Mémoire présenté le :
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA
et l'admission à l'Institut des Actuaires

Par : Bassirou BALDE

Titre**: Modélisation du taux de résiliation en Assurance MRH.**

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

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

Membre présents du jury de l'Institut des Actuaires

Brigitte DUBUS-THIRKELL
Lionel LAURENT

Membres présents du jury de l'ISFA

Pierre RIBEREAU
Véronique MAUME-DESCHAMPS

Secrétariat :

Mme Christine DRIGUZZI

Bibliothèque :

Mme Patricia BARTOLO

Entreprise

Nom : *BPCE Assurances*

Signature :

Directeur de mémoire en entreprise :
Nom : PAGANUS Jean François

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

Mots clés
Résiliation; Multirisques Habitation(MRH); Modèle logistique; Modèle de survie; Modèle de COX; Modèle de Kaplan-Meier; Loi Hamon; Modèle linéaire Généralisé; Modèle de Holt-Winters.

Résumé

Le système tarifaire des contrats Multirisques Habitation (MRH) de **BPCE Assurances** est actuellement basé sur une modélisation du risque par les Modèles Linéaires Généralisés et un système de surveillance de la sinistralité qui permet de pénaliser les clients sur sinistrés avec une hausse de leur prime voire une résiliation.
L'objectif principal de la modélisation du taux de résiliation est de participer à l'optimisation de ce système tarifaire. En effet en identifiant les profils de clients les plus aptes à résilier leur contrat, **BPCE Assurances** pourrait leur appliquer un ajustement tarifaire adéquat lors du passage de la prime technique à la prime commerciale, et ainsi affiner la segmentation du portefeuille.

Pour ce faire nous allons créer un score d'attrition avec la régression logistique. Dans un premier temps aucune précision ne sera faite sur l'horizon de la résiliation ce qui nous permettra de faire une première identification des profils les plus risqués pour la résiliation.
Cependant avec l'entrée en vigueur de la loi **Hamon**, la modélisation du taux de résiliation à un an reste plus pertinente pour d'une part produire un outil de projection de portefeuille annuel afin de se mettre dans les conditions d'application de cette loi et d'autre part intégrer les résultats dans la mise en place des tarifs annuels du produit MRH.
En outre afin de s'assurer des profils obtenus par la régression logistique, nous utiliserons les modèles de survie (modèle non paramétrique de kaplan-Meier et le modèle semi paramétrique de Cox) pour produire une nouvelle approche de la modélisation du taux de résiliation. En plus cette nouvelle modélisation nous permettra d'estimer la durée de vie d'un contrat MRH en portefeuille et sa durée de vie résiduelle suivant l'ancienneté et certains critères tarifaires. Notons que ces résultats pourront être utilisés dans les études de rentabilité du produit MRH.
Afin de se mettre dans les conditions d'application de la loi **Hamon**, nous ferons une projection de portefeuille sur un an. Celle-ci nécessite une estimation des affaires nouvelles (AFN) qui sera faite par la méthode de Holt-Winters et une application du modèle de résiliation sur un portefeuille actif.

Keywords
Termination ; Home insurance (“MRH”) ; Logistic model ; Survival model; Model of Cox; Kaplan-Meier model ; Hamon law; Generalized Linear Model; Holt Winters model.

Abstract

The fare system of the home insurance contracts of **BPCE Assurances** is currently based on a modelling of the risk by the Generalized Linear Models and a monitoring system of the loss ratio which allows to penalize the customers several disaster with an increase of their bonus even a termination.
The main objective of modelling of the rate of termination is to participate to the optimization of this tariff system. Indeed by identifying susceptible customer profiles to cancel their contract, **BPCE Assurances** could apply an adequate tariff adjustment during the passage of the technical price in commercial premiums, and by the way refine the portfolio segmentation.

In order to do so we are going to create a attribution score with logistic regression. At first no details will be made to the horizon of the termination. This will allow us to make a first identification of the profiles the most risked termination
However with application of the **Hamon** law, the modelling of rate termination in one year remains more relevant to produce in one hand a projection tool of annual portfolio to put itself in the conditions of application of this law and on the other hand integrate the results into the implementation of the annual lists price of the product “MRH”.
To make sure of profiles obtained by the logistic regression, we shall use the survival models (model non-parametric of Kaplan-Meier and the semi-parametric model of Cox) to produce a new approach of modelling the rate of termination.
This new modelling will allow us to estimate the cycle life of contract “MRH” in portfolio and its remaining life cycle in order of the seniority and certain tariff criteria. We can note that these results can be used in the studies of the profitability product of “MRH”.
To be in the enforcements conditions of **Hamon** law, we shall make a portfolio projection over one year. This one requires an estimation of the new subscriptions (“AFN”) which will be made by the method of Holt-Winters and the application of the termination model of an active portfolio.

Remerciements

Je tiens à remercier mon tuteur d'entreprise M. Jean François PAGANUS responsable du département Dommages aux biens et Risques Divers (DAB & RD), pour son encadrement, sa disponibilité et l'aide qu'il m'a portée tout au long de la réalisation de ce mémoire.

Je remercie M. Fréderic PLANCHET professeur à l'ISFA pour sa disponibilité et la clarté de ses réponses à mes diverses questions.

Enfin mes sincères remerciements à mes collègues du département **DAB & RD** pour leur accueil, leur sympathie et l'aide précieuse qu'ils m'ont apportée durant la réalisation de ce travail.

Sommaire

Introduction

Le marché de l'assurance non vie connait un essor considérable au fil de ces dernières années notamment depuis l'entrée de la bancassurance dans ce secteur.
En outre sa réglementation a connu récemment une forte évolution visant principalement à protéger les assurés.
En effet la **Loi Châtel** votée en **2005** et mise en application en **2012**, impose aux Assureurs de faire part aux assurés de leur avis d'échéance au moins 20 jours avant la date d'échéance de leur contrats, ce qui a pour but de les protéger de la tacite reconduction. Depuis décembre 2012, à la suite de la décision de la cour Européenne, il est interdit de discriminer les assurés sur la base de leur sexe. Cela a créé un véritable bouleversement pour les assureurs notamment dans la tarification en Assurance automobile. En plus, la **loi Hamon**, votée à l'Assemblée Nationale en juillet **2013,** permet depuis le 01 janvier 2015 aux assurés de pouvoir résilier à tout moment leurs contrats dommages après un an d'assurance.
Ces différents aspects conduisent les Assureurs à être de plus en compétitifs pour pouvoir garder leur rentabilité, et cela passe par l'optimisation de leur processus tarifaire.
La modélisation du taux de résiliation entre dans cette dynamique d'optimisation du tarif et devra permettre d'identifier les profils de clients les plus « fragiles » à la résiliation. Cela permettra à BPCE Assurances d'appliquer d'une part un tarif adéquat pour ces clients et d'autre part avoir une première visibilité de l'impact de la **loi Hamon** sur son portefeuille.

Dans un premier temps nous présenterons l'environnement de travail de ce mémoire en présentant le contexte et le périmètre de l'étude d'une part et d'autre part en décrivant la construction des bases d'études et le processus de sélection des variables explicatives.
Ensuite nous traiterons dans une deuxième partie de la modélisation du taux de résiliation par la régression logistique suivant les deux horizons retenus (horizon non fixé et à un an) et aussi par les modèles de survie en modélisant la durée de vie des contrats MRH que nous aborderons dans la troisième partie.
Enfin dans une quatrième partie nous ferons une comparaison des deux types de modélisation utilisés et une application de leurs résultats par une projection de portefeuille sur un horizon d'un an. Une estimation des affaires nouvelles par un modèle de série temporelle sera nécessaire pour cette partie.

Partie 1: Environnement de travail

1. Périmètre et contexte de l'étude

1.1. Périmètre de l'étude

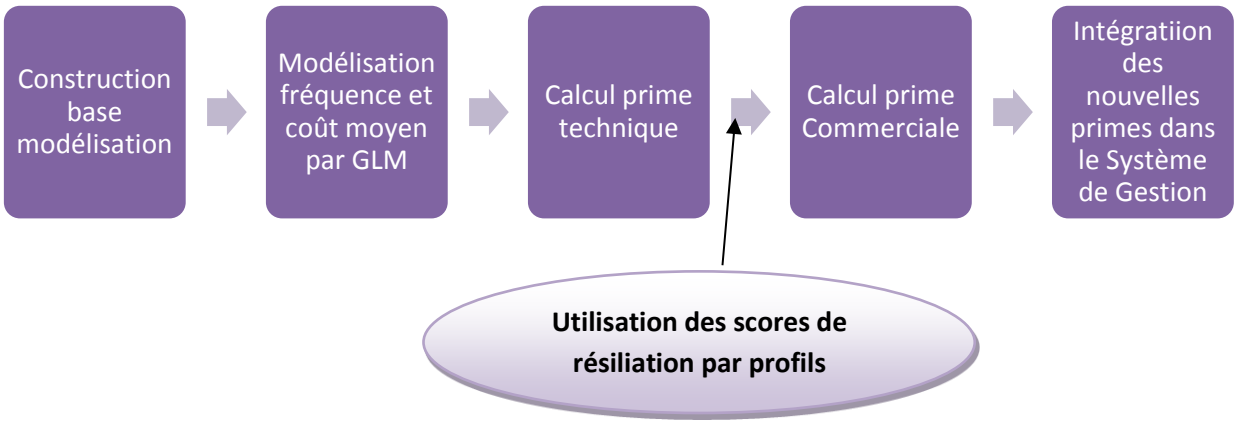
Notre étude va porter sur l'ensemble du produit MRH. Avec un chiffre d'affaire de 198,6 Millions en 2012 soit 31 % du CA globale de la compagnie, le produit MRH constitue ainsi un produit phare de BPCE Assurances qui en 2014 est classé 11^{ième} Assureur Habitation par le magazine **l'ARGUS de l'assurance.**

Notons que le contrat MRH de BPCE A dispose d'une large gamme de formules adaptées aux différents profils de clients (propriétaires, locataires, jeunes, résidences principales ou secondaire, logements en location).

La dernière génération de ce produit qui est actuellement commercialisé depuis Avril 2011 est la **MRH3**, les deux autres offres précédentes que sont **MRH1** et **MRH2** sont toujours présentes dans notre portefeuille. Ainsi dans la modélisation du taux de résiliation aucune restriction ne sera faite sur les différentes offres du produit MRH. De même, aucune restriction ne sera faite sur la nature de la résiliation (changement de domicile, déménagement souscripteur à l'étranger, changement d'offre…) car cette information n'est pas bien renseignée dans les bases de données.

1.2. Description du processus tarifaire

Comme dit plus haut la modélisation du taux de résiliation MRH a pour but de participer à l'optimisation du processus tarifaire de ce produit. Nous explicitons sur le schéma ci-dessous le processus de mise en place du tarif des contrats MRH.



Notons que c'est le modèle multiplicatif **Coût-fréquence** qui est utilisé. La modélisation est faite par les modèles linéaires généralisés pour les différentes garanties proposées dans les différentes formules.

La prime technique est obtenue en utilisant les résultats de la modélisation de la fréquence et du coût moyen en appliquant la formule **Prime technique=coût Moyen * fréquence.**

L'utilisation de la probabilité de résiliation sera faite entre le passage de la prime technique à la prime commerciale. En effet l'idée étant de pouvoir appliquer des ajustements tarifaires différenciés suivant le risque de résiliation pour les clients hors cible commerciale.

2. Construction des bases d'étude

Cette partie aborde la construction des bases d'étude utilisées dans les modélisations du taux de résiliation qui seront faites.

2.1. Base d'étude pour la modélisation du taux de résiliation à horizon non défini

2.1.1. Base d'entrée

Pour construire cette base d'étude nous utilisons la base **contrats** vue en janvier 2014. Celle-ci contient l'ensemble des contrats souscrits depuis 1999 avec leurs différents mouvements (situations). Nous y faisons les retraitements suivants :

- ⇨ Création de la variable cible **RESIL** prenant la valeur 1 ou 0 selon que le contrat a été résilié ou non.
- ⇨ Sélection de tous les produits MRH sauf le produit Responsabilité Civile vie privée (RCVP)
- ⇨ Suppression des contrats sans effet. Il s'agit des contrats avec une date d'effet ou de création du contrat égale à la date de fin du contrat. (date effet contrat = date de fin contrat ou date création= date fin de contrat)
- ⇨ Sélection de la situation la plus récente pour chaque contrat ce qui permet d'avoir la dernière vision du contrat et d'éliminer les doublons.

Afin d'augmenter le pouvoir prédictif du modèle sur le taux de résiliation qui sera construit, nous créons des variables potentiellement explicatives pour le risque résiliation que nous énumérons ci-après:

1. **ANCIENNETE** qui donne l'ancienneté du contrat en année par rapport à la date d'extraction de la base (31 janvier 2014) pour les contrats en cours et par rapport à la date de résiliation pour ceux qui sont résiliés.
2. **NOMBRE AVENANT** qui indique le nombre total d'avenant fait sur chaque contrat. Il existe deux types d'avenant : technique et administratif, l'un a un impact sur le tarif, et donc probablement sur le risque de résiliation et l'autre non car n'affectant en rien les termes du contrat. Donc nous décidons de ne retenir que les avenants techniques dans cette variable.
3. **NOMBRE CONTRATS MRH** qui compte le nombre de contrat MRH détenu par client. Ce calcul se fait grâce à la variable TITUL qui est un identifiant client. Comme notre base d'étude est construite par contrat, nous aurons ainsi sur chaque ligne un numéro de contrat, le nombre de contrat MRH détenu par le propriétaire du contrat. Notons qu'un client peut avoir un seul contrat sur sa résidence principale (RP) mais, plusieurs contrats en Résidence secondaire (RS) ou Propriétaire bailleur (PB).

4. **NOMBRE SINISTRES REGLES** correspond au nombre de sinistres MRH réglés pour chaque contrat. Pour créer cette variable nous utiliserons la base sinistre MRH pour compter le nombre de sinistre réglé pour chaque contrat et ensuite faire le rapprochement avec la base d'étude par numéro de contrat.

5. **NOMBRE SINISTRES SANS SUITE** contient le nombre de sinistre MRH clos sans suite (sinistre classé sans aucun règlement à l'assuré) par contrat; même principe de calcul que la variable **NOMBRE SINISTRES REGLES**

6. **NOMBRE SINISTRES CORPORELS** qui contient le nombre total de sinistre corporel pour chaque contrat MRH.

7. **NOMBRE SINISTRES CORPORELS REGLES** qui compte le nombre de sinistres corporels réglés pour chaque contrat MRH.

8. **NOMBRE CONTRAT IARD** donne, par client, le nombre de contrat total IARD détenu hors MRH sur les produits commercialisés par BPCE A : Garantie des Accidents de la Vie (GAV), Protection Juridique (PJ) et Assurance Automobile.

Comme pour la variable **NOMBRE CONTRATS MRH,** ces cinq dernières variables seront renseignées pour chaque contrat de notre base d'étude : à chaque contrat sera associé le nombre de contrats IARD détenu par son propriétaire, le nombre de sinistre réglés, le nombre de sinistres sans suite, le nombre de sinistres corporels total et réglés associés.

❖ Choix des variables de la base d'étude

En plus des variables construites précédemment nous devons choisir les variables à retenir dans notre base d'étude. Pour ce faire nous allons procéder d'abord à une sélection sur la pertinence des variables d'un point de vue «métier». En effet la base **contrats** contient 123 variables, il est judicieux de réduire ce nombre pour d'une part alléger l'exécution des programmes et d'autre part pour la robustesse du modèle qui devra être construit avec cette base.
Ainsi comme le but de cette étude est d'optimiser le tarif actuel du produit MRH, nous retiendrons d'abord les variables tarifaires auxquelles on ajoute les variables créées précédemment.
Voici la liste des variables tarifaires retenues :

Variables tarifaires	Description
type de contrat	Type d'offre souscrite (MRH1, MRH2, MRH3)
Type de résidence	Type de résidence couvert (RP, RS ou PB)
Qualité juridique	Indique si client est locataire ou propriétaire
Nombre de pièces	Nombre de pièces du logement assuré
Nombre de logements	Nombre de logement du Propriétaire Bailleur
Zone tarifaire	Zone de tarification
Zonier incendie	Crée à partir de la zone tarifaire
Zonier vol	Crée à partir de la zone tarifaire
Zonier dde	Crée à partir de la zone tarifaire
Franchise	Franchise
Surface dépendance	Surface des dépendances
Dépendance isolée	Possession d'une dépendance à une autre adresse ou non

Bois	Utilisation de bois dans la construction
Chaume	Utilisation de chaume dans la construction
Surface habitation	Surface d'habitation du logement assuré
Remise commerciale	Montant de la remise commerciale HT
Remise jeune	Réservé aux clients jeunes (-30 ans)
Option RC Equidé	Responsabilité Civil équidé en option
Option vélo	Assurance vol de vélo en option
Option Piscine	Assurance de la piscine en option
Option véranda	Assurance d'une véranda en option
Capital mobilier	Plafond d'indemnisation du capital mobilier
Type piscine	Type de piscine
Mesure de surveillance	Majoration tarifaire appliquée suite à une sur sinistralité
Logement	Type de logement (Appartement, Maison,..)
Formule	Formule souscrite

Tableau 1 : Variables Tarifaires

Variable créées
Ancienneté du contrat
Nombre d'avenants fait sur le contrat
Nombre de contrat MRH détenu par client
Nombre de sinistre MRH réglé par contrat
Nombre de sinistre sans suite MRH
Nombre de sinistre corporel par contrat
Nombre de sinistre corporel réglé par contrat
Nombre de contrat IARD hors MRH

Tableau 2 : Variables créées

❖ Retraitement des valeurs manquantes ou aberrantes

Dans cette section nous allons présenter les différents retraitements faits sur la base d'étude. Sous SAS nous créons trois macros variables **var_quali, var_dis** et **var_quant** qui contiennent respectivement les variables qualitatives, discrètes et les variables quantitatives continues. Cette distinction permet d'utiliser les procédures adéquates pour chaque type de variable pour les analyses statistiques que nous ferons par la suite.

➢ Variables qualitatives et discrètes

Pour ces variables nous utilisons sous SAS la procédure **freq** pour identifier les valeurs manquantes. Voici la liste des variables dont le pourcentage de valeurs manquantes est critique.

Variable	Pourcentage de Valeurs manquantes
Option vélo	**64,84 %**
Type piscine	**99,56 %**

Tableau 3 : Valeurs manquantes variables qualitatives et discrètes

Notons que les variables **type piscine** et **option vélo** ne sont renseignées que pour la dernière génération d'offre commercialisée (la MRH3). Ce qui explique cette forte absence de valeurs sur leurs observations.

Vu le nombre important de ces valeurs manquantes, nous décidons de supprimer ces deux variables de la base d'étude.

Nous faisons une recodification sur la variable formule en considérant les modalités F1, F2, F3, JEUNE qui diffèrent suivant les différentes options et garanties proposées dans chacune d'elle et Propriétaire Bailleur (PB) quel que soit le type de produit. Cela permet de diminuer le nombre de modalité de cette variable. En effet pour chaque produit hormis quelques particularités, nous retrouvons ces 5 types de formules et donc on peut enlever l'information du type de produit à cette variable. Par ailleurs nous pouvons retrouver l'information du type de produit dans la base par la variable **Type de produit**.
En plus les variables surface dépendance et dépendance isolée seront supprimées de notre base car elles sont mal renseignées.

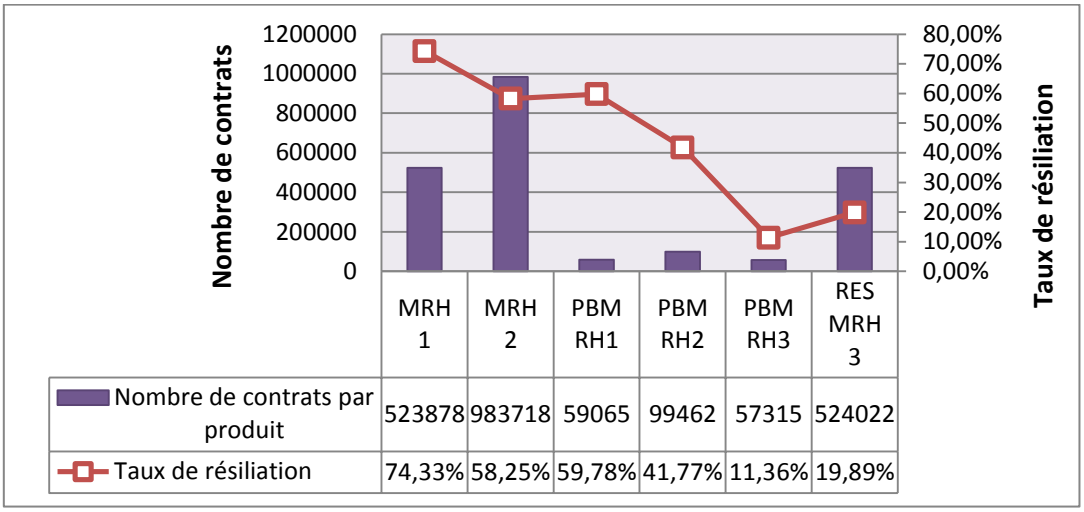
- **Variables quantitatives**

Nous utilisons la procédure **univariate** sous **SAS**. Nous obtenons que **surface habitation** présente un taux de valeurs manquantes de 24,82 % et que 90 % des valeurs indiquées sont à 0. Cela s'explique aussi par le fait qu'elle n'est renseignée que pour le produit MRH3. Nous supprimons cette variable de la base.
Pour **Nombre de Pièces** nous constatons des valeurs aberrantes en l'occurrence un nombre de pièces égales à 0 pour une portion de 0,03%. Par ailleurs nous constatons que cette proportion correspond à celles des valeurs manquantes pour les variables **Capital mobilier**, **qualité juridique, Type d'habitation**, **zone tarifaire**, **formule**, **chaume**, **bois**, **Surface dépendance** et après vérification on retrouve les mêmes observations. Donc nous décidons de supprimer ces observations.

2.1.2. Première analyse du taux de résiliation

Après les retraitements sur les différentes variables nous obtenons une base d'étude de **2 246 152** contrats. Dans cette section on se propose de faire des tris à plat sur certaines variables clés par rapport à la résiliation. Notons que le taux de résiliation au global c'est-à-dire le rapport entre nombre de contrat résiliés et le nombre total de contrats du portefeuille est de **51,17 %,** cela s'explique par le fait que nous n'avons fait aucune restriction sur le type de produit (MRH1 MRH2 MRH3), ni sur l'année d'exercice soit tous les contrats en portefeuille depuis 1999 année de début de commercialisation de la première offre (MRH1).

❖ **Tri à plat type de Produit**

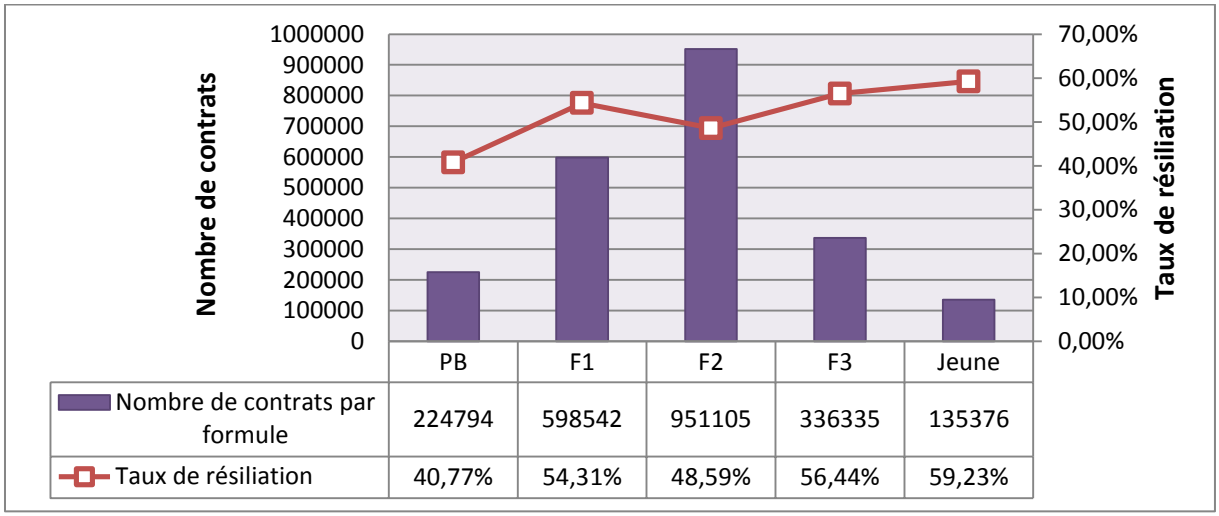


	MRH 1	MRH 2	PBM RH1	PBM RH2	PBM RH3	RES MRH 3
Nombre de contrats par produit	523878	983718	59065	99462	57315	524022
Taux de résiliation	74,33%	58,25%	59,78%	41,77%	11,36%	19,89%

Figure 1 : Taux de résiliation par type de produit

Nous remarquons sur ce graphe qu'il y a plus de résiliation sur le produit MRH 1 et MRH 2 par rapport à la MRH3 dont le taux reste en dessous des 35 %. Cela s'explique par le fait que depuis 2011 seule l'offre MRH3 est commercialisée et donc les autres produits tendent à disparaitre du portefeuille avec le temps.

❖ **Tri à plat Formule**



	PB	F1	F2	F3	Jeune
Nombre de contrats par formule	224794	598542	951105	336335	135376
Taux de résiliation	40,77%	54,31%	48,59%	56,44%	59,23%

Figure 2 : Taux de résiliation par formule

Pour la **formule**, nous observons plus de résiliations sur la F3 et Jeune et une forte présence de clients possédant une formule F2 avec un taux de résiliation moins important que les autres type de formule excepté PB qui présente le taux le plus bas. Le fort taux de résiliation de la formule JEUNE s'explique par la volatilité de la population qui la compose. En effet cette formule permet d'assurer des logements limités (2 pièces maximum) avec un tarif préférentiel pour les clients ayant moins de 30 ans.

❖ **Tri à plat qualité juridique**

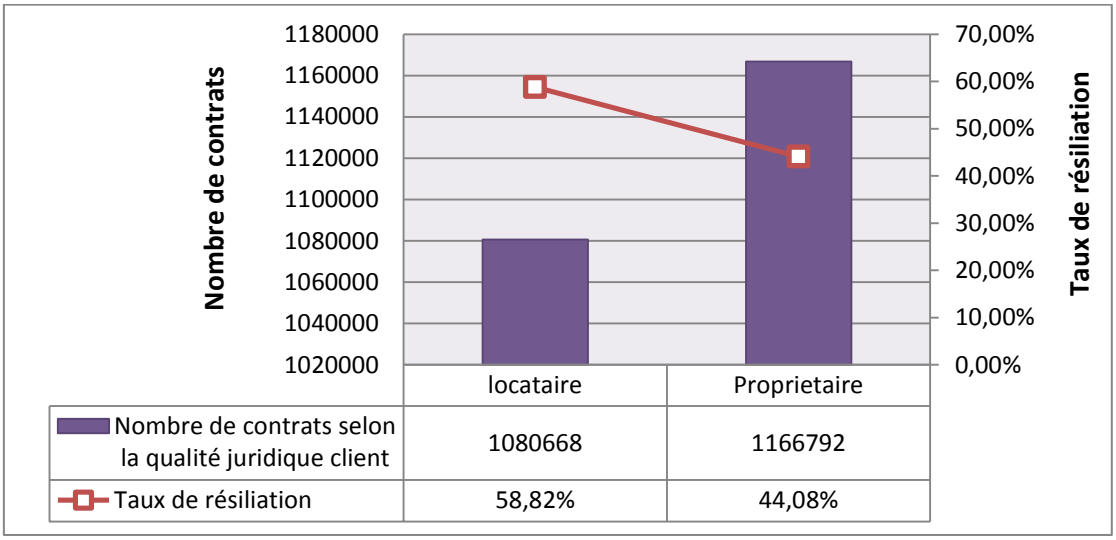


Figure 3 : Taux de résiliation par qualité juridique

Nous constatons que le taux de résiliation est plus important chez les locataires que les propriétaires ce qui parait assez logique car un locataire change plus souvent de logement ce qui constitue une occasion de changer d'assureur.

❖ **Tri à plat type d'habitation**

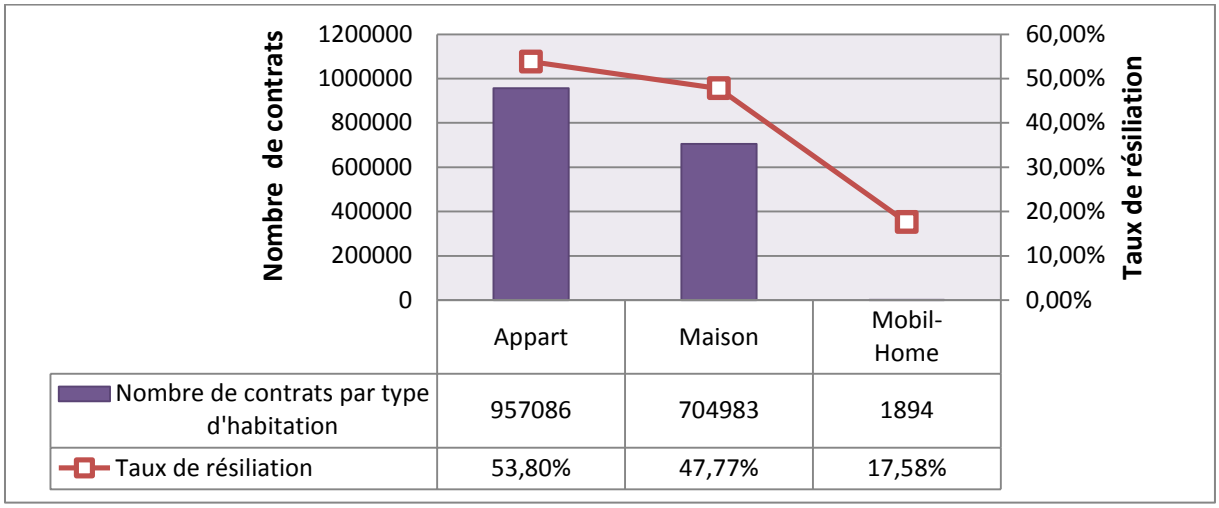


Figure 4 : Taux de résiliation par type d'habitation

Les appartements sont un peu plus touchés par la résiliation que les maisons, ce qui nous parait aussi logique car ce type de logement étant le plus souvent occupé par des locataires (cf. Tableau 4) qui ont tendance à plus résilier que les propriétaires. Notons que les contrats sur les mobil-homes ont commencé à être commercialisés en Avril 2011 d'où leur faible présence dans le portefeuille et donc un taux de résiliation moins important que pour les autres type de logement.

Qualité juridique	Type d'habitation	Répartition sur base modélisation
Propriétaire	Appartement	18,47 %
	Maison	33,37 %
	Mobil-Home	0,08 %
locataire	Appartement	38,28 %
	Maison	9,79 %
	Mobil-Home	0,01 %

Tableau 4 : Répartition qualité juridique client suivant le type d'habitation

2.1.3. Etude des corrélations

L'étude des corrélations se fera à l'aide de deux mesures: le V de cramer et la statistique du test de Kruskal Wallis.

