

Mémoire présenté le : 08 mars 2016

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

Par : Salma JAMAL

Titre : Construction du taux de rachat structurel en Epargne : approximation non linéaire
et agrégation de modèles

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

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

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

M. Frédéric PLANCHET
M. David VALLÉE

Signature

Entreprise :

Nom : Natixis Assurances

Signature :

Directeur de mémoire en entreprise :

Nom : Mme Sylvie BELLON

Signature :

Membres présents du jury de l'ISFA

M. Pierre-Emmanuel THÉRON

Invité :

Nom : Mme Gaëlle DOBROWOLSKI

Signature :

BPCE Vie
4 rue des Pirogues de Bercy
CS 61241
75580 PARIS CEDEX 12

*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

Secrétariat

Mme Christine DRIGUZZI

Bibliothèque :

Mme Patricia BARTOLO

Signature du candidat

[Signature]

Résumé

Mots clés : Rachats, Segmentation de portefeuille, Imputation des données manquantes, Critères discriminants, Analyse Factorielle des Données Mixtes, Arbres de régression, Réseaux de neurones, Bagging, Boosting.

Investisseurs aguerris, nouveaux rentiers et autres titulaires de biens à placer peinent à accorder leur pleine confiance aux marchés financiers. L'Épargne constitue, de fait et pour un large public, une niche d'investissement aux mécanismes rassurants. Cependant, la mise à disposition des montants épargnés n'est pas systématique. Les caractéristiques du contrat pour lequel opte chaque assuré conditionnent le désinvestissement total ou partiel de son épargne.

Plusieurs circonstances peuvent amener au rachat. Il y a celles qui sont exogènes : liées aux taux du marché, elles ne dépendent pas de l'environnement direct de l'assuré. Dans ce cas, lorsque la satisfaction de l'assuré vis-à-vis du taux de son contrat s'amointrit, il confie son épargne à un autre assureur. Le rachat est ici dit *conjoncturel*.

Dans ce mémoire, nous nous intéressons au cas des rachats qui sont exclusivement fonction de la situation intrinsèque de chaque assuré et qui sont dits *structurels*. L'objectif de l'étude présente consiste alors à constituer des groupes homogènes de risque grâce à des méthodes d'approximation non-linéaire permettant de détecter des relations complexes entre les facteurs de rachat. En effet et bien qu'elle soit incontournable en première approche, la vision linéaire atteint rapidement ses limites en présence d'un nombre important de facteurs à décrire. Elle ne décèle par exemple pas la contribution combinée des facteurs dans l'explication de variables d'intérêts.

Nous avons veillé à utiliser des méthodes adaptées à des données hétérogènes, que ce soit lors de l'imputation des données manquantes, de la description du portefeuille ou de la modélisation des taux de rachat. L'Analyse Factorielle des Données Mixtes est utilisée pour l'étape descriptive. S'adaptant à des bases volumineuses, cette méthode relève les relations linéaires entre des variables hétérogènes sans en altérer la nature. Elle permet alors de déceler de premières tendances comportementales chez nos assurés.

Le taux de rachat est ensuite désigné comme indicateur de l'intensité du rachat et est modélisé grâce à des arbres de régression et à des réseaux de neurones. Ces méthodes sont ensuite confrontées grâce à des agrégations de modèles (Bagging et Boosting).

Enfin, les modèles obtenus sont appliqués à des données récentes en guise de validation, étape commune de l'apprentissage statistique. Nous vérifions ainsi la cohérence des précédents résultats et observons l'évolution dans le temps du comportement de rachat des assurés.

La synthèse de ce mémoire vient expliquer l'apport et les limites de cette étude, avançant des voies d'amélioration pour de potentiels travaux à venir.

Abstract

Keywords : Lapses, Portfolio segmentation, Missing Data imputation, Discriminating criteria, Analysis Factor of Mixed Data, Regression trees, Neural networks, Bagging, Boosting.

At a time when financial markets health worry professionals and individuals, savings is such a secure investment nest. Indeed, it would seem logical to think that policyholders can, at any moment, recover all or a part of their savings. Actually, it is not systematic. The characteristics of each contract condition the divestment of its savings.

That said, it would be naive to consider that lapses occurred exclusively when policyholders need cash more or less urgently. Other circumstances may lead to it. It is the case, for instance, of some exogenous factors, i.e. factors that do not depend on the intrinsic situation of policyholders. This kind of factors can entail dynamic surrenders (or exceptional ones, according to the intensity of their manifestation) when policyholders decide to take benefit from interesting competitive rates.

In this study, we focus on structural lapses, that is to say those related to the intrinsic situation of each policyholder. The underlying assumption is to link the occurrence of these savings withdrawals only to the usual-everyday needs of policyholders independently of an eventual influence of financial markets health.

Thus, the purpose of this study is to identify homogeneous groups of risk reflecting the commercial history of the company and some macro-economic contextual elements.

To describe these study's data, we rely on Factor Analysis of Mixed Data. This method, tailored to large databases, detects linear relationships between heterogeneous factors without altering their nature.

We then choose to examine the redemption rate as an indicator of the lapses intensity. This target variable is modeled thanks to regression trees and neural networks. These models are then aggregated (with Bagging and Boosting methods) before being compared.

Finally, this paper synthesis explains, in the context of the study, the contribution of each analyzed method, suggesting improvement ways for potential future works.

Remerciements

Ce mémoire est le fruit d'une année d'apprentissage actuariel et humain. En parcourant ses pages, me revient le souvenir de longues journées (et de nuits) de recherches, de lectures, de programmation et d'échanges édifiants.

Elle fût la première main que j'ai serrée et le premier visage souriant que j'ai croisé chez Natixis Assurances. Ces heures où nous avons échangé sur le sujet de ce mémoire, sa patience, sa disponibilité (malgré ses nombreuses contraintes professionnelles) me sont restés. Sylvie Bellon, responsable du Département Epargne chez Natixis Assurances : une reconnaissance profonde pour m'avoir accueillie dans ton équipe.

De Gaëlle Dobrowolski, je garde l'empreinte d'un visage radieux, de consignes claires et d'une maturité professionnelle marquants. Auprès de toi, j'ai appris la construction de tables de mortalité Best Estimate, celle de lois de rachat et l'amour du thé.

Tu as été d'une bienveillance, d'une pédagogie et d'une gentillesse inégalables. Dalila Boughazi, pour ces riches journées où nous avons partagé les bonheurs de SAS, de l'analyse de données et de la bonne cuisine, une sincère gratitude.

Pour leur bonne humeur contagieuse, leurs interventions avisées, nos conversations inmanquablement instructives, je remercie François Penet, Sandrine Cadinot, Najwa Doulikian et tous les collaborateurs de la Direction de l'Actuariat.

Il m'est impensable de ne pas vous remercier pour cet inaltérable souvenir musical que vous m'avez donné la chance de vivre à vos côtés, Laurent Griveau et Patrick Gomis. Vous êtes la preuve vivante qu'il est possible d'être à la fois d'une compétence imposante et d'un incroyable sens du rythme.

C'est grâce à Pierre Théron et à Christian Robert que ce mémoire a pu emprunter de telles directions académiques et actuarielles. Le premier m'a menée sur la voie des approximateurs non-linéaires par le biais des arbres de régression, le second m'a guidée vers les lectures adéquates pour comprendre les rachats. Je vous remercie pour vos avis éclairés et vos remarques pertinentes.

A vous, mes parents, mon indéfectible soutien : il n'est propos qui puissent dire tout ce que je vous dois. Je vous dédie mon travail et plus encore.

Naturellement, une pensée particulière est adressée à tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce mémoire.

Table des matières

Introduction	5
I Aspects réglementaire et théorique du Rachat	7
1 Vision réglementaire	8
1.1 Le Rachat sous Solvabilité II	9
1.1.1 Calcul du SCR	9
1.1.2 Le Rachat dans les Spécifications Techniques	10
1.1.3 Le Rachat dans les ONC	11
1.2 Quid des enjeux fiscaux et de l'impact sur la gestion Actif-Passif?	12
2 Amorce théorique	14
2.1 Comment aborder le Rachat	14
2.2 Types de Rachat	14
2.3 Principaux travaux	16
2.4 Enjeux majeurs	17
2.5 Conclusion	18
II Structure de l'étude	19
Objectif de l'étude : Analyse des rachats en vision Best Estimate	20
Démarche et méthodes statistiques	20
III Description du portefeuille	22
3 Périmètre de l'étude	23
3.1 Base de données	23
3.1.1 Variables de l'étude	24
3.1.2 Premières observations :	26
3.2 Retraitements des données	28
3.2.1 Imputation des données manquantes	28
3.2.2 Règles d'agrégation	37
3.2.3 Taux de rachat	38
4 Analyse Factorielle des Données Mixtes	41
4.1 Cadre et hypothèses	41
4.2 L'analyse factorielle vue par les données mixtes	43
4.2.1 Analyse en Composantes Principales (ACP)	43
4.2.2 Analyse des Composantes Multiples(ACM)	46
4.2.3 Analyse Factorielle des Données Mixtes (AFDM)	47

5	Application au portefeuille de Natixis Assurances	50
5.1	Cercle des corrélations	50
5.2	Répartition des modalités	51
5.3	Interactions entre les variables	52
5.4	Conclusion	53
IV	Modélisation du taux de Rachat	54
6	Arbres de régression	55
6.1	Echantillonnage	55
6.2	Arbre de régression	56
6.2.1	Règles de segmentation	57
6.2.2	Affectation des individus	57
6.2.3	Elagage de l'arbre	58
6.2.4	Limites	58
7	Réseaux de neurones	60
7.1	Principaux travaux sur les réseaux de neurones	60
7.2	Formalisme d'un réseau de neurones	61
7.2.1	Types de réseaux de neurones	64
7.2.2	Structure d'un réseau de neurones	65
7.2.3	Types de réseaux	67
7.2.4	Ajustement des poids	68
7.2.5	Conclusion	71
8	Agrégation de modèles	72
8.1	Bagging	72
8.2	Boosting	73
8.3	Intérêt dans le cadre de l'étude	74
9	Application au portefeuille de Natixis Assurances	75
9.1	Arbres de régression	75
9.2	Réseaux de neurones	80
9.3	Agrégation de modèles	84
9.3.1	Forêt aléatoire	84
9.3.2	Boosting d'arbres de régression	86
9.3.3	Agrégation de réseaux de neurones	87
9.4	Conclusion	88
V	Validation des modèles : Application à une base récente	90
Analyse	Factorielle des Données Mixtes	92
Arbres	de régression	95
Réseau	de neurones	97
Agrégation	de modèles	97
Forêt	aléatoire	97
Boosting	d'arbres de régression	99
Agrégation	de réseaux de neurones	99
Conclusion	100

VI Conclusion	103
Annexes	107
Juridiction autour du Rachat	107
Textes législatifs	107
Contrats rachetables et contrats non rachetables	108
Abattement Fiscal	110
Compléments historiques sur le Rachat	111
Notion transverse aux approches du Rachat	111
Modélisation des taux de rachat	112
Principaux travaux sur les Rachats dynamiques	113
Analogie avec d'autres univers	115
Extension et graphiques autour des réseaux de neurones	116
Principaux réseaux de neurones	116
Tests de sensibilité sur un PCM à rétropropagation du gradient l'erreur	117
Boosting/Bagging - Graphiques	120
Table des figures	121
Bibliographie	123

Introduction

La loi de finances pour 2015ⁱ a ménagé le secteur de l'Assurance Vie, ne prévoyant que des mesures touchant à l'Assurance Maladie et aux organismes de Protection Sociale. Cette grâce, en 2015, s'entend au regard des nombreuses et fondamentales révisions apportées par la précédente réforme. En effet, le 13 décembre 2013, le Parlement adoptait définitivement la réforme de l'Assurance Vie dans le cadre du projet de loi de finance rectificative. Les contrats Euro-croissance, le fichier centralisé Ficovie, la vie génération et la fin de la commercialisation des contrats NSK ont en été les sujets phares. Ainsi, cette réforme, éclairée par le rapport Berger-Lefebvre, rappelait (et rappelle toujours) l'engouement que nourrissent les politiques pour l'Assurance Vie.

Avec près d'1.565,4Mds d'euros d'encours fin août 2015, celle qui reste le premier support de placement financier des Français engage près de 40 % de leur patrimoine financier brut. En l'espace des 8 premiers mois de l'année 2015, le cadre fiscal stable aidant, l'Assurance Vie a réalisé une collecte nette de 17,3Mds d'euros, dépassant celles, cumulées, du livret A et du Livret de Développement Durable (ces deux derniers supports ont par ailleurs dénoncé une décollecte de 4,06Mds d'euros pour un encours de 361,1Mds d'euros à fin août 2015)ⁱⁱ.

Les chiffres le martèlent, les encours de l'Assurance Vie ne sont pas figés : besoin urgent de liquidité, cadre fiscal évoluant, nouveaux produits sur le marché,... ce marché est sujet à des mouvements permanents tels les versements, les transferts ou encore les rachats. Ces derniers véhiculent des enjeux considérables pour les assureurs. Pour le comprendre, rappelons le cadre dans lequel ces opérations s'inscrivent :

Un contrat d'Assurance Vie permet à tout individu de confier un capital, versé en une ou en plusieurs fois et agrémenté de chargements fixés à la souscription, à un assureur qui s'engage, en cas de décès ou de survie de l'assuré, à restituer ce montant d'épargne diminué d'éventuels frais et augmenté d'un éventuel taux garanti à l'assuré ou à ses bénéficiaires. Le principal avantage de ce contrat est sa souplesse. Il offre à son détenteur une fiscalité intéressante, des supports d'investissement multiples et une gestion relativement libre des montants épargnés.

Par ailleurs, l'une des principales caractéristiques d'un contrat d'Assurance Vie tient à la notion d'engagement. Si celui de l'assureur est définitif, l'assuré, lui, bénéficie d'un droit de rétractation. En l'exerçant, il rachète son contratⁱⁱⁱ.

Ainsi, le rachat se manifeste lorsque un assuré souhaite récupérer la totalité ou une fraction de son épargne capitalisée, éventuellement diminuée de pénalités de rachat stipulées par le contrat. Ce retrait peut être motivé par des facteurs environnementaux (économiques tel le chômage, ou financiers tel le rendement des marchés financiers) ou expliqué par des

i. Loi n° 2014-1654 du 29 décembre 2014

ii. Source : site officiel de la FFSA et Articles de l'Argus de l'Assurance

iii. La notion de rachat n'est plus propre à l'Assurance Vie dans la mesure elle a, plus récemment, été étendue au cas des contrats pluriannuels d'Assurance Non Vie dans les Spécifications Techniques du QIS5

caractéristiques du contrat ou de l'assuré. Dans le premier cas, le rachat est dit *dynamique*, dans le second, *structurel*. D'autres fois encore, le rachat se produit dans des proportions spectaculaires, souvent déclenché par un effondrement global des marchés financiers. La panique entraîne un comportement grégaire des investisseurs et un rachat massif de leur épargne. Dans un tel contexte, le rachat est *exceptionnel*.

Les conséquences découlant du rachat de contrats d'Assurance Vie peuvent être désastreuses (cas du rachat exceptionnel par exemple). Aussi, pour contrôler ce risque, la réglementation définit un cadre précis pour le calcul des provisions techniques. Dans le cadre de la réforme Solvabilité 2 typiquement, chaque assureur doit estimer de façon très juste, et non plus excessivement prudente, le risque de rachat auquel son portefeuille l'expose.

Dans ce mémoire, nous proposons de modéliser les taux de rachats structurels constatés au sein du pôle Epargne de Natixis Assurances.

- * La 1^{ère} partie présente le contexte réglementaire et le cadre académique dans lesquels les rachats se sont développés et continuent à évoluer ;
- * Les 2^{ème} et 3^{ème} parties s'intéressent aux contrats sujets à rachat entre 2005 et 2010 :
 - La base de données est d'abord retraitée à l'aide, notamment, de méthodes d'imputation de données manquantes. Elle est ensuite décrite grâce à une Analyse Factorielle des Données Mixtes (AFDM).
 - Les taux de rachat sont ensuite analysés grâce à des arbres de régression et à des réseaux de neurones. Une agrégation de ces modèles permet d'en augmenter la robustesse et d'en fiabiliser les résultats.
- * La 4^{ème} partie est dédiée à l'application des méthodes retenues par le portefeuille 2005 - 2010 (AFDM, arbres de régression et réseaux de neurones) à une base de données récente (2013) afin de valider les modèles et de constater l'impact de l'évolution dans le temps des variables sur un portefeuille plus mûr.
- * Enfin, une synthèse permet de tirer des conclusions sur la pertinence des méthodes employées et propose des voies d'amélioration pour de potentiels travaux à venir.

Première partie

Aspects réglementaire et théorique
du Rachat

Chapitre 1

Vision réglementaire

Les contrats individuels en Assurance Vie présentent, entre autres, la particularité suivante : moyennant renonciation à la garantie de l'assureur, il est possible de se voir rembourser une partie de son contrat initial, la valeur de rachat. En France, les montants minimaux de rachat sont garantis et doivent, lors de la souscription, être déclarés dans la notice d'information ou dans les conditions généralesⁱ. Au moment du rachat, la provision mathématique du contrat définit alors l'assiette du remboursement. En effet, cette provision, désormais inutile pour l'assureur, peut (doit) être récupérée par l'assuré.

Le déclenchement des rachats peut découler de facteurs endogènes ou exogènes à la nature du contrat racheté. Dans un contexte économique équilibré où les marchés financiers proposeraient des prestations identiques en termes de taux de rendement des actifs, la décision du rachat serait uniquement liée à la situation intrinsèque de l'assuré. La réalité est plus nuancée : les assurés sont de plus en plus sensibilisés aux mouvements des marchés, notamment par le biais des médias. Les fluctuations boursières influencent nécessairement le comportement de ces assurés. Ainsi, quand les taux augmentent de façon plus ou moins marquée, les rachats peuvent se manifester par vagues conséquentes et sur des montants importants, entraînant potentiellement la ruine d'un assureur. A contrario, lorsque les taux d'intérêt chutent, les assurés ne voient plus l'intérêt de racheter leurs contrats. Par conséquent, l'assureur voit ses actions dévalorisées sur les marchésⁱⁱ et réalise des moins-values en vendant ses actifs.

Lorsque ces scénarios de hausse et de baisse des taux se réalisent, la survenance de vagues de rachats peut ne laisser vivre chez l'assureur qu'un effectif restreint de contrats en cours. L'équilibre du principe de compensations dans une mutualité est alors rompu. Une disposition législative permet alors à l'autorité de contrôle de suspendre le paiement des valeurs de rachatⁱⁱⁱ.

Par ailleurs, le risque de sélection adverse a également été désamorcé par la réglementation. En France, la loi interdit le rachat sur des contrats (tout en autorisant le transfert) dont la garantie est exclusivement en cas de vie (notamment pour les rentes viagères).

Enfin, puisqu'ils portent systématiquement sur des provisions considérables, la réforme Solvabilité II a induit une prise en compte plus juste et plus réaliste des rachats. En effet, sous le régime de cette nouvelle réforme, chaque assureur devra, entre autres mesures, constituer un montant de provisions reflétant exactement son exposition au risque de rachat.

i. A minima, les montants de rachat correspondants aux 8 premières années de vie du contrat

ii. voir mémoire d'A. Lamirault et de L. Vitry "Un contexte de taux bas : Quelle modélisation pour les rachats ?"

iii. Charte du contrôle de l'ACP - secteur assurance. Juin 2010

La section suivante aborde ainsi la gestion des rachats dans l'environnement assurantiel prochainement réformé.

1.1 Le Rachat sous Solvabilité II

En Europe, au vu des enjeux financiers que véhicule ce phénomène, le rachat est identifié comme étant le risque résiduel le plus important parmi tout autre risque de souscription^{iv}.

Ceci étant, sous la réforme Solvabilité II, la justesse prend le pas sur la prudence, en particulier en question de rachats. Une estimation au plus près des engagements de l'assureur est plus réaliste et plus légitime que la vision "apocalyptique" actuelle où, à tout instant, les probabilités de rachat sont maximales.

Cet objectif de précision est particulièrement délicat à atteindre dans la mesure où il requiert de cerner au mieux tous les risques gravitant autour des rachats et où l'exposition de chaque assureur à ces différents risques dépend de son historique commercial, politique, économique,....

La législation a donc naturellement défini un cadre normatif précis :

1.1.1 Calcul du SCR

Le SCR (*Solvency Capital Requirement*) correspond à une exigence de capital mise en place par la réforme Solvabilité II visant à limiter la probabilité de ruine^v des organismes d'assurance.

Il s'entend que les lignes d'activité d'un assureur définissent le périmètre des risques auxquels il est soumis. Son SCR est donc naturellement fonction de la composition de son portefeuille : il reflète alors son profil de risque.

Le montant minimal de fonds propres est soumis à un calcul rigoureusement encadré : il peut être déterminé grâce à une formule fermée (la **formule standard**) ou, après aval du superviseur, à un **modèle interne**. Si l'assureur choisit de s'appuyer sur la formule standard, il doit respecter la division bottom-up (ou modulaire) des différents modules liés aux facteurs de risque identifiés par le législateur (Figure 1.1). Une troisième option consiste à combiner ces deux méthodes.

Quant à la détermination de la part de SCR associée au risque de Rachat, les assureurs doivent calculer les 3 variations suivantes de la charge de capital :

- * *Lapse_{down}* : Pour les contrats dont la valeur de rachat est en dessous des provisions techniques, cette quantité est déterminée après calcul de la variation de la NAV^{vi} suite à une baisse permanente de 50% des taux de rachats (limitée à une baisse absolue de 20%);
- * *Lapse_{up}* : Pour les contrats dont la valeur de rachat dépasse les provisions techniques, cette valeur est estimée suite à une hausse permanente des taux de rachats (avec un plafond absolu de 100%);
- * *Lapse_{mass}* : Cette quantité, fixe, est attachée à toutes les polices. Elle vaut 30% de la perte en cas de chute des polices, si cette chute provoque une perte de valeur.

iv. EIOPA : EIOPA, report on the 5th Quantitative Impact Study(QIS5) for Solvency II. Mars 2011.

v. La probabilité de ruine devant être inférieure à 0,5% sur un horizon d'un an.

vi. C'est la différence entre la valeur des actifs et celle des passifs après application d'un choc causé par la hausse/baisse des taux d'intérêts.

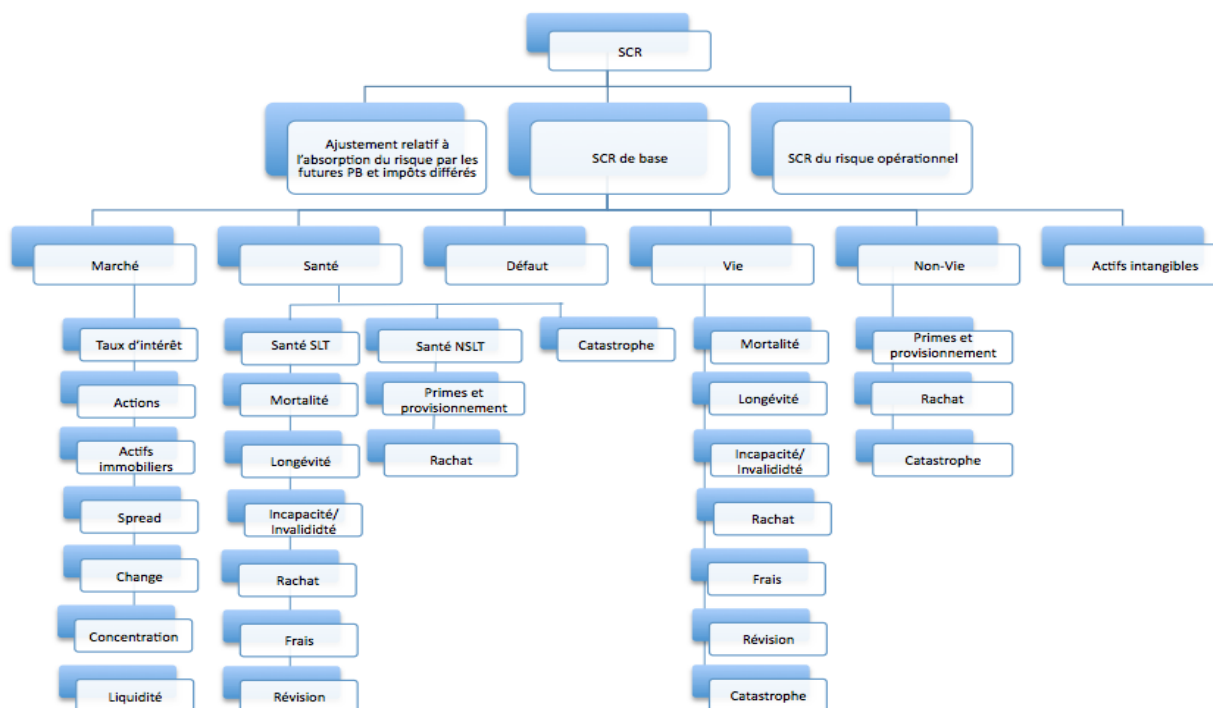


FIGURE 1.1 – Cartographie du SCR sous Solvabilité II

Le maximum de ces trois quantités constitue alors le SCR_{Rachat} .

Les SCR des différents modules de risques sont ensuite agrégés pour constituer le $BSCR$ (*Basic Solvency Capital Requirement*). Le BSCR est ainsi une combinaison des modules de risques (exprimés en pourcentage du BSCR) modulo une diminution de l'effet de diversification.

1.1.2 Le Rachat dans les Spécifications Techniques

L'EIOPA

L'EIOPA (l'Autorité Européenne des Assurances et Pensions Professionnelles (en français), ex-CEIOPS) voit le jour le 1^{er} janvier 2011. Cet organe de surveillance indépendant rassemble les autorités de contrôle des 28 Etats membres de l'Union Européenne et de l'Espace Economique Européen et veille sur les droits des assurés en éclairant la Commission sur la mise en place de la réglementation. Il oriente, produit des spécifications (normes) techniques et tient un rôle de conseil auprès du Parlement Européen, du Conseil de l'Union Européenne et de la Commission Européenne.

Les mesures définies par les spécifications techniques sont "testées" par l'EIOPA. Solliciter le concours d'organismes d'assurance européens permet à l'EIOPA de publier des études quantitatives d'impact (QIS) qui mesurent les répercussions de l'adoption des paramètres fixés par les spécifications techniques.

Les spécifications techniques de l'EIOPA accordent la définition suivante au risque de rachat :

"Le risque de Rachat correspond au risque de perte ou de variation défavorable à l'échelle des passifs du fait d'un changement des taux d'exercice prévus au niveau des options des assurés. Les options retenues pour leur pertinence sont celles liées aux droits juridiques ou contractuels des assurés quant à la conclusion totale ou partielle de leur contrat, l'avance, la réduction, la limitation/extension/suspension de la couverture (de la police) ou encore au maintien de la police jusqu'à son terme (échéance) contractuel." ^{vii}

Le niveau de précision des Spécifications Techniques sur le Rachat confirme l'intérêt que porte l'EIOPA à la bonne maîtrise de ce risque. Nous citons ainsi :

✱ **Les engagements de l'Assurance Vie (TP.2.64.)**

Si la valeur de rachat d'un contrat est supérieure à la somme de la marge de risque et de la valeur *Best Estimate* de ce contrat, la valeur du passif n'a pas à être valorisée pour atteindre cette valeur de rachat.

✱ **Option contractuelle (TP.2.95. et TP2.96)**

Le rachat anticipé correspond à une option contractuelle par le biais de laquelle le souscripteur peut récupérer tout ou partie de son épargne selon des conditions préalablement définies. L'exercice d'une option contractuelle (ici, le rachat anticipé) ne requiert pas le renouvellement des accords des parties.

✱ **L'option de Participation aux Bénéfices (TP.6.17. et TP.6.47.)**

La proposition d'une clause de Participation aux Bénéfices ajoute une dimension de complexité à l'étude du rachat d'un contrat d'Assurance Vie. Le montant de Participation aux Bénéfices versé aux assurés dépend de la marge annuelle engrangée par l'assureur, forcément variable. Le montant versé à l'assuré intervient très probablement dans sa décision de racheter son contrat. Les paramètres d'application de cette clause doivent donc être intégrés dans les modèles. De fait, l'utilisation d'une formule fermée est difficilement applicable au vu de la difficulté d'y inclure les décisions de gestion.

✱ **Motifs rationnels de variation des taux de rachat (TP.6.17.)**

Indépendamment de l'environnement économique, les caractéristiques de l'assuré contribuent à sa décision de racheter ou non son contrat. Chaque assureur doit initier une description de son portefeuille pour exhiber le lien entre le rachat et différents facteurs endogènes (essentiellement non économiques et opérationnels). Les facteurs discriminants relevés sont ensuite injectés dans les hypothèses de projection de flux afin d'observer l'étalement dans le temps d'un éventuel sur/sous provisionnement de certaines catégories d'assurés.

1.1.3 Le Rachat dans les ONC

Les Orientations Nationales Complémentaires (ONC) aux Spécifications Techniques sont des rapports rédigés par l'ACPR. Ces documents sont rattachés aux dernières études quantitatives d'impact (QIS5) : ils les complètent sans pour autant les remplacer.

Les ONC [17] s'intéressent, entre autres, aux rachats structurels et dynamiques. Alors que les rachats conjoncturels sont fonction de taux de marchés concurrentiels, les rachats structurels dépendent de paramètres liés à des tables d'expérience :

vii. *"Lapse risk is the risk of loss or adverse change in liabilities due to a change in the expected exercise rates of policyholder options. The relevant options are all legal or contractual policy holder rights to fully or partly terminate, surrender, decrease, restrict or suspend insurance cover or permit the insurance policy to lapse."* - Technical Specification for the Preparatory Phase (Part I). EIOPA April, 30th

- * S'agissant des **rachats structurels** : Un taux moyen de rachat peut être appliqué à des populations d'assurés appartenant à des groupes homogènes de risque à condition que l'entité démontre que, dans son cas, la distinction de variables discriminantes au rachat n'a pas d'impact sur le calcul des taux de rachat ;
- * Les **rachats dynamiques** sont sujets à des indications plus détaillées. La différence entre le taux servi par l'entité et le taux de rendement espéré par l'assuré définit des règles de calcul méthodiques du taux de rachat dynamique.

1.2 Quid des enjeux fiscaux et de l'impact sur la gestion Actif-Passif ?

Sans perte de généralité, il y a autant d'assurés que de potentielles motivations pour le rachat d'un contrat d'Épargne. Il est toutefois possible de reconnaître trois facteurs/contextes communs contribuant à la décision de rachat :

- Un besoin de liquidité plus ou moins imprévu ;
- Un contexte économique fluctuant (pouvant occasionner rachats dynamiques ou exceptionnels selon l'effet pandémique et l'ampleur de la baisse des taux dans les marchés financiers) ;
- L'évolution des avantages fiscaux liés au contrat racheté.

La fiscalité constitue en effet un pan important dans l'étude des rachats : lorsqu'un qu'un assuré souhaite effectuer un rachat sur son contrat d'Assurance Vie, les intérêts produits par son contrat lui sont imposables pour certains intervalles d'ancienneté dudit contrat. Après déduction des cotisations sociales (CSG, CRDS et prélèvements sociaux), ces intérêts peuvent être signalés par l'assuré dans le cadre de sa déclaration d'impôts sur le revenu (IR) (barème progressif) ou être taxés directement au moment du versement de l'épargne à l'assuré : c'est le principe du **Prélèvement Forfaitaire Libératoire (PFL)**, qui se manifeste sous forme d'un crédit d'impôt. Cependant, qu'il s'agisse du barème progressif ou du PFL, les taux appliqués aux intérêts diffèrent selon l'ancienneté et la date de souscription du contrat.

Notons que le PFL a, suite à la validation de la loi de finances de 2013, été supprimé pour les revenus du patrimoine. Cette mesure ne concerne que les revenus perçus en 2013 et ultérieurs, sujets à déclaration en 2014 et postérieurs. Ces derniers sont donc nécessairement soumis au barème progressif de l'IR.

Le tableau suivant résume les grandes lignes de la fiscalité appliquée lors du rachat d'un contrat d'Assurance Vie. A ce sujet, le mémoire de Fauvel et Le Pévédic [28] sur l'analyse de rachats d'un portefeuille de vie individuelle inclut une vision panoramique intéressante sur le lien entre les rachats et la fiscalité française.

