

**Mémoire présenté le :**

**pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA  
et l'admission à l'Institut des Actuaires**

Par : SOKHNA MAME BOUSSO MBACKE SAKHO

Titre Analyse et Modélisation des comportements de rachat d'un  
portefeuille d'épargne

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

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

*Membre présents du jury de l'Institut  
des Actuaires*

signature

*Entreprise :*

Nom :

*Signature :*

*Directeur de mémoire en entreprise :*

Nom :

*Signature :*

*Invité :*

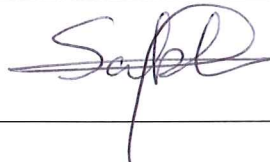
Nom :

*Signature :*

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

Signature du responsable entreprise

Signature du candidat





# Analyse et Modélisation des comportements de rachat d'un portefeuille d'épargne

Sokhna M. B. Mb. SAKHO

Institut de Science Financière et d'Assurances (ISFA)

Mémoire d'Actuariat

---

# Résumé

---

**Mots clés** : comportements de rachat, variables explicatives, rachats conjoncturels, mélange de GLM

Au cours de sa vie, un contrat d'épargne peut faire face à de multiples évènements pouvant modifier ses flux financiers futurs probables. C'est le cas notamment lors d'une modification du niveau de la prime ou de la structure de la prime ou lorsqu'il arrive à son échéance ou lors d'un rachat. Ce dernier évènement représente un risque majeur pour l'assureur au point qu'un module entier est dédié à son évaluation dans le cadre de la directive européenne Solvabilité II d'où la nécessité de sa bonne maîtrise. C'est dans ce contexte que s'inscrit ce mémoire. L'objectif est d'analyser puis de modéliser les comportements de rachat des assurés d'un portefeuille d'épargne. L'analyse et l'étude du portefeuille permettent d'abord de détecter les facteurs de risque susceptibles d'impacter la décision de rachat. Une méthodologie de sélection de variables est ensuite appliquée afin de choisir les variables pertinentes qui seront explicatives du comportement de rachat. Ceci a pour finalité l'implémentation de modèles de régressions (GLM) pour mesurer les effets de chacune de ses variables sur la variable d'intérêt (décision de racheter son contrat ou pas). Pour capter les rachats conjoncturels qui sont difficilement modélisables, un modèle mélange de GLM est introduit. Même si les résultats ne sont pas très satisfaisants liés au nombre et au choix de nos variables ainsi qu'à la dimension de notre portefeuille, les modèles implémentés ont permis néanmoins de dégager certaines caractéristiques et tendances qui justifieraient et expliqueraient les comportements de rachat du portefeuille.

---

# Abstract

---

**Key words** : surrender behaviours, predictor variables, mixture of logit, cyclical surrenders

During its lifetime, a saving insurance contract could face many events that should change its likely cash flow projections. This is true in particular when there is a change in the level of risk premiums or when the contract reaches the maturity or the policyholder decides to take all or part of the surrender value. The latter represents a great risk for the insurer so that a whole module is dedicated to the evaluation of the lapse risk under the European Solvency II directive. It is necessary to control this risk. That is why in this dissertation, our aim is to analyze and model the surrender behaviours of the policyholders of our saving portfolio. First, we analyze and study the portfolio so as to identify the main risk factors that could impact policyholders' decisions. Thus, a methodology of selecting variables is applied to choose the most relevant of them to explain the surrender behaviours. In fact, the chosen variables are taken into account during the implementation of the regression models in order to measure how much they affect the variable of interest (Decision of lapsing the contract). Furthermore, to understand the cyclical surrenders, we introduce a mixture of logit regressions. Even if our results have not been totally successful due to the failures regarding the portfolio and the predictor variables, the implemented models reveal us some characteristics and tendencies that justify and explain the surrender behaviours of the portfolio.

---

# Remerciements

---

Je remercie en premier toute l'équipe d'Optimum Vie de m'avoir accueillie au sein de la compagnie dans le cadre de mon apprentissage et de m'avoir proposé ce sujet de mémoire très intéressant. J'aimerais remercier plus particulièrement ma tutrice Delphine Alves et le Directeur Technique de l'Actuariat, Simon Bélanger.

Je voudrais remercier mon professeur Xavier Milhaud qui a beaucoup inspiré ce mémoire de par ses cours, ses articles et ses recherches. Je le remercie d'avoir été disponible pour répondre à toutes mes questions et pour ses précieux conseils.

Je remercie aussi l'ensemble du corps professoral de l'ISFA pour la qualité de la formation et pour toutes les connaissances que j'ai acquises tout au long de mon cursus.

Je ne saurais terminer sans remercier ma famille : ma mère, qui n'a ménagé aucun effort pour la réussite dans mes études, mon frère (Tapha), parrain, mentor, qui m'a toujours aidée, soutenue et conseillée depuis mes débuts en France, mon frère jumeau, pour ses encouragements et d'avoir toujours été là, et enfin mon amie(NFD) pour m'avoir soutenue et pour la correction du mémoire.

---

# Limites et Axes d'amélioration du mémoire

---

Dès l'entame de ce mémoire, nous tenons à signaler que le travail fourni ci-dessous n'est pas exhaustif d'où les limites observées. Certaines variables auraient pu être prises en compte dans la modélisation : le taux d'investissement Euros/UC du portefeuille et les réseaux de distribution de la compagnie d'assurance-vie. En effet, Optimum Vie travaillant avec un vaste réseau de courtiers, une bonne utilisation de cette information aurait pu nous édifier sur la relation entre l'assuré et le distributeur ainsi que l'impact du courtier sur les rachats et nous permettre de créer une variable pertinente à introduire dans les modèles.

Par ailleurs, le frein majeur qui constitue la limite principale des résultats des modèles, est la non-prise en compte de la temporalité tout au long de l'étude. Ladite variable permet de bien prendre en compte l'exposition des assurés à des contextes économiques différents. X. Milhaud dans [7] propose une façon pertinente de la considérer lors de la construction des échantillons d'apprentissage et de test. C'est une technique de validation temporelle qui consiste à construire les échantillons apprentissage et test, non plus aléatoirement comme nous avons pu le faire, mais en considérant la période étudiée dans le portefeuille.

L'étude et l'analyse du portefeuille réalisées dans la deuxième partie du mémoire, portent uniquement sur les bases de rachat. Ce choix d'étudier les comportements de rachat en se basant uniquement sur les rachats du portefeuille, est à bien égards discutable. Le portefeuille dans sa globalité aurait dû être pris en compte pour mieux interpréter les comportements observés et avoir une vue d'ensemble en termes de proportion sur les variables étudiées.

Pour finir, comme énoncé dans la conclusion, au-delà des facteurs de risque classiques, il convient maintenant d'introduire dans la modélisation des comportements de rachat des variables comme la mesure de la satisfaction client et les nouvelles réglementations. Mesurer la satisfaction client passe par l'analyse de son niveau de satisfaction par rapport au produit d'assurance et la bonne compréhension du fonctionnement de ce dernier et le suivi de l'évolution de sa relation avec le distributeur.

---

# Introduction

---

Placements favoris des français liés à leur fiscalité avantageuse, certains produits d'assurance-vie, les produits d'épargne notamment, offrent en plus aux assurés la possibilité de racheter leur contrat. Si pour les assurés, cette option de rachat représente un moyen fiable d'avoir la mainmise sur leur contrat et de prendre des décisions au gré de leurs besoins en liquidité ou de leur recherche de plus de flexibilité et de souplesse, du point de vue de l'assureur, la réalité est toute autre.

En effet, l'option de rachat constitue pour ce dernier un risque au sens où les comportements des assurés ne sont pas prévisibles : l'assuré peut racheter partiellement ou totalement à tout moment son contrat sous réserve de disponibilité de son épargne. En cas de rachats massifs, l'assureur pourra faire face à un fort risque de faillite de la compagnie. De plus, le risque de rachat modifiant les flux financiers futurs probables du contrat, se soulèvent donc des problématiques de gestion/actif pour la compagnie.

Par ailleurs, décrit souvent comme étant le deuxième risque des compagnies d'assurance, le risque de rachat est au cœur des nouvelles réglementations : son coût doit être désormais inclus dans les passifs d'assurance lors de leur évaluation qui doit être la plus juste et sincère possible en accord avec les normes comptables IFRS. Il a été aussi intégré au dispositif prudentiel Solvabilité II. Il devient alors impératif pour les assureurs de maîtriser ce risque, ce qui passe par une bonne modélisation des comportements de rachat des assurés.

C'est dans ce contexte que s'inscrit ce mémoire. La motivation première est que le sujet n'a jamais été abordé au sein d'Optimum Vie. En modélisant les comportements de rachat du portefeuille d'épargne, nous cherchons à expliquer, à justifier ou à comprendre les décisions des assurés à racheter leur contrat afin de pouvoir déployer différentes stratégies pour pouvoir bien gérer ce risque dans le cadre du pilotage de l'activité d'assurance.

Dans la première partie du mémoire, nous nous attardons d'abord sur les généralités régissant l'assurance-vie puis nous décrivons la compagnie OPTIMUM Vie et son portefeuille avant d'introduire notre problématique pour finir par mettre l'accent sur la fiscalité surtout celle des rachats.

Ensuite dans la deuxième partie, nous analysons et étudions notre portefeuille afin d'avoir des intuitions sur les facteurs de risque susceptibles d'influencer les comportements de rachat.

Enfin la dernière partie est consacrée à la modélisation qui passe par la présentation et la mise en pratique sur notre portefeuille d'épargne de deux modèles : une régression logistique et une approche par mélange de régressions logistiques.

---

# Sommaire

---

<b>Résumé</b>	<b>3</b>
<b>Abstract</b>	<b>4</b>
<b>Remerciements</b>	<b>5</b>
<b>Limites et Axes d'amélioration du mémoire</b>	<b>6</b>
<b>Introduction</b>	<b>7</b>
<b>1 Contexte du mémoire</b>	<b>10</b>
1.1 Cadre de l'assurance-vie . . . . .	10
1.1.1 Généralités . . . . .	10
1.1.2 Les différents types de contrats d'assurance-vie . . . . .	11
1.1.3 Caractéristiques et clauses des contrats d'assurance-vie . . . . .	12
1.2 Présentation d'Optimum Vie et de son portefeuille . . . . .	18
1.2.1 Présentation de la société . . . . .	18
1.2.2 Le portefeuille d'Optimum Vie . . . . .	18
1.3 Problématique du mémoire : les comportements de rachat en assurance-vie .	21
1.3.1 Le risque de rachat . . . . .	21
1.3.2 Intérêts de la connaissance de ce risque . . . . .	21
1.3.3 Approches de modélisation . . . . .	22
1.3.4 Le cadre règlementaire . . . . .	22
1.4 Paysage fiscal des contrats d'assurance-vie . . . . .	25
1.4.1 Généralités . . . . .	25
1.4.2 Fiscalité des rachats . . . . .	29
<b>2 Etude du portefeuille</b>	<b>44</b>
2.1 Présentation des données . . . . .	44
2.1.1 Description des données . . . . .	44
2.1.2 Retraitement des bases . . . . .	45
2.2 Statistiques Descriptives . . . . .	47
2.2.1 Les rachats entre 2005 et le 1er semestre 2017 . . . . .	47
2.2.2 Statistiques descriptives sur les variables retenues . . . . .	49

<b>3</b>	<b>Modélisation des comportements de rachat</b>	<b>64</b>
3.1	Le choix du modèle . . . . .	64
3.1.1	Les données . . . . .	64
3.1.2	La régression logistique (LR) . . . . .	66
3.2	Application . . . . .	77
3.2.1	Sélection des variables explicatives . . . . .	77
3.2.2	Mise en pratique de la régression logistique . . . . .	77
3.3	Modélisation par mélange de lois . . . . .	84
3.3.1	Généralités sur les modèles mélanges . . . . .	85
3.3.2	Quel modèle mélange pour notre problématique? . . . . .	90
	<b>Conclusion</b>	<b>101</b>
	<b>Bibliographie</b>	<b>103</b>
	<b>Annexe</b>	<b>105</b>

---

## CONTEXTE DU MÉMOIRE

---

### 1.1 Cadre de l'assurance-vie

#### 1.1.1 Généralités

Selon la définition<sup>1</sup> **IFRS** (International Financial Reporting Standards), un contrat d'assurance-vie est :

«Un contrat selon lequel une partie (l'assureur) accepte un risque d'assurance significatif d'une autre partie (le titulaire de la police) en convenant d'indemniser le titulaire de la police si un événement futur incertain spécifié (l'événement assuré) affecte de façon défavorable le titulaire de la police».

En se basant sur cette définition, certains contrats peuvent être requalifiés en instruments financiers (par exemple les contrats d'investissement avec un risque d'assurance non significatif). Le contrat d'épargne serait plutôt considéré comme un placement financier, plus proche d'un contrat de capitalisation. Toutefois, il est somme toute un contrat d'assurance vie du fait de son caractère aléatoire (dépendant de la durée de vie humaine). En effet, pour un contrat d'épargne, tous les versements ainsi que les intérêts générés vont constituer un capital. En cas de vie, on touchera ce capital épargné ; en cas de décès, ce capital sera transmis à la personne qui avait été préalablement désignée dans le contrat. C'est le principe de la clause bénéficiaire. Il faut noter aussi que le capital en cas de vie peut-être donné à l'assuré sous forme de rente. C'est le cas notamment lorsque le contrat d'épargne répond à un besoin de complément de retraite.

L'assurance-vie recouvre le périmètre décrit au paragraphe 3 de l'article 2 de la directive Solvabilité II :

- l'assurance en cas de vie, l'assurance en cas de décès, l'assurance mixte et l'assurance sur la vie avec contre-assurance (qui combinent des garanties en cas de décès et en cas de vie) ainsi que les assurances « nuptialité » et « natalité »,
- l'assurance de rente,
- les assurances complémentaires souscrites en complément d'une assurance vie (il s'agit généralement de garanties incapacité et invalidité),
- de contrats spécifiques comme les tontines ou les opérations de capitalisation.

C'est une spécificité française par rapport aux autres pays au sens où elle est le premier produit de placement des ménages français. Ce qui peut s'expliquer par le fait qu'elle s'est

---

1. Source : Cours Pierre Thérond, Normes IFRS, ISFA 2017

développée dans un contexte institutionnel et fiscal très favorable incitant les ménages à détenir le produit sur le long-terme. Par ailleurs, elle vient aussi pallier l'absence de fonds de pension en France.

## **1.1.2 Les différents types de contrats d'assurance-vie<sup>2</sup>**

### **1.1.2.1 les contrats en euros dits mono supports**

Ce sont des contrats destinés à ceux qui veulent investir sans prendre de risque. Leur principal atout est la sûreté du placement. Ils assurent pour l'épargnant une protection contre toute perte en capital. L'assureur s'engage à protéger le capital du souscripteur et à garantir un taux de rendement minimum. Pour ce faire, les fonds garantis en euros sont investis dans des titres sans risque, essentiellement des obligations. De plus, ces contrats bénéficient d'une double garantie :

- le capital investi est garanti : à l'échéance, il sera augmenté des intérêts capitalisés et d'une éventuelle participation aux bénéfices
- un effet cliquet, qui garantit à l'épargnant les gains de l'année écoulée. Les intérêts générés par son contrat lui sont définitivement acquis : il n'y a pas de risque de perte en capital.

### **1.1.2.2 les contrats en unités de compte**

Dans un contexte de baisse des taux entraînant la baisse du rendement des contrats d'assurance-vie en euros, de plus en plus de personnes se tournent vers les contrats en unités de compte. Ces derniers offrent des taux de rendement plus rémunérateurs, mais sont plus risqués. Le capital n'est pas garanti, contrairement aux contrats en euros. Une unité de compte est un support d'investissement financier tel que des parts ou des actions de valeurs mobilières ou immobilières (Société d'investissement à capital variable – SICAV, actions, obligations, parts de fonds communs de placement, parts de Société Civile Immobilière - SCI, parts de Société Civile de Placement Immobilier - SCPI). Les contrats en unités de compte permettent un investissement diversifié dans les marchés financiers et immobiliers. Ils sont destinés aux épargnants qui recherchent un investissement à long terme et qui sont prêts à accepter les risques inhérents aux fluctuations des marchés financiers (le capital n'est pas garanti, seul le nombre d'unités de compte l'est). Les investissements en unités de compte peuvent être très variés et personnalisés.

---

2. Source : Fédération Française de l'Assurance

### **1.1.2.3 les contrats dits multi supports**

Ces contrats combinent un fonds en euros pour sécuriser l'épargne placée et des fonds en unités de compte pour rechercher des performances. L'allocation (répartition) de l'épargne est fonction de la stratégie patrimoniale et du niveau de risque accepté par le souscripteur. Des arbitrages sont possibles à tout moment en fonction des résultats (opérations de transfert d'une fraction de l'épargne du support en euros vers le support en unités de compte et inversement).

### **1.1.2.4 Deux nouveaux types de contrats d'assurance-vie**

Ils ont vu le jour dans la loi de finance rectificative 2013, applicable au 1er Janvier 2014.

#### **1.1.2.4.1 Les fonds Euro-croissance**

Les contrats d'assurance-vie Euro-croissance sont des contrats intermédiaires entre ceux en euros et ceux en unités de compte. En effet, ils offrent aux épargnants l'espérance d'un rendement supérieur aux contrats en euros, avec une prise de risque plus faible que pour les contrats en unités de compte. Le contrat Euro-croissance peut comporter à la fois des supports en euros et/ou unités de compte et un fonds Euro-croissance ou croissance. Le capital investi dans ce fonds est garanti à **100%** au terme d'une durée d'au moins 8 ans. Les contrats Euro-croissance bénéficient de la fiscalité de l'assurance-vie classique.

#### **1.1.2.4.2 Les contrats Vie Génération**

Ils sont intégralement investis dans des supports en unités de compte à condition qu'au moins un tiers des versements soit placé dans des investissements «utiles à l'économie» comme le financement des PME-ETI, le logement social ou l'économie sociale et solidaire. En contrepartie de ces engagements, le ou les bénéficiaires, au décès du souscripteur de ce contrat, profiteront, en plus de l'abattement fixe de **152 500 €**, d'un abattement de **20%** des sommes soumises à l'imposition.

## **1.1.3 Caractéristiques et clauses des contrats d'assurance-vie**

### **1.1.3.1 Le Taux Minimum Garanti (TMG)**

Les assureurs s'engagent dans les conditions générales des contrats d'assurance-vie en euros à servir un taux minimum garanti. En réalité, ce dernier définit le taux de revalorisation minimum et obligatoire que l'assureur doit appliquer à l'encours de l'assuré chaque année lors de la revalorisation. C'est un engagement pour l'assureur, acté dans les conditions générales du contrat, auquel il ne peut se soustraire. Il est contraint de servir à minima ce TMG, voire un plus fort taux. Le rendement annuel du contrat de l'assuré ne peut donc pas être inférieur

au TMG. L'arrêté du 7 Juillet 2010 portant modification des modalités de garanties du taux minimum par les entreprises d'assurance encadre la réglementation du TMG depuis le 1er août 2010 suite aux nombreux abus des assureurs qui proposaient des taux excessifs pour appâter les clients :

- Le TMG doit être calculé annuellement et exprimé sous forme de pourcentage
- Il ne peut excéder **85%** du rendement des actifs de l'assureur sur les 2 dernières années
- Le taux minimum doit être garanti sur une durée d'au moins 6 mois et au plus égal à la période séparant la date d'effet de la garantie de la fin de l'exercice suivant
- La revalorisation des contrats bénéficiant d'un TMG ne doit pas nuire à la rémunération des autres produits
- Le plafond du TMG est fixé selon un indice qui dépend de la conjoncture des marchés obligataires

L'article A132-3 du code des assurances définit un seuil réglementaire que le TMG ne doit pas dépasser :

$$TMG_{max} = \min(150\% \times 75\% \times TME; \max(120\% \times 75\% \times TME; 110\% \times MTS))$$

TME désigne le Taux Moyen des Emprunts d'Etat, calculé sur une base semestrielle et est égal au plus élevé des deux taux suivants : taux à l'émission et taux de rendement sur le marché secondaire.

MTS représente la moyenne des taux moyens servis aux assurés.

Le choix du TMG est crucial pour les assureurs et doit être stratégique à bien des égards. Un taux minimum garanti élevé pour attirer plus de clients pourrait freiner l'assureur dans sa liberté d'investir. Il sera contraint d'investir dans des actifs à revenus certains du type obligations d'Etat afin de respecter ses engagements.

### 1.1.3.2 La participation aux bénéfices (PB) <sup>3</sup>

La loi prévoit que l'assureur doit calculer pour chaque exercice un montant minimal de participation aux bénéfices déterminé globalement à partir d'un compte de participation aux résultats. En vertu de l'article L.331-3 du Code des Assurances : «Les entreprises d'assurance sur la vie ou de capitalisation doivent faire participer les assurés aux bénéfices techniques et financiers qu'elles réalisent, dans les conditions fixées par arrêté du Ministre de l'Economie et des Finances». Ces profits sont distribués aux assurés par l'intermédiaire de la participation aux bénéfices. Cette dernière est reversée que sur les contrats en euros. L'allocation minimale est fixée de façon globale à hauteur de **85%** du solde du compte de résultat financier et de **90%** de celui du résultat technique. Il faut distinguer ces trois types de PB : la PB

---

3. Source : Présentation ISFA 2013 Camille GRACIANI, Assurance vie et participation aux bénéfices

réglementaire minimale (fixée par le Code des assurances), la PB contractuelle (clause de PB dans le contrat) et la PB discrétionnaire (pour des fins commerciales).

Le calcul se fait via un compte de participation aux résultats : Résultat technique (souvent faible) et Résultat financier = Produits financiers – Charges financières. A partir du compte de résultat, on sert d'abord les taux techniques, engendrés par la fixation du taux technique. Le reliquat est dévolu à la PB qui peut être immédiate ou différée (provision pour participation aux bénéficiaires, qui doit être distribuée sous 8 ans). Pour la PB réglementaire, le montant minimal est le solde d'un compte de participation aux résultats.

Solde technique = solde de souscription – charges d'acquisition et de gestion nettes

Ce solde technique tient compte des pertes sur les garanties accessoires de dommages corporels que l'on peut retrouver dans certains contrats collectifs.

Pour la PB contractuelle, d'après l'article L.132-5 du code des assurances, le contrat doit préciser les conditions d'affectation des bénéficiaires techniques et financiers. La clause doit spécifier le mode de calcul, les taux de participation aux résultats, la date d'attribution, la période d'incorporation (si la PB est différée) et les supports financiers.

En ce qui concerne la PB discrétionnaire, l'assureur peut décider de distribuer une part des produits financiers réalisés durant l'année à tous ou une partie des bénéficiaires des contrats d'assurance vie.

Définie par l'article R. 331-3 du code des assurances, la provision pour participation aux bénéficiaires (PPB) appartient aux assurés. Elle permet de lisser la rémunération des contrats et donc de piloter l'activité en fonction des contraintes commerciales (taux servis par les concurrents) et des contraintes réglementaires (redistribution sous 8 ans). La PPB peut permettre de servir les TMG court terme.

### 1.1.3.3 Les frais

Les frais sont nécessaires pour permettre à l'assureur d'amortir ses coûts (de gestion, d'acquisition. . . .). On distingue ainsi des :

- frais d'entrée et sur versements : il s'agit d'un pourcentage prédéfini, prélevé sur tous les versements initiaux, libres ou programmés.
- frais de gestion : prélevés sur l'épargne constituée temporellement, ils peuvent permettre à l'assureur de se constituer une marge
- frais de sortie composés des frais d'arrérage (déduits lors du versement d'une rente) et des frais de rachat prélevés sur les montants de rachat
- frais d'arbitrage entre les différents supports
- frais de commissions liés aux réseaux de distribution qui ne sont pas prélevés sur le contrat

#### **1.1.3.4 Les rachats et avances**

Dans un contrat d'épargne, le souscripteur a le droit, tout au long du contrat, de récupérer toute ou une partie de son épargne. C'est le droit de rachat. Il peut aussi demander en cas de besoin de liquidités une avance.

##### **1.1.3.4.1 Rachat partiel**

Il permet au souscripteur de retirer une partie de sa provision mathématique tandis que l'autre partie reste investie dans le contrat. Il ne met pas fin au contrat, l'antériorité fiscale est conservée et le solde du capital continue de fructifier.

##### **1.1.3.4.2 Rachat Total**

Il permet au souscripteur de se faire rembourser l'intégralité de la valeur de rachat du contrat avant son échéance. Il met fin au contrat et aux garanties qui y sont attachées.

##### **1.1.3.4.3 Avance**

L'avance est un prêt consenti par l'assureur au souscripteur en fonction de ses avoirs sur son contrat. Elle est d'autant plus intéressante qu'elle ne met pas fin au contrat et la valeur de rachat de ce dernier est inchangée. La valeur de rachat n'est en effet pas diminuée du montant de l'avance et les fonds investis sur le contrat continuent de se valoriser sur la même base qu'avant l'avance.

Comme tout prêt, l'avance fait l'objet d'un règlement qui varie en fonction des compagnies d'assurance. Le montant maximum de l'avance autorisé par contrat varie entre **60%** et **80%** de l'épargne investie selon les compagnies d'assurance. Le délai maximum de remboursement de l'avance est de 3 ans en général et peut être renouvelé d'une à deux fois. Le taux d'intérêt dû est fixé chaque année par la compagnie d'assurance. Il dépend soit du taux moyen des emprunts d'Etat (TME), soit du rendement du fonds en euros de l'année précédente et est majoré de **1** à **1,5%**.

#### **1.1.3.5 Les prélèvements sociaux**

Les intérêts générés par les contrats d'assurance vie (sauf PERP et Madelin) qui proposent un fonds en euros, sont soumis aux prélèvements sociaux chaque année au moment de leur inscription en compte (en général en Décembre). Depuis le 1er juillet 2011 (article 22 de la loi de finances pour 2011), il en va de même pour les intérêts des fonds en euros des contrats multi-supports. En effet, jusqu'à cette date, les produits des contrats d'assurance-vie multi-supports n'étaient soumis aux prélèvements sociaux que lors de leur dénouement en cas de vie (rachat partiel ou total) ou, depuis le 1er janvier 2010, au décès de l'assuré. Un mécanisme de restitution est prévu au rachat ou au décès, dans le cas où la somme des prélèvements acquittés sur le compartiment euro du contrat est supérieure au montant des prélèvements

sociaux calculés sur la totalité des produits du contrat à la date du rachat ou du décès. Dans ce cas, l'excédent de prélèvements sociaux déjà acquittés est restitué par l'intermédiaire de l'entreprise d'assurance.

L'article 8 de la loi de financement de la sécurité sociale 2014 a supprimé rétroactivement à compter du 26/09/2013 l'application de la méthode des taux historiques pour le calcul des prélèvements sociaux afférents aux produits exonérés d'impôt sur le revenu, pour la remplacer par un taux unique qui est celui en vigueur au moment du fait générateur (rachat, arrivée à échéance ou décès), soit actuellement **15,50 %** décomposé comme suit : **8,2 %** au titre de la CSG, **0,5%** au titre de la CRDS, **5,4%** au titre du prélèvement social, **0,3%** au titre du CAPS, et **1,1%** pour le financement du RSA.

Par ailleurs, pour les gains réalisés sur les supports en unités de compte, les prélèvements sociaux ne sont prélevés qu'à l'occasion des rachats ou encore lors du décès de l'assuré.

### 1.1.3.6 Les options de gestion

Ce sont des services financiers qui permettent de répartir automatiquement de l'épargne en fonction de l'évolution des marchés financiers. Les options de gestion peuvent se décomposer en trois grands groupes :

#### 1.1.3.6.1 La gestion profilée

L'assureur crée plusieurs profils de gestion et selon son appétence au risque, l'épargnant aura le choix entre ces différents supports sur lesquels seront investis ses versements. Par exemple, plusieurs assureurs proposent 4 profils :

- **Profil Prudence** : avec une composition à forte majorité d'obligations, ce profil est destiné aux épargnants ayant une faible tolérance au risque. On a une composition qui permet de réduire les risques liés aux marchés financiers
- **Profil Equilibre** : qui ménage un équilibre entre sécurité (obligations) et rentabilité (actions). Il est adapté pour les épargnants ayant une tolérance modérée au risque. Sa composition équilibrée permet de viser une certaine stabilité des rendements tout en profitant de la performance des marchés boursiers
- **Profil Diversifié** : avec une composition diversifiée entre actions et obligations, il permet de profiter des marchés boursiers internationaux tout en préservant un équilibre dans l'allocation des actifs afin de veiller à préserver au mieux le capital investi.
- **Profil Dynamique** : qui privilégie l'investissement en actions. Il est destiné aux épargnants ayant une forte tolérance au risque et pouvant supporter les risques liés aux fluctuations des marchés financiers

#### 1.1.3.6.2 La gestion libre

Elle permet au souscripteur de répartir librement son épargne entre les différents supports. Toutefois ce mode de gestion nécessite que l'investisseur soit averti pour éviter le risque de perte en capital sur les investissements en unités de compte.

#### 1.1.3.6.3 La gestion évolutive ou à horizon

Des arbitrages sont effectués à intervalles réguliers pour ventiler l'épargne en fonction du projet de l'épargnant. La gestion évolutive consiste en une sécurisation automatique du capital :

- La sécurisation des plus-values : un certain seuil est fixé et dès que les plus-values issues d'un support en unités de compte atteignent ce seuil, elles seront automatiquement transférées vers un support plus sécurisé comme le fonds euros.
- La dynamisation des plus-values du fonds en euros : chaque année, les plus-values générées sur le fonds en euro sont basculées à partir d'un certain seuil vers un support en unités de compte. Le capital restant sur le fonds en euros reste ainsi préservé des aléas des marchés financiers.
- L'investissement progressif : c'est une option qui permet d'investir progressivement sur les marchés financiers et de lisser dans le temps les fluctuations de ces derniers. Elle permet aussi, à partir d'un versement sur un support d'origine prudent (fonds euros par exemple) d'arbitrer selon une périodicité choisie vers un support plus dynamique
- Le rééquilibrage automatique : permet de maintenir la répartition de l'épargne choisie initialement et de la sauvegarder malgré les fluctuations du marché financier. Elle a pour but de réinvestir le capital vers des supports qui sont les moins impactés par les aléas du marché.

#### 1.1.3.7 Les garanties décès annexes

Contrairement aux contrats en euros qui garantissent contractuellement un rendement minimum, les contrats en unités de compte sans garantie décès font supporter à l'assuré l'intégralité du risque lié aux fluctuations des marchés financiers. Dans un contexte actuel de chute des marchés boursiers, les assureurs sont amenés à partager le risque de marché avec l'assuré pour rendre le contrat plus attractif. C'est ainsi que les assureurs ajoutent des garanties décès optionnelles aux contrats. Parmi ces garanties, on a :

- **La Garantie Plancher** : c'est la plus courante. Elle protège les bénéficiaires en cas de décès de l'assuré contre la baisse des marchés financiers. Ils recevront une prestation égale au maximum entre le cumul des primes investies nettes de frais et de rachats et le montant de la provision mathématique à la date de décès.
- **La Garantie Décès Indexée** : elle propose une prestation dont le montant est au minimum égal à la prime investie, capitalisée à un taux défini au préalable dans les

conditions générales.

