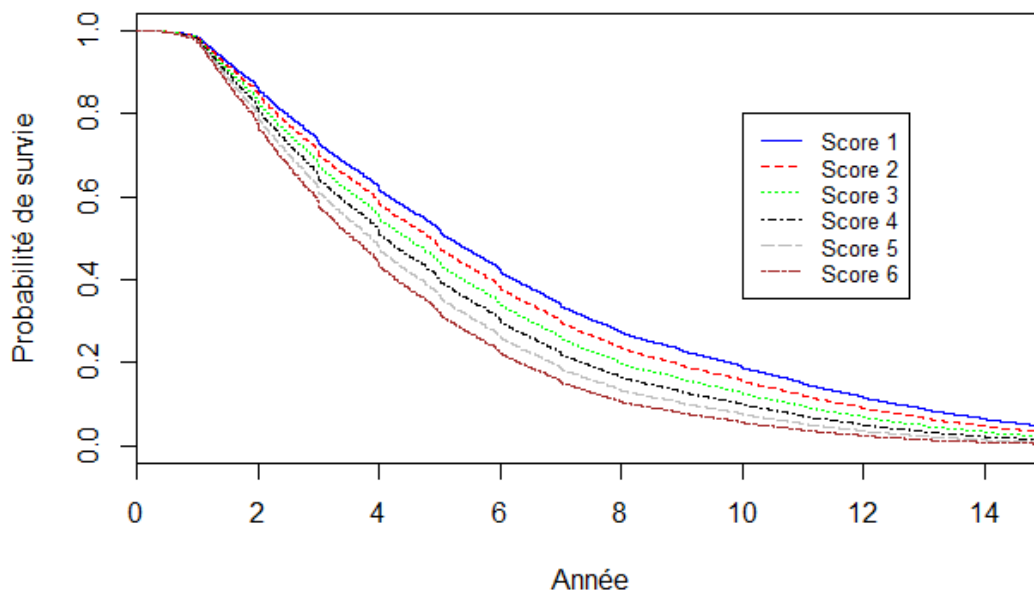


Figure 48 : Fonction de survie du modèle de Cox suivant la valeur client

La figure 48 expose la fonction de survie suivant la variable score de valeur client. Le taux de chute décroît plus vite pour les risques les moins bien notés. Ainsi, après 5 ans, l'écart de probabilité de

survie en portefeuille est de 21 points entre un assuré de score 1 (« bon client ») et de score 6 (« mauvais risque »).

Fonction de survie suivant l'indicatrice salarié groupe SG

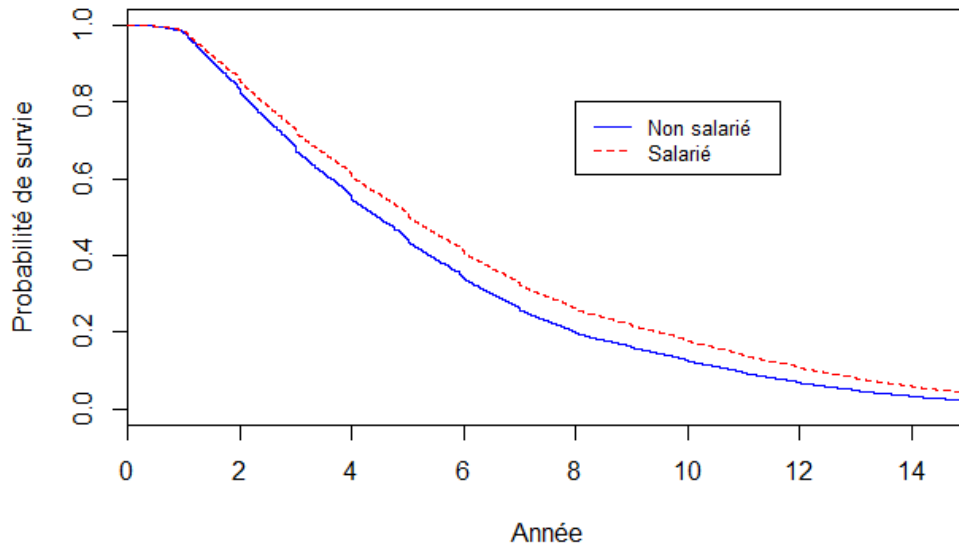


Figure 49 : Fonction de survie du modèle de Cox suivant l'indicatrice salarié SG

Le graphique ci-dessus présente la fonction de survie suivant l'indicatrice salarié du groupe Société Générale. La probabilité de survie est meilleure pour les salariés du groupe. Après 5 ans, la probabilité de rester en portefeuille est de 43% pour un client non salarié de la SG contre plus de 50% pour un salarié de la banque.

Graphique de la fonction de survie en fonction du statut d'occupation

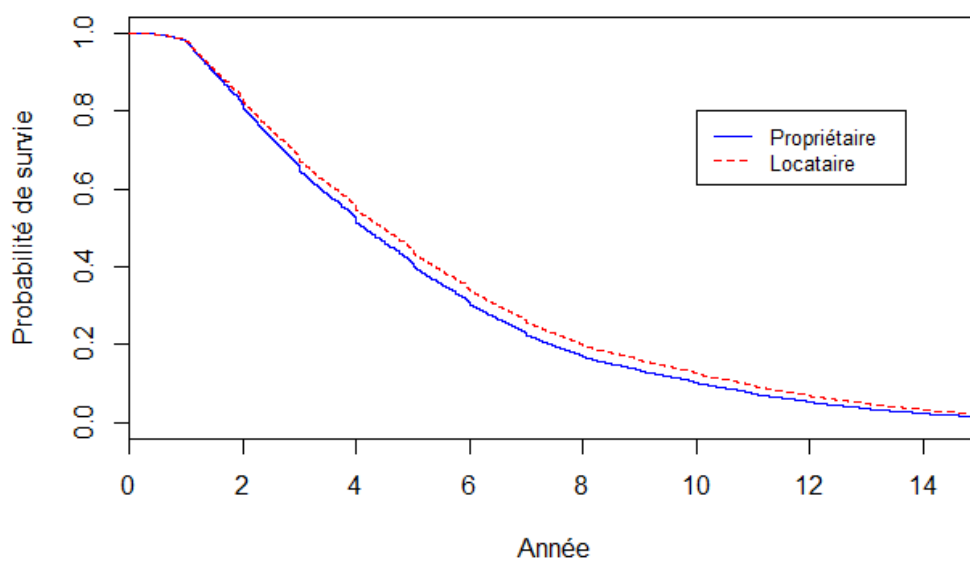


Figure 50 : Fonction de survie du modèle de Cox suivant le statut d'occupation

La figure 50 présente la fonction de survie suivant le statut d'occupation de l'assuré. La probabilité de survie ne diffère pas beaucoup entre un propriétaire et un locataire. Après 10 ans, la probabilité de rester en portefeuille est de 13% pour un locataire contre 10% pour un propriétaire.

d. Limites du modèle

Malgré un bon taux de concordance et des résultats proches de la régression logistique, certaines limites sont apparues lors de la modélisation de Cox.

L'hypothèse de proportionnalité des risques n'étant pas vérifiée pour certaines variables (notamment la majoration en euro), certains résultats obtenus sur le modèle final nous paraissent contre-intuitifs. Ainsi, le coefficient de la variable majoration en euro est égal à 0.99 (cf. figure 45) ce qui induirait une diminution de la résiliation de 1% pour chaque euro supplémentaire de majoration au terme.

En conséquence, ces limites nous conduisent à écarter le modèle de Cox et préférer le modèle de régression logistique pour l'étude de la sensibilité au prix.

3. Modèle de sensibilité au prix

Le modèle de sensibilité au prix a été développé sous *Addactis Pricing* à partir du modèle de régression logistique défini au chapitre III.1.d et les principes exposés dans le paragraphe suivant.

a. Théorie

Soit π_i la probabilité qu'un individu résilie sur une police i , on définit :

$$E(y_i) = P(y_i = 1) = \pi_i$$

En notant la majoration de prix au terme comme un facteur Prix (introduit en version polynomiale d'ordre q), le modèle s'écrit, pour la loi logit et l'observation i :

$$g(\pi_i) = g[E(y_i)] = \ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \sum_{j=0}^p \beta_j x_{ij} + \sum_{j=1}^q \alpha_j \text{Prix}_i^j$$

$$\Leftrightarrow \pi_i = \frac{1}{1 + \exp\left[-\left(\sum_{j=0}^p \beta_j x_{ij} + \sum_{j=1}^q \alpha_j \text{Prix}_i^j\right)\right]}$$

Etant donné que nous ne considérerons que les modalités des variables explicatives, l'indice i peut être omis dans la formule précédente. Pour une variable x_j fixée, la probabilité de résiliation s'exprime en fonction des caractéristiques de l'assuré et de la variation de prix :

$$\pi(\text{Prix}) = \frac{1}{1 + \exp \left[- \left(\sum_{j=0}^p \beta_j x_j + \sum_{j=1}^q \alpha_j \text{Prix}^j \right) \right]}$$

Et l'élasticité au prix au point P_1 est définie par l'expression suivante :

$$\varepsilon(P_1) = \frac{\frac{(\pi(P_2) - \pi(P_1))}{\pi(P_1)}}{\frac{(P_2 - P_1)}{P_1}}$$

Avec $P_2 > P_1$, $P_1 \neq 0$ et $P_2 - P_1 = c$ où c représente le pas de variation de la majoration de prix au terme en euro.

La formule est cohérente avec celle présentée plus en amont du mémoire (cf. I.3).

b. Résultats

Une synthèse des variables les plus sensibles au prix est présentée dans la figure suivante :

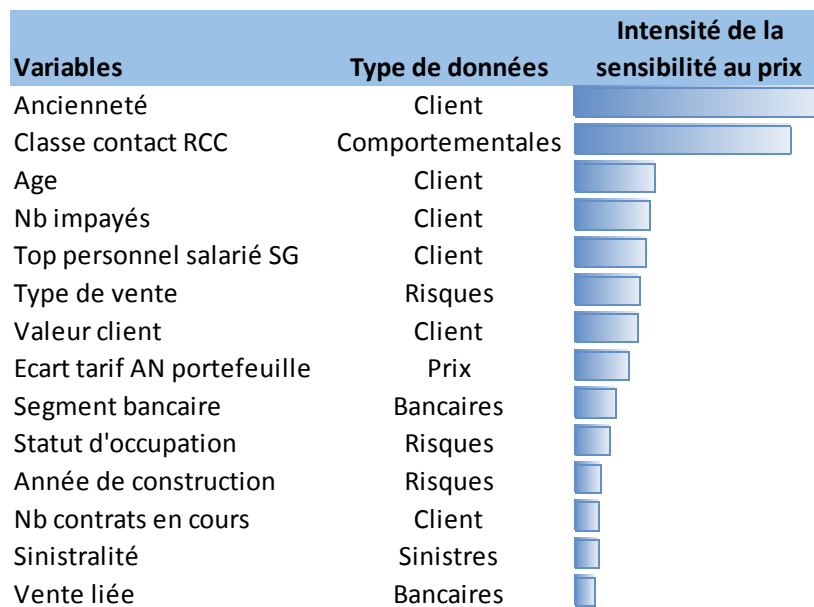


Figure 51 : Classement des variables les plus sensibles au prix

L'intensité de l'élasticité au prix a été analysée pour chaque variable et chaque modalité en comparant la variation du taux de résiliation entre une variation de la prime de -5€ à +50€ par pas de

5€. Ainsi, la variable ancienneté du contrat possède l'intensité de sensibilité au prix la plus importante. D'une manière générale, les variables les plus discriminantes dans la régression logistique se retrouvent en tête des variables les plus sensibles à une majoration tarifaire.

Illustrons les résultats obtenus pour les variables les plus significatives en terme d'élasticité au prix. Les graphiques 52 à 55 ont été construits en traçant les valeurs de π définies précédemment pour différentes valeurs de la variable variation de prix (P_1, P_2, \dots).

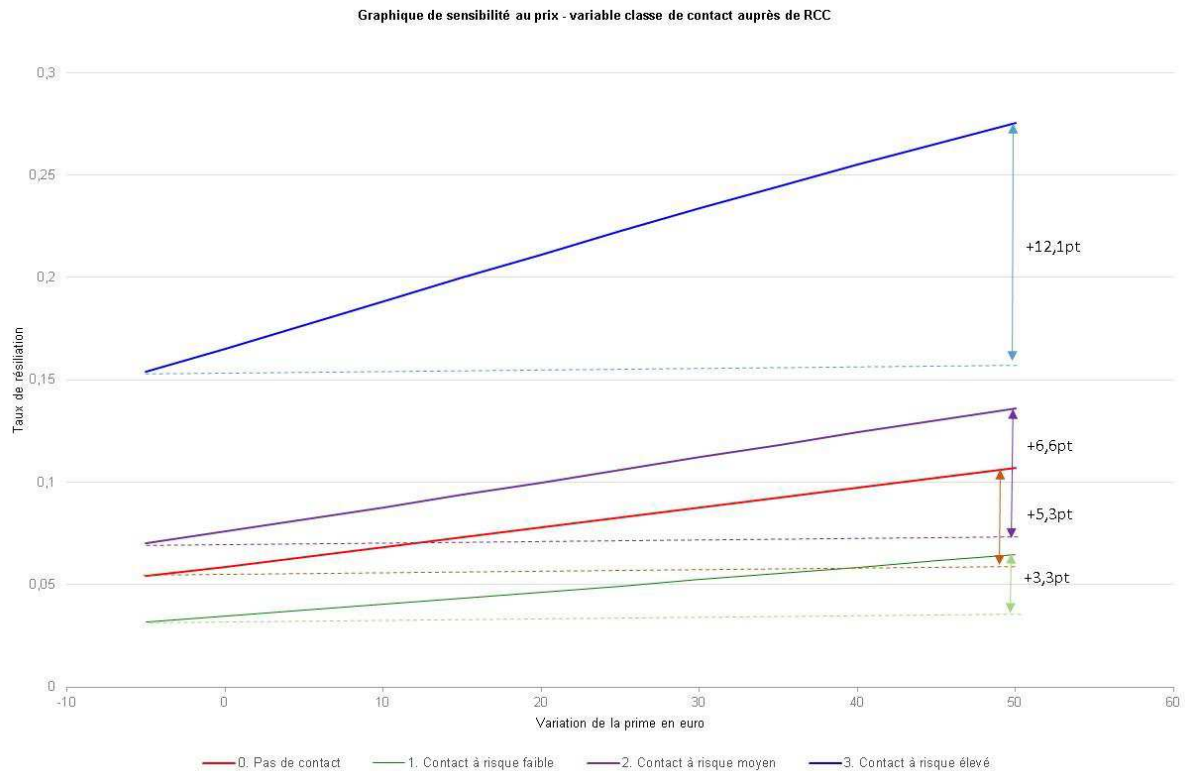


Figure 52 : Sensibilité au prix en fonction de la typologie de contact auprès de la plateforme de gestion

Pour une variation de la prime en euro de -5€ à +50€, nos assurés en portefeuille voient leur taux de résiliation augmenter de 12.1 points s'ils ont eu un contact à risque élevé, et de 5.3 points s'ils n'ont pas contacté la plateforme. Pour visualiser la sensibilité au prix, ce n'est pas le niveau du taux de résiliation qui importe mais la pente. L'allure des courbes nous renseigne sur l'élasticité de nos assurés, et non pas l'ordonnée à l'origine représentant ici le taux de résiliation.

Graphique de sensibilité au prix en fonction de l'âge de l'assuré

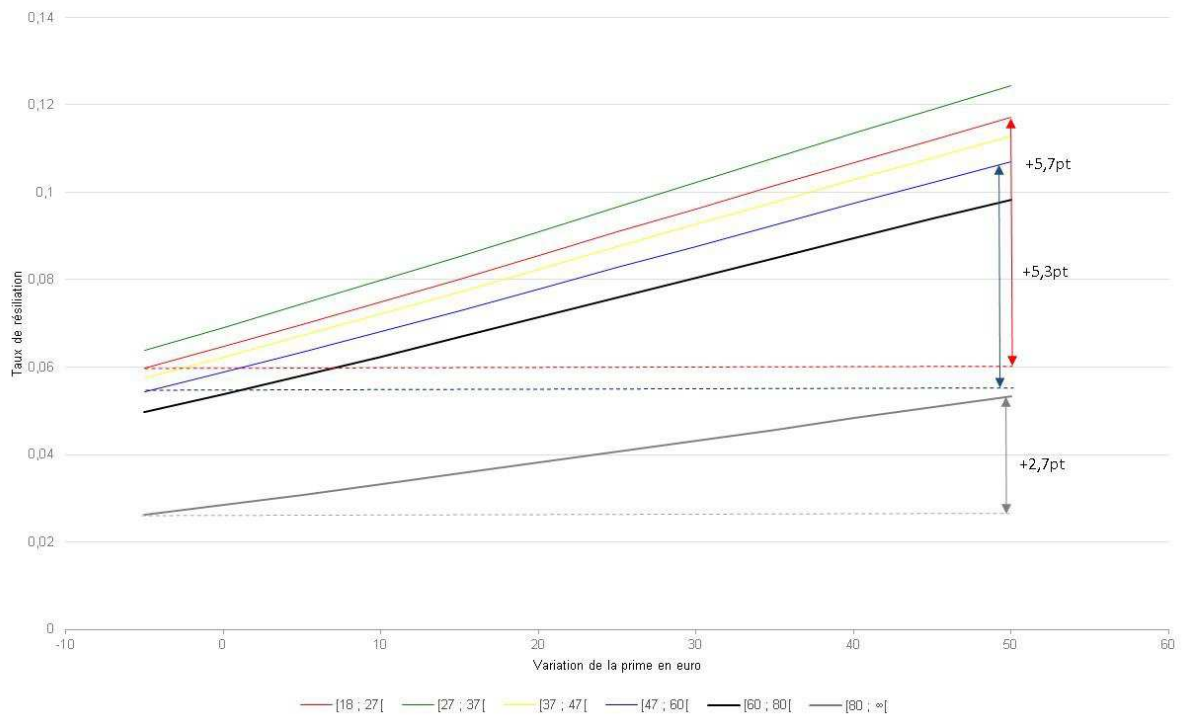


Figure 53 : Sensibilité au prix en fonction de l'âge du souscripteur

Plus l'assuré est jeune, plus il s'avère sensible au prix. Un écart de 3 points d'élasticité au prix est observé entre un assuré de 25 ans et un client âgé de plus de 80 ans.

Graphique de sensibilité au prix en fonction de l'ancienneté

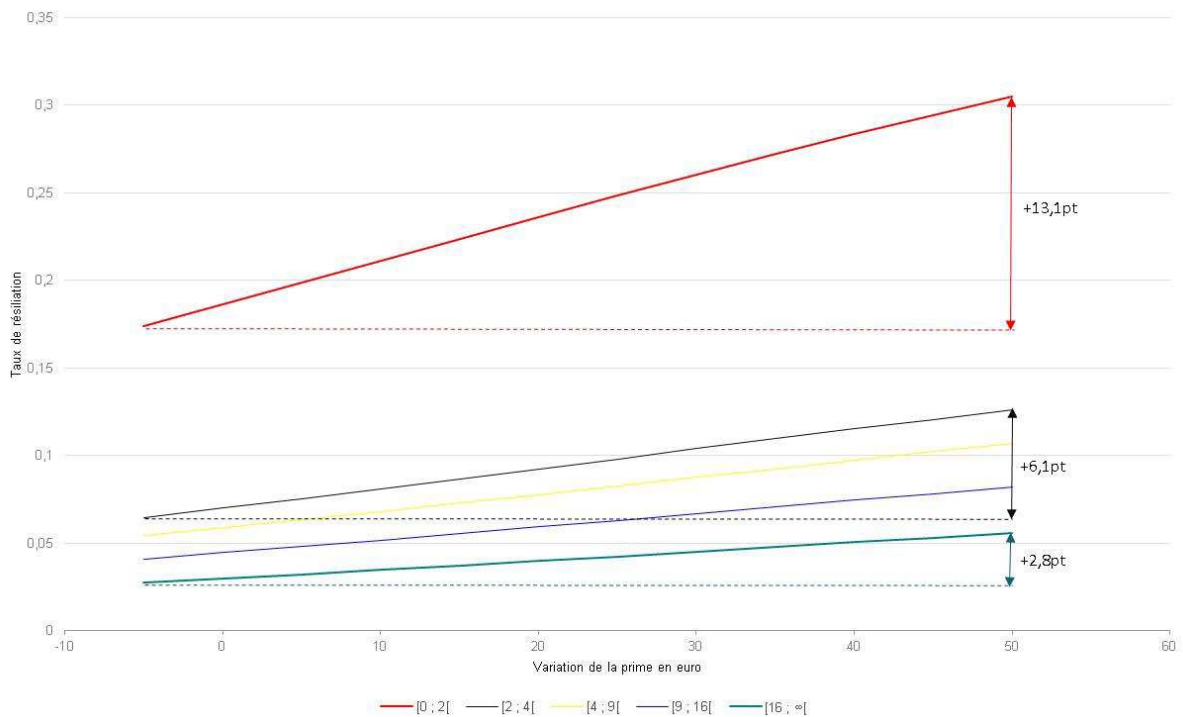


Figure 54 : Sensibilité au prix en fonction de l'ancienneté du contrat

Plus un contrat est ancien, moins il est sensible au prix. Un écart de plus de 10 points est constaté entre un contrat récent (hausse du taux de résiliation de 13.1 points entre -5€ et +50€ de majoration de prime pour un contrat de moins de 2 ans) et un contrat ayant plus de 16 ans d'ancienneté (hausse du taux de résiliation de 2.8 points pour les mêmes variations de prime). La hausse du taux de résiliation s'élève à 6.1 points pour un contrat ayant une ancienneté comprise entre 2 et 4 ans.

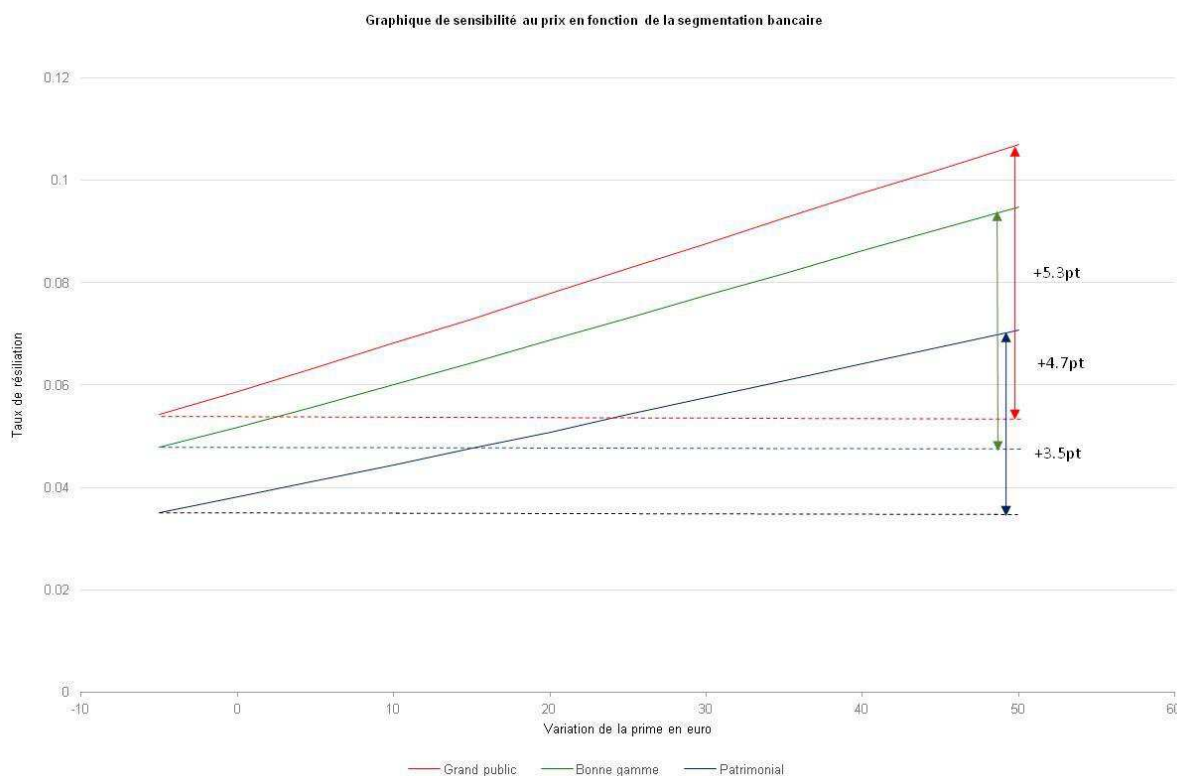


Figure 55 : Sensibilité au prix en fonction du segment bancaire

Les clients catégorisés comme « Patrimonial » sont moins sensibles au prix comparés aux clients « Grand public ». Un écart de près de 2 points de l'élasticité au prix est relevé entre ces deux typologies de clients.

D'autres graphiques sur l'élasticité au prix sur les variables salarié du groupe et statut d'occupation sont disponibles en annexe 9.

IV. Modélisation de la prime commerciale

1. Objectif

La prime termée à chaque renouvellement de contrat fonctionne sur un principe de revalorisation et non de retarification. Ainsi, la prime évolue de $x\%$ à l'échéance sans passer par l'étape de tarification. Au fil du temps, il peut donc exister un écart entre le niveau de la prime termée en portefeuille et celui du tarif affaires nouvelles.

L'objectif de ce chapitre est de construire un tarif technique, à partir d'un modèle GLM de fréquence et coût moyen par garantie, et d'identifier les segments à marge ou en déficit en comparant la prime commerciale modélisée et la prime termée en portefeuille.