Date de souscription du Contrat		
Avant le 01/01/1983	Entre le 02/01/1983 et le 31/12/1989	Après le 01/01/1990
Exonération d'impôts et application de prélèvements sociaux	<p>Uniquement pour les versements effectués avant le 26/09/1997, barème progressif (IR) ou PFL selon l'ancienneté du contrat :</p> <ul style="list-style-type: none"> — 45% pour une ancienneté < 2ans — 25% pour une ancienneté comprise entre 2 et 4 ans — 15% pour une ancienneté comprise entre 4 et 6 ans — 0% pour une ancienneté \geq 6 ans <p>Pour les versements effectués après le 25/09/1997, le PFL des contrats d'une ancienneté \geq 6 ans est de 7,5 % et le PFL est appliqué (toutes anciennetés de contrats confondus, après un abattement annuel de 4 600 €(resp. 9 200 €) pour une personne seule (resp. pour un couple soumis à une imposition légale commune)</p>	<p>Pour les versements effectués depuis la souscription, barème progressif (IR) ou PFL selon l'ancienneté du contrat :</p> <ul style="list-style-type: none"> — 35% pour une ancienneté < 4 ans — 15% pour une ancienneté comprise entre 4 et 8 ans — 0% pour une ancienneté \geq 8 ans <p>Pour les versements effectués après le 25/09/1997, le PFL des contrats d'une ancienneté \geq 8 ans est de 7,5 % et le PFL est appliqué (toutes anciennetés de contrats confondues, après un abattement annuel de 4 600 €(resp. 9 200 €) pour une personne seule (resp. pour un couple soumis à une imposition légale commune)</p>

D'autre part, il est intéressant d'évoquer le rapport entre la hausse des rachats et l'interaction Actif-Passif. Antoine Esquieu, associé Mazars, souligne dans un article de l'Argus de mars 2013 que la hausse des taux d'intérêt entraîne une moins-value des actifs des assureurs. Il en résulte une diminution du rendement des actifs et une vague de rachats massifs : les assureurs sont alors contraints de puiser dans leurs fonds propres. Pour limiter le risque lié à la hausse brutale des taux, les assureurs disposent d'un certain attirail de solutions plus ou moins probantes : le raccourcissement de la durée de l'actif obligataire, l'acquisition de couvertures, le recours à des obligations à taux variable ou la diminution choisie de la collecte, laquelle limiterait la dilution du rendement des actifs. Cette dernière solution, extrême, reste naturellement un dernier recours du fait de ses motivations difficiles à plaider d'un point de vue socio-économique.

Chapitre 2

Amorce théorique

Le rachat a d'abord été un phénomène dénonçant un ensemble de conditions économiques sous le joug desquelles le souscripteur d'une police d'Assurance Vie classique n'arrivait plus à verser de primes sur son contrat, entraînant alors la clôture de ce contratⁱ. Ensuite, force a été de constater que les assurés étaient également influencés par des éléments les définissant à titre personnel. Le rachat peut alors être associé à une réaction plus ou moins consciente des assurés. Le comportement de rachat est soit complètement volontaire (conscient), soit forcé par un besoin financier (inconscient), soit tenant à une incapacité à maintenir son contrat en vie (conscient ou non).

Loin de ces considérations comportementales, le formalisme mathématique du rachat s'est surtout appuyé sur des études économiques :

2.1 Comment aborder le Rachat

Quatre principales écoles analysent les rachats :

- * L'approche **statistique** repose sur des séries temporelles : une agrégation des contrats permet d'observer l'évolution des rachats dans le temps. Cette méthode permet de distinguer des pics (valeurs extrêmes) pouvant dénoncer des périodes économiques critiques (à l'échelle de l'entreprise ou à des niveaux plus importants) ;
- * Le point de vue **économique** se base sur la théorie de l'espérance d'utilité de Von Neumann et Morgenstern(1944) [51]. L'hypothèse fondamentale en est la rationalité des assurés ;
- * La vision **financière** se penche sur des modèles liés au rachat dynamique. Les contrats y sont considérés individuellement ;
- * Dans l'analyse **probabiliste**, les contrats sont approchés individuellement grâce à des modèles linéaires généralisés, des arbres de régression,...

Ces notions fondamentalement différentes reposent, entre autres, sur un principe commun : l'appétence au risque des assurés. Cette partie est développée en annexes VI.

2.2 Types de Rachat

Consiglio et De Giovanni(2010) [11] distinguent deux types de rachat liés au contexte qui entraîne la décision de rachat de l'assuré :

i. Modeling of Policyholders Behavior for Life Insurance and Annuity Products - A survey and Literature Review (2014), SOA

- Le premier type de rachat, jugé rationnel, est lié aux mouvements des marchés financiers : un investisseur met à profit son droit de racheter son contrat lorsque des opportunités d'arbitrage s'offrent à lui. Dans ce genre de contexte, le rachat est dit optimal car il offre au détenteur du contrat racheté le meilleur revenu possible au regard des mouvements du marché ;
- Le second est, quant à lui, expliqué par des indicateurs socio-économiques tels que le taux de chômage, le taux d'intérêt,... Le rachat, ici, est "irrationnel". Non pas du point de vue de l'assuré qui nourrit des arguments recevables à son sens lors du rachat de son contrat, mais au regard de la sphère boursière. En effet, le rachat peut survenir à un moment où l'assuré a des besoins de liquidité prononcés mais où les taux sont au plus bas dans les marchés.

Ce sont donc les facteurs incitant au rachat qui permettent d'en définir les deux principales manifestations :

Le Rachat structurel

Le rachat structurel est soumis à des facteurs humains relativement instables. En ce sens, il ne peut être schématisé par des modèles mathématiques infaillibles. A partir du moment où un facteur de rachat peut être intégré comme hypothèse mathématique quantifiable, le rachat n'est plus structurel. En général, tout ce qui est assimilable à un effet socio-économique dont les répercussions se ressentent à échelle individuelle est considéré comme motif structurel de rachat ⁱⁱ.

Peu d'ouvrages et d'articles traitent du rachat structurel : les informations qui constituent la matière essentielle pour l'étude de ces rachats sont celles, confidentielles, procurées par les contrats des assurés. En outre, l'erreur humaine ou les migrations informatiques peuvent entraîner une déperdition des données à un niveau interne, accentuant la problématique des données.

Le Rachat conjoncturel ou dynamique

Le rachat dynamique résulte souvent d'une hausse importante et durable des taux, laquelle aboutit à une opportunité d'arbitrage pour l'assuré (Bacinello (2005) [3]).

La littérature traitant de ce type de rachats est abondante. Associé à des indicateurs macro-économiques ou à des caractéristiques de l'entreprise, le rachat dynamique repose sur des données publiques.

Aparté :

Une troisième catégorie de rachat dite *exceptionnelle* peut être définie comme une manifestation rare et extrême du rachat. Alors que le rachat dynamique s'effectue dans un contexte d'opportunité d'arbitrage, l'assuré peut être confronté à une situation de perte de confiance totale en l'univers financier. Le rachat ne se fera pas ici pour bénéficier d'opportunités lucratives du marché mais pour en soustraire la totalité de son épargne. Ce comportement de panique, grégaire et massif, se produit généralement lors d'effondrements spectaculaires des marchés (Faillite de Lehman Brothers, crise de la dette publique grecque,...)

ii. Cependant, bien qu'appartenant à cette catégorie d'éléments, le PIB et le taux de chômage (par exemple) sont plutôt considérés comme des facteurs dynamiques.

Facteurs de Rachat

Toutes catégories de rachat confondues, les facteurs motivant la décision du rachat peuvent être classés comme suit :

Rachats Structurels		Rachats Dynamiques	Rachats Exceptionnels
Caractéristiques de l'assuré	Spécificités du contrat		
Âge Sexe Localisation géographique Segmentation commerciale (selon des avoirs bancaires et assurantiels)	Abattement fiscal Ancienneté du contrat Gamme de produit Type de contrat Frais de gestion Type de support Canal de distribution Pénalité de rachat Clause de Participation aux bénéfices Taux Moyen Garanti par le contrat	PIB Prix des garanties Etat de santé de l'assuré Taux de chômage Notation des agences de rating	Effondrement des marchés financiers

Les facteurs ci-dessus constituent une liste non exhaustive de facteurs motivant la décision de rachat. L'ensemble de ces critères n'apparaît naturellement pas comme déterminant dans une étude sur le rachat. La sélection de critères dans le cadre d'une étude expérimentale repose sur une réflexion économique validée par des tests statistiques.

2.3 Principaux travaux

Les études sur les rachats ont toujours dépendu de la nature, de la disponibilité et du volume des données analysées.

Dans la mesure où ce mémoire s'intéresse aux rachats structurels, nous présentons dans cette section les travaux menés sur ces rachats. Le lecteur intéressé est invité à consulter en Annexes VI les travaux qui se sont penchés sur les rachats dynamiques.

Rachats structurels

Puisque, par nature, les rachats structurels s'appuient sur des données non agrégées (donc nécessairement confidentielles), ils ont bénéficié de recherches et d'une littérature moins développées que pour les rachats conjoncturels.

En 1986, Renshaw et Haberman) [61] optent pour une régression logistique et un modèle binomial. Ils mettent en avant l'intervention de 4 critères dans la décision de rachat : l'âge à l'entrée dans le portefeuille, l'ancienneté du contrat, la réputation de l'entreprise et le type de contrat. Ils soulignent ainsi, à titre d'exemple, la corrélation entre le type et l'ancienneté du contrat. Leurs travaux montrent que la prise en compte d'une combinaison de critères est autrement plus pertinente que le fait de lier l'évolution des rachats à un seul facteur.

Le facteur économique (taux de chômage) est pondérant dans les études de Kagraoka(2005) [36] : ses conclusions, reposant sur la loi de Poisson et la loi Binomiale Négative, valident l'EFH (Emergency Fund Hypothesis, voir les annexes VI).

En 2009, l'ancienneté des contrats, l'année calendaire où le rachat se produit et la classe de produit sur lequel porte le contrat racheté sont présentés comme motifs incitant au rachat dans les études de Cerchiara et al.(2011) [10].

Par ailleurs, certains facteurs ont une contribution plus spectaculaire selon le pays où l'évolution du rachat est étudié. En France, la fiscalité avantageuse des contrats d'Assurance Vie est un facteur décisif au rachat. Ce constat a été mis en avant par Milhaud et al.(2010) [49]. Ils soulignent également l'importance de la prise en compte de l'ancienneté des contrats et de la clause de participation aux bénéfices dans la décision de rachat.

Enfin, l'âge des assurés ainsi que la fréquence de versement des primes sont également reconnus comme variables explicatives des rachats.

S'agissant d'études récentes sur le sujet : à l'instar de Milhaud (2011) [48] qui s'appuie sur la régression logistique et sur des arbres de régression (CART), des mémoires d'actuaire se sont penchés sur la question du rachat structurel en utilisant des méthodes originales et pertinentes :

- Les modèles de durée offrent un outil innovant pour analyser les rachats totaux. En assimilant ces derniers à des flux sortants, Berne (2013) [7] ramène l'étude des rachats totaux à celui d'un risque de survie en s'appuyant sur la méthode de Kaplan Meier. Seror et Nkihouabonga Yengue (2013) [69] optent pour leur part pour un modèle à hasards proportionnels de type Fine&Gray ;
- Fauvel et Le Pévédic(2011) [28] s'intéressent à une vision économique du rachat, plus précisément du point de vue de la théorie de l'utilité espérée de Von Neumann et Morgenstern. En empruntant, de plus, le raisonnement des arbres binomiaux, elles comparent lequel d'un rachat partiel ou d'un emprunt est le plus avantageux pour un assuré.

2.4 Enjeux majeurs

Les exigences réglementaires évoluant, les modèles doivent être constamment repensés. Quelles que soient les modifications qu'ils aient à subir, ces modèles doivent intégrer certaines problématiques incontournables :

- **Dans quelle mesure les comportements de rachat des assurés sont-ils liés ?**
Deux principaux acteurs influencent les décisions de rachat des assurés : les réseaux de distribution (voir Shumrak et al.(1999) [70]) et les médias (voir Anzili et De Cesare(2007) [1]). Crosby et Stephens (1987) [14] affirment que la satisfaction des assurés est intimement liée à la confiance qu'ils accordent à l'organisme qui gère leur épargne.
- **Quel est l'impact de l'anti-sélection sur les rachats ?**
Vannucci(2003) [75] met en avant le lien entre l'anti-sélection et le rachat en invoquant la question de l'asymétrie d'information entre un assureur et ses assurés (déséquilibre d'information en faveur, *in fine*, de ces derniers).
- **En cas de rachats exceptionnels, comment réagissent les assurés ?**
La notion de rachat exceptionnel, pour être convenablement traitée, doit être étudiée avec un historique de données important. Alors que Milhaud et Loisel(2011) [49] l'étudient d'un point de vue quantitatif, les réactions de panique des investisseurs sont étudiées d'un point de vue psychologique par d'autres auteurs (Fum et al.(2007) [30],

Pan et al.(2006) [57],...). Citons également Dhaene et al. [18] qui ont développé des recherches autour d'un indice de comportement grégaire (HIX), Luce (2010) [47] qui a étudié le lien entre périodes de crises et risquophilie des investisseurs ou encore Viscusi et al.(2011) [77] dont le rapport établit l'impact des informations diffusées en temps réel sur la santé des marchés financiers.

2.5 Conclusion

Un modèle parfaitement adapté à l'étude des rachats doit intégrer des éléments incontournables : les contraintes du législateur, les scénarios extrêmes, l'implication d'agents indépendants (courtiers, médias,...), les processus de prise de décision (approche économique des besoins de la société), l'anti-sélection, les hypothèses d'évolution des marchés, le périmètre d'informations financières auxquelles les assurés ont accès, ... et l'interaction entre tous ces éléments.

Il faut également être conscient du renouveau permanent dans le monde des assurances : de nouveaux produits sont voués à naître et les besoins économiques des agents à évoluer. Inclure ces facteurs incontrôlables sur un horizon de plus d'un an est difficilement envisageable mais il ne faut pas négliger le risque qu'ils véhiculent.

Ainsi, la prudence voudrait que les assureurs réalisent que leurs assurés acquièrent une vision éclairée du mouvement des marchés et adoptent, par là, un comportement de plus en plus dynamique et de plus en plus rationnel.

Deuxième partie

Structure de l'étude

Objectif de l'étude : Analyse des rachats en vision Best Estimate

Un risque est **couvrable** s'il peut être répliqué sur un marché financier. Les risques non répliquables sont donc particulièrement délicats à appréhender dans la mesure où ils ne peuvent être prévenus par aucune stratégie de couverture. De là, et à l'instar des risques soumis au principe de mutualisation (mortalité, ...), **le rachat est un risque non répliquable** puisqu'il dépend exclusivement du comportement des assurés. Par ailleurs, puisque la gestion des rachats dynamiques est harmonisée dans le QIS5, c'est en particulier le risque pesant sur les rachats structurels qui doit être évalué de façon autonome et robuste par chaque assureur.

Dans la mesure où la réforme Solvabilité 2 exige une vision Best estimate des risques de souscription, donc du rachat, il a été nécessaire pour Natixis Assurances de revoir ses critères d'agrégation afin de modéliser la meilleure estimation du comportement des assurés.

A ce jour, l'agrégation des contrats dans les modèles du service ALM (*Asset and Liability Management*) est basée sur la segmentation du portefeuille en trois catégories fiscales dépendant de l'âge des assurés et de l'ancienneté de leurs contrats (Cf . Figure 3.2). La segmentation des critères contributifs au rachat issue de cette étude a ainsi vocation à être intégrée dans les *Model Points* ALM (comme autorisé par le TP.1.70.C du QIS5).

Démarche et méthodes statistiques

Cette étude est réalisée en trois temps : **description du portefeuille** (partie III), **modélisation d'un indicateur du rachat** (partie IV) et **validation des modèles statistiques** (partie V).

- * La première étape consiste à **construire des groupes homogènes de risque**. La base de données utilisée est celle de contrats d'épargne rachetés. Ces groupes, par construction, vont refléter la sensibilité des assurés au rachat par association de facteurs discriminants au rachat. La méthode de segmentation appelée ici est une méthode d'approximation linéaire permettant de sonder dans un premier temps l'existence de relations simples entre les variables ;
- * Le taux de rachat est ensuite désigné comme **indicateur de l'intensité du rachat**. Ce n'est plus tant le rapprochement entre variables qui est mis en avant mais la contribution de chaque critère dans la décision du rachat.
 - Suite à la segmentation du portefeuille, la base de données est délestée des variables non significatives (au sens du portefeuille étudié) dans l'étude du rachat. Cela permet alors d'expliquer la construction du taux de rachat grâce à des techniques d'approximation non-linéaire : **arbres de régression** et **réseaux de neurones**.
 - Les méthodes d'approximation non-linéaire utilisées sont incontournables dans l'analyse de données complexes mais sont particulièrement sensibles et instables. Pour améliorer la qualité des modèles construits, deux méthodes d'**agrégation de modèles** sont ensuite utilisées et comparées : le Bagging et le Boosting.
- * Enfin, pour vérifier la pérennité et la cohérence des modèles construits, nous les appliquons, dans le cadre d'une étape de **validation**, à des données récentes.

Pour atteindre ces objectifs, nous faisons appel à la notion d'**apprentissage** statistique. La théorie de l'**apprentissage**, apparaît dans les années 60 avant d'être développée par Vapnik [76] en 1999. Les méthodes d'apprentissage s'inspirent des capacités cérébrales humaines à assimiler des informations pour les restituer ultérieurement ou pour les généraliser à des cadres différents.

Il existe plusieurs types d'apprentissage : supervisé, non/semi/partiellement supervisé, par renforcement,... Nous présentons ci-après les notions d'apprentissage supervisé et non-supervisé dans la mesure où, dans la suite de ce mémoire, nous faisons appel à des techniques attachées à celles-ci.

Apprentissage supervisé

Il est question d'apprentissage supervisé lorsqu'une étude vise à expliquer une variable (la variable d'intérêt) par d'autres variables de ladite étude.

Soient X_1, \dots, X_n, Z les variables d'une base de données, Z correspondant à la variable d'intérêt. L'objectif de l'apprentissage est de trouver une fonction f telle que $Z = f(X_1, \dots, X_n) + \eta$ où η représente le bruit associé à l'approximation de X_i par f .

L'apprentissage supervisé consiste donc à assimiler la structure de la base de données pour construire un modèle reproduisant cette structure. Ce modèle est alors capable de reproduire les valeurs de la variable d'intérêt ou du moins, de les approcher de très près.

La robustesse d'une méthode d'apprentissage est éprouvée en testant ses capacités de généralisation. La généralisation consiste à soumettre au modèle une fraction de la base de données pour laquelle il doit "prédire" les valeurs de la variable d'intérêt, en respectant un éventuel critère d'arrêt. L'erreur de prédiction évalue alors la fiabilité de la méthode.

Exemples : Scoring, Méthode des centres mobiles, Arbres de régression, perceptron multi-couches,...

Apprentissage non-supervisé

L'apprentissage non-supervisé s'apparente à une cartographie de l'espace des variables. Aucune variable n'est mise en avant a priori. L'espace des variables est découpé selon un critère prédéfini. L'apprentissage non-supervisé permet de rassembler les individus d'une base en groupes homogènes au regard d'un certain facteur. Les classes créées sont intra-homogènes mais sont distinctes les unes des autres.

Exemples : Analyse factorielle, Cartes de Kohonen, Classification Ascendante Hiérarchique,...

Troisième partie

Description du portefeuille

Chapitre 3

Périmètre de l'étude

Au sein d'une entreprise, les sources d'information sont multiples. Cependant, lorsque migrations ou erreurs humaines interviennent, les données peuvent rapidement devenir impropres à l'utilisation. En ce sens, le traitement des données est une étape cruciale : des techniques actuarielles révolutionnaires appliquées à des données erronées mèneront à des résultats inexploitable.

Par ailleurs, tout traitement de données est justifié par le dessein poursuivi par une étude.

La gestion des données manquantes, les hypothèses d'homogénéité du portefeuille ou la précision numérique des montants manipulés ne sont pas abordés de la même façon selon l'objectif que l'on se donne. Dans le cas d'une analyse comportementale, il faut déceler des tendances plus ou moins marquées. Pour cela, la profusion des données et la cohérence de la segmentation du portefeuille sont nécessaires.

Ce chapitre a donc vocation à expliquer les démarches suivies dans la manipulation des données au regard de l'objectif poursuivi.

3.1 Base de données

Les données relatives aux contrats d'Épargne de Natixis Assurances ont connu quelques périodes de migration plus ou moins étalées dans le temps, la plus notable datant de 2010. Aujourd'hui, l'ancien système de gestion n'abrite plus que les contrats collectifs et les contrats clôturés avant 2010. Le nouveau système de gestion des contrats est repris par GraphTalk A.I.Aⁱ. S'y trouve notamment toute la base de données de l'entité depuis la migration de 2010.

Le portefeuille étudié est celui de Natixis Assurances. Il est en particulier question des contrats d'Épargne sujets à rachat entre 2005 et 2010 et en 2013.

Dans cette étude, nous nous restreignons aux seuls cas des contrats individuels : nous mettons de côté les IFC (Indemnités de Fin de Carrière), Article 39, Article 82, contrats Madelin et PERP. Bien que le rachat des contrats collectifs soit légalement autorisé dans des cas dérogatoiresⁱⁱ, sa proportion dans le portefeuille étudié reste négligeable comparée à celle des contrats individuels.

i. Plateforme moderne de gestion des produits d'assurance individuelle et collective, éditée par CSC

ii. Dans le cas du rachat d'un contrat "Article 83", le rachat, sur la totalité du capital et en exonération d'impôt sur le revenu, est possible dans les cas suivants :

- Expiration des droits de l'assuré aux allocations chômage ;
- Cessation d'activité non salariée de l'assuré à la suite d'un jugement de liquidation ;
- Invalidité de l'assuré (de 2^{ème} et 3^{ème} catégories, article L. 341-4 du code de la Sécurité Sociale) ;
- Décès du conjoint ou du partenaire PACS ;

3.1.1 Variables de l'étude

Les variables retenues dans cette étude sont celles qui nous semblaient refléter le plus fidèlement l'état du portefeuille sans présenter une proportion excessive de données manquantes et/ou aberrantes (traitées par ailleurs lorsqu'elles ne sont pas légion (Cf section 3.2.1.3)).

Caractéristiques de l'assuré

- **Âge** : Les âges des assurés sont calculés par différences de millésimes. Le portefeuille des contrats rachetés est constitué d'assurés âgés entre 1 an (cas des PEE(Plan Epargne Enfant)) et 105 ans, avec une moyenne d'âge d'environ 60 ans ;
- **Localisation géographique** : Cette donnée, très peu souvent renseignée, va être éliminée de cette étude ;
- **Genre** : Distinction homme/femme ;
- **Segment de clientèle** : Le service Marketing associe chaque assuré à l'un des 4 segments de clientèle suivants : Aisé, Patrimonial, Fortuné et Grand Public. Cette segmentation est établie en référence aux encours bancaires et assurantiels (confondus) des assurés. La figure 3.1 explique cette répartition.

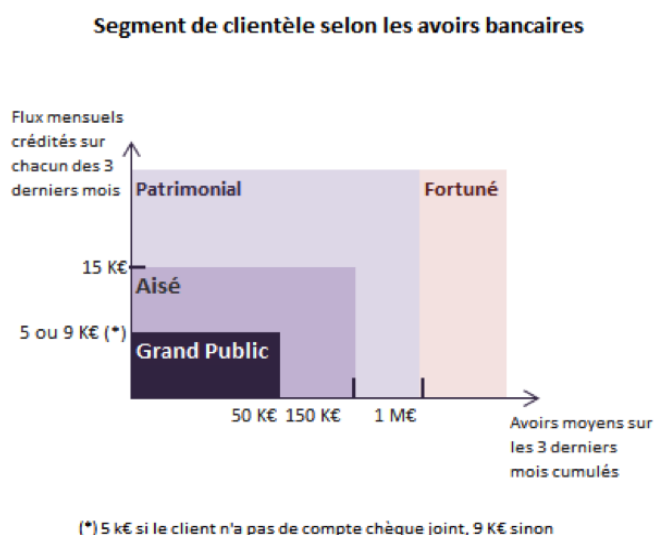


FIGURE 3.1 – Définition de classes d'assurés à partir de leurs encours bancaires et assurantiels

Spécificités du contrat

- **Montants et types de rachats** : Les montants de rachats sont liés à la date où le rachat est effectué et sont associés au type de rachat concerné (total, partiel libre ou programmé) ;
- **Provisions Mathématiques (PM)** : Les PM sont celles de l'année du rachat : les montants de PM ayant été impossibles à obtenir à un niveau mensuel pour les années comprises entre 2005 et 2010. C'est la raison pour laquelle l'étude des taux de

· Situation de surendettement de l'assuré.

rachat est faite ici à un niveau annuel. Par la suite, la notion de PM sera délaissée en faveur de celle d'assiette de rachat. Cette quantité est définie plus loin (Cf. section 3.2.3);

- **Gamme de produit :** En excluant les contrats type IFC, Article 39, Article 82, Madelin et PERP, seuls les contrats individuels sont retenus comme base de cette étude. Ces contrats sont divisés d'un point de vue commercial en produits destinés au Grand Public et en produits dédiés à la Gestion Patrimoniale;
- **Ancienneté du contrat :** A l'instar du calcul de l'âge de l'assuré, c'est également par différence de millésimes que les anciennetés des contrats sont déterminées. Les anciennetés de contrats varient entre 0 et 28 ans, avec une moyenne d'environ 8 ans;
- **Abattement fiscal :**
Les cotisations versées, pour celles dépassant un certain seuil, subissent une fiscalité différente selon l'âge auquel le souscripteur du contrat les a versées, sans que les intérêts capitalisés ne soient imposésⁱⁱⁱ. Le détail de cette fiscalité est développé en annexe VI.

Pour exhiber l'impact de la fiscalité sur la décision de rachat, le portefeuille est découpé selon les 3 catégories fiscales suivantes :

	Âge à la souscription	Âge actuel
Cat 1	< 70 ans	> 70 ans
Cat 2	> 70 ans	> 70 ans
Cat 3	< 70 ans	< 70 ans

FIGURE 3.2 – Catégories fiscales

- **Type de contrat :**
 - ✘ **Les produits Monosupport € :**
Les contrats Monosupport € sont principalement investis dans des obligations. Ils garantissent à l'assuré de toucher le montant de son épargne capitalisée au terme prévu par le contrat et ce, sans diminution possible du versement initial. L'avantage de ces contrats est leur côté sécuritaire et la liquidité du fonds. En revanche, les seules valorisations appliquées au capital sont le TMG (non nécessairement servi sur tous les contrats Monosupport €) et le versement d'une participation aux bénéfices (dont la distribution par l'assureur peut être faite dans un délai de 8 ans), ce qui représente un rendement assez limité en comparaison avec celui que peuvent apporter des contrats Multisupports;
 - ✘ **Les produits Multisupports :**
Une partie des contrats Multisupports est investie sur un fonds euros, l'autre sur des unités de compte. La partie en euros a un fonctionnement similaire à celui des contrats Monosupport €. La partie en unités de compte s'appuie sur des supports "risqués". Ces supports sont variés : OPCVM, SCI, SCPI, ... L'assuré choisit les unités de compte sur lesquelles il veut investir son épargne. Seul le nombre de parts des unités de compte sur lesquelles porte l'investissement est garanti par l'assureur. Dans ce genre de contrats, c'est l'assuré qui porte le risque de moins-value potentielle qui peut affecter son investissement.

iii. Le régime fiscal de l'assurance vie; *FFSA*

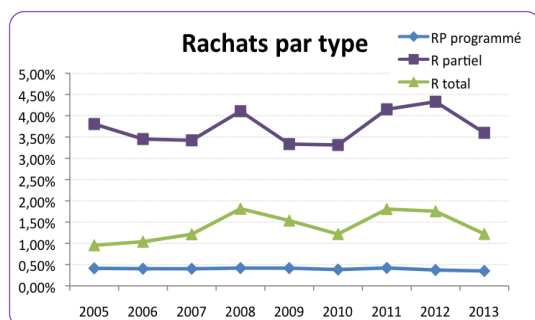
- **Type de support** : Il s'agit du support utilisé lorsque le contrat racheté est un contrat multisupports. Cette donnée n'est pas exploitée dans ce qui suit ;
- **Canal de distribution** : Cette donnée est rattachée au réseau de distribution gérant le contrat de l'assuré. Cette donnée, peu souvent renseignée, n'est pas exploitée ;
- **Taux Minimum Garanti (TMG)** : Dans le cas des contrats Monosupport €, le contrat mentionne la présence ou non d'un TMG et sa valeur s'il est proposé ;

Les variables retenues font l'objet d'une sélection dans les parties dédiées à l'analyse. Avant d'aborder cette partie, les prochaines sections présentent une première vue du portefeuille et expliquent les retraitements préalables effectués sur les données.

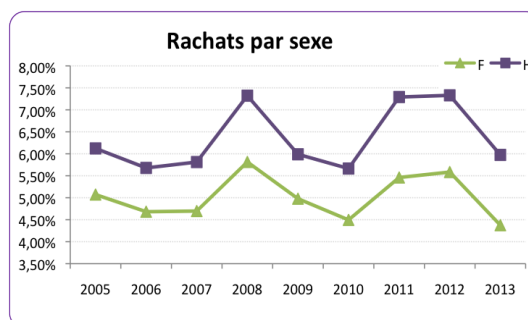
3.1.2 Premières observations :

Dans un premier temps, nous présentons l'évolution des taux de rachats sur des contrats agrégés selon certains critères : le type de rachat, le sexe de l'assuré, la gamme de produit du contrat racheté et la tranche d'âge de l'assuré. Cette dernière catégorie est définie grossièrement ici et fait l'objet d'une segmentation plus fine dans les chapitres suivants.

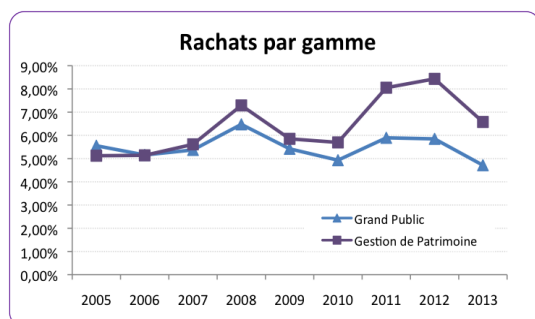
Les résultats présentés exposent également l'évolution des taux de rachats en 2011 et 2012. Du fait de la migration informatique, les données de ces deux années ont été retraitées indépendamment de cette étude. Les résultats des retraitements de ces deux années sont directement utilisés ici sans servir de base pour l'analyse des prochains chapitres.



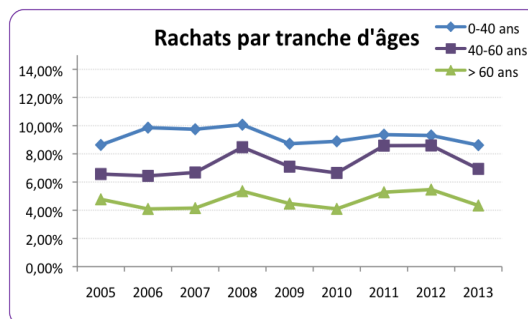
(a) Taux de rachat par Type de Rachat



(b) Taux de rachat par Sexe de l'assuré



(a) Taux de rachat par Gamme de produit



(b) Taux de rachat par Âge de l'assuré

Evolution des taux de rachats par type de rachat : (Figure 3.3a)

Le taux de rachat partiel est le taux le plus élevé (3,2% et 4,4%) qu'on ait plus observer sur la période étudiée (entre 0,9% et 1,8% et pour le rachat total et autour de 0,5% pour le rachat partiel programmé). En effet, les rachats partiels libres permettent au souscripteur du contrat d'éviter de clôturer son contrat, continuant ainsi à bénéficier de son antériorité fiscale ;

Outre les intervalles entre lesquels évoluent les taux pour le rachat total et le rachat partiel libre, la différence la plus remarquable au niveau du graphique est sans doute le sens de l'évolution entre 2005 et 2007. Alors que le taux de rachat total augmente, celui du rachat partiel diminue.

L'évolution du taux de rachat partiel programmé stagne, elle, autour des 0,5%, quelle que soit l'année.

Evolution des taux de rachats par sexe de l'assuré : (Figure 3.3b)

Les taux de rachat des hommes et des femmes ont exactement la même tendance. Cependant, alors que les taux de rachat des femmes sont compris entre 4,5% et 5,8%, ceux des hommes varient entre 5,5% et 7,4%. Ce mimétisme sera examiné dans la section dédiée à l'analyse des données 4.

Evolution des taux de rachats par gamme de produit : (Figure 3.4a)

En 2005 et en 2006, les taux de rachat des produits destinés au Grand Public dépassent ceux de la Gestion Patrimoniale, bien que restant globalement assez proches (entre 5% et 5,7%). La tendance s'inverse à partir de 2007 : les taux de rachat patrimoniaux atteindront même 8,8% en 2012.

L'explication la plus intuitive est la méfiance des investisseurs après la crise des subprimes. Les assurés les plus aisés ont maintenu leur défiance vis-à-vis de l'univers financier, d'où l'écart entre les taux de rachat des différentes gammes de produits.