❖ **V de Cramer**

1. Statistique de Khi deux

Soit X et Y deux variables aléatoires qualitatives. On considère leur tableau de contingence contenant les effectifs observés n_{lc} associé conjointement à la modalité y_l de la variable Y en ligne ($Y \in \{y_1, .., y_L\}$) et la modalité x_c de la variable X en colonne ($X \in \{x_1, \dots, x_C\}$). Notons par $n_{l.}$ et $n_{.c}$ les effectifs marginaux; $K = L \times C$ et n l'effectif globale du tableau.

Le χ^2 de Pearson permet de comparer les effectifs observés et les effectifs théoriques que l'on obtiendrait si les deux variables étudiées étaient indépendantes (l'hypothèse H0). Il s'obtient par la formule suivante:

$$\chi^2 = \sum_{k=1}^{K} \frac{(o_k - e_k)^2}{e_k}$$

où e_k correspond aux effectifs sous H_0 et o_k correspond aux effectifs observés ; $e_k = \frac{n_{l.} \times n_{.c}}{n}$

2. Test de khi deux

Ce test est basé sur les hypothèses suivantes:

$H_0 : X \text{ et } Y \text{ sont indépendants}$

$H_1 : X \text{ et } Y \text{ sont non indépendants}$

Sous H_0 la statistique de χ^2 suit asymptotiquement une loi de χ^2 à $(I-1)(J-1)$ degré de liberté. Avec I le nombre d'observation de la variable X et J celui de la variable Y.
Le test de χ^2 permet de déterminer si une liaison est significative mais ne quantifie pas l'intensité de la liaison. Le V de cramer permet de pallier à cela, et mesure cet intensité sur une échelle entre 0 et 1. Il est calculé par la formule ci-dessous :

$$V = \sqrt{\frac{\chi^2}{n \times \min(I-1; J-1)}}$$

❖ Test de Kruskal Wallis

C'est un test non paramétrique qui peut être étendu à K populations $(K \geq 2)$. Il est considéré comme l'alternative non paramétrique de l'analyse de la variance (ANOVA) dès que la distribution sous-jacente des données n'est plus gaussienne. Le test est basé sur les hypothèses suivantes :

H_0 : Les moyennes des K populations sont égales

H_1 : Il existe au moins une moyenne différente des autres

Comme dans chaque test non paramétrique et à la différence des tests paramétriques, le calcul ne porte pas sur les valeurs numériques des mesures issues des échantillons représentatifs des populations, mais sur leurs rangs attribués suite au classement des valeurs par ordre croissant.

Soit $\bar{r}$ la moyenne globale des rangs et $\bar{r}_k$ la moyenne des rangs pour les observations de la population k, la statistique de Kruskal-Wallis est définie de la manière suivante:

$$KW = \frac{12}{n(n+1)} \sum_{k=1}^{K} n_k (\bar{r}_k - \bar{r})^2$$

Si les effectifs sont importants ou si K > 6, la statistique de KW suit une loi de khi deux à K-1 degré de liberté.

Dans notre étude, le nombre de population est fixé à 2 (les contrats résiliés et ceux en cours) et donc nous utiliserons ce test pour comparer les distributions conditionnelles des variables quantitatives sur ces deux populations. Si l'hypothèse nulle est rejetée, alors plus la statistique KW est élevée, plus la variable considérée sera corrélée à la variable cible.

❖ Corrélation entre variables explicatives et variable cible

Pour aboutir à une modélisation robuste du taux de résiliation il est important de bien choisir les variables explicatives. Ce choix passe par une étude de la corrélation entre les variables explicatives et la variable cible mais aussi entre elles.

Nous utiliserons le V de Cramer comme mesure de dépendance pour les variables qualitatives-discrètes. Pour les variables quantitatives leur non normalité nous conduit à utiliser le test non paramétrique de Kruskal Wallis.

Ainsi selon la significativité de cette corrélation nous pouvons éliminer certaines variables de la modélisation. Cependant afin de ne négliger aucune information, nous ferons le choix de faire plusieurs modèles suivant que l'on supprime ou non des variables.

Voici la liste des variables qualitatives dont la corrélation à la variable cible est relativement faible. Nous retrouvons la liste complète de ces corrélations dans le Tableau 8.

Variable qualitative-discrètes	V de Cramer
Zonier Incendie	**0,00833**
Zonier vol	**0,00762**
Option RC équidé	**0,00708**
Nombre sinistre corporel réglés	**0,00404**
Bois	**0,00095**
Chaume	**0,00090**

Tableau 5 : Variables qualitatives moins corrélées à la variable cible

Pour les variable quantitatives le test est significatif au seuil de 5 % pour toute les variables et donc les moins corrélées à la variable cible sont celles dont la statistique de Kruskall Wallis est la plus petite.

Variables quantitatives	Statistique test de KruskalWallis
Prime	**108264**
Ancienneté	**35445**
Remise commerciale	**9007,56**

Tableau 6 : Variables quantitatives moins corrélées à la variable cible

Dans la construction de modèle prédictif il est préférable de discrétiser les variables continues afin d'augmenter la pouvoir prédictif du modèle et aussi pour faciliter l'interprétation des résultats. Ainsi nous discrétiserons la prime, l'ancienneté et le montant de la remise commerciale**,** les variables ainsi créées occupent respectivement les positions, $3^{ième}$,$4^{ième}$ et $15^{ième}$ sur 26 variables selon le V de Cramer.

❖ **Corrélation entre variables explicatives**

Dans cette partie nous nous intéressons à la colinéarité des variables explicatives. En effet l'étude de la corrélation linéaire entre les variables explicatives dans les modèles linéaires généralisés est indispensable.
Une règle empirique [1] que nous retiendrons pour le choix des corrélations acceptable est la suivante:

- Si coefficient de corrélation dépasse 0,9 corrélation inacceptable
- Si coefficient de corrélation dépasse 0,8 corrélation très dangereuse
- Si coefficient de corrélation dépasse 0,7 corrélation à surveiller.

Ainsi l'analyse de la corrélation entre les variables qualitatives et discrètes nous révèle trois corrélations inacceptables, deux dangereuses et trois à surveiller comme le montre le tableau suivant.

[1] Nous retrouvons cette règle dans **TUFFERY Stéphane [2010], p .94**

Variable 1	Variable 2	V Cramer
Zone tarifaire	**Zonier vol**	**1**
Zone tarifaire	**Zonier incendie**	**1**
Zone tarifaire	**Zonier DDE**	**1**
Avantage jeune	**Formule**	**0,89513**
Caisse	**Zonier incendie**	**0,82623**
Type produit	**Formule**	**0,77409**
Franchise	**Formule**	**0,70889**
Type de résidence	**Type produit**	**0,68194**

Tableau 7: Corrélation entre Variables qualitatives

Les **zoniers vol**, **dégâts des eaux** et **incendie** sont créées à partir de la variable **Zone tarifaire** ce qui explique cette corrélation parfaite entre ces variables et cette dernière. Pour la **formule** nous retrouvons l'information sur **type produit**, ainsi leur corrélation est aussi logique.

Ainsi pour éliminer ces corrélations dangereuses nous faisons le choix d'enlever l'une des variables de l'étude pour chaque paire. Vu que les variables **Zonier vol, Zonier incendie, Zonier DDE** sont moins significatives que **Zone tarifaire** par rapport à la variable cible, ce sont elles qui seront supprimées dans un premier temps.
Cependant la variable **Zone tarifaire** possède beaucoup de modalités, nous préférons dans un second temps garder les variables zoniers qui n'ont au maximum que 5 modalités pour la simplicité du modèle. Nous testerons les deux modèles selon que l'on retire ou garde les zoniers.
Pour la corrélation entre **Type produit** et **Formule** et entre **Avantage jeune** et **Formule** nous choisissons de retenir la variable **Formule**. Elle est moins significative que ces deux variables mais reste primordiale dans la détermination du profil des clients MRH.
Nous gardons **Formule** et **Franchise** dans le modèle, ce qui pourra être revu par la suite pour simplifier le modèle final.
Entre **Caisse** et **Zonier incendie** nous gardons l'une ou l'autre selon que l'on soit dans un modèle sans zonier ou avec zonier.
Pour les variables quantitatives nous n'observons pas de corrélations dangereuses.
Ainsi nous obtenons la liste des variables à retenir pour la suite.

VARIABLES	V Cramer avec RESIL
Type Produit	0,41057
Zone tarifaire	0,39042
Capital mobilier	0,24031
Prime discrétisée	0,2146
Ancienneté discrétisée	0,21374
Franchise	0,20495
Nombre avenant	0,18919
Caisse	0,18846
Remise commerciale	0,14999
Qualité juridique	0,14734
Nombre de Logement	0,11669

Formule	0,09872
Nombre de contrat IARD	0,08688
Nombre de Pièces	0,07428
Type d'habitation	0,06293
Nombre de sinistres sans suite	0,05238
Zonier DDE	0,04648
Nombre de contrat MRH	0,04128
Nombre de sinistres réglés	0,03715
TYPE Résidence	0,02514
Option véranda	0,02256
Nombre de sinistre corporels	0,01461
Mesure de Surveillance	0,01409
Zonier incendie	0,00833
Zonier vol	0,00762
Option RC Equidé	0,00708
Nombre de sinistres corporels réglés	0,00404
Bois	0,00095
Chaume	0,0009

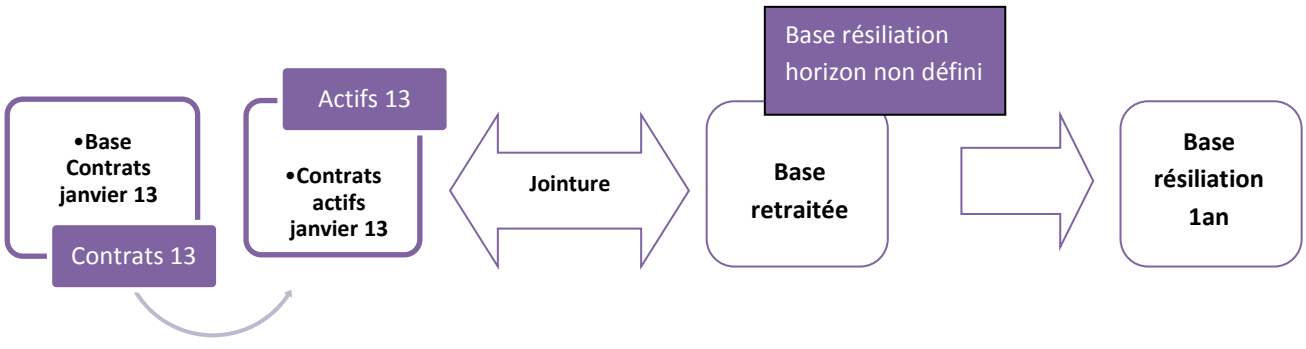
Tableau 8: Liste des variables explicatives

2.2. Base d'étude pour la modélisation du taux de résiliation à un an

2.2.1. Méthodologie

Le taux de résiliation à un an de l'année N se défini comme étant le nombre de contrats résiliés au cours de l'année N parmi ceux actifs en N-1.

Pour construire la base de modélisation pour la résiliation à un an, nous nous baserons sur cette définition en prenant comme base d'entrée celle que nous avons créée pour la première modélisation. Nous décrivons sur le schéma suivant le process appliqué :



La **jointure** permet de ne retenir que les contrats de la base retraitée qui étaient actif en 2013. Celle-ci correspond à la base d'étude construite pour la première modélisation du taux de résiliation.

Ainsi la base de modélisation du taux de résiliation à 1 an contient tous les contrats actifs de janvier 2013, avec une variable **RESIL** indiquant si ces derniers sont résiliés ou non en fin janvier 2014.

2.2.2. Choix et retraitement des variables

❖ **Choix des variables**

En plus des variables du Tableau 8, nous créons deux variables pour cette nouvelle base d'étude. Ces dernières indiquent la variation de la prime pour chaque contrat par rapport à la première prime connue que nous appellerons prime début ; et par rapport à la prime de l'année précédente (2013). Elles sont nommées respectivement **indicep1** et **indicep2** et calculées par les formules suivantes **:**

$$Indicep1 = \frac{pime\ en\ cours - prime\ début}{prime\ début} \qquad Indicep2 = \frac{pime\ en\ cours - prime\ 2013}{prime\ 2013}$$

❖ **Retraitement des variables**

Comme la base construite a déjà été retraitée, nous traiterons dans cette partie que les différentes recodifications faites sur les variables.

Nous décidons de discrétiser les variables **prime, ancienneté** et les deux indices de prix créés précédemment afin de faciliter l'interprétation des résultats et obtenir une segmentation plus précise par rapport au risque de résiliation.

Pour les variables **Nombre de contrat MRH**, **Nombre de contrat IARD**, **Nombre de logement**, **Nombre de sinistres réglés**, **Nombre d'avenants**, **Nombre de sinistres sans suite** nous faisons des regroupements de certaines modalités selon le niveau d'exposition au risque. Le tableau suivant indique les modalités prises par les variables recodifiées.

Variables	Modalités
Prime	0-204, 204-269, >=269
Ancienneté	0-1,4ans, 1,4-2 ans, 2-3 ans, 3-6ans, >=6ans
Indice 1 variation prime	[-0.18-0],] 0-0.1],] 0.1-0.7],] 0.7-2], Pas de variation
Indice 2 variation prime	[-0.1 - 0],] 0- 0.4], Pas de variation
Nombre de contrat MRH	1, 2, 3, 4, 5, 5+
	1, 2,

Nombre de contrat IARD	3, 4, 4+
Nombre de sinistre sans suite	1, 2, 3, 3+
Nombre de logement	1, 2, 3, 3+
Nombre de pièce	1, 2, 3, 4, 5, 6, 7, 7+
Nombre de sinistre réglés	1, 2, 2+

Tableau 9 : Modalités des variables recodifiées

2.2.3. Première analyse du taux de résiliation à un an

Nous faisons des tris à plat sur les variables **nombre de sinistres réglés**, **indicep1** et **indicep2**.

❖ **Tri à plat nombre de sinistres réglés**

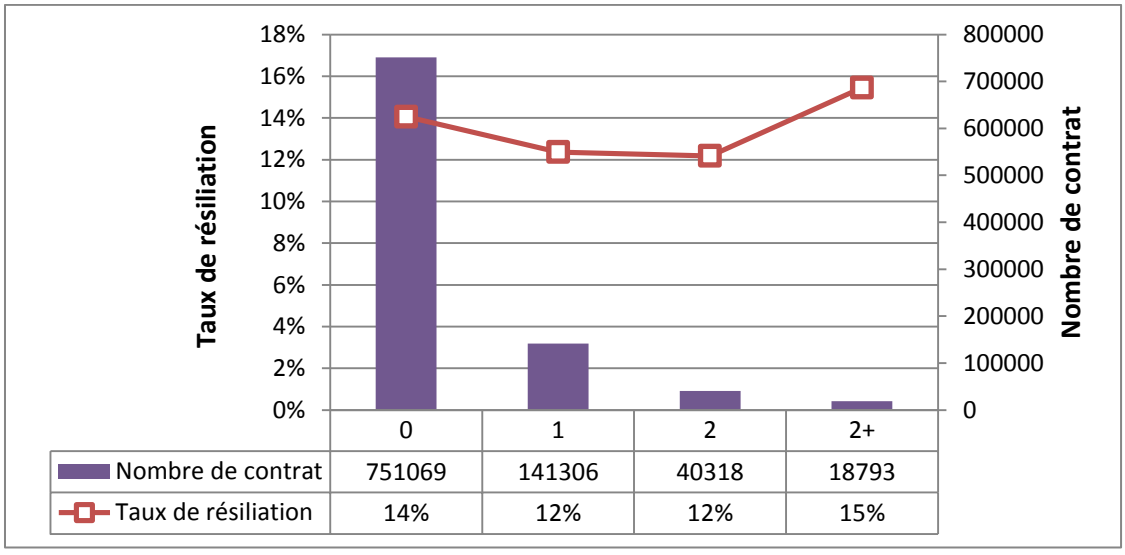
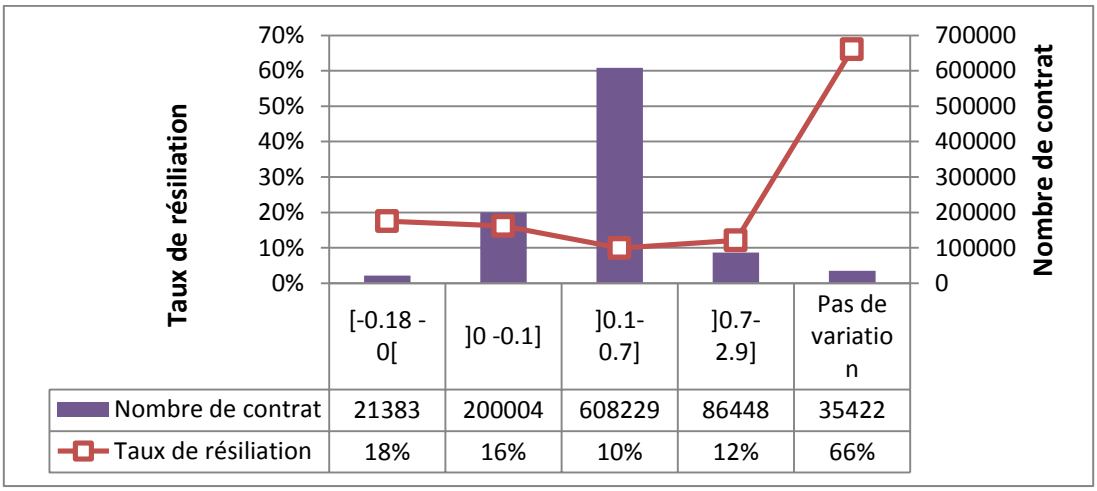


Figure 5 : Taux de résiliation par sinistres réglés

Nous voyons que les clients ayant plus de 2 sinistres réglés ont un taux de résiliation à un an plus important que ceux qui en ont eu moins. Ceux n'ayant pas de sinistres réglés ont un taux de résiliation relativement plus élevé que ceux ayant 1 ou 2 sinistres réglés.

❖ **Tri à plat indice variation prime par rapport prime début**

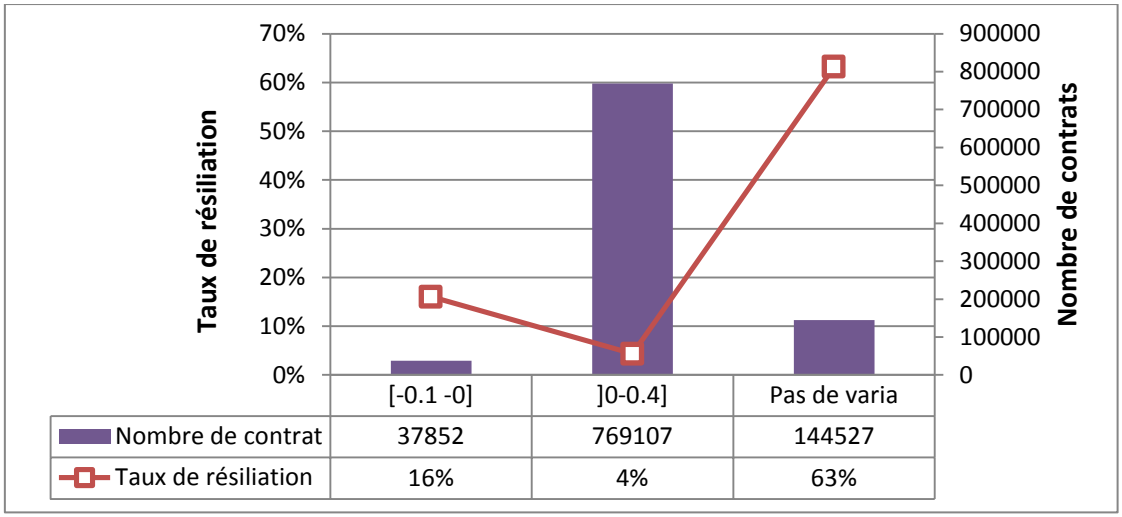


	[-0.18 - 0[	]0 -0.1]	]0.1- 0.7]	]0.7- 2.9]	Pas de variation
Nombre de contrat	21383	200004	608229	86448	35422
Taux de résiliation	18%	16%	10%	12%	66%

Figure 6 : Taux de résiliation par rapport à la variation de la prime début

Les contrats n'ayant pas eu de variation de prime par rapport à la première année d'assurance ont un taux de résiliation à un an largement plus élevé que les autres contrats bénéficiant d'une réduction ou d'une augmentation de prime.

❖ **Tri à plat indice variation prime par rapport à prime 2013**



	[-0.1 -0]	]0-0.4]	Pas de varia
Nombre de contrat	37852	769107	144527
Taux de résiliation	16%	4%	63%

Figure 7 : Taux de résiliation par rapport à la variation de la prime 2013

Comme pour le premier indice de variation de prime, les contrats n'ayant pas eu de changement de prime par rapport à rapport à la prime de 2013 ont un taux de résiliation largement plus élevés que les autres contrats.

2.2.4. Etude des corrélations

Nous faisons dans cette partie une première sélection des variables qui seront utilisées dans la l'élaboration des modèles. Cette sélection est basée sur la corrélation entre les différentes variables candidates et la variable cible et la corrélation entre les variables elles-mêmes.

❖ **Etude de la corrélation entre variables explicatives et variable cible**

Nous étudions dans cette partie la corrélation entre la variable cible et les autres variables de notre base d'étude. En ayant fait le choix de n'avoir que des variables qualitatives et discrètes dans le modèle, la mesure de dépendance qui sera utilisée est le V de cramer. Voici les valeurs obtenues pour chaque variable.

Variable	V de Cramer /RESIL
Indiceprix2	61,13%
Indiceprix1	30,84%
Qualité juridique	10,96%
Nombre de contrat MRH	10,28%
Type de Produit	9,72%
Formule	9,69%
Zone Tarifaire	9,29%
Ancienneté	7,07%
Nombre de contrat IARD	5,45%
Prime	5,41%
Nombre de logement	5,39%
Type de Résidence	5,33%
FRANCHISE	4,98%
Nombre de pièces	4,11%
Type d'habitation	3,96%
Nombre d'avenant	3,66%
Caisse	3,55%
Remise commerciale	2,73%
Avantage jeune	2,50%
Capital mobilier	2,34%
Nombre sinistres réglés	2,10%
Mesure de surveillance	1,92%
Zonier DDE	1,30%
Zonier vol	0,84%
Option piscine	0,78%
Option Véranda	0,68%
Zonier incendie	0,51%
Nombre sinistres corporels	0,38%
Nombre sinistre sans suite	0,25%
Nombre sinistres corporels réglés	0,24%
Bois	0,10%
Chaume	0,01%
Option RC Equidé	0,00%

Tableau 10 : Corrélation variables explicatives avec variable cible

Nous voyons que les deux indices de variation de primes créés sont fortement corrélés à la variable de résiliation. Les variables mise en rouge dans le tableau seront écartées du modèle à cause de la faible corrélation avec la variable modélisée.

- **Corrélation entre variables explicatives**

L'étude de la corrélation entre les variables explicative permet de détecter les colinéarités entre celles-ci. Pour les modèles de type modèle linéaire généralisé cette phase est indispensable. Nous regroupons dans le tableau suivant les plus fortes corrélations entre les variables explicatives.

Variable1	Variable2	Valeur absolue V Cramer
Zone Tarifaire	**Zonier incendie**	1
Zone Tarifaire	**Zonier DDE**	1
Zone Tarifaire	**Zonier vol**	1
CAISSE	**Zonier incendie**	0,82665
Avantage Jeune	**Formule**	0,75633
Type de résidence	**Type Produit**	0,70895
Type de résidence	**Formule**	0,59493
Nombre logement	**Type de Résidence**	0,5926
Nombre logement	**Type Produit**	0,57633
Nombre logement	**formule**	0,57518
Type Produit	**formule**	0,57439
Nombre de sinistre corporels	**Nombre de sinistre corporels réglés**	0,55148
Nombre de pièces	**Prime**	0,54443
Type Produit	**Ancienneté**	0,52813
FRANCHISE	**Type Produit**	0,51495
CAISSE	**Zonier vol**	0,50371
CAISSE	**Zonier DDE**	0,46392
Type Produit	**Capital mobilier**	0,45076

Tableau 11: Corrélation entre variables explicatives

La corrélation parfaite entre la variable **zone tarifaire** et les différents zoniers nous conduit à retenir que les **zoniers DDE** et **vol** dans la modélisation pour les mêmes raisons explicités dans la première modélisation. La variable **zonier incendie** est fortement corrélé à la variable **Caisse**, mais reste moins significative que celle-ci, donc nous la supprimerons des variables de modélisation. De même nous préférons supprimer les variables **avantage jeune**, **type de résidence** car étant respectivement très corrélées et moins significatives que **formule**, **type produit**.
Nous obtenons ainsi la première liste des variables qui alimenteront nos modèles.

Description Variable
type de produit
Type de résidence
Prime
Qualité juridique
Nombre de pièces
Nombre de logements
Caisse
Zonier VOL
Zonier DDE
Franchise
Remise commerciale
Capital mobilier
Logement
Ancienneté du contrat
Nombre de sinistre réglé par contrat
Nombre d'avenants fait sur le contrat
Nombre de contrat MRH détenu par client
Nombre de contrat IARD
Variation prime début et prime Janvier 2014
Variation prime Janvier 2013 et prime Janvier 14

Tableau 12 : Liste première sélection variables

2.3. Echantillonnage

L'échantillonnage est une phase importante dans la modélisation. Le but étant de construire deux bases d'étude, une base apprentissage sur laquelle se construit le modèle et une base de validation pour mesurer la stabilité et la robustesse du modèle retenu.

Pour construire ces deux bases nous utiliserons la méthode d'échantillonnage aléatoire stratifié.

Le principe de cette méthode est de diviser la population de départ en groupes homogènes (appelés strates), qui sont mutuellement exclusifs, puis on sélectionne à partir de chaque strate des échantillons indépendants.

Sous SAS nous disposons de la **proc surveyselect** pour réaliser cet échantillonnage. Nous faisons le choix de prendre 70 % des observations pour la base d'apprentissage et 30 % pour la base de validation. Un test de cohérence permet de s'assurer que le taux de résiliation reste inchangé sur ces deux bases.

Nous ferons cet échantillonnage pour les deux bases d'études construites précédemment.

Partie 2: Modélisation du taux de résiliation par Régression Logistique

1. Eléments théoriques

Afin de modéliser la variable cible RESIL qui prend les valeurs 0 si un contrat est en cours et 1 s'il est résilié nous utilisons dans cette partie un modèle de régression logistique. En effet l'idée est d'expliquer cette variable avec un nombre **p** de variables explicatives que nous avons choisi dans la partie précédente. Une première réponse à ce type de problématique est le modèle linéaire multiple. Cependant dans ce modèle l'hypothèse forte de la normalité des résidus ne peut s'appliquer pour la modélisation du taux de résiliation. En effet celle-ci suppose une normalité de la variable cible ce qui n'est pas le cas pour la variable **RESIL** qui est binomiale.

Une solution à cette problématique est d'utiliser les modèles linéaires généralisés qui permettent avec une fonction adapté de déterminer une relation entre l'espérance de la variable cible et les variables explicatives.

1.1. Rappels Modèles Linéaires Généralisés

❖ **Présentation**

Les modèles linéaires généralisés sont formés de trois composantes :

- Les variables explicatives

Définissent sous forme d'une combinaison linéaire la composante déterministe $\beta_0 + \sum_{i=1}^{p} \beta_i X_i$

- La fonction de lien

Décrit la relation entre la combinaison linéaire des variables explicatives et l'espérance de la variable cible notée μ. Par exemple la fonction de lien $G(\mu) = \log(\mu)$ permet de modéliser le logarithme de l'espérance.

- La variable de réponse

Précédemment appelée variable cible, elle est la composante aléatoire à laquelle est associée une loi de probabilité qui par ailleurs doit appartenir à la famille exponentielle. A toute loi de probabilité de la composante aléatoire est associée une fonction spécifique de l'espérance appelé fonction de lien canonique que nous noterons G_c. Par exemple pour une distribution normale, la fonction de lien canonique est la fonction identité.

Notons qu'une loi de probabilité appartient à la famille exponentielle si et seulement si sa densité est de la forme:

$$\boldsymbol{f(y_i, \theta_i, \omega_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)}\,\omega_i + c(y_i, \varphi, \omega_i)\right)}$$

- Les fonctions a, b, c sont spécifiées en fonction du type de loi exponentiel.
- Le paramètre canonique θ_i est inconnu. C'est une fonction de l'espérance : $\theta_i = G_c(\mu_i)$
- φ est appelé paramètre de dispersion. Il est supposé connu, sinon il est préalablement estimé et est noté φ_0. Le plus souvent $a(\varphi_0) = \varphi_0$.
- ω_i est un poids.

❖ **Estimation dans le modèle linéaire généralisé**

Les paramètres sont habituellement estimés par la méthode de vraisemblance. Dans la plupart des cas les équations obtenues sont non linéaires et donc les estimateurs sont calculés par des algorithmes itératifs de résolution d'équation non linéaire.
Pour les modèles logistiques binomiales et log linéaire de Poisson, l'algorithme utilisé est **l'algorithme de Newton-Raphson.**
Cet algorithme approxime le log de la fonction de vraisemblance dans un voisinage du paramètre initiale par une fonction polynomiale qui a la forme d'une parabole concave. Cette fonction a la même pente et la même courbure dans les conditions initiales que la fonction log vraisemblance et il est facile de déterminer le maximum de ce polynôme d'approximation. Avec ce maximum, on reprend la procédure décrite précédemment.
Les approximations successives convergent rapidement vers les estimateurs du maximum de vraisemblance.

❖ **Adéquation du modèle**

Nous utiliserons principalement deux statistiques pour juger de l'adéquation du modèle aux données:

➢ **La déviance normalisée**

A partir d'un modèle saturé, nous définissons cette statistique par la formule ci-dessous:

$$D^* = 2\log\frac{L(b_{max}, y)}{L(b\,, y\,)} = 2[l(b_{max}, y) - l(b, y)]$$

Avec L le log vraisemblance et l la vraisemblance.
Lorsque le modèle étudié est exact la déviance normalisée suit approximativement une loi de khi deux à $n - p$ degré de liberté, avec n nombre d'observations de la variable réponse et p le nombre de variables explicatives.

➢ **La statistique du khi-deux de Pearson**

Elle est définie par la formule suivante:

$$\chi^2 = \sum (y_i - \mu_i)^2 / Var(\mu_i)$$

Le khi-deux de Pearson normalisé est égale à : χ^2/φ.Comme pour la déviance cette statistique suit une loi de khi-deux à $n - p$ degré de liberté si le modèle étudié est exact.
Les statistiques de khi-deux de Pearson et la déviance peuvent être utilisés pour estimer le paramètre de dispersion φ leurs statistiques normalisées par leur moyenne égale à $n - p$.

$$D = D^* * \varphi \qquad ; \qquad \hat{\varphi} = \frac{D}{n-p} \qquad ou \quad \hat{\varphi} = \frac{\chi^2}{n-p}$$

1.2. Régression Logistique

La régression logistique est un cas particulier de modèles linéaires généralisés. Elle permet de modéliser les variables explicatives dichotomiques (binomiale) et de manière plus générale multinomiales.