- **La Garantie Décès Cliquet** : elle propose comme prestation minimum la plus haute valeur de l'unité de compte durant la période de référence.
- **Autre Garantie Décès** : versement d'un capital supplémentaire en cas de de décès accidentel (par exemple : montant fixe 10 000 € ou un % de l'épargne).

## 1.2 Présentation d'Optimum Vie et de son portefeuille

### 1.2.1 Présentation de la société<sup>4</sup>

Optimum Vie est l'une des filiales du groupe canadien Optimum. Fondé à Montréal au Canada par des actuaires en 1969, le Groupe Optimum est un groupe financier privé canadien d'envergure internationale qui exerce ses activités dans divers secteurs : assurance vie et réassurance vie, assurance de dommages, gestion d'actifs et de services financiers, actuariat-conseil et gestion immobilière. Il emploie plus de 500 personnes au Canada, aux Etats-Unis et en France.

Acquise en 1975, Optimum Vie demeure en France la seule société d'assurance vie de propriété canadienne. Elle offre principalement des contrats liés à des fonds d'investissement, de capitalisation, d'épargne et de retraite, ainsi que des produits traditionnels de prévoyance. Elle travaille également de pair avec ses clients pour développer des produits de niche ou d'extension de gamme, des solutions sur mesure de contrats affinitaires ou de produits en marque blanche. La gamme de produits est en constante évolution et Optimum Vie propose des solutions adaptées à la situation personnelle de ses clients. Elle collabore avec un réseau de courtiers indépendants, des courtiers grossistes, des groupements de courtiers mais aussi avec des Mutuelles et des assureurs non-vie.

### 1.2.2 Le portefeuille d'Optimum Vie

La société Optimum Vie conçoit des solutions d'assurance vie performantes qui répondent aux attentes et aux besoins spécifiques des clients. Chaque solution est développée au cas par cas. Optimum Vie offre des produits clés en main dans trois domaines d'expertise : l'épargne, la retraite et la prévoyance. Voici une liste non exhaustive des produits commercialisés :

#### 1.2.2.1 Les produits d'épargne

- **Optimum Projets** : c'est un contrat multi-support en UC à versements périodiques et/ou libres avec option de pilotage automatique vers le fonds équilibre à 5 ans du terme du contrat et vers le fonds prudent à 2 ans du terme du contrat. Il permet

---

4. Source : site internet Groupe Optimum

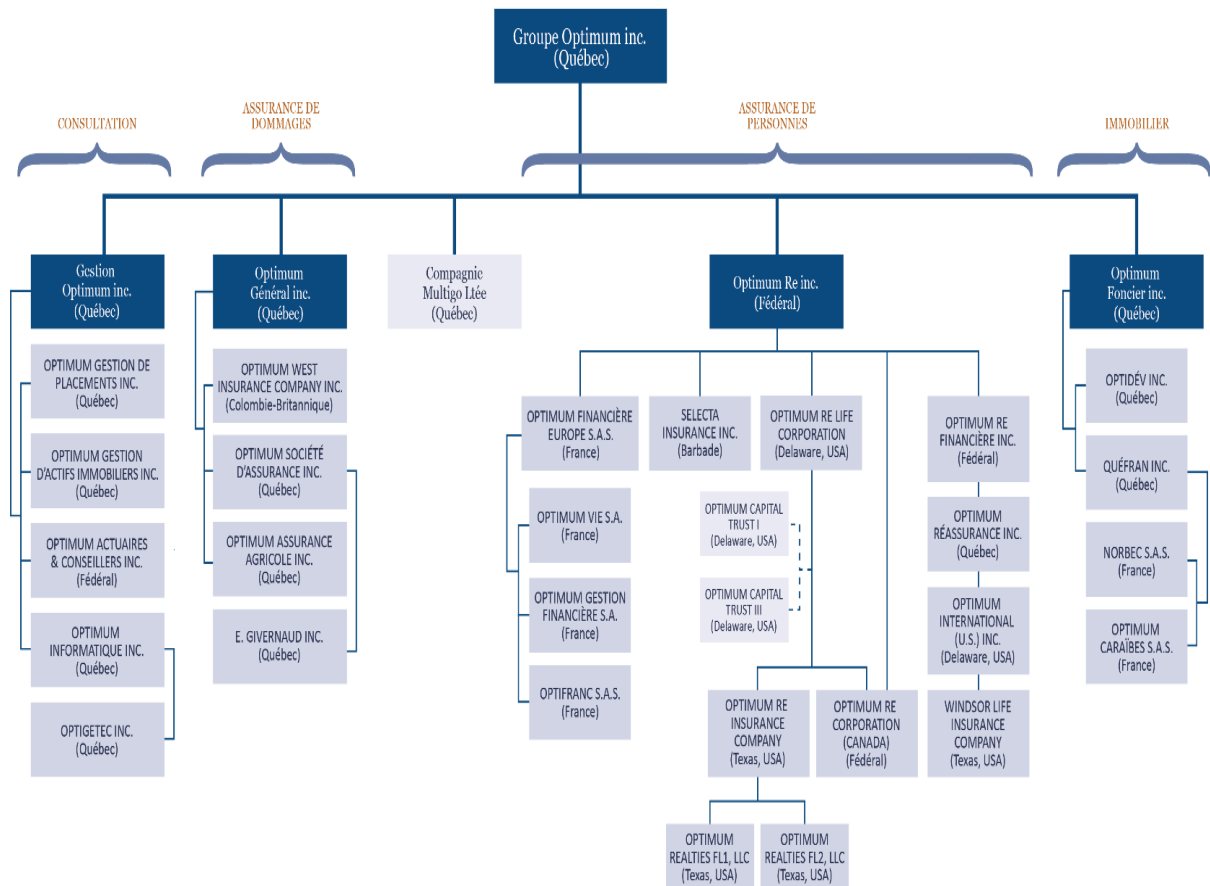


FIGURE 1.1 – Organigramme GROUPE OPTIMUM

la constitution d'un capital ou la construction d'un projet à long terme à partir de versements périodiques. Il comporte aussi une garantie complémentaire en cas de décès.

- Optimum Placements Privilèges est un contrat multi-support en UC et en euros à versements programmés et/ou libres avec une garantie complémentaire décès obligatoire. Il permet la constitution d'un capital et les clients bénéficient d'un investissement diversifié valorisant le long terme.
- Optimum Placements est un contrat multi-support en UC à versements programmés et/ou libres avec une garantie complémentaire décès obligatoire
- Optimum Capitalisation est un contrat en UC permettant la constitution d'un capital mais ne bénéficiant pas de la fiscalité assurance-vie. En réalité, il est assujéti au régime fiscal français des contrats de capitalisation. À ce titre, il ne se dénoue pas par le décès du souscripteur mais est transmis à ses héritiers. En cas de décès du Souscripteur, la valeur de rachat du contrat au jour du décès est soumise aux droits

de succession.

- Optimum Sécurité est un contrat en euros, une solution pour faire progresser son épargne sans risques. Il est éligible à l'épargne handicap.
- Solution Epargne Plus est un contrat en unités de compte à versements périodiques et/ou libres. Il permet la constitution d'un capital et comporte une garantie Décès.

Mis à part Optimum Projets et Solution Epargne Plus qui sont des contrats collectifs à adhésion individuelle et facultative, tous les autres contrats sont à caractère individuel.

### **1.2.2.2 Les produits de retraite**

- Optimum Retraite est un contrat multi-support en UC à versements périodiques et/ou libres avec option de pilotage automatique vers le fonds équilibre à 5 ans du terme du contrat et vers le fonds prudent à 2 ans du terme du contrat. Il permet de préparer une épargne indispensable pour faire face à la baisse des revenus. Il comporte une garantie Décès.
- Optimum Retraite Madelin est un contrat multi-support UC à versements périodiques et/ou libre qui permet la constitution d'un capital retraite payable sous forme de rente viagère. C'est un produit de retraite adapté aux travailleurs non-salariés pour disposer d'un complément de revenu significatif dans un cadre fiscal privilégié.
- Optimum Retraite PERP est un contrat multi-support souscrit dans le cadre de l'article L144-2 du Code des assurances ainsi que des dispositions du décret n°2004-342 du 21 avril 2004 relatif au Plan d'Épargne Retraite Populaire pour la constitution d'une rente viagère se dénouant au plus tôt à la date de liquidation de la retraite.

Les produits de retraite sont tous collectifs à adhésion individuelle et facultative.

### **1.2.2.3 Les produits de Prévoyance**

- Optimum Perspectives Vie Entière est un contrat en euros à primes périodiques. C'est un contrat d'assurance décès Vie Entière avec doublement du capital garanti en cas de décès accidentel et triplement en cas de décès lié au terrorisme et/ou transport en commun. Il permet de se prémunir contre les aléas de la vie et de protéger sa famille.
- Optimum Pérennité est aussi un contrat en euros à primes périodiques. C'est un contrat Vie entière à effet différé avec contre-assurance des primes en cas de décès pendant la période de différé.
- Optimum Prévoyance Essentielle est un contrat en euros à primes périodiques : Temporaire Décès jusqu'à la fin de période de paiement des primes, suivi d'une Vie Entière. C'est un contrat à prime nivelée avec une distinction fumeur/non-fumeur. Il permet de transmettre facilement son patrimoine.

Par ailleurs, spécialiste des solutions sur mesure, Optimum Vie travaille en étroite collaboration avec un vaste réseau de courtiers qui constituent les intermédiaires entre la société et les clients.

## 1.3 Problématique du mémoire : les comportements de rachat en assurance-vie<sup>5</sup>

### 1.3.1 Le risque de rachat

Comme vu précédemment, certains produits d'assurance-vie offrent la possibilité aux assurés d'effectuer des rachats sur leur contrat. A l'instar de l'option américaine, le rachat serait une option que l'assuré pourra exercer à son bon vouloir afin de retirer tout ou une partie de son épargne à tout moment avant la date d'échéance de son contrat. Le risque de rachat fait partie des risques les plus importants d'une compagnie d'assurance au sens où les rachats interfèrent dans leur gestion de placements et dans leur rentabilité. Du point de vue de l'assuré, l'option de rachat est très intéressante car elle lui permet d'avoir accès à son argent à tout moment et de disposer d'assez de liquidités au moment d'investir dans un projet ou en cas de difficultés. Alors que pour l'assureur, la réalité est tout autre. Il doit en effet faire face au risque de liquidité qui découle d'une vague des rachats suite à la levée des contraintes fiscales ou de sa faible rémunération des assurés par rapport au marché. Les assureurs sont ainsi contraints d'investir dans des actifs leur permettant d'assurer les versements lors des rachats mais aussi de pouvoir les vendre de façon imprévue.

On distingue généralement deux types de rachat : les rachats structurels et les rachats conjoncturels. Les rachats structurels sont ceux qui sont «anticipables ». Ils sont liés aux caractéristiques des assurés ; la décision de rachat est prise indépendamment du niveau des taux de marché. Les rachats conjoncturels sont quant à eux, liés à l'environnement concurrentiel, économique, financier ou à la réputation de la compagnie. Comparés aux rachats structurels, ils sont difficilement modélisables.

### 1.3.2 Intérêts de la connaissance de ce risque

Il est crucial pour les assureurs de maîtriser le risque de rachat afin de :

- Mettre en place une gestion Actif-Passif pour gérer le risque de liquidité
- Prédire la probabilité de rachat des individus de leur portefeuille
- Comprendre les comportements pour une meilleure segmentation de la population assurée

---

5. Source principale : Cours de Xavier Milhaud, Comportements de rachat en assurance-vie, ISFA 2017

- Adapter les caractéristiques des nouveaux produits lors de leur design pour mieux anticiper les comportements de rachat
- Gagner en parts de marché par l’optimisation de leur souscription
- Rendre plus strictes les incitations à conserver le contrat un minimum de temps

### 1.3.3 Approches de modélisation

Plusieurs approches de modélisation ont été proposées dans la littérature. Entre autres on a :

- l’approche Mathématiques Financières : le rachat est considéré comme une option américaine avec des hypothèses très importantes : agents rationnels et optimaux. La modélisation consiste en la valorisation de l’option de rachat.
- l’approche statistique série temporelle : modélisation collective du taux de rachat. Les données empiriques permettent de calibrer la fonction de rachat suivante :

$$r_d = r_0 * [1 - a * \ln(d + 1) * (\ln(d + 1) - b)]$$

avec  $d$  : la durée,  $r_0$  est observé et  $a$  et  $b$  sont les paramètres à calibrer.

- l’approche microéconomique : individuelle, maximisation de la théorie de l’utilité espérée
- l’approche probabiliste/économétrique : modélisation individualisée à travers des modèles d’intensité, des GLM ou des modèles de segmentation (arbres,...)

### 1.3.4 Le cadre réglementaire

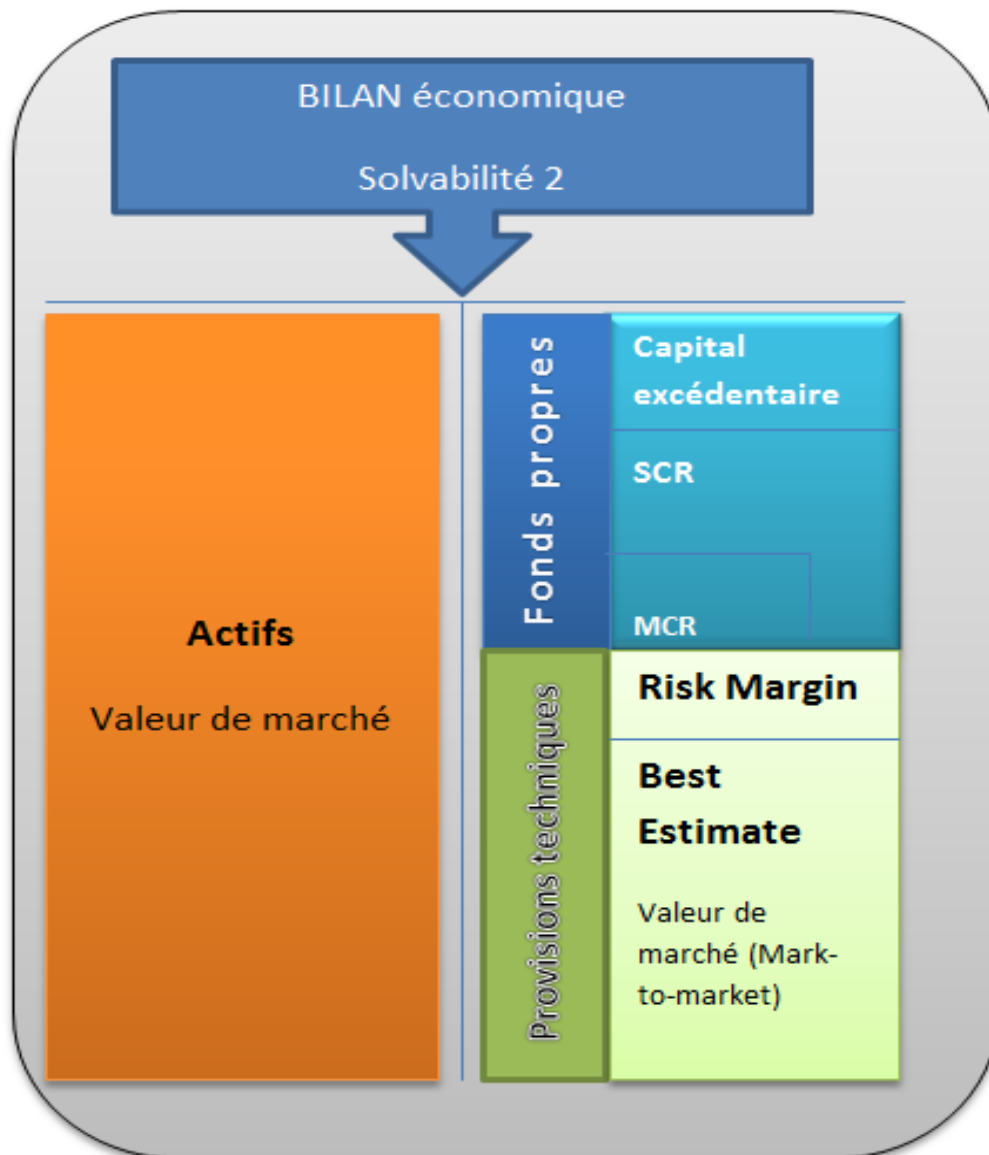
Le principe de base de la directive de Solvabilité II est une valorisation économique des éléments de bilan pour les besoins prudentiels et notamment pour le calcul des provisions techniques. Le tableau ci-dessous montre de façon schématique le bilan prudentiel sous le régime Solvabilité II<sup>6</sup>.

Le Best Estimate représente la valeur actualisée des cash flows futurs de l’assureur et repose sur des hypothèses réalistes et non nécessairement prudentes. Pour valoriser le passif en « Best Estimate », le modèle ALM, nécessaire pour toutes les sociétés d’assurance-vie commercialisant des contrats d’épargne, doit tenir compte des incertitudes liées aux comportements des assurés (rachats, décès,...).

L’article 79 de la Directive cadre Solvabilité 2 prévoit explicitement une évaluation de la

---

6. Source : Présentation sur Solvabilité II chez Optimum Vie



valeur de l'option de rachat.

Lorsqu'elles calculent leurs provisions techniques, les entreprises d'assurance et de réassurance tiennent compte de la valeur des garanties financières et de toute option contractuelle incluses dans leurs contrats d'assurance et de réassurance.

Pour le calcul du SCR (Solvency Capital Requirement), la méthode retenue est celle proposée par le QIS 5 (Quantitative Impact Studies), la méthode standard.

Les recommandations pour le calcul du SCR se font en 2 étapes :

(1) Chocs rep/down appliqués au taux de rachat structurel. Ils avaient été calibrés sur une étude empirique en 2003 aux U.K et en Pologne (sur un seul produit). On a :

$$LR_{up} = \text{Lapse Rate up} = \min(100\%, 150\% * LR)$$

où LR est le taux de rachat structurel, (*Lapse = Rachat*)

$$LR_{down} = \min(0, \max(50\% * LR, LR - 20\%))$$

Le risque de mauvaise estimation de LR et les rachats conjoncturels sont censés être intégrés.

(2) Mass Lapse Event : correspond à un évènement rare, catastrophe de type event. Il n'y a pas de données objectives pour calibrer correctement ce choc. L'idée est de considérer une perte de **30** % de l'ensemble des gains du portefeuille. Ce choc devrait s'ajuster au type de contrat et intégrer les rachats conjoncturels. Au final, c'est le maximum des deux étapes qui est retenu.

Il faut noter que parallèlement à cette formule standard SII, des recommandations ont été émises par l'ACPR pour le développement de modèles internes.

Par ailleurs, la publication de l'ACPR dans les Orientations Nationales Complémentaires (ONC) suggère une idée qui consiste en la séparation rachats structurels/ rachats conjoncturels. La stratégie est la suivante :

1) déterminer via un modèle interne le taux de rachat structurel

2) y ajouter un taux de rachat conjoncturel évalué comme une fonction du spread entre taux servi et taux référence marché.

Le taux de rachat serait ainsi égal à la somme des 2 composantes.

Remarque : en France, ce sont souvent des taux long terme pour la référence (exemple : OAT 10 ans).

$$\text{Rachat conjoncturel : RC} = \begin{cases} RC_{max} & \text{si } TS < TA < \alpha \\ RC_{max} * \frac{(TS-TA)-\beta}{\alpha-\beta} & \text{si } \alpha \leq TS - TA \leq \beta \\ 0 & \text{si } \beta \leq TS - TA \leq \gamma \\ RC_{min} * \frac{(TS-TA)-\gamma}{\delta-\gamma} & \text{si } \gamma \leq TS - TA \leq \delta \\ RC_{min} & \text{si } TS - TA > \delta \end{cases}$$

avec TS = taux servi et TA= taux attendu

- $\alpha$  est le seuil en-deçà duquel les rachats conjoncturels sont constants et fixés à  $RC_{max}$  (l'écart de taux n'explique plus ainsi le comportement des assurés) ;
- $\beta$  et  $\gamma$  sont respectivement les seuils d'indifférence à la baisse et à la hausse du taux servi (entre ces deux seuils le comportement de l'assuré n'est pas modifié) ;
- $\delta$  est le seuil au-delà duquel la diminution du taux de rachat structurel est constante et fixée à  $RC_{min}$  (ce n'est plus l'écart de taux qui explique le comportement des assurés).

Les paramètres qui sont proposés sont :

$RC_{min} \in [-6\%, 4\%]$  ,  $RC_{max} \in [20\%, 40\%]$

et  $\alpha \in [-6\%, 4\%]$ ,  $\beta \in [-2\%, 0\%]$ ,  $\gamma = 1$ , et  $\delta \in [2\%, 4\%]$

Cette méthode pose problème au sens où les seuils sont très subjectifs, elle n'intègre pas la corrélation entre les assurés et si toutes les compagnies adoptent ces préconisations dans leur modèle interne, il y aura un fort risque systémique.

Au final avec cette méthode, on prend :  $LR^{\text{choqué}} = \min(1, \max(0, RC + LR))$

## 1.4 Paysage fiscal des contrats d'assurance-vie

Le succès de l'assurance-vie s'explique en grande partie par sa fiscalité favorable. La fiscalité de l'assurance-vie dépend de plusieurs variables entre autres le caractère mono-support ou multi-support du contrat, la durée du contrat et le dénouement en cas de vie de l'assuré ou en cas de décès.

### 1.4.1 Généralités

#### 1.4.1.1 Fiscalité en cas de rachat du contrat d'assurance-vie

##### 1.4.1.1.1 Avant 8 ans

Pour les contrats souscrits après le 26 Septembre 1997, lors d'un rachat partiel ou total, les plus-values sont soit intégrées dans le revenu imposable, soit soumises à un prélèvement forfaitaire libératoire (PFL) lors du retrait. Ce dernier est de 35% avant 4 ans et de 15% entre 4 et 8 ans. Les situations de force majeure comme le licenciement et l'invalidité, permettent le rachat ou le dénouement anticipé sans imposition.

##### 1.4.1.1.2 Après 8 ans

La fiscalité est encore plus favorable car elle s'allège considérablement une fois que le contrat a atteint le cap des 8 ans. Après 8 ans, les plus-values sont imposables, soit par réintégration dans le revenu imposable, soit par prélèvement forfaitaire libératoire au taux de 7,5%. De

plus, l'assuré célibataire bénéficie d'un abattement fiscal de 4600 € et pour un couple marié ou pacsé, ce sera 9200 €.

**Remarque** : Pour les contrats souscrits avant le 1er janvier 1983, les plus-values sont totalement exonérées. Il en va de même pour les contrats ouverts avant le 26 septembre 1997 à condition que les versements aient été effectués avant cette date.

#### **1.4.1.1.3 Les prélèvements sociaux**

Ils sont appliqués différemment en fonction du type du contrat. Sur les fonds en euros, qu'ils soient souscrits dans un contrat mono-support ou multi-supports, les prélèvements sociaux sont prélevés chaque année sur les intérêts perçus (depuis 2011 pour les fonds euros des contrats multi-supports).

Pour les unités de compte, les prélèvements sociaux sont retenus au moment des retraits ou lors de la clôture, notamment en cas de décès de l'assuré. Le taux global des prélèvements sociaux s'élève à 15,5 % actuellement. (Approfondissement du cadre fiscal au 1.4.2))

#### **1.4.1.2 L'imposition de la rente viagère**

L'assuré peut choisir de percevoir régulièrement son épargne sous forme de rente jusqu'à son décès. C'est un moyen pour lui de se constituer par exemple un complément de retraite. Les montants perçus dans le cadre d'une rente viagère sont imposables. L'imposition dépend toutefois de la nature de la rente. Ainsi il faut faire la distinction entre les rentes versées à la sortie d'un produit d'épargne retraite de type (Perp, Préfon, Madelin) et celles versées à la sortie de tout autre placement proposant sur option la sortie en rente (Assurance-vie et Perco par exemple). Pour les premières, elles sont imposées comme les pensions de retraite c'est-à-dire soumises à l'impôt sur le revenu après application de l'abattement de 10% qui prévaut pour ces dernières. Elles sont aussi assujetties aux prélèvements sociaux (8,4%). On parle alors de fiscalité des rentes à titre gratuit. Pour la deuxième catégorie, c'est la fiscalité des rentes à titre onéreux qui s'applique. Ces rentes sont soumises à l'imposition sur le revenu et aux prélèvements sociaux (15,5%) mais seulement sur une fraction de l'épargne.

L'imposition de cette fraction se calcule en fonction de l'âge du bénéficiaire de la rente au moment où il perçoit son 1er versement : 70 % s'il a moins de 50 ans, 50% de 50 à 59 ans, 40 % de 60 à 69 ans et 30 % à partir et au-delà de 70 ans.

#### **1.4.1.3 Fiscalité de l'assurance-vie en cas de décès**

En cas de décès, le capital versé aux bénéficiaires désignés est exonéré de droits de succession dans les limites fiscales de l'assurance-vie. La loi TEPA en faveur du travail, de l'emploi et du pouvoir d'achat du 21/08/2007 prévoit l'exonération totale des droits de succession si le bénéficiaire de l'assurance-vie est :

- soit le conjoint ou le partenaire lié au défunt par un PACS
- soit un frère ou une sœur du défunt, célibataire, veuf divorcé ou séparé de corps, à la condition que la personne soit lors de l'ouverture de la succession à la fois :
  - âgée de plus de 50 ans ou atteinte d'une infirmité la mettant dans l'impossibilité de subvenir par son travail à ses besoins
  - domiciliée constamment avec le défunt pendant les 5 années qui ont précédé son décès

Pour les autres bénéficiaires, le traitement fiscal varie selon l'âge de l'assuré lors du versement des primes.

- **Avant 70 ans**

Le capital transmis en cas de décès n'est pas soumis aux droits de succession dans la limite de 152 500 € par bénéficiaire désigné. Pour les sommes supérieures à 152 500 € et inférieures à 700 000 € depuis le 01/07/2014, un taux d'imposition forfaitaire de 20 % est appliqué. Ce taux d'imposition passe à 31,25 % pour la fraction des sommes supérieures à 700 000 €.

- **Après 70 ans**

Les versements effectués après 70 ans sont exonérés de droits de succession à hauteur de 30 500 €. Au-delà, les sommes sont réintégrées dans la succession. Les plus-values générées sont, quant à elles, totalement exonérées de droits de succession.

**Remarque** : Les «vieux contrats» sont exonérés (cf. tableau récapitulatif)

Date d'ouverture de l'assurance-vie	Age de l'assuré lors du versement	Date à laquelle est intervenu le versement		Imposition des plus-values
		Avant le 12/10/1988	Après le 13/10/1988	
Jusqu'au 20/11/1991	Quel que soit l'âge de l'assuré	Exonération des droits de succession	Jusqu'à 902838 €, imposition à 20% du capital taxable, après un abattement de 152500 € / bénéficiaire. Au-delà, taxation forfaitaire au taux de 25%	<b>Contributions sociales</b> 8,2 % au titre de la <b>CSG</b> + 0,5% au titre de la <b>CRDS</b> + 5,4% au titre du prélèvement social + 0,3% au titre du <b>CAPS</b> + 1,1 % pour le financement du <b>RSA</b>
Entre le 21/11/1991 et le 12/10/1998	Moins de 70 ans	Exonération des droits de succession		
	Plus de 70 ans	Exonération à hauteur de 30500 € répartie entre les bénéficiaires (toutes assurances-vie confondues) puis taxation selon le degré de parenté avec l'assuré. Les intérêts accumulés sur l'assurance-vie sont exonérés		
Entre le 13/10/1988 et le 30/06/2014	Moins de 70 ans	Exonération jusqu'à 152 500 €, impositions à 20% du capital taxable jusqu'à 902 838 €, au-delà taxation forfaitaire au taux de 25%		
	Plus de 70 ans	Exonération à hauteur de 30 500 € répartie entre les bénéficiaires (toutes assurances-vie confondues) puis taxation selon le degré de parenté avec l'assuré. Les intérêts accumulés sur l'assurance-vie sont exonérés		
A partir du 01/07/2014	Moins de 70 ans	Exonération jusqu'à 152 500 €, puis imposition à 20% du capital taxable jusqu'à 700 000 € et 31,25% au-delà		
	Plus de 70 ans	Exonération à hauteur de 30 500 € répartie entre les bénéficiaires (toutes assurances-vie confondues) puis taxation selon le degré de parenté avec l'assuré. Les intérêts accumulés sur l'assurance-vie sont exonérés		

#### **1.4.1.4 L'impôt de solidarité sur la fortune (ISF)**

Pour les personnes soumises à l'impôt sur la fortune, leurs contrats d'assurance-vie font partie de leur patrimoine, de ce fait leur valeur de rachat doit être déclarée. Les capitaux et rentes perçus au terme du contrat entrent dans le patrimoine et sont donc imposables à ce titre. Le contribuable qui est assujéti à l'ISF, doit les incorporer à son capital taxable au 1er Janvier de l'année d'imposition.

### **1.4.2 Fiscalité des rachats <sup>7</sup>**

#### **1.4.2.1 Taxation des plus-values des contrats d'assurance-vie**

##### **1.4.2.1.1 Fiscalité liée au rachat (Hors prélèvements sociaux)**

Les plus-values rattachées aux contrats d'assurance-vie peuvent être imposables :

- lors d'un rachat total, la totalité des plus-values accumulées depuis l'ouverture du contrat peut être imposable
- lors d'un rachat partiel, c'est seulement le montant des plus-values compris dans le rachat qui peut être imposable.

---

7. Source : Fiscalité des produits d'épargne (Optimum Vie S.A)

	Adhésions aux contrats avant le 01/01/1983	Adhésions aux contrats entre le 01/01/1983 et le 31/12/1989	Adhésions aux contrats après le 01/01/1990
Versements effectués avant le 26/09/1997	Exonération	Exonération	Exonération
Versements effectués à partir du 26/09/1997		Si le contrat a entre 0 et 2 ans: Déclaration Impôt sur le revenu ou Prélèvement Forfaitaire Libératoire de 45%	Si le contrat a entre 0 et 4 ans: Déclaration Impôt sur le revenu ou Prélèvement Forfaitaire Libératoire de 35%
		Si le contrat a entre 2 et 4 ans: Déclaration Impôt sur le revenu ou Prélèvement Forfaitaire Libératoire de 25%	
		Si le contrat a entre 4 et 6 ans: Déclaration Impôt sur le revenu ou Prélèvement Forfaitaire Libératoire de 15%	
		Si le contrat a plus de 6 ans: Déclaration Impôt sur le revenu ou Prélèvement Forfaitaire Libératoire de 7,5%*	

Adhésion = date d'effet fiscale

\*Après abattement annuel de 4 600 € pour une personne seule ou 9 200 € pour un couple soumis à imposition commune. Exonération des plus-values générées sur les versements jusqu'à 30 489,80 € réalisés entre le 26/09/1997 et le 31/12/1997 sur les contrats souscrits avant le 26/09/1997.

#### 1.4.2.1.2 Assiette de taxation

Les méthodes de calcul de l'assiette de taxation diffèrent d'une société d'assurance à une autre mais donnent les mêmes résultats. L'administration fiscale autorise les assureurs à opter pour l'application d'une méthode globale qui a vocation à s'appliquer aux seuls contrats rachetables mentionnés au deuxième alinéa de l'article L. 131-1 du code des assurances c'est-à-dire aux contrats dont le capital ou la rente garantis sont exprimés totalement ou partiellement en unités de comptes.