Evolution des taux de rachats par âge de l'assuré : (Figure 3.4b)

Nous observons ici les taux de rachat de 3 générations différentes d'assurés. Plus un assuré est jeune, plus son taux de rachat est élevé. Les tendances, de hausse ou de baisse, ne sont pas très significatives. Nous notons qu'entre 2008 et 2010, tous les taux de rachat diminuent puis augmentent jusqu'en 2012 avant de baisser à nouveau.

Ne soulignant pas de mouvements brusques, ces courbes n'appellent pas une explication économique particulière, du moins pas à ce stade.

Pics de rachat et baisse des taux de rachat après 2012

Les 4 précédents graphiques mettent en avant des pics de rachat aux mêmes périodes : en 2008, en 2011 et en 2012. Alors que les premiers sont dus à la crise des subprimes de 2007, les seconds sont dus à l'instabilité des périodes de migration informatiques qu'a traversé l'entreprise.

Le retard de manifestation du rachat en 2008 s'explique par un réflexe différé des investisseurs. Face à un univers fragilisé où les conséquences économiques ont été aussi brusques que spectaculaires, la léthargie a rapidement laissé place à un comportement grégaire massif.

Enfin, notons que les taux de rachat, tous critères confondus, baissent après 2012 pour retrouver les valeurs autour desquelles ils évoluaient, hors périodes de migration et de crise financière.

3.2 Retraitements des données

Dans cette partie, nous expliquons les principaux retraitements appliqués aux données et la construction de la variable "taux de rachat".

3.2.1 Imputation des données manquantes

La question des données manquantes/aberrantes offre de nombreuses pistes de réflexion. La nature des valeurs manquantes/aberrantes guide souvent le choix de la méthode de retraitement à appliquer à ces données.

3.2.1.1 Valeurs aberrantes

Une valeur est aberrante lorsqu'elle est jugée atypique. Il peut s'agir de valeurs excessivement faibles/élevées au regard des valeurs moyennes prises par une variable, de quantités négatives alors que l'on s'attendrait à croiser des nombres strictement positifs (ou inversement), de valeurs exprimées dans une unité de mesure inadéquate,...

De telles valeurs ne sont pas légion dans le portefeuille : elles représentent environ 0,03% de l'ensemble du portefeuille sur le périmètre d'étude. Par ailleurs, les valeurs extrêmes qui ont été croisées ne sont pas assimilables à un rachat exceptionnel : elles sont très ponctuelles, ne se manifestent pas nécessairement dans les périodes de crises financières et ne caractérisent pas des contrats aux particularités remarquables.

Par précaution, les données aberrantes sont considérées comme des données manquantes.

3.2.1.2 Données manquantes

Les données manquantes ont des sources diverses : non saisie informatique, non disponibilité de l'information, migration informatique, valeurs aberrantes supprimées à des fins d'analyse,... Pour les retraiter convenablement, il faut les détecter, comprendre à quelles raisons leur présence est due, quelles variables elles concernent et dans quelle mesure leur retraitement au niveau de la base globale de données.

a. Types de données manquantes :

Little et Rubin(1987) [22] ont proposé plusieurs façons d'appréhender les données manquantes. Elles peuvent, selon ces auteurs, être distinguées selon leur répartition parmi l'ensemble des données :

- * Les données sont manquantes de façon complètement aléatoire (Missing Completely At Random (MCAR)) si l'absence de données pour une variable est indépendant des autres variables du jeu de données. L'absence de données ne dépendrait ainsi que d'éléments exogènes ;
- * Les données sont manquantes aléatoirement (Missing At Random (MAR)) si cette absence peut être liée aux valeurs observées d'autres variables du jeu de données ;
- * Le cas des données manquantes de façon non aléatoire (Missing Not At Random (MNAR)) correspond au cas où l'absence de données est directement liée à la nature de la variable étudiée. Autrement dit, si les données manquantes d'une variable ne peuvent être expliquées que par les données également manquantes d'autres variables du jeu de données.

Une illustration 3.5 de ces liens pourrait être la suivante : soit un jeu de données de 4 variables (X_0, X_1, X_2, X_3), où chacune présenterait des données manquantes, et soit Z l'ensemble des facteurs exogènes à la base de données.

Nous nous intéressons en particulier aux données manquantes de X_0 . Le tableau présenté ci-dessous lie l'absence de données de X_0 à la nature de la variable X_0 , aux autres variables du jeu de données et à des facteurs exogènes :

	Données manquantes de X_0		
	MCAR	MAR	MNAR
X_0	X	X	✓
X_1	X	✓	X
X_2	X	X	X
X_3	X	✓	X
Z	✓	X	X

FIGURE 3.5 – Lien entre données manquantes et données renseignées d'un jeu de données rudimentaire

L'illustration de Schafer et Graham dans le cas de données manquantes univariées^{iv} permet de comprendre cette distinction. Dans les lignes suivantes, X est assimilable à X_0 et Y à l'un des vecteurs X_1 ou X_2 ou X_3 :

Soient X la variable d'intérêt, Y une variable du même jeu de données que X , Z une variable exogène et R l'indicatrice associée à l'absence ou non de données ($R=1$ lorsque les valeurs sont absentes et $R=0$ indique qu'aucune donnée n'est manquante).

Les observations de X sont naturellement liées à celles de Y puisqu'elles font partie du même jeu de données.

Dans le cas :

- * MCAR, les données manquantes ($R=1$) sont dues à des facteurs exogènes (Z) ;
- * MAR, le facteur exogène (Z) intervient toujours mais les données manquantes ($R=1$) sont aussi imputables à la nature des observations (X) ;
- * MNAR, les observations (X), les variables (Y) et les facteurs exogènes (Z) sont autant de causes d'existence de données manquantes ($R=1$).

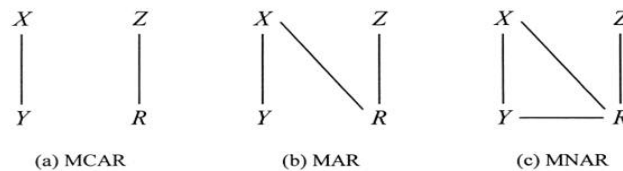


FIGURE 3.6 – Les données manquantes vues par Schafer et Graham

En formulant ces situations mathématiquement, si on note $L = (L_{obs}; L_{manq})$ l'ensemble des données, alors les données manquantes sont :

- * MCAR si $\mathbb{P}(R|L) = \mathbb{P}(R)$;
- * MAR si $\mathbb{P}(R|L) = \mathbb{P}(R|L_{obs})$, $\forall Y_{manq}$;
- * MNAR si la loi de R dépend de Y_{obs} et de Y_{manq} .

iv. Une valeur manquante est univariée si, pour une variable donnée, seule une observation est renseignée.

A chaque type de données manquantes correspond un traitement précis. Les lignes suivantes expliquent les méthodes de traitement des données manquantes et celles adoptées dans le cas des données de cette étude.

b. Gestion des données manquantes :

Les données manquantes sont traitées par l'une des méthodes suivantes :

*** Elimination des contrats aux données incomplètes :**

Cette mesure radicale ne doit être adoptée que si, en présence d'une base très fournie, les effectifs concernés sont peu nombreux et qu'ils ne concernent pas (dans le cas d'un assureur par exemple), des encours importants. Ils peuvent en revanche, être retraités séparément du reste de la base.

Cette méthode est à bannir pour toute étude où l'aspect quantitatif des données doit être mis en avant.

Par ailleurs, c'est la Loi Forte des Grands Nombres qui donne sa légitimité à la suppression de données. En effet, le tirage aléatoire d'un nombre important de données contient la même information que l'ensemble du portefeuille :

Loi Forte des Grands Nombres (de Kolmogorov)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires indépendantes, intégrables et de même distribution, alors :

$$\bar{X} = n^{-1} (X_1 + \dots + X_n) \xrightarrow[n \rightarrow +\infty]{p.s.} E[X_1]$$

*** Remplacement des données manquantes :**

Cette méthode consiste à remplacer les données manquantes par des valeurs qui transmettent quasiment la même information. Il peut s'agir de la moyenne ou d'une combinaison des données observées, la valeur la plus récurrente chez des individus similaires, la dernière donnée connue, le plus proche voisin, d'une valeur évaluée par un modèle de régression locale, ...

Ces méthodes d'imputation font l'objet de certaines critiques : l'article de Schafer et Graham sur les méthodes d'imputation des données manquantes [24] condamne le manque de précision qu'entraîne l'adoption de ces règles. Pour mesurer ce biais, les auteurs recommandent de calculer une marge d'erreur aléatoire présentant la fuite d'information inhérente à ces méthodes.

Enfin, certaines méthodes plus complexes reproduisent des résultats particulièrement satisfaisants en terme de réduction de la marge d'erreur aléatoire : l'imputation multiple, l'algorithme EMB (combinaison d'un algorithme de maximum de vraisemblance et d'une approche bootstrap),...

*** Maintien des valeurs non renseignées :**

Moyennant l'utilisation d'un logiciel pouvant réaliser des analyses en présence de données manquantes, cette possibilité est envisageable.

En effet, certaines méthodes sont, par construction, adaptées à la gestion des données manquantes. C'est en particulier le cas des **arbres de régression CART** qui s'appuient sur le principe des divisions de substitution (voir l'article de Feelders sur les divisions de substitution [29]). Une donnée manquante n'est tout simplement pas considérée comme candidate potentielle pour la division optimale d'un noeud.

3.2.1.3 Gestion des données manquantes dans le portefeuille de Natixis Assurances :

Dans la mesure où les données manquantes concernent toutes les variables, nous appliquons à nos bases quatre méthodes particulièrement adaptées à l'imputation de données hétérogènes manquantes : LOCF (Last Observation Carried Forward), kNN (Méthode des plus proches voisins), MICE (Multiple Imputation by Chained Equations) et l'imputation par Forêts aléatoires.

a. Méthodes retenues

LOCF (Last Observation Carried Forward)

Cette méthode d'imputation stationnaire est considérée comme étant la plus rudimentaire de toutes. Elle consiste à remplacer la valeur manquante par la dernière donnée connue de la variable.

kNN (Méthode des plus proches voisins)

La méthode des k plus proches voisins, mise en place par Troyanska et al. (2001) [53], s'appuie sur la distance observée entre les variables. C'est habituellement la distance euclidienne ou la distance de Mahalanobis qu'elle emploie. Cette méthode affecte aux données manquantes la moyenne (ou la médiane,...) des valeurs de ses k plus proches voisins au sens de cette distance, k étant un paramètre défini à l'appréciation de l'utilisateur de la méthode.

MICE (Multiple Imputation by Chained Equations)

L'imputation de données manquantes par chaînage d'équations, également connue sous le nom de FCS (Fully Conditional Specification), repose sur les travaux de Royston(2004) [63] et d'Oudshoorn et al. (1999) [35]. Cette méthode raisonne sur l'étude au cas par cas des lois conditionnelles des variables d'un jeu de données. Elle peut être expliquée comme suit :

Nous disposons d'un jeu de données de n variables $T = (T_1, T_2, \dots, T_n)$ où chaque variable présente des valeurs manquantes. Dans la mesure où cette méthode évalue les distributions des variables une par une, nous nous intéressons à la variable (quelconque) T_i où $i \in \llbracket 1, n \rrbracket$ où

$$T_i = (T_{i_1}^{manq}, T_{i_2}^{manq}, \dots, T_{i_m}^{manq}) \cup (T_{i_1}^{obs}, T_{i_2}^{obs}, \dots, T_{i_l}^{obs}) \quad \text{avec } m + l = n$$

Nous notons dans la suite $T_{-i} = T \setminus T_i$

MICE suppose que la distribution de T est complètement dépendante de celle d'un vecteur de paramètres inconnus que nous noterons $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ Pour avoir la distribution de β , l'algorithme commence d'abord par obtenir la distribution a posteriori des β_i par itérations successives du type : $P(T_i | T_{-i}, \beta_i)$ où $i \in \llbracket 1, n \rrbracket$.

A la j -ème itération, l'algorithme déduit les distributions des β_i^j et des T_i^j grâce aux distributions marginales des données observées :

$$\beta_i^j \sim P(\beta_i | T_i^{obs}, T_1^j, T_2^j, \dots, T_{i-1}^j)$$

Nous obtenons ainsi la i -ème distribution de T à la j -ème simulation :

$$T_i^j \sim P(T_i | T_i^{obs}, T_1^j, T_2^j, \dots, T_i^j, \beta_i^j)$$

Le même raisonnement est emprunté pour toutes les variables du jeu de données.

Forêts aléatoires

Initiée par Stekhoven et Bühlmann (2011) [26], cette méthode se base sur le principe d'apprentissage (voir section apprentissage). A l'instar des arbres de régression, son principal atout est de détecter les relations non linéaires entre les variables et de les respecter lors de l'imputation.

Elle s'effectue en 8 étapes :

- ✖ L'algorithme commence par apprendre le contenu du jeu de données : le type de ses variables, les éventuels liens les unissant,...
- ✖ Une variable Z est aléatoirement choisie puis les données sont découpées en 4 catégories :
 - Les données manquantes de Z ;
 - Les données renseignées de Z ;
 - Les données manquantes du jeu de données à l'exclusion de celles de Z ;
 - Les données renseignées du jeu de données à l'exclusion de celles de Z ;
- ✖ L'initialisation de l'algorithme nécessite d'appliquer une première imputation grâce à une quelconque méthode de complétion ;
- ✖ Les composantes de Z sont triées par ordre croissant. A chaque fois que Z présente une donnée manquante :
 - Une forêt aléatoire va s'ajuster sur les données connues du jeu de données (celles de Z et les autres) ;
 - Le modèle va appliquer l'ajustement précédent aux données manquantes de Z .
- ✖ Grâce aux valeurs nouvellement évaluées de Z , le modèle va généraliser l'ajustement de l'arbre à l'ensemble du jeu de données ;
- ✖ Les données manquantes du jeu de données, à l'exclusion de celles de Z , sont évaluées en utilisant les valeurs nouvellement calculées de Z ;
- ✖ L'ajustement suivant s'intéresse aux données incomplètes de l'ensemble du jeu de données, hormis celles de Z (dans la mesure où elles viennent d'être complétées).
- ✖ L'algorithme reprend depuis la première étape jusqu'à ce qu'un critère d'arrêt défini par l'utilisateur de l'algorithme soit atteint.

b. Résultats obtenus

La démarche suivie est la suivante : nous supprimons toutes les lignes pour lesquelles des données manquantes sont relevées. Les bases obtenues serviront de modèle pour tester nos méthodes de complétion. Des données manquantes sont aléatoirement et progressivement générées dans ces bases de données. Ainsi, 7 bases de données avec entre 10% et 80% (avec 10% de données manquantes supplémentaires générées d'une base à la suivante). Les méthodes de complétion plus tôt présentées sont alors appliquées.

Une fois la complétion effectuée, les données imputées sont confrontées aux données initiales.

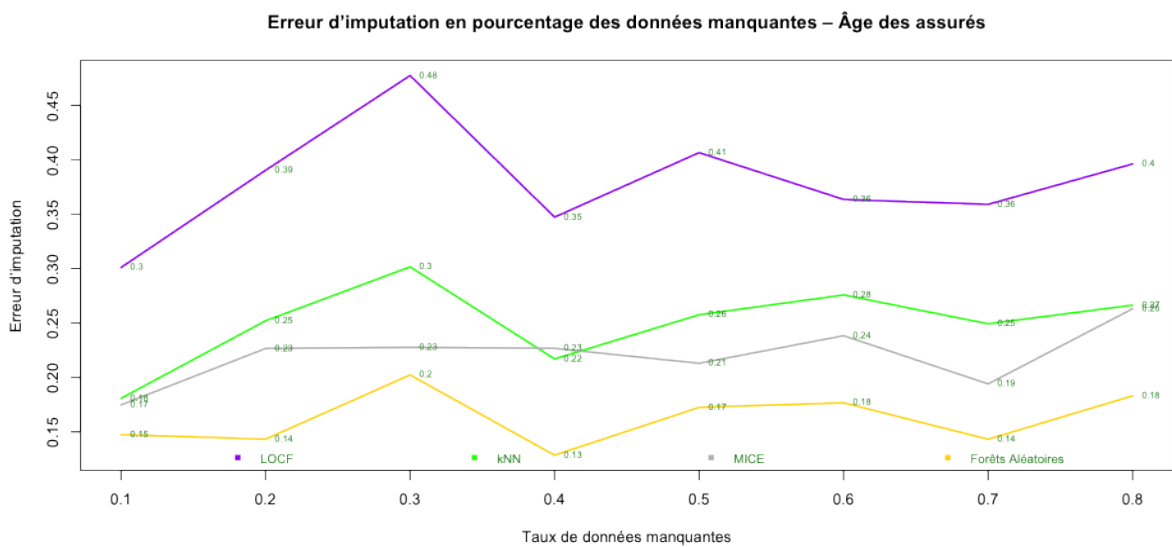
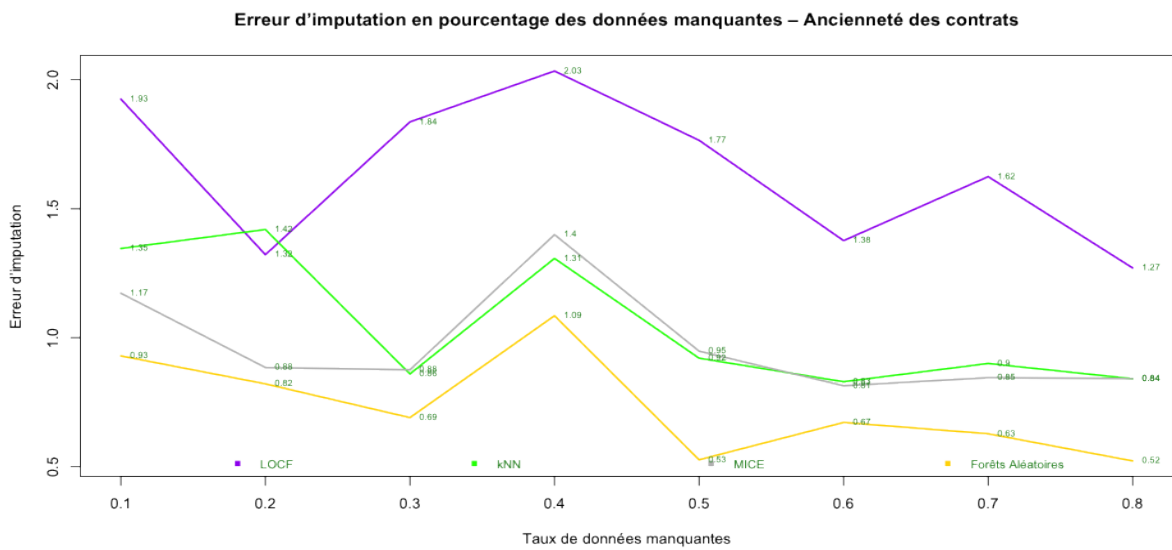
Dans le cas des données qualitatives, l'erreur d'imputation est évaluée en minimisant la distance de Hamming :

$$d_H(\hat{y}, \Sigma) = \min \left[\sum_i |\hat{y}_i - y_i| \right]$$

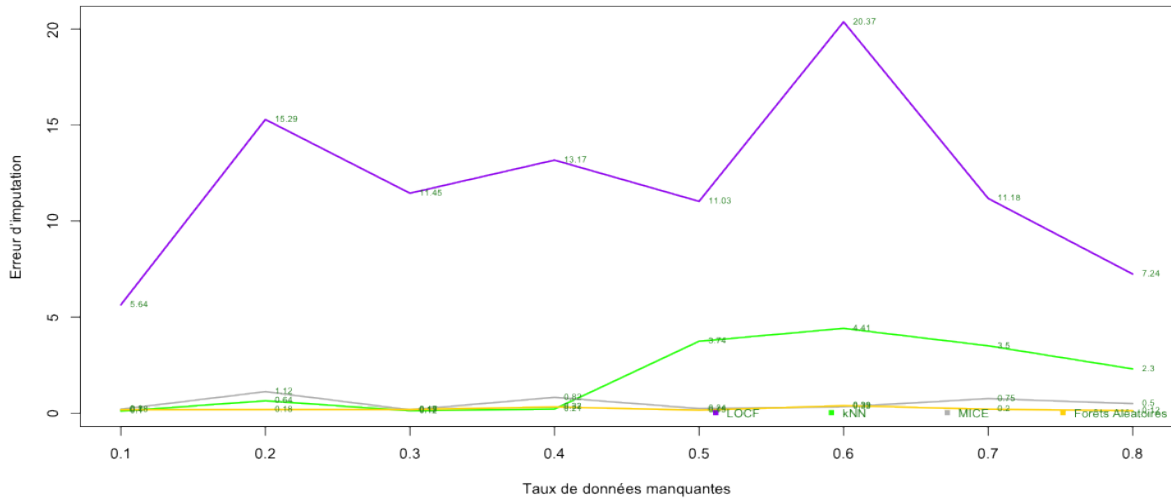
où Σ représente l'ensemble des observations renseignées, y_i les valeurs manquantes et \hat{y}_i les valeurs imputées pour les remplacer.

L'erreur d'imputation est, dans le cas des données quantitatives, évaluée en calculant la valeur absolue de la différence entre les données initiales et les données complétées.

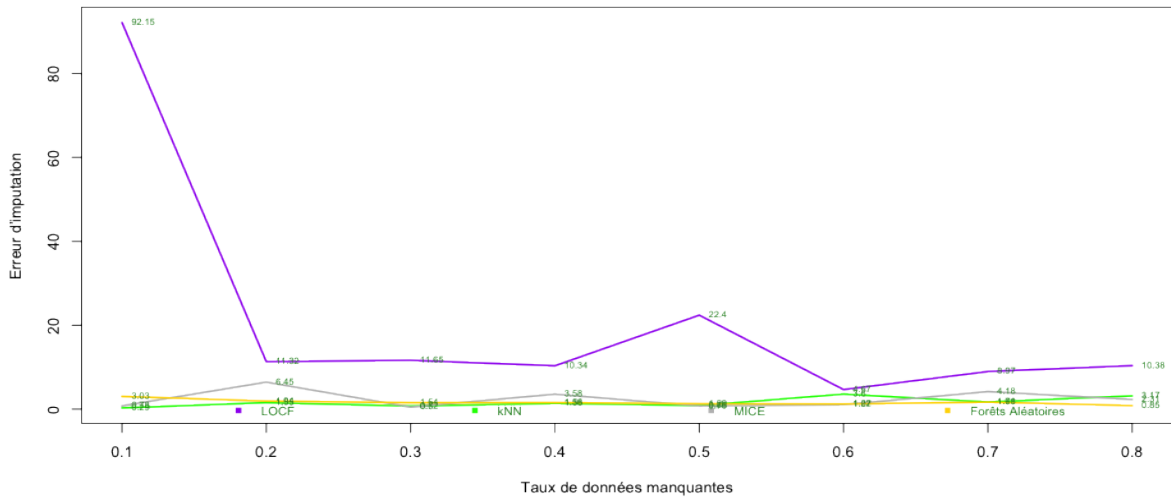
Le but est de chercher la méthode qui minimise l'erreur d'imputation afin de la généraliser à l'ensemble du jeu de données. Nous présentons ici les taux d'erreur obtenus pour différentes variables qualitatives et quantitatives :



Erreur d'imputation en pourcentage des données manquantes – PM moyenne



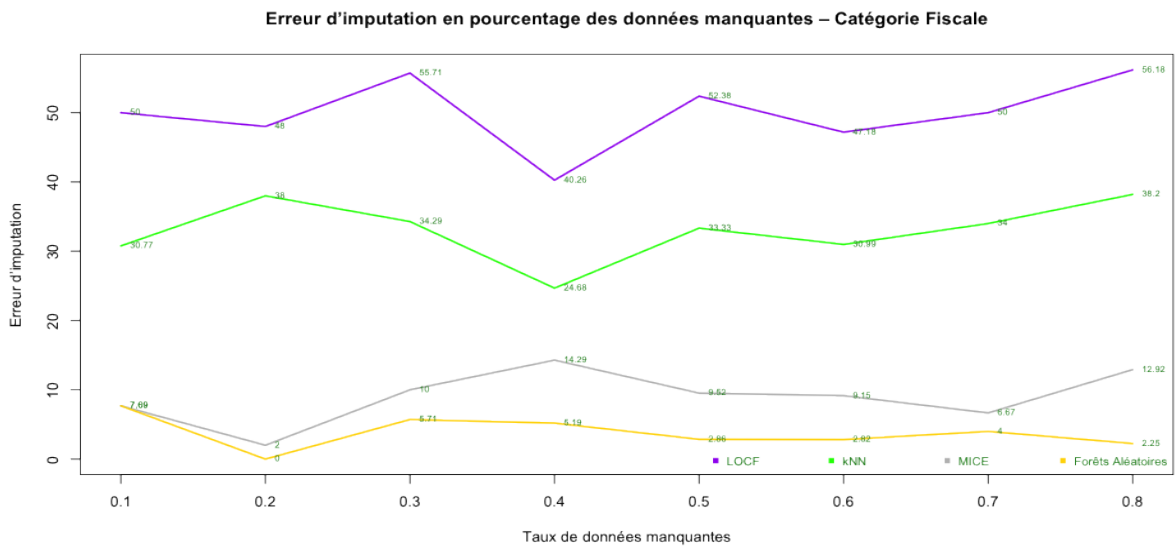
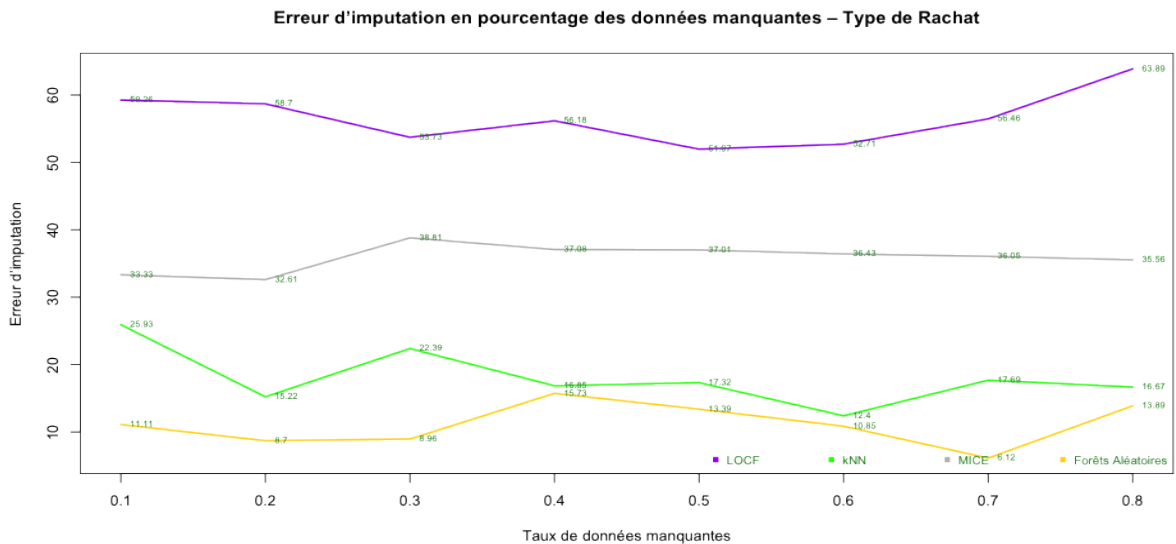
Erreur d'imputation en pourcentage des données manquantes – Montant des rachats



La méthode d'imputation par **forêts aléatoires**, représentée par la courbe jaune, est celle qui minimise le taux d'erreur d'imputation. Hormis pour le type de rachat où il atteint les 13%, il présente des valeurs extrêmement peu élevées comme le montrent les graphiques ci-dessus. Rappelons que, parmi les méthodes testées, c'est celle qui détecte les relations non linéaires. C'est un premier signe sur leur présence dans nos données.

La méthode **kNN** (utilisée ici avec $k = 5$), courbe verte, réduit l'erreur d'imputation en la confinant entre 0,32 % et 38,2 % (cas exceptionnel de la catégorie fiscale). Cette méthode repose sur le même principe que l'imputation par forêts aléatoires : la réduction de la distance entre les données observées et les données manquantes. Celle des forêts aléatoires est toutefois plus robuste puisque la convergence vers une solution optimale s'établit après un enchaînement de plusieurs étapes d'apprentissage des données.

Bien que l'imputation **MICE** ait été présentée comme un choix d'imputation théoriquement solide, elle semble ne pas convenir à notre jeu de données. L'imputation MICE est particulièrement adaptée au cas de données *MNAR*, or notre jeu de données présente



des données manquantes de façon particulièrement aléatoire, d'où l'inadéquation de cette méthode au cas présent.

La méthode **LOCF** est la moins probante : le taux d'erreur y est le plus élevé, toutes variables confondues. Cela n'est pas surprenant dans la mesure où, non seulement elle repose sur une relation linéaire entre les variables mais elle restreint cette relation à un voisinage extrêmement limité. Nous y retrouvons un pendant de la méthode des k plus proches voisins avec $k = 1$.

Enfin, nous évaluons la racine carrée normalisée de la somme des erreurs quadratiques moyennes NRMSE (Normalised Root Mean Squared Error) et le PFC (Proportion of Falsely Classified entries) pour chaque variable T_i :

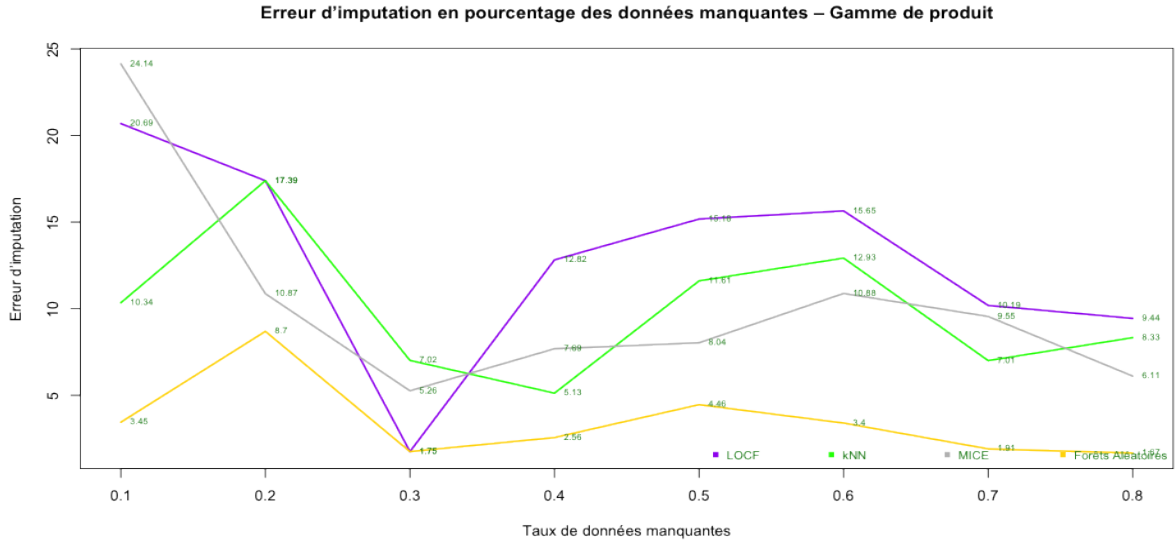


FIGURE 3.7 – Comparaison de l'évolution du taux d'erreur entre 3 méthodes d'imputation de données manquantes mixtes

$$\left\{ \begin{array}{l} NRMSE = \sum_N \sqrt{\frac{\mathbb{E}(T_{i_{initial}} - T_{i_{imput}})^2}{\mathbb{V}(T_{i_{imput}})}} \\ PFC = \sum_Q \frac{\mathbb{1}\{T_{i_{initial}} - T_{i_{imput}}\}}{Card_{valeursmanquantes}} \end{array} \right.$$

Dans ces formules, N (resp. Q) représente l'ensemble des variables numériques (resp. qualitatives).

Le graphique suivant représente le NRMSE et le PFC pour l'ensemble des variables qualitatives et quantitatives sur l'ensemble des jeux de données et ce pour les 3 méthodes d'imputation retenues :

L'imputation des données manquantes par la méthode des forêts aléatoires minimise l'erreur quadratique. Pour l'optimiser, et dans la mesure où cette méthode, par construction, nécessite une première complétion des données, nous l'appliquons à la base complétée par la méthode kNN pour le reste de cette étude.

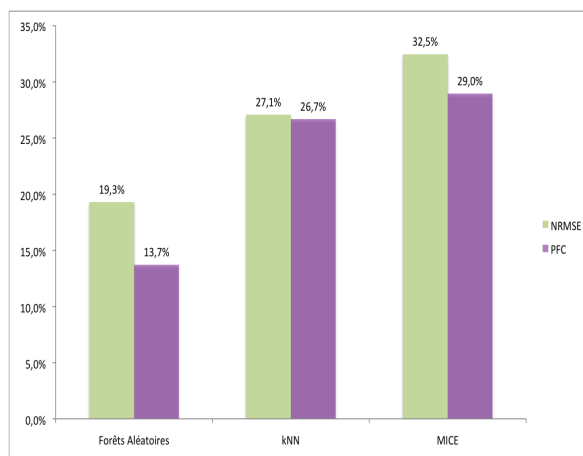


FIGURE 3.8 – Comparaison du taux d’erreur entre les méthodes d’imputations : kNN, MICE et MissForest

3.2.2 Règles d’agrégation

Cette section répond aux questions suivantes :

- Si un contrat est attaché à plusieurs assurés, auquel d’entre eux peut être associée la décision du rachat ?
- Si le contrat a subi plusieurs vagues de rachat en une année, quelle règle est adoptée pour le calcul du taux de rachat attaché à ce contrat ?