Précisons que l'objet principal du mémoire n'étant pas la construction d'un tarif Habitation, le modèle de tarification a été construit en utilisant des variables traditionnelles sans chercher à enrichir le modèle de données externes.

2. Produit Habitation

Le produit Habitation existe depuis le lancement de la compagnie d'assurances Sogessur en 1997. La dernière offre commercialisée date de 2010 et comprend cinq formules. Les conseillers de clientèle, Société Générale et Crédit du Nord, n'étant pas des agents d'assurance et le temps de souscription étant limité dû aux nombreux produits bancaires et assurantiels proposés, les offres MRH comprennent des packages de garanties avec relativement peu de modularité (les options existantes concernent les packs et le relèvement des capitaux assurés). Le principe de formules de garanties est résumé ci-après et détaillé en annexe 10 :

- Formule initiale (HA) comprenant les garanties essentielles (Responsabilité civile, Incendie, Tempête, Dégât des eaux, Attentat, Catastrophes naturelles, Défense Pénale et Recours suite à accident) ;
- Formule initiale+ (HB) couvrant en sus le Vol et Vandalisme ;
- Formule Confort (HC) enrichie avec le Bris de vitres ;
- Formule Confort+ (HD) ajoute les dommages électriques et une valeur à neuf limitée ;
- Formule Optimale (HE) comporte une valeur à neuf étendue et les dommages corporels.

3. Base de données

a. Période d'analyse

Afin de disposer de suffisamment d'informations et d'aboutir à des modèles statistiques qui convergent et conduisent à une tarification qui soit suffisamment robuste, une base sous SAS couvrant trois exercices sur la période de janvier 2015 à décembre 2017 a été construite.

Dans le but de disposer d'une vision « liquidée » des sinistres, les données sont vues à fin juillet 2018. Ainsi, la connaissance de la charge de sinistres est plus précise puisque la plupart des sinistres tardifs sont déclarés et une bonne partie des sinistres sont clos, ce qui permet d'obtenir un modèle de coût plus robuste.

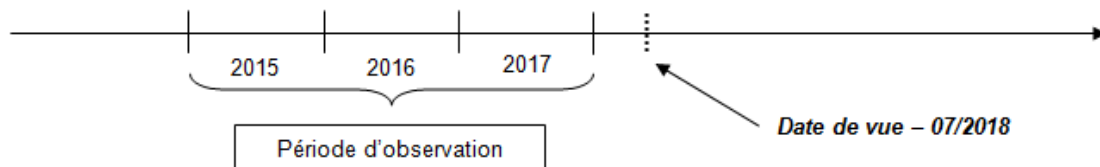


Figure 56 : Période d'étude des bases de données contrats et sinistres

Les bases contrats et sinistres excluent les formules Investisseur et Etudiant et l'ancienne offre pour des raisons de structures tarifaires différentes.

b. Données contrats

Chaque contrat ayant été exposé au moins un jour sur la période d'observation, de 2015 à 2017, est intégré dans la base des contrats. Cette table contient les données contrats et risques déjà détaillées dans le paragraphe II.1 avec en sus :

- La clé d'identification (numéro de contrat et numéro d'avenant) ;
- Les données risques :
 - Zone géographique (zonier défini au code INSEE) ;
 - L'étage du bien :
 - i. Rez de chaussée (RDC) ;
 - ii. Etage intermédiaire (ETI);
 - iii. Sous toiture (SST).
 - Présence d'un insert (élément d'une cheminée permettant de clore le foyer):
 - i. Non communiqué ou ne sait pas ;
 - ii. Pas d'insert ;
 - iii. Insert posé par un professionnel ;
 - iv. Insert non posé par un professionnel ;
 - v. Insert mais ne sait pas si posé par un professionnel.

- L'exposition des contrats sur la période (plus d'un million d'années police ou année risques observées entre 2015 et 2017). Cette donnée permettra de modéliser un nombre de sinistres pondéré des années risques.

c. Données sinistres

La base de données sinistres est constituée des variables suivantes :

- La clé d'identification des contrats ;
- Le numéro de sinistre ;
- La date de survenance du sinistre pour filtrer sur les années 2015 à 2017 ;
- Les garanties mises en jeu et détaillées dans le paragraphe 2 du présent chapitre (Dégât des eaux, Incendie, Vol, etc.) ;
- Le coût du sinistre : seuls les sinistres avec des montants de charges positifs sont conservés. En effet, il peut arriver que des sinistres présentent des coûts négatifs avec des recours supérieurs aux paiements. Par ailleurs, nous avons conservé les sinistres non garantis ou sans suite qui peuvent présenter des frais d'expertise même si le client n'est pas indemnisé in fine.

Cette base contient près de 100 000 sinistres et est rapprochée avec les données contrats. Elle va permettre la modélisation du coût moyen des dossiers en faisant le rapport entre la charge de sinistre et le nombre de sinistres.

4. Préambule technique

La théorie des modèles linéaires généralisés ou *general linear model (GLM)* est couramment utilisée dans l'assurance non vie pour modéliser l'impact d'une variable toutes choses égales par ailleurs. Il s'agit d'une extension des modèles linéaires classiques qui a été introduite dans les années 1970 par John Nelder et Robert Wedderburn.

a. Rappels théoriques

Un rappel sur le modèle linéaire simple permettra de mieux appréhender les modèles GLM. Le modèle linéaire simple décrit une variable aléatoire Y par une combinaison linéaire de variables explicatives X_1, \dots, X_p et ce pour n individus.

Le modèle linéaire simple s'écrit de la forme : $Y = X \cdot \beta + \epsilon$ avec les définitions suivantes :

- Y le vecteur des variables à expliquer de taille n ;
- X une matrice contenant les variables explicatives, de dimension $n \times (p + 1)$;

- β le vecteur des coefficients à estimer, de taille $p + 1$;
- ϵ le vecteur des résidus de taille n .

Cependant, le modèle linéaire simple ne permet pas de représenter des formes non linéaires. Par exemple, en assurance automobile, la prime pure n'est pas linéaire en fonction de l'âge du conducteur. Cette hypothèse de normalité dans la distribution des données peut être affranchie dans le cadre des modèles linéaires généralisés avec une variable réponse Y suivant une loi appartenant à la famille exponentielle.

Une distribution de probabilité appartient à la famille exponentielle si elle admet une fonction de densité s'écrivant sous la forme :

$$f_{\theta,\phi}(y) = c_{\phi}(y) \times \exp\left(\frac{y\theta - a(\theta)}{\phi}\right)$$

- θ est un paramètre de forme ;
- ϕ un paramètre dispersion ;
- $a(\theta)$ une fonction convexe ;
- $c_{\phi}(y)$ une fonction ne dépendant pas de θ .

Plusieurs lois connues appartiennent à la famille exponentielle et deux d'entre elles, la loi de Poisson et la loi Gamma, seront détaillées dans les paragraphes suivants.

Un modèle linéaire généralisé est un modèle défini par une combinaison linéaire de variables, une fonction de lien g et une loi de distribution de la variable à régresser appartenant à la famille exponentielle. Avec la notation μ pour l'espérance mathématique de Y et η_i le prédicteur linéaire, le modèle GLM s'écrit :

$$g[E(Y)] = g(\mu) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

$$\Leftrightarrow E(Y) = \mu = g^{-1}(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)$$

$$\Leftrightarrow E(Y_i)_{i=1,\dots,n} = \mu_i = g^{-1}\left(\beta_0 + \sum_{j=1}^p \beta_j x_{ij}\right) = g^{-1}(\eta_i)$$

b. Modèle de fréquence

La loi de Poisson sera utilisée pour le modèle de fréquence dans la mesure où il s'agit d'une loi de comptage permettant de calculer le nombre d'événements survenus au cours d'une période donnée. Elle prend ses valeurs dans les nombres entiers positifs, ce qui est cohérent avec l'estimation d'un nombre de sinistres.

Le modèle de fréquence dispose des caractéristiques suivantes :

- Variable réponse : nombre de sinistres ;
- Contrainte du modèle : exposition des contrats sur la période (ou années risques) ;

- Fonction de lien : logarithme $\eta_i = \ln(\mu_i)$;
- Loi de distribution : Poisson de paramètre μ avec pour densité : $f(y) = \frac{\exp(-\mu)\mu^y}{y!}$

Notons que l'espérance de la loi de Poisson est égale à sa variance : $E(Y) = Var(Y) = \mu = e^{X\beta}$

c. Modèle de coût moyen

La loi de Gamma sera utilisée pour le modèle de coût moyen étant donné qu'il s'agit d'une loi à valeurs strictement positives et marquée par une forte asymétrie. La queue de distribution à droite est plus étendue que celle de gauche, autorisant des valeurs de sinistres plus importants qu'une loi Normale par exemple.

Le modèle de coût moyen dispose des caractéristiques suivantes :

- Variable réponse : charge des sinistres ;
- Contrainte du modèle : nombre de sinistres ;
- Fonction de lien : inverse $\eta_i = \frac{1}{\mu_i}$;
- Loi de distribution : Gamma de paramètres α et β avec pour densité : $f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y}$

L'espérance de la loi de Gamma est égale à $E(Y) = \frac{\alpha}{\beta}$ et sa variance : $Var(Y) = \frac{\alpha}{\beta^2}$

d. Mesure de qualité du modèle

Plusieurs indicateurs existent pour déterminer la qualité d'un modèle GLM.

Déviance

La déviance permet de mesurer l'adéquation du modèle estimé aux données et généralise le coefficient de détermination R^2 bien connu pour le modèle linéaire. Il s'agit de comparer la fonction de vraisemblance L du modèle à la vraisemblance du modèle saturé, qui ajuste exactement toutes les valeurs de la variable réponse. La déviance est la quantité suivante :

$$D = -2\varphi \ln \left(\frac{L(\beta, \varphi)}{L(\beta_{saturé}, \varphi)} \right)$$

La déviance normalisée du modèle est définie par : $D^* = \frac{D}{\varphi}$

Cette information D^* suit une loi du chi-deux dont le nombre de degrés de liberté est celui qui sépare les deux modèles. La valeur attendue de D^* est donc approximativement son nombre de degrés de liberté dans la mesure où il s'agit de la valeur de l'espérance d'une loi du chi-deux.

Critères AIC et BIC

Les critères AIC et BIC ont déjà été définis dans le chapitre III.1.d.

5. Modélisation

a. Principe

La construction du modèle GLM a été réalisée à partir du logiciel *Addactis Pricing* module *pricing*.

Le modèle de fréquence et de coût moyen est construit pour chaque garantie, décrite dans le paragraphe IV.2, c'est-à-dire les garanties :

- Dégât des eaux ;
- Incendie ;
- Responsabilité civile (RC) ;
- Tempête Grêle Neige ;
- Bris de vitres ;
- Vol ;
- Vandalisme (ou dégâts suite à vol ou détériorations suite à vol) ;
- Dommages électriques.

Pour l'Incendie et la RC, des sinistres graves peuvent perturber la distribution des montants de sinistres et il est préférable, pour obtenir un modèle plus robuste, de découper la sinistralité en deux suivant un seuil d'écrêtement. Ainsi, la modélisation est réalisée à partir de la sinistralité de masse et la sinistralité de pointe. L'écrêtement des sinistres pour ces deux garanties a été positionné à 30 000€ correspondant à la segmentation des équipes de gestion de sinistres entre gestionnaires généralistes et spécialistes.

Le calcul de la prime pure (PP) par garantie s'écrira donc comme suit :

- Pour les garanties hors RC et Incendie :

$$PP = \text{Fréquence} \times \text{CoûtMoyen}$$
$$\Leftrightarrow E(S) = E(N) \times E(C)$$

Avec :

- S le coût total des sinistres ;
- N le nombre de sinistres ;
- C le coût unitaire de chaque sinistre.

- Pour les garanties RC et Incendie :

$$PP = \text{Fréqécrêté}_{GLM} \times \text{CMécrêté}_{GLM} + \text{Fréqexcédent}_{GLM} \times \text{CMexcédent}_{\text{Observé}}$$

Avec :

- $\text{Coût écrêté} = \text{Min}(\text{ChargeTotaleSinistre} ; 30\,000 \text{ €}) ;$
- $\text{CMexcédent}_{\text{Observé}}$ correspond au coût moyen des sinistres supérieurs à 30 000€, une modélisation du coût moyen sur ce segment étant caduque du fait d'une faible volumétrie de sinistres.

b. Significativité des variables

Le test de type III repose sur le test de nullité de q coefficients pour déterminer si les variables explicatives ont un impact significatif sur le modèle.

Pour une variable j donnée à q modalités, testons la pertinence de la variable dans le modèle en écrivant l'hypothèse suivante :

$$H_0: \beta_{j1} = \dots = \beta_{jq} = 0 \text{ contre } H_1: \exists k \in \{1, \dots, q\} \mid \beta_{jk} \neq 0$$

La significativité des variables explicatives est rendue possible par la théorie du maximum de vraisemblance donnant la loi asymptotique des estimateurs. Plusieurs tests existent :

- Test de Wald : le khi-deux de Wald est égal au carré du rapport du coefficient estimé à son erreur standard estimé. Il est similaire au carré du T de la loi de Student de la régression linéaire multiple ;
- Test LRT (*Likelihood Ratio Test*) : cette statistique calcule l'écart entre les vraisemblances des deux points $\hat{\beta}$ (estimateur du modèle sans contrainte) et β_{H_0} (β sous l'hypothèse nulle) et vérifie que la différence est égale à 0.

En combinant le rappel sur les procédés de sélection de variables (cf. III.1.d) et les éléments ci-dessus, la figure 57 présente la significativité des variables pour le modèle de fréquence de la garantie Dégât des eaux.

Criteria	Degrees of freedom	Chi-2	Pr > Chi-2
Zone géographique	9	4 266	-
Statut-type x étage	7	2 401	-
Nombre de pièces	9	1 770	-
Année de construction	4	1 122	0,00
Présence d'enfant	1	684	0,00
Type de résidence	1	129	0,00
Formule	4	125	0,00
Niveau des capitaux objets de valeur	3	100	0,00
Surface des dépendances	2	27	0,00
Présence d'un insert	4	29	0,00
Niveau des capitaux objets usuels	1	4	0,045

Figure 57 : Modèle de fréquence Dégât des eaux - significativité des variables

Ainsi, les variables explicatives détaillées ci-dessus sont toutes significatives puisque leur p -value est inférieur à 5%. Par ailleurs, la valeur de leur Khi-deux indique leur degré d'importance dans le modèle. Les variables zone géographique, statut-type croisé avec l'étage, nombre de pièces et année de construction apparaissent comme les plus discriminantes dans la modélisation de la fréquence Dégât des eaux.

c. Validation des modèles

Les résidus sont basés sur la différence entre les valeurs de la variable à expliquer $(Y_i)_{i=1,\dots,n}$ et leurs estimations $(\hat{Y}_i)_{i=1,\dots,n}$.

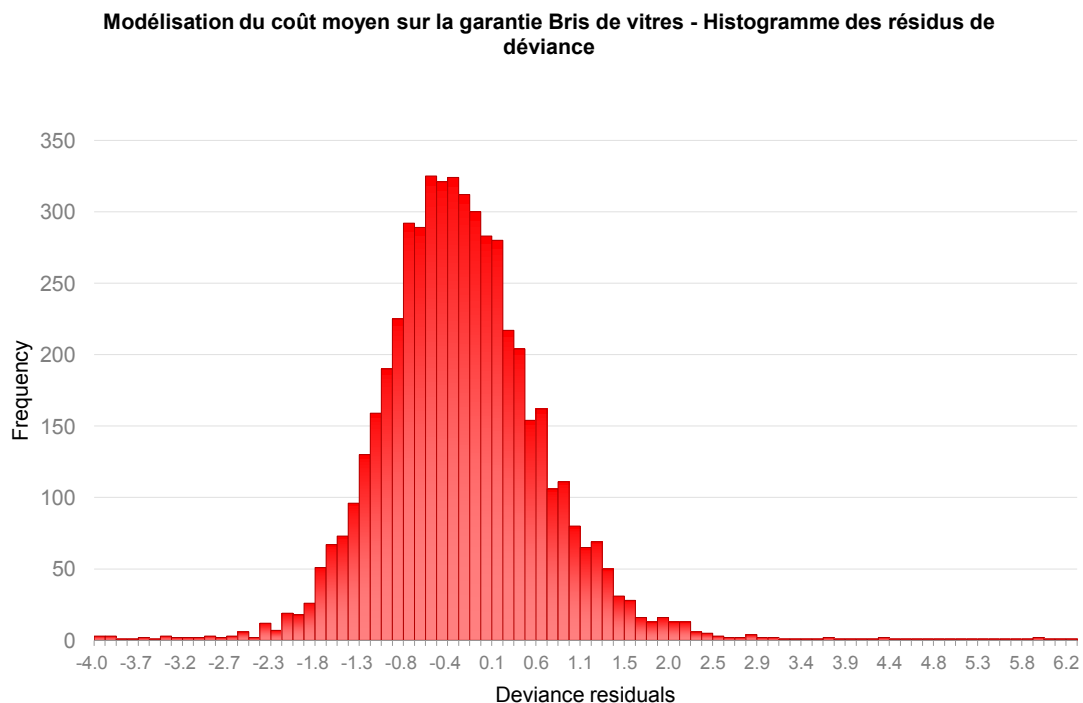
Contrairement à un modèle linéaire simple, dans un GLM, la normalité, la constance de la variance et l'additivité des effets ne sont plus un pré-requis. Les deux seules hypothèses du modèle GLM reposent sur l'appartenance à la famille exponentielle de la distribution de la variable à expliquer et l'effet additif des facteurs explicatifs sur la variable réponse à une transformation près.

L'analyse des résidus du GLM est principalement graphique et permet de vérifier que l'erreur reste aléatoire. Dans ce mémoire, les résidus calculés sont les résidus de déviance. Pour chaque observation i , ils s'écrivent :

$$r_{Di} = \text{signe}(y_i - \hat{y}_i) \times \sqrt{d_i} \quad \text{avec :}$$

- La déviance $D = \sum_i d_i$ définie au IV.4.d ;
- $\hat{y}_i = \mu_i = g^{-1}\left(\sum_{j=0}^p \hat{\beta}_j x_{ij}\right)$

Dans le cas d'une variable continue, le montant des sinistres, un histogramme des résidus de déviance est tracé (cf. figure 58).



Le résultat est conforme à l'attendu puisque l'histogramme des résidus de déviance s'avère typique d'une loi normale centrée en 0.

Pour le cas d'une variable discrète représentant le nombre de sinistres pour la modélisation de la garantie Dégât des eaux, on ne représente pas le nuage de points des résidus mais un regroupement par paquet de 300 valeurs de résidus ou *crunched residuals* (cf. annexe 11). Les déviances et leurs résidus sont calculés sur les valeurs moyennes des valeurs observées et prédites. Le modèle estimé apparaît conforme étant donné que la dispersion des points est symétrique autour de 0. Par ailleurs, il n'y a pas de tendance observée sur les résidus.

d. Modélisation de la fréquence

Les résultats obtenus sur la modélisation de la fréquence de quelques garanties et variables sont présentés dans ce paragraphe.

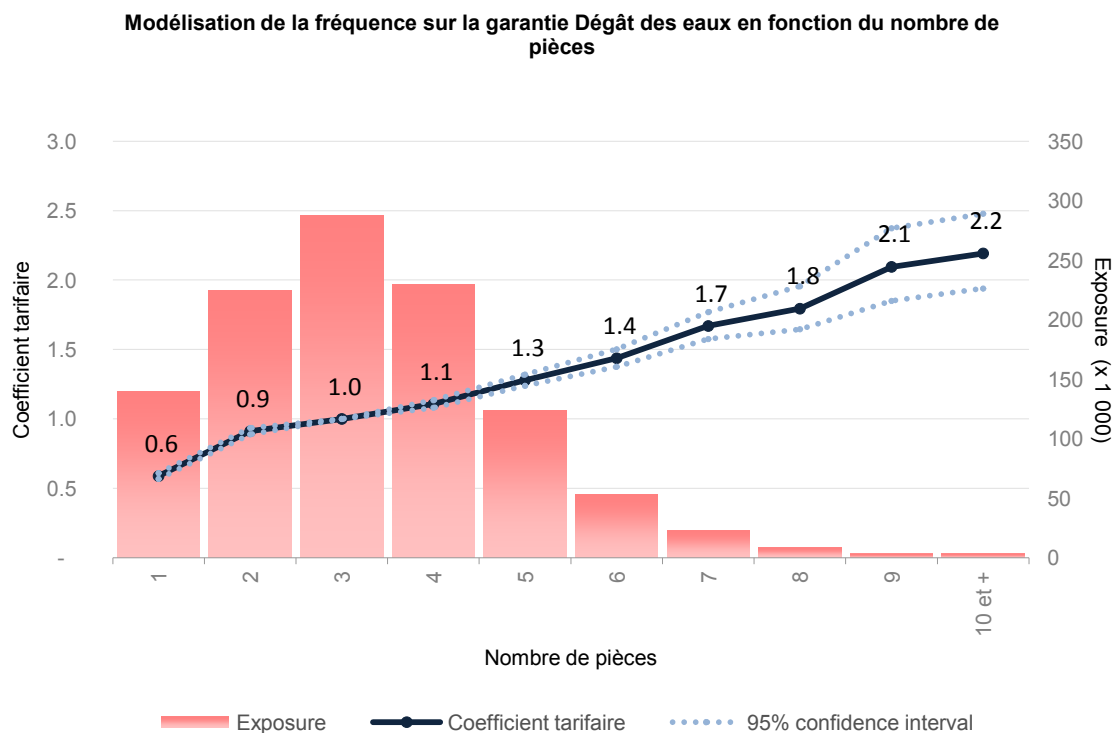


Figure 59 : Coefficient tarifaire de la fréquence Dégât des eaux par nombre de pièces

La pente tarifaire est quasiment linéaire pour le nombre de pièces sur la garantie Dégât des eaux. Ainsi, plus le nombre de pièces augmente, plus le coefficient multiplicatif est élevé. Ainsi une habitation de 6 pièces aura 40% de chances en plus d'être sinistrée que le niveau de référence correspondant à une habitation de 3 pièces. Remarquons que l'intervalle de confiance augmente dès lors que l'exposition des contrats, en rouge, diminue.

Modélisation de la fréquence sur la garantie Vol en fonction de la zone géographique

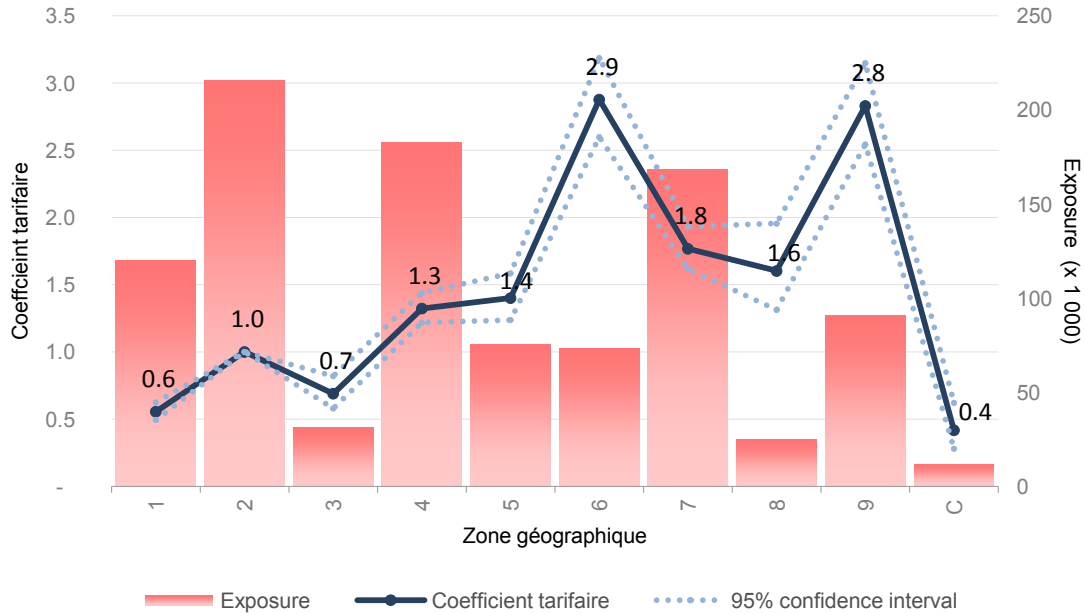


Figure 60 : Coefficient tarifaire de la fréquence Vol par zone géographique

La zone 1 correspondant aux communes rurales dispose de 40% de chances de moins d’avoir un sinistre Vol que le niveau de référence (zone 2 – villes secondaires). A l’inverse, la zone 9 relative aux arrondissements les plus sinistrés de Paris et Marseille a une probabilité supérieure à 180% d’être sinistrée en Vol par rapport à la zone 2. Notons également la cohérence du zonier qui discrimine bien les zones 3 et 4 (respectivement communes périurbaines sinistralité – et +), les zones 5 et 6 (respectivement grandes agglomérations sinistralité – et +) et les zones 8 et 9 (Paris et Marseille sinistralité – et +) puisque les coefficients tarifaires de ces zones sont bien distincts deux à deux.