La méthode globale est la méthode des compartiments fiscaux. Elle permet, après détermination du coefficient applicable à la valeur de rachat, d'affecter chaque versement aux deux compartiments fiscaux que sont : le compartiment exonéré et le compartiment imposé. Ainsi en fonction du compartiment affecté, les intérêts pourront être répartis à leur tour sur les deux compartiments. Cette méthode est susceptible d'être appliquée aux contrats soumis à des régimes fiscaux différents selon la date de versement des primes. Toutefois, il est nécessaire de faire la distinction entre les versements dont les intérêts sont exonérés et ceux pour lesquels les intérêts sont imposés (cf. tableau suivant)

Date d'ouverture de l'assurance-vie	Date à laquelle a été réalisé le versement	Montant du versement à l'origine	Imposition des plus-values	
			Abattement	Taux d'imposition
Jusqu'au 01/01/1983	Sans incidence	Sans incidence	Aucun	Exonération d'impôt sur les intérêts
Jusqu'au 25/09/1997	Jusqu'au 25/09/1997	< 30 490 €		
	Du 26/09/1997 au 31/12/1997	≥ 30 490 €	Pour les assurances-vie de plus de 8 ans: <ul style="list-style-type: none"> <li>• 4 600 € pour une personne</li> <li>• 9 600 € pour un couple marié.</li> </ul> Abattement annuel (toutes assurances-vie confondues) sur le montant des plus-values imposables si intégration aux revenus (ou en crédit d'impôt si PL)	Intégration des intérêts dans le revenu imposable (rubrique «valeurs et capitaux mobiliers - montant brut des revenus n'ouvrant pas droit à un abattement») ou Prélèvement libératoire de: <ul style="list-style-type: none"> <li>• 35% si la durée du contrat est inférieure à 4 ans               <ul style="list-style-type: none"> <li>• 15% pour un retrait effectué entre la 4e et la 8e année</li> </ul> </li> <li>• 7,50% après la 8e année</li> </ul>
	A partir du 01/01/1998			
A partir du 26/09/1997	Tous les versements	Sans incidence		

## Affectation en compartiments imposable et exonéré

Pour ce faire, nous optons ici pour la méthode globale et nous ne considérons que les contrats souscrits à partir du 01/01/1998. Le principe est de séparer la valeur de rachat entre compartiments imposable et exonéré. La première étape va consister à déterminer le coefficient d'imposition.

Notons :

$VRT_n$  : Valeur de rachat totale à la date  $n$

$PI_n$  : Prime Investie à la date  $n$

$PV_n$  : Prime Versée à la date  $n$

$f_s$  : frais

### • Calcul du coefficient

Soit  $\beta_n$  le coefficient de la méthode globale à la date  $n$ . Pour simplifier, nous posons  $\beta_{01/01/1988} = 0$ . Nous avons :

$$\beta_n = \frac{\beta_{n-1} \times VRT_n \text{ (avant versement)} + PI_n}{VRT_n \text{ (avant versement)} + PI_n}$$

avec  $PI_n = (1 - f_s) \times PV_n$

Ce coefficient appliqué à la valeur de rachat totale permet d'obtenir la valeur de rachat taxable à la date  $n$ . Par différence, on obtient la valeur de rachat exonérée.

$$VRI_n = \beta_n \times VRT_n$$

$$VRE_n = VRT_n - VRI_n$$

Pour séparer un rachat en part imposable et exonérée, on applique  $\beta_n$  au montant total du rachat (ce dernier s'entend tous supports confondus) :

$\text{Montant du rachat imposable} = \beta_n \times \text{Montant du rachat}$
$\text{Montant du rachat exonéré} = \text{Montant du rachat} - \text{Montant du rachat imposable}$

Les assiettes de calcul lors d'un rachat sont proportionnelles à celles déterminées en cas

de dénouement total (rachat total ou décès).

Le coefficient de proportionnalité utilisé lors des calculs est :

$$\text{coeff rachat} = \frac{\text{Montant du Rachat}}{\text{Epargne acquise au moment du rachat (avant rachat)}}$$

• Calcul des primes versées imposables et primées versées exonérées fiscales

**Après un versement**, on a :

Primes versées restantes imposables (après versement) = Primes versées restantes imposables (avant versement) + Montant du versement

Primes versées restantes exonérées (après versement) = Primes versées restantes exonérées (avant versement)

**Lors d'un rachat** (hors arbitrage) :

- Primes versées imposables rachetées = coeff rachat \* Primes versées restantes imposables (avant versement)

- Primes versées restantes imposables (après rachat) = Primes versées restantes imposables (avant rachat) – Primes versées imposables rachetées

- Primes versées totales rachetées = coeff rachat \* Primes versées restantes totales (avant rachat)

D'où Primes versées restantes totales (après rachat) = Primes versées restantes totales (avant rachat) – Primes versées totales rachetées

- Primes versées exonérées rachetées = Primes versées totales rachetées – Primes versées imposables rachetées

Et Primes versées restantes exonérées (après rachat) = Primes versées restantes totales (après rachat) – Primes versées restantes imposables (après rachat)

Il faut noter que nous traitons ici uniquement le cas des rachats partiels. En cas de décès, rachat total ou liquidation en rente totale, le coefficient équivalent à coeff rachat serait égal

à 1.

- Calcul de la plus-value fiscale en cas de rachat

La plus-value fiscale sera calculée comme suit (avant éventuelle restitution) :

Plus-Value Fiscale = Montant du Rachat – Primes versées restantes rachetées

Plus-value Fiscale Imposable = Montant du rachat imposable – Primes versées restantes imposables rachetées

Plus-value Fiscale exonérée = Montant du rachat exonéré – Primes versées restantes exonérées rachetées

Compartiment exonéré	Compartiment imposable
<b>Plus-value fiscale exonérée</b>	<b>Plus-value fiscale imposable</b>
Pas de retenue	Si option pour le PL: l'organisme payeur retient le pourcentage prévu de la plus-value imposable. L'effet de l'abattement de 4 600 € ou 9 200 € (après 8 ans) est restitué sous forme de crédit d'impôt
	Sinon prélèvement via déclaration d'impôt sur le revenu

- Cas d'exonération

Les produits des contrats d'assurance-vie sont exonérés d'impôt sur le revenu, quelle que soit la durée du contrat, lorsque le dénouement du contrat résulte :

- du licenciement de l'assuré ou de son conjoint
- de la mise à la retraite anticipée de l'assuré ou de son conjoint
- de l'invalidité de l'assuré ou de celle du conjoint correspondant au classement dans la 2ème ou 3ème catégorie
- pour les travailleurs indépendants, de la cessation d'activité non salariée du titulaire du contrat ou de son conjoint suite à un jugement de liquidation judiciaire

Pour les contrats avec option de sortie en rente viagère, les produits accumulés sur le contrat jusqu'à la sortie échappent à l'impôt sur le revenu quelle que soit leur durée même si les versements de la rente sont eux soumis à l'impôt sur le revenu.

#### 1.4.2.2 Calcul des prélèvements sociaux

##### 1.4.2.2.1 Les prélèvements sociaux au fil de l'eau pour les contrats euros

L'assiette imposable est constituée par le montant des produits (plus-values) du compartiment euro inscrit au contrat.

##### 1.4.2.2.1.1 Calcul des prélèvements sociaux (PS) sur les supports en euros lors de l'année de souscription N

**Assiette d'avance de PS 31/12/N** = Max (0 ; Produits € N)  
 = Max (0 ; Epargne € 31/12/N – versement initial € – autres versements € + rachat € brut de PS et frais)

**Avance de PS 31/12/N** = assiette d'avance PS 31/12/N × 15,5%

**Epargne € 31/12/N nette de PS** = Epargne € 31/12/N - avance de PS 31/12/N

##### 1.4.2.2.1.2 Calcul des prélèvements sociaux (PS) sur les supports en euros pour les années suivantes

**Assiette d'avance de PS 31/12/N+1** = Max (0 ; Produits € N+1)

= Max (0 ; Epargne € 31/12/N+1 – Epargne € 31/12/N nette de PS – versements € du 01/01/N+1 au 31/12/N+1 + rachat € brut de PS et frais du 01/01/N+1 au 31/12/N+1)

**Avance de PS 31/12/N+1** = assiette d'avance PS 31/12/N+1 × 15,5%

**Epargne € 31/12/N+1 nette de PS** = Epargne € 31/12/N+1 - avance de PS 31/12/N+1

#### 1.4.2.2.2 Les prélèvements sociaux au fil de l'eau pour les contrats d'assurance-vie multi-supports

L'article 22 de la loi de finances pour 2011 a aménagé les règles d'imposition aux prélèvements sociaux de la part des produits attachés aux droits exprimés en euros ou en devises des contrats d'assurance-vie multi-supports (article L.136-7 du Code de la sécurité sociale). Le rythme d'imposition des produits du compartiment euro des contrats d'assurance-vie multi-supports est aligné sur celui des contrats mono-supports exprimés en euros. Ce qui impose aux assureurs d'effectuer une «avance de trésorerie» à l'état, avance basée sur les seuls intérêts de la part euros des contrats sans tenir compte de l'éventuelle moins-value globale du contrat (avec supports UC).

##### 1.4.2.2.2.1 le fait générateur

Dans les contrats multi-supports, la part des produits exprimés en euros est, depuis le 1er juillet 2011, imposée lors de l'inscription en compte alors que les produits issus du compartiment en unités de compte sont imposés au dénouement du contrat.

##### 1.4.2.2.2.2 l'assiette

L'assiette imposable est constituée par le montant des intérêts du compartiment euro inscrits au contrat. L'assiette des prélèvements sociaux «au fil de l'eau» des supports euros des contrats multi-supports est le maximum entre «la PB sur encours nette de frais de gestion et brute de coûts de décès éventuels» et zéro.

L'assiette des prélèvements sociaux au fil de l'eau, correspondant à la PB sur encours totale calculée au taux servi, nette de Frais de Gestion (FG) est la suivante :

Assiette Avance  $PS_{31/12/N} = \text{Maximum (PB sur encours calculés au TS nette de frais de gestion ; 0)}$

avec **PB** : Participation aux Bénéfices, **TS** : Taux Servis

##### 1.4.2.2.2.3 Le montant des prélèvements sociaux au fil de l'eau

Le montant du prélèvement «au fil de l'eau» au 31/12/N associé à la taxe  $i$  sera :

Avance  $PS_{i,31/12/N} = \text{Dernier taux connu au 31/12/N du } PS_i * \text{Assiette Avance } PS_{31/12/N}$

Le montant des prélèvements sociaux est prélevé chaque année au 31/12/N directement sur le contrat après prélèvements des frais de gestion et avant coûts décès.

#### **1.4.2.2.2.4 Cas d'exonération**

Les assurés non-résidents ont droit à l'exonération des contributions sociales sur encours.

### **1.4.2.2.3 Les prélèvements sociaux en cas de sinistre (rachat partiel, total ou dénouement du contrat)**

#### **1.4.2.2.3.1 Le calcul de l'assiette**

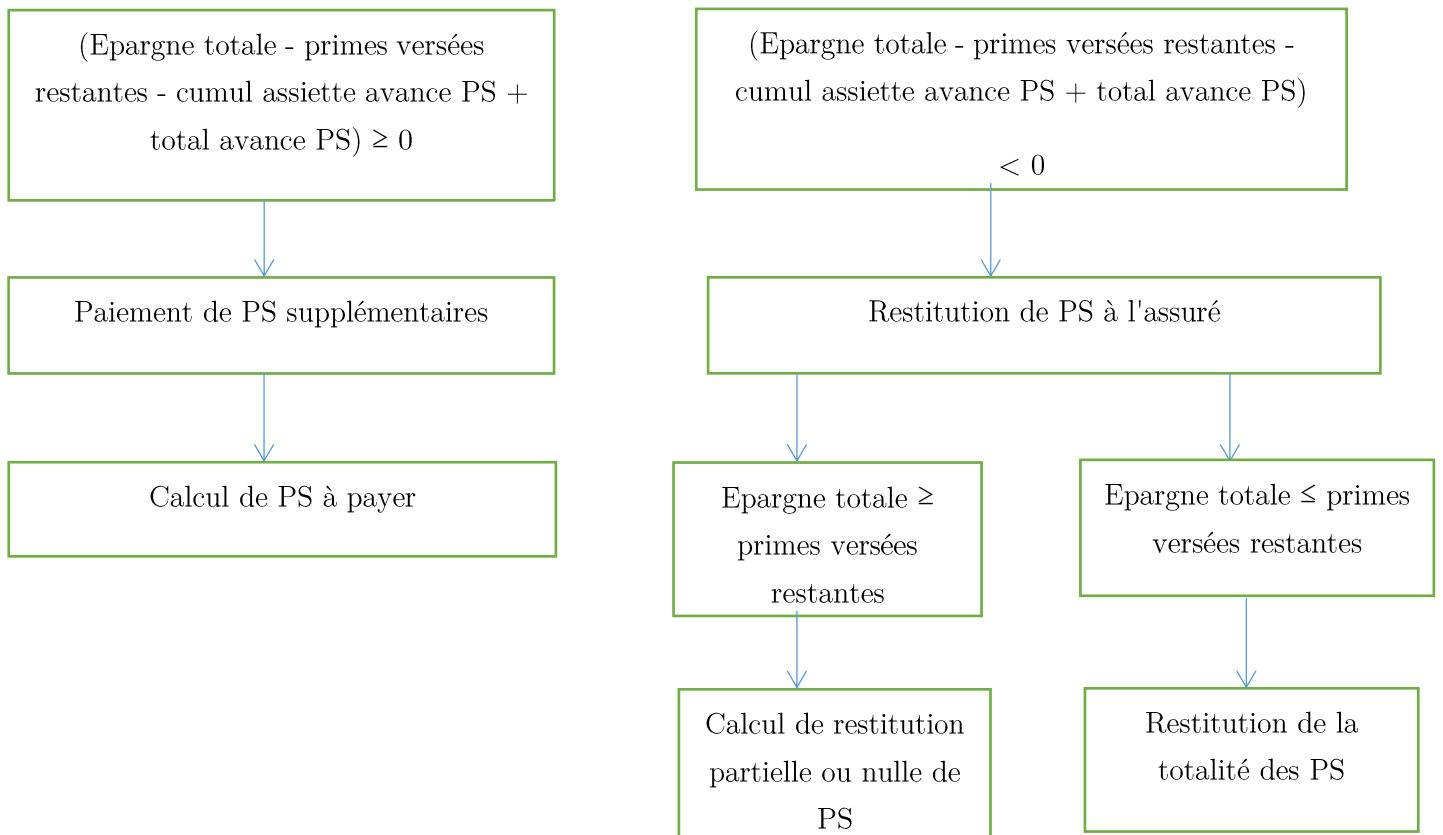
L'assiette des prélèvements sociaux est calculée en retranchant à l'épargne du contrat au jour du rachat, au décès de l'assuré ou à la liquidation totale du contrat, les primes versées restantes sociales et le cumul des assiettes d'avance de PS sur le compartiment euro, net de prélèvements sociaux avancés.

Si cette assiette est positive, l'assuré devra s'acquitter des prélèvements sociaux supplémentaires.

Si cette assiette est négative, 2 cas sont possibles :

- l'épargne totale du contrat est supérieure aux primes versées restantes : une restitution partielle ou nulle des prélèvements sociaux avancés sur l'euro sera effectuée ;
- l'épargne totale du contrat est inférieure aux primes versées restantes : une restitution totale des prélèvements sociaux avancés sur l'euro sera faite (en proportion du sinistre dans l'épargne contrat, en cas de rachat partiel ou de liquidation partielle)

## Tableau récapitulatif



### Remarque :

La restitution est assimilée à une restitution de trop payé non taxable à l'IR ni au 757 B et au 990 I. Le montant total de la restitution est versé à l'assuré le jour du règlement du sinistre.

La détermination du montant des prélèvements sociaux à payer ou à restituer en cas de dénouement (rachat total ou décès) et en cas de rachat partiel, passe par 3 étapes :

- détermination du cumul des assiettes Avance PS et Avance PS liés au sinistre
- calcul des PS totaux liés au sinistre
- calcul des PS à payer ou à restituer

Les assiettes de calcul lors d'un rachat partiel sont proportionnelles à celles déterminées en cas de dénouement total (rachat total ou décès). Le coefficient de proportionnalité est le Coeff sinistre défini par :

$$\text{Coeff sinistre} = \frac{\text{Montant du Sinistre}}{\text{Épargne acquise au moment du sinistre (avant le sinistre)}}$$

Par exemple en cas de rachat partiel, le montant du sinistre serait égal à la valeur de rachat partiel en cas de rachat partiel ou de liquidation partielle.

#### 1.4.2.2.3.2 Détermination des cumuls : Assiette d'Avance de PS et Avance de PS liés au sinistre

Pour chaque taxe  $i$  et taux  $t$ , on a :

##### **Cumul d'assiette d'avance de PS**

Cumul assiette avance  $PS_{i,t}$  (lié au sinistre) = Cumul assiette avance PS  $i,t$  (avant sinistre) \* Coeff sinistre

Cumul assiette avance  $PS_{i,t}$  (après sinistre) = Cumul assiette avance PS  $i,t$  (avant sinistre) - Cumul assiette avance  $PS_{i,t}$  (lié au sinistre)

Il faut noter qu'en cas de rachat total ou décès : Cumul assiette avance  $PS_{i,t}$  (après sinistre) = 0

Au final, on a : Cumul assiette avance PS (lié au sinistre) = MAX sur toutes les taxes  $i$  de  $\sum_t$  Cumul assiette avance  $PS_{i,t}$  (lié au sinistre)

Cumul assiette avance PS (après sinistre) = MAX sur toutes les taxes  $i$  de  $\sum_t$  Cumul assiette avance  $PS_{i,t}$  (après sinistre)

### **Avance de PS**

Total Avance PS<sub>i,t</sub> (lié au sinistre) = Total Avance PS<sub>i,t</sub> (avant sinistre) \* Coeff sinistre

Total Avance PS<sub>i,t</sub> (après sinistre) = Total Avance PS<sub>i,t</sub> (avant sinistre) - Total Avance PS<sub>i,t</sub> (lié au sinistre)

En cas de rachat total ou décès : Total Avance PS<sub>i,t</sub> (après sinistre) = 0

En somme on a : Total Avance PS (lié au sinistre) =  $\sum_i \sum_t$  Total Avance PS<sub>i,t</sub> (lié au sinistre)

Total Avance PS (après sinistre) =  $\sum_i \sum_t$  Total Avance PS<sub>i,t</sub> (après sinistre)

### • Calcul des PS totaux liés au sinistre

On détermine le montant de prélèvements sociaux calculés à partir de la valeur de l'épargne acquise en date de sinistre augmentée des prélèvements sociaux avancés.

Cette étape permet de déterminer le montant de prélèvements sociaux liés au sinistre qui aurait été dû s'il n'y avait pas eu de Prélèvements «au fil de l'eau».

En cas de Rachat total, partiel, décès de l'assuré ou liquidation en rente partielle ou totale :

Assiette PS totaux (liée au sinistre) = Montant sinistre + total avances PS (lié au sinistre)  
– Sommes versées sinistrées

Pour chaque taxe  $i$  et pour le dernier taux connu  $T_i$  on effectue le calcul de PS totaux :

PS( $i, T_i$ ) totaux (liés au sinistre) = Assiette PS totaux (liée au sinistre) \* Dernier taux connu de PS<sub>i</sub> en date de sinistre

Depuis le 26/09/2013, ce taux est fixé à 15,50%.

### • Calcul des PS à payer ou à restituer

#### Calcul de PS à payer

Si (Epargne totale - primes versées restantes - cumul assiette avance PS + total avance PS)  $\geq 0$  ; tous les montants étant bien entendu ceux avant sinistre alors il faut déterminer le montant de prélèvements supplémentaires à payer. En cas de rachat total,

partiel, décès de l'assuré ou liquidation en rente partielle ou totale :

Assiette PS à payer (liée au sinistre) = Assiette PS totaux (liée au sinistre) – Cumul assiette Avance PS (lié au sinistre)

Pour chaque taxe i et pour le dernier taux connu  $T_i$ , on a :

PS(i, $T_i$ ) à payer (liés au sinistre) = Assiette PS à payer (liée au sinistre) \* dernier taux connu du PSi en date de sinistre

Ainsi le montant de PS à payer sera égal à la somme des PS à payer par taxe :

$$\text{PS à payer (liés au sinistre)} = \sum_i PS_{i,T_i} \text{ à payer (liés au sinistre)}$$

### Calcul d'une restitution partielle ou nulle de PS

Si (**Epargne totale - primes versées restantes - cumul assiette avance PS + total avance PS**) < 0 et le contrat est globalement en gain, c'est-à-dire **Epargne Totale – Primes versées restantes** > 0 alors il y aura restitution partielle ou nulle de PS. Il faut déterminer un montant de prélèvements à restituer à l'assuré.

Ce montant peut être nul lorsque le montant de PS recalculé est supérieur à l'avance de PS, de ce fait il n'y aura pas de PS supplémentaires à verser.

En cas de rachat total, partiel, décès de l'assuré ou liquidation en rente partielle ou totale : pour chaque taxe i et taux t, on aura :

Assiette de restitution<sub>i,t</sub>(liée au sinistre) = Cumul assiette avance  $PS_{i,t}$ (lié au sinistre) – assiette PS totaux(liée au sinistre)

$$PS_{i,t} \text{ à restituer (liés au sinistre)} = \text{avance } PS_{i,t} \text{ (liée au sinistre)} - PS_{i,t} \text{ totaux (liés au sinistre)}$$

Le montant qui sera alors restitué à l'assuré est égal à :

$$\text{Restitution (liée au sinistre)} = \text{MAX} \left( \sum_{i,t} PS_{i,t} \text{ à restituer (liés au sinistre)} ; 0 \right)$$

### Restitution Totale

Si (**Epargne totale - primes versées restantes - cumul assiette avance PS + total avance PS**)  $< 0$  et le contrat est globalement en perte c'est-à-dire **Epargne Totale - Primes versées restantes**  $< 0$  alors il y aura restitution totale de PS à l'assuré. Pour déterminer ce montant, on a pour chaque taxe  $i$  et taux  $t$  :

Assiette de restitution $_{i,t}$  (liée au sinistre) = cumul assiette avance  $PS_{i,t}$  (lié au sinistre)

et  $PS_{i,t}$  à restituer = avance  $PS_{i,t}$  (lié au sinistre)

---

# ETUDE DU PORTEFEUILLE

---

L'objectif de cette deuxième partie est de réaliser les statistiques descriptives de notre portefeuille de rachats afin d'avoir une première intuition sur les facteurs de risque de rachat significatifs et de pouvoir aussi retenir les variables pertinentes pour la suite dans le cadre de la modélisation.

## 2.1 Présentation des données

La modélisation des comportements de rachat étant plus intéressante pour les produits d'épargne plutôt que pour les produits de prévoyance (essentiel du marché en collective), nous avons fait le choix de ne considérer que les produits d'épargne de notre portefeuille. Pour mener à bien les études de cette deuxième partie du mémoire, nous avons extrait un échantillon de notre base de rachats parce que le système informatique ne peut pas nous fournir toutes les informations pour certains rachats qui avaient été effectués dans le passé. Par mesure de précaution, nous utilisons pour cette partie un portefeuille composé uniquement des rachats survenus pour les contrats en Euros (8074), qui est géré de façon manuelle par le service Gestion de la compagnie. D'autre part, nous avons extrait du système informatique les rachats pour les contrats UC (6642).

### 2.1.1 Description des données

Pour chacune de nos bases de rachats (Euros ou UC), nous avons entre autres les informations classiques suivantes : N° Police, les Nom, Prénom, Date de Naissance et Code Postal de l'assuré, le Code Produit ainsi que le Nom du produit, la Date de Rachat, le type du rachat, le Montant Brut et le Montant Net du Rachat, les montants des Prélèvements Sociaux et des Prélèvements Libératoires, le Montant des Provisions Mathématiques (PM) avant rachat pour les contrats Euros et l'Épargne avant et après le rachat pour les contrats UC. De plus, pour les rachats Euros, nous avons aussi le montant des pénalités, les avances et quelques remarques sur certains rachats. Pour les rachats UC, nous avons le montant des plus-values lors du rachat, la fiscalité choisie et le montant de l'IFU (Imprimé fiscal Unique).

### 2.1.2 Retraitement des bases

Avant d'entamer l'étude, il a fallu au préalable retraiter les bases de rachats afin de ne disposer que de variables qui seront par la suite utiles dans la modélisation des comportements de rachat. Les retraitements suivants ont été effectués :

- Suppression des contrats avec des données clés manquantes et ceux qui sont souscrits par des sociétés dans la base Euros
- Analyse et retraitement des données aberrantes
- Création des variables Age de l'assuré et Ancienneté du contrat lors du rachat
- Traitement spécifique pour les rachats partiels programmés
- Répartition par classes des départements de résidence des assurés selon le critère du revenu moyen.

Finalement, nous nous retrouvons avec des bases restreintes : Rachats Euros avec 4684 données et Rachats UC (6642) et une liste de variables suivantes :

Rachats Euros	Rachats UC
Sexe assuré	NOM PRODUIT
Code postal	SEXE assuré
Code Département	Code postal
Classement Département	Code Département
Date Naissance Assuré	Classement Département
Age lors du rachat	Naissance
Ancienneté lors du rachat	Age lors du rachat
Date prise effet du contrat	Date effet
Nom produit	DATE RACHAT
Périodicité prime	Indicateur Rachat Programmé
Prélèvements sociaux	Ancienneté lors du rachat
Prélèvements libératoires	TYPE RACHAT
Montant brut du rachat	Périodicité prime
Montant Rachat Net	MONTANT BRUT
Montant Avance	MONTANT FRAIS
Motif	MONTANT NET
Remarques	MONTANT PLUS VALUE
Type de rachat	FISCALITÉ
Date de saisie du rachat	MONTANT PS
Pénalités	MONTANT PL
PM avant Rachat	MONTANT IFU
PM après Rachat	EPARGNE AVANT
Année de traitement	EPARGNE APRES
	Année de traitement

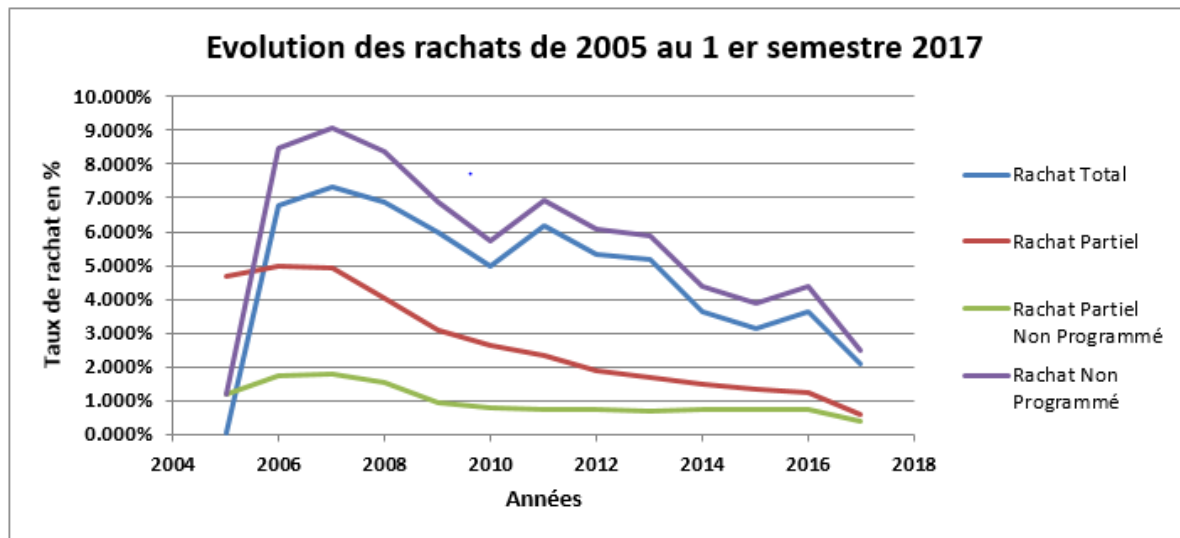
## 2.2 Statistiques Descriptives

Après avoir réalisé différents traitements sur nos bases, nous nous attelons à l'analyse des statistiques descriptives du portefeuille. Nous commencerons d'abord par l'étude des rachats et de leur évolution avant de présenter les statistiques de nos variables.

### 2.2.1 Les rachats entre 2005 et le 1er semestre 2017

Afin d'analyser l'évolution des rachats entre 2005 et le 1er semestre 2017, des calculs ont été effectués sur les différentes bases. Le choix de traiter séparément les contrats Euros et UC résulte du fait qu'ils aient des caractéristiques différentes. En effet comme le montre le tableau suivant, les rachats partiels sont beaucoup plus nombreux pour les contrats UC que pour les contrats Euros ; la grande majorité de ces rachats partiels UC sont des rachats programmés, résultante de l'option de rachat programmé intégrée dans la plupart des contrats UC. Dans un contexte d'étude des comportements de rachat dans ce mémoire, nous n'allons pas nous attarder sur les rachats programmés car ils reflètent des comportements prévisibles pour l'assureur. C'est pourquoi nous avons séparé les différents cas pour l'analyse de l'évolution des rachats.

Années	Nombre de contrats	Rachat Total				Rachat Partiel							Rachat Non Programmé	
		Euros	UC	Total Rachat Total	Rachat Total	Euros	UC	UC Non Programmé	Total Rachat Partiel	Total Rachat Partiel Non programmé	Rachat Partiel	Rachat Partiel Non Programmé	Total des Rachats Non Programmés	Rachat Non Programmé
2005	8367	2	0	2	0,024%	21	371	78	392	99	4,69%	1,183%	101	1,21%
2006	9133	479	138	617	6,76%	41	414	116	455	157	4,98%	1,719%	774	8,47%
2007	10118	579	160	739	7,30%	42	458	137	500	179	4,94%	1,769%	918	9,07%
2008	10727	495	240	735	6,85%	35	397	128	432	163	4,03%	1,520%	898	8,37%
2009	11066	367	293	660	5,96%	33	308	68	341	101	3,08%	0,913%	761	6,88%
2010	11334	298	264	562	4,96%	32	268	57	300	89	2,65%	0,785%	651	5,74%
2011	11568	358	354	712	6,15%	44	225	43	269	87	2,33%	0,752%	799	6,91%
2012	11911	343	293	636	5,34%	48	179	41	227	89	1,91%	0,747%	725	6,09%
2013	12430	406	238	644	5,18%	48	163	40	211	88	1,70%	0,708%	732	5,89%
2014	13565	288	203	491	3,62%	46	154	54	200	100	1,47%	0,737%	591	4,36%
2015	16341	321	188	509	3,11%	48	171	76	219	124	1,34%	0,759%	633	3,87%
2016	18984	196	493	689	3,63%	37	195	103	232	140	1,22%	0,737%	829	4,37%
2017	20403	62	365	427	2,09%	15	110	64	125	79	0,61%	0,387%	506	2,48%



Comme le montre la figure ci-dessus, nous avons illustré l'évolution des rachats au cas par cas : tout d'abord l'évolution des Rachats Totaux (RT), puis les Rachats Partiels (RP), les Rachats Partiels Non Programmés (RPNP) et enfin les Rachats Non Programmés (RNP). Ces derniers (RNP) représentent l'ensemble des rachats partiels et totaux qui ne sont pas programmés.

