

**Mémoire présenté devant le Master Droit économie et gestion
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 10/02/2021**

Par : Marie-Amélie Criqui

Titre : Le rachat total des contrats d'assurance-vie et de capitalisation :
comportement des assurés d'un portefeuille haut de gamme

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

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

Présidente du jury

Signatures

Entreprise :

Sandrine LEMERY

Nom : BNP Paribas Cardif

Directeur de mémoire en entreprise :

Nom : Matthieu Goas

Membres présents du jury de l'Institut
des Actuaires

Edith BOCQUAIRE

Nicolas BOURE

François-Xavier NEGRI

Florence PICARD

Sidick SIMRICK

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



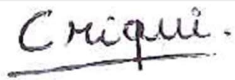
Membres présents du jury du Master
Mention Actuariat du CNAM

Olivier DESMETTRE

David FAURE

François WEISS

Signature du candidat



Secrétariat : Ghislaine ERNY

RESUME	4
ABSTRACT.....	4
REMERCIEMENTS	5
INTRODUCTION.....	6
PARTIE I : LES CONTRATS D'ASSURANCE-VIE ET DE CAPITALISATION.....	10
I LES DIFFERENTS TYPES DE PLACEMENTS	10
I.1 LE FONDS EN EUROS	10
I.2 LES SUPPORTS EN UNITES DE COMPTE OU « UC ».....	11
I.3 LE FONDS EURO-CROISSANCE/CROISSANCE	11
II LE FONCTIONNEMENT EN COURS DE VIE DU CONTRAT	17
II.1 LES VERSEMENTS.....	17
II.2 LES ARBITRAGES.....	17
II.3 LES CHARGEMENTS ET LES FRAIS.....	17
II.4 LES GARANTIES COMPLEMENTAIRES EN CAS DE DECES.....	18
III LE DENOUEMENT DU CONTRAT	19
III.1 LA SORTIE EN RENTE VIAGERE	19
III.2 LE DECES	20
III.3 LE RACHAT	21
PARTIE II : CARACTERISTIQUES DU PORTEFEUILLE ETUDIE ET SELECTION DES VARIABLES EXPLICATIVES.....	25
I STATISTIQUES DESCRIPTIVES DU PORTEFEUILLE ETUDIE	25
I.1 BASE DE DONNEES.....	25
I.2 STATISTIQUES DESCRIPTIVES.....	26
II ANALYSE DES CORRESPONDANCES MULTIPLES (ACM).....	42
II.1 PRESENTATION DU MODELE	42
II.2 DETERMINATION DU MODELE	47
II.3 RESULTATS DU MODELE	54
III LES ARBRES DE CLASSIFICATION.....	56
III.1 PRESENTATION DU MODELE	56
III.2 PRINCIPAUX RESULTATS	60
IV L'ALGORITHME FORETS ALEATOIRES (RANDOM FOREST)	61
IV.1 PRESENTATION DU MODELE	61
IV.2 SELECTION DES PARAMETRES DE LA FORET ALEATOIRE.....	62
IV.3 MESURE DE L'IMPORTANCE DES VARIABLES.....	66
V LA SELECTION DE VARIABLES VIA LA REGRESSION LOGISTIQUE	68
V.1 PRESENTATION DU MODELE	68
V.2 VIA LA SIGNIFICATIVITE DES COEFFICIENTS (TEST DE WALD).....	71
V.3 LA SELECTION PAR OPTIMISATION	72
VI SYNTHESE ET SELECTION DES VARIABLES	75
PARTIE III : VALIDATION DU MODELE DE REGRESSION LOGISTIQUE ET DE SON POUVOIR PREDICTIF.....	76
I RISQUE RELATIF ET ODDS-RATIO	76
I.1 DEFINITIONS.....	76

I.2	INTERPRETATION DES COEFFICIENTS	77
II	EVALUATION DU MODELE DE REGRESSION LOGISTIQUE.....	78
II.1	MATRICE DE CONFUSION ET PSEUDO-R ²	78
II.2	LA COURBE ROC	81
II.3	SYNTHESE.....	85
III	APPLICATION DU MODELE DE REGRESSION LOGISTIQUE SUR L'ANNEE 2016	87
III.1	PREMIERS RESULTATS.....	87
III.2	COMPARAISON AVEC LA METHODE ACTUELLE.....	87
IV	OUTIL DE PILOTAGE.....	89
	CONCLUSION.....	94
	BIBLIOGRAPHIE.....	96
	ANNEXES	97

Résumé

Il est stratégique, pour le pilotage d'activité d'une entreprise d'assurance de comprendre le comportement des assurés afin de pouvoir les anticiper. Ce mémoire a pour objectif d'étudier plus précisément le comportement des assurés pour les rachats totaux sur un portefeuille de contrats d'assurance-vie et de capitalisation. Le rachat trouve son origine :

- soit dans l'insatisfaction de l'assuré face à la revalorisation de son contrat, on parle alors de rachat *conjoncturel*,
- dans la situation unique de l'assuré, on parle alors de rachat *structurel*.

L'étude qui suit s'intéresse uniquement aux rachats structurels. L'objectif consiste à identifier des critères qui peuvent donner de l'information sur le comportement de rachat de l'assuré. Les résultats obtenus pourront ensuite servir à la segmentation des lois de rachats mais également à la création d'un outil fournissant « le risque » de rachat structurel anticipé. Cet outil pourra être une aide à la décision lors d'une réponse à un appel d'offre pour une souscription.

Afin d'identifier les différentes variables potentiellement discriminantes, nous allons analyser trois méthodes statistiques différentes et voir si elles aboutissent à des résultats similaires. En guise de validation, les modèles obtenus seront ensuite appliqués à des données n'ayant pas servi à la réalisation des modèles. La conclusion de ce mémoire viendra synthétiser l'apport de cette étude ainsi que ses limites d'applications et les potentiels travaux à venir.

Mots clés : Rachat structurel, Comportement d'assuré, épargne, Analyse des Correspondance Multiples, Arbre de Classification, Forêts Aléatoires, Régression Logistique.

Abstract

It is a priority, for an insurance company, to understand the policyholders' behavior so that it can be foreseen. This study focuses on policyholders' behavior regarding total surrenders from a sample of saving contracts.

Surrender can be triggered by:

- The policyholder's deception regarding the contract valorization: we call it dynamic surrender,
- The policyholder's specific situation: we call it structural surrender.

We have decided to focus on structural surrender only. Our aim is to identify criteria which could give us information re policyholder's behavior that leads him to redeem his savings. Our results will help us segment redemptions according to specific criteria and create a surrender risk tool. This tool will help decide or not to take on board a new client, through the RFI process.

In order to identify the variables, we will focus on three statistic methodologies and determinate if they lead to the same conclusions. In order to confirm our conclusions, the models will be applied to data that have not been used for the building of the models. Our study conclusion will wrap up the study input, the limits of its concrete application, the studies that still need to be lead.

Key words : Structural surrender, policyholders' behavior, savings, Multiple Correspondence Analysis, Classification Tree, Random Forest, Logistic Regression

Remerciements

Je tiens à remercier l'équipe d'actuariat épargne individuelle de BNP Paribas Cardif. Merci à Alexandra et Stéphanie de m'avoir fait confiance et qui grâce à elle j'ai pu approfondir mes connaissances d'actuariat.

Je remercie chaleureusement mes relecteurs notamment Emmanuelle, Matthieu et Delphine pour leurs remarques avisées.

Mes remerciements s'adressent également à l'ensemble du corps professoral de la formation d'actuariat du CNAM.

Je n'oublie pas mon mari Marc-Antoine, ma famille et mes amis pour leur soutien et leur patience.

Merci à tous et à toutes,

Introduction

Afin de garantir l'équilibre entre ses ressources et ses emplois et d'évaluer précisément les risques pris, l'assureur doit avoir une bonne connaissance des différents types de comportements des assurés. En effet, ces comportements ont une influence directe sur la collecte nette.

La collecte nette est le résultat de la différence entre les cotisations perçues et les prestations versées. **Il est donc primordial pour l'assureur de maîtriser ces deux facteurs : les cotisations perçues / les prestations servies.**

Concernant les cotisations perçues : L'assurance-vie est le placement préféré des Français : les cotisations perçues dans ce secteur ont en effet augmenté d'environ 5% entre 2014 et 2015. Ce produit ne semble pas avoir de concurrent pouvant allier comme lui rendement/fiscalité avantageuse/liquidité¹ de l'épargne placée (principalement pour le fonds en euros).

Malgré une augmentation des cotisations en unités de compte en 2015 (+32% par rapport à 2014), ces supports représentent « uniquement » 20% des cotisations de l'année : les 80% restants sont investis sur le fonds en euros² qui en raison de son taux garanti, est aujourd'hui le produit d'appel des assureurs.

Or, dans le contexte actuel de taux bas, une collecte brute importante sur le fonds en euros a pour conséquence d'augmenter le risque de liquidité pour l'assureur en cas de hausse des taux. En effet, l'investissement de ces nouvelles primes est confronté, du fait d'un environnement de taux bas, à un rendement obligataire dégradé. La croissance des encours en euros conduit de plus à diluer les plus-values latentes existantes sur ce périmètre. Dans la mesure où la participation aux bénéfices attribuée aux assurés décroît généralement moins vite que le rendement de l'actif général des assureurs-vie, le maintien de taux durablement bas perturbe l'équilibre financier des organismes.

En cas de remontée des taux d'intérêts, l'inertie du stock de placements progressivement accumulé pourrait ralentir la relance des performances financières, réduisant l'attractivité des fonds en euros existants. Ces contrats pouvant être rachetés à tout instant et leur valeur de rachat garantie, un scénario de sortie de masse d'encours est envisageable, qui pourrait peser très lourdement sur la liquidité de certains assureurs. Dans le contexte économique actuel, une forte collecte sur le fonds en euros n'est donc pas souhaitable pour les assureurs. L'ACPR leur recommande d'ailleurs de limiter les risques pris en « *adaptant, en permanence et de manière suffisante, leur politique commerciale et leur gestion, dans le respect des règles de protection de la clientèle.* ».

Il est très délicat, commercialement parlant, d'interdire un versement sur un fonds en euros. Cependant, du fait de la pérennisation des taux bas, certains assureurs ont mis en place une politique d'encadrement des versements complémentaires sur leur portefeuille comme par exemple un taux minimal d'investissement en unités de compte ou un montant maximal d'investissement sur le fonds en euros.

Afin de permettre une plus grande diversification de l'encours des produits d'assurance-vie, la présence d'un produit euro-diversifié s'est généralisée depuis le 26/06/2014 sur tous les contrats d'assurance-vie existants, via les supports euro-croissance³.

¹ Liquidité qui peut être altérée si activation des clauses de l'article L631-2-1 du code monétaire et financier introduit par la loi relative à la transparence, à la lutte contre la corruption et à la modernisation de la vie économique dite la loi « Sapin II ».

² Source FFA : http://www.ffsa.fr/sites/jcms/p1_1628124/fr/assurance-vie-collecte-nette-positive-en-2015?cc=fn_7348 Le support Euro croissance n'est pas pris en compte dans les statistiques fournies.

³ Ou support croissance lorsque la garantie à l'échéance est inférieure à 100% des primes nettes investies.

Ces supports proposent une garantie en capital, uniquement à l'échéance choisie à la souscription, égale à un pourcentage des cotisations versées sur le support, ce qui permet d'espérer des rendements à terme plus élevés qu'un support en euros, tout en présentant un moindre risque qu'un support en unités de compte (où la valeur de rachat n'est garantie à aucun moment).

L'assureur pourra ainsi proposer à ses clients des contrats d'assurance-vie à « trois branches » :

- Un fonds en euros,
- Des supports en unités de compte,
- Un fonds euro-croissance/croissance.

Or, encore une fois, le contexte actuel de taux bas limite le développement du support euro-croissance. En effet, le montant des investissements obligataires permettant d'assurer une garantie à 100%, au terme de 8 ans, est supérieur à 90% des nouveaux versements. Cela compromet la performance attendue des fonds et limite singulièrement les cas où l'orientation de l'épargne vers l'euro-croissance relèverait aujourd'hui, d'un conseil commercial opportun.

Pour relancer le développement de ce support, un décret est entré en vigueur le 13 juillet 2016, proposant une possibilité temporaire de transfert des plus-values latentes des fonds « euros » vers les fonds « euro-croissance ». A l'heure où nous rédigeons ce mémoire, nous manquons encore de recul sur l'efficacité de cette disposition.

Enfin, en ce qui concerne les supports en unités de compte, le distributeur peut rapidement se heurter à une problématique de devoir de conseil lorsqu'il s'agit de vouloir augmenter la part des investissements des clients vers les unités de compte.

Concernant les prestations servies :

Les prestations servies par un contrat d'assurance-vie mixte peuvent prendre trois formes distinctes :

- **La rente viagère,**
- **Le capital décès aux bénéficiaires,**
- **Le rachat partiel/total du contrat**

Contrairement à l'assurance non-vie où les primes sont versées pour couvrir un risque qui pourrait arriver selon une probabilité non nulle, l'essence même du contrat d'assurance-vie mixte est de reverser systématiquement un capital à l'heure du dénouement du contrat. Passer par la case « prestation » est inéluctable⁴. L'assureur a donc un intérêt stratégique de maîtriser le montant annuel probable de ces prestations futures afin d'établir la meilleure politique de gestion d'actif possible.

Même si la sortie en rente viagère est une possibilité prévue dans les conditions générales des contrats du portefeuille étudié dans ce mémoire, celle-ci est choisie de manière anecdotique par les assurés. Cela est dû principalement à une fiscalité dissuasive et à la crainte des assurés de se voir aliéner leur capital. Pourtant, la rente est censée répondre au besoin de l'assuré d'avoir un revenu de remplacement (par exemple : pour pallier la diminution de ses revenus suite à la retraite). En ce qui concerne l'assureur, la sortie en rente par l'assuré lui permet de dégager un potentiel bénéfique technique (si l'assuré décède avant la date probable de fin d'engagement estimée par l'assureur) contrairement à une sortie totale ou par rachats partiels programmés.

La sortie sous forme de capital décès est souvent présentée comme un avantage commercial des contrats d'assurance-vie car elle permet de répondre notamment à un besoin de transmission du patrimoine. Par ailleurs, la loi Eckert entrée en vigueur le 01/01/2016, qui prévoit une revalorisation

⁴ Elle l'est d'autant plus depuis l'obligation de transférer à la Caisse des dépôts et consignation le capital des contrats dits en déshérence.

minimale des capitaux décès ainsi que le transfert des capitaux à la Caisse des dépôts et consignations en cas de déshérence, incite l'assureur à rechercher plus activement les bénéficiaires d'assurés décédés. Cependant, le portefeuille d'Assurance Epargne Pension (AEP), sur lequel porte l'étude, est relativement « jeune » et comprend un faible nombre d'assurés : la réalisation d'une loi de mortalité reflétant au plus juste le comportement des assurés n'est pas aisée.

Enfin, la sortie sous forme de rachat partiel/total constitue la principale charge de prestations annuelles du portefeuille étudié et du marché de l'assurance-vie : ces derniers représentent plus de 60%⁵ des charges de prestations annuelles des assureurs. Le cadre d'étude de ce mémoire portera donc sur les rachats constatés sur le portefeuille étudié.

La littérature distingue généralement deux formes de rachats : les rachats « conjoncturels » et les rachats « structurels ». Les rachats « conjoncturels » sont le résultat d'une insatisfaction de l'assuré face au rendement servi par l'assureur, et d'une espérance de rendement chez un concurrent. Les rachats « structurels » répondent à des besoins/envies propres à l'assuré qui ne sont pas liés à l'environnement économique ni à la stratégie de revalorisation de l'assureur.

L'objectif de ce mémoire est de déterminer les variables les plus influentes pour déterminer une segmentation du portefeuille des assurés selon leurs comportements de rachats totaux et ensuite de proposer une modélisation des taux de rachats structurels constatés. L'utilisation principale souhaitée de cette modélisation est d'alimenter les réflexions commerciales sur le comportement des assurés afin d'identifier les mesures à mettre en place ou de cibler une clientèle précise à la souscription. Ainsi, la finalité de ce mémoire étant d'étudier les disparités qui peuvent exister selon les assurés du portefeuille, la réplique « parfaite » des comportements n'est pas nécessaire pour y parvenir. Par ailleurs, l'étude se concentre sur les raisons structurelles de rachats. Les éléments pouvant motiver l'assuré à racheter de manière conjoncturelle (les rendements estimés trop faibles, plus ou moins-values constatées sur le portefeuille, etc.) ne seront donc pas étudiés. Il en est de même des variables pour lesquelles un pilotage commercial est plus délicat tel que l'âge de l'assuré. Par ailleurs, la fiscalité successorale étant différente selon les primes versées avant ou après 70 ans, le critère de l'âge de l'assuré est le plus souvent utilisé dans la modélisation des rachats conjoncturels et est donc hors-périmètre de cette étude.

Cette étude est décomposée en trois parties. Dans un premier temps, nous présenterons la fiscalité des contrats d'assurance-vie et des contrats de capitalisation ainsi que leurs spécificités. La deuxième partie présentera le portefeuille étudié et les variables discriminantes déterminées. Enfin, nous étudierons l'impact des variables sur le comportement de rachat ainsi que la robustesse des modèles réalisés sur une année complète d'observation. Ainsi, nous étudierons la possibilité de réaliser un « *scoring* » à la souscription.

Remarque : La loi relative à la transparence, à la lutte contre la corruption et à la modernisation de la vie économique, dite loi « Sapin II », est entrée en vigueur le 11/12/2016. Un article⁶ de cette loi introduit un renforcement des pouvoirs du Haut Conseil de Stabilité Financière (HCSF). Il confère au HCSF le pouvoir de prendre des mesures conservatoires afin de préserver la stabilité du système financier ou de prévenir des risques présentant une menace grave et caractérisée pour la situation

⁵ Sources ACPR : Les chiffres du marché français de la banque et de l'assurance 2016

⁶ L.631-2-1 du Code monétaire et financier

financière de l'ensemble ou d'un sous-ensemble significatif d'organismes du secteur de l'assurance. Le HSCF peut alors, pour une durée de 3 mois renouvelables :

- Limiter temporairement l'exercice de certaines opérations ou activités, y compris l'acceptation de primes ou de versements ;
- Restreindre temporairement la libre disposition de tout ou partie des actifs ;
- Limiter temporairement, pour tout ou partie du portefeuille, le paiement des valeurs de rachat. Cette limitation ne peut pas faire l'objet d'une limitation temporaire de plus de 6 mois consécutifs ;
- Retarder ou limiter temporairement, pour tout ou partie du portefeuille, la faculté d'arbitrage ou le versement d'avances.

Cette limitation possible sur les rachats pourrait avoir un effet sur le comportement des assurés eux-mêmes et également sur la part des rachats « conjoncturels ». Cependant, l'impact de cette nouvelle réglementation ne sera pas étudié dans le cadre de ce mémoire faute de recul par rapport à sa date d'entrée en vigueur.

Partie I : Les contrats d'assurance-vie et de capitalisation

A toutes fins utiles, il convient tout d'abord de distinguer les contrats de capitalisation des contrats d'assurance-vie. En effet, l'évènement aléatoire lié à la durée de vie humaine ne s'applique pas sur les contrats de capitalisation. Le contrat de capitalisation est un contrat d'assurance sans contre-assurance décès. Malgré cette différence lors du dénouement par décès, les actes de gestion et les placements possibles sur ces deux contrats sont généralement similaires.

La notion d'assurance-vie englobe trois différents types de contrats :

- L'assurance en cas de vie : ce type de contrat permet de recevoir un capital ou une rente viagère lorsque le contrat arrive à son terme. Le terme est défini au moment de la souscription du contrat.
- L'assurance en cas de décès : ce type de contrat se dénoue également sous forme de capital ou de rente mais c'est le décès de l'assuré qui déclenche le dénouement et c'est donc le bénéficiaire désigné qui reçoit la prestation. Dans ce cas, trois types de contrat existent : l'assurance temporaire, l'assurance vie entière et les contrats obsèques.
- L'assurance mixte : ce type de contrat couvre à la fois le risque vie et le risque décès.

Dans la suite de ce mémoire, la notion d'assurance-vie fait référence aux contrats d'assurance sur la vie dits mixtes. C'est généralement à ce type de contrat qu'il est fait référence lorsque le terme d' « assurance-vie » est utilisé dans la vie courante.

I Les différents types de placements

Il existe à ce jour trois différents types de placements sur les contrats d'assurance-vie et de capitalisation.

Le contrat peut faire cohabiter en même temps ces trois types de placements – on parle alors de contrat « multi-supports » -, soit ne proposer qu'un seul type de placement – on parle alors de contrat « mono-support ».

I.1 Le fonds en euros

Ce support est libellé en euros. Les sommes qui y sont investies sont garanties à tout instant (diminuées des retraits éventuels du fonds et des éventuels frais de gestion pour les fonds en euros avec garantie brute de frais de gestion administrative). Elles capitalisent à un taux minimum annuel (fixe ou variable selon les conditions contractuelles), générant ainsi des intérêts qui sont acquis définitivement au contrat et vont capitaliser à leur tour. Ce mécanisme est appelé l'effet cliquet.

Le fonds en euros, parce que son capital est garanti à tout instant, est considéré sans risque pour l'assuré (modulo une garantie du capital brute/nette de frais de gestion administrative).

De plus l'entreprise d'assurance est tenue de faire participer les assurés aux bénéfices techniques et financiers réalisées par celle-ci (L.331-3 du Code des assurances). Le montant de participation aux bénéfices (PB) est au minimum de 90% des bénéfices techniques et 85% des résultats financiers

constatés au global de la compagnie. En revanche, la réglementation ne fixe pas les modalités de répartition entre les différents contrats. Celle-ci est donc réalisée à la discrétion de l'assureur. Les dispositions contractuelles peuvent alors prévoir un montant de PB supérieur/inférieur au taux indiqué dans la réglementation.

En contrepartie de la garantie en capital, le fonds en euros propose des rendements limités. Le contexte actuel de taux bas ne permet pas de présager une augmentation des rendements de ce support dans les prochaines années.

1.2 Les supports en unités de compte ou « UC »

Ces fonds peuvent prendre la forme d'actions, d'obligations, de parts d'OPCVM, de parts de SCPI etc. Les titres admissibles dans la catégorie « UC » du contrat sont définis par les articles L.131-1 et R.131-1 du Code des assurances.

Sur ce support, les assurés bénéficient directement du rendement des actifs financiers représentatifs des unités de compte. **La valeur de rachat fluctue donc à la hausse ou à la baisse en fonction de l'évolution des marchés financiers.** L'assureur s'engage sur le nombre d'unités de compte mais pas sur leur valeur. Le risque de perte est alors supporté par l'assuré.

En contrepartie de ce risque de placement, l'assuré peut espérer un rendement supérieur à celui proposé sur le fonds en euros.

1.3 Le fonds euro-croissance/croissance

Contrairement au fonds en euros, **la garantie des sommes investies est uniquement au terme** choisi par l'assuré qui doit être **de minimum 8 ans** (l'échéance du support peut être supérieure). Du fait de la complexité de ce support, une explication synthétique de son fonctionnement est proposée ci-dessous.

1.3.1 Engagements

La garantie au terme du contrat peut correspondre à la totalité ou à une partie des cotisations nettes de frais versées sur le support euro-croissance. Cette garantie peut également être nulle. Dans ce dernier cas, il s'agit d'un fonds représenté à 100% en provision de diversification (appelé communément « fond interne »).

Le niveau de garantie peut être fixé soit par chaque souscripteur soit dans les conditions mêmes du contrat.

Le montant garanti à terme est déterminé en fonction du taux garanti en vigueur au moment du versement sur le support euro-croissance.

Ainsi, selon le pourcentage de cotisations nettes versées garanties, la cotisation est répartie entre deux composantes :

- La provision mathématique

- La provision de diversification.

L'assureur détermine le montant de l'engagement au terme relatif à la part de la cotisation versée selon le pourcentage garanti de la cotisation, l'échéance fixée à la souscription (la durée restante jusqu'à l'échéance est notée n) et le taux garanti s'il existe :

$$Capital\ garanti = \%_{garanti} \times Primes_{nettes} \times (1 + TMG)^n$$

1.3.2 La provision mathématique

La provision mathématique (PM) correspond à la valeur actuelle de l'engagement au terme de l'investissement sur le support. Celle-ci est égale à la valeur du capital garanti au terme, actualisée de la date d'échéance prévue au contrat jusqu'à la date de calcul à un taux variant à chaque inventaire.

D'après l'article A.134-1 du Code des assurances, les provisions mathématiques sont calculées, pour chaque inventaire d'après un taux au plus égal à :

1. Pour chacun des engagements, 90% du dernier indice TEC n ⁷ publié par la Banque de France, où n correspond à l'échéance de la garantie du souscripteur. Une interpolation linéaire est réalisée entre les deux indices TEC n disponibles encadrant le plus directement l'échéance si cette dernière ne correspond pas à un indice TEC n disponible.
2. Par défaut, 90% du dernier indice TEC n publié par la Banque de France, où n correspond à la durée des engagements au passif du canton. Une interpolation linéaire sera effectuée si cette durée ne correspond pas à un indice TEC n disponible.

Pour une échéance ou une durée supérieure à la durée maximale disponible pour le TEC n , le taux retenu est celui du TEC de durée maximale et ne peut pas être négatif.

Le choix de la méthode est fait par l'entreprise d'assurance pour chaque canton. Cependant, si la méthode n°1 est retenue, ce choix est définitif.

L'article A.134-1 permet également que ce taux diffère de celui retenu lors du calcul du tarif.

Le taux peut varier d'un arrêté à l'autre : le montant de la provision mathématique peut de ce fait varier à la hausse comme à la baisse, suivant l'évolution du taux des emprunts d'Etat.

Lors du versement sur le support euro croissance, la part de cotisation affectée à la provision mathématique est calculée conformément à l'article A.335-1 du Code des assurances. La part de la cotisation qui n'est pas affectée à la provision mathématique est portée au compte du souscripteur sous forme de parts de provision de diversification.

L'assureur détermine le montant de la cotisation versée affecté à la provision mathématique :

$$PM = \frac{Capital\ garanti}{(1 + \text{taux d'actualisation}_d)^n}$$

⁷ Taux de l'échéance constante, pour n variant de 1 à 30, est le taux de rendement actuariel d'une valeur de Trésor fictive dont la durée de vie serait à chaque instant égale à n années. Ce taux est publié quotidiennement.

1.3.3 La provision de diversification

La provision de diversification (PD) est destinée à absorber les fluctuations de l'actif. Cette provision est détenue par chaque souscripteur sous forme individualisée de nombre de parts : lors d'un versement, le montant de la provision de diversification est égal à la différence entre le montant de la cotisation versée nette de frais et de la provision mathématique déterminée sous les conditions décrites en I.3.2.

L'assureur détermine le montant alloué à la provision de diversification :

$$PD = Prime_{nette} - PM$$

Le nombre de parts est obtenu en divisant ce montant par la valeur de provision de diversification en vigueur lors du mouvement.

D'après l'article R.134-5 I du Code des assurances, le contrat doit garantir une valeur, non nulle, minimale de la part de la provision de diversification⁸. Cette valeur doit être identique pour l'ensemble des souscripteurs, peu importe la date d'adhésion au contrat et ne doit pas être exprimée en pourcentage de la valeur de la part. Cette valeur constitue un engagement de l'entreprise d'assurance.

De plus, le contrat peut prévoir une possibilité de conversion en provision mathématique de la provision de diversification sous conditions décrites à l'article R.134-7 du Code des assurances.

1.3.4 L'affectation de la cotisation à l'engagement

Les engagements sont calculés en plusieurs étapes :

- L'assureur détermine le montant de l'engagement au terme relatif à la part de la cotisation versée selon le pourcentage garanti de la cotisation, l'échéance fixée à la souscription et le taux garanti s'il existe :

$$Capital\ garanti = \%_{garanti} \times Primes_{nettes} \times (1 + TMG)^n$$

- L'assureur détermine ensuite le montant de la cotisation versée qui doit être affecté à la provision mathématique :

$$PM = \frac{Capital\ garanti}{(1 + \text{taux d'actualisation}_d)^n}$$

- L'assureur détermine enfin le montant alloué à la provision de diversification :

$$PD = Prime_{nette} - PM$$

⁸ La règle énoncée par le décret du support euro croissance diffère légèrement de celle en vigueur lors de la création des produits eurodiversifiés. En effet, selon l'ancien article R.142-5 III du Code des assurances, la valeur minimale de la provision de diversification était déterminée en fonction de la valeur de provision de diversification à la date de création du fonds et ne pouvait pas être inférieure à 5% de cette dernière.

Le nombre de parts de PD détenu par le souscripteur est déterminé en divisant le montant de la PD obtenu par la valeur de la part de PD en vigueur lors du versement.

Le schéma suivant synthétise les étapes décrites plus haut :

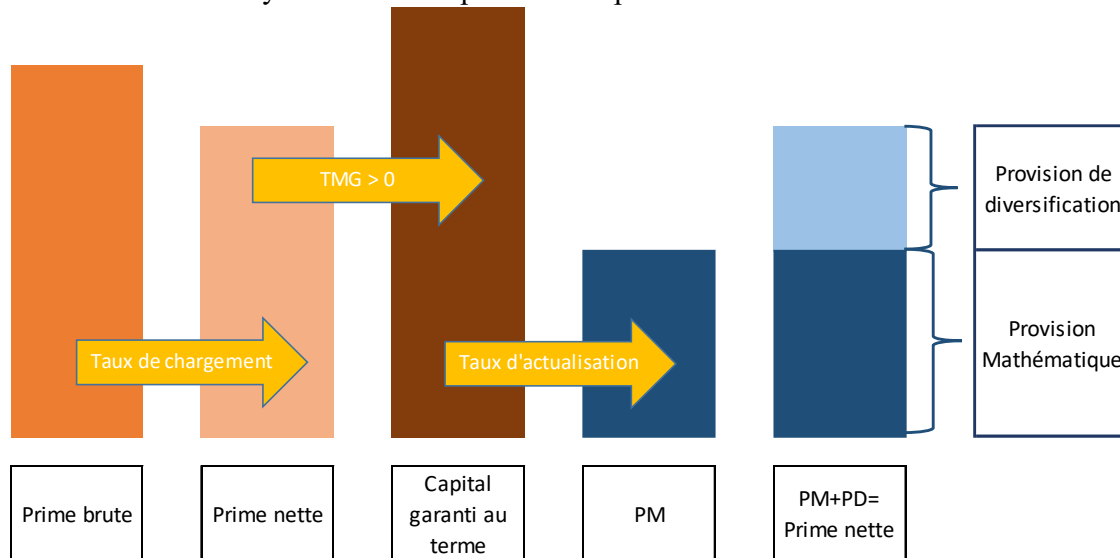


Figure 1: Calcul de la PM et de la PD au versement

Sur la base des hypothèses suivantes :

- Montant de la prime nette : 1 000€,
- Taux minimum garanti : 0%
- Pourcentage garanti au terme : 100%,
- Durée de l'engagement : 8 ans,
- Valeur liquidative de la part de PD à la souscription : 100€,
- Taux d'actualisation : 2,50%.

On a :

- $PM = \frac{1000}{(1+2,50\%)^8} = 820,75\text{€}$
- $PD = 1000 - 820,75 = 179,25\text{€}$
- $NbPD = \frac{179,25}{100} = 1,7925$

Le nombre de parts déterminé est acquis au souscripteur.

Les étapes effectuées lors du calcul de cet exemple sont :

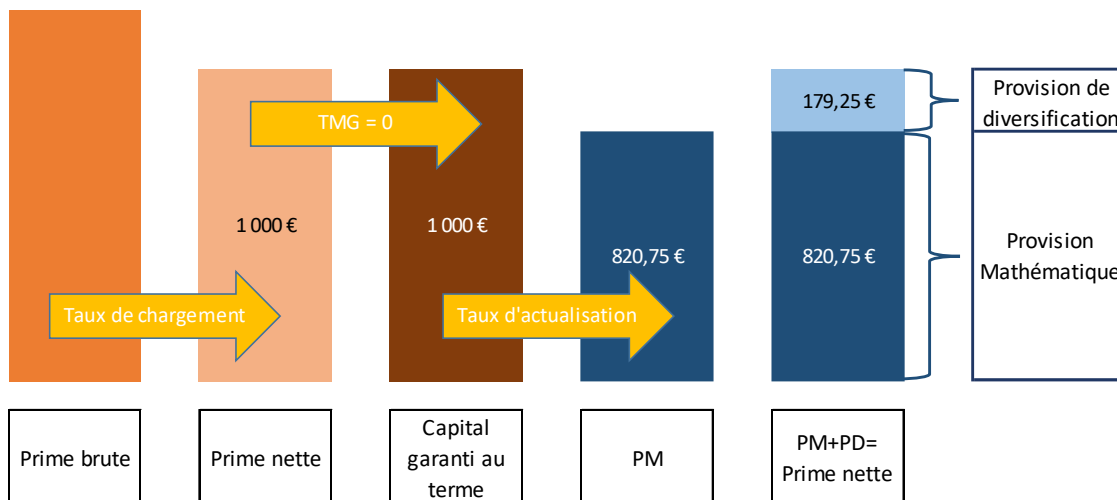


Figure 2 : Synthèse de l'exemple chiffré

1.3.5 Participation aux résultats

Le compte de participation aux résultats est défini à l'article A.331-4 II du Code des assurances. Ce compte doit être établi au minimum trimestriellement, voire hebdomadairement pour les fonds internes. Si les produits nets des placements sont négatifs ou si l'actif a réalisé une moins-value latente au cours de l'exercice, ces pertes sont imputées sur le compte de participation aux résultats. Son solde peut donc être créditeur ou débiteur. Lorsque ce compte présente un solde débiteur, le solde peut être reporté sur l'exercice suivant, déduction faite de la part du solde qui peut être compensée par une diminution de la valeur de la part de provision de diversification, dans la limite de la valeur minimale de cette provision, ou par la reprise de la provision collective de diversification différée (PCDD) ou encore par une combinaison de ces deux méthodes.

D'après l'article R.134-6 I du Code des assurances, plusieurs modes d'affectation des participations aux bénéfices sont possibles (une combinaison des quatre modes est également réalisable) :

- Revalorisation des engagements en euro : utilisable uniquement si le compte est créditeur et si les conditions définies à l'article R.134-6 II du Code des assurances sont respectées. C'est la seule affectation permettant de réaliser un effet cliquet à l'instar du fonds en euros.
- Revalorisation sous forme d'attribution de parts de provision de diversification : utilisable uniquement si le compte est créditeur. Les modalités de répartition sont à la discrétion de l'assureur.
- Revalorisation ou diminution de la valeur de la part de provision de diversification : la diminution de la valeur de la part est limitée à la valeur minimale définie au contrat.
- Dotation ou reprise de la provision collective de diversification différée : la dotation est possible uniquement si le compte est créditeur, et elle est limitée par la réglementation.

Lors de la création de ce support, les contrats commercialisés ne pouvaient proposer le fonds en euros comme support d'investissement en parallèle du fonds euro-croissance. Cette règle a été modifiée en janvier 2015. L'objectif de l'Etat était d'inciter les Français à investir leur épargne en

actions dans les PME⁹ (Petites et moyennes Entreprises) et ETI¹⁰ (Entreprise de taille intermédiaire), afin de répondre au besoin de financement des entreprises françaises.

Ainsi a été rendue possible la transformation des contrats fonds en euros + Unités de compte en contrats fonds en euros + Unités de compte + fonds euro-croissance, avec le maintien, en même temps, de l'antériorité fiscale (sous certaines conditions de conversion¹¹).

Cependant, les taux TEC étant extrêmement bas pour les échéances de courte et moyenne durée, cela ne permet pas lors d'un versement sur le support euro-croissance, d'allouer une part suffisamment importante de provision de diversification. La poche de captation des rendements d'actifs (la PD) est insuffisante pour espérer un rendement convenable face au risque pris.

Cela explique pourquoi la collecte observée sur le fonds euro-croissance est nettement inférieure aux deux autres supports présentés ci-dessus. Pour relancer le développement de ce support, un décret est entré en vigueur le 13 juillet 2016, proposant un mécanisme temporaire de transfert de plus-values latentes des fonds « euros » vers les fonds « euro-croissance ». Une brève présentation de ce mécanisme est proposée ci-dessous.

1.3.6 Les dispositions temporaires de transferts d'actifs

Jusqu'au 31 décembre 2019, les assureurs pouvaient transférer des actifs en représentation du fonds en euros vers des ou un fonds en euro-croissance, sous réserve d'une plus-value latente sur les actifs du fonds en euros.

Le montant des actifs transférés est inscrit à la provision collective de diversification différée. Ce montant n'est pas soumis à la limite de la dotation de la provision collective différée indiquée dans l'article A.132-11 du Code des assurances.

Le pourcentage de plus-value latente transféré est plafonné par la plus faible des deux valeurs suivantes :

- Le montant des prestations du fonds en euros versées lors de l'exercice divisé par la valeur totale des actifs de l'entreprise hors fonds en unités de compte et fonds cantonnés,
- Le montant des primes versées lors de l'exercice sur le fonds en euro-croissance divisé par la valeur totale des actifs de l'entreprise hors fonds en unités de compte et fonds cantonnés.

Si la collecte sur le fonds en euro-croissance est très faible, le transfert d'actifs possible sera aussi également très faible.

⁹ Est appelée PME, une entreprise présentant moins de 250 salariés et dont le chiffre d'affaires annuel n'excède pas 50 millions d'euros.

¹⁰ Est appelée ETI, une entreprise n'appartenant pas à la catégorie PME, présentant moins de 5000 salariés et dont le chiffre d'affaires annuel n'excède pas 1,5 milliards d'euro.

¹¹ La conversion doit être au minimum effectuée à hauteur de 10% des engagements en euros. Cette conversion est conditionnée par l'absence d'arbitrage par le souscripteur du fonds en euros sur un support en unités de compte dans les six mois précédant cette transformation.

II Le fonctionnement en cours de vie du contrat

II.1 Les versements

En cours de vie du contrat, l'assuré peut réaliser le versement de primes complémentaires, de façon périodique (versements programmés) ou de façon spontanée (versements libres).

II.2 Les arbitrages

L'opération d'arbitrage consiste à modifier l'allocation sur les supports, de tout ou partie de la valeur de rachat du contrat : l'assuré peut déplacer son investissement d'un support actuel vers un autre support (ex : du fonds en euros vers des supports en unités de compte, ou de supports en unités de compte vers d'autres supports en unités de compte, ou toute autre combinaison possible définie dans le contrat).

Les arbitrages peuvent être effectués :

- Soit de manière automatique, en suivant un profil de risque ou un horizon défini dans le contrat : on parle de gestion « profilée » ;
- Soit de manière individuelle par l'assuré : on parle de gestion « libre » ;
- Soit sur recommandation d'un conseiller financier, avec l'aval de l'assuré ; on parle enfin de gestion « accompagnée ».

II.3 Les chargements et les frais

L'assureur prélève des frais sur les opérations réalisées sur le contrat ou sur l'encours de celui-ci, afin de se rémunérer au titre des charges et de l'immobilisation de Fonds Propres qu'il subit. Par ailleurs, l'assureur doit, d'après l'article L.132-5-2 du Code des assurances, indiquer en caractères apparents ces frais dans une même rubrique de la note d'information (ou encadré) qui est remis à l'assuré avant la conclusion du contrat. Les frais généralement constatés sont :

- Frais sur versements : prélevés sur les sommes versées.
- Frais annuels sur encours : prélevés en fin de période (généralement en fin de mois). Ces frais sont calculés par rapport à l'encours constaté sur le contrat.
- Frais d'arbitrages : prélevés lors de la réalisation d'un arbitrage entre différents supports du contrat. Il est courant qu'un arbitrage « gratuit » par an soit prévu contractuellement. Les chargements associés à ces mouvements sont définis dans les dispositions contractuelles. Certains mouvements d'arbitrages peuvent être interdits ou limités par le contrat. Par exemple, le produit Cardif Essentiel interdit l'arbitrage du support euro-croissance vers un autre support, afin d'éviter que des mouvements de « sécurisation de plus-values » remettent en cause la mutualité et la politique de gestion d'actif du canton euro-croissance.
- Frais de sorties ou pénalités de rachat : Ces pénalités peuvent être fonction de l'ancienneté du contrat avec un maximum de 5% de l'encours. Si le rachat intervient dix ans après le premier versement, les pénalités sont nulles. Selon une étude sur la sensibilité de

l'assurance-vie aux taux d'intérêt en Europe, l'ACLI¹² estime que l'utilisation de ces pénalités est peu significative en France. En effet, en France, les pénalités de rachat sont essentiellement appliquées aux souscripteurs personnes morales soumis à l'impôt sur les sociétés.

- Coût des garanties optionnelles ou obligatoires qui peuvent être ajoutées au contrat. L'exemple le plus fréquent est la garantie complémentaire en cas de décès.

II.4 Les garanties complémentaires en cas de décès

En cas de décès de l'assuré, l'assureur garantit le versement d'un capital complémentaire à la valeur de rachat, dont le montant est fonction de la garantie souscrite en cas de décès :

- Lorsque le capital complémentaire est égal à la différence, uniquement positive, entre la valeur de rachat à la date d'effet du décès et le cumul des versements nets investis, déduction faite des rachats partiels bruts antérieurs, sous forme de réduction proportionnelle, on parle de garantie plancher « simple ».

La différence peut également être réalisée avec un montant de versements nets investis indexés à un taux annuel (garantie plancher indexée) ou multipliés par un coefficient supérieur à 100% (garantie plancher majorée).

Selon le contrat d'assurance-vie, la garantie en cas de décès peut être optionnelle ou obligatoire.

La tarification peut également être calculée de deux façons : une tarification *a priori* ou une tarification *a posteriori*. Lors d'une tarification *a priori*, les frais liés à cette garantie sont fixes durant toute la vie du contrat (ou durant toute la durée de validité de cette garantie). Lors d'une tarification *a posteriori*, des frais sont prélevés uniquement si un montant de capital sous risque est constaté lors du calcul des frais (le prélèvement de ces frais est généralement réalisé simultanément au prélèvement des frais sur encours).

Les garanties optionnelles en cas de décès ne sont pas accessibles notamment aux mineurs de moins de 12 ans ni aux personnes sous tutelle¹³. De plus, selon les dispositions contractuelles du contrat, certaines conditions peuvent mettre fin aux garanties et certaines causes de décès peuvent être exclues du périmètre couvert par ces garanties.

Remarque : Ces garanties n'existent pas sur les contrats de capitalisation qui ne garantissent pas le décès ni sur les contrats PERP sur lesquels elles sont interdites réglementairement.

¹² The American Council of Life Insurers. ACLI, Challenges of the life insurance industry in Europe in a low interest rate environment, Refocus Conference 2013, Session 1C, 4 March

¹³ L.132-3 du Codes des assurances

III Le dénouement du contrat ¹⁴

III.1 La sortie en rente viagère

La possibilité de sortie du contrat en rente viagère est soumise aux dispositions contractuelles du produit souscrit. Si les sorties en rente viagère est possible, des options sont généralement proposées telles que la réversion (totale ou partielle) ou un nombre d'annuités garanties. Le contrat est alors « transformé » en rente viagère.

III.1.1 La fiscalité de la rente viagère constituée à titre onéreux

Ce mémoire a pour objectif principal d'étudier les comportements de rachat des contrats d'assurance-vie ou de capitalisation souscrits à titre individuel. La fiscalité des produits PERP, Madelin et des contrats dits article 83 ou 39 n'est pas indiquée dans ce mémoire (rente viagère à titre gratuit).

Au moment de la sortie en rente viagère, c'est-à-dire au moment de la liquidation du contrat en rente, les produits financiers acquis sont exonérés d'impôt sur le revenu car ceux-ci sont considérés comme un revenu lors de la phase de versements des rentes mais uniquement pour une fraction de leur montant.

Cette fraction est déterminée forfaitairement¹⁵, pour toute la durée de la rente, d'après l'âge du crédirentier lors de l'entrée en jouissance de la rente :

- 70%, si l'intéressé était âgé de moins de 50 ans,
- 50% s'il était âgé de 50 à 59 ans inclus,
- 40% s'il était âgé de 60 à 69 ans inclus,
- 30% s'il était âgé de plus de 69 ans (c'est-à-dire plus de 70 ans révolus).

La fraction imposable est également soumise aux prélèvements sociaux selon le taux en vigueur lors du versement de l'arrérage (17,2% en 2018).

Les prélèvements sociaux sont également imputés lors de la transformation du contrat en rente viagère.

Globalement, ce mode de sortie est très rarement utilisé en assurance-vie individuelle. En effet, lors d'une étude réalisée en 2015 sur le portefeuille global de BNP Paribas Cardif, 72 sorties en rentes ont été constatées en 3 ans sur 1,95 millions de contrats en stock.

