

Mémoire présenté devant l'Université de Paris Dauphine
pour l'obtention du diplôme du Master Actuariat
et l'admission à l'Institut des Actuaires
le

Par : Houweida JEMLI

Titre : Étude du risque de rachat de produits d'épargne italiens par des données agrégées et individuelles

Confidentialité : Non Oui (Durée : 1 an 2 ans)

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

*Membre présent du jury de l'Institut
des Actuaires :*

Entreprise :

Nom : Crédit Agricole Assurances

Signature :

*Membre présent du Jury du Master
Actuariat de Paris Dauphine :*

Directeur de Mémoire en entreprise :

Nom : Honorine CALIXTE

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)*

Secrétariat :

Signature du responsable entreprise

Bibliothèque :

Signature du candidat

Résumé

Un contrat d'assurance vie est un engagement mutuel de deux parties : l'assuré et l'assureur. Bien que la décision de l'assureur de couvrir le risque soit généralement irrévocable, l'assuré jouit des options qui lui permettent de suspendre, réduire ou résilier la police en récupérant totalement ou partiellement le capital assuré. Ces options légales et contractuelles sont appelées options de rachat et présentent un risque pour l'assureur.

Le risque de rachat est un risque de souscription lié au comportement de l'assuré. Nous distinguons deux types de rachat. Le premier, appelé rachat structurel, est lié au besoin immédiat de liquidité par l'assuré. Le deuxième est le rachat conjoncturel ou dynamique. Ce dernier est induit par le comportement des assurés envers les variations des taux servis par l'assureur par rapport aux taux offerts sur le marché. En assurance vie, le risque de rachat est l'un des risques de souscription les plus conséquents. En effet, en cas d'un flux de rachat massif non prévu ou une mauvaise estimation de ce risque, l'assureur sera confronté à une sortie de liquidité importante.

Ce mémoire présente l'étude du risque de rachat. Il s'inscrit dans le cadre de révision des hypothèses liées au rachat de différents produits d'épargne et de prévoyance.

Dans ce mémoire, nous étudions le rachat par les estimateurs statistiques classiques et par l'approche de machine learning. Notre étude sera focalisée sur la loi et les hypothèses de rachats utilisées pour la projection des produits d'assurance vie des portefeuilles italiens. Nous utilisons dans notre étude des méthodes de machine learning à savoir les arbres de classification, l'algorithme XGBoost et le modèle de la régression pénalisée Elastic Net.

Mots-clés : assurance vie, rachat, XGBoost, arbre de classification, Théorie de crédibilité.

Abstract

An insurance policy is a contract between two parties: the insured and the insurer. Although the insurer can't, in most cases, renounce of his decision to cover the risk, the insurer can suspend, decrease or terminate the policy rights and recover fully or partially the sum insured. These legal or contractual options are called lapse options and present a risk to the insurer. The lapse risk is an underwriting risk linked to the behavior of the policyholder. Two types of lapse exist. The first, called the structural lapse, is linked to an immediate need for cash flow by the insured. The second is the dynamic lapse risk. The latter is the consequence of the policyholder's behavior toward the difference between the rates served by the insurer and the market rates. In life insurance, the lapse risk is one of the most significant underwriting risks. In fact, in the event of significant unexpected lapse rates or a poor estimation of these rates, the insurer will be facing a massive outflow of reserves. In this memoir, we will be studying the lapse rates using the classical statistical estimators and the approach of *machine learning*. We will be focusing on the assumptions and the methodologies used to study and forecast the life insurance products of an Italian portfolio. Therefore, we will put in use the machine learning methods such as the classification trees, the XGBoost algorithm and the penalized regression model.

Keywords: Life insurance, lapse, Surrender, Classification tree, XGBoost, Credibility theory, regression

Note de Synthèse

Objectifs

Ce mémoire constitue une étude du risque de rachat de la filiale italienne de Crédit Agricole Assurances, CA Vita. Cette étude a pour objectif la détermination de la loi de rachat des produits d'épargne de CA Vita et l'étude de l'évènement du rachat individuel. En fait, nous cherchons à estimer les taux de rachat de différents produits d'épargne, déterminer une méthode de lissage des taux bruts de rachat et remplacer la méthode d'extrapolation des courbes de rachat. Cette étude s'étend aussi sur l'étude du rachat individuel par la détermination des variables explicatives qui influencent le comportement des assurés liés au rachat et la prédiction de la probabilité individuelle de rachat. Contrairement à la France, la fiscalité italienne en matière d'assurance vie ne présente pas d'avantages considérables. Les deux pics observés auparavant sur les courbes de rachat des produits d'épargne en France sont inexistantes pour les courbes de rachats des produits d'épargne italiens. En fait, les variations des taux de rachat d'une ancienneté à l'autre sont dues majoritairement aux pénalités de rachats et aux options contractuelles qui diffèrent d'un produit à un autre. Sur le périmètre étudié, nous constatons que, pour la majorité des produits étudiés, le taux de rachat augmente brusquement pendant les premières années d'ancienneté et cela est dû au fait que le rachat, pour ce cas de produits, n'est pas permis avant une année d'ancienneté. Donc, après une année de suspension de l'option de rachat, les assurés qui ont une inclination pour racheter leurs contrats le font dès qu'ils sont capables à le faire.

Estimation des taux bruts de rachat

Le rachat est une option contractuelle qui donne à l'assuré la possibilité de récupérer totalement ou partiellement son capital investi dans la police d'épargne. L'étude de rachat est effectuée afin de construire la loi de rachat de chacun des produits du portefeuille. Les courbes de rachat construites sont utilisées pour la projection des produits d'épargnes pour effectuer les calculs des capitaux de solvabilité exigés. Le taux de rachat est étudié par des estimateurs non paramétriques issues des modèles de durée qui nous permettent de déterminer les taux de rachat d'un produit par ancienneté. Cela consiste à déterminer la fonction de survie du produit et à calculer les taux bruts de rachat de ce dernier. La fonction de survie nous sert comme outil pour déterminer la période d'observation ou encore les anciennetés à considérer pour le calcul des taux bruts de rachat.

L'estimateur de Kaplan Meier est l'estimateur le plus connu pour déterminer la fonction de survie. La définition de cet estimateur est assez intuitive et nous procure une formule mathématique de passage de la fonction de survie aux taux bruts de rachat. Nous utilisons cet estimateur comme outil de vérification et référence.

Nous utilisons aussi l'estimateur de Harrington et Fleming de la fonction de survie qui dérive de l'estimateur de la fonction du risque cumulé de Nelson Aalen. L'estimateur de Nelson Aalen estime le risque cumulé de rachat qui caractérise la probabilité cumulée de rachat pour chaque ancienneté. Nous utilisons cet estimateur pour des raisons de comparabilité et pour contrôler les résultats trouvés par l'estimateur de Kaplan Meier.

L'estimateur de taux brut de rachat le plus utilisé est l'estimateur de Hoem. C'est un estimateur facile à calculer et à implémenter. Pour chaque produit, nous déterminons le taux brut de rachat par ancienneté en se basant sur les montants rachetés et les montants totaux exposés au risque de rachat de chaque ancienneté.

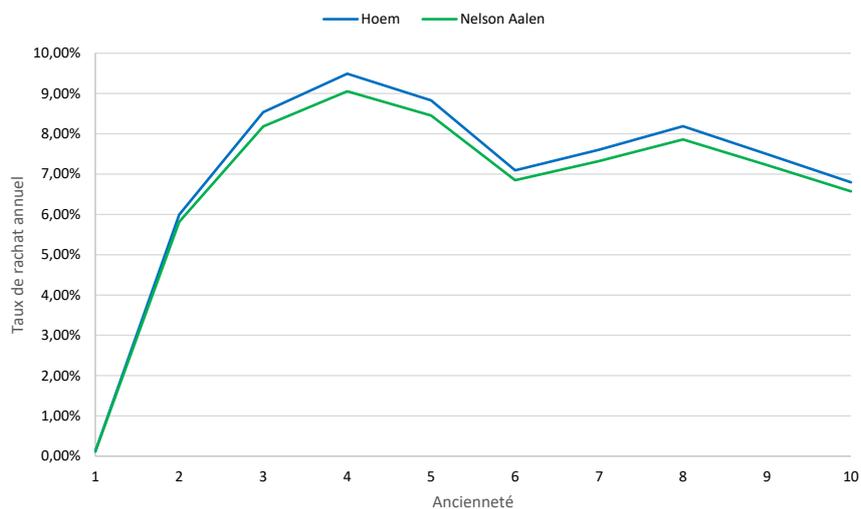


FIGURE 1 – Courbe des taux bruts de rachat estimés en fonction des anciennetés

L'estimation des taux bruts de rachat nous permet de construire une courbe de taux bruts de rachat par ancienneté pour chaque produit étudié. Ces courbes présentent souvent des fluctuations que nous devons ajuster et des points aberrants qu'on doit corriger. Nous utilisons ainsi les méthodes de lissage pour éliminer ces variations. Le lissage nous permet d'éliminer les fluctuations de la courbe de rachat tout en gardant la même allure et tendance de cette courbe. Un lissage ne doit pas être très prononcé c-à-d il doit garder l'allure des courbes ainsi que la signification des valeurs.

Lissage et extrapolation

Lissage

Afin de lisser les courbes de taux bruts de rachat, nous utilisons deux méthodes de lissage : le lissage par la moyenne mobile et la méthode de Whittaker-Henderson.

La moyenne Mobile La moyenne mobile est une méthode de lissage assez simple qui calcule les valeurs lissées comme étant une moyenne pondérée de valeurs consécutifs. La moyenne mobile garde la tendance de la courbe mais présente des limites pour le calcul des valeurs aux extrémités de la courbe à lisser.

Whittaker-Henderson Le lissage de Whittaker-Henderson est une méthode de lissage plus sophistiquée que la moyenne mobile. Cette méthode est simple à implémenter et elle est basée sur la minimisation de deux critères qui contrôlent la distance entre les valeurs lissées et les valeurs brutes et qui calculent la déviation des valeurs lissées par rapport aux valeurs brutes.

Lissage retenu En testant les deux méthodes de lissage sur le périmètre étudié, nous retenons le lissage de Whittaker-Henderson. En fait, le lissage par moyenne mobile conserve la tendance des courbes, mais tend à sous-estimer ou surestimer les taux de rachat en cas de fluctuations conséquentes de la courbe. Cette méthode peut donner des courbes peu lisses dans le cas où la courbe est très fluctuante. La méthode de Whittaker-Henderson est plus stable et garde un lissage assez satisfaisant pour n'importe quel type de courbe.

Extrapolation

Les courbes de rachat construites sont utilisées pour calibrer les hypothèses de rachat nécessaires pour effectuer les projections des produits d'épargne. Nous avons donc besoin de courbes de rachat pour des anciennetés allant jusqu'à quinze ans ou plus. Faute de données, nous utilisons les méthodes d'extrapolation pour compléter les courbes de rachat. La méthode utilisée par la filiale consiste à supposer que le rachat reste constant à partir de la dernière ancienneté observée et cette constante est égale à la dernière valeur de la courbe de rachat lissée. Cette hypothèse simpliste ne prend pas en considération ni l'historique des produits ni les spécificités du produit. Nous cherchons ainsi à améliorer l'extrapolation.

Nous disposons dans notre portefeuille de produits dont la courbe de rachat est assez complète c-à-d nous disposons de données sur le rachat jusqu'à quinze années d'ancienneté. Nous essayons donc à exploiter par la théorie de crédibilité ces données afin d'améliorer l'extrapolation pour les autres produits.

Théorie de crédibilité

La théorie de crédibilité repose sur le principe d'agrégation de données individuelles et collectives. Une estimation prend en considération l'historique individuel d'un produit et le comportement collectif de tout le portefeuille. Nous utilisons ainsi cette théorie pour estimer les taux de rachats des anciennetés non observées et extrapoler les courbes de rachat.

Nous testons trois modèles : Bühlmann, Bühlmann Straub et Hachemeister. Le modèle de Bühlmann donne une estimation sous forme d'une pondération entre le taux de rachat collectif et une estimation du taux de rachat individuel. Le modèle de Bühlmann Straub est une généralisation du modèle de Bühlmann. Nous affectons pour chaque produit et pour chaque ancienneté un poids spécifique qui

assigne un ordre d'importance à chaque observation. Le modèle de Hachemeister joint la régression à la théorie de crédibilité. Le coefficient de pondération est ainsi estimé par régression sur les années d'ancienneté. Ce modèle capte les tendances dans le temps de la courbe de rachat.

Nous testons les trois modèles sur notre portefeuille afin de déterminer le modèle le plus adapté aux spécificités du portefeuille étudié.

Modèle retenu Nous testons les modèles sur l'ensemble du périmètre. Nous présentons dans le tableau suivant les résultats pour trois produits du portefeuille.

Ancienneté	Produit B	Produit C	Produit D
1	0,00%	0,00%	1,20%
2	5,99%	2,51%	5,28%
3	8,54%	3,25%	4,84%
4	9,49%	4,65%	4,54%
5	8,83%	8,28%	5,71%
6	7,09%	9,45%	5,72%
7	7,61%	7,27%	4,41%
8	8,19%	2,95%	3,47%
Taux estimé par Bühlmann	6,84%	5,93%	6,64%
Taux estimé par Bühlmann Straub	7,02%	5,66%	5,87%
Taux estimé par Hachemeister	7,15%	4,01%	3,31%

TABLE 1 – Taux de rachat estimé par les trois modèles de la théorie de crédibilité

Le modèle de Hachemeister a été retenu après les résultats du Backtesting effectué sur tout le périmètre. Nous pouvons toutefois étudier les résultats en traçant les différentes courbes trouvées pour un des produits du portefeuille. Nous observons dans la figure suivante que le modèle de Hachemeister arrive à capter la tendance des dernières années du produit. En fait, après six années d'ancienneté les taux de rachat commencent à baisser. Les deux modèles de Bühlmann et Bühlmann Straub ne prennent pas en considération cette tendance et échouent à prédire un taux de rachat décroissant pour la huitième année d'ancienneté.

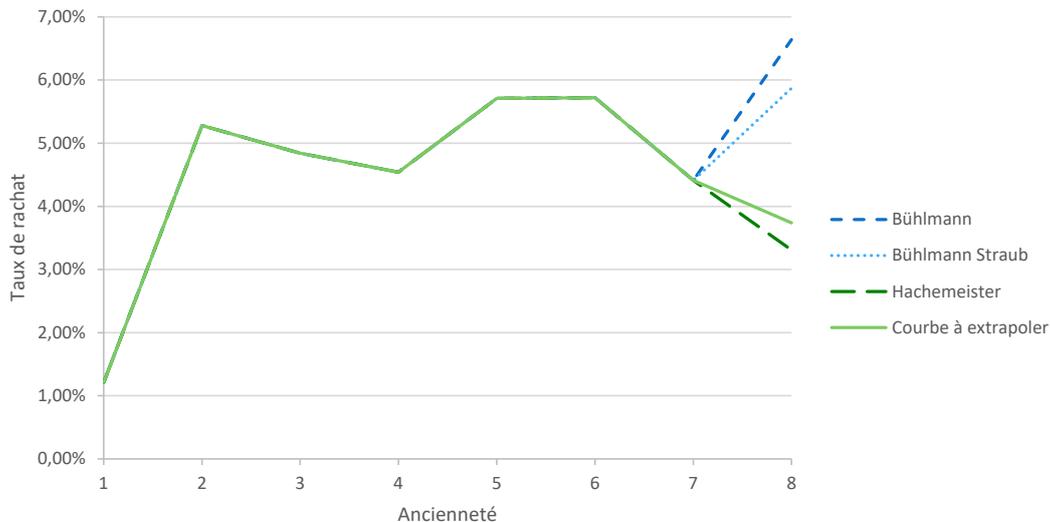


FIGURE 2 – Comparaison entre les différentes courbes extrapolées pour le produit D

Étude du rachat par les algorithmes du *Machine Learning*

L'étude du rachat par les estimateurs classiques nous ne permet pas de déterminer les variables qui influencent le comportement de rachat des assurés. Nous recourons aux algorithmes de *machine learning* qui nous permettent d'identifier les variables explicatives qui impactent le plus le comportement de rachat individuel. Ces algorithmes nous permettent aussi de prédire l'événement de rachat individuel et d'estimer une probabilité individuelle de rachat de chaque assuré du portefeuille.

Nous utilisons trois algorithmes : *Random Forest*, régression pénalisée *Elastic Net* et *XGBoost*. Nous déterminons que l'algorithme *XGBoost* est le plus performant pour prédire l'événement de rachat. Cependant, nous choisissons de retenir le modèle *Elastic Net* puisqu'il nous permet de mieux discerner les variables qui influencent le comportement des assurés.

Le tableau suivant compare les différents algorithmes de prédiction utilisés et présente les indicateurs de performance employés pour mesurer la performance de chaque algorithme.

Caractéristiques et indicateurs	Random Forest	Elastic Net	<i>XGBoost</i>
Pouvoir prédictif	Médiocre	Moyen	Bien
Interprétabilité	Moyen	Bien	Moyen
Temps de compilation	Médiocre	Moyen	Bien
Manipulation de données non pertinentes	Bien	Médiocre	Bien
TopDecileLift	5,61	6,643	7,154
LogLoss	0,44	0,156	0,053
AUC	0,96	0,97	0,996
RMSE	0,19	0,2	0,12

TABLE 2 – Tableau de comparaison des algorithmes de prédiction utilisés

Contribution des données individuelles L'exploitation des données individuelles, que l'assureur possède, pour prédire le rachat individuel permet d'une part d'affiner les estimations faites par les estimateurs statistiques et nous donne, d'une autre part, la possibilité de comprendre mieux le risque de rachat en maîtrisant mieux les variables qui influencent le comportement des assurés. Ces méthodes sont encore peu utilisées puisque le traitement de données qui précède l'application des modèles de prédiction est un traitement long et qui a un effet majeur sur la qualité des résultats. En fait, chaque modèle de prédiction est dépendant de la base de données qui a permis sa création. Ces algorithmes peuvent être aussi utilisés pour mieux commercialiser les contrats d'épargne et adapter la tarification au portefeuille d'assurés de la compagnie.

Synthesis note

Objectives

This memoir is a study of the lapse risk of the Italian subsidiary of *Crédit Agricole Assurances*, *CA Vita*. The purpose of this study is to determine the lapse risk distribution of the savings products of *CA Vita*. In fact, we want to estimate the lapse rates of the various products in the portfolio, determine a method of smoothing of the gross lapse rates and replace the method of extrapolation of the lapse curves. This study also extends to the study of the individual lapse event by determining the explanatory variables of the lapse event and the prediction of the individual probability of lapse.

Unlike France, the Italian tax regulations are not advantageous for the life insurance policies. The two peaks previously observed on the lapse curves of savings products in France are non-existent for the lapse curves of Italian savings products. In fact, variations in lapse rates are due mainly to lapse penalties and contractual options that differ from one product to another. An eye-catching observation is that, for most of the products studied, the lapse rate increases sharply during the first years of seniority and this is since the lapse, of the products we are studying, is not allowed before a year of seniority. So, the policyholders who are inclined to surrender their policies do so right after their lapse option is activated.

Estimation of gross lapse rates

The lapse option is a contractual option that gives the policyholder the possibility to fully or partially recover the capital invested in the savings policy. We are only studying the total lapse of *CA Vita*'s savings products. This study's object is to determine the lapse distribution of each product of the portfolio. The lapse curves determined are used for the projection of the savings products used for the calculation of solvency capital requirements.

We seek to determine the lapse rates by "amount", meaning that the data used represents the lapsed sums instead of the number of the lapsed policies. The lapse rates are estimated using non-parametric estimators that allows us to determine the lapse rates of each product by seniority. In practice we determine the survival function of the products and we calculate the gross lapse rates of these products. The survival function serves as a tool to determine the observation period or the seniorities taken into consideration to compute the gross lapse rates. In fact, by determining the survival function of a product we can determine the years of seniority which has sufficient and reliable data to

be used in the calculation of the gross lapse rates.

The Kaplan Meier estimator is the most known estimator used to estimate the survival function. The definition of this estimator is quite intuitive and provides us with a mathematical formula that allows us to determine the gross lapse rates. We use this estimator as a reference and to verify our calculations. We also use the Harrington and Fleming estimator of the survival function that derives from the cumulative hazard estimator of Nelson Aalen. We use this estimator to check our calculation by the Kaplan Meier estimator and as a way of comparison.

The most used estimator of rough lapse rates is the Hoem estimator. This estimator is easy to compute. For each product, we determine the lapse rates by seniority using both the date on the lapsed sums and the exposed sums for each given seniority.

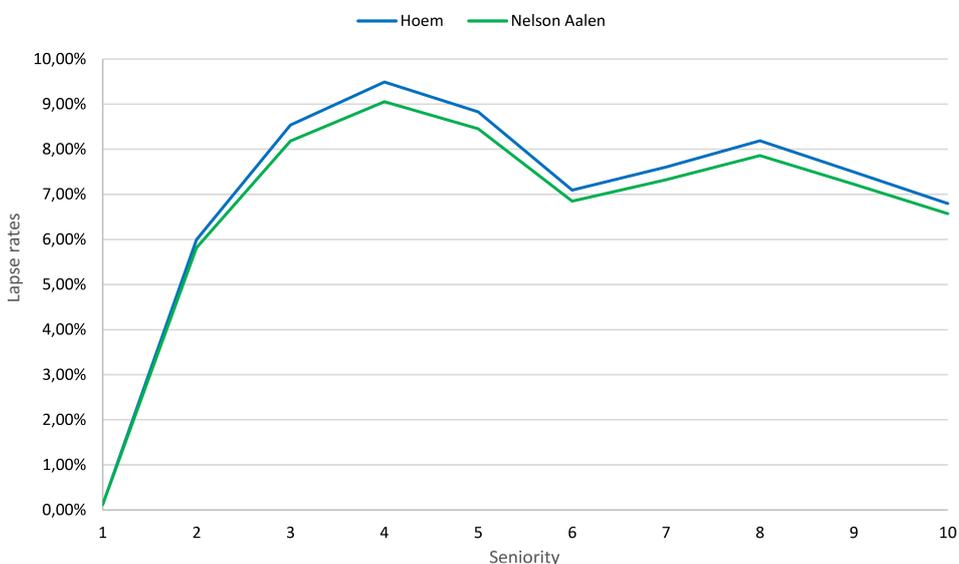


Figure 3 – Rough lapse rates curve estimated as a function of seniority

Estimating gross lapse rates allow us to construct a lapse curves by seniority. These curves have fluctuations and aberrant values that must be adjusted. Therefore, we use smoothing methods to eliminate these peculiarities. Smoothing allows us to eliminate the fluctuations of the lapse curve while keeping the same shape and trend of the curve.

Smoothing and et extrapolation

Smoothing

To obtain smoothed lapse curves, we use two smoothing methods: the moving average and the Whittaker-Henderson method.

The moving Average The moving average is a simple method that calculates the smoothed values as a weighted average of a series of consecutive values. The moving average conserves the trend of the curve but can't be applied to the values of the two ends of the curve.

Whittaker-Henderson The Whittaker-Henderson smoothing method is a more sophisticated smoothing method than the moving average. This method is easy to implement and is constructed around the minimization of two criteria that control the distance between the smoothed values and the raw values and that calculates the deviation of the smoothed values from the initial values.

Selected method After testing the two smoothing methods on the study perimeter, we retain the Whittaker-Henderson smoothing. In fact, the moving average smoothing method maintains the trend of the curves but tend to underestimate or overestimate the lapse rates in the event of significant fluctuations in the curve. Moreover, the moving average method produced poorly smoothed curves if the curve has various variations. The Whittaker-Henderson smoothing produces well smoothed curves and the smoothed values are quite close to the initial values.

Extrapolation

The lapse curves are used to calibrate the lapse hypothesis required to make projections of savings products. We, therefore, need lapse rates curves that go up to fifteen years or more. Due to lack of data, we use extrapolation methods to complete the missing values. The method used by CA Vita is to assume that the lapse remains constant from the last seniority observed and this constant is equal to the last value of the smoothed lapse curve. This simplistic assumption does not consider the product's history or its specificities.

We have products in our portfolio with a complete lapse curve, that is, we have data on lapsed sums up to fifteen years of seniority. In order to take advantage of these data, we use the credibility theory that allows us to incorporate data of different products to estimate the lapse rates of unobserved seniorities.

Credibility Theory

The theory of credibility is based on the principle of aggregation of individual and collective data. An estimated value by the credibility theory takes into consideration the individual history of a product and the collective comportment of the entire portfolio. We use this theory to estimate lapse rates for unobserved seniorities and to extrapolate the lapse curves.

We test three models: Bühlmann, Bühlmann Straub and the Hachemeister model. The Bühlmann model gives an estimation in the form of a weighting between the collective lapse rate and an estimate of the individual lapse rate. The Bühlmann Straub model is a generalized version of the Bühlmann model where for each product and for each seniority a specific weight is assigned. The model of Hachemeister combines the regression model with the credibility theory. The weighting coefficients are thus estimated by a regression on the years of seniority. This model captures the trends of the lapse curve.

Selected Model We test the three models to determine the most suitable model for the specificities of our portfolio.

Seniority	Product B	Product C	Product D
1	0,00%	0,00%	1,20%
2	5,99%	2,51%	5,28%
3	8,54%	3,25%	4,84%
4	9,49%	4,65%	4,54%
5	8,83%	8,28%	5,71%
6	7,09%	9,45%	5,72%
7	7,61%	7,27%	4,41%
8	8,19%	2,95%	3,47%
Rates estimated by Bühlmann	6,84%	5,93%	6,64%
Rates estimated by Bühlmann Straub	7,02%	5,66%	5,87%
Rates estimated by Hachemeister	7,15%	4,01%	3,31%

Table 3 – Estimated lapse rates by the credibility theory models

We plot the extrapolated curves by the three models and the observed lapse curve. We present the results of the tests done on three products to predict the lapse rates of the eighth seniority year using the figure below.

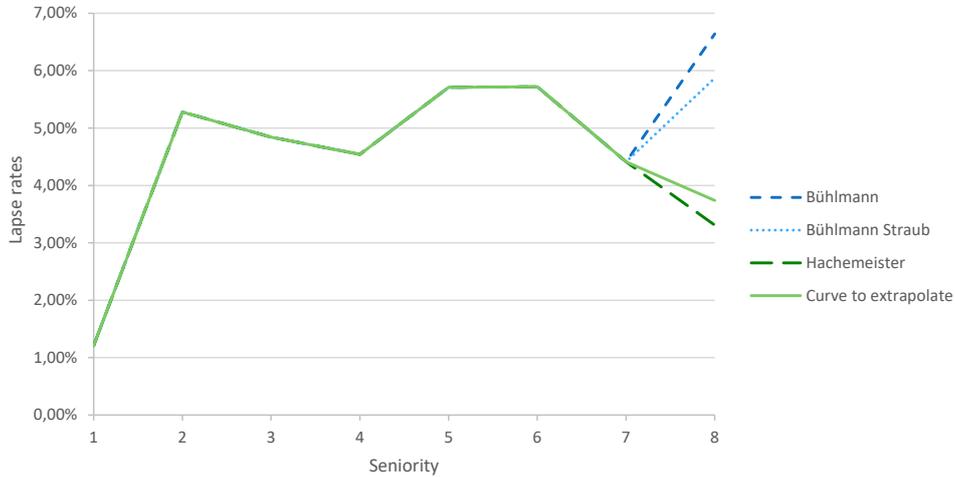


Figure 4 – Comparison of the different extrapolated curves for the product D

We observe that the Hachemeister model does integrate the trend of the curves in the prediction. This figure show that the Hachemeister predicts a value that is aligned with the last observed trend unlike the two other models. The Hachemeister model is selected among the three models tested after the backtesting results done on the whole portfolio.

Lapse study by the algorithms of *Machine Learning*

The study of the lapse by the classical estimators does not allow us to determine the variables that influence the lapse behavior of policyholders. We use the *machine learning* algorithms that allow us to identify the explanatory variables that have the greatest impact on individual lapse behavior. These algorithms also allow us to predict the individual lapse event and to estimate an individual probability of lapse for each insurer. We use three algorithms: *Random Forest*, penalized regression *Elastic Net* and *XGBoost*. We find that the *XGBoost* algorithm has the best performance in predicting the redemption event. However, we choose to keep the *Elastic Net* model since it allows us to better determine the

variables that influence the behaviour of policyholders.

The following table presents a comparison between the three algorithms used and the values of the performance indicators for each one.

Carateritics and performance indicators	Random Forest	Elastic Net	<i>XGBoost</i>
Predictive power	Poor	Average	Good
Interpretability	Average	Good	Average
Compilation time	Poor	Average	Good
Handling of irrelevant data	Good	Poor	Good
TopDecileLift	5,61	6,643	7,154
LogLoss	0,44	0,156	0,053
AUC	0,96	0,97	0,996
RMSE	0,19	0,2	0,12

Table 4 – Comparison table of the prediction algorithms

Contribution of individual data Using the individual data to predict the individual lapse event allows the insurer to improve its lapse rates estimations and to better understand the behavior of the policyholders which allows it to better control its lapse risk. However, these methods demand an enormous amount of time to prepare the data that will be used to build the prediction model. These models are dependent of the quality of the data used. We can use these algorithms to improve the customer targeting techniques and to improve the product pricing.

Remerciements

Je tiens à remercier l'équipe du service Actuariat et ALM de la direction de l'international de Crédit Agricole Assurances.

Tout d'abord, je voudrais remercier Monsieur BERGONZAT Antoine et Madame CALIXTE Honorine de leur aide et encadrement tout au long de mon stage.

Je souhaite ensuite adresser mes remerciements à Monsieur BEN YOUSSEF Mehdi, directeur de l'actuariat et du marketing international à Crédit Agricole Assurances Solutions, qui m'a donné l'opportunité d'effectuer mon stage au sein de son équipe.

Je remercie également mon tuteur académique Monsieur DUTANG Christophe , mon professeur Monsieur RAMMEH Hichem ainsi que Madame DHUIN Claudine pour leur suivie et leurs précieux conseils tout au long de mon stage.

Je voudrais enfin exprimer ma reconnaissance envers toutes les personnes qui m'ont soutenu et qui m'ont conseillé lors de la rédaction de ce mémoire.

Table des matières

Note de Synthèse	5
Synthesis note	11
Remerciements	17
Table des matières	19
Introduction	21
1 Assurance vie - Épargne	23
1.1 Épargne	23
1.2 Marché italien de l'assurance vie	25
1.3 Réglementation en Italie et autorités de contrôle	25
1.3.1 Les autorités régulatrices du secteur d'assurance en Italie	26
1.3.2 La fiscalité italienne	27
1.3.3 Types de produits en assurance vie	28
1.3.4 Le marché italien d'assurance vie	28
1.4 Données et approche	30
2 Loi de rachat : étude pour les données agrégées	33
2.1 Présentation	33
2.1.1 Définition du rachat	33
2.1.2 Étude du rachat	34
2.1.3 Notations et définitions	35
2.2 Estimateurs de la loi de rachat	36
2.2.1 Estimateur du taux brut de rachat	36
2.2.2 Estimateurs de la fonction de survie	36
2.2.3 Résultats et observations	39
2.3 Lissage	43
2.3.1 Moyenne mobile	43
2.3.2 Whittaker-Henderson	45
2.3.3 Résultats	48
2.4 Extrapolation	48

2.5	Théorie de crédibilité	50
2.5.1	Modèle de Bühlmann	50
2.5.2	Modèle de Bühlmann-Straub	53
2.5.3	Modèle de Hachemeister	55
2.5.4	Résultats	59
3	Prédiction du rachat par les données individuelles	63
3.1	Démarche de la prédiction par <i>machine learning</i>	63
3.2	Analyse de données	66
3.3	Modèles linéaires généralisés	72
3.3.1	Modèles linéaires	72
3.3.2	Modèles linéaires généralisés	73
3.4	Méthode d'agrégation d'arbre : Random Forest	75
3.4.1	Arbre de régression et de classification : CART	75
3.4.2	Random Forest	78
3.5	<i>eXtreme Gradient Boosting</i>	80
3.5.1	Les méthodes <i>Boosting</i>	80
3.5.2	<i>Gradient Boosting</i>	81
3.5.3	<i>eXtreme Gradient Boosting</i>	81
3.6	Optimisation des paramètres des modèles	82
3.7	Résultats	83
3.7.1	<i>Glmnet - Elastic Net</i>	83
3.7.2	<i>Random Forest</i>	85
3.7.3	<i>XGBoost</i>	85
3.7.4	Algorithme retenu	86
	Conclusion	89
	Bibliographie	91
	Annexes	93

Introduction

En assurance vie, les assureurs font face à divers types de risque lié aux spécificités de leurs produits. Les produits d'assurance vie comme les produits d'épargne, de prévoyance ou les produits décès sont liés à des risques spécifiques à l'assurance vie comme le risque de longévité, le risque de mortalité ou encore le risque de rachat. Nous nous intéressons au risque de rachat qui peut trop peser sur la solvabilité des assureurs vie.