Dans notre étude, la variable que nous cherchons à modéliser est la survenance de la résiliation d'un contrat MRH qui prend les valeurs 1 en cas de résiliation et 0 sinon.

- **Le modèle**

Soit $X = (X_1, X_1 \dots X_p)$ p variables explicatives d'Y. Notons $\pi(x) = E(Y|(X_i = x_1))$.

Nous avons la relation de régression linéaire suivante :

$$G(\pi) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i \quad \text{Avec G fonction de lien et } \beta_i \text{ coefficient réel.}$$

Comme Y suit une loi de Bernoulli, on a $\pi(x) = P(Y = 1|X_i)$, la probabilité de survenance de l'événement «résiliation» sachant que le facteur explicatif X_i.soit égal à x.
Le choix de la fonction de lien s'avère cruciale afin d'avoir des valeurs de $G(\pi)$ non bornées et que π reste dans $[0\ ,1]$ après transformation.
La fonction de lien qui sera utilisé est la fonction LOGIT $G(t) = \ln\left(\frac{t}{1-t}\right)$.

Ainsi $G(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$ $\quad donc \quad$ $\pi = \frac{e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_i}}{1 + e^{\beta_0 + \sum_{i=1}^{n} \beta_i X_i}}$ $\quad \epsilon\ [0\ ,1]$

Il existe d'autres fonctions de lien qui peuvent être utilisées dans la régression logistique que nous regroupons dans le tableau ci-dessous:

Modèle	Fonction de lien	Fonction de Transfert
Probit	$\Phi^{-1}(t)$ inverse fonction de répartition loi normale centrée réduite	$\int_{-\infty}^{t} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}}$
Log log	$\log[-\log(1-t)]$	$1 - e^{-e^t}$

- **Odds Ratio**

L'odds ratio d'une variable explicative mesure l'évolution du rapport des probabilités d'apparition de l'événement $Y = 1$par rapport à l'événement $Y = 0$ lorsque X_i (variable explicative) passe de x à $x + 1$ pour une variable continue et passe d'une modalité à une autre pour une variable qualitative.

Formule générale

$$\boldsymbol{OR = \frac{\pi(x+1)/[1-\pi(x+1)]}{\pi(x)/[1-\pi(x)]} = e^{\beta_i}}$$

Dans le cas particulier où X_i est binaire nous avons:

$$\boldsymbol{P(Y = 1 \,|X_i = 1) = \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}} \qquad \boldsymbol{P(Y = 1 \,|X_i = 0) = \frac{e^{\beta_0}}{1+e^{\beta_0}}}$$

$$OR = \frac{P(Y = 1|X_i = 1)/P(Y = 0|X_i = 1)}{P(Y = 1|X_i = 0)/P(Y = 0|X_i = 0)} = e^{\beta_i}$$

❖ **Estimation des paramètres**

L'estimation des paramètres du modèle est faite par la méthode du maximum de vraisemblance.
Soit l'échantillon $x(i) = (x_1, x_1 \dots x_n)$ de la variable X_i, les observations de Y sont $(y_{1,}, y_2, \dots y_{n,})$ prenant les valeurs 1 en cas résiliation et 0 sinon.
La fonction de vraisemblance s'écrit:

$$L_\beta = P(Y = y_{1,}|X_i = x_1) * P(Y = y_2|X_i = x_{2,}) * \dots * P(Y = y_n|X_i = x_{n,})$$

$$P(Y = y_{i,}|X_i = x_{i,}) = \pi(x_{i,}) \; si \; y_{i,} = 1 \; et \; P(Y = y_{i,}|X_i = x_{i,}) = 1 - \pi(x_{i,}) \; si \; y_{i,} = 0$$

Ainsi on peut avoir l'écriture suivante $P(Y = y_{i,}|X_i = x_i) = \pi(x_{i,})^{y_i}\left(1 - \pi(x_{i,})\right)^{1-y_i}$

$$\text{D' où } L_\beta = \prod_{i=1}^{n} \pi(x_{i,})^{y_i}\left(1 - \pi(x_{i,})\right)^{1-y_i}$$

La fonction log vraisemblance est donnée par

$$\boldsymbol{l_\beta = \sum_{i=1}^{n} y_{i,} \ln \pi(x_{i,}) + (1 - y_i) \ln\left(1 - \pi(x_{i,})\right)}$$

Les paramètres sont obtenus en maximisant cette fonction. Cependant, il y'a pas de solution analytique à ce problème En maximisant cette fonction par des algorithmes numériques de type Newton Raphson, on obtient les paramètres du modèle.

❖ **Tests sur les paramètres.**

Le but est de tester si l'apport de la variable X_i est significatif pour le modèle. Ce qui revient à faire le test d'hypothèse suivant:

$$H_0: \quad \beta_i = 0$$
$$H_1: \quad \beta_i \neq 0$$

➢ **Test de Wald**

Pour ce test on considère la statistique suivante

$\omega = \frac{\hat{\beta}_i}{\hat{\sigma}(\hat{\beta})}$ où $\hat{\sigma}(\hat{\beta})$ est l'estimation de l'écart type de l'estimateur β_i.

Sous H_0, ω^2 suit une loi de khi deux à un degré de liberté χ_1^2. Ainsi H_0 est rejetée si $\omega^2 \geq \chi_{1-\alpha}^2$, α seuil de confiance.

➢ **Test du rapport de vraisemblance**

L'apport de la variable X_i est mesuré par la statistique

$$G = -2 \log\left[\frac{Vraissembalce \; sans \; la \; variable \; X_i}{Vraissembalce \; avec \; la \; variable \; X_i}\right]$$

Sous H_0 , G suit asymptotiquement une loi de khi deux à un degré de liberté χ_1^2, donc on rejette l'hypothèse nulle si $G \geq \chi_{1-\alpha}^2$.

- **Test du score**

On calcule le score avec la formule suivante:

$$SCORE = \left[\frac{\partial l_\beta}{\partial \beta}\right]'_{\widehat{\beta}_{H_0}} J \left[\frac{\partial l_\beta}{\partial \beta}\right]_{\widehat{\beta}_{H_0}}$$

Avec:

l_β la vraisemblance,

$\hat{\beta}_{H_0}$ le vecteur des paramètres estimés sous l'hypothèse H_0,

J Matrice d'information de Fisher

❖ **Méthode de sélection des variables**

Les modèles logistiques seront construits à partir des variables présélectionnées des tableaux Tableau 8 et Tableau 12.

Pour s'assurer de ne retenir que les plus pertinentes nous ferons une seconde sélection basée sur une méthode de sélection automatique pas à pas. En effet il existe trois méthodes différentes que nous présentons brièvement ci-dessous.

- **Méthode ascendante (« FORWARD »)**

Ajoute à chaque étape la variable la plus significative parmi les candidates en partant de la constante.

- **Méthode descendante (« BACKWARD »)**

Cette méthode intègre toutes les variables dans le modèle et retire une à une les moins significatives.

- **Méthode Mixte (« STEPWISE »)**

C'est une combinaison des deux autres méthodes. Chaque étape de sélection «forward» est suivie d'une ou plusieurs étapes «backward» et ce jusqu'à ce qu'il y'ait plus de variable à entrer dans le modèle.

C'est la meilleure méthode car combinant les deux précédentes. C'est elle que nous retiendrons par la suite.

Sous SAS la régression logistique se fait par la **proc logistic** qui permet de mettre en option la méthode de sélection de variable de notre choix.

❖ **Qualité d'ajustement du modèle**

Les statiques suivantes permettent de juger globalement la qualité d'ajustement du modèle**.**

- **R2 Ajusté**

Il est compris en 0 et 1, plus il est proche de 1, meilleur est le modèle.

- **Déviance**

La déviance est mesurée par $-2\log(Vraissemblance\ modèle\ ajusté)$ dans le cas où la variable cible est binaire. Elle égale à la somme des carrés des résidus de déviances individuelles.[2]

- **AIC –BIC**

$\boldsymbol{AIC = -2\ln(L) + 2(p+1)}$ avec p le nombre de variables explicatives du modèle La valeur de l'AIC dépendent du nombre de variables explicatives: un modèle ayant plus de variables explicatives aura un AIC plus grand qu'un modèle avec moins de variables.

$\boldsymbol{BIC = -2\ln(L) + (p+1)\ln n}$ avec n le nombre d'observations. Ne prenant pas en compte uniquement le nombre de variables explicatives, le BIC reste meilleur que le critère AIC.

[2] TUFFERI Stéphane [2010], p 452

En plus de ces statistiques nous disposons d'outils permettant de mesurer le pouvoir prédictif du modèle.

- **Courbe ROC**

La courbe **ROC** (Receiver Operaor Characteristic curve) donne l'évolution du taux d'événement prédit comme tels (sensibilité) en fonction du taux de non-événement prédit comme tels (1-spécificité) lorsqu'on fait bouger le seuil s utilisé pour la prédiction.

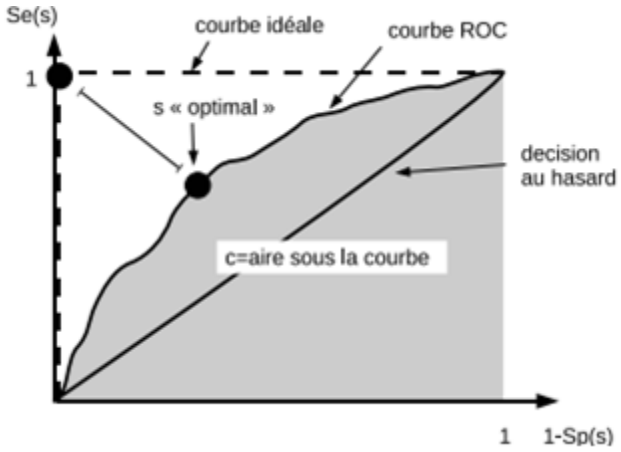


Figure 8 : Courbe ROC

- **Courbe LIFT et Indice de GINI**

La courbe LIFT représente la sensibilité en fonction du pourcentage d'individu ayant un score supérieur au seuil de prédiction.

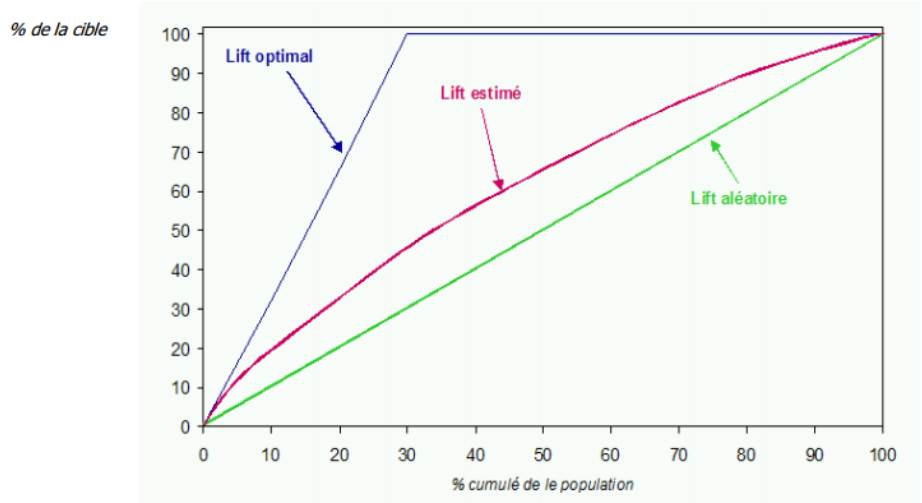


Figure 9 : Courbe LIFT

Le coefficient de **GINI** ou encore D'Sommer est défini comme étant le rapport entre la surface entre la courbe lift réelle et la diagonale et la surface entre le courbe lift idéale et la diagonale. Plus ce coefficient est proche de 1, meilleur est le modèle.

2. Résultats de la modélisation

Nous exposons dans cette section les résultats de la modélisation faite sur le taux de résiliation suivant qu'il n'y'ait pas d'horizon défini et sur un horizon d'un an.

2.1. Modélisation taux de résiliation à horizon non défini

A la suite de l'application de la méthode de sélection « stepwise », nous retirons les variables **Nombre de sinistres corporels**, **nombre de sinistres corporel réglés** et la variable **Chaume.**
Voici la liste des variables qui alimenteront les différents modèles que nous testerons par la suite.

Formule	**Type d'habitation**
Zone Tarifaire	**Nombre sinistres sans suite**
Capital mobilier	**Zonier DDE**
Prime	**Nombre de contrat MRH**
Ancienneté	**Nombre sinistres réglés**
FRANCHISE	**Option Piscine**
Nombre avenant	**Type de résidence**
Caisse	**Option véranda**
Remise commerciale	**Mesure de Surveillance**
Qualité juridique	**Zonier incendie**
Nombre de Logement	**Zonier vol**
Nombre de contrat IARD	**Option RC Equidé**
Nombre de Pièces	**Bois**

Tableau 13 : Variables retenues pour modélisation horizon non défini

En rouge les variables qui ne peuvent pas être mises-en même temps dans un modèle vu les fortes corrélations que nous avons notées précédemment entre elles. En bleu les variables quantitatives (continues) qui ont été discrétisées.
Nous nous proposons de comparer les modèles suivants selon l'utilisation ou non des variables quantitatives non discrétisées et de l'ajout des zoniers à la place de la variable zone tarifaire.

- **Modèle 1:** Variables quantitatives et présence variables **zone tarifaire, Caisse, zonier DDE et zonier Vol**
- **Modèle 2:** Variables quantitatives et présence variable **CAISSE, des zoniers DDE** et **zonier vol,** mais sans la variable **zone tarifaire.**
- **Modèle 3:** Variables quantitatives discrétisées, présence des **zoniers DDE** et **zonier Vol** et suppression de 9 variables explicatives.

Nous comparons ces différents modèles dans le tableau suivant selon différents critères. Notons que tous ces modèles sont stables sur la base de validation donc tous robustes.

Critère modèle	Modèle 1	Modèle 2	Modèle 3
R2 Ajusté	0,544	0,540	0,370
D Sommer	0,745	0,742	0,615
Analyse Type3 Var	Ok tous <0,001	Ok tous <0,001	Ok tous <0,001
% observations concordantes	87,3	87,1	80,7
% observations discordantes	12,7	12,9	19,3
C (courbe ROC)	0,873	0,871	0,807
AIC	1353553	1362535,9	1667295,8
BIC	1355442,2	1363885,4	1668510,4
-2-LOG L	1353245	1362315,9	1667097,8

Tableau 14 : Comparaison des modèles

Nous voyons que la suppression de la variable zone tarifaire et son remplacement par les variables **zonier DDE** et **zonier Vol** diminue légèrement la performance du modèle. Cependant vu le nombre de modalités importantes de cette variable (plus de 20 contre au maximum 5 pour les zoniers) sa suppression simplifie l'utilisation du modèle.

En outre l'utilisation des variables quantitatives ne permet pas de faire une bonne interprétation des résultats. En effet pour ces variables nous obtenons un seul paramètre pour la variable considérée et son utilisation commune dans le calcul de la probabilité de résiliation pour tous clients, diminue le niveau de segmentation des profils qui seront identifiés. Par exemple en prenant la variable Prime, il parait plus intéressant de savoir sur quelle tranche de prime les clients ont plus de probabilité de résilier que de ne pas avoir cette information.

C'est ainsi que nous retiendrons le modèle 3, il est un peu moins robuste que les deux autres mais plus simple d'utilisation au vu des raisons évoquées précédemment et du nombre de variables utilisées.

Dans le modèle 3 les variables qui ont été supprimées sont celles qui sont moins significatives et dont le retrait ne remet pas en cause la validité du modèle. Elles sont mises dans le tableau suivant:

Variables supprimées	V cramer/RESIL
Nombre de sinistres réglés	0,03715
Option piscine	0,02541
Type de résidence	0,02514
Option véranda	0,02256
Mesure surveillance	0,01409
Zonier incendie	0,00833
Zonier vol	0,00762
Option RC équidé	0,00708
Bois	0,00095

Tableau 15 : Variables supprimées

Voici la liste des 16 variables du modèle 3 :

Formule	**Nombre de Logement**
Captal mobilier	**Nombre contrat IARD**
Prime	**Nombre pièces**
Ancienneté	**Qualité juridique**
Franchise	**Type habitation**
Nombre d'avenant	**Nombre sinistres sans suite**
Caisse	**Zonier DDE**
Remise commercial	**Nombre contrats MRH**

Tableau 16 : Variables retenues

Afin d'aboutir à un modèle robuste et simple nous ferons par la suite le test de supprimer certaines variables du modèle 3. En outre sur ce modèle il existe certaines modalités des variables **Caisse**, **nombre de sinistre sans suite** et **nombre d'avenant** qui ne sont pas significatives. Pour corriger cela nous ferons des regroupements de modalités. Nous avons mis en annexe le comparatif des différents modèles d'optimisation faits sur le modèle 3 qui nous conduit au final à retenir un modèle ayant **14 variables explicatives** et dont toutes les modalités des variables sont significatives.

2.1.1. Résultats du modèle retenu

Nous exposons ici les résultats du modèle 3 optimisé.

- **Performance du modèle**

Pour juger de la performance du modèle nous allons analyser plusieurs tests et statistiques.
Nous commençons par le test global de significativité du modèle. Le tableau suivant montre que les tests Wald, score et rapport de vraisemblance sont tous significatifs au seuil de 5 %.

Test de l'hypothèse nulle globale : BETA=0			
Test	**Khi-2**	**DDL**	**Pr > Khi-2**
Rapp. de vrais.	495120,087	57	< ,0001
Score	426232,499	57	<,0001
Wald	317815,089	57	<,0001

Tableau 17 : Significativité globale du modèle

Ensuite le tableau des effets de type 3 de permet de juger de la significativité individuelle de chaque variable. Sur le tableau suivant nous voyons que toutes les variables du modèle sont significatives et que la variable **type d'habitation** est moins influente sur le taux de résiliation que les autres selon le khi2 de Wald.

Analyse des effets Type 3			
Effet	**DDL**	**Khi-2 de Wald**	**Pr > Khi-2**
Nombre contrat MRH	5	9977,4603	<,0001
Nombre contrats IARD	5	11134,4119	<,0001
Nombre logement	7	9619,3761	<,0001
Nombre pièces	11	24951,8740	<,0001
Nombre Avenant	5	25159,3003	<,0001
Prime	2	54587,5170	<,0001
ancienneté	4	111981,538	<,0001
formule	4	25479,1527	<,0001
Qualité juridique	1	4963,0345	<,0001
Type habitation	2	773,9139	<,0001
Zonier DDE	4	6592,4575	<,0001
Pourcentage remise commerciale	2	6309,2673	<,0001
Capital mobilier	2	59682,0366	<,0001
FRANCHISE	3	47103,7895	<,0001

Tableau 18 : Analyse effet de type 3 des variables

Le tableau qui suit montre que notre modèle a un bon taux de bonne prédiction. En effet l'aire sous la courbe ROC traduite par la valeur de l'AUC (c) est de **80,5 %.** De même les taux classements concordant et discordant et la valeur de l'indice de Gini qui est traduite par le D de Sommer sont de bonnes qualités. Pour rappel, plus l'AUC et l'indice de Gini sont proche de 1, meilleur est le modèle.

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	80,5	**D de Somers**	0,611
Pourcentage discordant	19,5	**Gamma**	0,611
Pourcentage lié	0,0	**Tau-a**	0,305
Paires	617688327600	**c**	0,805

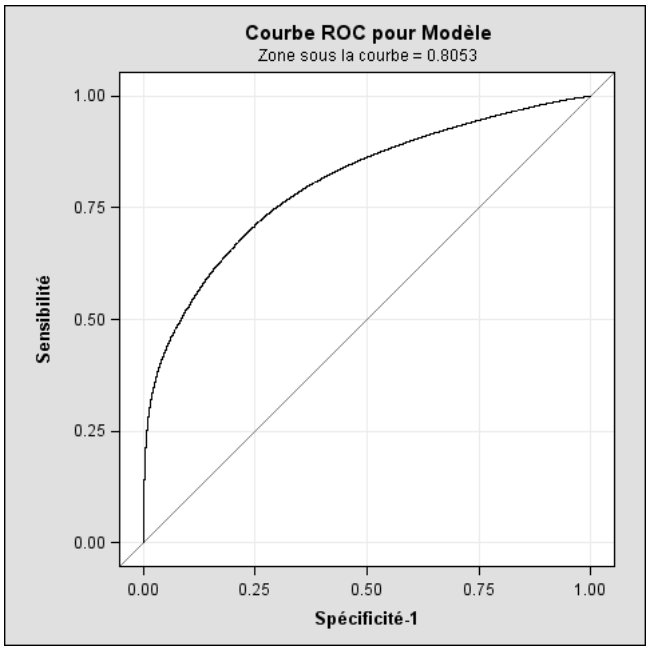


Figure 10 : Courbe ROC

Au vu de ces résultats nous pouvons dire que notre modèle est bien performant. Nous allons mesurer à présent son pouvoir prédictif. Pour ce faire nous construisons la matrice de confusion. Cette matrice représente le pourcentage d'événements prédits ou non selon un seuil de probabilité que nous aurons fixé. Pour faire le choix du seuil nous pouvons utiliser la table de classification qui donne une série de matrice de confusion obtenue pour un ensemble de seuil possible.

Voici un aperçu de cette table:

Table de classification									
	Correct		Incorrect		Pourcentages				
Niveau de proba.	Evénement	Non-événement	Evénement	Non-événement	Correct	Sensi-bilité	Spéci-ficité	Faux POS	Faux NEG
0.000	804E3	0	768E3	0	51.1	100.0	0.0	48.9	.
0.020	804E3	646	767E3	103	51.2	100.0	0.1	48.8	13.8
0.040	803E3	6971	761E3	886	51.5	99.9	0.9	48.7	11.3
0.060	801E3	22618	745E3	2949	52.4	99.6	2.9	48.2	11.5
0.080	798E3	42949	725E3	6672	53.5	99.2	5.6	47.6	13.4
0.100	792E3	68303	7E5	11850	54.7	98.5	8.9	46.9	14.8

Ce qui est important à relever c'est les valeurs de la sensibilité et de la spécificité. Elles traduisent respectivement le pourcentage d'événements prédits comme tels et le pourcentage de non-événements prédits comme tels (événement correspond au fait de résilier ou non).

Nous décidons de choisir le seuil **s** tel que la sensibilité = spécificité, donc ce seuil permet d'avoir un pourcentage de bonne prédiction sur la résiliation égale à celui de la bonne prédiction de non résiliation. Ce choix est difficile à faire sur cette table de classification. C'est pourquoi nous allons tracer les courbes de spécificité et de sensibilité et choisir **s** comme leur point de croisement.

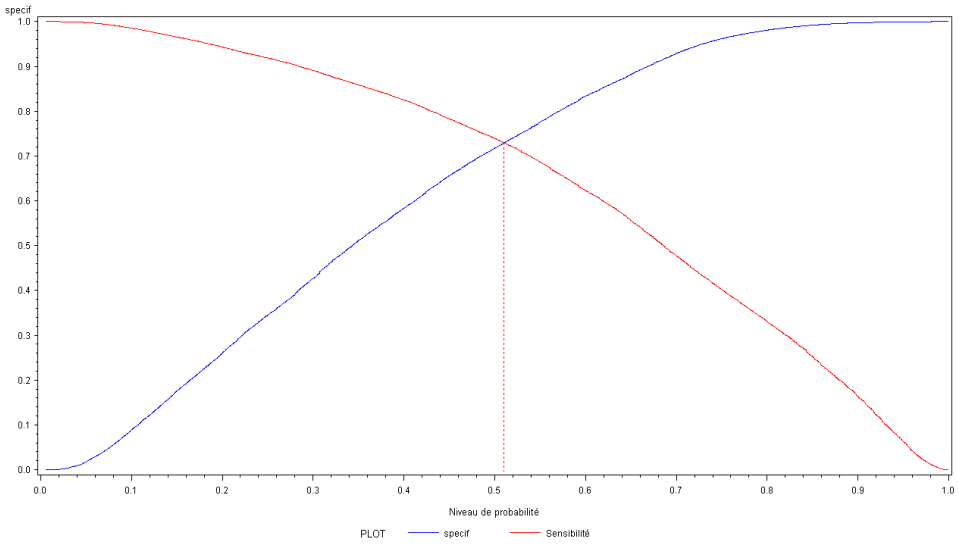


Figure 11 : Sensibilité et spécificité

Nous trouvons une valeur du seuil égale **0,5.** Nous l'utilisons pour avoir la matrice de confusion suivante :

		Prédit		Total
		1	0	
Observé	1	72,99 %	27,01 %	100 %
	0	27,09 %	72,91 %	100 %

Tableau 19 : Matrice de confusion sur la base apprentissage

Cette matrice indique que le modèle prédit bien 73,05 % des contrats résiliés contre 72,77 % des contrats non résiliés au seuil de 0,5. Donc notre modèle prédit bien l'événement de résiliation.

- **Stabilité et robustesse du modèle**

Le but est de voir si notre modèle reste stable sur la base de validation. Pour cela nous allons construire la matrice de confusion sur cette base et la comparer à celle obtenue sur la base d'apprentissage et tracer les courbes LIFT sur ces deux bases.

Nous commençons par analyser la matrice de confusion sur la base de validation.

		Prédit		Total
		1	0	
Observé	1	73,19 %	26,81 %	100 %
	0	27,10 %	72,90 %	100 %

Tableau 20 : Matrice de confusion sur la base de validation

Avec 73,19 % de bonne prédiction pour les contrats résiliés et 72,90 % pour les contrats non résiliés, notre modèle reste stable sur la base de validation avec une légère amélioration que sur la base d'apprentissage, donc il est robuste.

Cette robustesse du modèle se confirme par le graphique suivant qui représente les courbes LIFT sur les deux bases.

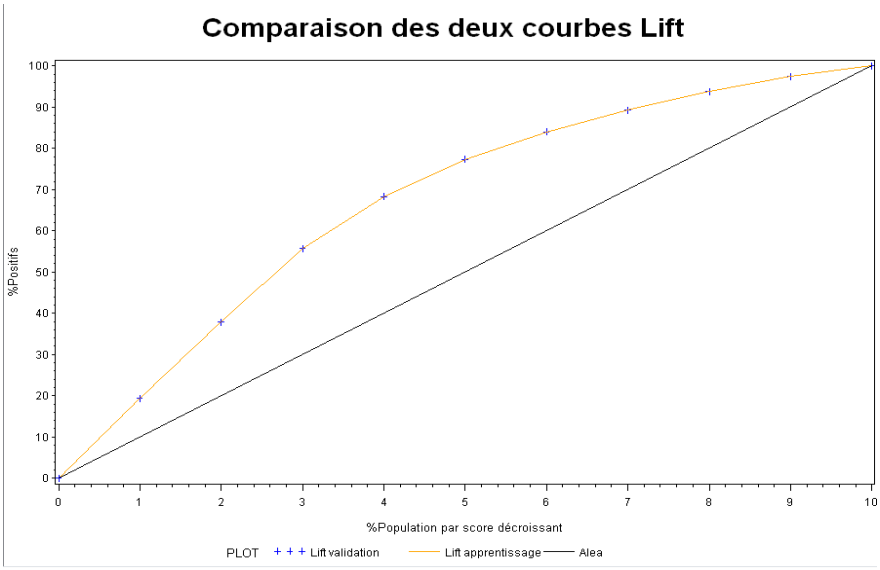


Figure 12 : Comparaison des courbes LIFT apprentissage et validation

Les deux courbes LIFT des deux bases sont presque confondus, donc notre modèle est stable.

- **Paramètres estimés et odds ratios**

Nous voyons sur le tableau suivant que tous les paramètres estimés sont significatif au seuil de 5 %.

Estimations par l'analyse du maximum de vraisemblance						
Variables	**Modalité**	**DDL**	**Valeur estimée**	**Erreur type**	**Khi-2 de Wald**	**Pr> Khi-2**
Constante		1	-2,4567	0,0326	2455,8361	<,0001
Nombre de contrat MRH	**1**					
	2	1	0,3542	0,00473	5599,3359	<,0001
	3	1	0,5308	0,00787	4553,8870	<,0001
	4	1	0,6169	0,0130	2247,0204	<,0001
	5	1	0,5791	0,0205	796,2256	<,0001
	+5	1	0,5919	0,0190	969,2351	<,0001
Nombre de contrat IARD	**0**					
	1	1	-0,4589	0,00503	8331,8657	<,0001
	2	1	-0,5048	0,00803	3946,9170	<,0001
	3	1	-0,4405	0,0151	856,4737	<,0001
	4	1	-0,2753	0,0246	125,1167	<,0001
	+4	1	-0,1130	0,0302	13,9836	0,0002
Nombre de logement	**0**		3,0225	0,0315	9197,5428	<,0001
	1	1				
	2	1	0,2185	0,0369	35,0769	<,0001
	3	1	0,6522	0,0530	151,1526	<,0001
	4	1	1,1007	0,0732	226,2622	<,0001
	5	1	1,1728	0,1199	95,7358	<,0001
	6	1	1,4931	0,1643	82,5562	<,0001

	+6	1	1,7345	0,2010	74,4760	<,0001
Nombre de pièces	1	1	-0,1599	0,00793	406,7898	<,0001
	2	1	-0,2459	0,00589	1743,5961	<,0001
	3					
	4	1	0,5069	0,00604	7034,6045	<,0001
	5	1	0,9318	0,00794	13769,9142	<,0001
	6	1	1,1255	0,0107	10981,0792	<,0001
	7	1	1,4617	0,0162	8099,8933	<,0001
	8	1	1,4247	0,0248	3299,5857	<,0001
	9	1	1,3846	0,0398	1209,5069	<,0001
	10	1	1,4789	0,0536	761,0372	<,0001
	11	1	1,3642	0,0961	201,3690	<,0001
	12	1	1,1299	0,0957	139,3324	<,0001
Nombre Avenants	0					
	1	1	-0,7114	0,00470	22939,7067	<,0001
	2	1	-0,6116	0,00809	5716,0482	<,0001
	3	1	-0,6066	0,0139	1892,0999	<,0001
	4	1	-0,6156	0,0235	688,1580	<,0001
	+4	1	-0,6305	0,0300	440,6728	<,0001
Prime	0-204.9					
	204.9-315.8	1	-0,9311	0,00585	25313,2829	<,0001
	>=315.8	1	-1,9910	0,00854	54395,5243	<,0001
Ancienneté	0-1ans	1	-0,0966	0,00539	321,2732	<,0001
	1-2ans	1	0,7260	0,00602	14560,8950	<,0001
	2-2.4ans	1	0,9249	0,00933	9831,9051	<,0001
	2.4-6ans					
	>=6ans	1	-1,5627	0,00627	62128,7634	<,0001
Formule	F1	1	-0,0219	0,00587	13,8803	0,0002
	F2					
	F3	1	0,9659	0,00676	20403,7050	<,0001
	Jeune	1	-0,4081	0,0104	1548,2662	<,0001
	PB	1	1,8083	0,0314	3314,1081	<,0001
Qualité juridique	Propriétaire					
	locataire	1	0,3444	0,00489	4963,0345	<,0001
Type d'habitation	Maison	1	0,1146	0,00521	484,2838	<,0001
	Appartement					
	Mobil-Home	1	-1,2369	0,0759	265,7362	<,0001
Zonier DDE	1					
	2	1	0,1287	0,00436	870,8258	<,0001
	3	1	0,2329	0,00591	1553,0223	<,0001
	4	1	0,6656	0,00877	5756,7248	<,0001
	5	1	0,4792	0,0138	1199,7253	<,0001
Remise commerciale	0.0000					
	0.0700	1	-0,4028	0,00510	6238,1448	<,0001
	0.2000	1	2,5644	0,3175	65,2528	<,0001
	K1					

Capital mobilier	**K2**	1	-1,4424	0,00594	59041,8126	<,0001
	K3	1	-2,9027	0,0952	928,7886	<,0001
FRANCHISE	**0.00**	1	0,0329	0,00596	30,6098	<,0001
	76.00	1	1,4963	0,00742	40646,2553	<,0001
	130					
	260.00	1	-1,2753	0,0213	3591,4962	<,0001

Tableau 21 : Paramètres estimés modèle final

Voici le tableau des Odds ratios des variables du modèle. Nous allons faire un zoom sur quelques une de ces variables pour l'interprétation. Notons que tous les odds peuvent être interprétés car leurs intervalles de confiance ne contiennent pas la valeur **1**. Les interprétations suivantes supposent que hormis la variable considérée, toutes les autres variables du modèle sont à modalités égales.