¹⁴ La fiscalité décrite dans ce chapitre est celle en vigueur lors de la rédaction de ce mémoire.

¹⁵ Détermination forfaitaire qui a pour objectif de représenter la plus-value « générée » durant la phase de rente.

III.2 Le décès

Le décès de l'assuré met fin au contrat d'assurance-vie¹⁶.

Les sommes stipulées payables lors du décès de l'assuré à un (ou des) bénéficiaire(s) déterminé(s) ne font pas partie de la succession de l'assuré, mais sont soumises à l'article 757B et/ou à l'article 990I du Code général des impôts.

Cependant, la loi TEPA a instauré la suppression des droits de succession pour le conjoint, le partenaire d'un PACS et les frères et sœurs résidant sous le même toit et remplissant certaines conditions¹⁷. Ainsi, ces bénéficiaires ne sont soumis ni à l'article 757B ni à l'article 990I.

Pour connaître la fiscalité applicable aux capitaux décès versés au titre d'un contrat d'assurance-vie, il convient de regarder :

- la date de souscription du contrat,
- la date de versement des primes,
- l'âge auquel les primes ont été versées (avant ou après 70 ans).

Ci-dessous un tableau synthétique indiquant la fiscalité à appliquer selon les caractéristiques indiquées ci-dessus :

		Contrat souscrit avant le 20 novembre 1991	Contrat souscrit à compter du 20 novembre 1991
Primes versées avant le 13 octobre 1998	Avant l'âge de 70 ans de l'assuré	Exonération totale des capitaux transmis	Exonération totale des capitaux transmis
	Après l'âge de 70 ans de l'assuré		757 B
Primes versées à compter du 13 octobre 1998	Avant l'âge de 70 ans de l'assuré	990 I	990 I
	Après l'âge de 70 ans de l'assuré		757 B

Tableau 1 : Synthèse de l'article fiscal appliqué en cas de décès selon les critères du contrat

III.2.1 La fiscalité au titre de l'article 990I

L'article 990I concerne uniquement les versements effectués à compter du 13 octobre 1998. Le capital versé au bénéficiaire excédant 152 500€ (tous contrats de tous assureurs confondus) sera taxé forfaitairement à hauteur de 20% jusqu'à 700 000€ puis 31,25% au-delà.

Lorsque les primes sont versées à la fois avant et après les 70 ans de l'assuré, ou à la fois avant et après le 13 octobre 1998, il convient de déterminer la part de la valeur de rachat prise en compte dans le calcul de l'assiette 990I. 2 méthodes peuvent être appliquées :

- Méthode globale simplificatrice définie dans le Bulletin Officiel des Impôts 2002
- Méthode impactant les rachats sur les versements soumis à l'article 990I.

¹⁶ Dans le cas d'une co-souscription avec dénouement au dernier décès, c'est le décès du dernier assuré qui met fin au contrat. Dans le cas d'un contrat de capitalisation, celui-ci ne se dénoue pas au décès du souscripteur.

¹⁷ Etre domicilié avec le défunt pendant les 5 années précédentes et être soit âgé de plus de 50 ans soit handicapé.

III.2.2 La fiscalité au titre de l'article 757B

L'article 757B ne concerne que les contrats souscrits à compter du 20 novembre 1991 et pour lesquels des versements ont été effectués après les 70 ans de l'assuré.

Ces primes sont imposées au titre des droits de mutation à titre gratuit (qui sont en fonction du lien de parenté entre l'assuré et le bénéficiaire) après abattement de 30 500€ par assuré et pour l'ensemble des contrats qu'il a souscrits¹⁸, quel que soit le nombre de bénéficiaires.

Du fait de la nécessité de prendre en compte le lien de parenté, c'est l'administration fiscale qui réalise le calcul de l'impôt dû au titre de l'article 757B. Les bénéficiaires, peuvent cependant demander à l'assureur de verser au service des impôts la taxe due au titre de l'article 757B. Dans ce cas, cette taxe est prélevée directement sur le capital décès dû.

III.3 Le rachat

Le rachat consiste à rembourser à l'assuré une partie ou la totalité de la valeur de rachat du contrat de façon anticipée, c'est-à-dire avant le terme fixé par le contrat ou avant le décès de l'assuré.

III.3.1 La fiscalité supportée en cas de rachat

La fiscalité supportée en cas de rachat a été modifiée au 01/01/2018. Or, les années d'observation de notre portefeuille vont de 2011 à 2016 : nous présenterons donc dans un premier temps la fiscalité en vigueur entre 2011 et 2016. Dans un deuxième temps, nous présenterons la nouvelle fiscalité entrée en vigueur au 01/01/2018.

III.3.1.1 Fiscalité en vigueur avant le 01/01/2018

Lors d'un rachat, le souscripteur personne physique peut au choix :

1. Réintégrer les produits du contrat dans sa déclaration annuelle de revenus (IRPP),
2. Opter pour le prélèvement forfaitaire libératoire (PFL) dont le taux varie selon l'ancienneté du contrat au moment du rachat.

Ci-dessous un tableau présentant les taux de prélèvement fiscaux appliqués par l'assureur lors du rachat :

Option fiscale / Ancienneté du contrat	IRPP	PFL
Entre 0 et 4 ans	0%	35%
Entre 4 et 8 ans		15%
Plus de 8 ans		7,5%

Tableau 2 : Taux des prélèvements fiscaux pris par l'assureur selon l'ancienneté du contrat

¹⁸ Dans le cas où la part du capital décès (net de prélèvements sociaux) correspondant aux primes versées après 70 ans est inférieure au montant global de ces primes brutes, l'assiette est limitée aux capitaux versés aux bénéficiaires.

Un contrat de plus de 8 ans bénéficiera d'un abattement de 4 600€ pour une personne seule et de 9 200€ pour un couple, par année civile, sur les produits acquis lors du rachat. Cet abattement est commun à l'ensemble des contrats d'assurance-vie/capitalisation détenus par le souscripteur. Il est réalisé par l'administration fiscale et non par l'assureur.

Le choix optimal entre les deux options IRPP ou PFL dépend du taux marginal d'imposition du souscripteur.

Les produits acquis lors d'un rachat sont également soumis aux prélèvements sociaux.

III.3.1.2 Fiscalité en vigueur à partir du 01/01/2018

Il convient de distinguer les produits des versements effectués avant et après le 27/09/2017 :

- Les produits des versements effectués avant le 27/09/2017 sont soumis au PFL ou à l'IRPP, selon le choix du souscripteur,
- Les produits des versements effectués à partir du 27/09/2017 sont soumis à un prélèvement forfaitaire obligatoire (PFO) non libératoire.

Ancienneté du contrat	Si le cumul des versements au 31/12/N-1 (net de l'éventuelle part rachetée) est égal ou inférieur à 150 000€	Si le cumul des versements au 31/12/N-1 (net de l'éventuelle part rachetée) est supérieur à 150 000€
Moins de 8 ans	12,8%	
Plus de 8 ans	7,5%	Prélèvement de l'assureur : 7,5% Fraction taxée à 7,5% ¹⁹ Solde taxé à 12,8%

Tableau 3 : Taux des prélèvements fiscaux pris par l'assureur selon l'ancienneté du contrat

En cas de rachat après 8 ans, les produits bénéficient d'un abattement annuel de 4 600€ pour une personne seule et de 9 200€ pour un couple (tous contrats d'assurance vie et de capitalisation confondus).

L'abattement s'applique en priorité :

- Aux produits attachés aux versements effectués avant le 27/09/2017,
- Aux produits attachés aux versements effectués à compter du 27/09/2017, qui sont imposés au taux de 7,5%,
- Aux produits attachés aux versements effectués à compter du 27/09/2017, qui sont imposés au taux de 12,8%.

Les produits acquis lors d'un rachat sont également soumis aux prélèvements sociaux.

¹⁹ La fraction taxée à 7,5 % correspondant au rapport :

- montant des produits x (150 000 – cumul des versements effectués avant le 27/09/2017, net de la part rachetée au 31/12/N-1),

- sur cumul des versements effectués à compter du 27/09/2017, net de la part rachetée au 31/12/N-1.

Le solde des produits est taxé à 12,8% par l'Administration fiscale (l'assureur ayant déjà prélevé 7,5% par un prélèvement forfaitaire non libératoire sur la totalité des produits contenus dans le rachat).

Cette modification de la fiscalité sur les contrats d'assurance vie et de capitalisation inquiète les assureurs. En effet, du fait de la diminution des incitations fiscales de ces contrats par rapport à des produits bancaires, les assureurs pensent que cette réglementation va transformer le passif de leur portefeuille en en diminuant la durée. Cette transformation sera très progressive car elle ne prendra effet que pour les contrats/versements souscrits depuis le 27/09/2017 et rachetés depuis le 01/01/2018. De ce fait, notre mémoire et notre étude du portefeuille d'Assurance Epargne Pension (AEP) ne prennent pas en compte l'impact de cette mesure fiscale.

III.3.2 Les comportements de rachat

On distingue deux types de rachats :

Les rachats structurels : ces rachats ne sont pas dus au contexte économique mais à un besoin spécifique de l'assuré (exemple : besoin de liquidité pour effectuer un achat/un investissement) et à ses caractéristiques propres (âge de l'assuré, type de contrat, etc.)

Les rachats conjoncturels : ces rachats sont en lien direct avec le rendement effectif du contrat. Le taux de rachat conjoncturel résulte principalement de l'écart constaté entre le taux de revalorisation servi et le taux de rendement espéré par l'assuré (on peut l'assimiler par exemple au taux moyen des emprunts d'Etat). Ces rachats sont généralement réinvestis chez un concurrent, avec l'espérance d'un meilleur rendement.

Soulignons l'existence d'une troisième catégorie de rachat : le rachat « exceptionnel ». Il s'agit de rachats massifs effectués par les assurés, non dans une recherche de rendement mais du fait d'une défiance totale à l'égard des marchés financiers ou par peur d'un effondrement du système existant.

Approfondir la connaissance du comportement des assurés de son portefeuille est donc primordiale afin de permettre une meilleure anticipation des taux de rachat et de favoriser un meilleur adossement actif-passif. Le nouveau cadre réglementaire Solvabilité II exige que soit pris en compte le risque de rachat. Les lois de rachats interviennent dans la valorisation des provisions *best estimate* et le calcul de certains modules du SCR comme le module « souscription vie ». Même si dans la réalité, distinguer les rachats conjoncturels des rachats structurels est complexe voire impossible, les assureurs modélisent généralement ces deux comportements de façon séparée.

III.3.3 Les alternatives au rachat

III.3.3.1 L'avance

D'après l'article L.132-21 du Code des assurances, l'assureur peut consentir une avance au souscripteur pour une durée déterminée. L'avance s'apparente à un prêt que l'assureur accorde à l'assuré sans remettre en cause l'antériorité fiscale du contrat ni de la répartition de sa valeur de rachat entre les supports investis.

Un engagement à caractère déontologique a été mis en place, relatif à l'utilisation des avances par les sociétés membres de la FFA.

Ainsi, l'avance consentie ne doit pas dépasser 80% du montant de la provision mathématique des contrats en euros et 60% de celle des contrats en supports unités de compte.

De plus, l'avance a une durée maximale de trois ans renouvelables deux fois.

Enfin, « le taux d'intérêt de l'avance doit être au moins égal au taux moyen des emprunts d'Etat, calculé sur une base au plus semestrielle, majoré du taux des frais de gestion du contrat et d'une rémunération normale de l'assureur.

Pour les contrats comportant un compartiment en euros, le taux doit être en outre au moins égal au taux de rémunération du compartiment euros du contrat de l'année précédant la demande d'avance majoré des frais de gestion du contrat et d'une rémunération normale de l'assureur. »

III.3.3.2 Le nantissement

Il s'agit de l'opération par laquelle le contrat d'assurance-vie et/ou de capitalisation est utilisé comme garantie de remboursement d'un prêt au créancier (article L.132-10 du Code des assurances). En présence d'un bénéficiaire acceptant²⁰, son accord est nécessaire pour effectuer une telle opération.

Cette opération permet de contracter un prêt auprès d'un organisme dédié tout en conservant son contrat d'assurance-vie ou de capitalisation. Cette possibilité est d'autant plus intéressante si in fine le montant de l'emprunt est inférieur au montant acquis sur le contrat d'assurance-vie ou de capitalisation.

²⁰ Personne physique ou morale désignée comme bénéficiaire du contrat d'assurance vie en cas de décès de l'Assuré et qui a manifesté sa volonté de recevoir le capital du contrat selon les modalités prévues à l'Article L.132-9 du Code des assurances.

Partie II : Caractéristiques du portefeuille étudié et sélection des variables explicatives

Actuellement, la loi de rachats totaux calculée sur le portefeuille est déterministe. Elle est réalisée, en nombre de contrats rachetés, sur la base des données historiques des contrats d'assurance-vie et de capitalisation, avec un unique découpage qui est fonction de l'ancienneté fiscale du contrat. Ceci est pour répondre aux exigences de la compagnie. Un découpage plus fin ne peut pas être utilisé lors de l'agrégation des données avec celles des autres portefeuilles gérés par la compagnie.

La loi de rachats partiels, quant à elle, est différenciée par support (fonds en euros ou supports en unités de compte), par ancienneté fiscale, et par typologie de souscription.

I Statistiques descriptives du portefeuille étudié

1.1 Base de données

Le portefeuille étudié est celui d'Assurance Epargne Pension (AEP), marque commerciale de l'entreprise d'assurance BNP Paribas Cardif. Jusqu'en 2010, l'établissement était une filiale de Dexia, sous le nom de Dexia Epargne Pension. AEP est depuis sa création en 2003 spécialisée dans les produits d'assurance-vie et de capitalisation dits « haut de gamme », vendus en marque blanche, à destination de la clientèle des banques privées et CGPI²¹.

Comme indiqué précédemment, cette étude est réalisée uniquement sur les contrats d'assurance-vie et capitalisation à souscription individuelle de type « multi supports ». Les contrats avec fonds euro-croissance sont également exclus du périmètre du fait de la présence d'une clause contractuelle de non-rachetabilité temporaire jusqu'au 31/03/2015. De plus, ces contrats représentent une faible volumétrie d'encours de la compagnie.

²¹ Conseiller en Gestion de Patrimoine Indépendant

1.2 Statistiques descriptives

Données au 31/12	2010	2011	2012	2013	2014	2015
Nombre de contrats	44 452	47 511	49 032	51 715	55 041	58 805
Encours total (M€)	4 192	4 453	5 004	5 935	6 583	7 154
Dont fonds Euro (M€)	2 257	2 434	2 618	3 080	3 447	3 795
Dont UC (M€)	1 935	2 019	2 386	2 855	3 136	3 360
Part d'UC moyenne (%)	46%	45%	48%	48%	48%	47%
Encours moyen (€)	94 300	93 728	102 060	114 769	119 604	121 657
Données au 31/12	2010	2011	2012	2013	2014	2015
Nombre de contrats	44 452	47 511	49 032	51 715	55 041	58 805
Encours total (M€)	4 191,85	4 453,09	5 004,20	5 935,29	6 583,10	7 154,02
Dont fonds Euro (M€)	2 256,80	2 434,25	2 617,94	3 080,15	3 446,97	3 794,52
Dont UC (M€)	1 935,04	2 018,84	2 386,25	2 855,15	3 136,13	3 359,50
Part d'UC moyenne (%)	46,16%	45,34%	47,69%	48,10%	47,64%	46,96%
Encours moyen (€)	94 300,49	93 727,65	102 059,80	114 769,25	119 603,59	121 656,73

Tableau 4 : Données de l'encours du portefeuille du 31/12/2010 au 31/12/2015

L'évolution de l'encours est généralement due à trois principales raisons :

- Les flux (entrants et sortants),
- La valorisation des contrats : sous forme de Participation au résultat pour l'euro ou de variation des valeurs liquidatives des unités de comptes,
- Les chargements divers.

Sur le portefeuille étudié, trois possibilités de rachat sont proposées aux clients :

- Le Rachat partiel (RCP) : remboursement partiel ponctuel de la valeur de rachat du contrat.
- Le Rachat total (RCT) : remboursement de la totalité de la valeur de rachat du contrat. Cet acte met fin au contrat souscrit.
- Les Rachats programmés (RPR) : remboursements partiels de la valeur de rachat du contrat effectués selon une périodicité et un montant déterminé par le souscripteur.

Montant Racheté (M€)	2011	2012	2013	2014	2015
Rachat Partiel	279,025	277,728	259,585	315,339	261,163
Rachat Total	155,247	233,897	127,956	170,212	145,175
Rachat programmé	10,644	12,060	14,184	14,750	15,310
Répartition des Rachats	2011	2012	2013	2014	2015
Rachat Partiel	63%	53%	65%	63%	62%
Rachat Total	35%	45%	32%	34%	34%
Rachat programmé	2%	2%	4%	3%	4%

Tableau 5 : Les différents rachats constatés de l'année 2011 à 2015

Au vu du faible impact sur le portefeuille des rachats programmés, ceux-ci seront agrégés aux Rachats Partiels dans la suite du document.

I.2.1 Les différents types de souscription

On distingue quatre typologies de cadres de souscription de contrats :

- Assurance-vie (souscrite par des personnes physiques, ou des personnes morales soumises à l'impôt sur le revenu),
- Contrats de capitalisation souscrit par des personnes morales soumises à l'impôt sur le revenu (IR),
- Contrats de capitalisation souscrit par des personnes morales soumises à l'impôt sur les sociétés (IS),
- Contrats de capitalisation souscrit par des personnes physiques.

Voici l'évolution de l'encours de ces typologies par rapport à l'encours total :

Typologie de Contrat	2010	2011	2012	2013	2014	2015
Assurance-Vie	61,70%	65,19%	67,99%	68,73%	68,22%	68,98%
Capitalisation	38,30%	34,81%	32,01%	31,27%	31,78%	31,02%
dont personne morale IR	3,31%	3,47%	3,52%	3,57%	3,37%	3,15%
dont personne morale IS	27,14%	23,51%	20,36%	19,47%	20,23%	19,38%
dont personne physique	7,86%	7,82%	8,12%	8,23%	8,18%	8,50%

Tableau 6 : Répartition de l'encours selon la typologie du contrat

Pour une personne physique, la principale différence entre un contrat d'assurance-vie et un contrat de capitalisation est le dénouement lors du décès du souscripteur. En effet, ces deux types de contrats proposent des supports équivalents et sont soumis à la même fiscalité en cas de rachat. On peut donc supposer que les raisons de rachat sont les mêmes sur ces deux types de contrats.

Les histogrammes ci-dessous présentent les différents types de rachats effectués de 2011 à 2015 selon la typologie du contrat :

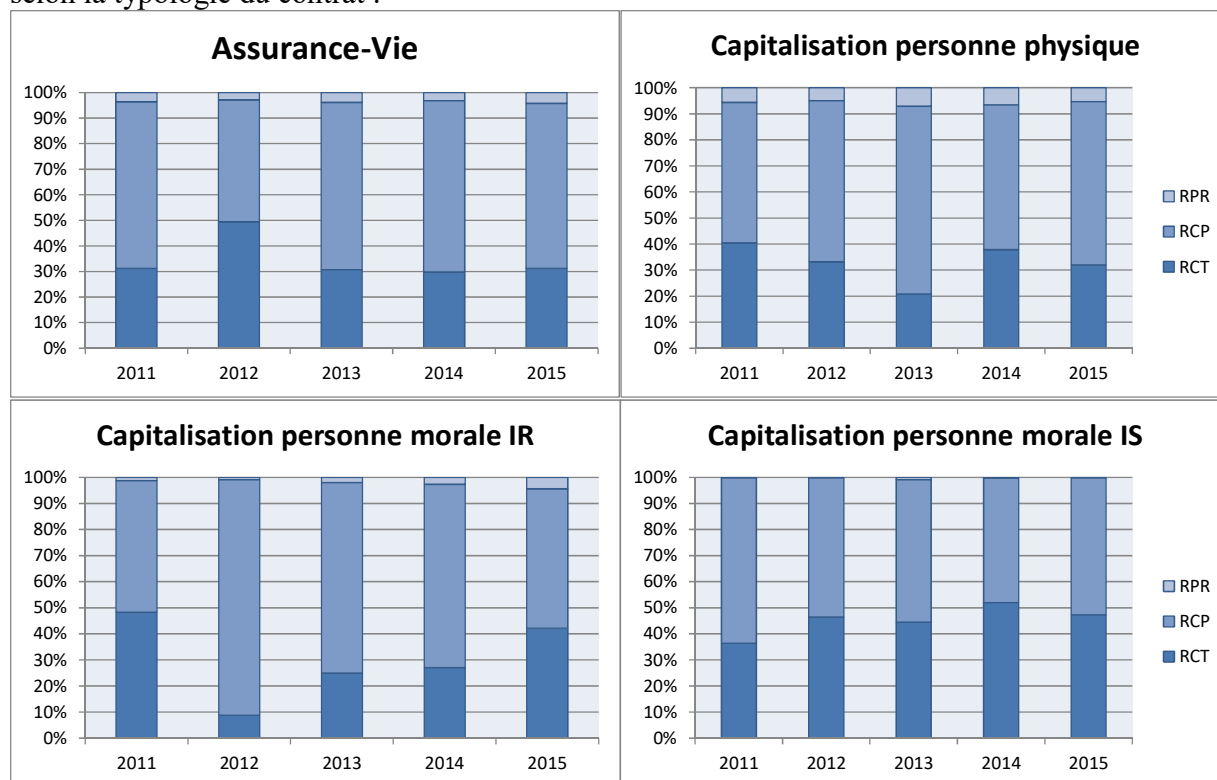


Figure 3 : Histogramme de répartition des rachats entre 2011 et 2015, selon la typologie du contrat

Nous pouvons constater que les contrats d'assurance-vie et les contrats de capitalisation personne physique affichent les mêmes proportions entre les différents types de rachats. Cela confirme l'hypothèse d'une équivalence de comportement de rachat sur ces deux typologies de contrats.

La répartition Rachat Partiel / Rachat total constatée sur les contrats personnes morales IS est différente de celle constatée sur la globalité du portefeuille. En effet, sur la globalité du portefeuille, la répartition (en encours) est en moyenne de 65% / 32% (les 3% restants correspondant aux rachats partiels programmés). Pour les contrats de capitalisation personnes morales IS, la répartition est aux alentours de 55% / 44%. On pourrait donc supposer que le comportement de ces assurés diffère de celui des assurés des autres types de souscriptions.

Voici l'évolution du stock de contrats au 31/12, de 2010 à 2015, selon la typologie du contrat :

Typologie de Contrat	2010	2011	2012	2013	2014	2015
Assurance-Vie	40 642	43 387	44 736	47 183	50 175	53 609
Capitalisation	3 810	4 124	4 296	4 532	4 866	5 196
Dont personne morale IR	363	402	438	461	477	483
Dont personne morale IS	1 719	1 625	1 496	1 498	1 545	1 592
Dont personne physique	1 728	2 097	2 362	2 573	2 844	3 121

Tableau 7 : Stock de contrats en vigueur du 31/12/2010 au 31/12/2015 selon la typologie du contrat

Ces deux indicateurs (encours et nombre de contrats) permettent de constater que la représentativité des contrats de capitalisation personne physique est assez faible par rapport aux autres typologies de contrats (même si elle est supérieure à celle des contrats de capitalisation souscrits par des personnes morales soumises à l'IR ou à l'IS).

Dans la suite de l'étude, les contrats d'assurance-vie et les contrats de capitalisation souscrits par des personnes physiques seront agrégés afin d'augmenter la robustesse des données.

1.2.2 Catégories socio-professionnelles

L'indication quant à la catégorie socio-professionnelle de l'assuré étant absente pour 90% des contrats observés (non renseignée ou indiquée « Inconnue »), celle-ci n'est pas utilisée dans le reste de l'étude. Pour autant, nous pensons qu'il peut exister une corrélation entre les revenus du ménage et le comportement de rachat.

Sur la base des données de l'étude INSEE « Revenu salarial net annuel moyen par sexe et PCS²² en 2013 : Comparaison régionale », nous déterminons cinq classes différentes selon la résidence fiscale du souscripteur :

- La classe 1 : représente les Non-résidents français,
- La classe 2_1 : représente les départements où le revenu salarial net moyen est supérieur ou égal à 21 000€,
- La classe 2_2 : représente les départements où le revenu salarial net moyen est supérieur ou égal à 20 000€ et inférieur à 21 000€,
- La classe 2_3 : représente les départements où le revenu salarial net moyen est supérieur ou égal à 19 000€ et inférieur à 20 000€,
- La classe 2_4 : représente les départements où le revenu salarial net moyen est inférieur à 19 000€.

La classification ci-dessus est réalisée de cette façon pour deux raisons principales :

- Distinction résidents / non-résidents : en cas de rachats, ces deux catégories ne sont pas fiscalement impactées de la même manière : on peut donc supposer qu'un résident et un non-résident n'auront pas le même comportement en cas de rachat.
- Distinction résident par revenu salarial : on peut supposer que les raisons de racheter son contrat sont différentes selon le revenu du foyer.

Voici l'évolution de l'encours de ces typologies par rapport à l'encours total :

Classification géographique	2010	2011	2012	2013	2014	2015
1	6,05%	6,32%	5,41%	5,17%	3,95%	3,56%
2_1	66,40%	65,93%	66,04%	66,50%	68,13%	68,17%
2_2	14,17%	14,05%	14,56%	14,33%	14,20%	14,50%
2_3	8,71%	8,77%	8,89%	9,03%	8,88%	9,25%
2_4	4,67%	4,93%	5,11%	4,97%	4,84%	4,53%

Tableau 8 : Répartition de l'encours selon la classification géographique

²² Professions et catégories socioprofessionnelles

Nous pouvons constater que le pourcentage de l'encours détenu par chaque classification reste relativement stable au cours des années excepté pour les non-résidents français où une légère diminution constante est remarquée.

Cela peut s'expliquer par le fait que même si les souscriptions des non-résidents ne sont pas interdites, la politique commerciale de l'entreprise ne cherche pas à prioriser ce type de souscription. Il y a donc une nette diminution du chiffre d'affaire entre 2010 et 2015 sur cette typologie de souscripteurs (malgré une augmentation du CA entre 2014 et 2015).

Pour les classifications représentant les souscripteurs résidents français, la stabilité de l'encours tend à faire penser que cette caractéristique n'influe pas la politique commerciale.

Afin d'avoir une idée du comportement de rachat des souscripteurs, nous proposons un indicateur de taux de rachat rapporté à la PM soit : $Taux = \frac{\text{Montant de Racha constatés dans l'année } N}{\frac{PM_{fin} + PM_{début}}{2}}$

Classification géographique	2011	2012	2013	2014	2015
1	7,78%	23,29%	10,97%	24,36%	7,07%
2_1	10,86%	10,80%	7,26%	7,19%	6,12%
2_2	10,36%	9,09%	6,96%	7,12%	4,99%
2_3	10,29%	8,87%	6,95%	7,36%	6,51%
2_4	5,55%	9,97%	6,51%	7,53%	8,44%

Tableau 9 : Montant de rachats rapporté à l'encours

Cet indicateur, bien qu'imprécis, nous permet d'obtenir une tendance sur les rachats effectués (tous types de rachats confondus) selon la catégorie étudiée.

Les taux de rachats constatés pour les classifications de résidents français semblent similaires entre classifications, même s'ils évoluent dans le temps. En revanche, le taux de rachat des non-résidents français semble plus volatile.

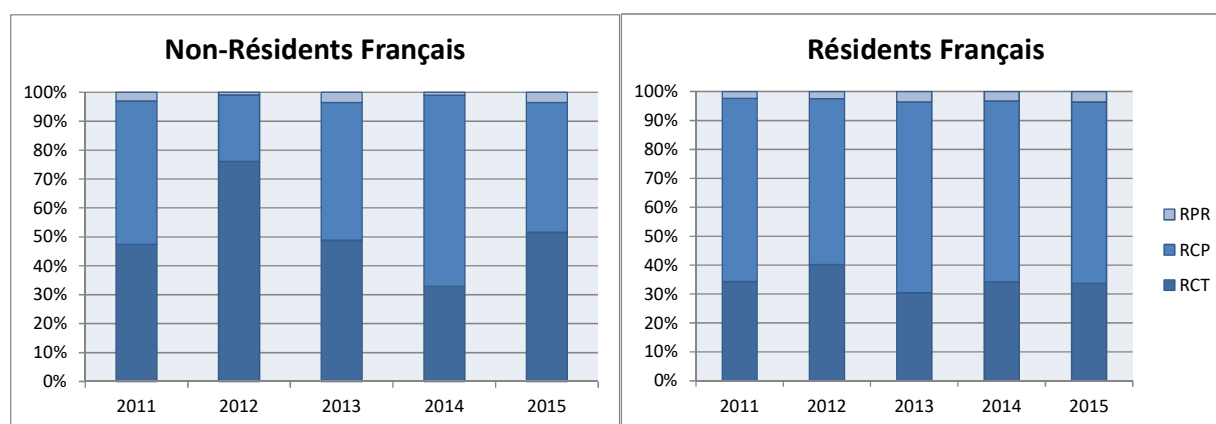


Figure 4 : Histogramme de répartition des rachats selon la résidence du souscripteur de 2011 à 2015

La proportion de rachats totaux constatée sur les souscriptions de non-résidents est bien plus importante que sur les souscriptions des résidents, ce qui laisse supposer que ces deux groupes n'ont pas le même comportement en cas de rachat.

Concernant la distinction des résidents français selon le revenu salarial moyen par département de résidence, aucun indicateur préliminaire ne laisse supposer un comportement différent entre ces classes. De plus, cette classification par revenu est implicite via la classification par réseaux de distribution.

Dans la suite de l'étude, nous effectuerons seulement la distinction : « résident » / « non-résident ».

1.2.3 Les différents réseaux de distribution

En étudiant l'encours depuis le 31/12/2010, nous constatons plus de 109 partenaires différents dont certains présentent un unique ou quelques contrats. L'utilisation brute de la liste des partenaires n'est pas appropriée car la segmentation réalisée ne présentera pas suffisamment de données pour prétendre à une étude assez robuste.

L'étude de l'encours moyen des clôtures entre 2010 et 2015, nous permet de répartir les contrats selon quatre segments de clientèle :

- Retail : Encours moyen < 20k€
- Aisé : 20k€ < Encours moyen < 100k€
- Patrimonial : 100k€ < Encours moyen < 500k€
- Gestion de Fortune : Encours moyen > 500k€

La direction de l'offre AEP classe les partenariats selon deux types :

- les partenariats auxquels il a été mis fin : nous les appellerons « Run-off »
- les partenariats toujours actifs : nous les appellerons « Partenariats ».

En croisant les quatre segments de contrats avec les deux types de partenariats, nous obtenons 7 groupes différents de réseaux de distribution :

Typologie_Partenariat	2010	2011	2012	2013	2014	2015
Partenariat_Aisé	9 094	11 040	12 293	12 906	13 318	13 915
Partenariat_Gestion de Fortune	1 238	1 443	1 640	2 001	2 357	2 776
Partenariat_Patrimonial	3 552	4 111	4 930	6 299	7 225	8 159
Run-off_Aisé	4 569	4 781	4 578	4 389	4 088	3 798
Run-off_Gestion de Fortune	250	236	213	204	193	181
Run-off_Patrimonial	1 516	2 403	2 246	2 139	2 004	1 847
Run-off_Retail	24 233	23 497	23 132	23 777	25 856	28 129
Total général	44 452	47 511	49 032	51 715	55 041	58 805

Tableau 10 : Nombre de contrats par segment de partenariat

Typologie_Partenariat	2010	2011	2012	2013	2014	2015
Partenariat_Aisé	19,69%	20,32%	21,37%	20,15%	19,01%	18,34%
Partenariat_Gestion de Fortune	29,43%	29,56%	29,76%	30,92%	33,32%	34,86%
Partenariat_Patrimonial	25,43%	25,94%	27,44%	30,65%	31,16%	31,90%
Run-off_Aisé	6,60%	5,57%	5,09%	4,33%	3,73%	3,27%
Run-off_Gestion de Fortune	7,29%	6,23%	5,30%	4,50%	4,07%	3,54%
Run-off_Patrimonial	7,87%	8,90%	7,71%	6,32%	5,43%	4,63%
Run-off_Retail	3,70%	3,48%	3,32%	3,13%	3,27%	3,47%
Total général	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%

Tableau 11 : Répartition de l'encours par segment de partenariat

Le tableau ci-dessus permet de constater que, depuis le 31/12/2010 plus de 75% de l'encours sont détenus par des contrats dont le partenariat est en cours. Ces statistiques démontrent également que la politique commerciale d'AEP, qui consiste à orienter l'encours du portefeuille vers une clientèle

dite « Haut de Gamme », semble fonctionner : l'encours sur les contrats Gestion de Fortune et Patrimonial ne cesse de croître depuis 2010 (ainsi que le nombre d'affaires nouvelles).

Typologie_Partenariat	2011	2012	2013	2014	2015
Partenariat_Aisé	8%	8%	6%	7%	6%
Partenariat_Gestion de Fortune	9%	12%	6%	6%	5%
Partenariat_Patrimonial	12%	11%	8%	10%	6%
Run-off_Aisé	17%	13%	12%	11%	13%
Run-off_Gestion de Fortune	11%	11%	5%	3%	6%
Run-off_Patrimonial	8%	12%	9%	9%	9%
Run-off_Retail	14%	14%	12%	9%	9%

Tableau 12 : Montant de rachats rapporté à l'encours

Au vu des premiers indicateurs, le taux de rachat est légèrement supérieur sur les contrats des partenariats en « Run-off », avec un taux plus élevé sur la classe « Run-off Aisé ».

En revanche, la répartition entre Rachats Totaux et Rachats partiels est clairement différente :

Typologie_Partenariat	2011	2012	2013	2014	2015
Partenariat_Aisé	34%	28%	34%	29%	28%
Partenariat_Gestion de Fortune	16%	49%	17%	36%	18%
Partenariat_Patrimonial	29%	34%	27%	24%	28%
Run-off_Aisé	50%	60%	50%	58%	72%
Run-off_Gestion de Fortune	81%	78%	45%	60%	82%
Run-off_Patrimonial	35%	42%	42%	54%	54%
Run-off_Retail	78%	78%	75%	70%	65%

Tableau 13 : Proportion des rachats totaux sur les montants de rachats constatés

La suite de l'étude nous permettra de savoir si la segmentation proposée est opportune ou si des agrégations sont possibles du fait d'un comportement de rachat similaire entre ces différents segments de partenariats.

1.2.4 Les différentes possibilités d'investissement

Le support d'investissement d'un contrat est, *a priori*, une variable déterminante dans le comportement de rachat d'un assuré. Un contrat investi sur des supports en unités de compte présente *a priori* peu de risque d'être racheté, surtout si la performance des supports investis est attractive. De même, un contrat investi sur des supports risqués est davantage considéré par les assurés comme un contrat de « placement » que comme un outil de trésorerie : l'assureur doit *a priori* s'attendre, sur ces contrats, à un rachat inférieur à celui des contrats investis uniquement sur du fonds en euro.

Sur notre portefeuille d'étude, nous constatons ci-dessous que le pourcentage d'encours investi sur les supports en unités de compte est supérieur à la moyenne constatée sur les contrats d'assurance-vie du marché (en 2015, les unités de compte représentaient 18% des encours du marché)²³ :

Supports	2010	2011	2012	2013	2014	2015
Fonds en Euros	53,83%	54,66%	52,31%	51,90%	52,36%	53,04%
Unités de Compte	46,17%	45,34%	47,69%	48,10%	47,64%	46,96%

Tableau 14 : Répartition de l'encours selon le support d'investissement

Les hypothèses formulées ci-dessus, semblent se confirmer au vu du taux de rachat plus faible constaté sur les supports unités de compte :

Supports	2011	2012	2013	2014	2015
Fonds en Euros	13,45%	12,80%	9,58%	9,92%	7,82%
Unités de Compte	6,55%	9,09%	4,92%	5,90%	4,26%

Tableau 15 : Montant de rachats rapporté à l'encours

Etant donné le contexte actuel de taux bas, et afin de diminuer le risque porté par l'assureur, la collecte sur des supports en unités de compte est mise en avant par l'assureur. Cependant, l'aversion au risque des assurés fait perdurer l'existence de contrats investis uniquement sur du fonds en euros. Les tableaux ci-dessous présentent la répartition de l'encours selon le profil d'investissement de l'assuré :

Investissements	2010	2011	2012	2013	2014	2015
Mono-support Euros	27,64%	27,13%	27,35%	27,38%	26,80%	27,01%
Mono-support UC	18,46%	17,91%	20,16%	19,25%	17,45%	16,49%
Multi-supports	53,90%	54,96%	52,49%	53,37%	55,74%	56,50%

Tableau 16 : Répartition de l'encours selon le support d'investissement

Multi-Supports	2010	2011	2012	2013	2014	2015
Fonds en Euros	48,58%	50,10%	47,57%	45,93%	45,85%	46,08%
Unités de Compte	51,42%	49,90%	52,43%	54,07%	54,15%	53,92%

Tableau 17 : Répartition de l'encours entre le fonds en euros et les unités de compte pour les contrats détenant ces deux supports

Les contrats détenant à la fois du fonds en euros et des supports en unités de compte affichent une prépondérance de l'encours investi sur des supports en unités de compte.

Par ailleurs, les taux de rachats constatés selon le profil d'investissement corroborent nos hypothèses de comportement émises plus haut :

Investissements	2011	2012	2013	2014	2015
Mono-support Euros	14,22%	16,18%	13,26%	15,29%	13,68%
Mono-support UC	6,72%	9,49%	9,82%	12,90%	7,81%
Multi-supports	9,51%	10,95%	7,13%	9,27%	8,43%

Tableau 18 : Montant de rachats rapporté à l'encours

²³Source : <https://www.ffa-assurance.fr/content/tableau-de-bord-de-assurance-en-2016>

Nous continuerons donc d'effectuer la distinction entre ces différents profils d'investissement pour analyser le comportement de rachat des assurés de notre portefeuille d'étude.

Au-delà de la distinction fonds en euros ou supports en unités de compte, différentes offres commerciales/possibilités d'investissements particulières sont proposées aux assurés. Par exemple, lors d'un investissement sur le fonds en euros, il est possible de bénéficier d'une offre commerciale généralement limitée dans le temps. Commercialement, il est intéressant de savoir si le fait d'investir sur un support particulier peut influencer le comportement de rachat du contrat.

On peut alors différencier les offres commerciales proposées sur le fonds en euros et les investissements sur des supports en unités de compte qui diffèrent des OPCVM habituels.

1.2.4.1 Les offres sur le fonds en euros en « Run-off »

- Taux Technique : Certains contrats du portefeuille bénéficient de taux minimum garantis viagers. Ces offres ne sont plus commercialisées. Les contrats bénéficiant d'un taux technique sur leur fonds en euros représentaient 2,52% au 31/12/2010 sont passés à 1,16% en 2015. Cela s'explique par le fait que l'ancienneté de ces contrats était supérieure à la moyenne constatée du portefeuille car il s'agissait uniquement de contrats migrés depuis un portefeuille de contrats d'assurance-vie plus ancien. La représentativité de cette offre est trop peu significative pour identifier un comportement spécifique. **Par conséquent, nous ne prendrons pas en compte dans notre étude le critère de la présence oui/non du taux technique sur le contrat.**
- L'offre Europtimal+ : En avril de l'année N+1, l'entreprise calcule la moyenne arithmétique des taux de rendement de l'année N, nets de frais de gestion administrative, des fonds en euros de contrats multi-supports d'un panier défini contractuellement. Sont exclus du panier, les contrats des compagnies du groupe BNP Paribas Cardif et les contrats fermés aux nouvelles adhésions/souscriptions. Si cette moyenne calculée est supérieure au taux de rendement, net de frais de gestion administrative, attribué l'année N, au versement dans le cadre de l'offre Europtimal+, une contribution exceptionnelle sera attribuée le 30/04 de l'année N+1 pour la part du versement affectée au fonds en euros dans le cadre de l'offre Europtimal+ qui subsistait sur ce support au 31/12 de l'année N. La contribution exceptionnelle correspond à la restitution de tout ou partie des frais de gestion administrative prélevés au titre du contrat. Cette réduction de frais sera égale, dans la limite d'un plafond défini lors de la souscription, à la différence positive exprimée en pourcentage entre :
 - o la moyenne des taux de rendement nets des fonds en euros du panier pour l'année N,
 - o et le taux de rendement net de frais de gestion administrative attribué l'année N (issu du taux de revalorisation du fonds en euros ou Taux Minimum Annuel) au versement investi dans le cadre de l'offre Europtimal+.

Le montant de la contribution exceptionnelle s'obtient en appliquant la réduction de frais à la part de la valeur de rachat au 31/12/N affectée au fonds en euros issue du versement dans le cadre de l'offre Europtimal+ plafonnée à l'encours moyen de l'année N sur ce support

(tenant ainsi compte des versements et des rachats effectués durant l'année N sur le fonds en euros).

Cette offre est accessible sous condition du versement (initial ou complémentaire) d'un montant minimum défini contractuellement ; elle est valable pour les 4 clôtures annuelles suivant l'investissement.

Cette offre a été commercialisée de 2010 à 2013 (et combinée avec une offre TMA durant la première année).

Les contrats bénéficiant d'une offre Europtimal+ non couplée avec un taux TMA, représentent 0,01% de l'encours total des années de commercialisations au 31/12/2015 (le dernier versement de contribution exceptionnelle pouvait être réalisé le 30/04/2017). La représentativité de cette offre est trop peu significative pour identifier un comportement spécifique. **Par conséquent, nous ne prendrons pas en compte dans notre étude le critère de la présence oui/non de cette offre au contrat.**

1.2.4.2 Les offres sur le fonds en euros en cours de commercialisation

a. Alternative au fonds en euros classique

- L'offre Euroconviction : la valeur de rachat investie bénéficiant de cette offre est adossée aux actifs du fonds en euros pour une part comprise entre 75% et 100%, le reste étant investi sur une poche d'actifs dynamiques. Les sommes qui y sont investies sont garanties à tout instant (diminuées des retraits éventuels du fonds) et toute participation aux bénéfices attribuée est définitivement acquise. Cette offre permet donc de bénéficier de l'effet cliquet du fonds en euros tout en pouvant prétendre (mais sans garantie contractuelle) à un taux de revalorisation annuelle supérieur à celui du fonds en euros du fait de la part investie selon une allocation plus dynamique.

b. Offres commerciales temporaires

- Offre Taux Minimum Annuel (TMA) : Cette offre est accessible lors du versement initial ou complémentaire d'un montant minimal défini contractuellement. Le versement permettant de prétendre à cette offre peut être conditionné par un investissement minimal sur des unités de compte. Le taux de revalorisation définie au 31/12 de l'année N sera égal au maximum entre le TMA et le taux de revalorisation du fonds en euros du contrat. Les possibilités commerciales de cette offre étant intimement corrélées avec le taux moyen des emprunts d'Etat du fait de la réglementation qui encadre ces taux (article A132-3 du Code des assurances), et ces taux étant devenus très faibles, cette offre commerciale n'est plus attractive. D'où la diminution significative de l'investissement sur ces offres depuis 2014 ainsi que le pourcentage d'encours de contrats bénéficiant de cette offre (22% au 31/12/2013 vs 7% au 31/12/2015). Cette offre a également été commercialisée en combinaison avec l'offre Europtimal+ présentée ci-dessus et avec l'offre Bonus.
- Offre Bonus : Cette offre est accessible lors du versement initial ou complémentaire d'un montant minimal défini contractuellement. Le versement permettant de prétendre à l'offre Bonus doit être investi selon un pourcentage minimum, défini contractuellement, sur des

supports en unités de compte. La part du versement investie sur le fonds en euros bénéficiera d'une majoration exceptionnelle, brute de frais de gestion administrative, du taux de revalorisation brut de frais de gestion administrative du fonds en euros du contrat. Cette majoration est définie contractuellement selon le pourcentage investi sur des supports en unités de compte. Par exemple, le contrat bénéficiera d'une majoration du taux de participation aux bénéfices de 0,40% si le versement est réalisé sur au moins 25% d'unités de compte.

Le tableau ci-dessous présente la répartition de l'encours des contrats bénéficiant d'un TMA et/ou d'une offre Bonus, ou de l'offre Euroconviction :

Offres commerciales	2010	2011	2012	2013	2014	2015
TMA	7,52%	14,53%	18,54%	22,10%	14,11%	7,28%
Bonus	0,00%	0,00%	0,00%	3,71%	10,58%	9,83%
TMA et/ou Bonus	7,52%	14,53%	18,54%	22,10%	21,42%	16,78%
Euroconviction	1,33%	3,85%	3,96%	4,13%	4,73%	5,30%

Tableau 19 : Proportion de l'encours bénéficiant d'une offre commerciale

Le pourcentage indiqué en 2013 pour la ligne TMA et /ou Bonus est équivalent au pourcentage des contrats bénéficiant d'une offre TMA car en 2013 l'offre Bonus était proposée uniquement en combinaison avec un TMA.