3.2.2.1 Contrats à deux têtes :

Les contrats à deux têtes sont soit des contrats conjugaux, soit des PEE (Plan Epargne Enfant). Dans le cas des :

- ✘ **Contrats conjugaux** : le contrat est affecté au conjoint le moins âgé, autrement dit à celui qui a le plus de chances de survivre à son conjoint et donc plus le temps de pouvoir exercer son droit au rachat ;
- ✘ **PEE** : tant que la majorité n’est pas atteinte, un assuré n’a pas la faculté de racheter son contrat, c’est son tuteur légal qui détient ce droit. C’est donc à l’assuré le plus âgé que revient la décision du rachat.

3.2.2.2 Rachats multiples :

Sur un exercice donné, un contrat peut connaître plusieurs mouvements de rachats :

- Plusieurs rachats partiels (libres et/ou programmés) ;
- Plusieurs rachats partiels (libres et/ou programmés) suivis d’un rachat total ;
- Un rachat total ;

L’approche choisie consiste à, **par exercice**, lier chaque contrat racheté à un type de rachat. Un contrat ayant subi plusieurs vagues de rachat par exercice doit donc voir ses montants de rachat agrégés et affectés à un type de rachat.

Cette étape est particulièrement importante pour la définition de l’assiette du rachat.

Ainsi, pour un exercice donné :

- ✘ En cas de vagues de rachats partiels (libres ou programmés, le "ou" étant exclusif), les sommes de rachats sont agrégées et le type de rachat maintenu est celui du rachat partiel libre ou programmé selon les cas ;

- ✘ Si un contrat a été sujet à au moins un rachat partiel programmé et à au moins un rachat partiel libre, les montants de rachats de l'exercice sont agrégés et le type de rachat retenu est celui de rachat partiel ;
- ✘ Si un contrat a été sujet à un rachat total et à au moins un rachat partiel (libre et/ou programmé), les montants de rachats sont agrégés et seule est préservée la mention de rachat total.

Le tableau suivant (Cf figure 3.9) résume ces règles :

Type de rachat par survenance	Rachat Total		x	x	x			
	Rachat Partiel	Libre	x	x		x	x	
		Programmé	x		x	x		x
Type de rachat maintenu pour l'exercice			Rachat Total			Rachat Partiel Libre		Rachat Partiel Programmé

FIGURE 3.9 – Règle d'affectation d'un type de rachat à un contrat en cas de vagues multiples durant un exercice

3.2.3 Taux de rachat

Le taux de rachat est un indicateur de l'intensité de manifestation du rachat. Il est défini comme suit :

$$\text{Taux de rachat} = \frac{\text{Montant du rachat}}{\text{Assiette du rachat}}$$

Cette étude s'intéresse aux profils d'assurés susceptibles de racheter leurs contrats. Le taux de rachat ainsi défini a donc été construit en pensant principalement à cette population.

Ainsi, cette quantité ne reflète que la proportion de la population d'assurés rachetant son contrat.

Dans le cadre de l'évaluation du risque de rachat auquel est soumis un assureur, cette quantité doit, pour être exploitable, être rapportée à l'encours global, i.e. à la provision mathématique des contrats rachetés et non rachetés.

Cette opération est notamment effectuée dans la partie dédiée à la modélisation du taux de rachat.

Le calcul du taux de rachat dépend de l'assiette du rachat :

- ✘ Dans le cas d'un **rachat total**, le taux de rachat est unitaire puisque l'assuré récupère toute son épargne (Montant du rachat = Assiette du rachat) ;
- ✘ Lors d'un **rachat partiel**, le calcul du taux de rachat dépend de l'assiette du taux. Cette dernière dépend de **l'existence ou non de versements dans l'année**. Ainsi, nous définissons :

$$\text{Assiette RP} = \begin{cases} \frac{\text{PM d'ouverture} + (\text{PM de clôture} + \text{Montant du rachat})}{2} & \text{Si Montant du rachat} < (\text{PM d'ouverture} + \text{PM de clôture}) \\ \text{PM de clôture} + \text{Montant du Rachat} & \text{Sinon} \end{cases}$$

La distinction de ces deux formules est nécessaire pour éviter des taux de rachat aberrants ($> 100\%$). Nous allons le démontrer.

Notons :

- La formule Assiette RP = $\frac{\text{PM d'ouverture} + (\text{PM de clôture} + \text{Montant du rachat})}{2}$: formule ① ;
- La formule Assiette RP = (PM de clôture + Montant du Rachat) : formule ②.

Nous disposons de 3 contrats sur lesquels s'est effectué un rachat partiel (le rachat partiel correspond à la définition exposée à la section 3.2.2.2). Ces contrats présentent les informations suivantes :

	1 ^{er} cas	2 ^{ème} cas	3 ^{ème} cas
PM d'ouverture	100	10	100
PM de clôture	20	20	10
Montant du rachat	110	100	110

FIGURE 3.10 – Evaluation du taux de rachat partiel par variation de l'assiette de calcul - cas pratique (1/2)

Dans le 1^{er} cas, la formule ① permet de lisser les primes éventuelles sur l'année alors que la formule ② a tendance à ne pas prendre en compte celles qui ont été faites avant la date du rachat.

Dans la mesure où cette étude porte sur des taux **annuels**, la formule ① est, dans ce cas, jugée plus prudente.

Le 2^{ème} cas illustre la légitimité de la formule ②. Un versement effectué dans l'année justifie les flux identifiés sur ce contrat. Les versements n'ont pas pu être remontés dans les bases de données de cette étude. C'est pour cela qu'ils n'ont pas pu être explicitement invoqués dans la formule de calcul du taux de rachat. Les versements doivent être implicitement pris en compte : c'est le rôle de la formule ②.

Le 3^{ème} cas correspond à la situation où le montant de rachat vaudrait la somme de la PM d'ouverture et de la PM de clôture. La formule ① conduirait à un taux de rachat de 100% alors que la formule ② amènerait à un taux de rachat d'environ 92%, plus cohérent avec un rachat partiel.

		1 ^{er} cas	2 ^{ème} cas	3 ^{ème} cas
PM d'ouverture		100	10	100
PM de clôture		20	20	10
Montant du rachat		110	100	110
Assiette du rachat	1 ^{ère} formule	115	65	110
	2 ^{ème} formule	130	120	120
Taux de rachat	1 ^{ère} formule	0,96	1,54	1
	2 ^{ème} formule	0,85	0,83	0,92
Somme des primes versées dans l'année		30	110	20

FIGURE 3.11 – Evaluation du taux de rachat partiel par variation de l'assiette de calcul - cas pratique (2/2)

Finalement, le taux de rachat se voit affecter la formule suivante :

$$\text{Taux de rachat} = \begin{cases} \frac{\text{Montant du rachat}}{\text{Assiette RP}} & \text{Si le rachat est partiel (libre ou programmé)} \\ 1 & \text{Si le rachat est total} \end{cases}$$

La section suivante propose une description linéaire des bases retraitées grâce à une méthode d'analyse où la nature des variables n'est pas altérée par une quelconque conversion : l'Analyse Factorielle des Données Mixtes.

Chapitre 4

Analyse Factorielle des Données Mixtes

Les bases de données volumineuses ont cela de commun que les interactions entre les variables, de quelque intensité qu'elles soient, y sont invisibles à l'oeil nu. L'analyse de données est une technique permettant de déceler certaines de ces interactions.

Ainsi, en projetant l'ensemble des observations sur des axes factoriels de manière à limiter la perte d'information, l'analyse de données permet de mettre en avant des populations au comportement homogène relativement à une ou à plusieurs variables.

4.1 Cadre et hypothèses

Il existe deux façons d'étudier des données mixtes : en altérant la nature de l'un quelconque des types de variables (ACM ou ACP) ou en préservant la nature de chaque variable (AFDM).

Le cadre décrit ci-après est commun aux trois types d'analyse présentées dans cette section (ACP, ACM et AFDM).

Soient :

- Σ l'ensemble du portefeuille étudié ;
- n le nombre d'individus dans le portefeuille ;
- L'espace \mathbb{R}^n est ici doté de la métrique D suivante :

$$D_{(i,j)} = w_i \mathbb{1}_{\{i=j\}}$$

où w_i désigne le poids de chaque individu.

Par simplification, chaque individu est affecté du même poids $\frac{1}{n}$, autrement dit :

$$\forall (i, j), D_{(i,j)} = D = \frac{1}{n} \mathbf{I}_d$$

- $X = (X_p)_{p \in \llbracket 1; P \rrbracket}$ le vecteur des variables quantitatives ;
- $Y = (Y_l)_{l \in \llbracket 1; L \rrbracket}$ le vecteur des variables qualitatives ;
- M_l le nombre de modalités de chaque variable qualitative l .

Les modalités s'écrivent donc comme des variables $(Y_{l,m_l})_{(l,m_l) \in \llbracket 1; L \rrbracket \times \llbracket 1; M_l \rrbracket}$

- S le nombre total de modalités, toutes variables à modalités confondues. Autrement dit :

$$\# \{ \text{Ensemble des variables} \} = \sum_{l=1}^L M_l + P = S$$

- Les observations associées à chaque variable sont notées :
 - ⊠ x_i^j pour la i -ème observation de la j -ème variable **quantitative** ;
 - ⊠ y_i^j pour la i -ème observation de la j -ème variable **qualitative**.
- n_{l,m_l} la fréquence de la modalité Y_{l,m_l} ;
- Le symbole $'$ à droite d'un vecteur (ligne) indique qu'il a été transposé ;
- Le symbole $\bar{}$ au-dessus d'une variable signifie qu'elle est centrée-réduite.

Le portefeuille initial prend donc la forme matricielle suivante :

$$\begin{aligned} \Sigma &= \begin{pmatrix} x_1^1 & \cdots & x_1^P & y_1^1 & \cdots & y_1^L \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^P & y_n^1 & \cdots & y_n^L \end{pmatrix} \\ &= \begin{pmatrix} x_1^1 & \cdots & x_1^P & y_1^{1,1} & \cdots & y_1^{1,M_1} & \cdots & y_1^{L,1} & \cdots & y_1^{L,M_L} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_n^1 & \cdots & x_n^P & y_n^{1,1} & \cdots & y_n^{1,M_1} & \cdots & y_n^{L,1} & \cdots & y_n^{L,M_L} \end{pmatrix} \end{aligned}$$

où, $\forall i \in \llbracket 1; n \rrbracket, \forall j \in \llbracket 1; L \rrbracket, \forall l \in \llbracket M_1; M_j \rrbracket$:

$$y_i^{j,l} = \begin{cases} 1 & \text{si l'individu } i \text{ présente la modalité } l \text{ de la variable qualitative } j \\ 0 & \text{sinon} \end{cases}$$

Autrement dit, $y_i^{j,l}$ représente la i -ème observation de la l -ème modalité pour la j -ème variable **qualitative**.

La dernière égalité correspond à un codage binaire des variables qualitatives. Ce type d'algorithme permet de convertir les variables à modalités en variables numériques. Le tableau obtenu est un **tableau disjonctif complet** qui indique si un individu présente ou non une modalité donnée.

4.2 L'analyse factorielle vue par les données mixtes

Dans cette section, ce qui nous intéresse, ce n'est pas tant la version classique des méthodes d'analyses de données (en particulier pour l'ACP et l'ACM) que l'approche qu'elles proposent dans le cas de données hétérogènes.

4.2.1 Analyse en Composantes Principales (ACP)

Tenenhaus(1977) [73] puis Escoffier(1979) [21] ont été les premiers à proposer le codage des variables qualitatives.

Cette méthode qui fait ses preuves dans des espaces de faible dimension, aboutit très souvent à des optimums locaux dans des espaces de dimension importante. Nous allons en formaliser les étapes.

4.2.1.1 Principe du codage optimal

L'ACP s'intéresse initialement à des variables numériques. Pour l'appliquer à un jeu de données mixtes, il faut optimiser le codage des variables à modalités.

La réalisation d'une ACP dans le cadre des données mixtes tient au choix d'une métrique. Ainsi, pour toute variable qualitative et toute modalité de cette variable, nous considérons une fonction de codage α_l qui, appliquée aux modalités $Y_{l,m}$, va permettre de se ramener à une variable quantitative.

La variable codée s'écrit alors :

$$\begin{aligned} H_l &= \alpha_l \circ Y_l \\ &= \sum_{m=1}^{M_l} \alpha_l(m) Y_{l,m} \\ &= \sum_{m=1}^{M_l-1} \alpha_l(m) \left(Y_{l,m} - \frac{n_{l,m}}{n_{l,M_l}} Y_{l,M_l} \right) \end{aligned}$$

Ces variables codées se voient imposer deux conditions pour limiter leur dégénérescence :

$$\begin{cases} \sum_{y_i^{j,l} \in \Sigma} H_l(y_i^{j,l}) = 0 \\ \sum_{y_i^{j,l} \in \Sigma} (H_l(y_i^{j,l}))^2 = n \end{cases}$$

L'objectif de l'ACP est alors de maximiser la quantité :

$$\sum_{i=1}^L \sum_{j=1}^{\gamma} \text{cor}^2(H_i, R_j) + \sum_{p=1}^P \sum_{j=1}^{\gamma} \text{cor}^2(X_p, R_j)$$

Ici, les $(R_j)_{j \in \llbracket 1, \gamma \rrbracket}$ désignent des variables **non corrélées** caractérisant Σ . Les R_j puisent leurs valeurs dans X et Y, autrement dit :

$$R_j \in \{X_1, \dots, X_P, Y_1, \dots, Y_L\}, \forall j \in \llbracket 1, \gamma \rrbracket$$

L'un des enjeux principaux de l'ACP est alors de déterminer γ , où $\gamma \leq P + L$.

Notons :

$$\begin{cases} \beta_l = [\alpha_l(m) \dots \alpha_l(M_l - 1)]' \\ \Theta_l = \left[Y_{l,1} - \frac{n_{l,1}}{n_{l,M_l}} Y_{l,M_l}, \dots, Y_{l,(M_l-1)} - \frac{n_{l,(M_l-1)}}{n_{l,M_l}} Y_{l,M_l} \right] \end{cases}$$

Ainsi, les variables codées H_l sont centrées réduites si et seulement si :

$$\begin{cases} \alpha_l \circ Y_l = \Theta_l \beta_l \\ \beta_l' \left(\frac{1}{n} \Theta_l' \Theta_l \right) \beta_l = 1 \end{cases} \quad (4.2.1)$$

Finalement, le problème de maximisation s'écrit :

$$\underset{\beta_1, \dots, \beta_L}{Max} \left[\sum_{l=1}^L \sum_{j=1}^{\gamma} cor^2(\Theta_l \beta_l, R_j) + \sum_{p=1}^P \sum_{j=1}^{\gamma} cor^2(X_p, R_j) \right]$$

où

$$\begin{cases} H_l \text{ est centrée-réduite, } \forall l \in \llbracket 1; L \rrbracket \\ \text{Les } R_j \text{ sont centrées-réduites et } \mathbf{non\ corrélées}, \forall j \in \llbracket 1, \gamma \rrbracket \end{cases}$$

L'algorithme de l'ACP repose sur des postulats nécessairesⁱ portant sur les variables nominales :

1^{ère} proposition :

Avec les notations introduites précédemment le maximum de la fonction $cor^2(\Theta_l \beta_l, R_j)$ est atteint sur les γ premières composantes principales réduites des L variables centrées-réduites $\alpha_l \circ Y_l$.

De plus, ce maximum est égal à la variance expliquée par les γ premières composantes principales.

2^{ème} proposition :

Supposons que les $(R_j)_{j \in \llbracket 1, \gamma \rrbracket}$ soient déterminées de telle sorte à ce qu'elles soient centrées-réduites et non corrélées. Alors, l'analyse canonique entre les variables formées par les colonnes de Θ_i et par les R_j permet d'obtenir des vecteurs $(\beta_j^*)_{j \in \llbracket 1, \beta \rrbracket}$ qui constituent des candidats pour le maximum de la fonction $cor^2(\Theta_l \beta_l, R_i)$.

Ces vecteurs, qui vérifient la deuxième condition de (4.2.1), sont en réalité les vecteurs canoniques associés à Θ_l et sont assimilés à la 1^{ère} corrélation canonique $\sqrt{\lambda_l}$.

Autrement dit, les λ_i correspondent à la **communauté** (i.e. à la qualité de la représentation) de $\Theta_l \beta_l$.

i. Les démonstrations de ces résultats sont données par Michel Tenenhaus dans *Analyse en composantes principales d'un ensemble de variables nominales ou numériques*, Revue de statistique appliquée, tome 25, n°2 (1977), p. 43-45

Ainsi :

$$Max_{\beta_1, \dots, \beta_L} \sum_{l=1}^L \sum_{i=1}^{\gamma} cor^2(\Theta_l \beta_l^*, R_i) = \sum_{i=1}^{\gamma} \lambda_i$$

3^{ème} proposition :

Soient $(e_i)_{i \in [1, \gamma]}$ les vecteurs canoniques de \mathbb{R}^{γ} , V la matrice de variance-covariance totale et V_{int} la matrice de variance-covariance inter-classes.

Si, $\forall l \in [1, L]$, $\forall i \in [1, \gamma]$, β_l et e_i réalisent :

$$\begin{aligned} \beta_l' \left(\frac{1}{n} \Theta_l' \Theta_l \right) \beta_l &= e_i' e_i \\ &= 1 \end{aligned}$$

Alors le maximum de $cor^2(\Theta_l \beta_l, R e_i)$ est atteint pour e_i^* qui est le vecteur propre associé à la plus grande valeur λ_i^* de $V^{-1} V_{int}$ (dans le cas présent, V est la matrice identité).

Le maximum est également atteint pour

$$\beta_l^* = \frac{1}{\sqrt{\lambda_l}} (\bar{R}(l, 1), \dots, \bar{R}(l, M_l))' e_i^* = \frac{1}{\sqrt{\lambda_l}} (\Theta_l' \Theta_l^{-1}) \Theta_l' R e_i^*$$

où $\bar{R}(i, j) = (\bar{R}_1(i, j), \dots, \bar{R}_{\gamma}(i, j))'$ est le vecteur moyenne des R_l sur les observations $x \in Y_i^{-1}(j)$

4.2.1.2 Etapes du codage

L'algorithme de codage s'appuie sur une itération effectuée sur les β_l^i et les $R_j^i \forall i \in \mathbb{N}$, $\forall l \in [1, L]$ et $\forall j \in [2, m]$.

Concernant la **solution initiale** β_l^0 , celle-ci peut être définie soit arbitrairement (en respectant les conditions (4.2.1)), soit, de façon plus élaborée, en posant :

$$\begin{cases} \overset{\circ}{\Theta} = [\Theta_1, \dots, \Theta_L, \bar{X}_1, \dots, \bar{X}_P]; \\ T = \text{Diag} [\Theta_1' \Theta_1, \dots, \Theta_L' \Theta_L, \bar{X}_1' \bar{X}_1, \dots, \bar{X}_P' \bar{X}_P] \\ = \text{Diag} [\Theta_1' \Theta_1, \dots, \Theta_L' \Theta_L, n, \dots, n] \end{cases}$$

La diagonalisation de la matrice $T^{-1} \Theta' \Theta$ permet d'obtenir les R_j .

Les vecteurs $\beta_j^{(0)}$ peuvent alors être calculés grâce à la proposition 2 II.

Supposons ensuite que le $(k-1)$ ^{ème} indice (i.e. β_l^{k-1}) vérifie les conditions (4.2.1). Cette hypothèse est vérifiée puisque l'étape précédente veille à ce que la construction de β_l^{k-1} respecte ces conditions.

Les R_j^k doivent alors représenter les composantes principales des variables $\Theta_1 \beta_1^{k-1}, \dots, \Theta_L \beta_L^{k-1}, \bar{X}_1, \dots, \bar{X}_P$, c'est-à-dire :

$$R_j^k = Max \sum_{l=1}^L \sum_{i=1}^P cor^2(\Theta_l \beta_l^{k-1}, \bar{X}_i)$$

Ainsi, $\forall j \in \llbracket 1, \gamma \rrbracket$, la valeur optimale des R_j^k correspond à la somme des valeurs propres de $\frac{1}{n}BB'$.

Nous retrouvons dans cette somme la variance des L variables centrées-réduites $\Theta_l\beta_l$, expliquée par les γ premières composantes principales. C'est bien ce qu'annonçait la proposition 1 II.

Quant aux β_j^k , ils doivent maximiser $cor^2(\Theta_l\beta_l, \bar{A}_l)$ tout en respectant la contrainte

$$\beta_l^k \left(\frac{1}{n} \Theta \Theta' \right) \beta_l^k = 1$$

Ici, $\forall j \in \llbracket 1, \gamma \rrbracket$ et comme l'avancéait la proposition 2 II, la valeur optimale des β_j^k est donnée par la somme des carrés de la 1^{ère} corrélation canonique.

Finalement, la condition nécessaire et suffisante pour que l'algorithme converge tient à la convergence simultanée des R_j^k et des β_j^k vers une même limite. Au demeurant, cette convergence unique mène plus probablement à un maximum local que global : c'est là la grande faiblesse des méthodes itératives.

4.2.1.3 Limites de la méthode

La première est d'ordre technique, la seconde relève davantage de l'interprétation des résultats graphiques de l'ACP.

Il faut d'abord signaler que les méthodes itératives ont un défaut considérable lorsqu'il est question de convergence des variables : la limite a de grandes chances de n'être atteinte que localement. Cela rend la généralisation des résultats à tout l'espace des observations particulièrement délicate.

D'autre part, la lecture graphique des résultats repose sur la proximité des centres de gravité des individus et des variables. Les individus aux comportements foncièrement différents (cas extrêmes) risquent d'être négligés dans cette analyse, surtout s'ils ne sont pas légion.

4.2.2 Analyse des Composantes Multiples(ACM)

L'idée d'exploiter l'ACM dans le cas de variables hétérogènes est attribuable à B. Escoffier [20]. Sa réflexion part de ce constat : une variable numérique prend ses valeurs dans un sous-espace de \mathbb{R}^n de dimension 1 alors qu'une variable modale puise ses valeurs dans un sous-espace de \mathbb{R}^n de dimension supérieure à 1 au sens large.

4.2.2.1 Logique du codage

Le principe de cette méthode est d'aborder les variables quantitatives du point de vue d'une ACP normée et les variables qualitatives de celui d'une ACM. La manipulation de chaque type de variable est, de plus, soumis à des règles précises.

Etapes :

Les variables sont traitées en deux temps, selon leur nature. Il s'agit :

- de coder les variables qualitatives grâce à un tableau disjonctif complet 4.1.

- d'affecter aux variables quantitatives deux valeurs : $0.5(1 \pm \bar{x}_i^j)$. Ces points sont, par construction, à égale distance du centre de gravité du nuage associé au portefeuille global.

Nous nous ramenons ainsi, à un coefficient $\frac{1}{n}$ près, à la métrique identité. Il est alors possible de localiser le centre de gravité G du nuage de l'ensemble des variables (qualitatives et quantitatives confondues) au niveau de la droite de vecteur directeur $(1, \dots, 1)$.

Objectif :

Le but principal de cette analyse est de déterminer les vecteurs de \mathbb{R}^n qui maximisent la quantité $\sum_{l=1}^L \cos^2 \alpha_l$ où α_l est l'angle entre chaque vecteur et le sous-espace dont dépend la $l^{\text{ème}}$ variable. Les vecteurs qui maximisent la somme des $(\cos^2 \alpha_l)_{l \in [1;L]}$ sont appelés **facteurs** de l'ACM.

L'inertie associée à chaque couple de points $(0.5(1 - \bar{x}_i^j), 0.5(1 + \bar{x}_i^j))$ vaut $\frac{1}{P+L}$. De fait, l'inertie de chaque facteur vaut $\frac{1}{P+L} \left(\sum_{l=1}^L \cos^2 \alpha_l + \sum_{l=1}^{2 \times P} \cos^2 \alpha'_l \right)$ où α'_l désigne l'angle entre chaque facteur et chaque variable numérique X'_p .

4.2.2.2 Avantages et Inconvénients

L'ACM permet donc de mettre en avant les variables contribuant le plus à l'inertie d'une base de données mixtes. Il suffit de considérer chaque variable de cette base, de suivre les étapes de codage expliquées plus haut puis d'évaluer la somme des carrés des corrélations entre cette variable et le reste des variables (préalablement traitées comme présenté plus tôt). Plus cette somme est élevée, plus forte est la contribution de la variable dans l'explication de l'inertie globale du portefeuille.

Ceci étant, dans cette méthode, le partitionnement des données numériques présente deux inconvénients majeurs : la perte d'information et le choix arbitraire (et donc non forcément rigoureux) du nombre de classes. L'approche théorique est intéressante mais n'apporte cependant pas de conclusions probantes sur des cas proposant un nombre de variables $S = P + L \geq 10$. Ceci est dû au recours à une quantité géométrique (le cosinus) qui a une fiabilité relative dans des espaces de grande dimension.

En outre, cette méthode reste très peu concluante dès que le nombre d'individus se chiffre en centaines. C'est en particulier le cas dans cette étude.

4.2.3 Analyse Factorielle des Données Mixtes (AFDM)

L'ACM et L'ACP proposent une base théorique intéressante pour aborder des données hétérogènes mais leurs lacunes respectives, notamment en présence d'un grand nombre d'individus ou de déséquilibre entre le nombre de variables qualitatives et quantitatives, rend leur application dans le cadre de l'Assurance Vie discutable.

En combinant les avantages de ces deux méthodes, Pagès [56] réussit à estomper leurs faiblesses. L'analyse à laquelle il a recours porte le nom d'**Analyse Factorielle des Données Mixtes (AFDM)**.

4.2.3.1 Principe de l'AFDM

L'AFDM est un procédé qui emprunte le raisonnement de l'ACM pour l'appliquer au cas d'une ACP (voir la figure 4.1). L'un des avantages de l'AFDM est sans doute le suivant :

Quel que soit le nombre de modalités qu'elle présente, si une variable qualitative est projetée sur le sous-espace engendré par le nuage de points de ses modalités, son inertie sera toujours unitaire.

Cela signifie que peu importe le nombre de modalités, une variable qualitative, dans ce contexte, aura toujours la même inertie qu'une variable quantitative. Il n'y a donc pas de déséquilibre de représentation des types de variables, ce qui constituait l'un des inconvénients majeurs dans l'usage de l'ACP et de l'ACM.

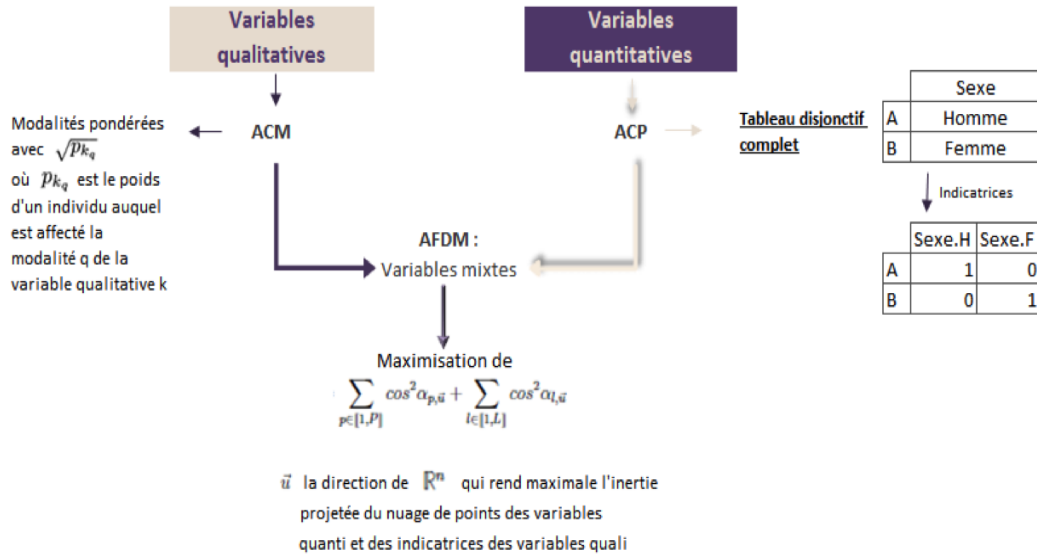


FIGURE 4.1 – Principe de l'AFDM

Formalisme de la méthode :

Nous avons présenté l'ACP et l'ACM dans le cas simplifié où les individus étaient associés à un même poids. Ce n'est pas forcément le cas en réalité. Aussi, pour se placer dans un contexte plus général, nous supposons dans la suite que les poids w_i ne sont pas nécessairement unitaires.

L'analyse de variables hétérogènes impose une première mesure : chaque variable numérique doit être centrée-réduite. Désormais, nous faisons donc systématiquement appel aux $(\bar{X}_p)_{p \in [1, P]}$ et aux $(Y_{l,m})_{(l,m) \in [1, L] \times [1, M_l]}$ tels que présentés au début de ce chapitre.

Alors que les variables quantitatives sont assimilées à des vecteurs de norme 1, les variables qualitatives sont représentées par les nuages de points des indicatrices des $Y_{l,m}$.

Il faut que ces nuages de points aient les mêmes caractéristiques inertielles que dans le cas d'une ACM : c'est ce qui permet de créer l'analogie entre ACM et ACP. Pour cela il suffit que l'indicatrice de modalité m_l soit affectée du coefficient $\frac{1}{\sqrt{w_{m_l}}}$.

Une fois l'inertie de chaque nuage de points (associé à une variable qualitative) projetée sur une variable centrée, nous obtenons le carré du rapport de corrélation entre la variable qualitative et la variable centrée. C'est la contrainte de l'affectation du coefficient $\frac{1}{\sqrt{w_{m_l}}}$ qui permet d'obtenir ce résultat.

Si nous notons \vec{u} la direction de \mathbb{R}^n qui rend maximale l'inertie projetée du nuage de points des variables numériques et des indicatrices des variables qualitatives, alors la propriété énoncée revient à maximiser la quantité :

$$\sum_{p \in \llbracket 1, P \rrbracket} r^2(p, \vec{u}) + \sum_{l \in \llbracket 1, L \rrbracket} \eta^2(l, \vec{u}) = \sum_{p \in \llbracket 1, P \rrbracket} \cos^2 \alpha_{p, \vec{u}} + \sum_{l \in \llbracket 1, L \rrbracket} \cos^2 \alpha_{l, \vec{u}}$$

Nous retrouvons, géométriquement, le cadre établi dans le cas de l'ACM.

Quant à la comparaison avec une ACP, elle est envisageable à partir du moment où les variables qualitatives sont codées grâce à un tableau disjonctif complet 4.1. Les indicatrices appelées dans la construction d'un tableau disjonctif complet engendrent des sous-espaces de dimensions égales au nombre de modalités de la variable transformée, moins 1.

L'évocation du tableau disjonctif complet n'est pas sans nous rappeler le raisonnement de l'ACM. En réalité, la jonction entre ACP et ACM ne peut se faire qu'à deux conditions : les modalités de chaque variable qualitative doivent avoir le même nombre d'effectifs et les variables quantitatives doivent être centrées et réduites. Si ces hypothèses sont vérifiées, alors la représentation des modalités via l'ACP est similaire à celle de l'ACM, à un facteur près : $\frac{1}{\sqrt{d-1}}$ où d représente le nombre **commun** des modalités des variables qualitatives.

L'équilibre dans les effectifs des modalités n'est cependant pas un phénomène systématiquement croisé en pratique. Il faut alors étudier la corrélation entre les indicatrices des modalités, pour une variable qualitative donnée. Notons α et β deux modalités d'une même variable qualitative et f_α et f_β leurs fréquences respectives. Le coefficient de corrélation entre α et β vaut :

$$C_{(\alpha, \beta)} = - \left((f_\alpha f_\beta)^{\frac{1}{2}} ((1 - f_\alpha)(1 - f_\beta))^{-\frac{1}{2}} \right)$$

Si les modalités des variables ne portent pas sur des effectifs de même nombre, alors plus l'effectif auquel les modalités sont associées sera faible, plus le coefficient de corrélation révélera leur orthogonalité. A l'inverse, ce coefficient montre que pour des effectifs importants, les modalités d'une variable s'éloignent les unes des autres.

Formulé autrement : plus une modalité présentera un effectif important, plus elle aura de chances de contribuer à la constitution des axes principaux de l'ACP.

En somme, pourquoi l'AFDM ?