Modélisation de la fréquence sur la garantie RC en fonction de la présence d'enfants

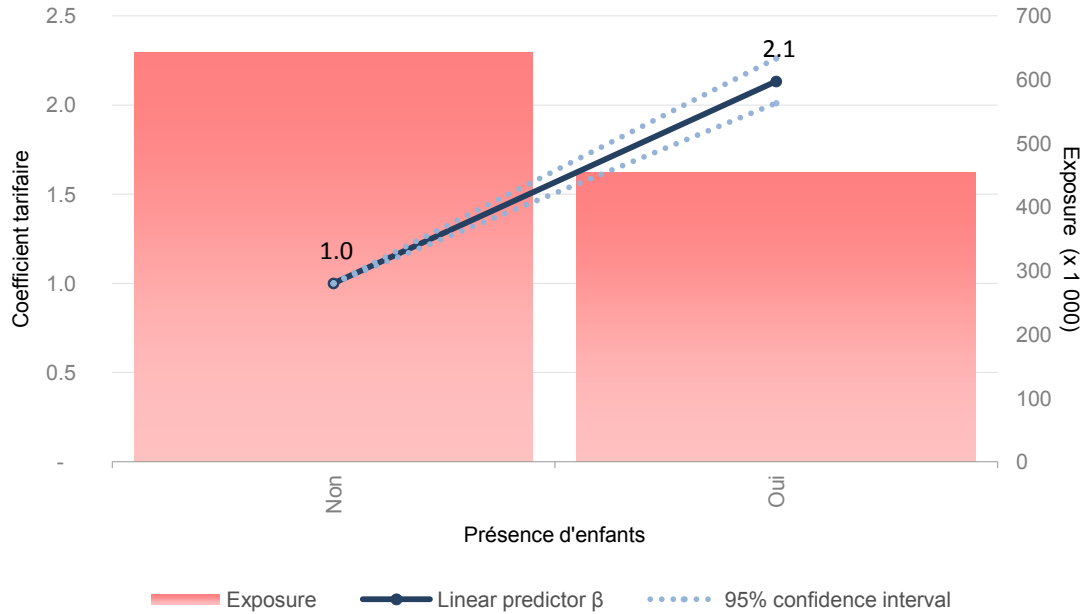


Figure 61 : Coefficient tarifaire de la fréquence RC suivant la présence d'enfants

Pour la modélisation de la fréquence de la garantie Responsabilité Civile (RC), le graphique ci-dessus présente le coefficient tarifaire en fonction de la présence d'enfants dans l'habitation. Le choix de cette variable n'est pas laissé au hasard dans la mesure où elle dispose de la plus grande significativité dans le modèle. La présence de jeunes enfants implique une augmentation de la probabilité d'être sinistré en garantie RC de +110% du à leur inattention et aux accidents domestiques.

e. Modélisation du coût moyen

Les résultats obtenus sur la modélisation du coût moyen sur quelques garanties et variables sont présentés dans ce paragraphe.

Modélisation du coût moyen sur la garantie Vol en fonction du niveau des capitaux objets de valeur

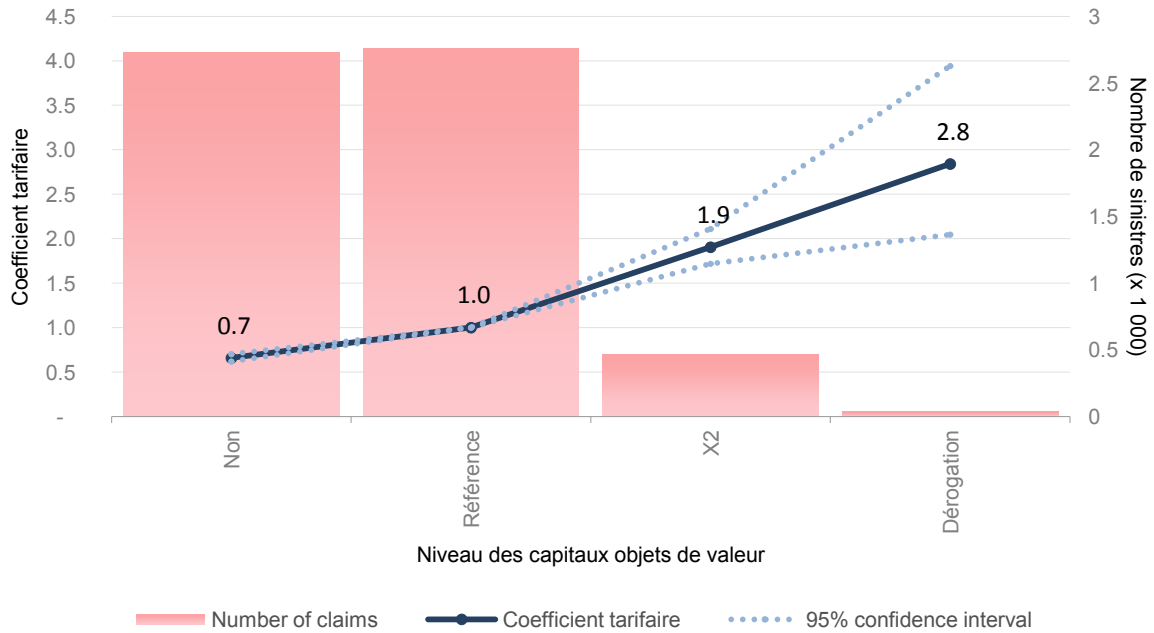


Figure 62 : Coefficient tarifaire du coût moyen Vol suivant les capitaux objets de valeur

Le graphique ci-dessus présente la modélisation du coût moyen sur la garantie Vol en fonction du niveau des capitaux objets de valeurs (bijoux, montres, tableaux, etc.) souscrit par le client. Ces capitaux dépendent de la formule et du nombre de pièces du client. Le client a la possibilité de doubler ses capitaux, voire d'accéder à des niveaux supérieurs sur demande auprès de la plateforme de gestion de contrats RCC. Ainsi, la modalité X2 engendre un quasiment doublement (+90%) du coût d'un sinistre Vol par rapport à la modalité de référence. A l'inverse, si le client recherche un meilleur positionnement tarifaire, il peut choisir une valeur nulle pour ses objets de valeur et ainsi le coût moyen diminuera d'environ 30% par rapport au niveau de référence. Enfin, une certaine volatilité sur le niveau des coefficients est observée pour les modalités peu représentées (X2 et Dérogation).

Modélisation du coût moyen sur la garantie Incendie en fonction de la présence d'un insert

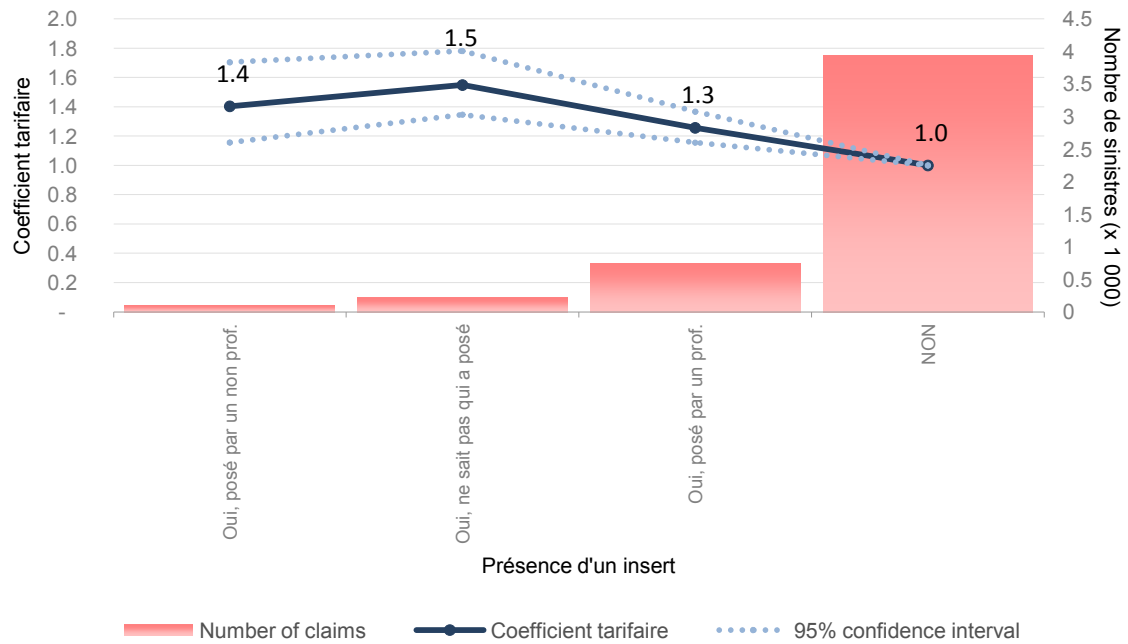


Figure 63 : Coefficient tarifaire du coût moyen Incendie suivant la présence d'un insert

La présence d'insert sur le coût moyen d'un sinistre Incendie est discriminante, comme l'indique le graphique supra. La présence d'un insert dans l'habitation a un impact compris entre +30% et +50% sur le coût moyen Incendie d'une habitation sans insert. La segmentation entre insert posé par un professionnel ou un non professionnel présente également un impact sur le coût moyen en question supérieur à +10%.

f. Modélisation de la prime pure

Une fois modélisée séparément la fréquence et le coût moyen, nous procédons à la consolidation de la prime pure par garantie. Les mêmes facteurs explicatifs ont été conservés pour chaque couverture.

Quelques résultats sont présentés ci-dessous.

Modélisation de la prime pure sur la garantie Vol en fonction du statut-type et de l'étage

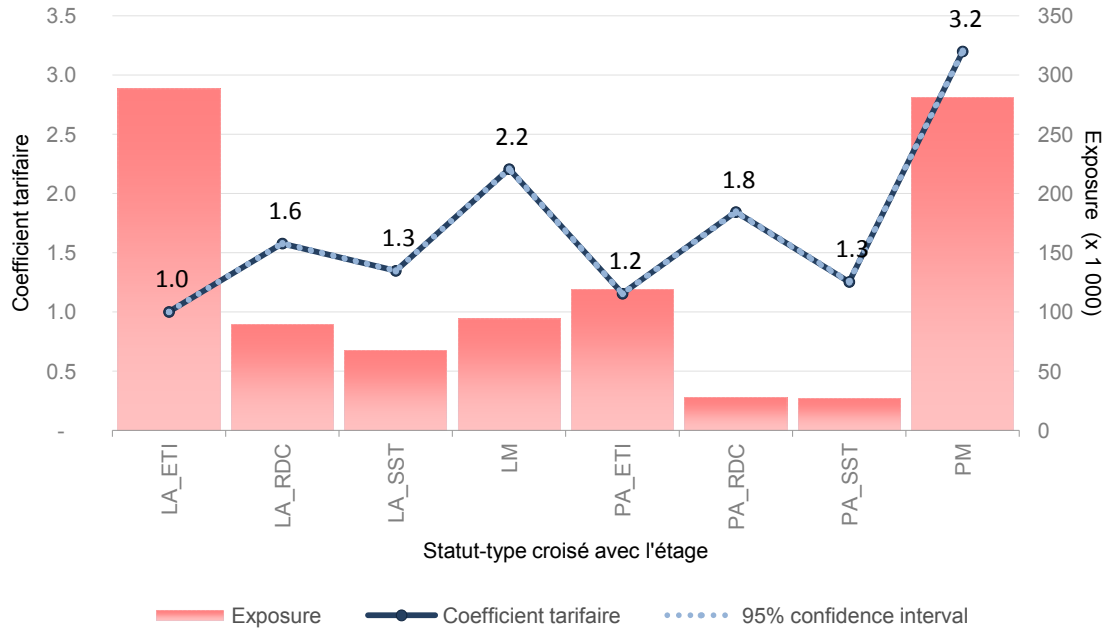


Figure 64 : Coefficient tarifaire de la prime pure Vol en fonction du statut-type et de l'étage

L'analyse de ce graphique permet d'aboutir à plusieurs constats :

- Les maisons, locataires et propriétaires, présentent un coût du risque supérieur, respectivement de +120% et +220%, à la modalité de référence (locataire d'appartement en étage intermédiaire). En effet, les maisons sont plus vulnérables que les appartements face au risque de vol ;
- Par ailleurs, les logements situés au rez de chaussée engendrent une augmentation du coût du risque d'environ +60% (pour les locataires et propriétaires d'appartements) par rapport à un logement situé en étage intermédiaire. Ce phénomène s'explique par l'accès plus aisé aux fenêtres (côté rue ou côté jardin) ;
- Enfin, les appartements sous les toits présentent un risque intermédiaire, supérieur aux logements en étage intermédiaire mais inférieur aux habitations en rez de chaussée. Ici aussi, le cambrioleur peut préférer une habitation au dernier étage car il y aura moins d'aller et venue qu'une habitation en étage inférieur.

Modélisation de la prime pure sur la garantie Détériorations suite à vol en fonction du type de résidence

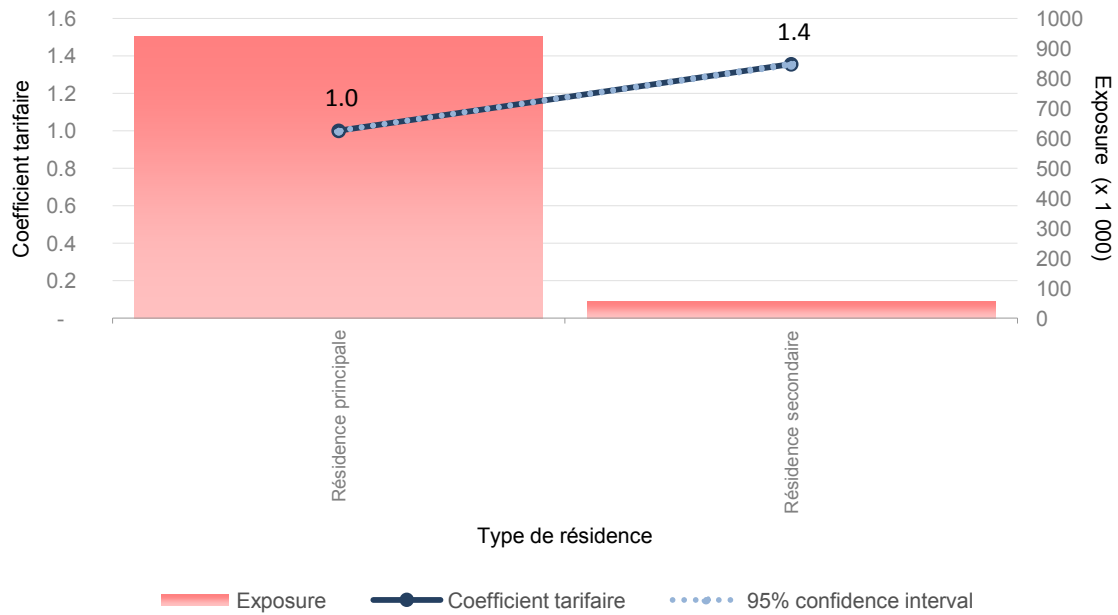


Figure 65 : Coefficient tarifaire de la prime pure Dégâts suite à vol suivant le type de résidence

La prime pure de la garantie Détériorations suite à vol augmente de +40% pour une résidence secondaire par rapport à une résidence principale. La fréquence et le coût moyen de sinistres de type vandalisme progresse pour les résidences secondaires dans la mesure où elles sont, par définition, moins habités par leurs occupants. Ainsi, un cambrioleur disposera de davantage de temps pour « explorer » l'habitation contrairement à une résidence principale.

Modélisation de la prime pure sur la garantie Dégât des eaux en fonction de l'année de construction

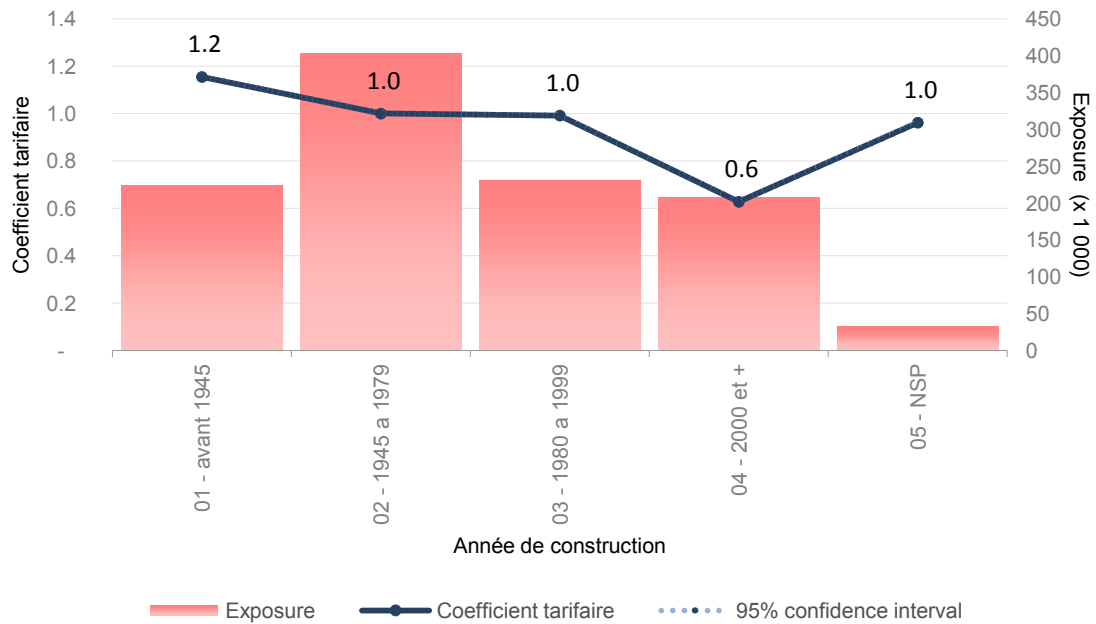


Figure 66 : Coefficient tarifaire de la prime pure Dégât des eaux suivant l'année de construction

La garantie Dégât des eaux en fonction de l'année de construction présente une augmentation du coût du risque en fonction de la vétusté de l'habitation. Pour les immeubles et maisons d'avant 1945, le coût du risque est 20% supérieur à une habitation dont la construction est comprise entre 1945 et 1979. Pour les logements datant de 2000 et plus, le risque dégât des eaux diminue d'environ 40%. La vétusté d'une habitation peut engendrer des fuites plus fréquentes dans les canalisations contrairement à un logement récent.

Modélisation de la prime pure sur la garantie Tempête Grêle Neige en fonction de la superficie des dépendances

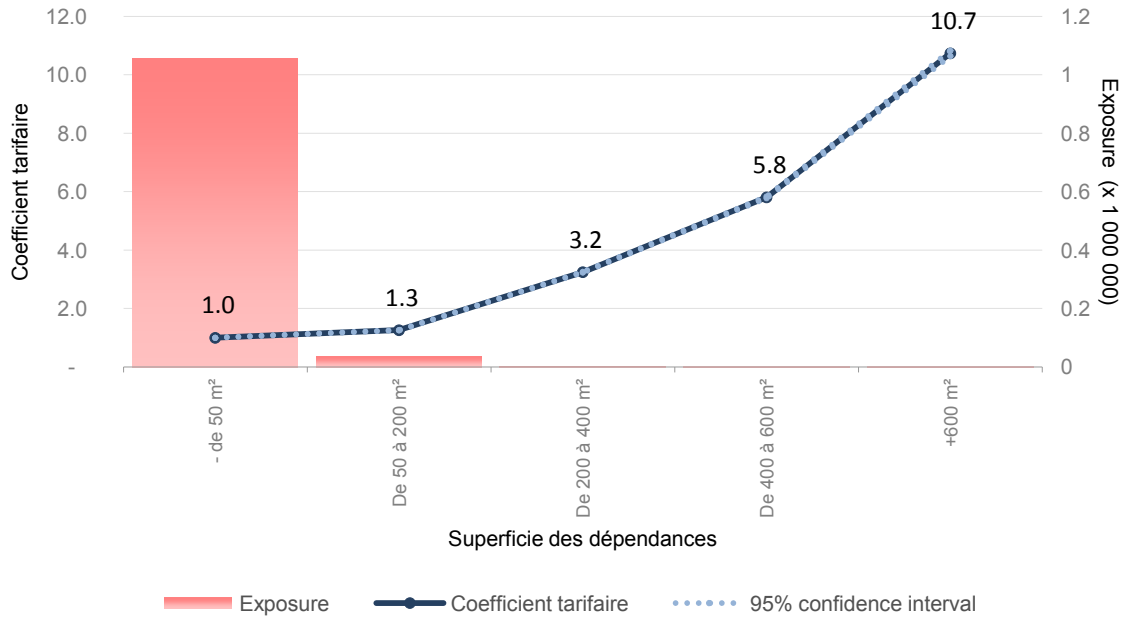


Figure 67 : Coefficient tarifaire de la prime pure Tempête grêle neige suivant la taille des dépendances

Le coût du risque de la garantie Tempête, Grêle, Neige progresse en fonction de la superficie des dépendances. Pour des dépendances de 50 à 200 m², la prime pure augmente de 30% par rapport à la modalité de référence (- de 50 m²). La croissance est exponentielle pour les superficies de dépendances supérieures. L'augmentation du coût du risque de la garantie tempête en fonction de la superficie des dépendances est logique puisque la taille de l'habitation assurée s'accroît. Une explication supplémentaire peut provenir de la vulnérabilité des dépendances qui peut être supérieure à un logement traditionnel (du fait d'une construction moins solide) et donc amplifier le coût des sinistres tempête sur cette typologie de locaux.

En synthèse, une comparaison entre la prime pure observée et la prime pure modélisée a été réalisée et conclut à une bonne approximation des données par la modélisation (écart de -0.1% entre les deux primes pures).

Garantie	Observé							Modélisé	
	Années risques	Nb sinistres	Charge totale	Fréquence	Coût moyen	Prime pure	Prime pure	Ecart	
Bris de vitres	742 478	5 138	2 689 540	0.7%	523 €	3.6	3.6	-0.7%	
Dégât des eaux	1 096 856	57 713	55 730 153	5.3%	966 €	50.8	50.0	-1.7%	
Dommmages électriques	327 242	1 925	1 284 840	0.6%	667 €	3.9	3.9	-0.4%	
Détériorations suite à vol	994 613	6 307	7 774 962	0.6%	1 233 €	7.8	7.5	-3.8%	
Incendie et événements assimilés	1 096 856	5 019	41 646 402	0.5%	8 298 €	38.0	38.8	2.1%	
Responsabilité civile	1 096 856	6 347	15 559 438	0.6%	2 451 €	14.2	14.3	0.8%	
Tempête, grêle, neige	1 096 856	7 607	11 238 449	0.7%	1 477 €	10.2	9.9	-3.6%	
Vol	994 613	5 995	15 903 722	0.6%	2 653 €	16.0	16.4	2.8%	
Total	1 096 856	96 051	151 827 506	9.5%	1 456 €	138.4	138.2	-0.1%	

Figure 68 : Synthèse de la prime pure observée et modélisée par garantie

6. Calcul de la prime commerciale

Au modèle de prime pure (PP) détaillé précédemment s'ajoute les garanties et chargements suivants :

- Garanties ajoutées au modèle de tarif :
 - Garantie Catastrophes naturelles : le tarif de cette garantie est réglementaire et se calcule comme suit : $12\% * (prime\ pure_{Incendie} + prime\ pure_{Dégât\ des\ eaux} + prime\ pure_{Vol} + prime\ pure_{Vandalisme} + prime\ pure_{Dommages\ électriques} + prime\ pure_{Bris\ de\ vitres} + prime\ pure_{Tempête\ grêle\ neige})$;
 - Garantie Dommages corporels (DC) : étant donné les faibles volumes visualisés sur la période d'analyse, la prime pure de cette garantie est égale à la sinistralité dossier à dossier observé ;
 - Garanties Assistance et DPRSA : ces garanties n'étant pas assurées par Sogessur, leur tarif est déterminé par les partenaires et est fixe par formule de garanties.
- Chargements :
 - Commission distributeur (ComDist) : la commission diffère suivant qu'il s'agisse d'une affaire nouvelle ou du renouvellement du contrat. Elle s'établit respectivement à $x\%$ et $y\%$ de la prime avec $x > y$;
 - Réassurance (pour la sinistralité de pointe et la réassurance de la garantie Catastrophes Naturelles avec la CCR) : $x\%$ de la prime ;
 - Frais généraux (FG) : compris entre $x\%$ et $y\%$ de la prime suivant les formules de garanties. Cette variabilité s'explique par le fait que les actes de gestion seront plus nombreux sur un contrat haut de gamme du fait d'une fréquence de sinistres supérieure ;
 - Produits financiers (Pfi) : $-x\%$ de la prime ;
 - Taxe d'assurances : 9% excepté pour les garanties Incendie (30%) et Assistance (20%). La taxe attentat de 5.90€ est également ajoutée au tarif commercial.

L'expression mathématique pour la formule Optimale du tarif commercial TTC s'écrit comme suit (avec une simplification sur la présentation du taux de taxe qui est calculé en réalité pour chaque garantie dans le mémoire) :

$$Prime_{com}^{TTC} = \frac{(PP_{modèle} + PP_{Cat.Nat.} + PP_{DC} + PP_{Assistance} + PP_{DPRSA})}{1 - ComDist - Réass - FG + Pfi} \times (1 + TxTaxe_{moyen}) + Taxe_{Att}$$

7. Estimation de la marge

Le tarif construit dans les paragraphes IV.5 et IV.6 est ensuite appliqué ligne à ligne aux critères de risques de notre portefeuille de contrats termés en 2017 à partir d'une macro développée en *Visual Basic*.

Cette routine applique les coefficients du modèle GLM aux caractéristiques spécifiques de chaque contrat et ajoute les garanties additionnelles (Catastrophes naturelles, Dommages corporels, Assistance et Défense pénale et recours), les chargements et taxes définis précédemment.

La modélisation de la prime commerciale ayant été limitée au portefeuille des contrats de la nouvelle offre et aux échéances de l'année 2017, l'analyse porte sur une base de plus de 310 000 lignes pour un chiffre d'affaires de près de 92 M€.