Pour les rachats totaux, le faible taux en 2005 pourrait faire penser à un portefeuille immature mais le taux de rachat partiel dans la même année vient contrebalancer cette idée. Après 2005, nous assistons à une hausse fulgurante du taux de rachat total qui atteint son apogée en 2007. Ce qui s'explique par le contexte de déséquilibre financier qui régnait dans le monde et qui s'est soldé par la crise des *subprimes* en 2007. Globalement l'évolution du taux de rachat total est rythmée par les périodes de crise avec des pics marqués en 2007 et en 2011 (crise de la dette grecque). Nous pouvons donc constater qu'une partie de nos rachats totaux sont des rachats conjoncturels. Pour les rachats non programmés, nous notons une évolution similaire à celle des rachats totaux à la différence près que les taux sont plus élevés dans ce deuxième cas. Ce qui se justifie amplement par la faible proportion de rachats partiels non programmés par rapport aux rachats totaux.

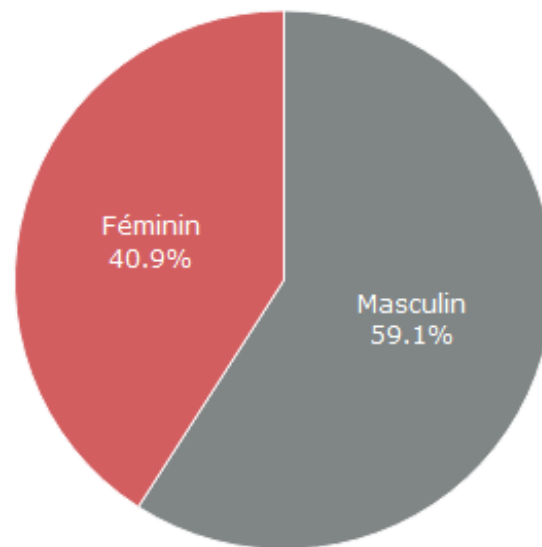
Pour les rachats partiels, en tenant compte ou non des rachats programmés, la tendance est plutôt à la baisse après le pic de 2007. Pour les rachats partiels non programmés, nous observons même une certaine stabilisation depuis 2010. Sans les rachats programmés, nous pouvons dire que chez Optimum Vie les taux de rachat partiel ne sont pas élevés (entre 0.70% et 1.5%).

## 2.2.2 Statistiques descriptives sur les variables retenues

### 2.2.2.1 Analyse exploratoire des variables qualitatives

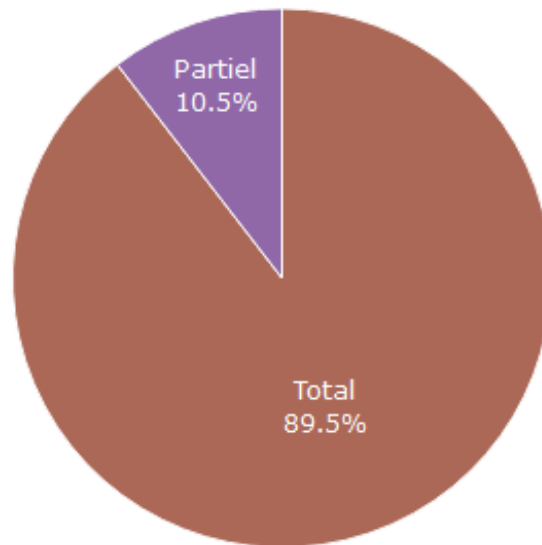
Pour les contrats Euros, la répartition par sexe nous révèle que les femmes rachètent moins que les hommes. Elles sont plus prudentes.

Répartition par Sexe des rachats Euros



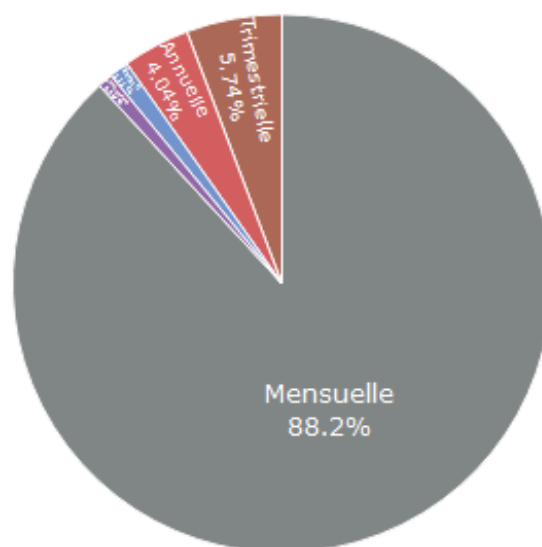
Comme nous avons pu le remarquer auparavant, les rachats totaux prédominent les rachats partiels pour les contrats Euros. Le rendement des contrats Euros étant garanti mais faible, les assurés ne rachètent que s'ils ont vraiment des besoins de liquidité urgents.

### Type des rachats Euros



En ce qui concerne la périodicité de la prime, la plupart des contrats Euros rachetés sont à primes périodiques versées mensuellement. Ce qui peut être assez révélateur. En effet, dès lors que l'assuré verse régulièrement ses primes (mensuellement ou trimestriellement), il est plus à même de connaître l'évolution de son épargne acquise et pourra ainsi exercer son option de rachat à tout moment pour n'importe quel motif.

### Périodicité de la Prime des rachats Euros



Après avoir classé les départements de résidence des assurés selon le critère du revenu moyen déclaré<sup>8</sup>, nous avons par la suite réparti les rachats Euros selon le classement réalisé.

Classes
Tranches de revenu
Moyenne des revenus déclarés
Départements classés selon le montant moyen en euros des revenus déclarés par foyer à l'administration fiscale en 2015
1
[41000 ; 27500]
32749
75, 92, 78, 74, 91, 94, 01, 68, 69, 77
2
]27500 ; 24000]
25556
95, 31, 67, 38, 06, 44, 73, 51, 25, 21, 35, 45, 60, 28, 33, 37, 64, 83, 90, 13, 27, 63, 56, 57, 54, 14, 41, 29, 39, 49, 76, 85
3
]24000 ; 22000]
23035
17, 53, 26, 22, 72, 40, 10, 71, 2A, 86, 42, 79, 34, 50, 05, 89, 59, 87, 18, 80, 19, 84, 04, 16, 07,30, 12, 46, 81, 55, 70, 43, 32, 88
4
]22000 ; 21000]
21505
61, 02, 52, 36, 65, 47, 82, 03, 08 , 58, 24, 2B, 62, 15, 48, 93, 66
5
< 21000
17385
11, 09, 23, 972, 974, 973, 971, 976
999
-
-
Autres : étrangers et départements ne faisant pas partie du classement

8. Source : [www.journaldunet.com/economie/impots/classement/departements/revenu-fiscal](http://www.journaldunet.com/economie/impots/classement/departements/revenu-fiscal)

Suivant ce classement, nous avons obtenu la répartition suivante :

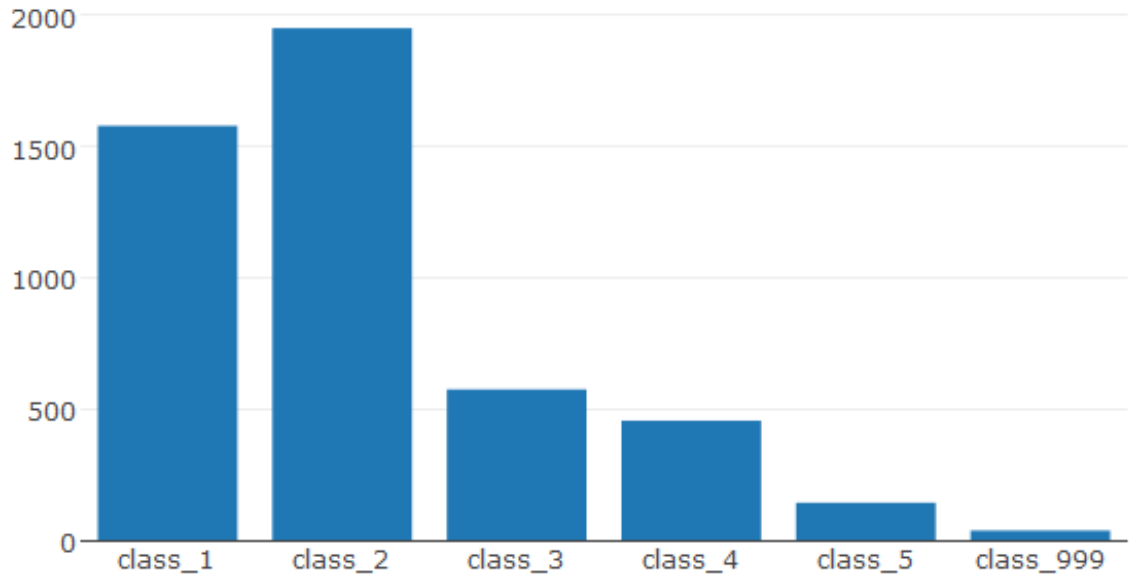
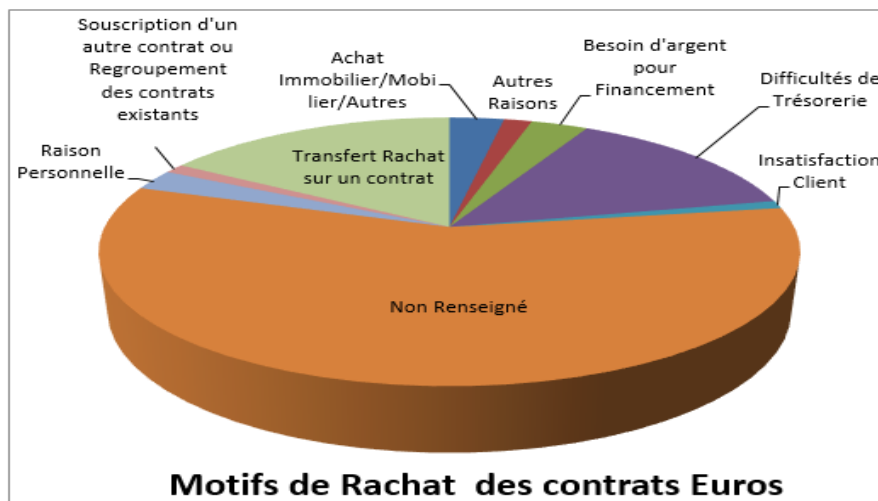


FIGURE 2.1 – Répartition par département de résidence de l'assuré des Rachats Euros

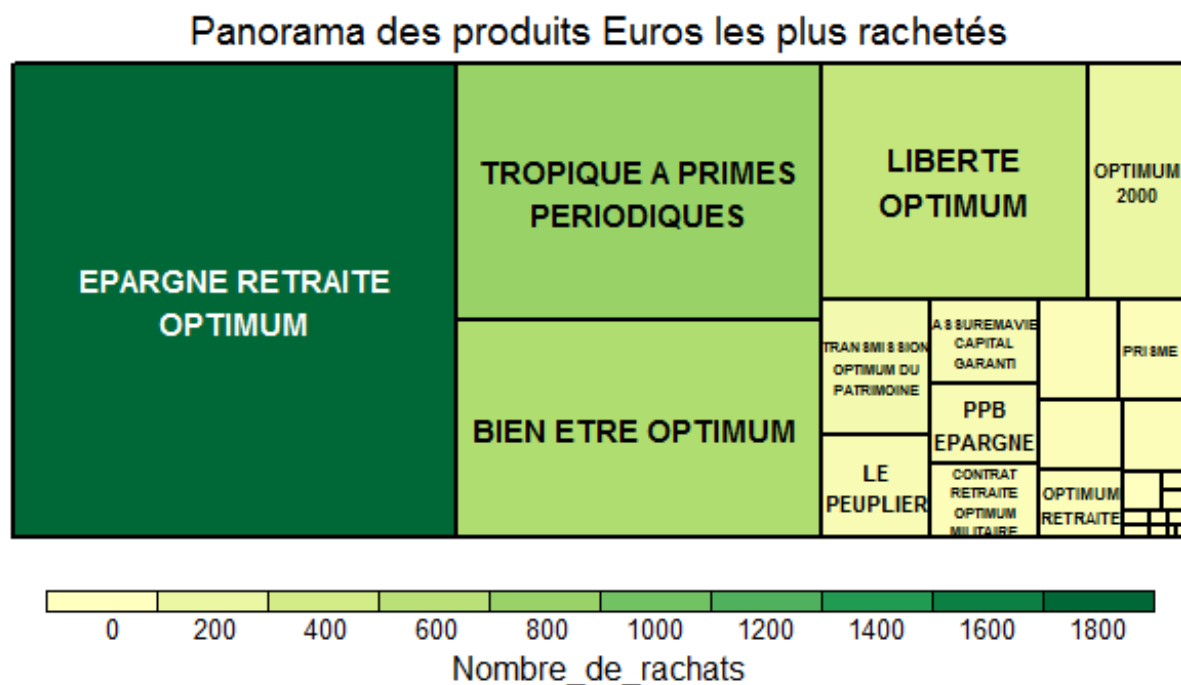
Ce graphe nous révèle que ce sont plutôt les assurés aux revenus élevés qui rachètent. Ce qui fait jaillir une première idée selon laquelle le comportement de rachat des assurés est fortement corrélé à leur patrimoine. Notons juste que la classe 999 représente les départements étrangers et les départements des îles mineures de la France qui n'ont pas été classés selon leur revenu moyen. Leur nombre n'étant pas conséquent dans notre portefeuille pour de multiples raisons, on pourrait comprendre qu'il en va de même pour leurs rachats.

Disposant de quelques motifs de rachat pour les contrats Euros, nous avons construit le graphe suivant :



Ces informations nous permettent de valider l'hypothèse de fonds d'urgence comme le motif principal du comportement de rachat des assurés. En réalité selon cette hypothèse, l'assuré rachète son contrat lorsqu'il est confronté à un choc négatif non anticipé. C'est le cas notamment ici lorsqu'il a des difficultés de trésorerie ou qu'il a besoin d'argent urgemment pour financer des projets à court terme. Il faut toutefois noter que chez Optimum Vie, il y a une part non-négligeable des rachats qui est dédiée au financement d'autres contrats par le biais du transfert. Nous pouvons supposer que ce comportement est peut-être lié à une recherche d'un meilleur rendement ou à des effets marketing.

Par ailleurs, nous avons exploré les différents produits Euros qui sont rachetés.

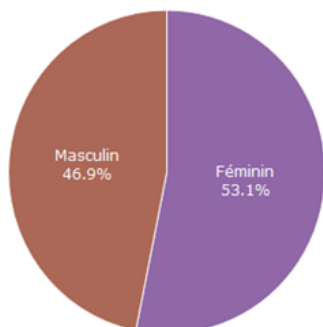


Au départ, nous avons 24 types de produits Euros dans notre base de rachats dont certains avec un faible effectif. C'est pourquoi nous avons privilégié ce type de graphe afin qu'on ne puisse observer que les produits les plus rachetés. Deux grands produits phares sont distingués :

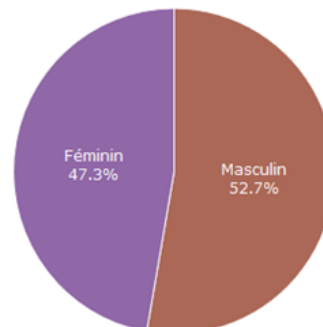
- Epargne Retraite Optimum : c'est un contrat d'épargne en Euros à versements périodiques, rachetable à tout moment. Il permet la constitution d'un complément de retraite par capitalisation. Sa durée de détention ne peut être inférieure à 15 ans, ce qui peut impliquer une vague massive de rachats au-delà de cette limite. A contrario, cette limite peut être aussi source de rachats totaux anticipés puisqu'elle peut s'avérer très élevée pour certains qui ont des besoins urgents. Il faut noter aussi que c'est un contrat très ancien qui bénéficiait d'un taux minimum garanti assez élevé à l'époque de sa commercialisation.
- Tropiques à primes périodiques : est un contrat d'épargne et de retraite. Il peut être racheté à tout moment et sans pénalités, ce qui peut justifier largement le nombre important de rachats. Son taux d'intérêt technique avait été fixé à plus de 3%.

Pour les rachats UC, nous avons adopté une démarche comparative entre la base des rachats UC (initiale) et la base des rachats UC non programmés (base initiale sans les rachats programmés).

Répartition par sexe des rachats UC

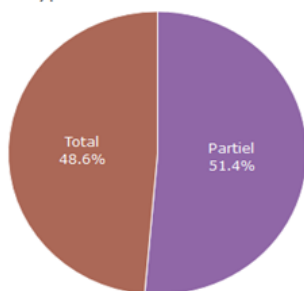


Répartition par sexe des rachats UC non programmés

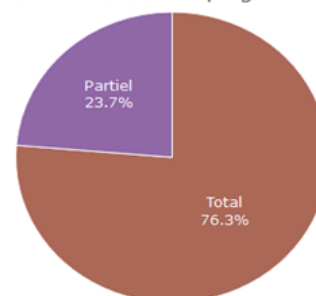


Il découle de l'analyse de ces graphes que les femmes sont «risquobobes» c'est-à-dire qu'elles sont assez prudentes en ce qui concerne les rachats. En effet, elles préfèrent les rachats partiels programmés en accord avec l'assureur et ne pas prendre beaucoup de risques. La connaissance de leur situation matrimoniale aurait pu nous édifier sur ce choix prudent qui peut-être lié à des besoins ponctuels comme par exemple une célibataire ou une divorcée avec des enfants à charge.

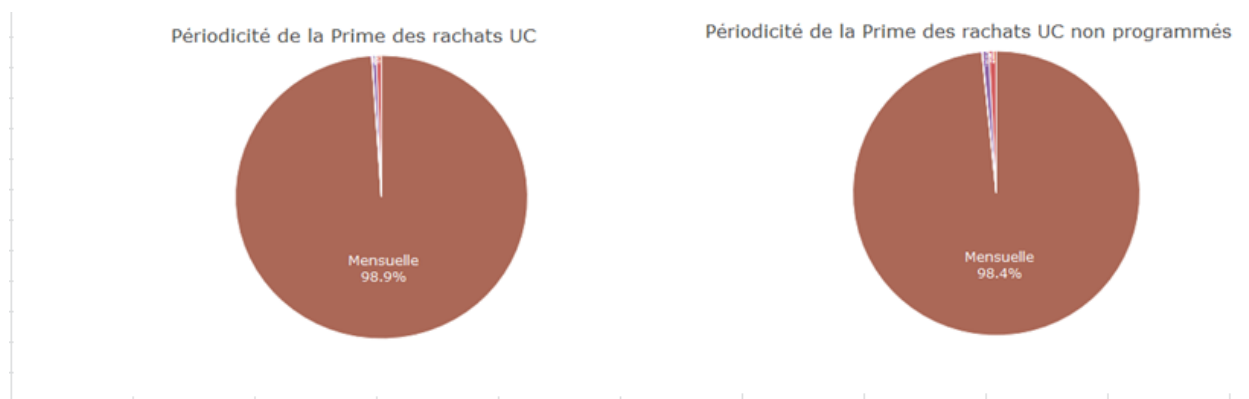
Type des rachats UC



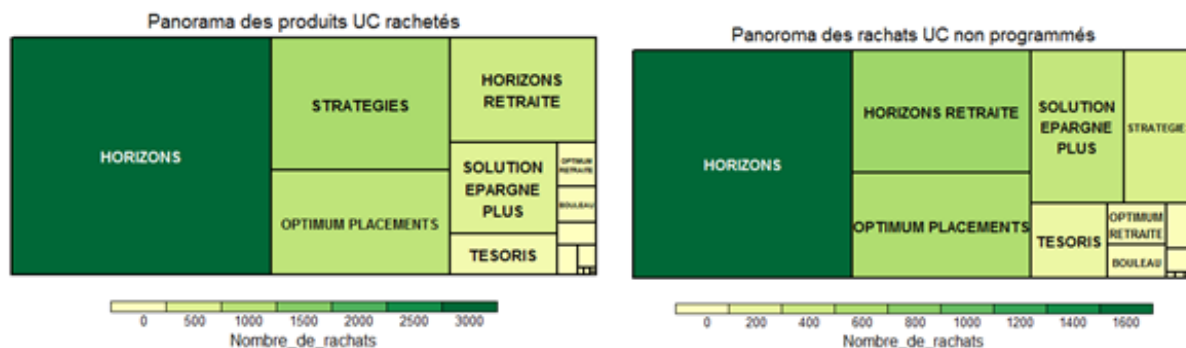
Type des rachats UC non programmés



Pour le type du rachat, nous remarquons que plus de 2/3 des rachats UC sont des rachats totaux, ce qui confirme la tendance chez Optimum Vie. Les rachats partiels programmés déséquilibrent le portefeuille des rachats UC et peuvent biaiser dans une moindre mesure l'analyse. Comme pour les rachats Euros, il y a très peu de rachats partiels non anticipables.



La périodicité de la prime ne nous fournit pas de plus amples informations. La périodicité mensuelle prédomine. Nous comprenons que dès lors qu'ils versent mensuellement des primes, ils disposent d'un meilleur suivi de l'évolution de leur encours et qu'ils peuvent racheter au gré de l'augmentation et de la diminution des plus-values ou de leurs besoins.



Pour les produits, il y a un fort impact selon que l'option de rachat programmé soit permise ou pas. Globalement, on a 5 grands produits UC phares qui sont rachetés :

- Horizons est un contrat multi-supports, multi-profils, libellé en unités de compte
- Stratégies est aussi un contrat multi-supports, multi-profils, libellé en UC. Le rachat total ou partiel est possible à tout moment mais il y a un montant minimum du rachat partiel et un solde minimum de l'épargne constituée après rachat partiel. Ce qui explique la préférence des assurés pour les rachats programmés d'où ce produit englobe plus du quart de ces derniers.
- Optimum Placements est un contrat multi-support en UC à versements programmés et/ou libres avec une garantie complémentaire décès obligatoire. Entre 2012 et 2013, la

compagnie a adopté en accord avec les courtiers une politique proposant aux anciens contrats un transfert sur Optimum Placements, le but étant de garder les assurés en portefeuille plus longtemps. Les principaux contrats qui étaient concernés par ce transfert sont : Tropiques à primes périodiques, Liberté Optimum, Bien être Optimum et Contrat Retraite Optimum Militaire. Il faut noter que ces contrats sont tous des contrats Euros. De ce fait, on pourrait penser que les nombreux rachats du produit (OPL) sont peut-être liés au fait que les assurés ne sont pas satisfaits de la rentabilité du produit ou ne sont pas habitués aux fluctuations de leur épargne sachant que leurs anciens contrats Euros leur proposaient des taux minimums garantis assez élevés.

- Horizons Retraite est un contrat d'assurance vie de groupe à capital variable. Il permet à l'adhérent assuré de constituer une épargne retraite à durée fixe, au moyen de versements programmés et éventuellement des versements exceptionnels sur des supports en UC correspondant à des OPCVM. La spécificité de ce contrat est que la faculté de rachat est exceptionnelle. En réalité, le but de l'adhésion est exclusivement le versement des prestations liées à la cessation d'activité professionnelle et donc ne comporte pas de valeur de rachat (article L 132-23 du Code des Assurances). Toutefois, le droit de rachat demeure en cas de cessation d'activité de l'adhérent à la suite d'un jugement de liquidation judiciaire ou des invalidités de 2ème ou 3ème catégorie prévues à l'article L 341-4 du Code de la Sécurité Sociale.
- Solution Epargne Plus est un contrat d'assurance-vie de groupe à adhésion individuelle et facultative. Il cible tout prospect ayant un objectif d'épargne à long terme en vue de la constitution d'un capital. Le montant minimum d'un rachat partiel s'élève à 750€.

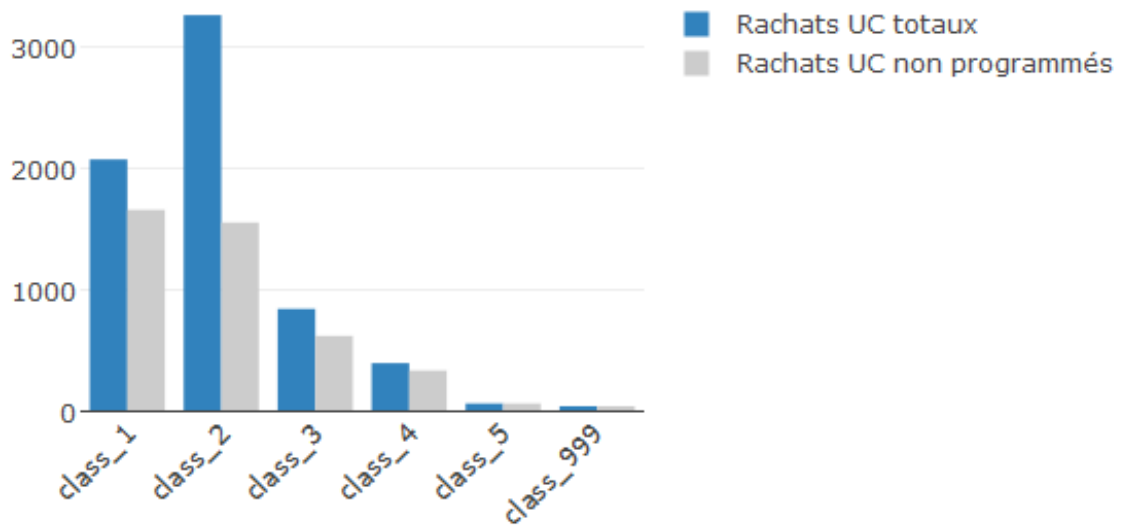


FIGURE 2.2 – Répartition des Rachats UC et des Rachats UC non programmés par département de résidence de l'assuré

Nous pouvons déduire de ce graphe que les assurés des classes 5 et 999 (respectivement les assurés en provenance des départements aux revenus faibles et des départements étrangers) ne rachètent pas beaucoup. Ceux du classe 2 sont quant à eux plus prompts à effectuer des rachats programmés ; cette classe comporterait sûrement une grande proportion de population féminine. Cette répartition valide encore l'hypothèse de corrélation du comportement de rachat avec le niveau de richesse détenu.

### 2.2.2.2 Etude des variables quantitatives

Variabes / Rachats Euros	Médiane	Moyenne	Ecart-type	Minimum	Maximum
Age lors du rachat	45	46,08	10,78	15	92
Ancienneté lors du rachat	10	10,71	6,89	0	44
Montant brut du rachat	4638	7335	10780,44	1	320328
PM avant rachat	5330	8893	18468,08	1	420430

Les écarts entre la moyenne et la médiane d'une part et l'écart type et la moyenne d'autre part des Montants Bruts de rachat et des Provisions Mathématiques juste avant le rachat,

prouvent que le portefeuille des rachats Euros n'est pas homogène et qu'il pourrait même être délimité en classes, ce dont attestent les sommes élevées des montants maximum rachetés par rapport à la moyenne. Les faibles montants rachetés correspondent à des rachats d'office suite au non-paiement de la prime ou à une annulation de souscription, le tout avec une faible provision mathématique. Ces rachats représentaient 1.58% de la base des rachats Euros et ils ont été considérés comme des rachats totaux lors du traitement. Par ailleurs, les variables Age et Ancienneté lors du rachat sont assez bien réparties dans le portefeuille. Il faut noter aussi qu'il y a des assurés qui détiennent leur contrat sur le long terme.

Variables/ Rachats UC	Médiane	Moyenne	Ecart-type	Minimum	Maximum
Age lors du rachat	55	56,05	17,06	5	105
Ancienneté lors du rachat	4	4,39	3,42	0	16
Montant brut du rachat	2500	10716,76	35344,44	2,6	1136085
Epargne avant	14407,1	57100,4	134390,8	2,6	1606634

Pour les rachats UC, le même phénomène qu'avec les rachats Euros est observé. Le portefeuille est inégalement réparti avec de très fortes disparités sur les encours, ce qui peut être la conséquence des fluctuations dont peut faire l'objet ce type de contrat. Les variables Age et Ancienneté lors du rachat restent bien réparties dans l'ensemble. Toutefois, la plupart des rachats surviennent avant la levée des contraintes fiscales à 8 ans, ce qui assez révélateur sur la corrélation des comportements de rachat UC par rapport à la fiscalité. Comparés aux contrats en Euros, ceux en UC ne sont pas détenus très longtemps par les assurés, l'ancienneté maximale étant de 16 ans contre 44 ans. Ce qu'on pourrait justifier par les fluctuations de l'épargne qui pousseraient certains assurés averses au risque à racheter.

Par la suite, les taux de rachats ont été représentés en fonction de l'âge et de l'ancienneté.

Pour les rachats Euros, nous observons deux phases qui se matérialisent par une hausse des taux à partir de 28 ans jusqu'à l'âge 47 et puis par une baisse pour les âges suivants. Les assurés ont un fort besoin de liquidité sur ce premier intervalle de temps pour financer l'achat ou la construction de leur résidence ou la concrétisation d'autres projets de jeunesse. Après 47 ans, les taux décroissent et se stabilisent après 81 ans. Les assurés âgés rachètent

moins que les plus jeunes.

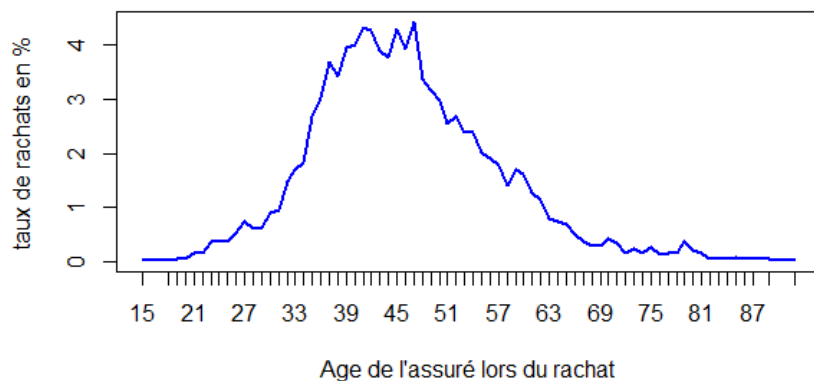


FIGURE 2.3 – Taux de rachat en fonction de l'Age de l'assuré lors du rachat

Par ailleurs, le succès de l'assurance-vie lui venant en grande partie de sa fiscalité avantageuse, l'ancienneté est une variable très importante dans la modélisation des rachats. Nous observons un premier pic à 2 ans qui représente les rachats précoces mais aussi l'impact des rachats partiels. Sans surprise, nous avons un deuxième pic à 10 ans qui matérialise le saut fiscal suite à la levée des contraintes après 8 ans d'ancienneté. Au-delà de 20 ans d'ancienneté, la tendance est plutôt à la baisse et se stabilise après 30 ans.