Le recours à l'AFDM s'explique par une volonté de combiner les avantages de méthodes performantes dans leurs périmètres respectifs mais qui révèlent des lacunes si ces périmètres sont étendus. L'ACM impose des contraintes sur le nombre d'observations et sur le découpage en classes des variables quantitatives. L'ACP fait appel à un tableau disjonctif complet, autrement dit à une bijection transformant les variables qualitatives en variables quantitatives. Si les données ne sont pas réduites (cas typique des variables exprimées dans une même unité de mesure), il y a un risque à ce que l'analyse ne soit pas précise.

L'idée de la bijection est intéressante. Ce qui manque au codage par tableau disjonctif complet est une pondération des individus présentant une modalité. Cette pondération doit, en tout état de cause, faire intervenir les fréquences des modalités, compensant ainsi l'éventuelle non-réduction des données. Le fait de s'appuyer sur des fréquences n'est pas anodin : c'est l'ACM qui l'inspire. Une fois le codage adopté dans ce sens, opter pour une ACP semble tout à fait convenable. C'est là le principal apport de l'AFDM qui combine bien des principes de l'ACP et de l'ACM.

La section suivante propose l'application de l'AFDM au cas du portefeuille Epargne de Natixis Assurances. Le choix de cette méthode dans cette étude s'explique par l'objectif visé : constituer des groupes homogènes de risque à partir des données de contrats rachetés.

Chapitre 5

Application au portefeuille de Natixis Assurances

Si le principe théorique est rigoureusement défini, l'interprétation graphique n'en demeure pas moins délicate. Puisqu'elle combine les principes de construction de l'ACP et de l'ACM, l'AFDM doit traduire graphiquement les effets d'interactions entre les variables quantitatives et qualitatives. Il ne sera donc pas surprenant de devoir analyser des graphiques familiers aux utilisateurs des ACM et des ACP.

Résultats graphiques et interprétation des résultats

L'AFDM a été réalisée grâce au logiciel R, en utilisant notamment la fonction *PCAmix*. Cette fonction a la particularité de proposer une rotation des axes factoriels afin de limiter l'orientation de toutes les variables vers le premier axe (effet taille connu particulièrement en ACP). La rotation ne modifie pas l'inertie totale. Parmi les sorties graphiques, nous en retenons trois qui nous semblent particulièrement pertinentes dans le cadre de cette étude.

5.1 Cercle des corrélations

Le cercle des corrélations représente les liens entre les variables quantitatives. Il faut observer la direction des vecteurs portant les variables. Ainsi, nous relevons, dans l'ordre des ellipses colorées, que :

- L'ancienneté du contrat augmente avec l'âge des assurés ;
- Le rachat s'effectue sur un contrat d'autant plus tard (ancienneté élevée lors du rachat) que le TMG servi par le contrat est élevé ;
- Les assurés âgés sont ceux dont l'assiette du rachat est la plus importante ;
- Plus l'assiette du rachat est importante, plus le taux du rachat est faible. Ce dernier a été défini comme valant le rapport des montants de rachat par leur assiette. La relation indiquée par le cercle de corrélation signifie que lorsque l'assiette augmente, le montant de rachat diminue et, conséquemment, le taux de rachat diminue.

Lorsque les vecteurs portant les variables sont orthogonaux ou presque, cela signifie qu'il n'existe pas de relation linéaire simple entre les variables concernées. C'est typiquement le cas des couples de variables suivants (taux de rachat ; TMG), (taux de rachat ; ancienneté), (Assiette ; TMG) et (TMG ; Âge). **Cela ne signifie pas que les variables en question soient indépendantes** : des relations non linéaires peuvent exister entre elles.

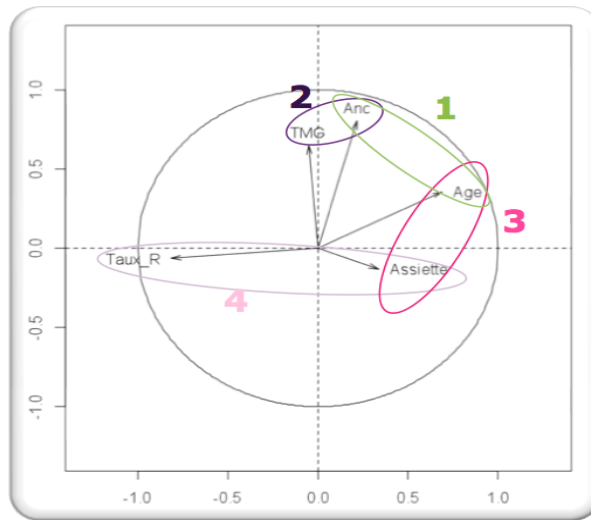


FIGURE 5.1 – Répartition des variables quantitatives après rotation

5.2 Répartition des modalités

Le graphique représentant les modalités des variables qualitatives doit être abordé comme en ACM. Nous lisons dans ce graphique la part d'inertie restituée par chaque axe.

Voici une interprétation que nous proposons pour celui obtenu à l'issue de l'AFDM :

- Trois des quatre modalités du segment de clientèle (Aisé, Patrimonial et Fortuné, encadré rouge) sont très proches. Elles sont situées de l'autre côté de l'axe principal par rapport à la quatrième modalité (Grand Public). Le même éloignement est observé entre les modalités de gamme de produits (Grand Public et Gestion Patrimoniale, encadré fushia). Ceci indique que les assurés rachètent différemment selon leur richesse et selon les produits auxquels ils ont souscrit ;
- Les rachats totaux (encadré bleu) sont plus remarquables auprès des assurés âgés de moins de 70 ans et ayant souscrit avant l'âge de 70 ans (3^{ème} catégorie fiscale, encadré vert). Ces rachats concernent en particulier les contrats Multi-supports ;
- Les rachats partiels programmés (encadré bleu) sont effectués par une clientèle aisée (encadré rouge), âgée et ayant souscrit tardivement (2^{ème} catégorie fiscale, encadré vert) ;
- les rachats partiels (encadré bleu) s'effectuent majoritairement sur des contrats concernant des produits Grand Public (encadré fushia) ;
- Les assurés ayant souscrit avant leur 70^{ème} anniversaire et âgés en 2010 de plus de 70 ans (1^{ème} catégorie fiscale, encadré vert) rachètent surtout des contrats en monosupport euro ;
- Les modalités homme et femme sont proches et, qui plus est, proches du centre de gravité. Le sexe n'est donc pas une variable discriminante.

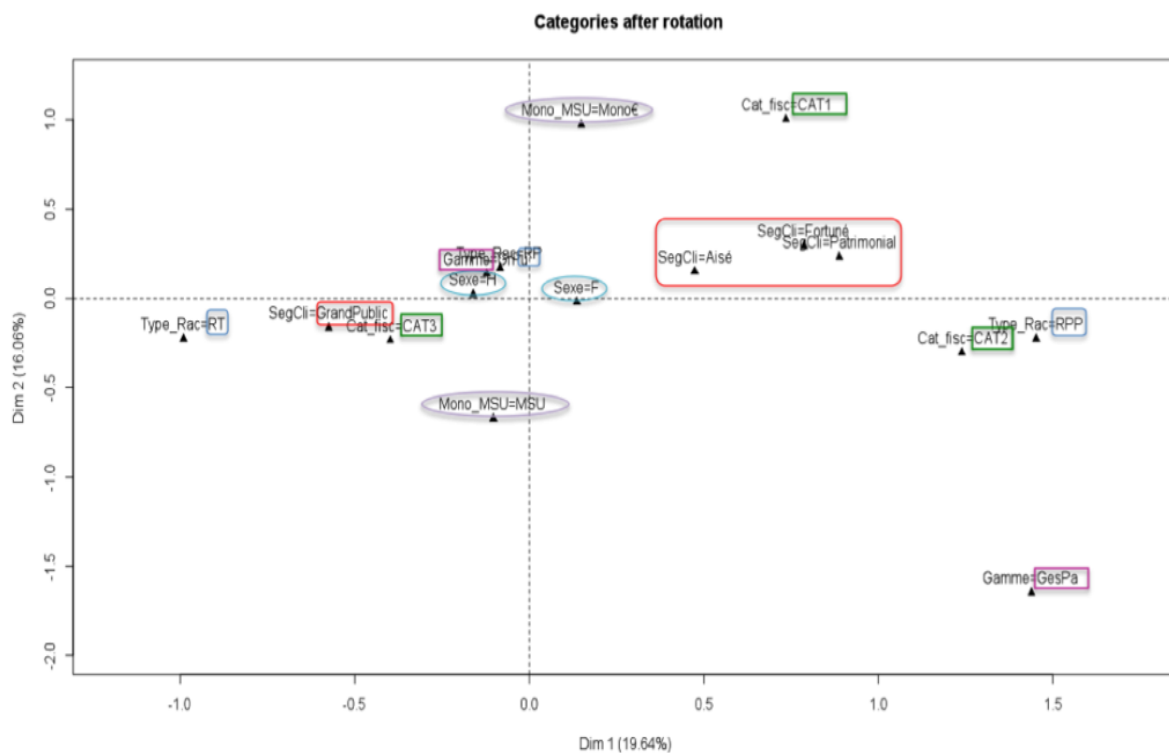


FIGURE 5.2 – Répartition des variables après rotation

5.3 Interactions entre les variables

L'AFDM permet une lecture superposée des deux précédents graphiques par le biais d'un troisième graphique. Toutefois, le cercle des corrélations exhibe des tendances contraires (par exemple) non observées sur le graphique de l'AFDM. Quand à la répartition des modalités, elle permet de constater les interactions entre les modalités, toutes variables qualitatives confondues, non observées sur le graphique de l'AFDM.

Pour ce qui est de la lecture croisée des résultats, son intérêt principal est de faciliter l'interprétation des axes :

- Le type de rachat, le taux de rachat et le segment de clientèle contribuent de façon importante à la constitution de l'axe principal. Ces variables évoluent dans le même sens. Le taux de rachat, nous le constatons plus tôt, dépend de la situation financière de l'assuré. L'assiette par ailleurs, même si timidement, est présente sur cet axe. Nous pouvons donc le traduire comme un axe de **Richesse** ;
- Le type de support du contrat racheté est associé à l'ancienneté du contrat. Des recherches historiques au niveau du portefeuille ont dévoilé que les contrats les plus anciens investissent davantage sur des contrats Mono-support €. Avec le temps, la tendance s'inverse en faveur des contrats MSU. Par ailleurs, les TMG sont servis sur les contrats Mono-support € et sur le fond général des Multi-supports, ce qui explique le lien entre les variables concernées. Nous interprétons donc l'axe secondaire comme un axe du **Temps** ;
- Les variables proches du centre de gravité ne sont pas contributives dans l'interprétation des axes. C'est le cas ici du sexe et, dans une certaine mesure, de l'assiette du rachat ;

- La présence de variables autour de la première bissectrice explique l'équi-contribution de ces variables à la formation des deux axes. C'est en particulier le cas de la gamme de produits. La catégorie fiscale et l'âge, liées, semblent contribuer légèrement plus à l'explication de l'axe principal.

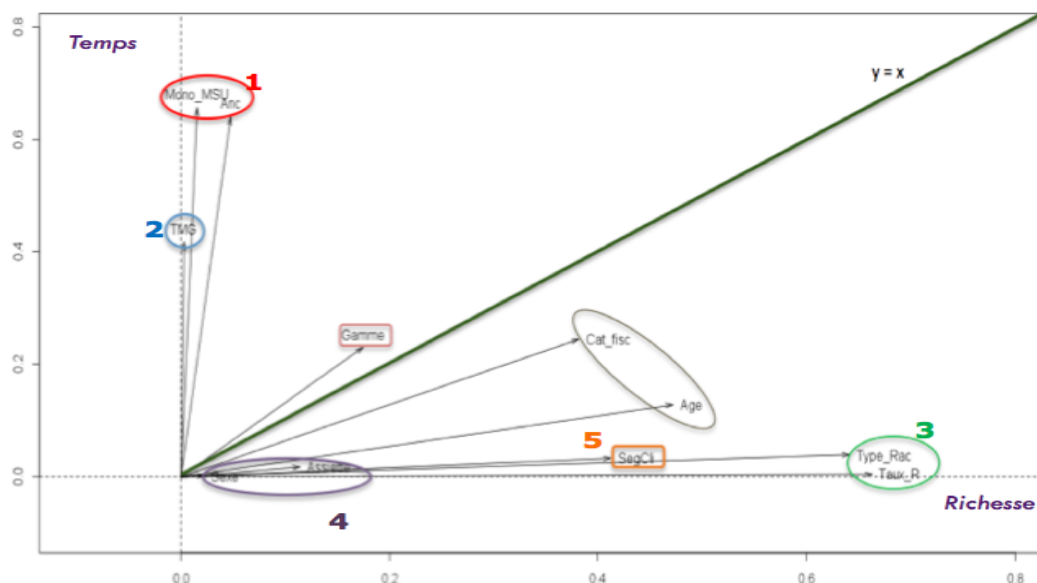


FIGURE 5.3 – Analyse croisée des variables

5.4 Conclusion

L'analyse factorielle a offert une vision à plat de la base de données. Elle a permis de souligner de premières tendances comportementales mais surtout d'isoler des critères a priori non discriminants (Sexe et Assiette) et d'en exhiber de nombreux autres qui expliquent le phénomène du rachat chez nos assurés.

La prochaine partie de ce mémoire est dédiée à la modélisation du taux de rachat par le biais d'approximateurs non-linéaires : les arbres de régression et les réseaux de neurones. Les modèles issus de ces méthodes seront agrégés pour en augmenter la robustesse. Ces méthodes permettront de présenter la contribution de chaque critère dans l'explication du taux de Rachat.

Quatrième partie

Modélisation du taux de Rachat

Chapitre 6

Arbres de régression

Des résultats aisément lisibles et interprétables, une sensibilité aux relations non linéaires, une gestion intégrée des données manquantes, une capacité à s'adapter à des données mixtes,... les **arbres de décision** sont définitivement un outil de modélisation et de classification incontournable. Lorsque sa variable d'intérêt est numérique, un arbre de décision (resp. qualitative), est dit de **arbre de régression** (resp. arbre de classification).

Dans le cas de cette étude, les arbres de régression vont permettre d'expliquer les taux de rachat structurels grâce à d'autres variables de la base de données.

L'étude incontournable autour de cette problématique est sans doute celle, brillante, menée par X. Milhaud (2014) [48]. Afin de déterminer le profil des assurés susceptibles de racheter leur contrat d'épargne, Milhaud propose d'utiliser arbres de décision, régression logistique et modèle GLM. Pour appliquer les premiers, il sélectionne un nombre de variables explicatives pouvant conditionner la décision du rachat (variable d'intérêt binaire). En s'appuyant sur des arbres de classification, il exhibe les critères qui, dans le cas du portefeuille analysé, expliquent le rachat (ancienneté, type du contrat, richesse de l'assuré sont notamment retenus).

Dans l'étude présente, nous relevons, entre autres, deux principales différences d'approche avec celle de Milhaud :

- * Les arbres utilisés sont des arbres de régression qui restituent des taux de rachat et non une probabilité (binaire) de rachat ;
- * Les bases étudiées sont celles de contrats rachetés : nous savons que les contrats sont rachetés mais nous souhaitons associer à chaque taux de rachat (et donc à chaque type de rachat puisque par construction, les deux sont liés) un profil d'assurés.

Les sections suivantes sont à caractère pratique : elles détaillent les étapes suivies dans la construction d'un arbre de régression. Le lecteur intéressé par le détail mathématique est invité à consulter l'introduction méthodique du mémoire de E. Scornet [68] ainsi que le mémoire de X. Milhaud.

6.1 Echantillonnage

L'échantillonnage est une technique statistique qui consiste à découper un jeu de données de telle sorte à ce que les échantillons construits soient utilisés à différentes finalités. Généralement, il est d'usage de constituer un échantillon d'apprentissage, un échantillon de test et un échantillon de validation. Le 1^{er} sert à entraîner le modèle à apprendre les différentes structures qui caractérisent le jeu de données. Le second sert à ajuster les paramètres du modèle. Le dernier, enfin, est utilisé à des fins de prévision : autrement dit, à vérifier les capacités de généralisation du modèle.

L'échantillonnage n'est possible que si le phénomène à exhiber n'est pas d'une rareté exceptionnelle. Si c'est le cas, cette opération risque de ne pas faire apparaître le phénomène étudié au niveau de l'échantillon de test. Il existe des méthodes permettant de mettre en avant la représentativité des échantillons avant de procéder à l'apprentissage : test de Student, test du Khi-deux,...

Par ailleurs, l'échantillonnage n'est réalisable que si la population est suffisamment importante. Lorsque ce n'est pas le cas, la *validation croisée* permet d'obtenir des résultats satisfaisants. Cette technique consiste à diviser une base de données en n segments : $n - 1$ serviront à apprendre les données et un segment sera utilisé lors de la phase de test. Le découpage est répété aléatoirement n fois. Le taux d'erreur de cette opération est calculé comme la moyenne des taux d'erreurs sur chacun des découpages.

Notons également que plusieurs techniques d'échantillonnages existent. Celle choisie doit être cohérente avec l'objectif de l'étude. La méthode d'échantillonnage la plus élémentaire consiste à tirer aléatoirement un nombre N d'individus dans une base de données. Il est également possible de constituer de premières partitions du jeu de données, par ordre alphabétique des noms de familles des assurés par exemple, puis de tirer aléatoirement dans chacune de ces partitions (*échantillonnage stratifié*). Dans les études de sondages, le recensement consiste à faire dans un premier temps un choix déterministe (quartiers d'une ville à sonder,...), puis à étudier l'ensemble des sous-groupes désignés (*échantillonnage par grappes*). D'autres techniques d'échantillonnages existent, aléatoires : systématique, avec une probabilité proportionnelle à la taille,... et non aléatoires : échantillonnage au jugé, par quotas,...

Dans le cas présent, la question de la représentativité ne se pose pas dans la mesure où le jeu de données étudié est celui de contrats rachetés. Nous effectuons un échantillonnage simple sur nos jeux de données avec la répartition usuelle : 80% pour l'échantillon d'apprentissage, 20% pour l'échantillon de test. L'étape de validation sera réalisée grâce à des données récentes : celles de l'année 2013.

6.2 Arbre de régression

Un arbre de régression est un arbre de décision dont la variable d'intérêt est numérique (il est dit de classification si cette variable est qualitative). Les arbres de décision sont, comme leur nom l'indique, un outil d'aide à la décision. En disposant d'une base de données où les critères pouvant expliquer la construction d'une variable précise sont légion, l'arbre de décision permet de distinguer ceux de ces critères qui contribuent le plus au comportement de la variable d'intérêt.

Il existe plusieurs types d'algorithmes d'arbres de décision, les plus connus étant le CART (Classification and Regression Trees), le C5.0 et le CHAID (CHi-squared Automatic Interaction Detector). Ils diffèrent par les types de variables qu'ils peuvent gérer et par leurs algorithmes respectifs.

Dans la suite, nous nous intéressons en particulier aux arbres CART, popularisés par Breiman et al. [42], appliqués au cas de la régression.

L'intérêt principal des arbres de régression réside dans la simplicité avec laquelle ils peuvent être mis en place. L'arbre est construit en détectant, dans un premier temps, les variables les plus contributives dans l'explication de la variable d'intérêt. Des règles de classification sont alors fixés. Ces règles définissent l'arborescence de l'arbre : chaque nœud/extrémité (les extrémités sont également appelées *feuilles*) de l'arbre est caractérisé

par un ensemble de règles. Les individus sont répartis sur les extrémités de l'arbre selon ces règles : un individu ne peut être affecté qu'à un nœud/feuille.

La ramification de l'arbre cesse s'il n'y a plus qu'un individu par nœud (classes de singletons), si un critère d'arrêt initialisé par l'utilisateur est atteint, si la pureté des nœuds est jugée suffisante,...

Nous détaillons ci-après les phases de construction d'un arbre CART de régression :

6.2.1 Règles de segmentation

La segmentation d'un arbre est réalisée par des choix successifs de variables explicatives contributives dans l'explication de la variable d'intérêt. Le nombre de branches dépend alors du type de la variable explicative sélectionnée à chaque nœud. Il y a :

- $k - 1$ branches formées si la variable explicative est continue à k valeurs possibles ;
- $2^{k-1} - 1$ branches formées si la variable explicative est qualitative à k modalités.

La segmentation se poursuit jusqu'à ce que, par exemple, il n'y ait qu'un individu par nœud ou qu'un critère d'arrêt soit atteint. Ce dernier peut être contrôlé grâce à des indicateurs de précision du modèle.

Dans le cas des arbres CART, c'est l'indice de Gini qui permet de mesurer cette précision en appréciant la distribution uniforme des classes au niveau de chaque nœud. L'indice de Gini mesure l'*impureté* de chaque nœud : plus il est bas, plus le nœud est pur. Cet indice s'écrit :

$$I_{Gini} = \sum_{i \neq j} P(i|t)P(j|t) \quad (6.2.1)$$

où $P(i|t)$ désigne la proportion d'éléments de la classe i affectés au nœud t .

L'algorithme d'un arbre CART vise à réduire l'impureté en se déplaçant d'un nœud père vers ses nœuds fils. Autrement dit, l'arbre cherche toujours à maximiser : $I_{\text{nœud père}} - (I_{\text{nœud 1er fils}} + I_{\text{nœud 2ème fils}})$.

6.2.2 Affectation des individus

Une fois les segments de l'arbre construits grâce aux règles définies plus tôt et à la maximisation de la pureté de chaque nœud, les individus sont répartis dans chaque nœud de l'arbre. Un individu ne peut être affecté qu'à un nœud.

Cette répartition peut, dans le cas des arbres CART, être réalisée en faisant appel à un coût de mauvais classement des individus, directement intégré dans le calcul de l'indice de Gini. La formule (6.2.1) devient :

$$I_{Gini} = \sum_{i \neq j} C(i|j)P(i|t)P(j|t)$$

où $C_i(j)$ le coût du mauvais classement d'une classe i dans la classe j .

L'affectation des individus aux nœuds consiste à réduire le coût de mauvaise affectation associé à chaque nœud en réduisant celui de chaque individu affecté à ce nœud.

6.2.3 Elagage de l'arbre

L'élagage de l'arbre consiste à délester l'arbre des feuilles qui n'apporteraient rien à l'analyse de l'arbre : le nombre d'individus est trop faible pour la pertinence de l'étude, la qualité de l'arbre n'augmente plus de façon significative avec la ramification,... Cette opération est délicate car elle ne doit pas priver l'arbre de ses capacités prédictives en supprimant un nombre trop important de nœuds.

Les arbres CART intègrent un processus d'élagage dans leur algorithme. En effet, l'algorithme construit plusieurs arbres et sélectionne celui dont le taux d'erreur est le moins important.

Pour autant, il existe plusieurs façons d'élaguer un arbre (en prenant particulièrement garde à la significativité de ses effectifs). Dans le cadre de cette étude, notre choix se porte sur la règle de l'écart-type de Breiman et al.(1984) qui consiste à réduire le facteur de complexité de l'arbre final en choisissant celui qui comporte le moins de branches terminales. Cette technique sera expliquée plus en détail dans l'application des arbres CART à notre base de données.

6.2.4 Limites

Nous l'évoquions plus tôt, les avantages des arbres de décision sont nombreux : interprétation aisée des résultats, manipulation informatique intuitive, adaptation à des données mixtes/manquantes, temps de calcul raisonnables,... cependant cet outil de classement et de prédiction souffre d'un certain manque de robustesse lié en particulier à sa sensibilité au risque de surapprentissage, au côté décisif du premier nœud de l'arbre (il conditionne en effet toutes les ramifications de l'arbre) ou encore à la dépendance entre variable à expliquer et variables explicatives.

Surapprentissage

Le surapprentissage survient lorsqu'un modèle parcourt des données un nombre important de fois jusqu'à s'imprégner de toutes les caractéristiques de ces données, en particulier de traits particulièrement exceptionnels. Il retient alors ces traits exceptionnels au même titre que des comportements réguliers et les généralise à l'ensemble de la base étudiée. Ce phénomène peut être croisé lorsque les populations ne sont pas importantes, que les variables explicatives sont très nombreuses,...

Dans le cas des arbres de régression, le risque de surapprentissage est accru lorsque le nombre de feuilles/nœud est important. Deux procédés permettent de limiter ce risque : l'élagage de l'arbre et l'agrégation. En effet, alors que le premier permet de mettre à l'écart des individus atypiques non forcément pertinents, le second vérifie, par la construction de nombreux arbres, la redondance d'un critère sur l'ensemble des arbres construits et mesure sa contribution dans la construction de la variable d'intérêt dans chaque arbre.

Sensibilité aux optimums locaux

C'est là l'une des plus importantes faiblesses des arbres de régression. L'algorithme étudie les variables explicatives successivement. Les nœuds sont construits de façon enchaînée et un critère choisi pour figurer à un emplacement de l'arbre n'est plus réétudié par la suite. Cela entend que modifier en amont la construction d'un critère fort remet en question la construction de l'intégralité de l'arbre. Ici aussi, l'agrégation de modèles permet de remonter les variables les plus déterminantes dans l'explication de la variable d'intérêt.

Dépendance entre variable d'intérêt et variables explicatives

Comme dans toute méthode d'apprentissage supervisé (i.e. où une variable à expliquer est identifiée) il faut être particulièrement vigilant quant aux liens possibles entre la variable d'intérêt et les variables explicatives. Par lien, nous entendons : construction. Si la construction de la variable d'intérêt est étroitement liée à une variable explicative, cela doit être pris en compte dans l'interprétation des résultats.

Les avantages liés aux arbres de régression en font un outil incontournable. Les défauts énumérés ci-dessus peuvent être floutés par l'agrégation de modèles (au détriment de la facilité de lecture des résultats des arbres) et par une attention particulière aux variables explicatives. La section suivante présente un autre outil de modélisation dédié à l'analyse de données mixtes : les réseaux de neurones.

Chapitre 7

Réseaux de neurones

Si les réseaux de neurones sont naturellement liés aux domaines de la médecine et de la biologie, le concept de ces modèles sophistiqués a rapidement gagné d'autres univers (sécurité, énergies, télécommunication,...). Un réseau de neurones fait d'abord référence à l'architecture du cortex humain. La complexité de cette architecture a de fait été associée à tout phénomène dont le fonctionnement élaboré dissimulerait des connexions invisibles à l'oeil nu.

Au-delà d'une simple lecture d'informations, les réseaux de neurones peuvent donner un poids aux variables explicatives d'une base de données. Ils sont aussi bien capables de décrire (fonction de classification) que d'analyser (fonction de prédiction). Pourtant, le paramétrage difficilement maîtrisable de ses arguments, l'interprétation délicate de ses résultats ou encore l'hermétisme de ses conditions d'application font des réseaux de neurones un outil particulièrement sensible à manipuler. Cette sensibilité peut être contrôlée en faisant appel aux réseaux de neurones parallèlement à d'autres méthodes de classification/prédiction non-linéaires afin de soutenir/infirmer des postulats préalablement établis.

Quant à l'usage des réseaux de neurones en Actuariat, ce sont aujourd'hui les branches Santé et Prévoyance qui exploitent en particulier cet outil. L'approche décrite dans ce mémoire emprunte l'approche de certains mémoires tout en s'appuyant sur le contexte de l'Épargne pour expliquer les résultats observés.

Dans les sections suivantes, les aspects historiques de l'utilisation d'un réseau de neurones sont détaillés en premier lieu. Le formalisme mathématique des réseaux de neurones vient ensuite. Enfin, les limites de ces modèles sont présentés.

7.1 Principaux travaux sur les réseaux de neurones

Les premiers modèles basés sur les réseaux de neurones ont été établis par McCulloch et Pitts (1943) [27]. Les auteurs analysent le fonctionnement de réseaux de neurones humains et tentent d'expliquer comment le cerveau humain réussit à produire des informations complexes à partir de la simple connexion existant entre les neurones. Le neurone, élément de base mis en exergue par leurs travaux, prend le nom de **neurone MCP**. Ils montrent qu'une combinaison appropriée de neurones MCP peut traduire n'importe quelle assertion logique pour peu qu'elle puisse être exprimée en termes binaires. Leurs travaux ont particulièrement contribué à faire avancer le domaine électronique.

Les réseaux de neurones souffrent de leur hermétisme : on ne comprend pas comment les mettre en pratique et ce qu'ils apportent concrètement dans des études statistiques. Un essoufflement et une déception générale expliquent alors la traversée du désert que cet outil va connaître. Ce n'est qu'à partir de 1967 que les travaux de A. Shun-Ichi (1967) [71] sur

la réduction du taux d'erreur lié à la pondération des variables explicatives d'un réseau de neurones vont leur valoir un regain d'intérêt. A. H. Klopff(1972) commence par entraîner des données biologiques sur un modèle neuronal alors que P. Werbos (1974) [78] met en lumière le phénomène le plus populaire en matière de réseaux de neurones : la rétropropagation du gradient dans les perceptrons multicouches.

Depuis, et bien que leur accessibilité soit régulièrement remise en question, les réseaux de neurones suscitent de plus en plus l'intérêt d'académiciens et de professionnels qui y trouvent un outil prometteur.

En milieu actuariel, les récents papiers rédigés sur l'utilisation des réseaux de neurones concernent les milieux de la Santé et de l'allocation d'actifs. Dans son mémoire, J.-M. Aouizerate [2] (extrêmement bien documenté sur les modèles basés sur les réseaux de neurones), le modèle tarifaire Santé de Gras Savoye est construit grâce à un modèle perceptrons multi-couches avec l'algorithme de rétropropagation du gradient de l'erreur. Le modèle a permis de prédire des tarifs plus réalistes que ceux obtenus via des méthodes plus standards (Modèles Linéaires Généralisés et Memory Based Reasoning). Anne-Cécile Richard [62], quant à elle, met en exergue dans le milieu de l'allocation stratégique, la pertinence de cet outil dans le cadre de la prévision de rendements financiers.

Nous allons, par la suite, expliquer les aspects rudimentaires d'un réseau de neurones.

7.2 Formalisme d'un réseau de neurones

Il existe une notion biologique et une notion mathématique du neurone. La seconde doit son formalisme à la première :

- Le **neurone biologique** est composé d'une tige fine (l'axome) qui lie "physiquement" les neurones et de petits filaments (les dendrites) entourant le corps du neurone. Le signal nerveux est transmis depuis les dendrites vers les axomes et ce, par le biais des synapses qui les joignent ;

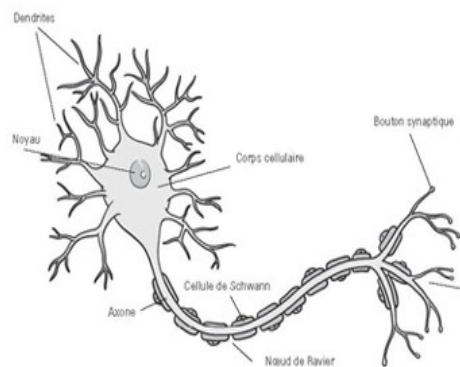


FIGURE 7.1 – Composition d'un neurone biologique

- Le **neurone formel**, également appelé *nœud*, est un concept mathématique reprenant le mécanisme du neurone biologique. D'un neurone formel émane une information qui sera reçue par un autre neurone. Au sein d'un réseau, ces connexions sont multiples et de natures différentes.

Un réseau de neurones est, de façon très sommaire, composé d'une couche d'entrée, d'une couche de sortie et d'éventuellement une ou plusieurs couches intermédiaires, dite couches cachées. Chaque couche est composée de plusieurs neurones.

C'est la présence de couches cachées qui donne cette dimension non-linéaire aux sorties d'un réseau de neurones. Ne pas inclure de couches intermédiaires reviendrait à réaliser un système linéaire simple !

Les neurones de la couche d'entrée représentent les variables explicatives de la base de données. Il y a autant de neurones dans la couche d'entrée que de variables discrètes/continues et que de modalités de variables qualitatives.

Les neurones de la couche de sortie dépendent du type du réseau : si le réseau est prédictif, il y a autant de nœuds dans la couche de sortie que de variables explicatives (ou que de modalités de variables à expliquer si ces dernières sont qualitatives).

La figure 7.2 illustre l'exemple d'un réseau de neurones contenant 4 neurones dans la couche d'entrée, 3 neurones dans la couche intermédiaire et un neurone dans la couche de sortie.