Le calcul de la marge se détermine, ligne à ligne, en faisant la différence entre la prime terminée actuellement en portefeuille et la prime commerciale modélisée correspondant à l'équilibre technique.

La marge totale s'établit à 1.2 M€. Cette marge moyenne par contrat de 4€, à rapporter à une prime moyenne TTC de 296€, représente 1.3% de la prime. La marge calculée aboutit ainsi à un ratio combiné technique de 99%, niveau qui apparaît en ligne avec les estimations de rentabilité du produit Habitation issues des travaux de provisionnement en vision économique.

Après agrégation des niveaux de marges moyens par contrat suivant différentes variables rapportés aux primes moyennes, les graphiques 69 et 70 peuvent être construits :

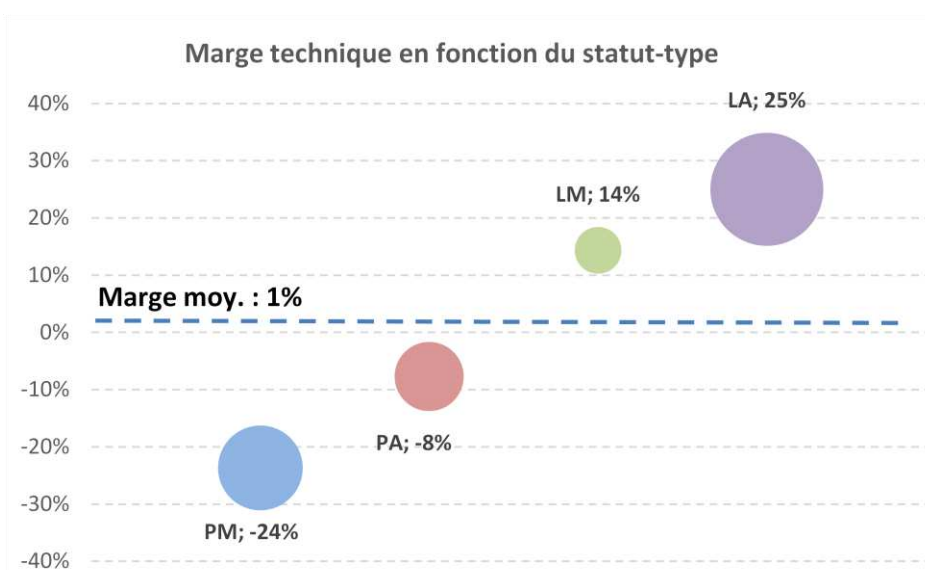


Figure 69 : Marge moyenne en fonction du statut-type

La marge moyenne segmentée par statut d'occupation croisée avec la nature du bien aboutit aux constats suivants :

- Les propriétaires de maisons et d'appartements ne sont pas rentables à hauteur respectivement de -24% et -8%. La justification liée à cette perte sur cette typologie de clients est liée à plusieurs effets. Tout d'abord, le segment des propriétaires de maisons est particulièrement touché par les événements climatiques et les catastrophes naturelles. Or, certains lissages dans les coefficients tarifaires associés à ce statut-type ont atténué l'effet

des tempêtes et donc dégradé la rentabilité. Par ailleurs, ce segment des propriétaires est une cible privilégiée par les distributeurs bancaires car il s'agit de clients généralement plus fidèles et plus patrimoniaux ;

- A l'inverse, les locataires de maisons et d'appartements apparaissent sur-tarifés, respectivement de +14% et +25%. Le taux de transformation élevé et les primes moyennes plus faibles sur cette typologie de clients sont des éléments de justification de ce niveau tarifaire, tout comme le principe de mutualisation avec les propriétaires ;
- Les poids de chaque statut-type sont représentés par la taille des bulles de la figure 69. Ainsi, les locataires d'appartements sont le segment le plus visible dans la base.

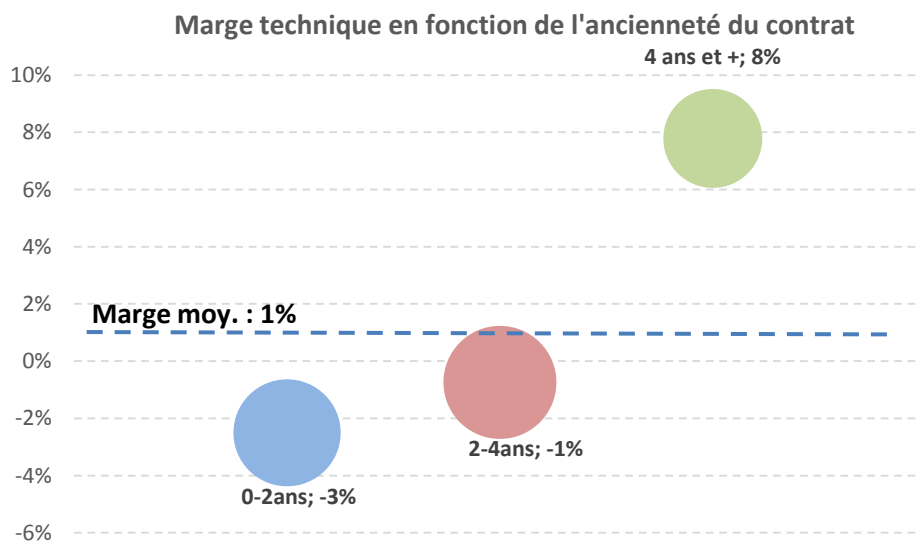


Figure 70 : Marge moyenne en fonction de l'ancienneté du contrat

La marge moyenne en fonction de l'ancienneté du contrat nous amène aux constats suivants. Les contrats anciens en portefeuille, à partir de 4 ans, financent les affaires nouvelles. En effet, la compagnie perd en moyenne 3% de la prime pour les contrats récents (inférieurs à 2 ans). Entre 2 et 4 ans, le point d'équilibre est quasiment atteint (-1%) et après 4 ans, la rentabilité s'améliore nettement (+8%). Le marché de l'assurance habitation étant concurrentiel en France, les marges de manœuvres sur le tarif affaires nouvelles sont limitées pour les assureurs. Ainsi, la rentabilité s'acquière au fil des années grâce au renouvellement des contrats et aux politiques tarifaires associées.

Le volume des contrats ayant une ancienneté supérieure à 4 ans n'est pas si important dans la mesure où la base étudiée dans ce chapitre ne contient que des contrats de l'offre commercialisée depuis fin 2010 donc avec une ancienneté maximum de 7 ans vue à fin 2017.

Un troisième graphique de marge présentant l'analyse du statut-type croisé avec la formule de garanties est disponible en annexe 12. Il indique une stabilité de la rentabilité ou du déficit des statut-type en fonction des formules.

V. Optimisation de la stratégie de renouvellement

L'outil *Addactis Pricing* module *simulation* est utilisé pour optimiser la politique tarifaire au terme pour l'année N+1 à partir des informations connues en N. Le modèle de sensibilité au prix issu de la régression logistique défini au chapitre III sert d'input pour l'optimisation du chiffre d'affaires sous contrainte d'un nombre de contrats renouvelés. Pour optimiser la marge sous contrainte d'un volume d'affaires ou du chiffre d'affaires, il est ajouté le modèle tarifaire GLM défini au IV. La figure 71 résume les modèles utilisés précédemment et les optimisations de la politique tarifaire qui seront développées dans ce chapitre.

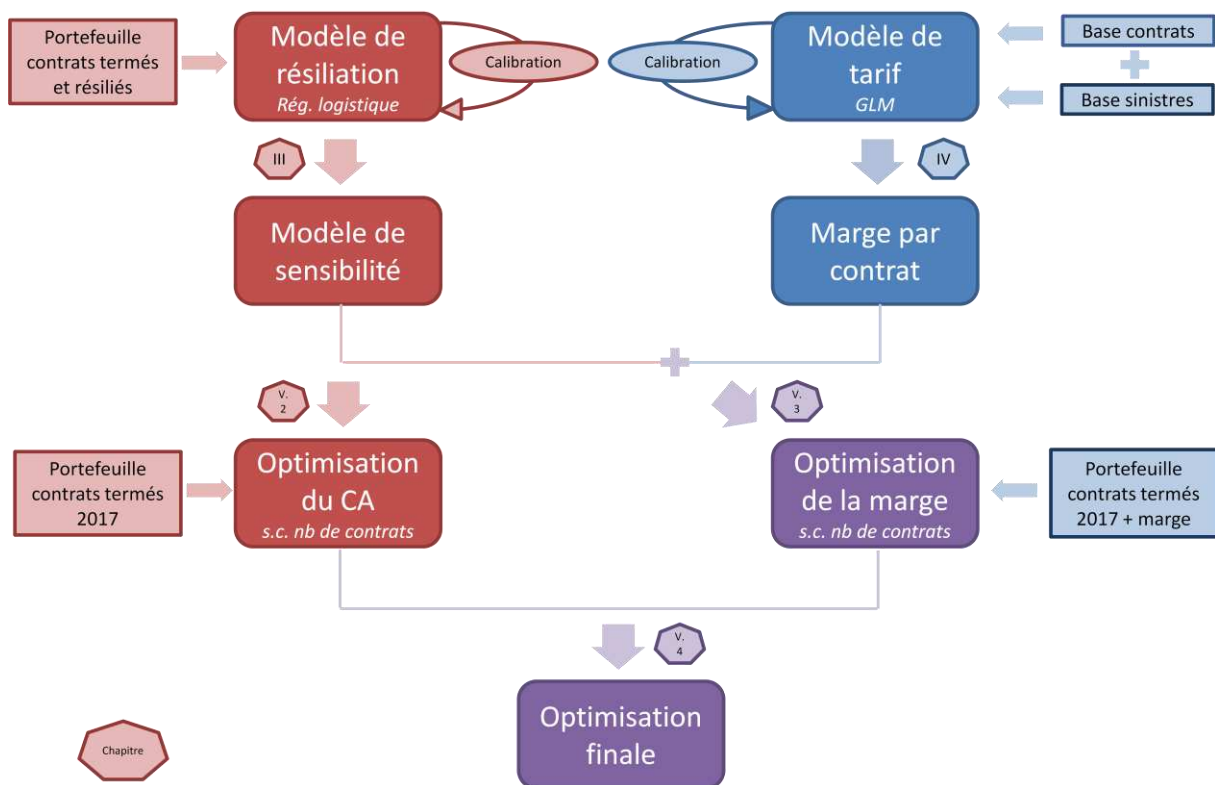


Figure 71 : Synthèse des modèles et des étapes pour l'optimisation de la stratégie de renouvellement

1. Définition des approches d'optimisation

Deux approches d'optimisations seront étudiées dans ce paragraphe : une approche non paramétrique et l'approche paramétrique actuellement en vigueur chez *Societe Generale Insurance*.

a. Approche non paramétrique

L'approche non paramétrique est une méthode quadratique qui consiste à transformer le problème en un programme d'optimisation sous contraintes. Sous respect d'un niveau de contrats terminés et de

l'encadrement de la majoration de prime entre un minimum et un maximum définis à dire d'expert, le programme va proposer la majoration adéquate pour chaque client en portefeuille tenant compte de ses caractéristiques propres. Cette optimisation conduit à des résultats optimaux même si en pratique elle semble difficile à mettre en œuvre pour un bancassureur puisque chaque client verra sa prime varier individuellement dans de fortes amplitudes. Dans ce cas, l'explication de la politique tarifaire auprès du distributeur serait ardue (cf. 1.3).

Mathématiquement, le programme d'optimisation Z s'écrit sous la forme :

$$Z \begin{cases} \text{Max}_{PS_1, \dots, PS_n} f(PS_1, \dots, PS_n) \\ \text{Sous contraintes : } \forall i \in [1, \dots, n] \sum_{i=1}^n (1 - \pi_i) \geq \text{Nombre de contrats} \\ \text{et } PS_{inf_i} \leq PS_i \leq PS_{sup_i} \end{cases}$$

Avec :

- f la fonction d'optimisation (chiffre d'affaires ou marge) ;
- PS_i la prime simulée pour l'individu i ;
- π_i la probabilité de résiliation d'un individu i ;
- PS_{inf_i} et PS_{sup_i} respectivement les primes minimums et maximums pour chaque individu (ie : la prime terminée au quelle on ajoute le minimum ou maximum de majoration en euro).

Le programme Z peut également s'écrire comme suit :

$$Z \begin{cases} \text{Max}_{\tilde{y}} f(\tilde{y}) \\ \text{Sous contraintes : } \forall j \in [1, \dots, m] g_j(\tilde{y}) \leq c_j \end{cases}$$

Où :

- g_j les fonctions contraintes ;
- $c_j \in \mathbb{R}$;
- m correspond aux nombres de contraintes.

On introduit la fonction Lagrangien, fonction permettant l'étude des problèmes d'optimisation avec contrainte, définie par :

$$L(\tilde{y}, \tilde{\lambda}) = f(\tilde{y}) - \sum_{j=1}^m \lambda_j [g_j(\tilde{y}) - c_j]$$

Avec : λ_j les multiplicateurs de Lagrange associés aux contraintes.

Plusieurs méthodes de résolution existent pour résoudre ce programme d'optimisation (algorithme du simplexe, du gradient, du lagrangien,...) mais ne seront pas détaillées dans ce mémoire.

b. Approche paramétrique

L'approche paramétrique repose sur le choix par l'utilisateur des variables ordinales (statut d'occupation, ancienneté du contrat, âge de l'assuré, ...) qui permettront de construire la matrice de terme et les majorations associées. La prime est toujours encadrée entre un minimum et un maximum en euro. Le choix des variables sera donc primordial et reposera sur les analyses effectuées dans le chapitre relatif à la sensibilité au prix. En se fixant une contrainte d'un taux de résiliation donné, la maximisation du chiffre d'affaires ou de la marge sera déterminée en traçant la frontière efficiente.

Cette notion de frontière efficiente a été introduite par Harry Max Markowitz en 1952 dans le cadre d'un portefeuille de gestion d'actifs où l'on représente le couple (risque, rendement). L'idée est de déterminer un portefeuille qui maximise le rendement attendu pour un niveau de risque donné.

2. Optimisation du chiffre d'affaires sous contrainte d'un nombre de contrats renouvelés

a. Présentation de la base de simulation

La base de données utilisée dans l'optimisation du chiffre d'affaires est celle détaillée dans le chapitre II.1 avec les amendements suivants :

- Réintégration des affaires nouvelles de 2017 dans la base car l'intérêt est de simuler une politique tarifaire N+1 donc comprenant les contrats exposés l'année N (affaires nouvelles ou contrats renouvelés) ;
- Limitation aux contrats terminés en 2017 (les termes précédents ne nous intéressent plus car l'objectif est de partir de l'exposition la plus récente pour récupérer la dernière prime payée par le client et ensuite simulée une nouvelle prime) ;
- Suppression des résiliations de la base pour tenir compte des contrats en cours uniquement.

La base constituée comprend ainsi plus de 394 000 lignes et présente un chiffre d'affaires y compris taxes de 130 M€.

b. Fonction d'optimisation du chiffre d'affaires

La fonction d'optimisation cherche à maximiser le chiffre d'affaires (CA) sous contrainte d'un taux de résiliation (équivalent à un nombre de contrats renouvelés) et d'un encadrement de la majoration en euro par contrat.

Le programme d'optimisation s'écrit :

$$\left\{ \begin{array}{l} \text{Optimisation}_{CA} = \text{Prime}_{\text{simulée}} \times [1 - \text{modele}_{\text{sensi}}(\text{Prime}_{\text{simulée}} - \text{Prime}_{\text{avant}_{\text{terme}}})] \quad (1) \\ \text{Sous contraintes (s.c.) : } Tx_{\text{renouvelé}} = [1 - \text{modele}_{\text{sensi}}(\text{Prime}_{\text{simulée}} - \text{Prime}_{\text{avant}_{\text{terme}}})] \quad (2) \\ \text{et } \text{Prime}_{\text{inf}} \leq \text{Prime}_{\text{simulée}} \leq \text{Prime}_{\text{sup}} \quad (3) \end{array} \right.$$

La formule (1) correspond au chiffre d'affaires pondéré par la probabilité de renouveler son contrat. La $\text{Prime}_{\text{simulée}}$ est la prime calculée lors de la simulation sous contrainte de l'encadrement de prime défini en (3). Le paramètre en argument du modèle de sensibilité correspond bien à une variation de prime en euro puisque il s'agit de la différence de deux primes (prime simulée et prime terminée lors du précédent terme).

La formule (2) est un indicateur donnant une probabilité de renouveler son contrat et correspond à la contrainte de volume de contrats restants en portefeuille.

La contrainte (3) sert à définir un niveau minimum et maximum de la nouvelle prime pour éviter que le modèle n'alloue des variations trop importantes aux segments à optimiser. En effet, majorer de +10% un client payant une prime de 650€ reviendrait à augmenter sa police de 65€, ce qui paraît élevé. Ainsi, le minimum de revalorisation en euro défini dans la matrice actuelle et dépendant de la formule et de la sinistralité a été paramétré à 0€ (car l'arbre de décision construit dans la partie III nous a indiqué que les assurés pouvaient être sensibles à partir de 1€ de majoration) et le maximum de revalorisation en euro à 40€ pour ne pas trop contraindre le modèle d'optimisation.

c. Matrice paramétrique actuelle

La matrice actuellement en production chez *Societe Generale Insurance* croise les variables suivantes : statut-type, sinistralité et ancienneté.

Statut-type	Sinistralité	Ancienneté	Majorations
Locataire d'appartement	Non sinistré	0-3 ans	3.0%
		3 ans et +	3.6%
	Sinistré	0-3 ans	5.8%
		3 ans et +	6.5%
Locataire de maison	Non sinistré	0-3 ans	3.0%
		3 ans et +	3.6%
	Sinistré	0-3 ans	5.8%
		3 ans et +	6.5%
Propriétaire d'appartement	Non sinistré	0-3 ans	4.0%
		3 ans et +	3.4%
	Sinistré	0-3 ans	6.8%
		3 ans et +	6.3%
Propriétaire de maison	Non sinistré	0-3 ans	4.0%
		3 ans et +	3.4%
	Sinistré	0-3 ans	6.8%
		3 ans et +	6.3%

Figure 72 : Optimisation du CA - approche paramétrique – allocations de la matrice actuelle

Cette stratégie de majorations aboutit à une revalorisation moyenne de +3.7% du chiffre d'affaires du portefeuille pour un taux de résiliation de 6.6%. Dans la suite de la partie V.2, ce taux de résiliation sera utilisé comme contrainte et une optimisation du chiffre d'affaires sera recherchée.

d. Matrice paramétrique actuelle optimisée

Dans ce paragraphe, la contrainte est de conserver les mêmes variables que celles utilisées actuellement dans la matrice de politique tarifaire (notamment pour des raisons de délais ou de coûts informatiques induits en utilisant de nouvelles variables). Sous ces conditions, une optimisation du gain au terme est-elle possible ?

Statut-type	Sinistralité	Ancienneté	Majorations
Locataire d'appartement	Non sinistré	0-3 ans	7.0%
		3 ans et +	7.4%
	Sinistré	0-3 ans	6.0%
		3 ans et +	6.8%
Locataire de maison	Non sinistré	0-3 ans	7.0%
		3 ans et +	7.3%
	Sinistré	0-3 ans	6.0%
		3 ans et +	6.6%
Propriétaire d'appartement	Non sinistré	0-3 ans	1.0%
		3 ans et +	2.0%
	Sinistré	0-3 ans	1.0%
		3 ans et +	2.0%
Propriétaire de maison	Non sinistré	0-3 ans	1.0%
		3 ans et +	2.0%
	Sinistré	0-3 ans	1.0%
		3 ans et +	2.0%

Figure 73 : Optimisation du CA - approche paramétrique – allocations optimisées sous contraintes des mêmes variables

Cette matrice utilise les résultats du modèle de résiliation et de sensibilité au prix. La segmentation définie ci-dessus propose de majorer davantage les locataires puisque les propriétaires sont plus élastiques au prix. Par ailleurs, les clients anciens en portefeuille (plus de 3 ans) sont également mis à contribution.

Cette stratégie de majorations aboutit à une revalorisation moyenne de +4.0% du chiffre d'affaires du portefeuille pour un taux de résiliation de 6.6%, soit un gain de chiffre d'affaires de +0.6M€ par rapport à la matrice actuelle présentée dans le paragraphe précédent.

e. Matrice paramétrique optimisée sans contrainte sur les variables

Pour la construction de la matrice optimisée, les variables les plus sensibles au prix définies au III.3.b ont été testées :

- Age x classe contact RCC x ancienneté x top personnel salarié SG ;
- Age x classe contact RCC x ancienneté x statut d'occupation.

Le fait d'être salarié du groupe Société Générale n'a pas été retenu dans l'optimisation dans la mesure où cela engendrait une majoration supérieure pour les salariés du groupe par rapport à un client lambda, ce qui irait à l'encontre de la politique interne de promotion des offres d'assurance auprès des salariés.

Ainsi, la matrice optimisée retenue correspond aux croisements des variables âge avec la typologie de contact auprès de RCC, l'ancienneté du contrat et le statut d'occupation. Par ailleurs, l'ajout de la variable statut d'occupation (propriétaires, locataires) s'explique par son côté discriminant dans l'identification des segments à marge (cf. IV.7).

La simulation a conduit à 23 040 combinaisons de majorations pour les modalités des variables précitées. Par pas de 0.5pt entre +0% et +10% (au-delà le niveau de majoration peut être considéré comme trop élevé pour l'assuré), l'outil teste toutes les combinaisons et les affiche dans le graphique 74. Cette figure présente l'ensemble de ces combinaisons et le point optimum choisi sur la frontière efficiente correspondant à un nombre de contrats renouvelés de 368 100, soit un taux de résiliation de 6.6%.

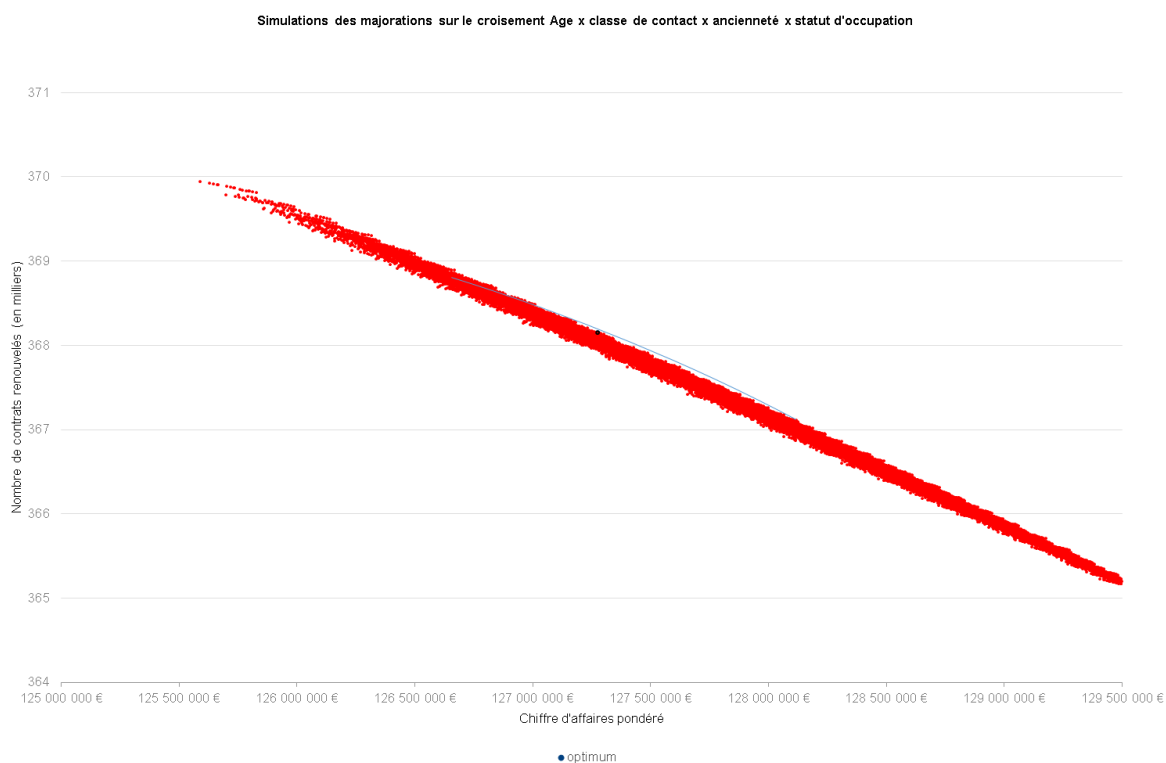


Figure 74 : Optimisation du CA - approche paramétrique – simulations de majorations sans contraintes

Pour le taux de résiliation défini supra, la stratégie de majoration optimum associée à ces quatre variables est détaillée ci-dessous :

Variables	Modalités	Majorations
Age	18-37 ans	0%
	37-60 ans	0%
	60-80 ans	4.0%
	80 ans et +	6.9%
Type contact RCC	Pas de contact / risque faible	0.5%
	Contacts à risque moyen	0%
	Contacts à risque élevé	0%
Ancienneté	0-2 ans	0%
	2-4 ans	0%
	4-9 ans	1.0%
	9-16 ans	5.0%
	16 ans et +	6.0%
Statut d'occupation	Locataire	3.0%
	Propriétaire	0%

Figure 75 : Optimisation du CA - approche paramétrique – allocation optimisée sans contraintes

Un assuré âgé se verra appliquer une majoration supérieure à une personne plus jeune dans la mesure où ce segment est moins élastique au prix. Il en est de même pour les clients anciens en portefeuille, moins sensibles au prix, qui reçoivent des augmentations supérieures ainsi que les locataires. Enfin, les assurés ayant contacté récemment la plateforme de gestion pour des motifs à risque moyen ou élevé sont également ménagés.