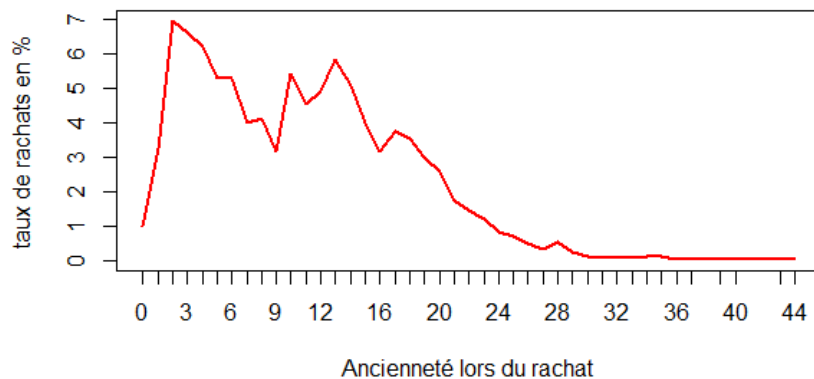


FIGURE 2.4 – Taux de rachat en fonction de l'Ancienneté lors du rachat

Pour les rachats UC, l'impact des rachats programmés rend les taux de rachat assez volatiles en fonction de l'âge surtout pour les âges élevés. Jusqu'à l'âge 45, la tendance est plutôt à la hausse pour faire face après à de multiples variations.

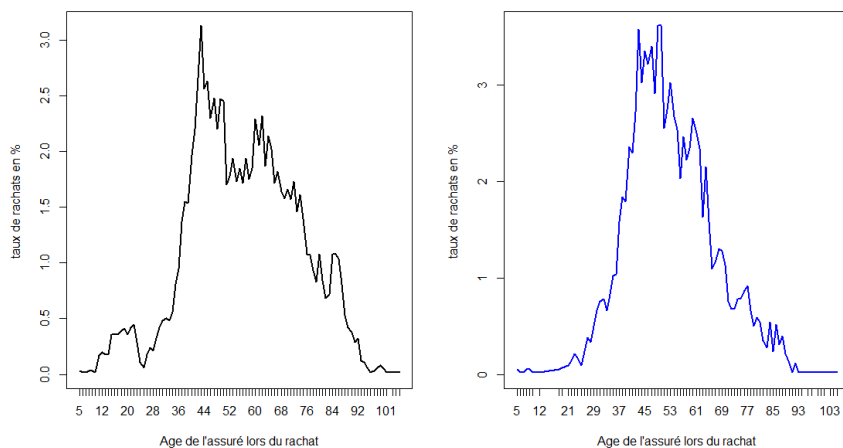


FIGURE 2.5 – Taux de rachat en fonction de l'Age de l'assuré pour la base Rachats UC (gauche) et les UC non programmés (droite)

Pour ce qui est l'ancienneté, aussi invraisemblable que cela puisse paraître, le pic intervient au bout d' 1 an d'ancienneté. De plus les taux ont tendance à décroître en fonction de l'ancienneté avec deux légers sursauts à 8 ans et 10 ans.

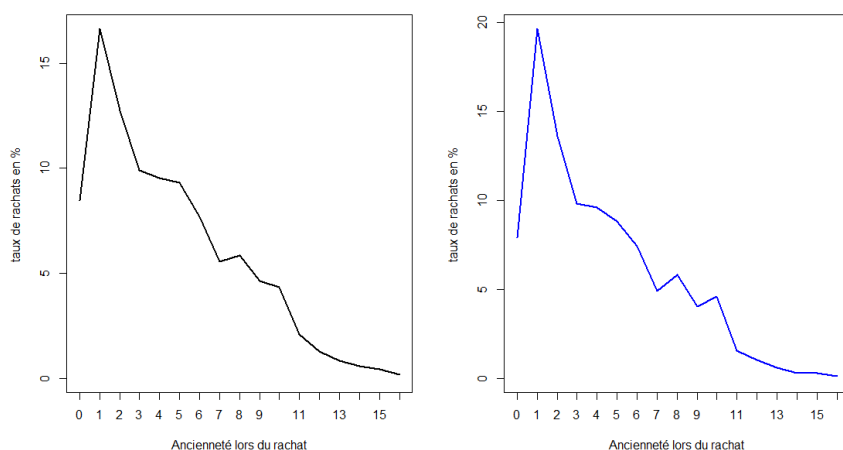


FIGURE 2.6 – Taux de rachat en fonction de l'ancienneté pour les UC (gauche) et les UC non programmés (droite)

Ce pic au bout d'1 d'ancienneté suscitant beaucoup d'interrogations sur les possibles causes de ces rachats massifs, le graphe de l'évolution des plus-values en fonction de l'ancienneté a été réalisé. Ce dernier montre que les rachats massifs à cette période sont dus à des moins-values très conséquentes qui poussent les assurés averses au risque à racheter leur contrat. On pourrait aussi penser que le léger saut fiscal observé précédemment est lié au pic culminant des plus-values à 10 ans. En effet, ce montant élevé des plus-values a peut-être amorti les rachats massifs auxquels on aurait pu s'attendre après 8 ans d'ancienneté. L'hypothèse de recherche d'un meilleur rendement pourrait prévaloir ici comme justificatif des comportements de rachat des produits UC.

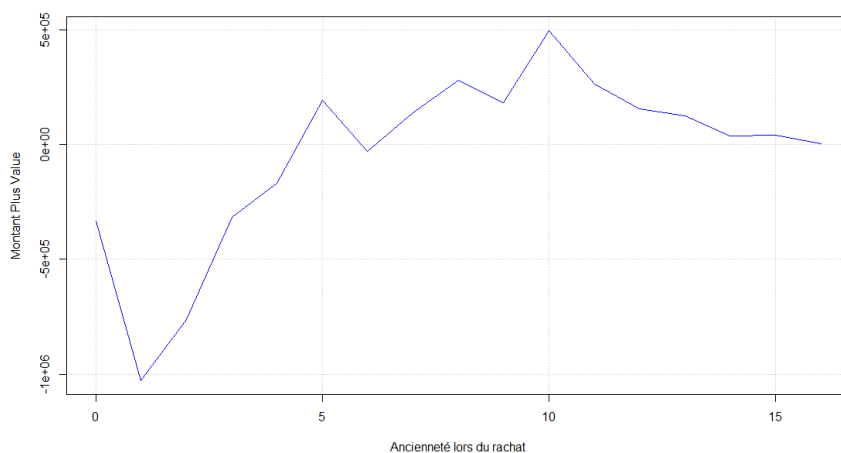


FIGURE 2.7 – Evolution du montant des plus-values en fonction de l'Ancienneté lors du rachat

D'après la Direction Technique, ces rachats massifs peuvent s'expliquer aussi par une incompréhension du fonctionnement des contrats par les assurés. En réalité, les primes investies en première année de souscription servant en général à financer le coût de la garantie plancher décès incluse dans la plupart des contrats UC, les assurés souvent, ne comprennent pas leur faible épargne acquise lorsqu'ils reçoivent le bilan de l'évolution de leur contrat à la fin de la première année.

Tout au long de l'analyse des statistiques descriptives, nous avons pu distinguer des facteurs de risque de rachat d'origine structurelle comme conjoncturelle. En effet, lorsque la conjoncture économique se dégrade, les assurés ont tendance à racheter leur contrat, phénomène très observé pour les contrats UC où les taux ne sont pas garantis et que leur rendement est faible en cas de crise financière. Pour les contrats Euros où les taux sont garantis, nous avons pu observer des sorties anticipées qui peuvent être liées à une recherche d'un meilleur TMG proposé par la concurrence. Par ailleurs, le rachat structurel correspond aussi à un besoin de liquidité immédiat de l'assuré. C'est le cas lorsqu'il a des difficultés de trésorerie ou qu'il veut financer un projet comme l'achat d'un immobilier. Ces comportements sont indépendants des conditions de marché.

Ces études statistiques nous ont permis d'identifier certains facteurs de risque susceptibles de fortement influencer le comportement de rachat des assurés. Ces facteurs seront par la suite considérés dans le cadre de la modélisation, ce qui permettra de confirmer ou d'infirmer nos intuitions.

---

# MODÉLISATION DES COMPORTEMENTS DE RACHAT

---

Comme il a été mentionné au 1.3), le risque de rachat représente un grand enjeu pour les assureurs-vie d'où la nécessité d'une modélisation précise des comportements de rachat des assurés afin d'anticiper, de prévoir et de maîtriser ce risque. Le but de cette partie est donc de proposer des méthodes pour la modélisation comportementale avec les informations dont nous disposons sur notre portefeuille d'épargne. En réalité, intuitivement la décision de racheter ou non son contrat par un assuré dépend de plusieurs facteurs : à titre illustratif, soit il a besoin à un instant  $t$  de trésorerie pour financer un projet et il dispose d'assez d'encours sur son contrat d'assurance pour ce faire, soit la conjoncture économique se dégrade et il rachète son contrat motivé par la recherche d'un meilleur rendement. Modéliser ce comportement revient donc à l'expliquer par les facteurs de risque qui pourront influencer l'assuré. La modélisation comportementale apparaît donc comme étant un processus qui permet d'étudier le lien entre une variable à expliquer (Rachat ou non de son contrat au cours de l'année) et un ensemble d'autres variables qui seront explicatives. Un premier modèle intuitif qui semble adapté pour ce type de modélisation est le cas particulier de la régression logistique des GLMs (Modèles Linéaires Généralisés). Une autre approche par modèles mélange a été développée très récemment dans la littérature.

## 3.1 Le choix du modèle

### 3.1.1 Les données

La deuxième partie du mémoire nous a permis d'avoir des intuitions sur les facteurs de risque susceptibles d'expliquer les comportements de rachat. Nous avons ainsi reconstitué un nouveau portefeuille d'épargne avec de nouvelles variables. En effet, en se basant sur certains travaux qui ont été effectués sur le même sujet et sur l'analyse de la deuxième partie, nous avons retenu d'emblée des variables.

Nous disposons ainsi d'un portefeuille de 25 942 contrats d'épargne de 2005 à la fin du 1er semestre 2017 avec 12 variables. Parmi ces dernières, nous avons celles qui concernent l'assuré que sont :

- l'âge de souscription qui est aussi l'âge de tarification,
- le genre (Féminin ou Masculin),
- le classement de son département de résidence selon le critère qui a été évoqué à la partie 2) et
- la prime investie

Nous avons aussi des variables qui sont propres au contrat : entre autres

- le numéro d'identification,
- la date d'effet,
- l'ancienneté,
- la périodicité de la prime,
- le taux de PB,
- le taux servi
- la date de changement de statut.

Pour les contrats sortis, la date de changement de statut correspond à la date de sortie et pour les contrats en vigueur, elle est fixée au 30/06/2017.

Les variables :

- Indicateur Rachat Total,
- Option PB
- Nombre de Rachats Partiels antérieurs

ont été aussi créées.

Au départ, nous voulions modéliser les rachats totaux ainsi que les rachats partiels mais nous nous sommes contentées finalement de ne considérer que l'absence ou non de Rachat Total comme étant la variable d'intérêt et pour prendre en compte les rachats partiels, la variable Nombre de Rachats Partiels a été créée.

La variable Indicateur Rachat Total représente la variable à expliquer. C'est une variable qualitative dichotomique. Si on la note  $Y$ , on a  $Y \in \{0, 1\}$

$$Y_i = \begin{cases} 1 & \text{si Rachat au cours de l'année} \\ 0 & \text{sinon.} \end{cases}$$

Les autres variables seront celles explicatives. Ces variables étant choisies de manière

intuitive ou en se basant sur d'autres travaux, il faudra effectuer des tests statistiques afin de vérifier statistiquement l'existence d'un lien entre ces variables et celle que l'on doit expliquer.

### 3.1.2 La régression logistique (LR)

Nous souhaitons expliquer notre variable d'occurrence (Rachat au cours de l'année ou non) par les autres variables que nous avons à notre disposition. Une régression logistique peut mettre en évidence l'effet de l'ensemble des variables sur notre variable d'intérêt.

#### 3.1.2.1 Définition

Appartenant à la classe des modèles linéaires généralisés (GLM), la régression logistique permet de modéliser la probabilité d'occurrence d'une variable dichotomique à partir de variables qualitatives ou quantitatives. Elle est le plus souvent appliquée en santé pour identifier des facteurs liés à une maladie ou pour rechercher les causes de décès ou de survie des patients. La régression logistique permet d'analyser l'effet de chaque variable explicative sur la variable d'intérêt. Le modèle peut être aussi utilisé à des fins prédictives.

Soit  $Y$  une variable à valeurs dans  $\{0, 1\}$  à expliquer par  $p$  variables explicatives  $X = (X_1, X_2, \dots, X_p)^T$ . Le modèle logistique propose une modélisation de la loi de  $Y|X = x$  par une loi de Bernoulli de paramètres :

$$p_\beta(x) = \mathbb{P}_\beta(Y = 1|X = x)$$

telle que :

$$\log \frac{p_\beta(x)}{1-p_\beta(x)} = \beta_0 + \beta_1 + \dots + \beta_p x_p = x^T \beta$$

$$\mathbf{logit} p_\beta(x) = x^T \beta$$

avec  $\mathbf{logit}$  désignant la fonction bijective et dérivable de  $]0,1[$  dans  $\mathbb{R} : p \rightarrow \log\left(\frac{p}{(1-p)}\right)$

On peut également écrire :

$$p_\beta = \mathbb{P}_\beta(Y = 1|X = x) = \frac{\exp(x^T \beta)}{1 + \exp(x^T \beta)}$$

Dans le modèle logistique, deux choix doivent être effectués pour définir le modèle :

- i) Le choix d'une loi pour  $Y|X = x$ , ici la loi de Bernoulli
- ii) Le choix de la modélisation de  $\mathbb{P}(Y = 1|X = x)$  par :

$$\text{logit } \mathbb{P}_\beta(Y = 1|X = x) = x^T \beta$$

### 3.1.2.2 Estimation des paramètres

Soit :

$$Y_i = \begin{cases} 1 & \text{si Rachat au cours de l'année} \\ 0 & \text{sinon.} \end{cases}$$

la variable qu'on souhaite expliquer à l'aide de  $p$  variables explicatives  $X = (X_1, \dots, X_p)$ .

Comme nous l'avons vu précédemment le modèle logistique est défini par :

$$\text{logit } p_\beta(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p = x^T \beta$$

avec  $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)$  et  $x = (x_1, x_2, \dots, x_p)$

La méthode du Maximum de Vraisemblance nous permet d'estimer ces paramètres.

Notons  $L_n(\beta)$  la fonction de vraisemblance.  $(Y|X = x)$  suit une loi de Bernoulli alors la fonction de vraisemblance comme dans [1] s'écrit :

$$L_n(\beta) = \prod_{i=1}^n \mathbb{P}_\beta((Y = y_i | x_i)) = \prod_{i=1}^n p_\beta(x_i)^{y_i} (1 - p_\beta(x_i))^{1-y_i}$$

En passant au logarithme, nous obtenons :

$$\mathcal{L}_n(\beta) = \sum_{i=1}^n y_i \log \left( \frac{p_\beta(x_i)}{1 - p_\beta(x_i)} \right) + \log(1 - p_\beta(x_i))$$

L'estimateur du maximum de vraisemblance (s'il existe) est solution de l'équation du

score suivante :

$$S(\beta) = \nabla \mathcal{L}_n(\beta) = X^T (\mathbb{Y} - \mathbb{P}_\beta) = 0$$

avec  $\nabla \mathcal{L}_n(\beta)$  le vecteur du gradient au point  $\beta$  défini par :

$$\nabla \mathcal{L}_n(\beta) = \left[ \frac{\partial \mathcal{L}_n}{\partial \beta_0}(\beta), \dots, \frac{\partial \mathcal{L}_n}{\partial \beta_p}(\beta) \right]^T$$

et qui s'obtient par dérivation de la fonction log-vraisemblance :

$$\begin{aligned} \frac{\partial \mathcal{L}_n}{\partial \beta_j}(\beta) &= \sum_{i=1}^n \left[ y_i x_{ij} - \frac{x_{ij} \exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)} \right] \\ &= \sum_{i=1}^n [x_{ij} (y_i - p_\beta(x_i))] \end{aligned}$$

$$\nabla \mathcal{L}_n(\beta) = \sum_{i=1}^n [x_i (y_i - p_\beta(x_i))] = \mathbf{X}^T (\mathbb{Y} - \mathbb{P}_\beta)$$

où

$$\mathbb{Y} = (y_1, \dots, y_n)^T \text{ et } \mathbb{P}_\beta = (p_\beta(x_1), \dots, p_\beta(x_n))^T.$$

L'estimateur du maximum de vraisemblance  $\hat{\beta}$  est le  $\beta$  qui annule l'équation du score. Cette dernière se réécrit comme suit :

$$\begin{cases} x_{11}y_1 + \dots + x_{n1}y_n = x_{11} \frac{\exp(\beta_1 x_{11} + \dots + \beta_p x_{1p})}{1 + \exp(\beta_1 x_{11} + \dots + \beta_p x_{1p})} + \dots + x_{n1} \frac{\exp(\beta_1 x_{n1} + \dots + \beta_p x_{np})}{1 + \exp(\beta_1 x_{n1} + \dots + \beta_p x_{np})} \\ \vdots \\ x_{1p}y_1 + \dots + x_{n1}y_n = x_{11} \frac{\exp(\beta_1 x_{11} + \dots + \beta_p x_{1p})}{1 + \exp(\beta_1 x_{11} + \dots + \beta_p x_{1p})} + \dots + x_{np} \frac{\exp(\beta_1 x_{n1} + \dots + \beta_p x_{np})}{1 + \exp(\beta_1 x_{n1} + \dots + \beta_p x_{np})} \end{cases}$$

Ce système d'équation n'admet pas de solution explicite. Sa résolution requiert donc l'utilisation d'algorithmes d'optimisation tels que celui de Newton-Raphson.

Après avoir obtenu les estimations des coefficients de régression  $\hat{\beta}$  nous pouvons en déduire l'estimation de la probabilité de rachat de chaque assuré.

### 3.1.2.3 Interprétation des coefficients du modèle

Pour interpréter les coefficients du modèle, on a recours à l'Odds-ratio qui est une mesure de quantification du risque. L'Odds-ratio mesure l'évolution du rapport des probabilités d'apparition de l'évènement  $Y = 1$  contre  $Y = 0$ , lorsque  $X_j$  passe de  $x_j$  à  $x_{j+1}$ . Il sert aussi à mesurer le contraste entre les effets d'une variable qualitative.

Considérons l'étude du comportement de rachat d'un contrat (variable expliquée en 0/1) en fonction du genre de l'assuré (variable explicative binaire également). Soient  $p_1$  la probabilité de racheter son contrat si l'assuré est de genre Masculin et  $p_0$  la probabilité de racheter si l'assuré est de genre Féminin. L'Odds-ratio (OR) se calcule selon la formule suivante :

$$OR = \frac{\frac{p_1}{(1-p_1)}}{\frac{p_0}{(1-p_0)}}$$

De manière générale, dans un modèle de régression logistique, pour une variable  $Y \in \{0, 1\}$  à expliquer et une variable  $X_j$  explicative à valeur binaire  $\{0,1\}$ , l'Odds-Ratio correspond à l'exponentielle du coefficient de régression  $\beta_j$  :

$$OR = \frac{P(Y=1|X_j=1)/(1-P(Y=1|X_j=1))}{P(Y=1|X_j=0)/(1-P(Y=1|X_j=0))} = e^{\beta_j}$$

Pour une variable qualitative à plusieurs modalités, l'exponentielle du coefficient de régression correspond à l'odds-ratio entre la modalité de référence et la modalité considérée.

Un  $OR < 1$  indique une influence négative de  $X_j$  sur  $Y$  et un  $OR > 1$ , une influence positive de  $X_j$  sur  $Y$ . L'intervalle de confiance de l'OR au niveau 95% est :

$$[e^{\hat{\beta}-1.96S_{X_j}}, e^{\hat{\beta}+1.96S_{X_j}}]$$

### 3.1.2.4 Sélection et validation de modèle

Le principe de sélection de modèle consiste à soit choisir en présence de  $M_1, M_2, \dots, M_n$  modèles, un seul, le « meilleur » parmi eux par rapport à notre échantillon ou à sélectionner automatiquement le modèle adéquat. La validation constitue, quant à elle la seconde étape pour vérifier statistiquement si le modèle retenu est le « bon ».

#### 3.1.2.4.1 Sélection d'un modèle

Dans le cadre de modèles logistiques, plusieurs méthodes ont été étudiées dans la littérature pour sélectionner le modèle adéquat. D'après L.Rouviere (Rennes2) toujours dans [1], il n'existe pas de critère universel qui permet de définir la notion de meilleur modèle. Un meilleur modèle est toujours évalué par rapport à un critère donné. Parmi les nombreux critères, ceux de l'AIC et BIC sont les plus utilisés. L'idée est de comparer deux modèles donnés en utilisant leur vraisemblance. En réalité, plus la vraisemblance est grande, plus la log-vraisemblance est grande et meilleur est le modèle en terme d'ajustement. En se basant sur cette philosophie, on serait amené à choisir le modèle qui maximise la vraisemblance à savoir le modèle saturé, mais ce dernier sur-ajuste les données car il est sur-paramétré. C'est pourquoi il est nécessaire pour choisir un modèle parcimonieux, de pénaliser la vraisemblance par une fonction du nombre de paramètres.

L'AIC (Akaike Information Criterion) est défini pour un modèle logistique  $M$  à  $p$  paramètres par :

$$AIC(M) = -2\mathcal{L}_n(\hat{\beta}_n) + 2p$$

avec  $\hat{\beta}_n$  l'EMV (estimateur du Maximum de Vraisemblance) des paramètres du modèle.

Le BIC (Bayesian Information Criterion) du modèle  $M$  est quant à lui défini par :

$$BIC(M) = -2\mathcal{L}_n(\hat{\beta}_n) + p \log n$$

Pour chaque modèle, ces deux critères sont calculés et le modèle qui sera retenu sera celui qui minimise l'AIC ou le BIC.

Par ailleurs, un autre critère est la capacité de prédiction du modèle. L'idée est de chercher à comparer les pouvoirs de prédiction des différents modèles logistiques et de choisir celui qui prédit le mieux. Pour ce critère, une règle de prévision doit être définie au préalable pour

le modèle logistique. Pour un modèle  $M_\beta$  fournissant une estimation  $\widehat{p}_\beta(x) = p_{\widehat{\beta}}$ , une règle de prévision  $\widehat{g}_\beta$  évidente à partir de cette estimation est :

$$\widehat{g}_\beta(x) = \begin{cases} 1 & \text{si } \widehat{p}_\beta(x) \geq s \\ 0 & \text{sinon.} \end{cases}$$

avec  $s \in [0, 1]$  un seuil fixé par l'utilisateur et qui prend par défaut la valeur 0.5 dans la plupart des logiciels statistiques.

Il existe plusieurs critères pour mesurer la performance d'une règle de prévision  $\widehat{g}$ . Le plus classique est d'estimer les probabilités d'erreur.

Soit un échantillon  $D_n = (X_1, Y_1), \dots, (X_n, Y_n)$  et  $\widehat{g} : \mathbb{R}^{n+1} \rightarrow \{0, 1\}$  une règle de prévision construite à partir de cet échantillon, la probabilité d'erreur de  $\widehat{g}$  est définie par :

$$L(\widehat{g}) = \mathbb{P}(\widehat{g}(X) \neq Y | D_n)$$

Etant donné  $N$  règles  $\widehat{g}_n$   $n=1, \dots, N$ , l'approche consiste à :

- i) estimer les probabilités d'erreur de toutes les règles candidates à l'aide de l'échantillon
- ii) choisir la règle qui possède la plus petite estimation

La difficulté réside dans le choix d'un bon estimateur pour  $L(\widehat{g})$ . Intuitivement, une première idée consisterait à utiliser :

$$\widehat{L}(\widehat{g}) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\widehat{g}(X_i) \neq Y_i}$$

Mais ceci n'est pas un bon estimateur de  $L(\widehat{g})$  ; en effet, la loi des grands nombres ne peut pas s'appliquer car les variables  $\mathbb{1}_{\widehat{g}(X_i) \neq Y_i}$  ne sont pas indépendantes. De plus, le principal problème est que l'échantillon  $D_n$  est utilisé deux fois pour calculer  $\widehat{g}$  puis pour estimer  $L(\widehat{g})$ .

Pour pallier ce problème, les praticiens utilisent la procédure d'apprentissage/validation qui consiste à découper l'échantillon  $D_n$  en deux :

- un échantillon d'apprentissage  $D_l = (X_i, Y_i), i \in J_l$  de taille  $l$ , utilisé pour construire et estimer les paramètres de différents modèles logistiques afin d'en déduire les règles de prévision  $\widehat{g}$ .

- un échantillon test ou de validation  $D_m = (X_i, Y_i), i \in J_m$  de taille  $m$  (avec  $m+l = n$ ), utilisé pour estimer les probabilités d'erreur de chaque règle  $L(\hat{g})$  par :

$$L_n(\hat{g}) = \frac{1}{m} \sum_{i \in J_m} \mathbb{1}_{\hat{g}(X_i) \neq Y_i}$$

avec  $J_l \cup J_m = \{1, \dots, n\}$  et  $J_l \cap J_m = \emptyset$ .

Il faut noter que  $L_n(\hat{g})$  est un estimateur sans biais de  $L(\hat{g})$ .

Une autre approche pour sélectionner un modèle est la sélection automatique. Autant pour les deux méthodes précédentes, on est en présence de modèles déjà construits et il fallait en choisir un, cette approche de sélection consiste à chercher parmi les variables explicatives  $X_1, X_2, \dots, X_p$ , celles qui expliquent le mieux  $Y$  pour un critère donné (AIC ou BIC par exemple). Toutefois, toujours selon L. Rouvière [1], le nombre de sous-ensembles de  $p$  variables étant de  $2^p$  si  $p$  est grand, sélectionner le meilleur modèle selon le critère qu'on s'est fixé parmi les  $2^p$  modèles logistiques sera coûteux en temps de calcul. C'est pourquoi il convient d'utiliser les méthodes de recherche pas à pas.

Il existe trois grandes méthodes de recherche pas à pas : la méthode ascendante (Forward selection), la méthode descendante (Backward selection) et la méthode progressive (Stepwise selection).

- Méthode Backward : elle part du modèle initial avec toutes les variables disponibles. Au fur et à mesure une variable explicative est retirée si son retrait contribue à une variation optimale et significative (interprétable) du critère de choix qu'on s'est donné. Nous nous arrêtons lorsque toutes les variables sont retirées ou lorsque le retrait d'une variable n'a plus d'impact significatif sur le critère.
- Méthode Forward : c'est le contraire de la méthode Backward. A chaque pas, une variable explicative est ajoutée au modèle.
- Méthode Stepwise : elle combine les deux précédentes. C'est la même méthode que celle ascendante mais des variables déjà introduites peuvent être retirées. En effet, au fur et à mesure qu'on introduit de nouvelles variables, il se peut que certaines variables déjà introduites ne soient plus significatives.

### 3.1.2.4.2 Validation d'un modèle

Après avoir sélectionné ou choisi un modèle, il est nécessaire de mener une étude pour le valider ou l'affiner en mesurant sa qualité d'ajustement globale par rapport aux données observées. Plusieurs tests ont été proposés dans la littérature.

#### 3.1.2.4.2.1 Le test d'adéquation de la déviance

L'idée ici est de se baser sur la vraisemblance. Plus celle-ci est proche de 1, plus le modèle est proche des données. La déviance d'un modèle  $\mathcal{M}$  étant définie comme suit :

$$D_{\mathcal{M}} = 2 \left( \mathcal{L}_{sat} - \mathcal{L}_n \left( \hat{\beta}_n \right) \right)$$

avec  $\hat{\beta}_n$  l'EMV des paramètres et  $\mathcal{L}_n$  la log-vraisemblance du modèle, elle nous permet de comparer la vraisemblance du modèle avec celle du modèle saturé qui est parfait en termes d'adéquation aux données. Ainsi une déviance faible implique une bonne adéquation. Le test d'adéquation de la déviance pose les deux hypothèses suivantes :

$H_0$  : le modèle est adéquat, les données sont bien générées selon par exemple notre modèle logistique contre l'hypothèse

$H_1$  : il ne l'est pas.

En présence de données répétées, la déviance suit une loi du  $\chi_{T-p}^2$  sous  $H_0$  lorsque  $n_t \rightarrow \infty$ ,  $t = 1, \dots, T$ .

On rejettera ainsi  $H_0$  si  $D_{\mathcal{M},obs}$  est plus grande que le quantile d'ordre  $1 - \alpha$  de la loi  $\chi_{T-p}^2$ .

#### 3.1.2.4.2.2 Le test d'adéquation de Pearson

Avec les mêmes hypothèses que le test de déviance, il s'applique toujours dans le cadre des données répétées. La statistique du test est :

$$P = \sum_{t=1}^T \frac{\left( y_t - n_t p_{\hat{\beta}_n}(X_t) \right)^2}{n_t p_{\hat{\beta}_n}(X_t) \left( 1 - p_{\hat{\beta}_n}(X_t) \right)}$$

$P$  suit une loi du  $\chi_{T-p}^2$  sous  $H_0$  en présence de données répétées lorsque  $n_t \rightarrow \infty$ ,

$t = 1, \dots, T$ .

$H_0$  sera rejetée si  $P_{obs}$  est plus grande que le quantile d'ordre  $1 - \alpha$  de la loi  $\chi_{T-p}^2$ .

Remarquons que ces deux tests d'adéquation sont asymptotiques et utilisables uniquement dans le cas de données répétées.

### 3.1.2.4.2.3 Le test de Hosmer-Lemeshow

En présence de données individuelles, c'est le test de Hosmer-Lemeshow qui est utilisé. Il consiste à évaluer la concordance entre les valeurs prédites et observées des observations regroupées en quantiles, typiquement des déciles. Il dépend du nombre de groupes fixés a priori, et il est peu puissant en cas de mauvaise spécification. Plusieurs approches ont été proposées pour construire ce test. Nous présentons ici la démarche proposée par Laurent Rouvière. En présence de données individuelles  $(x_1, Y_1), \dots, (x_n, Y_n)$ , la statistique du test se construit comme suit :

- i) Les probabilités estimées  $p_{\hat{\beta}_n}(x_i)$  sont classées par ordre croissant.
- ii) Ces probabilités ordonnées sont ensuite séparées en  $K$  groupes de taille égale (on prend souvent  $K = 10$  si  $n$  est suffisamment grand).
- iii) Si on note :
  - $m_k$  les effectifs du groupe  $k$ ,
  - $o_k$  le nombre de succès ( $Y = 1$ ) observé dans le groupe  $k$
  - $\mu_k$  la moyenne des  $\hat{p}_{\beta}(x_i)$  dans le groupe  $k$

alors la statistique est définie par :

$$C^2 = \sum_{k=1}^K \frac{(o_k - m_k \mu_k)^2}{m_k \mu_k (1 - \mu_k)}$$

$C^2$  suit approximativement sous  $H_0$  un  $\chi_{K-2}^2$ .

Une autre construction très intéressante du test a été proposée dans [2].

Après avoir retenu le modèle de régression logistique pour expliquer le comportement de rachat des assurés, tout le problème consiste maintenant en la sélection des variables explicatives à intégrer dans la régression. Cette sélection passe par une recherche systématique de toutes les variables statistiquement liées à notre variable d'intérêt. Des tests statistiques seront donc réalisés sur les variables que nous avons retenues.