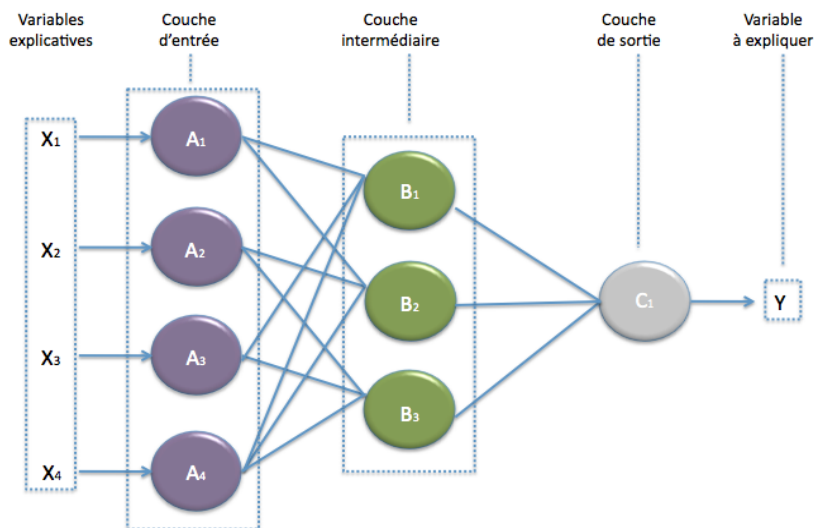


FIGURE 7.2 – Exemple d'un réseau de neurones élémentaire

Le lien entre deux neurones est appelé **poids synaptique**. C'est ce poids qui contient l'information partagée par ces deux neurones. Dans la mesure où il représente une probabilité de réponse, il est naturellement compris entre -1 et 1.

Il existe cependant d'autres moyens de capturer cette information : via la distance entre les neurones de la couche d'entrée et des centres des nœuds de la couche intermédiaires lorsqu'il n'en existe qu'une, par exemple.

Au niveau de la couche d'entrée, chaque nœud est naturellement attaché à des poids synaptiques (les w_i de la figure 7.3). Ensuite, les poids sont liés aux nœuds suivants via une **fonction de combinaison**. Cette fonction, appliquée en entrée des nœuds de la couche d'entrée (et en entrée des couches intermédiaires si le réseau en dispose) rattache les valeurs en entrées des neurones aux connexions (les poids) qui existent entre eux. En général, cette fonction est une somme pondérée des valeurs des nœuds par les poids synaptiques auxquels elles sont rattachées.

A cette combinaison est ensuite appliquée une **fonction d'activation** qui va estimer l'activité du nœud à partir de cette somme. Les fonctions d'activation neuronales sont nombreuses et sont au libre choix de l'utilisateur. Elles sont détaillées dans la section 7.2.2.3.

Enfin, notons l'existence de **biais** (également appelé *seuil d'activation des neurones*) au niveau de chaque nœud. Un biais est un neurone dont la valeur, soumise à la fonction d'activation, vaut toujours 1. Un biais, associé à une couche donnée, est connecté à l'ensemble des neurones de la couche suivante mais à aucun de la couche auquel il est rattaché. Dans la figure 7.3 par exemple, le biais B_1 est rattaché à la couche d'entrée mais uniquement connecté aux neurones de la couche intermédiaire.

Les poids associés à un biais sont rajoutés à la sortie de la fonction de combinaison appliquée aux neurones de la couche à laquelle est connectée le biais. La présence d'un biais a un impact sur la définition de la fonction d'activation comme expliqué dans la section 7.2.2.3.

L'ensemble des biais, des données en entrée transformées par la fonction de combinaison puis par la fonction d'activation, est appelé **noyau**.

La figure 7.3 illustre l'intervention de ces deux fonctions dans la lecture de l'information à partir de la couche d'entrée d'un réseau de neurones élémentaire.

La fonction de combinaison est notée f et la fonction d'activation g . Dans la mesure où cet exemple fait appel à une couche cachée, les valeurs finales au niveau de la couche de sortie sont une pondération des valeurs de sortie de la couche intermédiaire auxquelles est à nouveau appliquée la fonction d'activation.

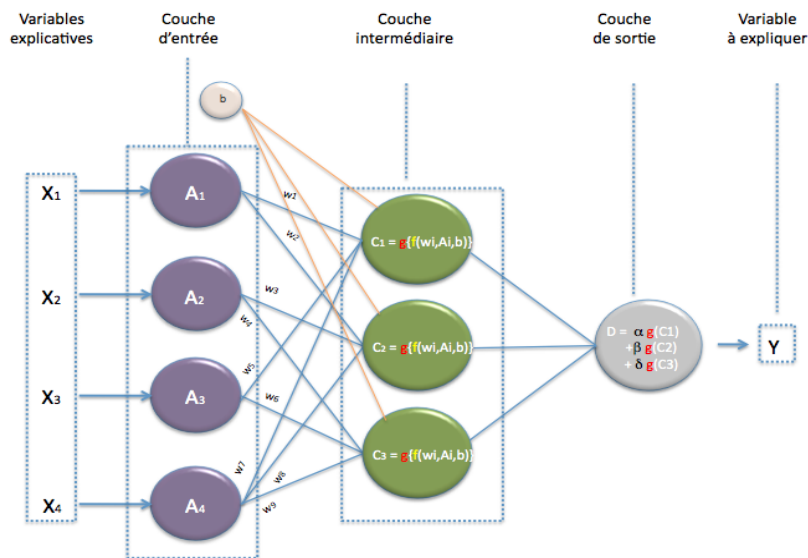


FIGURE 7.3 – Fonction de combinaison et d'activation dans un réseau de neurones

Pour résumer, un réseau de neurones est composé, au niveau et entre les couches qui le composent, des éléments suivants :

- De données en entrée et en sortie du modèle, ainsi que de transformations intermédiaires des variables d'entrée. Ces différentes variables sont représentées par des **neurones** ;
- De **biais** qui contribuent à l'ajustement du réseau par l'unique moyen des poids sans devoir manipuler des paramètres de la fonction d'activation ;

- De **poïds** qui traduisent la probabilité de transmission de l'information entre les neurones ;
- De **noyaux** rassemblant l'application de la fonction de combinaison à des données en entrée (données initiales ou en sortie d'une couche cachée), l'application de la fonction d'activation aux sorties de la fonction de combinaison et l'effet du biais.

7.2.1 Types de réseaux de neurones

L'information, dans un réseau de neurones, peut circuler de deux façons :

- Soit en "sens unique" en se déplaçant de la couche d'entrée vers la couche de sortie (en passant éventuellement par des couches intermédiaires) mais sans retour en arrière possible ;
- Soit en permettant aux valeurs en sortie des nœuds de revenir vers des couches précédentes, afin de repenser leurs valeurs de sortie. Lorsque le déplacement peut se faire dans les 2 sens, il est question de "cycle".

Dans le premier cas, le réseau est dit non-bouclé, dans le second : il est bouclé.

7.2.1.1 Réseaux non bouclés

Un réseau de neurones non bouclé est un réseau de neurones où l'information est conduite dans un seul sens : des entrées vers la variable réponse (Cf figure 7.2). On dit que ce type de réseaux est *acyclique* : ne dépendant pas du temps, ces réseaux sont également dit "statiques".

7.2.1.2 Réseaux bouclés (ou réseaux récurrents)

Il est question de réseau bouclé lorsque la réponse d'un nœud peut être utilisée comme valeur d'entrée pour le nœud d'une couche précédente. Une telle opération est possible uniquement en introduisant une notion de temporalité. Ainsi, la sortie utilisée en entrée d'une couche précédente est nécessairement accompagnée d'un facteur de retard strictement positif. Ces réseaux sont donc "dynamiques". La figure 7.4 illustre le déplacement de l'information dans un réseau bouclé.

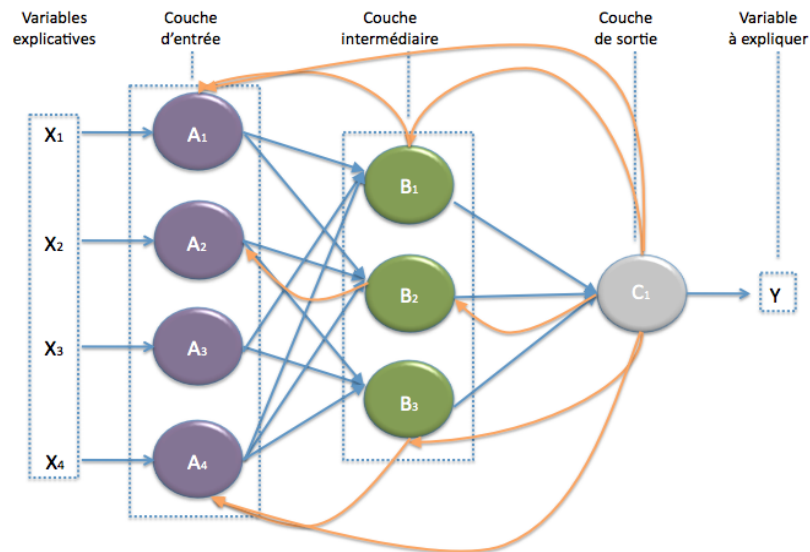


FIGURE 7.4 – Illustration d'un réseau de neurones bouclé

Nerrand (1993) [52] montre qu'il est toujours possible de mettre sous forme canonique un réseau de neurones bouclé faisant intervenir un réseau non bouclé. Grâce à des facteurs de retard, certaines valeurs de sortie de ce dernier peuvent être utilisées comme valeurs d'entrée.

Autrement dit : tout part d'un réseau de neurones non bouclé.

7.2.2 Structure d'un réseau de neurones

L'architecture d'un réseau de neurones est définie par :

- La fonction du réseau ;
- Le nombre de neurones et de couches du réseau ;
- Les connexions entre les neurones ;
- L'ajustement des connexions.

7.2.2.1 Fonction du réseau

L'apprentissage peut, notamment, être supervisé ou non supervisé (Cf chapitre II).

Les réseaux de neurones peuvent s'inscrire dans l'un ou l'autre de ces cadres. Nous abordons le cas du perceptron multi-couches et du réseau à fonctions de base radiales (apprentissage supervisé) puis celui des cartes de Kohonen (apprentissage non-supervisé) dans les prochaines sections afin d'illustrer l'inscription des réseaux de neurones dans ces différentes approches statistiques.

7.2.2.2 Nombre de neurones et de couches du réseau

Dans un réseau de neurones, il existe toujours une couche d'entrée et une couche de sortie. Des couches intermédiaires peuvent exister entre les deux.

Si le jeu de données dispose de $X_N = (X_{N_1}, \dots, X_{N_n})$ variables explicatives numériques et de $X_Q = (X_{Q_1}, \dots, X_{Q_m})$ variables explicatives qualitatives à $(l_i)_{i \in \llbracket 1, m \rrbracket}$ modalités chacune, alors la couche d'entrée disposera de $n + l \times m$ nœuds dans sa couche d'entrée. De façon symétrique, la couche de sortie a 1 neurone si la variable à expliquer est numérique. Si cette dernière est qualitative à k modalités, la couche de sortie aura k neurones.

La question du nombre de neurones est décisive lorsqu'une (ou plusieurs) couche cachée est intégrée au réseau. Le nombre de neurones de cette couche doit être fixé par l'utilisateur. Ce nombre peut dépendre du nombre d'individus, du nombre de classes à prédire, du bruit qui peut être contenu dans la variable à prédire, de l'architecture du réseau, de l'algorithme d'apprentissage, du nombre d'itérations,... Plusieurs théories existent sur le réglage progressif de ce nombre (Blum (1992) [9], Swingler (1996) [72], Berry and Linoff (1997) [23],...), la plus répandue consiste à tester plusieurs nombres et de choisir celui qui minimise le taux d'erreur. Barron (1993) [4] a d'ailleurs démontré le résultat suivant concernant la relation entre nombres de neurones dans la (ou les) couche cachée et la vitesse de décroissance du taux d'erreur :

Commençons par définir la notion d'erreur intégrée, notée RISE pour *Root Integrated Squared Error* :

Soit g une fonction continue sur un espace de dimension finie, ν une mesure de probabilité définie sur la boule fermée $B_a = \{x : |x| \leq a\}$. Alors pour tout $n \geq 1$, il existe une combinaison $g_n(x)$ de fonctions sigmoïdes telles que $RISE = \int_{B_a} (g(x) - g_n(x))^2 \nu(dx)$

Ensuite, soit un jeu de données disposant de α entrées ($\alpha = n + l \times m$ dans l'exemple précédent). Si nous construisons un réseau de neurones à l'aide d'une seule couche intermédiaire à β neurones, alors Barron affirme que, sous certaines hypothèses, la vitesse de décroissance de la racine carrée de l'erreur intégrée sera de l'ordre de $\frac{1}{\sqrt{\beta}}$.

7.2.2.3 Connexions entre neurones

Les neurones n'utilisent pas l'information brute des inputs : elle est d'abord pondérée par les poids via la fonction de combinaison. Une fonction d'activation, appliquée à la sortie de la fonction de combinaison, permet d'introduire une non-linéarité dans la réponse accordée par le nœud.

a. Fonction de combinaison

La fonction de combinaison permet de combiner les valeurs en entrée d'un nœud et les poids des nœuds connectés. Les réseaux de neurones utilisent généralement l'une des deux fonctions de combinaisons suivantes :

- Une combinaison linéaire entre les poids et les entrées des nœuds, autrement dit :

$$f(w_i, A_i, b) = \sum_i w_i A_i + b$$

où w_i , A_i et b représentent, respectivement, les poids, les valeurs d'entrée et le biais liés à une couche donnée (couche d'entrée ou couche intermédiaire). Ce type de fonction caractérise par exemple les perceptrons multi-couches ;

- La distance euclidienne (plus rarement la distance de Mahalanobis) entre les observations d'une couche initiale et les centres des nœuds de la couche suivante. La fonction d'activation prend alors la forme :

$$f(w_i, A_i, b) = (\|A_i - w_i\|_2)^2 + b = \sum_i (A_i - w_i)^2 + b$$

Cette forme de fonction d'activation caractérise en particulier les réseaux à fonctions de base radiales.

La fonction de combinaison est parfois appelée potentiel. Elle est alors notée :

$$f(w_i, A_i, b) = \nu_i$$

Nous utilisons cette notation dans la suite de ce mémoire.

b. Fonction d'activation

La fonction d'activation permet d'intégrer une non-linéarité aux valeurs en sortie de nœuds. Parmi les fonctions d'activations les plus utilisées, nous citons :

La fonction Identité : $g(x) = x$

La fonction linéaire seuillée : $g(x) = \begin{cases} 0 & \text{si } x \leq x_{min} \\ \alpha x + \beta & \text{si } x_{min} < x < x_{max} \\ 1 & \text{si } x \geq x_{max} \end{cases}$

La fonction Sigmoidé : $g(x) = \frac{1}{1+e^{-\alpha x}}$

La fonction Gaussienne : $g(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Bien qu'il n'existe pas de règle imposant le choix d'une fonction d'activation en particulier, la fonction sigmoïde est celle qui est le plus souvent utilisée. En effet, les sorties d'une fonction sigmoïde ne sont pas linéaires. Elle permet donc d'ajuster le moins de paramètres possibles en phase d'apprentissage et de test. On dit que la fonction sigmoïde permet une approximation *parcimonieuse*.

7.2.3 Types de réseaux

Les réseaux de neurones sont classés selon leur algorithme d'apprentissage. Il existe de nombreux modèles de réseaux de neurones, nous en citons un nombre important en Annexe. Nous présentons ci-après les deux exemples les plus utilisés dans le cas d'un apprentissage supervisé : le perceptron multi-couches et le réseau à fonctions radiales de base, et un modèle de réseau utilisé dans le cas de la classification : la carte de Kohonen.

7.2.3.1 Le Perceptron Multi-Couches

Le Perceptron Multi-Couches (PMC) est un réseau de neurones non bouclé, utilisé dans le cadre d'un apprentissage supervisé, dont la fonction d'activation est une sigmoïde. L'une des dernières améliorations introduites dans l'étude des PMC, visant à accélérer l'apprentissage, consiste à remplacer la fonction sigmoïde classique présentée plus tôt par la fonction tangente hyperbolique : $g(x) = \frac{2\alpha}{1+e^{-\beta x}} - \alpha$ avec $\alpha = 1,716$ et $\beta = 0,667$.

Un PMC est composé d'au moins une couche intermédiaire en plus des couches d'entrée et de sortie. Dans un PMC, chaque nœud d'une couche est lié à l'ensemble des nœuds de la couche précédente mais pas à celle qui la précède (en cas de multiples couches intermédiaires).

L'intérêt de ce réseau tient à ce qu'une somme de sigmoïdes peut approcher n'importe quelle fonction pour peu qu'elle soit suffisamment régulière.

Le modèle de PMC le plus populaire est celui qui utilise l'algorithme de rétropropagation du gradient pour l'apprentissage du modèle. C'est celui pour lequel nous optons dans le cadre de ce mémoire.

7.2.3.2 Le RBF

Les réseaux RBF (*Radial Basis Functions*) sont des réseaux non bouclés, utilisés dans le cadre d'un apprentissage supervisé, à fonction d'activation gaussienne. Dans ces réseaux, il n'y a qu'une couche cachée en plus de la couche d'entrée et de la couche de sortie.

Les réseaux RBF utilisent la distance (euclidienne ou plus rarement celle de Mahalanobis) entre les valeurs d'entrée et le centre des nœuds de la couche suivante. De plus, la fonction d'activation procurant les valeurs finales du modèle est toujours égale à la fonction identité.

Une autre différence notable entre les réseaux RBF et les PMC concerne sans doute la vitesse d'apprentissage et de test. Alors que les premiers sont plus rapides à l'étape d'apprentissage, les seconds le sont en phase de test. En effet, dans un réseau RBF, l'indépendance des nœuds cachés les uns par rapport aux autres leur permet de faire converger l'algorithme plus rapidement vers une solution optimale. Parallèlement à cela, dans la mesure où il ne dispose que d'une couche cachée, les neurones de cette couche sont sensiblement plus importants que dans un PMC, d'où un temps plus important pour la généralisation du modèle en phase de test.

7.2.3.3 La carte auto-organisatrice de Kohonen

La carte auto-organisatrice de Kohonen est un réseau neuronal bouclé basé sur un algorithme à apprentissage compétitif non-supervisé. Un réseau est dit compétitif lorsque l'ensemble de ses neurones est en compétition pour véhiculer une information. Dans ce type de réseaux, l'absence d'un quelconque neurone n'empêche pas la transmission de l'information.

Ce réseau opère une séparation "topologique" de l'espace en étudiant les distances séparant les classes d'individus selon un certain critère. Dans ces cartes, il n'y a pas de connexions entre les nœuds de la couche de sortie : les neurones y sont organisés dans une grille, le plus souvent à deux dimensions. Nous parlons ici uniquement de voisinage entre les nœuds.

A chaque itération, le réseau ajuste les poids de sorte à ce que la distance de chaque individu par rapport à son voisinage soit la moindre possible. Lorsque toute la population de l'échantillon d'apprentissage a été soumise à l'algorithme et que tous les poids ont été ajustés, l'apprentissage est abouti.

Les réseaux de Kohonen auraient pu offrir une alternative intéressante à la segmentation de portefeuille proposée par AFDM dans ce mémoire.

7.2.4 Ajustement des poids

Afin d'éviter les phénomènes de surapprentissage et de permettre au réseau d'optimiser son pouvoir de généralisation lors de la prédiction, nous devons fixer au réseau des règles d'apprentissage : vitesse de parcours des données, ... Ces règles sont définies par l'**algorithme d'apprentissage**.

C'est en particulier cet algorithme qui ajuste les poids durant l'apprentissage en minimisant une fonction appelée **fonction de coût**. Cette fonction représente la différence entre les sorties réelles de la base et les valeurs que l'on obtient par le biais du modèle. Cette fonction donne ainsi une idée plus ou moins précise de la qualité et des capacités prédictives du modèle.

Sa forme diffère selon que l'on souhaite approcher une fonction continue (régression) ou séparer l'espace des variables prédictives (classification).

Dans le cas de la régression, il est d'usage que la fonction de coût soit égale à l'erreur quadratique :

$$C(w) = \sum_{\alpha} (y_{\alpha} - y_{\alpha}^w)^2 \quad (7.2.1)$$

ou, plus rarement, à l'entropie croisée :

$$C(w) = - \sum_{\alpha} (y_{\alpha} \ln y_{\alpha}^w + (1 - y_{\alpha}) \ln(1 - y_{\alpha}^w))$$

où w est le vecteur des poids, les $y_{\alpha}(w)$ sont les sorties du modèle pour le neurone α et les y_{α} représentent les sorties réelles de la base de données. Les sorties suivent donc, par hypothèse, une loi gaussienne.

Nous ferons référence à la fonction quadratique de coût dans ce mémoire.

Un algorithme d'apprentissage réagit par rapport à sa fonction de coût. Dans l'espace des arguments de cette fonction, l'algorithme cherche une direction qui lui permet de diminuer la valeur de cette fonction. Il se déplace dans le sens de cette direction et réexécute la même opération afin d'atteindre un critère d'arrêt défini. Chaque algorithme a ses propres pas et direction de descente.

Le cas des réseaux de neurones est particulier. Dans la mesure où un réseau de neurones est un approximateur non-linéaire, la fonction de coût qui lui est associée n'est pas quadratique par rapport aux paramètres de ce réseau. Pour minimiser la fonction de coût en s'inspirant des méthodes des moindres carrés, les réseaux de neurones font appel à des méthodes itératives s'appuyant sur la notion de **gradient**.

Il existe plusieurs types d'apprentissage : algorithme de Levenberg-Marquardt, algorithme de descente du gradient,... Nous présentons ci-après l'un des algorithmes d'apprentissage les plus connus : l'algorithme de rétropropagation du gradient.

Algorithme de rétropropagation du gradient

Le **gradient** d'un neurone désigne sa contribution à l'erreur globale imputable au modèle. Si on note E l'erreur globale du réseau, alors le gradient est noté $\frac{\partial E}{\partial w_{ij}}$.

Le fait de s'appuyer sur cette quantité pour apprécier les ajustements à apporter à un réseau n'est pas très robuste. En effet, la dimension de cette différentielle n'indique pas la proportion dans laquelle les poids doivent être modifiés.

L'algorithme de rétropropagation, développé par Rumelhart et al. (1986) [64] est un algorithme permettant de calculer la valeur du gradient de la fonction de coût à chaque itération. Il s'appuie pour cela sur le fait que le gradient de la fonction de coût est égal à la somme des fonctions de coût partielles $C^j(w)$:

$$\nabla C(w) = \sum_{\alpha} \nabla C^{\alpha}(w) \quad (7.2.2)$$

Nous calculons ensuite les dérivées partielles comme suit :

$$\left(\frac{\partial C^{\alpha}}{\partial w_{ij}} \right)_{\alpha} = \left(\frac{\partial C^{\alpha}}{\partial \nu_i} \right)_{\alpha} \left(\frac{\partial \nu_i}{\partial w_{ij}} \right)_{\alpha} = \delta_i^{\alpha} x_{ij}^{\alpha}$$

où, pour chaque exemple α :

- $\left(\frac{\partial C^{\alpha}}{\partial \nu_i} \right)_{\alpha} = \delta_i^{\alpha}$ représente la valeur du gradient de la fonction partielle de coût par rapport au potentiel lié au $i^{\text{ème}}$ neurone.
- $\frac{\partial \nu_i}{\partial w_{ij}} = x_{ij}^{\alpha}$ représente la valeur de la dérivée partielle du potentiel par rapport au paramètre w_{ij} . C'est également la valeur de l'entrée j du neurone i .

Le développement de δ_i^{α} diffère selon que l'on se trouve sur une couche interne ou externe :

- Dans le cas des couches externes, nous pouvons écrire :

$$\begin{aligned} \delta_i^{\alpha} &= \left(\frac{\partial C^{\alpha}}{\partial \nu_i} \right)_{\alpha} \\ &= \left(\frac{\partial C^{\alpha}}{\partial y_{\alpha}^w} \right)_{\alpha} \left(\frac{\partial y_{\alpha}^w}{\partial \nu_i} \right)_{\alpha} \\ &= -2(y_{\alpha} - y_{\alpha}^w) \left(\frac{\partial y_{\alpha}^w}{\partial \nu_i} \right)_{\alpha} \\ &= -2(y_{\alpha} - y_{\alpha}^w) y_{\alpha}^w (1 - y_{\alpha}^w) \end{aligned}$$

En considérant la fonction sigmoïde pour $y_{\alpha}^w = (1 + e^{-w})^{-1}$.

- Dans le cas des cellules cachées, l'entrée de la $i^{\text{ème}}$ cellule est fonction des k neurones cachés auxquels elle est attachée. Ceci s'écrit comme suit :

$$\begin{aligned}
\delta_i^\alpha &= \left(\frac{\partial C^\alpha}{\partial \nu_i} \right)_\alpha \\
&= \sum_k \left(\frac{\partial C^\alpha}{\partial \nu_k} \right)_\alpha \left(\frac{\partial \nu_k}{\partial \nu_i} \right)_\alpha \\
&= \sum_k \left(\frac{\partial C^\alpha}{\partial \nu_k} \right)_\alpha \left(\frac{\partial \nu_k}{\partial y_i^w} \right)_\alpha \left(\frac{\partial y_i^w}{\partial \nu_i} \right)_\alpha \\
&= \sum_k \left(\frac{\partial C^\alpha}{\partial \nu_k} \right)_\alpha w_{ki} y_i^w (1 - y_i^w) \\
&= y_i^w (1 - y_i^w) \sum_k w_{ki} \delta_k^\alpha
\end{aligned}$$

Finalement, pour un quelconque exemple de la base, nous avons :

$$\frac{\partial C}{\partial w_{ij}} = \begin{cases} 2(y - y^w)x_{ij}y^w(1 - y^w) & \text{dans la couche externe} \\ y_i^w(1 - y_i^w)x_{ij} \sum_k w_{ki} \delta_k & \text{dans les couches internes} \end{cases}$$

En réalité, l'algorithme de rétropropagation de l'erreur ajuste les poids du réseau selon les changements de signe du gradient plutôt que des valeurs que prennent ces poids.

Autrement dit, si on note $w_{i,j}$ le poids lié au $i^{\text{ème}}$ individu de la $j^{\text{ème}}$ classe, alors l'algorithme de rétropropagation du gradient propose de considérer que les variations des poids s'écrivent :

$$\Delta w_{i,j} = \epsilon \frac{\partial E}{\partial w_{ij}} x_{i,j}$$

où ϵ est le **taux d'apprentissage** lié au modèle. Ce paramètre contrôle la vitesse de déplacement du gradient. Ce taux contrôle la propension dans laquelle les poids varient durant la phase d'apprentissage.

De plus, il est d'usage d'ajouter, à chaque itération t , un terme d'inertie, appelé **moment** afin de réduire les oscillations du modèle. Nous le notons \mathbf{m} :

$$\Delta w_{ij}^{t-1} = \epsilon \frac{\partial E}{\partial w_{ij}} x_{i,j} + m w_{ij}^{t-1}$$

L'algorithme de rétropropagation du gradient est donc basé sur :

- Une étape de propagation durant laquelle l'entrée de chaque neurone est calculée à partir des neurones de la couche précédente en utilisant une fonction de combinaison et une fonction d'activation adaptées ;
- Une étape de rétropropagation où s'effectue un calcul récursif des gradients partiels de la fonction de coût. La somme de ces gradients partiels permet ensuite d'obtenir la fonction de coût du modèle.

7.2.5 Conclusion

Il ne s'agit pas, dans ce mémoire, d'occulter les défauts inhérents à l'utilisation des réseaux de neurones. Ce ne sont pourtant pas des inconvénients propres à ces modèles mais à toutes les techniques d'apprentissage basées sur une approche non-linéaire. Les plus notables dans le cas des réseaux neuronaux sont le risque de surajustement, le risque d'arrêt aux minima locaux et l'opacité de traduction des résultats. Les deux premiers ont été abordés dans la partie relative aux arbres de régression. Quant à l'opacité des résultats, elle peut être rattrapée par une étude préalable de la base de données et par une modélisation parallèle via une autre méthode prédictive.

Nous soulignons enfin que les réseaux de neurones, bien que peu répandus dans le milieu actuariel, possèdent des qualités qui en font un outil de prédiction et de modélisation très intéressants :

7.2.5.1 La parcimonie

La parcimonie d'un modèle représente sa capacité à venir à bout de la fonction pour laquelle il a été construit en manipulant le moins de paramètres possible.

En ce sens, les réseaux de neurones sont des modèles d'approximation non-linéaire parcimonieux. En effet, lors de la phase d'ajustement des poids, le réseau, sans nécessairement éprouver ses arguments (nombre de couches cachées, ...) s'adapte aux données et épouse leur structure. Ceci est particulièrement lié au sens de variation des poids qui évolue de manière linéaire par rapport au nombre de variables à approcher.

7.2.5.2 L'approximation

Les réseaux de neurones sont des approximateurs universels. Cybenko (1989) [15], Funahashi (1989) [31] et Hornick (1989) [34] ont démontré la propriété suivante :

" Toute fonction bornée suffisamment régulière peut être approchée uniformément, avec une précision arbitraire, dans un domaine fini de l'espace de ses variables, par un réseau de neurones comportant une couche de neurones cachés en nombre fini, possédant tous la même fonction d'activation et un neurone de sortie linéaire. "

Cette assertion n'indique pour autant pas comment approcher une fonction qui respecterait de telles conditions. En effet, l'appréciation des paramètres du modèle est à la charge du constructeur de ce modèle, cette propriété n'étant qu'un théorème d'existence.

La prochaine section aborde le sujet de l'agrégation de modèles. Les méthodes d'agrégation sont particulièrement salutaires dans le cas des approximateurs peu robustes (les méthodes non linéaires en général) puisqu'elles permettent de réduire le risque de surapprentissage et celui du choix d'optimum locaux.

Chapitre 8

Agrégation de modèles

La qualité d'un modèle statistique est liée à sa précision et à sa robustesse : augmenter la première consiste à diminuer le biais du modèle, la seconde, à baisser sa variance. Il existe plusieurs techniques pour affiner ces deux points : en agrégeant des modèles issus d'une même méthode, en combinant des modèles construits grâce à différents procédés, en mélangeant ces techniques,... Ces méthodes sont particulièrement intéressantes dans le cas de modèles instables tels que les arbres de décision et les réseaux de neurones.

Le principe de l'agrégation de modèles consiste à bousculer le déroulement de l'apprentissage en modifiant l'échantillon d'apprentissage ou en faisant varier les paramètres de l'apprentissage sur un même échantillon.

Les modèles que l'on souhaite agréger ne doivent pas présenter des résultats sensiblement divergents : l'agrégation n'aurait pas de sens. Ainsi, la confrontation de modèles agrégés tendrait à mutualiser leurs conclusions, le but étant d'optimiser leurs algorithmes pour les faire converger vers une solution commune. Les méthodes d'agrégation de modèles les plus populaires, que nous présentons dans ce chapitre, sont le Bagging et le Boosting.

8.1 Bagging

Le Bagging est un processus aléatoire d'agrégation de modèles développé en 1996 par Breiman [42]. Grâce à un classement ou à une régression, le Bagging trouve le modèle "moyen" approchant le mieux les données d'intérêt. Le modèle issu du Bagging n'est pas forcément celui qui présente les meilleures performances parmi tous les modèles construits mais ses performances sont nettement supérieures aux performances moyennes.

L'une des principales caractéristiques du Bagging est sa capacité à diminuer la variance d'un modèle. Cela s'explique par le fait que, pour n échantillons décorrés, de même variance V , la variance du modèle issu du Bagging vaut $\frac{V}{n}$. Cette moyenne est naturelle dans le cas d'un modèle où la variable d'intérêt est numérique. Dans le cas où elle serait qualitative, le modèle agrégé établit une moyenne des probabilités de réponse de chaque modalité.

Enfin, cette réduction de variance entraîne la réduction de l'erreur de prévision.

En réalité, les échantillons concernés par le Bagging ne sont pas indépendants : ce sont des échantillons bootstraps obtenus par des tirages aléatoires avec remise de n individus parmi n . C'est l'**erreur out-of-bag** qui permet, durant l'exécution de l'algorithme de Bagging, de contrôler le nombre de ces échantillons (voire de l'optimiser). En effet, l'erreur out-of-bag, particulièrement sensible au nombre d'échantillons bootstrap, varie avec l'évolution de ce nombre et se stabilise dès qu'est atteint le nombre nécessaire d'échantillons à l'agrégation.

Cas particulier : Forêts aléatoires

Une forêt aléatoire est une agrégation d'arbres de décision. La particularité d'une forêt aléatoire par rapport à un Bagging classique (réputé inefficace sur les *stumps* ou arbres binaires) consiste à introduire une étape supplémentaire de tirage aléatoire sur l'ensemble des variables prédictives. En effet, sur chaque échantillon bootstrap, l'algorithme de construction de la forêt va :

- Tirer aléatoirement k variables parmi les K variables explicatives de la base de données. En général, k est plus ou moins égal à la racine carrée (*resp.* du tiers) du nombre de variables explicatives du modèle dans le cas d'une classification (*resp.* d'une régression) ;
- Chercher le nœud optimal au sens de la méthode choisie (voir le paragraphe ci-après) ;
- Itérer le processus pour un nombre N d'arbres, initialisé par l'utilisateur ;
- Calculer, sur l'ensemble des arbres construits, la moyenne (ou la moyenne des probabilités de chaque modalité) de la variable d'intérêt.