Offres commerciales	2011	2012	2013	2014	2015
TMA	7,31%	8,26%	6,40%	10,61%	3,05%
Bonus	0,00%	0,00%	2,08%	6,21%	4,59%
TMA et/ou Bonus	7,31%	8,26%	6,40%	8,97%	4,15%
Euroconviction	4,79%	7,34%	4,61%	3,97%	5,36%
Aucune offre spécifique	12,24%	13,34%	8,12%	7,72%	7,19%

Tableau 20 : Montant de rachats rapporté à l'encours

L'agrégation des deux offres commerciales permettrait d'avoir un flux continu depuis 2011 de contrats bénéficiant d'une offre commerciale de majoration de PB en fin d'année. Nous constatons également que les contrats bénéficiant d'une offre commerciale, ou investis sur le support Euroconviction, ont un taux de rachat inférieur aux contrats qui ne sont pas investis sur ces dispositifs. Cela confirme, pour le moment, notre intuition sur le comportement des souscripteurs. Lorsque nous comparons la répartition de la typologie de rachats effectués par les contrats bénéficiant d'une offre TMA ou d'une offre Bonus, nous remarquons que la répartition des types de rachats est similaire :

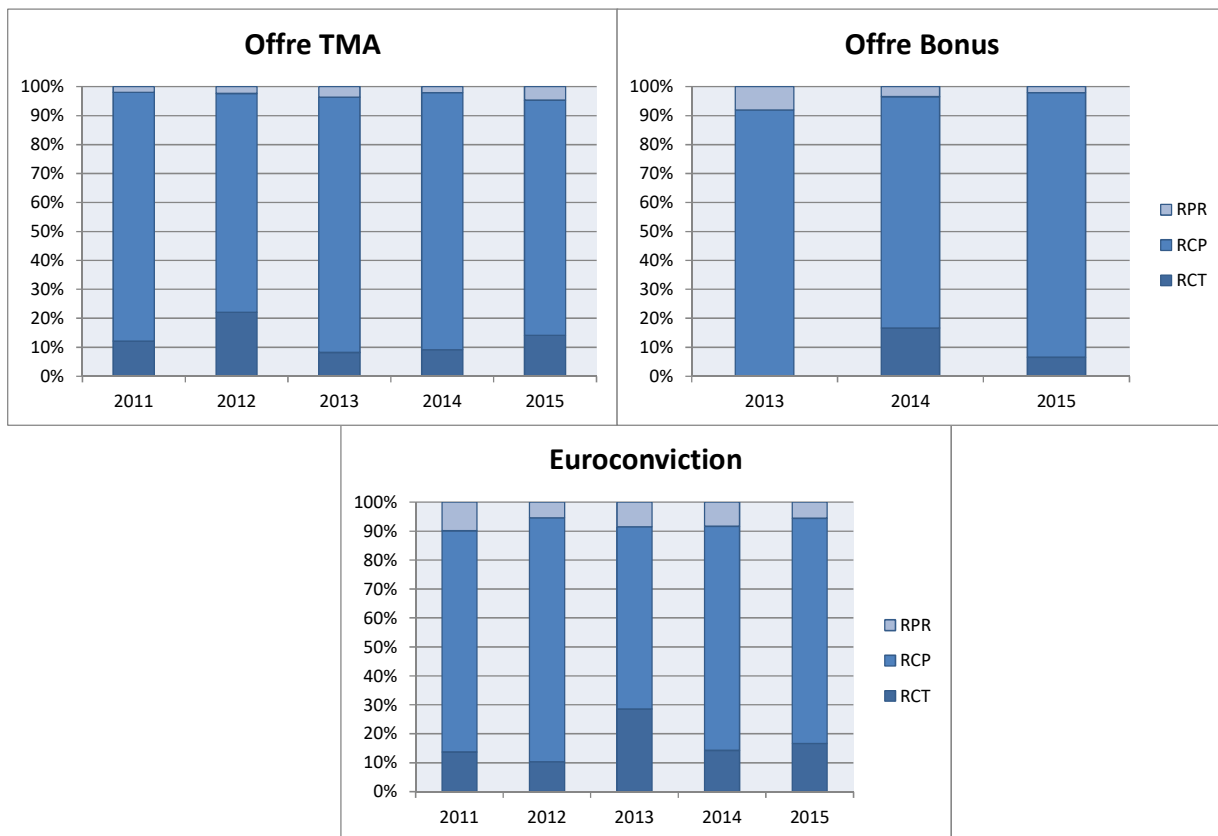


Figure 5 : Histogramme de répartitions des rachats de 2011 à 2015, selon l'offre commerciale

Afin d'avoir une homogénéité des données, c'est-à-dire de ne pas avoir de « saut » du fait des offres commerciales limitées dans le temps, ce qui permet d'avoir une plus grande robustesse de nos données, **il est décidé pour la suite de l'étude d'utiliser les critères discriminants :**

- **Bénéficiaire d'une offre TMA et/ou Bonus ou ne bénéficie pas d'une de ces offres,**
- **Investissement ou non sur le fonds Euroconviction.**

1.2.4.3 Gestion d'unités de compte spécifiques

Lors d'un investissement sur des unités de compte, il est possible de souscrire :

- Sur des Supports Immobiliers :
 - o Les SCPI (Sociétés Civiles de Placement Immobilier) : La plupart des SCPI ont pour vocation à investir dans de l'immobilier tertiaire, bureaux et commerces en tête. Le but est de dégager un rendement via l'acquisition et la gestion d'un patrimoine immobilier locatif. Les SCPI peuvent effectuer à la marge des travaux d'aménagements, acheter des biens d'équipement et revendre de façon ponctuelle certains immeubles mais elles ont interdiction formelle de construire des biens immobiliers. Au 31/12/2015, 95% de l'investissement en supports immobiliers est réalisé sur des SCPI.

- Les SCP (Sociétés Civiles de Portefeuille) de SCPI : Les SCP investies dans le portefeuille étudié sont des sociétés civiles à capital variable composées principalement de parts de SCPI d'entreprise. Au 31/12/2015, 3,8% de l'investissement en supports immobiliers était réalisé sur des SCP de SCPI.
- Les SCI (Sociétés Civiles immobilières) : Les SCI permettent à plusieurs personnes, physiques ou morales de détenir ensemble un ou des biens immobiliers. Les SCI disponibles mêlent dans leur portefeuille des immeubles, des parts de SCPI ou même des parts de sociétés foncières. Au 31/12/2015, 0,66% de l'investissement en supports immobiliers était réalisé sur des SCI.
- Les OPCI (Organismes de Placement Collectif Immobilier) : si la majorité de son actif doit être investi en fonds immobilier, elles disposent d'une enveloppe pouvant être placés sur les marchés financiers (obligation d'avoir 40% d'actifs liquides dans une OPCI). Cette diversification permet à ce fonds une liquidité intrinsèque plus forte qu'une part de SCPI. Cette catégorie immobilière est disponible en investissement sur le portefeuille depuis 2016.

Chez AEP, l'investissement sur cette catégorie de supports est accepté uniquement en versements complémentaires et l'allocation globale en supports immobiliers ne peut pas excéder 30% de la valeur de rachat totale du contrat. De plus, les conditions contractuelles définissent un montant minimal d'investissement et un montant maximal d'investissement total.

Enfin, des frais de sorties sont appliqués (3% sur le prix de retrait) si des rachats sont réalisés sur les unités de compte SCPI avant le troisième anniversaire de l'investissement. Ces frais de sorties ne sont pas appliqués sur les supports SCI/SCP ou OPCI.

Du fait de la liquidité (*a priori*) plus faible de ces supports et des frais de sorties appliqués sur les supports SCPI, on pourrait légitimement supposer que le comportement de rachat des assurés ayant investi sur un de ces supports soit différent.

Remarque : L'investissement sur les OPCI étant possible uniquement depuis 2016 sur nos contrats, nous ne pouvons pas savoir, à ce stade si le comportement de rachats des assurés investis sur ces supports s'apparentera davantage à celui constaté sur les supports en unités de compte « classiques » ou immobiliers. Ce point pourra être étudié lors d'une actualisation de l'étude, lorsque nous aurons plus de recul sur la commercialisation de ces supports.

Mode de gestion sous mandat des supports en unités de compte (« UC Mandat ») : Dans ce cas, le souscripteur délègue sa faculté de sélection et d'arbitrages entre les supports en unités de compte à un tiers agréé, en signant un mandat d'arbitrage.

Offres commerciales	2010	2011	2012	2013	2014	2015
Immobiliers	10,18%	10,83%	11,94%	12,84%	13,30%	13,55%
UC Mandat	18,66%	21,42%	21,45%	22,31%	22,21%	20,71%

Tableau 21 : Proportion de l'encours ayant une unité de compte « spécifique »

Offres commerciales	2011	2012	2013	2014	2015
Supports Immobiliers	7,42%	8,14%	4,76%	4,97%	4,92%
UC Mandat	7,64%	11,79%	5,14%	5,35%	4,57%
Aucune offre spécifique	9,68%	9,11%	7,14%	8,54%	5,65%

Tableau 22 : Montant de rachats rapporté à l'encours

Dans la suite de l'étude, nous conserverons les critères discriminants suivants :

- Bénéficie d'un support immobilier ou ne bénéficie pas d'un support immobilier,
- Bénéficie d'une gestion sous mandat ou ne bénéficie pas d'une gestion sous mandat.

1.2.5 Ancienneté du contrat au moment du rachat

Comme indiqué au chapitre III.3.1 de la première partie, la fiscalité supportée en cas de rachat diffère selon l'ancienneté du contrat. Ce critère est donc considéré comme le facteur de risque principal. La fiscalité étant différente entre les personnes physiques et les personnes morales, ce facteur de risque est analysé par la suite selon le type de souscripteur.

1.2.5.1 Personnes physiques

Pour les personnes physiques, nous ne disposons pas, dans notre portefeuille, des données d'ancienneté supérieure à 13 ans suffisantes pour être interprétées. Nous agrégeons donc ces données en une seule variable.

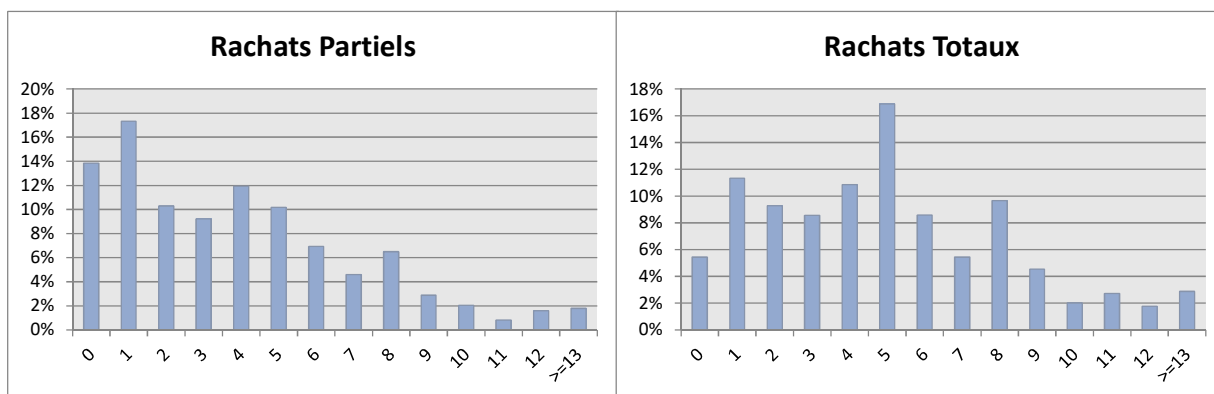


Figure 6 : Histogramme de répartition des rachats de 2011 à 2015 (cumulé) selon l'ancienneté

Nous constatons sur l'histogramme ci-dessus, quelques points d'accumulation, notamment sur les premières années. Nous observons également le pic des 8 ans, conséquence de la fiscalité de l'assurance-vie.

1.2.5.2 Personnes Morales

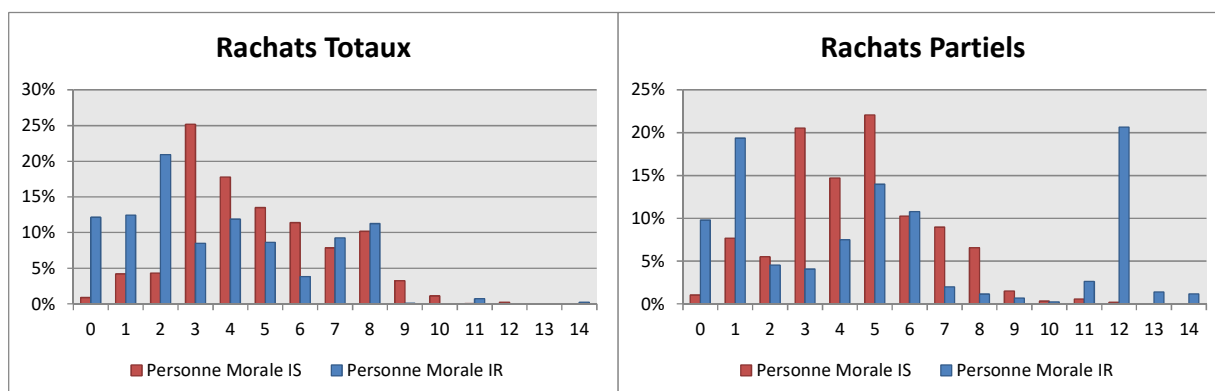


Figure 7 : Histogramme de répartition des rachats de 2011 à 2015 (cumulé) selon l'ancienneté

Concernant les personnes morales IS, les rachats sont nettement plus importants à partir de la troisième année d'ancienneté.

Cependant, si on regarde la répartition des rachats selon l'ancienneté par année, nous constatons que le comportement des assurés change au fil des ans. En effet, de 2011 à 2013, la proportion de rachat la plus importante concerne les contrats supérieurs à 3 ans²⁴. A partir de 2014, la proportion de rachats la plus importante est sur la quatrième année du contrat. Cela peut s'expliquer du fait de la modification des règles de dissuasion de rachats sur fonds en euros pour les contrats personnes morale IS. En effet, depuis 2011, les souscriptions sont soumises à un engagement à caractère déontologique²⁵ qui instaure le fait que la rémunération afférente au moins aux douze premiers mois des engagements en euros soit attribuée au contrat uniquement au terme du quatrième anniversaire de la souscription. Cette décision avait pour objectif d'inciter les assurés à conserver leur contrat pendant au moins quatre ans et ainsi à ne pas utiliser ces contrats pour des opérations de trésorerie. Avant la mise en place de cet engagement, des taux de pénalités de rachats dégressifs en fonction de l'ancienneté (0% après 3 ans d'ancienneté) s'appliquaient. Du fait d'un manque de recul sur cet engagement mis en place à depuis 2011 (les premiers contrats soumis à cet engagement ont atteint leur quatrième anniversaire en 2015), cette modification de comportement ne peut pas être totalement validée pour le moment. Cependant, cet axe de lecture devra être pris lors de l'actualisation de cette étude afin de corroborer ou non cette inflexion de comportement.

Concernant les contrats personnes morales IR, au vu des premiers constats, l'ancienneté ne semble pas être un axe de lecture discriminant pour les comportements de rachats.

²⁴ Cf. Histogrammes en annexe des rachats constatés année par année.

²⁵ Engagement relatif aux contrats d'assurance-vie et aux contrats de capitalisation souscrits par des personnes morales.

1.2.6 Tableaux de synthèse

L'étude qui suit portera uniquement sur les cas de Rachat total constatés sur les contrats. Les variables présélectionnées seront les suivantes :

Constat d'un Rachat Total sur le contrat dans l'année d'observation	Rachat Total = 1 si oui 0 sinon
Typologie souscription	3 possibilités : <ul style="list-style-type: none">- Assurance vie ou capitalisation PP- Capitalisation personnes morales IR- Capitalisation personnes morales IS
Réseaux de distribution	7 possibilités : <ul style="list-style-type: none">- Partenariat Aisé- Etc.
Présence de fonds en euros	Variable binaire (Oui/Non) ²⁶
Présence d'unités de compte	Variable binaire (Oui/Non)
Offre commerciale TMA et/ou Bonus	Variable binaire (Oui/Non)
Fonds Euroconviction	Variable binaire (Oui/Non)
Supports immobiliers	Variable binaire (Oui/Non)
UC Mandat	Variable binaire (Oui/Non)
Résidence	1 (Non-Résidents) ou 2 (Résidents)
Ancienneté du contrat	14 possibilités : <ul style="list-style-type: none">- 0 à 12- Supérieure ou égale à 13

Pour les variables binaires, celles-ci sont déterminées :

- sur la valeur de rachat à la date de l'évènement : pour les contrats avec sortie définitive en cours de l'année d'observation (Rachat Total/ Décès etc.)
- sur l'encours présent au 31/12/N-1 : pour les contrats en vigueur au 31/12/N-1 et 31/12/N
- sur l'encours présent au 31/12/N : pour les contrats souscrits au cours de l'année N et encore en vigueur au 31/12/N

Une fois effectuée cette agrégation de classifications en raison de la taille de notre portefeuille, notre objectif consiste maintenant à sélectionner le minimum de données apportant le maximum d'informations. Les chapitres suivants ont pour objectif de sélectionner les variables les plus significatives sur la décision de rachat. Plusieurs méthodes statistiques sont possibles. Afin de valider un modèle, il est nécessaire de réaliser des méthodes distinctes pour mettre au jour les variables convergentes de chaque modèle.

La suite de l'étude va présenter trois modèles de sélection. Ces différents modèles sont appliqués pour chaque année d'observation et sur une base agrégée.

L'année 2016 sera utilisée pour tester la viabilité des modèles réalisés et leur pouvoir de prédiction.

²⁶ Dans notre data frame, Oui=1 / Non=0 pour toutes les variables binaires

II Analyse des Correspondances Multiples (ACM)

II.1 Présentation du modèle

L'Analyse des Correspondances Multiples (ACM) appartient à la famille des méthodes d'analyse factorielle. Elle permet d'expliquer les dépendances entre différentes variables, décrivant des individus, à l'aide de « nouvelles variables » créées, appelées facteurs. Cette méthode est adaptée uniquement lorsque toutes les variables sont qualitatives.

L'ACM permet d'étudier une population de I individus décrits par J variables qualitatives. Pour chaque variable qualitative, il existe plusieurs modalités de réponse de la variable.

Un individu est considéré dans l'ensemble de ses modalités. L'ensemble de ses ressemblances et de ses différences est appelé variabilité des individus. Ainsi, le premier objectif de l'ACM est d'étudier la variabilité des individus du point de vue multidimensionnel.

L'ACM s'intéresse également aux liaisons entre les variables. Pour cela, on visualise les associations entre modalités. Deux variables qualitatives sont liées si les modalités de l'une s'associent de façon particulière aux modalités de l'autre. Cela permet une visualisation d'ensemble des associations entre les modalités et donc une visualisation d'ensemble des liaisons entre les variables. Cette visualisation globale des liaisons sera établie à partir de variables synthétiques quantitatives utilisées pour construire les plans factoriels. En analyse factorielle, l'idée majeure est de choisir des variables synthétiques ayant la plus forte variance possible.

II.1.1 Représentation des données

Les données peuvent être rassemblées dans un tableau *Individus* \times *Variables*.

La structure du tableau des données brutes initiales est :

- Chaque ligne i correspond à un individu (statistique) ;
- Chaque colonne j correspond à une variable (qualitative) ;
- A l'intersection de la ligne i et de la colonne j se trouve la valeur x_{ij} qui correspond à la modalité possédée par l'individu i pour la variable j .

Cette structure, fournie aux logiciels d'ACM, est appelée « tableau de codage condensé » :

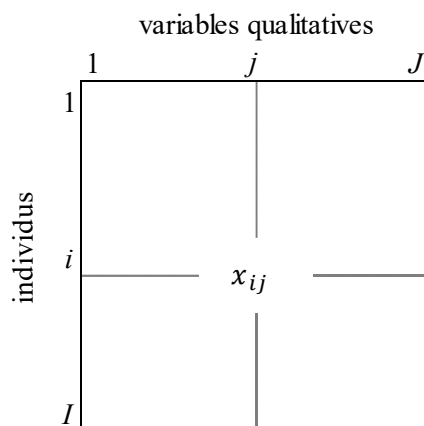


Figure 8 : Tableau de codage condensé

Le tableau sur lequel se fonde le programme d'ACM est le tableau disjonctif complet (TDC). Soit la variable j possédant k modalités : le TDC comprend à l'intersection de la ligne i et de la colonne k la valeur x_{ik} qui vaut 1 si l'individu i possède la modalité k et vaut 0 sinon. L'origine de la terminologie « Tableau Disjonctif Complet » vient du fait que l'ensemble des valeurs x_{ik} d'un même individu, pour les modalités d'une même variable, comporte la valeur 1 une fois (complet) et une fois seulement (disjonctif).

On définit :

- K_j nombre et ensemble des modalités de la variable j .
- $K = \sum_{j=1}^J K_j$ nombre et ensemble des modalités toutes variables confondues.

Le Tableau Disjonctif Complet peut se représenter sous la forme suivante :

		variable 1				variable j			variable J				
		1				1	k	K_j	K				Somme
individus	1												J
	I	0	1	0	0	x_{ik}			0	0	1	0	J
Somme		I_1				I_k			I_k				IJ

Figure 9 : Représentation du Tableau Disjonctif Complet

Un TDC possède non seulement une nature différente de celle d'un tableau de contingence mais aussi des propriétés numériques particulières :

- Les valeurs dans le tableau ne sont que des 0 et des 1 ;
- Les colonnes peuvent être regroupées par bloc (qui correspondent chacun à une variable) dont la somme est une colonne composée de 1 ;
- La somme des nombres d'une même ligne est constante et égale J .

II.1.2 Nuage des individus

Le poids d'un individu vaut $\frac{1}{I}$. En effet, dans l'étude, nous considérons que tous les individus ont le même degré d'importance.

Un individu est représenté par les modalités qu'il possède. Deux individus se ressemblent s'ils présentent globalement les mêmes modalités. La distance entre deux individus i et l est définie par :

$$d^2(i, l) = \sum_k \frac{IJ}{I_k} \left(\frac{x_{ik}}{J} - \frac{x_{lk}}{J} \right)^2 = \frac{1}{J} \sum_k \frac{I}{I_k} (x_{ik} - x_{lk})^2$$

L'expression $(x_{ik} - x_{lk})^2$ vaut 0 ou 1 et ne diffère de 0 que pour les modalités k possédées par un seul des deux individus i et l . La distance $d(i, l)$ croît donc avec le nombre de modalités qui diffèrent pour les individus i et l .

La modalité k intervient dans cette distance avec le poids $\frac{I}{I_k}$: la présence d'une modalité rare éloigne son ou ses possesseurs de tous les autres individus.

La distance entre un individu i et le centre de gravité du nuage des individus, G_I est :

$$d^2(i, G_I) = \frac{I}{J} \sum_K \frac{1}{I_k} - 1$$

Le centre de gravité du nuage des individus, G_I , a pour coordonnée pour la modalité k , $\frac{I_k}{IJ}$, proportion, au coefficient J près, des individus ayant choisi la modalité k . Il peut s'interpréter comme un individu théorique « moyen ». Un individu est d'autant plus éloigné de G_I qu'il possède des modalités rares.

Par ailleurs, l'inertie totale du nuage des individus est égale à : $\sum_k \frac{I_k}{I} d^2(i, G_I) = \frac{K}{J} - 1$

L'inertie totale ne dépend pas du contenu du tableau mais du nombre moyen de modalités par variable. Cette donnée n'intervient pas dans l'interprétation des résultats de l'ACM.

II.1.3 Nuage des modalités

Le poids de la modalité k est proportionnel à l'effectif I_k et vaut $\frac{I_k}{IJ}$.

La distance entre deux modalités k et h est définie par :

$$\begin{aligned} d^2(k, h) &= \sum_i I \left(\frac{x_{ik}}{I_k} - \frac{x_{ih}}{I_h} \right)^2 \\ &= \frac{I}{I_k I_h} [\text{nombre d'individus possédant une et une seule des modalités } h \text{ et } k] \end{aligned}$$

Cette distance croît avec le nombre d'individus possédant une et une seule des modalités h et k , et décroît avec l'effectif de chacune de ces modalités. Deux modalités d'une même variable sont obligatoirement assez éloignées l'une de l'autre dans l'espace. Deux modalités possédées par les mêmes individus sont confondues. Les modalités rares sont éloignées de toutes les autres.

La distance entre le centre de gravité G_K et une modalité k est égale à :

$$d^2(k, G_K) = \left(\frac{I}{I_k} \right) - 1$$

Celle-ci indique l'influence de l'effectif d'une modalité sur sa distance au point moyen.

L'inertie de la modalité k vaut : $\frac{I_k}{IJ} \times d^2(k, G_K) = \frac{1}{J} \times \left(1 - \frac{I_k}{I}\right)$

Une modalité a donc une influence *a priori* d'autant plus grande qu'elle est rare. Par exemple, une modalité présente dans 1% seulement de la population possède une influence presque deux fois plus grande qu'une modalité présente dans 50% de la population. Ce résultat est assez intuitif, pour différencier les individus, les modalités très fréquentes ont peu d'intérêt ; une modalité rare attire l'attention sur le(s) individu(s) qui la possède(nt). Dans la pratique, l'étude globale vise à mettre en exergue des phénomènes généraux : il est fortement conseillé de regrouper les modalités rares.

II.1.4 Variables qualitatives

Jusqu'ici, les variables qualitatives n'ont été considérées qu'au travers de leurs modalités. En effet, un individu est caractérisé par les modalités qu'il possède ; la liaison entre deux variables qualitatives s'analyse au travers des associations entre leurs modalités. Cependant, les variables en tant que telles interviennent également dans l'interprétation.

L'inertie totale des K_j modalités k de la variable j est égale : $\frac{K_j - 1}{J}$. De ce fait, cette inertie est d'autant plus grande que la variable possède beaucoup de modalités.

En fait, cette inertie doit être mise en regard de la dimension $K_j - 1$ du sous-espace engendré par les modalités de la variable j : plus l'inertie est élevée, plus elle se répartit selon un grand nombre de dimensions.

En fait, l'inertie des modalités de la variable j est égale à la constante $\frac{1}{J}$ en projection dans n'importe quelle direction du sous-espace qu'elles engendrent. Ainsi, dans la recherche du premier axe d'inertie, aucune variable n'est privilégiée. En revanche, une variable possédant deux modalités (oui/ non) ne peut être liée fortement qu'à un seul axe (opposant donc ces deux réponses), alors qu'une variable possédant plusieurs modalités peut être liée à plusieurs dimensions.

Si l'on projette la modalité k sur un vecteur unitaire centré v de \mathbb{R}^I (dont la i^e coordonnée v_i est la valeur de la fonction v pour l'individu i), la longueur de cette projection vaut (en notant \bar{v}_k la moyenne de la fonction v pour les individus présentant la modalité k) :

$$\langle k, v \rangle = \sum_i \frac{1}{I} \left(\frac{Ix_{ik}}{I_k} - 1 \right) v_i = \bar{v}_k$$

Ensuite l'inertie projetée de K_j sur v vaut : $\frac{1}{J} \sum_k \frac{I_k}{I} \bar{v}_k^2 = \frac{1}{J} \eta^2(j, v)$

Au coefficient $\frac{1}{J}$ près, cette inertie est égale à l'inertie interclasses de la fonction v pour la partition des individus définis par la variable j . L'inertie totale de v valant 1, cette inertie interclasses peut être vue comme le pourcentage d'inertie (de v) « expliqué » par la variable qualitative j . Elle est donc confondue avec le carré du rapport de corrélation, noté $\eta^2(j, v)$, entre d'une part la variable qualitative j et d'autre part la fonction numérique v . Cet indicateur s'interprète géométriquement comme le carré du cosinus de l'angle entre la variable quantitative v et le sous-espace engendré par la variable qualitative j : c'est la mesure de liaison classiquement utilisée entre une variable quantitative et une variable qualitative.

II.1.5 Projection

Le produit scalaire entre deux vecteurs u et v de \mathbb{R}^K s'écrit :

$$\langle u, v \rangle_M = u' M v = \sum_k \frac{I_k}{IJ} u_k v_k$$
$$\langle u, u \rangle_M = u' M u = \sum_k \frac{I_k}{IJ} u_k^2$$

Avec M la matrice diagonale des poids.

Avec les notations suivantes :

F_s : L'ensemble des projections de tous les points du nuage d'individus sur le s^e axe factoriel u_s

G_s : L'ensemble des projections des variables sur le s^e axe factoriel v_s .

Les relations de transition de l'ACM s'écrivent :

$$F_s(i) = \frac{1}{\sqrt{\lambda_s}} \sum_{k \in K} \frac{x_{ik}}{J} G_s(k)$$
$$G_s(k) = \frac{1}{\sqrt{\lambda_s}} \sum_{i \in I} \frac{x_{ik}}{I_k} F_s(i)$$

Ces relations signifient qu'en projection sur l'axe s , l'individu i est placé au coefficient $\frac{1}{\sqrt{\lambda_s}}$ près au barycentre des modalités qu'il possède. De même, la modalité k est placée au coefficient $\frac{1}{\sqrt{\lambda_s}}$ près au barycentre des individus qui la possèdent. Cela signifie qu'une modalité représente, à une dilatation près, la moyenne des individus qui la possèdent.

Par ailleurs, la quantité maximisée par les axes factoriels dans l'espace \mathbb{R}^I est l'inertie projetée du nuage de l'ensemble des modalités. En regroupant les modalités d'une même variable, ce critère n'est autre que la moyenne des carrés des rapports de corrélation entre le facteur et chacune des variables. Il en résulte que les facteurs F_s de l'ACM sont les variables numériques les plus liées à l'ensemble des variables qualitatives étudiées et, en ce sens, constituent les variables synthétiques annoncées.

II.1.6 Contribution à un axe factoriel

La contribution d'une modalité k peut être définie de la façon suivante :

$$CTR_s(k) = \frac{I_k}{IJ} G_s(k)^2$$

On obtient, en notant $\bar{F}_s(k)$, la coordonnée du barycentre des individus possédant la modalité k , dans \mathbb{R}^K :

$$CTR_s(k) = \frac{I_k}{IJ} \frac{1}{\lambda_s} \bar{F}_s(k)^2$$

La contribution d'une modalité k peut donc être exprimée aussi dans \mathbb{R}^K : elle est proportionnelle à l'inertie du barycentre des individus possédant la modalité k . Cela signifie que si l'on considère, dans \mathbb{R}^K , le nuage des barycentres associés aux modalités, il apparaît que les axes de l'ACM séparent aussi « au mieux » ces barycentres.

II.1.7 L'usage de variables supplémentaires

On peut distinguer deux groupes de variables : les variables actives qui déterminent les axes et les variables supplémentaires qui ne participent pas au calcul des valeurs et des vecteurs propres. Mis à part cela, les éléments supplémentaires sont traités comme les variables actives sur la représentation de celles-ci dans les dimensions de l'ACM.

II.2 Détermination du modèle

Dans notre étude, la variable Rachat Total (oui/non) est considérée comme une variable supplémentaire dans les résultats de l'ACM. Toutes les autres variables sont considérées pour le moment comme des variables actives de l'analyse.

Comme indiqué précédemment, les modalités « rares » sont à éviter pour avoir des résultats globaux assez robustes. Une représentation de la fréquence des modalités est réalisée pour savoir si des regroupements supplémentaires à ceux déjà effectués dans la partie descriptive sont nécessaires. Pour cela, le seuil de représentativité pris en compte est de 1% de la population observée.

Concernant la variable présentant le plus de modalités de l'étude (i.e. la variable représentant l'ancienneté du contrat), voici ci-dessous les fréquences de chaque ancienneté :

Ancienneté \ Année étudiée	2011	2012	2013	2014	2015
0	12,0%	10,6%	11,4%	12,0%	12,0%
1	12,1%	11,4%	10,0%	10,5%	11,0%
2	10,6%	11,0%	10,3%	9,0%	9,4%
3	9,7%	9,4%	9,9%	9,3%	8,0%
4	14,1%	8,7%	8,3%	8,7%	8,3%
5	15,8%	12,4%	7,5%	7,2%	7,5%
6	12,2%	14,0%	10,7%	6,6%	6,3%
7	7,7%	11,0%	12,4%	9,5%	6,0%
8	3,9%	6,8%	9,9%	11,2%	8,4%
9	0,8%	3,2%	5,6%	8,1%	9,4%
10	0,3%	0,6%	2,7%	4,6%	6,8%
11	0,2%	0,2%	0,5%	2,2%	4,0%
12	0,1%	0,2%	0,2%	0,5%	1,9%
13+	0,6%	0,6%	0,7%	0,8%	1,1%

Tableau 23 : Fréquence des anciennetés des contrats observées

Afin de pouvoir comparer les résultats de chaque année d'observation, les regroupements doivent être similaires pour tous les jeux de données. De plus, le nombre important de dimensions de l'analyse est dû au fait que cette variable possède beaucoup plus de modalités que toutes les autres

variables étudiées. Afin de gagner en lisibilité dans les résultats de l'étude de l'ACM, la variable Ancienneté sera dans un premier temps une variable supplémentaire (et non active) de l'ACM réalisée.

Aucune modalité des variables binaires de l'étude ne présente une fréquence inférieure à 1% sur toutes les années d'observations. Le tableau ci-dessous indique la fréquence uniquement pour la modalité la moins représentative de chaque variable :

Modalité des variables \ Année étudiée	2011	2012	2013	2014	2015
Absence de fonds en euros	51,9%	46,4%	42,3%	38,0%	34,3%
Absence d'unités de compte	17,4%	18,6%	19,2%	19,0%	18,4%
Présence de Supports immobiliers	10,2%	11,0%	12,1%	12,3%	11,8%
Présence d'une Gestion sous Mandat	2,4%	2,5%	2,7%	2,8%	2,8%
Présence du fonds Euroconviction	2,3%	3,7%	3,9%	4,3%	4,9%
Présence d'une Offre commerciale TMA et/ou Bonus	1,2%	3,0%	4,9%	5,2%	3,9%
Non-Résidents	1,4%	1,5%	1,5%	1,3%	1,2%

Tableau 24 : Fréquence des modalités les moins représentées

Selon le cadre fiscal du contrat :

Cadre Fiscal \ Année étudiée	2011	2012	2013	2014	2015
Assurance-Vie	95,7%	96,0%	96,2%	96,3%	96,5%
Capitalisation pers. morale IS	3,5%	3,1%	2,9%	2,8%	2,7%
Capitalisation pers. morale IR	0,8%	0,9%	0,9%	0,9%	0,8%

Tableau 25 : Fréquence des modalités du cadre fiscal

La modalité Capitalisation personnes morales soumises à l'IR n'est donc pas assez représentée. Cela pourrait induire une trop grande influence de cette modalité. Celle-ci va être agrégée avec la modalité la plus fréquente c'est à dire le cadre fiscal Assurance-Vie. En effet, nous avons préféré agrégés avec cette modalité pour deux raisons :

- la modalité de souscription de ces cadres fiscaux est similaire, contrairement aux contrats souscrits par des personnes morales soumises à l'IS (prime unique et frais de sorties conformément à l'engagement à caractère déontologique) ;
- cela permet de ne pas biaiser les résultats que nous pourrions constater sur les souscriptions personnes morales soumises à l'IS.

Enfin, selon le canal de distribution du contrat :

Réseau de distribution \ Année étudiée	2011	2012	2013	2014	2015
Partenariat_Aise	22,1%	23,9%	24,1%	23,5%	22,9%
Partenariat_Patrimonial	8,4%	9,8%	11,8%	12,9%	13,7%
Partenariat_Gestion de Fortune	2,9%	3,3%	3,7%	4,2%	4,6%
Run-off_Retail	51,2%	48,6%	47,2%	47,9%	48,5%
Run-off_Aise	10,0%	9,4%	8,6%	7,6%	6,7%
Run-off_Patrimonial	4,9%	4,6%	4,1%	3,7%	3,2%
Run-off_Gestion de Fortune	0,5%	0,5%	0,4%	0,3%	0,3%

Tableau 26 : Fréquence des réseaux de distribution

La modalité « Run-off Gestion de Fortune » est alors agrégée avec la modalité « Run-off Patrimonial » car c'est la modalité ayant les caractéristiques les plus proches de cette dernière.

Après cette étape préliminaire, pour plus de lisibilité du document, seuls les résultats de l'ACM sur les données sur le cumul des observations faites entre 2011 et 2015 seront présentés en détail. Par ailleurs, les résultats de l'ACM sont homogènes sur toutes les années d'observation prises séparément ; les informations communiquées ci-dessous sont valables pour tous les jeux de données. Les résultats et les sorties graphiques du logiciel sur les autres années d'observation sont indiqués en Annexe.

Nous réalisons l'ACM avec la fonction *MCA* de la librairie *FactoMineR* du logiciel *R*. Sur les données observées, $J = 9$ et $K = 22$ (8 variables binaires et 1 variable à 6 modalités). L'ACM est donc sur 13 dimensions ($K - J$).

L'inertie maximum expliquée par un axe dans notre étude est de : $\frac{J}{(K-J)} = 69,23\%$ et l'inertie moyenne est de : $\frac{1}{J} = \frac{1}{9}$.

Tout d'abord, nous devons sélectionner le nombre d'axes à étudier pour notre analyse. Ci-dessous l'histogramme présentant les valeurs propres de chaque dimension :

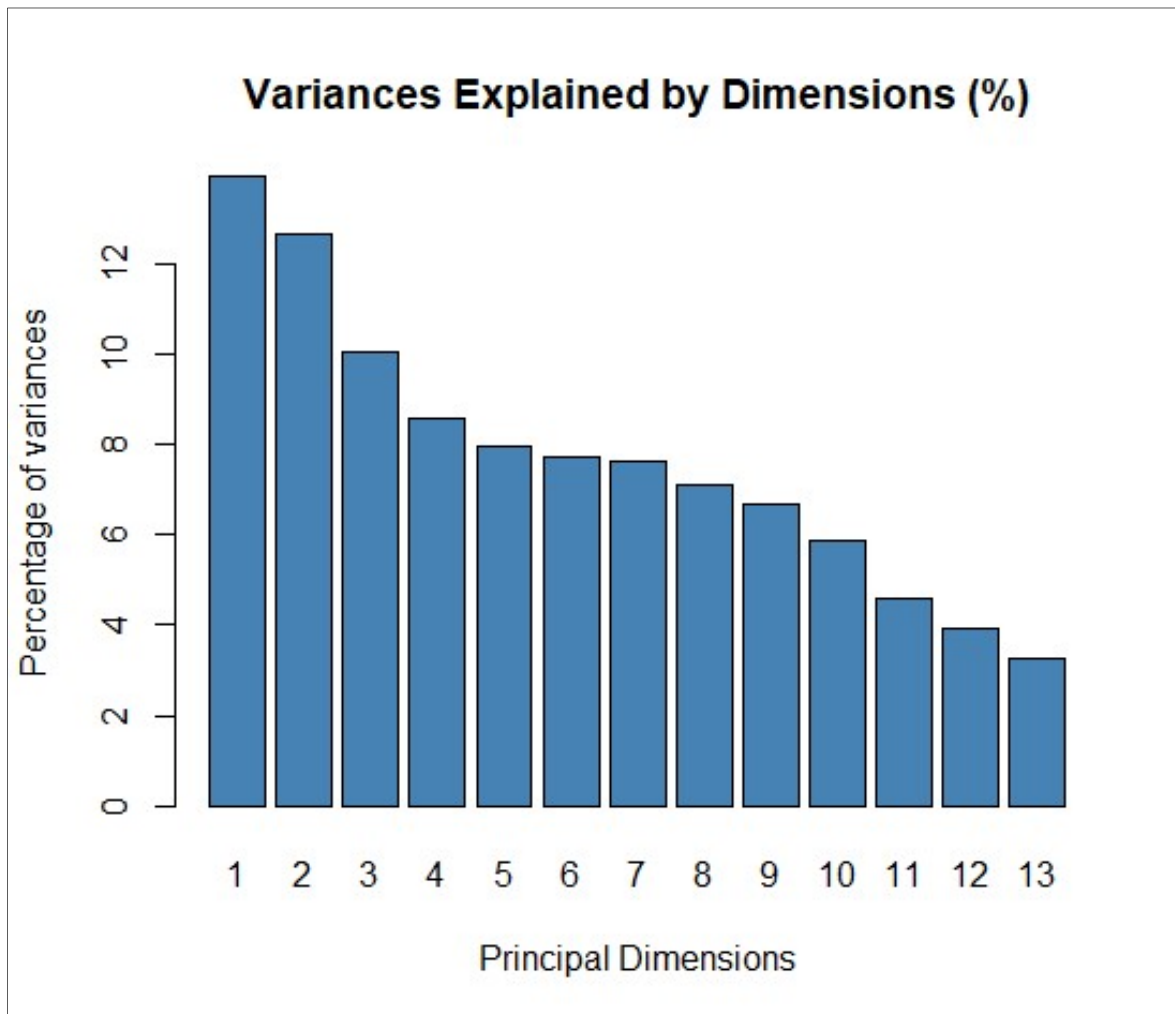


Figure 10 : Graphique décroissant des valeurs propres de l'ACM sur les données de 2011 à 2015

En pratique, la décision du nombre d'axes conservés est principalement basée sur des règles empiriques. La somme des valeurs propres est égale à l'inertie totale du nuage. Ainsi, chaque axe est caractérisé par le pourcentage d'inertie qu'il permet d'expliquer ; sont alors retenus uniquement les axes avec les plus fortes valeurs propres. Au vu de la délicatesse de ce choix, plusieurs règles ont été testées :

- **Critère de Kaiser** : Ce critère propose de garder les axes factoriels ayant des valeurs propres plus élevées que la moyenne des valeurs propres (moyenne = Inertie totale/nombre d'axes = $1/J$). En général, ce critère engendre un grand nombre d'axes retenus dans l'ACM. **Dans notre exemple, cela nous amènerait à conserver 6 axes.**
- **Critère du Coude** : si on observe un décrochement (ou coude) dans la décroissance, c'est-à-dire une accélération de cette dernière, suivie d'une décroissance plus régulière, alors les axes sélectionnés sont ceux situés avant le décrochement. Par analogie avec la recherche d'un point d'inflexion, ce critère propose de conserver les axes jusqu'au changement de signe dans la différence au second ordre. **Dans notre exemple, cela nous amènerait à conserver 1 seul axe.**

- Critère de l'inertie minimale : on sélectionne les premiers axes afin d'atteindre un pourcentage donné d'inertie expliquée. **Dans notre exemple, si nous souhaitons atteindre un minimum de 50 %, cela nous amènerait à conserver 5 axes.**
- Critère du bon sens : les plans et les axes sont analysés et on ne retient que ceux qui sont interprétables.

Par application du critère du bon sens, voici ci-dessous, la représentation graphique des premières dimensions de l'ACM. Sur ces graphiques, les modalités sont représentées par une couleur selon la valeur du \cos^2 .

$$\cos^2 = \frac{G_s(k)^2}{d^2(k, G_K)}$$

Cette valeur représente la qualité de la projection d'une modalité sur l'axe ; la somme des \cos^2 sur l'ensemble des axes vaut 1. Si une catégorie d'une variable donnée est bien représentée par deux dimensions, la somme des \cos^2 est proche de 1. Plus le \cos^2 est faible, plus il faut interpréter avec prudence la modalité sur le plan.