Le risque de rachat, comme défini dans la partie 1 des spécifications techniques de la phase préparatoire de l'EIOPA[EIOPA, 2014], est *"un risque de perte et de déviation adverse dans le passif due à un changement dans le taux d'exercice attendu des options de l'assuré. Ces options sont des droits légaux ou contractuels des assurés à résilier la police totalement ou partiellement, abandonner, décroître, restreindre ou suspendre la couverture d'assurance ou permettre l'expiration de la police"*.

Étant un risque conséquent sur la solvabilité de l'assurance vie, le CEIOPA[CEIOPA, 2007] a recommandé dans "autres conseils à la commission européenne sur les questions du pilier 1 ", l'inclusion explicite des exigences en SCR (formule standard) pour le risque de rachat. Étant concernée par la directive Solvabilité II, CA Vita, la filiale italienne de Crédit Agricole Assurances, veille à bien modéliser le risque de rachat. Ce mémoire présente l'étude du rachat faite pour cette entité.

La modélisation du rachat repose sur la détermination de la loi de rachat de chaque produit du portefeuille de la filiale. Nous devons ainsi disposer de données agrégées de chaque produit pour plusieurs années ce qui peut présenter un défi. Nous pouvons aussi exploiter les données individuelles pour étudier le rachat.

Notre objectif est d'étudier les estimateurs utilisés pour la détermination des taux de rachat des produits d'épargne italiens. Nous présentons les contraintes rencontrées et les méthodes adoptées afin de construire des courbes de rachat complètes.

Nous présentons d'abord, dans le premier chapitre, le cadre de l'assurance vie en Italie. Ensuite nous exposons dans le deuxième chapitre les différents estimateurs des taux de rachat, les méthodes de lissage des courbes de taux bruts de rachat ainsi que la théorie de crédibilité utilisée pour extrapoler les courbes de rachat. Le troisième chapitre traitera les algorithmes de *machine learning* qui nous permettent d'exploiter les données individuelles des assurés pour affiner l'étude du rachat.

Chapitre 1

Assurance vie - Épargne

Ce chapitre présente le cadre d'étude des rachats. Notre étude de rachat sera menée sur un portefeuille italien de CA Vita, une filiale de Crédit Agricole Assurances en Italie. Nous présentons donc le marché italien de l'assurance vie ainsi que les entités régulatrices de ce marché.

1.1 Épargne

L'épargne est un contrat d'assurance en cas de vie présentant des garanties en cas de décès. Le contrat d'épargne constitue un placement fait au profit de l'assuré par l'assureur afin de faire fructifier un capital investi. C'est une protection sociale qui permet de se prémunir contre tous les aléas de la vie.

Un contrat d'épargne est caractérisé par les partis concernées, le type de supports d'investissement, le type de versement, le type de gestion et le type de garanties.

a) Acteurs concernés

Un contrat d'épargne implique différents partis concernées :

- L'assureur : c'est la compagnie d'assurance prestataire du service. L'assureur est la partis qui offre le contrat d'assurance et qui propose les modalités du contrat.
- Le souscripteur : c'est la personne qui souscrit le contrat d'assurance et qui s'engage à payer les primes en contrepartie du service proposé par l'assureur.
- L'assuré : c'est la personne sur laquelle repose le risque. En général, l'assuré est le souscripteur.
- Les bénéficiaires : ce sont les personnes qui recevront les prestations garantis en cas de décès de l'assuré. Ils sont désignées par le souscripteur.

b) Types de versement

Le versement des primes peut s'effectuer de différents façons :

- Versement unique : Le souscripteur verse à l'assureur une seule prime.
- Versement périodique : Le versement se fait de façon régulière et périodique (mensuellement, trimestriellement, annuellement, etc.)

- Versement libre : Le souscripteur verse une prime initiale lors de la souscription du contrat. Ensuite, il peut augmenter son épargne en versant des primes supplémentaires. Le souscripteur a la liberté de choisir quand et combien il veut verser.

c) Supports d'investissement

Nous distinguons deux types de supports d'investissement sur lequel contrat est investi.

- Support en euros : le contrat à support en euros est un contrat où le capital est placé dans un fonds en euros. Le fonctionnement de ces contrats est basé sur la capitalisation. Les intérêts générés sont ajoutés au capital et sont irrévocablement acquis par l'assuré. En plus du taux de rendement minimum garanti, le contrat en fonds d'euro verse d'éventuelles participations aux bénéficiaires. Le capital est donc sécurisé et l'assureur est donc le seul porteur du risque.
- Support en unités de compte : le contrat en unités de compte est un contrat d'épargne où l'investissement du capital est indexé à une ou plusieurs unités de compte. Aucune garantie n'est offerte sur le capital initial. L'assuré subit les moins-values et les plus-values du placement fait. À la maturité du contrat l'assuré reçoit le capital acquis. Ce type de contrat présente un rendement espéré plus élevé que les contrats en euro et un risque sur la valeur finale du capital pour l'assuré.
 - Index linked : ce sont des contrats dont la prime est investie dans des instruments dont le rendement est lié au rendement à un ou plusieurs indices du marché boursier.
 - Unit linked : ce sont des contrats dont la prime est investie dans des sociétés d'investissement à capital variable et lié à des fonds et des véhicules d'investissement.

Le nombre de support d'investissement d'un contrat d'épargne diffère d'un produit à l'autre. Nous distinguons deux types de contrats :

- Les contrats mono-support : un contrat mono-support est un contrat constitué exclusivement de fonds en euros ou des fonds en unités de compte.
- Les contrats multi-supports : l'assuré est libre d'investir dans un ou plusieurs types de fonds. Il peut ainsi placer une partie du capital dans des fonds en euros et placer le reste dans des fonds en unités de compte. Le risque est ainsi réparti entre l'assuré et l'assureur selon le pourcentage d'investissement choisi.

d) Types de gestion

Dans le cas des contrats d'épargne multi-support, le souscripteur a le choix entre différents types de gestion de leur investissement :

- Gestion libre : Le souscripteur a la liberté de choisir la proportion de placement dans chaque support. La gestion libre est, en général, proposée aux souscripteurs ayant une connaissance des marchés financiers et qui sont capables de choisir et adapter l'allocation optimale de leur épargne.
- Gestion pilotée : l'assureur propose aux souscripteurs différents profils d'investissement selon leur besoin et aversité au risque. Un souscripteur privilégiant le risque peut, par exemple, choisir une allocation majoritairement investie sur les actifs risqués.

- Gestion évolutive : c'est un type de gestion automatique qui s'adapte au profil de l'assuré. Une allocation initiale est choisie par le souscripteur. la répartition évolue ensuite en fonction de l'âge du contrat. Un contrat dont la maturité approche aura une proportion investie en fond euro la plus importante afin de sécuriser l'épargne.

e) Types de garanties

Un contrat d'épargne offre plusieurs options et garanties valable pendant la durée du contrat, par exemple l'option d'arbitrage, l'option de prorogation ou encore l'option de transformation du capital en rente. Le contrat d'épargne offre aussi des garanties en cas de décès de l'assuré :

- Garantie plancher : cette garantie permet de contrôler le risque de perte sur les unités de compte dans le cas où le décès survient quand le rendement des UC est assez faible. Cela garantie pour les bénéficiaires un capital minimum non affecté par l'évolution de la valeur des UC.
- Garantie majorée : les bénéficiaires reçoivent, en cas de décès de l'assuré, au minimum le capital investi ou le capital majoré d'un pourcentage diminué des frais.
- Garantie cliquet : option qui garantie aux bénéficiaires un capital égal à la valeur de rachat la plus haute atteinte par les versements à une date quelconque.

1.2 Marché italien de l'assurance vie

Le marché italien de l'assurance est constitué de 215 compagnies (dont 104 sont des filiales des compagnies européennes). Ce marché est assez développé et a un fort potentiel de croissance puisque les Italiens sont sous-assurés en assurance vie. En effet, en 2017 L'Italie est classée 4ème en Europe et 8ème mondialement dans le marché d'assurance en matière de recouvrement des primes.

En Italie la bancassurance présente le canal principal de distribution des contrats d'assurance vie. La fiscalité en Italie ne favorise pas l'assurance épargne et prévoyance. Ce marché est le deuxième marché du Crédit Agricole Assurances après la France.

1.3 Réglementation en Italie et autorités de contrôle

Le secteur de l'assurance et de la réassurance en Italie est régi essentiellement par trois cadres légaux :

- Le Code civil qui détermine les grandes lignes de l'assurance vie et non-vie ainsi que les règles générales des contrats d'assurance.
- Le Code d'assurance (décret législatif 209/2005) qui met en place le cadre légal d'exercice des différents fournisseurs de services d'assurance.
- La loi financière consolidée (décret législatif 58/1998) qui présente le cadre légal de fonctionnement des émetteurs et des distributeurs des produits d'assurance.

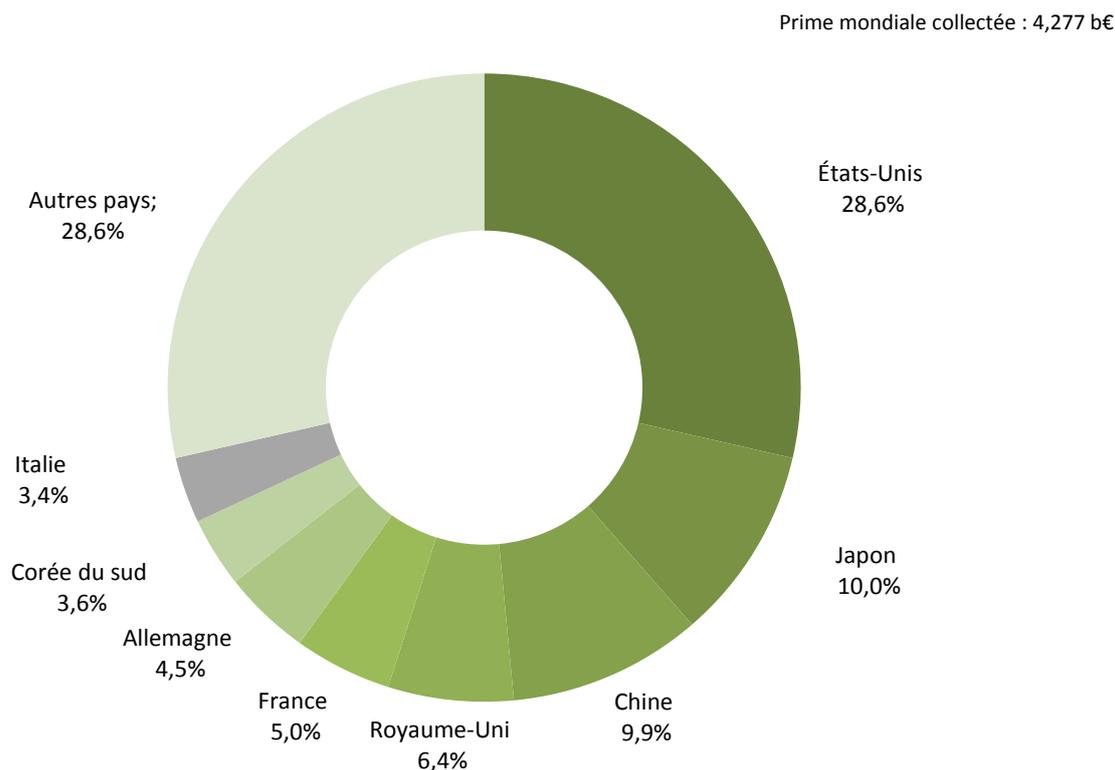


FIGURE 1.1 – Distribution mondiale des primes collectées par pays en 2016

Source : Swiss Re, Sigma n°3/2017 - 2016 data

1.3.1 Les autorités régulatrices du secteur d'assurance en Italie

En Italie le secteur d'assurance et de réassurance est réglementé par les lois italiennes et les lois européennes. Des corps légaux italiens supervisent en permanence pour le marché afin de garantir une application adéquate des réglementations. Ces entités de supervision protègent les droits des assurés et s'assurent d'une gestion prudente et adéquate des compagnies d'assurance et de réassurance.

a) L'institut de supervision des assurances : IVASS

L'institut de supervision des assurances représente le corps légal moniteur du marché italien d'assurance et de réassurance. L'institut est intégré dans la Banque d'Italie et elle est le successeur de l'ISVAP, l'ancien superviseur du secteur d'assurance italien, ce depuis 2013. L'institut surveille les compagnies d'assurance et de réassurance opérant en Italie et assure les droits des détenteurs de polices d'assurance. L'institut tient aussi le RUI (le registre unique des intermédiaires d'assurance et de réassurance) et veille à la bonne application de la réglementation par tous les acteurs du marché. Cette autorité assure la transposition et l'application des directives et des réglementations européennes publiées par l'EIOPA.

b) La commission nationale pour les entreprises et la bourse : CONSOB

La commission nationale pour les entreprises et la bourse est l'autorité italienne a pour vocation de surveiller le marché financier italien. Cette autorité supervise l'application par les entreprises d'assurances et de réassurance des réglementations lors de l'émission des produits à contenu financier.

c) L'association nationale des compagnies d'assurance : ANIA

L'association nationale des compagnies d'assurance est le représentant des compagnies d'assurance opérant en Italie. Cette association a été fondée en 1944 et regroupe 170 membres en 2017. Elle a pour objectif de veiller sur les intérêts des sociétés d'assurance tout en prenant en considération l'intérêt général du pays. Elle intervient ainsi sur différentes problématiques techniques, législatives ou encore économiques.

1.3.2 La fiscalité italienne

Les entreprises exerçant des activités en Italie sont soumises à un impôt sur les sociétés IRES (imposta sul reddito delle società) au taux de 24%. Un taux de production régionale IRAP (imposta regionale sulle attività produttive) est aussi appliqué au taux de 3.9% (le taux IRAP peut varier selon les régions et selon les types d'entités).

a) Taxe sur les conventions d'assurance

Les polices d'assurance vérifiant les exigences de territorialité fiscale (personne résidente en Italie ou risque assuré en Italie etc.) sont soumises à la taxe sur les conventions d'assurance.

L'assureur prélève cette taxe et est dans l'obligation de la réserver aux autorités fiscales italiennes. Le taux de prélèvement de la taxe sur les conventions d'assurance varie de 2.5% à 21.25% selon le type de risque couvert par la police. Cependant, les contrats d'assurance vie profitent d'une exemption spécifique.

Selon des situations spécifiques, comme l'emplacement de l'assureur, la taxe sur les conventions d'assurance doit être payée soit par l'assureur soit par l'assuré. Ici l'assureur peut être un assureur italien, une filiale italienne d'une assurance étrangère, une assurance étrangère opérant en Italie sous le régime de « liberté de services » ou un réassureur si la taxe n'a pas été payée par l'assureur.

b) Taxe sur les réserves mathématiques

Les assureurs appliquant la taxe de substitution de 26% sont sujets à une taxe de 0.45% sur les provisions mathématiques déclarées dans le rapport financier annuel. Une exemption est appliquée pour les provisions issues des polices couvrant les risques de mortalité ou d'invalidité ou certains fonds de pension.

c) Autres taxes

En souscrivant à une police d'assurance, les assurés sont sujets à une taxe de substitution à la taxe sur le revenu des particuliers qui s'élève à 26% sur les profits réalisés dans le cadre du rachat partiel

ou total de leurs contrats d'assurance (une réduction de 12.5% est appliqué sur la partie investie dans des obligations d'État ou dans des actifs équivalents).

Cette taxe de substitution est prélevée par les assureurs (assureurs italiens ou filiales italiennes des assureurs étrangers). Pour les polices souscrites auprès d'un assureur étranger opérant sous le régime de « liberté de services », la taxe doit être déclarée par l'assuré dans sa déclaration de revenus annuelle s'il n'y a pas un prélèvement par l'assuré.

d) Taxes diverses

Des contributions supplémentaires et des surtaxes peuvent être imposées aux assureurs.

Le code des assurances prévoit que les assureurs, commercialisant des polices d'assurance auto, doivent payer une contribution annuelle à la gestion des fonds de garantie pour les victimes d'accidents de la route.

e) Taxe sur la valeur ajoutée

Les transactions d'assurance et de réassurance ainsi que les services fournis par les intermédiaires des assurances et des réassurances sont exemptés de la taxe sur la valeur ajoutée.

1.3.3 Types de produits en assurance vie

Le secteur d'assurance vie italien est partitionné en 6 branches ou « Rami » :

- Ramo I : produits traditionnels (non-linked) comme les contrats vie entière ou les contrats vie temporaire. Ces produits constituent la part du marché italien d'assurance vie la plus importante.
- Ramo II : produits liés à la natalité et la nuptialité. Ces produits n'ont jamais été activés.
- Ramo III : produits en unités de compte UL (unit-linked) ou contrats sur indices IL (index-linked).
- Ramo IV : produits d'assurance de soins prolongés.
- Ramo V : produits de capitalisation.
- Ramo VI : retraite et pension.

1.3.4 Le marché italien d'assurance vie

Dans le cadre des taux d'intérêt faibles, les assurances font face à un défi important en matière de besoin en capital. Les assurances en Europe essaient de trouver un équilibre entre investir dans des actifs risqués, afin de garantir les rendements octroyés à leurs assurés, et garder un capital conforme aux exigences de la directive Solvabilité II.

En Italie, les taux garantis sont relativement bas, ce qui limite l'exposition au risque de taux. Toutefois, les rendements fournis aux assurés décroissent en permanence suite à la baisse des taux d'intérêt. La part des produits servant des taux entre 0% et 1% est passée de 29% à 46% entre fin 2014 et fin 2016*. Dans ce cadre des taux faibles, les assureurs commercialisent de plus en plus les produits UL et IL qui

*. Source : Banca d'Italia (Financial Stability Report, November 2016), Moody's Investors Service

ne présentent pas de risque pour les assureurs. Par conséquent, la part des contrats Ramo III dans le marché de l'assurance vie en Italie a augmenté considérablement ces dernières années.

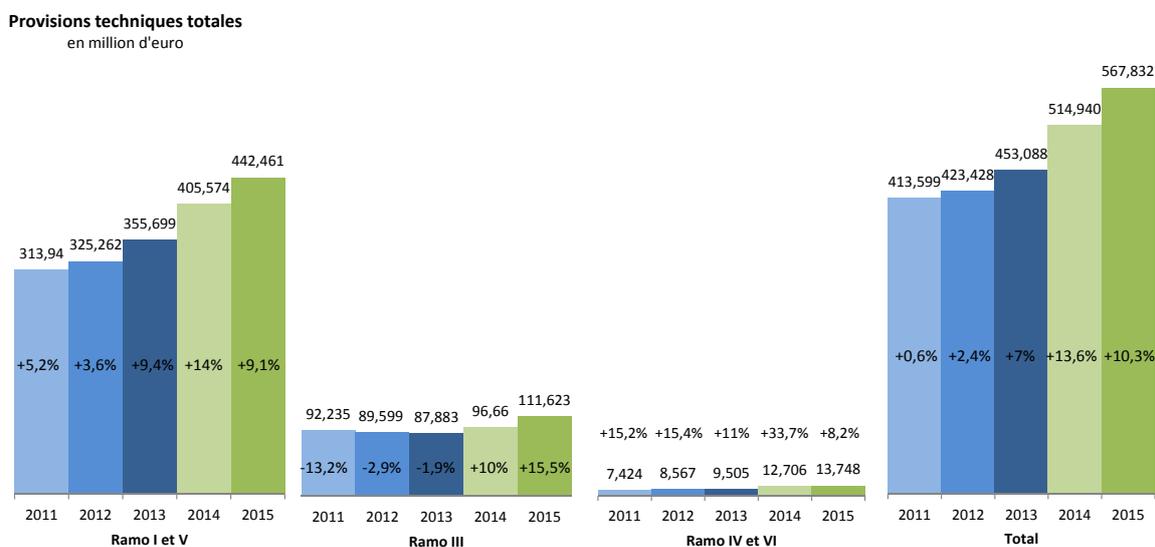


FIGURE 1.2 – Évolution des provisions techniques en Italie par branche

Le graphique ci-dessus présente l'évolution des provisions techniques totales d'assurance vie en Italie. La dominance des produits traditionnels est constatée sur le marché depuis 2011 mais la vitesse de croissance a diminué en 2015 pour passer de +14% en 2014 à +9.1% en 2015. Une croissance globale de 8.2% sur le marché d'assurance a été observée entre 2011 et 2015 ainsi qu'une augmentation des taux de croissance d'une année à une autre.

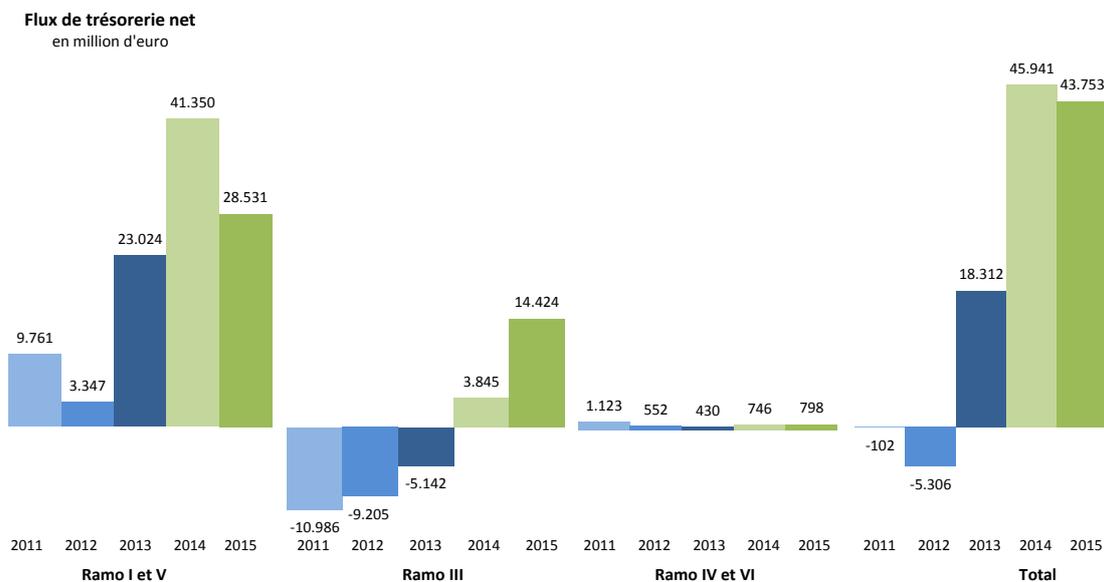


FIGURE 1.3 – Évolution du flux de trésorerie en Italie par branche

Nous observons dans le graphique ci-dessus les flux de trésorerie nets entre 2011 et 2015 classés

par types de « Ramo ». Le passage de -10.986 M€ à +14.424 M€ en 4 ans du flux de trésorerie des contrats du Ramo III montre la croissance de ces contrats dans le marché italien.

L'augmentation de la part des Ramo III dans le marché italien peut être expliquée par la convergence des assurés vers la commercialisation des produits multi-supports. Ces produits, joignant les produits classiques et les produits unit-linked et index-linked, présentent moins de risque pour l'assureur.

1.4 Données et approche

Dans ce mémoire, nous utilisons deux approches pour l'étude de rachat. Pour chaque approche, nous nous servons de données de type différents mais qui concernent les mêmes produits. Nous disposons de deux types de données : données agrégées et données individuelles. Nous utilisons les données agrégées pour déterminer la loi de rachat de chaque produit étudié. Les données agrégées concernent les montants totaux rachetés de chaque produit pour chaque année d'observation et pour chaque ancienneté. Les données individuelles servent à construire un modèle de prédiction. Ce modèle sert à prédire la probabilité de rachat de chaque assuré en fonction de ces caractéristiques.

La base de données agrégées reçues est construite à partir des données de 1995 jusqu'à 2017 et concerne dix produits d'épargne de CA Vita. N'est pas considéré comme rachat tout police résiliée pour une erreur lors de son émission ou résiliés dans les 30 jours après son émission les calculs faits par l'entité sont basés sur le montant de l'achat. Une étude fait au préalable par entité sur la pertinence des données et de la complétude de celle-ci a permis de déterminer la période d'observation à considérer pour chaque produit. Cette période diffère d'un produit à un autre selon le volume de données disponibles et peut être différentes de la période d'observation disponible.

Les dix produits étudiées sont des produits d'épargne en mono-support et en multi-supports. Les garanties de chaque produits sont différents ainsi que les taux de pénalité en cas de rachat. La plupart des produits sont à maturité fixe de 15 à 20 ans.

L'étude du rachat par les données agrégées nous permet de déterminer le taux de rachat pour chaque ancienneté de chaque produit étudié. Cela nous permet de prédire d'une manière globale les montants qui seront rachetés de chaque produit. Cette vision globale du rachat peut être améliorer si l'étude de rachat est plus segmentée c-à-d si nous estimons le rachat police par police. En fait, le comportement de chaque assuré varie non seulement en fonction du produit auquel il est souscrit mais aussi en fonction des caractéristiques de l'assuré (âge, catégorie socio-professionnelle, etc...). De ce fait, L'estimation du rachat individuel peut nous fournir des estimations plus affines et un aperçu sur le comportement de chaque assuré.

L'étude du rachat individuel est effectuée en utilisant les données individuelles des assurés. La base de données individuelle utilisée contient des informations décrivant les assurés, comme l'âge à la sortie ou le sexe, et des informations concernant la police souscrite par l'assuré comme l'ancienneté à la sortie et la performance du contrat. Nous utilisons ses données afin de construire un modèle de prédiction de l'événement de rachat individuel qui nous permet de déceler les variables qui expliquent au mieux

le comportement de rachat des assurés. Ces modèles servent aussi à calculer une probabilité de rachat pour chaque assuré du portefeuille. Une étude détaillée de la base utilisée est effectuée et présentée dans le troisième chapitre.

Chapitre 2

Loi de rachat : étude pour les données agrégées

Dans ce chapitre, nous présentons les différentes méthodes utilisées, pour le calcul des taux de rachat, dans le cadre de revue des hypothèses de rachats servant à la projection des produits d'épargne de l'entité CA Vita. Nous traitons également des méthodes de lissage de la courbe de rachat. La dernière partie de ce chapitre concerne la théorie de crédibilité. Cette théorie nous servira à répondre à un besoin spécifique à définir par la suite.

2.1 Présentation

2.1.1 Définition du rachat

Le rachat est une option contractuelle offerte au porteur de police d'assurance. Cette option donne le droit à l'assuré de résilier sa police, de récupérer totalement ou partiellement le capital investi ou encore le droit de réduire le montant des primes payées pour financer sa police.

Le rachat est un événement aléatoire lié aux comportements des assurés qui présente un risque éventuel pour l'assureur. Nous pouvons distinguer deux catégories de rachat : le rachat structurel et le rachat dynamique :

- **Rachat structurel** : c'est le rachat liés aux facteurs relatives aux caractéristiques des assurés et du contrat. C'est factures sont appelés facteurs structurels. Le rachat structurel peut être dû à un besoin de liquidité immédiat. La modélisation de ce type de rachat est faite sur la base du comportement passés des assurés. Nous citons ci-dessous quelques facteurs structurel :
 - Caractéristiques de l'assuré : âge, sexe, catégorie socio-professionnelle, nombre d'enfant, richesse, etc.
 - Caractéristiques de la police : ancienneté, performance, taux minimum garanti, réseau de distribution, périodicité de versement des primes, etc.
- **Rachat dynamique** : c'est le rachat induit par le comportement des assurés généralement en réponse à un écart constaté entre le taux servi par l'assureur et les taux offerts sur marché. Le

rachat peut être aussi dû à un changement de comportement des assurés suite à d'autres facteurs dites conjoncturels. Nous citons ci-dessous quelques uns :

- Fiscalité : un changement de la fiscalité peut affecter le comportement des assurés.
- Taux de rendement : les assurés peuvent racheter leur contrats si le taux servi chez un concurrent est plus intéressant.
- facteurs économiques : le contexte économique et politique peut affecter considérablement le comportement de rachat des assurés. Le taux de chômage, l'inflation, instabilité politique, etc.

Dans ce mémoire, nous étudions le rachat structurel. Nous nous intéressons au comportement passé des assurés et nous nous servons des estimateurs statistiques afin de déterminer les taux de rachat par année d'ancienneté.

Modélisation du rachat structurel

Le rachat structurel dépend de plusieurs paramètres (âge de l'assuré, ancienneté de la police, pénalité de rachat, etc.). Ce type de rachat exige une modélisation adéquate qui prend en considération la spécificité du périmètre étudié ainsi que le portefeuille d'assuré de l'assureur. Afin de bien modéliser ce type de rachat, nous devons en premier lieu définir le périmètre de notre étude. Nous serons ensuite amenés à étudier la qualité des données sur le périmètre sélectionné. La préparation de la base de données est une étape cruciale à l'étude puisque la qualité des données et la fiabilité de ces derniers affectent l'exactitude et la crédibilité des résultats.

Après l'analyse de données, nous choisissons l'estimateur à utiliser pour déterminer les taux bruts de rachat. Nous pouvons effectuer les calculs pour chaque produits ou nous pouvons déterminer des groupes homogènes d'assurés. Dans notre cas, l'étude est faite par produit puisque les résultats obtenues seront utilisé pour les projections des produits afin de calculer les exigences en capital de solvabilité. La détermination des taux bruts de rachat doit être suivi par l'application des méthodes de lissage. En fait, les courbes de rachat obtenus présentent souvent des pics irréguliers dû à des impuretés de données ou à des événements aberrants.

Les durées d'observation de l'historique des rachats étant souvent insuffisantes, nous utilisons des méthodes d'extrapolation qui nous permettent de prolonger la courbe de rachat. L'application d'une méthode d'extrapolation nous donne ainsi une courbe de rachat complète. La loi de rachat structurel est ainsi déterminée et peut être utilisé pour calibrer les modèles de projection.

2.1.2 Étude du rachat

Le contrôle de la qualité des données et la détermination des périodes d'observation de chaque produit sont effectuées au préalable. Nous nous disposons ainsi de données agrèges pour chaque produit. Nous cherchons dans un premier temps à déterminer la loi de rachat sur le périmètre étudié. Nous nous basons sur différentes méthodes d'estimation de rachat afin d'optimiser les hypothèses de rachat servant aux calculs Solvabilité II. Nous présentons les méthodes utilisées par l'entité ainsi que les méthodes applicables sur les données disponibles. Dans notre étude, le rachat est fonction de l'ancienneté. Nous calculons donc la probabilité de rachat d'un contrat pour chaque année d'ancienneté.

Analogie avec l'étude de la mortalité

Nous utilisons, de manière analogique, les méthodes d'estimation des taux de mortalité pour le calcul des taux de rachat. Nous supposons que nous étudions, au lieu de la durée de vie des assurés, la durée de vie des contrats. Nous exposons ainsi toutes les méthodes par rapport à leur application pour les probabilités de mortalité.

2.1.3 Notations et définitions

a) Fonction de survie S

Notons par F la fonction de répartition de la variable aléatoire étudiée X , représentant la durée de survie, et par S sa fonction de survie. Nous avons :

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= \mathbb{P}(X > t) \quad \forall t \geq 0 \end{aligned}$$

La fonction de survie peut être vue comme la probabilité d'un assuré (resp.d'un contrat) de "survivre" jusqu'à l'instant t (resp. l'ancienneté t).

b) Taux de hasard ou risque instantané λ

Par définition, pour t fixé, le risque instantané caractérise la probabilité de chute (décès ou rachat) juste après t , conditionnellement au fait d'avoir survécu jusqu'au temps t :

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(t \leq X < t + h | X \geq t)}{h} = \frac{f(t)}{S(t)} = -\ln(S(t))'$$

avec f la densité de probabilité de X . Si F est dérivable au point $t \geq 0$, alors $F' = f(t)$.

c) Taux de hasard cumulé Λ

Le taux de hasard cumulé Λ est défini comme l'intégrale du risque instantané λ :

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln(S(t))$$

Nous trouvons donc :

$$S(t) = \exp(-\Lambda(t))$$

Nous déduisons par cette expression que $H(X)$ suit la loi exponentielle de paramètre 1. En effet :

$$\mathbb{P}(H(T) \leq t) = \mathbb{P}(T \leq H^{-1}(t)) = S(H^{-1}(t)) = e^{-t}$$

d) Censure et troncature

Dans l'analyse des données de survie, nous faisons face souvent à des données incomplètes, censurées ou tronquées. Les censures et les troncatures sont des observations qui ne s'étalent pas sur la totalité de notre période d'étude et que nous devons prendre en considération.

Les données censurées sont les données dont la période d'observation effective est inférieure à la période d'étude. La sortie du sujet observé se fait avant la fin effective de la période d'observation considérée. Les données tronquées sont les données dont la date d'entrée en observation ne coïncide pas avec le début de la période d'étude.

2.2 Estimateurs de la loi de rachat

Nous désignons par t une année d'ancienneté et par i un produit étudié. Nous considérons une partition annuelle de la période d'observation. Nous supposons que t varie entre 1 et T et que i varie entre 1 et n .