Plusieurs constructions de forêts ont été déclinées selon le moment et l'endroit à partir duquel on introduit de l'aléatoire dans la construction de la forêt. Ainsi, dans un *random subspace*, une sélection aléatoire est réalisée au niveau de chaque nœud pour sélectionner α variables, dans un *radom select split*, la sélection s'opère au niveau des séparations des nœuds : la meilleure séparation est choisie par β séparations possibles au niveau de chaque nœud, ...

Enfin, les forêts aléatoires sont particulièrement intéressantes pour leur vitesse de convergence vers une solution optimale [8] dans le cas d'un nombre limité de variables explicatives pertinentes.

8.2 Boosting

Proposée en 1990 par Schapire [66] puis développée en 1996 par Freund et Shapire [65], le Boosting est un processus adaptatif (à chaque itération le modèle est une version adaptative du précédent), souvent déterministe, d'agrégation de modèles.

Le Boosting a cela de commun avec le Bagging qu'il promet une meilleure qualité de prédiction en agrégeant plusieurs modèles. Cependant, plusieurs points le distinguent du Bagging :

- Utilisation de l'échantillon d'apprentissage entier et non d'échantillons bootstrap, hormis dans le cas du boosting dit *Arcing* (voir la fin de ce chapitre) ;
- Ajustement chaîné des poids des individus. A chaque itération, l'algorithme de Boosting se concentre sur les individus les plus difficiles à classer. Il augmente leurs poids et ne réagit pas par rapport au reste des individus : leur ajustement ne se fera qu'à l'itération suivante sans que les individus ayant été préalablement classés ne le soient à nouveau ;
- Le Boosting permet de réduire le biais d'un modèle. Dans le cas du Bagging, la moyenne de la somme des espérances des α échantillons bootstrapés est égale à l'espérance d'un seul modèle : le biais du modèle final est inchangé contrairement au cas du Boosting ;
- Le modèle agrégé est basé sur une pondération de chaque modèle par sa qualité d'ajustement (hormis dans le Boosting dit *Arcing* où tous les modèles ont le même poids). L'étape de pondération des modèles peut aboutir à une croissance du risque d'apprentissage et donc de l'erreur de prédiction, ce risque est toutefois atténué en l'absence de bruit ;

- L'algorithme de Boosting est sensible au bruit et aux valeurs extrêmes. Ceci est particulièrement dû à la concentration de l'algorithme sur des populations limitées à chaque itération et à la pondération des modèles par leurs qualités d'ajustements.

Il existe plusieurs techniques de Boosting, adaptées à la nature de la variable d'intérêt. Les plus populaires sont :

- L'*AdaBoost* (Adaptative Boosting : historiquement, premier modèle de Boosting) permet d'approcher une variable binaire ;
- L'*Arcing* (Adaptively Resample and Combine) représente une version aléatoire de Boosting où, ce n'est plus l'échantillon d'apprentissage entier qui est utilisé mais des échantillons constitués à chaque itération, à l'aide d'un tirage avec remise et selon des probabilités inversement proportionnelles à la qualité d'ajustement du précédent tirage.

8.3 Intérêt dans le cadre de l'étude

La confrontation de méthodes de Boosting et de Bagging dans le cadre de ce mémoire s'est inspirée d'une étude menée par Optiz et Maclin [25].

Dans cette étude, les auteurs font appel à l'AdaBoosting et à l'Arcing comme méthodes de Boosting ainsi qu'aux forêts aléatoires et au Bagging de réseaux de neurones.

Ils démontrent que :

- L'agrégation par Bagging est presque toujours plus performante qu'un modèle simple ;
- L'agrégation par Boosting est globalement plus performante que celle par Bagging (et donc qu'un modèle simple) ;
- En présence de bruit, il y a un risque important de sur-apprentissage lié à l'agrégation par Boosting ;
- L'agrégation par Boosting dépend dans une certaine mesure de la base étudiée, cette dépendance est en tout cas plus affirmée que dans le cas de l'agrégation par Bagging ;
- L'importance des variables explicatives dans les modèles obtenus par agrégation est liée à un ensemble limité de variables explicatives ;
- L'agrégation par Boosting des arbres de décision performe d'autant plus que l'échantillon d'apprentissage est important.

Les auteurs concluent alors en mettant en avant l'agrégation par Bagging comme plus performante dans le cas des arbres de décision et des réseaux de neurones bien que l'agrégation par Boosting ait abouti à une augmentation de la précision dans certains cas.

Chapitre 9

Application au portefeuille de Natixis Assurances

L'Analyse Factorielle des Données Mixtes a souligné qu'au regard des autres variables de l'étude, le sexe et l'assiette du rachat contribuent faiblement à l'inertie globale du portefeuille étudié. Elles sont donc mises à l'écart pour l'exécution de nos méthodes d'apprentissage.

9.1 Arbres de régression

Nous souhaitons constituer des populations aux lignes définies. Cependant, l'âge, l'ancienneté et le TMG sont des variables numériques pouvant prendre un nombre conséquent de valeurs. En l'état, ces variables risquent d'étouffer rapidement les nœuds d'un arbre de régression. Nous optons alors pour une distribution plus transparente des variables citées plutôt :

- Les âges de nos assurés sont répartis en deux catégories : les plus et les moins de 70 ans, reprenant ainsi les classes des catégories fiscales ;
- Les anciennetés sont distinguées jusqu'à 7 ans, puis regroupées pour celles supérieures ou égales à 8 ans ;
- Les TMG sont distingués en trois catégories : nuls, supérieurs et inférieurs au Taux de Rendement des Actifs (TRA).

La construction d'arbres de régression a été réalisée ici grâce à la librairie *rpart* du logiciel R.

La construction de l'arbre repose sur la détermination d'un coefficient de pénalisation (noté cp par la suite) optimal. Ce facteur définit la complexité de l'arbre : plus le cp est important, moins l'arbre aura de segmentations. Ce coefficient sera ajusté en fonction des résultats d'un premier arbre.

L'arbre est construit sur un échantillon d'apprentissage de 98 156 assurés et se présente comme suit (Figure 9.1) :

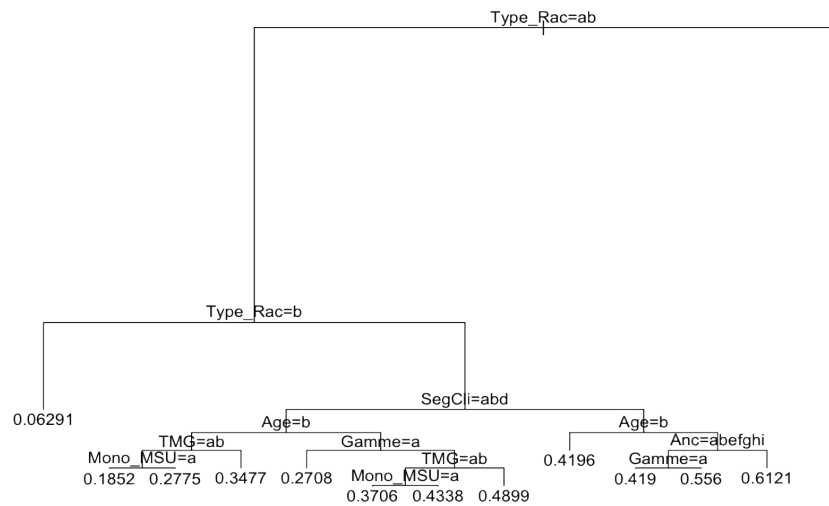


FIGURE 9.1 – Arbre de régression sur le Taux de rachat - Initialisation du coefficient de pénalisation

La décroissance du cp en fonction de l'erreur relative (Figure 9.2) donne des pistes pour l'élagage de l'arbre. En effet, l'erreur augmente lorsque le cp augmente et donc lorsque le nombre de nœuds diminue.

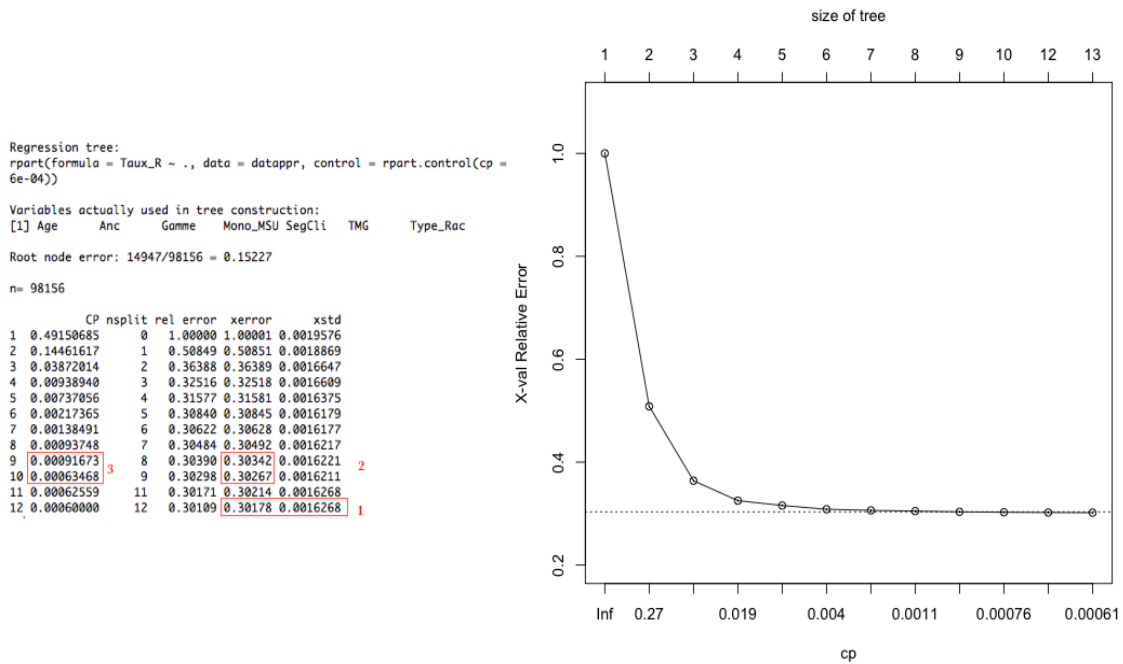


FIGURE 9.2 – Coefficients de pénalité - Elagage de l'arbre de régression

La règle de Breiman [42] recommande un seuil d'erreur inférieur à

$$0,30178 + 0,0016268 = 0,3034068$$

Autrement dit, un cp compris, strictement, entre 0,00063468 et 0,00091673.

Nous choisissons un cp de 0,0007

En effet, en dessous d'un certain seuil (situé du côté du cp de 0,00217365), l'erreur relative commence à se stabiliser. Le cp choisi correspond bien à une valeur pour laquelle l'erreur n'évolue plus de façon significative.

L'arbre élagué prend la forme suivante :

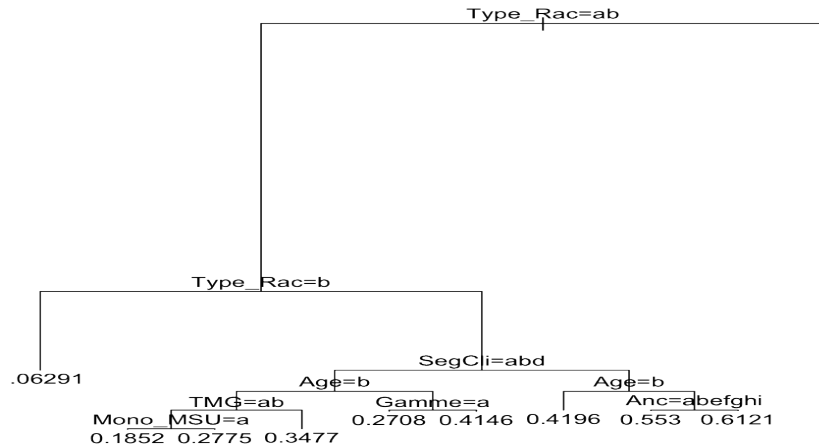


FIGURE 9.3 – Arbre de régression élagué

L'arbre retenu est constitué de 9 segmentations, il comporte donc 10 feuilles contre 13 pour l'arbre initial. Les quantités observées au niveau des feuilles de l'arbre correspondent à la valeur prédite moyenne du taux de rachat pour les populations de chaque feuille.

Les taux de rachats indiqués par l'arbre correspondent à une moyenne de taux de rachats des Model Points représentés. Il faut, pour avoir une vision plus réaliste des taux engagés, se ramener au niveau du portefeuille global. Pour cela, nous calculons, pour les groupes homogènes de risque identifiés, des taux de rachat en rapportant les montants de rachats de ces populations à la provision mathématique moyenne du portefeuille puis à la provision mathématique moyenne de chaque groupe. Les résultats obtenus sont exposés sur le tableau ci-aprèsⁱ :

i. Pour des raisons de confidentialité, un coefficient multiplicateur a été appliqué aux deux premières colonnes de ce tableau

en lisant l'arbre de gauche à droite

Population d'assurés ayant racheté leur police en 2010	Taux cohorte ayant racheté	Taux rachat cohorte sur PM Globale	Taux affiché par l'arbre
RPP	5,21%	0,33%	6,29%
RP, Fortuné/Aisé/Patrimonial, >70ans, TMG<TRA ou TMG=0, EUR	17,62%	0,24%	18,52%
RP, Fortuné/Aisé/Patrimonial, >70ans, TMG<TRA ou TMG=0, MSU	26,49%	0,25%	27,75%
RP, Fortuné/Aisé/Patrimonial, >70ans, TMG>TRA	27,53%	0,08%	34,77%
RP, Fortuné/Aisé/Patrimonial, <70ans, Gestion de Patrimoine	31,58%	0,45%	27,08%
RP, Fortuné/Aisé/Patrimonial, <70ans, Grand Public	34,72%	1,10%	41,46%
RP, Clientèle Grand Public, >=70ans	46,51%	0,28%	41,96%
RP, Clientèle Grand Public, <70ans, anciennetés 0, 1 et supérieures ou égales à 4	61,96%	0,88%	55,30%
RP, Clientèle Grand Public, <70ans, anciennetés 2 et 3	54,18%	0,16%	61,21%
RT	120,00%	1,39%	100,00%

FIGURE 9.4 – Taux de rachat par populations d'assurés

L'arbre élagué peut être également présenté sous forme détaillée (Figure 9.5) :

```
n= 98156
node), split, n, deviance, yval
* denotes terminal node

1) root 98156 14946.5900 0.50623790
 2) Type_Rac=RP,RPP 75101 7600.2390 0.35465960
 4) Type_Rac=RPP 18977 223.6302 0.06290536 *
 5) Type_Rac=RP 56124 5215.0900 0.45330940
 10) SegCli=Aisé, Fortuné, Patrimonial 22107 2022.0720 0.32734480
 20) Age=>=70ans 9150 679.8275 0.23253210
 40) TMG=Sans TMG, TMG<TRA 7817 533.0153 0.21289090
 80) Mono_MSU=Monoc€ 5471 365.0888 0.18516650 *
 81) Mono_MSU=MSU 2346 153.9143 0.27754580 *
 41) TMG=TMG>=TRA 1333 126.1125 0.34771200 *
 21) Age=<70ans 12957 1201.9050 0.39429990
 42) Gamme=GesPa 1829 114.6477 0.27078610 *
 43) Gamme=GrPu 11128 1054.7680 0.41460060 *
 11) SegCli=GrandPublic 34017 2614.2840 0.53517130
 22) Age=>=70ans 6641 524.1841 0.41962880 *
 23) Age=<70ans 27376 1979.9350 0.56320020
 46) Anc=>8ans, 0, 3, 4, 5, 6, 7 22630 1656.1750 0.55295480 *
 47) Anc=1, 2 4746 310.0583 0.61205240 *
 3) Type_Rac=RT 23055 0.0000 1.00000000 *
```

FIGURE 9.5 – Arbre de régression élagué - version développée

C'est cette représentation qui permet de tirer des conclusions sur les relations non linéaires détectées par l'arbre, permettant ainsi d'identifier des groupes homogènes :

- * Le type de rachat est le critère le plus discriminant du jeu de données étudié. Les rachats totaux caractérisent une population foncièrement différente de celle qui effectue des rachats partiels et partiels programmés. Cette population constitue 23,49 % de la population étudiée.

A ce niveau de l'arbre, seuls les rachats partiels semblent candidats à une segmentation plus fine ;

- * Le segment de clientèle est le second critère le plus discriminant. Il est intéressant de remarquer qu'à l'instar de l'analyse factorielle, l'arbre de régression détecte des comportements différents entre les segments de clientèle Aisé, Fortuné et Patrimonial (abd au niveau du nœud n°10) et les clients Grand Public ;
 - Au niveau des clients aisés, l'âge oriente la décision de rachat : alors que les assurés âgés de plus de 70 ans rachètent selon la nature de leur contrat (Mono euro/Multi-supports) et du TMG servi, les plus jeunes se concentrent davantage sur le type de produit pour lequel ils ont souscrit ;
 - Les clients "Grand Public" (modalité du segment de clientèle) rachètent différemment leur contrat selon leur âge. Ceux âgés de moins de 70 ans ont deux comportements distincts selon leur ancienneté : nous notons une population plus importante ayant une ancienneté entre 1 et 2 ans. L'explication de ce phénomène est principalement commerciale. En effet, les conseillers bancaires sont objectivés par le nombre d'affaires nouvelles sans être pénalisés par les sorties. Ce phénomène est d'autant plus accentué auprès d'une jeune clientèle : ceci est constaté commercialement mais c'est également une tendance détectée par l'arbre.

L'arbre ainsi construit doit prouver ses capacités de généralisation. Pour cela, nous estimons l'erreur de prévision de l'arbre retenu en l'appliquant à l'échantillon de test.

Nous proposons dans un premier temps d'évaluer une matrice de confusion sur un cas précis. Nous estimons ainsi les erreurs commises en prédisant les taux de rachat inférieurs à 15% : au niveau du portefeuille, ce sont globalement les taux correspondants aux rachats partiels programmés.

Valeurs estimées \ Valeurs observées	Taux de rachat \geq 15%	Taux de rachat $<$ 15%
Taux de rachat \geq 15%	16 539	3 252
Taux de rachat $<$ 15%	339	4 409

TABLE 9.1 – Evolution de l'erreur relative en fonction du facteur de complexité

En disposant de 19 791 contrats dont le taux de rachat excède les 15%, le modèle a un risque d'erreur de $16.43\% = \frac{3252}{16539+3252}$. En disposant de 4 748 contrats dont le taux de rachat est inférieur à 15%, le modèle a un risque d'erreur de $0.7\% = \frac{339}{339+4409}$.

Cet exemple précis ne suffit pas à juger des capacités de généralisation de l'arbre. Nous calculons donc l'erreur quadratique moyenne de prévision et obtenons un taux de 10,4%.

Le modèle apporte une estimation intéressante de l'erreur mais jugée encore insatisfaisante dans la mesure où un seul arbre a été utilisé pour accorder ces résultats. La partie dédiée à l'agrégation de modèles permettra de consolider ou d'infirmer le bilan de l'arbre.

Conclusion

Les premiers constats confortent les observations établies par l'AFDM (voir chapitre 5) :

- Le **segment de clientèle** ainsi que le **type de rachat** sont des critères fortement discriminants au rachat d'un contrat d'épargne. **Ce sont des éléments à mettre en avant dans les études de rentabilité** ;
- La **gamme de produits** semble mise à l'écart par cette étude alors qu'est souligné le lien entre l'**ancienneté** des contrats et le **type de contrat** (Monosupport €/Multisupports) racheté ;
- L'**âge** des assurés oriente leur décision de rachat : à une tranche donnée correspondent des besoins et une réalité financière qui conditionnent les critères du rachat.

9.2 Réseaux de neurones

Nous rappelons que dans le cadre de cette étude, le modèle choisi est celui du **perceptron multi-couches à rétropropagation du gradient de l'erreur**.

Nous présentons les étapes de modélisation puis les résultats de l'analyse.

Etapes de modélisation

L'implémentation de notre réseau de neurones est réalisée selon les étapes suivantesⁱⁱ :

- Prétraitement des données :
 - ⊗ Nous appliquons la fonction logarithme aux variables continues avant de les normaliser. Le passage au logarithme sert à limiter l'écrasement des valeurs moyennes par de potentielles données extrêmes ;
 - ⊗ Nous créons des variables pour les modalités des variables qualitatives, à l'image de la technique du tableau disjonctif complet (Cf paragraphe 4.1) ;
 - ⊗ Les variables discrètes qui prennent leurs valeurs entre 1 et $n \in \mathbb{N}$, sont transformées en remplaçant chaque $i^{\text{ème}}$ valeur par $\sum_{j=0}^{i-1} 2^{-j}$.
- Initialisation des poids :

Dans le cas des algorithmes utilisés dans cette étude (les fonctions appelées des packages *neuralnet*, *caret* notamment), les poids sont initialisés aléatoirement (ils seront ajustés au cours de l'apprentissage par l'algorithme, ici, celui de rétropropagation du gradient de l'erreur) ;
- Un échantillon d'apprentissage est soumis à l'algorithme afin qu'il en assimile la structure. Durant cette étape, l'algorithme :
 - ⊗ applique la fonction de combinaison et la fonction d'activation à la couche d'entrée et aux couches intermédiaires ;
 - ⊗ évalue la fonction de coût globale du modèle (Cf équation 7.2.2) ;
 - ⊗ rétropropage l'erreur pour calculer les erreurs individuelles du modèle ;

ii. voir TUFFERY S., *Data Mining et statistique décisionnelle - L'intelligence des données*, 2007

- ✘ ajuste les poids des neurones en partant des neurones terminaux vers les neurones intermédiaires ;
- ✘ répète l'opération jusqu'à atteindre un critère d'arrêt (seuil pour les dérivées partielles de la fonction d'erreur) ou un nombre défini d'itérations ;
- ✘ applique le modèle retenu à l'échantillon test pour en vérifier la généralisation.

Analyse des résultats

Les réseaux de neurones ont, notamment, deux particularités : ce sont des approxima-
teurs parcimonieux aux résultats graphiques limités.

En effet, une fois de premiers paramètres fixés, c'est le réseau qui s'adapte aux données. En revanche, l'interprétation des résultats ne bénéficie pas d'une panoplie de graphiques aussi riche que dans le cas des arbres de régression.

Nous nous reposons donc principalement sur l'analyse des résultats numériques que propose l'algorithme.

Nous utilisons les fonctions des packages *neuralnet* et *caret* pour notre étude.

Les sorties de modèles peuvent se présenter de manière graphique comme sur la figure 9.6. Dans ce graphique, nous observons les nœuds attachés aux inputs du modèle, la variable d'intérêt, les neurones cachés, les couches intermédiaires auxquels ils sont rattachés, les biais (en bleu) et enfin, les poids obtenus suite à la phase finale de l'apprentissage.

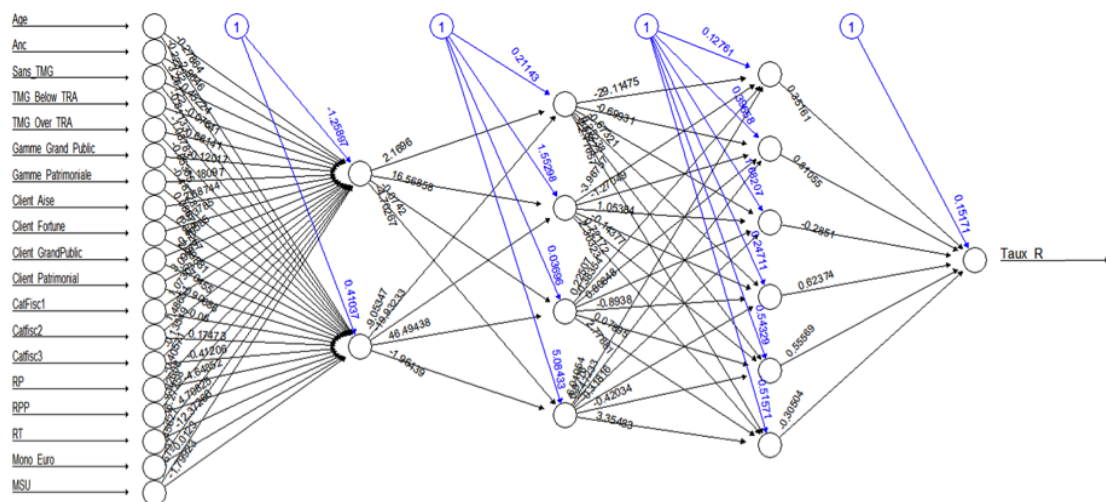


FIGURE 9.6 – Réseau de Neurones - Exemple d'un modèle à 3 couches intermédiaires de, successivement, 2, 4 et 6 neurones cachés

Les paramètres permettant d'obtenir ces modèles sont modulables en phase d'initialisation. Le réseau les adapte ensuite aux données. Parmi ceux auxquels nous appliquons des tests de sensibilité, jugés les plus pertinents dans le cas du PMC à rétropropagation du gradient : la fonction d'activation, le nombre de neurones cachés, le nombre de couches intermédiaires et le taux d'apprentissage. La fonction de combinaison est choisie égale à une combinaison linéaire de poids et d'entrées de neurones, faisant également intervenir un biais à chaque couche.

Les tableaux présentés ci-aprèsⁱⁱⁱ déclinent, pour chaque critère, l'erreur d'apprentissage, le nombre d'itérations nécessaires avant la stabilisation du modèle et l'erreur de prévision exécutée sur l'échantillon de test :

- * Le 1^{er} test concerne le **nombre de couches intermédiaires** (voir tableau de la figure 9.7) . Nous l'augmentons de 1 à 3. Pour chaque couche cachée, le nombre de neurones cachés est fixé à 6 : lorsque le réseau comporte deux (resp. trois) couches intermédiaires, chacune est composée de 3 neurones (resp. deux).

Plus le nombre de couches intermédiaires augmente, moins l'erreur d'apprentissage est importante. Le nombre d'itérations semble, lui, erratique, signe qu'il n'est pas nécessairement lié au nombre de couches intermédiaires.

Ce test met en avant le risque de surapprentissage lié à l'augmentation du nombre de couches intermédiaires. Plus ce nombre augmente, plus le réseau a tendance à mémoriser la structure des données qui lui sont présentées : l'erreur d'apprentissage baisse puisque le réseau assimile de mieux en mieux les données. Cela impactera toutefois la qualité des prévisions puisque le pouvoir de généralisation du réseau s'en trouve proportionnellement réduit.

Aussi, lors du choix du nombre de couches, nous gardons à l'esprit que l'intérêt d'un PMC réside dans la possibilité de choisir plus d'une couche intermédiaire mais que ce choix doit se faire au regard de l'étape de prévision.

	Nombre de couches intermédiaires **		
	1	2	3
Erreur d'apprentissage	43,386	43,335	43,213
Nombre d'itérations	14 109	43 422	23 594
Erreur de prévision	0,36	0,14	0,09

** avec 2 neurones cachés par couche

FIGURE 9.7 – Réseau de Neurones - Sensibilité du PMC au nombre de couches intermédiaires

- * Le 2nd test de sensibilité (voir tableau de la figure 9.8) concerne le **nombre de neurones cachés**. Pour une seule couche intermédiaire, nous le faisons varier entre 2, 4 et 6.

Ce test est délicat dans le sens où les théories autour de l'effet du nombre de neurones d'un réseau sur son pouvoir de généralisation ne sont pas unanimes. Nous nous appuyons ici sur les résultats de l'erreur d'apprentissage et de test :

Dans ce tableau, l'erreur d'apprentissage baisse lorsque le nombre de neurones augmente alors que l'erreur de prévision suit la tendance inverse. Ceci étant, l'évolution de l'erreur de prévision est plus timide que celle de l'erreur d'apprentissage, nous en déduisons que le risque de surapprentissage n'est pas accru du fait d'un choix important de neurones (contrairement au cas des couches intermédiaires).

- * Le 3^{ème} test (voir tableau de la figure 9.7) concerne la **fonction d'activation**. Nous l'indiquons plus tôt (voir paragraphe 7.2.2.3), la fonction sigmoïde assure des sorties non linéaires. Une telle fonction d'activation fait baisser l'erreur d'apprentissage mais fait surtout baisser le nombre d'itérations avant atteinte d'un modèle optimal : le gain de temps en apprentissage est 5 fois meilleur que dans le cas d'une fonction linéaire !

iii. les graphiques correspondant à chaque paramètre éprouvé sont présentés en Annexes VI

	Nombre de neurones cachés *		
	2	4	6
Erreur d'apprentissage	43,925	43,732	43,213
Nombre d'itérations	25 926	8 287	14 109
Erreur de prévision	0,34	0,37	0,42

* avec une couche cachée

FIGURE 9.8 – Réseau de Neurones - Sensibilité du PMC au nombre de neurones cachés

	Fonction d'activation	
	Linéaire	Sigmoïde ***
Erreur d'apprentissage	43,732	43,702
Nombre d'itérations	25 926	4 536
Erreur de prévision	0,37	0,12

*** tangente hyperbolique

FIGURE 9.9 – Réseau de Neurones - Sensibilité du PMC à la fonction d'activation

* Dans le 4^{ème} test (voir tableau de la figure 9.10), nous exhibons le pouvoir de l'un des paramètres les plus sensibles d'un réseau de neurones : le **taux d'apprentissage**. Ce taux représente la vitesse à laquelle le réseau va parcourir les données. Nous le fixons d'abord à 0,9 puis le faisons diminuer à 0,5 puis à 0,1. Il faut revenir sur le principe d'utilisation du taux d'apprentissage par le réseau : le réseau prend compte du taux d'apprentissage qui lui a été indiqué en paramètre et parcourt une première fois l'ensemble des données. Plus ce taux est grand, plus la vitesse de parcours des données est importante. Le réseau fait diminuer ce taux au fur et à mesure que s'enchaînent les itérations. La baisse progressive du taux d'apprentissage est ce qui permet au réseau de détecter les optimum locaux sur lesquels il ne s'était pas arrêté lorsque sa vitesse (le taux d'apprentissage) était élevée. Nous comprenons l'enjeu majeur derrière ce taux : s'il est très élevé, le réseau peut ne détecter aucun optimum, s'il est très petit, il y a risque de surapprentissage puisque le réseau peut s'arrêter à un optimum local et le désigner comme optimum global. La généralisation du modèle s'en trouve alors faussée.

Dans le cas présent, pour un taux d'apprentissage élevé (0,9), le réseau parcourt rapidement les données (nombre d'itérations modeste). Sa vitesse de parcours est moins importante et conséquemment, son nombre d'itérations est plus grand lorsque ce taux est bas.

	Taux d'apprentissage		
	0,9	0,5	0,1
Erreur d'apprentissage	43,529	43,532	43,534
Nombre d'itérations	40 653	65 339	98 054
Erreur de prévision	0,66	0,40	0,02

FIGURE 9.10 – Réseau de Neurones - Sensibilité du PMC au taux d'apprentissage

Au regard des résultats obtenus, nous construisons un modèle avec 2 couches intermédiaires à 6 neurones cachés (3 par couche). La fonction sigmoïde (tangente hyperbolique) est celle utilisée et le taux d'apprentissage est fixé à 0,25.

Les résultats obtenus pour ce modèle sont présentés sur le tableau de la figure 9.11.

Erreur d'apprentissage	43,040
Nombre d'itérations	101 582
Erreur de prévision	0,04

FIGURE 9.11 – Réseau de Neurones - Résultats d'apprentissage et de prévision sur un PMC {3 3} à fonction d'action sigmoïde et à taux d'apprentissage de 25%

Le modèle montre un taux d'erreur 9.12 intéressant bien que le temps d'apprentissage soit un peu lourd du fait du taux d'apprentissage peu élevé. Ainsi, et bien que le modèle présente un taux d'erreur intéressant, il n'en demeure pas moins sensible à des minima locaux. L'étape d'agrégation permet d'optimiser les chances d'arriver à un minima global. C'est l'objet de la section suivante.

```
> MSE.nn <- sum((test.r - pr.nn_)^2)/nrow(datestr_1)
> print(paste(MSE.nn))
[1] "0.0443613878986652"
```

FIGURE 9.12 – Réseau de Neurones - Erreur de prévision sur un PMC {3 3}

9.3 Agrégation de modèles

Dans cette partie, nous consolidons les résultats obtenus par les arbres de régression et par les réseaux de neurones, l'objectif étant d'augmenter la robustesse de chaque méthode en les faisant converger vers des optimums globaux.

Nous agrégeons plusieurs modèles de même type dans un premier temps. Nous nous prononçons en conclusion de ce chapitre sur l'efficacité de chaque méthode au regard de l'erreur de prévision qu'elles procurent.

9.3.1 Forêt aléatoire

La 1^{ère} technique d'agrégation proposée est celle des forêts aléatoires. Nous utilisons pour cela la fonction *randomForest* du package du même nom.