Estimations des rapports de cotes			
Effet	**Valeur estimée du point**	**Intervalle de confiance de Wald à 95 %**	
Nombre contrat MRH 2 vs 1	**1,425**	1,412	1,438
Nombre contrat MRH 3 vs 1	**1,700**	1,674	1,727
Nombre contrat MRH 4 vs 1	**1,853**	1,806	1,901
Nombre contrat MRH 5 vs 1	**1,784**	1,714	1,858
Nombre contrat MRH +5 vs 1	**1,807**	1,741	1,876
Nombre contrat IARD 1 vs 0	**0,632**	0,626	0,638
Nombre contrat IARD 2 vs 0	**0,604**	0,594	0,613
Nombre contrat IARD 3 vs 0	**0,644**	0,625	0,663
Nombre contrat IARD 4 vs 0	**0,759**	0,724	0,797
Nombre contrat IARD +4 vs 0	**0,893**	0,842	0,948
Nombre logement 0 vs 1	**20,543**	19,312	21,852
Nombre logement 2 vs 1	**1,244**	1,157	1,338
Nombre logement 3 vs 1	**1,920**	1,730	2,130
Nombre logement 4 vs 1	**3,006**	2,605	3,470
Nombre logement 5 vs 1	**3,231**	2,554	4,086
Nombre logement 6 vs 1	**4,451**	3,225	6,142
Nombre logement +6 vs 1	**5,666**	3,821	8,402
Nombre pièces 1 vs 3	**0,852**	0,839	0,866
Nombre pièces 2 vs 3	**0,782**	0,773	0,791
Nombre pièces 4 vs 3	**1,660**	1,641	1,680
Nombre pièces 5 vs 3	**2,539**	2,500	2,579
Nombre pièces 6 vs 3	**3,082**	3,018	3,147
Nombre pièces 7 vs 3	**4,313**	4,178	4,453
Nombre pièces 8 vs 3	**4,157**	3,959	4,364
Nombre pièces 9 vs 3	**3,993**	3,694	4,317
Nombre pièces 10 vs 3	**4,388**	3,951	4,874
Nombre pièces 11 vs 3	**3,913**	3,241	4,724
Nombre pièces 12 vs 3	**3,095**	2,566	3,734
Nombre Avenant 1 vs 0	**0,491**	0,486	0,495
Nombre Avenant 2 vs 0	**0,542**	0,534	0,551
Nombre Avenant 3 vs 0	**0,545**	0,531	0,560

Nombre Avenant 4 vs 0	**0,540**	0,516	0,566
Nombre Avenant +4 vs 0	**0,532**	0,502	0,565
Prime 204.9-315.8 vs 0-204.9	**0,394**	0,390	0,399
prime >=315.8 vs 0-204.9	**0,137**	0,134	0,139
ancienneté 0-1ans vs 2.4-6ans	**0,908**	0,898	0,918
ancienneté 1-2ans vs 2.4-6ans	**2,067**	2,043	2,091
ancienneté 2-2.4ans vs 2.4-6ans	**2,522**	2,476	2,568
ancienneté >=6ans vs 2.4-6ans	**0,210**	0,207	0,212
formule F1 vs F2	**0,978**	0,967	0,990
formule F3 vs F2	**2,627**	2,593	2,662
formule Jeune vs F2	**0,665**	0,652	0,679
formule PB vs F2	**6,100**	5,736	6,487
Qualité juridique locataire vs Propriétaire	**1,411**	1,398	1,425
Type habitation Maison vs Appart	**1,121**	1,110	1,133
Type habitation Mobil-Home vs Appart	**0,290**	0,250	0,337
Zonier dde 2 vs 1	**1,137**	1,128	1,147
Zonier dde 3 vs 1	**1,262**	1,248	1,277
Zonier dde 4 vs 1	**1,946**	1,912	1,979
Zonier dde 5 vs 1	**1,615**	1,572	1,659
% remise commerciale 0.0700 vs 0.0000	**0,668**	0,662	0,675
% remise commerciale 0.2000 vs 0.0000	**12,992**	6,974	24,205
Capital mobilier K2 vs K1	**0,236**	0,234	0,239
Capital mobilier K3 vs K1	**0,055**	0,046	0,066
FRANCHISE 0.00 vs 130.00	**1,033**	1,022	1,046
FRANCHISE 76.00 vs 130.00	**4,465**	4,401	4,531
FRANCHISE 260.00 vs 130.00	**0,279**	0,268	0,291

Tableau 22 : ODDS ratios modèle horizon non défini

Au vu du Tableau 22 nous voyons que tous les odds de la variable nombre de contrat MRH sont strictement supérieurs à 1, et nous constatons que globalement ils sont croissants suivant les modalités. Ainsi nous pouvons dire que les clients qui ont plus de contrats MRH ont entre **42 %** et **85 %** plus de chance de résilier leur contrat que ceux qui n'en possède qu'un.
Ce résultat permet d'affirmer que la multi détention de contrats MRH augmente le risque de résiliation chez le client. A priori nous nous attendions à un résultat contraire; cependant ce résultat peut s'expliquer par un besoin de réduction de la charge de prime pour le client. En effet prenons l'exemple d'un client possédant en plus de sa résidence principale, 3 résidences secondaires et des logements en PB, ce dernier aurait tendance à regarder plus la concurrence pour alléger au mieux sa charge sur l'ensemble de ses contrats que le client ne possédant qu'une résidence principale à assurer.

Pour la variable nombre de contrat IARD, contrairement à la multi détention mono produit, on observe que plus le client possède de contrat IARD moins il a de chance de résilier par rapport à celui qui n'en possède pas. Donc la multi détention de plusieurs produits réduit bien le risque de résiliation.

La variable nombre de logement n'est renseignée que pour les contrats PB ; pour les autres formules elle est à 0. En prenant comme valeur de référence 1 logement, nous voyons que plus un client PB a

de logements, plus son risque de résiliation est élevé (jusqu'à 5 fois plus).Par contre les contrats hors PB ont 20 fois plus de chance d'être résiliés que ceux en PB.

Quant au nombre de pièces, nous pouvons dire que les clients possédant moins de 3 pièces ont près de **11 % et 22 %** moins de chance de résilier que ceux qui en possèdent 3 alors que ceux qui en possède plus de 3 ont entre **66 %** plus de chance de résilier voir 4 fois plus.

Pour les clients ayant fait des avenants (1 à plus de 4) ont tous près de **46 %** moins de chance de résilier que ceux qui n'ont pas fait d'avenant. Ce résultat ne semble pas intuitif mais reste cohérent par rapport à l'observation faite sur notre base comme le montre le tri à plat mis en **annexe 2** de cette variable.

Pour la prime, nous voyons que les clients qui paient le plus résilient moins que ceux qui paient moins. En effet ceux ayant une prime comprise entre 204,9 € et 315 € et ceux dont la prime est supérieur à 315 € ont respectivement 61 % et 87 % moins de chance de résilier que les clients qui paient une prime inférieur à 204 €.

Les clients qui ont une ancienneté entre 1 et 2,4 ans ont une probabilité de résiliation de **+100 %** plus importante que ceux qui ont entre 2,4 et 6 ans d'ancienneté, alors que ceux qui ont moins d'un an ou plus de 6 ans d'ancienneté ont respectivement **10 %** et**79 %** moins de chance de résilier que la tranche 2,4 - 6 ans.

Enfin les clients ayant souscrit à une formule F1 ou jeune ont entre **67 % et 95 %** moins de chance de résilier que ceux de la formule F2. Tans dis que les clients possédant une formule PB ou F3 ont près de 2 à 3 fois plus de chance de résilier que ceux de la formule F2.

2.2. Modélisation taux de résiliation à un an

Nous procédons maintenant à la modélisation du taux de résiliation à un an. Comme pour le taux de résiliation sans horizon défini, nous utilisons un modèle de régression logistique avec le même principe d'échantillonnage que pour la première étude.

La méthode de sélection « stepwise », nous permet de faire une deuxième sélection des variables de modélisation. Les variables retenues sont listée dans le tableau ci-dessous :

Description Variable
type de contrat
Formule
Caisse
Prime
Qualité juridique
Type habitation
Ancienneté du contrat
Nombre de sinistre réglé par contrat
Nombre d'avenants fait sur le contrat
Nombre de contrat MRH détenu par client
Nombre de contrat IARD
Variation prime début et prime Janvier 14
Variation prime janvier13 et prime janvier 14

Tableau 23 : Variables sélectionnées pour la modélisation

Nous élaborons spécialement trois modèles suivant la suppression de trois variables au maximum.

- Modèle 1: toutes les variables retenues
- Modèle 2: toutes les variables sauf **CAISSE**, **type d'habitation**, **Nombre avenant**
- Modèle 3: toutes les variables sauf **CAISSE** et **Type d'habitation**

Le choix de la suppression de ces différentes variables dans les modèles s'explique par la faible corrélation à la variable cible. En plus de cela nous observons pour la variable CAISSE un nombre important de modalités, pour la variable type habitation une non significativité de la modalité mobil-home et pour le Nombre d'avenant une non significativité de certaines de ses modalités également.

Critère modèle	Modèle 1	Modèle 2	Modèle 3
R2 Ajusté	0,4905	0,4846	0, 4903
D Sommer	0,76	0,744	0,742
Analyse Type3 Var	Ok tous <0,001	Ok tous <0,001	Ok tous <0,001
% observations concordantes	88	87,9	88
% observations discordantes	12	12,1	12
AUC (courbe ROC)	0,88	0,87	0,88
AIC	323700,87	326623,49	323826,27
BIC	324533,74	327034,21	324579,27
-2-LOG L	323554,87	326551,49	323694,27

Tableau 24 : Comparaison des modèles

Nous retiendrons le modèle 2, malgré que le BIC soit supérieur à celui du modèle 3 et que le R2 ajusté soit relativement plus petit. Cependant le pourcentage d'observation concordante et discordante et le coefficient de l'AUC reste meilleur que celui du modèle 3 et il reste simple d'utilisation et d'interprétation au vu des variables retenues.

2.2.1. Résultats du modèle retenu

Nous exposons dans cette partie les résultats du modèle 2 décrit plus haut.

- **Performance du modèle**

Pour juger de la performance du modèle nous analysons d'abord le test global de significativité du modèle.

Test de l'hypothèse nulle globale : BETA=0			
Test	**Khi-2**	**DDL**	**Pr > Khi-2**
Rapp. de vrais.	206920,868	35	<,0001
Score	266791,707	35	<,0001
Wald	146418,385	35	<,0001

Tableau 25 : Significativité globale du modèle

Au vu du Tableau 25, les tests de score, Wald et du rapport de vraisemblance sont significatifs au seuil de 5 %.

Dans le tableau suivant nous voyons que toutes les variables du modèle sont significatives et contribue bien à la performance. La variable **nombre de sinistres réglés** est la moins influente du modèle au sens du Khi2 de Wald.

Analyse des effets Type 3			
Effet	**DDL**	**Khi-2 de Wald**	**Pr > Khi-2**
Ancienneté	4	7986,6841	<,0001
prime	2	138,4903	<,0001
indiceprix2	2	110794,459	<,0001
indiceprix1	4	700,3073	<,0001
Nombre contrat MRH	5	1562,9308	<,0001
Nombre contrat IARD	5	1646,2408	<,0001
Nombre sinistre réglé	3	134,3674	<,0001
Type de Produit	5	13664,3322	<,0001
Formule	4	421,0343	<,0001
Qualité juridique	1	1141,7337	<,0001

Tableau 26 : Analyse effet de type 3 des variables

Le tableau suivant nous montre que le modèle a un taux élévé de bonne prédiction. L'aire sous la courbe ROC est de 87,9 % et l'indice de GINI est proche de 1. Les pourcentages d'observations concordantes et discordantes sont de bonnes qualités.

Association des probabilités prédites et des réponses observées			
Pourcentage concordant	87,9	**D de Somers**	**0,758**
Pourcentage discordant	12,1	**Gamma**	0,759
Pourcentage lié	0,0	**Tau-a**	0,180
Paires	52623623910	**c**	**0,879**

Pour mesurer le pouvoir prédictif du modèle, nous traçons la courbe sensibilité et spécificité afin de trouver le seuil s pour la construction de la matrice de confusion.
Nous décidons de choisir le seuil **s** tel que la sensibilité = spécificité, ce qui se fera en traçant les courbes sensibilité et spécificité et en identifiant le point de croisement.

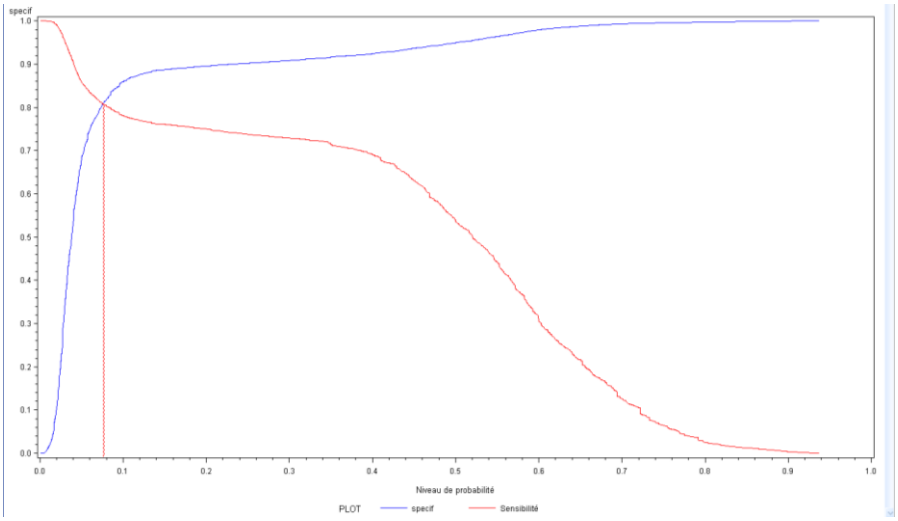


Figure 13 : Sensibilité et Spécificité Modèle résiliation à un an

Le seuil trouvé est de 0,08, comme le montre la figure 13. La matrice de confusion ci-dessous est construite à ce seuil. Nous obtenons un taux de bonne prédiction de 80,88 % pour les contrats résiliés et 81,51 % de contrats non résiliés. Donc le modèle est bien performant.

		Prédit		Total
		1	0	
Observé	1	80,88 %	19,12 %	100 %
	0	18,43 %	81,51 %	100 %

Tableau 27 : Matrice de confusion base apprentissage

- **Stabilité du modèle**

Pour vérifier la stabilité du modèle sur la base de validation, nous construisons la matrice de confusion et comparons les courbes lifts sur les deux bases (apprentissage et validation).

		Prédit		Total
		1	0	
Observé	1	80,80 %	19,20 %	100 %
	0	18,48 %	81,52 %	100 %

Tableau 28 : Matrice de confusions base validation

Nous avons 80,80 % de bonne prédiction pour les contrats résiliés et 81,52 % pour les contrats non résiliés. C'est presque les mêmes taux obtenus sur la base d'apprentissage. Cela prouve que le modèle est stable et cette stabilité est confirmée par les courbes lifts représentés ci-dessous.

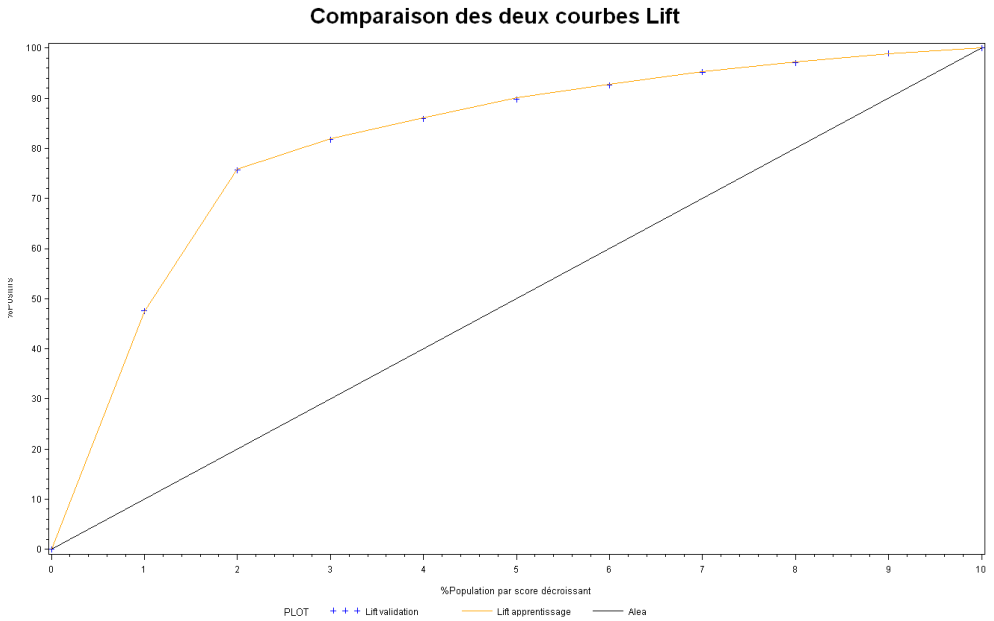


Figure 14 : Comparaison courbe LIFT apprentissage et validation modèle résiliation un an

Nous voyons que les courbes LIFT sur les bases d'apprentissage et de validation sont confondues. Le modèle retenue est bien stable et robuste.

- **Paramètres estimés et odds ratios**

Voici le tableau des paramètres estimés de toutes les modalités de chaque variable.

Estimations par l'analyse du maximum de vraisemblance						
Variable	**Modalité**	**DDL**	**Valeur estimée**	**Erreur type**	**Khi-2 de Wald**	**Pr > Khi-2**
Constante		1	**-3,5651**	0,0200	31653,1885	<,0001
Ancienneté	**0-1,4ans**	1	-1,9905	0,0255	6070,5563	<,0001
	1,4-2ans	1	-1,9582	0,0247	6303,7759	<,0001
	2-3ans	1	-0,7491	0,0195	1482,3042	<,0001
	3-6ans	1	-0,3483	0,0155	506,5742	<,0001
	>=6ans					
prime	**204-269**	1	0,0559	0,0137	16,5674	<,0001
	>=269	1	0,1679	0,0145	134,0213	<,0001
	0-204					
Taux variation prime par rapport prime 13	**Pas de variation**	1	3,9407	0,0119	110229,393	<,0001
	[-0,1 -0]	1	1,5011	0,0196	5866,1221	
	[0,1-0,4]					
Taux variation prime par rapport prime départ	**Pas de variation**	1	-0,1378	0,0211	42,7443	<,0001
	[-0,18 -0[	1	-0,4273	0,0289	218,2624	<,0001
	]0 -0,1]	1	0,1812	0,0142	162,3809	<,0001
	] 0,7- 2,9]	1	0,1554	0,0180	74,8710	
	] 0,1- 0,7]					

Nombre de contrat MRH	2	1	0,3717	0,0111	1116,9890	<,0001
	3	1	0,4382	0,0182	580,7379	<,0001
	4	1	0,4479	0,0300	222,9470	<,0001
	5	1	0,3876	0,0484	64,1520	<,0001
	+5	1	0,4925	0,0422	135,9299	<,0001
	1					
Nombre de contrat IARD	1	1	-0,3386	0,0126	726,0400	<,0001
	2	1	-0,5279	0,0207	649,3281	<,0001
	3	1	-0,7479	0,0429	304,2249	<,0001
	4	1	-0,7870	0,0722	118,9006	<,0001
	+4	1	-1,1497	0,0943	148,7852	<,0001
	0					
Nombre de sinistre réglé	1	1	0,1440	0,0140	105,3750	<,0001
	2	1	0,0724	0,0246	8,6351	0,0033
	+2	1	0,2116	0,0325	42,4854	<,0001
	0					
Type de Produit	MRH1	1	-0,5018	0,0199	634,8351	<,0001
	MRH3	1	2,1453	0,0191	12581,2439	<,0001
	PBMRH1	1	1,8911	0,2389	62,6686	<,0001
	PBMRH2	1	2,5170	0,2371	112,6555	<,0001
	PBMRH3	1	4,1071	0,2385	296,6531	<,0001
	MRH2					
Formule	F1	1	0,1482	0,0125	140,1325	<,0001
	F3	1	-0,1249	0,0172	53,0048	<,0001
	Jeune	1	0,2685	0,0216	154,0099	<,0001
	PB	1	-2,4605	0,2366	108,1768	<,0001
	F2					
Qualité juridique	locataire	1	0,3911	0,0116	1141,7337	<,0001
	Propriétaire					

Tableau 29 : **Paramètres estimés du modèle final**

Toutes les estimations des paramètres sont significatives au seuil de 5 %.

Afin de pouvoir analyser l'influence de chaque variable par modalités sur le risque de résiliation à un an nous analysons le tableau des odds ratios. Notons qu'aucun intervalle de confiance des valeurs des odds ne contient la valeur 1, donc ces derniers sont tous interprétables.

Estimations des rapports de cotes			
Effet	Valeur estimée du point	Intervalle de confiance de Wald à 95 %	
ancienneté 0-1,4ans vs >=6ans	0,137	0,130	0,144
ancienneté 1,4-2ans vs >=6ans	0,141	0,134	0,148
ancienneté 2-3ans vs >=6ans	0,473	0,455	0,491
ancienneté 3-6ans vs >=6ans	0,706	0,685	0,728
prime 204-269 vs 0-204	1,057	1,029	1,086
prime >=269 vs 0-204	1,183	1,150	1,217
indiceprix2 Pas de variation vs]0-0,4]	51,454	50,271	52,665
indiceprix2 [-0,1 -0] vs]0-0,4]	4,486	4,317	4,662
indiceprix1 Pas de variation vs]0,1- 0,7]	0,871	0,836	0,908

indiceprix1 [-0,18 -0[vs]0,1- 0,7]	0,652	0,616	0,690
indiceprix1]0 -0,1] vs]0,1- 0,7]	1,199	1,166	1,233
indiceprix1]0,7- 2,9] vs]0,1- 0,7]	1,168	1,128	1,210
Nombre contrat MRH 2 vs 1	1,450	1,419	1,482
Nombre contrat MRH 3 vs 1	1,550	1,496	1,606
Nombre contrat MRH 4 vs 1	1,565	1,476	1,660
Nombre contrat MRH 5 vs 1	1,473	1,340	1,620
Nombre contrat MRH +5 vs 1	1,636	1,506	1,778
Nombre contrat IARD 1 vs 0	0,713	0,695	0,731
Nombre contrat IARD 2 vs 0	0,590	0,566	0,614
Nombre contrat IARD 3 vs 0	0,473	0,435	0,515
Nombre contrat IARD 4 vs 0	0,455	0,395	0,524
Nombre contrat IARD +4 vs 0	0,317	0,263	0,381
Nombre de sinistre réglés 1 vs 0	1,155	1,124	1,187
Nombre de sinistre réglés 2 vs 0	1,075	1,024	1,128
Nombre de sinistre réglés +2 vs 0	1,236	1,159	1,317
Type Produit MRH1 vs MRH2	0,605	0,582	0,630
Type Produit MRH3 vs MRH2	8,544	8,230	8,871
Type Produit PBMRH1 vs MRH2	6,626	4,149	10,583
Type Produit PBMRH2 vs MRH2	12,392	7,785	19,724
Type Produit PBMRH3 vs MRH2	60,770	38,082	96,977
formule F1 vs F2	1,160	1,132	1,189
formule F3 vs F2	0,883	0,853	0,913
formule Jeune vs F2	1,308	1,254	1,365
Qualité juridique locataire vs Propriétaire	1,479	1,445	1,513

Tableau 30 : Odds variables du modèle résiliation à un an

Les interprétations suivantes supposent que hormis la variable considérée toutes les autres variables du modèle sont à modalités constantes.

Nous retiendrons que pour la résiliation à un an, les contrats d'ancienneté inférieure à 6 ans ont tous moins de chance d'être résiliés que ceux d'ancienneté de 6 ans ou plus.

Nous observons que la probabilité de résiliation à un an est croissante par rapport à la prime. En effet les contrats dont la prime est comprise entre 204 € et 269 € ou supérieur à 269 € ont respectivement 5 % et 18 % plus de chance d'être résiliés que ceux avec une prime inférieure à 204 €.

Les clients n'ayant pas eu de variation de prime par rapport à la prime de 2013 ont 50 fois plus de chance de résilier que ceux qui ont eu une augmentation inférieure à 40 %. De même les clients ayant eu une diminution de prime inférieur à 18 % en une année ont 4 fois plus de chance de résilier que ceux qui ont eu une augmentation inférieure à 40 %.

Les contrats ayant bénéficiés d'une réduction de prime maximale de 10 % par rapport à la prime de départ et ceux qui n'ont pas eu de variation de prime ont respectivement 38 % et 13 % moins de chance d'être résilié que ceux bénéficiant d'une augmentation comprise entre 10 % et 70 %. Par contre les clients ayant eu une augmentation de prime inférieure à 10 % ou comprise entre 70 % et

presque 300 % par rapport à la prime de départ, ont respectivement 19 % et 16 % plus de chance de résilier que ceux dont l'augmentation est comprise entre 10 % et 70 %.

Nous remarquons aussi que la possession de plusieurs contrats MRH augmentait la probabilité de résiliation chez les clients.

En revanche la multi détention de produit IARD chez le client diminuait la probabilité de résiliation en comparaison au client muni d'un seul contrat MRH.

Le résultat sur la variable **nombre de sinistres réglés** nous semble étonnant. En effet nous voyons que les clients ayant eu des sinistres réglés ont plus de chance de résilier leur contrat que ceux n'ayant aucun sinistre réglés. Pour une analyse plus exacte il serait intéressant de créer une variable traduisant la proportion de sinistre déclarés et de sinistres réglés, cela permettrait d'avoir une meilleure interprétation des résultats.

Les contrats de l'offre MRH1 hors PB ont 40 % moins de chance de résilier que ceux de la MRH2, alors que ceux de MRH3 ont 8 fois plus de chance de résilier.
Les contrats PB de toutes les offres ont entre 6 à 60 fois plus de chance d'être résilié que les contrats MRH2 hors PB.

Enfin nous voyons que les locataires ont 47 % plus de chance de résilier que les propriétaires.

Partie 3: Modélisation durée de vie des contrats MRH

Dans cette partie nous proposons de faire une nouvelle approche de modélisation du taux de résiliation en utilisant la durée de vie des contrats MRH. L'idée étant d'estimer la probabilité de survie d'un contrat en portefeuille à chaque date donnée suivant certaines variables tarifaires. Pour ce faire nous allons utiliser les modèles de survie.

1. Les modèles de survie

La durée de survie désigne le temps qui s'écoule depuis un instant initial jusqu'à la survenue d'un événement précis (par exemple décès, guérison). Dans notre étude l'événement considéré est la résiliation d'un contrat MRH.

Les données de durée présentent les particularités suivantes :

- Les données ne peuvent être observés que sur un sous ensemble de $[0, +\infty[$, on dit qu'il y'a troncature
- Pour certains individus, il peut arriver que l'événement ne se produise pas pendant la durée d'observation, dans ce cas on parle de données censurées.

1.1. Distribution de survie

Nous considérons la variable aléatoire de durée de vie T prenant ses valeurs dans $[0, +\infty[$ On note par F sa fonction de répartition et f sa densité.

- **Fonction de survie**

La fonction de survie est le complément à 1 de la fonction de répartition.

$$\boldsymbol{S(t) = 1 - F(t) = P(T > t)}$$

C'est une fonction décroissante tel que $\boldsymbol{S(0) = 1}$ et $\boldsymbol{\lim_{t\to\infty} S(t) = 0}$

- **Survie conditionnelle**

On s'intéresse à la durée de survie après un instant t d'un individu sachant qu'il est en vie jusqu'en t. Cette fonction est donnée par la formule suivante :

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

- **Fonction de hasard**

La fonction de hasard ou fonction de risque (ou taux de défaillance ou taux de panne) se défini ainsi :

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

Elle fournit la probabilité que la durée de vie soit comprise entre t et $t + dt$ sachant qu'elle est plus grande que *t.* La fonction de hasard représente le taux instantané de sortie de l'état que l'on observe. Si l'on s'intéresse à la durée de vie des contrats MRH, il représente le taux instantané de résiliation suivant une ancienneté donnée

Ces différentes fonctions caractérisent à chacune la distribution de la variable T.

1.2. Censure/Troncature

❖ **Censure**

La censure est le phénomène le plus couramment rencontré lors de la collecte des données de survie. Cette censure peut être à droite si pour certain individu la date de survenance de l'événement n'est pas connue sur la période d'observation ou à gauche si la date de survenance de l'événement est déjà connue pour certain individus.

En assurance la censure à droite est la situation la plus rencontrée (étude durée de vie humaine, durée de vie contrats…).

❖ **Troncature**

Il y'a troncature gauche (respectivement. droite) lorsque la variable d'intérêt n'est pas observable lorsqu'elle est inférieure à un seuil $c > 0$ (respectivement. supérieure à un seuil > 0).

Les troncatures diffèrent ainsi des censures car elles concernent l'échantillonnage lui-même. En cas de troncature on perd complètement l'information sur les observations en dehors d'une certaine plage alors que pour la censure les durées de vie ne sont pas toutes observées ; pour certaines d'entre elles, on sait seulement qu'elles sont supérieures à une certaine valeur connue.

❖ **Description base d'étude**

Pour la modélisation de la durée de vie des contrats MRH nous allons utiliser les bases de modélisation construites pour la régression logistique. Ces bases contiennent des contrats de 1999 jusqu'à janvier 2014, avec une variable **ancienneté** calculée suivant que le contrat soit résilié ou en cours. Ainsi nous sommes en présence d'une censure à droite fixe, c'est-à-dire que pour les contrats en cours en janvier 2014, leur date de résiliation n'est pas connue car la base est arrêtée à ce mois, donc l'ancienneté maximale ou censure est de 15 ans.

Pour la suite nous noterons «base 1» la base d'étude pour la résiliation à horizon non défini et «base 2» celle de la résiliation à un an.