### 3.1.2.4.3 Les tests statistiques

#### 3.1.2.4.3.1 Le test du chi 2

Le test du chi 2 permet de tester l'indépendance des variables qualitatives à l'aide d'un tableau de contingence.

Le tableau de contingence associé aux variables sous l'hypothèse d'indépendance est obtenu en effectuant le produit des fréquences marginales. Il faut ensuite comparer la distribution empirique obtenue avec le tableau de contingence et la distribution théorique calculée en évaluant la distance entre les deux distributions bidimensionnelles désignées sous les notations :

$$\{f_{ij} : 1 \leq i \leq I; 1 \leq j \leq J\} \text{ et } \{f_i \times f_j : 1 \leq i \leq I; 1 \leq j \leq J\}$$

La distance est mesurée via la quantité :

$$\chi_{(I-1)(J-1)}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{i,j} - n_{..} f_i \times f_j)^2}{n_{..} f_i \times f_j}$$

appelée distance du chi 2 à  $ddl = (I - 1)(J - 1)$  degrés de liberté.

Le test d'indépendance du chi-deux permet de prendre une décision quant à l'hypothèse d'indépendance. Cela revient à tester l'hypothèse d'indépendance :

$$H_0 : f_{ij} = f_i \times f_j, \forall 1 \leq i \leq I \text{ et } \forall 1 \leq j \leq J$$

ou encore :

$$H_0 : f_{ij} - f_i \times f_j = 0, \forall 1 \leq i \leq I \text{ et } \forall 1 \leq j \leq J$$

l'hypothèse nulle contre l'hypothèse de non indépendance

$$H_1 : f_{ij} \neq f_i \times f_j, \text{ pour au moins un couple } (i, j)$$

l'hypothèse alternative.

Notons juste que la principale limite du test chi-2 est qu'il est sensible à la taille de l'échantillon et au nombre de degrés de liberté.

Après avoir sélectionné les variables statistiquement liées à la variable à expliquer, il convient de détecter parmi celles-ci, les variables fortement inter corrélées. En effet, une forte corrélation entre les variables explicatives sélectionnées peut nuire à la qualité de la régression par la suite. Comme précédemment, le test du  $\chi^2$  peut nous permettre d'établir l'existence d'un effet entre nos variables explicatives qualitatives si on les croise dans un tableau de contingence. Toutefois, il faudrait fixer une limite au-delà de laquelle on peut considérer que les variables sont fortement associées. Pour éviter de faire ce choix non judicieux parfois à bien des égards, nous privilégions le test du  $V$  de Cramer. Ce dernier permet de comparer l'intensité du lien ou le degré de relation entre deux variables qualitatives.

### 3.1.2.4.3.2 Le test du $V$ de Cramer

Soient  $A$  et  $B$  deux variables qualitatives à  $p$  et  $q$  modalités respectivement,  $n$  le nombre d'individus sur lesquels  $A$  et  $B$  ont été observés. La table de contingence est un tableau croisé de  $q$  colonnes et  $p$  lignes. Si on note  $n_{ij}$  le nombre d'individus possédant à la fois la modalité  $i$  de la variable  $A$  et la modalité  $j$  de la variable  $B$  alors on a :

	B1	...	Bj	...	Bq	total
A1	$n_{11}$	...	$n_{1j}$	...	$n_{1q}$	$n_{1.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Ai	$n_{i1}$	...	$n_{ij}$	...	$n_{iq}$	$n_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Ap	$n_{p1}$	...	$n_{pj}$	...	$n_{pq}$	$n_{p.}$
total	$n_{.1}$	...	$n_{.j}$	...	$n_{.q}$	$n_{..}$

Le  $V$  de Cramer est la racine carrée du  $\chi^2$  divisé par le  $\chi^2$  max. Le  $\chi^2$  max théorique correspond à l'effectif multiplié par le plus petit côté du tableau (nombre de lignes ou de colonnes) moins 1.

$$V = \sqrt{\frac{\chi^2}{n \times \min(p-1, q-1)}}$$

## 3.2 Application

### 3.2.1 Sélection des variables explicatives

Nous avons retenu en fonction des statistiques descriptives de la partie 2) et en s’inspirant des travaux qui ont été réalisés sur le même sujet les covariables suivantes :

- Age lors de la souscription,
- Genre de l’assuré,
- Ancienneté du contrat,
- Classement département de résidence,
- Prime Investie,
- Périodicité de la prime,
- Clause de PB
- Nombre de Rachats Partiels.

L’étape préalable a consisté à vérifier statistiquement l’existence d’un lien entre chacune de ces variables et la variable Indicateur de Rachat Total qu’on veut expliquer. Toutes les variables ont été catégorisées avant de réaliser le test d’indépendance du  $\chi^2$ . Les p-values obtenues nous ont permis de rejeter l’hypothèse nulle  $H_0$  d’indépendance de nos variables avec la variable d’intérêt. Toutes les p-values sont très inférieures au seuil 5% sauf pour la covariable Genre de l’assuré qui avoisine 4.47%. Ce qui est assez révélateur de l’impact du genre de l’assuré sur son comportement de racheter ou non. Toutefois, nous l’avons considéré comme une variable significative.

Par la suite, puisque toutes les variables choisies au départ ont été retenues, nous avons étudié le degré de corrélation entre elles afin d’éviter les problèmes de multicollinéarité dans la régression. Avec le test V de Cramer, nous avons obtenu des coefficients très proches de 0 (tous inférieurs à 0.2) d’où nous en déduisons que nos covariables ne sont pas dépendantes.

En somme, toutes les covariables qu’on avait retenues vont être considérées comme des variables explicatives à intégrer dans la régression logistique.

### 3.2.2 Mise en pratique de la régression logistique

#### 3.2.2.1 Implémentation du modèle et Résultats

Les variables explicatives retenues ayant chacune 2, 3 modalités ou plus, notre premier réflexe a été de construire un modèle logistique complet avec toutes nos covariables à l’aide de la fonction `glm` de R. Le but est de pouvoir détecter les modalités non significatives qui

peuvent découler d'une mauvaise répartition de notre portefeuille.

Sortie R Modèle Complet				
Coefficients				
	Estimate	Std. Error	Pr(> z )	Indc
(Intercept)	0.52086	0.04869	< 2e-16	***
AGE_SOUSCRIPTION ]35,60]	0 (référence)			
AGE_SOUSCRIPTION[0,35]	0.18994	0.034	2.31E-08	***
AGE_SOUSCRIPTION(60,102]	-0.22469	0.05401	3.18E-05	***
GENRE_ASSUREF	0 (référence)			
GENRE_ASSUREM	-0.13388	0.0296	6.09E-06	***
ANCIENNETE]5, 10]	0 (référence)			
ANCIENNETE[0, 5]	-0.47993	0.04433	< 2e-16	***
ANCIENNETE]10, 20]	-0.78353	0.04734	< 2e-16	***
ANCIENNETE>20	-0.80101	0.07249	< 2e-16	***
Classement_Département1	0 (référence)			
Classement_Département2	-0.07599	0.03396	0.0252	*
Classement_Département3	-0.47205	0.04417	< 2e-16	***
Classement_Département4	-0.23198	0.05128	6.06E-06	***
Classement_Département5	-0.47431	0.09502	5.99E-07	***
Classement_Département999	0.27017	0.17611	0.125	
PRIME.INVESTIE]1000-5000]	0 (référence)			
PRIME.INVESTIE< 1000	-2.32877	0.07078	< 2e-16	***
PRIME.INVESTIE]5000-10000]	0.59569	0.03947	< 2e-16	***
PRIME.INVESTIE>10000	0.9616	0.0413	< 2e-16	***
PERIODICITE_PRIME12	0 (référence)			
PERIODICITE_PRIME1	-0.9828	0.08665	< 2e-16	***
PERIODICITE_PRIME2	-1.29529	0.17367	8.76E-14	***
PERIODICITE_PRIME4	-0.60755	0.07874	1.20E-14	***
PERIODICITE_PRIMEU	-2.0816	0.26625	5.36E-15	***
Clause_PBPB_faible	0 (référence)			
Clause_PBNO_PB	-0.96565	0.04263	< 2e-16	***
Clause_PBPB_élevé	14.74634	90.88547	0.8711	
Clause_PBPB_moyen	0.85049	0.05884	< 2e-16	***
Nb_Rachat_Partiel0	0 (référence)			
Nb_Rachat_Partiel1	0.5322	0.09135	5.68E-09	***
Nb_Rachat_Partiel2	0.3968	0.16492	0.0161	*
Nb_Rachat_Partiel3	0.32568	0.26035	0.211	
Nb_Rachat_Partiel4	0.58567	0.35294	0.097	.
Nb_Rachat_Partiel5	0.53887	0.48602	0.2675	
Nb_Rachat_Partiel6	-2.13195	1.10399	0.0535	.
Nb_Rachat_Partiel7	13.82988	535.41117	0.9794	
Nb_Rachat_Partiel8	-0.4906	1.00652	0.626	
Nb_Rachat_Partiel9	0.59955	1.24812	0.631	
Nb_Rachat_Partiel10	-14.36146	535.41117	0.9786	
Nb_Rachat_Partiel12	-14.43745	535.41117	0.9785	
Nb_Rachat_Partiel13	-12.66624	535.41123	0.9811	

Ce premier modèle permet de voir que certaines modalités d'une même variable ne sont pas significatives d'où la nécessité de s'interroger sur leur éventuel regroupement avec d'autres modalités. Des régressions logistiques simples ont été respectivement effectuées entre la variable d'intérêt et les variables : Nombre de rachats partiels, Clause de PB et Classement Département. En changeant les modalités de référence et en analysant les sorties de régression, nous avons pu grâce à la fonction *LinearHypothesis* de la librairie **car** sous R effectuer des tests de nullité ou d'égalité de coefficients de régression afin de pouvoir regrouper certaines modalités. On obtient ainsi de nouvelles covariables avec un nombre de modalités réduit.

Ensuite, nous avons privilégié la méthode de sélection automatique des variables explicatives pertinentes. Au préalable, nous avons subdivisé notre base en deux : base d'apprentissage (learning) 65% et base de validation (testing) 35%. En partant du modèle de base (régression sur la constante), la fonction **stepAIC** de R nous a permis d'appliquer la méthode de recherche pas à pas Stepwise. La sélection finale faite par l'algorithme est la suivante :

Modèle Stepwise					
Coefficients:	Estimate	Std. Error	z value	Pr(> z )	Indc
(Intercept)	0.48477	0.05996	8.085	6.22E-16	***
AGE_SOUSCRIPTION(35, 60]	0 (référence)				
AGE_SOUSCRIPTION[0,35]	0.18834	0.04182	4.503	6.69E-06	***
AGE_SOUSCRIPTION(60,102]	-0.28326	0.06677	-4.242	2.21E-05	***
GENRE_ASSUREF	0 (référence)				
GENRE_ASSUREM	-0.11865	0.03652	-3.249	0.001159	**
ANCIENNETE]5, 10]	0 (référence)				
ANCIENNETE[0, 5]	-0.49132	0.0544	-9.031	< 2e-16	***
ANCIENNETE]10, 20]	-0.75735	0.05859	-12.927	< 2e-16	***
ANCIENNETE>20	-0.73863	0.08943	-8.259	< 2e-16	***
Classement_Département1	0 (référence)				
Classement_Département2	-0.06767	0.04189	-1.615	0.106246	
Classement_Département3	-0.4728	0.05461	-8.658	< 2e-16	***
Classement_Département4	-0.2469	0.06342	-3.893	9.91E-05	***
Classement_Département5	-0.36693	0.1064	-3.449	0.000563	***
PRIME.INVESTIE[1000, 5000]	0 (référence)				
PRIME.INVESTIE< 1000	-2.32162	0.08587	-27.035	< 2e-16	***
PRIME.INVESTIE]5000-10000]	0.57927	0.04883	11.862	< 2e-16	***
PRIME.INVESTIE>10000	0.91565	0.0511	17.92	< 2e-16	***
PERIODICITE_PRIME12	0 (référence)				
PERIODICITE_PRIME1	-0.99245	0.10873	-9.128	< 2e-16	***
PERIODICITE_PRIME2	-1.33014	0.23075	-5.764	8.19E-09	***
PERIODICITE_PRIME4	-0.62166	0.09705	-6.405	1.50E-10	***
PERIODICITE_PRIMEU	-1.9303	0.30307	-6.369	1.90E-10	***
Clause_PBPB_faible	0 (référence)				
Clause_PBNO_PB	-0.88052	0.05232	-16.83	< 2e-16	***
Clause_PBPB_moyen	0.9618	0.07234	13.296	< 2e-16	***
Nb_Rachat_Partiel0	0 (référence)				
Nb_Rachat_Partiel1	0.51692	0.11445	4.516	6.29E-06	***
Nb_Rachat_Partiel2	0.3631	0.14746	2.462	0.013801	*

Notons que l'AIC obtenu avec ce modèle est inférieur à celui du modèle précédent et

toutes les variables sont significatives.

Par ailleurs, afin d'interpréter ces résultats, nous avons analysé les odds-ratio et la valeur absolue de  $(\beta_k \setminus \sigma(\beta_k))$  (rapport du coefficient de régression sur son écart-type) de chacun de nos facteurs de risque  $k$ . Plus cette valeur absolue est grande, plus la variable explicative influe sur la variable d'intérêt.

	OR	2.50%	97.50%	p	Estimate/Std. Error
(Intercept)	1.623803	1.444014	1.8267	6.22E-16	8.084889927
AGE_SOUSCRIPTION(35, 60]	0 (référence)				
AGE_SOUSCRIPTION[0,35]	1.207242	1.112198	1.3103	6.69E-06	4.503586801
AGE_SOUSCRIPTION(60,102]	0.753323	0.660554	0.8582	2.21E-05	4.242324397
GENRE_ASSUREF	0 (référence)				
GENRE_ASSUREM	0.888119	0.826747	0.954	0.0011592	3.24890471
ANCIENNETE]5, 10]	0 (référence)				
ANCIENNETE[0, 5]	0.611821	0.54991	0.6806	< 2.2e-16	9.031617647
ANCIENNETE]10, 20]	0.468905	0.417963	0.5259	< 2.2e-16	12.92626728
ANCIENNETE>20	0.47777	0.400988	0.5694	< 2.2e-16	8.259308957
Classement_Département1	0 (référence)				
Classement_Département2	0.934567	0.860894	1.0146	0.1062458	1.615421342
Classement_Département3	0.623254	0.559879	0.6935	< 2.2e-16	8.65775499
Classement_Département4	0.781216	0.68977	0.8845	9.91E-05	3.893093661
Classement_Département5	0.692859	0.561577	0.8524	0.0005635	3.448590226
PRIME.INVESTIE[1000, 5000]	0 (référence)				
PRIME.INVESTIE< 1000	0.098115	0.082672	0.1158	< 2.2e-16	27.03645045
PRIME.INVESTIE]5000-10000]	1.784737	1.621902	1.9641	< 2.2e-16	11.86299406
PRIME.INVESTIE>10000	2.4984	2.260595	2.762	< 2.2e-16	17.91878669
PERIODICITE_PRIME12	0 (référence)				
PERIODICITE_PRIME1	0.370667	0.298971	0.458	< 2.2e-16	9.127655661
PERIODICITE_PRIME2	0.26444	0.165111	0.4093	8.19E-09	5.764420368
PERIODICITE_PRIME4	0.537052	0.4437	0.6492	1.50E-10	6.405564142
PERIODICITE_PRIMEU	0.145104	0.077894	0.2575	1.90E-10	6.369155641
Clause_PBPB_faible	0 (référence)				
Clause_PBNO_PB	0.414566	0.374141	0.4593	< 2.2e-16	16.8295107
Clause_PBPB_moyen	2.616413	2.271579	3.0164	< 2.2e-16	13.2955488
Nb_Rachat_Partiel0	0 (référence)				
Nb_Rachat_Partiel1	1.67686	1.341609	2.1019	6.29E-06	4.516557449
Nb_Rachat_Partiel2	1.43778	1.078257	1.9234	0.0138005	2.462362675

Les facteurs de risque qui présentent une grande valeur absolue sont les variables : **Prime Investie**, **Clause de PB** et **Ancienneté du contrat**. Ces variables ont donc une forte influence sur la probabilité de rachat.

En outre, la prime investie étant un indicateur de richesse, ce sont les assurés avec un faible patrimoine qui rachètent le plus leur contrat. Ce comportement peut-être lié au fait qu'ils ne disposent pas d'assez d'argent pour payer les primes ou qu'ils aient des besoins ponctuels de liquidité qui les poussent à désinvestir leurs primes jusqu'à ce qu'elles deviennent insuffisantes.

En ce qui concerne l'ancienneté, son influence sur le comportement de rachat est liée à

l'essence même de l'assurance-vie en France qui est sa fiscalité avantageuse. C'est ainsi que le risque de rachat évolue dans le portefeuille au gré des pénalités de rachat, des prélèvements sociaux et de la levée des contraintes fiscales.

La clause de Participation aux bénéfices impacte aussi fortement la décision de rachat. Les contrats avec un faible taux de PB sont les plus rachetés, ce qui se justifie par la recherche d'un meilleur rendement. Les assurés rachètent leur contrat pour investir dans un autre produit avec un taux de PB plus élevé. Il faut aussi noter qu'ils peuvent aussi être influencés par un effet marketing. C'est le cas notamment lorsqu'un courtier leur propose un transfert vers un produit avec un taux plus élevé.

Après ces trois variables, toujours en analysant les résultats, la périodicité de la prime semble aussi impacter le comportement de rachat. Comme on l'avait remarqué lors de l'analyse du portefeuille, plus ils paient fréquemment leurs primes (mensuellement ou trimestriellement), plus ils sont à même de connaître l'évolution de leur épargne et ainsi racheter en cas de besoin.

Les jeunes-adultes rachètent plus leur contrat que les plus âgés : ils ont besoin de liquidité pour financer les études, acheter des biens mobiliers ou investir sur des biens immobiliers.

L'étude a montré que plus les assurés effectuent des rachats partiels, moins ils rachètent totalement leur contrat. En effet, le rachat partiel est un moyen pour pallier le rachat total et il est conditionné par la détention de suffisamment d'encours.

Enfin le genre de l'assuré ne semble pas avoir un très grand impact sur les comportements de rachat.

### 3.2.2.2 Validation du modèle

Ayant à notre disposition des données individuelles, nous avons effectué le test d'Hosmer-Lemeshow pour évaluer la pertinence du modèle logistique retenu. Sous R, nous avons utilisé la fonction `hoslem.test` de la librairie (*ResourceSelection*). La p-valeur obtenue n'étant pas supérieure au seuil  $\alpha = 0.05$ , on admet alors que le modèle n'est pas bien adapté aux données. Toutefois, nous avons aussi évalué le pouvoir prédictif du modèle. Ce dernier peut se mesurer sous la forme d'une matrice de confusion évaluant le taux de bien et mal classés sur l'échantillon d'apprentissage tandis que l'erreur de prévision est estimée à partir de l'échantillon test.

Erreur d'ajustement		
	0	1
0	7733	2686
1	2294	4149

Erreur de prévision		
	0	1
0	4181	1389
1	1197	2313

Le taux d'erreur estimé sur l'échantillon d'apprentissage est de 29.53%. Le taux d'erreur estimé sur l'échantillon test est du même ordre : 28.48%. Afin de pouvoir analyser les résultats fournis par la matrice de confusion, nous avons calculé certains indicateurs :

- Le taux de succès correspondant à la probabilité de bon classement du modèle
- La sensibilité ou encore Taux de Vrais Positifs (TVP) qui indique la capacité du modèle à retrouver les positifs (rachats)
- La précision qui indique la proportion de vrais positifs parmi les individus qui ont été classés positifs
- La spécificité, qui à l'inverse de la sensibilité, indique la proportion de négatifs détectés
- Le taux de faux positifs (TFP) qui correspond à la proportion de négatifs qui ont été classés positifs

Taux de succès	Sensibilité (Se)	Précision	Spécificité (Sp)	Taux de Faux Positifs
71.52%	65.90%	62.48%	75.06%	24.94%

Généralement un « bon » modèle doit avoir un taux d'erreur et un taux de faux positif proches de 0 et une sensibilité et une spécificité proches de 1. Nos résultats bien qu'ils soient représentatifs et satisfaisants, ne nous permettent pas d'attester que notre modèle est bon. En effet, un modèle peut être jugé comme adéquat que si on le compare à d'autres modèles. L'idéal serait d'implémenter d'autres modèles pour pouvoir faire une comparaison. Néanmoins, la méthode de régression logistique nous a permis non seulement de mettre en avant des profils de risque de nos assurés mais aussi de sélectionner les variables les plus pertinentes pour expliquer leurs comportements de rachat. Notons que tout au long de l'analyse, les variables qui ont été utilisées sont représentatives des caractéristiques des assurés ou de leurs contrats, de ce fait nous n'avons pas tenu compte de l'impact du contexte économique et financier. En réalité, lorsque l'environnement économique se dégrade ou la réputation de la compagnie s'entache, il devient difficile d'anticiper les comportements de rachat. On pourrait assister à un phénomène de rachats massifs, ce qui sous-entend que les comportements de rachat entre les agents sont corrélés. Il devient donc nécessaire de prendre en compte dans la modélisation autant les facteurs endogènes ou structurels que ceux exogènes (conjuncturels). Xavier Milhaud dans [3] a montré qu'une simple régression logistique comme on n'a pu l'implémenter ne suffit plus si on doit prendre en compte les variables conjoncturelles. Il introduit alors l'approche modèle mélange pour tenir compte potentiellement des effets de corrélation et de l'hétérogénéité provenant des phénomènes inobservables comme les décisions de politique ou le Marketing produit.

### 3.3 Modélisation par mélange de lois

Lors de notre modélisation précédente, nous avons considéré les assurés individuellement, chacun présentant ses propres caractéristiques. Toutefois, des phénomènes de rachats massifs ont été observés liés par exemple à un changement de conjoncture économique. Ce qui laisse penser que les décisions de rachat des assurés peuvent être dans une moindre mesure corrélées. Si on veut modéliser les comportements de rachat en prenant en compte certains facteurs externes, une idée serait de partitionner notre population en classes sous-jacentes c'est-à-dire regrouper les individus ayant les mêmes caractéristiques (qui seraient donc susceptibles de réagir de la même façon) par classes homogènes. Cette recherche d'une partition peut se faire en optimisant un critère visant à regrouper automatiquement les individus dans des classes, chacune la plus homogène possible et, entre elles, les plus distinctes possibles. On parle de partitionnement ou en anglais de *clustering*. C'est le principe de la classification non supervisée. Des approches géométriques ont été développées comme méthodes de classification : Clustering Hiérarchique (Classification Ascendante Hiérarchique) et Clustering par affectation dynamique (K-Means). L'idée générale commune à ces deux méthodes est d'utiliser les distances entre les points. Ces méthodes ne seront pas détaillées puisqu'elles présentent des limites. En effet, il ne suffit pas seulement de définir rigoureusement des sous-groupes homogènes, il faut aussi pouvoir comparer deux partitions d'un même ensemble, or cela passe par une comparaison du critère d'homogénéité sur lequel repose chacune d'elles. Les méthodes géométriques ne permettent pas cette comparaison. Une bonne partition doit être composée de classes interprétables. C'est ainsi qu'une nouvelle approche de modélisation probabiliste est privilégiée par rapport aux approches géométriques. Ce sont les modèles mélange qui constituent une alternative puissante.

La première utilisation des modèles de mélange a été faite par Karl Pearson qui a créé un mélange de deux densités de probabilités gaussiennes lors d'une étude concernant des mesures de proportions de la taille du crâne par rapport à la longueur du corps de 1000 crabes. Depuis, l'étude des mélanges de lois a connu un grand succès et il existe beaucoup d'ouvrages dans la littérature s'y afférents. Les modèles mélange se sont ainsi développés dans des domaines très variés tels que la biologie, la génétique, l'ingénierie ou encore l'économie. Dans le cadre d'une classification par modèles de mélange, le sens donné à l'homogénéité diffère des considérations géométriques. Il s'appuie en réalité sur l'analyse de la distribution de probabilité de la population. L'idée est que les individus d'une même classe constituent un échantillon de la même loi de probabilité. Cette approche nous permet non seulement de disposer des probabilités d'appartenance des individus aux différentes classes mais elle propose aussi des solutions théoriques au choix du nombre de classes (qui est inconnu) par l'utilisation des critères classiques de sélection de modèles. Comme bien remarqué dans [4], la classification à base de mélanges peut prendre plusieurs formes selon l'information disponible. Si certaines données sont déjà classées, la classification consistera alors à répartir les

autres données (non classées) dans les classes existantes. On parlera alors de classification semi-supervisée. On parle de classification non supervisée lorsqu'aucune des données n'est classée. Nous allons présenter ici le cadre général des modèles mélange dans un contexte non supervisé.

### 3.3.1 Généralités sur les modèles mélanges

#### 3.3.1.1 Formalisation des modèles de mélange

Cette partie théorique a été présentée dans [5]. Considérons  $n$  individus pour lesquels on dispose d'observations pour une variable  $x$  qui sont notées  $x_1, \dots, x_n$ . Ces individus sont supposés issus de  $K$  classes. En général, le nombre de classes  $K$  est supposé connu et les observations  $x_1, \dots, x_n$  sont des réalisations de  $n$  variables aléatoires  $X_1, \dots, X_n$  qui sont chacune issue d'une distribution propre à la classe à laquelle appartient l'individu associé. Introduisons la variable  $Z$ , appelé étiquette (label) pour chaque individu  $t$  à classer. Ainsi chaque  $X_t$  est associée à un vecteur de dimension  $K$ , noté  $Z_t = Z_{t1}, \dots, Z_{tK}$  tel que :

$$Z_{tk} = \begin{cases} 1 & \text{si l'individu } t \text{ appartient à la classe } k \\ 0 & \text{sinon} \end{cases}$$

Notons  $\pi_k$  la probabilité que l'individu appartienne à la classe  $k$  c'est-à-dire :

$$\pi_k = P(Z_{tk} = 1)$$

Cette probabilité ne prenant pas en compte l'information dont on dispose, à savoir l'observation  $x_t$ , alors c'est une probabilité *a priori* d'appartenance à la classe  $k$  : en d'autres termes, la probabilité qu'une observation prise au hasard appartienne à la classe  $k$ . De plus,  $\sum_{k=1}^K \pi_k = 1$

Nous en déduisons ainsi que les variables aléatoires  $Z_t$  ont pour distribution une loi multinomiale de paramètres les probabilités *a priori* :

$$Z_t \sim \mathcal{M}(1; \pi_1, \dots, \pi_K)$$

Après avoir défini l'appartenance des individus aux classes, il convient de définir la distribution des observations dans chacune des classes. Notons la distribution de  $X_t$  sachant

que l'individu  $t$  appartient à la classe  $k$  par :

$$X_t | Z_{tk} = 1 \sim f_k(x_t)$$

avec  $f_k$  la distribution de probabilité attribuée à la classe  $k$ . Dans un cadre paramétrique,  $f_k$  doit appartenir à une famille de lois paramétrées :

$$f_k(\cdot) = f(\cdot; \theta_k)$$

où  $\theta_k$  sont les paramètres de la distribution  $f$  dans la classe  $k$ . Le choix de la famille d'appartenance de  $f$  est crucial au sens où il aura des conséquences sur la décision finale et que les paramètres  $\theta_k$  permettent de différencier les classes. Un choix de famille complexe pourra entacher la précision de l'estimation des paramètres.

La loi du couple  $(X_t, Z_t)$  est ainsi définie par :

$$f(x_t, z_t) = P(z_t) f(x_t | z_t) = \pi_{z_t} f(x_t | z_t)$$

avec  $\pi_{z_t} = P(Z_t = z_t)$  et  $f(x_t | z_t)$  est la densité de  $X_t$  conditionnellement à  $Z_t = z_t$ . A partir de cette loi, on peut en déduire la loi marginale de  $X_t$ .

Soit l'individu  $t$  appartient à la classe 1 avec probabilité  $\pi_1$  de distribution  $f(\cdot; \theta_1), \dots$ , soit l'individu appartient à la classe  $K$  avec probabilité  $\pi_K$  de distribution  $f(\cdot; \theta_K)$ , la distribution de  $X_t$  s'écrit alors sous la forme :

$$\begin{aligned} f(x_t; \phi) &= P(Z_{t1} = 1) \times f(x_t; \theta_1) + \dots + P(Z_{tK} = 1) \times f(x_t; \theta_K) \\ &= \sum_{k=1}^K P(Z_{tk} = 1) \times f(x_t; \theta_k) \\ &= \sum_{k=1}^K \pi_k f(x_t; \theta_k) \end{aligned}$$

C'est un mélange de distributions : la somme des distributions des  $K$  classes pondérées par leur taille  $\pi_k$  (la proportion d'individus appartenant à la classe). Chaque classe  $k$  est ainsi caractérisée dans le modèle par  $\pi_k$  et par  $\theta_k$ . On notera par la suite  $\phi = (\phi_1, \dots, \phi_K)$  l'ensemble des paramètres du modèle avec pour chaque classe  $k$  :

$$\phi_k = (\pi_k, \theta_k)$$

Dans un modèle mélange, nous avons donc les variables  $X_t$  et  $Z_t$  qui sont représentées sous forme d'un couple  $Y_t = (X_t, Z_t)$  appelé données complètes. Dans le cadre de notre modélisation, seules les  $X_t$  sont observées, nous sommes donc en présence de données incomplètes. Nous devons reconstruire les variables d'appartenance  $Z_{tk}$ .

### 3.3.1.2 Formule de Bayes

L'objectif étant de classer l'individu  $t$  dans la classe dont il a le plus de chance d'être issu eu égard de sa valeur  $x_t$  observée et des caractéristiques des classes, nous nous intéressons donc à la probabilité qu'il appartienne à la classe  $k$  sachant que l'on a observé pour lui la valeur  $x_t$  de  $X$ . Cette probabilité est notée  $\tau_{tk}$  :

$$\tau_{tk} = P(Z_{tk} = 1 | X_t = x_t)$$

C'est la probabilité *a posteriori* que l'individu soit dans la classe  $k$  (l'information  $x_t$  dont on dispose étant bien prise en compte). Avec la formule de Bayes, cette probabilité se réécrit :

$$\tau_{tk} = \frac{P(Z_{tk} = 1) f(x_t; \theta_k)}{f(x_t)}$$

On obtient au final :

$$\tau_{tk} = \frac{\pi_k f(x_t; \theta_k)}{\sum_{l=1}^K \pi_l f(x_t; \theta_l)}$$

Pour classer les individus, c'est la règle du **Maximum A Posteriori** (MAP) qui est utilisée. Elle consiste à classer l'individu  $t$  dans la classe  $k$  correspondant à sa probabilité  $\tau_{tk}$  maximale. Si cette dernière est proche de 1 pour une classe donnée alors l'individu est classé de façon certaine dans cette classe, sinon il sera difficile de classer avec certitude. Afin de pouvoir classer les individus, il est nécessaire de connaître les paramètres des classes ( $\pi_k$  et  $\theta_k$ ). Il faudra donc les estimer.