Nous posons pour chaque ancienneté $t \in \llbracket 1 ; T \rrbracket$, les quantités suivantes :

r_t : nombre de contrats rachetés en t

c_t : nombre de contrats censurés sur la période $]t - 1, t]$

e_t : nombre de contrats tronqués sur la période $]t - 1, t]$

n_t : nombre de contrats "à risque" juste avant t (nombre de contrats exposés dans le portefeuille étudié après traitement des censures et des troncatures).

Nous avons :

$$n_t = n_{t-1} - r_{t-1} + e_t - c_t$$

2.2.1 Estimateur du taux brut de rachat

a) Estimateur de Hoem

L'estimateur de Hoem est un estimateur de taux brut. Il est défini, pour chaque ancienneté t , comme suit :

$$\hat{q}_t = \frac{Lapse_t}{Exposure_t} \quad \forall t > 0$$

avec :

$Lapse_t$: le montant ou le nombre de contrats d'ancienneté t rachetés

$Exposure_t$: l'exposition en montant ou en nombre en t . $Exposure_t = n_t + e_t - c_t$

2.2.2 Estimateurs de la fonction de survie

a) Estimateur Nelson Aalen du risque cumulé

L'estimateur de Nelson Aalen est un estimateur du taux de hasard cumulé Λ . Cet estimateur est donné par l'expression suivante :

$$\hat{\Lambda}_{NA}(t) = \sum_{k \leq t} \frac{r_k}{n_k}$$

Sa variance est estimée en utilisant la théorie de comptage et une approximation par la loi de Poisson :

$$\widehat{Var}(\hat{\Lambda}_{NA}(t)) = \sum_{k \leq t} \frac{r_k}{n_k^2}$$

b) Estimateur de Harrington et Fleming

Nous rappelons la relation entre le taux de hasard cumulé Λ et la fonction de survie S :

$$\Lambda(t) = -\ln(S(t)) \quad \forall t > 0$$

L'estimateur de la fonction de survie de Harrington et Fleming \widehat{S}_{HF} est dérivé de l'estimateur de Nelson Aalen. Il est donné par :

$$\begin{aligned} \widehat{S}_{HF}(t) &= \exp\left(-\widehat{\Lambda}_{NA}(t)\right) \\ &= \prod_{k \leq t} \exp\left(-\frac{r_k}{n_k}\right) \quad \forall t > 0 \end{aligned}$$

L'estimateur de la variance de \widehat{S}_{HF} est donné par la delta méthode comme suit :

$$\widehat{Var}\left(\widehat{S}_{HF}(t)\right) = \left(\widehat{S}_{HF}(t)\right)^2 Var\left(\widehat{\Lambda}_{NA}(t)\right)$$

c) Estimateur de Kaplan Meier

L'estimateur de Kaplan-Meier est construit sur l'idée suivante : survivre après un temps t est équivalent à être en vie juste avant t et ne pas mourir au temps t . Alors, q_t , la probabilité de décès à l'âge t ou encore la probabilité qu'un contrat soit racheté après t années d'ancienneté, s'exprime comme suit :

$$q_t = \mathbb{P}(X \leq t + 1 | X > t)$$

Par la formule de Bayes, nous obtiendrons :

$$\begin{aligned} q_t &= \frac{\mathbb{P}(t < X \leq t + 1)}{\mathbb{P}(X > t)} \\ &= \frac{F(t + 1) - F(t)}{S(t)} \\ &= 1 - \frac{S(t + 1)}{S(t)} \end{aligned}$$

Donc :

$$q_t = 1 - \frac{S(t + 1)}{S(t)} \quad \text{et} \quad S(t) = \prod_{k \leq t} (1 - q_k)$$

L'estimateur de Kaplan Meier prend en considération les censures et les troncatures et est donné par la formule suivante :

$$\widehat{S}_{KM}(t) = \prod_{k \leq t} \left(1 - \frac{r_k}{n_k}\right)$$

Estimateur de Greenwood de la variance de l'estimateur de Kaplan-Meier

Nous estimons la variance de l'estimateur de la fonction de survie de Kaplan-Meier par l'estimateur de Greenwood.

$$\widehat{Var}\left(\widehat{S}_{KM}(t)\right) = \widehat{S}_{KM}^2(t) \sum_{k \leq t} \frac{r_k}{n_k(n_k - r_k)}$$

Cette estimation est dérivée des résultats suivants :

$$\begin{aligned}\widehat{Var}(\log(\widehat{S}_{KM})) &\approx \sum_{k \leq t} \frac{r_k}{n_k(n_k - r_k)} \\ &\approx \frac{1}{\widehat{S}_{KM}(t)} Var(\widehat{S}_{KM}(t)) \quad (*)\end{aligned}$$

La formule (*) découle d'une part de l'hypothèse de normalité asymptotique de l'estimateur de Kaplan Meier et d'autre part de l'application du delta méthode pour $\widehat{Var}(\log(\widehat{S}_{KM}))$:

$$Var(f(Z)) \approx (f'(\mathbb{E}(Z)))^2 Var(Z)$$

L'intervalle de confiance (approximation par la loi normale dans le cas où la LFGN est applicable) de niveau de confiance $1 - \alpha$ de l'estimateur de Kaplan Meier est :

$$\left[\widehat{S}_{KM}(t) \pm q_{\alpha/2} \sqrt{\widehat{Var}(\widehat{S}_{KM}(t))} \right]$$

avec q_α le $\alpha^{\text{ème}}$ quantile de la loi normale centrée réduite. Si $\widehat{S}_{KM}(t)$ est très proche de 0 ou de 1, nous utilisons plutôt l'intervalle de confiance de Rothman :

$$\frac{K}{K + q_{\frac{\alpha}{2}}^2} \left[\widehat{S}_{KM}(t) + \frac{q_{\frac{\alpha}{2}}^2}{2K} \pm q_{\frac{\alpha}{2}} \sqrt{\widehat{Var}(\widehat{S}_{KM}(t)) + \frac{q_{\frac{\alpha}{2}}^2}{4K^2}} \right]$$

avec $K = \frac{\widehat{S}_{KM}(t)(1 - \widehat{S}_{KM}(t))}{\widehat{Var}(\widehat{S}_{KM}(t))}$

2.2.3 Résultats et observations

Contrairement à la France, il n'existe pas de fiscalité spécifique par période de rétention de la police d'assurance. Tous pics de la courbe de rachat ne sont pas liés à la fiscalité mais aux options contractuelles des produits comme la pénalité de rachat.

Nous présentons les courbes du taux brut de rachat des quatre produits estimés par la méthode de Hoem et Nelson Aalen ainsi que les courbes de la fonction de survie.

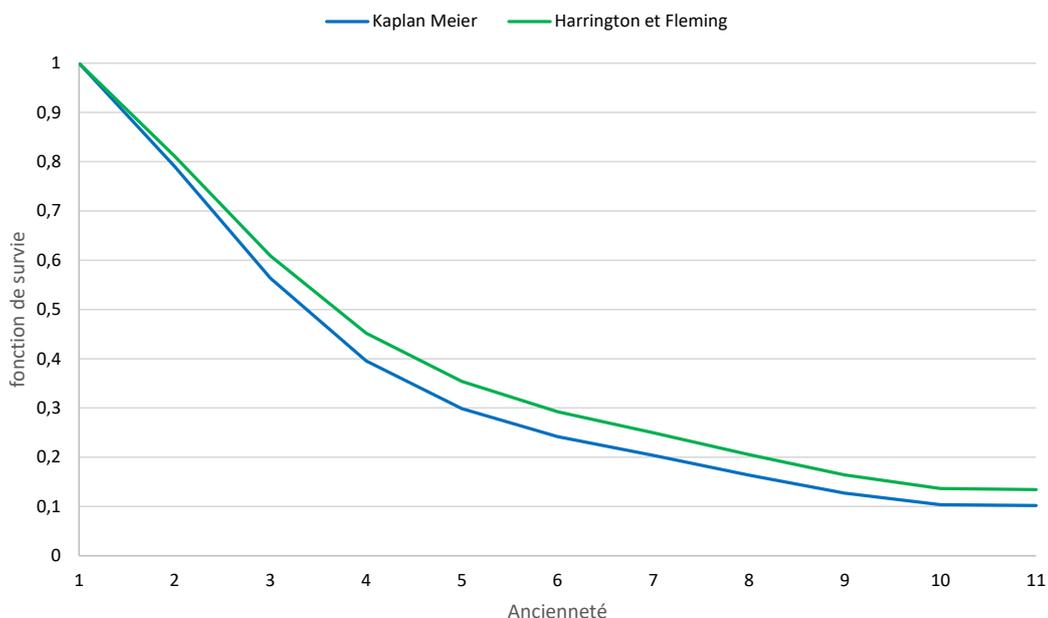


FIGURE 2.1 – Courbe de survie du produit A estimée par l'estimateur Kaplan Meier et Harrington et Fleming

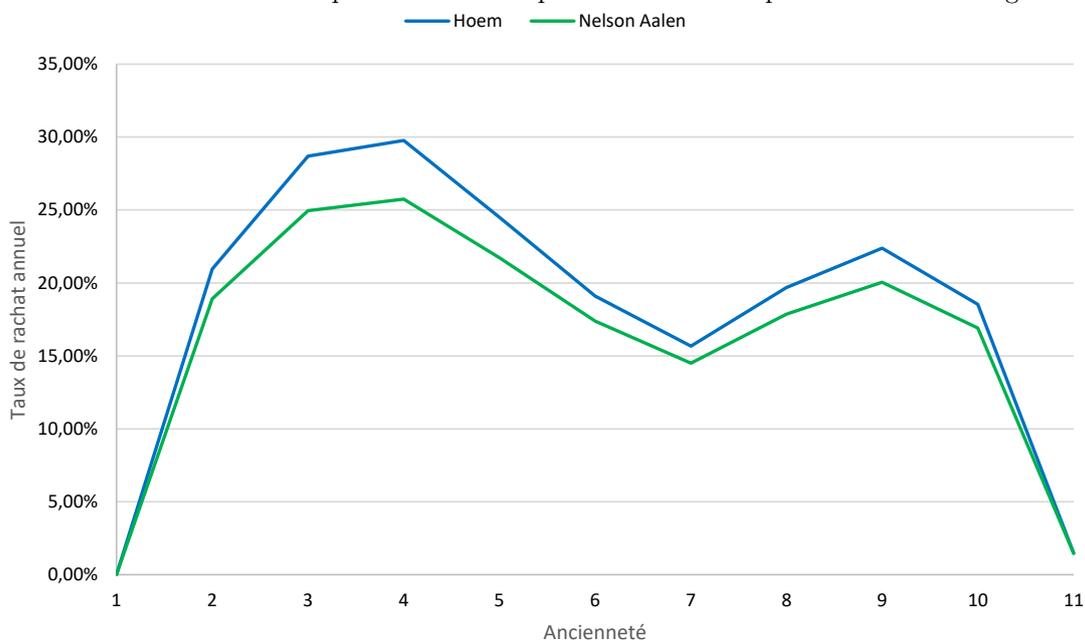


FIGURE 2.2 – Courbe de taux de rachat du produit A estimée par estimateur Hoem et Nelson Aalen

La fiscalité italienne n'est pas avantageuse pour les contrats d'assurance vie et d'épargne. Cela explique les taux assez élevés du rachat observé pour beaucoup des contrats étudiés. Le taux de rachat de la première année d'ancienneté du produit A est nul car ce produit n'autorise pas le rachat avant une année d'ancienneté. Pour ce produit, nous ne disposons que de données sur une période de douze années. La valeur de la dernière année a été négligée puisque le volume de données de la dernière année est faible et ils sont peu fiables. Comme le rachat n'est pas autorisé qu'après une année d'ancienneté, nous observons un pic de rachat des premières années de la vie de ce produit.

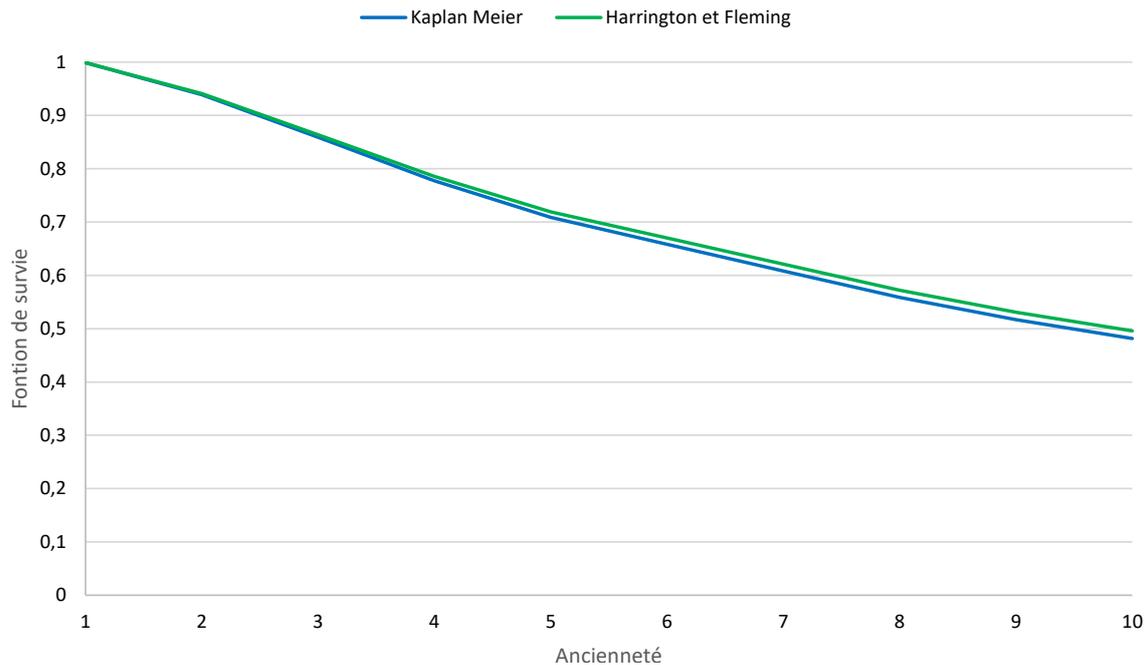


FIGURE 2.3 – Courbe de survie du produit B estimée par l'estimateur Kaplan Meier et Harrington et Fleming

La courbe ci-dessous représente la courbe de taux du rachat du produit B. L'historique disponible sur ce produit nous permet de construire une courbe de rachat des dix premières années d'ancienneté. Comme le produit A, l'option de rachat, du produit B, n'est applicable qu'après une année d'ancienneté. Après la première ancienneté le taux de rachat augmente rapidement de 0% à plus de 9%. Les taux se stabilisent après la sixième année d'ancienneté. En effet, la pénalité de rachat du produit B est invariable après la sixième année d'ancienneté.

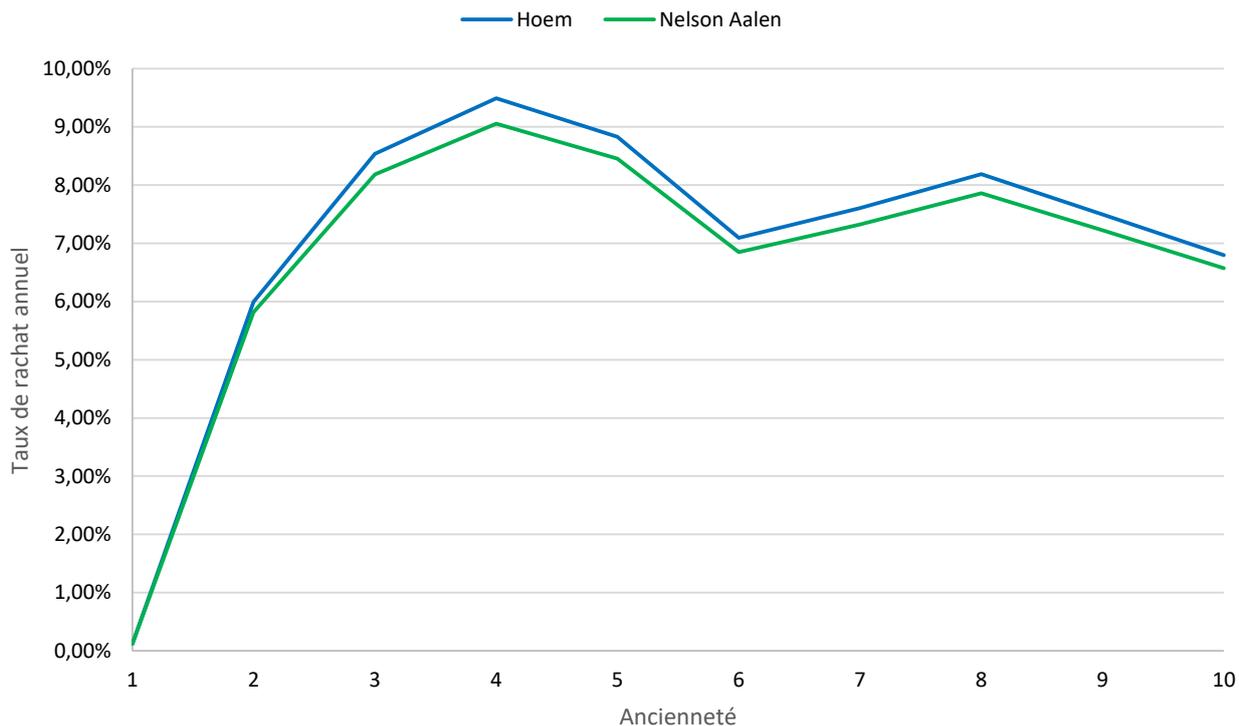


FIGURE 2.4 – Courbe de taux de rachat du produit B estimé par estimateur Hoem et Nelson Aalen

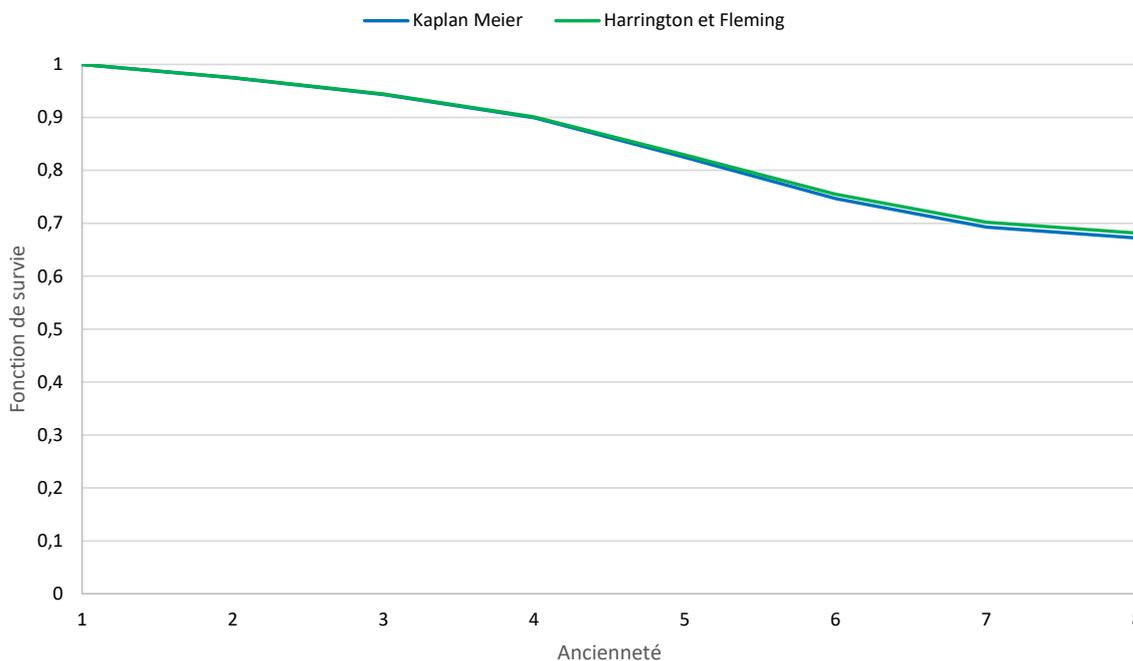


FIGURE 2.5 – Courbe de survie du produit C estimé par l’estimateur Kaplan Meier et Harrington et Fleming

Le produit C est un produit d’épargne lié à des indices de marché. La courbe de rachat est croissante jusqu’à la sixième année d’ancienneté. Les options contractuelles de ce produit n’encourage pas la rétention du contrat pour plus que six années afin d’envisager les taux de placement faibles.

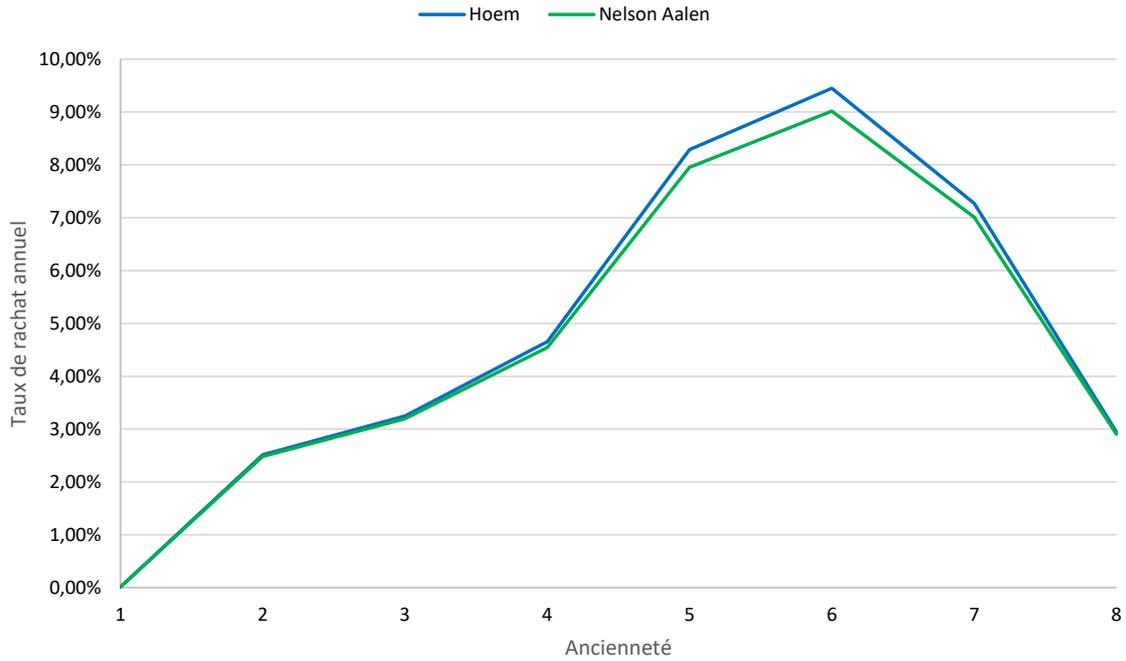


FIGURE 2.6 – Courbe de taux de rachat du produit C estimé par estimateur Hoem et Nelson Aalen

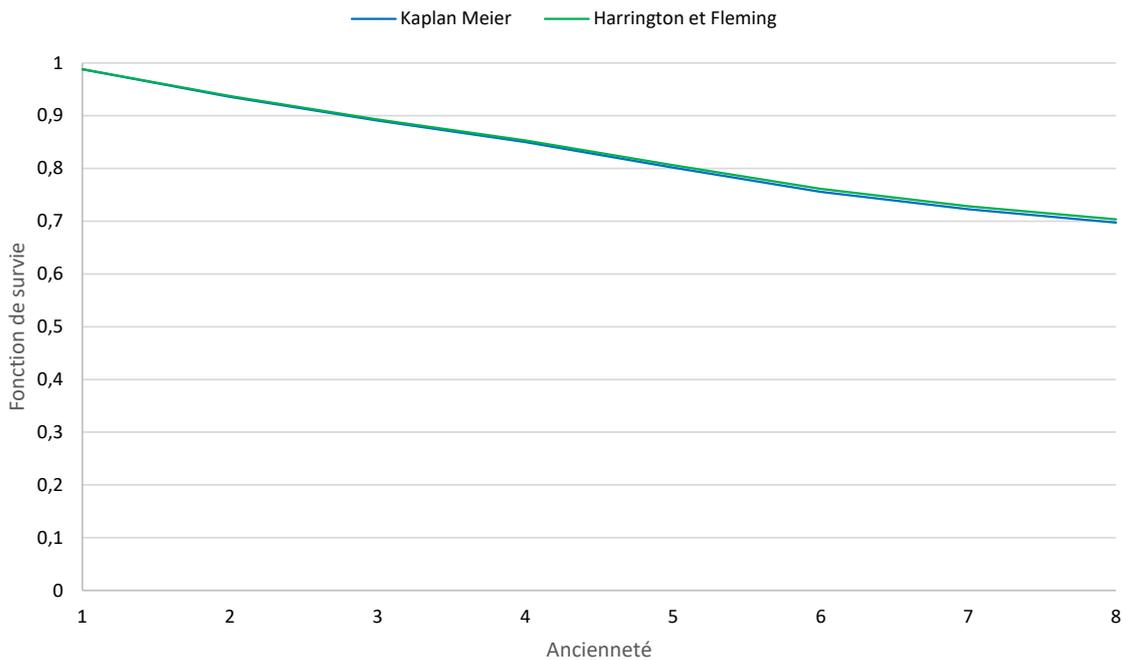


FIGURE 2.7 – Courbe de survie du produit D estimé par l'estimateur Kaplan Meier et Harrington et Fleming

Le taux de rachat de la première année d'ancienneté est assez faible puisque la pénalité de rachat est la plus conséquente la première année de la vie du contrat. Le produit D est un produit d'épargne multisupports. CA Vita favorise la commercialisation de ce produit pour son risque relativement faible. Le rachat est réduit fortement à travers les différents avantages et options contractuelles du produit. Cela explique les taux faibles de rachat de ce produit.

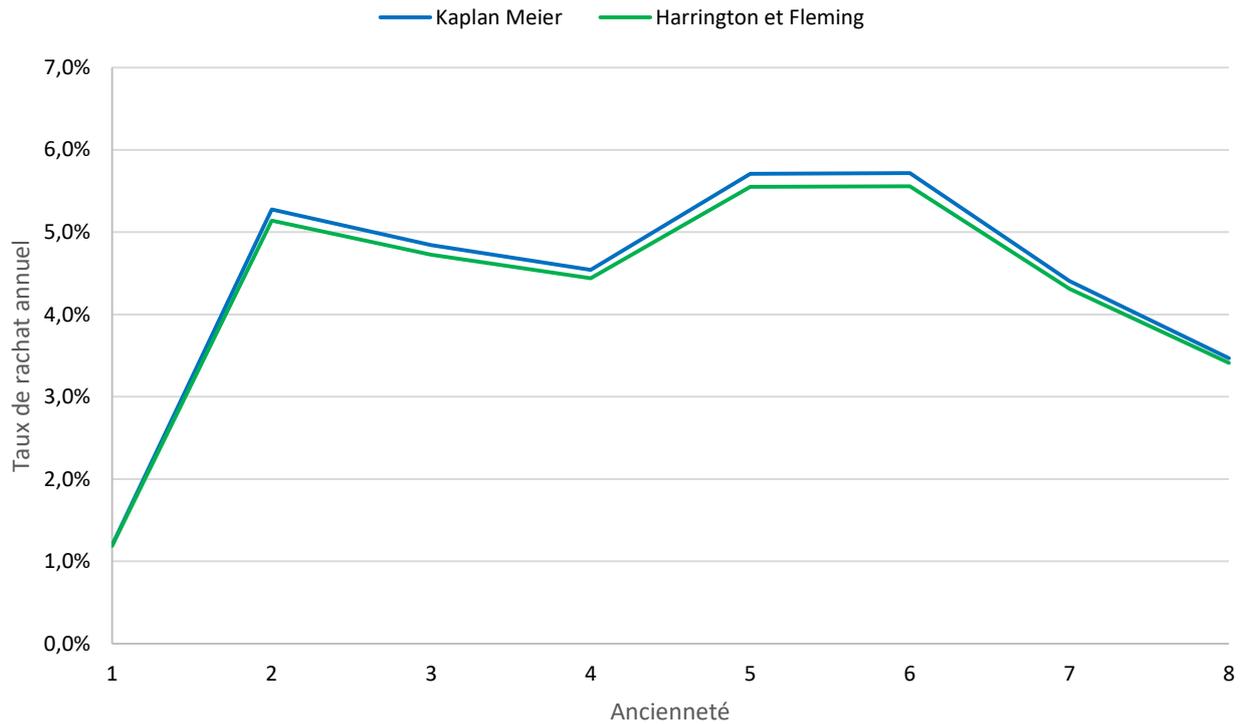


FIGURE 2.8 – Courbe de taux de rachat du produit D estimée par estimateur Hoem et Nelson Aalen

2.3 Lissage

2.3.1 Moyenne mobile

La méthode de lissage par une moyenne mobile est une méthode assez simple à appliquer mais elle présente l'inconvénient de sa sensibilité aux valeurs fluctuantes ou extrêmes. De plus, l'application de cette méthode aux premières et aux dernières valeurs de la courbe à lisser est limitée. La formule utilisée est la suivante :

$$q_t = \sum_{i=-r}^{+r} a_i \hat{q}_{t+i}$$

avec r le nombre de périodes à considérer pour le lissage et les a_i les poids de chaque période.

Nous utilisons le plus souvent le lissage par la moyenne mobile sur 1 année et sur 2 années c'est-à-dire nous considérons une période avant t et une période après t (resp. 2 périodes avant t et 2 périodes après t).

Les poids les plus utilisés sont ceux qui nous donnent une moyenne mobile symétrique : $a_t = a_{-t}$

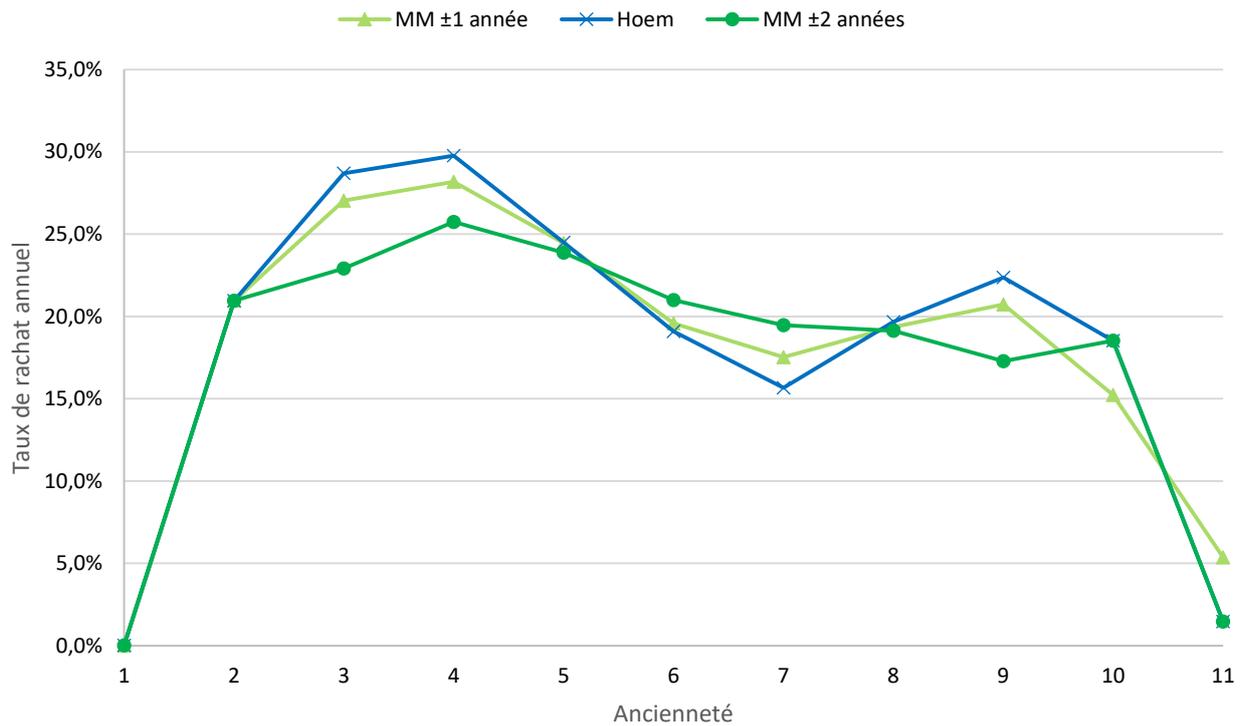


FIGURE 2.9 – Courbe de rachat du produit A lissée par la méthode de la moyenne mobile

Les deux lissages effectués donnent une courbe en dessous de la courbe de taux de rachat pour la plupart des ancienneté.