2. Modélisation

Pour modéliser la durée de vie des contrats nous utilisons principalement deux méthodes: une estimation non paramétrique et une estimation semi-paramétrique. Nous faisons le choix de ne pas utiliser des modèles paramétriques pour éviter la problématique du choix de la distribution de survie qui dans notre cas n'est pas connue à priori.

2.1. Modèle non paramétrique

Les modèles non paramétriques permettent d'estimer l'une des différentes fonctions caractérisant la distribution de la variable de durée T sans faire aucune hypothèse a priori sur celle-ci.

❖ **Estimation de Kaplan-Meier**

L'estimateur de Kaplan-Meier découle de l'idée suivante : survivre après un temps $t > s$ c'est être en vie juste avant t et ne pas mourir au temps t :

$$S(t) = P(T > t \,|T > s)P(T > s) = P(T > t \,|T > s)S(s)$$

En renouvelant l'opération on fait apparaitre des produit de terme en $P(T > t \,|T > s)$.

En considérant des temps d'évènement (décès ou censure) distinctes $T_{(i)}$ $(i = 1, .., n)$ rangés par ordre croissant, on obtient :

$$P\big(T > T_{(j)}\big) = \prod_{k=1}^{j} P(T > T_{(k)} \ \big| T > T_{(k-1)})$$

Soit r_i le nombre d'individu à risque de subir l'évènement avant le temps $T_{(i)}$ et d_i le nombre de décès en $T_{(i)}$ .Soit p_i la probabilité de survivre sur l'intervalle $\left]\, T_{(i-1)}, T_{(i)}\,\right]$, $q_i = 1 - p_i$ peut être estimé par $\hat{q}_i = \frac{d_i}{r_i} = \frac{d_i}{n-i+1}$.

Nous notons D_i la fonction définie par :

$D_i = \begin{cases} 1 \ si \ X_i < c \\ 0 \quad sinon \end{cases}$ avec C une censure fixe et X_i la durée de survie observée.

En $T_{(i)}$ si $D_i = 1$ alors 'il y'a décès donc $d_i = 1$ et dans le cas contraire l'observation est censurée et $d_i = 0$.L'estimateur de Kaplan-Meier s'écrit donc

$$\boldsymbol{\widehat{S}(t) = \prod_{T_{(i)} \leq t} \left(1 - \frac{D_i}{r_i}\right) = \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n - i + 1}\right)^{D_i}}$$

En présence d'ex aequo et en supposant par convention que les observations non censurées précèdent toujours les observations censurées, nous obtenons l'expression suivante :

$$\boldsymbol{\widehat{S}(t) = \prod_{T_{(i)} \leq t} \left(1 - \frac{d_i}{r_i}\right)}$$

- **Propriétés de l'estimateur**

Sans présence de censure l'estimateur de Kaplan-Meier correspond à l'estimateur empirique de la fonction de répartition. En outre c'est un estimateur qui possède de bonnes propriétés: il est convergent, asymptotiquement gaussien, cohérent et est également un estimateur du maximum de vraisemblance généralisé. Toutefois, cet estimateur est biaisé positivement. (cf. F.PLANCHET [2014][3]).

- **Estimation fonction de survie de kaplan-Meier**

Dans ce qui suit nous utiliserons l'estimateur de Kaplan-Meier sur des intervalles de temps d'un an. En effet actuellement la résiliation de contrat MRH ne peut se faire qu'à l'échéance (annuelle) du contrat. La base de données utilisée dans cette partie est celle de la résiliation à horizon non défini.

L'estimation de la fonction de survie peut être faite suivant une variable déterminée ou pour l'ensemble du portefeuille. Sous SAS cette estimation est produite par la procédure **LIFETEST**.

Nous faisons en premier l'estimation de la fonction de survie sur l'ensemble du portefeuille, c'est-à-dire sur toute notre base d'étude. Nous obtenons le graphique de la fonction de survie estimée ci-dessous :

[3] **Frédéric PLANCHET [2014], «** Modèle de durée, application actuarielle », 4-L'estimation non paramétrique en présence de données censurées : estimateur de Kaplan-Meier et estimateur de Nelson-Aalen

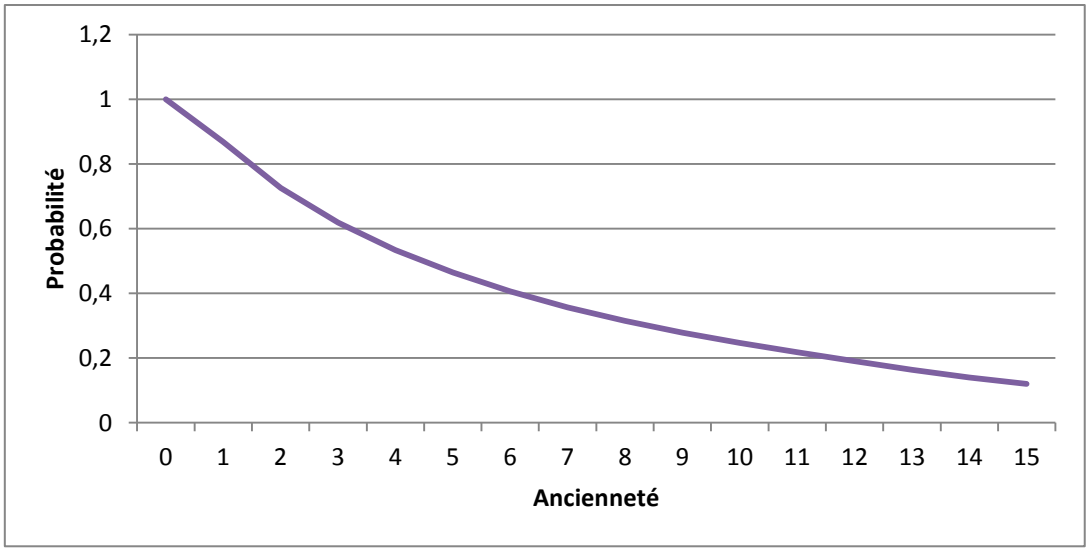


Figure 15 : Survie Portefeuille Kaplan Meier

La fonction de survie du portefeuille permet d'avoir une idée de la durée de survie des contrats en portefeuille. Nous voyons sur ce graphe que la probabilité qu'un contrat reste en portefeuille diminue de manière presque linéaire suivant l'ancienneté.

Ainsi la probabilité qu'un contrat MRH reste encore en portefeuille après 15 ans d'ancienneté est presque de 0,1.

Avec l'estimation non paramétrique nous pouvons aussi déterminer la survie suivant les caractéristiques des contrats.

Nous voyons sur la Figure 16 que les contrats en formule PB ont une probabilité de survie plus importante que le reste des formules.

La fonction de survie de la formule Jeune décroit très rapidement par rapport aux autres formules : après 4 ans d'ancienneté la probabilité de survie des Jeunes est de 0,2 alors que toutes les autres formules sont à plus de 0,4.

Il faut noter que c'est une formule valable uniquement pour les clients de moins de 30 ans; nous observons qu'après 6 ans d'ancienneté la probabilité de survie est presque nulle, ce qui peut s'expliquer par cette contrainte sur l'âge du client et une forte volatilité de ces profils.

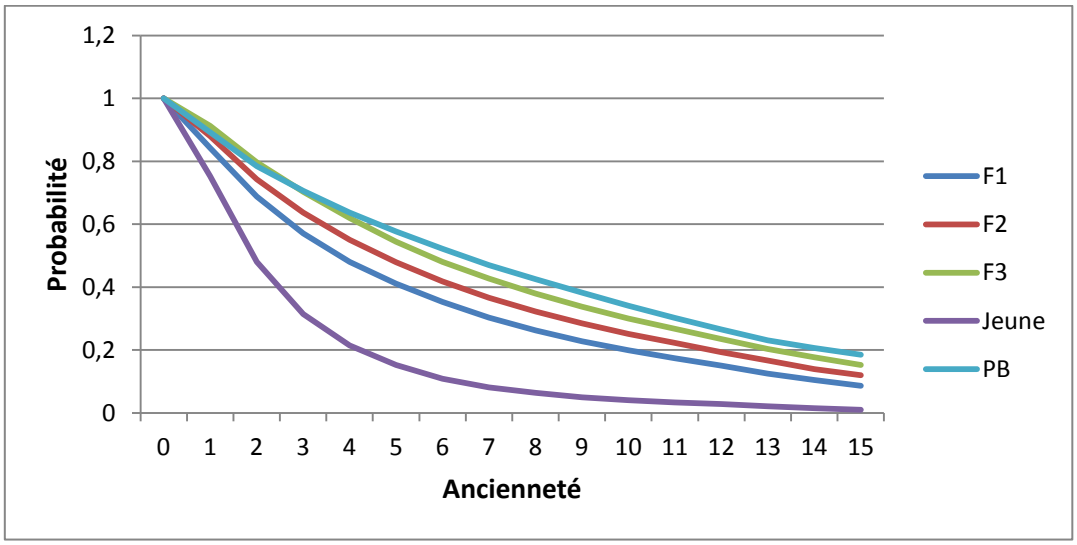


Figure 16 : Fonctions de survie des modalités de la variable Formule

❖ **Durée de vie résiduelle par formule**

Pour calculer la durée de vie résiduelle des contrats suivant la formule, nous utilisons la formule suivante :

$$E(T_k) = \sum_{i=0}^{\infty} \frac{S(k+i)}{S(k)}$$

Avec

- T_k la variable aléatoire correspondant à la durée de vie résiduelle pour une ancienneté k.
- S fonction de survie de la variable de durée T.

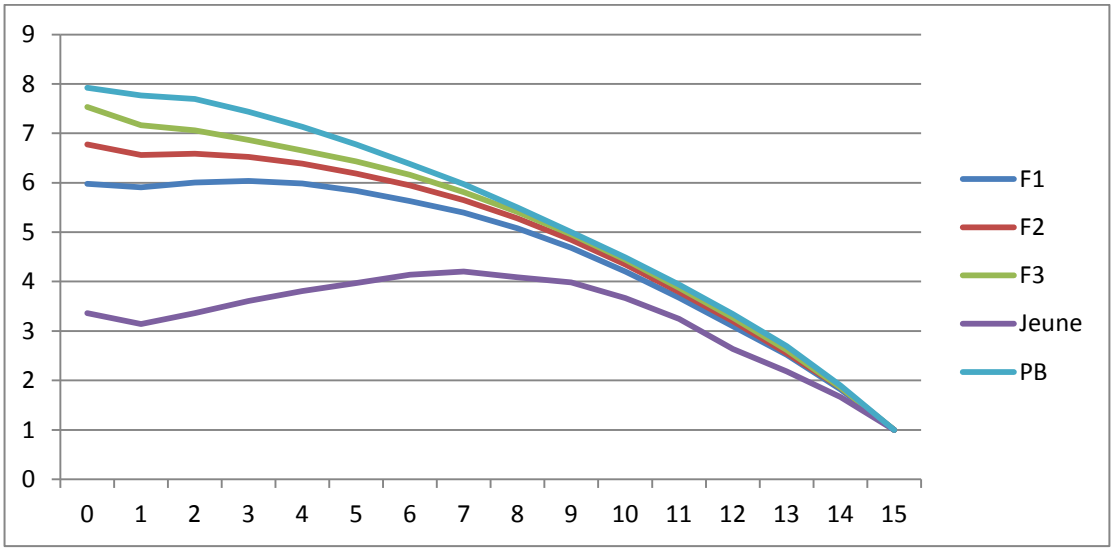


Figure 17 : Espérance de vie résiduelle Formule

Nous retrouvons comme pour les fonctions de survie, une espérance de vie à la souscription de 8 ans pour les PB. Elle reste supérieure à celles des formules F1, F2, F3 jusqu'à 7 ans d'ancienneté, et ensuite les espérances de vie de ces formules sont presque identiques et décroissent progressivement jusqu'à 15 ans d'ancienneté.

La formule Jeune par contre a une espérance de vie assez atypique, en effet à la souscription elle est de 3 ans et augmente progressivement entre les anciennetés 1 et 7 ans pour atteindre 4 ans, et entre 7 ans et 15 ans d'ancienneté l'espérance de vie diminue progressivement.
Les contrats en formule Jeune ont une espérance de vie nettement inférieure à celles des autres formules quel que soit l'ancienneté du contrat. En outre c'est après 7 ans d'ancienneté que les contrats en formule Jeune atteignent l'espérance de vie maximale contrairement aux autres formules pour qui elle est maximale à la souscription.

2.2. Modèle semi paramétrique

Les modèles semi-paramétriques cherchent à estimer la fonction de hasard de la variable durée de vie T en tenant compte de l'influence des facteurs exogènes.

Ce sont des modèles dit à hasards proportionnels. Pour ces modèles nous retrouvons la notion de hasard de base qui donne la forme générale du hasard et qui est commune à tous les individus.

La fonction de hasard de ces modèles se caractérise par la relation suivante :

$$\forall\, \boldsymbol{t} > 0 \qquad h(\boldsymbol{t}|\boldsymbol{Z}.\boldsymbol{\theta}) = \boldsymbol{h_0}(\mathbf{t})\, \mathbf{f}(\boldsymbol{Z};\boldsymbol{\theta}) \text{ avec}$$

- h_0 fonction de hasard de base
- f fonction positive
- Z vecteur des covariables
- θ vecteur des paramètres

2.2.1. Le modèle de Cox

Parmi ces modèles à hasards proportionnels le plus utilisé est celui de COX. Dans ce modèle aucune hypothèse n'est faite sur le hasard de base. La fonction de hasard est de la forme :

$$\boldsymbol{h}(\boldsymbol{t}|\boldsymbol{Z}.\boldsymbol{\theta}) = \boldsymbol{h_0}\,(\mathbf{t})\, \boldsymbol{exp}(\boldsymbol{\theta}'\boldsymbol{Z}) \quad \forall\, \boldsymbol{t} > 0$$

C'est un modèle à risques proportionnels c'est-à-dire que le rapport de risque défini pour une variable X par $\frac{\boldsymbol{h}(\boldsymbol{t}|\boldsymbol{X_i}.\boldsymbol{\theta})}{\boldsymbol{h}(\boldsymbol{t}|\boldsymbol{X_j}.\boldsymbol{\theta})} = \frac{\exp(\boldsymbol{X_i}\boldsymbol{\theta})}{\exp(\boldsymbol{X_j}\boldsymbol{\theta})}$ est constant quel que soit t.

❖ **Estimation des paramètres**

La méthode d'estimation utilisée dans le modèle de Cox est la vraisemblance partielle. Cette méthode consiste à estimer uniquement le coefficient de la régression θ en considérant le hasard de base h_0 comme un paramètre de nuisance. Cette vraisemblance partielle encore appelée vraisemblance de Cox est donnée par l'expression suivante où n correspond au nombre d'individu.

$$\boldsymbol{L_{COX}} = \prod_{\boldsymbol{i=1}}^{\boldsymbol{n}} \left[\frac{\boldsymbol{exp}(\boldsymbol{\theta}'\boldsymbol{Z_i})}{\sum_{\boldsymbol{j=1}}^{\boldsymbol{n}} \boldsymbol{exp}(\boldsymbol{\theta}'\boldsymbol{Z_i})\mathbf{1}_{\{\boldsymbol{T_i} \le \boldsymbol{T_j}\}}} \right]^{\boldsymbol{d_i}}$$

Nous retrouvons le détail du calcul de cette vraisemblance dans Fréderic PLANCHET [2014][4].

❖ **Tests du modèle de Cox**

Nous retrouvons principalement deux types de test pour ce modèle:

➢ **Validation hypothèse de proportionnalité**

Pour tester cette hypothèse nous utiliserons une méthode graphique qui consiste à tracer la fonction $\log(-\log(\mathrm{S_t}))$ en fonction de $\log(\text{ancienneté})$ pour toutes les modalités de chaque variable.
En effet nous avons la relation suivante:

$$\mathrm{S(t)} = S_0(\mathrm{t})^{\boldsymbol{exp}(\boldsymbol{\theta}'\boldsymbol{Z})} \Rightarrow \log\big(-\log\big(\boldsymbol{S}(\boldsymbol{t})\big)\big) = \boldsymbol{\theta}'\boldsymbol{Z} + \log\big(-\log\big(\boldsymbol{S_0}(\boldsymbol{t})\big)\big)$$

[4] **Frédéric PLANCHET [2014] «** Modèle de durée, application actuarielle », 2- Statistique des modèles de durée paramétriques et semi-paramétriques, http://www.ressources-actuarielles.net/

Dans le cas de risques proportionnels le tracé de la fonction $\log(-\log(S_t))$ devrait se traduire par des courbes translatées.

- **La nullité globale des paramètres**

Nous retrouvons trois tests pour vérifier l'hypothèse nulle $H_0 : \theta = 0$; notons $\theta^{(0)}$ le vecteur des paramètres sous cette hypothèse et p le nombre de variable du modèle.

1. Test de vraisemblance dont la statistique est

$$\xi^R = 2\left[2(\log L_{Cox}(\theta) - \log L_{Cox}(\theta^{(0)})\right] \underset{H_0}{\rightarrow} \chi^2(p)$$

2. Test de Wald avec la statistique suivante:

$$\xi^W = \left(\hat{\theta} - \theta^{(0)}\right)' I\left(\hat{\theta}\right)^{-1}\left(\hat{\theta} - \theta^{(0)}\right) \underset{H_0}{\rightarrow} \chi^2(p)$$

3. Et enfin le Test du Scor avec la statistique

$$\xi^W = \left(\frac{\partial \log L_{Cox}}{\partial \theta}\right)'_{\theta=0} I\left(\hat{\theta}\right)^{-1}\left(\frac{\partial \log L_{Cox}}{\partial \theta}\right)_{\theta=0} \underset{H_0}{\rightarrow} \chi^2(p)$$

2.2.1.1. Modèle de COX avec covariables constantes

Pour étudier l'impact des variables (caractéristiques des contrats MRH) sur la durée de vie des contrats représentée par la variable ancienneté, nous devons d'abord étudier la corrélation entre ces variables et cette dernière.

La première sélection faite sur la partie de la modélisation avec la régression logistique nous a permis d'éliminer toutes les variables qui sont liées entre elles. En partant de cette liste et en ajoutant l'hypothèse que les variables doivent être constantes au cours du temps nous retiendrons la liste suivante :

Variable	Corrélation avec l'ancienneté
Qualité juridique	29 %
Capital mobilier	28 %
Franchise	25 %
Formule	17 %
Remise commerciale	16 %
Type habitation	13 %
Zonier DDE	7 %
Nombre de pièces	7 %

Tableau 31 : Variables candidates modèle Cox

Pour utiliser le modèle de Cox, nous devons vérifier d'abord l'hypothèse principale de proportionnalité des risques sur les variables explicatives à intégrer encore appelé covariables.

Nous vérifions cette hypothèse sur toutes les variables listées dans le Tableau 31.

- **Test de l'hypothèse de proportionnalité**

Nous exposons dans cette partie les résultats du test sur quelques variables, le reste est mis en annexe 5.

Sur le graphe du $\log(-\log(\text{survie}))$ de la variable **Formule** de la Figure 18, nous constatons que les courbes des différentes modalités sont bien translatées, et donc l'hypothèse de proportionnalité est bien vérifiée pour cette variable. Cette hypothèse est aussi vérifiée pour les variables **franchise** et **Nombre de pièces**; pour cette dernière une recodification des modalités a été faite en considérant une modalité pour chacun des trois cas suivants: nombre de pièce égale à 3, nombre de pièces supérieur à 3 et enfin nombre de pièces inférieur à 3.

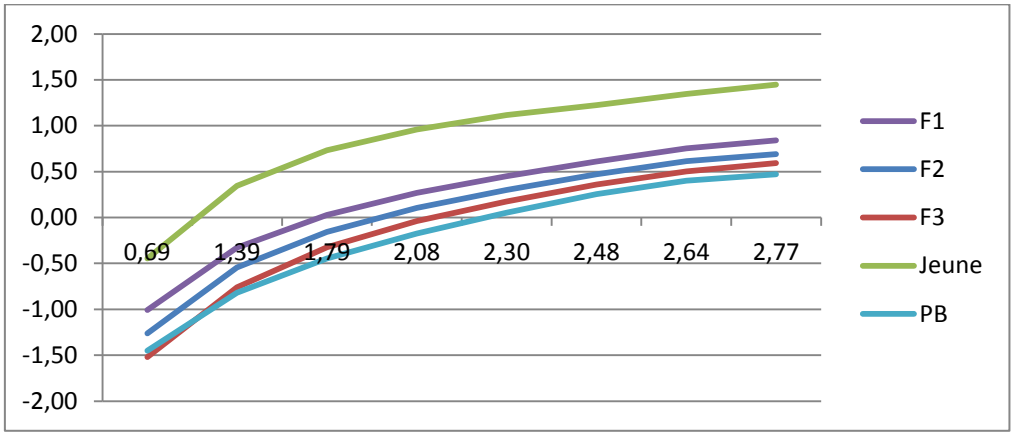


Figure 18 : Log (-log (survie)) variable Formule

Nous voyons sur le graphe ci-dessous que pour la variable **type habitation** l'hypothèse de proportionnalité est aussi vérifiée sauf pour la modalité mobil-home, que nous préférerons supprimer pour la suite.

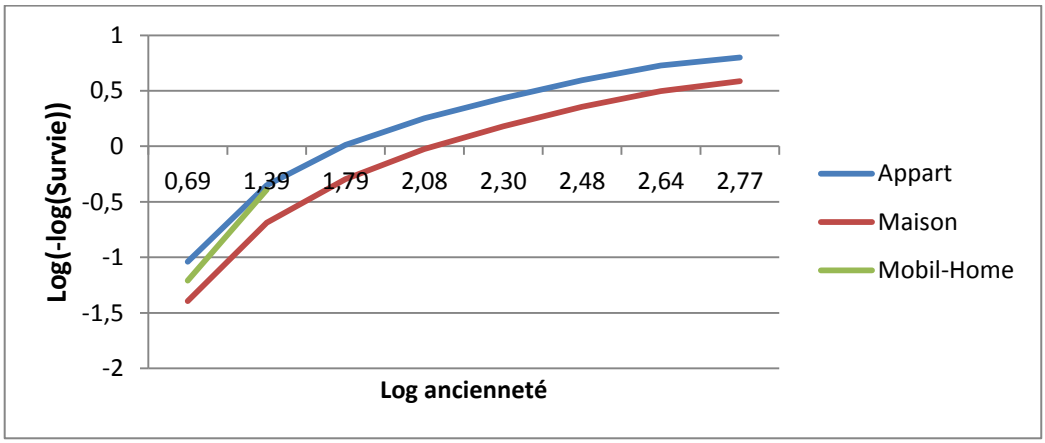


Figure 19: Log (-log (survie)) variable Type habitation

Pour la variable capital mobilier l'hypothèse de proportionnalité n'est pas vérifiée comme le montre le graphe ci-dessous et donc elle sera écartée de la modélisation. De même que la variable **zonier DDE** (nous trouverons en annexe 5 le graphique de test pour cette variable).

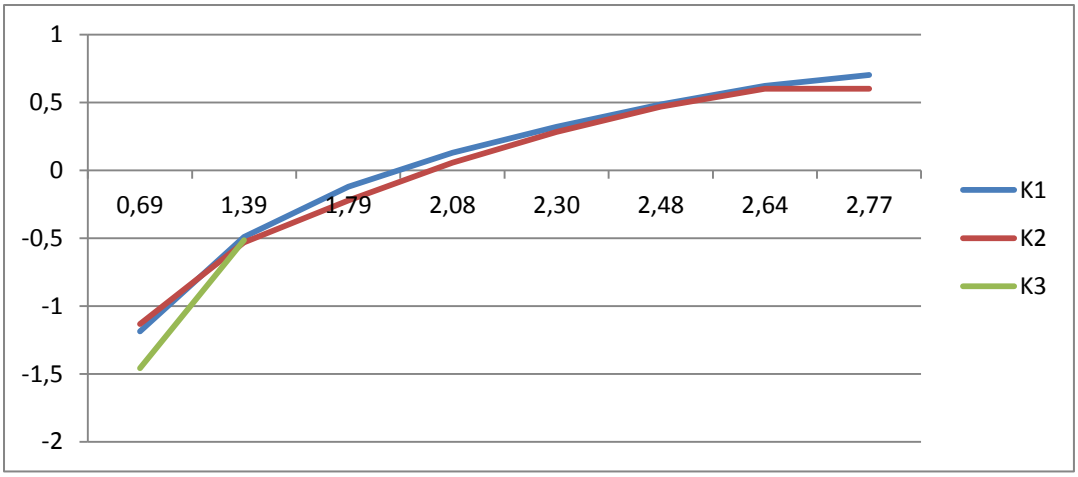


Figure 20 : Log (-log (survie)) variable Capital mobilier

Pour la variable **remise commerciale**, seule la modalité « 0,2 » ne respecte pas l'hypothèse de proportionnalité. Comme les contrats ayant ce taux de remise font 2 % de la base d'étude, nous décidons de les supprimer. Ainsi la variable pourcentage de la remise sera retenu dans le modèle mais sans la modalité « 0,2 ».

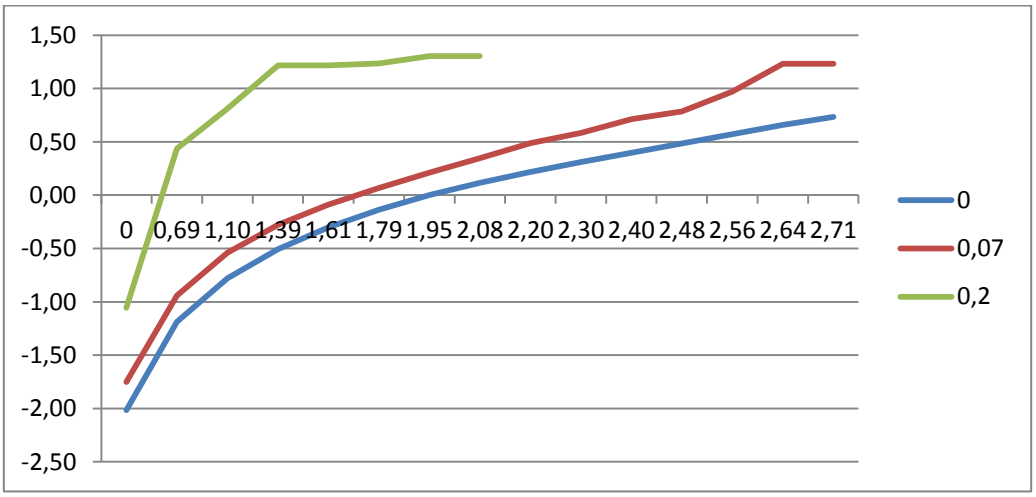


Figure 21 : Log (-log (survie)) variable remise commerciale

Nous considérerons que pour la variable nombre de pièce l'hypothèse de proportionnalité est vérifiée.

Au final nous retiendrons les variables suivants pour le modèle: **franchise**, **Type habitation**, **qualité juridique**, **Remise commerciale**, **nombre de pièce**, **formule**.

Le modèle de Cox possède une autre spécificité qui porte sur les variables explicatives qualitatives. En effet pour pouvoir l'appliquer les variables explicatives qualitatives doivent toutes être binaires. Ainsi pour les variables qualitatives ayant un nombre de modalités supérieur 2, une recodification est nécessaire. Celle-ci consiste à créer une variable binaire pour chaque modalité qui prendra la valeur 1 si l'individu a la modalité en question et 0 sinon.

Nous regroupons dans le tableau suivant ces nouvelles variables :

Variable initiale	Nouvelles variables
Franchise	Franchise 1 (0 €)
	Franchise 2(70 €)
	Franchise 3 (classe de référence 130 €)
	Franchise 4 (260 €)
Nombre de pièces	Nombre de pièce 1 < 3 pièces
	Nombre de pièce 2 >3 pièces
	Nombre de pièce 3(Classe de référence)
Formule	Formule1 (Jeune)
	Formule2 (F1) (classe de référence)
	Formule3 (F2)
	Formule4 (F3)
	Formule5 (PB)

Tableau 32 : Liste des nouvelles variables correspondantes aux modalités

Les variables qualité juridique, type d'habitation et remise commerciale sont binaires avec respectivement comme modalité de référence « locataire », « Maison » et « 7 % ».

Notons que les variables représentant les modalités de référence ne seront pas introduites dans le modèle, elles permettent de calculer la survie de base S_0 et permettrons aussi d'interpréter les résultats obtenues sur les autres modalités.

❖ Modèle Cox sur base 1

Sous SAS le modèle de Cox se réalise par la procédure **PHREG.** En introduisant les variables retenues nous obtenons les résultats suivants:

Statistiques d'ajustement du modèle		
Critère	**Sans covariables**	**Avec covariables**
-2 LOG L	30298870	30159871
AIC	30298870	30159895
SBC	30298870	30160038

Test de l'hypothèse nulle globale : BETA=0			
Test	**Khi-2**	**DDL**	**Pr>Khi-2**
Likelihood Ratio	138999,123	12	<,0001
Score	155154,865	12	<,0001
Wald	147607,830	12	<,0001

Les tests de nullité globale des paramètres nous conduisent tous à rejeter cette hypothèse donc notre modèle est pertinent.

Paramètre	DDL	Valeur estimée des paramètres	Erreur type	Khi-2	Pr>Khi-2	Rapport de risque
Remise commerciale	1	-0,22328	0,00267	6990,3926	<,0001	0,800
Qualité juridique	1	-0,50929	0,00243	43967,5655	<,0001	0,601
Type Habitation	1	-0,01170	0,00258	20,5441	<,0001	0,988
franchise1	1	-0,03210	0,00341	88,8258	<,0001	0,968
franchise2	1	-0,10227	0,00275	1380,0824	<,0001	0,903
franchise4	1	0,28623	0,01476	376,0805	<,0001	1,331
formule1	1	0,61830	0,00447	19135,2897	<,0001	1,856
formule2	1	0,02672	0,00283	89,1124	<,0001	1,027
formule4	1	-0,07801	0,00354	485,1071	<,0001	0,925
formule5	1	-0,06197	0,00417	220,8273	<,0001	0,940
Nbpe1	1	0,10806	0,00268	1620,9263	<,0001	1,114
Nbpe2	1	0,01558	0,00265	34,5018	<,0001	1,016

Tableau 33 : Estimation des Paramètres Modèle de Cox

Le tableau ci-dessus nous montre que tous les paramètres sont significatifs au seuil de 5 %.

- **Interprétation des résultats**
  - ✓ **Ratio de Risque**

Les paramètres ainsi estimés permettent de mesurer le rapport de risque qui est égale à e^{β_i}, $\beta_i$est le paramètre de la variable i.

- **Remise commerciale**

Avec un rapport de risque de **0,80,** pour toutes choses égales par ailleurs aux valeurs de référence, les contrats n'ayant pas de remise commerciale ont un risque de résiliation de 20 % moins que celui des contrats ayant eu une remise de 7 %.

- **Qualité juridique**

La valeur du ratio de risque pour la variable qualité juridique est de **0,601,** donc en considérant toutes choses égales par ailleurs aux valeurs de référence nous pouvons dire que le risque de résiliation des propriétaires est 40 % moins que celui des locataires.

- **Type habitation**

Pour le type habitation le rapport de risque vaut **0,988**. Donc nous pouvons dire que pour toutes choses égales par ailleurs aux valeurs de référence les contrats en maison ont un risque de résiliation légèrement (1 %) supérieur à ceux en appartement.