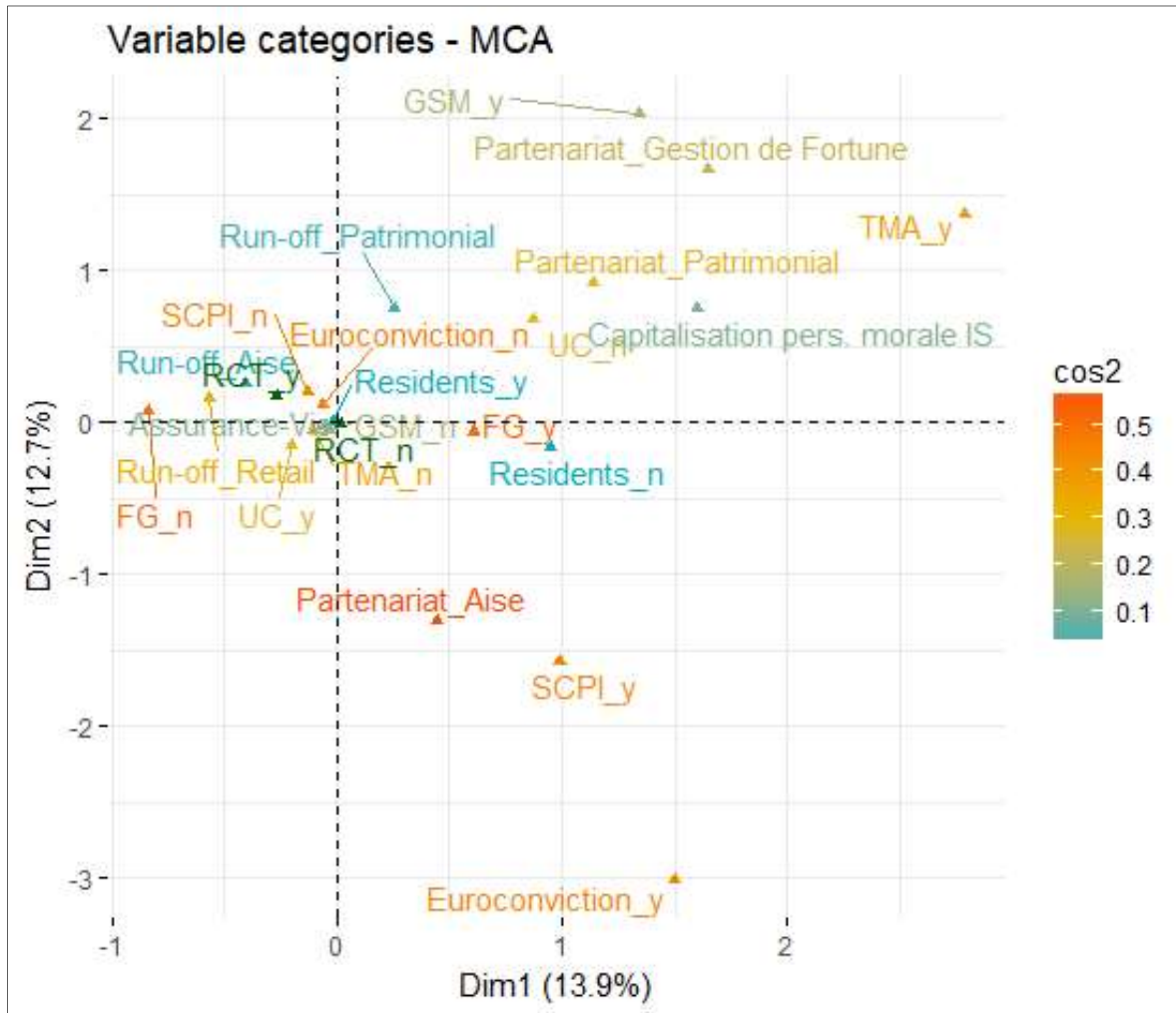


Figure 11 : Représentation des 2 premières dimensions de l'ACM sur les données de 2011 à 2015

La première dimension sépare les contrats selon la présence ou non de fonds en euros (FG_y et FG_n). Logiquement, les modalités de présence d'offres commerciales sur le fonds en euros (resp. les modalités d'absence d'offres commerciales sur le fonds en euros) se situent du même côté de l'axe que la modalité FG_y (resp. FG_n).

La deuxième dimension distingue les contrats selon qu'ils ont été souscrits via des partenaires toujours actifs ou non. D'un côté les contrats inférieurs à 100K€, et de l'autre ceux supérieurs à 100K€.

On peut remarquer que les contrats de capitalisation PM IS sont des contrats dont l'encours est supérieur à 100K€ (Patrimonial et Gestion de Fortune). Par ailleurs, ces contrats sont plus souvent investis à 100% sur le fonds en euros (proche de la modalité UC_no) et avec une offre commerciale (TMA_y). De même, les supports immobiliers et l'offre Euroconviction sont davantage souscrits par une clientèle « Aisé ».

Concernant la variable supplémentaire de Rachat Total, la principale information qui se dégage concerne les contrats non investis sur le fonds en euros. Pour ceux-ci la réalisation d'un rachat est plus prépondérante que sur les autres contrats. Hormis cela, les deux modalités de la variable

Rachat Total sont proches l'une de l'autre ; la distinction de ces deux dimensions fournit peu d'informations supplémentaires sur des variables discriminantes dans la réalisation d'un Rachat Total.

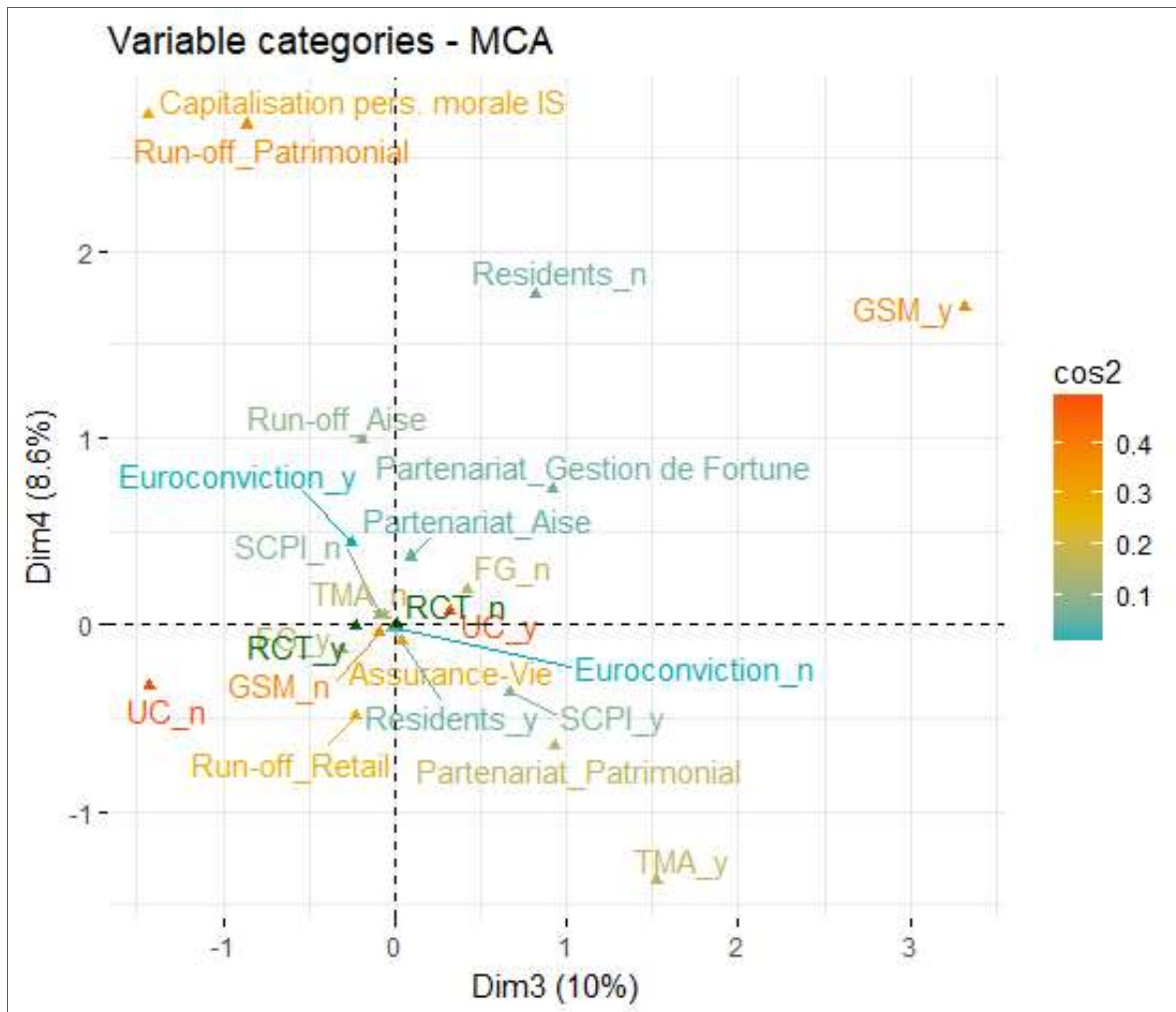


Figure 12 : Représentation des dimensions 3 et 4 de l'ACM sur les données de 2011 à 2015

Sur le graphique ci-dessus, la dimension 3 peut être décrite comme celle séparant les contrats présentant ou non des supports en unités de compte. Celle-ci fournit de façon logique les mêmes informations que la dimension n°1. La dimension 4 est celle séparant les contrats de capitalisation PM IS et les contrats d'Assurance Vie.

Concernant la variable supplémentaire de Rachat Total, c'est la même information qui se dégage : les contrats non investis sur le fonds en euros auraient tendance à racheter plus.

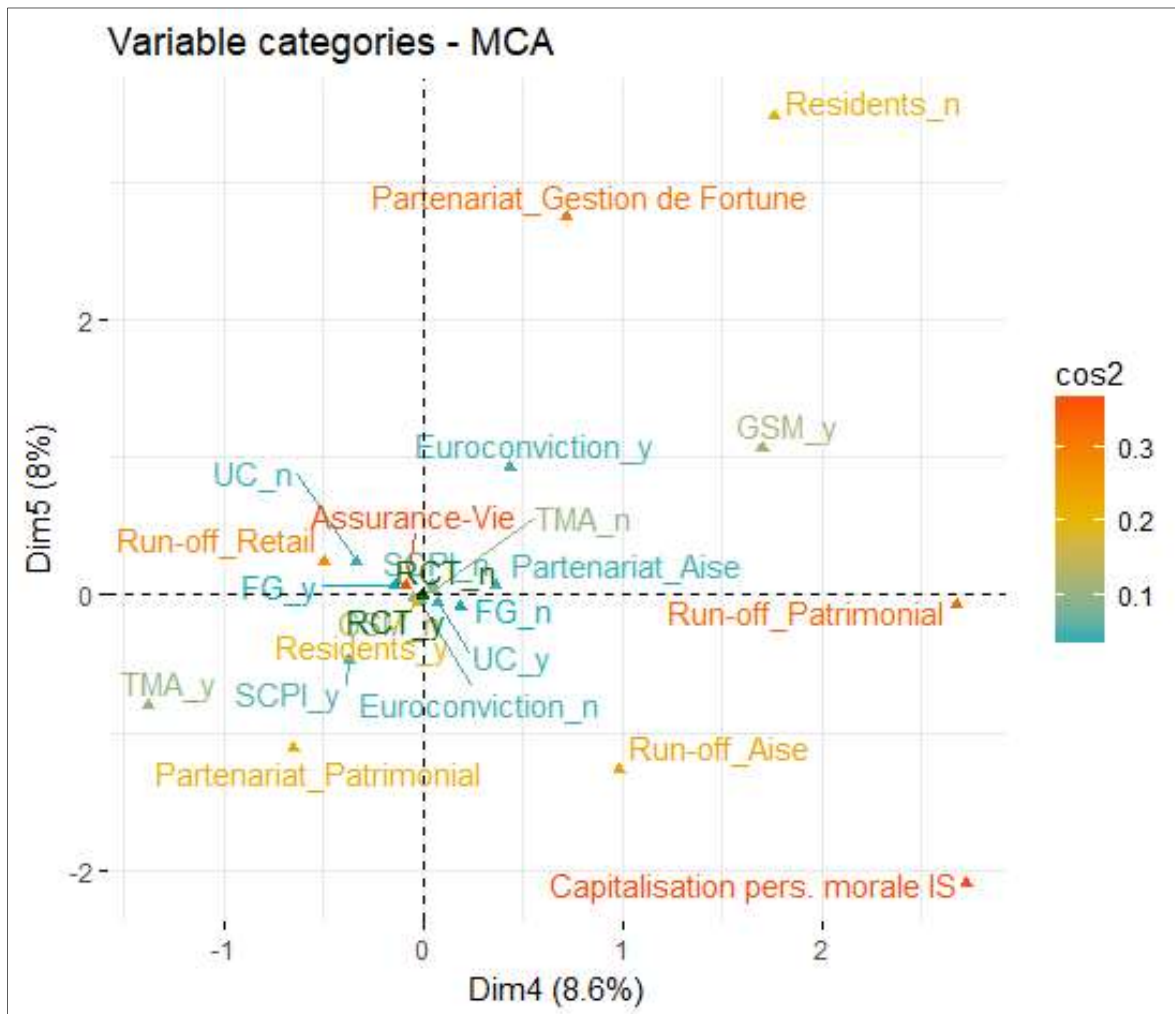


Figure 13 : Représentation des dimensions 4 et 5 de l'ACM sur les données de 2011 à 2015

Lors de la lecture du graphique ci-dessus, il apparaît que la dimension 5 n'apporte pas d'informations complémentaires compréhensibles sur les liaisons entre les modalités.

De ce fait, les dimensions retenues sont les dimensions 1 à 4.

II.3 Résultats du modèle

D'après cette première analyse, nous réalisons un graphique des contributions des variables dans la réalisation des 4 premières dimensions de l'ACM. La ligne en pointillé rouge indique la valeur moyenne attendue sous l'hypothèse nulle. Nous pouvons considérer, de ce fait, que toutes les modalités en dessous n'apportent pas une contribution significative à la classification des individus (la contribution moyenne est l'inverse du nombre de modalités actives, $\frac{1}{K}$) :

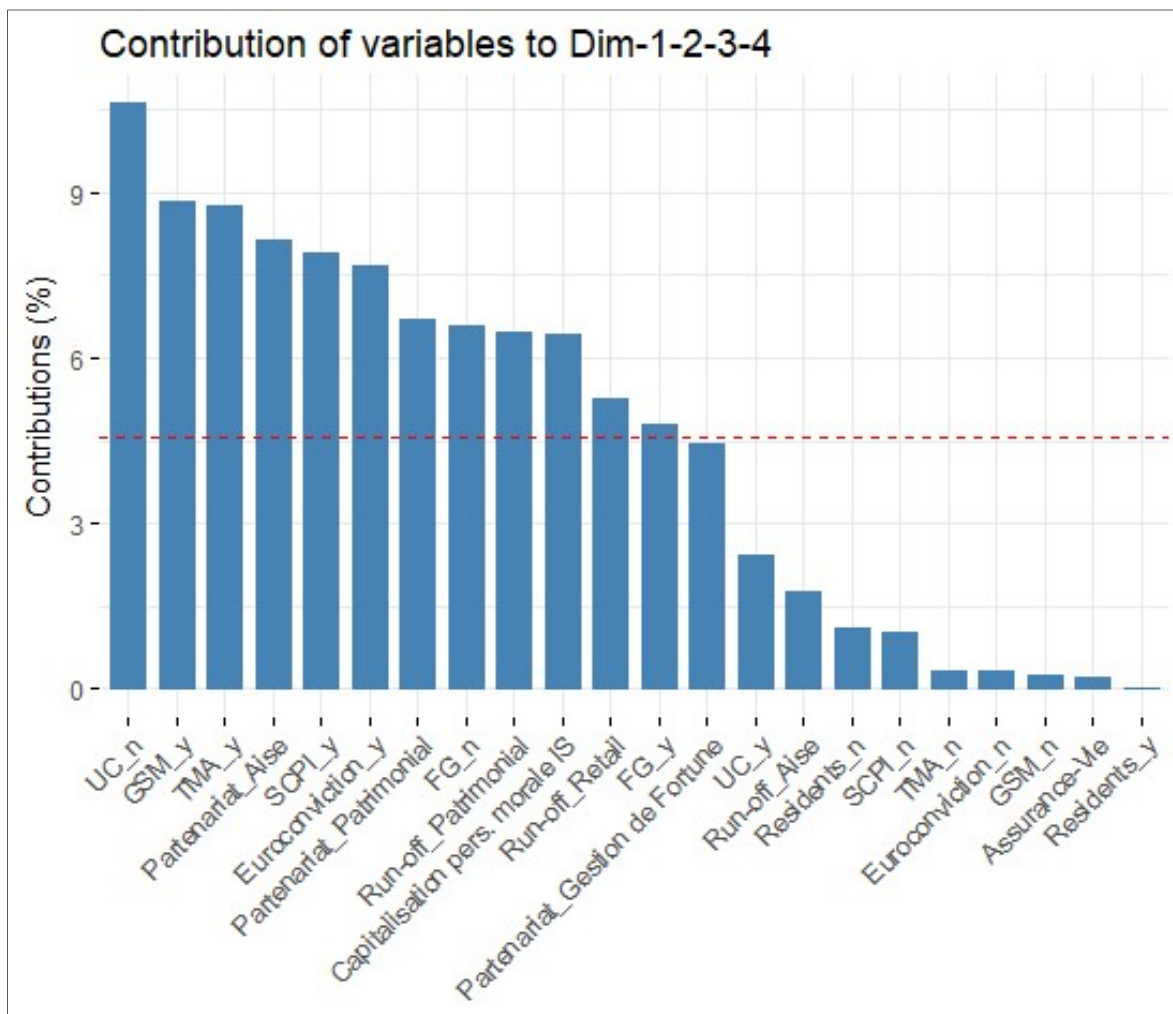


Figure 14 : Contribution des modalités aux dimensions de l'ACM

Ce graphique représente les modalités des variables mais nous ne pouvons pas conserver une seule modalité de la variable si celle-ci est binaire. La variable représentant le statut de résident fiscal (oui/non) a toutes ses modalités en dessous de la limite. Nous pouvons donc considérer cette variable comme non significative. Par ailleurs, il peut être intéressant de regrouper tous les contrats « Run-off », quel que soit le montant de l'encours, au vu de la faible contribution des modalités « Run-off_Aisé » et « Run-off_Retail » pour créer des catégories de comportements des contrats du portefeuille.

Hormis sur l'investissement 100% fonds en euros / 100% UC, cette étude n'a pas apporté beaucoup d'informations sur l'existence de variables discriminantes significatives concernant le rachat total. Pour compléter ces premiers éléments, tentons une approche différente : la méthode des arbres de classification.

III Les arbres de classification

III.1 Présentation du modèle

La création d'un arbre de classification est une méthode à but explicatif qui résout les problèmes de discrimination et de régression en divisant successivement l'échantillon en sous-groupes. Il s'agit de sélectionner parmi les variables explicatives celle qui est la plus liée à la variable à expliquer. Cette variable fournit une première division de l'échantillon en plusieurs sous-ensembles appelés segment. Puis on réitère cette procédure à l'intérieur de chaque segment en recherchant la deuxième meilleure variable, et ainsi de suite... Il s'agit donc d'une classification descendante à but prédictif opérant par sélection de variables : chaque classe doit être la plus homogène possible vis-à-vis de la variable à prédire.

La méthode mise en exergue dans cette étude est l'algorithme CART (Classification and Regression Tree) remis en lumière après les travaux de Breiman & al²⁷.

La méthode CART permet de construire un arbre de décision binaire par divisions successives de l'échantillon en deux sous-ensembles.

Les divisions possibles dépendent de la nature statistique de la variable :

- variable binaire : une division possible,
- variable nominale N (k modalités) : $2^{k-1} - 1$ divisions possibles,
- variable ordinale²⁸ O (k modalités) : $k - 1$ divisions possibles,
- variable quantitative Q (q valeurs distinctes) : $q - 1$ divisions possibles.

Une observation de la table de données appartient à une unique feuille : les sous-ensembles créés sont disjoints.

L'indice utilisé dans la règle de divisions de l'échantillon par défaut sous le logiciel R est l'indice de Gini (qui peut également être appelé mesure de l'impureté du nœud ou indice de diversité).

La mesure de l'impureté associée au segment t est :

$$i(t) = \sum_r^k \sum_s^k P(r/t)P(s/t) \text{ avec } r \neq s$$

Notre étude comporte uniquement deux modalités, nous avons donc : $i(t) = 2 \times p \times (1 - p)$

Plus le mélange des classes dans le segment est important plus $i(t)$ est élevé.

Chaque division d'un nœud entraîne de facto une réduction de l'impureté :

$$\Delta i(s, t) = i(t) - p_g i(t_g) - p_d i(t_d)$$

La fonction $i(t)$ étant concave, l'impureté moyenne ne peut que décroître par division d'un nœud.

La meilleure division d'un nœud est celle qui assure la plus grande réduction de l'indice en passant du nœud à ses segments descendants.

Lors de la création de l'arbre, il n'y a pas de règle d'arrêt dans le processus de division des segments : une procédure d'élagage est réalisée et permet de supprimer les branches les moins informatives après l'obtention de l'arbre complet.

²⁷ Breiman, Friedman, Olshen et Stone [1984], *Classification And Regression Trees*. Wadsworth and Brooks.

²⁸ Pour rappel, la différence entre une variable ordinale et une variable nominale est qu'il existe un ordre entre les différentes modalités de la première variable mentionnée.

III.1.1 Elagage

L'arbre est réalisé grâce à la fonction *rpart* du logiciel *R*. L'arbre est construit à partir des paramètres suivants :

- coefficient de pénalisation²⁹ nul ($cp = 0$) : permet d'avoir l'arbre maximal présentant toutes les feuilles possibles,
- le nombre minimal d'observations dans une feuille est fixé à 1.

Sous *R*, le post élagage doit être réalisé par l'utilisateur du logiciel. La commande *printcp(.)* permet de réaliser cette étape.

```
Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2015, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete          Cadre_fiscal          Classification.Reseau  Classification_Geographique
[5] EuroConviction      FondsG                SCPI                  TMA_Bonus
[9] UC                  UC.Mandat

Root node error: 2797/62733 = 0.044586

n= 62733

      CP nsplit rel error  xerror  xstd
1 3.1284e-04  0  1.00000 1.00000 0.018482
2 2.3835e-04  9  0.99714 0.99964 0.018479
3 1.7876e-04 12  0.99642 0.99893 0.018473
4 1.3407e-04 14  0.99607 1.00000 0.018482
5 1.3001e-04 23  0.99464 1.00107 0.018491
6 1.1918e-04 36  0.99285 1.00107 0.018491
7 8.9381e-05 45  0.99178 1.00215 0.018501
8 7.1505e-05 57  0.99070 1.00536 0.018529
9 6.5005e-05 62  0.99035 1.00679 0.018542
10 5.9588e-05 80  0.98892 1.00858 0.018557
11 4.4691e-05 104 0.98749 1.00930 0.018564
12 2.5538e-05 117 0.98677 1.01001 0.018570
13 0.0000e+00 131 0.98641 1.01001 0.018570
```

Figure 15 : Affichage de la commande *Rpart(.)* réalisée sur les données de l'année 2015

Cette commande décrit des séquences d'arbres en mettant en relation le coefficient de pénalisation avec :

- *nsplit* : le nombre de segmentation (nombre de feuilles-1),
- *rel error* : l'erreur calculée sur l'échantillon d'apprentissage (normalisée de manière à ce que l'erreur sur la racine soit égale à 1),
- *xerror* : l'erreur calculée en validation croisée (normalisée également / $K = 10$ par défaut sous *R*). Celui-ci est considéré comme un meilleur estimateur de l'erreur réelle,
- *xstd* : l'écart-type de l'erreur calculée en validation croisée

Zoom sur le calcul de *xerror* : *Rpart(.)* construit l'arbre maximal avec la totalité de l'échantillon. Ensuite, la commande calcule les séquences d'arbres emboîtés en modulant le paramètre *CP*. Pour chaque valeur de *CP* sont calculées les données décrites ci-dessus. *Rpart(.)* partitionne les données en $K = 10$ portions (paramétrable). Il calcule un arbre sur les 9/10èmes, et utilise la fraction restante 1/10 pour estimer l'erreur. En faisant tourner les sous échantillons, il dispose donc de 10 arbres. Il utilise alors les valeurs de *CP* de la table précédente pour réduire chaque arbre et calculer l'erreur sur la fraction test qui leur est associée. Ainsi, pour une valeur spécifiée de *CP*, il dispose

²⁹ Toute division de l'arbre qui ne diminue pas ce coefficient n'est pas envisagée par la machine. Ce paramètre a pour rôle principal d'économiser du temps de calcul en supprimant les divisions qui ne sont pas valables par rapport au critère.

en interne de 10 arbres avec leur taux d'erreur respectif. La moyenne de ces valeurs constitue le taux d'erreur en validation croisée : *xerror*.

L'arbre élagué est choisi en prenant celui dont le coefficient de pénalisation minimise la variable *xerror*.

Le plus petit arbre correspondant à ce critère dans notre exemple est le n°3 avec 12 segmentations. Pour obtenir cet arbre, nous devons utiliser la commande *prune(.)* avec :

$$1,7876 \times 10^{-4} < cp \leq 2,3835 \times 10^{-4}$$

On prend pour exemple : $cp = 2 \times 10^{-4}$

```
Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2015, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete          Cadre_fiscal          Classification.Reseau SCPI

Root node error: 2797/62733 = 0.044586

n= 62733

      CP nsplit rel error  xerror   xstd
1 0.00031284     0  1.00000  1.00000  0.018482
2 0.00023835     9  0.99714  0.99964  0.018479
3 0.00020000    12  0.99642  0.99893  0.018473
```

Figure 16 : Résultat suite à l'élagage de l'arbre de décision

Les variables utilisées pour réaliser l'arbre de décision sont alors l'ancienneté, le cadre fiscal, le réseau de distribution et la présence de supports immobiliers.

L'arbre de décision obtenu est le suivant :

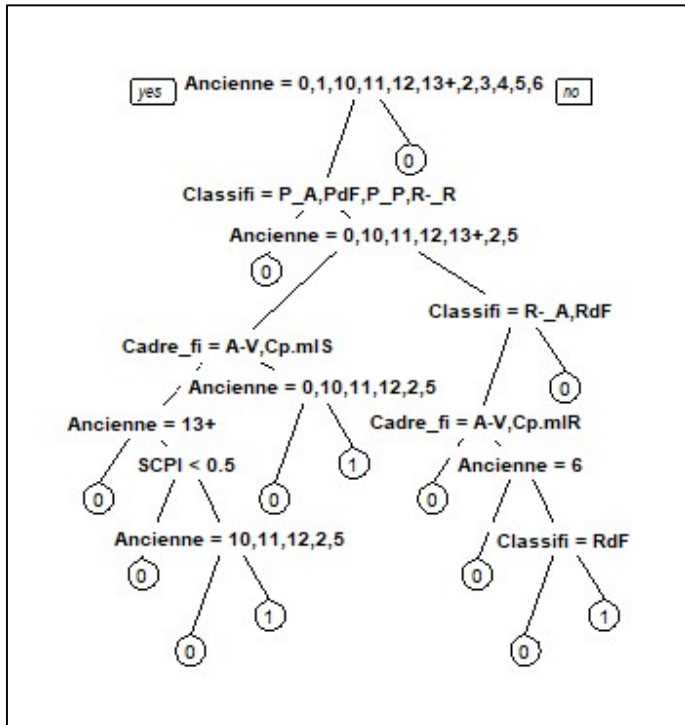


Figure 17 : Arbre de décision obtenu

On applique l'arbre sur les données observées en 2016 (via la commande *predict(.)*) et on réalise la matrice de confusion suivante :

Observations \ Prédications	Prédications	
	0	1
0	59139	15
1	2882	0

Le taux d'erreur en test est : $\frac{2882+15}{59139+2882+1} = 4,67\%$

III.2 Principaux résultats

Prédictions \ Observations	2011		2012		2013		2014		2015		Cumul	
	0	1	0	1	0	1	0	1	0	1	0	1
0	56 029	3 125	58 855	229	58 869	285	58 871	283	59 139	15	58 860	294
1	2 593	289	2 638	244	2 644	238	2 646	236	2 882	0	2 646	236

Variables	Année étudiée					Données Cumulées
	2011	2012	2013	2014	2015	
Cadre fiscal	Non	Oui	Non	Non	Oui	Oui
Réseaux de distribution	Oui	Oui	Oui	Oui	Oui	Oui
Présence de fonds en euros	Oui	Oui	Oui	Non	Non	Oui
Présence d'unités de compte	Non	Oui	Oui	Oui	Non	Oui
Offre commerciale TMA et/ou Bonus fonds Euroconviction	Non	Oui	Non	Non	Non	Non
Supports immobiliers	Oui	Oui	Oui	Non	Oui	Oui
UC Mandat	Non	Non	Non	Non	Non	Non
Résidence	Non	Oui	Oui	Oui	Non	Oui
Ancienneté du contrat	Oui	Oui	Oui	Oui	Oui	Oui
Taux de mauvaise classification de l'année 2016	9,22%	4,73%	4,72%	4,72%	4,67%	4,74%
Taux de Rachat Total calculé	5,50%	0,88%	0,84%	0,84%	0,027%	0,85%

Tableau 27 : Significativité des variables dans le modèle

Concernant les variables discriminantes :

- Le critère du réseau de distribution et celui de l'ancienneté sont utilisés pour tous les arbres réalisés.
- Hormis pour l'année 2014, le critère d'investissement sur un support immobilier est pris en compte pour la création des arbres de décision,
- Hormis pour l'année 2014, le critère d'investissement sur le fonds Euroconviction n'est pas pris en compte pour la création des arbres de décision,
- Hormis pour l'année 2012, la présence d'une offre commerciale temporaire sur le fonds en euros n'a pas été sélectionné comme critère de décision pour les autres arbres.
- Avoir un mode de gestion sous mandat n'est sélectionné par aucun des arbres de décision.
- Pour les autres critères : cadre fiscal, présence de fonds en euros, présence d'unités de compte et lieu de résidence, l'utilisation est plus nuancée.

Remarque : Si nous n'appliquions aucune segmentation, nous aurions, sur les données de 2016, la matrice de confusion suivante :

Observations \ Prédictions	0	1
	0	59154
1	2882	0

Le taux de mauvaise classification (similaire au taux de rachat total constaté) serait : $\frac{2882}{59154+2882} = 4,65\%$

Ce taux est inférieur à tous ceux calculés par les arbres de décision sur toutes les années observées. En revanche, le taux de rachat total calculé via les arbres de décision est largement inférieur à ce qui peut être constaté sur l'année 2016.

Ici, l'arbre de décision n'est pas utilisé seulement dans un but prédictif mais également pour déterminer des tendances de variables discriminantes significatives.

Les avantages de la méthode CART sont nombreux :

- L'arbre de décision binaire est lisible par tout utilisateur et constitue un moyen de communication très efficace.
- La validation par une méthode de ré-échantillonnage est une technique très compréhensible pour l'utilisateur.
- La mixité des variables qu'accepte la procédure offre une latitude non négligeable lors de la réalisation d'études de données.

Cet outil présente néanmoins certains points faibles :

- Un léger biais lors de la sélection des variables avec une sélection plus fréquente des variables explicatives offrant le plus de divisions possibles (avec plus de modalités),
- La réalisation en séquentiel de l'arbre peut avoir pour conséquence de cacher l'effet d'une variable explicative par une autre lors du choix de règle de division. Cela peut se produire par exemple lorsque deux variables explicatives sont très corrélées entre elles.
- Un manque de robustesse dans les résultats : une légère modification des données peut engendrer différents arbres de classification.

Face à ces inconvénients de la méthode, et en partant du principe de mutualisation, des algorithmes d'agrégations de modèles apparaissent en littérature afin de pallier cette problématique. Cela a amené à la création des forêts aléatoires par Breiman³⁰ que nous présentons par la suite.

IV L'algorithme Forêts aléatoires (*Random Forest*)

IV.1 Présentation du modèle

L'algorithme *Random Forest* est un cas particulier de *Bagging*³¹ pour *bootstrap aggregating* spécifique aux modèles définis par des arbres binaires tel que CART.

Les principes du *Bagging* s'appliquent à toute méthode de modélisation mais n'ont d'intérêt que dans le cas de modèles « instables » donc non linéaires. L'implémentation de ces algorithmes manque de sens avec la régression ou l'analyse discriminante.

Soit la variable aléatoire Y à expliquer. Dans notre étude, celle-ci possède deux modalités $\{1,0\}$ où la modalité $\{1\}$ correspond à la survenance du rachat total dans l'année d'observation.

Le fichier comporte J descripteurs $\{X_1, X_2, \dots, X_j\}$.

Le vecteur de valeurs pour un individu ω s'écrit $(X_1(\omega), X_2(\omega), \dots, X_j(\omega))$

Soit $f(x)$ un modèle de fonction de x : $f(x) = \{x_1, x_2, \dots, x_j\}$.

L'échantillon d'observation Ω est de taille n et on note z un échantillon de loi F :

$z = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

³⁰ L.Breiman, *Random forests*, Machine Learning 45 [2001],5-32

³¹ L.Breiman, *Bagging predictors*, Machine Learning 26 [1996], n°2, 123-140.

Considérant B échantillons indépendants notés $\{z_b\}_{b=1,\dots,B}$, la prévision par agrégation de modèles est alors dans notre étude :

$$\hat{f}_B(\cdot) = \operatorname{argmax}_j \operatorname{card}\{b \mid \hat{f}_{z_b}(\cdot) = j\}^{32}$$

ce qui correspond à un vote à la majorité simple des modèles.

Le *Bagging* permet finalement de trouver le modèle « moyen » approchant le mieux les données d'intérêt. Moyenner les prévisions de plusieurs modèles indépendants permet de réduire la variance et donc de réduire l'erreur de prévision.

Dans les faits, les B échantillons ne sont pas indépendants : ce sont des échantillons *bootstraps* obtenus par des tirages aléatoires avec remise de n individus dans l'échantillon d'observation Ω .

Afin de contrôler, voire d'optimiser le nombre d'échantillons nécessaires, nous pouvons utiliser l'erreur *out-of-bag*³³. Pour chaque observation (x_i, y_i) , on considère l'ensemble des échantillons *bootstraps* ne contenant pas (x_i, y_i) ainsi que tous les arbres associés à ces échantillons. On prévoit la valeur \hat{y}_i en utilisant uniquement ces arbres. L'erreur moyenne de prévision associée est alors calculée. L'erreur *out-of-bag* varie dès lors avec le nombre d'échantillons *bootstraps* existants, puis se stabilise dès qu'elle atteint le nombre nécessaire d'échantillons à l'agrégation.

La particularité de l'algorithme *Random Forest* comparé à un *Bagging* classique est l'ajout d'un tirage aléatoire sur l'ensemble des variables explicatives.

Le principe de l'algorithme est le suivant :

- Chaque arbre est appris sur un sous-ensemble de la population tiré aléatoirement avec remise,
- Chaque construction d'un nœud de l'arbre est réalisée sur un sous-ensemble d'un nombre m de variables explicatives tirées aléatoirement.
- Pour l'estimation finale, on choisit la catégorie la plus fréquemment utilisée.

La sélection aléatoire d'un nombre réduit de variables explicatives à chaque étape de la construction d'un arbre accroît la variabilité en mettant en avant nécessairement d'autres variables. Chaque modèle de base est donc moins optimal, mais l'agrégation conduit finalement à de bons résultats. Pour optimiser m , une procédure de validation croisée est opérée.

IV.2 Sélection des paramètres de la forêt aléatoire

Parmi les principaux paramètres des forêts aléatoires, nous pouvons noter :

- la taille des arbres de décision (*nodesize* = nombre minimal d'observation dans chaque feuille),
- le nombre d'arbres agrégés,
- le nombre de variables explicatives à chaque scission de nœud.

Concernant la taille des arbres de décision, nous avons laissé la valeur par défaut (*nodesize* = 1). En effet, une valeur plus grande permettrait une plus grande robustesse de chaque arbre mais risquerait d'augmenter la corrélation entre les arbres et de diminuer le gain de l'agrégation.

³² Ensemble des points en lesquels une expression atteint sa valeur maximale

³³ Signifie finalement « en dehors du bootstrap »

IV.2.1 Le nombre d'arbres agrégés

Le nombre d'arbres générés par la fonction *Randomforest* sous *R* est de 500 par défaut. Les forêts n'étant pas assujetties au problème de sur-apprentissage, nous essayons de déterminer si ce nombre est suffisant ou si son augmentation est bénéfique pour le modèle. En effet, si l'augmentation du nombre de générations d'arbres n'est pas utile, il n'est pas nécessaire de perdre du temps machine pour un calcul qui n'augmente pas le pouvoir de prédiction par rapport à la valeur par défaut.

```
Call:
  randomForest(formula = Rachat_total ~ Classification_Geographique + UC + EuroConviction + Cadre_fiscal + SCPI + TM
A_Bonus + Classification.Reseau + FondsG + UC.Mandat + Anciennete, data = donnees_2015, importance = TRUE, keep
p.forest = T)
  Type of random forest: classification
  Number of trees: 500
  No. of variables tried at each split: 3

  OOB estimate of error rate: 4.46%
Confusion matrix:
  0 1 class.error
0 59936 0 0
1 2797 0 1
```

Figure 18 : Résultat de la fonction *RandomForest* avec les paramètres par défaut sur l'année 2015

Nombres d'arbres générés	Taux de mauvais classement des Rachats Totaux					
	2011	2012	2013	2014	2015	Données cumulées
100	99,4%	100%	98,9%	92,5%	100%	97,7%
300	99,4%	100%	100%	97,7%	100%	99%
500	99,5%	100%	100%	96,6%	100%	94,7%
1000	99,5%	100%	100%	96,4%	100%	94,7%
1200	99,4%	100%	100%	93,6%	100%	94,8%

Tableau 28 : Taux de mauvais classement selon le nombre d'arbres de la forêt

Nombres d'arbres générés	Erreur <i>out-of-bag</i>					
	2011	2012	2013	2014	2015	Données cumulées
100	7,05%	6,69%	5,45%	4,97%	4,46%	5,69%
300	7,05%	6,69%	5,5%	5,08%	4,46%	5,7%
500	7,06%	6,69%	5,49%	5,07%	4,46%	5,66%
1000	7,06%	6,69%	5,49%	5,1%	4,46%	5,66%
1200	7,05%	6,69%	5,49%	5,02%	4,46%	5,65%

Tableau 29 : Taux *out-of-bag* selon le nombre d'arbres de la forêt

L'augmentation du nombre d'itérations n'a pas d'incidence significative sur le pouvoir prédictif des modèles au vu des faibles variations de l'erreur *out-of-bag* et du taux de mauvais classement des contrats rachetés. La représentation graphique de l'évolution du taux d'erreur *out-of-bag* en fonction du nombre d'arbres agrégés via les données de l'année 2015 met en exergue une stabilité très rapide de ce taux.

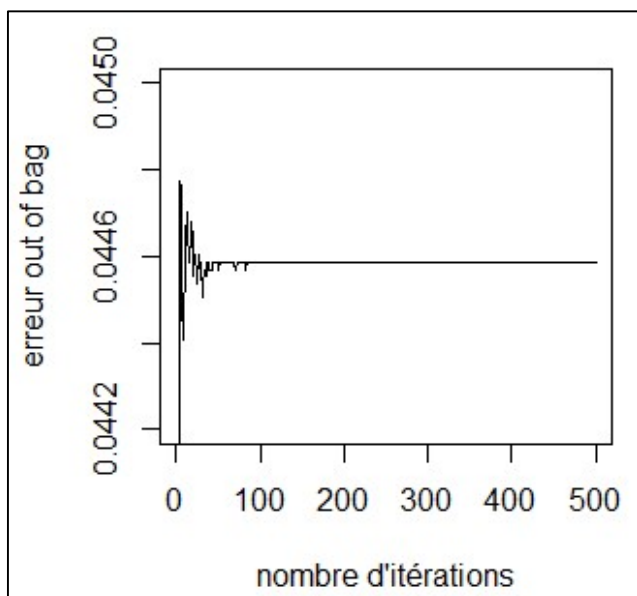


Figure 19 : Evolution du taux d'erreur *out-of-bag* du modèle réalisé via les données de l'année 2015.

La valeur par défaut du nombre d'arbres agrégés est donc suffisante pour avoir des résultats significatifs (i.e. 500 arbres).

IV.2.2 Le choix du nombre de variables

Le nombre de prédicteurs sélectionnés pour la détermination de chaque nœud ne doit pas être trop grand : une valeur trop grande augmente la corrélation entre les modèles et diminue le gain de l'agrégation de ces derniers. A contrario, une valeur trop petite conduit à des modèles unitaires trop biaisés, insuffisamment prédictifs, surtout lorsque de nombreux prédicteurs sont faiblement discriminants.

Notre base de données présente 10 variables explicatives pour le modèle de forêts aléatoires. En effet, l'algorithme des forêts aléatoires ne réalise pas de variables disjonctives pour les prédicteurs présentant plus de 2 modalités.

Le nombre de variables sélectionnées aléatoirement pour chaque réalisation de nœud (*mtry* sous *R*) vaut par défaut *entier*($\sqrt{10}$) = 3.

Nous testons plusieurs fois l'algorithme en modifiant ce paramètre entre 1, 3, 5 et 7, et nous regardons l'erreur *out-of-bag* calculée pour chaque modèle.

Nombres de variables	Taux de mauvais classement des Rachats Totaux					
	2011	2012	2013	2014	2015	Données cumulées
1	100%	100%	100%	100%	100%	100%
3	99,5%	100%	100%	96,6%	100%	94,7%
5	98,7%	99,6%	92,2%	89,6%	99,7%	94,6%
7	98,4%	97,2%	92%	89,5%	99,5%	94,6%

Tableau 30 : Taux de mauvais classement selon le nombre de variables aléatoires à chaque nœud

Nombres de variables	Erreur <i>out-of-bag</i>					
	2011	2012	2013	2014	2015	Données cumulées
1	7,09%	6,69%	5,49%	5,14%	4,46%	5,71%
3	7,06%	6,69%	5,49%	5,07%	4,46%	5,66%
5	7,05%	6,74%	5,37%	4,92%	4,45%	5,7%
7	7,08%	6,72%	5,39%	4,95%	4,48%	5,72%

Tableau 31 : Taux d'erreur *out-of-bag* selon le nombre de variables aléatoires à chaque nœud

En augmentant le nombre de variables randomisées à chaque nœud, le taux de mauvais classement des contrats rachetés s'améliore (qui est à 100% avec la valeur de *mtry* par défaut pour la moitié de nos jeux de données) tout en ayant un taux d'erreur *out-of-bag* très proche voire meilleur.

Le passage de la variable *mtry* = 5 à 7 n'améliore pas significativement nos taux d'erreurs. Le nombre de variables randomisées pris en compte dans la suite de l'étude sera de 5.

Remarque : Pour augmenter le nombre de contrats prédits « rachetés », on peut utiliser un vecteur *cutoff* de longueur égale au nombre de modalités à prédire (ici 2 pour notre étude), de sorte que la classe prédite ne soit pas nécessairement celle qui remporte la plus grande proportion des votes mais celle dont le ratio « proportion des votes/*cutoff* » est maximal.

Voici ci-dessous une synthèse de différentes valeurs de ce paramètre :

Cut off	Erreur <i>out-of-bag</i>					
	2011	2012	2013	2014	2015	Données cumulées
0.2	7,12%	6,75%	5,39%	4,97%	4,5%	5,77%
0.25	7,09%	6,73%	5,38%	4,95%	4,49%	5,75%
0.3	7,07%	6,74%	5,37%	4,94%	4,47%	5,74%
0.4	7,06%	6,77%	5,37%	4,93%	4,46%	5,72%
0.5	7,05%	6,74%	5,37%	4,92%	4,45%	5,7%

Tableau 32 : Taux d'erreur *out-of-bag* selon la valeur Cut off

Cut off	Taux de mauvais classement des Rachats Totaux					
	2011	2012	2013	2014	2015	Données cumulées
0.2	98,2%	96,8%	91,88%	89,3%	99,4%	94,4%
0.25	98,2%	97%	91,98%	89,4%	99,5%	94,4%
0.3	98,4%	97,2%	92,02%	89,5%	99,6%	94,5%
0.4	98,5%	99%	92,15%	89,5%	99,6%	95,5%
0.5	98,7%	99,6%	92,2%	89,6%	99,7%	94,6%

Tableau 33 : Taux de mauvais classement selon la valeur Cut off

La très légère amélioration du taux de mauvais classement des Rachats Totaux se fait au détriment du taux d'erreur *out-of-bag* (*OOB*) constaté. Ces variations ne sont pas assez significatives pour modifier les paramètres par défaut. **Nous resterons donc avec la valeur par défaut (i.e. 0,5).**

IV.3 Mesure de l'importance des variables

Pour déterminer l'importance d'une variable dans une forêt aléatoire, Breiman propose deux critères.

Pour chaque arbre k , on mesure le taux d'erreur OOB_k de cet arbre noté E_k . Parmi les individus *out-of-bag*, on permute aléatoirement les valeurs de la variable considérée (on échange la valeur de cette variable pour chaque individu par la valeur de cette variable d'un autre individu) et on mesure le taux d'erreur E'_k de l'arbre sur ces individus ainsi perturbés. Plus la différence entre E'_k et E_k est grande, plus la variable « mélangée » est déterminante dans la classification des individus. En effet, si la différence est faible, cela signifie que le classement des individus est très proche entre E et E' et donc que la permutation n'a pas perturbé la classification. Le critère d'importance qui en découle « *Mean Decrease Accuracy* » est alors la moyenne de $E'_k - E_k$ divisée par son écart type.