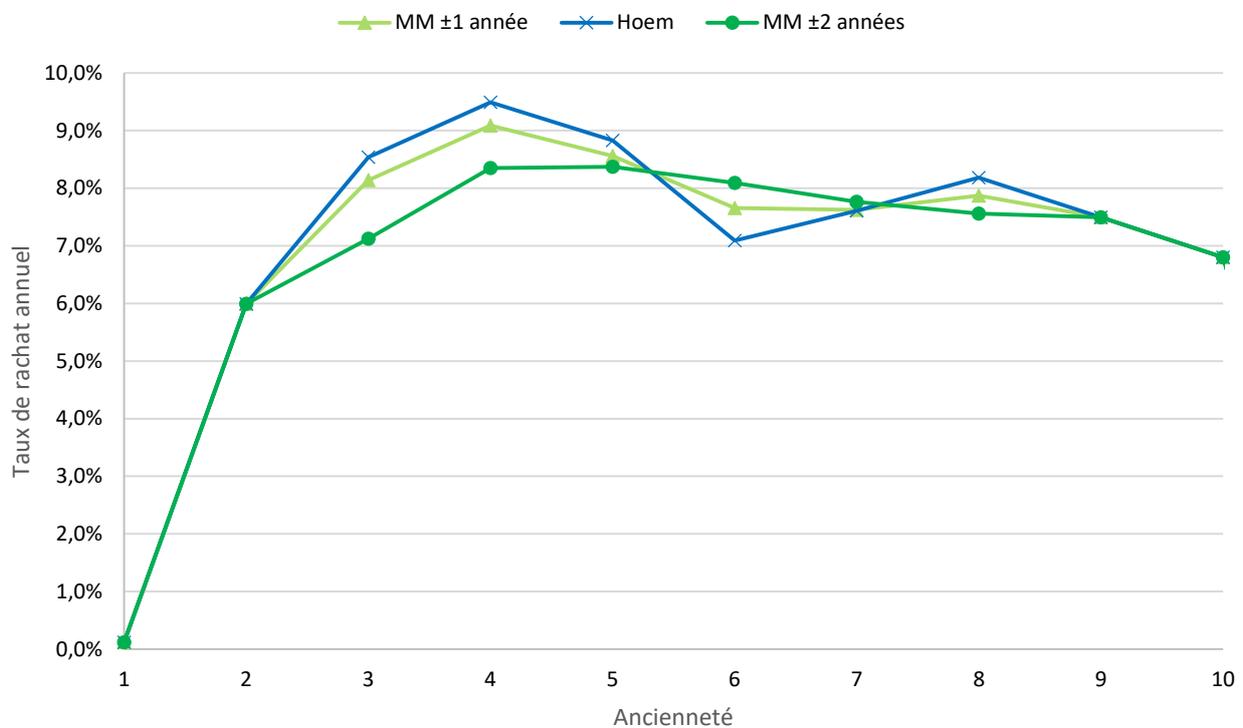


FIGURE 2.10 – Courbe de rachat du produit B lissée par la méthode de la moyenne mobile

Le lissage par la moyenne mobile à une année donne une courbe assez proche de la courbe de taux puisque la courbe initiale est assez lisse.

2.3.2 Whittaker-Henderson

Le lissage de Whittaker Henderson se base sur la minimisation d'une combinaison linéaire de deux critères : le critère de fidélité et le critère de régularité. Le critère de régularité est pondéré par un paramètre h et dépend d'un paramètre z qui influence le degré de lissage.

a) Notations

Avant la définition des deux critères, nous définissons d'abord les poids $(w_t)_{t \in \llbracket 1 ; T \rrbracket}$ comme suit :

$$\forall t \in \llbracket 1 ; T \rrbracket \quad w_t = \frac{Exposure_t}{\sum_i Exposure_i}$$

Ces poids sont dits a posteriori. Ils sont utilisés dans le cas où nous remarquons que certaines valeurs estimées importent plus dans le lissage par rapport à l'effectif présent à cet âge. Ce choix de pondération permet de limiter le poids des données aberrantes.

Nous notons qu'il existe différents choix de pondération. Nous choisissons la pondération afin d'obtenir une courbe lissée que nous jugeons adéquate.

Nous définissons maintenant le critère de fidélité :

$$F = \sum_{i=1}^T w_i * (q_i - \hat{q}_i)^2$$

tel que q_i désigne le taux à ajuster et le \hat{q}_i le taux lissé.

Nous définissons ensuite le critère de régularité comme suit :

$$R = \sum_{i=1}^{t-z} (\Delta^2 q_i)^2$$

avec z : un paramètre à fixer.

De plus, Δ^z est l'opérateur différence avant d'ordre z défini comme suit :

$$\Delta^z f(x) = \Delta^{z-1}(\Delta f(x)) \quad \text{avec} \quad \Delta f(x) = f(x) - f(x-1)$$

b) Moyenne de Whittaker-Henderson

Les taux de rachat lissés sont les taux qui minimisent la moyenne de Whittaker-Henderson M .

Pour calculer cette moyenne nous fixons un paramètre h qui détermine le poids du critère de régularité.

$$M = F + h * R = \sum_{i=1}^T w_i * (q_i - \hat{q}_i)^2 + h * \sum_{i=1}^{T-z} (\Delta^2 q_i)^2$$

c) Lissage Whittaker-Henderson

Les taux q_t lissés sont les taux tels que les dérivées partielles de M sont nulles (ce qui est suffisant puisque M est convexe).

Nous cherchons ainsi les q_t tels que :

$$\forall t \quad \frac{\partial M}{\partial q_t} = 0$$

Pour une meilleure manipulation de la résolution des T équations, nous utilisons la présentation matricielle du problème de minimisation.

$$L = \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_T \end{bmatrix} \quad Q = \begin{bmatrix} \hat{q}_1 \\ \hat{q}_2 \\ \vdots \\ \hat{q}_T \end{bmatrix}$$

$$W = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & w_T \end{bmatrix} \quad \Delta^z V = \begin{bmatrix} \Delta^z q_1 \\ \Delta^z q_2 \\ \vdots \\ \Delta^z q_{T-z} \end{bmatrix}$$

Nous réécrivons R et F comme suit :

$$R = (\Delta^z V)^t \cdot \Delta^z V \quad F = (L - Q)^t \cdot W \cdot (L - Q)$$

Nous notons que nous désignons par A^t la transposée de la matrice A.

Nous posons K_z la matrice de dimension $(T-z) \times T$ telle que $\Delta^z V = K_z \cdot V$.

Nous présentons l'exemple pour $z=3$:

$$K_3 = \begin{bmatrix} 1 & -3 & 3 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & -3 & 3 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -3 & 3 & 1 & 0 \end{bmatrix}$$

Nous réécrivons donc M sous forme matricielle :

$$M = F + h * R$$

$$= L^t \cdot W \cdot V - 2L^t \cdot W \cdot Q + Q^t \cdot W \cdot Q + hV^t \cdot K_z^t \cdot K_z \cdot V$$

Alors :

$$\frac{\partial M}{\partial L} = 2W \cdot V - 2W \cdot Q + 2hK_z^t \cdot K_z \cdot V$$

Nous cherchons maintenant le vecteur L tel que $\frac{\partial M}{\partial L} = 0$:

$$\frac{\partial M}{\partial L} = 0 \Rightarrow C \cdot L = W \cdot Q \quad \text{avec} \quad C = W + hK_z^t \cdot K_z$$

Nous trouvons, dans le cas où C est inversible, le vecteur des taux lissés :

$$L = C^{-1} \cdot W \cdot Q$$

d) Choix des paramètres

Nous choisissons les paramètres z et h qui nous donnent une courbe lissée satisfaisante. Dans la pratique, le paramètre z est fixé à 3 indifféremment de l'allure de la courbe à lisser. Dans ce cas, nous appliquons une différence d'avance d'ordre 3 qui nous permet de tenir compte d'un nombre suffisant de données consécutives et de lisser ainsi les irrégularités. Le paramètre h est fixé de manière à ne pas avoir un lissage très marqué. Le paramètre h est pris entre 1 et 5 pour le lissage de la courbe de rachat.

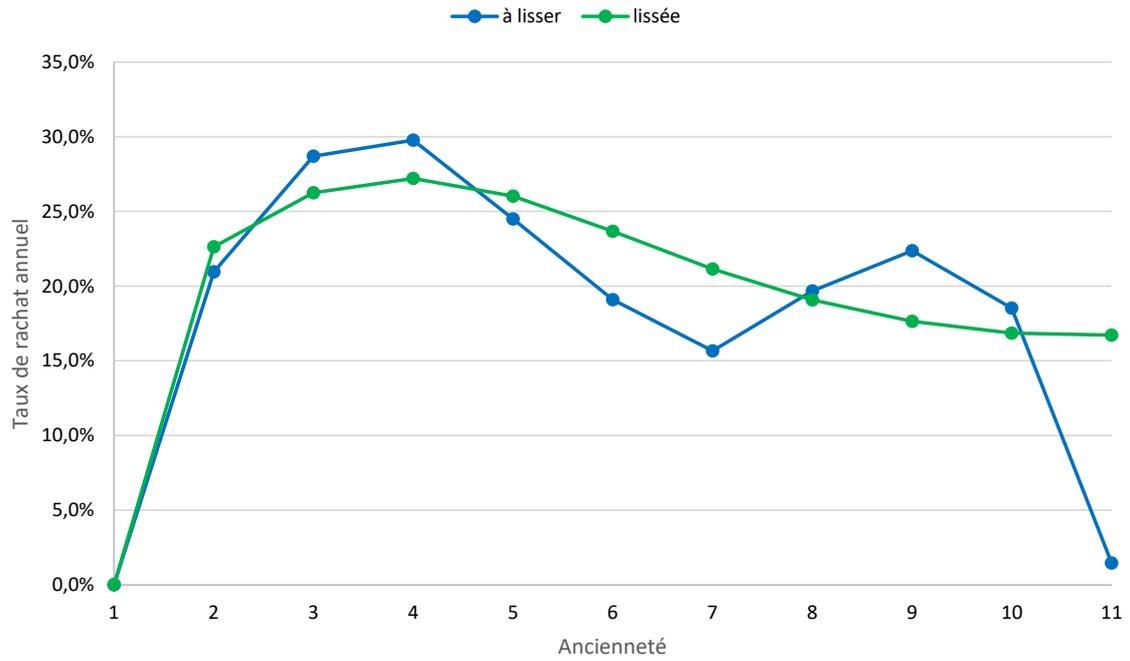


FIGURE 2.11 – Courbe lissée par la méthode de Whittaker-Henderson du produit A

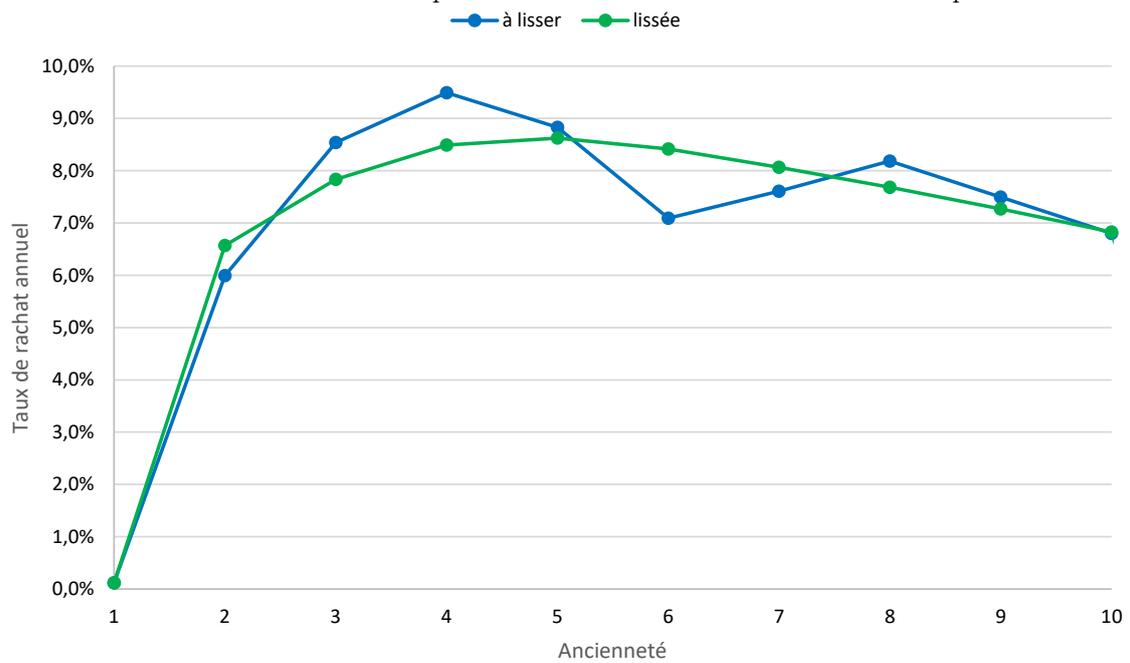


FIGURE 2.12 – Courbe lissée par la méthode de Whittaker-Henderson du produit B

2.3.3 Résultats

Le lissage par la moyenne mobile est une méthode simpliste de lissage et facile à appliquer. Toutefois, les résultats obtenus avec la moyenne mobile montrent la faiblesse de la méthode par rapport aux autres méthodes de lissage. La limite d'application de cette méthode sur les extrémités des courbes de rachat affecte l'efficacité du lissage. De plus, les valeurs lissées par la moyenne mobile sont affectées beaucoup par la variation des valeurs historiques et les fluctuations de la courbe de rachat.

Le lissage de Whittaker-Henderson donne des résultats très satisfaisants. Les courbes lissées par la méthode de Whittaker-Henderson ont une allure assez lisse. De plus, les valeurs lissées sont assez proches des valeurs brutes. En fait, en calculant le critère de fidélité, qui détermine la distance entre les valeurs brutes et les valeurs lissées, nous trouvons que le lissage de Whittaker-Henderson présente, pour la majorité des produits étudiés, des valeurs plus faibles que le lissage par la moyenne mobile ce qui montre que les valeurs lissées de Whittaker-Henderson sont les plus fiables et les plus proches des valeurs estimées.

Au final, nous retenons le lissage de Whittaker-Henderson qui nous procure des courbes de rachat assez lissées et qui garde en moyenne l'allure et les valeurs de la courbe des taux bruts.

2.4 Extrapolation

Afin de construire une courbe de rachat complète, nous avons besoin des données pour tout l'intervalle de temps étudié. Cela n'étant pas toujours le cas, nous recourons à l'extrapolation de la courbe de rachat. Nous utilisons le principe de la moyenne mobile pour l'extrapolation. Nous supposons que le taux de rachat pour une année d'ancienneté est une moyenne pondérée des années passées.

Nous présentons l'extrapolation jusqu'à quinze années d'ancienneté de la courbe de taux brut de rachat et de la courbe lissée par Whittaker-Henderson des deux produits A et B.

Pour le produit A, nous éliminons la valeur de la dernière année d'ancienneté puisque la valeur de l'onzième année affecte considérablement les valeurs extrapolées dans le cadre de l'hypothèse que le rachat garde en moyenne les mêmes valeurs que celles observées dans le passé.

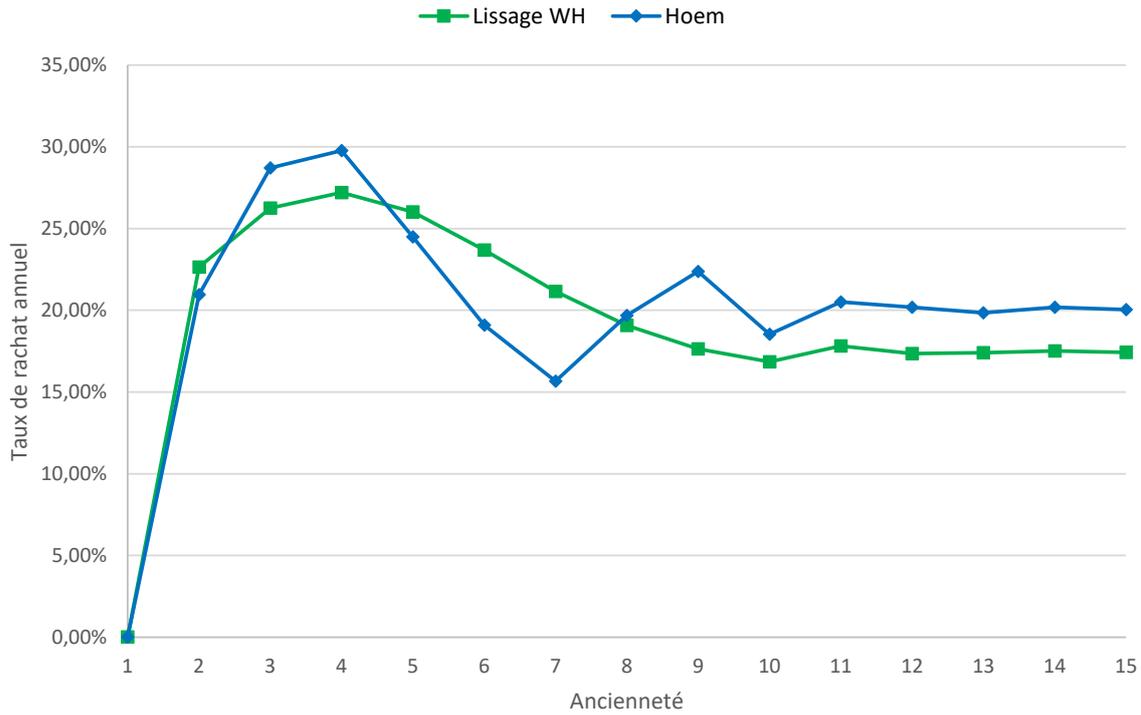


FIGURE 2.13 – Extrapolation par moyenne mobile de la courbe de taux brut de rachat et de la courbe de rachat lissée par W-H du produit A

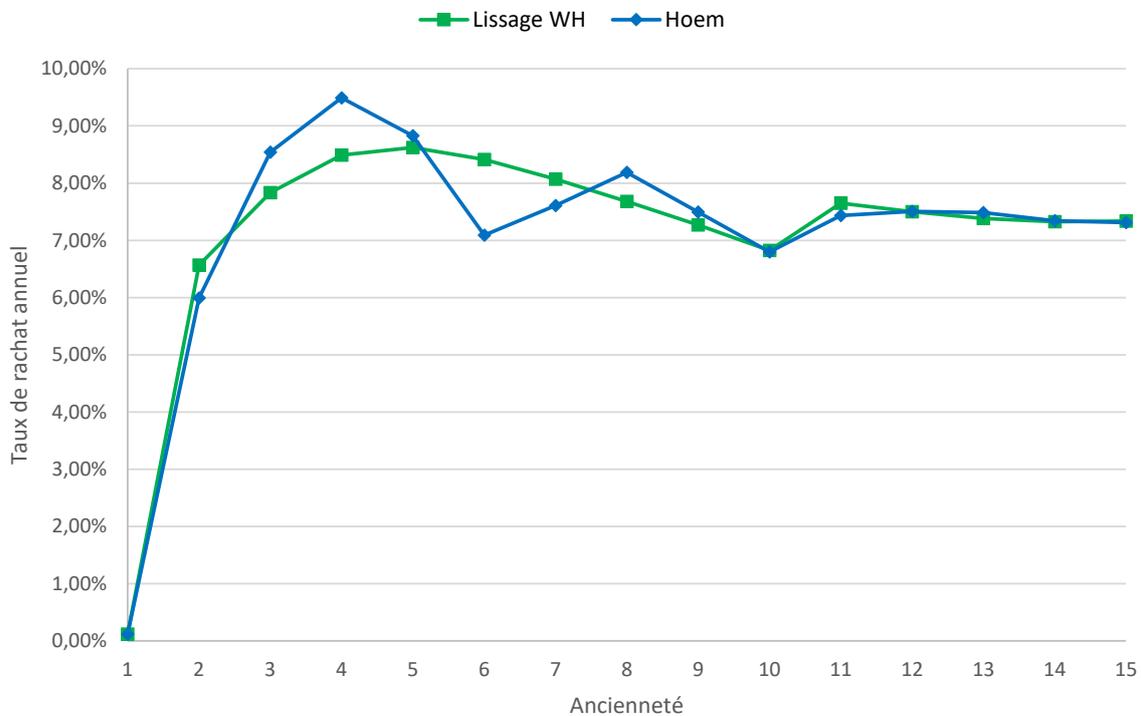


FIGURE 2.14 – Extrapolation par moyenne mobile de la courbe de taux brut de rachat et de la courbe de rachat lissée par W-H du produit B

2.5 Théorie de crédibilité

La théorie de crédibilité est une approche utilisée principalement pour calculer la prime à tarifier en assurance non vie. La théorie repose sur les principes de l'inférence bayésienne et sur l'hypothèse réaliste de l'hétérogénéité des produits. Elle permet la détermination des primes individuelles de différents contrats en tenant compte de l'expérience collective et individuelle de ces contrats.

Nous appliquerons les méthodes de la théorie de la crédibilité pour répondre à deux besoins. Le premier est de prédire les taux de rachat manquants pour les années d'ancienneté non observées. Le deuxième consiste à prédire les taux de rachat des produits ayant un historique d'observation à court terme en exploitant les données déjà observées sur d'autres produits.

2.5.1 Modèle de Bühlmann

Le modèle de Bühlmann présente une pondération entre la prime a priori (collective) et la prime individuelle propre à chaque risque. Elle donne ainsi une prédiction sur la base des données a priori et celles observées.

Pour le modèle de Bühlmann tous les données ont le même poids donc la même influence sur la prédiction. Nous utilisons ce modèle afin de prédire le taux de rachat de l'année à venir en nous basant sur la base historique pour un portefeuille donnée.

a) Notations

Nous notons n le nombre de contrats étudiés dans le portefeuille et T la dernière période observée. Nous avons pour $i \in [1, n]$ et $t \in [1, T]$:

X_{it} : la variable aléatoire représentant l'observation relative au risque i au temps t

Nous définissons la variable aléatoire (non observable) Θ qui représente le facteur de risque. Le paramètre Θ_{it} est relatif à chaque risque i et décrit les caractéristiques de ce dernier dans la période t .

Si nous observons Θ_i , pour $i \in [1, n]$, la prime chargée au risque i serait la prime de risque $\mu(\Theta_i)$:

$$\mu(\Theta_i) = \int x dF(x|\Theta_i) = \mathbb{E}(X|\Theta_i)$$

En revanche nous chargeons à tous les risques la prime collective μ :

$$\mu = \mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\Theta))$$

Ces deux quantités sont estimées par les deux estimateurs suivants :

$$\hat{\mu}(\Theta_i) = \bar{X}_i = \frac{1}{T} \sum_{t=1}^T X_{it}$$

$$\hat{\mu} = \bar{X} = \frac{1}{N} \sum_{i=1}^n \bar{X}_i$$

b) Hypothèses

- $\forall i \in [1, n]$, conditionnellement à Θ_i , les $(X_{it})_{t \in \llbracket 1 ; T \rrbracket}$ sont indépendantes et identiquement distribuées (*i.i.d*) de loi $F(\cdot | \Theta_i)$.

Les moments conditionnels d'ordre 1 et 2 sont comme suit :

$$\begin{aligned} \mu(\Theta_i) &= \mathbb{E}(X_{it} | \Theta_i) & \tau^2 &= \text{Var}(\mu(\Theta_i)) \\ s^2(\Theta_i) &= w_{it} \text{Var}(X_{it} | \Theta_i) & \sigma^2 &= \mathbb{E}(s^2(\Theta_i)) \end{aligned}$$

- Les couples $(X_1, \Theta_1), (X_2, \Theta_2), \dots$ sont indépendants et identiquement distribués (*i.i.d*).

c) Modèle de Bühlmann

Dans le modèle Bühlmann nous cherchons, $\forall i \in [1, n]$, le meilleur estimateur linéaire de $\mu(\Theta_i)$ qui minimise l'espérance de l'erreur.

Nous cherchons :

$$\hat{\mu}(\Theta_i) = \mathbb{E}(X_{i,T+1} | X_{i1}, \dots, X_{iT}) = a_{i0} + \sum_{t=1}^T a_{it} X_{it}$$

tel que $\mathbb{E}((\mu(\Theta) - \hat{\mu}(\Theta))^2) = \mathbb{E}((\hat{\mu}(\Theta_i) - X_{i,T+1})^2)$ soit minimal.

Autrement, nous cherchons :

$$(\hat{a}_{i0}, \dots, \hat{a}_{iT}) = \underset{a_{i0}, \dots, a_{iT}}{\operatorname{argmin}} \mathbb{E} \left(\left(\mu(\Theta) - a_{i0} - \sum_{t=1}^T a_{it} X_{it} \right)^2 \right)$$

L'estimateur de crédibilité est donné par :

$$\hat{\mu}(\Theta_i) = Z_i \bar{X}_i + (1 - Z_i) \mu$$

avec

$\mu = \mathbb{E}(\mu(\Theta_i))$: l'espérance globale des observations pour le risque individuel i

$Z_i = \frac{T}{T + \sigma^2/\tau^2}$: le facteur de crédibilité

Remarques :

- Lorsque $T \rightarrow \infty$ alors $Z_i = \frac{1}{1 + \frac{\sigma^2}{T\tau^2}} \rightarrow 1$. C'est-à-dire que plus nous avons d'expérience sur un contrat moins nous avons besoin de l'expérience collective.
- $\tau^2 = \text{Var}(\mu(\Theta_i))$ représente la variance de l'espérance conditionnelle appelée inter-variance. Si τ est assez grande alors nous avons un portefeuille hétérogène et Z_i est proche de 1. Cela signifie que l'expérience individuelle est assez différente que l'expérience collective.
- $\sigma^2 = \mathbb{E}(s^2(\Theta_i))$ représente l'espérance de la variance conditionnelle appelée intra-variance. Si σ^2 augmente, Z_i converge vers 0 ce qui signifie qu'il est préférable de considérer l'expérience collective lorsque la variance intrinsèque du risque i est grande.

d) Estimation des paramètres

Afin de calculer l'estimateur de crédibilité de Bühlmann nous devons estimer les quantités μ , σ^2 et τ^2 .

Nous cherchons à estimer ces paramètres par des estimateurs "naturels" sans biais.

Estimateur de μ

Nous avons $\mu = \mathbb{E}\left(\mathbb{E}\left(X_{it}|\Theta_i\right)\right)$

Donc un estimateur "naturel" de μ est :

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \bar{X}_i = \frac{1}{Tn} \sum_{i=1}^n \sum_{t=1}^T X_{it}$$

Cet estimateur est sans biais car :

$$\begin{aligned} \mathbb{E}(\hat{\mu}) &= \mathbb{E}\left(\mathbb{E}\left(\frac{1}{Tn} \sum_{i=1}^n \sum_{t=1}^T X_{it}|\Theta_i\right)\right) \\ &= \mathbb{E}\left(\mathbb{E}(X_{it}|\Theta_i)\right) \\ &= \mathbb{E}(\mu(\Theta_i)) \\ &= \mu \end{aligned}$$

Estimateur de σ^2

Afin d'estimer $\sigma^2 = \mathbb{E}(s^2(\Theta_i))$, nous cherchons d'abord un estimateur de s^2 .

Nous pouvons considérer l'estimateur sans biais suivant :

$$\hat{s}^2(\Theta_i) = \frac{1}{T-1} \sum_{t=1}^T (X_{it} - \bar{X}_i)^2$$

L'estimateur de σ^2 est donc le suivant :

$$\hat{\sigma}^2 = \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T (X_{it} - \bar{X}_i)^2$$

Cet estimateur est sans biais. En effet,

$$\begin{aligned} \mathbb{E}(\hat{\sigma}^2) &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \mathbb{E}\left((X_{it} - \bar{X}_i)^2\right) \\ &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \mathbb{E}\left(\left(X_{it} - \mu(\Theta_i)\right)^2 + (\mu(\Theta_i) - \bar{X}_i)^2\right) \\ &= \mathbb{E}(s^2(\Theta_i)) \\ &= \sigma^2 \end{aligned}$$

Estimateur de τ^2

Un estimateur intuitif de l'inter-variance $\tau^2 = Var(\mu(\Theta_i))$ est :

$$\frac{1}{n-1} \sum_{i=1}^n (\bar{X}_i - \hat{\mu})^2$$

Toutefois, cet estimateur est biaisé. En effet :

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left((\bar{X}_i - \hat{\mu})^2 \right) &= \mathbb{E} \left((\bar{X}_1 - \hat{\mu})^2 \right) \\ &= Var(\bar{X}_1 - \hat{\mu}) + \mathbb{E}(\bar{X}_1 - \hat{\mu})^2 \\ &= Var \left(\frac{n-1}{n} \bar{X}_1 - \frac{1}{n} \sum_{i=2}^n \bar{X}_i \right) \\ &= \frac{(n-1)^2}{n^2} Var(\bar{X}_1) - \frac{n-1}{n^2} Var(\bar{X}_1) \\ &= (n-1) Var(\bar{X}_1) \\ &= (n-1) \left(Var(\mathbb{E}(\bar{X}_1|\Theta)) + \mathbb{E}(Var(\bar{X}_1|\Theta)) \right) \\ &= \frac{n-1}{n} \left(\tau^2 + \frac{\sigma^2}{T} \right) \end{aligned}$$

Nous ajustons l'expression de l'estimateur afin d'obtenir l'estimateur sans biais suivant :

$$\hat{\tau}^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{X}_i - \hat{\mu})^2 - \frac{1}{T} \hat{\sigma}^2$$

2.5.2 Modèle de Bühlmann-Straub

Le modèle de Bühlmann-Straub est une généralisation du modèle de Bühlmann. Ce dernier ne désigne pas de poids spécifique à chaque contrat pour chaque date. Avec le modèle de Bühlmann-Straub, nous avons la possibilité de donner un poids différent à certaines observations. Nous pouvons aussi affecter un poids nul aux données manquantes.

a) Notations et hypothèses

- Conditionnellement à Θ_i , les $(X_{it}, t \in \llbracket 1 ; T \rrbracket)$ sont indépendants et nous avons

$$\mathbb{E}(X_i|\Theta_i) = Y_i \beta(\Theta_i) = \mu(\Theta_i)$$

avec $\beta(\Theta_i)$ le vecteur de régression $\in M_{T,1}(T \leq n)$, de composantes linéairement indépendantes
 Y_i est la matrice *design*.

$$Cov(X_i, X_i^t|\Theta_i) = \Sigma_i(\Theta_i)$$

- Les couples $(\Theta_1, X_1), \dots, (\Theta_n, X_n)$ sont indépendants et $\Theta_1, \dots, \Theta_n$ sont *i.i.d*

Nous conservons les mêmes notations et hypothèses que le modèle de Bühlmann et nous ajoutons aux formules les poids spécifiques. Nous désignons par w_{it} le poids du contrat i à la date t .

Nous avons ainsi :

$\forall i \in [1, n]$, Conditionnellement à Θ_i , les $(X_{it})_{t \in \llbracket 1 ; T \rrbracket}$ sont indépendants de loi $F(\cdot | \Theta_i)$. De plus, les couples $(X_1, \Theta_1), (X_2, \Theta_2), \dots$ sont indépendants et $\Theta_1, \dots, \Theta_n$ sont *i.i.d.*

$$\begin{aligned} \mu(\Theta_i) &= \mathbb{E}(X_{it} | \Theta_i) & \tau^2 &= \text{Var}(\mu(\Theta_i)) \\ s^2(\Theta_i) &= w_{it} \text{Var}(X_{it} | \Theta_i) & \sigma^2 &= \mathbb{E}(s^2(\Theta_i)) \end{aligned}$$

b) Modèle de Bühlmann-Straub

Dans le modèle de Bühlmann-Straub nous cherchons à trouver le meilleur estimateur linéaire de $\mu(\Theta_i)$ qui minimise l'espérance de l'erreur quadratique.