- **Nombre de pièce**

Le ratio du nombre de pièces inférieur à 3 est de 1,114, donc pour toutes choses égales par ailleurs, les contrats ayant cette modalité ont plus de chance de résilier que ceux possédant un nombre de pièces à 3. Pour ceux par contre ayant un nombre de pièce supérieur à 3, leur risque de résiliation est légèrement supérieur (1 %) à celui des contrats avec 3 pièces.

- **Formule**

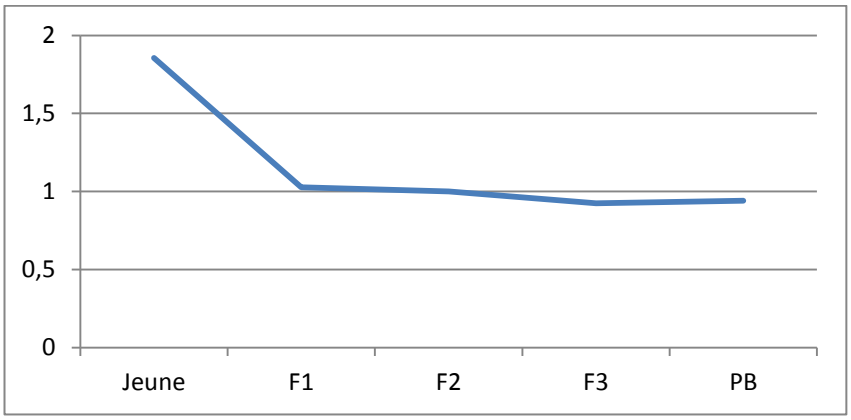


Figure 22 : Rapport de risque Formule

Nous voyons nettement que pour toutes choses égales par ailleurs aux valeurs de référence que les contrats en formule Jeune ont une probabilité de résiliation de 85 % plus importante que les contrats en formule F2. Les contrats souscrits aux autres formules ont une probabilité de résiliation presque identique à ceux de la formule F2.

- **Franchise**

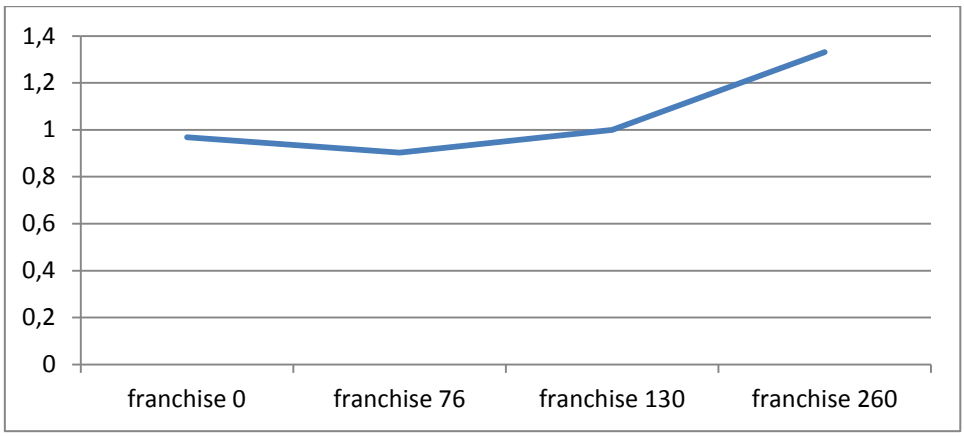


Figure 23 : Rapport de risque Franchise

Ce graphique des rapports de risque sur la variable franchise permet de dire, pour toutes choses égales par ailleurs aux valeurs de référence, que les contrats ayant une franchise de 0 € ou de 76 € ont une probabilité de résiliation légèrement inférieure à celle des contrats ayant une franchise de 130 €, alors que ceux ayant une franchise de 260 € ont une probabilité de 52 % plus que la franchise à 130 €.

➢ **Fonction de survie**

L'estimation de la fonction de survie dans le modèle de Cox est basée sur les valeurs des covariables et des paramètres estimés. Par défaut ce sont les valeurs de référence qui sont prises.

$$\hat{S}(t\backslash z) = \widehat{S_0}(t)^{\exp(\hat{\beta}_{QualJURI}+\hat{\beta}_{RemisecOM}+\hat{\beta}_{Typehabi}+\hat{\beta}_{Franchise0}+\hat{\beta}_{Franchise76}+\hat{\beta}_{Franchise260})}$$

Avec cette formule nous obtenons la fonction de survie estimée de notre portefeuille à l'aide de la fonction **« survfit »** de R que nous représentons sur le graphe ci-dessous.

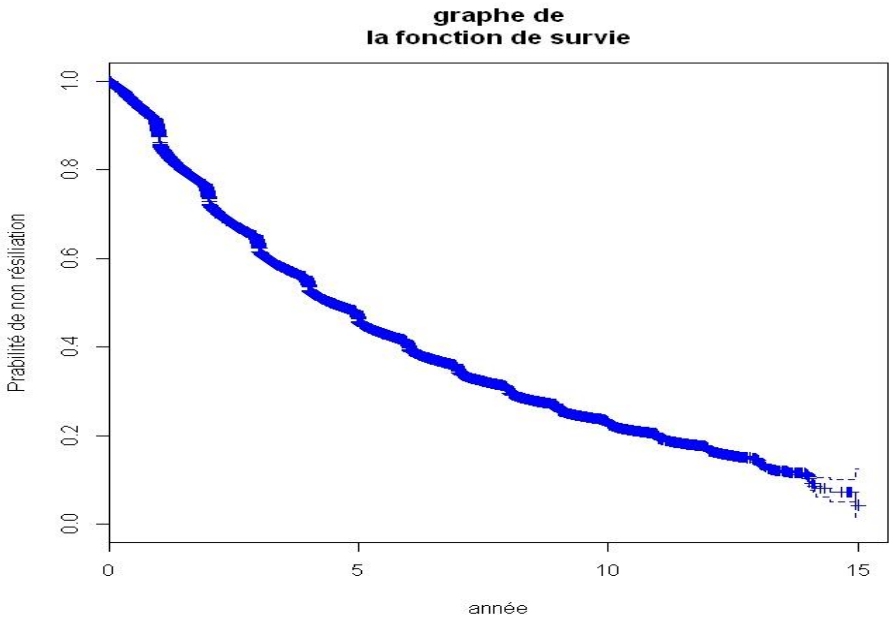


Figure 24 : Fonction de survie modèle de Cox

Nous voyons que comme pour l'estimation de Kaplan-Meier, la fonction de survie est presque linéaire et décroissante suivant l'ancienneté. La probabilité de rester en portefeuille après 15 ans est à un peu moins de 0,1 presque égale à celle dans l'estimation non paramétrique.

Afin de déterminer l'espérance de vie et l'espérance de vie résiduelle des contrats suivant les variables retenues dans le modèle, nous nous intéressons à l'estimation de la fonction de survie suivant ces critères.

- **Fonction de survie par variable**

Pour estimer la fonction de survie par critère, on se place dans le cas où seules les valeurs de la variable considérée changent ,les autres variables étant fixées à leur valeur de référence. C'est ce qui permet de faire une comparaison des fonctions de survie des différentes modalités de la variable.

- ✓ **Qualité juridique**

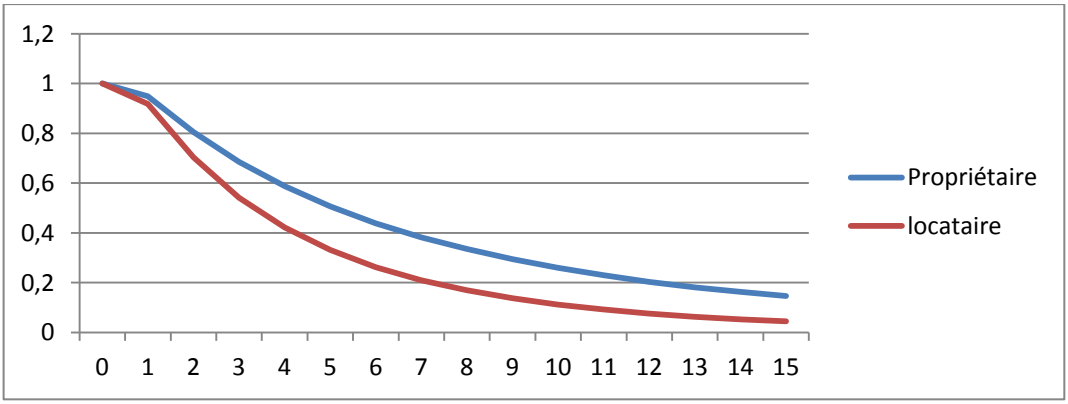


Figure 25 : Survies modalités variable Qualité juridique

Nous voyons sur la Figure 25 que les locataires ont une probabilité de survie nettement inférieure à celle des propriétaires pour toutes choses égale par ailleurs aux modalités de référence. Nous

calculons dans ces conditions l'espérance de vie d'un locataire et propriétaire à la souscription d'un contrat MRH et l'espérance de vie résiduelle suivant l'ancienneté de celui-ci :

	Espérance de vie
Locataire	7 ans
Propriétaire	9,43 ans

Tableau 34 : Espérance de vie suivant la Qualité juridique

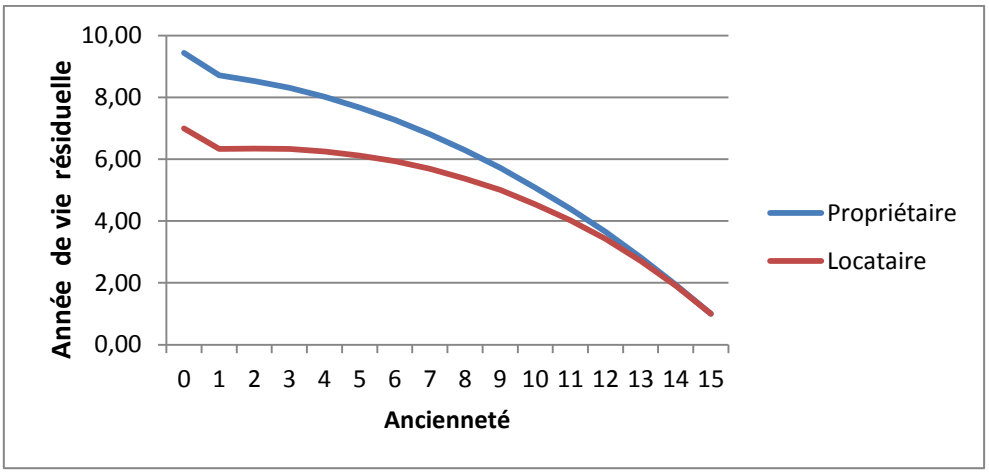


Figure 26 : Espérance de vie résiduelle variable qualité juridique par modalité

L'espérance de vie à la souscription est nettement plus élevée pour les propriétaires que les locataires et cette tendance est observée jusqu'à 13 ans d'ancienneté.

✓ **Type Habitation**

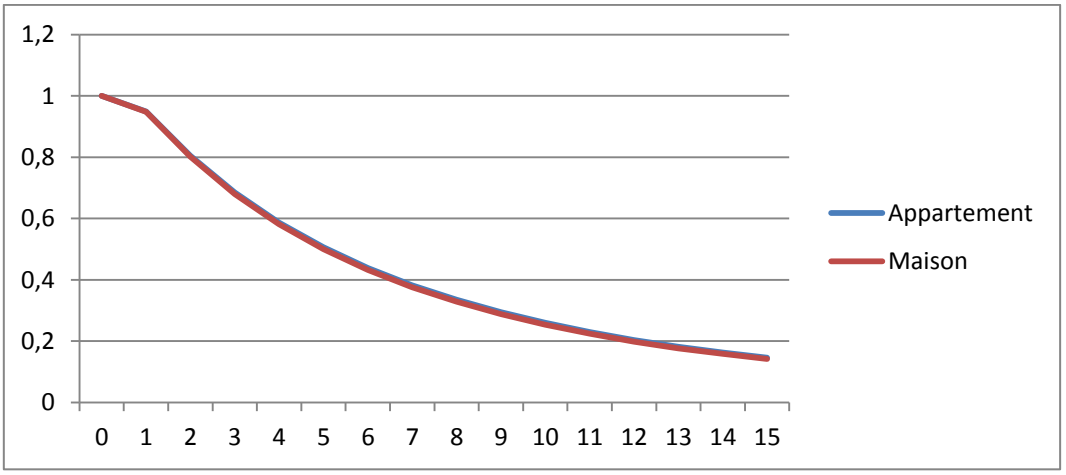


Figure 27 : Survies modalités variable Type Habitation

Pour le critère type habitation nous voyons que les probabilités de survie sont presque identiques entre les appartements et les maisons ce qui se traduit aussi par une espérance de vie à la souscription presque identique.

	Espérance de vie
Maison	9,57 ans
Appartement	9,43 ans

Tableau 35 : Espérance de vie suivant Type habitation

Nous représentons dans le graphique suivant l'espérance de vie résiduelle suivant l'ancienneté des contrats pour les différentes modalités de la variable type habitation.

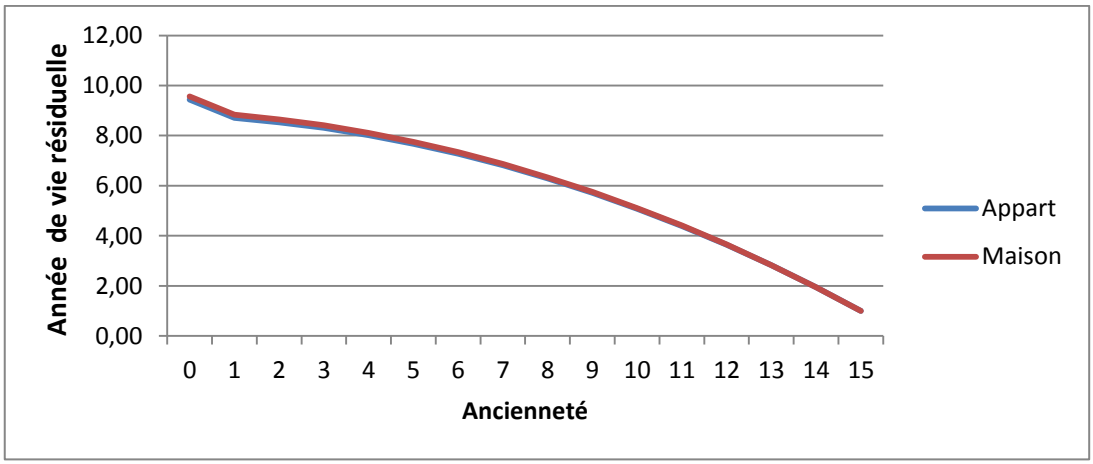


Figure 28 : Espérance de vie résiduelle suivant Type Habitation

❖ **Modèle de Cox sur base 2**

Dans cette section nous désirons uniquement mesurer le risque de résiliation à un an par rapport aux différentes variables retenues dans la section précédente par le modèle de Cox. Pour cela, nous faisons tourner modèle de Cox retenu à la section précédente sur la base de donnée du taux de résiliation à un an. Nous obtenons les résultats suivants :

Test de l'hypothèse nulle globale : BETA=0			
Test	**Khi-2**	**DDL**	**Pr > Khi-2**
Likelihood Ratio	76815,9187	11	<,0001
Score	90876,9563	11	<,0001
Wald	75493,1206	11	<,0001

Estimations par l'analyse du maximum de vraisemblance						
Paramètre	**DDL**	**Valeur estimée des paramètres**	**Erreur type**	**Khi-2**	**Pr>Khi-2**	**Rapport de risque**
Remise Commerciale	1	-0,42962	0,00693	3844,0117	<,0001	0,651
Qualité juridique	1	-0,54078	0,00661	6687,0766	<,0001	0,582
franchise1	1	0,17927	0,00866	428,2402	<,0001	1,196
franchise2	1	-1,66422	0,01225	18446,3660	<,0001	0,189
franchise4	1	1,36170	0,02289	3537,9403	<,0001	3,903
formule1	1	1,24079	0,01316	8890,2380	<,0001	3,458
formule2	1	-0,04552	0,00835	29,7246	<,0001	0,955
formule4	1	-1,13273	0,01072	11156,0450	<,0001	0,322
formule5	1	-0,06472	0,01244	27,0814	<,0001	0,937
Nbpe1	1	0,06242	0,00785	63,2504	<,0001	1,064
Nbpe2	1	-0,03408	0,00718	22,5405	<,0001	0,966

Tableau 36 : Estimation des paramètres

Les tests de nullité globale des paramètres sont tous significatifs au seuil de 1 % ; de même que les paramètres estimés ci-dessus.

Comme pour les résultats de la section précédente nous voyons que pour toutes choses égales par ailleurs aux valeurs de références, les contrats ayant une remise commerciale de 7 % résilient plus que ceux n'ayant pas de remise. De même, les locataires résilient plus sur un an que les propriétaires; les contrats avec moins de 3 pièces résilient plus que ceux avec 3 ; qui eux ont un risque similaire à 1 % près des contrats ayant un nombre de pièces supérieur à 3. Pour les variables franchise et formule nous représentons sur les graphes ci-dessous les rapports de risque :

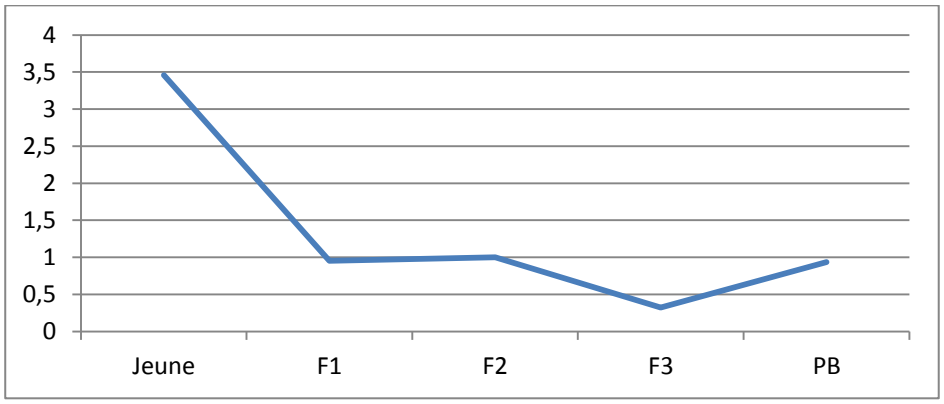


Figure 29 : Rapport de risque de la variable Formule

Pour toutes choses égales par ailleurs aux valeurs de références, la probabilité de résiliation à un an de la formule jeune est 2,5 fois plus importante que celle de la formule F2. Les contrats en formule F3 résilient près de 65 % moins que ceux en formule F2.Les contrats en F1 et les PB ont presque la même probabilité de résiliation que les contrats en F2.

Nous représentons ensuite le graphe des rapports de risque de la variable franchise.

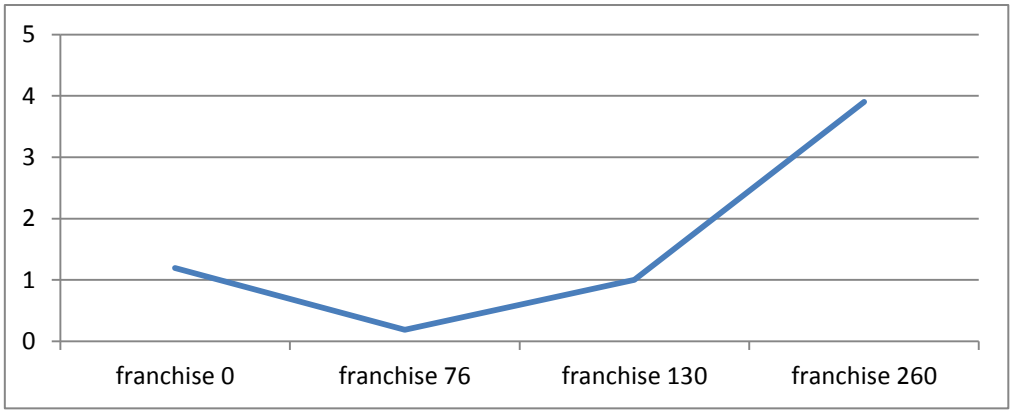


Figure 30 : Rapport de risque de la variable Franchise

Les contrats avec une franchise à 260 € ont près de 3 fois plus de chance d'être résiliés sous un an que les contrats en franchise 130 € ; pour toutes choses égale par ailleurs aux valeurs de référence. Les contrats avec une franchise à 76 € ont une probabilité de résiliation sous un an de près de 80 % moins importante que celle des contrats en franchise 130 €.

2.2.1.2. Modèle de Cox: Prise en compte d'une variable temporelle

La prise en compte d'une covariable dépendante du temps permet d'affiner d'avantage notre modèle.
En effet nous avons supposé dans les modèles de Cox présentés précédemment que les variables utilisées sont constantes dans le temps ; ce qui n'est pas toujours le cas. C'est pourquoi dans cette section nous intégrons dans le modèle la variable **prime** dont les valeurs varient suivant les années.
Dans ce contexte l'hypothèse de proportionnalité des risques ne peut être vérifiée.

- **Construction de la base de modélisation**

Pour prendre en compte cette variabilité de la prime suivant les années, nous devons récupérer la prime annuelle des contrats de la base de modélisation construite dans la première modélisation.

Pour ce faire nous construisons une base « prime » à partir de la base contrats de janvier 2014. Cependant l'historique des situations de la majorité des contrats de cette base s'arrêtent en 2005. Donc nous ne retiendrons que les contrats dont la prime est renseignée de 2005 à 2014. Cela correspond à des contrats des offres MRH1 ou MRH2.

Afin d'éviter une censure gauche, nous supprimons les contrats de la base de modélisation de la première section (« base 1 ») qui sont résiliés avant 2005. A cette nouvelle base nous faisons correspondre la base « prime » pour obtenir une base contenant pour chaque ligne un contrat avec ses caractéristiques et les différentes primes versées sur la période 2005-2014 selon sa situation.
Notons qu'une variable prime sera créée pour chaque année (prime1, prime2,…, prime10) de même une variable année (année1,…, année10).

- **Résultats**

Avec cette nouvelle base nous obtenons les résultats suivants avec le modèle de Cox :

Statistiques d'ajustement du modèle		
Critère	Sans covariables	Avec covariables
-2 LOG L	3407112,0	3388525,4
AIC	3407112,0	3388541,4
SBC	3407112,0	3388620,5

Test de l'hypothèse nulle globale : BETA=0			
Test	Khi-2	DDL	Pr > Khi-2
Likelihood Ratio	18586,5979	8	<,0001
Score	18811,8905	8	<,0001
Wald	18429,9153	8	<,0001

Les tests de nullité globale des paramètres nous conduisent tous à rejeter cette hypothèse donc notre modèle est pertinent.

Estimation par l'analyse du maximum de vraisemblance						
Paramètre	DDL	Valeur estimée des paramètres	Erreur type	Khi-2	Pr > Khi-2	Rapport de risque
Prime	1	-0,00209	0,0000374	3129,7880	<,0001	0,998
Remise Commerciale	1	-0,58162	0,01306	1983,1099	<,0001	0,559
Qualité juridique	1	-0,50785	0,00591	7375,3152	<,0001	0,602
Type habitation	1	0,01327	0,00687	3,7315	0,0534	1,013
franchise1	1	0,39215	0,00933	1765,2976	<,0001	1,480
franchise2	1	0,38840	0,00891	1900,3945	<,0001	1,475
Nbpe1	1	0,05597	0,00737	57,6696	<,0001	1,058
Nbpe2	1	0,18248	0,00764	569,8194	<,0001	1,200

Tableau 37 : Estimation des paramètres du modèle de Cox sur base 2

Dans ce modèle les rapports de risque permettent d'affirmer comme dans le premier modèle que pour toutes choses égales par ailleurs aux valeurs de référence les propriétaires ont moins de chance de résilier que les locataires (40 %).

Le rapport de risque de la variable **Prime** est de 0,998, donc on peut dire qu'un contrat ayant une augmentation d'un euro a 0,2 % moins de chance de résilier qu'un contrat n'ayant pas eu cette augmentation pour toutes choses égales par ailleurs aux valeurs de référence. Donc nous dirons que l'augmentation d'un euro à la prime d'un contrat n'augmente pas sa probabilité de résiliation.

Ce résultat rejoint ceux obtenus par les modèles logistiques, ainsi que les observations faites sur la résiliation par rapport à la prime.

Ce résultat s'expliquerait dans ce modèle par le fait que l'information apportée par les variables tarifaires (**qualité juridique**, **type habitation**, **nombre de pièces**) se retrouve dans la prime ce qui limite l'influence de celle-ci sur l'évènement résiliation en présence de ces variables. En effet la prime d'un contrat MRH est obtenue à partir des variables retenue dans la modélisation des coûts et des fréquences à l'aide d'un modèle GLM; et, d'année en année une augmentation uniforme par génération d'offre est appliquée sur le portefeuille. Etant donné que nous avons dans ce modèle de COX les valeurs de la **Prime** sur 10 ans, l'information apportée par cette variable se résumera donc à l'information contenue dans les primes modélisées et donc un reflet de celle apportée par les variables tarifaires, d'où le résultat sur la **Prime**.

Cependant pour déterminer précisément l'influence de la prime sur le comportement des clients MRH face au risque de résiliation, il serait intéressant d'étudier l'élasticité du taux de résiliation au prix.

Partie 4: Comparaison des modèles et Projection de portefeuille

1. Comparaison des modèles utilisés

Pour comparer les deux types de modèles utilisés dans la modélisation du taux de résiliation, nous nous appuyons sur l'effet des variables sur le risque résiliation dans chaque type de modèle. Pour cela nous considérons uniquement les variables qui sont présentes dans les deux types de modèle.

- **Résiliation horizon non défini**

Dans cette partie nous allons comparer les résultats de la régression logistique à ceux du modèle de durée construit sur la base 1. Nous retenons pour cela les variables **Formule**, **Type habitation**, **Nombre de pièces**, **qualité juridique**, **franchise** qui sont communes aux deux modèles.

Le Tableau 22 des odds ratios nous avait permis de conclure que dans le modèle de résiliation à horizon non défini, les contrats en maison avaient plus de chance d'être résiliés que ceux en appartement et que les locataires résilient plus que les propriétaires. Nous retrouvons ces résultats dans le modèle de Cox sur la base 1. Les proportions du risque de résiliation ne sont pas égales dans les deux modèles mais sont dans un même ordre de grandeur.
Pour la variable **nombre de pièces**, nous avons vu que dans le modèle de Cox, les contrats ayant un nombre de pièces supérieur à 3 avaient un peu plus de chance d'être résiliés que ceux à 3 pièces. Dans la modèle logistique nous retrouvons ce résultat pour toutes les modalités de la variable Nombre de pièce supérieur à 3, avec une probabilité de résiliation plus importante que dans le modèle de Cox. En outre pour les contrats avec un nombre de pièce inférieur à 3, nous avons une probabilité de résiliation moins importante que les contrats à 3 pièces dans la régression logistique alors que dans le modèle de Cox nous avons un résultat inverse.
De même pour les variables **Formule** et **Franchise**, les résultats obtenus sur les deux modèles ne sont pas identiques.
En effet pour la **Formule**, dans la régression logistique nous trouvons que les contrats en formule jeunes et F1 résiliaient moins que ceux en formule F2 et que les PB et les formule F3 résilient plus que les contrats en F2. Pour le modèle de Cox les résultats sont presque à l'opposé pour les formules Jeune, F3 et PB; en effet les contrats en formule jeune résilient plus que ceux en F2, les contrats en F3 résilient moins alors que les PB ont un risque de résiliation presque identique à la F2.
Quant à la variable Franchise, les résultats entre les deux modèles différent sur les modalités 76 € et 260 €. En effet nous avons vu pour la régression logistique que les contrats avec une franchise de 76 € résiliaient plus que ceux avec franchise 130 € et que ceux avec une franchise de 260 € résiliaient moins que les contrats avec la franchise de référence. Pour le modèle de Cox nous obtenons un résultat inverse.
Donc sur l'horizon non défini les résultats de la modélisation obtenus sur les deux modèles ne sont pas totalement conformes. Cela peut pourrait s'expliquer par la différence existant sur les variables retenues dans les deux modèles. Cependant les résultats du modèle de Cox restent plus cohérents avec l'observation empirique faite sur le portefeuille; donc nous les privilégions pour cet horizon de résiliation.

➢ **Résiliation à horizon d'un an**

A présent faisons la même approche de comparaison entre le modèle logistique à un an et le modèle de Cox construit sur la « base 2 ».
Nous retiendrons pour cela que les variables **qualité juridique** et **formule** qui sont communes aux deux modèles. D'après le modèle logistique, les locataires ont une probabilité de résiliation à un an plus importante que les propriétaires, ce que nous retrouvons dans le modèle de Cox.
Pour la variable formule, dans le modèle de Cox, nous avons conclu que les contrats en formule Jeune ont plus de chance de résilier en un an que ceux en F2, alors que les contrats en formule F1, PB et F3 ont moins chance d'être résiliés que ceux de la formule F2 voir presque identique pour les deux dernier formules.
Nous retrouvons le même résultat dans le modèle logistique pour les formules Jeune, PB et F3 par rapport à la formule F2. Pour les contrats en PB, la probabilité de résiliation est nettement inférieure à celle de la formule F2 dans ce modèle.
Par contre les contrats en formule F1 ont légèrement plus de chance d'être résilié en un an que les formule F2 dans ce modèle.
Donc les résultats exposés plus haut sur la résiliation à un an, sont globalement identiques dans les deux modèles et restent cohérents par rapport aux observations faites sur le portefeuille MRH.

2. Application: Projection du portefeuille à un an

Pour mettre en application les résultats de la modélisation du taux de résiliation, nous allons dans cette partie faire une projection de portefeuille sur le produit MRH à horizon d'un an.

Pour ce faire nous choisissons d'utiliser les résultats de la modélisation par régression logistique. Ce choix se base sur la simplicité de la prédiction obtenue avec ce type de modèle.
Cette projection de portefeuille va prendre en compte deux facteurs: la résiliation sur un an et le nombre d'affaires nouvelles (AFN) souscrites sur la même période. Nous allons dans un premier temps faire une prédiction des AFN entre janvier 2014 et janvier 2015 avant d'appliquer le modèle de résiliation sur la base des contrats actifs à janvier 2014.

2.1. Estimation des AFN

Nous disposons d'un historique mensuel des AFN entre 1999 et 2013.L'idée est de proposer un modèle de série temporelle adéquat qui nous permette de faire une prédiction sur ces données.

2.1.1. Présentation des séries temporelles

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

- La tendance ou trend (f_t) d'une série représente son évolution à long terme. Elle traduit son comportement «moyen». Elle peut se définir comme une combinaison linéaire de m fonctions du temps choisie à priori.
- La saisonnalité (s_t) correspond à un phénomène qui se répète à intervalle de temps régulier.

- Les résidus ou bruit (ε_t) correspondent à des fluctuations irrégulières, en générale de faible intensité mais de nature aléatoire.

L'étude d'une série temporelle nécessite l'identification de ces différentes composantes; nous utiliserons la méthode des moyennes mobiles pour la décomposition de la série des AFN.