Le deuxième critère « *Mean Decrease Gini* », spécifique aux arbres CART, est la baisse d'impureté des nœuds résultant de la scission de la variable, baisse moyennée sur l'ensemble des arbres. Ces deux critères peuvent être représentés de la manière graphique suivante :

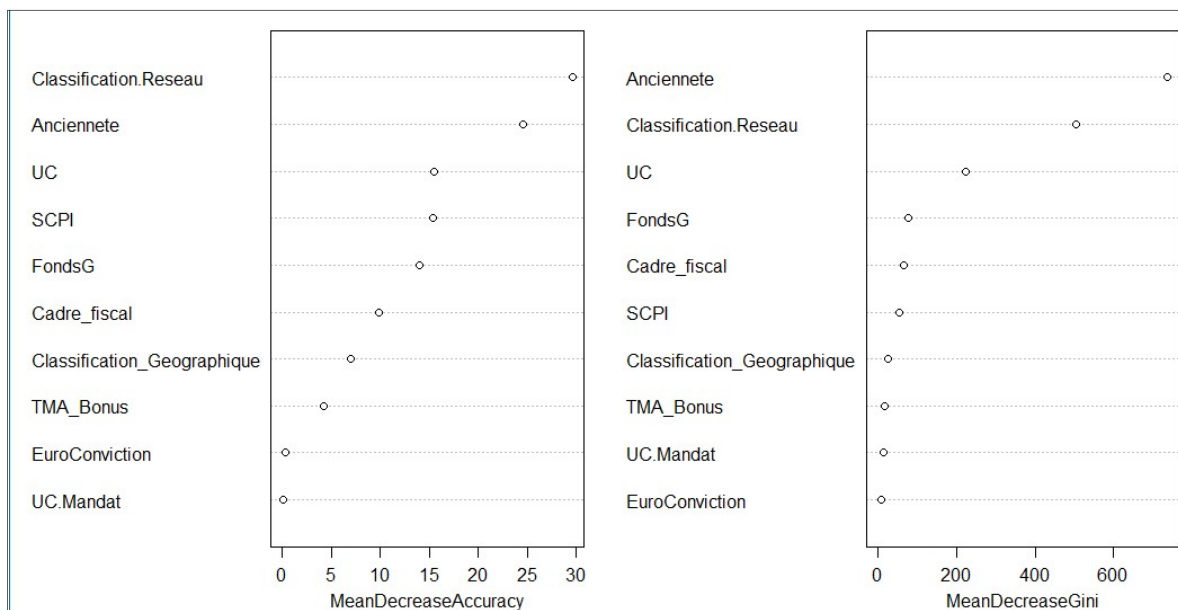


Figure 20 : Importance des variables déterminée sur la table des données cumulées

Pour les variables très discriminantes et celles qui le sont très peu, ces deux critères fournissent des résultats assez équivalents.

De la même manière, du passage d'un arbre de classification à l'implémentation d'une forêt aléatoire, pour fiabiliser les résultats, plusieurs forêts aléatoires (ici 30) sont réalisées en déterminant l'importance de chaque variable dans chacune des forêts, puis son importance moyenne dans l'ensemble des forêts aléatoires.

Voici ci-dessous la représentation graphique des résultats calculés sur les données de l'année 2015 (*Mean Decrease Accuracy* à gauche et *Mean Decrease Gini* à droite) :

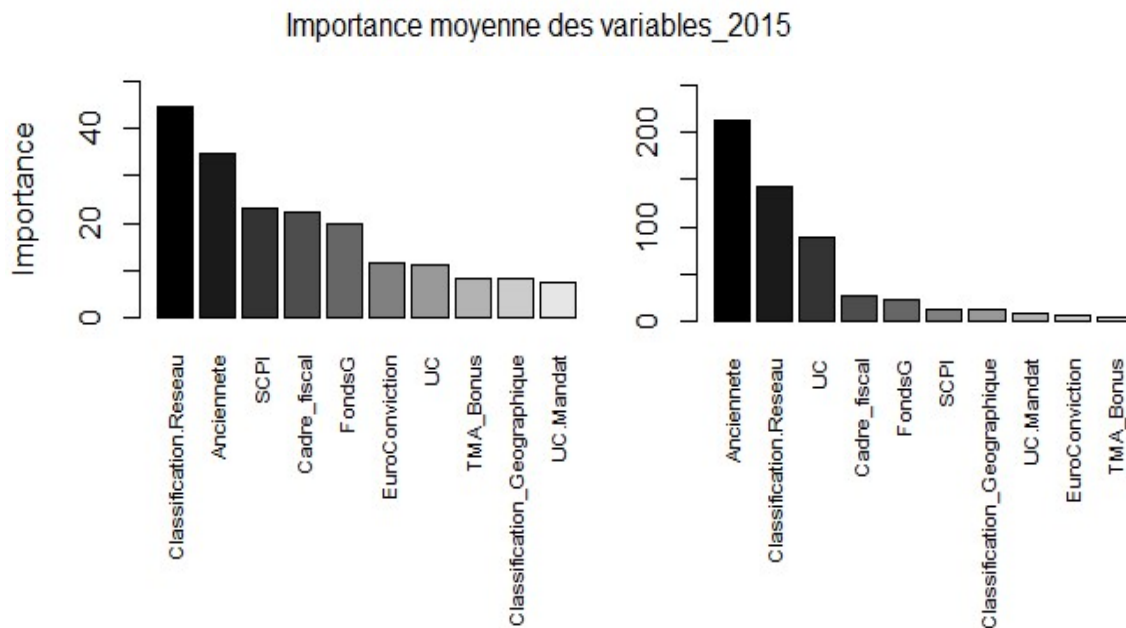


Figure 21 : Importance moyenne des variables via les données de 2015

Le tableau ci-dessous synthétise les résultats obtenus pour tous les jeux de données étudiés (les résultats graphiques sont fournis en annexe) :

Variables \ Année étudiée	2011	2012	2013	2014	2015	Données Cumulées
Cadre fiscal	Oui	Oui	Ne se prononce pas	Ne se prononce pas	Oui	Oui
Réseaux de distribution	Oui	Oui	Oui	Oui	Oui	Oui
Présence de fonds en euros	Oui	Oui	Oui	Oui	Oui	Oui
Présence d'unités de compte	Oui	Oui	Oui	Oui	Ne se prononce pas	Oui
Offre commerciale TMA et/ou Bonus	Non	Ne se prononce pas	Non	Non	Non	Non
fonds Euroconviction	Non	Non	Non	Non	Ne se prononce pas	Non
Supports immobiliers	Oui	Oui	Ne se prononce pas	Ne se prononce pas	Ne se prononce pas	Oui
UC Mandat	Non	Non	Non	Non	Non	Non
Résidence	Non	Non	Non	Non	Non	Non
Ancienneté du contrat	Oui	Oui	Oui	Oui	Oui	Oui
Taux de mauvaise classification de l'année 2016	8,54%	4,65%	4,71%	4,71%	4,65%	4,98%
Taux de Rachat Total calculé	4,79%	0,01%	0,83%	0,83%	0%	1,01%

Tableau 34 : Significativité des variables dans le modèle

Cette agrégation d'arbres confirme les résultats que nous avons obtenus via l'arbre de classification sur le rôle discriminant de certaines variables, notamment le réseau de distribution et l'ancienneté du contrat.

La présence d'une gestion sous mandat, du fonds Euroconviction ou d'une offre TMA/Bonus est également considérée comme non significative pour la détermination du comportement de rachats totaux.

Par ailleurs, hormis sur les données cumulées, l'agrégation d'arbres permet d'améliorer légèrement le taux de mauvaise classification des données 2016. En revanche, ces taux restent toujours supérieurs au taux constaté de 2016 (4,65%) et les taux de rachat sont toujours nettement inférieurs au taux constaté sur les données de 2016.

Complétons alors l'approche de l'estimation des taux de rachats totaux et des variables discriminantes par une observation des données via la régression logistique.

V La sélection de variables via la régression logistique

V.1 Présentation du modèle

La régression logistique binaire est utilisée lorsque la valeur de la variable aléatoire Y possède deux modalités $\{1,0\}$. Dans notre étude, la modalité $\{1\}$ correspond à la survenance du rachat total dans l'année d'observation.

L'échantillon d'observation Ω est de taille n . La valeur prise par Y pour un individu ω est notée $Y(\omega)$.

Le fichier comporte J descripteurs $\{X_1, X_2, \dots, X_j\}$. Le vecteur de valeurs pour un individu ω s'écrit $(X_1(\omega), X_2(\omega), \dots, X_j(\omega))$. La probabilité, à priori, pour un individu d'effectuer le rachat total de son contrat dans l'année d'observation est : $P[Y(\omega) = 1] = p(\omega)$.

La probabilité, à posteriori, d'un individu ω d'effectuer le rachat total de son contrat dans l'année d'observation, i.e en sachant les valeurs prises par les descripteurs est notée $P[Y(\omega) = 1 / X(\omega)] = \pi(\omega)$. C'est cette probabilité que l'on cherche à modéliser en apprentissage supervisé.

Pour simplifier les formules, lorsqu'il n'existe pas d'ambiguïté, on pourra poser $\pi(\omega) = \pi$.

La fonction de répartition utilisée dans ce modèle est la fonction logistique définie par : $F(x) = \frac{e^x}{1+e^x}$

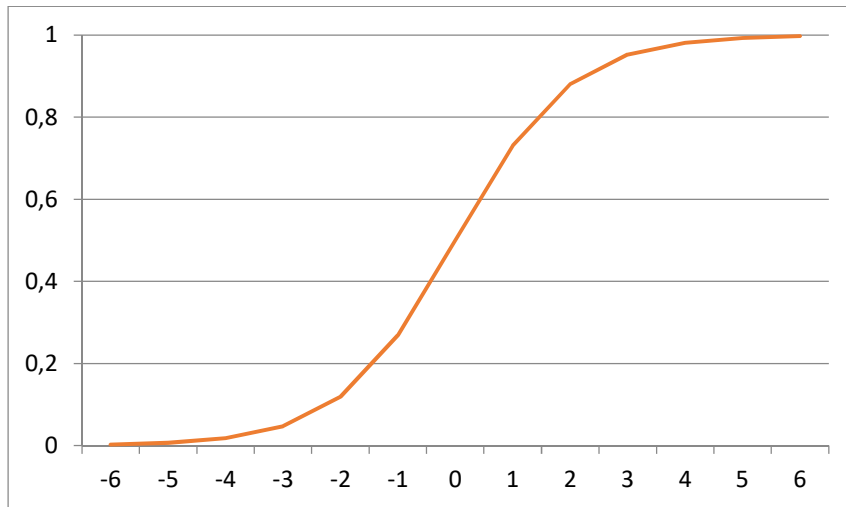


Figure 22 : Fonction logistique

Cette fonction est bien adaptée à la modélisation de probabilités, car elle prend ses valeurs entre 0 et 1 selon une courbe en S.

On peut écrire :

$$\pi(\omega) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j}}$$

$$\text{où } \text{Log} \left(\frac{\pi(\omega)}{1 - \pi(\omega)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j$$

$\beta_0, \beta_1, \dots, \beta_j$ sont les paramètres que l'on souhaite estimer à partir des données observées.

V.1.1 Estimation des paramètres par la maximisation de la vraisemblance

On modélise la probabilité en utilisant comme hypothèse de distribution la loi binomiale $\mathcal{B}(1, \pi)$.

Ce qui donne : $P[Y(\omega) / X(\omega)] = \pi(\omega)^{Y(\omega)} \times (1 - \pi(\omega))^{(1 - Y(\omega))}$.

La vraisemblance d'un échantillon Ω correspond à la probabilité d'obtenir l'échantillon Ω à partir d'un tirage dans la population. Elle varie entre 0 et 1 et s'écrit :

$$L = \prod_{\omega} \pi(\omega)^{Y(\omega)} \times (1 - \pi(\omega))^{(1 - Y(\omega))}$$

La méthode du maximum de vraisemblance consiste à produire les paramètres de la régression logistique qui maximisent la probabilité d'observer cet échantillon.

Afin de faciliter le traitement, il est souvent plus simple de travailler avec la log-vraisemblance :

$$LL = \sum_{\omega} Y(\omega) \times \ln \pi + (1 - Y(\omega)) \times \ln(1 - \pi)$$

Le logarithme est une fonction monotone. Les paramètres qui maximisent la vraisemblance sont les mêmes que ceux qui maximisent la log-vraisemblance. La log-vraisemblance est une fonction

convexe, ce qui implique une unique solution aux estimateurs de β_j . Comme il n'existe pas d'expression analytique de ces estimateurs, nous sommes contraints de passer par des procédures numériques pour les déterminer.

La régression logistique est réalisée via la fonction *GLM* du logiciel *R*. N'étant pas le sujet même de notre étude, l'algorithme de détermination des paramètres utilisés est décrit de manière synthétique.

V.1.2 Algorithme de la fonction GLM sous R

L'algorithme utilisé est celui de Newton-Raphson. Le système à résoudre est :

$$f_i(x) = f_i(x_1, x_2, \dots, x_n) = 0 \text{ avec } i = 1, 2, \dots, n$$

On part du développement de Taylor suivant :

$$f_i(x + dx) = f_i(x) + \sum_{j=1}^n \frac{\delta f_i}{\delta x_j} dx_j + O(dx^2) = 0 \text{ avec } i = 1, 2, \dots, n$$

Avec $J_{ij} = \frac{\delta f_i}{\delta x_j}$, on obtient :

$$f_i(x + dx) = f_i(x) + J dx + O(dx^2) = 0 \text{ avec } i = 1, 2, \dots, n$$

L'algorithme s'obtient ensuite en annulant $f_i(x + dx)$ et en négligeant les termes d'ordre supérieur :

$$x^{k+1} = x^k + dx \text{ avec } J dx = -f(x^k)$$

Cet algorithme nécessite la résolution d'un système linéaire à chaque étape.

V.1.3 Conditions de validité de la régression logistique

Avant d'utiliser la régression logistique pour déterminer un modèle, assurons-nous que les conditions de validité sont réunies. On considère que les résultats du modèle sont valides s'il y a au moins 5 à 10 événements constatés par variable explicative : si la variable à expliquer présente un nombre d'évènements trop faible par rapport au nombre de variables explicatives du modèle, les résultats risquent d'être mal évalués.

Année d'observation	2011	2012	2013	2014	2015
Nombre de Rachats Totaux constatés	3651	3548	3031	3014	2797
Nombre de variables explicatives*10	28*10=280				

La condition de validité est vérifiée. Le modèle peut par conséquent être appliqué sur toutes les années d'observation et les résultats peuvent être considérés comme fiables.

V.2 Via la significativité des coefficients (test de Wald)

V.2.1 Variables catégorielles

Pour le modèle de régression logistique, les variables qualitatives à plusieurs classes sont recodées en plusieurs variables binaires : k modalités sont recodées en $k - 1$ variables binaires. Les coefficients calculés s'appliquent en comparant les $k - 1$ modalités avec la modalité de référence. Dans ce cas, il est important que la modalité de référence soit une de celle des plus représentées. Dans le cas des variables catégorielles, il est nécessaire de tester la significativité statistique de chacune des $k - 1$ variables binaires mais également la significativité globale de la variable catégorielle. Celle-ci est testée grâce à la fonction *drop* de *R*.

La modalité de référence utilisée est la modalité la plus représentée dans les données observées. L'effet global de la variable catégorielle par la fonction *drop1* permet de constater si, au global, cette variable est statistiquement significative sur les rachats totaux.

V.2.2 Principaux résultats de la régression logistique

On considère une variable statistiquement significative lorsque *p-value* est inférieure à 5%.

Variables \ Année étudiée	2011	2012	2013	2014	2015	Données cumulées
Cadre Fiscal	Oui	Oui*	Oui*	Oui*	Oui*	Oui
Réseaux de distribution	Oui	Oui	Oui**	Oui**	Oui**	Oui
Présence de fonds en euros	Oui	Oui	Non	Oui	Non	Oui
Présence d'unités de compte	Non	Oui	Oui	Oui	Oui	Oui
Offre commerciale TMA et/ou Bonus	Non	Oui	Oui	Oui	Oui	Oui
fonds Euroconviction	Non	Non	Non	Oui	Oui	Oui
Supports immobiliers	Oui	Oui	Oui	Oui	Non	Oui
UC Mandat	Non	Non	Non	Non	Non	Non
Résidence	Non	Non	Non	Oui	Oui	Oui
Ancienneté du contrat	Oui***	Oui***	Oui***	Oui	Oui	Oui

Tableau 35 : Synthèse de la significativité des variables dans le modèle de régression logistique

* : La variable cadre fiscal est globalement significative, mais la modalité « Capitalisation personnes morales IR » n'est pas significative par rapport à la modalité de référence « Assurance-Vie ». Les données de ces deux cadres fiscaux vont par conséquent être agrégées.

** : La variable réseaux de distribution est globalement significative, mais à partir de 2013, il existe toujours au moins une modalité type « Run-off » qui n'est pas statistiquement significative par rapport à la modalité de référence (« Run-Off Retail »). Tous les réseaux de distribution en « Run-off » vont finalement être agrégés.

*** : La variable ancienneté du contrat est globalement significative, mais certaines années (à partir de la 10^{ème} année) ne le sont pas par rapport à la modalité de référence (8^{ème} année d'ancienneté). Cela est certainement dû à la faible représentation de ces années fiscales du fait de l'ancienneté globale du portefeuille étudié. On remarque que depuis 2014, toutes les modalités de la variable ancienneté fiscale sont significatives. C'est pourquoi nous décidons de conserver cette variable telle quelle.

Agrégation des réseaux Run-off et agrégation des modalités Assurance vie et capitalisation personnes morales soumises à l'IR :

Variables \ Année étudiée	2011	2012	2013	2014	2015	Données cumulées
Cadre Fiscal	Oui	Oui	Oui	Oui	Oui	Oui
Réseaux de distribution	Oui	Oui	Oui	Oui	Oui	Oui
Présence de fonds en euros	Oui	Oui	Non	Oui	Non	Oui
Présence d'unités de compte	Non	Oui	Oui	Oui	Oui	Oui
Offre commerciale TMA et/ou Bonus fonds Euroconviction	Non	Oui	Oui	Oui	Oui	Oui
Supports immobiliers	Non	Oui	Oui	Oui	Non	Oui
UC Mandat	Oui	Non	Non	Oui	Non	Oui
Résidence	Oui	Non	Non	Oui	Non	Non
Ancienneté du contrat	Oui*	Oui*	Oui	Oui	Oui	Oui

Tableau 36 : Synthèse de la significativité des variables dans le modèle de régression logistique n°2

* : La variable ancienneté du contrat est globalement significative, mais pas pour certaines modalités de la variable.

Concernant les variables discriminantes :

- Le cadre fiscal, le réseau de distribution et l'ancienneté du contrat sont des variables statistiquement significatives sur tous les échantillons de données observées.
- Hormis pour l'année 2011, le critère d'investissement sur des unités de compte est statistiquement significatif,
- Hormis pour l'année 2011 et 2014, le critère d'investissement sur des supports immobiliers est statistiquement significatif,
- Hormis pour l'année 2011 et 2014, le lieu de résidence n'est pas considéré comme une variable significative pour expliquer le comportement de rachat des assurés.

V.3 La sélection par optimisation

La sélection par optimisation (également appelée régression logistique pas à pas) consiste à trouver le sous-ensemble de variables prédictives qui minimise un critère. Ce critère ne peut pas être la déviance ($-2LL$) puisqu'elle diminue de manière mécanique lorsque l'on ajoute de nouvelles variables, à l'instar de la somme des carrés des résidus dans la régression linéaire.

Il faut par conséquent un critère qui :

- contrebalance la réduction de la déviance,
- comptabilise la qualité de l'ajustement,
- comptabilise la complexité du modèle par le biais d'un indicateur.

Lorsque nous ajoutons des variables pertinentes, le critère doit continuer à décroître ; lorsque nous ajoutons des variables qui ne sont pas en rapport avec la prédiction ou qui sont redondantes par rapport aux variables déjà choisies, le critère doit augmenter.

Deux critères répondent à ces spécifications :

- le critère AIC d'Akaike : $AIC = -2LL + 2 \times (J + 1)$
- ou le critère BIC de Schwartz : $BIC = -2LL + \ln(n) \times (J + 1)$

($J + 1$) est le nombre de paramètres à estimer, avec J le nombre de variables explicatives.

Les différentes possibilités offertes par le logiciel *R* sont :

- Sélection *FORWARD* par critère *AIC* (test du score) : la formule part du modèle trivial (modèle uniquement avec une constante), puis ajoute une à une des variables explicatives jusqu'à ce que la règle d'arrêt soit déclenchée. Les résultats peuvent être différents de ceux réalisés en analysant les rapports du test de Wald.
- Sélection *BACKWARD* (Méthode Descendante par le Test de Wald) : cette méthode part du modèle complet puis supprime une à une les variables tant que le critère *AIC* décroît. Le processus s'arrête dès que l'*AIC* stagne ou croît. Contrairement à la sélection *FORWARD*, cette méthode permet de conserver le lien si une variable est plus explicative du fait de la présence d'une autre variable explicative dans le modèle.
- Sélection *STEPWISE* : dite méthode progressive. C'est une méthode qui alterne celle de la sélection *FORWARD* et celle de la sélection *BACKWARD*.

Dans cette étude, nous expliciterons uniquement les résultats de la méthode *STEPWISE* :

Variables \ Année étudiée	2011	2012	2013	2014	2015	Données cumulées
Cadre fiscal	Oui	Oui	Oui	Oui	Oui	Oui
Réseaux de distribution	Oui	Oui	Oui	Oui	Oui	Oui
Présence de fonds en euros	Oui	Oui	Non	Oui	Non	Oui
Présence d'unités de compte	Non	Oui	Oui	Oui	Oui	Oui
Offre commerciale TMA et/ou Bonus	Non	Oui	Oui	Oui	Oui	Oui
fonds Euroconviction	Non	Non	Non	Oui	Oui	Oui
Supports immobiliers	Oui	Oui	Oui	Oui	Oui	Oui
UC Mandat	Oui	Non	Non	Oui	Non	Oui
Résidence	Oui	Non	Non	Oui	Oui	Oui
Ancienneté du contrat	Oui	Oui	Oui	Oui	Oui	Oui

Tableau 37 : Résultat de la méthode STEPWISE

Agrégation des capitalisations personnes morales IR et des réseaux Run-off :

Variables \ Année étudiée	2011	2012	2013	2014	2015	Données cumulées
Cadre fiscal	Oui	Oui	Oui	Oui	Oui	Oui
Réseaux de distribution	Oui	Oui	Oui	Oui	Oui	Oui
Présence de fonds en euros	Oui	Oui	Non	Oui	Non	Oui
Présence d'unités de compte	Oui	Oui	Oui	Oui	Oui	Oui
Offre commerciale TMA et/ou Bonus	Non	Oui	Oui	Oui	Oui	Oui
fonds Euroconviction	Oui	Non	Non	Oui	Oui	Oui
Supports immobiliers	Oui	Oui	Oui	Oui	Oui	Oui
UC Mandat	Oui	Oui	Non	Oui	Oui	Oui
Résidence	Oui	Non	Non	Oui	Oui	Non
Ancienneté du contrat	Oui	Oui	Oui	Oui	Oui	Oui

Tableau 38 : Résultat de la méthode STEPWISE n°2

En ce qui concerne les variables discriminantes : les critères du cadre fiscal, des réseaux de distribution, de l'ancienneté du contrat, d'un investissement sur les supports en unités de compte ou sur les supports immobiliers sont considérés comme significatifs.

VI Synthèse et sélection des variables

Afin de déterminer les variables les plus discriminantes du comportement de rachat d'un assuré, nous avons suivi différentes méthodes statistiques. Nous croisons leurs résultats :

Méthodologie Variables	Analyse des correspondances Multiples	Forêts aléatoires	Régression logistique	
			Significativité des coefficients	Sélection Pas à Pas
Cadre fiscal	Oui	Oui	Oui	Oui
Réseaux de distribution	Oui	Oui	Oui	Oui
Présence de fonds en euros	Oui	Oui	Oui	Oui
Présence d'unités de compte	Oui	Oui	Oui	Oui
Offre commerciale TMA et/ou Bonus	Oui	Non	Oui	Oui
fonds Euroconviction	Oui	Non	Ne se prononce pas	Oui
Supports immobiliers	Oui	Ne se prononce pas	Oui	Oui
UC Mandat	Oui	Non	Ne se prononce pas	Oui
Résidence	Non	Non	Non	Ne se prononce pas
Ancienneté du contrat	Non étudié	Oui	Oui	Oui

Tableau 39 : Synthèse des résultats des différents modèles

Le modèle sur lequel nous nous basons pour réaliser la projection des données 2016 comprend toutes les variables explicatives de l'étude hormis la Résidence. Nous testerons la sensibilité de l'agrégation des données en « Run-off » et de l'agrégation du cadre fiscal « personnes morales IR » avec le cadre fiscal d'Assurance-Vie.

Fort de ces résultats, nous avons sélectionné les variables discriminantes dans le comportement de rachat total des assurés du portefeuille. Comme indiqué lors de l'analyse des arbres de décision, le *back-testing* sur les données 2016 ne fournissait pas des résultats concluants. La prochaine partie se concentre alors sur la robustesse du modèle de régression logistique sur les prévisions des données 2016.

Partie III : Validation du modèle de régression logistique et de son pouvoir prédictif

Nos données ont été séparées en deux échantillons :

- Les données permettant de créer le modèle de prédiction. Cet échantillon dit « d'apprentissage » contient les observations du portefeuille à pas de temps annuel de 2011 à 2015.
- Les données permettant de valider le modèle ainsi que son pouvoir prédictif. Cet échantillon dit de « validation » contient les observations du portefeuille de l'année 2016.

Cette technique de validation « temporelle » permet de rendre compte de l'exposition des assurés à des contextes conjoncturels différents et ainsi de tester la qualité du modèle et également son pouvoir prédictif de simulations de futurs taux de rachat.

I Risque Relatif et Odds-ratio

1.1 Définitions

On appelle risque relatif (*RR*) le surcroît de chances, pour le groupe exposé, d'être positif par rapport au groupe témoin.

Partons de l'exemple ci-dessous :

	Angine	
Maladie cardiaque	Oui	Non
Oui	A = 3	B = 3
Non	C = 2	D = 12

$$\text{Nous définissons, } RR = \frac{\frac{a}{(a+c)}}{\frac{b}{(b+d)}} = \frac{\frac{3}{5}}{\frac{3}{15}} = 3$$

Les personnes qui ont une angine ont 3 fois plus de chances que celles qui n'en ont pas de développer une maladie cardiaque. Si $RR = 1$, la variable n'a pas d'incidence sur l'évènement étudié.

L'Odds ou rapport de chances est défini comme un rapport de probabilités dans un groupe.

$$\text{Par exemple : } odds(1) = \frac{\frac{a}{(a+c)}}{\frac{c}{(c+a)}} = \frac{\frac{3}{5}}{\frac{2}{5}} = 1.5$$

Dans le groupe des personnes ayant une angine de poitrine, on a 1,5 fois plus de chances d'avoir une maladie cardiaque que de ne pas en avoir.

L'Odds-ratio est égal au rapport entre l'Odds du groupe exposé et l'Odds du groupe témoin. :

$$OR = \frac{odds(1)}{odds(0)} = 6$$

La notion d'Odds-ratio (OR) est privilégiée car celle-ci possède la propriété d'être invariante au mode d'échantillonnage contrairement au Risque Relatif. Si la probabilité de l'évènement est rare alors ces deux ratios prennent des valeurs similaires.

Dans la régression logistique :

Lorsque nous procédons à un codage binaire simple, les coefficients de la régression logistique correspondent aux logs Odds-ratios de chaque modalité par rapport à la modalité de référence (celle qui a été exclue du recodage).

1.2 Interprétation des coefficients

Nous présentons ci-dessous les Odds-ratios sur le modèle réalisé à partir des variables sélectionnées dans la Partie II.

Afin de simplifier la lecture lorsqu'un Odds-ratio, lorsque celui-ci est inférieur à 1, nous le remplaçons par la valeur $\frac{-1}{OR}$. Exemple : si le log Odds-ratio = 0,664 alors le coefficient indiqué est -1,5 (ce qui signifie qu'il y a 1,5 fois moins de risque de Rachat Total) :

Variables \ Année	2011	2012	2013	2014	2015	Cumul
Classification.ReseauPartenariat_Aise	-3,48	-3,12	-2,32	-1,75	-1,75	-2,48
Classification.ReseauPartenariat_Gestion de Fortune	-5,73	-1,90	-2,96	-1,88	-2,95	-2,33
Classification.ReseauPartenariat_Patrimonial	-2,73	-2,03	-1,96	-1,65	-2,25	-2,01
Classification.ReseauRun-off_Aise	-1,66	-1,33	-1,12	1,13	1,28	-1,10
Classification.ReseauRun-off_Gestion de Fortune	-1,88	-1,80	-3,73	-3,48	-2,80	-1,96
Classification.ReseauRun-off_Patrimonial	-3,17	-1,47	-1,81	-1,36	-1,12	-1,54
Anciennete.c0	-3,21	-6,87	-7,75	-6,41	-8,53	-6,33
Anciennete.c1	-1,58	-2,12	-2,74	-3,19	-4,13	-2,66
Anciennete.c2	-1,24	-2,04	-2,40	-2,60	-3,04	-2,18
Anciennete.c3	1,40	-1,73	-2,41	-2,47	-3,36	-1,77
Anciennete.c4	1,16	1,04	-2,56	-2,55	-2,52	-1,48
Anciennete.c5	-1,36	-1,20	-1,32	-3,28	-3,14	-1,62
Anciennete.c6	-1,63	-1,74	-1,84	-1,55	-4,28	-1,84
Anciennete.c7	-1,51	-1,69	-2,58	-2,28	-1,40	-1,86
Anciennete.c9	-1,71	-2,00	-1,51	-1,95	-1,52	-1,70
Anciennete.c10	7,45	-1,50	-1,48	-1,70	-2,72	-1,94
Anciennete.c11	1,69	-2,06	-7,02	-5,57	-3,12	-3,66
Anciennete.c12	3,85	1,31	-6,83	-50,37	-4,62	-4,33
Anciennete.c13+	1,69	-3,67	-2,45	-2,57	-5,81	-2,81
Fonds en euros	-1,19	-1,11	1,05	-1,10	-1,00	-1,19
UC.Mandat	-1,31	-1,09	1,13	-1,40	-1,19	-1,12
UC	-1,06	-1,22	-1,52	-2,04	-2,30	-1,59
EuroConviction	-1,47	-1,07	1,01	-2,45	-2,12	-1,56
Cadre_fiscal_v2Capitalisation pers. morale IR	1,92	-1,10	1,49	1,13	1,15	1,31
Cadre_fiscal_v2Capitalisation pers. morale IS	2,55	2,49	1,80	1,59	1,59	1,85
SCPI	-1,24	-1,38	-1,74	-1,43	-1,12	-1,28
TMA Bonus	-1,60	-2,13	-3,19	-2,72	-3,04	-2,85

Tableau 40 : Odds-ratios du modèle de régression logistique

Nous constatons que les contrats bénéficiant d'une offre TMA/Bonus sur le fonds en euros ou investissant sur des unités de compte sont des contrats deux fois moins rachetés que les contrats ne présentant pas ces caractéristiques.

Par ailleurs, malgré ces coefficients significatifs, il apparaît que l'absence d'un fonds en euros et l'absence et d'une gestion sous mandat présentent un risque de rachat total supplémentaire très faible par rapport aux contrats présentant ces modalités (toutes choses égales par ailleurs).

Les valeurs de ces Odds-ratios corroborent également les premiers résultats par rapport à la significativité du cadre fiscal Capitalisation personnes morales IR par rapport aux contrats de cadre fiscal Assurance-Vie.

Ainsi, nous testerons également les régressions sans prendre en compte ces variables sur l'année 2016.

II Evaluation du modèle de régression logistique

II.1 Matrice de confusion et pseudo-R²

Pour rappel, si nous n'appliquons aucune segmentation sur les données de 2016, le taux de mauvaise classification (similaire au taux de rachat total constaté) serait de 4,65%.

Par ailleurs, en se basant sur la matrice de confusion suivante :

Valeur Observée \ Valeur Prédite	0	1
0	Vrai Positifs (VP)	Faux Négatifs (FN)
1	Faux Positifs (FP)	Vrais Négatifs (VN)

Tableau 41 : Matrice de Confusion

Les notions suivantes sont introduites :

- La sensibilité, dit également le taux de vrais positifs (TVP), indique la capacité du modèle à retrouver les positifs : $\frac{VP}{(VP+FN)}$
- La précision indique la proportion de vrais positifs parmi les individus qui ont été classés positifs : $\frac{VP}{VP+FP}$ On parle de valeur prédictive positive.
- La spécificité, à l'inverse de la sensibilité indique la proportion de négatifs détectés : $\frac{VN}{(VN+FP)}$

Le taux de faux positifs (TFP) correspond à la proportion de négatifs qui ont été classés positifs $TFP = 1 - \text{spécificité}$.

Sous ces conditions, le modèle sélectionné devra présenter un taux d'erreur faible et des valeurs élevées de sensibilité, précision et spécificités. Cependant, lorsqu'on oriente l'apprentissage afin d'améliorer la sensibilité, la précision et la spécificité sont souvent dégradées. Un modèle présentant les meilleurs résultats sur ces deux typologies est celui à retenir.

En complément, nous introduisons l'indicateur des pseudos- R^2 afin de confronter les taux de mauvaise classification. Les pseudos- R^2 résultent de l'opposition, sous différentes formes, de la vraisemblance du modèle étudié avec celle du modèle trivial. Cela permet de quantifier si le modèle réalisé « surperforme » le modèle trivial. Dans notre étude, nous réaliserons un pseudo- R^2 basé sur le taux d'erreur : $R_e^2 = 1 - \frac{Erreur_{Modèle}}{Erreur_{Trivial}}$

Si le modèle a un taux d'erreur nul, $R_e^2 = 1$.

A contrario, si le modèle ne sait pas faire mieux que le modèle par défaut alors $R_e^2 = 0$.

Dans un premier temps, ci-dessous ces indicateurs sur les différents modèles réalisés uniquement sur une année d'observation et sur un modèle agrégeant les données de 2011 à 2015 :

Année étudiée	2011	2012	2013	2014	2015	2011 à 2015
Toutes Variables						
Mauvaise Classification	9,72%	5,77%	4,48%	4,59%	4,29%	4,57%
Sensibilité	0,93	0,97	0,98	0,98	0,99	0,98
Précision	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,42	0,38	0,36	0,33	0,32	0,37
Pseudo- R^2	-1,09	-0,24	0,03	0,01	0,08	0,02

Année étudiée	2011	2012	2013	2014	2015	2011 à 2015
Sans Fonds en €						
Mauvaise Classification	9,05%	5,08%	3,98%	4,23%	4,01%	4,03%
Sensibilité	0,93	0,97	0,99	0,99	0,99	0,99
Précision	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,50	0,46	0,42	0,37	0,35	0,43
Pseudo- R^2	-0,95	-0,09	0,14	0,09	0,14	0,13

Année étudiée	2011	2012	2013	2014	2015	2011 à 2015
Sans UC Mandat						
Mauvaise Classification	9,62%	5,71%	4,42%	4,57%	4,26%	4,52%
Sensibilité	0,93	0,97	0,98	0,98	0,99	0,98
Précision	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,42	0,39	0,37	0,34	0,32	0,37
Pseudo- R^2	-1,07	-0,23	0,05	0,02	0,08	0,03

Année étudiée	2011	2012	2013	2014	2015	2011 à 2015
Sans Fds € & UC Mandat						
Mauvaise Classification	9,62%	5,71%	4,42%	4,57%	4,26%	4,00%
Sensibilité	0,93	0,97	0,99	0,99	0,99	0,99
Précision	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,51	0,46	0,42	0,37	0,35	0,43
Pseudo- R^2	-0,93	-0,08	0,15	0,09	0,14	0,14

Tableau 42 : Taux de Classification de différents modèles

Nous constatons que les modèles basés sur les données de 2011 et 2012 impliquent une bien moins bonne classification des contrats alors que les résultats des modèles 2013/2014/2015 et les données cumulées déterminent mieux la classification des contrats que si nous considérons qu'aucun rachat ne soit réalisé dans l'année.

Par ailleurs, la sensibilité et la précision des modèles sont très élevées (proche de 1) alors que la spécificité ne dépasse pas 0,5.

De plus, les modèles sans les variables représentant la présence de fonds en euros et/ou la gestion sous mandat améliorent les résultats de classification et la spécificité.

L'année 2011 présente des résultats très distincts des autres années mais nous remarquons que le modèle présentant le plus de données d'apprentissage semble plus performant. Ainsi pour la suite, nous nous présenterons les modèles réalisés sur les données agrégées :

- 2011 à 2015 : utilisation de toutes les données à disposition
- 2012 à 2015 : exclusion de l'année 2011 qui présente les plus grosses différences de caractéristiques,
- 2013 à 2015 : utilisation uniquement des données où les modèles « battent » le modèle trivial.

Voici maintenant les résultats, selon les variables discriminantes conservées, obtenus :

- sans aucune agrégation de modalités,
- avec agrégat des modalités Assurance-Vie et Capitalisation personnes morales IR,
- avec agrégat des segments de clientèle des modalités « Run-Off* »

Toutes Variables	2011*	2012*	2013*	2011* AssVie/Capi	2012* AssVie/Capi	2013* AssVie/Capi	2011* Run off	2012* Run off	2013* Run off
Mauvaise Classification	4,57%	4,47%	4,37%	4,56%	4,47%	4,36%	4,43%	4,35%	4,25%
Sensibilité	0,98	0,98	0,99	0,98	0,98	0,99	0,98	0,99	0,99
Précision	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,37	0,36	0,35	0,37	0,36	0,35	0,38	0,37	0,35
Pseudo-R ²	0,02	0,04	0,06	0,02	0,04	0,06	0,05	0,06	0,08

Sans Fonds en €	2011*	2012*	2013*	2011* AssVie/Capi	2012* AssVie/Capi	2013* AssVie/Capi	2011* Run off	2012* Run off	2013* Run off
Mauvaise Classification	4,03%	3,99%	3,99%	4,02%	3,99%	3,98%	3,84%	3,84%	3,86%
Sensibilité	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99
Précision	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,43	0,41	0,39	0,43	0,41	0,39	0,44	0,42	0,40
Pseudo-R ²	0,13	0,14	0,14	0,13	0,14	0,14	0,17	0,17	0,17

Sans UC Mandat	2011*	2012*	2013*	2011* AssVie/Capi	2012* AssVie/Capi	2013* AssVie/Capi	2011* Run off	2012* Run off	2013* Run off
Mauvaise Classification	4,52%	4,42%	4,33%	4,51%	4,42%	4,32%	4,38%	4,31%	4,23%
Sensibilité	0,98	0,98	0,99	0,98	0,98	0,99	0,98	0,99	0,99
Précision	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,37	0,37	0,35	0,38	0,37	0,35	0,38	0,37	0,36
Pseudo-R ²	0,03	0,05	0,07	0,03	0,05	0,07	0,06	0,07	0,09

Sans Fonds en € et sans UC Mandat	2011*	2012*	2013*	2011* AssVie/Capi	2012* AssVie/Capi	2013* AssVie/Capi	2011* Run off	2012* Run off	2013* Run off
Mauvaise Classification	4,00%	3,97%	3,96%	4,00%	3,97%	3,95%	3,83%	3,83%	3,85%
Sensibilité	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99	0,99

Précision	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97	0,97
Spécificité	0,43	0,41	0,39	0,43	0,41	0,39	0,44	0,42	0,40
Pseudo-R ²	0,14	0,15	0,15	0,14	0,15	0,15	0,18	0,18	0,17

Tableau 43 : Taux de Classification de différents modèles agrégés

Nous constatons que les modèles :

- sans la variable Fonds en euros
- sans la variable Fonds en euros et sans la variable UC Mandat

présentent les meilleurs taux de mauvaises classifications et de spécificités. Cependant, dans ces modèles, celui effectué sur les données à partir de 2013 classe moins bien les contrats que les modèles effectués avec les données de 2012 et 2011.

Par ailleurs, les modèles avec une agrégation des modalités « Run-off* » présentent une amélioration de la classification. En revanche, l'indicateur « spécificité » reste relativement faible.

Grâce aux indicateurs fournis via la matrice de confusion, plusieurs modèles « sortent » du lot. Cependant, un des écueils de la matrice de confusion est que celle-ci est sensible à l'importance relative des groupes. Si les groupes sont déséquilibrés, les résultats fournis sont moins pertinents. Dans notre échantillon, le groupe de la population rachetant leur contrat est en grande infériorité. Nous couplons ces premiers résultats avec l'étude d'un deuxième indicateur ne présentant pas ce biais : la courbe ROC et le critère qui lui est associé l'AUC.

II.2 La courbe ROC

La courbe ROC (*Receiver Operating Signal* : Transmission de Signal) est un outil très intéressant pour l'évaluation et la comparaison des performances des modèles :

- Les résultats peuvent être fournis sous forme de graphique,
- Elle permet de déterminer si un modèle surpasse un autre,
- Elle est opérationnelle dans le cas des distributions très déséquilibrées,
- Un indicateur synthétique interprétable lui est associé : le critère AUC.

Pour l'ensemble des observations, il y a une valeur observée Y et un score \hat{S} qui permet de prédire \hat{Y} . La règle d'affectation est alors réalisée en fixant un seuil s et :

- Si $\hat{S} > s$, alors \hat{Y} est « positif »,
- Si $\hat{S} < s$, alors \hat{Y} est « négatif ».

En faisant varier s de 0 à 1, nous déterminons, via la matrice de confusion associée, le taux de vrais positifs (TVP, la sensibilité) et le taux de faux positifs (TFP = 1 – spécificité). La courbe ROC correspond au graphique du nuage de points qui relie les couples (TVP, TFP). Les deux situations encadrant les courbes possibles :

- Tous les positifs sont situés devant les négatifs, la courbe ROC est aux extrémités du repère.
- Le modèle attribue des valeurs au hasard, les positifs et les négatifs sont mélangés : la courbe ROC se confond avec la première bissectrice.

Ci-dessous un graphique³⁴ présentant différents résultats de courbe ROC :

³⁴Extrait du support de cours STA 201 du Cnam de Gilbert Saporta : http://cedric.cnam.fr/~saporta/Sensibilite_specificiteSTA201.pdf

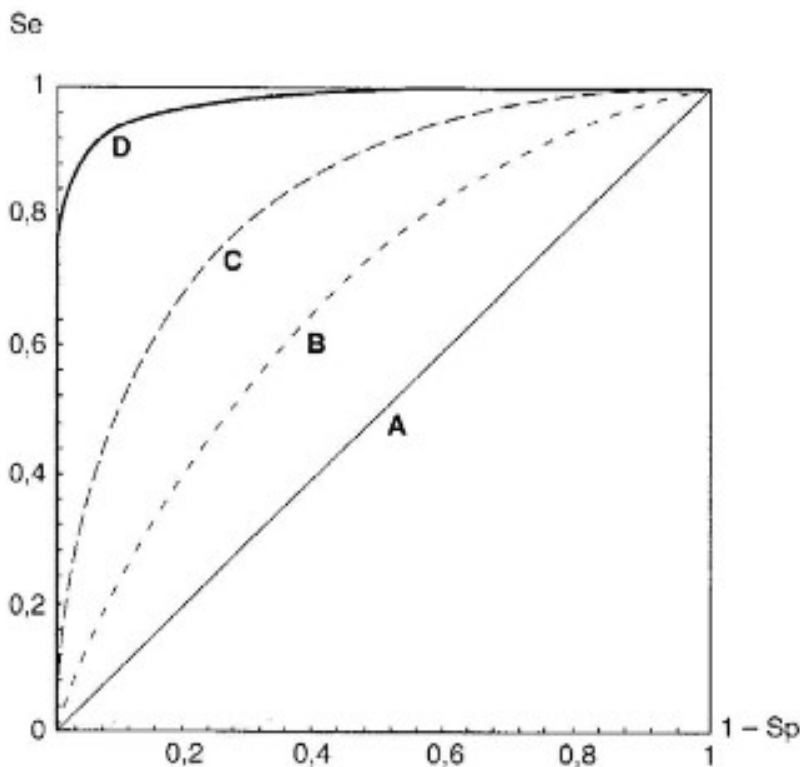


Figure 3-7 Différentes courbes ROC. A : test sans intérêt diagnostique. B : mauvais test diagnostique. C : test meilleur que B. D : bon test diagnostique.

Figure 23 : Différentes courbes ROC

On caractérise numériquement la courbe ROC en déterminant le critère AUC (Area Under the Curve) qui correspond à la surface située sous la courbe. Elle exprime la probabilité de placer un individu positif devant un négatif.