D'après le modèle de Bühlmann-Straub le meilleur estimateur linéaire de crédibilité est l'estimateur $\hat{\mu}(\Theta_i)$ suivant

$$\hat{\mu}(\Theta_i) = Z_i \bar{X}_i + (1 - Z_i) \mu$$

avec :

$$\begin{aligned} Z_i &= \frac{w_{i.}}{w_{i.} + \frac{\sigma^2}{\tau^2}} \\ \bar{X}_i &= \sum_{t=1}^T \frac{w_{it} X_{it}}{w_{i.}} \\ w_{i.} &= \sum_{t=1}^T w_{it} \end{aligned}$$

c) Estimateurs des paramètres

Estimateur de σ^2

Nous trouvons de la même manière que précédemment un estimateur de σ^2 :

$$\hat{\sigma}^2 = \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T w_{it} (X_{it} - \bar{X}_i)^2$$

Cet estimateur est sans biais. En effet :

$$\begin{aligned} \mathbb{E}(\hat{\sigma}^2) &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T w_{it} \mathbb{E} \left(\mathbb{E} \left((X_{it} - \bar{X}_i)^2 | \Theta_i \right) \right) \\ &= \frac{1}{n(T-1)} \sum_{i=1}^n \sum_{t=1}^T \mathbb{E} \left(w_{it} \text{Var}(X_{it} | \Theta_i) \right) \\ &= \sigma^2 \end{aligned}$$

Estimateur de τ^2

Un estimateur sans biais de τ^2 est :

$$\hat{\tau}^2 = c \left(\sum_{i=1}^n w_{i.} (\bar{X}_i - \bar{X})^2 - (n-1) \hat{\sigma}^2 \right)$$

avec

- $c = \left(w_{..} - \sum_{i=1}^n \frac{w_{i.}^2}{w_{..}} \right)^{-1}$
- $\bar{X} = \frac{1}{n} \sum_{i=1}^n \frac{w_{i.}}{w_{..}} \bar{X}_i$
- $w_{..} = \sum_{i=1}^n w_{i.}$

En effet :

$$\begin{aligned} \mathbb{E}((\bar{X}_i - \bar{X})^2) &= \mathbb{E}((\bar{X}_i - \mu + \mu - \bar{X})^2) \\ &= \text{Var}(\bar{X}_i) + \text{Var}(\bar{X}) - 2\text{Cov}(\bar{X}_i, \bar{X}) \\ &= \sigma^2 \left(\frac{1}{w_{i.}} - \frac{1}{w_{..}} \right) + \tau^2 \left(1 - 2 \frac{w_{i.}}{w_{..}} + \sum_{i=1}^n \frac{w_{i.}^2}{w_{..}^2} \right) \end{aligned}$$

Donc nous trouvons

$$\mathbb{E} \left(\sum_{i=1}^n w_{i.} (\bar{X}_i - \bar{X})^2 \right) = \tau^2 \underbrace{\left(w_{..} - \sum_{i=1}^n \frac{w_{i.}^2}{w_{..}} \right)}_{c^{-1}} + (n-1)\sigma^2;$$

ou encore

$$\mathbb{E}(\hat{\tau}^2) = c \left(\sum_{i=1}^n w_{i.} \mathbb{E}((\bar{X}_i - \bar{X})^2) - (n-1)\mathbb{E}(\hat{\sigma}^2) \right) = \tau^2$$

Comme l'estimateur τ^2 peut être négatif nous considérons l'estimateur $\max(\tau^2, 0)$

2.5.3 Modèle de Hachemeister

Dans les deux modèles précédents nous prenons l'hypothèse d'homogénéité dans le temps des données étudiés. Pour les données où nous observons une tendance dans le temps, le modèle de Hachemeister sera plus adapté. Ce modèle est un modèle de régression qui se base sur le principe de la théorie de crédibilité.

Nous avons recours à cette méthode afin d'exploiter les données sur le rachat des produits ayant un historique assez complet afin de calculer les taux de rachat des nouveaux produits qui ne disposent pas d'un historique suffisant. De plus, une tendance en fonction de l'ancienneté a été constatée pour certains produits que nous étudierons dans cette partie.

a) La régression linéaire

Le modèle de régression linéaire est donnée par l'expression suivante :

$$Y = X\beta + \xi$$

où :

Y : Vecteur à prédire $(y_1 \ y_2 \ \dots \ y_n)^t$

$$X : \text{Vecteur explicative} \begin{bmatrix} 1 & x_{11} & \cdots & \cdots & x_{1T} \\ 1 & x_{21} & \cdots & \cdots & x_{2T} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \cdots & \cdots & x_{nT} \end{bmatrix}$$

β : vecteur des paramètres du modèle $[\beta_0 \ \beta_1 \ \dots \ \beta_T]^t$

ξ : bruit aléatoire avec $\mathbb{E}(\xi) = 0$ et $Var(\xi) = \Sigma$

Dans le cas de la méthodes des moindres carrés pondérés, nous avons :

$\Sigma = \sigma^2 W^{-1}$ avec $W = \text{Diag}(w_1, w_2, \dots, w_n)$ la matrice des poids et σ^2 un paramètre fixe mais inconnu.

L'estimateur linéaire des moindres carrés non biaisé est

$$\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Y$$

la matrice de covariance de $\hat{\beta}$ est égale à $\Sigma_{\hat{\beta}} = (X' \Sigma^{-1} X)^{-1}$

Le problème de régression linéaire se résume au problème d'estimation de β . Le meilleur estimateur est l'estimateur des moindres carrés qui minimise l'erreur quadratique.

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \mathbb{E} \left((Y - X\beta)^2 \right)$$

L'estimateur des moindres carrés pondérés est donc

$$\hat{\beta} = (X' W X)^{-1} X' W Y$$

de matrice de covariance égale à $\Sigma_{\hat{\beta}} = (X' W X)^{-1}$

b) Modèle de Hachemeister

En examinant la courbe des taux de rachat, nous constatons une tendance dans le temps de quelques produits. Nous utilisons donc le modèle de Hachemeister qui prend en considération la tendance des données dans le temps.

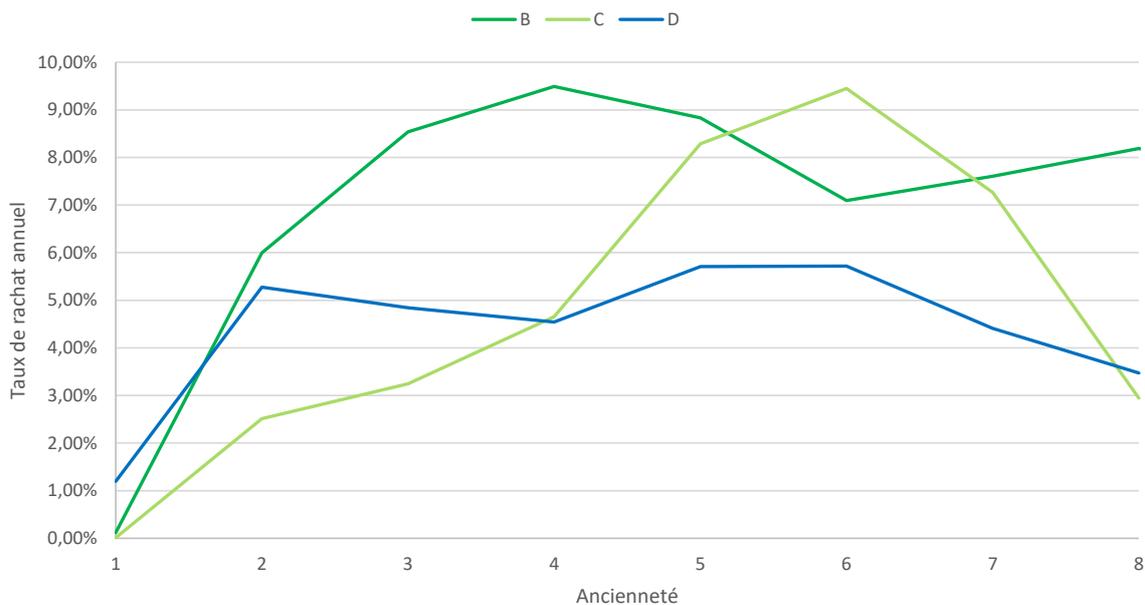


FIGURE 2.15 – Taux de rachat de trois produits pour huit années d'ancienneté

c) Notations et hypothèses

Le modèle de Hachemeister suppose les hypothèses suivantes :

- Conditionnellement à Θ_i , les $(X_{it}, t \in \llbracket 1 ; T \rrbracket)$ sont indépendants et nous avons

$$\mathbb{E}(X_i|\Theta_i) = Y_i\beta(\Theta_i) = \mu(\Theta_i)$$

avec $\beta(\Theta_i)$ le vecteur de régression $\in M_{T,1}(T \leq n)$, de composantes linéairement indépendantes

Y_i est la matrice design.

$$\text{Cov}(X_i, X_i^t|\Theta_i) = \Sigma_i(\Theta_i)$$

- Les couples $(\Theta_1, X_1), \dots, (\Theta_n, X_n)$ sont indépendants et $\Theta_1, \dots, \Theta_n$ sont *i.i.d*

L'estimateur des moindres carrés est solution du problème de minimisation suivant :

$$\min_{\beta_1, \dots, \beta_T} \sum_{i,j=1}^n \left((X_i - (Y\beta)_i) \Sigma_{ij}^{-1} (X_j - (Y\beta)_j) \right)$$

Conditionnellement à Θ_i , le meilleur estimateur linéaire des moindres carrés de $\beta(\Theta_i)$ est

$$\beta(\hat{\Theta}_i) = \left(Y_i' \Sigma_i(\Theta_i)^{-1} Y_i \right)^{-1} Y_i' \Sigma_i(\Theta_i)^{-1} X_i$$

de matrice de covariance

$$\Sigma_{\hat{\beta}(\Theta_i)} = \left(Y_i' \Sigma_i(\Theta_i)^{-1} Y_i \right)^{-1}$$

Nous avons Θ_i inconnu alors, $\Sigma_i(\Theta_i)$ est inconnu. Posons ainsi $S_i = \mathbb{E}(\Sigma_i(\Theta_i))$. En remplaçant la matrice de covariance par S_i nous obtenons l'analogie avec le modèle de Bühlmann-Straub où le paramètre structure est $\sigma^2 = \mathbb{E}(\sigma^2(\Theta_i))$

Théorème de compression des données :

Conditionnellement à Θ_i , le meilleur estimateur linéaire sans biais de \hat{B}_i à partir des X_i est

$$\hat{B}_i = \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} X_i$$

En effet, cette expression ne dépend pas de Θ_i alors c'est bien la compression optimale des données. De plus, nous avons

$$\begin{aligned} \mathbb{E}(\hat{B}_i|\Theta_i) &= \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} \mathbb{E}(X_i|\Theta_i) \\ &= \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} Y_i \beta(\Theta_i) \\ &= \beta(\Theta_i) \quad \text{alors l'estimateur est sans biais conditionnellement à } \Theta_i \end{aligned}$$

Sa matrice d'erreur quadratique moyenne est

$$\mathbb{E} \left(\left(\hat{B}_i - \beta(\Theta_i) \right) \left(\hat{B}_i - \beta(\Theta_i) \right)^t \right) = \left(Y_i' S_i^{-1} Y_i \right)^{-1}$$

Formule de Hachemeister

La formule de Hachemeister de l'estimateur de crédibilité de $\beta(\Theta_i)$ est donnée par :

$$\hat{\beta}(\Theta_i) = Z_i B_i + (I - Z_i) \beta$$

Avec :

$$B_i = \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} X_i$$

$$S_i = \mathbb{E}(\Sigma_i(\Theta_i))$$

$$Z_i = T \left(T + \left(Y_i' S_i^{-1} Y_i \right)^{-1} \right)^{-1}$$

$$T = Cov(\beta(\Theta_i), \beta(\Theta_i)')$$

$$\beta = \mathbb{E}(\beta(\Theta_i))$$

Son erreur quadratique moyenne est $\mathbb{E} \left(\left(\hat{\beta}(\Theta_i) - \beta(\Theta_i) \right) \left(\hat{\beta}(\Theta_i) - \beta(\Theta_i) \right)' \right) = (I - Z_i) T$

La régression linéaire avec modèle de tendance linéaire

La formule de Hachemeister dans le cas simple de la régression linéaire présente un problème puisque la courbe issue du modèle ne se trouve pas forcément entre la courbe individuelle et celle collective comme attendu. En fait, lors de la conception du modèle de crédibilité de la régression Hachemeister trouve que pour un contrat la pente de la courbe de crédibilité est inférieure à celle de la courbe collective et individuelle. Le problème a été résolu par Bühlmann & Gisler (1997) en modélisant l'ordonnée à l'origine au centre de la plage de temps au lieu de l'origine du temps.

Nous étudions, en annexe, le modèle avec le barycentre du temps au lieu de l'origine du temps. Nous présentons avant le modèle à l'origine du temps afin de voir la source d'inconvénients de la méthode et sa solution mathématique.

Nous trouvons deux formules de crédibilité semblables à celle du modèle de Bühlmann-Straub.

$$\hat{\beta}_0(\Theta) = Z_{11} B_0 + (1 - Z_{11}) \beta_0$$

$$\hat{\beta}_1(\Theta) = Z_{22} B_1 + (1 - Z_{22}) \beta_1$$

telles que β_0 et β_1 les composantes de β et

$$B_0 = \mathbb{E}_w(X_t) \quad Z_{11} = w_{it} Var(X_{it} | \Theta_i)$$

$$B_1 = \frac{w.}{w. + \frac{\sigma^2}{\tau_0^2}} \quad Z_{22} = \frac{w. Var_w(t)}{w. Var_w(t) + \frac{\sigma^2}{\tau_1^2}}$$

Nous rappelons que nous avons utilisé le centre de gravité collectif $t_0 = \sum_{k=1}^T k \frac{w.k}{w.}$ au lieu du centre de

gravité individuel $t_{i0} = \sum_{k=1}^T k \frac{w_{ik}}{w_i}$. La différence dans l'utilisation de l'un ou de l'autre est négligeable.

Remarques : Dans ce modèle nous sommes capable d'étudier l'influence des paramètres sur les résultats. Nous remarquons par exemple que, si τ_1^2 est assez petit par comparaison avec $\frac{\sigma^2}{w. Var_w(t)}$ alors Z_{22} sera petite ce qui donne une pente de la courbe de crédibilité qui converge vers la pente de la courbe collective.

2.5.4 Résultats

Nous appliquons la théorie de crédibilité sur les produits B, C et D car les taux de rachat observés pour ces produits sont de même ordre de grandeur. Nous comparons dans les tableaux ci-dessous les résultats estimés de la huitième ancienneté avec les taux observés.

Ancienneté	Produit B	Produit C	Produit D
1	0,00%	0,00%	1,20%
2	5,99%	2,51%	5,28%
3	8,54%	3,25%	4,84%
4	9,49%	4,65%	4,54%
5	8,83%	8,28%	5,71%
6	7,09%	9,45%	5,72%
7	7,61%	7,27%	4,41%
8	8,19%	2,95%	3,47%
Taux estimé par Bühlmann	6,84%	5,93%	6,64%

TABLE 2.1 – Taux de rachat estimé par le modèle Bühlmann

Le modèle Bühlmann donne des valeurs loin des valeurs observées sur la huitième ancienneté. L'erreur quadratique moyenne s'élève à 0,21%. Comme le modèle de Bühlmann donne le même poids à tous les produits et à toutes les années d'ancienneté, nous pouvons espérer que le modèle Bühlmann Straub peut contourner les défauts de ce dernier.

Ancienneté	Produit B	Produit C	Produit D
1	0,00%	0,00%	1,20%
2	5,99%	2,51%	5,28%
3	8,54%	3,25%	4,84%
4	9,49%	4,65%	4,54%
5	8,83%	8,28%	5,71%
6	7,09%	9,45%	5,72%
7	7,61%	7,27%	4,41%
8	8,19%	2,95%	3,47%
Taux estimé par Bühlmann Straub	7,02%	5,66%	5,87%

TABLE 2.2 – Taux de rachat estimé par le modèle Bühlmann Straub

Le modèle de Bühlmann Staub donne des estimations plus affines que le modèle Bühlmann. Cependant l'écart entre les valeurs observées et les valeurs estimées reste important.

Ancienneté	Produit B	Produit C	Produit D
1	0,00%	0,00%	1,20%
2	5,99%	2,51%	5,28%
3	8,54%	3,25%	4,84%
4	9,49%	4,65%	4,54%
5	8,83%	8,28%	5,71%
6	7,09%	9,45%	5,72%
7	7,61%	7,27%	4,41%
8	8,19%	2,95%	3,47%
Taux estimé par Hachemeister	7,15%	4,01%	3,31%

TABLE 2.3 – Taux de rachat estimé par le modèle Hachemeister

Le modèle Hachemeister semble le meilleur à estimer des taux de rachat proches des valeurs estimées. Ce modèle capte la tendance des courbes de rachat des différents produits. Nous remarquons que le modèle estime que les taux vont décroître ce qui s’aligne avec les observations actuelles. En effet, la courbe de rachat des trois produits a une tendance décroissante sur les dernières anciennetés et l’estimateur prédit un taux de rachat qui suit la même tendance.

Modèle retenu Comme les courbes de rachat construites présentent des tendances dans le temps, le modèle Hachemeister doit normalement convenir à notre portefeuille. Un backtesting sur l’ensemble des produits étudiés a été effectué pour déterminer le modèle le plus adapté à notre portefeuille. Les résultats du backtesting effectué sur l’ensemble de produits montrent que ce modèle est le plus adéquat. Le modèle de Hachemeister parvient à capter la tendance des courbes et s’adapte mieux que les deux autres modèles aux fluctuations et aux spécificités des courbes.

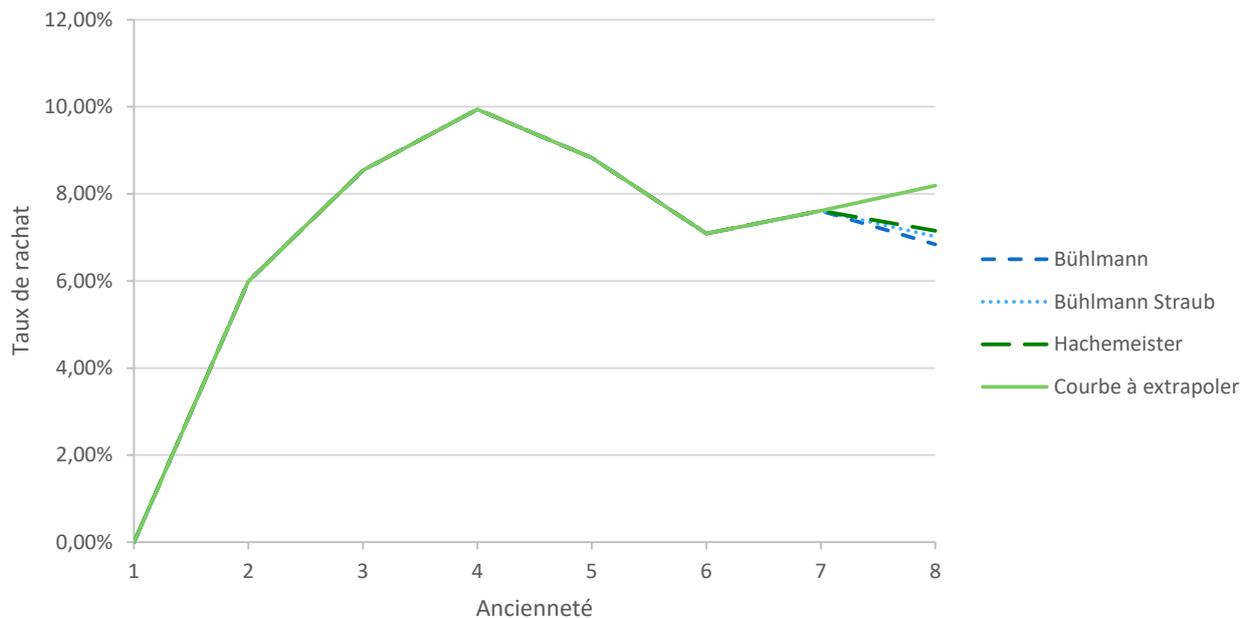


FIGURE 2.16 – Comparaison entre les différentes courbes extrapolées pour le produit B

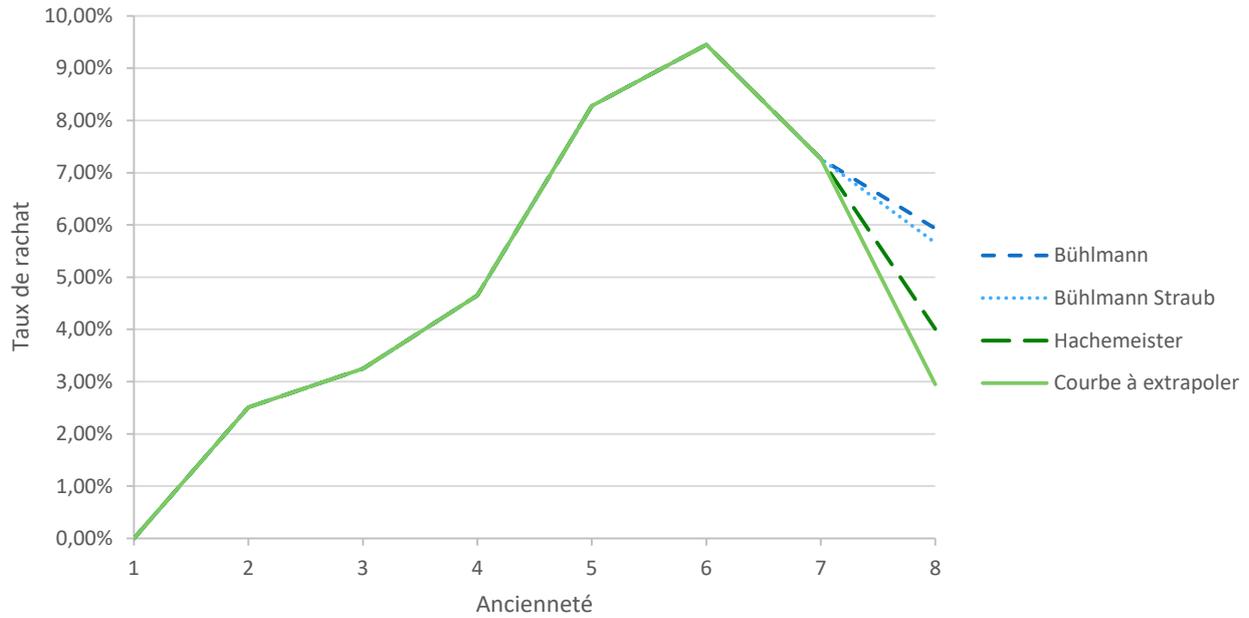


FIGURE 2.17 – Comparaison entre les différentes courbes extrapolées pour le produit C

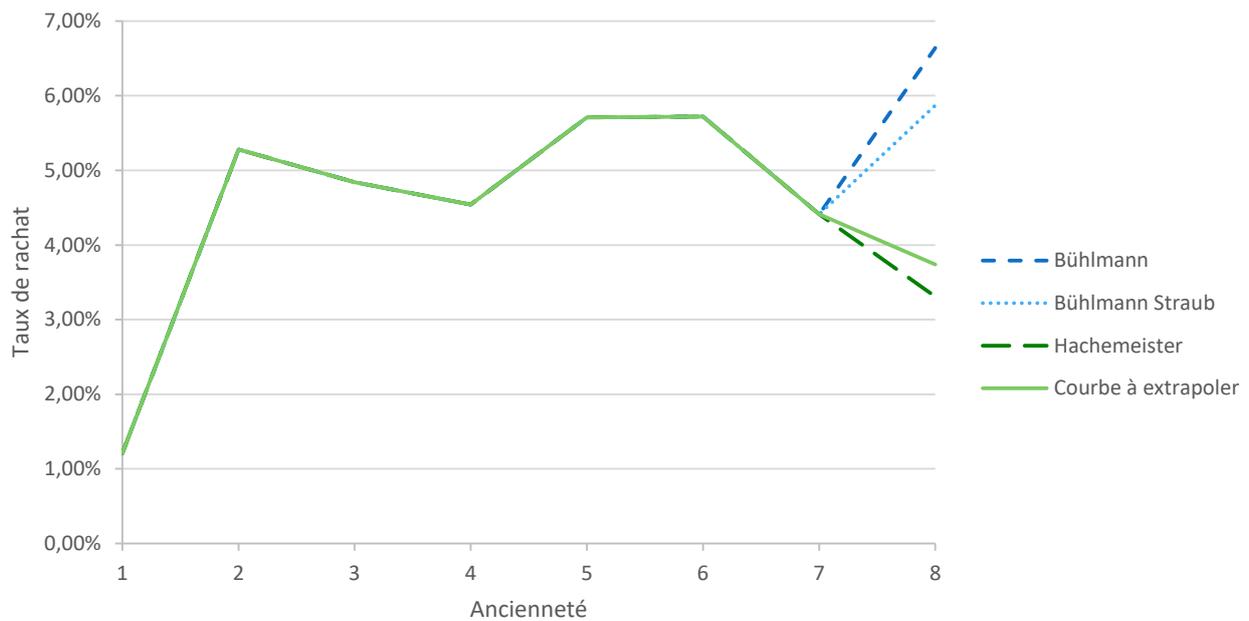


FIGURE 2.18 – Comparaison entre les différentes courbes extrapolées pour le produit D

Les résultats présentés dans les tableaux ci-dessus reflètent les résultats et les observations faite pour tout le périmètre. les courbes ci-dessus nous permettent de comparer visuellement les trois modèles et d'exploiter plus facilement les résultats présentes dans les tableaux.

Le modèle Hachemeister est le modèle le plus adapté aux produits étudiés puisqu'il capte la tendance observée sur l'ensemble des produits et donne des estimations assez proches de la valeur observée pour la huitième année d'ancienneté. Le modèle Bühlmann donne des résultats médiocres qui ne reflètent pas la réalité des produits. Il ne prend pas en considération le poids de chaque produit et de chaque ancienneté passée dans l'estimation. Pour plus de visibilité des résultats et pour comparer les trois

modèles nous traçons les résultats affichés dans les tableaux. Le modèle de Hachemeister suit toujours la tendance des courbes

Chapitre 3

Prédiction du rachat par les données individuelles

Dans ce chapitre, nous étudions les méthodes de *machine learning* utilisées pour prédire l'évènement de rachat individuel des assurés. Nous présentons la base de données sur laquelle nous avons appliqué les modèles de prédiction, ensuite nous développons les méthodes utilisées pour prédire le rachat individuel. Nous cherchons à créer des groupes homogènes d'assurés en se basant sur leur probabilité de rachat. Cette classification sera incorporée dans la tarification des produits.

3.1 Démarche de la prédiction par *machine learning*

La démarche suivie pour la prédiction du rachat est la démarche classique de tout projet de *machine learning*. Nous explorons d'abord les données extraites. Nous analysons les variables explicatives issues de la base d'apprentissage brute afin de s'assurer de la qualité des données. Ensuite, nous appliquons les étapes de *data engineering* qui consistent à "nettoyer" et à "réparer" les données erronées. Après ce traitement, nous appliquons un découpage de la base. Les modèles d'apprentissage seront appliqués sur la base d'apprentissage et le modèle qui ressort sera testé sur l'échantillon appelé *test*. La dernière étape consiste à comparer et à tester la robustesse des modèles. Pour appliquer ces démarches nous avons utilisé le logiciel R.

a) Exploration des données

Le premier pas dans un projet de *machine learning* est l'analyse des données. Cette étape consiste en une étude préliminaire de la robustesse et de la qualité des données. Une étude de la complétude des données sera menée afin de choisir le traitement des données manquantes. Nous déterminons aussi les corrélations entre les différentes variables de la base ce qui nous aidera dans la prochaine étape de *data engineering*.

b) *Data Engineering*

La phase de *data engineering* consiste à corriger, ajouter, supprimer ou modifier les valeurs aberrantes ou manquantes. Nous créons aussi des nouvelles variables si nécessaire et nous supprimons les

variables qui n'apportent pas une information pertinente sur la variable à expliquer ou qui contribue par une information similaire à d'autres variables de la base.

c) Échantillonnage

L'échantillonnage est une étape importante en *machine learning*. Elle consiste à diviser la base de données en deux échantillons appelés base *train* et base *test*. La répartition classique est de 70% pour la base d'apprentissage appelée *train* et 30% pour la base de test. Plusieurs méthodes d'échantillonnage existent pour découper la base : échantillonnage aléatoire, échantillonnage temporelle, etc. Les deux bases doivent être comparables et l'échantillonnage doit être choisi de sorte que les modèles d'apprentissage ne seront ni en sur-apprentissage ni en sous-apprentissage.

Nous choisissons d'utiliser l'année de souscription comme variable de séparation de la base. Ainsi, la base d'apprentissage rassemble les observations dont la police est souscrite avant ou en 2015 (63% de la base initiale) et la base de test rassemble le reste de la base (37% de la base initiale).

d) Apprentissage et prédiction

La phase de prédiction repose sur des algorithmes de *machine learning* choisis en fonction de la taille de la base, le nombre des variables explicatives ou encore le temps de calculs et les besoins en puissance machine. Nous choisissons d'utiliser les méthodes d'arbres aléatoires, l'algorithme XGboost et le modèle linéaire généralisé. Nous détaillons ensuite les modèles utilisés et la formulation mathématique derrière chaque modèle et nous exposons les résultats relatifs à chaque méthode.

e) Comparaison des modèles

A fin de choisir le modèle le plus performant et le plus adapté aux variables explicatives disponibles nous utilisons des indicateurs de performance des modèles de prédiction. Nous présentons dans ce qui suit les indicateurs choisis pour faire la comparaison des modèles.

Taux de bien classés Dans le cas de la classification où la valeur à expliquer prend deux modalités que nous appelons "positif" et "négatif", nous pouvons calculer le taux de bien classés. Ce taux est appelé *précision* et est défini comme suit :

$$\tau = \frac{\text{Nombre de cas bien classés}}{\text{Nombre total de cas}}$$

L'indicateur de précision est en fait insuffisant. Nous donnons l'exemple d'un événement rare qui se produit avec une probabilité de 5%. Un prédicteur qui ne prédit jamais cet événement, aura une précision de 95%.

Pour contourner l'insuffisance du critère de précision, deux autres quantités sont introduites. Le premier est le taux de vrais positifs appelé *sensibilité*. Le deuxième critère est le taux de vrais négatifs appelé *spécificité*.

$$\text{Sensibilité} = \frac{\text{Nombre de vrais positifs}}{\text{Nombre de positifs}}$$

$$\text{Spécificité} = \frac{\text{Nombre de vrais négatifs}}{\text{Nombre de négatifs}}$$

Ces deux indicateurs, sensibilité et spécificité, sont interprétés comme étant les indicateurs de précision restreint, respectivement, aux cas positifs et négatifs.

Nous introduisons aussi les deux quantités suivantes :

$$\text{Taux de faux positifs} = 1 - TVN = \frac{\text{Nombre de faux positifs}}{\text{Nombre de négatifs}}$$

$$\text{Taux de faux négatifs} = 1 - TVP = \frac{\text{Nombre de faux négatifs}}{\text{Nombre de positifs}}$$

Ces quatre quantités sont utilisées pour calculer la matrice de confusion.

	+	-
+	Taux vrais positifs	Taux faux positifs
-	Taux faux négatifs	Taux vrais négatifs

Courbe ROC La courbe ROC (caractéristique de fonctionnement du récepteur) est un outil très utilisé dans différents domaines. C'est un outil graphique qui permet de comparer le comportement des prédicteurs de manière globale. De plus, la courbe ROC est associée à un indicateur plus synthétique et facile à interpréter, Le critère AUC (Aire sous la courbe).

La courbe ROC est la courbe du taux de vrais positifs en fonction du taux de faux positifs ou encore la courbe de la *sensibilité* en fonction de 1- *spécificité*.

Dans le cadre de la régression logistique, la classification est effectuée par rapport à un score. Ce score est la valeur de la fonction de régression pour une réalisation donnée. Ainsi, si la valeur de ce score dépasse un seuil s , alors 1 est affecté à la variable de sortie. La courbe ROC trace les valeurs de la sensibilité en ordonnée et les valeurs de 1- spécificité en abscisse pour différentes valeurs de s entre 0 et 1. L'interprétation de la pertinence du prédicteur est basée sur la position de la courbe par rapport à la première bissectrice.

Les deux figures suivantes présentent deux cas extrêmes. La figure à gauche présente le cas de discrimination parfaite. Tous les cas positifs sont situés devant les négatifs si on les trie selon un score décroissant. La figure à droite présente le cas d'absence de discrimination où les valeurs sont attribuées au hasard.

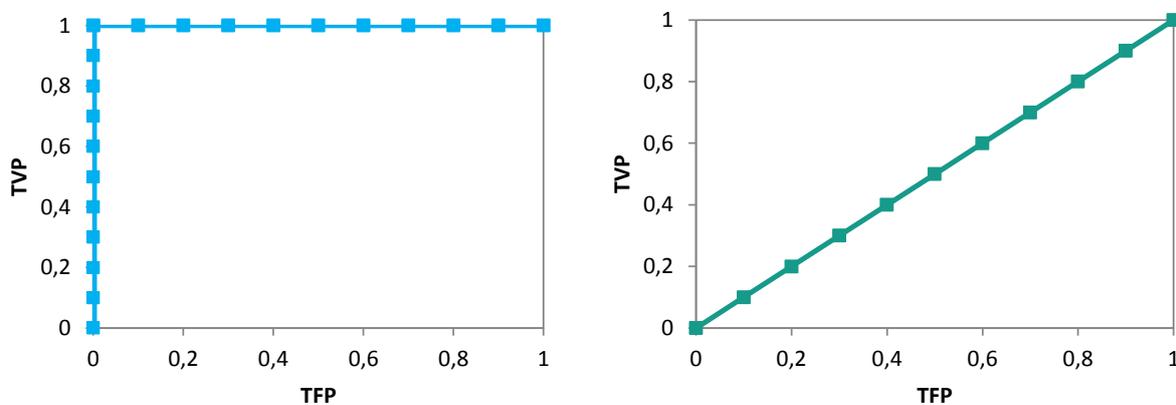


FIGURE 3.1 – Illustration des deux cas extrêmes de la courbe ROC

Le critère AUC L'AUC est un indicateur qui peut remplacer la courbe ROC pour effectuer une comparaisons rapide des classifieurs. Cet indicateur présente la probabilité de placer un individu positif devant un autre négatif. Il se calcule comme la surface sous la courbe ROC.