- **Moyennes Mobiles**

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

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

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

Une moyenne mobile est centrée si $m_1 = m_2 = m$; elle est symétrique si et seulement si $\theta_{-i} = \theta_i$

Une **moyenne mobile arithmétique** est une moyenne mobile centrée d'ordre impaire et telle que $\boldsymbol{\theta_i = \frac{1}{2m+1}}$.

- **Effet de la moyenne mobile sur la tendance et la saisonnalité**

L'application d'une moyenne mobile arithmétique ne modifie pas une tendance constante et conserve une tendance linéaire.
Si une série possède une saisonnalité de période P, alors l'application d'une moyenne mobile d'ordre P supprime cette saisonnalité.

2.1.2. Décomposition et ajustement de la série temporelle des AFN

Nous représentons ci-dessous la série temporelle correspondant à notre jeu de données observé sur 14 ans.

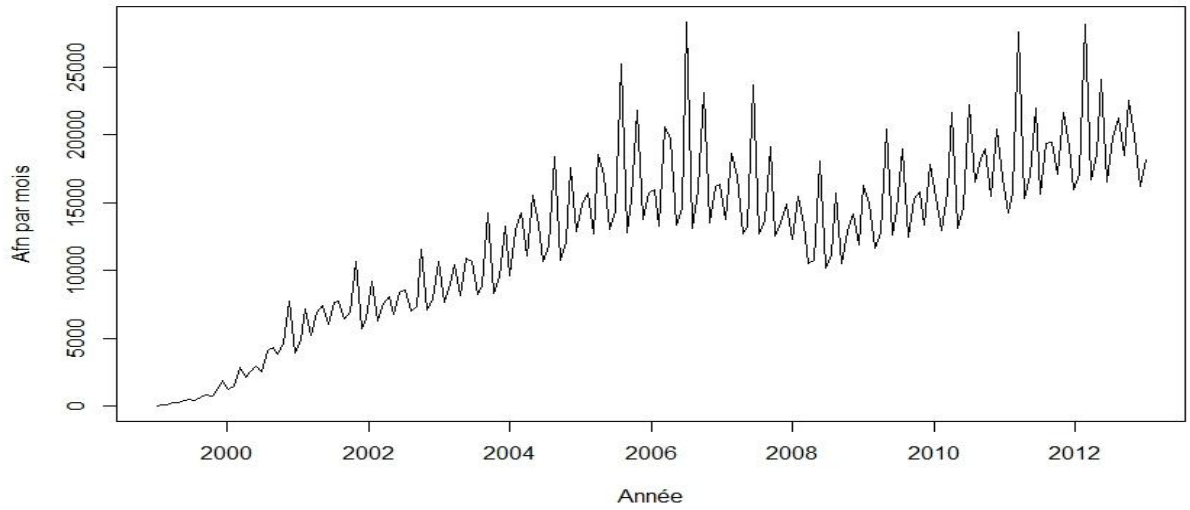


Figure 31 : Série des affaires nouvelles

Au vu de ce graphique ci-dessus, une décomposition additive semble adaptée à la série des AFN. Nous la considérons de la forme suivante:

$$X_t = f_t + s_t + \varepsilon_t$$

Avec f_t la saisonnalité et ε_t les résidus.

Pour s'assurer de cette décomposition nous utilisons les moyennes mobiles pour représenter les différentes composantes de la série. Avec la fonction décompose de R nous obtenons le graphe ci-dessous:

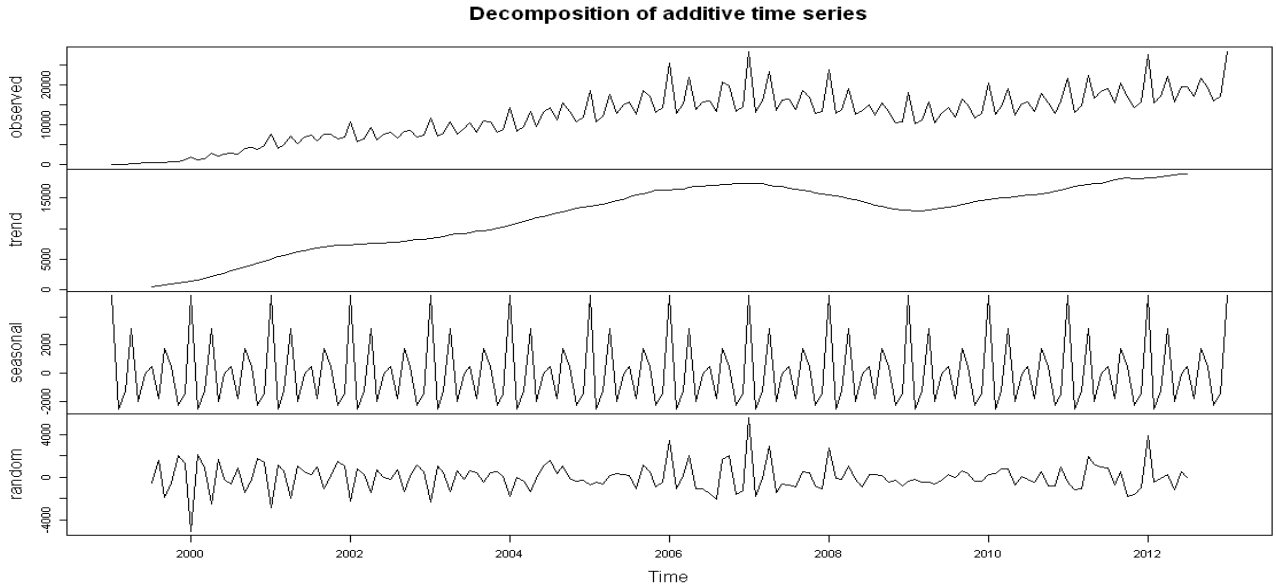


Figure 32 : Décomposition additive par Moyenne mobile série AFN

La tendance est globalement linéaire avec une saisonnalité de 12 mois. L'ajustement que nous proposons de faire sur cette série est le lissage exponentiel Holt Winters avec une saisonnalité additive.

- **Les lissages exponentiels**[5]

Les méthodes de lissage exponentiel sont des techniques empiriques de prévision qui accordent une grande importance aux observations passées d'une série temporelle. Nous retrouvons principalement les trois méthodes suivantes:

- **Le lissage exponentiel simple**

Permet d'effectuer des prévisions pour des séries dont la tendance est constante et sans saisonnalité. Soit X_t une tel série avec T premières observations $X_1, \dots X_T$; nous cherchons à prévoir X_{T+h}.

Soit β un réel tel que $0 < \beta < 1$; $\widehat{X}_{T+h}$ est solution du problème des moindre des carré suivant :

$$\min_a \sum_{j=0}^{T-1} \beta^j (X_{T-j} - a)^2$$

Et la solution est donné par

$$\boldsymbol{\widehat{X}_{T+h} = (1-\beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}}$$

- **Le lissage exponentiel double**

Le lissage exponentiel double généralise le lissage simple au cas où la série peut être ajustée par une droite au voisinage de T mais sans saisonnalité. La prévision de X_{T+h} est de la forme: $\hat{X}_{T+h} = \hat{a}_T \mathrm{h} + \hat{b}_T$ ou $(\hat{a}, \hat{b})$ minimise la fonction:

$$\sum_{j=0}^{T-1} \beta^j (X_{T-j} - (aj + b))^2$$

En notant respectivement $S_1 = (1-\beta) \sum_{j=0}^{T-1} \beta^j X_{T-j}$ et $S_2 = (1-\beta) \sum_{j=0}^{T-1} \beta^j S_1(t-j)$ la série lissée et la série doublement lissée.

La prévision $\hat{X}_{T+h}$ par la méthode de lissage exponentiel double est donnée par $\boldsymbol{\widehat{X}_{T+h} = \widehat{a}_T \mathrm{h} + \widehat{b}_T}$

où β est la constante de lissage et le couple $(\hat{a}, \hat{b})$ est donné par:

$$\begin{cases} \hat{a}_T = \dfrac{1-\beta}{\beta} (S_1(T) - S_2(T)) \\ \hat{b}_T = 2S_1(T) - S_2(T) \end{cases}$$

- **Le lissage de Holt-Winters**

C'est une méthode qui s'applique à la fois au série avec et sans saisonnalité. Elle diffère de la méthode du lissage exponentiel double par les formules de mise à jours. Ces formules diffèrent aussi

[5] Les éléments de cette section sont inspirés du cours de **LAGNOUX Agnès** , « Séries chronologiques ».

selon le type de saisonnalité du modèle considéré (sans saisonnalité, saisonnalité additive ou multiplicative). Nous étudierons uniquement le cas de la saisonnalité additive.

- **Holt Winters avec saisonnalité additive**

On considère les T premières observations de la série (X_t) , et on suppose que cette série puisse être approchée au voisinage de T par $a(t-T)+b+s_t$ où s_t représente la saisonnalité de période P. L'estimation de a, b, s_t proposée par la méthode Holt Winters est donnée par les formules de mise à jour suivantes:

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

Avec α, β, γ des constantes de lissage appartenant à $]0,1[$

La première formule de mise à jour s'interprète comme une moyenne pondérée de la différence des niveaux estimés aux instants T et $T-1$ et la pente estimée à l'instant $T-1$.
La deuxième s'interprète comme une moyenne pondérée de l'observation X_T(à laquelle on a retranché la composante saisonnière estimée à l'étape précédente) et l'estimation de la tendance faite à l'instant $T-1$.
La troisième s'interprète comme une moyenne pondérée de l'observation X_T (à laquelle on a retranché le niveau calculé à l'instant T) et de la composante saisonnière calculée à l'instant $T-P$).

Notons que dans le cas du modèle de la série des AFN que nous avons retenu, l'hypothèse que les résidus soit un bruit blanc est nécessaire à l'application de la méthode de Holt-Winters.

- **Etude des résidus de la décomposition de la série des AFN.**

Nous vérifions dans cette partie si les résidus obtenus par la décomposition des moyennes mobiles de la série étudiée sont bien un bruit blanc c'est-à-dire un processus aléatoire stationnaire et centré.

- Stationnarité

Pour rappel un processus X_t est stationnaire si son espérance et ses autocovariances sont indépendants du temps.
La fonction d'autocovraiance est défini par $\forall\ h\ \in \mathbb{Z}, \gamma(h) = Cov(X_t, X_{t-h})$ et l'autocorrélogramme par $\forall\ \ h\ \in \mathbb{Z}$, $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$.
La p value du test de **Box-Pierce** est de 0,848 ce qui permet de valider l'homoscédasticité des résidus et le test de non stationnarité de Dickey-fuller produit une p value de 0,01, donc les résidus sont stationnaires.

- Normalité

Nous étudions la normalité des résidus en faisant un ajustement par une loi normale avec une moyenne et variance égale à celle des résidus. Voici la comparaison graphique des densités des résidus et loi normale théorique.

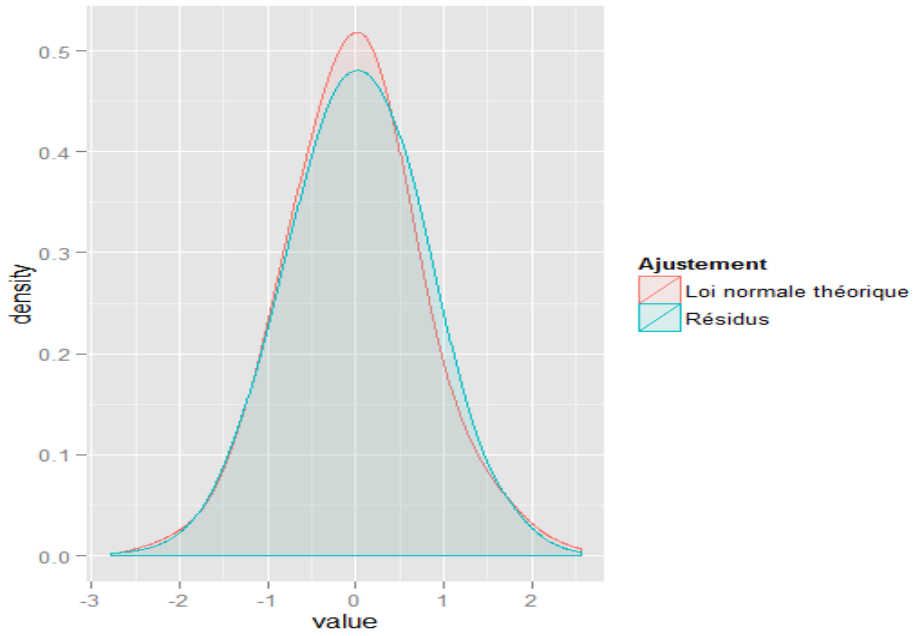


Figure 33 : Ajustement résidus par loi normale

Au vu de ce graphe nous pouvons considérer que les résidus suivent bien une loi normale centrée. Ce qui est confirmé par le test de normalité de **Shapiro Wilk** dont la p value de 0,9496.

Au final les résidus de la décomposition sont un bruit blanc gaussien.

- **Prévisions des AFN avec Holt Winters**

Nous avons vu que la tendance de la série des AFN était linéaire ou localement linéaire et que les résidus de la décomposition sont un bruit blanc gaussien, donc nous pouvons appliquer la méthode de prévision de Holt-Winters sur cette série.
Nous faisons une prévision de 35 mois, de février 2013 à Décembre 2015, que nous représentons sur le graphe ci-dessous :

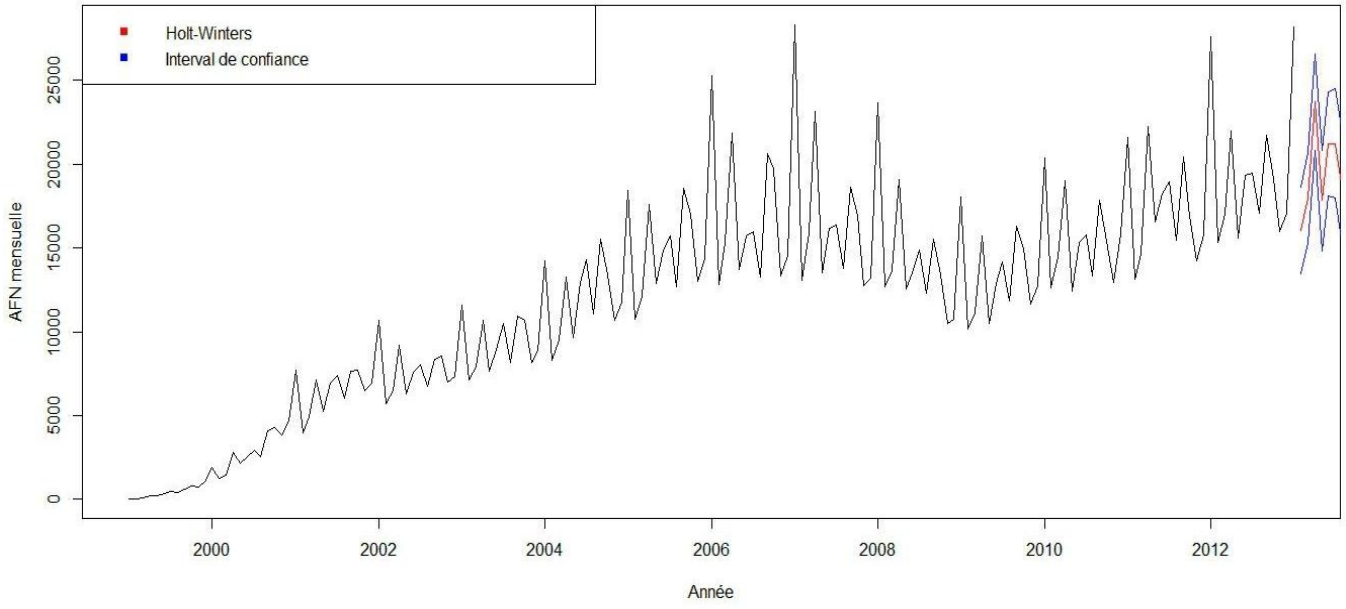


Figure 34:Résultat Prévisions AFN

Les valeurs prédites sont bien cohérentes avec les observations passées. De plus les intervalles de confiance à 95 % des prévisions sont assez proches des valeurs prédites. Donc l'erreur de l'estimation est assez faible.

Pour s'en convaincre, nous trouverons ci-après un tableau de comparaison des valeurs observées et prédites en 2013.

Période	observée 2013	Prédiction 2013	Prédit-Observé
févr-13	16720	16058	-662
mars-13	18600	17970	-630
avr-13	24136	23705	-431
mai-13	16570	17821	1251
juin-13	19828	21228	1400
juil-13	21226	21241	15
août-13	18514	18490	-24
sept-13	22560	22941	381
oct-13	19864	20103	239
nov-13	16194	16985	791
déc-13	18119	18440	321

Tableau 38 : Comparaison AFN Prédites et observées en 2013

La différence maximale entre l'observé et le prédit est de 1400 AFN soit 7,06 % des AFN observées pour le mois de juin. Sur toute l'année 2013, cette différence s'élève à 2651 AFN soit 1 ,25 % des AFN observées sur cette année. Cette erreur de prédiction est relativement faible; de plus vu son signe (nous prédisons un peu plus que l'observé) et la pente de la tendance (cf. Figure 32) entre 2010 - 2013, elle pourrait nous permettre d'être au plus juste sur les prédictions des années à venir.

2.2. Application du modèle de résiliation

Après l'estimation des affaires nouvelles, nous allons dans cette section prédire le nombre de résiliations sur un horizon d'un an.
Pour ce faire, nous allons utiliser le modèle logistique à horizon d'un an construit à la partie 3.
En considérant la base d'étude construite pour ce modèle, nous ne retenons que les contrats actifs en janvier 2014 (les contrats RCVP étant exclus à la construction de la base).
Après application du modèle sur cette base, en considérant un seuil pour la probabilité de résiliation de 0,075 (déterminé lors de la construction du modèle) nous obtenons un taux de résiliation de **18,45 %.** Ce taux de résiliation à un an prédit est légèrement supérieur aux taux de résiliations généralement observés qui sont de 14 % sur l'ensemble du portefeuille actif MRH.
Cette différence s'explique par la non prise en compte des contrats RCVP, qui ont généralement un taux de résiliation très faible, de la suppression de certains contrats présentant des anomalies et enfin la marge d'erreur provenant du modèle.
En outre nous pouvons prédire les résiliations par profil que nous définissons par les variables **nombre de pièce**, **type habitation**, **franchise**, **qualité juridique et formule**. En prenant les 10 profils majoritaires de la base des contrats actifs de janvier 2014, nous obtenons les taux de résiliation à un an suivant :

Profil contrats actifs janv. 14	Taux résiliation prédit
3 pièces –Appart – 130 – locataire - F1	**27,2 %**
2 pièces - Appart - 130 - locataire- F1	**31,6 %**
4 pièces – Maison – 0 – Propriétaire - F2	**12,0 %**
5 pièces - Maison- 0- Propriétaire - F2	**14,0 %**
2 pièces – Appart – 130 – Propriétaire - PB	**6,6 %**
4-pièces - Maison - 130 - Propriétaire - F2	**9,3 %**
3pièces – Appart – 0 – locataire - F2	**31,0 %**
5 pièces- Maison- 0- Propriétaire- F3	**11,2 %**
5 pièces – Maison – 130 – Propriétaire - F2	**10,7 %**
4 pièces - Maison- 0 - Propriétaire - F3	**9,6 %**

Tableau 39 : Profils majoritaires et leurs taux de résiliation à un an prédit

Nous voyons que les profils « propriétaires » ont un taux de résiliation à un an moins important que ceux des « locataires » quelques soit les valeurs des autres variables du profil.
Une campagne marketing pourrait être envisagée en faveur des clients locataires afin de leur garder davantage en portefeuille.
Le nombre de contrats MRH qui seront résiliés sous un an est obtenu en multipliant le taux de résiliation par le nombre de contrats actif en janvier 2014.

2.3. Résultats de la projection

Après avoir obtenu les estimations pour le nombre d'affaires nouvelles et le nombre de résiliations sur un an, nous pouvons à présent estimer le nombre de contrats actifs en janvier 2015. Pour ce faire nous faisons l'hypothèse que le taux de résiliations sur les contrats RCVP est nul. Comme vu plus haut nous avons une légère sur estimation de la résiliation par rapport aux observations antérieures et donc cette hypothèse nous permet de ne pas l'accentuer.

Pour juger de la pertinence de la projection réalisée nous la comparons à celle qui est actuellement faite dans le département Dommages aux Biens & Risques Divers basée sur une méthode déterministe et aussi par rapport aux données observées. Voici le tableau comparatif:

Méthode d'estimation	Nombre AFN Fin janv. 2015	Nombre de contrats actifs janvier 2015	Ecart observés -Prédits
Holt-Winters	258 924	1 166 760	4 %
Méthode déterministe	250 000	1 206 971	8 %
Observé	258 850	1 118 031	-

Tableau 40 : Comparaison des méthodes de projection

Cette comparaison nous permet de confirmer que l'estimation faite par Holt-Winter est cohérente avec ce qui existe actuellement et reste plus proche de l'observé.

D'après une enquête parue dans **l'ARGUS de l'assurance**[6]**,** les clients MRH sont les plus fidèles à leurs assureurs et seraient moins enclin à utiliser des comparateurs de prix sur internet pour faire leurs souscriptions. S'ils venaient à garder ce comportement après la mise en application de la loi Hamon, l'impact de celle-ci sur notre portefeuille se mesurerait par la méthode élaborée ci-dessus sans aucune hypothèse supplémentaire.
Dans le cas contraire il serait nécessaire de modifier la construction des bases de modélisation pour observer uniquement la résiliation des contrats avec une ancienneté au moins égale à 1 an.

[6] L'ARGUS de l'Assurance, 19 Septembre 2014 N°7376, pages 38-39

Conclusion

Nous avons traité dans ce mémoire la modélisation du taux de résiliation du produit MRH en faisant d'abord une première modélisation basée sur la régression logistique. Pour ce faire nous avons choisi de traiter deux horizons de résiliation, la première qui n'est pas fixé à priori englobe donc tout l'historique de notre portefeuille soit 15 ans, et un horizon d'un an qui nous permet d'être dans l'environnement d'application de la loi Hamon.
L'objectif premier de ces modélisation était d'identifier les profils de clients les plus aptes à résilier leur contrats MRH. Les résultats obtenus confirment souvent des faits observés ou des idées intuitives sur le phénomène de résiliation. Par exemple nous avons pu voir que les clients en formule Jeune résiliaient plus que les autres, de même que les clients résidants en appartement résilient plus que ceux en maison.
Ensuite une nouvelle modélisation du taux de résiliation par les modèles de survie a été faite pour d'une part, confirmer les résultats de la régression logistique et d'autre part déterminer la durée de vie en portefeuille d'un contrat MRH de BPCE A.
Les résultats de la modélisation seront utilisés comme un outil d'aide à la décision lors du passage de la prime technique à la prime commerciale lors de la mise en place des tarifs annuels. L'idée étant pour les profils n'étant pas une cible marketing d'appliquer une évolution tarifaire différenciée suivant le risque de résiliation. Par exemple appliquer une majoration plus importante pour les profils « locataires » que pour les profils « propriétaires ».
Enfin nous avons fait une projection de portefeuille à horizon un an pour se mettre dans les conditions d'application de la loi Hamon. Pour ce faire nous avons d'abord estimé le nombre d'affaires nouvelles sur un an par Holt-Winters et ensuite appliqué le modèle de résiliation sur le portefeuille actif de janvier 2014. Cela permet d'une part de quantifier la perte éventuelle qui proviendrait de l'application de cette loi en faisant l'hypothèse plus ou moins forte que la hausse de la résiliation qui proviendrait de cette loi est contenue dans l'erreur de prédiction de notre modèle. Et d'autre part d'avoir une idée plus précise de la structure du portefeuille dans un an.
Nous avons mesuré l'impact de la prime sur le risque de résiliation et globalement le résultat qui ressort dans les deux types de modélisation faites, est que la hausse de la prime n'augmente pas le risque de résiliation. Cependant ce résultat assez fort pourrait s'expliquer par le fait que l'information apportée par les variables tarifaires se retrouve dans la prime ce qui limite l'influence de celle-ci sur l'évènement résiliation en présence de ces variables suivant les modèles.
Pour une mesure plus exacte de cette influence une étude de la sensibilité de la prime par rapport à la résiliation serait nécessaire.

Bibliographie

CHARBONNIER François, GRUET Pierre [2012] « Étude de la sensibilité et de la déformation d'un portefeuille de prévoyance », Mémoire actuariat : http://www.ressources-actuarielles.net/

CHIHI Ossama [2011] « Sensibilité du taux de résiliation au prix en assurance MRH occupant et simulation du portefeuille », Mémoire actuariat : http://www.ressources-actuarielles.net/

FONTAINE Léonard [2011] « La modélisation de la valeur contrat (PNPV) par la refonte du modèle de résiliation », Mémoire actuariat : http://www.ressources-actuarielles.net/

LABIT HARDY Héloïse [2012] « Modélisation de la durée de vie des contrats d'assurance santé », Mémoire actuariat : http://www.ressources-actuarielles.net/

LAGNOUX Agnès « Série chronologique », Université de Toulouse LE MIRAIL, ISMAG, http://www.math.univ-toulouse.fr/~lagnoux/Poly_renf.pdf

PLANCHET Frédéric [2014] « Modèle de durée, application actuarielle », 2- Statistique des modèles de durée paramétriques et semi-paramétriques http://www.ressources-actuarielles.net/

PLANCHET Frédéric [2014] « Modèle de durée, application actuarielle », 4-L'estimation non paramétrique en présence de données censurées : estimateur de Kaplan-Meier et estimateur de Nelson-Aalen, http://www.ressources-actuarielles.net/

SAPORTA Gilbert [2009] « Régression logistique et analyse discriminante: Comparaison théorique et Pratique », Conservatoire National des Arts et Métier, http://cedric.cnam.fr/~saporta/discriminante.pdf

SAINT PIERRE Philippe [2014] « Introduction à l'analyse des durées de survie», Université Pierre et Marie Curie

TUFFERI Stéphane [2010] « Data Mining et Statistique décisionnelle: l'intelligence des données », 3ième Edition

Table des figures

Annexes

Annexe 1:Taux de résiliation horizon non défini: Corrélation entre variables explicatives

Variable 1	Variable 2	Valeur absolue V Cramer
Zone Tarifaire	Zonier incendie	1
Zone Tarifaire	Zonier vol	1
Zone Tarifaire	Zonier dde	1
Avantage jeune	Formule	0,89513
Caisse	Zonier incendie	0,82623
Type de Produit	Formule	0,77409
Type de résidence	Type de Produit	0,70889
Franchise	Formule	0,68194
Type de résidence	Formule	0,59634
Nombre de Logement	Type de Résidence	0,59292
Nombre de pièces	Prime	0,56955
Nombre de sinistre corporel	Nombre de sinistre corporel réglé	0,55098
Type de Produit	Ancienneté	0,52424
Franchise	Type de Produit	0,51388
Caisse	Zonier vol	0,50273
Ancienneté	Formule	0,49111
Capital mobilier	Formule	0,47989
CAISSE	Zonier dde	0,46393
formule	Qualité juridique	0,46171
formule	Prime	0,45719
Type de Produit	Capital mobilier r	0,45079
Type de Produit	Zonertarifaire	0,44841
Nombre de Logement	Type de Produit	0,44649
Nombre de pièce	Type habitation	0,44081
Mesure de Surveillance	Nombre de sinistre réglé	0,43816
Type habitation	Qualité juridique	0,4215
Caisse	Zone tarifaire	0,42002
Zonier dde	Zonier vol	0,40958
Type de Résidence	Zone Tarifaire	0,38307
Type habitation	Prime	0,36249
Franchise	Zone Tarifaire	0,35647
Prime	Qualité juridique	0,3546
Ancienneté	Capital mobilier	0,35194
Nombre de pièces	Qualité juridique	0,34024

Type Produit	Qualité juridique	0,33996
Type habitation	Zonier Tarifaire	0,32465
Zone Tarifaire	Formule	0,3146
Franchise	Ancienneté	0,31118
Zone Tarifaire	Qualité juridique	0,30423
Nombre de Logement	Qualité juridique	0,30413
Type de résidence	Qualité juridique	0,30077
Nombre de logement	Prime	0,29544
Type de Produit	Prime	0,29073
Nombre de logement	Formule	0,28758
Remise Commerciale	Prime	0,27729
Avantage jeune	Nombre de pièce	0,27574
Type habitation	Zonier vol	0,274
Franchise	Prime	0,26938
Avantage jeune	Type de Produit	0,26122
Type de résidence	Prime	0,254
Type habitation	Formule	0,25196
Zone tarifaire	Ancienneté	0,24457
Remise Commerciale	Formule	0,23564
Zonier vol	Formule	0,22876
Avantage jeune	Ancienneté	0,21526
Nombre sinistre corporel réglé	Nombre de sinistre réglé	0,21426
Type de résidence	Nombre contrat MRH	0,21179
Remise Commerciale	Nombre contrat MRH	0,20431
Nombre pièce	Zonier vol	0,20396
Zone Tarifaire	Prime	0,20375
Remise Commerciale	Nombre de pièce	0,2037
CAISSE	Type habitation	0,20096

Tableau 41 : Corrélations entre variables explicatives limité à 20%

Annexe 2:Optimisation modèle résiliation à horizon non défini

Le but de cette partie est de simplifier notre modèle tout en gardant sa robustesse .Pour ce faire nous faisons 2 modifications sur le modèle retenu:

- ✓ **Modification 1:** Suppression des variables **CAISSE**, **Nombre de sinistres sans suite**
- ✓ **Modification 2:** Suppression des variables CAISSE **Nombre de sinistres sans suite** et le **type d'habitation.**

Critère modèle	Modèle retenu	Modification 1	Modification 2
R2 Ajusté	0,37	0,3586	0,3581
D Sommer	0,615	0, 607	0,607
Analyse Type3 Var	Ok toutes <0,001	Ok toutes <0,001	Ok toutes <0,001
% obs concordante	80,7	80,4	80,3
% obs discordante	19,3	19,6	19,7
AUC (Aire courbe ROC)	0,807	0, 804	0,803
AIC	1667655,3	1686499,2	1687303,6
BIC	1668943,4	1687394,8	1688174,6
-2-LOG L	1667445,3	1686353,2	1687161,6

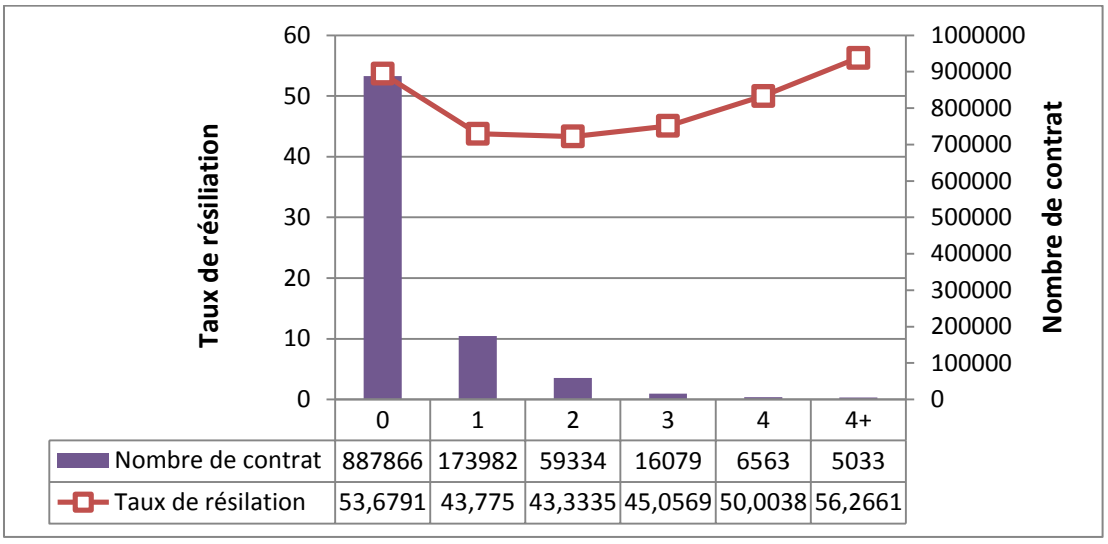
Tableau 42 : Comparatif des modèles obtenus après modifications

Au vu de ces résultats, les deux modèles obtenus après modification ont sensiblement la même performance, ainsi nous faisons le choix de retenir la première modification.
En effet la variable CAISSE a beaucoup de modalités qui ne peuvent être regroupés et comme présence du zonier peut déterminer la position géographique du client, sa suppression permet de simplifier notre modèle. La variable Nombre de sinistre sans suite reste la moins significative au vu de l'analyse de type 3, et par rapport à la variable Type d'habitation reste moins accessible.