Le tableau se lit de la manière suivante : un propriétaire de 60 ans, n'ayant pas eu de contact avec la plateforme de gestion et avec une ancienneté de contrat de 8 ans se verra appliquer la majoration suivante : $(1+4\%) \times (1+0.5\%) \times (1+1\%) \times (1+0\%) = +5.5\%$

Cette stratégie de majorations aboutit à un taux de résiliation de 6.6% et une prime après terme de 136.1 M€ (+4.6% de revalorisation) représentant un gain de chiffre d'affaires de +1.2 M€ par rapport à la matrice actuellement en production.

f. Approche non paramétrique

La construction de l'approche non paramétrique aboutit à des variations différentes pour chaque ligne de la base puisque le modèle essaie d'optimiser chaque caractéristique des variables présentes dans le modèle de sensibilité au prix. Sous contrainte d'un encadrement de la prime minimum et maximum, la majoration d'un contrat peut évoluer de +0% à +69% (cf. figure 76).

Optimisation du chiffre d'affaires- distribution des majorations de prime pour la simulation non paramétrique

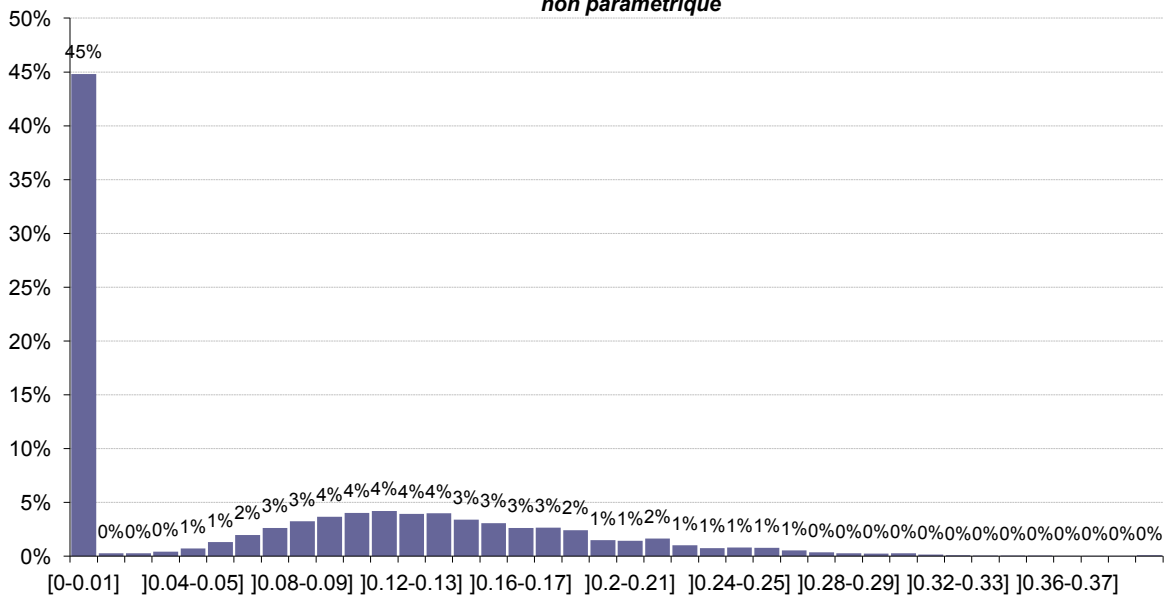


Figure 76 : Optimisation du CA - approche non paramétrique - distribution des majorations de primes

Le graphique 77 présente quelques résultats pour les variables équipement du client, type de résidence et salarié du groupe Société Générale. Les clients les moins sensibles au prix (habitant en résidence secondaire, client équipé de plusieurs contrats d’assurances ou salarié du groupe) se voient appliquer les majorations les plus fortes.

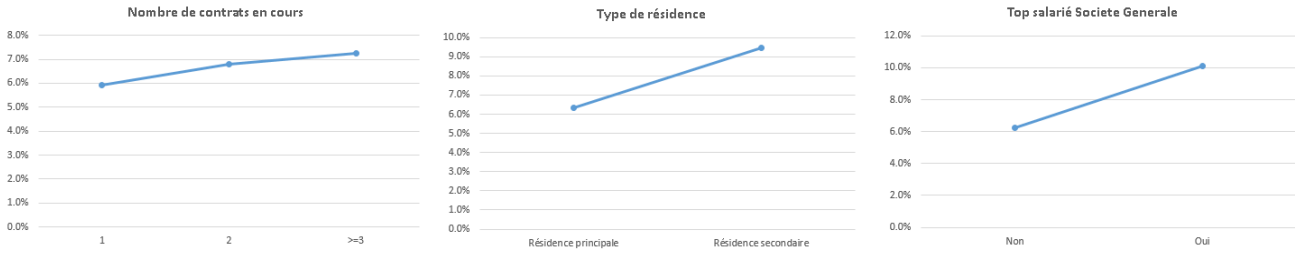


Figure 77 : Optimisation du CA - approche non paramétrique - zoom sur quelques variables

Cette approche non paramétrique aboutit à une revalorisation moyenne de +6.5% du chiffre d’affaires du portefeuille pour un taux de résiliation de 6.6%.

g. Résultats des stratégies pour l'optimisation du chiffre d'affaires

En synthèse, quatre stratégies de majorations pour l'optimisation du chiffre d'affaires ont été étudiées dans cette partie et sont résumées dans le graphique et tableau ci-dessous :

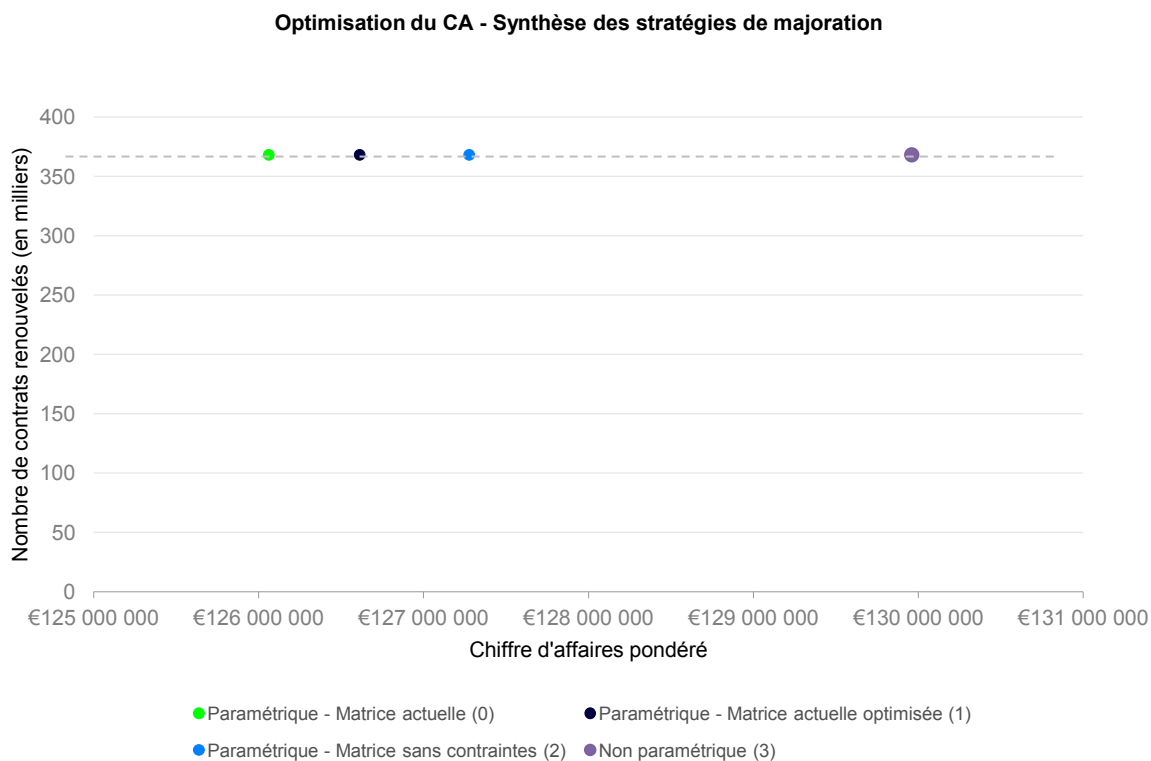


Figure 78 : Synthèse graphique des stratégies de majorations suivant l'axe CA

Pour un même volume de contrats renouvelés (environ 368 100 correspondant à un taux de résiliation de 6.6%), ce graphique permet de comparer l'étendue de chiffre d'affaires supplémentaires entre les scénarios :

Stratégie	Libellé	Type	Taux résiliation	Prime après terme	Revalorisation du CA	CA pondéré	Gain CA pondéré
0	Matrice actuelle		6.6%	134.9M€	3.7%	126.1M€	
1	Matrice actuelle optimisée	Paramétrique	6.6%	135.3M€	4.0%	126.6M€	.6M€
2	Matrice sans contraintes		6.6%	136.1M€	4.6%	127.3M€	1.2M€
3		Non paramétrique	6.6%	138.6M€	6.5%	130.0M€	3.9M€

Figure 79 : Synthèse des stratégies de majorations suivant l'axe CA et gains associés

Le scénario 2 correspondant à la matrice sans contraintes sur les variables permet d'afficher un gain comparé au scénario 0 de +0.9 point sur la revalorisation représentant un surplus de chiffre d'affaires pondéré par la probabilité de renouvellement de +1.2 M€.

Le scénario 3, non paramétrique, permettrait d'obtenir de meilleurs résultats puisque le gain en matière de revalorisation s'élève à +2.8 points, soit +3.9 M€. Néanmoins, comme expliqué précédemment, ce scénario aboutit à une majoration spécifique pour chaque client puisqu'elle dépend des caractéristiques propres de chaque assuré et serait donc ardue à expliquer aux distributeurs bancaires.

3. Optimisation de la marge sous contrainte d'un nombre de contrats renouvelés

Regardons désormais sous l'angle de l'optimisation de la marge sous contrainte d'un nombre de contrats renouvelés. En effet, optimiser le chiffre d'affaires est intéressant pour une compagnie d'assurances mais conserver des contrats rentables peut s'avérer une stratégie plus judicieuse.

a. Présentation de la base de simulation

La base de données utilisée dans l'optimisation de la marge est celle détaillée dans le paragraphe V.2. avec les modifications suivantes :

- Exclusion des contrats souscrits avant novembre 2010 pour être cohérent avec la construction du modèle tarifaire du chapitre IV qui concernait uniquement la nouvelle offre ;
- Ajout de la variable prime commerciale modélisée (ou $Prime_{model_{tarif}}$ définie dans la formule (4) ci-dessous) construite au chapitre IV.

La base constituée comprend ainsi plus de 271 000 lignes et présente un chiffre d'affaires y compris taxes de près de 81 M€.

b. Fonction d'optimisation

La fonction d'optimisation détaillée ci-dessous cherche à maximiser la marge sous contrainte d'un taux de résiliation (équivalent à un nombre de contrats renouvelés) et d'un encadrement de la majoration en euro par contrat. Le programme d'optimisation s'écrit alors :

$$\begin{aligned} \{Optimisation_{Marge} \\ = (Prime_{simulée} - Prime_{model_{tarif}} \times Coeff_{inflation}) \\ \times [1 - modele_{sensi}(Prime_{simulée} - Prime_{avant_{terme}})] \quad (4) \end{aligned}$$

Sous contraintes : (2) et (3) définis au V.2

La formule (4) correspond au calcul de la marge pondérée par la probabilité de renouveler son contrat. Le premier paramètre correspond bien au calcul d'une marge puisqu'il fait la différence entre la $Prime_{simulée}$ (prime subissant la majoration de prime à l'échéance) et la variable $prime_{model_{tarif}}$ (prime modélisée correspondant à l'équilibre technique et définie au chapitre IV). Pour être cohérent avec l'objectif de la politique tarifaire de répercuter les dérives de sinistralité observées, il a été introduit un coefficient d'inflation de +2% pour tenir compte d'une hausse de coût moyen observée sur le portefeuille habitation.

Les contraintes (2) et (3) relatives au nombre de contrats renouvelés et à l'encadrement de la prime après terme restent identiques à celles définies dans le paragraphe V.2.

c. Matrice paramétrique actuelle optimisée

La matrice actuelle en production est simulée avec les mêmes allocations de majorations que celles définies au paragraphe V.2.c et aboutit à une revalorisation moyenne de +3.7% du chiffre d'affaires du portefeuille et une marge de 1.9 M€ pour un taux de résiliation de 7.5%. Le taux de résiliation obtenu ici est supérieur à celui de la partie V.2 (6.6%) dans la mesure où la base est réduite aux contrats de la nouvelle offre souscrits depuis fin 2010 qui disposent d'une ancienneté inférieure.

Dans la suite du paragraphe V.3, ce taux de résiliation sera utilisé comme contrainte et une optimisation de la marge sera recherchée.

Statut-type	Sinistralité	Ancienneté	Majorations
Locataire d'appartement	Non sinistré	0-3 ans	5.0%
		3 ans et +	7.0%
	Sinistré	0-3 ans	6.0%
		3 ans et +	6.9%
Locataire de maison	Non sinistré	0-3 ans	5.0%
		3 ans et +	6.9%
	Sinistré	0-3 ans	6.0%
		3 ans et +	6.7%
Propriétaire d'appartement	Non sinistré	0-3 ans	0.0%
		3 ans et +	6.9%
	Sinistré	0-3 ans	5.0%
		3 ans et +	6.7%
Propriétaire de maison	Non sinistré	0-3 ans	0.0%
		3 ans et +	6.6%
	Sinistré	0-3 ans	4.9%
		3 ans et +	6.2%

Figure 80 : Optimisation de la marge s.c. d'un volume de contrats - approche paramétrique – allocations optimisées sous contrainte des mêmes variables

Cette matrice est plutôt similaire à celle présentée au paragraphe V.2.d, en majorant davantage les locataires contrairement aux propriétaires. Par ailleurs, les clients récents en portefeuille sont davantage épargnés.

Pour un taux de résiliation de 7.5%, cette allocation conduit à un niveau de marge pondéré par la probabilité de renouveler de 2.2 M€.

d. Matrice paramétrique optimisée sans contrainte sur les variables

La matrice optimisée retenue correspond aux croisements des variables définies au paragraphe V.2.e, c'est-à-dire l'âge de l'assuré, la typologie de contact auprès de la plateforme de gestion, l'ancienneté du contrat et le statut d'occupation.

La simulation a conduit à 196 020 combinaisons de majorations pour les modalités des variables énoncées ci-dessus. Après sélection du point optimum choisi sur la frontière efficiente pour le taux de résiliation défini en amont, l'allocation de majorations suivante a été retenue :

Variables	Modalités	Majorations
Age	18-37 ans	0.0%
	37-60 ans	0.0%
	60-80 ans	1.0%
	80 ans et +	6.9%
Type contact RCC	Pas de contact / risque faible	0.5%
	Contacts à risque moyen	0.0%
	Contacts à risque élevé	0.0%
Ancienneté	0-2 ans	0.0%
	2-4 ans	0.0%
	4 ans et +	8.5%
Statut d'occupation	Locataire	1.0%
	Propriétaire	0.0%

Figure 81 : Optimisation de la marge s.c. d'un volume de contrats - approche paramétrique – allocation optimisée sans contraintes

L'allocation de majorations aboutit à des majorations plus importantes pour les clients âgés et anciens en portefeuille et n'ayant pas eu des contacts « risqués » avec la plateforme de gestion. Le déficit de rentabilité observé sur le segment des propriétaires semble compenser l'effet élastique de cette population puisque l'optimisation aboutit à un quasi statu-quo sur le statut d'occupation (légère majoration pour les locataires).

Pour un nombre de contrats renouvelés de l'ordre de 250 600 (soit un taux de résiliation de 7.5%), le niveau de marge pondéré par la probabilité de renouveler s'établit à 2.5 M€.

e. Résultats des stratégies pour l'optimisation de la marge sous contrainte d'un nombre de contrats renouvelés

Pour l'optimisation de la marge sous contrainte d'un nombre de contrats renouvelés, deux stratégies d'allocations de majorations ont été présentées dans cette partie mais les quatre stratégies définies au paragraphe V.2 ont été étudiées et sont résumées dans la figure 82. Ces stratégies sont toutes construites sous contrainte d'un taux de résiliation de 7.5% (niveau observé sur la stratégie 0).

Stratégie	Libellé	Type	Taux résiliation	Prime après terme	Revalorisation du CA	Marge pondérée	Gain marge pondérée
0	Matrice actuelle		7.5%	83.8M€	3.7%	1.9M€	
1	Matrice actuelle optimisée	Paramétrique	7.5%	84.2M€	4.3%	2.2M€	.3M€
2	Matrice sans contraintes		7.5%	84.5M€	4.6%	2.5M€	.6M€
3		Non paramétrique	7.5%	86.3M€	6.8%	4.3M€	2.4M€

Figure 82 : Synthèse des stratégies de majorations suivant l'axe marge s.c. d'un volume de contrats et gains associés

La stratégie 0 présente la matrice de la dernière politique tarifaire appliquée dans la compagnie et aboutit à une revalorisation du chiffre d'affaires de +3.7% pour un niveau de marge pondérée par la probabilité de renouvellement de 1.9 M€. Pour rappel, la marge calculée sur le portefeuille avant simulations d'une nouvelle politique tarifaire était de 1.2 M€ (cf. IV.7).

La stratégie 1, matrice actuelle en conservant les mêmes variables mais en allouant différemment les majorations, présente une amélioration de la marge comparée à la stratégie initiale évaluée à +0.3 M€.

La stratégie 2, présentant la matrice sans contraintes sur l'utilisation des variables, aboutit à une marge de 2.5 M€, soit un gain de +0.6 M€ par rapport à la stratégie 0 et à une revalorisation du chiffre d'affaires de +4.6%.

Comme indiqué dans le paragraphe V.2.f, la construction de l'approche non paramétrique (stratégie 3) engendre des variations spécifiques pour chaque contrat de la base de simulation. La distribution des majorations de l'approche non paramétrique est présentée en annexe 13. Cette approche non paramétrique aboutit à une revalorisation moyenne de +6.8% du chiffre d'affaires du portefeuille et un niveau de marge de 4.3 M€ (+2.4 M€ par rapport à la simulation 0).

4. Optimisation finale

L'optimisation de la marge sous contrainte d'un niveau de chiffre d'affaires a été étudiée mais aboutit à une fonction croissante, donc sans optimum.

Dans ce chapitre, l'optimisation de l'allocation des majorations a été étudiée suivant les angles développés précédemment :

- Optimisation du chiffre d'affaires sous contrainte d'un nombre de contrats renouvelés (V.2) ;
- Optimisation de la marge sous contrainte d'un nombre de contrats renouvelés (V.3).

L'objet de ce paragraphe est de définir une matrice optimale des allocations de majorations pour la base de simulation détaillée au paragraphe V.3.

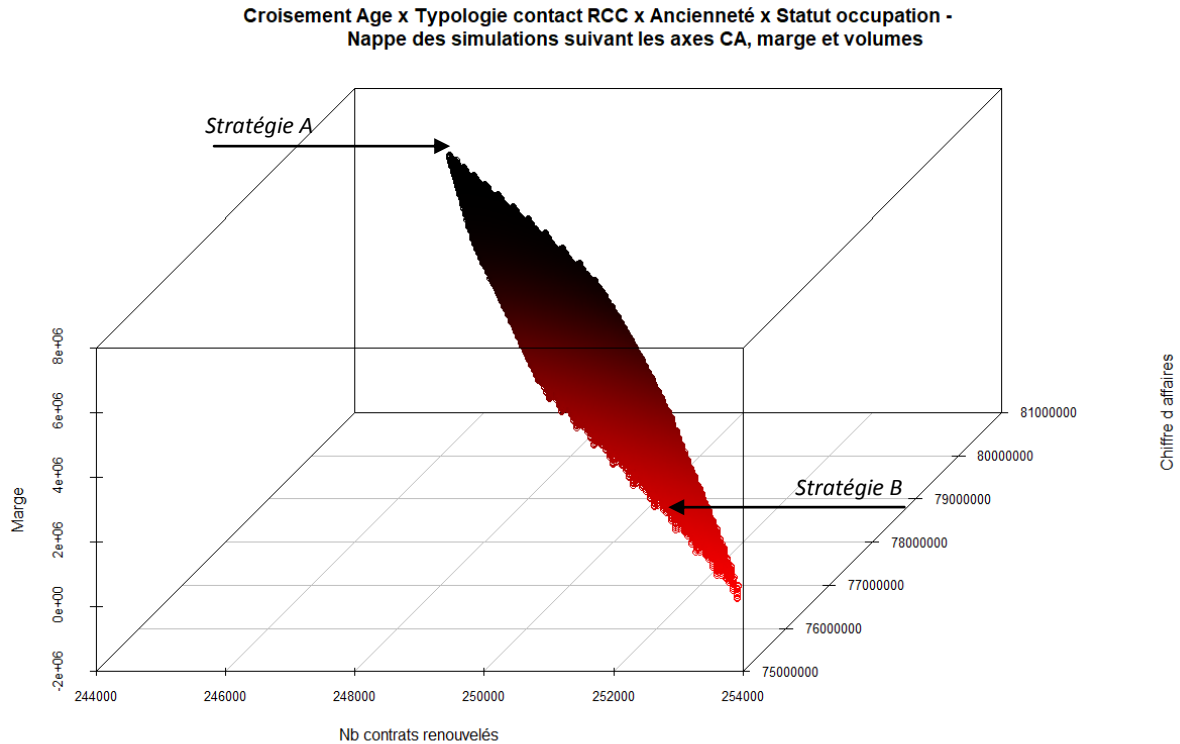


Figure 83 : Nappe des simulations de majorations pour l'allocation optimisée sans contraintes

La figure 83 présente une nappe construite à partir des 196 020 simulations de l'approche paramétrique pour l'allocation optimisée sans contraintes définie au paragraphe précédent. Sur l'axe x se trouve le nombre de contrats renouvelés, sur l'axe y le chiffre d'affaires et l'axe z , la marge. Chaque point représente le croisement des majorations sur l'âge, la typologie de contact auprès de la plateforme de gestion, l'ancienneté du contrat et le statut d'occupation.

Plus le chiffre d'affaires augmente, plus le niveau de marge croît et ce au détriment du volume de contrats en portefeuille.

Deux stratégies, A et B, sont détaillées ci-dessous.

La stratégie A, conduisant au choix du point le plus élevé sur la nappe, correspond aux coordonnées suivantes :

- Nombre de contrats renouvelés : $x = 245\,800$, soit un taux de résiliation de 9.3% ;

- Chiffre d'affaires pondéré $y = 80.5$ M€, soit un taux de revalorisation moyen du portefeuille de +10% ;
- Marge pondérée par la probabilité de renouveler : $z = 6.7$ M€.

Cette stratégie aboutit à un niveau élevé de revalorisation moyenne de +10% du chiffre d'affaires qui semble difficilement acceptable par le distributeur.

La stratégie B correspond au point optimal permettant l'optimisation du chiffre d'affaires et de la marge tout en limitant la hausse du taux de résiliation de 7.5% défini précédemment :

- Nombre de contrats renouvelés : $x = 250\ 400$, soit un taux de résiliation de 7.6% ;
- Chiffre d'affaires pondéré $y = 77.9$ M€, soit un taux de revalorisation moyen du portefeuille de +4.6% ;
- Marge pondérée par la probabilité de renouveler : $z = 2.6$ M€.

La stratégie B est obtenue avec les allocations de majorations suivantes :

Variables	Modalités	Majorations	Mix contrats N	Mix contrats N+1	Ecart mix
Age	18-37 ans	0%	32%	32%	-0.2pt
	37-60 ans	0%	48%	48%	-0.1pt
	60-80 ans	1.0%	17%	17%	+0.1pt
	80 ans et +	9.8%	3%	3%	+0.1pt
Type contact RCC	Pas de contact / risque faible	0.5%	87%	88%	+0.9pt
	Contacts à risque moyen	0%	7%	7%	-0.1pt
	Contacts à risque élevé	0%	6%	5%	-0.7pt
Ancienneté	0-2 ans	0%	18%	17%	-0.9pt
	2-4 ans	1%	58%	58%	+0.7pt
	4 ans et +	6.0%	24%	25%	+0.3pt
Statut d'occupation	Locataire	0%	54%	55%	+0.7pt
	Propriétaire	2%	46%	45%	-0.7pt

Figure 84 : Optimisation finale – allocation sur la nappe de simulations pour la stratégie B

Cette allocation aboutit à des majorations élevées pour les personnes âgées et les clients anciens en portefeuille. Les clients ayant contactés avec des motifs risqués la plateforme de gestion sont épargnés. Les propriétaires sont également mis à contribution puisqu'ils doivent rattraper leur retard en matière de rentabilité.

Concernant le changement de répartition dans le mix portefeuille, avec ces majorations, les propriétaires affichent un taux de résiliation de 9% contre 6.4% pour les locataires. Ainsi, on constate une baisse du poids des propriétaires dans le portefeuille de 0.7 point au profit des locataires.

Cette stratégie B apparaît donc préférable dans la mesure où elle engendre une légère dégradation du taux de résiliation (+0.1 point par rapport à la stratégie de la matrice actuelle définie dans la figure 83) mais permet une amélioration du chiffre d'affaires probabilisé (+ 0.6 M€) et de la marge (+0.7 M€).