La valeur $AUC = 0.5$ est considérée comme la valeur de référence où le prédicteur affecte les scores de manière aléatoire. Le cas où $AUC = 1$ présente le cas de discrimination parfaite.

Un prédicteur doit donner un $AUC > 0.5$, il doit performer mieux qu'une classification au hasard. Plus la valeur de l'AUC s'approche de 1 plus le classifieur est performant.

Log Loss Le *Log Loss* est définit comme suit :

$$LL = -\frac{1}{n} \sum_{i=1}^n \sum_{k=1}^L z_{ik} \log(\pi_{ik})$$

z_{ik} prend la valeur 1 si l'individu prend la modalité k de la variable Y à expliquer, 0 sinon. π_{ik} est la probabilité estimée que l'individu i prenne la modalité k de Y. Dans notre cas, où Y est binaire, $L = 2$ et $z_{ik} = y_i$.

Le critère *Log Loss* n'est pas borné et prend la valeur 0 si toutes les observations sont bien classées avec une probabilité 1.

Autres indicateurs Nous disposons en plus des indicateurs de performance présentés, d'autres indicateurs de performance classiques. Ces indicateurs cherchent à calculer l'erreur de prédiction présentée par l'écart observé entre la valeur prédite et la valeur observée de la variable de sortie.

— Erreur quadratique moyenne MSE : $\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$

— Écart quadratique moyen $RMSE$: $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

— Erreur absolue moyenne MAE : $\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$

— Corrélation de Pearson : ρ^2

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

3.2 Analyse de données

Notre étude sera menée sur un portefeuille de différents produits d'épargne de CA Vita. Les polices étudiées sont souscrites entre 1995 et 2018. La base ne contient pas des polices en vigueur. Le tableau suivant présente les variables contenues dans la base de données. Nous notons que quelques variables ont été éliminées préalablement puisqu'ils donnent des informations postérieures à l'évènement de rachat étudié. Sauf indication, les variables sont observées à la date de sortie. Ainsi, la prédiction n'est pas faite par rapport à une année d'observation fixe mais par rapport à la date de sortie de police.

Cela élimine le problème de la censure puisque chaque police est observée sur la totalité de sa période d'existence dans le portefeuille.

Nom de la variable	Description de la variable
Policy	Numéro de la police
Policy.Status	Status de la police : en vigueur, annulée ou expirée
Comercial.tariff	Numéro du type de produit
Distribution.network	Réseau de distribution
LoB	Branche d'activité liée au produit
Sex	Sexe du détenteur de la police
Age.at.subscription	Âge à la souscription de la police
Effect.date	Date de souscription
Reversal.effect.date	Date de rachat/sortie
Expiry.Date	Date de maturité de la police
Initial.premium	prime payée à la souscription
Premium.sum	Somme des primes payées
Initial.performance	Capital investi revalorisé à la première date d'anniversaire annuelle de la police
Performance.to.date	Capital investi revalorisée à la dernière date d'anniversaire annuelle de la police
Reversal.reason	Raison de sortie/rachat : 12 modalités dont une indique le rachat total de la police
Age.at.reversal	Âge à la sortie/rachat
Reversal_Year	Année de sortie/rachat
Holding.Period	Ancienneté de la police
Duration	Durée jusqu'à la maturité
Duration.still.to.go	Période restante de la vie du contrat
Diff-CPI	Différence de l'indice des prix à la consommation entre l'année de souscription et l'année de sortie/rachat
Difference	Différence en pourcentage entre la performance entre l'année de souscription et l'année de sortie/rachat
Diff-Inflation	Différence du taux d'inflation entre l'année de souscription et l'année de sortie/rachat
Diff-GNI	Différence du revenu national brut entre l'année de souscription et l'année de sortie/rachat

TABLE 3.1 – Tableau des variables explicatives du modèle

Le graphique ci-dessous présente la matrice de corrélation calculée sur R.

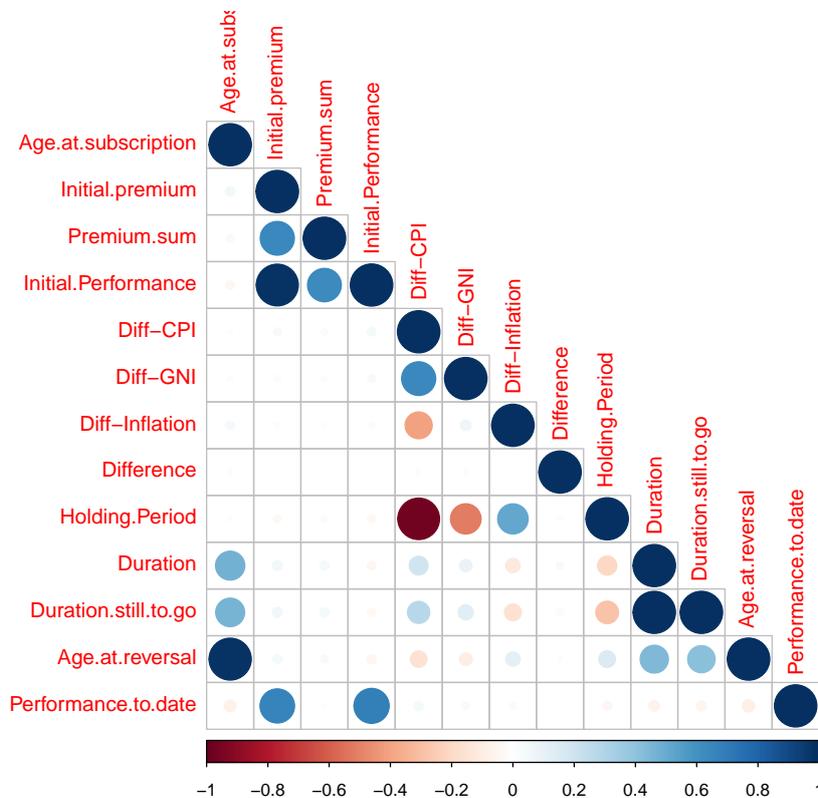


FIGURE 3.2 – Matrice de corrélation des variables explicatives numériques

Nous observons une forte corrélation entre l'âge à la souscription et l'âge à la sortie/rachat. De plus, une forte corrélation entre la prime initiale et la performance initiale est observée. De ce fait, nous fusionnons les deux variables pour créer la variable "Performance" qui est égale à la différence entre ces deux variables.

Les variables "Policy" et "Policy.Status" ont été éliminées car elles ne contribuent pas à la prédiction du rachat.

a) Âge à la souscription

La base de données étudiée présente des assurés âgés entre 1 et 99 à la souscription.

La base contient aussi des polices dont l'âge n'est pas renseigné, nous affectons à ces observations la valeur -1.

La figure suivante représente l'histogramme de l'âge à la souscription hors polices d'âge non renseigné et la courbe de densité d'une loi normale. Nous remarquons qu'on nous avons peu d'observations pour les jeunes âgés moins de 20 et pour les assurés âgés.

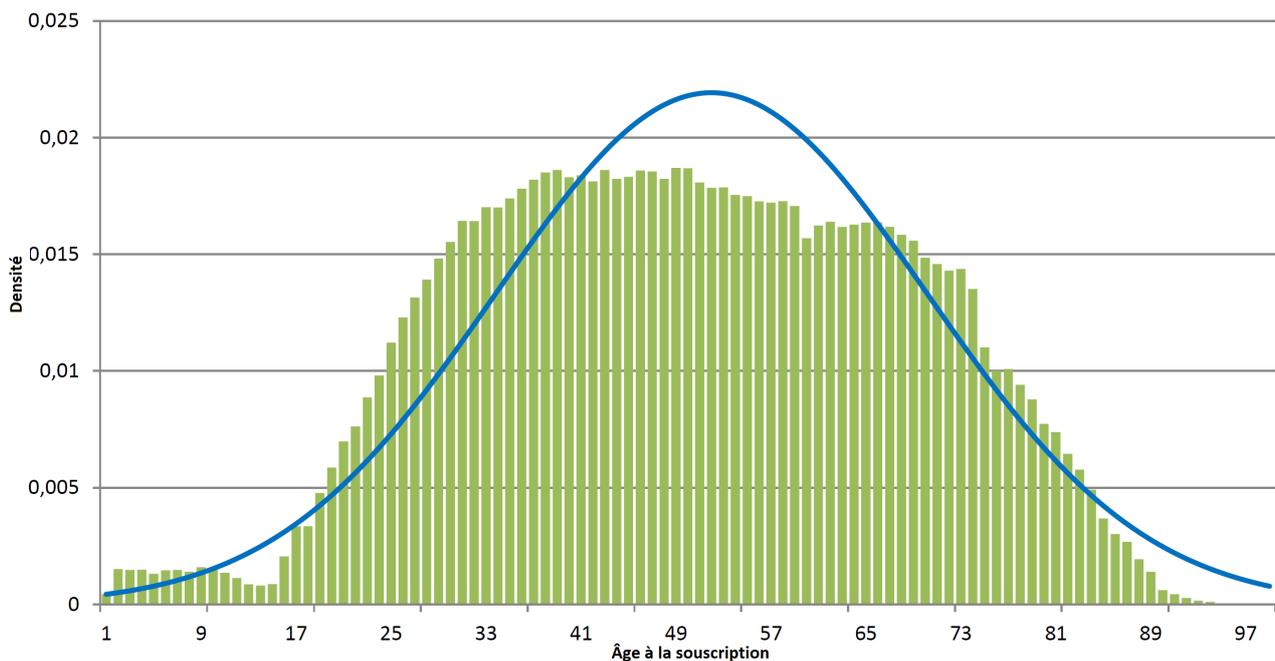


FIGURE 3.3 – Distribution des âges à la souscription de toutes les polices

b) Âge au rachat

L'histogramme suivant présente la distribution des âges des polices lors du rachat. La valeur -1 présente les polices dont on ne connaît pas l'âge de l'assuré. Cet histogramme nous permet de connaître la proportion de polices rachetées dont l'âge n'est pas renseigné.

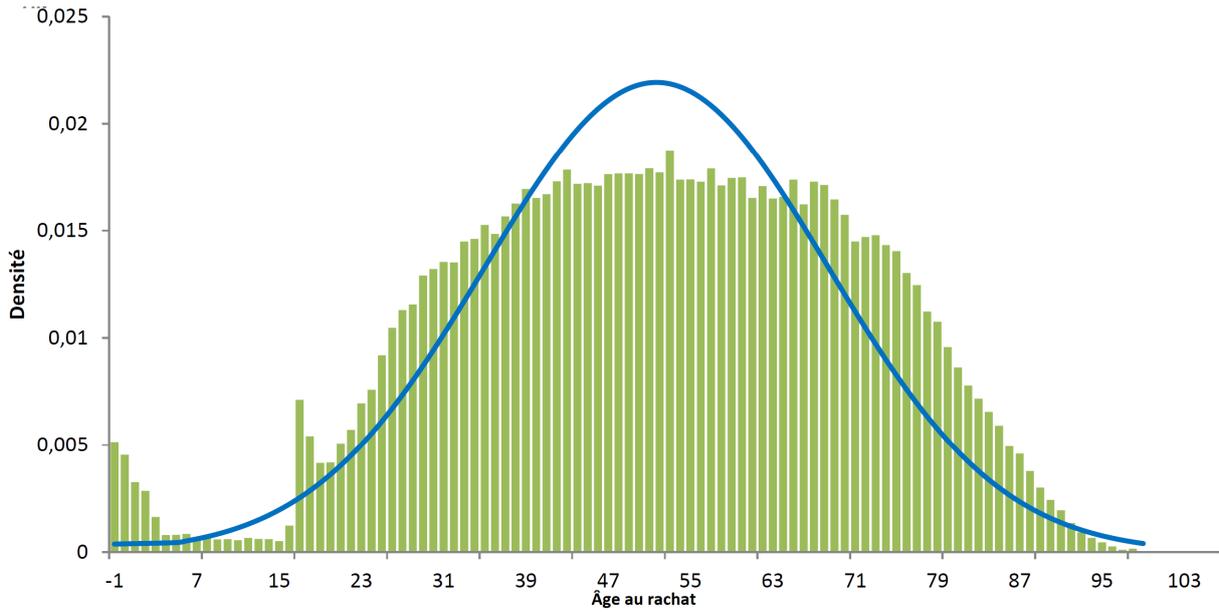


FIGURE 3.4 – Distribution des âges au rachat des polices rachetées

c) Sexe

Nous étudions la répartition des polices par sexe afin de s’assurer de l’uniformité de la répartition des polices. La figure ci-dessous présente le pourcentage des polices pour les deux sexes. Nous remarquons que la répartition de contrats entre les deux sexes est équilibrée. Sur toute la base nous avons presque 41 % des polices souscrites par des femmes et 45% des polices souscrites par des hommes. Pour le reste des polices dont le sexe n’est pas renseigné. La valeur "Inconnu" est affectée à la variable sexe pour les polices dont le sexe est inconnu.

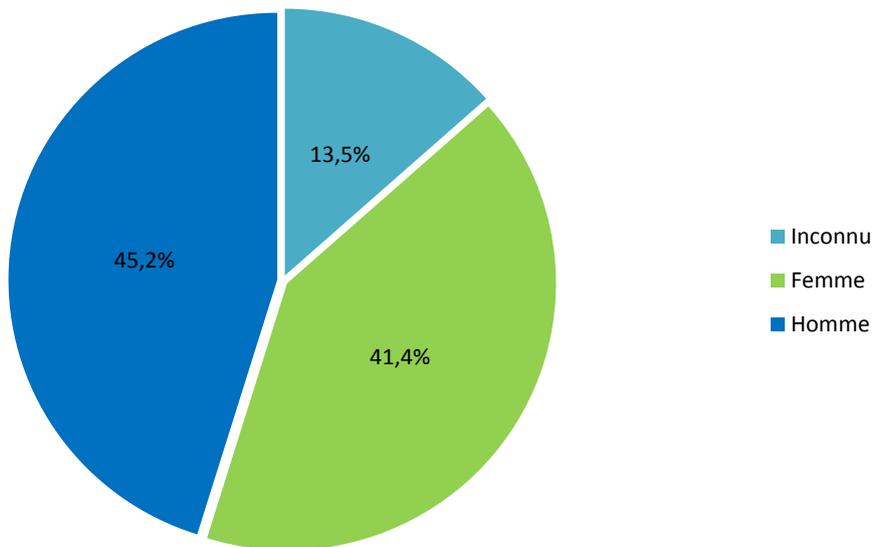


FIGURE 3.5 – Répartition des assurés par sexe

d) Réseau de distribution

Les polices du portefeuille étudié sont distribuées par 9 chaînes de distribution. Elles sont distribuées principalement par des bancassurances (81%).

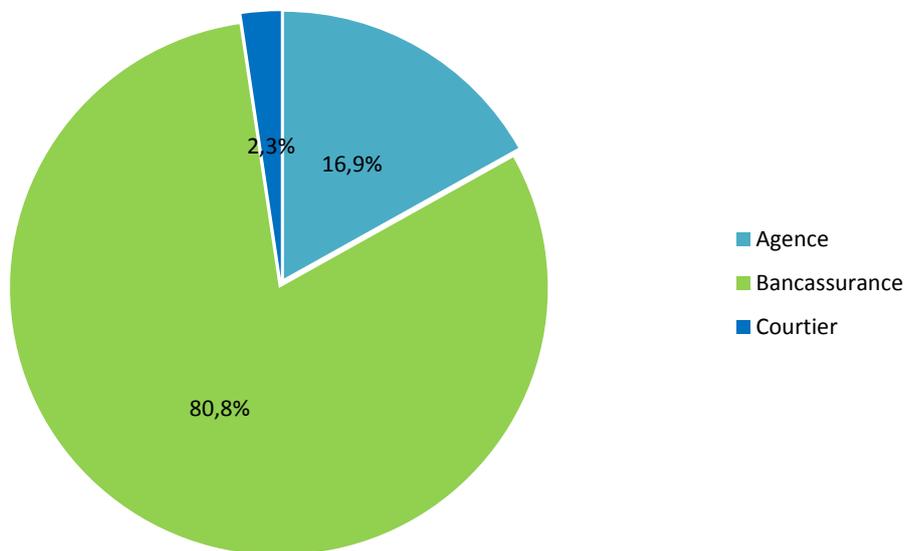


FIGURE 3.6 – Répartition des polices par réseau de distribution

e) Année de souscription

La base étudiée contient des polices issues de 1995 à 2018. L'histogramme suivant montre la répartition du nombre de polices rachetées par année de souscription.

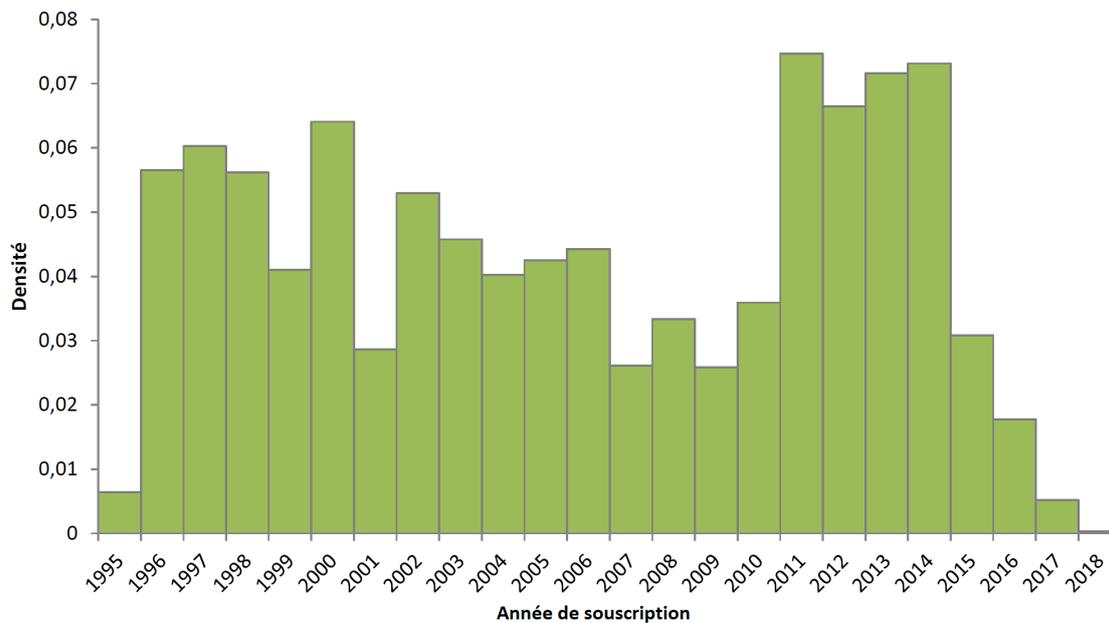


FIGURE 3.7 – Distribution des polices rachetées par année de souscription

f) Ancienneté

D'après l'histogramme suivant, les polices rachetées ont en général entre un et cinq années d'ancienneté. Nous notons que la plupart des polices étudiées n'autorisent pas le rachat avant une année d'ancienneté ce qui explique la faible densité des polices rachetées avant une année d'ancienneté.

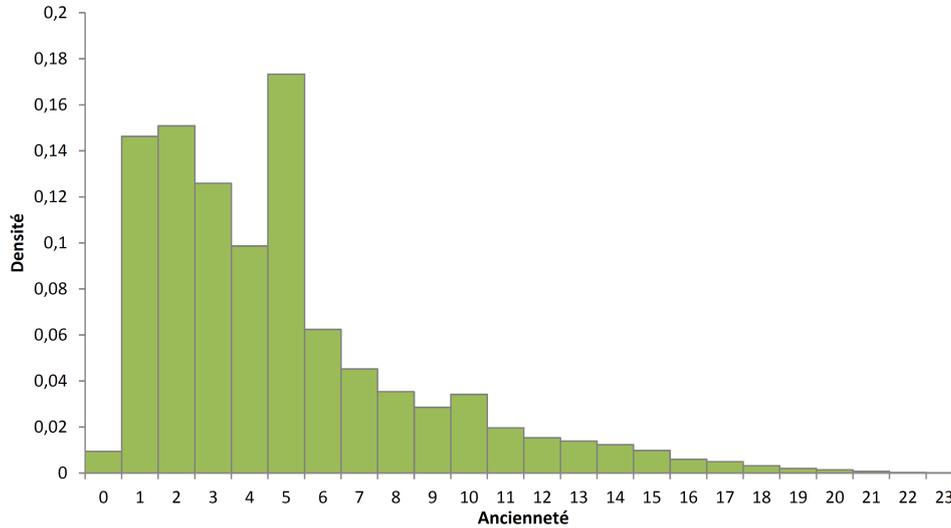


FIGURE 3.8 – Répartition des polices rachetées par ancienneté

g) Taux de rachat en fonction de l'âge au rachat

Comme le nombre d'observations pour les âges inférieurs à 20 ans et supérieurs à 80 ans est très faible, nous observons une très grande fluctuation du taux de rachat pour ces âges. Les âges pour lesquels nous disposons d'un nombre suffisant d'observations, le taux de rachat est relativement stable et varie autour de 23%.

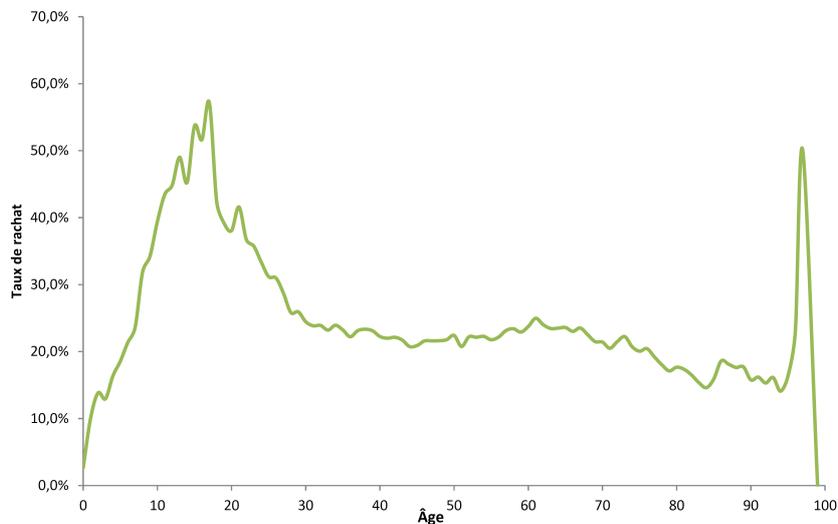


FIGURE 3.9 – Courbe du taux de rachat en fonction de l'âge au rachat

3.3 Modèles linéaires généralisés

3.3.1 Modèles linéaires

Les modèles linéaires sont conçus pour prédire et expliquer les valeurs d'une variable aléatoire Y dite variable à expliquer ou variable réponse à partir d'autres variables aléatoires X_1, \dots, X_p dites variables explicatives. On cherche ainsi une relation entre les n variables explicatives et la variable réponse Y .

Si $Y \in L^1 \left(\mathbb{E}(|Y|) < \infty \right)$, nous définissons la fonction de régression de Y sur $X = (X_1, \dots, X_p)$ par :

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ x &\longrightarrow \mathbb{E}(Y|X = x) \end{aligned}$$

La régression linéaire repose sur l'hypothèse de linéarité de la fonction de régression par rapport aux variables explicatives. Dans le cas où le modèle linéaire est non applicable, par exemple si la variable Y n'est pas continue, nous utilisons la généralisation des modèles linéaires où la linéarité est supposée entre les variables explicatives et une fonction lien appliquée à la fonction de régression de Y sur X . Nous présentons d'abord le modèle de régression linéaire simple ensuite nous exposons le modèle linéaire généralisé, le modèle de régression logistique.

a) La régression linéaire

Nous notons par Y la variable réponse qu'on souhaite prédire et par X_1, \dots, X_p les variables explicatives. Nous supposons que nous disposons de n observations. Si nous disposons des variables explicatives qualitatives, nous n'avons qu'à ajouter autant de variables explicatives que de modalités qui prennent les valeurs 0 ou 1.

Dans le cadre de la régression linéaire, f est supposée linéaire tel que pour tout x_1, \dots, x_p :

$$f(x_1, \dots, x_p) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

où $\beta = (\beta_0, \beta_1, \dots, \beta_p)^t$ constitue le vecteur des paramètres de la fonction f .

Résidus Soit ξ une variable aléatoire représentant l'erreur stochastique de la prédiction de Y . Cette erreur regroupe aussi l'impact des variables non prises en compte dans le modèle. Nous prenons souvent $\xi_i \sim \mathcal{N}(0, \Sigma)$

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \xi$$

Un estimateur du vecteur des paramètres β est l'estimateur des moindres carrées $\hat{\beta}_{MC}$ défini comme suit :

$$\hat{\beta}_{MC} = \underset{\beta \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \|Y - X\beta\|^2$$

Si X est inversible, alors nous avons,

$$\hat{\beta}_{MC} = (X^t X)^{-1} X^t Y$$

3.3.2 Modèles linéaires généralisés

La généralisation du modèle linéaire remplace l'hypothèse de linéarité de la fonction f par l'hypothèse de linéarité d'une fonction g de l'espérance conditionnelle telle que :

$$g(\mathbb{E}(Y|X)) = X\beta$$

avec $Y = (Y_1, \dots, Y_n)^t$

$X = (X_1, \dots, X_p)$ et $X_i = (1, X_{i1}, \dots, X_{in})^t \quad \forall i = 1, \dots, p$

g est une fonction inversible appelée fonction de lien. Le modèle linéaire classique est retrouvé si on prend g la fonction identité. Nous cherchons à estimer la fonction de régression :

$$\mathbb{E}(Y|X) = g^{-1}(X\beta)$$

a) La régression logistique : modèle logit

Dans le modèle de régression logistique La variable Y ne prend que deux valeurs 0 ou 1. La fonction de lien est la fonction logit définie comme suit :

$$g(x) = \log\left(\frac{x}{1-x}\right)$$

La loi de $Y|X$ est la loi Bernoulli de paramètre :

$$\pi(x) = \mathbb{P}(Y = 1|X = x) = \mathbb{E}(Y|X = x)$$

De plus :

$$\mathbb{E}(Y|X = x) = g^{-1}(X\beta) = \frac{e^{X\beta}}{1 + e^{X\beta}}$$

Fonction de vraisemblance La fonction de vraisemblance du modèle est définie par :

$$L(\beta) = \prod_{i=1}^n f_{\beta}(x_i)$$

avec f_{β} la fonction de densité de $Y|X$:

$$f_{\beta}(x) = \pi(x)^y (1 - \pi(x))^{1-y}$$

Donc la vraisemblance est :

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i}$$

La fonction de log-vraisemblance est donc définie comme suit :

$$\begin{aligned} \mathcal{L}(\beta) &= \log(L(\beta)) \\ &= \sum_{i=1}^n \left(y_i \log(\pi(x_i)) + (1 - y_i) \log(1 - \pi(x_i)) \right) \\ &= \sum_{i=1}^n \left(y_i (x_i \beta - \log(1 + e^{x_i \beta})) - (1 - y_i) \log(1 + e^{x_i \beta}) \right) \\ &= \sum_{i=1}^n \left(y_i x_i \beta - \log(1 + e^{x_i \beta}) \right) \end{aligned}$$

Estimateur du maximum de vraisemblance des paramètres Nous estimons les paramètres du modèle $(\beta_0, \beta_1, \dots, \beta_p)$ par l'estimateur de maximum de vraisemblance $\hat{\beta}_{MV}$ solution de l'équation suivante :

$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta} = 0$$

Nous calculons d'abord le gradient de la log-vraisemblance :

$$\begin{aligned} \forall j \in \llbracket 0 ; p \rrbracket \quad \frac{\partial \mathcal{L}(\beta)}{\partial \beta_j} &= \sum_{i=1}^n \left(y_i x_{ij} - \frac{x_{ij} e^{x_i \beta}}{1 + e^{x_i \beta}} \right) \\ &= \sum_{i=1}^n x_{ij} \left(y_i - \pi(x_i) \right) \end{aligned}$$

Pour $j = 0$ nous avons :
$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta_0} = \sum_{i=1}^n \left(y_i - \pi(x_i) \right).$$

Nous vérifions la convexité de la Hessienne de la log-vraisemblance.

$$\frac{\partial \mathcal{L}(\beta)}{\partial \beta_j \partial \beta_l} = - \sum_{i=1}^n x_{ij} x_{il} \frac{e^{x_i \beta}}{(1 + e^{x_i \beta})^2} \quad \forall j \in \llbracket 0 ; p \rrbracket, \quad \forall l \in \llbracket 0 ; p \rrbracket$$

La solution à l'équation $\nabla \mathcal{L}(\beta) = 0$ n'a pas une forme analytique. Des méthodes numériques sont utilisées pour approximer la solution telles que l'algorithme de Newton Raphson. Dans le cadre des GLM, La méthode itérative de Newton Raphson consiste à résoudre un système de moindres carrés pondérés. Cet algorithme est appelé les moindres carrés pondérés itérés et est implémenté sous R.

Un vecteur $\beta^0 = (\beta_0^0, \beta_1^0, \dots, \beta_p^0)$ est fixé au départ, ensuite, l'algorithme utilise la relation de récurrence suivante :

$$\beta^{i+1} = \beta^i - \left(\frac{\partial \mathcal{L}(\beta)}{\partial \beta^i \partial \beta^{it}} \right)^{-1} \times \left(\frac{\partial \mathcal{L}(\beta)}{\partial \beta^i} \right)$$

L'algorithme s'arrête lorsque la quantité $|\beta^{i+1} - \beta^i|$ est suffisamment faible.

b) Régression pénalisée

L'estimateur des moindres carrés présente des inconvénients. Nous rappelons que cet estimateur est défini comme suit :

$$\hat{\beta}_{MC} = \operatorname{argmin}_{\beta \in \mathbb{R}^{p+1}} \|Y - X\beta\|^2$$

Cet estimateur est non satisfaisant en raison de :

- L'interprétabilité du modèle : Nous sommes souvent amené à sélectionner un sous-ensemble de variables explicatives afin de pouvoir ressortir l'ordre d'importance des variables explicatives. Nous cherchons ainsi un estimateur parcimonieux (ayant beaucoup de coefficients nuls).
- La précision de l'estimation : le biais de l'estimateur des moindres carrés est petit tandis que sa variance est assez importante. Une réduction du nombre de coefficient du modèle peut améliorer la précision de l'estimateur en réduisant la variance.

La régression pénalisée est utilisée afin de contrôler le paramètre du modèle. Nous présentons dans ce qui suit la régression LASSO, Ridge et Elastic Net.

Régression LASSO Le LASSO (*Least Absolute Shrinkage and Selection Operator*) est une méthode de réduction des coefficients de régression. Cette méthode, présentée en 1996 par Robert Tibshirani [Tibshirani, 1996], est appliquée aux modèles GLM. Elle consiste à appliquer une pénalisation proportionnelle à la norme L_1 du vecteur de paramètres β lors du calcul de la fonction de vraisemblance ($\|\beta\|_{L_1} = \sum_i |\beta_i|$).

Dans le cadre du modèle de régression logistique, la nouvelle fonction de vraisemblance à minimiser est la suivante :

$$\sum_{i=1}^n \left(y_i x_i \beta - \log(1 + e^{x_i \beta}) \right) + \lambda \|\beta\|_{L_1}$$

Régression Ridge Afin de maîtriser l'amplitude des valeurs des coefficients β_0, \dots, β_p , une pénalité à la norme L_2 du vecteur de coefficient β est utilisée ($\|\beta\|_{L_2}^2 = \sum_i \beta_i^2$). Ce paramètre permet de pénaliser les grandes valeurs de β .

L'estimateur de ridge est donc la solution du problème de minimisation de la fonction suivante

$$\sum_{i=1}^n \left(y_i x_i \beta - \log(1 + e^{x_i \beta}) \right) + \lambda \|\beta\|_{L_2}^2$$

Régression Elastic Net L'Elastic net est une combinaison entre la régression Ridge et Lasso. L'estimateur Elastic Net est donc défini comme suit :

$$\hat{\beta} = \sum_{i=1}^n \left(y_i x_i \beta - \log(1 + e^{x_i \beta}) \right) + \lambda \left(\alpha \|\beta\|_{L_1} + (1 - \alpha) \|\beta\|_{L_2}^2 \right)$$

Nous retrouvons pour $\alpha = 0$ la méthode LASSO, et la méthode Ridge pour $\alpha = 1$. Nous notons qu'une transformation préalable est à faire avant d'appliquer l'algorithme de la régression pénalisée. Cette transformation consiste à normaliser les variables. En fait, nous devons transformer tous les variables en variables centrées réduites. Donc toutes les variables seront d'espérance nulle et de variance égale à 1. Cela permet à l'algorithme de comparer les différents variables puisqu'ils s'avèrent sur la même échelle.