### 3.3.1.3 Estimation des paramètres du modèle mélange

La première méthode utilisée par Pearson pour l'estimation fut basée sur l'étude des moments. Il existe aussi des méthodes graphiques mais les plus utilisées sont les méthodes bayésiennes et celle du maximum de vraisemblance. Nous allons exposer ici cette dernière méthode. Nous devons estimer les paramètres  $\phi_k = (\pi_k, \theta_k)$  de notre modèle. En tenant compte des notations précédentes et de l'indépendance des observations, la vraisemblance

s'écrit :

$$\begin{aligned} L(X_1, \dots, X_n; \phi) &= \prod_{t=1}^n f(X_t; \phi) \\ &= \prod_{t=1}^n \left\{ \sum_{k=1}^K \pi_k f(X_t; \theta_k) \right\} \end{aligned}$$

On obtient par passage au logarithme :

$$\log L(X_1, \dots, X_n; \phi) = \sum_{t=1}^n \log \left\{ \sum_{k=1}^K \pi_k f(X_t; \theta_k) \right\}$$

Avec cette expression de la log-vraisemblance, il nous sera difficile de chercher les paramètres du modèle en annulant sa dérivée. C'est pourquoi nous aurons recours à des algorithmes itératifs de recherche de maximum. Les algorithmes les plus puissants sont ceux du type EM («Expectation-Maximisation»).

### - L'Algorithme EM

Du à Dempster, Laird et Rubin (1977), il a pour but de fournir un estimateur lorsqu'il est impossible de calculer la solution en raison de la présence de données cachées ou manquantes ou que l'expression de la vraisemblance est analytiquement impossible à maximiser. C'est un algorithme itératif qui, à chaque itération, opère deux étapes : phase « **Expectation** » désignée comme l'étape E et la phase « **Maximisation** » ou étape M d'où son nom EM. Les aspects théoriques de cet algorithme sont détaillés dans [6]. Rappelons que nous sommes ici en présence de données incomplètes en classification non supervisée. Or dans la littérature, la vraisemblance des données incomplètes est décrite souvent comme n'étant pas simple à manipuler. C'est dans ce contexte que la vraisemblance des données complètes présente un grand intérêt pour l'algorithme EM. Ce dernier tire en effet partie de l'information manquante en considérant l'espérance conditionnelle de la log-vraisemblance de l'échantillon complet sachant les données observées. La vraisemblance des données complètes est donnée par :

$$\begin{aligned} \log L(X, Z; \phi) &= \log \left\{ \prod_{t=1}^n \prod_{k=1}^K [\pi_k f(X_t; \theta_k)]^{Z_{tk}} \right\} \\ &= \sum_{t=1}^n \sum_{k=1}^K Z_{tk} \log \{ \pi_k f(X_t; \theta_k) \} \end{aligned}$$

Les variables  $Z_{tk}$  n'étant pas observées, l'idée a été de les remplacer par leur meilleure

prédiction sachant les données observées  $X_t$ , à savoir  $(Z_{tk} | X_t = x_t) = \tau_{tk}$ . La vraisemblance est équivalente alors à :

$$\sum_{t=1}^n \sum_{k=1}^K \tau_{tk} \log\{\pi_k f(X_t; \theta_k)\}$$

La connaissance des probabilités  $\tau_k$  implique l'obtention des valeurs des estimations des paramètres en maximisant la vraisemblance. C'est ce que fait exactement l'algorithme EM.

En réalité d'après [6], lors de la phase « Expectation », l'algorithme procède à l'estimation des données inconnues sachant les données observées et la valeur des paramètres déterminée à l'itération précédente. Par la suite dans la phase « Maximisation », il procède à la maximisation de la vraisemblance en utilisant l'estimation des données inconnues effectuée à l'étape E et met à jour la valeur du ou des paramètres pour la prochaine itération. Si on reprend notre cas et on note toujours comme dans [5],  $\phi^{(h)} = (\pi^{(h)}, \theta^{(h)})$  la valeur du paramètre courant, l'**étape E** de l'algorithme EM va consister à l'itération  $(h + 1)$  au calcul des probabilités *a posteriori*  $\tau_k$  à partir de la valeur courante des paramètres  $\phi^{(h)}$  :

$$\tau_{tk}^{(h+1)} = \frac{\pi_k^{(h)} f(X_t; \theta_k^{(h)})}{\sum_{l=1}^K \pi_l^{(h)} f(X_t; \theta_l^{(h)})}$$

Et à l'**étape M**, on actualise les paramètres en maximisant la vraisemblance où on remplace les  $\tau_k$  par les valeurs obtenues à l'étape E :

$$\phi^{(h+1)} = \underset{\phi}{\text{Argmax}} \sum_{t=1}^n \sum_{k=1}^K \tau_{tk}^{(h+1)} \{\log \pi_k f(X_t; \theta_k)\}$$

### 3.3.1.4 Choix de modèle

En classification non supervisée ou classification automatique, la principale difficulté réside dans l'évaluation du bon nombre de classes  $K$ . La classification avec les lois mélange fournit cependant des outils puissants pour déterminer le nombre de classes avec des critères de choix de modèle qui sont très répandus. Une première approche se présente sous forme d'un test d'hypothèse. On teste l'hypothèse nulle  $H_0$  d'un nombre de classes  $K_0$  contre l'hypothèse alternative  $H_1$  d'un nombre de classes  $K_1$  supérieur. C'est le LRT (Likelihood Ratio Test) qui a pour statistique le log-rapport de la vraisemblance maximale du modèle sous chacune des hypothèses. Malheureusement ce test présente plusieurs inconvénients. Parmi eux, on a la non-connaissance de la distribution asymptotique de la statistique du test. D'autres détails sur certains des inconvénients de ce test sont explicités dans [4] et [7]. Par ailleurs, lors de l'estimation des paramètres du modèle mélange dans la partie précédente, le nombre

de classes avait été fixé au préalable. L'idée était d'estimer les paramètres pour plusieurs valeurs de  $K$  et après de choisir un modèle adéquat. Une deuxième approche serait donc de choisir l'un des modèles mélange créés avec les différentes valeurs de  $K$  selon un critère. Les critères les plus utilisés sont l'AIC (Akaike Information Criterion) ou le BIC (Bayesian Information Criterion).

### 3.3.2 Quel modèle mélange pour notre problématique ?

#### 3.3.2.1 Les mélanges de régressions logistiques

Revenons-en à notre problématique de modélisation des comportements de rachat de notre portefeuille d'épargne. Notre variable réponse, Indicateur de rachat total est défini comme suit :

$$Y_i = \begin{cases} 1 & \text{si Rachat total} \\ 0 & \text{sinon} \end{cases}$$

C'est notre variable à expliquer. Nous avons à notre disposition d'autres variables qui sont celles explicatives. En réalité, notre objectif est de mesurer les effets ou l'influence de ces variables explicatives sur la décision de rachat de nos assurés. Nous avons modélisé en premier lieu ce comportement avec une régression logistique qui était adaptée. Dans un contexte de lois mélange, un mélange de régressions logistiques serait donc adapté pour prendre en compte dans la modélisation l'hétérogénéité non observée de notre portefeuille.

Soit les notations suivantes :

$i = 1, \dots, n$  le nombre d'individus de notre portefeuille

$k = 1, \dots, c$  le nombre de classes

$l = 1, \dots, t$  les variables explicatives

Considérons  $Y = (Y_1, Y_2, \dots, Y_n)$  un vecteur de variables aléatoires de loi de Bernoulli. La construction du modèle mélange de régressions logistiques que nous allons présenter, a été introduite dans [8] : nous supposons que chaque individu suit une distribution mélange fini de  $c$  classes de proportions  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_c)$  avec  $\sum_{k=1}^c \alpha_k = 1$  sachant que nous n'avons pas connaissance de la classe d'appartenance de chaque individu. La densité conditionnelle de chaque individu  $Y_i = y_i$  issu de la classe  $k$  est alors définie par :

$$f_{i|k}(y_i|p_k) = (p_k)^{y_i} (1 - p_k)^{1-y_i} = (1 - p_k) \exp\left(y_i \ln \frac{p_k}{1 - p_k}\right)$$

avec  $p_k$  la probabilité d'appartenance de l'individu  $i$  à la classe  $k$ .

Par ailleurs les mélanges de régressions logistiques étant un cas particulier des mélanges de GLMs( Modèles linéaires Généralisés) qui ont cette forme générale :

$$f_{i|k}(y_i|p_k) = a(p_k) b(y_i) \exp[y_i Q(p_k)]$$

En conciliant cette formule avec celle précédente, nous obtenons par identification :

$$a(p_k) = 1 - p_k$$

$$b(y_i) = 1$$

$$Q(p_k) = \ln \frac{p_k}{1-p_k}$$

Spécifions le prédicteur linéaire  $\eta_k$  et la fonction lien  $g(\cdot)$  de la classe  $k$  :

$$\eta_k = g(p_k(x))$$

Soit  $x = (x_1, x_2, \dots, x_t)$  nos variables explicatives et  $\beta = (\beta_1, \beta_2, \dots, \beta_c)$  un vecteur de coefficients inconnus avec  $\beta_k = (\beta_{kl})$  de la classe  $k$ . Le prédicteur linéaire s'exprime en fonction des  $t$  variables explicatives et des  $\beta_k$  :

$$\eta_k = \sum_{l=1}^t x_l \beta_{kl}$$

Pour chaque individu  $i$ , nous avons :

$$\eta_{ki} = g(p_k(x_i))$$

avec  $x_i = (x_{i1}, \dots, x_{it})$ . Ce qui implique que

$$\eta_{ki} = \sum_{l=1}^t x_{il} \beta_{kl}$$

Nous pouvons donc affirmer que conditionnellement à la classe d'appartenance  $k$ , le modèle linéaire généralisé consiste en la spécification de la variable  $Y_i$ , du prédicteur  $\eta_{ki}$  et de la fonction lien  $g(\cdot)$ . MuCullagh and Nelder fournissent un lien canonique avec

$$Q(p_k(x_i)) = \eta_{ki}$$

Ainsi, on a :

$$\eta_{ki} = \ln \frac{p_k(x_i)}{1 - p_k(x_i)}$$

Ce qui implique que

$$\ln \frac{p_k(x_i)}{1 - p_k(x_i)} = \sum_{l=1}^t x_{il} \beta_{kl}$$

Avec la transformation suivante :

$$\exp \left( \sum_{l=1}^t x_{il} \beta_{kl} \right) = \frac{p_k(x_i)}{1 - p_k(x_i)}$$

On obtient la formule pour  $p_{ki}$  :

$$p_{ki} = p_k(x_i) = \frac{\exp \left( \sum_{l=1}^t x_{il} \beta_{kl} \right)}{1 + \exp \left( \sum_{l=1}^t x_{il} \beta_{kl} \right)}$$

Nous venons en effet d'exprimer les  $p_k$  en fonction des variables explicatives. La densité conditionnelle de chaque individu  $Y_i = y_i$  issu de la classe  $k$  peut alors se réécrire sous la forme suivante :

$$f_{i|k}(y_i|p_k) = (1 - p_{ki}) \exp \left( y_i \ln \frac{p_{ki}}{1 - p_{ki}} \right)$$

D'où nous déduisons la fonction de densité :

$$f_{i|k}(y_i|\beta_k) = \exp \left\{ y_i \sum_{l=1}^t x_{il} \beta_{kl} - \ln \left[ 1 + \exp \left( \sum_{l=1}^t x_{il} \beta_{kl} \right) \right] \right\}$$

Au final, la fonction de densité de  $Y_i$  peut être exprimée sous la forme de mélange fini :

$$f(y_i|\alpha, \beta) = \sum_{k=1}^c \alpha_k f_{i|k}(y_i|\beta_k)$$

avec

$$\sum_{k=1}^c \alpha_k = 1$$

L'objectif maintenant est d'estimer les paramètres  $\beta_k$  et  $\alpha_k$  sachant  $y_i$  et  $x_{il}$  par la maximisation de la vraisemblance du modèle :

$$\begin{aligned} L(\alpha, \beta|Y = y) &= \prod_{i=1}^n f(y_i|\alpha, \beta) = \prod_{i=1}^n \sum_{k=1}^c \alpha_k f_{i|k}(y_i|\beta_k) \\ &= \prod_{i=1}^n \sum_{k=1}^c \alpha_k \exp \left\{ y_i \sum_{l=1}^t x_{il} \beta_{kl} - \ln \left[ 1 + \exp \left( \sum_{l=1}^t x_{il} \beta_{kl} \right) \right] \right\} \end{aligned}$$

avec l'algorithme EM dont la procédure a été décrite dans la partie précédente.

### **3.3.2.2 Application du modèle mélange de logit à notre portefeuille**

#### **3.3.2.2.1 Les données utilisées**

Dans cette deuxième partie de la modélisation, nous avons créé une nouvelle base de notre portefeuille d'épargne en ajoutant de nouvelles variables économiques et financières. En plus des variables qui avaient été introduites dans la régression logistique simple, nous avons aussi : la saisonnalité, la variation du TME, la variation du CAC40 et la variation du chômage, ces trois dernières représentant les facteurs exogènes ou conjoncturels. Notons aussi que la variable Clause de PB a été remplacée par la Classe des Taux Servis.

La saisonnalité est une variable temporelle correspondant à la date de sortie pour les rachats et les autres sorties et pour les contrats en cours, elle a été fixée à Juin-2017. En effet, elle a été introduite comme variable parce que l'assuré, selon certaines périodes de l'année, peut décider de racheter son contrat ou pas. Cette variable a été délimitée en classes.

Pour les trois variables conjoncturelles, nous avons considéré leur variation mensuelle de Janvier 2005 à Juin 2017. En partant de Janvier 2005 comme référence, nous les avons estimées à la hausse ou à la baisse. Par ailleurs, pour la variable Classes Taux Servis, elle a été introduite à la place de la clause de PB parce qu'après avoir analysé les taux et discuté avec la Direction Technique, il s'est avéré que la distribution de PB repose sur une stratégie spécifique de la compagnie. Cette dernière consiste à servir un taux de PB très faible aux contrats ayant déjà un taux garanti assez élevé afin de ne point se retrouver avec des taux de revalorisation très élevés pour certains contrats et d'assurer aussi un certain équilibre pour pouvoir faire face aux engagements. C'est pourquoi, le taux crédité au contrat (taux de participation aux bénéfices + taux garanti) nous semble plus pertinent pour expliquer la décision de rachat que le taux de PB seulement.

### 3.3.2.2 Implémentation du modèle mélange et résultats

L'étape préalable a consisté à effectuer des tests statistiques afin de trouver un lien existant entre notre variable d'intérêt (la décision de racheter) et les nouvelles variables explicatives qui ont été introduites. Les p-values très inférieures au seuil fixé 5% du test d'indépendance du chi2 nous ont permis d'affirmer la dépendance de notre variable d'intérêt avec nos variables conjoncturelles. Le test du V de Cramer a été aussi effectué pour tester le degré de liaison entre nos variables explicatives : nous avons déduit des faibles coefficients proches de 0 obtenus, que nos variables explicatives ne sont pas dépendantes.

Pour l'implémentation du modèle mélange de logit, nous avons utilisé le package **flexmix** de R. Au départ, ne connaissant pas le nombre de composantes (classes)  $K$ , nous avons commencé avec le modèle sans les variables explicatives. Avec les fonctions *FLXMRglm* et *stepFLexmix*, l'estimation de ce premier modèle va nous fournir le nombre de composantes en sélectionnant automatiquement un modèle parcimonieux selon le critère BIC.

iter	converged	k	k <sub>0</sub>	logLik	AIC	BIC	ICL
2	TRUE	1	1	-13569.21	27140.42	27148.16	27148.16
2	TRUE	2	2	-13569.21	27144.42	27167.62	50495.35
4	TRUE	3	3	-13569.22	27148.43	27187.09	64122.27
5	TRUE	4	4	-13569.21	27152.43	27206.56	73825.52
4	TRUE	5	5	-13569.21	27156.43	27226.02	81284.85
5	TRUE	6	6	-13569.22	27160.43	27245.49	87060.79
5	TRUE	7	7	-13569.21	27164.43	27264.95	92449.94

Le modèle ainsi sélectionné est celui avec une seule composante. L'étape suivante a consisté donc à *fitter* un modèle à une composante en introduisant nos variables explicatives. C'est le modèle global. Avec un BIC de 22376,82, ce deuxième modèle est plus parcimonieux que celui précédent. Toutefois en analysant les coefficients des régressions ainsi que les p-value de notre sortie de modèle avec la fonction *refit()*, nous en déduisons que les variables Ancienneté, Prime Investie et Périodicité de la Prime restent toujours pertinentes et explicatives de la décision de rachat. Pour autant que les variations du TME et du CAC 40 impactent la variable d'intérêt, les variable Age Souscription, Nombre de Rachats Partiels et Taux de chômage ne semblent pas beaucoup influencer sur la décision de rachat (voir Annexe).

C'est ainsi qu'un troisième modèle mélange, toujours à une composante, a été estimé en supprimant ces trois dernières variables afin de mesurer l'impact des variables restantes.

Avec un BIC de 22334,26, ce modèle nous permet d'observer une nette différence dans la significativité des variables Classes des Taux Servis et Variation du TME. La saisonnalité et la Variation du CAC 40 restent des variables pertinentes. Puisque les effets des variables Ancienneté, Prime Investie et Périodicité de la Prime restent inchangés pour les deux modèles précédents, la fonction **FLXMRglmfix** nous permet de les fixer afin d'estimer un quatrième modèle.

Ce dernier modèle présente les mêmes résultats que le troisième. Avec la fonction `refit()`, nous avons pu analyser la significativité des variables explicatives.

Modalités	Estimate	Std. Error	z value	Pr(> z )	Ind
ANCIENNETE ]5, 10 ]	0 (référence)				
ANCIENNETE]0, 5]	-0.508203	0.045093	-11.2700	< 2.2e-16	***
ANCIENNETE]10, 20]	-0.384665	0.049889	-7.7105	1.253e-14	***
ANCIENNETE>20	-0.606310	0.072808	-8.3275	< 2.2e-16	***
PRIME.INVESTIE ]1000-5000]	0 (référence)				
PRIME.INVESTIE< 1000	-0.720089	0.035058	-20.5399	< 2.2e-16	***
PRIME.INVESTIE]5000-10000]	0.279690	0.040038	6.9857	2.835e-12	***
PRIME.INVESTIE>10000	0.267350	0.040035	6.6779	2.424e-11	***
PERIODICITE_PRIME12	0 (référence)				
PERIODICITE_PRIME1	-0.578290	0.078037	-7.4104	1.259e-13	***
PERIODICITE_PRIME2	-1.078160	0.119368	-9.0322	< 2.2e-16	***
PERIODICITE_PRIME4	-0.835251	0.072406	-11.5357	< 2.2e-16	***
PERIODICITE_PRIMEU	-1.078860	0.217853	-4.9522	7.336e-07	***
(Intercept)	2.889908	0.075415	38.3199	< 2.2e-16	***
Classes_TxServistaux_moyen	0 (référence)				
Classes_TxServisNO_tx	-0.135848	0.049249	-2.7584	0.005808	**
Classes_TxServistaux_élevé	0.405806	0.062759	6.4660	1.006e-10	***
Classes_TxServistaux_faible	-0.261825	0.048079	-5.4457	5.161e-08	***
Classes_SAISONNALITfin_d'annee	0 (référence)				
Classes_SAISONNALITEcourant	-0.631849	0.051083	-12.3691	< 2.2e-16	***
Classes_SAISONNALITEdébut_d'année	-0.080235	0.065738	-1.2205	0.222269	
Classes_SAISONNALITEété	0.152336	0.064226	2.3719	0.017697	*
Variation_TMEBL	0 (référence)				
Variation_TMEB	-0.920533	0.046091	-19.9721	< 2.2e-16	***
Variation_TMEH	0.082857	0.066520	1.2456	0.212917	
Variation_TMEHL	-0.172408	0.058728	-2.9357	0.003328	**
Variation_TMES	0.484342	0.179330	2.7008	0.006916	**
Variation_CAC40H	0 (référence)				
Variation_CAC40B	-0.460874	0.042651	-10.8056	< 2.2e-16	***
Variation_CAC40BL	0.130531	0.105939	1.2321	0.217900	
Variation_CAC40HL	-0.090099	0.070182	-1.2838	0.199215	

#### Sortie modèle mélange à une composante

Les agents réagissent automatiquement lors d'une forte baisse du CAC 40, ce comportement est à bien des égards rationnel. Ils rachètent aussi lors qu'on assiste à une variation à la baisse du TME. Pourtant, nous observons des comportements assez mitigés lorsque le TME hausse légèrement ou qu'il se stabilise. L'effet de la saisonnalité est assez faible. Ils rachètent plus au courant de l'année qu'en été ou encore en début d'année. En ce qui concerne les taux servis, ceux qui n'en perçoivent pas ont de faibles probabilités de racheter leur contrat. Globalement le modèle mélange à une composante ne nous a pas permis de classer nos

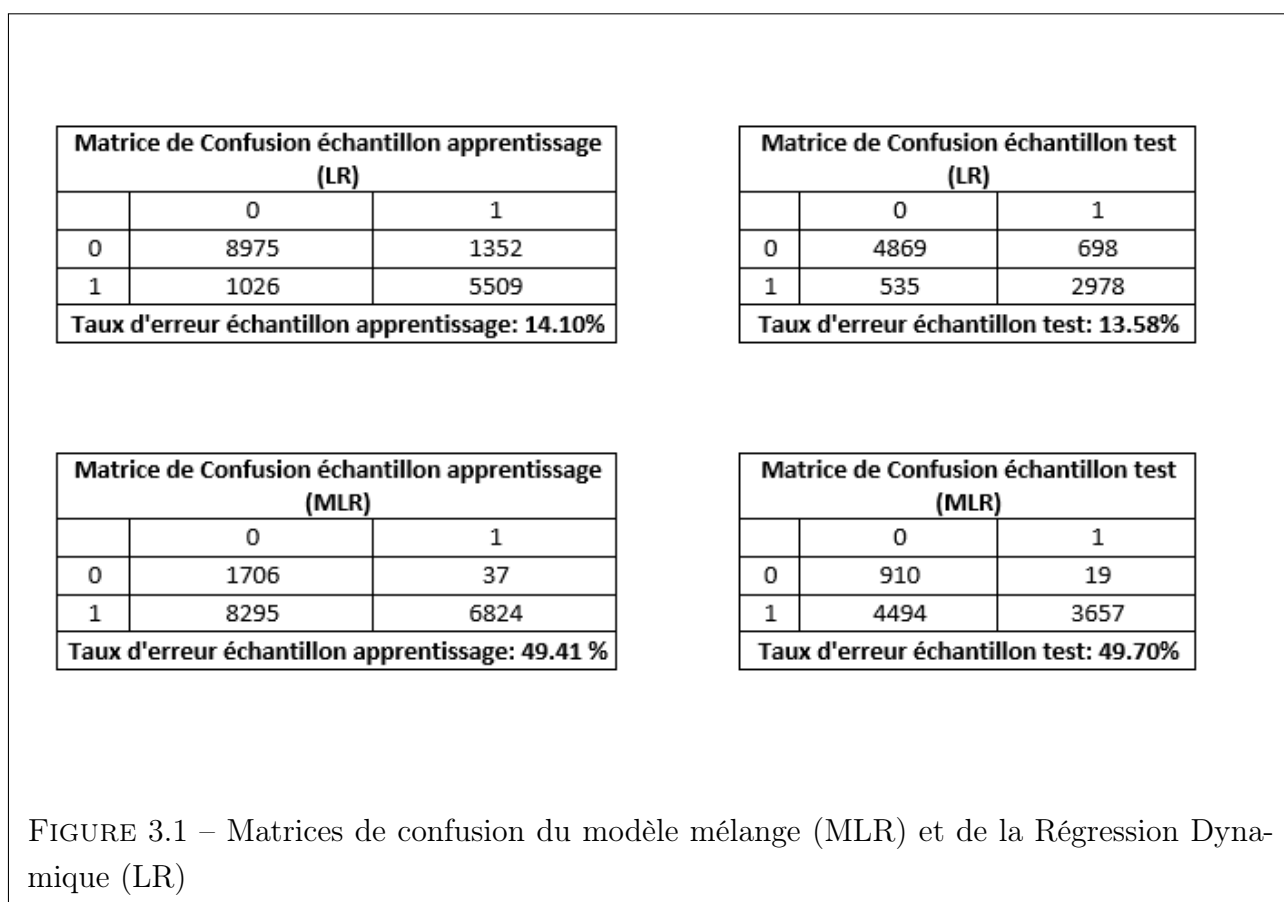
individus en groupes homogènes et donc d'observer une certaine hétérogénéité de notre portefeuille, ce qui est peut-être lié à la dimension de ce dernier. Nous aurions voulu mesurer l'impact de nos variables explicatives selon les différents groupes.

Ce modèle mélange de logit à une composante pourrait ainsi s'apparenter à une simple régression logistique dénommée régression dynamique (avec les variables conjoncturelles) dans la littérature. C'est ainsi que dans cette dernière section du mémoire, nous allons dans un but comparatif implémenter une nouvelle régression logistique avec les variables retenues du modèle mélange. Nous avons les résultats des odds-ratio et de la valeur absolue de  $(\beta_k/\sigma(\beta_k))$ .

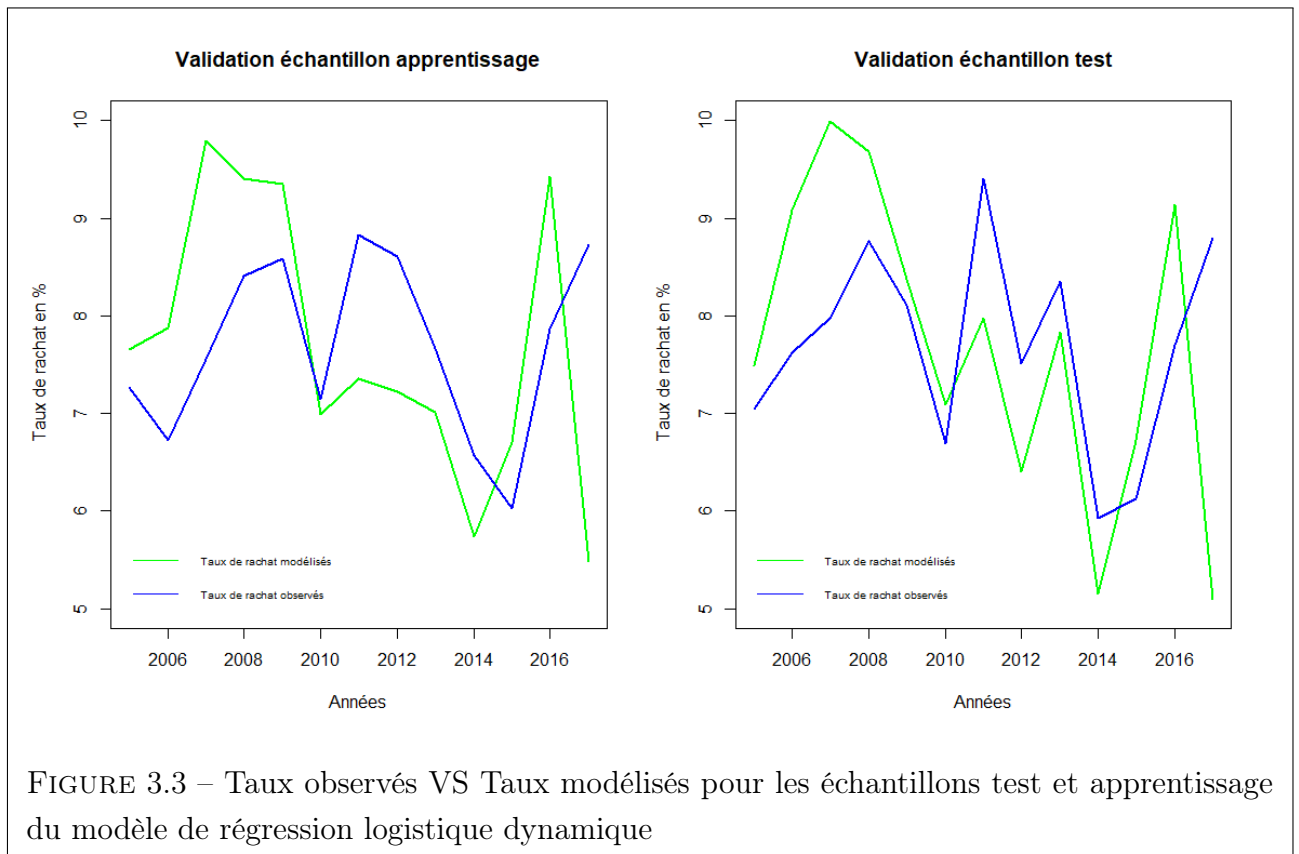
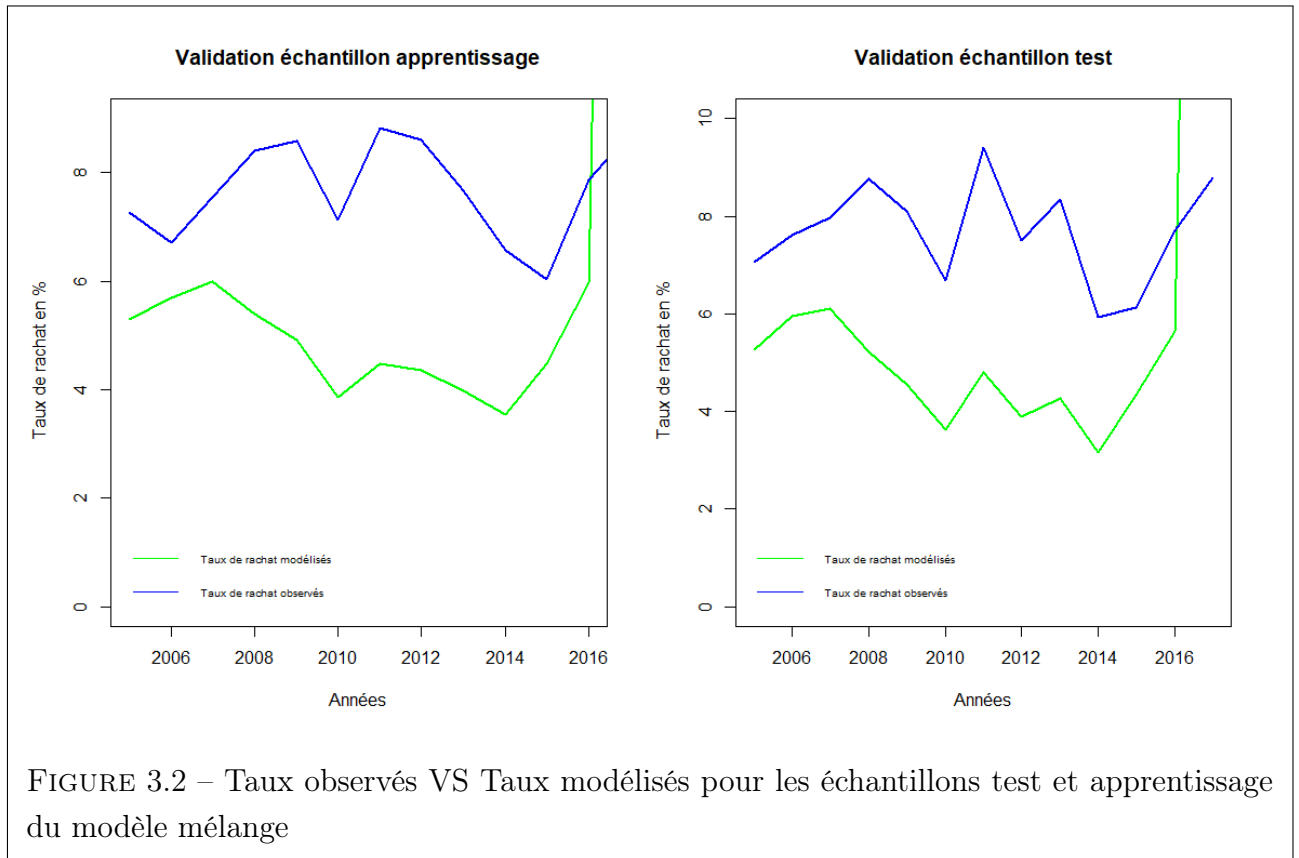
	OR	2.5 %	97.5 %	p	Estimate/Std. Error
<b>(Intercept)</b>	33.630322	27.136444	41.7932	< 2.2e-16	31.91203704
<b>Variation_TMEB</b>	0.138917	0.121965	0.1580	< 2.2e-16	29.87106538
<b>Variation_TMEH</b>	1.311268	1.091217	1.5774	0.003933	2.883485848
<b>Variation_TMEHL</b>	0.790239	0.672470	0.9286	0.004242	2.859467995
<b>Variation_TMES</b>	1.279460	0.827535	2.0425	0.283881	1.071664637
<b>PRIME.INVESTIE&lt; 1000</b>	0.038806	0.031977	0.0469	< 2.2e-16	33.34534072
<b>PRIME.INVESTIE]5000-10000]</b>	2.013956	1.780534	2.2788	< 2.2e-16	11.12329202
<b>PRIME.INVESTIE&gt;10000</b>	2.148073	1.893140	2.4378	< 2.2e-16	11.85379845
<b>Classes_SAISONNALITEcourant</b>	0.271768	0.235567	0.3131	< 2.2e-16	17.94504132
<b>Classes_SAISONNALITEdébut_d'année</b>	0.853718	0.712659	1.0229	0.086227	1.715665003
<b>Classes_SAISONNALITEété</b>	1.436719	1.204719	1.7139	5.584e-05	4.02980427
<b>PERIODICITE_PRIME1</b>	0.210718	0.160713	0.2757	< 2.2e-16	11.30975379
<b>PERIODICITE_PRIME2</b>	0.054839	0.034097	0.0863	< 2.2e-16	12.28775182
<b>PERIODICITE_PRIME4</b>	0.146664	0.116852	0.1838	< 2.2e-16	16.60705943
<b>PERIODICITE_PRIMEU</b>	0.052554	0.025184	0.1069	1.016e-15	8.024843367
<b>Variation_CAC40B</b>	0.337186	0.299235	0.3797	< 2.2e-16	17.88909001
<b>Variation_CAC40BL</b>	1.194114	0.895111	1.6043	0.233111	1.192364565
<b>Variation_CAC40HL</b>	0.885985	0.729759	1.0770	0.222634	1.219625227
<b>Classes_TxServisNO_tx</b>	0.490778	0.425874	0.5654	< 2.2e-16	9.847260653
<b>Classes_TxServistaux_élevé</b>	1.998372	1.682583	2.3759	3.658e-15	7.866492444
<b>Classes_TxServistaux_faible</b>	0.660453	0.574712	0.7588	4.806e-09	5.853393537
<b>ANCIENNETE[0, 5]</b>	0.420737	0.368917	0.4796	< 2.2e-16	12.93708906
<b>ANCIENNETE]10, 20]</b>	0.375724	0.322867	0.4369	< 2.2e-16	12.6882696
<b>ANCIENNETE&gt;20</b>	0.344651	0.279375	0.4251	< 2.2e-16	9.946960501

Tableau 1: Odds-Ratio et Valeurs Absolues de la Régression Logistique Dynamique

Nous observons les mêmes résultats qu'avec le modèle mélange d'un point de vue interprétabilité des coefficients de régression. L'effet de la variable CAC 40 reste assez faible, toutefois les effets des variables TME et saisonnalité sont plus accentués avec ce modèle et diffèrent légèrement des observations du modèle mélange. Afin de comparer les deux modèles, nous avons estimé leur pouvoir prédictif et mesuré leur ajustement à nos données.