Conclusion

La sensibilité au prix appliquée à l'assurance non vie n'est pas un sujet nouveau dans le domaine de l'actuariat. Néanmoins, avec l'arrivée récente de la loi Hamon, les règles du jeu évoluent puisque l'assuré a la faculté de résilier à tout moment. Avec une rotation des portefeuilles plus importante, il apparaît que la rétention des clients en portefeuille est un enjeu majeur pour les compagnies d'assurances.

L'enjeu du mémoire était l'étude du comportement des assurés en habitation lors d'une majoration à l'échéance pour mesurer leur sensibilité au prix. En parallèle, un modèle de prime pure a été développé afin de définir les populations à marge et ainsi optimiser le chiffre d'affaires mais également la marge lors du renouvellement du contrat.

Les deux premières parties ont été consacrées à la présentation du contexte de l'étude puis à l'élaboration de la base de données. Des analyses descriptives sur les principales variables et une analyse en correspondances multiples ont permis de mettre en exergue un effet classe sociale sur nos données.

La troisième partie a porté sur l'analyse de la résiliation au terme. Une première étude a permis de construire un arbre de décision pour qualifier les variables les plus discriminantes expliquant la résiliation au terme. Le développement d'un modèle de régression logistique est venu conforter le choix des variables âge, ancienneté du contrat, majoration en euro au terme et typologie de contact auprès de la plateforme de gestion. Un modèle de survie, de type Cox PH, a également été construit pour analyser le taux de chute de nos assurés dans le portefeuille. Par exemple, chaque augmentation de l'âge d'une année diminue la probabilité de résilier de 2%. Puis, un modèle de sensibilité au prix a été développé à partir de la régression logistique et nous a permis de définir les typologies d'assurés les plus élastiques au prix. Par exemple, un contrat subissant une majoration de 50 euros à l'échéance verra son taux de résiliation croître de plus de 13 points s'il a moins de 2 ans d'ancienneté contre une hausse modérée du taux de résiliation de moins de 3 points pour un contrat étant resté plus de 16 ans en portefeuille.

En parallèle, une modélisation de la prime pure a été réalisée via un modèle linéaire généralisé de fréquence et coût moyen par garantie. Une fois ajoutés les chargements, cette prime commerciale technique a été comparée à la prime terminée en portefeuille afin d'identifier les segments de populations à marge ou déficitaires. Il en résulte que les locataires présentent des marges positives au détriment des propriétaires et que les contrats anciens en portefeuille « financent » les affaires nouvelles.

Dans la dernière partie, des travaux d'optimisation du chiffre d'affaires et de la marge ont été réalisés. A partir du modèle de résiliation et de sensibilité au prix et des calculs de rentabilité, il a été possible d'optimiser la matrice actuelle de politique tarifaire. Les gains en matière de chiffre d'affaires et de marge s'élèvent respectivement à +0.6 M€ et +0.7 M€ pour la base de contrats limitée à la nouvelle offre. En considérant la base totale, les gains se rapprocheraient du million d'euros. L'approche non paramétrique aboutit à des résultats encore plus significatifs mais son allocation, spécifique pour chaque caractéristique du contrat, serait difficilement acceptable par les distributeurs car complexe à expliquer au client.

L'intégration de nouvelles variables issues des données bancaires ou de gestion a été un élément essentiel de l'étude mais il aurait pu également être étudié l'ajout de tarifs concurrents ou leurs annonces de majorations à l'échéance.

Par ailleurs, une autre application de l'étude de la sensibilité au prix au terme pourrait être utilisée pour la politique de surveillance du portefeuille et ainsi déterminer si les majorations appliquées pour les clients sur-sinistrés sont adéquates.

Enfin, la conclusion de ce mémoire se terminera sur une question : est-ce que l'effet prix est le seul facteur de risque pouvant expliquer le départ d'un assuré? D'autres critères, plus qualitatifs, tels que le service apporté au client au moment de la souscription ou la rapidité du règlement d'un sinistre, pourraient également influencer la décision du client à rester ou non au sein de la compagnie.

Liste des figures

Figure 1 : Résiliation loi Châtel	9
Figure 2 : Résiliation loi Hamon	10
Figure 3 : Répartition des résiliations Habitation par typologie de 2014 à 2017	10
Figure 4 : Evolution des taux de résiliations depuis la loi Hamon - écart entre la date d'échéance d'avril et de résiliation	11
Figure 5 : Temps forts de la politique tarifaire chez Societe Generale Insurance	14
Figure 6 : Taux de résiliation vu à l'ultime par échéance principale entre 2015 et 2017	15
Figure 7 : Nappe de résiliation MRH pour les échéances 2016 et en fonction de l'écart entre la date d'échéance et la date de résiliation	15
Figure 8 : Extrait de quelques variables de la base de données	17
Figure 9 : Détail du score de valeur client	19
Figure 10 : Classification des contacts clients auprès de la plateforme de gestion	20
Figure 11 : Stock de contrats et primes acquises MRH entre 2011 et 2017	21
Figure 12 : Comparaison du taux d'affaires nouvelles entre Sogessur et le marché	22
Figure 13 : Comparaison du volume de prêts immobiliers et des taux d'obligation d'Etat à 10 ans entre 2013 et 2017	23
Figure 14 : Taux de résiliation en fonction de l'année de l'échéance	24
Figure 15 : Taux de résiliation en fonction de l'écart entre le terme et l'observation des données	25
Figure 16 : Taux de résiliation en fonction de la variation de prime au terme	26
Figure 17 : Taux de résiliation en fonction de la prime client	27
Figure 18 : Taux de résiliation en fonction de l'âge de l'assuré	28
Figure 19 : Taux de résiliation en fonction du segment de rentabilité	29
Figure 20 : Taux de résiliation en fonction de la typologie des contacts clients auprès de RCC	30
Figure 21 : Taux de résiliation en fonction du statut d'occupation	31
Figure 22 : Taux de résiliation en fonction du niveau du coefficient de surveillance du portefeuille	32
Figure 23 : Taux de résiliation en fonction de la segmentation bancaire	33
Figure 24 : Taux de résiliation en fonction de la souscription de l'assurance avec un prêt immobilier	33
Figure 25 : Histogramme des valeurs propres de l'ACM	36
Figure 26 : Représentation des modalités des variables dans le plan factoriel des axes 1 et 2	37
Figure 27 : Représentation du cercle des corrélations des axes 1 et 3	38
Figure 28 : Fonction de répartition de la loi logit	39
Figure 29 : Arbre de décision modèle simplifié	42
Figure 30 : Erreur commise en validation croisée en fonction de la taille de l'arbre	43
Figure 31 : Arbre de décision complet - classement des variables les plus significatives	43
Figure 32 : Matrice de corrélation de la base d'étude	44
Figure 33 : Régression logistique - classement des variables les plus significatives	46
Figure 34 : Significativité des variables du modèle logistique	47
Figure 35 : Histogramme des prédictions des taux de résiliations sur la base d'apprentissage et de validation	47
Figure 36 : Modèle logistique de résiliation suivant l'âge du souscripteur	48
Figure 37 : Modèle logistique de résiliation suivant l'ancienneté	48
Figure 38 : Modèle logistique de résiliation suivant la typologie de contact RCC	49

<i>Figure 39 : Modèle logistique de résiliation suivant l'année de construction du bien</i>	<i>50</i>
<i>Figure 40 : Modèle logistique de résiliation suivant le statut d'occupation</i>	<i>50</i>
<i>Figure 41 : Test de Hosmer et Lemeshow</i>	<i>51</i>
<i>Figure 42 : Matrice de confusion</i>	<i>52</i>
<i>Figure 43 : Courbe ROC du modèle de régression logistique.....</i>	<i>53</i>
<i>Figure 44 : Comparaison des courbes ROC de la régression logistique entre apprentissage et validation</i>	<i>54</i>
<i>Figure 45 : Rapport des risques instantanés du modèle de Cox</i>	<i>58</i>
<i>Figure 46 : Fonction de survie et de hasard du modèle de Cox</i>	<i>59</i>
<i>Figure 47 : Fonction de survie du modèle de Cox suivant l'âge du client</i>	<i>60</i>
<i>Figure 48 : Fonction de survie du modèle de Cox suivant la valeur client</i>	<i>60</i>
<i>Figure 49 : Fonction de survie du modèle de Cox suivant l'indicatrice salarié SG</i>	<i>61</i>
<i>Figure 50 : Fonction de survie du modèle de Cox suivant le statut d'occupation.....</i>	<i>61</i>
<i>Figure 51 : Classement des variables les plus sensibles au prix.....</i>	<i>63</i>
<i>Figure 52 : Sensibilité au prix en fonction de la typologie de contact auprès de la plateforme de gestion</i>	<i>64</i>
<i>Figure 53 : Sensibilité au prix en fonction de l'âge du souscripteur</i>	<i>65</i>
<i>Figure 54 : Sensibilité au prix en fonction de l'ancienneté du contrat.....</i>	<i>65</i>
<i>Figure 55 : Sensibilité au prix en fonction du segment bancaire</i>	<i>66</i>
<i>Figure 56 : Période d'étude des bases de données contrats et sinistres</i>	<i>68</i>
<i>Figure 57 : Modèle de fréquence Dégât des eaux - significativité des variables.....</i>	<i>73</i>
<i>Figure 58 : Résidus de la modélisation du coût moyen Bris de vitres.....</i>	<i>74</i>
<i>Figure 59 : Coefficient tarifaire de la fréquence Dégât des eaux par nombre de pièces.....</i>	<i>75</i>
<i>Figure 60 : Coefficient tarifaire de la fréquence Vol par zone géographique.....</i>	<i>76</i>
<i>Figure 61 : Coefficient tarifaire de la fréquence RC suivant la présence d'enfants</i>	<i>77</i>
<i>Figure 62 : Coefficient tarifaire du coût moyen Vol suivant les capitaux objets de valeur.....</i>	<i>78</i>
<i>Figure 63 : Coefficient tarifaire du coût moyen Incendie suivant la présence d'un insert.....</i>	<i>79</i>
<i>Figure 64 : Coefficient tarifaire de la prime pure Vol en fonction du statut-type et de l'étage</i>	<i>80</i>
<i>Figure 65 : Coefficient tarifaire de la prime pure Dégâts suite à vol suivant le type de résidence.....</i>	<i>81</i>
<i>Figure 66 : Coefficient tarifaire de la prime pure Dégât des eaux suivant l'année de construction.....</i>	<i>82</i>
<i>Figure 67 : Coefficient tarifaire de la prime pure Tempête grêle neige suivant la taille des dépendances</i>	<i>83</i>
<i>Figure 68 : Synthèse de la prime pure observée et modélisée par garantie.....</i>	<i>83</i>
<i>Figure 69 : Marge moyenne en fonction du statut-type</i>	<i>85</i>
<i>Figure 70 : Marge moyenne en fonction de l'ancienneté du contrat</i>	<i>86</i>
<i>Figure 71 : Synthèse des modèles et des étapes pour l'optimisation de la stratégie de renouvellement</i>	<i>87</i>
<i>Figure 72 : Optimisation du CA - approche paramétrique – allocations de la matrice actuelle.....</i>	<i>90</i>
<i>Figure 73 : Optimisation du CA - approche paramétrique – allocations optimisées sous contraintes des mêmes variables.....</i>	<i>91</i>
<i>Figure 74 : Optimisation du CA - approche paramétrique – simulations de majorations sans contraintes.....</i>	<i>92</i>
<i>Figure 75 : Optimisation du CA - approche paramétrique – allocation optimisée sans contraintes.....</i>	<i>93</i>
<i>Figure 76 : Optimisation du CA - approche non paramétrique - distribution des majorations de primes</i>	<i>94</i>

<i>Figure 77 : Optimisation du CA - approche non paramétrique - zoom sur quelques variables</i>	<i>94</i>
<i>Figure 78 : Synthèse graphique des stratégies de majorations suivant l'axe CA</i>	<i>95</i>
<i>Figure 79 : Synthèse des stratégies de majorations suivant l'axe CA et gains associés</i>	<i>95</i>
<i>Figure 80 : Optimisation de la marge s.c. d'un volume de contrats - approche paramétrique – allocations optimisées sous contrainte des mêmes variables</i>	<i>97</i>
<i>Figure 81 : Optimisation de la marge s.c. d'un volume de contrats - approche paramétrique – allocation optimisée sans contraintes</i>	<i>98</i>
<i>Figure 82 : Synthèse des stratégies de majorations suivant l'axe marge s.c. d'un volume de contrats et gains associés</i>	<i>99</i>
<i>Figure 83 : Nappe des simulations de majorations pour l'allocation optimisée sans contraintes</i>	<i>100</i>
<i>Figure 84 : Optimisation finale – allocation sur la nappe de simulations pour la stratégie B.....</i>	<i>101</i>

Lexique

ACM : analyse en correspondances multiples

CA : chiffre d'affaires correspondant aux primes émises cumulées sur une année

CM : coût moyen, charge totale des sinistres rapportée au nombre de sinistres

EP : échéance principale, correspondant à la date du terme

FFA : Fédération Française de l'Assurance

GAFA : Google, Amazon, Facebook, Apple

GLM : *general linear model* ou modèle linéaire généralisé

LA : locataire d'appartement

LC : loi Châtel

LH : Loi Hamon

LM : locataire de maison

MRH : multirisques habitation

PA : propriétaire d'appartement

PM : propriétaire de maison

PNO : propriétaire non occupant

PP : prime pure ou espérance du coût du risque. Elle s'écrit : fréquence x coût moyen ou charge totale des sinistres divisée par les années risques

RCC : Relation clients contrats. Plateforme de gestion des contrats dommages de *Societe Generale Insurance*

SDP : surveillance du portefeuille, politique permettant d'analyser les clients ayant une sinistralité atypique ou des comportements frauduleux

Taux de transformation : rapport entre le nombre de contrats souscrits et le nombre de devis.

Bibliographie

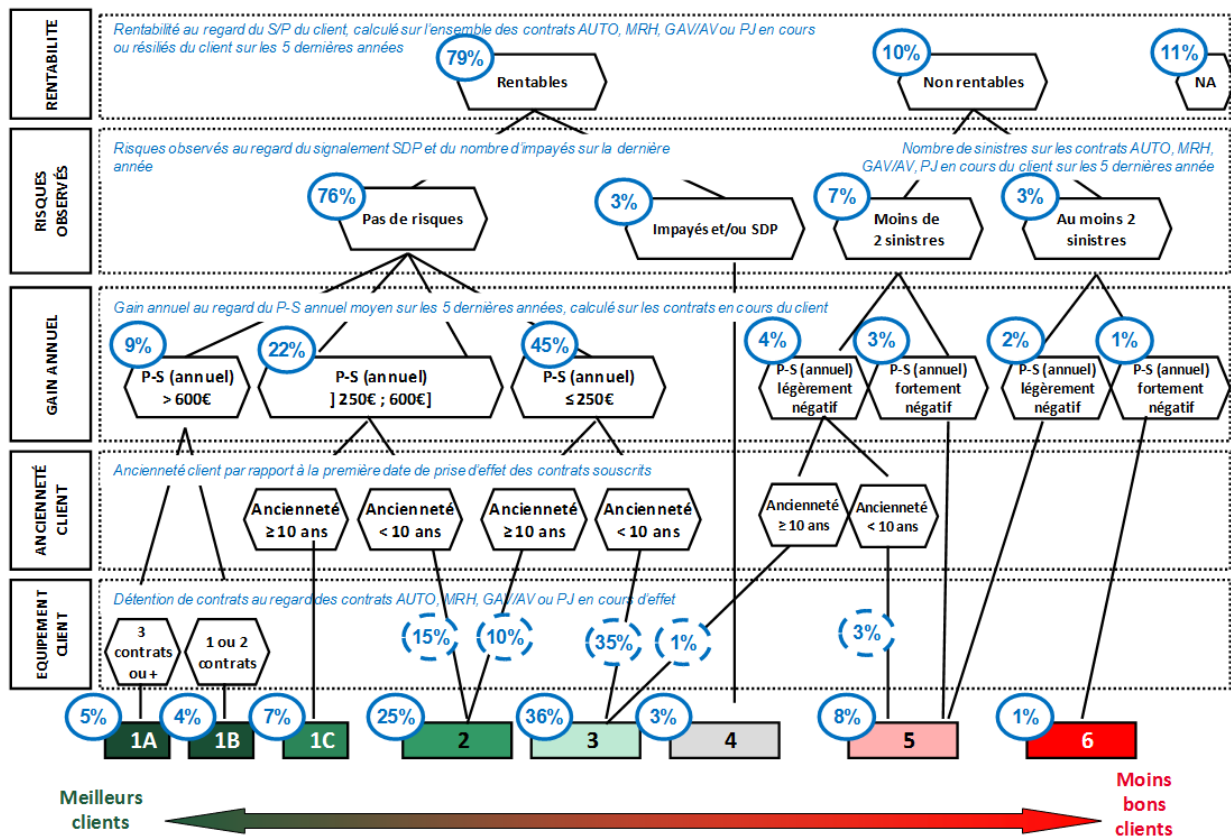
- [1] SAINT PIERRE P. (2014), Introduction à l'analyse des durées de survie, Université Pierre et Marie Curie
- [2] PLANCHET F. (2018), Modèles de durée, Institut de science financière et d'assurances
- [3] SAPORTA G. (2006), Probabilités, analyse des données et statistique
- [4] OHLSSON E., JOHANSSON B. (2010), Non-life insurance pricing with generalized linear models
- [5] ROUVIERE L. (2014), Régression logistique avec R, Université Rennes 2
- [6] RAKOTOMALALA R. (2017), Pratique de la régression logistique, Université Lumière Lyon 2
- [7] GEFFRAY S. (2009), Analyse des durées de vie avec le logiciel R, Université de Strasbourg
- [8] BALDE B. (2015), Modélisation du taux de résiliation en assurance MRH
- [9] DUTANG G. (2011), Regression models of price elasticity in non-life insurance
- [10] BOUQUET C., MENARD P. (2011), Assurance automobile – Optimisation des ressources à l'échéance
- [11] Note technique tarif MRH 2010 – Document interne
- [12] Documentation technique Addactis Pricing, Behavior, Simulation joyjconac-wY

Annexes

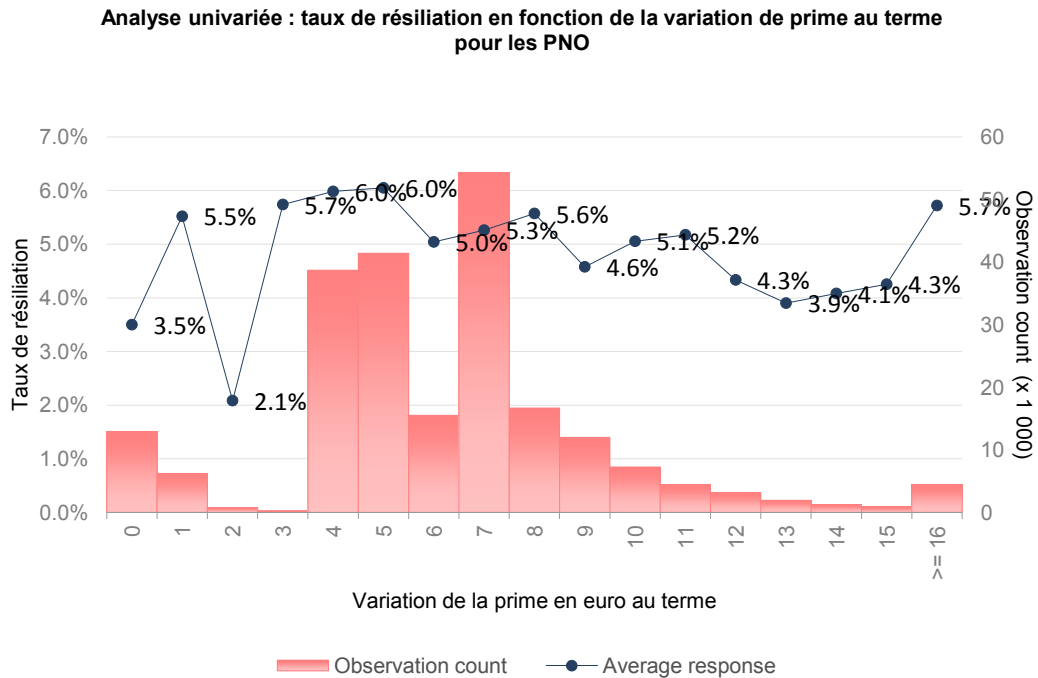
1. Triangle de résiliations cumulé par mois et année d'échéance

Triangle de résiliations co		LC	M	M+1	M+2	M+3	M+4	M+5	M+6	M+7	M+8	M+9	M+10	M+11
3	2014	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%
4	2014	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%	3.0%
5	2014	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%
6	2014	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%
7	2014	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%
8	2014	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%
9	2014	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%	3.1%
10	2014	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%	3.2%
11	2014	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%	3.3%
12	2014	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%	4.0%
1	2015	2.9%	2.9%	2.9%	2.9%	3.2%	3.5%	3.7%	4.0%	4.3%	4.4%	4.6%	4.8%	5.0%
2	2015	2.8%	2.8%	2.8%	3.1%	3.4%	3.6%	3.9%	4.1%	4.3%	4.5%	4.7%	4.9%	5.1%
3	2015	2.9%	2.9%	3.1%	3.5%	3.6%	4.0%	4.2%	4.3%	4.6%	4.8%	5.1%	5.3%	5.4%
4	2015	3.2%	3.2%	3.6%	4.0%	4.3%	4.6%	4.8%	5.1%	5.4%	5.6%	5.9%	6.1%	6.4%
5	2015	3.2%	3.2%	3.6%	3.9%	4.2%	4.4%	4.7%	4.9%	5.2%	5.4%	5.6%	5.9%	6.0%
6	2015	2.9%	3.0%	3.5%	3.8%	4.1%	4.4%	4.6%	4.9%	5.1%	5.3%	5.6%	5.9%	6.1%
7	2015	3.2%	3.3%	3.7%	3.9%	4.2%	4.6%	4.8%	5.1%	5.3%	5.6%	5.8%	6.0%	6.2%
8	2015	3.0%	3.1%	3.5%	3.9%	4.2%	4.5%	4.7%	5.0%	5.3%	5.6%	5.9%	6.1%	6.3%
9	2015	2.6%	2.6%	3.0%	3.5%	3.8%	4.0%	4.3%	4.6%	4.8%	5.0%	5.2%	5.5%	5.7%
10	2015	2.9%	2.9%	3.5%	3.9%	4.1%	4.4%	4.7%	5.0%	5.2%	5.5%	5.8%	6.0%	6.2%
11	2015	3.1%	3.1%	3.7%	4.0%	4.3%	4.6%	5.0%	5.2%	5.5%	5.7%	6.0%	6.1%	6.2%
12	2015	3.0%	3.1%	3.5%	3.9%	4.3%	4.7%	5.0%	5.3%	5.6%	5.8%	6.0%	6.3%	6.4%
1	2016	2.3%	2.5%	3.0%	3.4%	3.8%	4.1%	4.4%	4.8%	5.1%	5.3%	5.6%	5.9%	6.1%
2	2016	2.3%	2.6%	3.1%	3.5%	3.9%	4.2%	4.5%	4.8%	5.0%	5.3%	5.5%	5.8%	5.9%
3	2016	2.2%	2.5%	3.0%	3.4%	3.7%	4.1%	4.3%	4.5%	4.9%	5.2%	5.4%	5.7%	5.9%
4	2016	2.4%	2.8%	3.3%	3.8%	4.2%	4.5%	4.6%	5.0%	5.3%	5.7%	5.9%	6.3%	6.6%
5	2016	2.2%	2.6%	3.1%	3.6%	3.9%	4.1%	4.5%	4.8%	5.1%	5.4%	5.7%	6.0%	6.3%
6	2016	2.2%	2.6%	3.1%	3.5%	3.7%	4.1%	4.4%	4.6%	4.9%	5.2%	5.6%	6.0%	6.2%
7	2016	2.1%	2.4%	2.9%	3.2%	3.6%	3.9%	4.3%	4.6%	4.9%	5.3%	5.7%	6.0%	6.3%
8	2016	2.5%	2.9%	3.3%	3.7%	4.1%	4.4%	4.7%	5.0%	5.5%	5.9%	6.2%	6.5%	6.7%
9	2016	2.1%	2.3%	2.8%	3.2%	3.6%	3.8%	4.2%	4.5%	4.8%	5.2%	5.4%	5.7%	6.0%
10	2016	2.1%	2.4%	2.9%	3.2%	3.6%	3.9%	4.2%	4.6%	4.9%	5.2%	5.5%	5.8%	6.0%
11	2016	2.5%	2.9%	3.4%	3.8%	4.1%	4.6%	5.0%	5.4%	5.7%	6.0%	6.3%	6.5%	6.8%
12	2016	2.3%	2.7%	3.2%	3.7%	4.2%	4.6%	5.0%	5.3%	5.7%	6.0%	6.2%	6.5%	6.8%
1	2017	2.3%	2.7%	3.2%	3.7%	4.1%	4.5%	4.9%	5.2%	5.6%	5.8%	6.1%	6.4%	6.7%
2	2017	2.2%	2.5%	3.0%	3.5%	3.9%	4.2%	4.6%	4.8%	5.1%	5.4%	5.7%	6.0%	6.2%
3	2017	2.0%	2.4%	3.1%	3.5%	3.9%	4.2%	4.5%	4.8%	5.1%	5.4%	5.7%	6.0%	6.2%
4	2017	2.1%	2.7%	3.3%	3.7%	4.1%	4.4%	4.6%	4.9%	5.3%	5.5%	5.8%	6.2%	6.4%
5	2017	1.8%	2.3%	2.8%	3.2%	3.4%	3.7%	4.0%	4.3%	4.6%	4.9%	5.2%	5.4%	5.7%
6	2017	1.9%	2.3%	3.0%	3.4%	3.6%	3.9%	4.2%	4.5%	4.8%	5.1%	5.4%	5.7%	5.9%
7	2017	1.6%	2.0%	2.5%	2.8%	3.1%	3.5%	3.8%	4.1%	4.3%	4.6%	4.8%	5.1%	5.3%
8	2017	2.2%	2.6%	3.1%	3.5%	3.9%	4.2%	4.6%	4.9%	5.2%	5.5%	5.8%	6.1%	6.4%
9	2017	2.1%	2.4%	3.0%	3.4%	3.8%	4.1%	4.5%	4.8%	5.1%	5.4%	5.7%	6.0%	6.2%
10	2017	1.9%	2.3%	2.8%	3.2%	3.6%	3.9%	4.2%	4.5%	4.8%	5.0%	5.3%	5.6%	5.8%
11	2017	1.9%	2.4%	2.9%	3.3%	3.7%	4.0%	4.3%	4.6%	4.9%	5.2%	5.5%	5.8%	6.0%
12	2017	1.8%	2.1%	2.6%	2.9%	3.2%	3.5%	3.8%	4.1%	4.3%	4.6%	4.8%	5.1%	5.3%