On peut déterminer la valeur de l'AUC via la méthode des trapèzes :

$$AUC = \sum_{i=1}^n s_i \text{ Où } s_i = (TFP_i - TFP_{i-1}) \times \frac{TVP_i + TVP_{i-1}}{2}$$

Ainsi, dans le cas d'une discrimination parfaite $AUC = 1$. A contrario, si le modèle attribue des scores au hasard, $AUC = 0,5$. Le modèle doit alors faire mieux.

Différents paliers³⁵ sont proposés pour évaluer la qualité de la discrimination :

Valeur de l'AUC	Commentaire
$AUC = 0,5$	Pas de discrimination
$0,7 \leq AUC < 0,8$	Discrimination acceptable
$0,8 \leq AUC < 0,9$	Discrimination excellente
$AUC \geq 0,9$	Discrimination exceptionnelle

Tableau 44 : Interprétation des valeurs de l'AUC

³⁵ Paliers fournis dans l'ouvrage Pratique de la Régression Logistique, Ricco Rakotomalala

Ci-dessous, la valeur du critère AUC des différents modèles testés :

AUC \ Année étudiée	2011	2012	2013	2014	2015	2011 à 2015
Toutes Variables	0,677	0,787	0,834	0,807	0,836	0,835
Sans Fonds en €	0,674	0,786	0,832	0,810	0,837	0,842
Sans UC Mandat	0,678	0,785	0,835	0,806	0,837	0,835
Sans Fonds en € et sans UC Mandat	0,675	0,788	0,834	0,810	0,838	0,842

Année étudiée \ AUC	2011*	2012*	2013*	2011* AssVie/Capi	2012* AssVie/Capi	2013* AssVie/Capi	2011* Run off	2012* Run off	2013* Run off
Toutes Variables	0,835	0,840	0,842	0,835	0,840	0,842	0,839	0,843	0,844
Sans Fonds en €	0,842	0,845	0,844	0,841	0,844	0,844	0,847	0,847	0,848
Sans UC Mandat	0,835	0,840	0,842	0,834	0,840	0,842	0,838	0,842	0,844
Sans Fonds en € et sans UC Mandat	0,842	0,844	0,844	0,841	0,844	0,844	0,847	0,847	0,847

Tableau 45 : Taux de Classification de différents modèles agrégés

En se basant sur les seuils indiqués ci-dessus, nous constatons que tous les modèles sur les données agrégées ont une valeur d'AUC très satisfaisante. Les écarts sont même parfois très minimes.

Les modèles présentant les meilleurs AUC sont ceux où les modalités de la typologie d'un contrat Run-Off sont agrégées sans la variable Fonds en euros, soit sans la variable de la gestion sous mandat en complément. Les taux entre les modèles réalisés sur les données partant de 2011, de 2012 et de 2013 jusqu'à 2015 sont quasi-identiques.

Ci-dessous, 2 graphiques représentant les courbes ROC associées aux modèles sans Fonds en euros puis aux modèles sans Fonds en euros et sans gestion sous Mandat.

La courbe verte présente le modèle réalisé sur les données de 2011 à 2015.

La courbe bleue présente le modèle réalisé sur les données de 2012 à 2015.

La courbe rouge présente le modèle réalisé sur les données de 2013 à 2015.

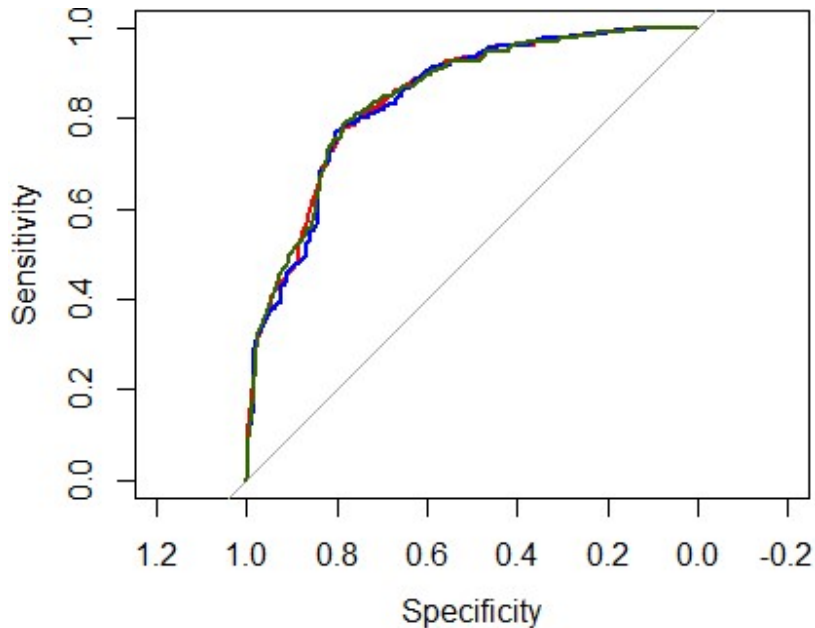


Figure 24 : Courbes ROC des modèles « Run-Off* » sans la variable Fonds en euros

Il apparaît que les courbes ROC verte et rouge se chevauchent à certains endroits. Malgré une valeur AUC légèrement plus forte, le modèle 2013 n'est pas le plus performant en tous points de seuils testés.

En revanche, la courbe bleue est inférieure ou égale en tous points aux valeurs d'un autre modèle.

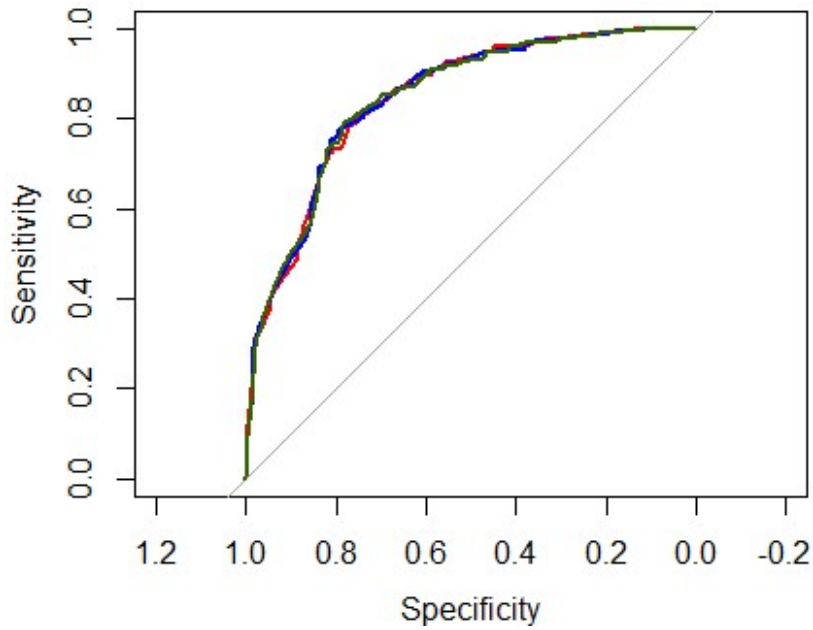


Figure 25 : Courbes ROC des modèles « Run-Off* » sans les variables Fonds en euros et Gestion sous mandat

Tout comme les modèles réalisés sans la variable Fonds en euros, les modèles ci-dessus présentent une valeur d'AUC identique, cependant les courbes sont très légèrement différentes.

Par ailleurs, en se basant sur les résultats de la matrice de confusion, le modèle de 2011 sans les variables Fonds en euros et Gestion sous mandat présentait de meilleurs résultats d'indicateurs que le modèle de 2013 sans la variable Fonds en euros.

En comparant les 2 courbes ci-dessous :

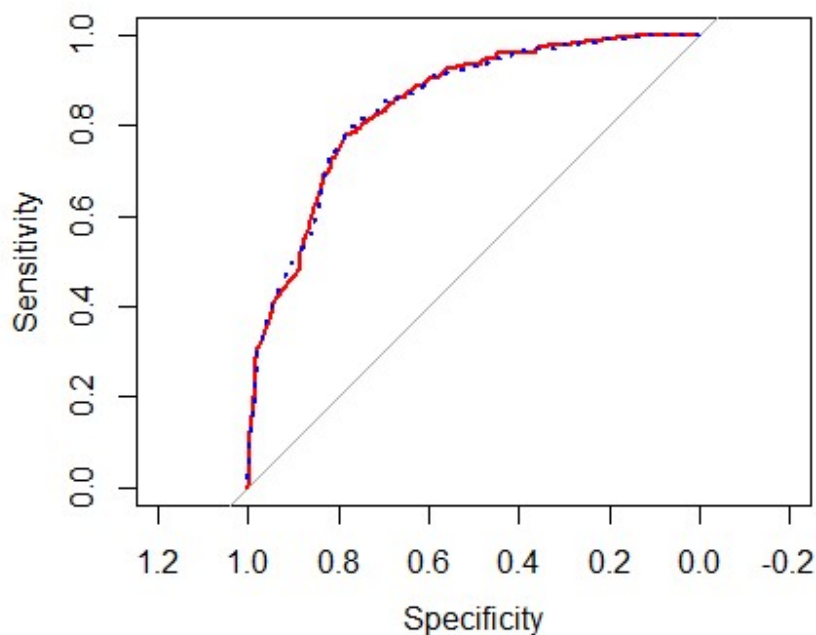


Figure 26 : Courbe ROC « Run-Off* » données 2011 et 2013

La courbe bleue présente le modèle réalisé sans les variables Fonds en euros et Gestion sous mandat sur les données 2011.

La courbe rouge présente le modèle réalisé sans la variable Fonds en euros sur les données 2013. Les courbes sont extrêmement similaires, se confondant parfois et se chevauchant très légèrement à certains endroits.

II.3 Synthèse

Le modèle qui nous semble le plus approprié suite aux différents critères d'évaluations est le suivant :

- Population : Données de 2011 à 2015. Cela est cohérent du fait de l'augmentation de la profondeur des observations ce qui permet une plus grande robustesse des modèles.
- Non prise en compte de l'indication sur la présence du Fonds en euros : malgré la significativité de cette variable, l'Odds-ratio est très faible et les modèles sans cette variable présentent de meilleurs indicateurs de performances. Par ailleurs, cette indication est sous-jacente des autres variables (absence d'UC implique la présence du Fonds en euros / La présence d'une offre Commerciale TMA/Bonus implique la présence du Fonds en euros, etc.).
- Sur la base du principe de raisonnement du rasoir d'Ockham, l'indication sur la présence d'une Gestion sous mandat ne sera pas prise en compte, les modèles présentant des indicateurs de performances quasi-identiques à ceux dont la variable est présente.

Enfin, nous vérifions que les coefficients du modèle sélectionné sont significatifs :

```

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -1.61831    0.02724  -59.411 < 2e-16 ***
Reseau.distribPartenariat_Aise -0.86577    0.02806  -30.855 < 2e-16 ***
Reseau.distribPartenariat_Gestion de Fortune -0.82130    0.06303  -13.030 < 2e-16 ***
Reseau.distribPartenariat_Patrimonial -0.67398    0.03325  -20.271 < 2e-16 ***
Anciennete.c0     -1.86004    0.04884  -38.083 < 2e-16 ***
Anciennete.c1     -0.98510    0.03699  -26.633 < 2e-16 ***
Anciennete.c10    -0.66032    0.04944  -13.357 < 2e-16 ***
Anciennete.c11    -1.29668    0.08963  -14.467 < 2e-16 ***
Anciennete.c12    -1.49379    0.15298   -9.765 < 2e-16 ***
Anciennete.c13+  -1.13129    0.14049   -8.052 8.13e-16 ***
Anciennete.c2     -0.77819    0.03597  -21.634 < 2e-16 ***
Anciennete.c3     -0.57445    0.03437  -16.714 < 2e-16 ***
Anciennete.c4     -0.39258    0.03216  -12.206 < 2e-16 ***
Anciennete.c5     -0.47828    0.03199  -14.951 < 2e-16 ***
Anciennete.c6     -0.61798    0.03298  -18.738 < 2e-16 ***
Anciennete.c7     -0.62476    0.03343  -18.687 < 2e-16 ***
Anciennete.c9     -0.53097    0.03811  -13.932 < 2e-16 ***
UC                -0.33574    0.02028  -16.558 < 2e-16 ***
EuroConviction    -0.50376    0.08149   -6.182 6.33e-10 ***
Cadre_fiscal_v2Capitalisation pers. morale IR  0.16749    0.09963   1.681  0.0927 .
Cadre_fiscal_v2Capitalisation pers. morale IS  0.45939    0.04336  10.594 < 2e-16 ***
SCPI              -0.33598    0.03543   -9.483 < 2e-16 ***
TMA_Bonus         -1.06632    0.10727  -9.940 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 123019 on 281113 degrees of freedom
Residual deviance: 116950 on 281091 degrees of freedom
AIC: 116996

```

Figure 27 : Coefficients du modèle de régression logistique sélectionné

Le modèle sélectionné présente des coefficients significatifs et des indicateurs de performances très satisfaisants aux vues de la matrice de confusion et de la courbe ROC.

Cependant, la courbe ROC ainsi que le critère AUC permettent de déterminer si le modèle identifie assez bien les positifs mais en revanche ne permettent pas de juger si la probabilité prédite est pertinente.

Ainsi, dans la suite de l'étude, nous comparons les résultats fournis par le modèle de régression logistique choisi avec le modèle utilisé actuellement basé sur l'estimateur Kaplan-Meier afin d'étudier la pertinence des résultats du modèle.

III Application du modèle de régression logistique sur l'année 2016

III.1 Premiers résultats

Pour rappel le taux de rachat constaté sur les données de 2016 est de 4,65%.

Le taux de rachat déterminé des données 2016 via le modèle est de 3,28%. Ce taux est en cohérence avec l'indicateur de spécificité (détection des négatifs) relativement faible du modèle.

Pour affiner la comparaison, nous déterminons, sur la base des données 2016, le taux de rachat de chaque ancienneté fiscale ainsi que les taux réels constatés en 2016 :

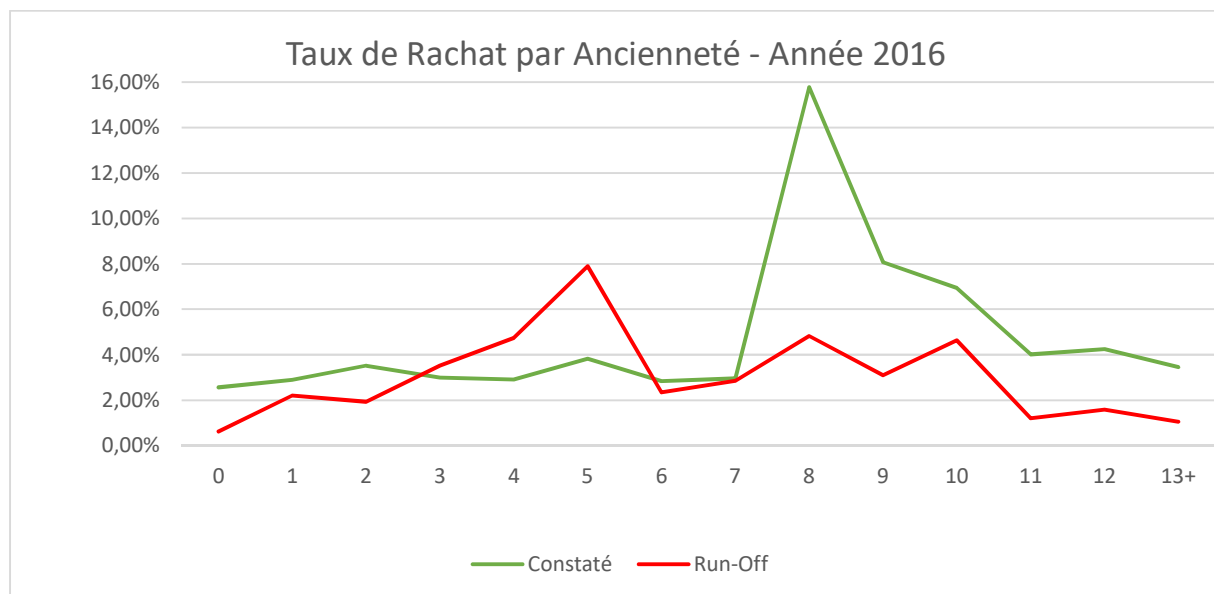


Figure 28 : Graphique du taux de Rachat Total par Ancienneté

III.2 Comparaison avec la méthode actuelle

A l'heure actuelle, la loi de rachats structurels utilisée dans l'entreprise est définie par ancienneté fiscale et par réseau de distribution. Le portefeuille étudié, lui, est considéré comme un réseau de distribution distinct.

La méthode utilisée, pour l'analyse des contrats en cours et sortis, est l'estimateur de Kaplan Meier. C'est une méthode d'analyse des données de survie dites "censurées" ; elle permet d'inclure tous les contrats d'une étude de survie, quel que soit le recul d'observation. Le calcul des taux de survie repose sur le principe des probabilités conditionnelles. Les taux sont évalués dès que survient un événement (rachat total ou décès).

Définition d'une variable censurée

Au moment où nous analysons des données, nous disposons en général d'un échantillon composé de différents types d'informations :

- Certains contrats sont sortis du portefeuille, via l'évènement étudié (exemple : rachetés totalement) et nous disposons alors d'une information complète, de la forme $(T = t)$, où T est la variable aléatoire « durée de survie ».
- D'autres contrats, au contraire, sont en portefeuille et l'évènement étudié ne s'est pas déclenché pendant la période d'observation (exemple : en vigueur ou décédé pour l'étude des comportements de rachat total). Tout ce que nous connaissons de la durée de survie de ces contrats est que celle-ci est supérieure au temps qui s'est écoulé entre l'instauration du traitement et le moment de l'analyse. Nous avons ainsi une information de la forme $(T > t)$.

Une donnée dont nous ne connaissons qu'une borne est dite **censurée**.

Un échantillon comportant, en plus des données « habituelles » (c'est à dire dont la valeur est observée/observable), des données censurées, est qualifié d'« échantillon censuré ».

Définition de l'estimateur de Kaplan-Meier

Si tous les contrats ayant subi une censure sont supprimés de la base de données, il y a une perte d'information importante pour l'estimation des taux de rachat.

L'estimateur de Kaplan-Meier évite cette perte d'information dans le cas d'une censure à droite.

Être encore en portefeuille après un instant t c'est :

- être en portefeuille juste avant t ,
- et ne pas sortir en t .

Définissons les variables suivantes :

- n_j : nombre de contrats en cours, juste avant le temps j ,
- m_j : nombre de sorties observées au temps j ,
- q_j : proportion observée de contrats en cours au jour j , parmi ceux qui étaient en cours juste avant le jour j : $q_j = \frac{n_j - m_j}{n_j}$

Soit les hypothèses suivantes :

- Si $m_j = 0$ alors $q_j = 1$
- Seuls interviennent effectivement dans les calculs les jours où des sorties ont été observées.
- La fonction de survie est constante entre deux temps de sorties (rachats/renoncations sans effet) consécutifs.
- Les données censurées affectent n_j .

On estime ainsi la fonction de survie, $S(t) = \prod_{j=0}^t q_j = \prod_{j < t} \frac{n_j - m_j}{n_j}$.

Voici ci-dessous un graphique permettant de comparer la loi calculée via l'estimateur Kaplan-Meier, les taux de rachat constatés et le modèle déterminé appliqué aux données 2016 :

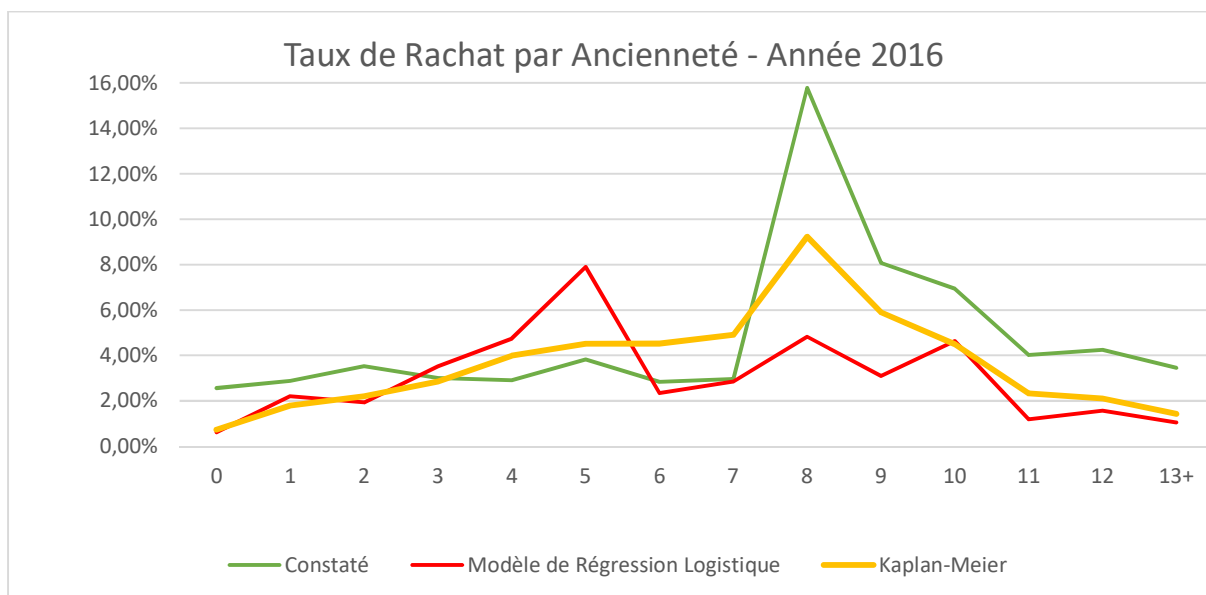


Figure 29 : Graphique du taux de Rachat Total par ancienneté – Régression logistique vs Kaplan Meier

Les taux déterminés par l’estimateur de Kaplan-Meier impliquent une surestimation des rachats constatés des contrats d’ancienneté de 4 à 7 alors que le modèle de régression logistique a une tendance à la sous-estimation par rapport aux taux effectifs (hormis les anciennetés de 4 et 5 ans). Par ailleurs, le pic des 8 ans est mieux estimé que celui du modèle réalisé par la régression logistique. A partir de l’ancienneté 10, nous pouvons considérer que les taux des 2 méthodes sont assez similaires.

Malgré le pic des 8 ans présent mais sous-estimé, le modèle de régression logistique propose globalement une estimation pertinente des taux de rachats. En outre, le modèle de régression logistique permet une plus grande diversification de critères discriminants lors de l’étude de comportement de rachat total, ce qui est intéressant pour l’appui commercial lors de réalisations d’affaires nouvelles qui est l’objectif principal de l’étude.

Suite à la validation de ce modèle et à la significativité de ces résultats, nous pouvons déterminer un « profilage » des contrats proposés à la souscription et ainsi estimer à l’avance le comportement potentiel de rachat.

IV Outil de Pilotage

Ci-dessous les Odds-ratio du modèle sélectionné :

Variabes	Cumul
Partenariat_Aise	-2,38
Partenariat_Gestion de Fortune	-2,27
Partenariat_Patrimonial	-1,96
Anciennete.c0	-6,42
Anciennete.c1	-2,68
Anciennete.c2	-2,18
Anciennete.c3	-1,78
Anciennete.c4	-1,48
Anciennete.c5	-1,61

Anciennete.c6	-1,86
Anciennete.c7	-1,87
Anciennete.c9	-1,70
Anciennete.c10	-1,94
Anciennete.c11	-3,66
Anciennete.c12	-4,45
Anciennete.c13+	-3,10
UC	-1,40
Euroconviction	-1,65
Cadre_fiscal_v2Capitalisation pers. morale IR	1,18
Cadre_fiscal_v2Capitalisation pers. morale IS	1,58
SCPI	-1,40
TMA_Bonus	-2,90

Tableau 46 : Odds-ratios du modèle de régression logistique choisi

Ces rapports confirment les points suivants :

- Les contrats dont les distributeurs sont commercialement actifs sont moins rachetés que les contrats dont la relation avec le partenaire n'existe plus. Plus l'encours moyen est important, plus le rapport faiblit (et plus la probabilité de rachat augmente). Les contrats de capitalisation souscrits par des personnes morales soumises à l'IS présentent un risque plus fort que les contrats d'assurance-vie (donc souscrits par des personnes physiques).
- Globalement, les contrats présentant une offre commerciale font moins l'objet de rachats que les contrats sans une offre (UC ou Fonds en euros).

En complément de ces rapports de chances pour chaque catégorie d'une variable, l'objectif principal de cette étude est de mettre à disposition un outil d'aide à la décision lors d'appels d'offres pour une nouvelle souscription concernant les probabilités de rachats totaux de manière « anticipée ». On considère qu'un rachat est anticipé s'il est réalisé dans les 8 premières années du contrat (en se basant sur les avantages fiscaux). Par ailleurs, ce profilage étant utilisé lors de la souscription, ces données ne sont pas utiles pour la population des partenaires devenue commercialement inactive. De ce fait, nous conservons uniquement la notion d'encours moyen des contrats suivis par des partenaires (Aisé / Patrimonial / Gestion de Fortune).

En se basant sur les coefficients calculés par le modèle, nous déterminons les différents taux ci-dessous :

Cadre fiscal	Encours Moyen	UC	SCPI	Euroconviction	TMA_Bonus	Proba de Rachat
Capitalisation pers. morale IS	Patrimonial	0	0	0	0	39%
Assurance-Vie	Patrimonial	0	0	0	0	32%
Capitalisation pers. morale IS	Aisé	1	0	0	0	30%
Capitalisation pers. morale IS	Patrimonial	1	0	0	0	29%
Assurance-Vie	Aisé	0	0	0	0	26%
Assurance-Vie	Gestion de Fortune	0	0	0	0	25%
Assurance-Vie	Patrimonial	1	0	0	0	24%
Capitalisation pers. morale IR	Aisé	1	0	0	0	24%
Capitalisation pers. morale IR	Patrimonial	1	0	0	0	24%
Assurance-Vie	Gestion de Fortune	1	0	0	0	21%
Assurance-Vie	Aisé	1	0	0	0	21%
Capitalisation pers. morale IS	Patrimonial	0	0	0	1	19%
Assurance-Vie	Patrimonial	1	1	0	0	18%
Capitalisation pers. morale IR	Patrimonial	0	0	0	0	18%
Capitalisation pers. morale IS	Gestion de Fortune	0	0	0	1	16%
Assurance-Vie	Aisé	1	1	0	0	15%
Capitalisation pers. morale IS	Patrimonial	1	0	0	1	14%
Capitalisation pers. morale IS	Gestion de Fortune	1	0	0	0	14%
Capitalisation pers. morale IR	Patrimonial	0	0	0	1	13%
Capitalisation pers. morale IS	Aisé	0	0	0	0	13%
Assurance-Vie	Patrimonial	0	0	0	1	13%
Capitalisation pers. morale IS	Aisé	0	0	0	1	12%
Capitalisation pers. morale IR	Gestion de Fortune	0	0	0	1	11%
Capitalisation pers. morale IR	Gestion de Fortune	1	0	0	0	11%
Assurance-Vie	Gestion de Fortune	0	0	0	1	11%
Capitalisation pers. morale IS	Aisé	1	1	0	0	11%
Assurance-Vie	Aisé	1	0	1	0	11%
Capitalisation pers. morale IR	Patrimonial	1	0	0	1	11%
Capitalisation pers. morale IS	Aisé	1	0	0	1	11%
Capitalisation pers. morale IS	Gestion de Fortune	0	0	0	0	11%
Assurance-Vie	Aisé	0	0	0	1	11%
Capitalisation pers. morale IS	Gestion de Fortune	1	0	0	1	10%
Assurance-Vie	Patrimonial	1	0	0	1	9%
Assurance-Vie	Aisé	1	1	1	0	9%
Capitalisation pers. morale IR	Gestion de Fortune	1	0	0	1	8%
Capitalisation pers. morale IR	Aisé	0	0	0	1	8%
Assurance-Vie	Gestion de Fortune	1	0	0	1	8%

Capitalisation pers. morale IR	Aisé	1	0	0	1	8%
Assurance-Vie	Aisé	0	0	1	0	8%
Assurance-Vie	Aisé	1	0	0	1	8%
Assurance-Vie	Patrimonial	1	1	0	1	7%
Capitalisation pers. morale IR	Aisé	1	1	0	0	7%
Assurance-Vie	Aisé	0	0	1	1	7%
Capitalisation pers. morale IS	Aisé	1	0	1	1	6%
Capitalisation pers. morale IR	Patrimonial	1	1	0	1	6%
Assurance-Vie	Aisé	1	1	0	1	6%
Capitalisation pers. morale IS	Aisé	0	0	1	1	5%
Assurance-Vie	Gestion de Fortune	1	1	0	1	5%
Assurance-Vie	Aisé	1	0	1	1	5%
Capitalisation pers. morale IS	Aisé	0	0	1	0	5%
Capitalisation pers. morale IS	Patrimonial	1	1	0	0	4%
Capitalisation pers. morale IS	Aisé	1	1	0	1	4%
Capitalisation pers. morale IR	Aisé	1	1	0	1	4%
Capitalisation pers. morale IS	Patrimonial	1	1	0	1	4%
Capitalisation pers. morale IR	Aisé	0	0	1	1	4%
Assurance-Vie	Aisé	1	1	1	1	3%
Capitalisation pers. morale IR	Aisé	1	0	1	1	3%
Capitalisation pers. morale IR	Aisé	1	1	1	1	3%
Capitalisation pers. morale IR	Patrimonial	1	1	0	0	3%
Capitalisation pers. morale IS	Gestion de Fortune	1	1	0	1	2%
Capitalisation pers. morale IR	Aisé	1	0	1	0	2%
Assurance-Vie	Gestion de Fortune	1	1	0	0	2%
Capitalisation pers. morale IR	Gestion de Fortune	1	1	0	1	1%
Capitalisation pers. morale IS	Aisé	1	1	1	1	1%

Tableau 47 : Probabilité de rachat total sur les différentes typologies

Sur la base de la totalité des observations, la probabilité de rachat au cours des 7 premières années est de 23,5%. Dans ce cas, il apparaît que 9 typologies de populations présentent un taux de rachat supérieur à celui du portefeuille global. Aucune ne présente un investissement sur une offre commerciale particulière (sur le fonds en euros ou sur les UC).

Pour 4 de ces classifications, les contrats sont investis à 100% sur le fonds en euros.

Ainsi, les opérations commerciales :

- incitant l'investissement sur les UC (selon le profil de risque du client) via également des catégories différentes d'un OPC classique,
- proposant une revalorisation supplémentaire sur le fonds en euros ;

permettent réellement une conservation plus longue du contrat.

En revanche, malgré l'engagement à caractère déontologique sur la souscription des personnes morales soumises à l'IS, cette catégorie présente un risque de rachat plus important. Il est donc opportun de se pencher sur des leviers supplémentaires de dissuasion du rachat (des pénalités de

rachats plus importantes par exemple) ou de limiter la souscription de ces contrats (ces contrats étant proposés uniquement à versement unique).

Pour les contrats d'assurance-vie, outre le fait d'être plus sélectif à l'entrée, nous pourrions proposer d'encadrer les versements complémentaires et de mettre en place des pénalités de rachat (dans la limite des maxima réglementaires).

Conclusion

C'est notre volonté d'approfondir la connaissance du passif afin d'appréhender au mieux la politique d'investissement de l'entreprise qui a déclenché cette étude. Dans ce contexte, cette étude avait pour objectif d'améliorer la connaissance du comportement des assurés en analysant différents facteurs sur l'évènement de rachat total et en particulier lors d'un rachat pour des raisons structurelles.

Les différents modèles étudiés ont permis de dégager certaines variables communes qui ont un impact sur la décision de rachat total par l'assuré. Notre étude a également permis de valider certaines « intuitions » que nous avons sur le comportement du portefeuille, notamment celles-ci :

- malgré l'engagement à caractère déontologique, les contrats souscrits par des personnes morales soumises à l'IS,
- les contrats dont le partenariat commercial a cessé,
- les contrats investis uniquement sur le fonds en euros,

sont plus rachetés que les autres contrats ne présentant pas ces caractéristiques.

Ces comportements sont également constatés sur le marché global de l'assurance-vie, le taux de rachat constaté en 2017 étant de 5%³⁶ des contrats en euros (vs. 3% des contrats en unités de compte) et de 8% pour les contrats de capitalisation³⁷.

A contrario, le fait de présenter :

- un investissement sur des unités de comptes,
- ou d'avoir présenté un versement lié à une offre TMA/Bonus,
- toutes offres particulières en général.

permet de diminuer la probabilité de rachat.

En ce qui concerne les modèles statistiques, autant l'algorithme de forêts aléatoires n'a pas dégagé de résultats satisfaisants lors du *back-testing*, autant le modèle de régression logistique a, quant à lui, fourni des résultats statistiquement significatifs, exploitables et performants. Le modèle définitif est basé sur les variables discriminantes suivantes :

Méthodologie	Régression logistique
Variables	
Cadre fiscal	Oui
Réseaux de distribution	Oui
Présence de fonds en euros	Non
Présence d'unités de compte	Oui
Offre commerciale TMA et/ou Bonus	Oui
fonds Euroconviction	Oui
Supports immobiliers	Oui
UC Mandat	Non
Résidence	Non
Ancienneté du contrat	Oui

³⁶ Sources ACPR : Analyse et Synthèses n°98, La rentabilité technique des organismes d'assurance-vie et mixtes établis en France

³⁷ Taux déterminés par un montant de prestations sur encours du portefeuille : le niveau des taux ne sont pas totalement comparables au taux de rachat total de notre étude déterminé sur un nombre de contrats rachetés.

Actuellement, dans l'entreprise, les lois de comportement sont calculées en fonction du réseau de distribution et de l'ancienneté des contrats. Au vu des résultats de l'étude, la précision de ces lois pourrait être améliorée en utilisant d'autres paramètres propres aux contrats.

De plus, via cette approche, nous avons réalisé un outil de pilotage pouvant être utilisé lors du processus d'acceptation des dossiers de souscription (et donc de l'acceptation du risque probable). En fonction du « score » calculé par l'outil de pilotage, la réponse à l'appel d'offre pourrait imposer la contrainte d'investir un montant minimum sur les supports en unités de compte et/ou un montant maximum sur le fonds en euros ; elle pourrait aussi contenir une clause instaurant des pénalités de sortie.

Comme indiqué plus haut, l'axe de lecture de ce mémoire est focalisé sur les rachats pour raisons structurelles. Cependant, il faut faire preuve de bon sens : il est impossible de complètement dissocier les rachats structurels des rachats conjoncturels car les rachats même structurels ne peuvent pas être complètement indépendants de l'environnement économique. Par exemple, l'environnement de taux bas dans lequel l'étude a été menée a pour conséquence de diminuer le rendement des fonds en euros des contrats d'assurance-vie rendant de ce fait moins attractif ces produits par rapport à d'autres investissements. Lors d'un besoin d'argent, les souscripteurs pourraient alors être plus enclins à racheter leurs contrats d'assurance-vie plutôt que clôturer un autre produit. Ces typologies de rachats restent dans le domaine « structurel », car liés à un besoin propre de l'assuré mais sont également déterminés du fait de l'environnement économique. Ces éléments n'ont pas pu être quantifiés dans cette étude du fait de manque de données sur les motifs de rachats de nos données.

Afin de conclure définitivement cette étude, évoquons le fait que les résultats de cette étude pourraient être amenés à évoluer :

- Concernant les contrats de capitalisation souscrits par des personnes morales soumises à l'impôt sur les sociétés : l'engagement à caractère déontologique ayant été mis en place sur le portefeuille à partir de 2011 et principalement en 2012, les contrats soumis à cette contrainte atteignent tout juste leur quatrième année de souscription. Même si les données nous montrent une inflexion des rachats sur cette catégorie de souscripteurs, les données ne permettent pas un recul suffisant pour savoir si le comportement sur ces contrats après 4 ans de souscription, consiste à racheter massivement ou si, au contraire, l'engagement a permis une fidélisation plus longue car le contrat de capitalisation n'est plus utilisé comme placement de liquidités par les entreprises. Il serait profitable, dans les années à venir, d'étudier plus précisément cette catégorie du portefeuille.
- Concernant la « *flat tax* » : la nouvelle fiscalité appliquée à partir du 01/01/2018 risque de fortement modifier le comportement des assurés en termes de rachat à moyen et long terme. En effet, l'assurance-vie perd son attractivité fiscale qui poussait les souscripteurs les plus aisés à rester longtemps investi sur leur contrat.

Bibliographie

- FREY Laure : Eclairages de l'enquête Patrimoine sur les comportements de rachat en assurance-vie. Analyses et Synthèses n°59 de l'ACPR
- <http://www.lesechos.fr/finance-marches/banque-assurances/0211329190634-assurance-vie-ce-que-la-loi-sapin-ii-pourrait-changer-pour-les-epargnants-2030388.php>
- ACPR : Les chiffres du marché français de la banque et de l'assurance 2016
- ACLI : Challenges of the life insurance industry in Europe in a low interest rate environment, Refocus Conference 2013, Session 1C, 4 March
- INSEE : Revenu salarial net annuel moyen par sexe et PCS³⁸ en 2013 : Comparaison régionale
- FFA : <https://www.ffa-assurance.fr/content/tableau-de-bord-de-assurance-en-2016>
- FFA : [Recueil des engagements à caractère déontologique des entreprises d'assurance membres de la FFA](https://www.ffa-assurance.fr/la-federation/publications/deontologie/recueil-des-engagements-caractere-deontologique-des) : <https://www.ffa-assurance.fr/la-federation/publications/deontologie/recueil-des-engagements-caractere-deontologique-des>
- AHADO Frédéric et ARIAS Liliana : La rentabilité technique des organismes d'assurance-vie et mixtes établis en France. Analyses et Synthèses n°98 de l'ACPR
- BREIMAN Leo : Random forests, Machine Learning 45 [2001],5-32
- BREIMAN Leo : Bagging predictors, Machine Learning 26 [1996], n°2, 123-140.
- MOOC : Introduction à la statistique avec R de l'université UPSUD

Ouvrages

- RAKOTOMALALA Ricco : Pratique de la régression logistique
- TUFFERY Stéphane : Modélisation prédictive et apprentissage statistique avec R
- TUFFERY Stéphane: Data Mining and Statistics for Decision Making
- DROESBEKE Jean-Jacques, LEJEUNE Michel, SAPORTA Gilbert : Modèles statistiques pour données qualitatives
- BREIMAN, FRIEDMAN, OLSHEN et STONE: Classification And Regression Trees.