Ces résultats démontrent le très mauvais pouvoir prédictif du modèle mélange de logit comparé à celui de la régression logistique. En effet, pour autant que le modèle mélange classe bien les rachats, les non-rachats sont très mal prédits d'où les taux d'erreur sont très élevés.



Ces graphes de validation démontrent la mauvaise qualité d'ajustement des deux modèles. En effet, le modèle mélange sous-estime fortement les probabilités de racheter. Aucun des deux modèles ne s'ajuste bien aux données. Néanmoins, la courbe d'évolution des rachats prédits pour le modèle mélange suit la même trajectoire que celle des rachats observés même si les taux sont fortement sous-estimés. Ceci nécessite alors que l'on s'interroge sur la pertinence de la démarche adoptée pour implémenter le modèle depuis la délimitation du portefeuille jusqu'à la sélection des variables et de leur nombre.

Finalement, nous pouvons dire que pour notre portefeuille d'épargne, les comportements de rachat sont mieux prédits par la régression logistique dynamique que par le modèle mélange de logit. Les variables structurelles comme la fiscalité, le niveau de richesse de l'assuré ainsi que la fréquence de paiement des primes impactent fortement les comportements de rachat. Ces derniers deviennent difficiles à modéliser lorsqu'on prend en compte des variables conjoncturelles et qu'ils deviennent assez personnalisés.

---

# Conclusion

---

L'objectif de ce mémoire a été d'analyser, de comprendre, d'expliquer et de prédire les comportements de rachat du portefeuille d'épargne de la compagnie, ce qui passe d'abord par l'identification des facteurs de risque susceptibles d'influencer la décision de rachat et qui se poursuit par la mesure de leurs effets.

Des statistiques descriptives ont permis d'identifier les facteurs de risque à prendre en compte dans la modélisation. Pour cette dernière, intuitivement, ce sont les modèles de régression qui s'imposent, que ce soient des GLM ou des modèles de durée. Notre variable d'intérêt étant dichotomique, c'est une régression logistique qui a été utilisée en premier lieu avant l'introduction d'un modèle mélange de régressions logistiques. Un volet entier a été dédié à la sélection des variables à introduire dans les modèles pour réduire la dimension des variables explicatives.

Il découle de l'analyse et de la modélisation que certaines variables à effets structurels comme l'Ancienneté, le niveau de richesse avec la variable Prime Investie et la fréquence de paiement des primes influencent fortement la décision de rachat des assurés. La recherche d'un meilleur rendement par le biais des taux servis a aussi un fort impact sur les comportements de rachat des assurés. La prise en compte des variables conjoncturelles est problématique au sens où elles sont difficilement modélisables. Le modèle mélange de logit, introduit pour la première fois dans un contexte de modélisation des rachats par Xavier Milhaud dans [7] n'a pas fourni de résultats satisfaisants pour notre portefeuille. Nous nous attendions à ce qu'il différencie notre portefeuille en classes homogènes composées d'assurés qui se comporteraient de la même façon selon l'évolution de certaines variables afin de nous permettre de capturer les effets de certains facteurs de risque. Ce qui démontre encore une fois toute la difficulté de modéliser un comportement humain qui relève parfois de facteurs cachés. Cela nous amène à nous interroger sur la pertinence des variables choisies. Avons-nous pris en compte des facteurs représentatifs des tendances de notre portefeuille surtout en ce qui concerne les variables conjoncturelles ? On ne saurait affirmer si les résultats seraient pareils si on avait considéré par exemple comme une variable conjoncturelle le taux des obligations d'Etat 10 ans. D'autres facteurs difficiles à mesurer comme la réputation de la compagnie, la relation entre les courtiers qui sont les principaux intermédiaires d'Optimum Vie et les clients et la réelle satisfaction des clients auraient pu aussi nous édifier sur les comportements de rachat des assurés. Par ailleurs, notre méthode de sélection de variables a-t-elle été pertinente ? En effet, une bonne méthode de sélection des variables améliore notre connaissance du phénomène de causalité entre les variables explicatives et la variable d'intérêt. La

sélection d'un nombre de variables réduit aurait pu améliorer la qualité de prédiction des différents modèles.

Au-delà des caractéristiques propres à l'assuré et au contrat et des variables économiques et financières, il convient de nos jours d'intégrer dans la modélisation des comportements de rachat les nouvelles contraintes réglementaires. Avec une flat tax de 30% pour les contrats d'assurance-vie de plus de 150 000 €, une probable augmentation des contributions sociales (passage du taux fixe de 15,5 % à 17,2 % ) et une disposition de la loi Sapin 2 qui prévoit dans certains cas exceptionnels de limiter, voir bloquer tout versement, arbitrage et retrait sur les contrats d'assurance-vie pendant une période maximum de 6 mois, l'impact sur les comportements de rachat des assurés est plus qu'apparent et significatif. En effet, si la flat tax à 30 % encourage les sorties avant 4 ans rendant ainsi l'épargne longue moins avantageuse, si l'augmentation des prélèvements sociaux dissuade les agents à effectuer des rachats partiels ou totaux et qu'en plus la loi Sapin 2 limite leur souplesse vis-à-vis de leur option de rachat, les assureurs devront s'attendre selon la composition de leur portefeuille à un renversement des courbes de tendance en ce qui concerne les comportements de rachat des assurés.

---

# Bibliographie

---

- [1] ROUVIERE LAURENT (2015) *Régression Logistique avec R*, Université Rennes 2, UFR Sciences Sociales
- [2] RAKOTOMALA RICCO (2017) *Pratique de la Régression Logistique*, Université Lumière Lyon 2
- [3] MILHAUD XAVIER (2012) *Exogenous and endogenous risk factors management to predict surrender behaviors*, Université Claude Bernard Lyon 1, Institut de Science Financière et d'Assurances
- [4] LOURME ALEXANDRE (2011) *Contribution à la Classification par Modèles de Mélange et Classification Simultanée d'Echantillons d'Origines Multiples*, Université Lille 1
- [5] LEBARBIER E., MARY-HUARD T. (2012) *Classification non Supervisée*, AgroParisTech
- [6] SANTOS FREDERIC (2015) *L'algorithme EM : une courte présentation*, Université de Bordeaux
- [7] MILHAUD XAVIER (2012) *Mélanges de GLMs et nombre de composantes : application au risque de rachat en Assurance Vie*, Université Claude Bernard Lyon 1
- [8] CHEN HWEI-MING, YANG Miin-SHEN (2002) *Fuzzy Class Logistic Regression Analysis*, Department of Applied Mathematics, Chung Yuan Christian University
- [9] GRUN BETTINA, LEISCH FRIEDRICH (2007) *Applications of finite mixtures of regression models*, Computational Statistics & Data Analysis, 51(11) :5247-5252
- [10] LEISCH FRIEDRICH (2004) *A General Framework for Finite Mixture Models and Latent Class Regression in R*, Journal of Statistical Software, 11(8), 1-18, October 2004
- [11] OHNOUNA Emmanuel (2008) *Evaluation Best Estimate de contrats d'épargne en euros*, Mémoire d'actuaire, Université Louis Pasteur Strasbourg
- [12] LADIAS NICOLAS (2012) *Analyse des causes de rachats sur des contrats d'assurance vie*, Mémoire d'actuaire, Institut de Science Financière et d'Assurances
- [13] SEROR DEBORAH, YENG DAVID (2013) *Modélisation des comportements de rachats dans un cadre de risques compétitifs*, Mémoire d'actuaire, ENSAE Paris Tech
- [14] RAKAH NAFOULAH (2012) *Modélisation des rachats dans les contrats d'épargne*, Centre des Etudes Actuarielles

- [15] GRILLI L., RAMPICHINI C., VARRIALE R. (2013) *Binomial mixture modelling of university credits*, Communications in Statistics - Theory and Methods

# Annexe

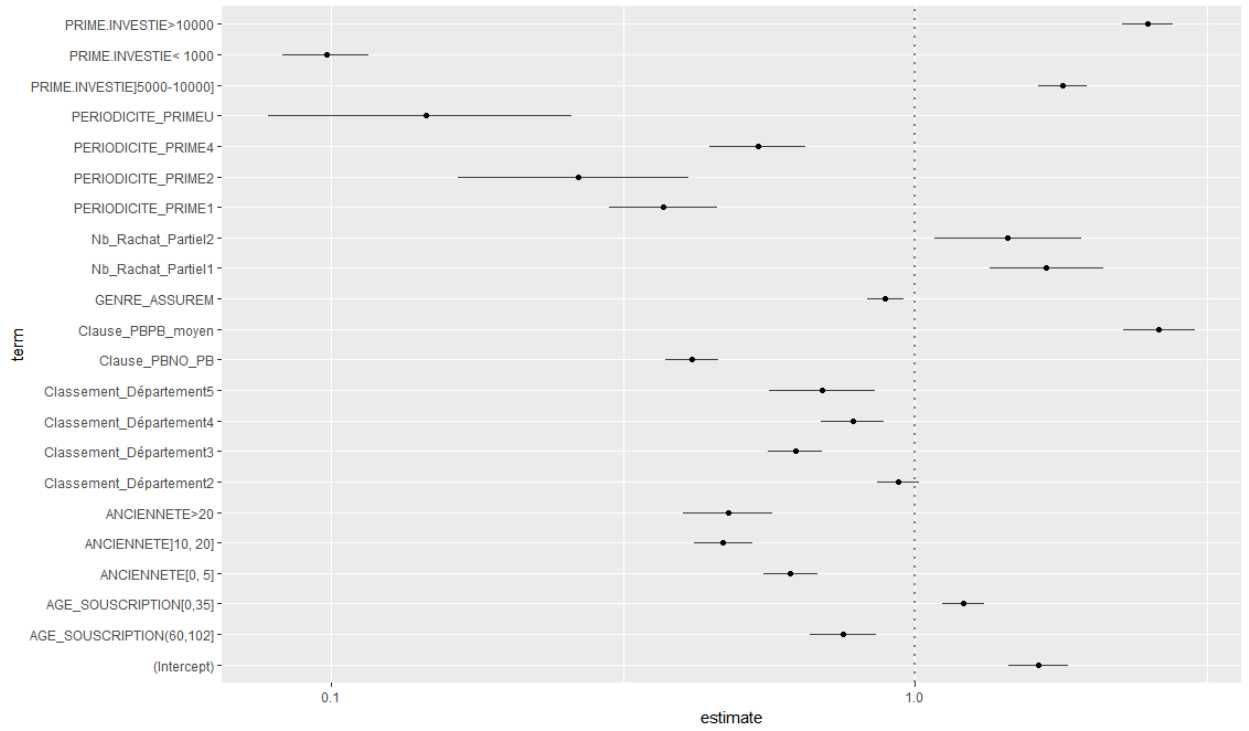


FIGURE 3.4 – Représentation graphique des odds-ratio pour la régression logistique sans les variables conjoncturelles

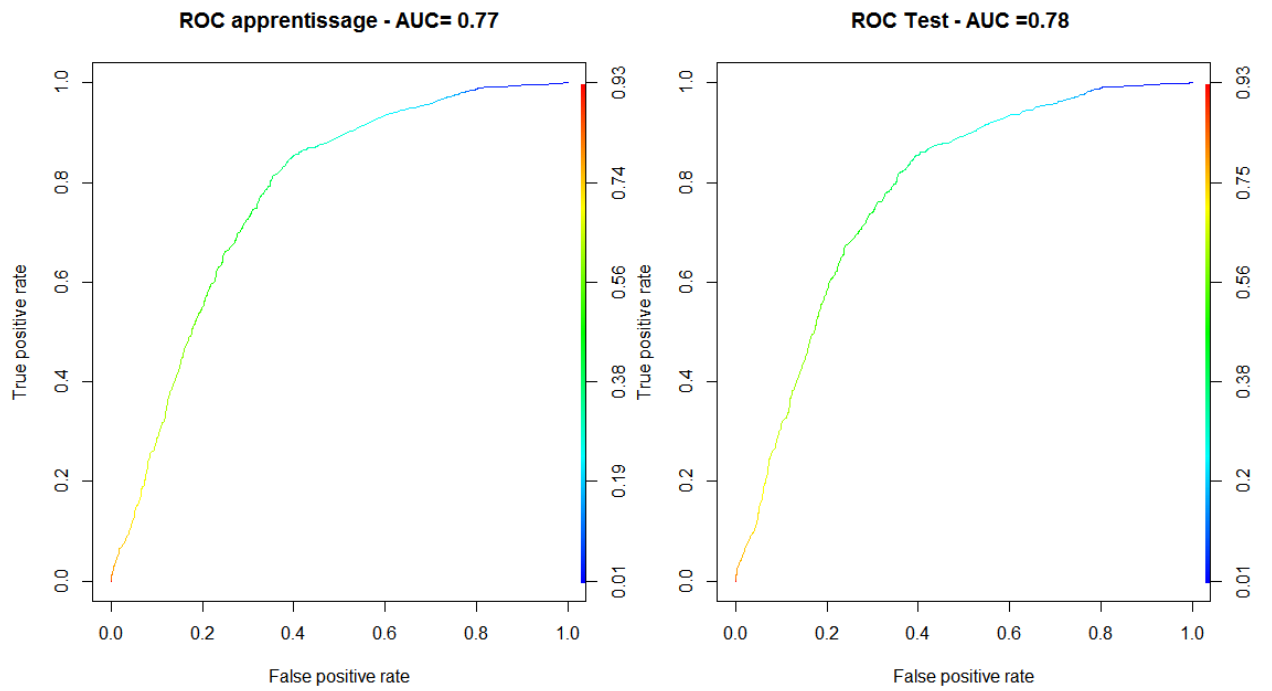


FIGURE 3.5 – Courbes ROC des échantillons apprentissage et test pour la régression logistique sans les variables conjoncturelles

	Estimate	Std. Error	z value	Pr (> z )	Indc
<b>(Intercept)</b>	2.8921839	0.0818005	35.3565	< 2.2e-16	***
<b>AGE_SOUSCRIPTION (35, 60]</b>	(0 référence)				
<b>AGE_SOUSCRIPTION[0, 35]</b>	0.0281921	0.031289	0.901	0.36758	
<b>AGE_SOUSCRIPTION(60, 102]</b>	-0.094297	0.0440491	-2.1407	0.0323	*
<b>ANCIENNETE ]5, 10]</b>	(0 référence)				
<b>ANCIENNETE[0, 5]</b>	-0.5015624	0.0454694	-11.0308	< 2.2e-16	***
<b>ANCIENNETE]10, 20]</b>	-0.3893053	0.0500598	-7.7768	7.44E-15	***
<b>ANCIENNETE&gt;20</b>	-0.6202985	0.0736912	-8.4175	< 2.2e-16	***
<b>PRIME.INVESTIE ]1000-5000]</b>	(0 référence)				
<b>PRIME.INVESTIE&lt; 1000</b>	-0.7204798	0.035183	-20.478	< 2.2e-16	***
<b>PRIME.INVESTIE]5000-10000]</b>	0.2762917	0.0401672	6.8785	6.05E-12	***
<b>PRIME.INVESTIE&gt;10000</b>	0.2638309	0.0414367	6.3671	1.93E-10	***
<b>PERIODICITE_PRIME12</b>	(0 référence)				
<b>PERIODICITE_PRIME1</b>	-0.5643601	0.0782605	-7.2113	5.54E-13	***
<b>PERIODICITE_PRIME2</b>	-1.0702443	0.1194125	-8.9626	< 2.2e-16	***
<b>PERIODICITE_PRIME4</b>	-0.8296472	0.0725468	-11.436	< 2.2e-16	***
<b>PERIODICITE_PRIMEU</b>	-1.0328761	0.2188536	-4.7195	2.36E-06	***
<b>Classes_TxServistaux_moyen</b>	(0 référence)				
<b>Classes_TxServisN0_tx</b>	-0.1087224	0.0505246	-2.1519	0.03141	*
<b>Classes_TxServistaux_élevé</b>	0.4083951	0.0628372	6.4993	8.07E-11	***
<b>Classes_TxServistaux_faible</b>	-0.2494087	0.0484087	-5.1522	2.58E-07	***
<b>Nb_Rachat_Partie10</b>	(0 référence)				
<b>Nb_Rachat_Partie11</b>	0.187318	0.0998135	1.8767	0.06056	.
<b>Nb_Rachat_Partie12</b>	-0.0059696	0.134659	-0.0443	0.96464	
<b>Classes_SAISSONNALITEfin_d'annee</b>	(0 référence)				
<b>Classes_SAISSONNALITEcourant</b>	-0.5267668	0.0670998	-7.8505	4.14E-15	***
<b>Classes_SAISSONNALITEdébut_d'année</b>	-0.0394025	0.0698255	-0.5643	0.57255	
<b>Classes_SAISSONNALITEété</b>	0.1884901	0.06701	2.8129	0.00491	**
<b>Variation_TMEBL</b>	(0 référence)				
<b>Variation_TMEB</b>	-0.9197643	0.0461821	-19.916	< 2.2e-16	***
<b>Variation_TMEH</b>	0.0794179	0.0666086	1.1923	0.23314	
<b>Variation_TMEHL</b>	-0.1668168	0.0588085	-2.8366	0.00456	**
<b>Variation_TMES</b>	0.4489506	0.1801243	2.4924	0.01269	*
<b>Variation_CAC40H</b>	(0 référence)				
<b>Variation_CAC40B</b>	-0.4626635	0.042796	-10.8109	< 2.2e-16	***
<b>Variation_CAC40BL</b>	0.1303799	0.1064218	1.2251	0.22053	
<b>Variation_CAC40HL</b>	-0.0855435	0.0706151	-1.2114	0.22574	
<b>Variation_ChomageH</b>	(0 référence)				
<b>Variation_ChomageB</b>	-0.1322911	0.0565482	-2.3394	0.01931	*
<b>Variation_ChomageS</b>	-0.0109182	0.0849473	-0.1285	0.89773	

Tableau 1: Sortie R du modèle mélange à une composante avec tous les variables explicatives

	Estimate	Std. Error	z value	Pr(> z )	Indc
(Intercept)	3.51401	0.12199	28.806	< 2e-16	***
AGE_SOUSCRIPTION (35, 60]	0 (référence)				
AGE_SOUSCRIPTION[0,35]	0.07698	0.05402	1.425	0.154139	
AGE_SOUSCRIPTION(60,102]	-0.50767	0.08747	-5.804	6.48E-09	***
ANCIENNETE ]5, 10]	0 (référence)				
ANCIENNETE[0, 5]	-0.81673	0.06777	-12.051	< 2e-16	***
ANCIENNETE]10, 20]	-0.93763	0.07793	-12.032	< 2e-16	***
ANCIENNETE>20	-1.2823	0.11163	-11.487	< 2e-16	***
PRIME.INVESTIE [1000-5000]	0 (référence)				
PRIME.INVESTIE< 1000	-3.38314	0.10026	-33.743	< 2e-16	***
PRIME.INVESTIE]5000-10000]	0.6546	0.06339	10.327	< 2e-16	***
PRIME.INVESTIE>10000	0.77867	0.06673	11.669	< 2e-16	***
PERIODICITE_PRIME12	0 (référence)				
PERIODICITE_PRIME1	-1.46212	0.13668	-10.698	< 2e-16	***
PERIODICITE_PRIME2	-2.65184	0.23508	-11.28	< 2e-16	***
PERIODICITE_PRIME4	-1.80377	0.11943	-15.103	< 2e-16	***
PERIODICITE_PRIMEU	-3.2708	0.35783	-9.141	< 2e-16	***
Classes_TxServistaux_moyen	0 (référence)				
Classes_TxServisNO_tx	-0.59193	0.07592	-7.796	6.37E-15	***
Classes_TxServistaux_élevé	0.80345	0.08923	9.004	< 2e-16	***
Classes_TxServistaux_faible	-0.39142	0.07267	-5.386	7.18E-08	***
Nb_Rachat_Partiel0	0 (référence)				
Nb_Rachat_Partiel1	0.45077	0.14597	3.088	0.002014	**
Nb_Rachat_Partiel2	0.37244	0.1889	1.972	0.048653	*
Classes_SAISONNALITEfin_d'annee	0 (référence)				
Classes_SAISONNALITEcourant	-0.97022	0.09785	-9.916	< 2e-16	***
Classes_SAISONNALITEdébut_d'année	0.07781	0.09991	0.779	0.43614	
Classes_SAISONNALITEété	0.54527	0.09453	5.768	8.02E-09	***
Variation_TMEBL	0 (référence)				
Variation_TMEB	-2.08361	0.06843	-30.45	< 2e-16	***
Variation_TMEH	0.12563	0.09529	1.318	0.187363	
Variation_TMEHL	-0.31266	0.08449	-3.7	0.000215	***
Variation_TMES	0.01063	0.22245	0.048	0.961899	
Variation_CAC40H	0 (référence)				
Variation_CAC40B	-1.12519	0.06232	-18.054	< 2e-16	***
Variation_CAC40BL	-0.0497	0.15058	-0.33	0.741344	
Variation_CAC40HL	-0.15258	0.10107	-1.51	0.131119	
Variation_ChomageH	0 (référence)				
Variation_ChomageB	-0.30462	0.08139	-3.743	0.000182	***
Variation_ChomageS	0.0721	0.1207	0.597	0.550254	

Tableau 2 : Sortie R du modèle logit dynamique avec toutes les variables explicatives

	prior	size	post>0	ratio
Comp.1	0.503	16742	16862	0.99288
Comp.2	0.497	120	16862	0.00712

FIGURE 3.7 – Résultat essai modèle mélange à deux composantes avec certaines variables fixées

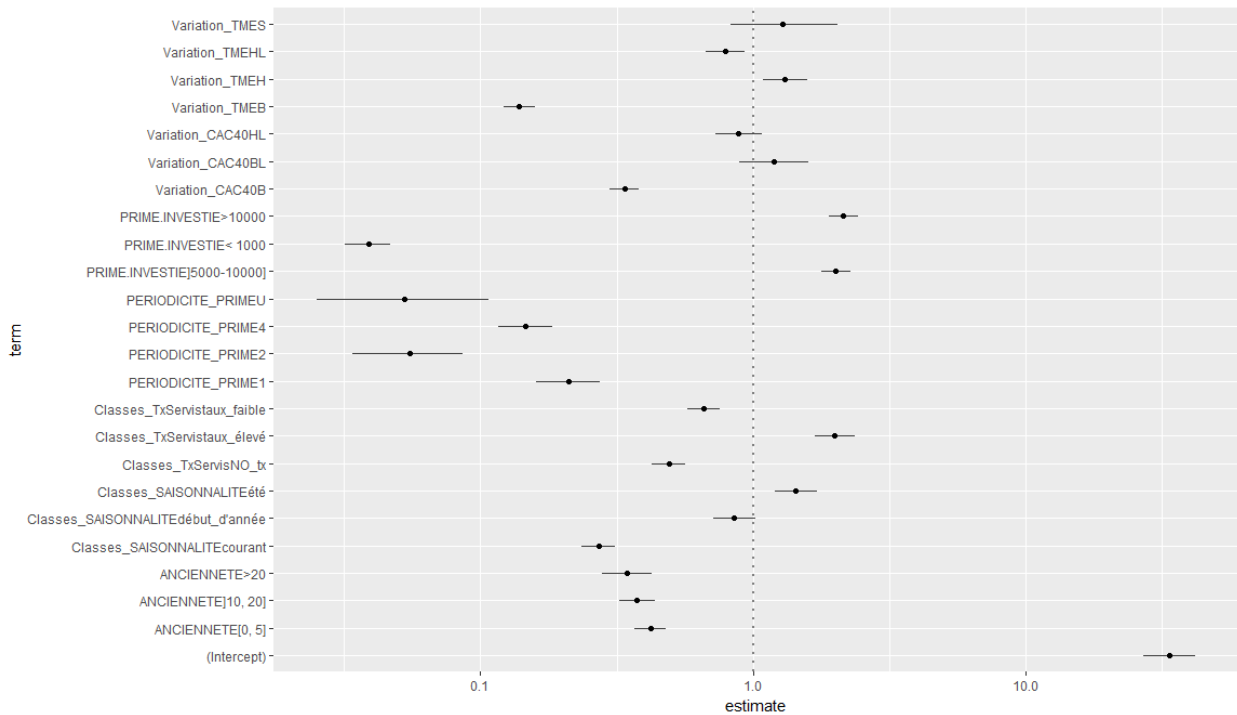


FIGURE 3.6 – Graphe des odds-ratio du modèle logit dynamique retenu

Aire sous la courbe ROC	LR	MLR
apprentissage	0.9086633	0.9019293
test	0.9110612	0.9053417

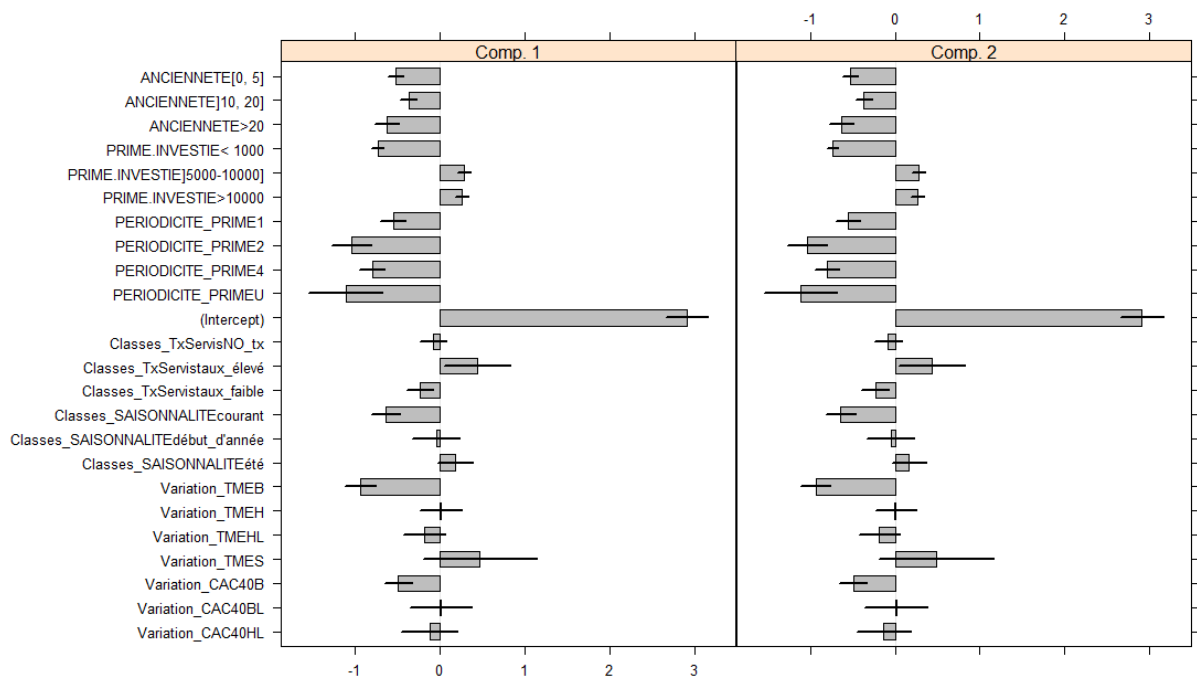


FIGURE 3.8 – Coefficients de régression du modèle mélange à deux composantes