2. Mécanisme de construction de la valeur client assurantiel sur le périmètre dommages

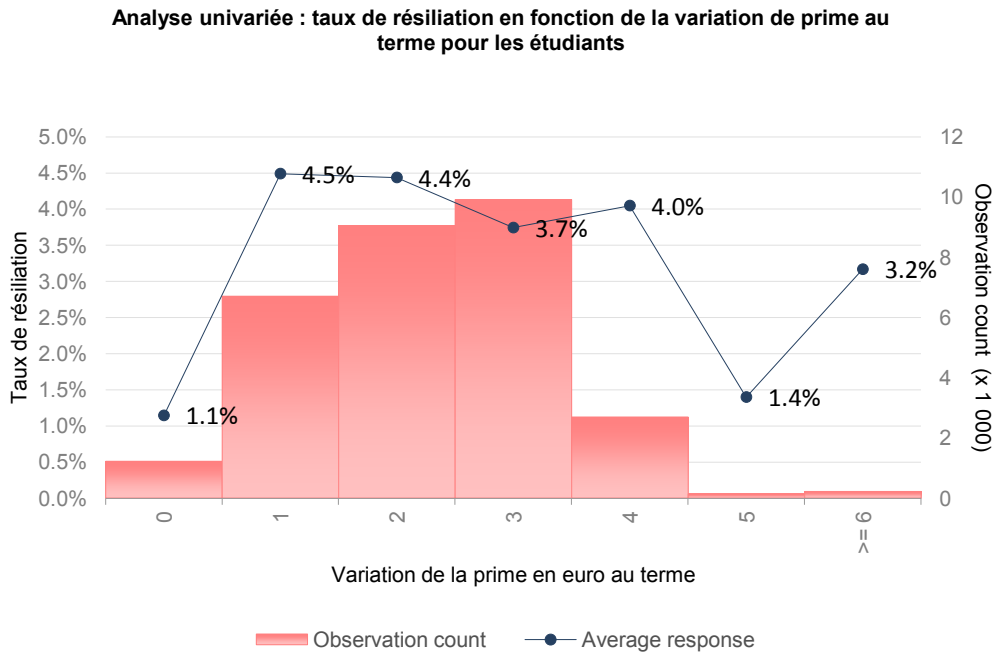


3. Taux de résiliation croisé avec la majoration en euro au terme et la formule Investisseur



- Primes faibles, déduction fiscale de la prime d'assurance dans les charges locataires du propriétaire investisseur → pas de variation significative du taux de résiliation pour cette population

4. Taux de résiliation croisé avec la majoration au terme et la formule Etudiant

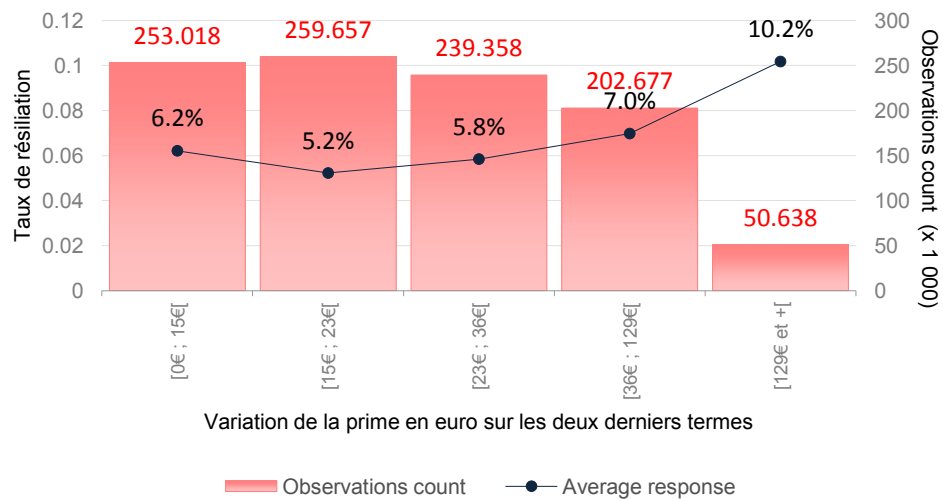


- Primes faibles, faible durée de vie de ces contrats en portefeuille → pas de variation significative du taux de résiliation pour cette population

5. Analyse descriptive - taux de résiliation en fonction d'autres variables explicatives

- Données prix : La variation de la prime en euro cumulée sur les deux derniers termes

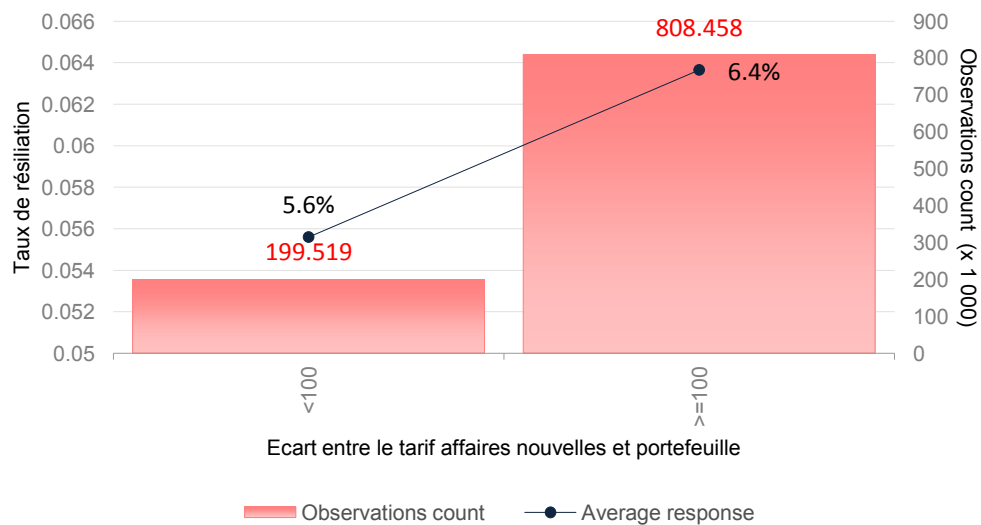
Analyse univariée : taux de résiliation en fonction de la variation de prime par rapport aux deux derniers termes



➔ Progression du taux de résiliation en fonction de la majoration en euro cumulée sur les deux derniers termes

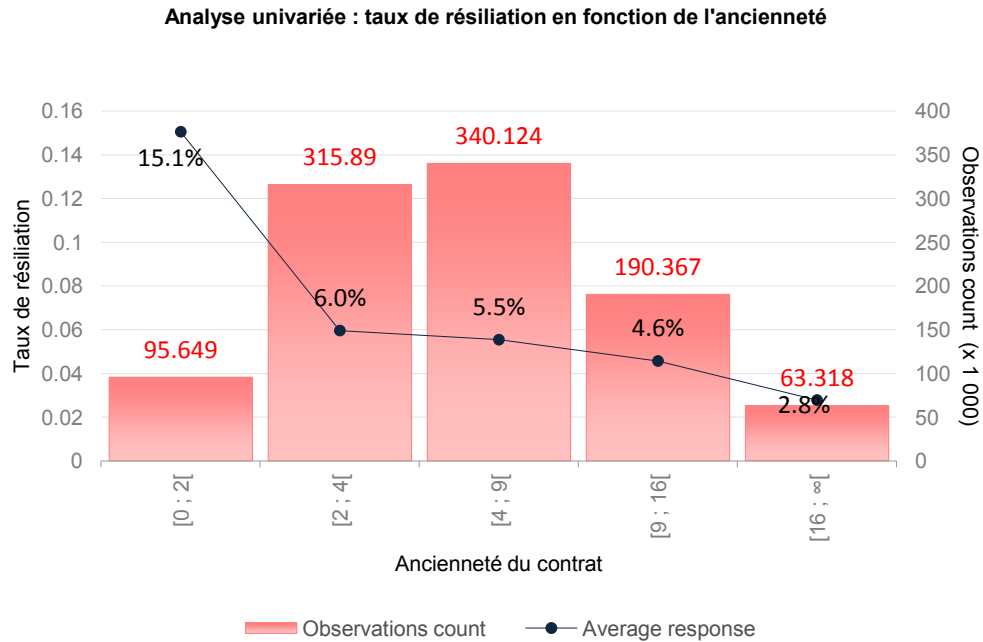
- Données prix : l'écart au tarif entre affaires nouvelles et portefeuille

Analyse univariée: taux de résiliation en fonction de l'écart au tarif entre affaires nouvelles et portefeuille



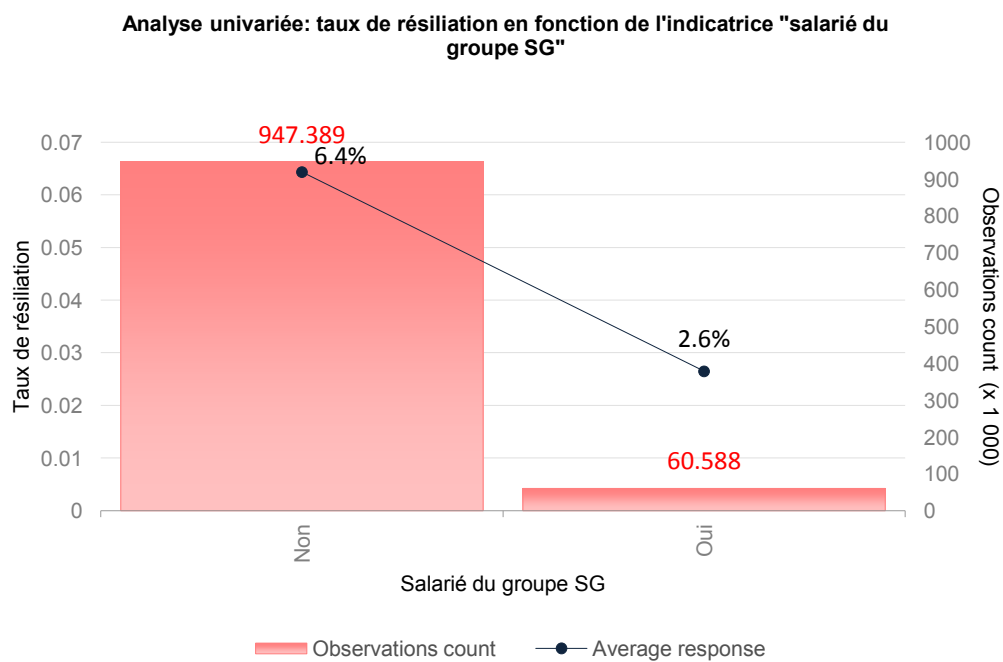
➔ Plus l'écart entre la prime terminée et la prime en vision affaires nouvelles, toutes choses égales par ailleurs, s'accroît, plus le taux de résiliation progresse.

- Données client : l'ancienneté du contrat



➔ Le taux de résiliation diminue avec l'ancienneté du contrat

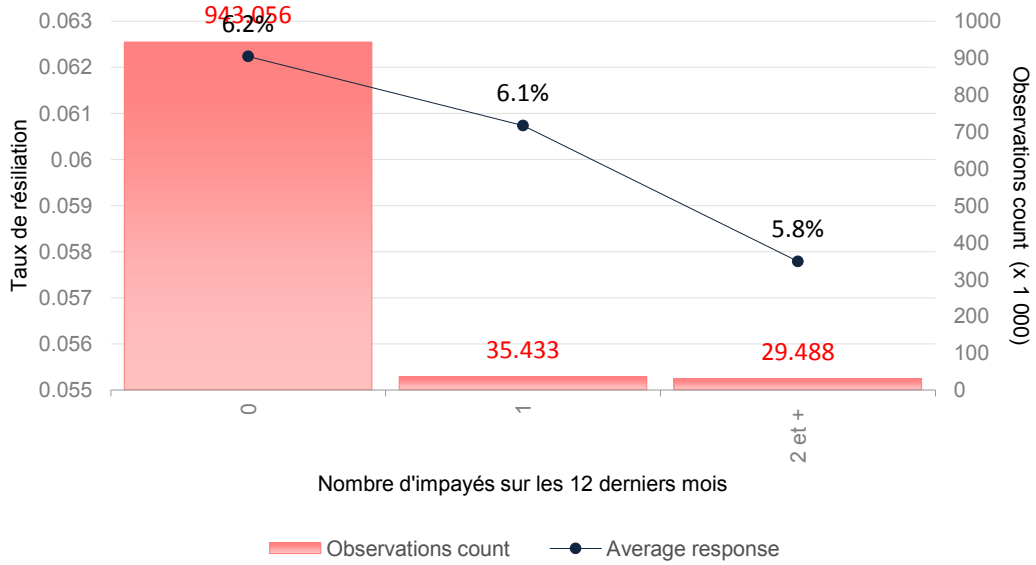
- Données client : l'indicatrice salarié du groupe Société Générale



➔ Un salarié du groupe Société Générale a tendance à moins résilier qu'un client non salarié SG

- Données client : le nombre d'impayés sur les 12 derniers mois précédant le terme

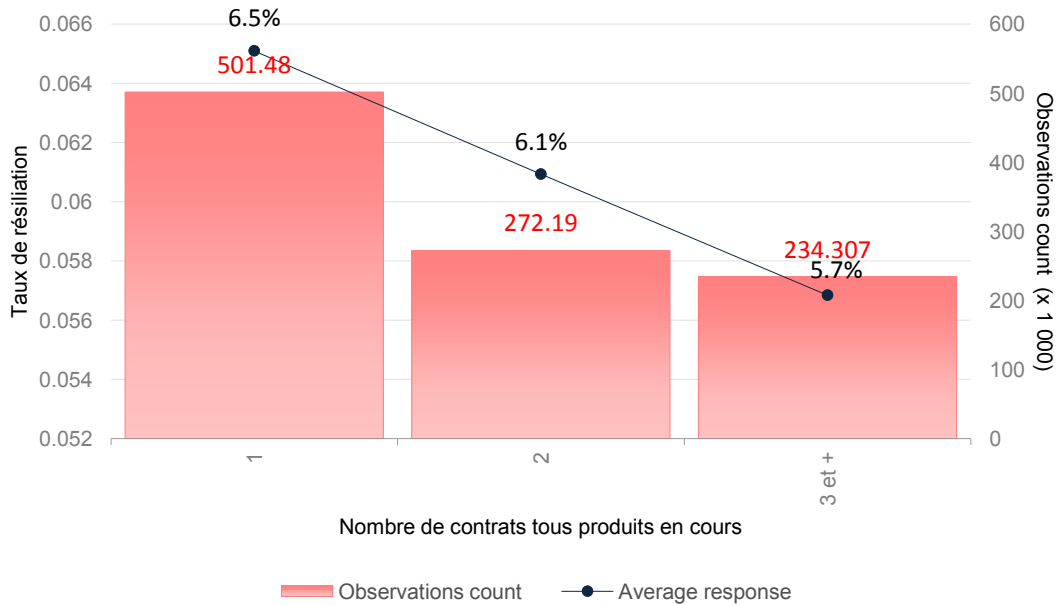
Analyse univariée : taux de résiliation en fonction du nombre d'impayés dans la dernière année



➔ Plus le client a d'impayés, moins il résilie

- Données client : nombre de contrats, tous produits, détenus par le client

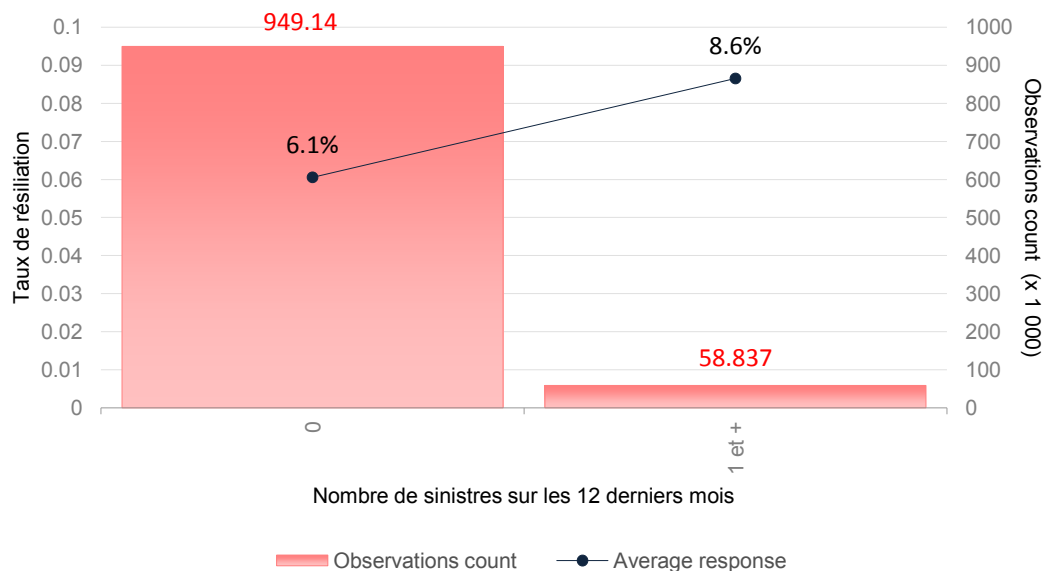
Analyse univariée : taux de résiliation en fonction de l'équipement du client



➔ Plus le client est équipé en contrats dommages (Automobile, MRH, Protection Juridique, etc.), plus sa probabilité de résilier diminue

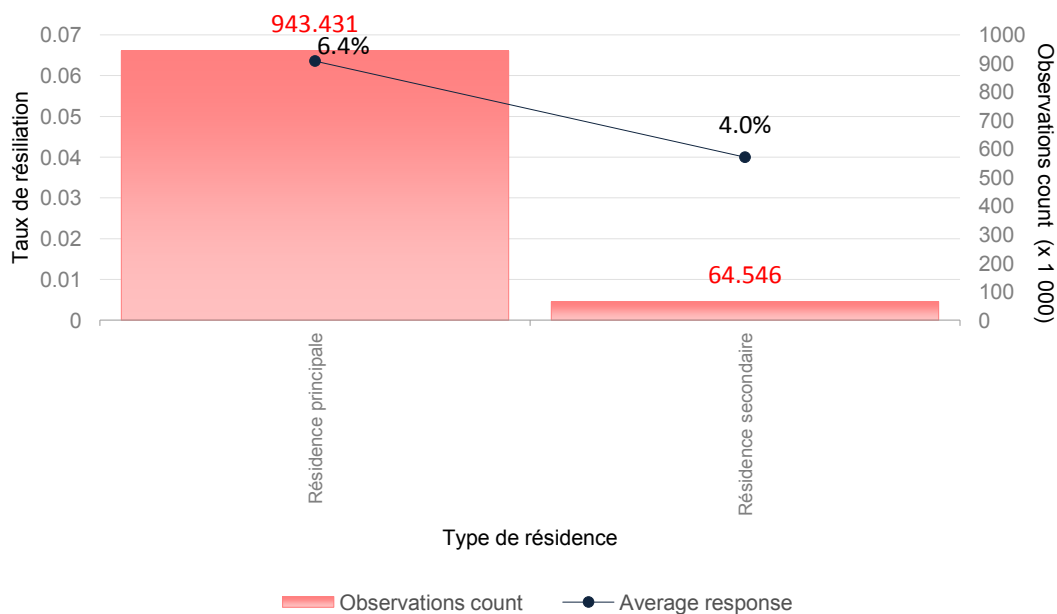
- Données risque : le nombre de sinistres sur les 12 derniers mois avant le terme

Analyse univariée: taux de résiliation en fonction du nombre de sinistres dans la dernière année



- ➔ La présence d'un sinistre dans les 12 derniers mois augmente la probabilité de résilier probablement du au fait que la matrice de politique tarifaire tient compte de la sinistralité et majore davantage ce type de client
- Données risque : le type de résidence

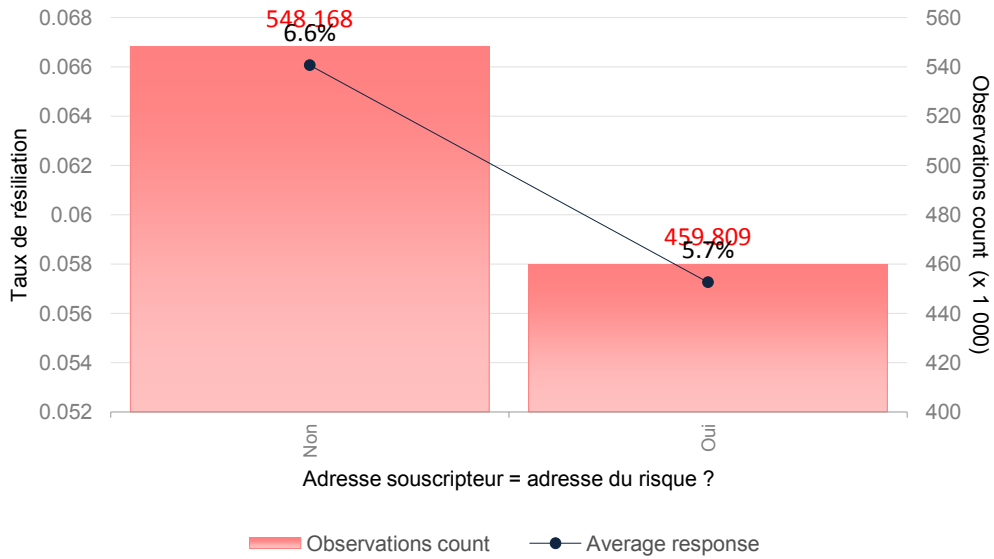
Analyse univariée : taux de résiliation en fonction du type de résidence



- ➔ Le taux de résiliation est plus faible pour les résidences secondaires par rapport aux résidences principales

- Données risque : l'adresse du souscripteur est-elle l'adresse du risque ?

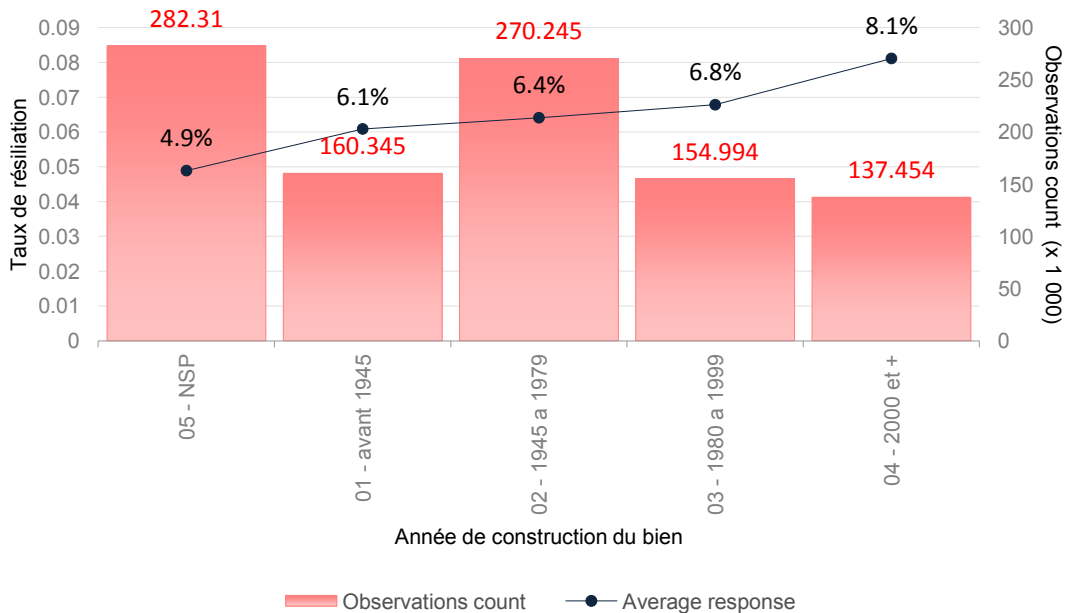
Analyse univariée : taux de résiliation en fonction de l'adresse du souscripteur



➔ Le taux de résiliation diminue d'un point si l'adresse du souscripteur est la même que l'adresse du bien assuré

- Données risque : année de construction du bien assuré

Analyse univariée: taux de résiliation en fonction de l'année de construction du bien



➔ Le taux de résiliation progresse plus le bien est récent.

6. Analyse en correspondances multiples : rappels théoriques et inertie

Analyse factorielle des correspondances (AFC)

L'analyse factorielle des correspondances permet d'analyser les fréquences relatives dans la population des modalités de deux variables qualitatives x et y .