Nous obtenons ainsi un modèle avec 14 variables explicatives. Afin que tous les paramètres de toutes les modalités des variables soient significatifs, nous ferons des regroupements de modalité sur le nombre d'avenants et nombre de contrats IARD.

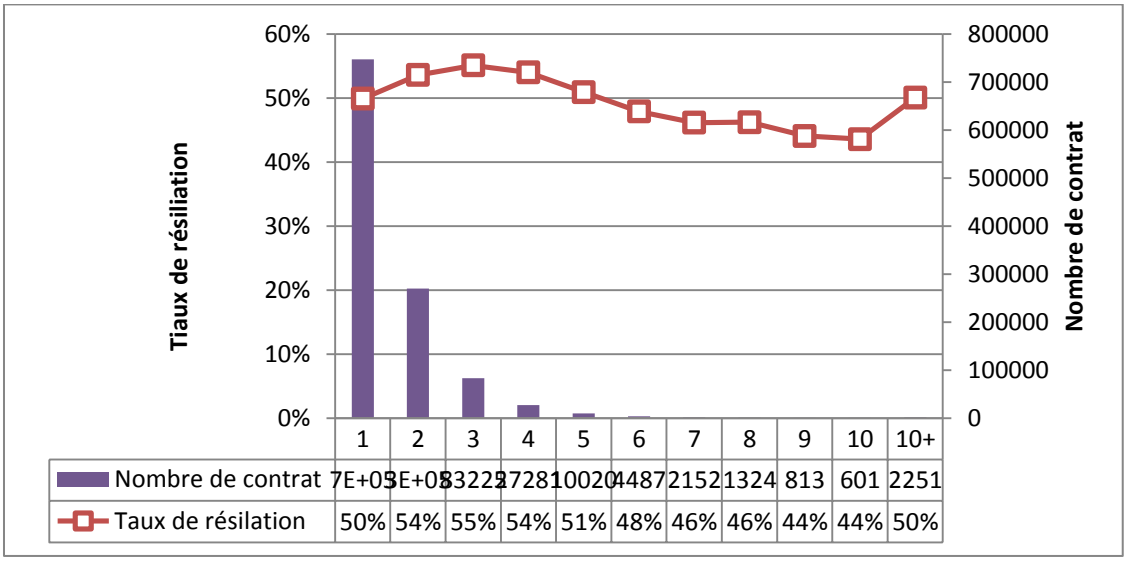
Annexe 3:Analyse taux de résiliation horizon non défini sur quelques variables

❖ **Nombre de contrat IARD**



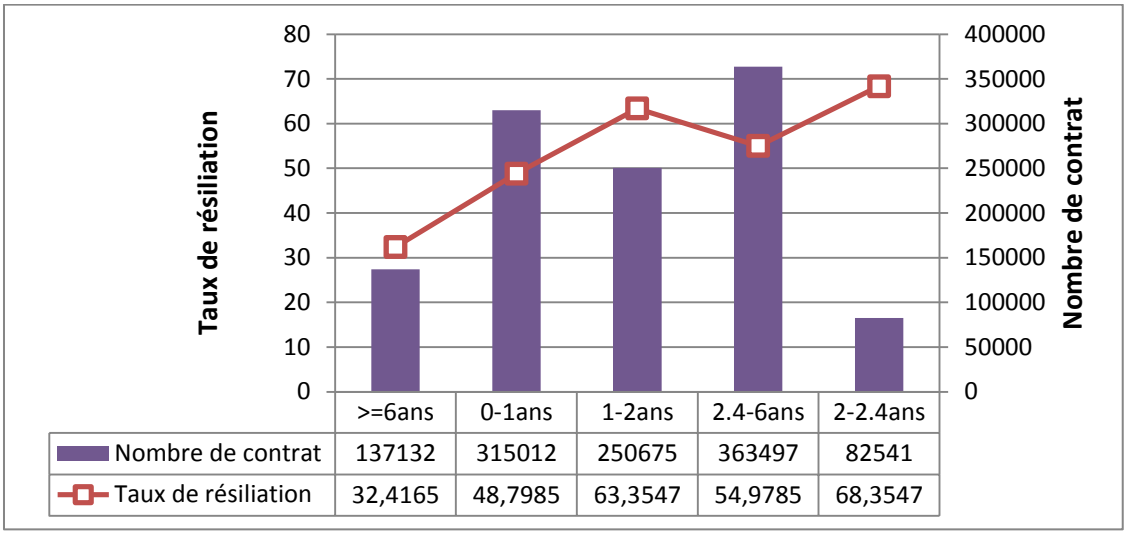
	0	1	2	3	4	4+
Nombre de contrat	887866	173982	59334	16079	6563	5033
Taux de résilation	53,6791	43,775	43,3335	45,0569	50,0038	56,2661

❖ **Nombre de contrat MRH**



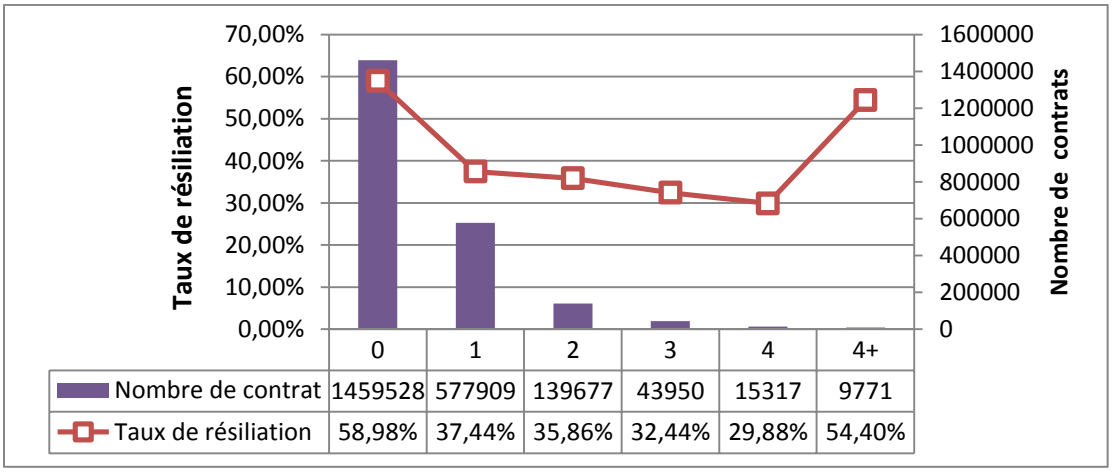
	1	2	3	4	5	6	7	8	9	10	10+
Nombre de contrat	7E+05	3E+05	83225	27281	10020	4487	2152	1324	813	601	2251
Taux de résilation	50%	54%	55%	54%	51%	48%	46%	46%	44%	44%	50%

❖ **Ancienneté du contrat**



	>=6ans	0-1ans	1-2ans	2.4-6ans	2-2.4ans
Nombre de contrat	137132	315012	250675	363497	82541
Taux de résiliation	32,4165	48,7985	63,3547	54,9785	68,3547

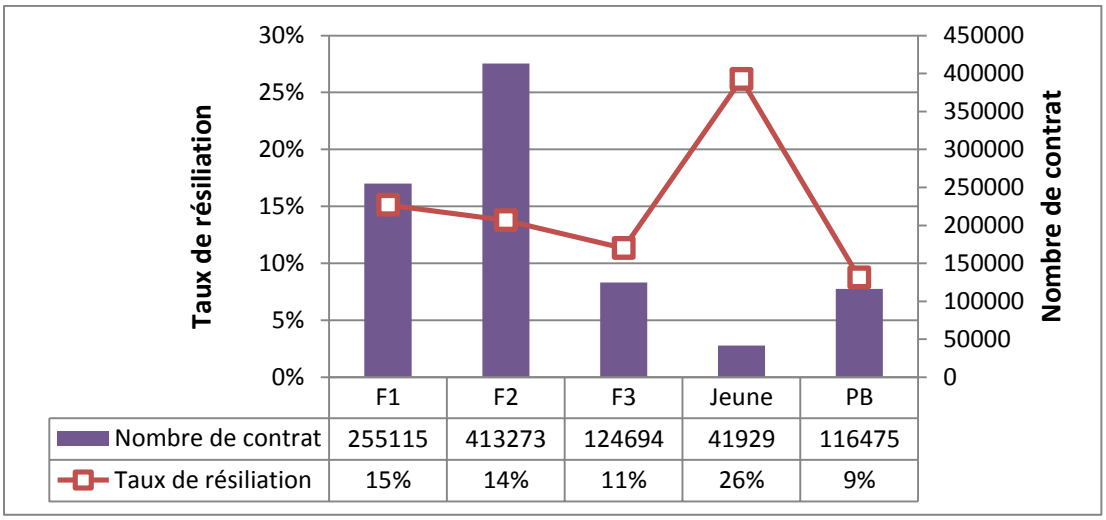
❖ **Nombre d'avenant**



	0	1	2	3	4	4+
Nombre de contrat	1459528	577909	139677	43950	15317	9771
Taux de résiliation	58,98%	37,44%	35,86%	32,44%	29,88%	54,40%

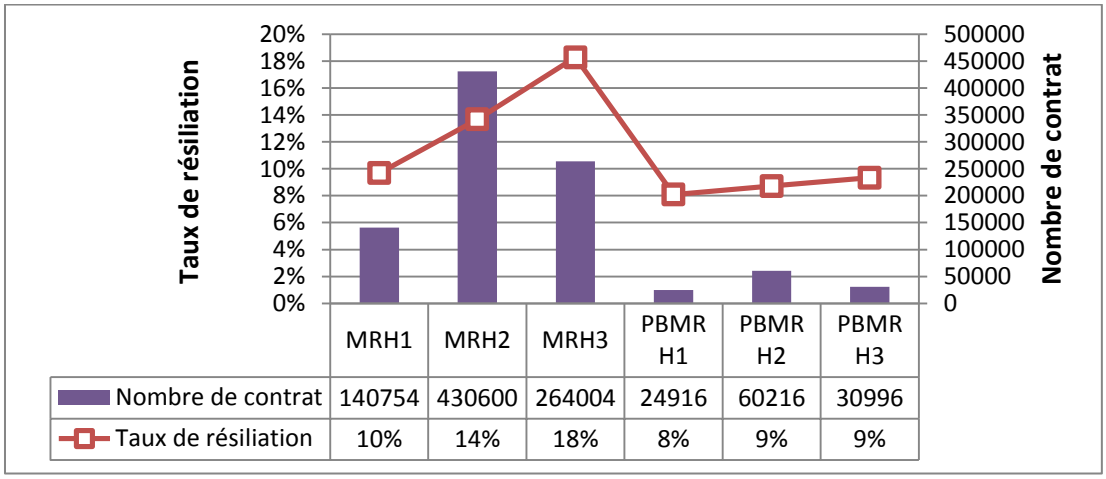
Annexe 4:Analyse taux de résiliation à un an sur quelques variables

❖ **Tri à plat formule**



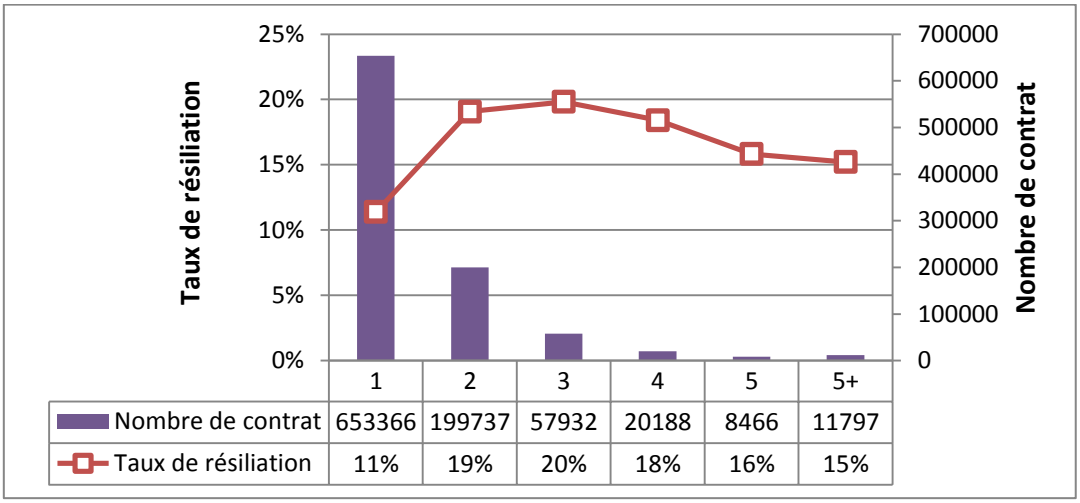
	F1	F2	F3	Jeune	PB
Nombre de contrat	255115	413273	124694	41929	116475
Taux de résiliation	15%	14%	11%	26%	9%

❖ **Tri plat produit**



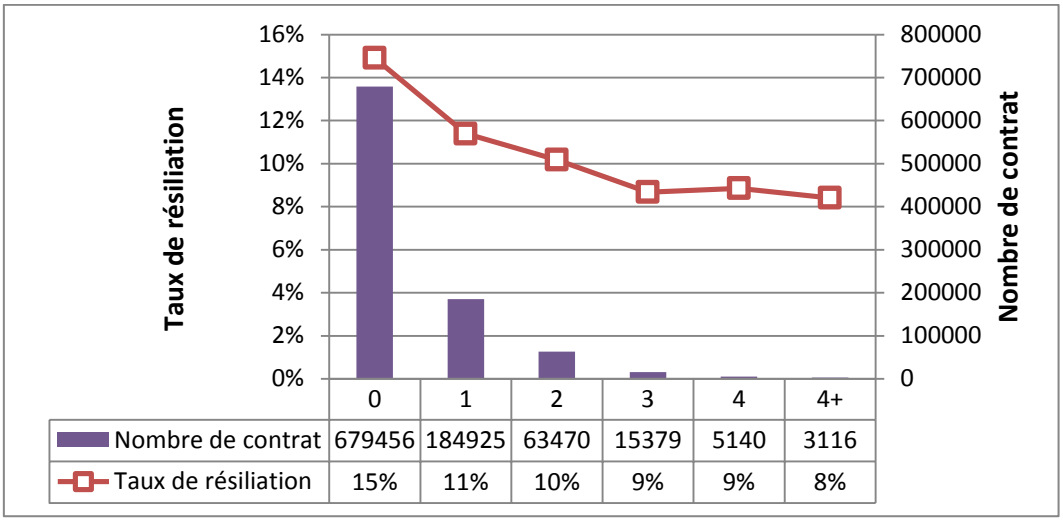
	MRH1	MRH2	MRH3	PBMR H1	PBMR H2	PBMR H3
Nombre de contrat	140754	430600	264004	24916	60216	30996
Taux de résiliation	10%	14%	18%	8%	9%	9%

❖ **Tri à plat nombre de contrat MRH**



	1	2	3	4	5	5+
Nombre de contrat	653366	199737	57932	20188	8466	11797
Taux de résiliation	11%	19%	20%	18%	16%	15%

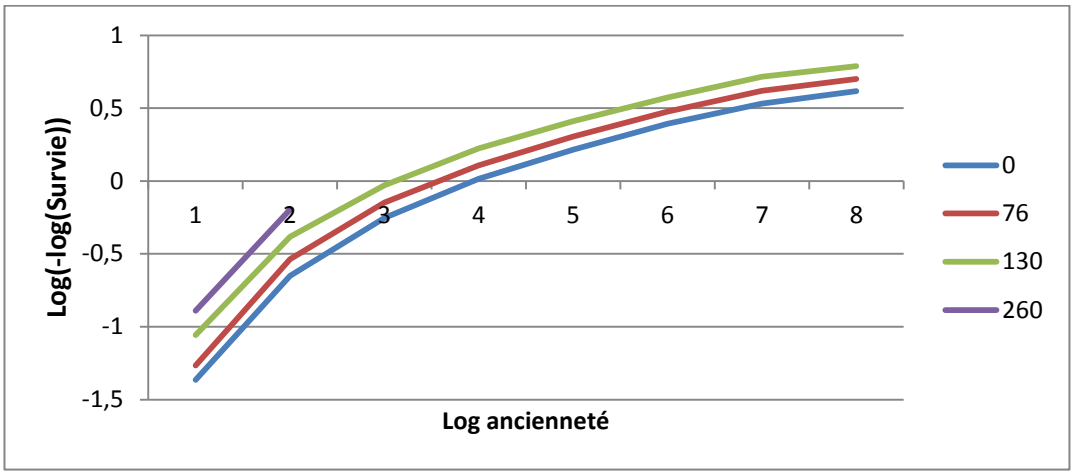
❖ **Tri à plat nombre de contrat IARD**



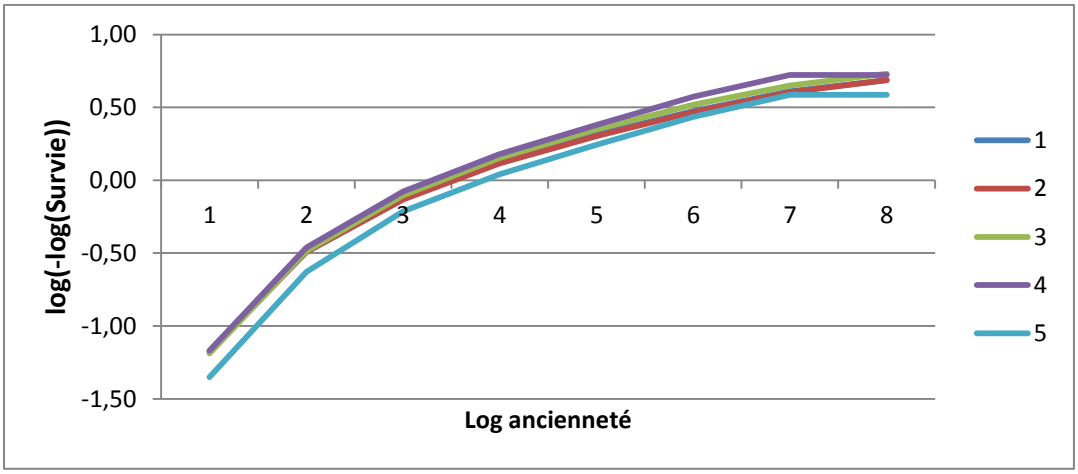
	0	1	2	3	4	4+
Nombre de contrat	679456	184925	63470	15379	5140	3116
Taux de résiliation	15%	11%	10%	9%	9%	8%

Annexe 5:Test graphique de proportionnalité sur les variables candidates

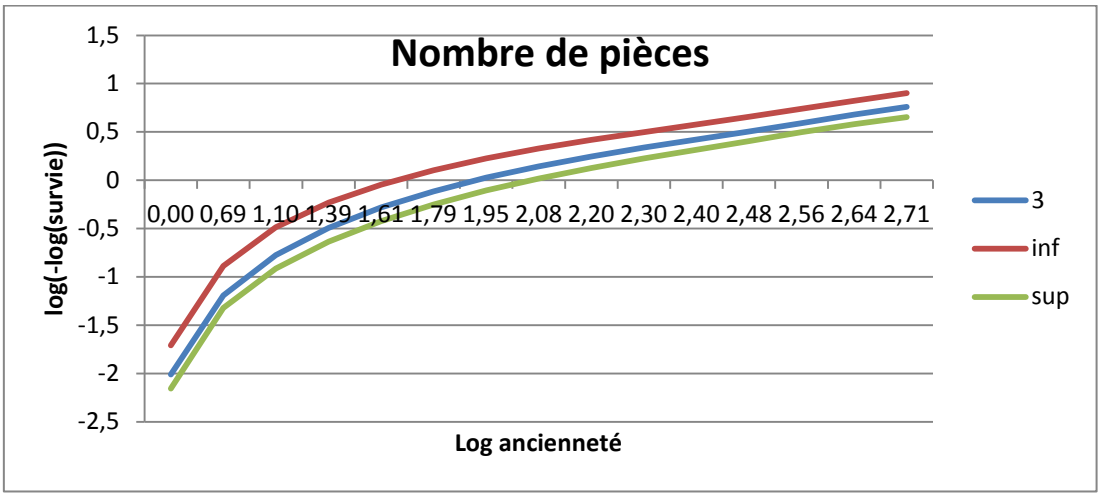
❖ **Franchise**



❖ **Zonier DDE**



❖ **Nombre de pièces**



Annexe 5:Tests statistiques

1. Test de Box Pierce

Ce test permet d'identifier les bruits blancs (ε_t) ; sa statistique permet de tester si $cov(\varepsilon_t, \varepsilon_{t-h}) = 0$ $\forall\ h$.Le test s'écrit :

$$\begin{cases} H_0 : \rho(1) = \rho(2) = \cdots = \rho(h) = 0 \\ H_1 : il\ esxiste\ i\ tel\ que\ \rho(i) \neq 0 \end{cases}$$

Pour effectuer ce test on utilise la statistique de Box-Pierce donnée par

$$Q_h = T \sum_{k=1}^{h} \widehat{\rho_k}^2$$

Où h est le nombre de retards, T est le nombre d'observation et $\widehat{\rho_k}$ l'autocorrélation empirique. Sous H_0 , Q_h suit asymptotiquement une loi de χ^2 à h degrés de liberté. Donc H_0 est rejeté si Q est supérieur au quantil d'ordre $(1-\alpha)$ de la loi χ_h^2 (α seuil d'erreur).

2. Test Shapiro-Wilk

Soit un échantillon $x = (x_1, x_2, \ldots . x_n)$, le test de Shapiro Wilk permet de tester si x suit une loi normale (H_0) ou non. La statistique du test est donnée par :

$$W = \frac{\left[\sum_{i=1}^{\left[\frac{n}{2}\right]} a_i (x_{(n-i+1)} - x_{(i)})\right]^2}{\sum_i (x_i - \bar{x})^2}$$

Où

- $x_{(i)}$ correspond à la série des données triées,
- $\left[\frac{n}{2}\right]$ est la partie entière du rapport $\left[\frac{n}{2}\right]$
- a_i constantes générés à partir de la moyenne et de la matrice de variance covariance des quantiles d'un échantillon de taille n suivant la loi normale.

La statistique W compare la forme de la distribution de l'échantillon à la distribution normale. Plus W est élevé, plus la compatibilité avec la loi normale est crédible. La région critique ou rejet de l'hypothèse nulle s'écrit $W < W_\alpha$.Les valeurs de W_α sont lues dans la table Shapiro-Wilk suivant la valeur de α et la taille de l'échantillon.
La probabilité critique ou p-value est calculé par $p - value = P(W_\alpha < W)$.
Donc si $p - value < \alpha$ on rejette H_0 sinon on garde l'hypothèse H_0.

Annexe 6: Décret loi Hamon

Décrets, arrêtés, circulaires

TEXTES GÉNÉRAUX

MINISTÈRE DES FINANCES ET DES COMPTES PUBLICS

Décret n° 2014-1685 du 29 décembre 2014 relatif à la résiliation à tout moment de contrats d'assurance et portant application de l'article L. 113-15-2 du code des assurances

NOR : *FCPT1410391D*

Publics concernés : *personnes physiques en dehors de leurs activités professionnelles, entreprises d'assurance.*
Objet : *conditions et modalités d'application du droit de résiliation à tout moment introduit par la loi n° 2014-344 du 17 mars 2014 relative à la consommation.*
Entrée en vigueur : *le présent décret entre en vigueur le lendemain de sa publication.*
Notice : *le décret vient préciser les conditions d'application du droit de résiliation à tout moment de contrats d'assurance, défini par l'article 61 de la loi n° 2014-344 du 17 mars 2014 relative à la consommation et codifié à l'article L. 113-15-2 du code des assurances. Il définit les branches dont relèvent les contrats auxquels s'appliquent ce nouveau droit et ses modalités d'exercice. En particulier, il organise son articulation avec les autres droits de résiliation déjà prévus dans le code des assurances, et il établit les modalités spécifiques de résiliation pour les contrats d'assurances mentionnés au quatrième alinéa de l'article L. 113-15-2 (contrats d'assurance de responsabilité civile automobile et de responsabilité locative).*
Références : *le présent décret est pris en application de l'article L. 113-15-2 du code des assurances. Le code des assurances modifié par le présent décret peut être consulté, dans sa rédaction issue de cette modification, sur le site Légifrance (http://www.legifrance.gouv.fr).*

Le Premier ministre,

Sur le rapport du ministre des finances et des comptes publics,

Vu le code des assurances, notamment son article L. 113-15-2 ;
Vu la loi n° 2014-344 du 17 mars 2014 sur la consommation, notamment son article 61 ;
Vu le décret n° 2011-144 du 2 février 2011 relatif à l'envoi d'une lettre recommandée par courrier électronique pour la conclusion ou l'exécution d'un contrat ;
Vu l'avis du comité consultatif de la législation et de la réglementation financières en date du 12 novembre 2014 ;
Le Conseil d'Etat (section des finances) entendu,

Décrète :

Art. 1er. – Le chapitre III du titre Ier du livre Ier du code des assurances est complété par deux articles ainsi rédigés :

« *Art. R. 113-11.* – Relèvent de l'article L. 113-15-2 les contrats d'assurance tacitement reconductibles suivants, couvrant les personnes physiques en dehors de leurs activités professionnelles :
« 1° Les contrats relevant des branches mentionnées au 3 ou au 10 de l'article R. 321-1, incluant une garantie responsabilité civile automobile définie à l'article L. 211-1 ;
« 2° Les contrats relevant des branches mentionnées au 8, au 9 ou au 13 de l'article R. 321-1, incluant une garantie couvrant la responsabilité d'un propriétaire, d'un copropriétaire ou d'un occupant d'immeuble ;
« 3° Les contrats relevant des branches mentionnées au 9, au 13, au 16 *c* ou au 16 *j* de l'article R. 321-1, constituant un complément d'un bien ou d'un service vendu par un fournisseur.

« *Art. R. 113-12.* – I. – Pour les contrats mentionnés à l'article R. 113-11, lorsque sont remplies les conditions de résiliation prévues à l'article L. 113-15-2, l'assureur applique les dispositions de cet article :
« 1° Lorsque l'assuré dénonce la reconduction tacite du contrat en application de l'article L. 113-15-1, postérieurement à la date limite d'exercice du droit de dénonciation du contrat ;
« 2° Lorsque l'assuré demande la résiliation du contrat en se fondant sur un motif prévu par le code des assurances dont l'assureur constate qu'il n'est pas applicable ;
« 3° Ou lorsque l'assuré ne précise pas le fondement de sa demande de résiliation.

« II. – Pour les contrats mentionnés à l'article R.113-11, dès réception de la demande de résiliation, que cette demande émane de l'assuré ou qu'elle soit effectuée pour le compte de ce dernier par le nouvel assureur selon les modalités définies au III, l'assureur communique par tout support durable à l'assuré un avis de résiliation l'informant de la date de prise d'effet de la résiliation, en application du premier alinéa de l'article L. 113-15-2. Cet avis rappelle à l'assuré son droit à être remboursé du solde mentionné au troisième alinéa de l'article L. 113-15-2 dans un délai de trente jours à compter de cette date.

« III. – L'assuré qui souhaite procéder à la résiliation de contrats visés au quatrième alinéa de l'article L. 113-15-2, en vue de contracter avec un nouvel assureur, en transmet la demande à ce dernier par lettre ou tout support durable. Dans sa demande, l'assuré manifeste expressément sa volonté de résilier son contrat en cours et de souscrire un nouveau contrat auprès du nouvel assureur. Ce dernier doit être en mesure de justifier de la demande qui lui est adressée par l'assuré, avant de procéder aux formalités prévues à ce quatrième alinéa.

« Le nouvel assureur notifie alors au précédent assureur la résiliation du contrat de l'assuré par lettre recommandée, y compris électronique. La notification mentionne le numéro du contrat, le nom du souscripteur, le nom du nouvel assureur choisi par l'assuré. Elle rappelle que le nouvel assureur s'assure de la continuité de la couverture de l'assuré durant l'opération de résiliation. La date de réception de la notification de résiliation est présumée être le premier jour qui suit la date d'envoi de cette notification telle qu'elle figure sur le cachet de la poste de la lettre recommandée ou, s'il s'agit d'une lettre recommandée électronique, sur la preuve de son dépôt selon les modalités prévues à l'article 2 du décret n° 2011-144 du 2 février 2011 relatif à l'envoi d'une lettre recommandée par courrier électronique pour la conclusion ou l'exécution d'un contrat.

« Le nouveau contrat ne peut prendre effet avant la prise d'effet de la résiliation de l'ancien contrat.

« Pour les contrats d'assurance mentionnés au 1° de l'article R. 113-11, lorsque l'assuré le lui demande, l'ancien assureur transmet dans les meilleurs délais, et au maximum dans un délai de quinze jours, au nouvel assureur le relevé d'information prévu à l'article 12 de l'annexe à l'article A. 121-1.

« IV. – Lorsque, pour les contrats visés au quatrième alinéa de l'article L. 113-15-2, la demande de résiliation est adressée directement par l'assuré à l'ancien assureur, ce dernier l'informe, par tout support durable, dès réception de cette demande, de son droit à résiliation dans les conditions prévues à ce même quatrième alinéa. »

Art. 2. – Le ministre des finances et des comptes publics, le ministre de l'économie, de l'industrie et du numérique et la secrétaire d'Etat chargée du commerce, de l'artisanat, de la consommation et de l'économie sociale et solidaire sont chargés, chacun en ce qui le concerne, de l'exécution du présent décret, qui sera publié au *Journal officiel* de la République française.

Fait le 29 décembre 2014.

MANUEL VALLS

Par le Premier ministre :

Le ministre des finances et des comptes publics,
MICHEL SAPIN

Le ministre de l'économie, de l'industrie et du numérique,
EMMANUEL MACRON

La secrétaire d'Etat chargée du commerce, de l'artisanat, de la consommation et de l'économie sociale et solidaire,
CAROLE DELGA