3.4 Méthode d'agrégation d'arbre : Random Forest

L'algorithme d'apprentissage de la forêt aléatoire est une technique d'agrégation d'arbres de décision. L'algorithme construit une large collection d'arbres de décision non corrélés et renvoie comme prédiction la moyenne des prédictions des arbres de la forêt.

Nous introduisons d'abord les arbres de décision CART, ensuite nous présentons l'algorithme de la forêt aléatoire.

3.4.1 Arbre de régression et de classification : CART

Les arbres de décision CART constituent un outil d'apprentissage statistique conçu pour la régression et la classification selon le type de variable à expliquer. Ce sont des méthodes non paramétriques, c'est-à-dire qu'elles ne nécessitent pas d'hypothèse sur la distribution des données. Elles bénéficient

d'une structure hiérarchique qui les rend lisibles et faciles à interpréter.

Un arbre de décision est construit à l'aide d'une séquence récursive de règle de division exécutée pour chaque noeud de l'arbre à partir de la racine, qui regroupe tous les individus de la base, jusqu'à arriver à une condition d'arrêt prédéfinie. À chaque étape d'itération, une variable aléatoire et un critère de séparation sont désignés pour optimiser l'homogénéité des deux régions issues de la division.

Les deux nouvelles régions sont disjointes. De ce fait, Chaque individu ou observation n'appartient qu'à un unique noeud terminal. Le prédicteur de l'arbre retourne pour cet individu la valeur agrégée du groupe auquel il appartient.

Après la création de l'arbre, K régions ou noeuds R_1, \dots, R_K sont obtenus et la régression de Y est la fonction suivante.

$$f(x) = \sum_{j=1}^K c_j \mathbb{1}_{R_j}(x)$$

où les c_j $j = 1, \dots, K$ sont des constantes à estimer qui présentent la valeur attribuée à un individu appartenant au noeud R_j .

L'estimateur de la constante c est la solution du problème de minimisation de la somme des carrés des erreurs :

$$\min_c \sum_{i=1}^n (y_i - f(x))^2$$

$\forall j \in \llbracket 1 ; K \rrbracket$, un estimateur de c_j est donné par :

$$\forall j \in \llbracket 1 ; K \rrbracket \quad c_j = \frac{1}{N_j} \sum_{x_i \in R_j} y_i$$

avec $N_j = \#\{i | x_i \in R_j\}$: cardinal du noeud R_j .

La procédure de création de l'arbre commence par la division de la racine en deux noeuds fils. Un couple de division est d'abord choisi. Le couple est constitué par une variable explicative et un critère de séparation ou de division relatif à cette variable. Nous obtenons, après le découpage, deux partitions plus homogènes. Le découpage crée deux régions plus homogènes que le noeud dont ils sont originaires.

a) Fonction d'hétérogénéité

Le but de l'arbre de décision est de trouver, pour chaque noeud, la division, ou encore la variable explicative et la règle de séparation, qui minimise l'hétérogénéité des deux noeuds fils. Nous définissons ainsi la fonction d'hétérogénéité.

Dans le cas de la régression, l'hétérogénéité d'un noeud R_j , $\forall j \in \llbracket 1 ; K \rrbracket$, est défini de la manière suivante :

$$D_{R_j} = \frac{1}{N_j} \sum_{i \in R_j} (y_i - \hat{c}_j)^2$$

b) Critère de division

Une division est dite admissible si aucun des deux noeuds issus du découpage n'est vide. L'algorithme de l'arbre de décision privilégie la sélection de variables explicatives ayant beaucoup de

modalités. Cela est dû à la souplesse procurée à l'algorithme lors de la construction des deux sous-régions. Nous cherchons à diminuer le nombre de modalités des variables qualitatives car cela pourrait conduire à un sur-apprentissage.

Le critère de division choisi pour un noeud donné est basé sur la fonction d'hétérogénéité. L'algorithme récupère la division optimale, parmi toutes les divisions admissibles, qui assure la minimisation de l'hétérogénéité de deux noeuds fils issus de la division. Une division est définie par un couple (X_j, s) avec

- X_j est la variable aléatoire sélectionnée pour le découpage
- s représente le critère ou la frontière de séparation du noeud.
 - Si X_j est quantitative : s représente le seuil de séparation entre les deux régions. Si nous prenons R_1 et R_2 les deux noeuds issus du découpage, ils seront définis comme suit :

$$R_1 = \{i | x_{ij} \leq s\} \qquad R_2 = \{i | x_{ij} > s\}$$

- Si X_j est qualitative : s représente l'ensemble des modalités possible de X_j . Ainsi, Les deux noeuds issus du découpage R_1 et R_2 , ils seront définis comme suit :

$$R_1 = \{i | x_{ij} \in s\} \qquad R_2 = \{i | x_{ij} \notin s\}$$

Nous effectuons la sélection du couple du découpage de façon à minimiser l'hétérogénéité des deux noeuds fils.

$$\min_{c_1} \sum_{x_i \in R_1} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2} (y_i - c_2)^2 \quad (\bullet)$$

La minimisation de la quantité ci-dessus conduit à la maximisation de l'homogénéité des deux noeuds fils. Les constantes c_1 et c_2 qui minimisent les deux quantités sont les suivantes :

$$\hat{c}_1 = \frac{1}{N_1} \sum_{x_i \in R_1} y_i \qquad \hat{c}_2 = \frac{1}{N_2} \sum_{x_i \in R_2} y_i$$

Nous pouvons calculer pour chaque variable aléatoire explicative le seuil correspondant à un découpage optimal pour la variable en question. Ensuite, nous déterminons le couple (X_j, s) qui minimise la quantité (\bullet) ci-dessus. Nous itérons les étapes de recherche de coupure optimale pour chaque nouveau noeud jusqu'à atteindre la règle d'arrêt choisie au préalable.

c) Critère d'arrêt

La croissance de l'arbre est un processus qui s'arrête lorsqu'une condition d'arrêt est déclenchée. Cette condition peut être activée lorsqu'il n'existe plus de division admissible ou lorsque le nombre d'observations des noeuds terminaux est inférieur à une valeur seuil.

Le choix de la condition d'arrêt est crucial. En fait, si l'arbre est très profond, nous nous retrouvons dans le cas de sur-apprentissage tandis qu'un arbre peu profond peut conduire à un sous-apprentissage. Afin de contourner ce problème nous pouvons adopter différentes approches. Une première approche consiste à itérer le découpage jusqu'à ce que la somme des carrés des erreurs ne décroisse plus suffisamment. Cette approche présente un inconvénient car un premier découpage peut ne pas améliorer

la prédiction mais peut être suivi par un autre découpage assez pertinent.

Une autre approche consiste à construire un arbre de grande taille \mathcal{T}_{max} . Le découpage ne s'arrête que lorsque la taille de chaque noeud terminal excède une taille limite. Un élagage est ensuite appliqué. L'élagage permet d'obtenir un arbre optimal de taille entre celle de l'arbre trivial d'une seule feuille et l'arbre de taille maximale \mathcal{T}_{max} .

d) Élagage

L'élagage est une technique qui consiste à construire une série emboîtée de sous-arbres à partir de l'arbre maximal \mathcal{T}_{max} . Un sous arbre \mathcal{T} de \mathcal{T}_{max} ($T \subset \mathcal{T}_{max}$) est un arbre obtenu par suppression d'un nombre arbitraire de noeuds non terminaux. L'arbre optimal sera choisi parmi les arbres de la suite construite.

Nous posons $K_{\mathcal{T}}$ le nombre de noeuds terminaux de l'arbre \mathcal{T} et nous cherchons, pour tout $\alpha \geq 0$, le sous arbre \mathcal{T}_{α} qui minimise C_{α} , défini comme suit :

$$C_{\alpha}(\mathcal{T}) = \sum_{i=1}^{|\mathcal{T}|} N_i D_i(\mathcal{T}) + \alpha K_{\mathcal{T}}$$

avec $|\mathcal{T}|$: nombre de noeuds de l'arbre \mathcal{T}

α est un paramètre de pénalisation qui présente le compromis entre la taille de l'arbre et l'adéquation de l'ajustement aux données. Pour $\alpha = 0$, la solution est l'arbre maximal \mathcal{T}_{max} qui maximise l'ajustement aux données et donc minimise l'hétérogénéité des noeuds. Nous retenons comme arbre minimal celui qui correspond à la valeur de α minimisant le critère d'hétérogénéité.

3.4.2 Random Forest

Les algorithmes de la forêt aléatoire sont des algorithmes d'apprentissage supervisé introduits par Léo Breiman[Breiman, 2001]. L'algorithme de la forêt aléatoire est connu par la précision de sa prédiction et sa capacité d'exécution sur les bases de données au nombre d'observations réduit et de nombre de variables explicatives élevé. Cette technique est utilisée afin d'améliorer le pouvoir prédictif d'un seul arbre de décision.

L'algorithme construit des arbres de manière indépendante et en utilisant à chaque fois un ré-échantillonnage aléatoire de la base d'apprentissage appelé *Boosting*.

a) Ré-échantillonnage aléatoire : *Bagging*

La technique de ré-échantillonnage est utilisée lors de la construction des arbres de la forêt aléatoire. L'algorithme de *Bagging* sert à construire un grand nombre d'arbres à partir plusieurs échantillons bootstrapés. En fait, un tirage équiprobable avec remis des observations de la base est effectué afin de créer un nouvel échantillon de même taille que l'échantillon initial. Cette opération est répétée B fois (avec B très grand). Les B arbres de la forêt sont développés à partir des échantillons tirés.

De plus, lors de la division d'un noeud, un sous-ensemble de variables explicatives est sélectionné. Cette sélection est utilisée afin de diminuer la corrélation entre les arbres particulièrement lorsque certaines variables possèdent plus de pouvoir explicatif que d'autres variables ce qui fait qu'ils sont inclus dans la plupart des arbres.

b) Algorithme de la forêt aléatoire

Les étapes de l'algorithme de la forêt aléatoire sont présentées ci-dessous :

1. Pour $b = 1, \dots, B$ avec B : nombre total de bootstraps.
 - a. n observations sont sélectionnées avec remise parmi les n observations de la base. Cette étape présente la technique de *Bagging*.
 - b. Un arbre de décision aléatoire \mathcal{T}_b est créé en utilisant l'échantillon bootstrapé. Les étapes décrites ci-dessous sont itérées jusqu'à atteindre la condition d'arrêt.
 - l variables sont sélectionnées aléatoirement à partir des p variables de la base.
 - Une variable est choisie parmi les l variables ainsi qu'une condition de division. Le couple est choisi de façon à minimiser l'hétérogénéité des noeuds fils.
 - Deux noeuds fils sont créés.
2. Une forêt d'arbre est créée : $\{\mathcal{T}_b\}_{b=1, \dots, B}$
3. Le prédicteur construit est ainsi \hat{f}_B :

$$\hat{f}_B(x) = \frac{1}{B} \sum_{b=1}^B \hat{\mathcal{T}}_b(x)$$

avec $\hat{\mathcal{T}}_b$ le prédicteur de l'arbre \mathcal{T}_b

L'agrégation des arbres effectuée par l'algorithme de *Bagging* permet de réduire la variance du modèle. De plus, la sélection aléatoire des variables utilisées pour la construction des arbres permet de minimiser la corrélation entre les arbres. En effet, si nous supposons que les arbres ont une variance de σ^2 et qu'ils sont deux à deux corrélés par ρ , alors :

$$\begin{aligned} \text{Var}(\hat{f}_B(x)) &= \text{Var}\left(\frac{1}{B} \sum_{b=1}^B \hat{\mathcal{T}}_b(x)\right) \\ &= \frac{1}{B^2} \sum_{b=1}^B \left(\sigma^2 + \sum_{b'=1, b' \neq b}^B \rho \sigma^2\right) \\ &= \frac{\sigma^2}{B^2} \sum_{b=1}^B (1 + (B-1)\rho) \\ &= \rho \sigma^2 + \frac{(1-\rho)\sigma^2}{B} \end{aligned}$$

Plus le nombre d'itérations croît plus la variance décroît.

c) Importance des variables : Erreur *Out-of-bag*

La construction de la forêt se fait à l'aide d'arbres indépendants ce qui nous procure la possibilité de mesurer l'erreur par validation croisée en profitant du bootstrap déjà réalisé pour classifier l'importance des variables et leur contribution.

Soit une observation (x_i, y_i) , $i \in \llbracket 1 ; n \rrbracket$. Nous notons par F_{-i} la sous-forêt construite à partir d'un

échantillon ne contenant pas l'observation (x_i, y_i) .

L'erreur *Out-of-bag* est défini comme l'erreur de prédiction moyenne des arbres composant la forêt F_{-i} . Cette erreur nous permet de déterminer l'ordre d'importance des variables explicatives. Cela nous permet de déterminer l'importance de l'information marginale qu'une variable aléatoire apporte à la prédiction de la forêt.

d) Hyperparamétrage de la forêt aléatoire

Le paramétrage du modèle est une étape primordiale puisque la performance de l'algorithme est étroitement dépendante du jeu de paramètre choisi.

Nous présentons dans ce qui suit les paramètres qui affectent le plus le modèle.

- *Nombre d'arbres* : Afin d'obtenir une prédiction stable et robuste, un très grand nombre d'arbre est nécessaire. Cependant le temps de calcul augmente avec le nombre d'arbres ce qui peut rendre l'algorithme très lent. Une approche utilisée pour remédier à ce problème est de choisir un grand nombre d'arbre à construire B et d'arrêter le processus de génération d'arbre lorsque l'erreur *Out-of-bag* ne décroît pas suffisamment.
- *Dimension de l'échantillon de création des arbres* K : la dimension de l'échantillon à bootstraper est un paramètre à prendre en considération. Un équilibre doit être fait entre l'adéquation des données pour l'arbre à créer et l'indépendance des arbres.
- *Nombre de variables de création des arbres* l : Le nombre de variables utilisées pour déterminer le meilleur découpage est choisie de façon à ne pas se retrouver dans le cas de sur-échantillonnage ou de sous échantillonnage. En fait, si $l = n$ nous serons confrontés au problème de sur-échantillonnage. Cependant, si $l = 1$ l'algorithme sera totalement aléatoire.
- *La profondeur maximale des arbres* : Si la profondeur des arbres est très grande, alors l'échantillon est parfaitement répliqué par l'arbre (sur-échantillonnage). Si cette profondeur est fixée à 1, alors pour chaque arbre la racine sera divisée en deux sous noeuds et la prédiction pour chaque région sera égal à la moyenne des valeurs observées pour de chaque noeud.

3.5 *eXtreme Gradient Boosting*

3.5.1 Les méthodes *Boosting*

Les méthodes de prédiction *Boosting* sont basées, comme le Random Forest, sur l'agrégation d'arbre de décision. La différence avec le Random Forest consiste dans le fait que la génération des arbres se fait en série. En effet, Le *Boosting* génère des arbres reliés c'est-à-dire que chaque arbre a accès à son prédécesseur. La création des arbres a donc pour objectif d'améliorer la prédiction de l'arbre précédent en affectant plus de poids aux observations mal classées de son prédécesseur. Ces méthodes s'occupent mieux des observations difficiles à classer et évitent en même temps le problème de sur-apprentissage en agrégeant les arbres.

Différents algorithmes de *Boosting* existent. Ils sont utilisés pour agréger des modèles ou agréger les arbres. Ils diffèrent selon la méthode d'agrégation d'arbre, la méthode de traitement des observations

mal classées ou leur fonction de perte.

Nous utilisons dans notre étude la méthode de XGBoost (*eXtreme Gradient Boosting*) qui est une variante de l'algorithme de *Gradient Boosting*. Nous présentons dans ce qui suit la méthode de *Gradient Boosting*.

3.5.2 Gradient Boosting

Le modèle *Gradient Boosting* est appelé ainsi puisqu'un algorithme de descente du gradient est utilisé pour minimiser la fonction de perte lors de la génération des arbres. Soit \mathcal{T}_b un arbre construit à l'étape b . Soit f_b le modèle à l'étape b :

$$f_b = \sum_{k=1}^b \mathcal{T}_k$$

Soit E la fonction de perte du modèle. Nous considérons comme fonction de perte la fonction des moindres carrées E défini comme suit :

$$E = \sum_{i=1}^n \frac{(Y_i - f(X_i))^2}{2}$$

Le gradient de E par rapport à $f(X_i)$ est :

$$\frac{\partial E}{\partial f(X_i)} = \frac{1}{2} \sum_{k=1}^n \frac{\partial (Y_k - f(X_k))^2}{\partial f(X_i)} = f(X_i) - Y_i$$

L'algorithme compile de la manière suivante :

- Un premier arbre de décision \mathcal{T}_1 est construit. Le résidu de la valeur observée Y_i est $res_1 = Y_i - f_1(X_i) = -\frac{\partial E}{\partial f_1(X_i)}$
- Pour $b = 2, \dots, B$ itérations, les étapes suivantes sont itérées.
 - L'algorithme construit l'arbre \mathcal{T}_b qui prédit les résidus de l'étape précédente afin de combler les manques de l'arbre \mathcal{T}_{b-1} . Le modèle construit à l'étape b est donc $f_b(X) = \mathcal{T}_{b-1}(X) + \mathcal{T}_b(X)$.
 - L'arbre \mathcal{T}_b est égal à $\mathcal{T}_b(X_i) = (Y_i - f_{b-1}(X_i)) + (f_b(X) - Y_i)$. Le résidu est donc : $res_b = Y_i - f_b(X_i) = -\frac{\partial E}{\partial f_b(X_i)}$

3.5.3 eXtreme Gradient Boosting

L'algorithme *eXtreme Gradient Boosting* est une version de l'algorithme *Gradient Boosting*. Cet algorithme est connu par son capacité prédictive et son vitesse d'exécution. Comme les méthodes *Boosting*, *XGBoost* vise à réduire le biais, tout en contrôlant la variance. Nous rappelons que les algorithmes de prédiction utilisent une fonction appelée la fonction objective afin de mesurer la robustesse de l'algorithme. Cette fonction est la somme de deux fonctions : la fonction de perte et la fonction de régularisation. La fonction de perte mesure la capacité prédictive du modèle tandis que la fonction de régularisation contrôle la complexité du modèle et est utilisée afin d'éviter le problème de sur-apprentissage.

l'algorithme XGBoost construit des arbres de manière séquentielle. Si le modèle final contient B arbres, la prédiction finale de Y pour une observation x_i est :

$$\hat{y}_i = \sum_{b=1}^B \mathcal{T}_b(x_i)$$

L'algorithme initialise la valeur à prédire à 0.

$$\hat{y}_i^0 = 0$$

Comme décrit pour le *Gradient Boosting* la prédiction pour l'étape 1 est :

$$\hat{y}_i^1 = \hat{y}_i^0 + f_1(x_i)$$

Arrivant à l'itération b, nous avons :

$$\hat{y}_i^b = \hat{y}_i^{b-1} + \mathcal{T}_b(x_i) = \sum_{k=1}^b \mathcal{T}_k(x_i)$$

À chaque étape, nous déterminons $\mathcal{T}_b(x_i)$ L'arbre à choisir pour chaque étape est celui qui minimise la fonction d'objectif. L'arbre choisi pour chaque étape dépend de la forme de la fonction de régularisation et de la fonction de perte choisie.

a) XGBoost en pratique

L'algorithme de XGBoost a été implémenté sous R en utilisant le package "xgboost". Étant donné le nombre de paramètres de l'algorithme nous n'optimiserons que les paramètres suivants :

- *max_depth* : profondeur maximale d'un arbre.
- *nthread* : nombre de coeurs utilisés pour l'exécution en parallèle
- *eta* : Réduction de la taille d'étape utilisées lors de l'étape de mise à jour afin d'éviter le surajustement.
- *objective* : paramètre qui spécifie le type de fonction objectif utilisée. *objective*="binary :logistic" est spécifié pour la régression logistique pour la classification binaire.
- *metrics* : la métrique utilisée pour évaluer la validation croisée. *metrics* peut prendre les valeurs : "auc", "mse", "LogLoss", "rmse", etc.

3.6 Optimisation des paramètres des modèles

Afin de choisir la meilleure combinaison des paramètres nous pouvons tester toutes les combinaisons possibles. Cette démarche coûte énormément en temps de compilation et est quasi impossible à exécuter.

Une approche d'optimisation utilisée est la grille de recherche d'hyperparamétrage aléatoire. Cette méthode consiste à construire aléatoirement une grille de différentes combinaisons possibles des paramètres. Les différentes combinaisons sont ensuite testées jusqu'à ce que la mesure d'erreur ne décroisse plus suffisamment.

3.7 Résultats

Nous présentons dans cette partie les résultats donnés par les modèles de prédiction utilisés : *Elastic Net*, *Random Forest* et *XGBoost*. Nous rappelons que toutes les étapes du *machine learning* ont été appliquées en utilisant le logiciel R. La démarche de *data engineering* a été effectuée avant d'exécuter les modèles. Cette opération est la même pour les deux algorithmes de *Random Forest* et *XGBoost*.

3.7.1 *Glmnet - Elastic Net*

L'implémentation du modèle Elastic Net est faite par le package "*glmnet*" qui nous permet d'appliquer la régression pénalisée. Avant d'appliquer ce modèle nous effectuons une autre fois l'étape de *data engineering* afin de transformer quelques variables quantitatives en variables à modalités. Cette opération aide à améliorer notre prédiction ainsi qu'à rendre le modèle plus facile à interpréter. Les variables à transformer sont les suivantes : "*Age.at.reversal*", "*Duration.still.to.go*", "*Holding.period*" et "*Age.at.subscription*". Pour chacune de ces variables, nous avons créé des modalités en se basant sur l'analyse de données effectuée sur la base des données. Cette opération nous a permis d'améliorer considérablement la prédiction. Nous citons l'exemple du log Loss qui est passé de 0.27 à 0.09 pour la prédiction appliquée sur la base test.

Variables qualitatives ajoutées La transformation des quatre variables, présentée dans le paragraphe précédent, est effectuée après l'analyse de la distribution de ces variables et leur relation avec le rachat.

Nom de la variable	Nom de la nouvelle variable	Modalités de la nouvelle variable
Holding.Period	HP	<=1
		2-6
		7-10
		>10
Age.at.reversal	AaR	<=20
		20-50
		51-80
		>80
Age.at.subscription	AaS	<=20
		20-50
		51-80
		>80
Duration.still.to.go	DSTG	<=20
		21-80
		>80

TABLE 3.2 – Tableau des nouvelles variables issues de la phase *data engineering*

Nous présentons les résultats du modèle *Elastic Net* pour une seule combinaison de valeurs. Cette combinaison a été choisie après exécution de l'algorithme pour une grille de valeur.

La figure suivante présente l'importance des variables pour l'algorithme *Elastic Net*

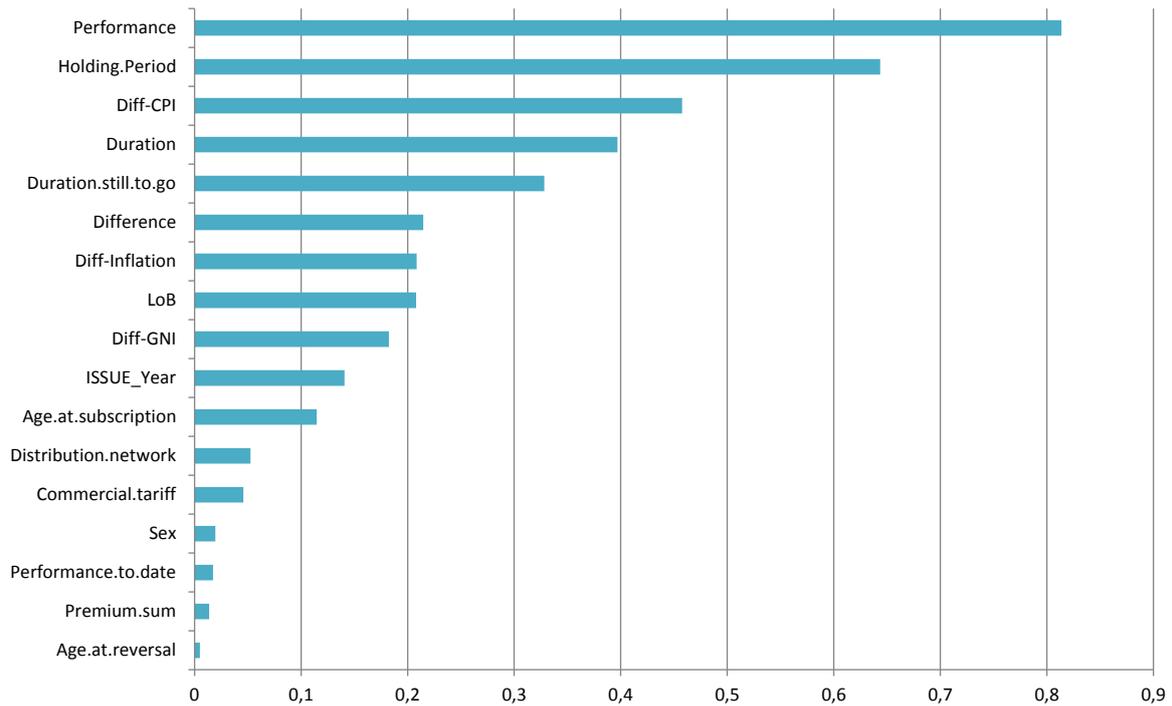


FIGURE 3.10 – Importance des variables explicatives dans le modèle *Elastic Net*

Nous présentons dans le tableau suivant les coefficients de quelques variables explicatives.

Nom	Coefficient
(Intercept)	-1,738804497
Premium.sum	-0,006374593
Performance.to.date	-0,004964805
Performance	0,338879616
Diff-CPI	-0,326733573
Issue_Year	-0,003531719
Difference	-0,104713162
Duration	0,330230313
Sex.F	-0,102108608
Sex.M	-0,063983419
LoB.MULTIRAMO	0,173278484
HP.7-10	0,968153495
AaS.>80	-0,2467145
AaR.20-50	0,133993137
DSTG.<=20	-0,467554012

TABLE 3.3 – Tableau des coefficients de l'algorithme Elastic Net

Les indicateurs de performance du modèle *ElasticNet* sont calculés en utilisant la base test. le tableau suivant résume les valeurs trouvées :

Modèle	TopDecileLift	LogLoss	AUC	RMSE
<i>Elastic Net</i>	6.643	0.156	0.97	0.2

TABLE 3.4 – Tableau des indicateurs de performance de l'algorithme *Elastic Net* appliqué sur la base *Test*

3.7.2 *Random Forest*

L'algorithme *Random Forest* est utilisé avec le package "*ranger*". Une validation croisée avec cinq folds (échantillons) est utilisée. La figure suivante présente l'importance des variables dans l'algorithme de *Random Forest*.

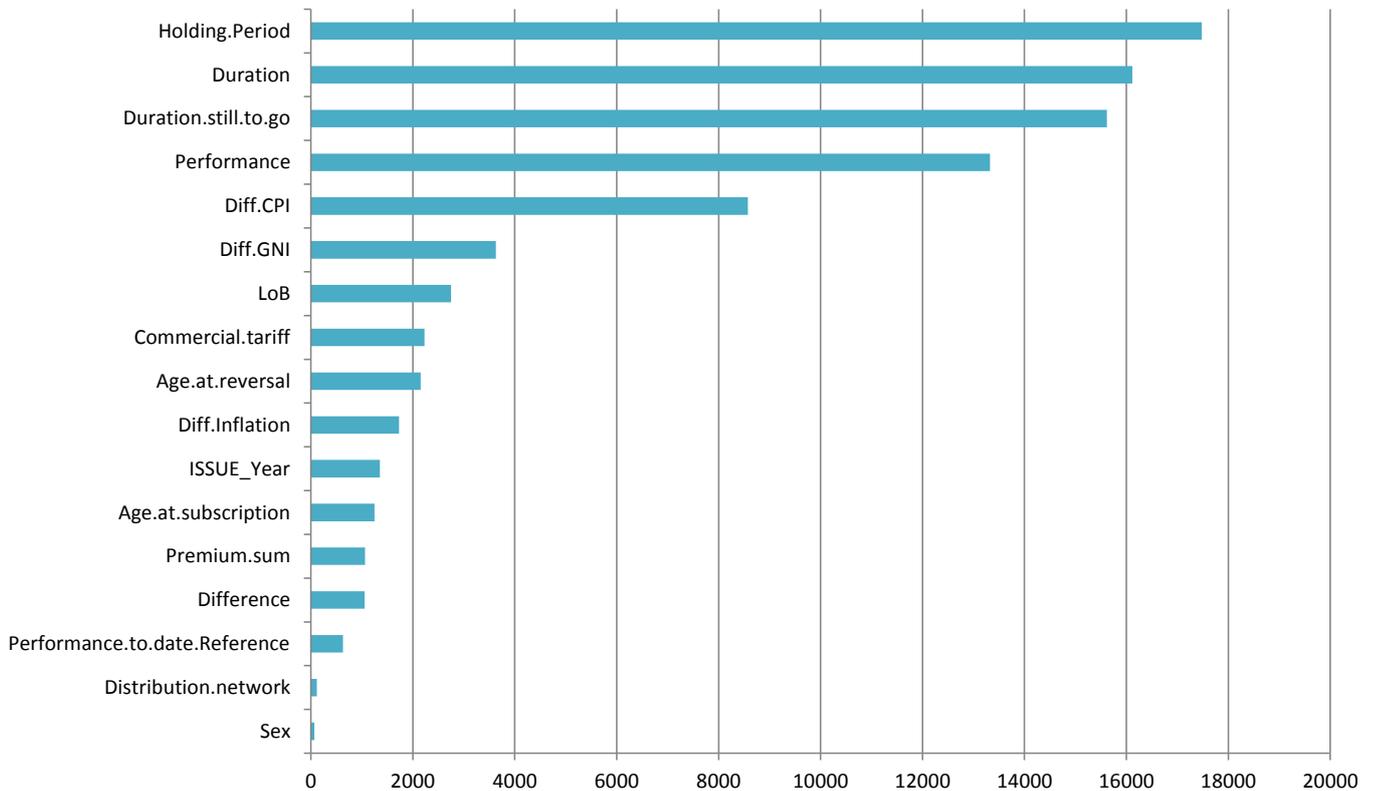


FIGURE 3.11 – Importance des variables explicatives dans le modèle *Random Forest*

La variable " *Holding Period* " semble être la variable la plus importante lors de la création de la forêt aléatoire. L'algorithme privilégie les variables donnant des informations sur la période restante dans la vie des polices. Les deux variables " *Sex* " et " *Distribution.network* " ont une importance très faible dans la prédiction. Les variables économiques ont une importance significative dans l'algorithme de *Random Forest*.

Les indicateurs de performance du modèle *Random Forest* sont présentés dans le tableau suivant :

Modèle	TopDecileLift	LogLoss	AUC	RMSE
<i>Random Forest</i>	5.61	0.44	0.96	0.19

TABLE 3.5 – Tableau des indicateurs de performance de l'algorithme *Random Forest* appliqué sur la base *Test*

3.7.3 *XGBoost*

L'algorithme *XGBoost* est exécuté pour un jeu de paramètre choisi afin de maximiser la prédiction. Nous présentons dans la figure suivante l'importance des variables dans l'algorithme.