On définit le tableau de contingence qui compte le nombre d'individus prenant simultanément la modalité X_i de x et Y_j de y de la manière suivante :

$$K_{(I,J)} = \begin{pmatrix} \vdots & & \\ \cdots & n_{i,j} & \cdots \\ \vdots & & \end{pmatrix} \begin{array}{c} \boxed{n_{i,\cdot}} \\ \vdots \\ \boxed{n_{\cdot,j}} \end{array}$$

Avec les sommes marginales : $n_{i,\cdot} = \sum_{j=1}^J n_{i,j}$ et $n_{\cdot,j} = \sum_{i=1}^I n_{i,j}$ et $n_{i,j}$ l'effectif à l'intersection de la i -ème ligne et de la j -ème colonne.

Puis on définit une loi de probabilité empirique en construisant le tableau des fréquences :

$$F_{(I,J)} = \begin{pmatrix} \vdots & & \\ \cdots & f_{i,j} & \cdots \\ \vdots & & \end{pmatrix} \text{ avec } f_{i,j} = \frac{n_{i,j}}{n} \text{ et } n = \sum_{i=1}^I \sum_{j=1}^J n_{i,j}$$

Ainsi, la probabilité qu'un individu de notre population prenne les modalités X_i et Y_j s'écrit :

$$P(x = X_i \text{ et } y = Y_j) = \frac{n_{i,j}}{n} = f_{i,j}$$

Intéressons-nous maintenant à la probabilité conditionnelle de y sachant x :

$$\mathbb{P}[x = X_i, y = Y_j | x = X_i] = \frac{n_{i,j}}{n_{i,\cdot}} = \frac{f_{i,j}}{f_{i,\cdot}}$$

Cette probabilité conditionnelle représente le nombre d'individus ayant les modalités X_i et Y_j par rapport à la sous population qui a la modalité X_i , il s'agit des profils-lignes.

De la même manière, on définit les profils-colonnes : $\mathbb{P}[x = X_i, y = Y_j | y = Y_j] = \frac{n_{i,j}}{n_{\cdot,j}} = \frac{f_{i,j}}{f_{\cdot,j}}$

Si les variables x et y sont indépendantes alors :

$$\begin{aligned} P(x = X_i \text{ et } y = Y_j) &= P(x = X_i) \times P(y = Y_j) \\ \Leftrightarrow \frac{n_{i,j}}{n} &= \frac{n_{i,\cdot}}{n} \times \frac{n_{\cdot,j}}{n} \\ \Leftrightarrow n_{i,j} &= \frac{n_{i,\cdot} \cdot n_{\cdot,j}}{n} \end{aligned}$$

Ainsi, l'effectif observé $n_{i,j}$ est égal à l'effectif théorique $\frac{n_{i.}n_{.j}}{n}$ en cas de situation d'indépendance.

Cette démonstration nous permet d'introduire le test d'indépendance non paramétrique pour des variables discrètes : le test du Khi-deux (χ^2) :

$$\sum_{i,j} \frac{\left(n_{i,j} - \frac{n_{i.}n_{.j}}{n}\right)^2}{\frac{n_{i.}n_{.j}}{n}} \sim \chi^2(I-1, J-1)$$

Cette formule compare la différence entre l'effectif observé et théorique et on divise les écarts par la fréquence moyenne de chaque modalité afin de ne pas donner plus d'importance aux modalités les plus fréquentes.

Si la statistique est inférieure au quantile à un seuil de 5% (par exemple), on accepte l'hypothèse H0 (les variables sont indépendantes) sinon on rejette l'hypothèse H0.

La distance entre deux profils-lignes s'écrit alors :

$$d_{\chi^2}^2(i, i') = \sum_{j=1}^J \left[\frac{n}{n_{.j}} \times \left(\frac{n_{i,j}}{n_{i.}} - \frac{n_{i',j}}{n_{i'.}} \right)^2 \right]$$

La distance entre deux profils-colonnes se définit de manière analogue.

Enfin, on définit l'inertie du nuage de points de la façon suivante :

$$I(\text{nuage}) = \sum_{i=1}^I \sum_{j=1}^J \frac{(f_{i,j} - f_{i.}f_{.j})^2}{f_{i.}f_{.j}} \sim \frac{1}{n} \chi^2(I-1, J-1)$$

L'inertie mesure l'intensité de la liaison entre les variables. Plus l'inertie est grande, plus l'intensité est élevée.

Variables de l'ACM et inertie du nuage

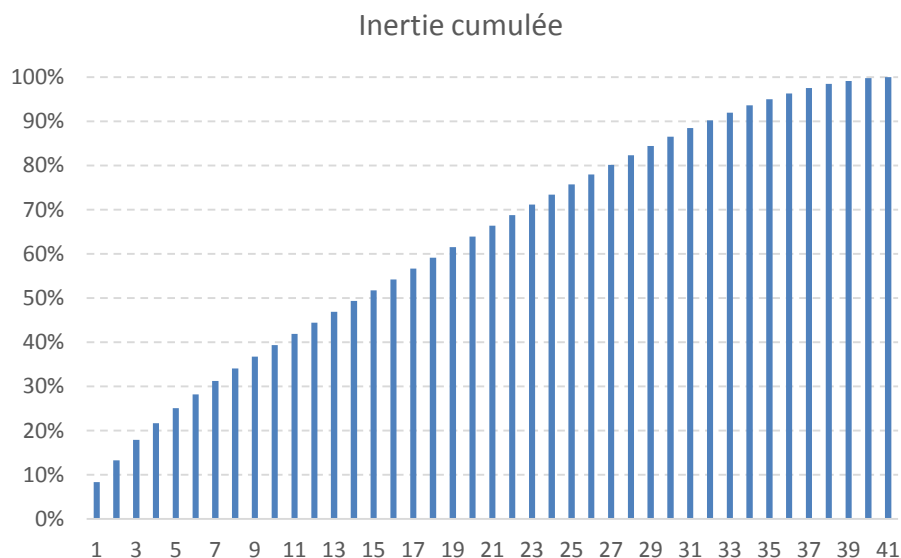
Le détail des variables introduites dans l'ACM est précisé dans le tableau infra.

Variable	Nombre de modalités -1
Age	6
revenu	6
CSP	8
Statut-type	3
Prime client	6
Ancienneté	2
Jeune actif	1
Segment bancaire	2
Fractionnement	3
Nombre pièces	3
Vente liée	1
Total	41

L'inertie du nuage est égale à 3.73 et n'a pas d'interprétation statistique dans la mesure où elle dépend du nombre de modalités de chaque variable. L'inertie se calcule de la manière suivante :

$$I = \frac{1}{M} \sum_{n=1}^M (p_m - 1) = \frac{1}{11} \times 41 = 3.73$$

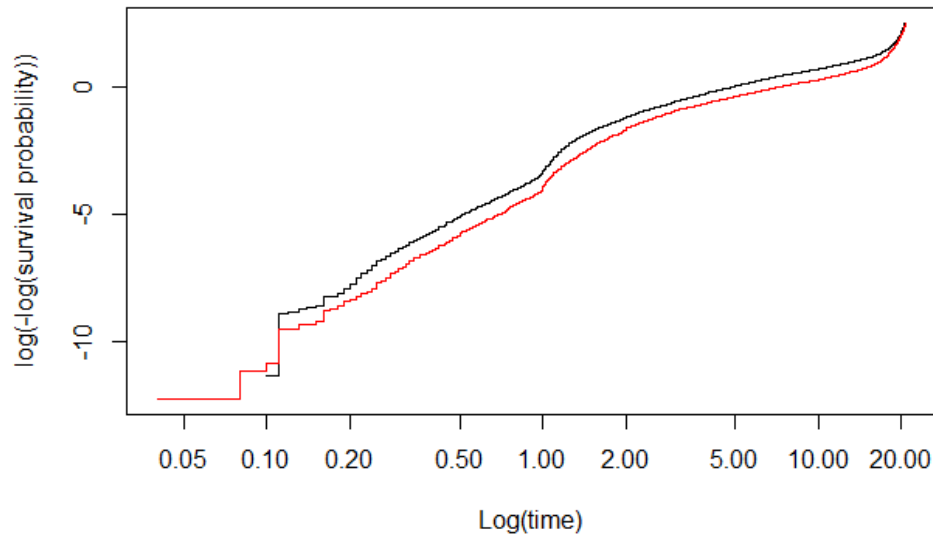
Inertie cumulée



➔ Ce graphique présente l'inertie cumulée de l'ACM.

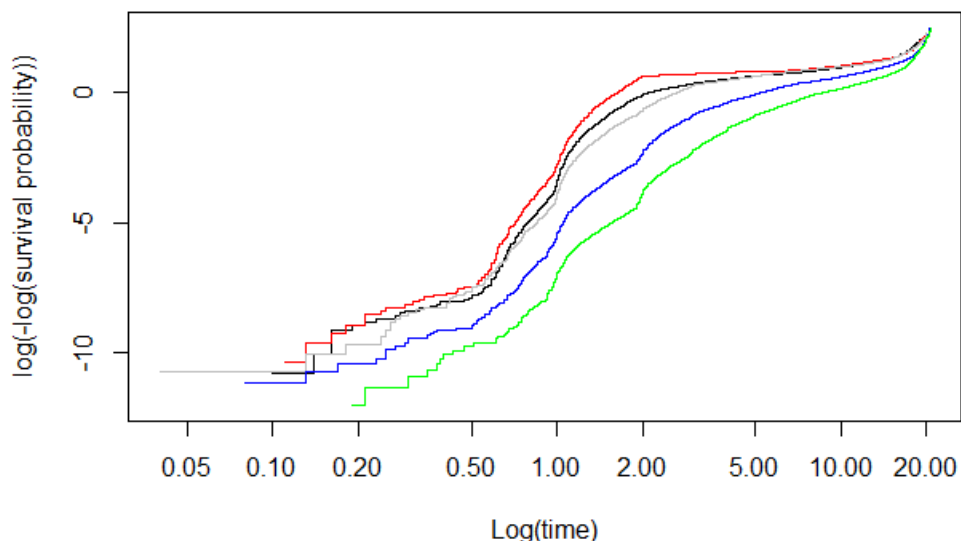
8. Modèle de Cox : graphiques d'hypothèse des risques proportionnels

log-log de la fonction de survie sur le statut d'occupation



- Le graphique du $\log(-\log(\text{survie}))$ de la variable *Statut d'occupation* présente les deux modalités de la variable qui se traduisent en fonction du log de l'ancienneté. Par conséquent, l'hypothèse de proportionnalité des risques est vérifiée pour cette variable.

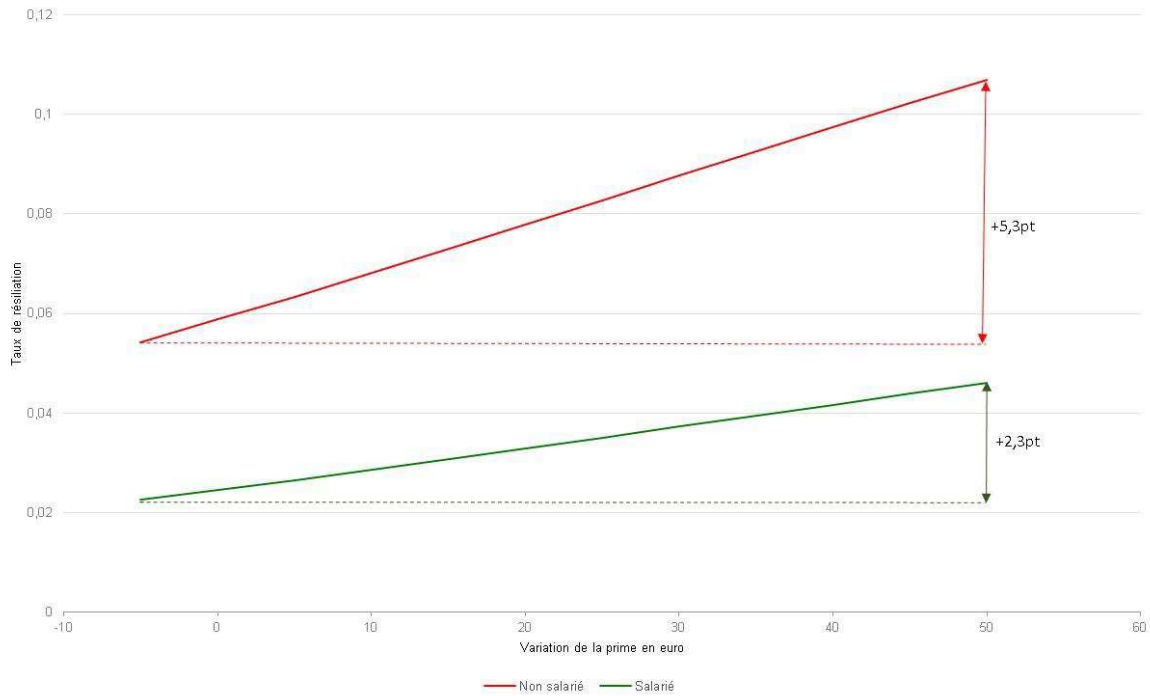
log-log de la fonction de survie en fonction de majoration en euro



- Le graphique du $\log(-\log(\text{survie}))$ de la variable *Majoration en euro* présente les cinq tranches de la variable discrétisée. Les courbes ne sont plus parallèles et se croisent à certains moments rendant caduque l'hypothèse de proportionnalité des risques pour cette variable.

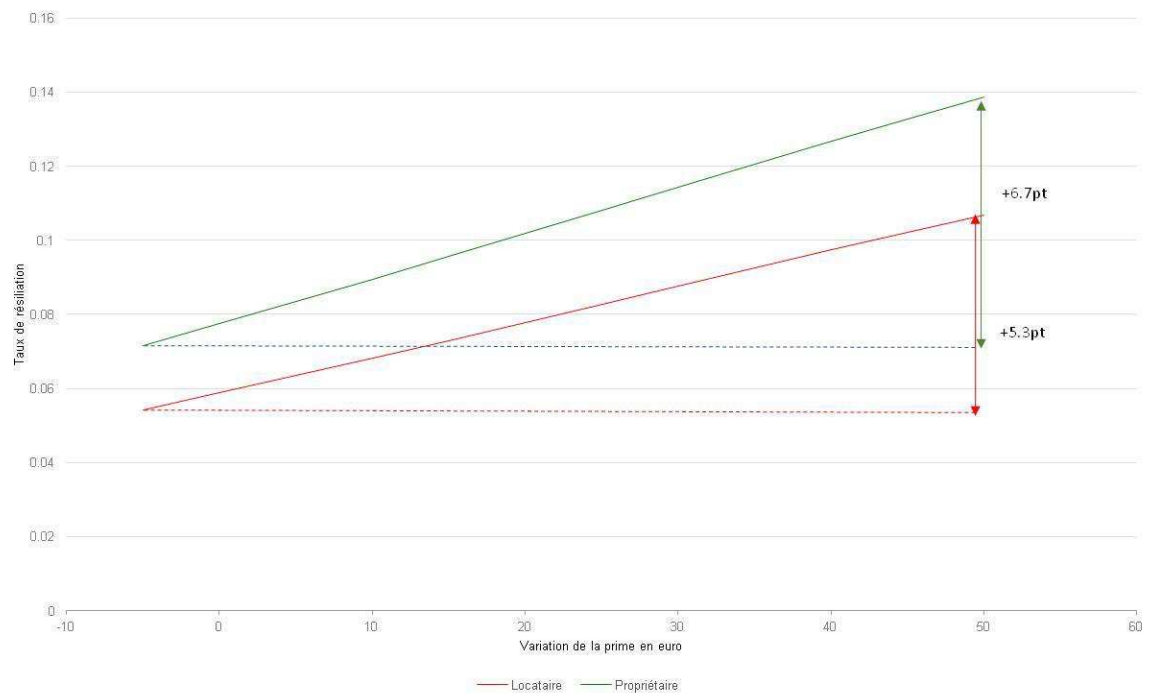
9. Sensibilité au prix : graphiques sur d'autres variables significatives

Graphique de sensibilité au prix en fonction de l'indicatrice salarié SG



➔ Les clients qui travaillent dans le groupe SG sont moins sensibles au prix. Un écart de 3 points sur l'élasticité au prix est constaté entre un client salarié et non salarié du groupe.

Graphique de sensibilité au prix en fonction du statut d'occupation



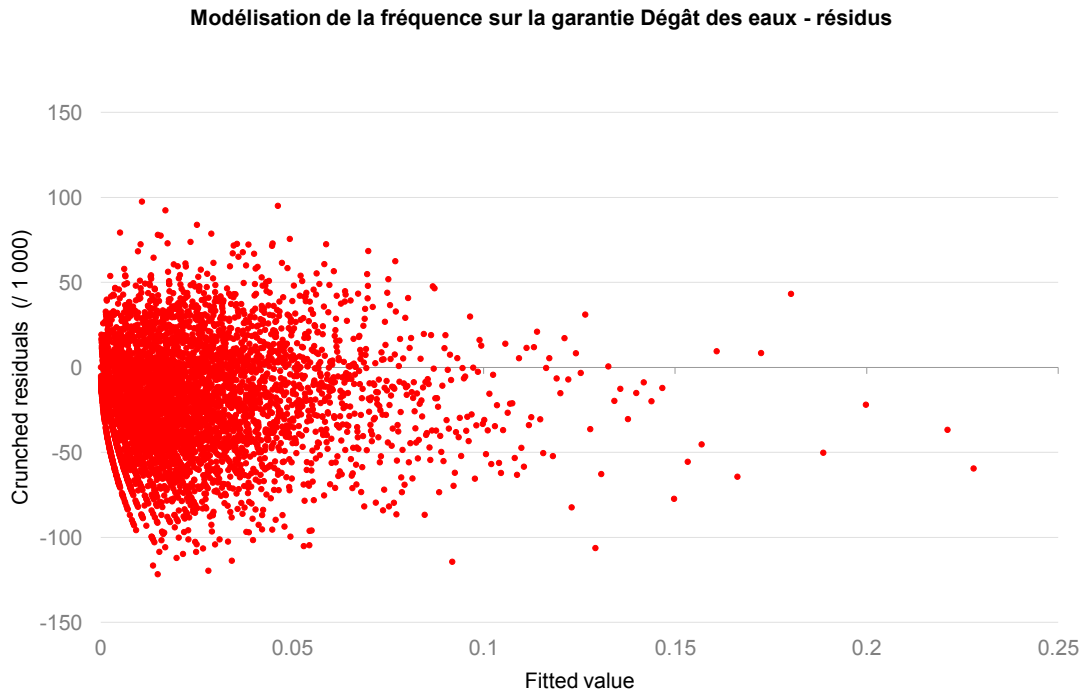
➔ Les propriétaires sont plus sensibles au prix (progression du taux de résiliation de 6.7 points pour une variation de prime entre -5€ et +50€ contre 5.3 points pour les locataires). Un écart de 1.4 points sur l'élasticité au prix est relevé entre ces deux segments.

10. Modélisation de la prime commerciale : formules de garanties

Votre bien	Appartement ou maison		
	1 à 4 pièces	15 pièces maximum	
Votre formule	INITIALE	CONFORT	OPTIMALE
Vos principales garanties			
Responsabilité civile	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Incendie et événements assimilés	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dégâts des eaux et gel	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Tempête, grêle, neige	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Attentats ou actes de terrorisme	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Catastrophes naturelles et technologiques	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Défense Pénale et Recours suite à accident	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Vol et détériorations suite à vol	option	<input type="checkbox"/>	<input type="checkbox"/>
Vol sur le lieu de vacances		<input type="checkbox"/>	<input type="checkbox"/>
Bris de glaces		<input type="checkbox"/>	<input type="checkbox"/>
Bris de glaces sur les meubles			<input type="checkbox"/>
Dommmages électriques et Valeur à neuf jusqu'à 2 ans sur matériel électroménager, HI-FI, vidéo et informatique		option	<input type="checkbox"/>
Valeur à neuf étendue : 4 ans sur matériel électroménager, HI-FI, vidéo et informatique 6 ans sur tous les autres biens (hors objets de valeur)			<input type="checkbox"/>
Dommmages corporels			<input type="checkbox"/>
Vos services d'assistance			
Assistance en cas d'urgence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Assistance au quotidien		<input type="checkbox"/>	<input type="checkbox"/>
Assistance électroménager, HI-FI et TV			<input type="checkbox"/>
Services Experts			<input type="checkbox"/>
Vos Packs à la carte*			
Pack Responsabilité civile Plus : - Assistante Maternelle - Chambres d'hôtes - Ânes, poneys ou chevaux		option	
Pack Piscine et Spa			
Pack Jardin			
Pack Loisirs			
Pack Environnement			
Vos extensions de garantie (souscription uniquement via Allô Contrats)			
Responsabilité civile de l'enfant à l'étranger		option	
Cave à vin en dépendances			
Vol en dépendances – plafond de 3 200 euros			
Vol en dépendances – plafond de 8 000 euros			
Les montants maximums pour lesquels vous êtes couvert			
Pour vos objets usuels	Se référer aux Conditions Particulières		
Pour vos objets de valeur			

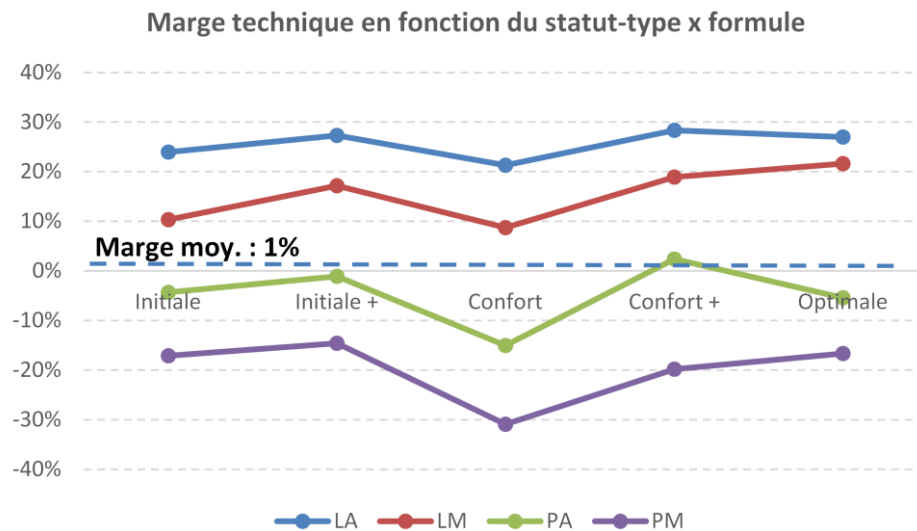
Inclus dans la formule. *Les packs que vous avez souscrits sont mentionnés dans vos Conditions Particulières.

11. Modélisation de la prime commerciale : résidus de la modélisation de la fréquence de la garantie dégât des eaux



➔ Le modèle estimé sur la fréquence de la garantie Dégât des eaux apparaît conforme étant donné que la dispersion des points est symétrique autour de 0 et il n'y a pas de tendance observée sur les résidus.

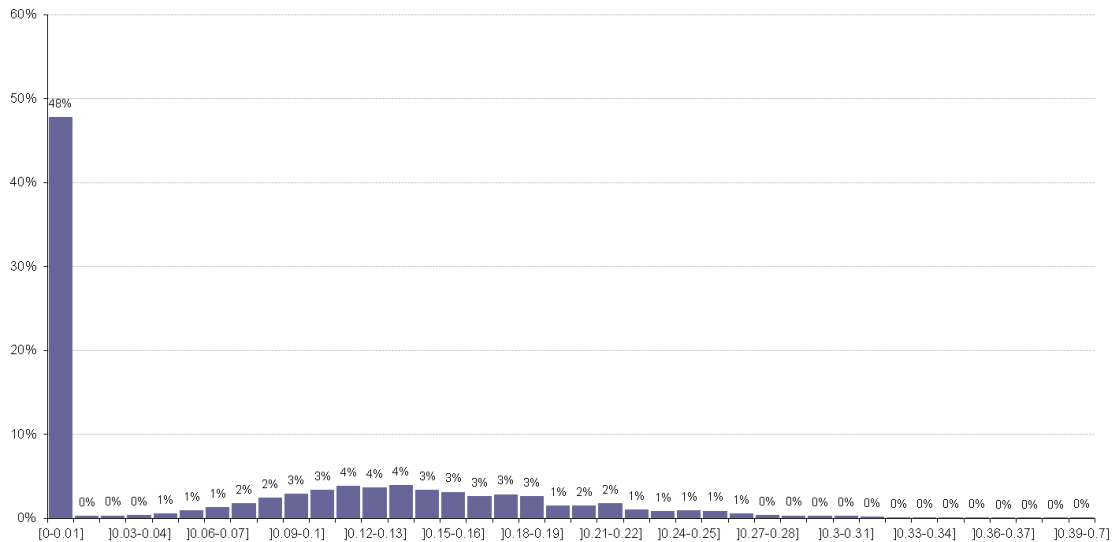
12. Modélisation de la prime commerciale : marge en fonction du statut-type croisée avec la formule



→ Le graphique indique une relative stabilité de la marge suivant la formule de garanties. Les locataires sont rentables pour l'ensemble des formules tandis que les propriétaires d'appartements sont proches de l'équilibre pour certaines formules. A l'inverse, les propriétaires de maison présentent un déficit technique pour l'ensemble des formules. A noter que la formule Confort présente un profil particulier avec une détérioration de la marge ou une aggravation du déficit pour l'ensemble des statut-types.

13. Optimisation de la marge : distribution des majorations de prime pour l'approche non paramétrique

Optimisation de la marge - distribution des majorations de prime pour la simulation non paramétrique



➔ Variabilité des majorations en pourcentages importantes pour l'approche non paramétrique de l'optimisation de la marge