Mémoires d'actuariat

- BERNE Ophélie : Estimation des lois de rachat et de mortalité en assurance-vie
- HAJ TAIEB Olfia et NIEBEL Caroline : Rachats totaux dans le cadre de contrats d'épargne d'un portefeuille Haut de Gamme
- JAMAL Salma : Construction du taux de rachat structurel en Epargne
- LADIAS Nicolas : Analyse des causes de rachats sur des contrats d'assurance-vie
- MILHAUD Xavier : Segmentation et modélisation des comportements de rachats en Assurance Vie
- RAKAH Naoufal : Modélisations des rachats dans les contrats d'épargne

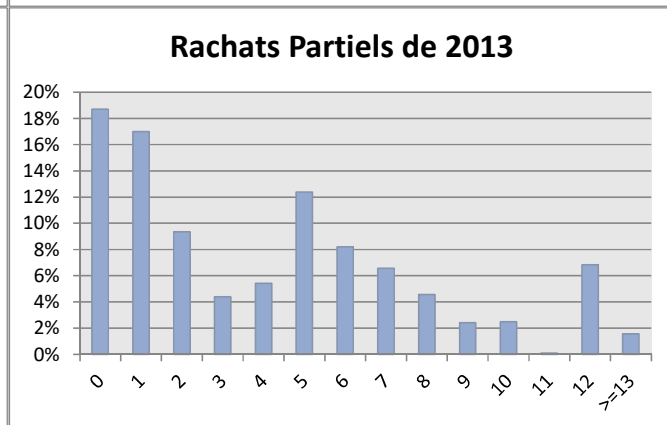
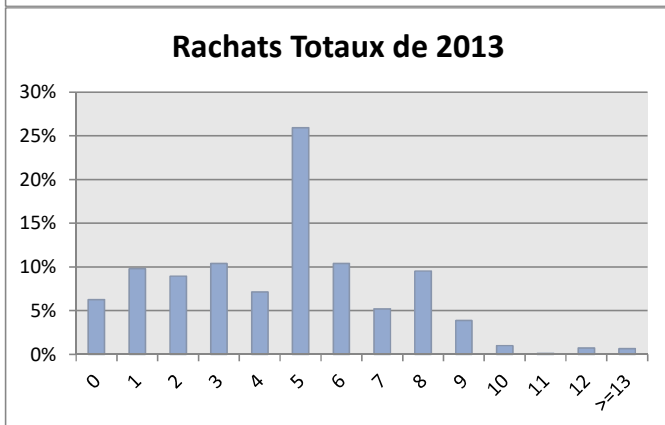
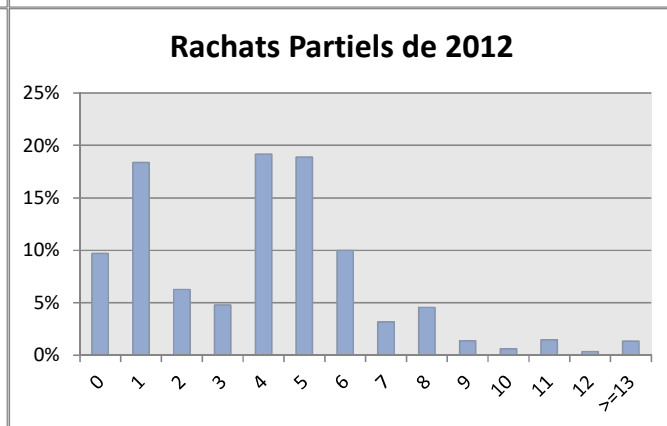
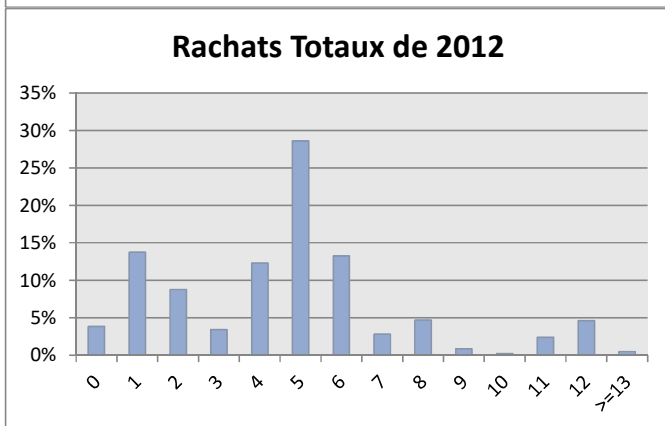
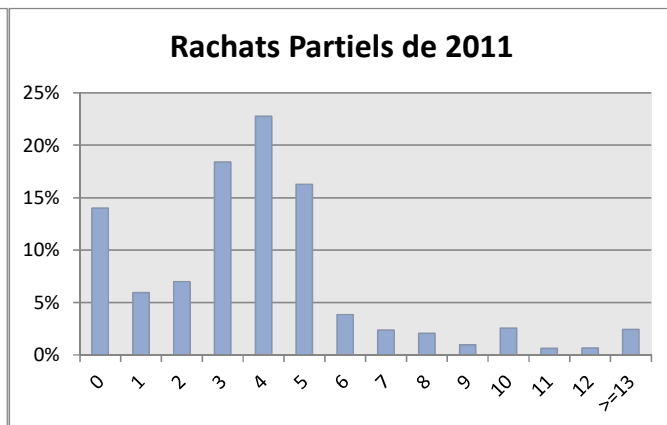
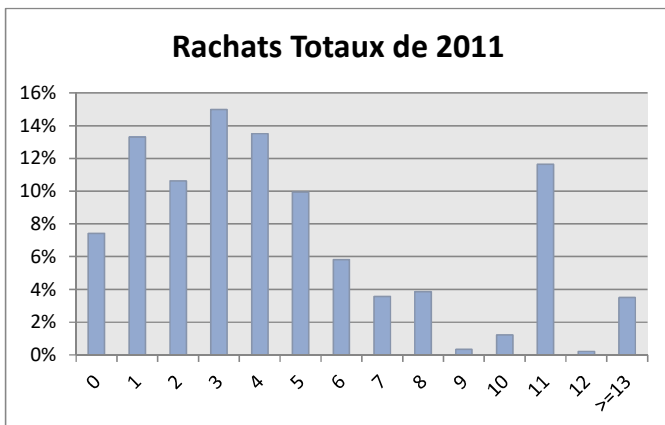
³⁸ Professions et catégories socio-professionnelles

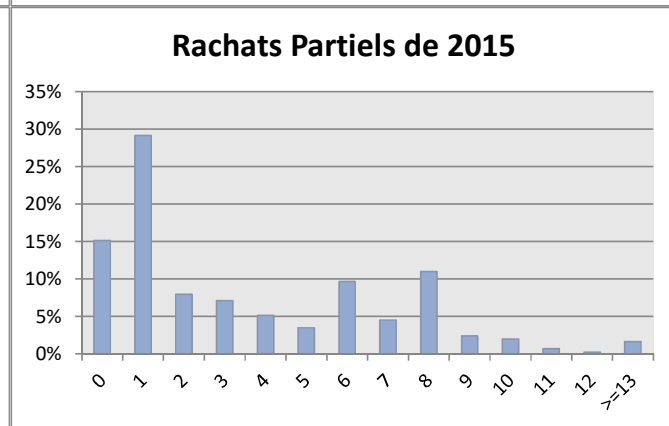
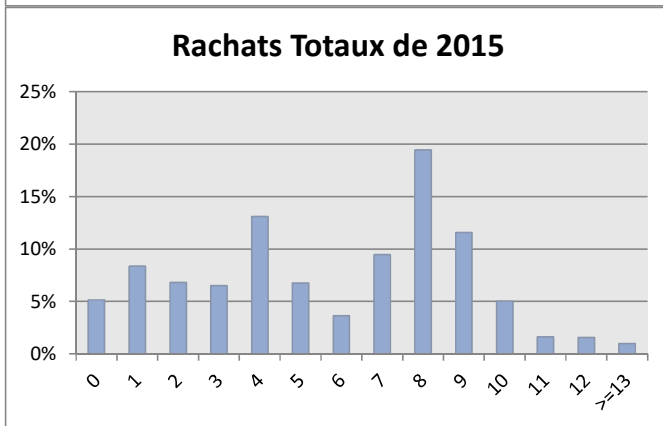
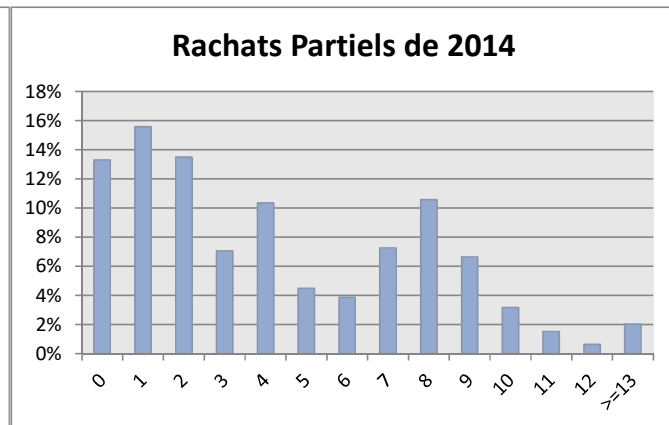
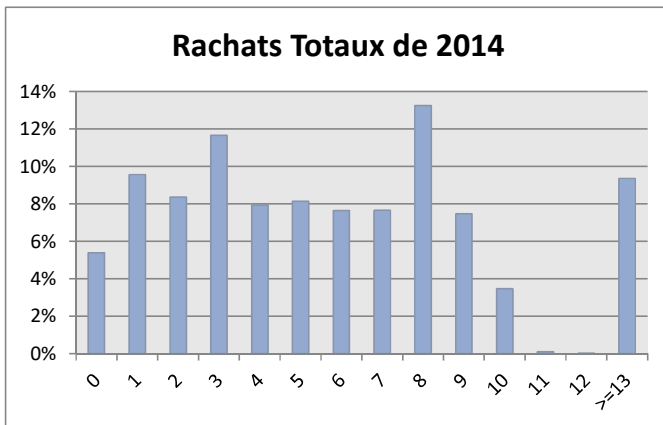
Annexes

A. Lexique de notation

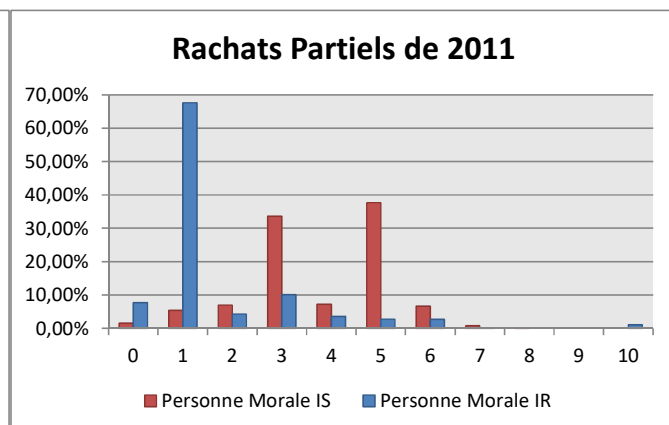
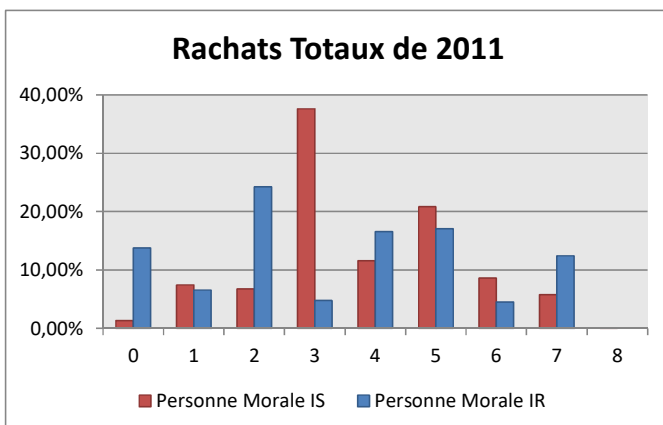
Notation	Définition
TMG	Taux minimum garanti sur le support déterminé lors du versement
PM	Montant de la provision mathématique calculée sur un contrat lors d'un versement sur le support euro croissance
PM_d	Montant de la provision mathématique calculé à la date d
$PM_d(0)$	Montant de la provision mathématique calculé à la date d au taux d'actualisation de 0%
PD	Montant de la provision de diversification calculé sur un contrat lors d'un versement sur le support euro croissance
PD_d	Montant de la provision de diversification calculé à la date d
$NbPD$	Nombre de parts de provision de diversification
$NbPD_d$	Nombre de parts de provision de diversification calculé à la date d
VL_d	Valeur liquidative de la part de provision de diversification en vigueur à la date d
VL_{min}	Valeur liquidative minimale de la part de provision de diversification
Analyse des Correspondances Multiples	
I	Nombre d'individus i étudiés lors d'une Analyse des Correspondances Multiples
J	Nombre de variables qualitatives j
K_j	Nombre et ensemble des modalités de la variable j
K	Nombre et ensemble des modalités toutes variables confondues
$d^2(i, l)$	Distance entre deux individus i et l
G_I	Centre de gravité du nuage des individus
F_s	Ensemble des projections de tous les points du nuage d'individus sur le s^e axe factoriel u_s
G_s	Ensemble des projections des variables sur le s^e axe factoriel v_s
$CTR_s(k)$	Contribution d'une modalité k
Méthode CART	
$i(t)$	Indice de Gini, indice utilisé dans la règle de division de l'échantillon
Régression logistique	
ω	Individu de l'échantillon d'observation
$p(\omega)$	Probabilité, à priori, pour un individu d'effectuer le rachat total de son contrat dans l'année d'observation
$\pi(\omega)$	Probabilité, à posteriori, d'un individu ω d'effectuer le rachat total de son contrat dans l'année d'observation
L	Vraisemblance d'un échantillon Ω
LL	Log-vraisemblance d'un échantillon Ω
RR	Risque relatif
OR	Odds-ratio
Estimation Kaplan Meier	
$S(t)$	Fonction de survie

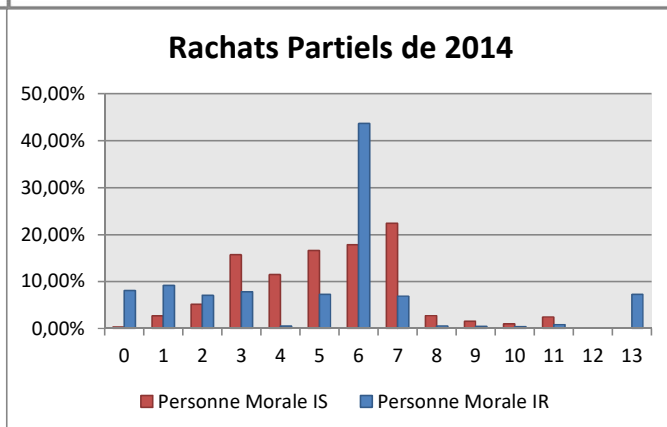
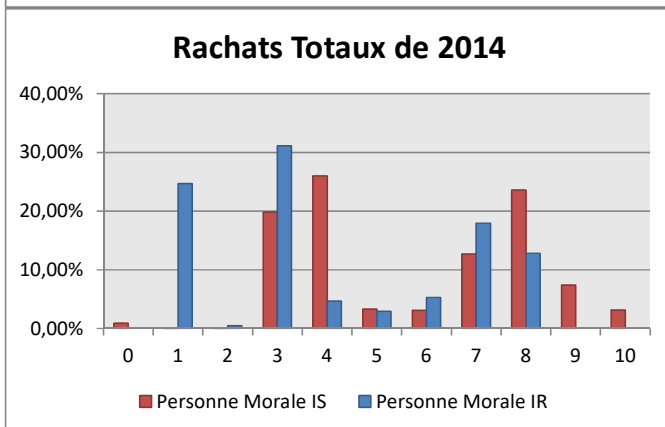
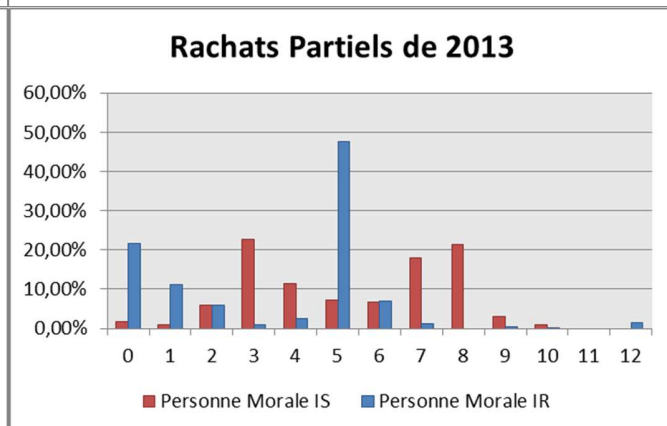
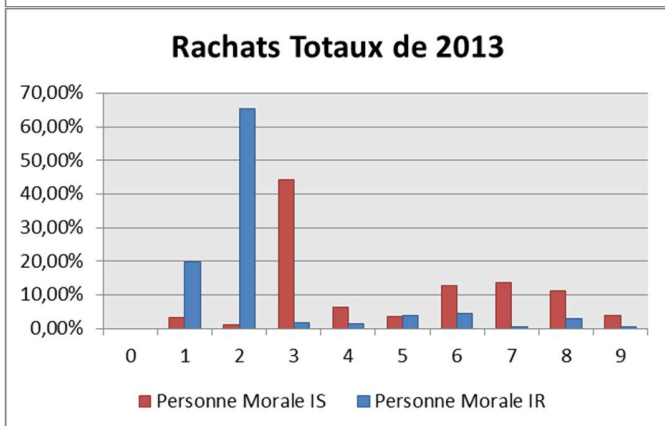
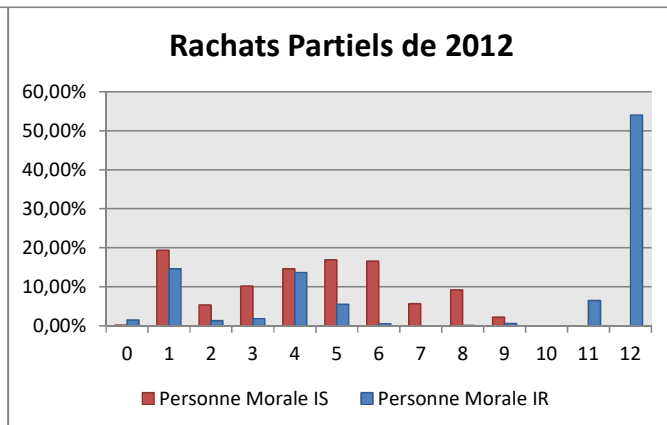
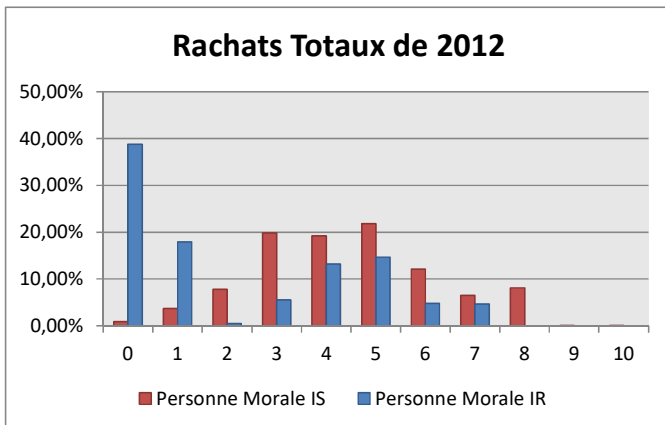
B. Statistiques sur les Rachats Totaux et Rachats Partiels des personnes physiques de 2011 à 2015

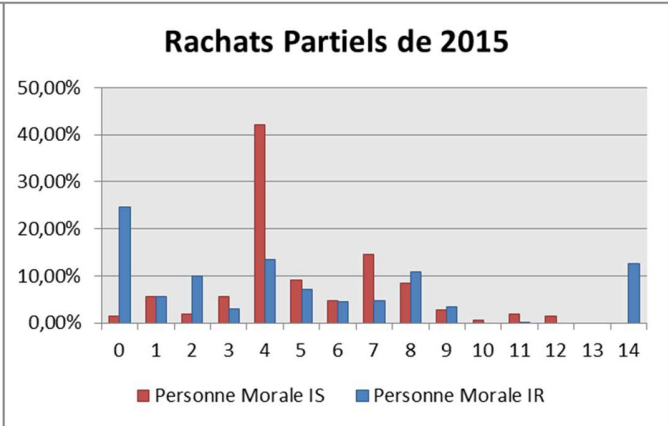
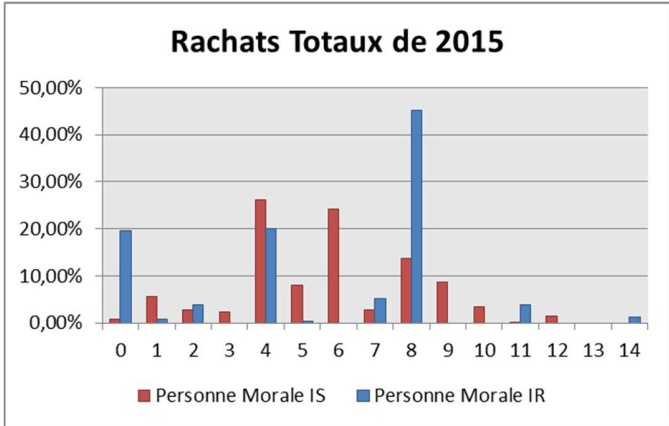




C. Statistiques sur les Rachats Totaux et Rachats Partiels des personnes morales de 2011 à 2015







D. Valeurs propres de l'ACM sur le modèle des données 2011 à 2015 agrégée :

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.20144593	13.946257	13.94626
dim 2	0.18334915	12.693403	26.63966
dim 3	0.14505636	10.042363	36.68202
dim 4	0.12411891	8.592848	45.27487
dim 5	0.11497820	7.960029	53.23490
dim 6	0.11165486	7.729952	60.96485
dim 7	0.11016504	7.626810	68.59166
dim 8	0.10250750	7.096673	75.68834
dim 9	0.09634877	6.670300	82.35864
dim 10	0.08490281	5.877887	88.23652
dim 11	0.06642450	4.598619	92.83514
dim 12	0.05648290	3.910354	96.74550
dim 13	0.04700951	3.254504	100.00000

E. Résultats année par année de l'ACM :

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.20147324	13.948147	13.94815
dim 2	0.17807616	12.328349	26.27650
dim 3	0.14277361	9.884327	36.16082
dim 4	0.12187154	8.437260	44.59808
dim 5	0.11376212	7.875839	52.47392
dim 6	0.11176969	7.737901	60.21182
dim 7	0.10999959	7.615357	67.82718
dim 8	0.10329305	7.151057	74.97824
dim 9	0.09980746	6.909747	81.88798
dim 10	0.09129313	6.320294	88.20828
dim 11	0.06842707	4.737259	92.94554
dim 12	0.06280389	4.347962	97.29350
dim 13	0.03909390	2.706501	100.00000

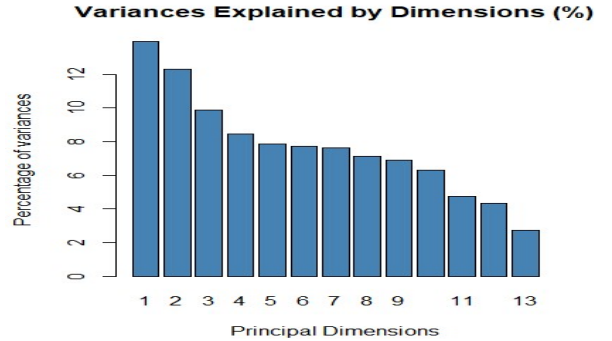


Figure 30: Valeurs propres 2011

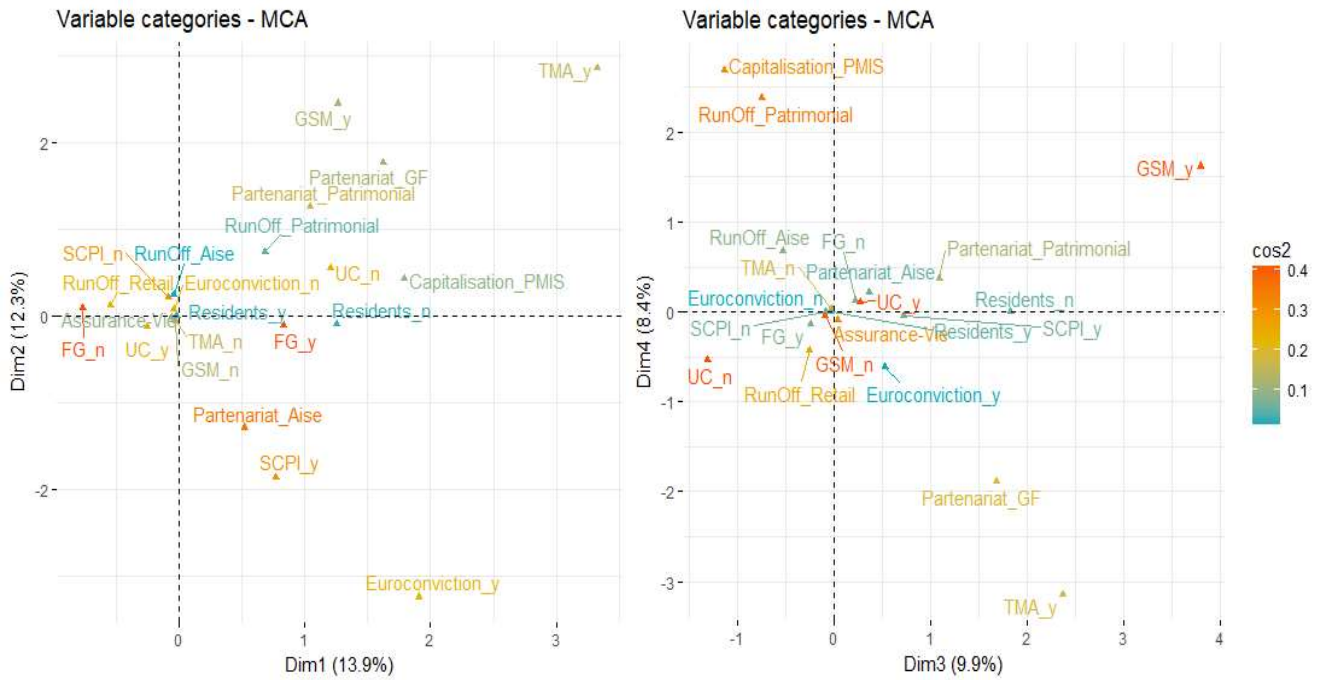


Figure 31: 4 premières dimensions de l'ACM 2011

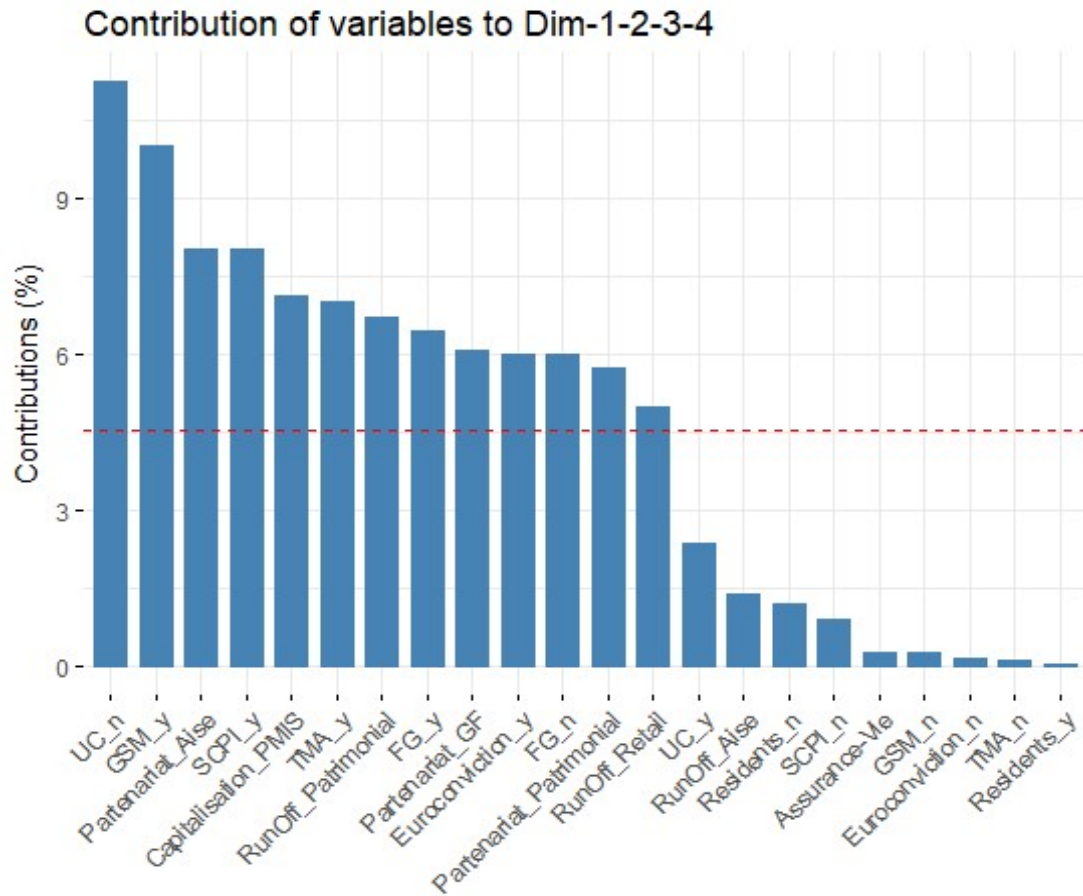


Figure 32: Contributions des variables aux 4 premières dimensions de l'ACM 2011

	eigenvalue	percentage of variance	cumulative percentage of variance	
dim 1	0.20483374	14.180798	14.18080	
dim 2	0.18556161	12.846573	27.02737	
dim 3	0.14218171	9.843349	36.87072	
dim 4	0.12453390	8.621578	45.49230	
dim 5	0.11400150	7.892412	53.38471	
dim 6	0.11227973	7.773212	61.15792	
dim 7	0.10828043	7.496337	68.65426	
dim 8	0.10243052	7.091344	75.74560	
dim 9	0.09600414	6.646440	82.39204	
dim 10	0.08827040	6.111028	88.50307	
dim 11	0.06615356	4.579862	93.08293	
dim 12	0.05730302	3.967132	97.05006	
dim 13	0.04261017	2.949935	100.00000	

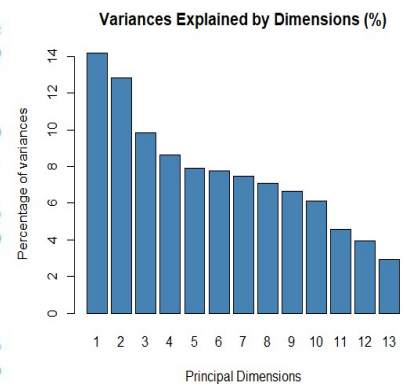


Figure 33 : Valeurs propres 2012

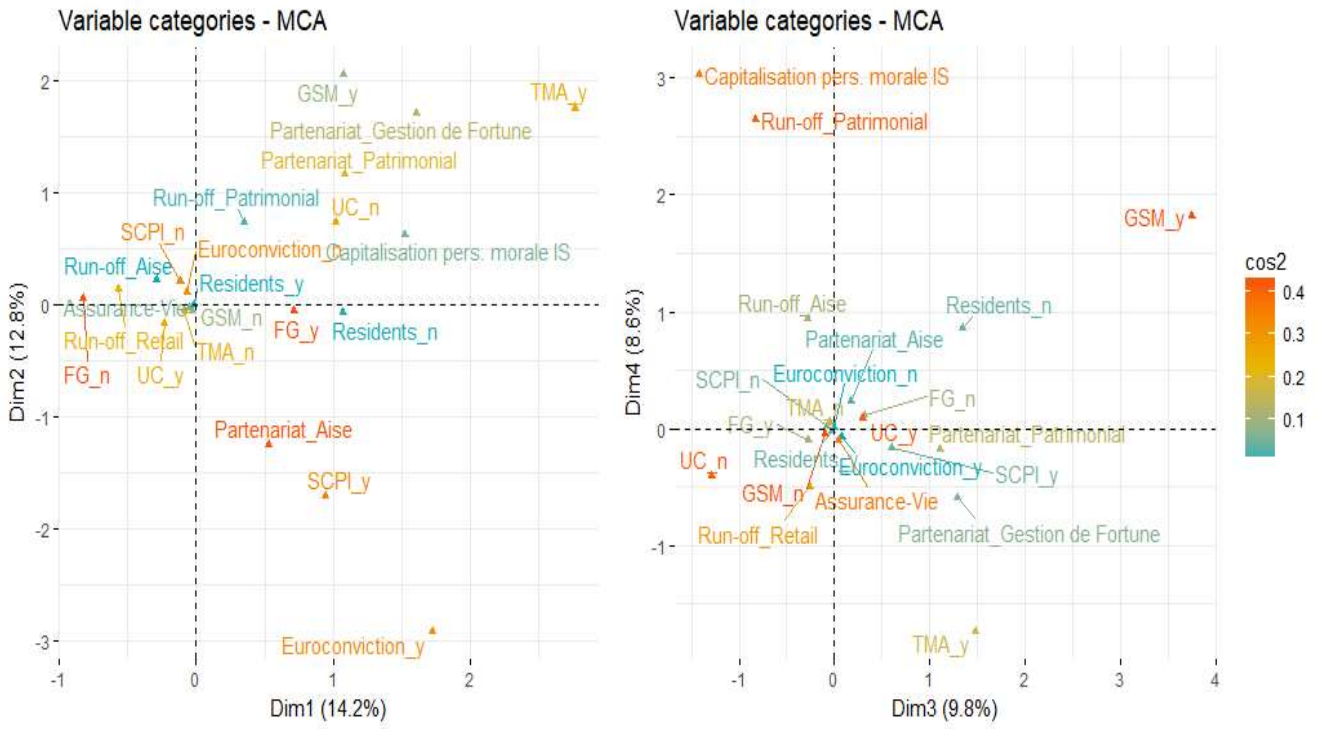


Figure 34 : 4 premières dimensions de l'ACM 2012

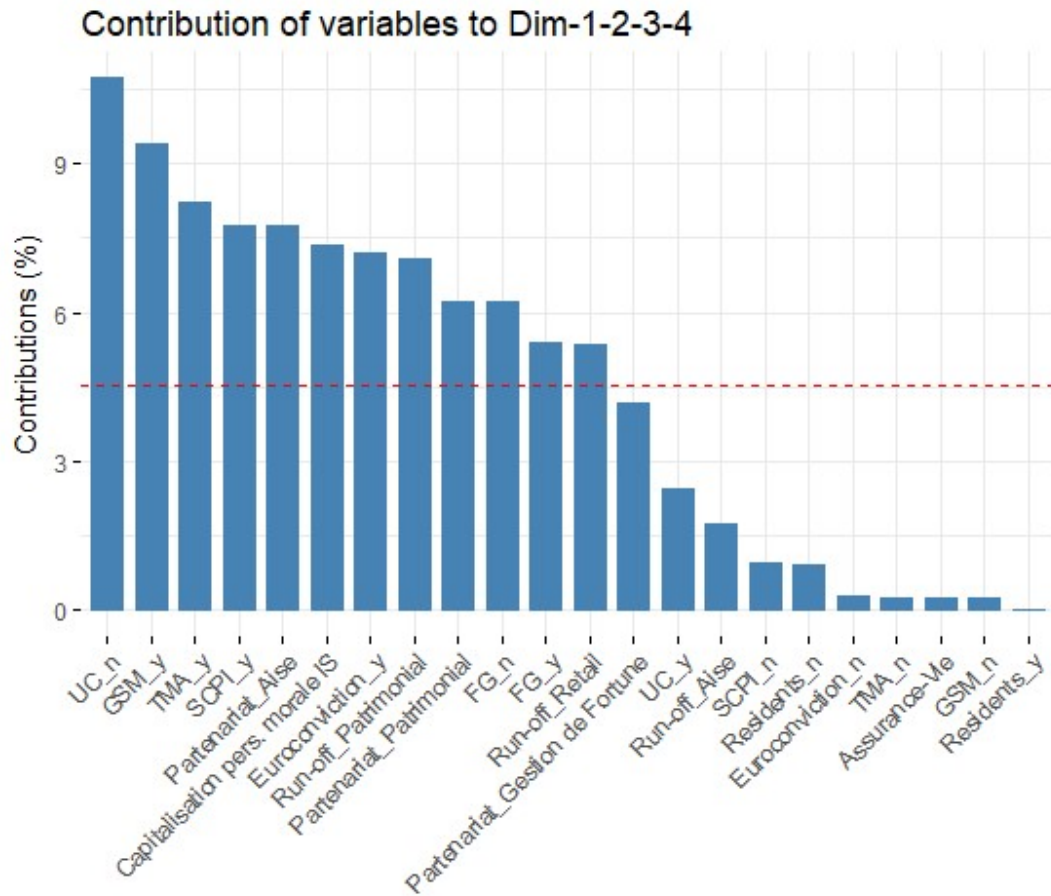


Figure 35 : Contribution des variables aux 4 premières dimensions de l'ACM 2012

	eigenvalue	percentage of variance	cumulative percentage of variance
dim 1	0.20606793	14.266242	14.26624
dim 2	0.18518736	12.820664	27.08691
dim 3	0.14485178	10.028200	37.11511
dim 4	0.12565811	8.699407	45.81451
dim 5	0.11699425	8.099602	53.91411
dim 6	0.11164176	7.729045	61.64316
dim 7	0.10946995	7.578689	69.22185
dim 8	0.10101151	6.993104	76.21495
dim 9	0.09506833	6.581654	82.79661
dim 10	0.08446037	5.847257	88.64386
dim 11	0.06639790	4.596778	93.24064
dim 12	0.05252206	3.636142	96.87678
dim 13	0.04511312	3.123216	100.00000

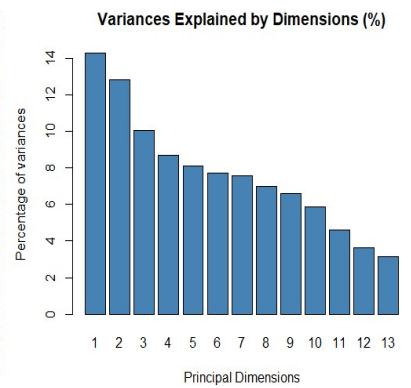


Figure 36 : Valeurs propres 2013

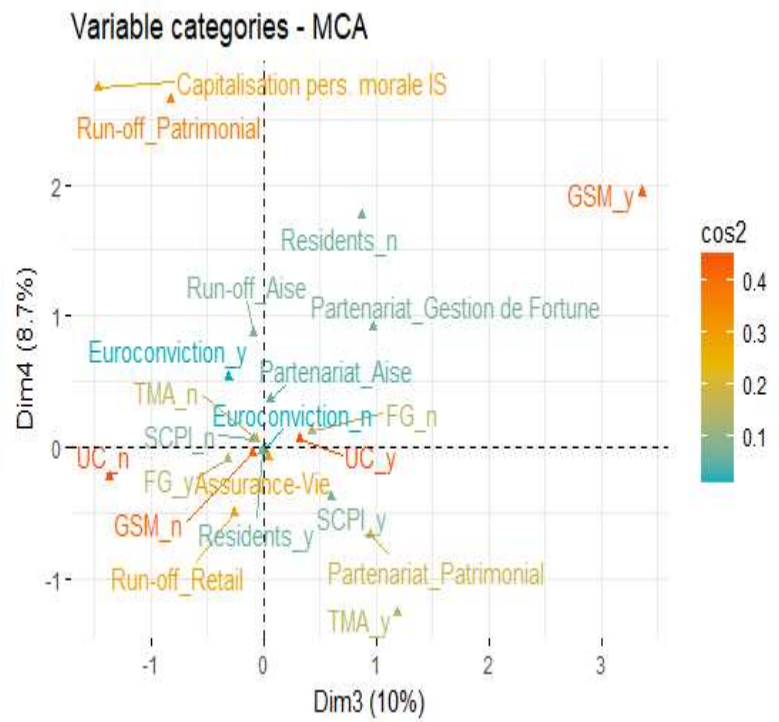
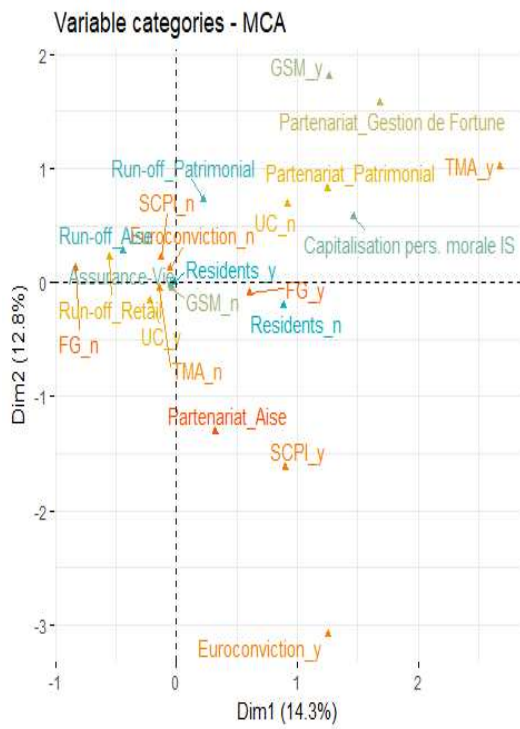


Figure 37 : 4 premières dimensions de l'ACM 2013

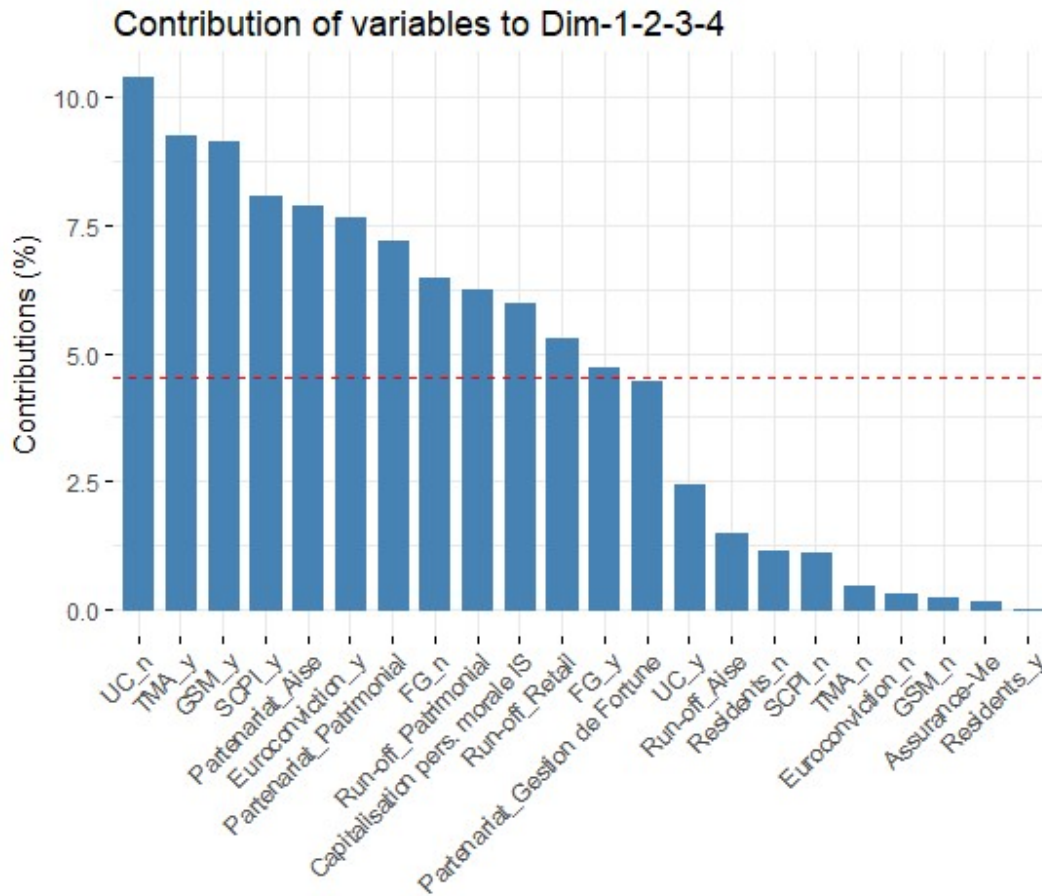


Figure 38 : Contribution des variables aux 4 premières dimensions de l'ACM 2013

	eigenvalue	percentage of variance	cumulative percentage of variance	variance
dim 1	0.20228812	14.004562	14.00456	14.00456
dim 2	0.18484893	12.797233	26.80180	26.80180
dim 3	0.14598384	10.106574	36.90837	36.90837
dim 4	0.12560345	8.695623	45.60399	45.60399
dim 5	0.11634129	8.054397	53.65839	53.65839
dim 6	0.11144046	7.715108	61.37350	61.37350
dim 7	0.11089274	7.677190	69.05069	69.05069
dim 8	0.10155789	7.030931	76.08162	76.08162
dim 9	0.09485130	6.566628	82.64825	82.64825
dim 10	0.08222915	5.692787	88.34103	88.34103
dim 11	0.06616713	4.580801	92.92184	92.92184
dim 12	0.05370756	3.718215	96.64005	96.64005
dim 13	0.04853260	3.359949	100.00000	100.00000

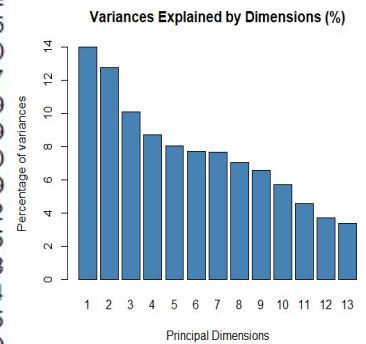


Figure 39 : Valeurs propres 2014

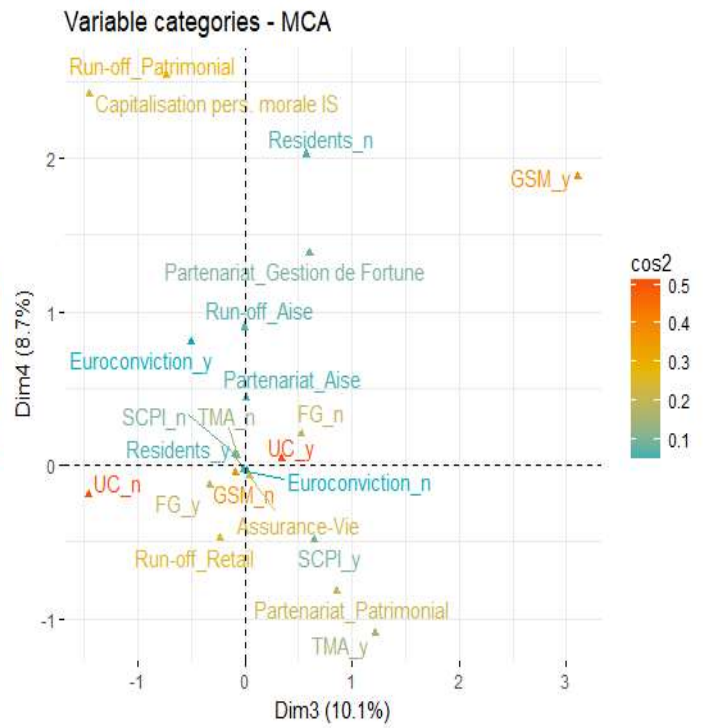
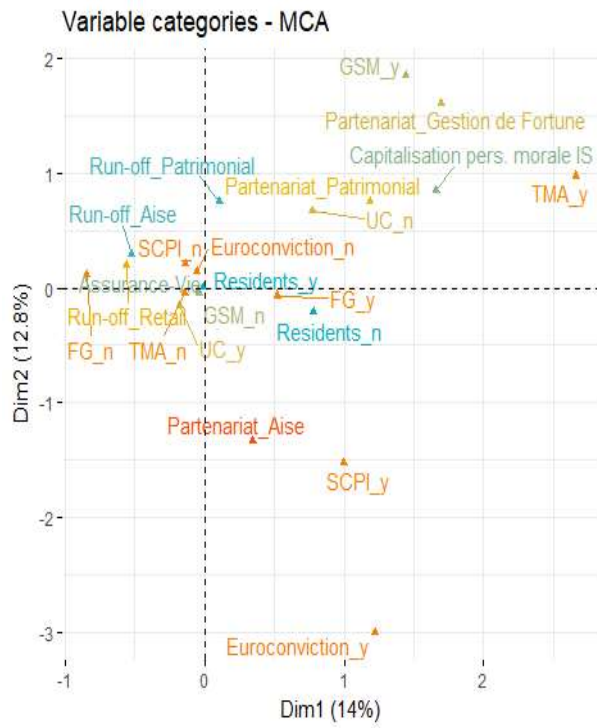