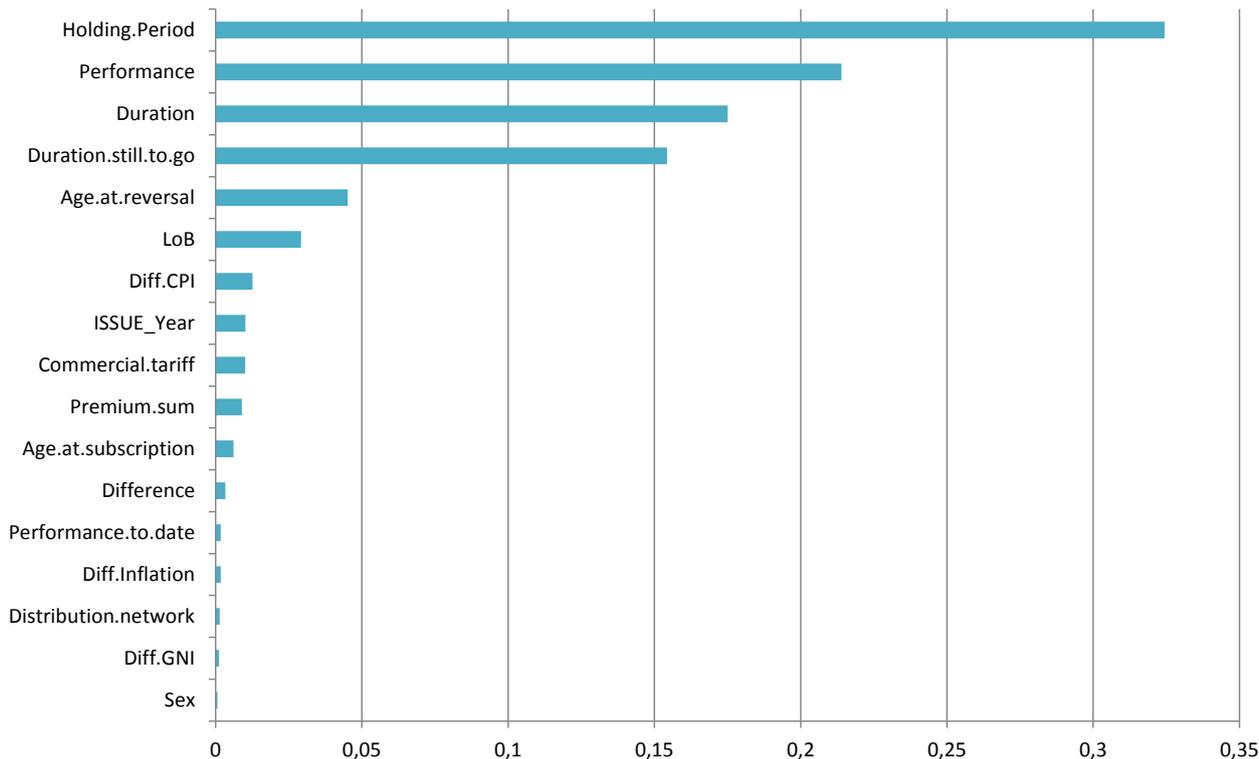


FIGURE 3.12 – Importance des variables explicatives dans le modèle *XGBoost*

L'algorithme de *XGBoost* affecte la même importance aux variables que l'algorithme *Random Forest*. Cela peut être expliqué par le fait que les deux algorithmes se basent sur les modèles d'arbre de décision. Le *XGBoost* accorde une très grande importance à la variable "Performance" que nous avons ajoutée. Il donne plus d'importance à l'âge à la souscription que le *Random Forest*.

Comme nous avons indiqué, l'algorithme *XGBoost* est connu pour sa robustesse et ses pouvoirs prédictifs. Le tableau suivant contenant les indicateurs de performance du modèle montre que l'algorithme *XGBoost* donne les meilleurs résultats pour tous les indicateurs utilisés.

Modèle	TopDecileLift	LogLoss	AUC	RMSE
<i>XGBoost</i>	7.154	0.053	0.996	0.12

TABLE 3.6 – Tableau des indicateurs de performance de l'algorithme *XGBoost* appliqué sur la base *Test*

3.7.4 Algorithme retenu

D'après les indicateurs de performance présentés ci-dessus, l'algorithme le plus performant est l'algorithme *XGBoost*. Cet algorithme est connu par ses pouvoirs prédictifs. Cet algorithme sert à prédire l'évènement de rachat individuel. Nous pouvons ainsi déterminer la probabilité de rachat de chaque police. Toutefois, ce le modèle Elastic Net surpasse l'algorithme *XGBoost* dans la détermination et l'étude des variables explicatives du rachat. En fait, Grace à sa facilité d'interprétation et de manipulation, le modèle Elastic Net est le modèle retenu pour l'étude des variables explicatives du rachat. Ce modèle nous permet de déterminer les variables les plus influentes sur le comportement de

rachat des assurés. De plus, le modèle de régression Elastic Net présente des résultats assez satisfaisants en matière de prédiction de l'évènement de rachat. L'algorithme Random Forest est utilisé comme modèle référence appelé "benchmark".

Caractéristiques et indicateurs	Random Forest	Elastic Net	<i>XGBoost</i>
Pouvoir prédictif	Médiocre	Moyen	Bien
Interprétabilité	Moyen	Bien	Moyen
Temps de compilation	Moyen	Médiocre	Bien
Manipulation de données non pertinentes	Bien	Médiocre	Bien
TopDecileLift	5,61	6,643	7,154
LogLoss	0,44	0,156	0,053
AUC	0,96	0,97	0,996
RMSE	0,19	0,2	0,12

TABLE 3.7 – Tableau de comparaison des algorithmes de prédiction utilisés

Conclusion

Le rachat est un risque qui peut compromettre la solvabilité des compagnies d'assurance. Dans le cadre de la directive Solvabilité II, les assureurs sont invités à étudier et bien évaluer leur exposition à ce risque.

Ce mémoire s'inscrit ainsi dans le cadre de révision des hypothèses de rachat de CA Vita, l'entité italienne de Crédit Agricole Assurances, et de l'étude du rachat par différentes approches selon le type de donnée à disposition. Les courbes de rachat obtenues seront utilisées dans la modélisation des produits d'épargne de l'assurance. Les projections faites permettent de déterminer les capitaux de solvabilité exigés.

Nous disposons de deux types de données : données agrégées et données individuelles. Nous utilisons des estimateurs comme l'estimateur de Hoem pour déterminer les taux de rachat par ancienneté de chaque produit par les données agrégées. Nous exploitons aussi les méthodes de lissage et d'extrapolation afin d'obtenir des courbes de rachat complètes. Les données individuelles sont traitées par les algorithmes de *machine learning* afin de déterminer la probabilité individuelle de rachat et déterminer les variables qui influencent le plus le comportement des assurés.

Les méthodes de *machine learning* nous permettent d'affiner les estimations effectuées par les estimateurs statistiques classiques. Elles nous donnent aussi la possibilité de mieux comprendre le comportement individuel des assurés. Toutefois, ces algorithmes sont trop dépendants des données utilisées pour la construction des modèles de prédiction. En fait, la qualité des prédictions est étroitement liée à la qualité des données. De plus, l'interprétabilité de ces modèles peut s'avérer compliquée.

L'étude du rachat par les algorithmes de *machine learning* peut être exploitée afin d'améliorer les stratégies de commercialisation des produits ou la tarification des produits d'assurance.

Bibliographie

- [ANIA, 2016] ANIA (2015-2016). Italian insurance in figures. Technical report.
- [Breiman, 2001] Breiman, L. (2001). Random forests. *Machine Learning*, pages 5–32.
- [Bühlmann and Gisler, 2005] Bühlmann, H. and Gisler, A. (2005). *A Course in Credibility Theory and its Applications*. Springer.
- [CAA, 2012] CAA (2012). Review documents ca vita, lapse risk.
- [Catherine,] Catherine, H. Analyse des durées de survie.
- [CEIOPA, 2007] CEIOPA (2007). Autres conseils à la commission européenne sur les questions du pilier 1. Technical report.
- [Christophe Dutang,] Christophe Dutang, Vincent Goulet, M. P. actuar : An r package for actuarial science. *Journal of Statistical Software*.
- [D.W. Hosmer, 2000] D.W. Hosmer, S. L. (2000). *Applied Logistic Regression, Second Edition*.
- [EIOPA, 2014] EIOPA (2014). Technical Specification for the Preparatory Phase. Technical report.
- [Frédéric, 2017] Frédéric, P. (2017). Modèles de durée, support de cours 2017-2018.
- [Hardy and Panjer, 1998] Hardy, M. and Panjer, H. (1998). A credibility approach to mortality risk. *ASTIN Bulletin*, 28(2) :269–283.
- [Lutz and Biernat, 2015] Lutz, M. and Biernat, E. (2015). *Data science : fondamentaux et études de cas : Machine learning avec Python et R*. Editions Eyrolles.
- [SwissRe, 2012] SwissRe (2012). The italian insurance market. Technical report.
- [Tibshirani, 1996] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1) :267–288.
- [Wikistat, a] Wikistat. Agrégation de modèles. [En ligne].
- [Wikistat, b] Wikistat. Sélection de modèle en régression linéaire. [En ligne].
- [Xavier, 2011] Xavier, M. (2011). Modèles de crédibilité avec régression au barycentre du temps. *Rapport de projet de fin d'études, INP Grenoble*.
- [Éric Samuel, 2016] Éric Samuel, N. N. (2016). Prédiction des comportements de rachat en épargne individuel : une approche machine learning.

Annexes

Démonstration du modèle de Bühlmann

Dans le modèle Bühlmann nous cherchons, $\forall i \in [1, n]$, le meilleur estimateur linéaire de $\mu(\Theta_i)$ qui minimise l'espérance de l'erreur.

Nous cherchons :

$$\hat{\mu}(\Theta_i) = \mathbb{E}\left(X_{i,T+1} | X_{i1}, \dots, X_{iT}\right) = a_{i0} + \sum_{t=1}^T a_{it} X_{it}$$

tel que $\mathbb{E}\left((\mu(\Theta) - \hat{\mu}(\Theta))^2\right) = \mathbb{E}\left((\hat{\mu}(\Theta_i) - X_{i,T+1})^2\right)$ soit minimal.

Autrement, nous cherchons :

$$(\hat{a}_{i0}, \dots, \hat{a}_{iT}) = \underset{a_{i0}, \dots, a_{iT}}{\operatorname{argmin}} \mathbb{E} \left(\left(\mu(\Theta) - a_{i0} - \sum_{t=1}^T a_{it} X_{it} \right)^2 \right)$$

L'estimateur de crédibilité est donné par :

$$\hat{\mu}(\Theta_i) = Z_i \bar{X}_i + (1 - Z_i) \mu$$

avec

$\mu = \mathbb{E}(\mu(\Theta_i))$: l'espérance globale des observations pour le risque individuel i

$Z_i = \frac{T}{T + \sigma^2/\tau^2}$: le facteur de crédibilité

Démonstration

Pour $i \in [1, n]$ fixé, la condition de premier ordre s'écrit comme suit :

$$\frac{\partial}{\partial a_{i,j}} \mathbb{E} \left(\left(\mu(\Theta_i) - a_{i0} - \sum_{t=1}^T a_{it} X_{it} \right)^2 \right) = 0 \quad \forall j \in [0, T]$$

Nous avons, $\mathbb{E}(X_{it}) = \mathbb{E}(\mathbb{E}(X_{ij} | \Theta_i)) = \mathbb{E}(\mu(\Theta_i)) = \mu$.

Donc, pour $t = 0$, nous avons :

$$\begin{aligned} \frac{\partial}{\partial a_{i,0}} \mathbb{E} \left(\left(\mu(\Theta_i) - a_{i0} - \sum_{t=1}^T a_{it} X_{it} \right)^2 \right) &= \mathbb{E} \left(\mu(\Theta_i) - a_{i0} - \sum_{t=1}^T a_{it} X_{it} \right) \\ &= \left(1 - \sum_{t=1}^T a_{it} \right) \mu - a_{i,0} \\ &= 0 \end{aligned}$$

Nous avons ainsi :

$$a_{i,0} = \left(1 - \sum_{t=1}^T a_{it}\right)\mu \quad (*)$$

$\forall j \in [1, T]$, nous obtenons :

$$\begin{aligned} \frac{\partial}{\partial a_{i,j}} \mathbb{E} \left((\mu(\Theta_i) - a_{i0} - \sum_{t=1}^T a_{it} X_{it})^2 \right) &= \mathbb{E} \left(X_{ij} (\mu(\Theta_i) - a_{i0} - \sum_{t=1}^T a_{it} X_{it}) \right) \\ &= \mathbb{E} \left(X_{ij} \mu(\Theta_i) \right) - a_{i0} \mu - \sum_{\substack{t=1 \\ t \neq j}}^T a_{it} \mathbb{E} \left(X_{it} X_{ij} \right) - a_{ij} \mathbb{E} \left(X_{ij}^2 \right) (**) \\ &= 0 \end{aligned}$$

Calculons d'abord les quantités individuelles :

$$\begin{aligned} \text{Var}(X_{it}) &= \mathbb{E} \left(\text{Var}(X_{it} | \Theta) \right) + \text{Var} \left(\mathbb{E}(X_{it} | \Theta) \right) \\ &= \mathbb{E} \left(s^2(\Theta_i) \right) + \mathbb{E} \left(\mu(\Theta_i)^2 \right) - \mathbb{E} \left(\mu(\Theta_i) \right)^2 \\ &= \sigma^2 + \text{Var} \left(\mu(\Theta_i) \right) \\ &= \sigma^2 + \tau^2 \end{aligned}$$

$$\begin{aligned} \mathbb{E}(X_{it}^2) &= \text{Var}(X_{it}) + \mathbb{E}(X_{it})^2 \\ &= \text{Var} \left(\mathbb{E}(X_{it} | \Theta_i) \right) + \mathbb{E} \left(\text{Var}(X_{it} | \Theta_i) \right) + \mathbb{E} \left(\mathbb{E}(X_{it} | \Theta_i) \right)^2 \\ &= \sigma^2 + \tau^2 + \mu^2 \end{aligned}$$

De plus, $\forall t \neq j$ nous avons :

$$\begin{aligned} \mathbb{E}(X_{ij} X_{it}) &= \mathbb{E} \left(\mathbb{E}(X_{ij} X_{it} | \Theta_i) \right) \\ &= \mathbb{E} \left(\mu(\Theta_i)^2 \right) \\ &= \text{Var} \left(\mu(\Theta_i) \right) + \mathbb{E} \left(\mu(\Theta_i) \right)^2 \\ &= \tau^2 + \mu^2 \end{aligned}$$

$$\begin{aligned} \text{Enfin} \quad \mathbb{E} \left(X_{ij} \mu(\Theta_i) \right) &= \mathbb{E} \left(\mathbb{E}(X_{ij} \mu(\Theta_i) | \Theta_i) \right) \\ &= \mathbb{E} \left(\mu(\Theta_i)^2 \right) \\ &= \tau^2 + \mu^2 \end{aligned}$$

Alors (**) se réécrit comme suit :

$$\mathbb{E} \left(X_{ij} \mu(\Theta_i) \right) - a_{i0} \mu - \sum_{t=1, t \neq j}^T a_{it} \mathbb{E} \left(X_{it} X_{ij} \right) - a_{ij} \mathbb{E} \left(X_{ij}^2 \right) = (\tau^2 + \mu^2) \left(1 - \sum_{t=1}^T a_{it} \right) - a_{ij} \sigma^2 - a_{i0} \mu$$

En utilisant la formule (*), nous obtenons l'équation suivante :

$$a_{ij}\sigma^2 = \tau^2 \left(1 - \sum_{t=1}^T a_{it} \right)$$

Nous remarquons que la formule trouvée ne dépend pas de j .

Donc, $\forall i \in [1, n]$

$$a_{i1} = a_{i2} = \dots = a_{iT} \quad \text{et} \quad a_{i1} = \frac{\tau^2}{\sigma^2 + T\tau^2}$$

De plus,

$$a_{i0} = \left(1 - \frac{T\tau^2}{\sigma^2 + T\tau^2} \right) \mu$$

Nous trouvons ainsi l'estimateur linéaire de crédibilité

$$\begin{aligned} \hat{\mu}(\Theta_i) &= a_{i0} + \sum_{t=1}^T a_{it} X_{it} \\ &= \left(1 - \frac{T\tau^2}{\sigma^2 + T\tau^2} \right) \mu + \frac{T\tau^2}{\sigma^2 + T\tau^2} \bar{X}_i \end{aligned}$$

Démonstrations du modèle de Hachemeister

Rappelons d'abord le modèle de Hachemeister et ses hypothèses. Ensuite nous exposons les différentes démonstrations à savoir dans ce modèle.

Notations et hypothèses

Le modèle de Hachemeister suppose les hypothèses suivantes :

- Conditionnellement à Θ_i , les $(X_{it}, t \in [1; T])$ sont indépendants et nous avons

$$\mathbb{E}(X_i | \Theta_i) = Y_i \beta(\Theta_i) = \mu(\Theta_i)$$

avec $\beta(\Theta_i)$ le vecteur de régression $\in M_{T,1}(T \leq n)$, de composantes linéairement indépendantes
 Y_i est la matrice design.

$$\text{Cov}(X_i, X_i^t | \Theta_i) = \Sigma_i(\Theta_i)$$

- Les couples $(\Theta_1, X_1), \dots, (\Theta_n, X_n)$ sont indépendants et $\Theta_1, \dots, \Theta_n$ sont *i.i.d*

L'estimateur des moindres carrés est solution du problème de minimisation suivant :

$$\min_{\beta_1, \dots, \beta_T} \sum_{i,j=1}^n \left((X_i - (Y\beta)_i) \Sigma_{ij}^{-1} (X_j - (Y\beta)_j) \right)$$

Conditionnellement à Θ_i , le meilleur estimateur linéaire des moindres carrés de $\beta(\Theta_i)$ est

$$\beta(\hat{\Theta}_i) = \left(Y_i' \Sigma_i(\Theta_i)^{-1} Y_i \right)^{-1} Y_i' \Sigma_i(\Theta_i)^{-1} X_i$$

de matrice de covariance

$$\Sigma_{\hat{\beta}(\Theta_i)} = (Y_i' \Sigma_i(\Theta_i)^{-1} Y_i)^{-1}$$

Nous avons Θ_i inconnu alors, $\Sigma_i(\Theta_i)$ est inconnu. Posons ainsi $S_i = \mathbb{E}(\Sigma_i(\Theta_i))$. En remplaçant la matrice de covariance par S_i nous obtenons l'analogie avec le modèle de Bühlmann-Straub où le paramètre structure est $\sigma^2 = \mathbb{E}(\sigma^2(\Theta_i))$

Théorème de compression des données :

Conditionnellement à Θ_i , le meilleur estimateur linéaire sans biais de \hat{B}_i à partir des X_i est

$$\hat{B}_i = \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} X_i$$

En effet, cette expression ne dépend pas de Θ_i alors c'est bien la compression optimale des données. De plus, nous avons

$$\begin{aligned} \mathbb{E}(\hat{B}_i | \Theta_i) &= \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} \mathbb{E}(X_i | \Theta_i) \\ &= \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} Y_i \beta(\Theta_i) \\ &= \beta(\Theta_i) \quad \text{alors l'estimateur est sans biais conditionnellement à } \Theta_i \end{aligned}$$

Sa matrice d'erreur quadratique moyenne est

$$\mathbb{E} \left(\left(\hat{B}_i - \beta(\Theta_i) \right) \left(\hat{B}_i - \beta(\Theta_i) \right)^t \right) = \left(Y_i' S_i^{-1} Y_i \right)^{-1}$$

Formule de Hachemeister

La formule de Hachemeister de l'estimateur de crédibilité de $\beta(\Theta_i)$ est donnée par :

$$\hat{\beta}(\Theta_i) = Z_i B_i + (I - Z_i) \beta$$

Avec :

$$B_i = \left(Y_i' S_i^{-1} Y_i \right)^{-1} Y_i' S_i^{-1} X_i$$

$$S_i = \mathbb{E}(\Sigma_i(\Theta_i))$$

$$Z_i = T \left(T + \left(Y_i' S_i^{-1} Y_i \right)^{-1} \right)^{-1}$$

$$T = \text{Cov}(\beta(\Theta_i), \beta(\Theta_i)')$$

$$\beta = \mathbb{E}(\beta(\Theta_i))$$

Son erreur quadratique moyenne est $\mathbb{E} \left(\left(\hat{\beta}(\Theta_i) - \beta(\Theta_i) \right) \left(\hat{\beta}(\Theta_i) - \beta(\Theta_i) \right)^t \right) = (I - Z_i) T$

La régression linéaire avec modèle de tendance linéaire

La formule de Hachemeister dans le cas simple de la régression linéaire présente un problème puisque la courbe issue du modèle ne se trouve pas forcément entre la courbe individuelle et celle collective comme attendue. En fait, lors de la conception du modèle de crédibilité de la régression Hachemeister trouve que pour un contrat la pente de la courbe de crédibilité est inférieure à celle de la courbe collective et individuelle. Le problème a été résolu par Bühlmann & Gisler (1997) en modélisant l'ordonnée à l'origine au centre de la plage de temps au lieu de l'origine du temps.

Nous étudions, dans ce qui suit, le modèle avec le barycentre du temps au lieu de l'origine du temps. Nous présentons avant le modèle à l'origine du temps afin de voir la source d'inconvénients de la méthode et sa solution mathématique.

Modèle avec l'ordonnée à l'origine du temps

Le modèle de régression linéaire avec le temps comme covariable est donnée par l'équation suivante :

$$\mathbb{E}(X_{it} | \Theta_i) = \underbrace{\beta_0(\Theta_i)}_{\substack{\text{ordonnée à l'origine} \\ (\text{intercept})}} + t \underbrace{\beta_1(\Theta_i)}_{\substack{\text{pente} \\ (\text{slope})}}$$

Cette expression peut être exprimée en notation matricielle comme suit :

$$\mathbb{E}(X_i|\Theta_i) = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \end{pmatrix}}_{Y_i} \underbrace{\begin{pmatrix} \beta_0(\Theta_i) \\ \beta_1(\Theta_i) \end{pmatrix}}_{\beta(\Theta_i)}$$

Nous supposons que nous sommes dans le modèle standard de régression. Nous avons ainsi

$$S_i = \mathbb{E}\left(\text{Cov}(X_i, X_i'|\Theta_i)\right) = \sigma^2 W_i^{-1}$$

avec
$$W_i = \begin{pmatrix} w_{i1} & 0 & \dots & 0 \\ 0 & w_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{iT} \end{pmatrix}$$

Pour des raisons de simplification de notation, nous omettons l'indice i et nous introduisons les quantités suivantes. En considérant la distribution $\left(\frac{w_t}{w.}\right)_{t \in \llbracket 1 ; T \rrbracket}$ nous posons :

- $\mathbb{E}_w(t) = \sum_{t=1}^T t \frac{w_t}{w.}$
- $\text{Var}_w(t) = \mathbb{E}_w(t^2) - \left(\mathbb{E}_w(t)\right)^2 = \sum_{t=1}^T t^2 \frac{w_t}{w.} - \left(\sum_{t=1}^T t \frac{w_t}{w.}\right)^2$
- $\mathbb{E}_w(X_t) = \sum_{t=1}^T X_t \frac{w_t}{w.}$

Nous posons aussi la quantité $V = Y' W Y$

$$\begin{aligned} V &= \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 2 & 3 & \dots & T \end{pmatrix} \begin{pmatrix} w_1 & 0 & \dots & \dots & 0 \\ 0 & w_2 & \dots & \dots & 0 \\ 0 & 0 & w_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & w_T \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & T \end{pmatrix} \\ &= w. \begin{pmatrix} \sum_{t=1}^T \frac{w_t}{w.} & \sum_{t=1}^T t \frac{w_t}{w.} \\ \sum_{t=1}^T t \frac{w_t}{w.} & \sum_{t=1}^T t^2 \frac{w_t}{w.} \end{pmatrix} \\ &= w. \begin{pmatrix} 1 & \mathbb{E}_w(t) \\ \mathbb{E}_w(t) & \mathbb{E}_w(t^2) \end{pmatrix} \end{aligned}$$

Pour calculer l'estimateur de $\beta(\Theta)$ par la formule de Hachemeister nous calculons les quantités Z et B. Nous commençons par B :

$$\begin{aligned}
B &= V^{-1}Y'WX \\
&= \frac{1}{w} \begin{pmatrix} 1 & \mathbb{E}_w(t) \\ \mathbb{E}_w(t) & \mathbb{E}_w(t^2) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=1}^T w_t X_t \\ \sum_{t=1}^T t w_t X_t \end{pmatrix} \\
&= \frac{1}{Var_w(t)} \begin{pmatrix} \mathbb{E}_w(t^2) & -\mathbb{E}_w(t) \\ -\mathbb{E}_w(t) & 1 \end{pmatrix} \begin{pmatrix} \mathbb{E}_w(X_t) \\ \mathbb{E}_w(tX_t) \end{pmatrix} \\
&= \frac{1}{Var_w(t)} \begin{pmatrix} \mathbb{E}_w(t^2) \mathbb{E}_w(X_t) - \mathbb{E}_w(t) \mathbb{E}_w(tX_t) \\ \mathbb{E}_w(tX_t) - \mathbb{E}_w(t) \mathbb{E}_w(X_t) \end{pmatrix}
\end{aligned}$$

Pour calculer la matrice de crédibilité Z nous prenons l'hypothèse intuitive d'indépendance entre l'ordonnée à l'origine et la pente. De ce fait, nous avons :

$$T = Cov(\beta(\Theta_i), \beta(\Theta_i)') = \begin{pmatrix} \tau_0^2 & 0 \\ 0 & \tau_1^2 \end{pmatrix}$$

Nous avons alors :

$$\begin{aligned}
Z &= T(T + \sigma^2 V^{-1})^{-1} \\
&= (I + \sigma^2 V^{-1} T^{-1})^{-1}
\end{aligned}$$

Nous calculons ainsi l'inverse de T :

$$\begin{aligned}
T^{-1} &= \begin{pmatrix} 1/\tau_0^2 & 0 \\ 0 & 1/\tau_1^2 \end{pmatrix} \\
&= \frac{1}{\sigma^2} \begin{pmatrix} \kappa_0 & 0 \\ 0 & \kappa_1 \end{pmatrix} \text{ avec } \kappa_0 = \frac{\sigma^2}{\tau_0^2} \text{ et } \kappa_1 = \frac{\sigma^2}{\tau_1^2}
\end{aligned}$$

Nous trouvons donc :

$$I + \sigma^2 V^{-1} T^{-1} = \frac{1}{w \cdot Var_w(t)} \begin{pmatrix} \kappa_0 \mathbb{E}_w(t^2) & -\kappa_1 \mathbb{E}_w(t) \\ -\kappa_0 \mathbb{E}_w(t) & \kappa_1 \end{pmatrix}$$

Nous posons $c = w \cdot Var_w(t)$:

$$\begin{aligned}
Z &= \frac{w \cdot Var_w(t)}{(\kappa_1 + c)(\kappa_0 \mathbb{E}_w(t^2) + c) - \kappa_1 \kappa_0 \mathbb{E}_w(t)^2} \begin{pmatrix} \kappa_1 + w \cdot Var_w(t) & \kappa_1 \mathbb{E}_w(t) \\ \kappa_0 \mathbb{E}_w(t) & \kappa_0 \mathbb{E}_w(t^2) + w \cdot Var_w(t) \end{pmatrix} \\
&= \frac{w}{D} \begin{pmatrix} \kappa_1 + w \cdot Var_w(t) & \kappa_1 \mathbb{E}_w(t) \\ \kappa_0 \mathbb{E}_w(t) & \kappa_0 \mathbb{E}_w(t^2) + w \cdot Var_w(t) \end{pmatrix}
\end{aligned}$$

avec $D = \kappa_1 \kappa_0 + \kappa_1 w + w \cdot \kappa_0 \mathbb{E}_w(t^2) + w^2 Var_w(t)$

Modèle au barycentre du temps

La matrice de crédibilité Z n'est pas diagonale ce qui nous mène au problème du modèle présentée par Hachemeister. Pour remédier à ce problème l'ordonnée à l'origine est pris non pas à l'origine du temps mais au "barycentre du temps" du rang de temps considéré. Ce centre est défini comme suit :

$$t_0 = \mathbb{E}_w(t) = \sum_{k=1}^T k \frac{w_k}{w}.$$

La matrice de design est donc

$$Y = \begin{pmatrix} 1 & 1 - t_0 \\ 1 & 2 - t_0 \\ \vdots & \vdots \\ 1 & T - t_0 \end{pmatrix}$$

Une transformation linéaire de l'axe du temps n'a pas d'effet sur l'estimateur de crédibilité. La transformation appliquée change $\beta_0(\Theta)$ qui ne représente plus l'ordonnée à l'origine du temps mais l'ordonnée au "centre de gravité" du temps. Nous modélisons ainsi de nouvelles quantités et nous supposons aussi que la matrice de covariance T reste diagonale. Nous reprenons les calculs faits en prenant le fait que les covariances sont invariantes par translation en plus $\mathbb{E}_w(t') = \mathbb{E}_w(t - E_w(t)) = 0$ et nous trouvons ainsi les résultats suivants :

$$\begin{aligned} B &= \frac{1}{Var_w(t)} \begin{pmatrix} \mathbb{E}_w((t - E_w(t))^2) & \mathbb{E}_w(X_t) \\ \mathbb{E}_w((t - E_w(t))X_t) & \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{E}_w(X_t) \\ \frac{Cov_w(t, X_t)}{Var_w(t)} \end{pmatrix} \end{aligned}$$

La matrice de crédibilité Z est désormais diagonale car $\mathbb{E}_w(t') = 0$.

Enfin :

$$\begin{aligned} D &= \kappa_1 \kappa_0 + \kappa_1 w + w \kappa_0 \mathbb{E}_w(t^2) + w^2 Var_w(t) \\ &= (\kappa_0 + w)(Var_w(t) + \kappa_1) \end{aligned}$$

$$Z = \begin{pmatrix} \frac{w}{w + \frac{\sigma^2}{\tau_0^2}} & 0 \\ 0 & \frac{w Var_w(t)}{w Var_w(t) + \frac{\sigma^2}{\tau_1^2}} \end{pmatrix}$$

Nous retrouvons ainsi deux formules de crédibilité semblables à celle du modèle de Bühlmann-Straub.

$$\begin{aligned} \hat{\beta}_0(\Theta) &= Z_{11} B_0 + (1 - Z_{11}) \beta_0 \\ \hat{\beta}_1(\Theta) &= Z_{22} B_1 + (1 - Z_{22}) \beta_1 \end{aligned}$$

telles que β_0 et β_1 les composantes de β et

$$\begin{aligned} B_0 &= \mathbb{E}_w(X_t) & Z_{11} &= w_{it} Var(X_{it} | \Theta_i) \\ B_1 &= \frac{w}{w + \frac{\sigma^2}{\tau_0^2}} & Z_{22} &= \frac{w Var_w(t)}{w Var_w(t) + \frac{\sigma^2}{\tau_1^2}} \end{aligned}$$

Nous rappelons que nous avons utilisé le centre de gravité collectif $t_0 = \sum_{k=1}^T k \frac{w_{.k}}{w_{..}}$ au lieu du centre de gravité individuel $t_{i0} = \sum_{k=1}^T k \frac{w_{ik}}{w_i}$. La différence dans l'utilisation de l'un ou de l'autre est négligeable.

Table des figures

1	Courbe des taux bruts de rachat estimés en fonction des anciennetés	6
2	Comparaison entre les différentes courbes extrapolées pour le produit D	8
3	Rough lapse rates curve estimated as a function of seniority	12
4	Comparison of the different extrapolated curves for the product D	14
1.1	Distribution mondiale des primes collectées par pays en 2016	26
1.2	Évolution des provisions techniques en Italie par branche	29
1.3	Évolution du flux de trésorerie en Italie par branche	29
2.1	Courbe de survie du produit A estimé par l'estimateur Kaplan Meier et Harrington et Fleming	39
2.2	Courbe de taux de rachat du produit A estimé par estimateur Hoem et Nelson Aalen .	39
2.3	Courbe de survie du produit B estimé par l'estimateur Kaplan Meier et Harrington et Fleming	40
2.4	Courbe de taux de rachat du produit B estimé par estimateur Hoem et Nelson Aalen .	41
2.5	Courbe de survie du produit C estimé par l'estimateur Kaplan Meier et Harrington et Fleming	41
2.6	Courbe de taux de rachat du produit C estimé par estimateur Hoem et Nelson Aalen .	42
2.7	Courbe de survie du produit D estimé par l'estimateur Kaplan Meier et Harrington et Fleming	42
2.8	Courbe de taux de rachat du produit D estimé par estimateur Hoem et Nelson Aalen .	43
2.9	Courbe de rachat du produit A lissée par la méthode de la moyenne mobile	44
2.10	Courbe de rachat du produit B lissée par la méthode de la moyenne mobile	44
2.11	Courbe lissée par la méthode de Whittaker-Henderson du produit A	47
2.12	Courbe lissée par la méthode de Whittaker-Henderson du produit B	47
2.13	Extrapolation par moyenne mobile de la courbe de taux brut de rachat et de la courbe de rachat lissée par W-H du produit A	49
2.14	Extrapolation par moyenne mobile de la courbe de taux brut de rachat et de la courbe de rachat lissée par W-H du produit B	49
2.15	Taux de rachat de trois produits pour huit années d'ancienneté	56
2.16	Comparaison entre les différentes courbes extrapolées pour le produit B	60
2.17	Comparaison entre les différentes courbes extrapolées pour le produit C	61
2.18	Comparaison entre les différentes courbes extrapolées pour le produit D	61

3.1	Illustration des deux cas extrêmes de la courbe ROC	65
3.2	Matrice de corrélation des variables explicatives numériques	67
3.3	Distribution des âges à la souscription de toutes les polices	68
3.4	Distribution des âges au rachat des polices rachetées	69
3.5	Répartition des assurés par sexe	69
3.6	Répartition des polices par réseau de distribution	70
3.7	Distribution des polices rachetées par année de souscription	70
3.8	Répartition des polices rachetées par ancienneté	71
3.9	Courbe du taux de rachat en fonction de l'âge au rachat	71
3.10	Importance des variables explicatives dans le modèle <i>Elastic Net</i>	84
3.11	Importance des variables explicatives dans le modèle <i>Random Forest</i>	85
3.12	Importance des variables explicatives dans le modèle <i>XGBoost</i>	86