Figure 40 : 4 premières dimensions de l'ACM 2014

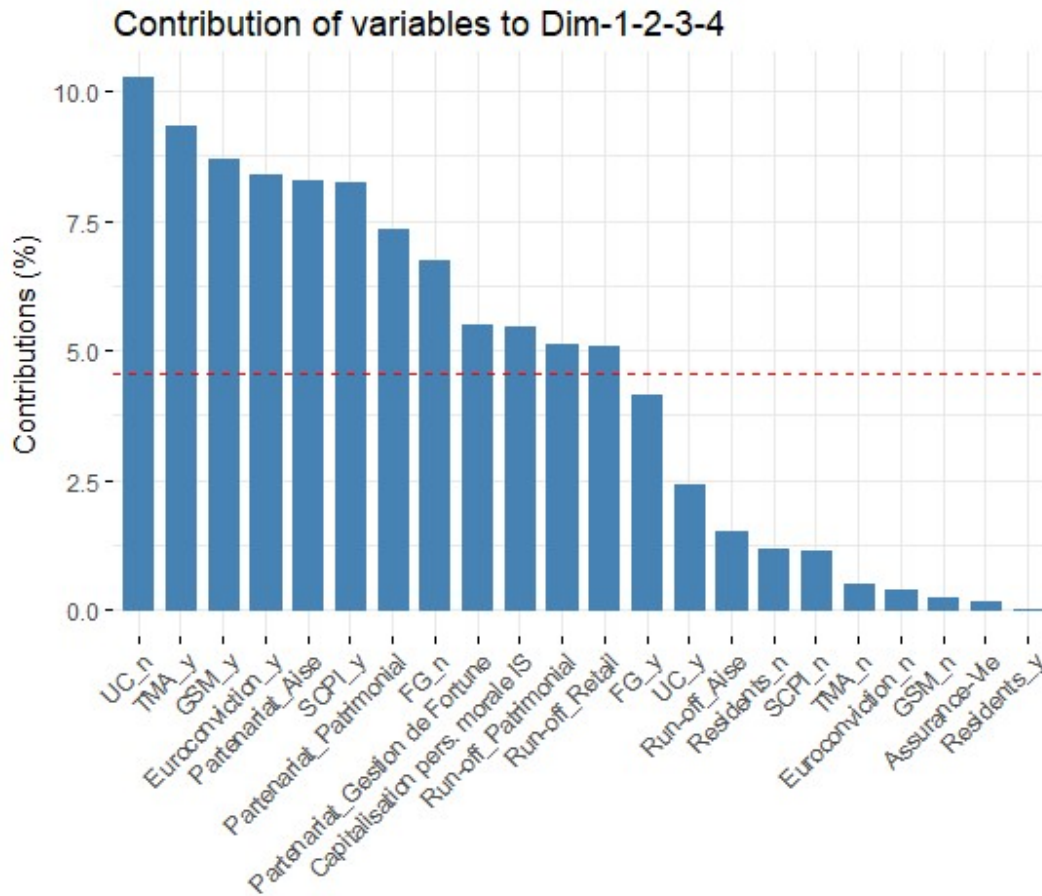


Figure 41 : Contribution des variables aux 4 premières dimensions de l'ACM 2014

	eigenvalue	percentage of variance	cumulative percentage of variance	variance
dim 1	0.19375461	13.413781	13.41378	13.41378
dim 2	0.18502853	12.809667	26.22345	26.22345
dim 3	0.14727049	10.195649	36.41910	36.41910
dim 4	0.12559239	8.694858	45.11396	45.11396
dim 5	0.11518894	7.974619	53.08857	53.08857
dim 6	0.11238898	7.780775	60.86935	60.86935
dim 7	0.11050593	7.650411	68.51976	68.51976
dim 8	0.10179044	7.047031	75.56679	75.56679
dim 9	0.09546375	6.609029	82.17582	82.17582
dim 10	0.08049612	5.572808	87.74863	87.74863
dim 11	0.06721672	4.653465	92.40209	92.40209
dim 12	0.05744256	3.976793	96.37889	96.37889
dim 13	0.05230499	3.621114	100.00000	100.00000

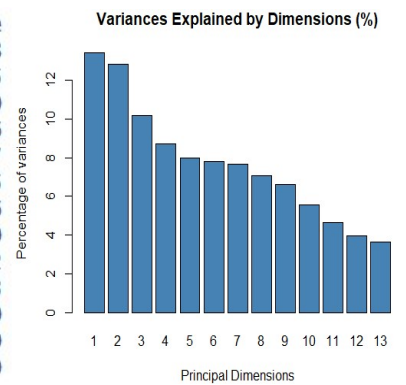


Figure 42 : Valeurs Propres 2015

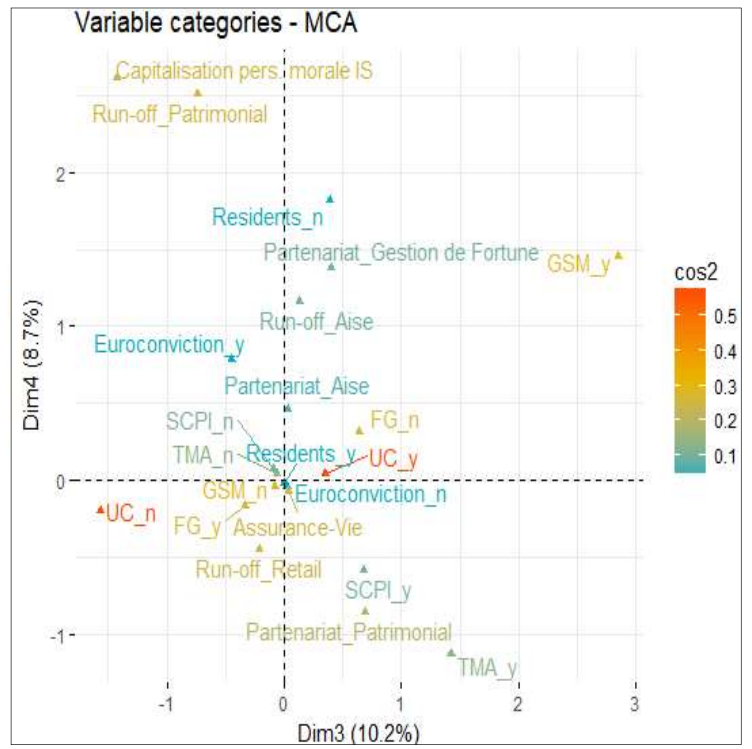
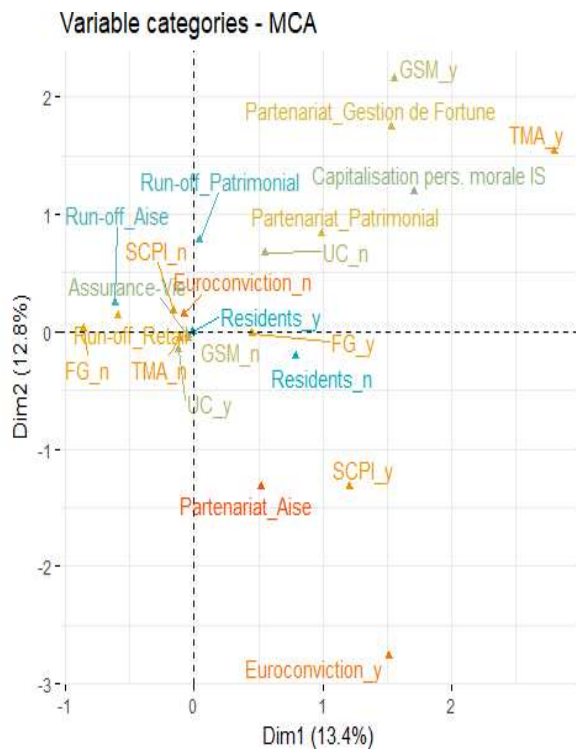


Figure 43 : 4 premières dimensions de l'ACM 2015

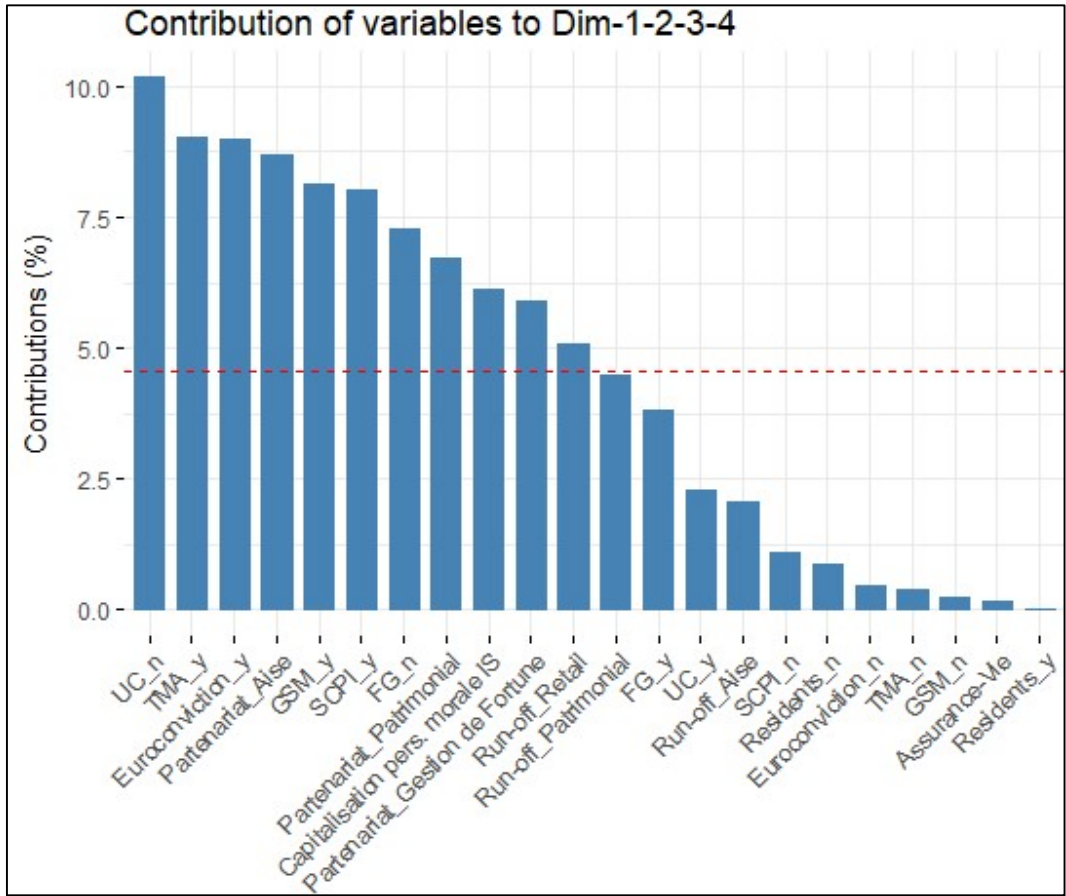


Figure 44 : Contribution des variables aux 4 premières dimensions de l'ACM 2015

F. Résultats année par année de la méthode CART :

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2011, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete           Cadre_fiscal           Classification.Reseau   Classification_Geog
raphique
[5] FondsG                SCPI                   TMA_Bonus              UC
[9] UC.Mandat

Root node error: 3651/51472 = 0.070932

n= 51472

      CP nsplit rel error  xerror  xstd
1  9.1299e-04    0  1.00000  1.00000  0.015952
2  2.7390e-04    6  0.99151  0.99178  0.015891
3  1.3695e-04   15  0.98904  0.99260  0.015897
4  1.0956e-04   19  0.98850  0.99233  0.015895
5  1.0271e-04   24  0.98795  0.99233  0.015895
6  9.1299e-05   34  0.98685  0.99260  0.015897
7  6.8474e-05   52  0.98521  0.99397  0.015908
8  6.0866e-05   56  0.98494  0.99617  0.015924
9  5.4780e-05   65  0.98439  0.99753  0.015934
10 4.5650e-05   75  0.98384  0.99836  0.015940
11 3.4237e-05   81  0.98357  0.99808  0.015938
12 0.0000e+00   89  0.98329  0.99808  0.015938

```

Figure 45 : Arbre 2011

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2011, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete           Classification.Reseau  FondsG                SCPI

Root node error: 3651/51472 = 0.070932

n= 51472

      CP nsplit rel error  xerror  xstd
1 0.00091299    0  1.00000  1.00000  0.015952
2 0.00027500    6  0.99151  0.99178  0.015891

```

Figure 46 : Arbre élagué 2011

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2012, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete           Cadre_fiscal           Classification.Reseau   Classification_Geographique
[5] Euroconviction       FondsG                  SCPI                    TMA_Bonus
[9] UC                    UC.Mandat

Root node error: 3548/53038 = 0.066895

n= 53038

      CP nsplit rel error  xerror  xstd
1  3.7580e-04    0  1.00000  1.00000  0.016217
2  2.8185e-04    9  0.99662  1.00395  0.016247
3  2.1139e-04   16  0.99436  1.00451  0.016251
4  1.8790e-04   20  0.99352  1.00028  0.016219
5  9.3950e-05   23  0.99295  0.99887  0.016209
6  7.6868e-05   29  0.99239  1.00338  0.016243
7  7.0462e-05   40  0.99154  1.00395  0.016247
8  5.6370e-05   44  0.99126  1.00564  0.016259
9  4.6975e-05   78  0.98929  1.00676  0.016268
10 4.3361e-05   100 0.98816  1.00676  0.016268
11 4.0264e-05   119 0.98732  1.00705  0.016270
12 0.0000e+00   126 0.98703  1.00733  0.016272

```

Figure 47 : Arbre 2012

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2012, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete           Cadre_fiscal           Classification.Reseau   Classification_Geographique
[5] FondsG              SCPI                  TMA_Bonus              UC

Root node error: 3548/53038 = 0.066895

n= 53038

      CP nsplit rel error  xerror  xstd
1 0.00037580    0  1.00000 1.00000 0.016217
2 0.00028185    9  0.99662 1.00395 0.016247
3 0.00021139   16  0.99436 1.00451 0.016251
4 0.00018790   20  0.99352 1.00028 0.016219
5 0.00015000   23  0.99295 0.99887 0.016209

```

Figure 48 : Arbre élagué 2012

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2013, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete           Cadre_fiscal           Classification.Reseau   Classification_Geographique
[5] EuroConviction       FondsG                 SCPI                   TMA_Bonus
[9] UC                   UC.Mandat

Root node error: 3031/55188 = 0.054921

n= 55188

      CP nsplit rel error  xerror  xstd
1 4.0251e-03    0  1.00000 1.00000 0.017658
2 3.2992e-04    5  0.97987 0.97987 0.017490
3 2.4744e-04    9  0.97855 0.98086 0.017498
4 1.6496e-04   13  0.97757 0.97921 0.017484
5 1.3197e-04   26  0.97493 0.98053 0.017495
6 1.0997e-04   36  0.97229 0.97987 0.017490
7 8.2481e-05   42  0.97163 0.97987 0.017490
8 6.5985e-05   54  0.97064 0.98119 0.017501
9 5.4987e-05   74  0.96932 0.98086 0.017498
10 3.6658e-05   86  0.96866 0.98152 0.017503
11 3.2992e-05   95  0.96833 0.98185 0.017506
12 0.0000e+00  105  0.96800 0.98185 0.017506

```

Figure 49 : Arbre 2013

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2013, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete           Classification.Reseau   Classification_Geographique FondsG
[5] SCPI                 UC

Root node error: 3031/55188 = 0.054921

n= 55188

      CP nsplit rel error  xerror  xstd
1 0.00402507    0  1.00000 1.00000 0.017658
2 0.00032992    5  0.97987 0.97987 0.017490
3 0.00024744    9  0.97855 0.98086 0.017498
4 0.00020000   13  0.97757 0.97921 0.017484

```

Figure 50 : Arbre élagué 2013

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2014, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete          Cadre_fiscal          Classification.Reseau  Classification_Geographique
[5] EuroConviction      FondsG                SCPI                  TMA_Bonus
[9] UC                  UC.Mandat

Root node error: 3014/58683 = 0.051361

n= 58683

      CP nsplit rel error  xerror   xstd
1 1.0119e-02    0  1.00000 1.00000 0.017741
2 6.6357e-04    4  0.95952 0.95952 0.017397
3 3.3179e-04    5  0.95886 0.95952 0.017397
4 1.2442e-04    6  0.95853 0.95919 0.017394
5 1.2065e-04   40  0.95388 0.96483 0.017443
6 1.1849e-04   52  0.95222 0.96483 0.017443
7 1.1060e-04   67  0.95023 0.96483 0.017443
8 8.2946e-05   73  0.94957 0.96483 0.017443
9 7.1097e-05   93  0.94791 0.96583 0.017451
10 5.5298e-05  107  0.94691 0.96649 0.017457
11 0.0000e+00  113  0.94658 0.96715 0.017463

```

Figure 51 : Arbre 2014

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_2014, method = "class",
      control = parametres)

Variables actually used in tree construction:
[1] Anciennete          Classification.Reseau  Classification_Geographique EuroConviction
[5] UC

Root node error: 3014/58683 = 0.051361

n= 58683

      CP nsplit rel error  xerror   xstd
1 0.01011944    0  1.00000 1.00000 0.017741
2 0.00066357    4  0.95952 0.95952 0.017397
3 0.00033179    5  0.95886 0.95952 0.017397
4 0.00020000    6  0.95853 0.95919 0.017394

```

Figure 52 : Arbre élagué 2014

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_cumulees,
      method = "class", control = parametres)

Variables actually used in tree construction:
[1] Anciennete          Cadre_fiscal          Classification.Reseau  Classification_Geographique
[5] EuroConviction      FondsG                SCPI                  TMA_Bonus
[9] UC                  UC.Mandat

Root node error: 16041/281114 = 0.057062

n= 281114

      CP nsplit rel error  xerror   xstd
1 1.9450e-03    0  1.00000 1.00000 0.0076670
2 5.4548e-04    5  0.99027 0.99027 0.0076319
3 1.8702e-04    9  0.98809 0.98884 0.0076267
4 1.2468e-04   10  0.98791 0.98809 0.0076240
5 3.1170e-05   11  0.98778 0.98797 0.0076235
6 1.7812e-05   13  0.98772 0.98822 0.0076244
7 1.5585e-05   20  0.98759 0.98884 0.0076267
8 1.2468e-05   36  0.98734 0.98890 0.0076269
9 8.9058e-06   67  0.98691 0.98971 0.0076298
10 7.7925e-06   74  0.98685 0.99034 0.0076321
11 6.9267e-06   82  0.98678 0.99046 0.0076326
12 0.0000e+00   91  0.98672 0.99052 0.0076328

```

Figure 53 : Arbre sur les années d'observations cumulées

```

Classification tree:
rpart(formula = Rachat_total ~ Classification_Geographique +
      UC + EuroConviction + Cadre_fiscal + SCPI + TMA_Bonus + Classification.Reseau +
      FondsG + UC.Mandat + Anciennete, data = donnees_cumulees,
      method = "class", control = parametres)

Variables actually used in tree construction:
[1] Anciennete          Cadre_fiscal          Classification.Reseau  Classification_Geographique
[5] FondsG              SCPI                  UC

Root node error: 16041/281114 = 0.057062

n= 281114

   CP nsplit rel error  xerror   xstd
1 0.00194502    0  1.00000 1.00000 0.0076670
2 0.00054548    5  0.99027 0.99027 0.0076319
3 0.00018702    9  0.98809 0.98884 0.0076267
4 0.00012468   10  0.98791 0.98809 0.0076240
5 0.00010000   11  0.98778 0.98797 0.0076235

```

Figure 54 : Arbre élagué sur les années d'observations cumulées

G. Résultats graphiques de l'importance des variables dans l'agrégation de forêts aléatoires

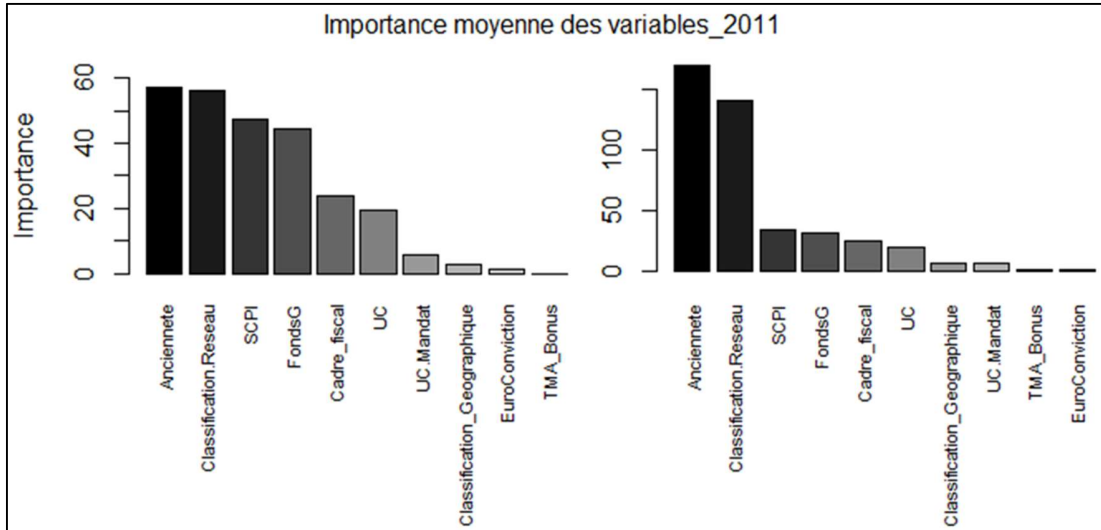


Figure 55 : Données 2011

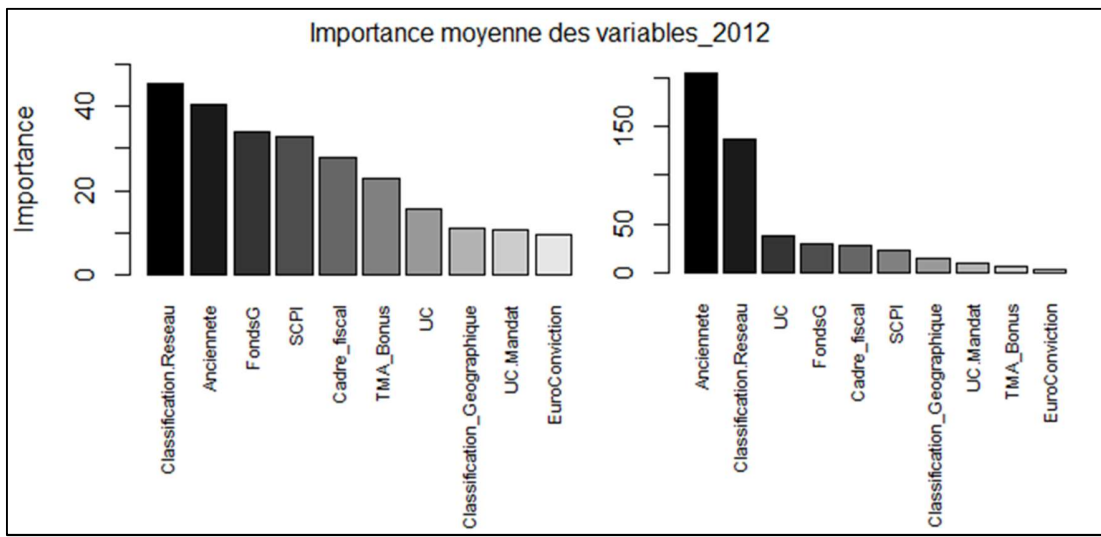


Figure 56 : Données 2012

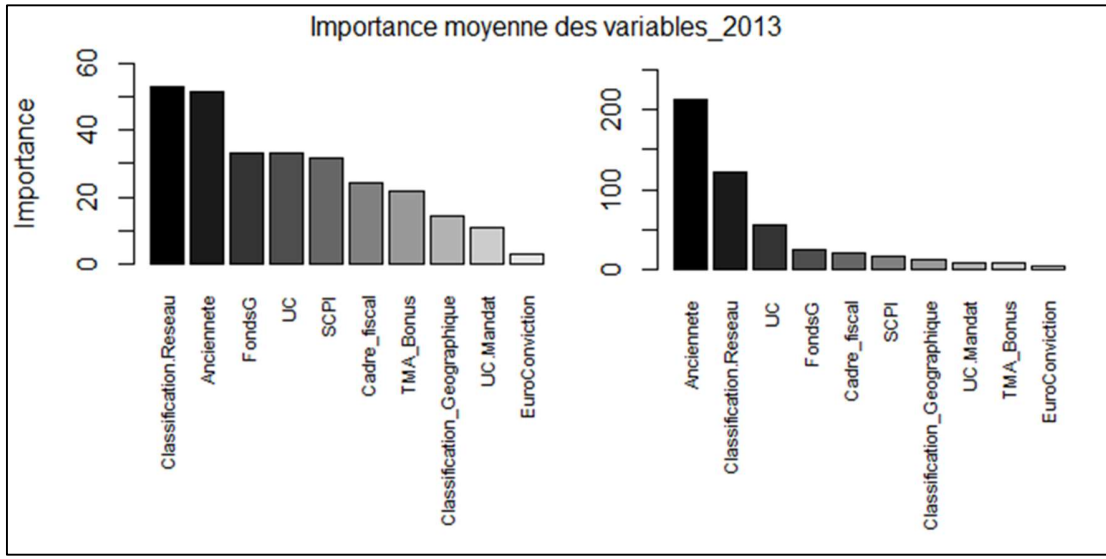


Figure 57 : Données 2013

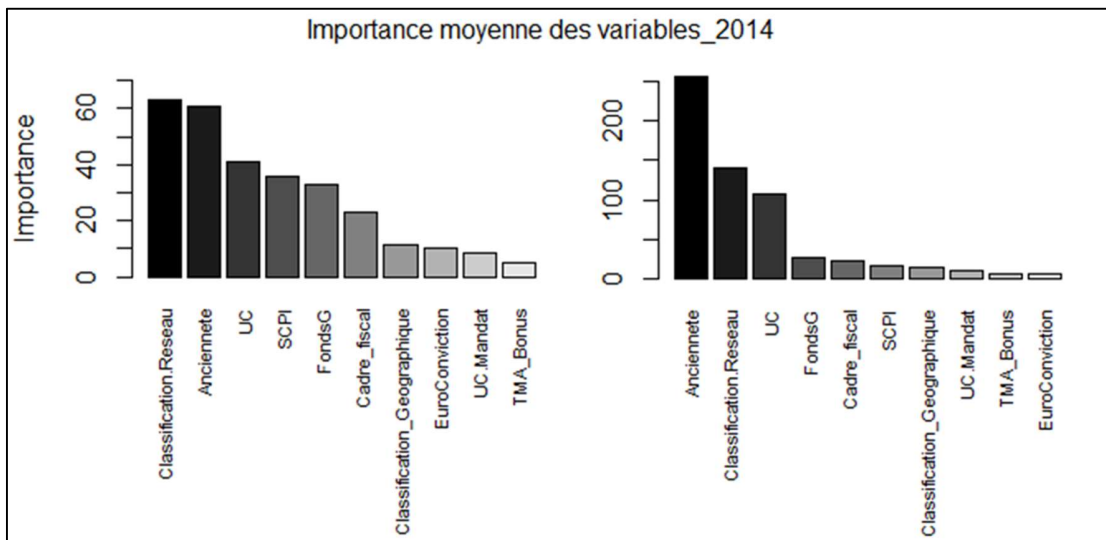


Figure 58 : Données 2014

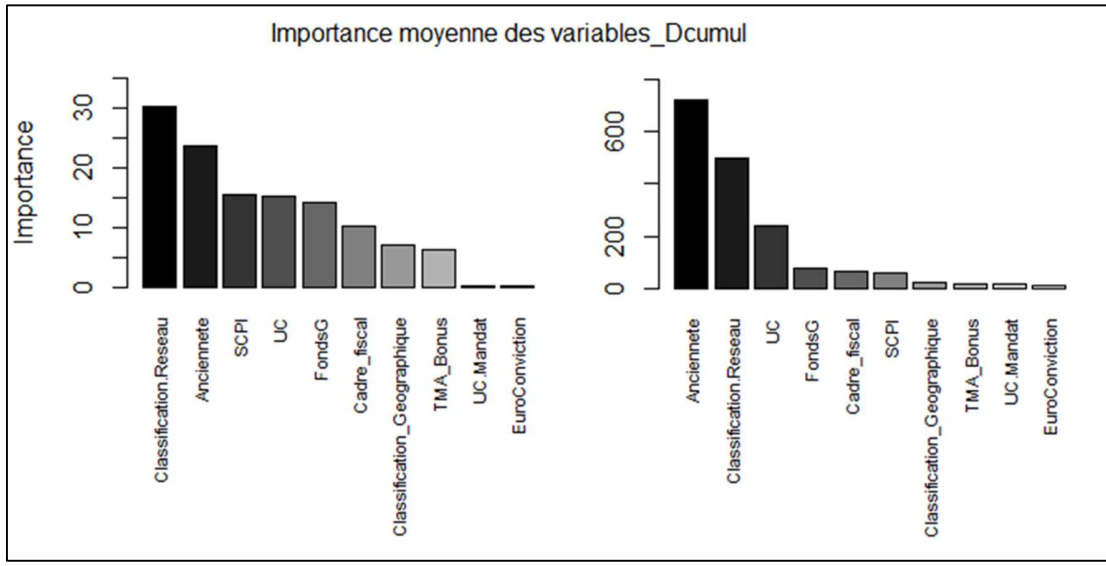


Figure 59 : Données cumulées

H. Résultat de la régression logistique par année d'observation

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		24846	24904		
Classification.Reseau	6	25466	25512	619.97	< 2.2e-16 ***
Anciennete2	13	25345	25377	498.71	< 2.2e-16 ***
FondsG	1	24858	24914	11.51	0.0006911 **
UC.Mandat	1	24849	24905	2.66	0.1026763
Classification_Geographique	1	24850	24906	3.94	0.0472102 *
UC	1	24847	24903	0.78	0.3760835
EuroConviction	1	24848	24904	2.01	0.1566741
Cadre_fiscal_v2	2	24928	24982	82.07	< 2.2e-16 ***
SCPI	1	24853	24909	6.91	0.0085658 **
TMA_Bonus	1	24848	24904	1.66	0.1972639

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.70941	0.07648	-22.351	< 2e-16 ***
Classification.ReseauPartenariat_Aise	-1.23991	0.06746	-18.379	< 2e-16 ***
Classification.ReseauPartenariat_Gestion de Fortune	-1.73061	0.18753	-9.228	< 2e-16 ***
Classification.ReseauPartenariat_Patrimonial	-0.99434	0.08681	-11.454	< 2e-16 ***
Classification.ReseauRun-off_Aise	-0.50202	0.06192	-8.107	5.19e-16 ***
Classification.ReseauRun-off_Gestion de Fortune	-0.62679	0.21811	-2.874	0.004056 **
Classification.ReseauRun-off_Patrimonial	-1.14476	0.10875	-10.527	< 2e-16 ***
Anciennete20	-1.30880	0.09680	-13.521	< 2e-16 ***
Anciennete21	-0.60520	0.07183	-8.426	< 2e-16 ***
Anciennete210	1.86689	0.22568	8.272	< 2e-16 ***
Anciennete211	0.37687	0.47713	0.790	0.429596
Anciennete212	1.21094	0.62993	1.922	0.054563 .
Anciennete213+	0.39832	0.32214	1.237	0.216273
Anciennete22	-0.36179	0.07102	-5.094	3.50e-07 ***
Anciennete23	0.19252	0.06094	3.159	0.001582 **
Anciennete25	-0.44850	0.05831	-7.692	1.45e-14 ***
Anciennete26	-0.63027	0.06592	-9.561	< 2e-16 ***
Anciennete27	-0.55506	0.07453	-7.448	9.49e-14 ***
Anciennete28	-0.14319	0.08538	-1.677	0.093545 .
Anciennete29	-0.67734	0.22959	-2.950	0.003175 **
FondsG	-0.16927	0.05032	-3.364	0.000768 ***
UC.Mandat	-0.27461	0.17341	-1.584	0.113289
Classification_GeographiqueNon-Residents	-0.42948	0.23077	-1.861	0.062738 .
UC	-0.05616	0.06343	-0.885	0.375963
EuroConviction	-0.38597	0.28813	-1.340	0.180384
Cadre_fiscal_v2Capitalisation pers. morale IR	0.64189	0.20570	3.120	0.001806 **
Cadre_fiscal_v2Capitalisation pers. morale IS	0.93077	0.09797	9.500	< 2e-16 ***
SCPI	-0.21620	0.08349	-2.589	0.009614 **
TMA_Bonus	-0.47035	0.39106	-1.203	0.229065

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 26358 on 51471 degrees of freedom				
Residual deviance: 24846 on 51443 degrees of freedom				
AIC: 24904				

Figure 60: Résultat de la régression logistique sur les données observées en 2011

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		24624	24682		
Classification.Reseau	6	25013	25059	389.22	< 2.2e-16 ***
Anciennete2	13	25131	25163	507.52	< 2.2e-16 ***
FondsG	1	24629	24685	5.13	0.0235049 *
UC.Mandat	1	24624	24680	0.40	0.5248120
Classification_Geographique	1	24624	24680	0.02	0.8842024
UC	1	24636	24692	12.20	0.0004788 ***
EuroConviction	1	24624	24680	0.14	0.7053563
Cadre_fiscal_v2	2	24706	24760	82.26	< 2.2e-16 ***
SCPI	1	24639	24695	15.86	6.82e-05 ***
TMA_Bonus	1	24635	24691	11.31	0.0007712 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-1.78815	0.07477	-23.916	< 2e-16	***
Classification.ReseauPartenariat_Aise	-1.13816	0.06747	-16.870	< 2e-16	***
Classification.ReseauPartenariat_Gestion de Fortune	-0.63886	0.13542	-4.718	2.39e-06	***
Classification.ReseauPartenariat_Patrimonial	-0.70861	0.07898	-8.972	< 2e-16	***
Classification.ReseauRun-off_Aise	-0.28578	0.06035	-4.736	2.18e-06	***
Classification.ReseauRun-off_Gestion de Fortune	-0.58856	0.22384	-2.629	0.008554	**
Classification.ReseauRun-off_Patrimonial	-0.38194	0.08414	-4.539	5.65e-06	***
Anciennete20	-1.74635	0.12387	-14.098	< 2e-16	***
Anciennete21	-0.57216	0.07896	-7.246	4.29e-13	***
Anciennete210	-0.22757	0.21078	-1.080	0.280303	
Anciennete211	-0.54085	0.51677	-1.047	0.295291	
Anciennete212	0.45051	0.38369	1.174	0.240337	
Anciennete213+	-1.12057	0.40060	-2.797	0.005154	**
Anciennete22	-0.53433	0.07493	-7.131	9.99e-13	***
Anciennete23	-0.36633	0.07620	-4.808	1.53e-06	***
Anciennete24	0.22020	0.06517	3.379	0.000727	***
Anciennete26	-0.37110	0.06307	-5.884	4.00e-09	***
Anciennete27	-0.34267	0.06698	-5.116	3.12e-07	***
Anciennete28	0.18029	0.06878	2.621	0.008758	**
Anciennete29	-0.51431	0.11170	-4.605	4.13e-06	***
Fondsg	-0.10731	0.04759	-2.255	0.024143	*
UC.Mandat	-0.08959	0.14220	-0.630	0.528667	
Classification_GeographiqueNon-Residents	-0.02569	0.17700	-0.145	0.884610	
UC	-0.20203	0.05779	-3.496	0.000472	***
EuroConviction	-0.06626	0.17659	-0.375	0.707515	
Cadre_fiscal_v2Capitalisation pers. morale IR	-0.09228	0.25958	-0.355	0.722221	
Cadre_fiscal_v2Capitalisation pers. morale IS	0.91030	0.09333	9.753	< 2e-16	***
SCPI	-0.32232	0.08296	-3.885	0.000102	***
TMA_Bonus	-0.75818	0.24942	-3.040	0.002367	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
(Dispersion parameter for binomial family taken to be 1)					
Null deviance: 26045 on 53037 degrees of freedom					
Residual deviance: 24624 on 53009 degrees of freedom					
AIC: 24682					

Figure 61: Résultat de la régression logistique sur les données observées en 2012

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		22115	22173		
Classification.Reseau	6	22394	22440	279.72	< 2.2e-16 ***
Anciennete.c	13	22701	22733	586.28	< 2.2e-16 ***
FondsG	1	22116	22172	0.79	0.3733
UC.Mandat	1	22115	22171	0.59	0.4409
Classification_Geographique	1	22116	22172	1.22	0.2695
UC	1	22169	22225	54.46	1.583e-13 ***
EuroConviction	1	22115	22171	0.00	0.9686
Cadre_fiscal_v2	2	22143	22197	27.92	8.637e-07 ***
SCPI	1	22158	22214	43.31	4.682e-11 ***
TMA_Bonus	1	22147	22203	32.70	1.076e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.494871	0.072746	-20.549	< 2e-16 ***
Classification.ReseauPartenariat_Aise	-0.844734	0.065978	-12.803	< 2e-16 ***
Classification.ReseauPartenariat_Gestion de Fortune	-1.098121	0.169100	-6.494	8.36e-11 ***
Classification.ReseauPartenariat_Patrimonial	-0.678112	0.082606	-8.209	2.23e-16 ***
Classification.ReseauRun-off_Aise	-0.112283	0.063769	-1.761	0.078275 .
Classification.ReseauRun-off_Gestion de Fortune	-1.318448	0.339203	-3.887	0.000102 ***
Classification.ReseauRun-off_Patrimonial	-0.600399	0.097079	-6.185	6.22e-10 ***
Anciennete.c0	-2.047197	0.118418	-17.288	< 2e-16 ***
Anciennete.c1	-1.007908	0.086285	-11.681	< 2e-16 ***
Anciennete.c10	-0.392548	0.107662	-3.646	0.000266 ***
Anciennete.c11	-1.948891	0.417118	-4.672	2.98e-06 ***
Anciennete.c12	-1.921182	1.009759	-1.903	0.057091 .
Anciennete.c13+	-0.898010	0.361754	-2.482	0.013051 *
Anciennete.c2	-0.876865	0.081228	-10.795	< 2e-16 ***
Anciennete.c3	-0.879946	0.078279	-11.241	< 2e-16 ***
Anciennete.c4	-0.939458	0.085199	-11.027	< 2e-16 ***
Anciennete.c5	-0.280485	0.072598	-3.864	0.000112 ***
Anciennete.c6	-0.607419	0.069968	-8.681	< 2e-16 ***
Anciennete.c7	-0.946309	0.072309	-13.087	< 2e-16 ***
Anciennete.c9	-0.409023	0.081286	-5.032	4.86e-07 ***
FondsG	0.044256	0.049635	0.892	0.372595
UC.Mandat	0.120650	0.154560	0.781	0.435035
Classification_GeographiqueNon-Residents	0.193207	0.170441	1.134	0.256975
UC	-0.418372	0.056641	-7.386	1.51e-13 ***
EuroConviction	0.006122	0.155406	0.039	0.968577
Cadre_fiscal_v2Capitalisation pers. morale IR	0.406059	0.215213	1.887	0.059191 .
Cadre_fiscal_v2Capitalisation pers. morale IS	0.591198	0.110298	5.360	8.32e-08 ***
SCPI	-0.556665	0.088532	-6.288	3.22e-10 ***
TMA_Bonus	-1.156429	0.236099	-4.898	9.68e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 23483 on 55187 degrees of freedom
Residual deviance: 22115 on 55159 degrees of freedom
AIC: 22173

Figure 62: Résultat de la régression logistique sur les données observées en 2013

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		22285	22343		
Classification.Reseau	6	22451	22497	166.44	< 2.2e-16 ***
Anciennete.c	13	22957	22989	672.39	< 2.2e-16 ***
FondsG	1	22289	22345	4.05	0.0441539 *
UC.Mandat	1	22289	22345	3.99	0.0458963 *
Classification_Geographique	1	22295	22351	10.34	0.0013010 **
UC	1	22457	22513	172.09	< 2.2e-16 ***
EuroConviction	1	22311	22367	26.17	3.129e-07 ***
Cadre_fiscal_v2	2	22302	22356	17.83	0.0001345 ***
SCPI	1	22303	22359	18.02	2.183e-05 ***
TMA_Bonus	1	22320	22376	35.22	2.950e-09 ***

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.26128	0.06983	-18.063	< 2e-16 ***
Classification.ReseauPartenariat_Aise	-0.57110	0.06287	-9.083	< 2e-16 ***
Classification.ReseauPartenariat_Gestion de Fortune	-0.66694	0.13538	-4.926	8.37e-07 ***
Classification.ReseauPartenariat_Patrimonial	-0.51805	0.07413	-6.989	2.78e-12 ***
Classification.ReseauRun-off_Aise	0.11409	0.06439	1.772	0.076432 .
Classification.ReseauRun-off_Gestion de Fortune	-1.25681	0.35603	-3.530	0.000415 ***
Classification.ReseauRun-off_Patrimonial	-0.32272	0.09522	-3.389	0.000701 ***
Anciennete.c0	-1.85239	0.10013	-18.500	< 2e-16 ***
Anciennete.c1	-1.15882	0.08470	-13.681	< 2e-16 ***
Anciennete.c10	-0.53021	0.08695	-6.098	1.07e-09 ***
Anciennete.c11	-1.72051	0.19561	-8.795	< 2e-16 ***
Anciennete.c12	-3.91822	1.00036	-3.917	8.97e-05 ***
Anciennete.c13+	-0.95054	0.28971	-3.281	0.001035 **
Anciennete.c2	-0.95460	0.08419	-11.338	< 2e-16 ***
Anciennete.c3	-0.90680	0.07950	-11.406	< 2e-16 ***
Anciennete.c4	-0.93557	0.07803	-11.989	< 2e-16 ***
Anciennete.c5	-1.18810	0.09108	-13.045	< 2e-16 ***
Anciennete.c6	-0.43670	0.07406	-5.897	3.71e-09 ***
Anciennete.c7	-0.82387	0.07320	-11.254	< 2e-16 ***
Anciennete.c9	-0.66640	0.07276	-9.159	< 2e-16 ***
FondsG	-0.10019	0.04991	-2.007	0.044723 *
UC.Mandat	-0.32678	0.17037	-1.918	0.055112 .
Classification_GeographiqueNon-Residents	0.50436	0.14727	3.425	0.000615 ***
UC	-0.71396	0.05450	-13.101	< 2e-16 ***
EuroConviction	-0.89024	0.19596	-4.543	5.55e-06 ***
Cadre_fiscal_v2Capitalisation pers. morale IR	0.13620	0.22427	0.607	0.543640
Cadre_fiscal_v2Capitalisation pers. morale IS	0.47552	0.10796	4.405	1.06e-05 ***
SCPI	-0.35712	0.08662	-4.123	3.74e-05 ***
TMA_Bonus	-0.99060	0.18792	-5.271	1.35e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 23767 on 58682 degrees of freedom
Residual deviance: 22285 on 58654 degrees of freedom
AIC: 22343

Figure 63: Résultat de la régression logistique sur les données observées en 2014

	Df	Deviance	AIC	LRT	Pr(>Chi)
<none>		21123	21181		
Classification.Reseau	6	21396	21442	273.18	< 2.2e-16 ***
Anciennete.c	13	21939	21971	815.81	< 2.2e-16 ***
FondsG	1	21123	21179	0.00	0.9495742
UC.Mandat	1	21124	21180	0.99	0.3189048
Classification_Geographique	1	21127	21183	3.63	0.0566666 .
UC	1	21361	21417	237.30	< 2.2e-16 ***
EuroConviction	1	21147	21203	23.28	1.404e-06 ***
Cadre_fiscal_v2	2	21139	21193	15.46	0.0004401 ***
SCPI	1	21125	21181	1.85	0.1733416
TMA_Bonus	1	21149	21205	25.93	3.540e-07 ***

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.255756	0.075681	-16.593	< 2e-16 ***
Classification.ReseauPartenariat_Aise	-0.567666	0.065188	-8.708	< 2e-16 ***
Classification.ReseauPartenariat_Gestion de Fortune	-1.101138	0.149078	-7.386	1.51e-13 ***
Classification.ReseauPartenariat_Patrimonial	-0.819845	0.078748	-10.411	< 2e-16 ***
Classification.ReseauRun-off_Aise	0.243783	0.064061	3.806	0.000142 ***
Classification.ReseauRun-off_Gestion de Fortune	-1.035963	0.329356	-3.145	0.001658 **
Classification.ReseauRun-off_Patrimonial	-0.123831	0.094361	-1.312	0.189414
Anciennete.c0	-2.141068	0.112485	-19.034	< 2e-16 ***
Anciennete.c1	-1.417252	0.090731	-15.620	< 2e-16 ***
Anciennete.c10	-1.001367	0.088199	-11.354	< 2e-16 ***
Anciennete.c11	-1.136702	0.115528	-9.839	< 2e-16 ***
Anciennete.c12	-1.532933	0.185661	-8.257	< 2e-16 ***
Anciennete.c13+	-1.759779	0.278810	-6.312	2.76e-10 ***
Anciennete.c2	-1.110910	0.089377	-12.429	< 2e-16 ***
Anciennete.c3	-1.211545	0.094280	-12.851	< 2e-16 ***
Anciennete.c4	-0.927234	0.084183	-11.015	< 2e-16 ***
Anciennete.c5	-1.144900	0.088511	-12.935	< 2e-16 ***
Anciennete.c6	-1.454632	0.104773	-13.884	< 2e-16 ***
Anciennete.c7	-0.335169	0.075609	-4.433	9.30e-06 ***
Anciennete.c9	-0.418515	0.067677	-6.184	6.25e-10 ***
FondsG	-0.003282	0.051902	-0.063	0.949576
UC.Mandat	-0.169178	0.173229	-0.977	0.328761
Classification_GeographiqueNon-Residents	0.334055	0.167806	1.991	0.046511 *
UC	-0.834020	0.054091	-15.419	< 2e-16 ***
EuroConviction	-0.750796	0.170250	-4.410	1.03e-05 ***
Cadre_fiscal_v2Capitalisation pers. morale IR	0.148564	0.229758	0.647	0.517884
Cadre_fiscal_v2Capitalisation pers. morale IS	0.470738	0.115051	4.092	4.28e-05 ***
SCPI	-0.113663	0.084228	-1.349	0.177189
TMA_Bonus	-1.107228	0.253921	-4.361	1.30e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 22867 on 62732 degrees of freedom
Residual deviance: 21123 on 62704 degrees of freedom
AIC: 21181

Figure 64: Résultat de la régression logistique sur les données observées en 2015