Nous appliquons cette fonction à l'échantillon d'apprentissage (figure 9.13).

```
> rf<-randomForest(Taux_R~.,data=datappr,importance=T,proximity=T)
> rf

Call:
randomForest(formula = Taux_R ~ ., data = datappr, importance = T, proximity = T)
Type of random forest: regression
Number of trees: 500
No. of variables tried at each split: 2

Mean of squared residuals: 0.04771883
% Var explained: 68.51
```

FIGURE 9.13 – Forêt Aléatoire - Premiers paramètres

Arbitrairement, le nombre d'arbres utilisé est égal à 500. La fonction teste deux variables à la fois à chaque division de nœud. La variance expliquée par le modèle est de 68,5%. La moyenne de la somme des carrés des erreurs (MSE) vaut 4,77%.

Nous faisons varier deux paramètres par la suite : le nombre d'arbres et le nombre de variables à chaque nœud de l'arbre.

Le nombre d'arbres est initialisé à 500 puis augmenté par pas de 500 jusqu'à atteindre les 1500 arbres. Celui des variables à chaque nœud varie entre 2 et 4. Nous obtenons ainsi :

Nombre d'arbres	MSE	% de la variance expliquée
500	4,77%	68,51%
1000	4,84%	68,05%
1500	4,42%	68,18%

Nombre de variables à chaque nœud	MSE	% de la variance expliquée
2	4,77%	68,51%
3	4,6%	70,14%
4	4,64%	69,85%

Concernant le nombre d'arbres, les quantités observées n'évoluent pas de façon significative. Cette stabilité commence en réalité avant le seuil des 500 arbres. La figure 9.14 montre en effet qu'hormis de très légers pics, le seuil des 300 arbres est le point de départ de cet état de stabilité.

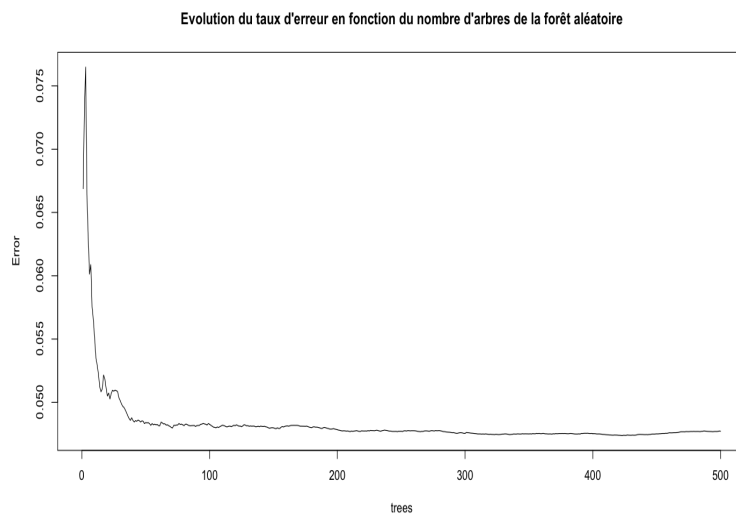


FIGURE 9.14 – Forêt Aléatoire - Taux d'erreur et nombre d'arbres

Par ailleurs, la variation du nombre de variables à chaque nœud ne montre pas d'évolution particulière dans la variation du MSE et du pourcentage de la variance expliquée. Nous choisissons toutefois de garder 3 variables par nœuds dans la mesure où nous observons pour ce nombre de variable les quantités les plus intéressantes pour ces deux indicateurs.

Ainsi, pour la suite de cette analyse, nous fixons le nombre d'arbres à 500 et le nombre de variables à chaque nœud à 3.

L'intérêt est ensuite porté à l'importance des variables explicatives (figure 9.15).

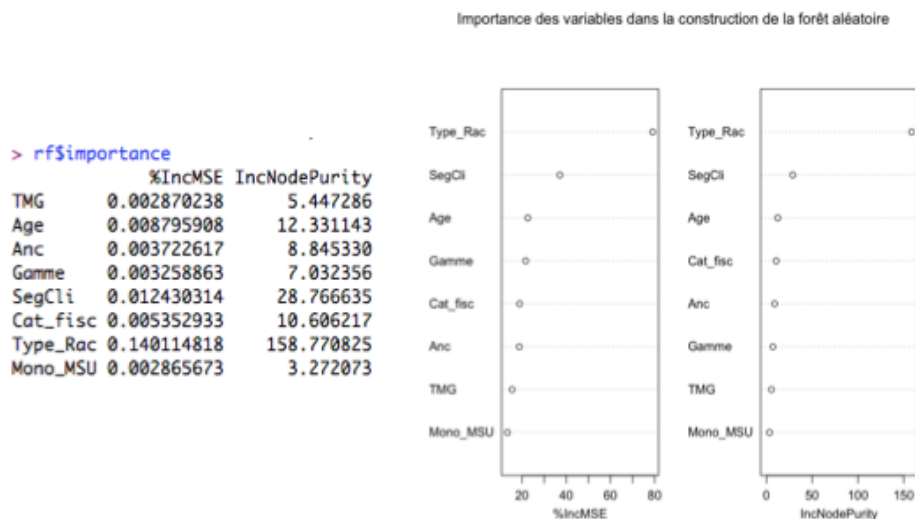


FIGURE 9.15 – Forêt Aléatoire - Importance des variables

Deux quantités sont calculées dans le cas de la régression : la **décroissance moyenne de la précision** ($\%IncMSE$) et la **décroissance moyenne de la pureté des nœuds** ($IncNodePurity$).

Le calcul est réalisé selon les étapes suivantes :

Le modèle évalue l'erreur de prédiction avec un positionnement initial des variables explicatives. Les valeurs de ces dernières sont ensuite permutées. L'erreur de prédiction est à nouveau évaluée. Ce calcul est effectué pour les différents arbres de la forêt. Pour chaque variable, les différences entre les erreurs avant et après permutations, ainsi que la somme des carrés des erreurs, sont normalisées (par le nombre d'arbres) et divisées par l'écart-type des différences si cet écart-type est non nul.

Les deux indicateurs évoluent globalement dans le même sens hormis pour la variable Gamme. Lorsque, durant le processus d'apprentissage, les modalités de la variable Gamme (Grand Public, Gestion de Patrimoine) sont permutées par le modèle, la pureté des nœuds ne décroît pas à la même vitesse que la somme des carrés des erreurs. L'instabilité de cette variable avait été mise en exergue lors de l'étude des arbres de régression.

Par ailleurs, il est naturel de voir apparaître de façon prédominante le type de rachat en guise de variable contributive dans la construction du taux de rachat (souvenons-nous de la phase de construction de cette variable (chapitre 3)). Le Segment de Clientèle et l'Âge des assurés s'imposent ensuite dans l'ordre des variables explicatives du taux de rachat.

9.3.2 Boosting d'arbres de régression

Pour obtenir un modèle de Boosting appliqué à des arbres de régression, nous utilisons les packages *gbm* et *caret* du logiciel R.

Le boosting d'arbres de régression est sensible à trois paramètres : le nombre d'arbres de la forêt, la profondeur de la forêt et le taux d'apprentissage.

- Nous fixons le nombre initial d'arbres à 50 et l'incrémentons progressivement de 50 jusqu'à atteindre 500 arbres, seuil que nous avons fixé lors de la modélisation des forêts aléatoires ;

- La profondeur de la forêt désigne ici la complexité des arbres qui la composent. Cette complexité est relative au nombre de nœuds à chaque segmentation des arbres. Nous proposons 5 niveaux de complexité afin d'évaluer la sensibilité du modèle à ce paramètre ;
- Le taux d'apprentissage est ici fixé à 0,001 pour éviter le risque de surapprentissage, bien qu'il soit réduit dans le cas des modèles agrégés.

L'agrégation par Boosting se fait ici sur 200 modèles bootstrap construits pour chaque groupe d'arbres : 200 modèles bootstrap pour les 50 premiers arbres, puis 200 modèles bootstrap sur les 100 arbres obtenus après incrémentation du nombre d'arbres, etc.

Dans le cas d'une régression, le modèle calcule l'erreur quadratique moyenne (RMSE) et le R^2 à chaque itération.

L'algorithme sélectionne le modèle minimisant l'erreur quadratique moyenne.

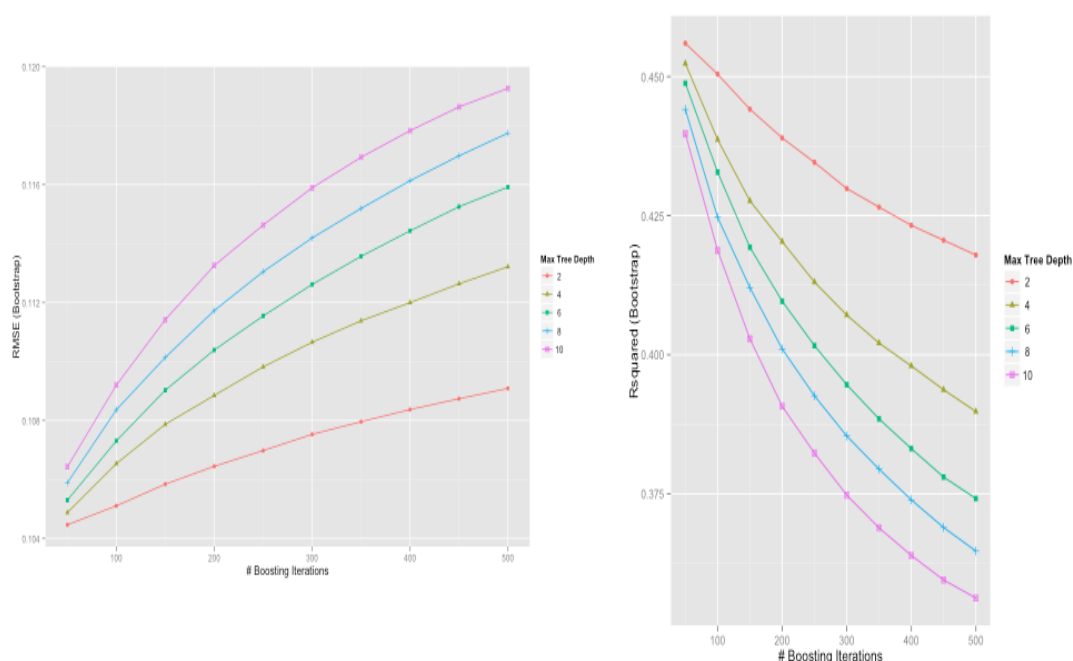


FIGURE 9.16 – Boosting Arbres de régression - RMSE et R^2 d'un Boosting d'arbres de régression

Le RMSE est compris entre 0,1045 et 0,1995 et est plus sensible à la profondeur de la forêt qu'au nombre d'arbres la composant.^{iv}

Dans le cas présent, le RMSE croît avec la profondeur de l'arbre et avec le nombre d'arbres le composant, le R^2 réagit inversement.

9.3.3 Agrégation de réseaux de neurones

L'agrégation des réseaux de neurones, que ce soit dans le cas du bagging ou du boosting, requiert le renseignement de trois paramètres : le nombre d'itérations de l'apprentissage, la taille du réseau, autrement dit son nombre de couches intermédiaires et de neurones cachés, et un facteur de contrôle, le *decay*, qui pénalise les valeurs prises par les poids synaptiques

iv. D'autres visuels sont proposés en annexe VI

en les contraignant à décroître exponentiellement vers 0. Le contrôle de ce paramètre agit sur le temps de convergence du modèle vers une solution globale.

La réduction du sur-apprentissage est directement liée à la diminution du nombre de neurones cachés et à l'augmentation du *decay*.

Nous fixons les valeurs suivantes :

Nombre d'itérations	Decay	Nombre de neurones cachés
200	0.9	6

A l'instar des modèles simples, l'agrégation de réseaux de neurones n'offre pas de résultats graphiques. En ce sens, nous nous basons sur les taux d'erreur obtenus par l'algorithme au cours de son exécution :

```
> summary(resamps)

Call:
summary.resamples(object = resamps)

Models: bag_nn, boost_nn
Number of resamples: 200

RMSE
      Min. 1st Qu. Median  Mean 3rd Qu.  Max. NA's
bag_nn  0.1038 0.1117 0.1134 0.1134 0.1155 0.1230  0
boost_nn 0.1046 0.1114 0.1135 0.1134 0.1155 0.1207  0

Rsquared
      Min. 1st Qu. Median  Mean 3rd Qu.  Max. NA's
bag_nn  0.3584 0.4449 0.4618 0.4647 0.4866 0.5469  0
boost_nn 0.3825 0.4390 0.4620 0.4647 0.4896 0.5455  0
```

FIGURE 9.17 – Agrégation de réseaux de neurones - Comparaison Bagging/Boosting

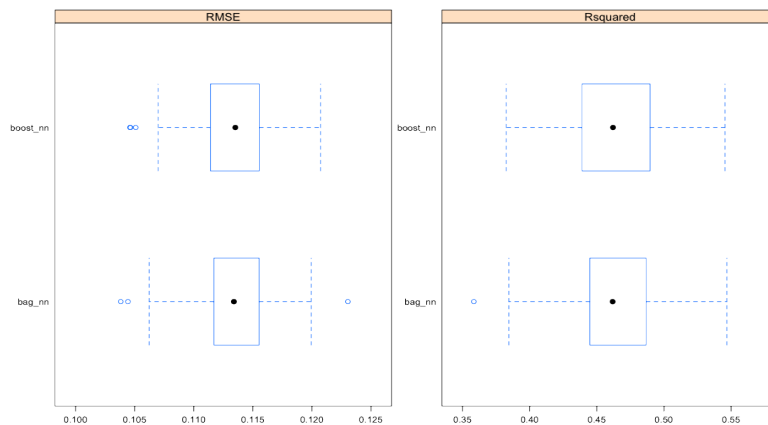


FIGURE 9.18 – Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Boxplot RMSE et R^2

Les niveaux du RMSE et du R^2 du bagging et du boosting obtenus sont très proches et présentent des valeurs également proches de celles du boosting d'arbres de régression.

9.4 Conclusion

Comparons les 4 agrégations de modèles grâce à leurs erreurs de prévisions respectives et rappelons les valeurs analogues dans les modèles simples :

Arbres de régression		
Modèle Simple	Bagging	Boosting
10,4%	3,18%	3,6%

Réseaux de neurones		
Modèle Simple	Bagging	Boosting
4,43%	1,97%	3,88%

De ces 4 modèles, l'erreur la moins importante est celle du bagging de réseaux de neurones.

Nous sélectionnons ce modèle et observons l'importance des variables 9.19 :

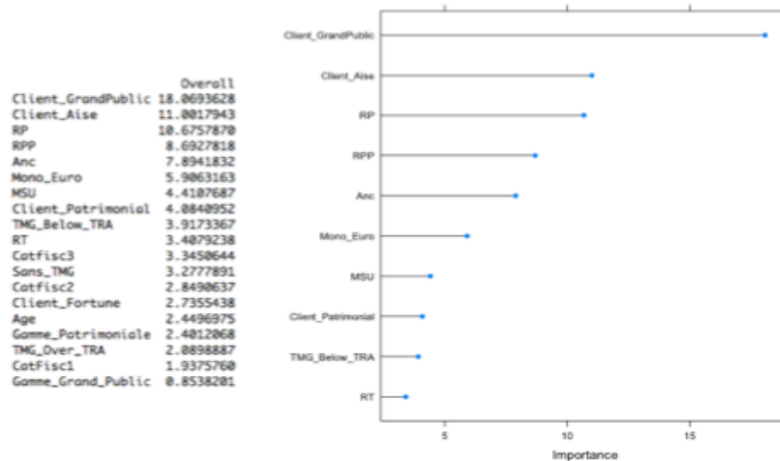


FIGURE 9.19 – Bagging de Réseaux de Neurones - Importance des variables

Les résultats obtenus nous rappellent ceux constatés au niveau des arbres de régression : les segments de clientèle sont clairement mis en avant, notamment les clients dits "Grand Public" et "Aisé". Le type de rachat intervient ensuite avant l'ancienneté et le type de support. Bien que les variables ne soient pas appelées dans le même ordre que dans les arbres de régression, ce sont les mêmes qui dominent et qui expliquent la construction du taux de rachat.

Cinquième partie

Validation des modèles : Application à une base récente

Ce chapitre clôt ce mémoire. Rappelons la procédure qui a été suivie tout son long et ce que ce chapitre propose de démontrer :

- **Bases de données** : Nous nous appuyons, dans cette étude, sur deux bases de données de même périmètre mais couvrant des périodes différentes. Ces bases correspondent à des contrats d'Épargne individuelle sur lesquels ont été effectués des rachats. Le type de rachat concerné ici est celui, structurel, non lié à la conjoncture économique mais aux caractéristiques des contrats et des assurés.

La 1^{ère} base couvre un historique de 6 années : de 2005 à 2010 (ce sont des années où le rachat s'est effectué et non des années de souscription). Pour des raisons logistiques, les données des années 2011 et 2012 n'ont pas pu être exploitées.

La 2^{ème} base concerne les contrats rachetés en 2013 : c'est celle qui sera utilisée dans ce chapitre.

- **Méthodes statistiques** Pour chacune de ces bases, nous faisons appel à une Analyse Factorielle des Données Mixtes (AFDM) pour sélectionner de premières variables discriminantes au Rachat.

Le taux de rachat est ensuite retenu comme variable indicatrice de l'intensité du rachat.

Les variables retenues par l'AFDM sont appelées pour modéliser ce taux. Nous employons pour cela deux types d'approximateurs non-linéaires : les arbres de régression et les réseaux de neurones. Ces modèles sont intéressants à maints égards (voir chapitres précédents) mais souffrent de leur manque de robustesse. Aussi, nous agrégeons ces modèles pour en fiabiliser les sorties.

- **Validation des méthodes** : L'étape de validation consiste enfin à appliquer les modèles retenus par la première base aux données de la seconde, plus récentes. Ce procédé est assimilable à la validation de modèle appliquée dans la théorie de l'échantillonnage.

Nous vérifions ainsi le pouvoir prédictif des modèles retenus, tant en terme de précision, de robustesse que de pérennité.

Les données 2013 correspondent au portefeuille en vigueur diminué des sorties (rachats totaux, arrivées à échéance,..) et alimenté par de nouveaux contrats. Il s'agit donc majoritairement des contrats vieillissant et étudiés préalablement entre 2005 et 2010.

En ce sens, l'intérêt de cette partie consiste à vérifier d'une part que les résultats obtenus en 2013 sont cohérents avec ceux obtenus précédemment et, d'autre part, à étudier l'évolution dans le temps du comportement lié au rachat.

Analyse Factorielle des Données Mixtes

Cercle des corrélations

L'application de l'AFDM aux données 2013 permet d'avancer les observations et analyses suivantes :

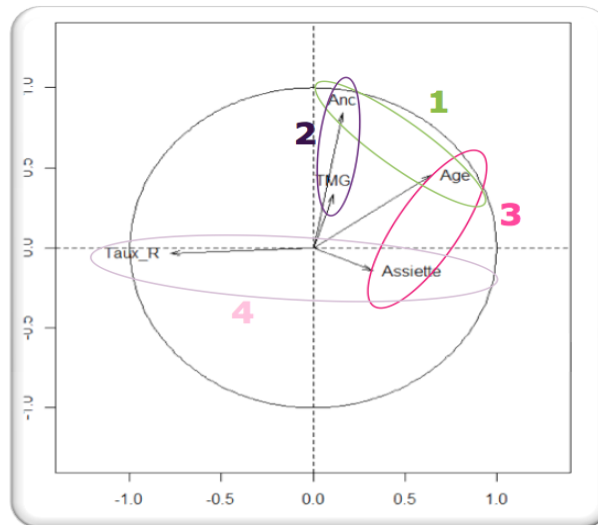


FIGURE 9.20 – Répartition des variables quantitatives après rotation - Portefeuille 2013

- L'âge des assurés, l'ancienneté de leurs contrats, le taux de rachat et l'assiette des rachats ont quasiment les mêmes coordonnées que dans le cas des données 2010 ;
- La différence la plus remarquable concerne le TMG (2^{ème} ovale sur la figure 9.20). Entre 2010 et 2013, la variable liée au TMG a tendance à s'éloigner de celle du taux de rachat. En 2010, l'angle droit quasi parfait entre ces deux variables indiquait qu'il n'existait pas de relation linéaire entre elles. Leur éloignement indiquerait a priori une tendance d'évolution inverse.

Pour comprendre ce phénomène, il faut l'aborder historiquement : depuis 2010, Natixis Assurances ne commercialise plus de produits Monosupport € proposant une option de Taux Minimum Garanti (le dernier contrat de ce type a été souscrit en 2010 et proposait un TMG de 2,75%. Alors que la base continue à être alimentée par divers autres contrats, ceux offrant un TMG voient leur nombre baisser (arrivée à échéance, rachat total, décès de l'assuré,...)).

L'ancienneté des contrats est donc fortement corrélée à la proposition ou non d'un TMG : les contrats proposant cette option font partie des plus anciens contenus dans

le portefeuille en 2013. Une relation linéaire simple apparaît avec le temps entre le taux de rachat et le TMG : lorsque le TMG augmente, le taux de rachat diminue.

- Les autres liens entre les variables sont identiques à ceux observés en 2010 :
 - ✘ Les assiettes de rachat progressent positivement avec l'âge des assurés (en moyenne, les assurés les plus âgés sont les plus riches (3^{ème} ovale sur la figure 9.20)), ce sont également ces mêmes assurés qui rachètent leurs contrats le plus tardivement (1^{er} ovale sur la figure 9.20);
 - ✘ Un fort taux de rachat est lié à une assiette faible, ce qui est rassurant et prévisible compte tenu de la définition du taux de rachat (4^{ème} ovale sur la figure 9.20);
 - ✘ Il y'a persistance (dans le sens où nous l'observions déjà sur les données 2005 à 2010) d'une relation non linéaire entre le taux de rachat et l'ancienneté des contrats. Le lien entre ces deux variables est certainement complexe et ne peut être souligné par une analyse factorielle qui ne peut mettre en avant que des relations linéaires. Les arbres de régression et les réseaux de neurones nous permettront de vérifier l'existence d'un autre type de lien entre ces variables.

Répartition des modalités

A l'instar du cercle des corrélations, la répartition des modalités est globalement inchangée entre la première base de données et celle de 2013 :

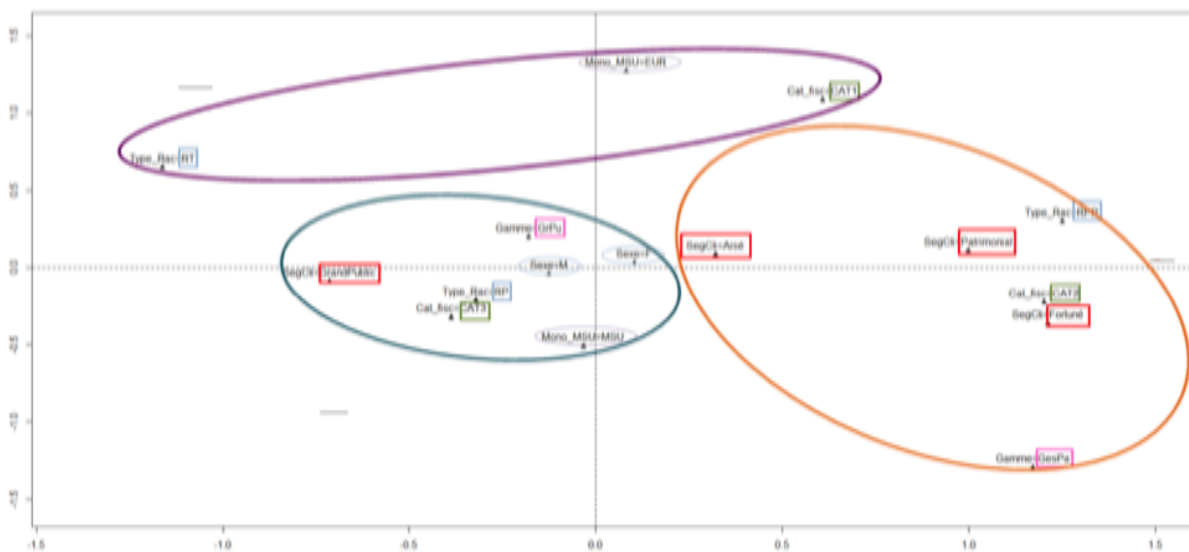


FIGURE 9.21 – Répartition des modalités après rotation - Portefeuille 2013

- Le sexe apparaît toujours comme variable neutre du point de vue de l'inertie globale (voir intersection des deux axes sur la figure 9.21);
- Les clients Grand Public (modalité de la variable Segment de clientèle) ont un comportement opposé à celui de clients plus aisés (modalités : Aisé, Fortuné, Patrimonial de la variable Segment de clientèle) (voir encadrés rouge de la figure 9.21);

- La variable liée au rachat total s'éloigne du type de support Mono€. Il serait prématuré d'en déduire une tendance à ce stade. Nous laissons aux modèles linéaires le soin de présenter ou non d'éventuels liens entre ces variables ;
- Le rachat des contrats Multisupports est toujours effectué de façon importante par une population âgée ayant souscrit avant ses 70 ans, correspondant à la 3^{ème} catégorie fiscale (encadré vert et ovale violet en haut de la figure 9.21) ;
- Le changement principal concerne les clients fortunés et patrimoniaux. Les clients fortunés semblent être les plus âgés et être ceux qui souscrivent le plus tardivement (2^{ème} catégorie fiscale). Les clients fortunés et patrimoniaux ont une forte tendance à opter pour les rachats partiels programmés (ovale de droite sur la figure 9.21).

Interactions entre les variables

La proximité des variables par rapport à l'un ou à l'autre des deux axes est globalement identique à celle que nous observons sur les données 2005 à 2010 : l'axe des abscisses est toujours assimilé à un axe de richesse, celui des ordonnées, à un axe du temps.

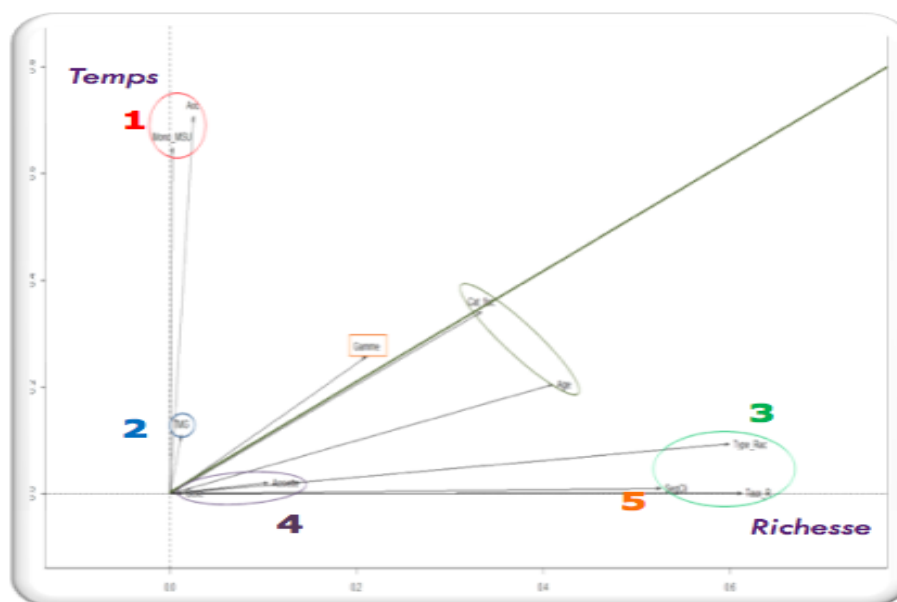


FIGURE 9.22 – Analyse croisée des variables - Portefeuille 2013

Par ailleurs, nous constatons que :

- De même qu'au niveau des cercles des corrélations, l'analyse croisée des variables indique que le TMG a de moins en moins d'impact dans l'inertie globale du portefeuille. Nous maintenons tout de même cette variable pour le reste de l'étude afin d'étudier l'évolution du comportement des assurés rachetant leur contrat au regard de la présence ou non d'un TMG sur ledit contrat ;
- Les variables Gamme, Age et Catégorie fiscale gravitent toujours autour de la 1^{ère} bissectrice. Ces variables sont aussi sensibles au passage du temps qu'au niveau de richesse ;

- La variable liée au segment de clientèle est plus proche du taux de rachat ;
- La variable liée au type de rachat est toujours proche de celle du taux de rachat.

Ces deux derniers points laissent à penser que le type de rachat et le segment de clientèle interviennent dans l'explication du taux de rachat. Les arbres de régression et les réseaux de neurones confirmeront cette remarque.

Après avoir écarté le sexe et l'assiette du rachat comme variables non discriminantes, nous appliquons aux données 2013, les modèles retenus par les données 2005 à 2010.

Arbres de régression

Dans le cas des données 2005 à 2010, nous avons retenu un modèle d'arbre dont le cp, représentant un critère de pénalisation de l'arbre, valait 0,0007.

Ce cp, dans le cas des données 2013, permet d'obtenir un arbre à 8 segmentations :

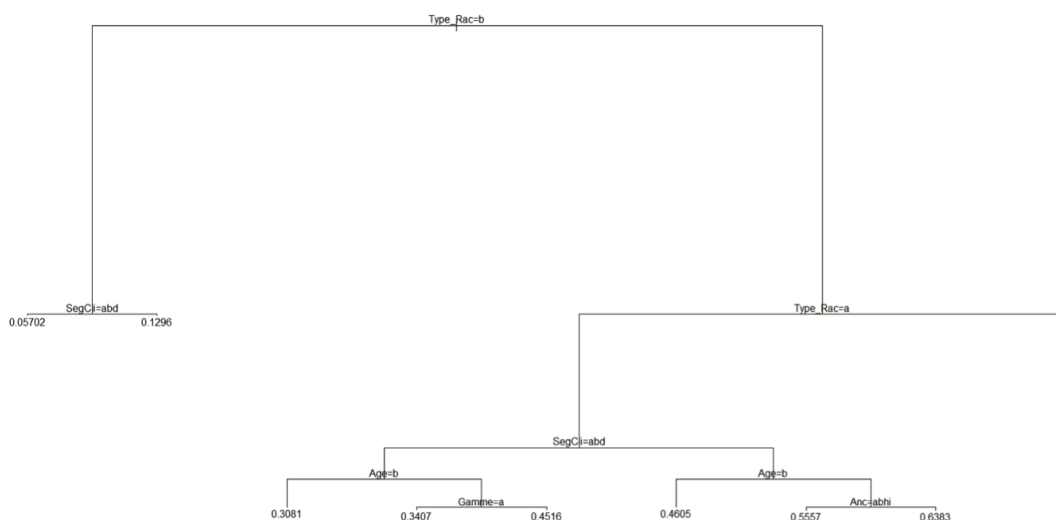


FIGURE 9.23 – Arbre de régression sur le Taux de rachat - 2013

De même que pour les résultats établis entre 2005 et 2010, nous présentons dans le tableau 9.24 les taux de rachats des groupes de risques identifiés par l'arbre, recalculés en rapportant les montants de rachats de ces groupes à la provision mathématique moyenne du portefeuille puis à la provision mathématique moyenne de chaque groupe^v.

L'arbre retenu permet d'énoncer les constats suivants :

v. Pour des raisons de confidentialité, un coefficient multiplicateur a été appliqué aux deux premières colonnes de ce tableau

en lisant l'arbre de gauche à droite

Population d'assurés ayant racheté leur police en 2013	Taux cohorte ayant racheté	Taux rachat cohorte sur PM Globale	Taux affiché par l'arbre
RPP, Aisé/Fortuné/Patrimonial	4,95%	0,33%	5,70%
RPP, Clientèle GrandPublic	11,62%	0,02%	12,96%
RP, Fortuné/Aisé/Patrimonial, >70 ans	27,14%	0,70%	30,81%
RP, Fortuné/Aisé/Patrimonial, <70 ans, Gestion de Patrimoine	35,78%	0,94%	34,07%
RP, Fortuné/Aisé/Patrimonial, <70 ans, Grand Public	42,25%	1,45%	45,16%
RP, Clientèle GrandPublic, >70 ans	49,15%	0,16%	46,05%
RP, Clientèle GrandPublic, <70 ans, ancienneté de 0 à 1 an puis supérieure ou égale à 7 ans	63,05%	0,41%	55,57%
RP, Clientèle GrandPublic, <70 ans, ancienneté de 2 à 6 ans	67,01%	0,25%	63,83%
RT	120,00%	0,48%	100,00%

FIGURE 9.24 – Taux de rachat par populations d'assurés

- Le type de rachat et le segment de clientèle sont les facteurs les plus contributifs dans l'explication du taux de rachat.

Le segment de clientèle qui n'apparaissait pas dans l'explication du taux de rachat partiel programmé dans le cas des données 2005-2010, remonte immédiatement en haut de l'arbre. Il prend cette même position pour les cas du rachat partiel et du rachat total ;

- Les assurés rachetant partiellement leurs contrats ont deux comportements distincts : selon leur segment de clientèle d'abord, puis selon leur âge.

- ✘ Parmi les assurés les plus aisés (segment de clientèle "Aisé", "Fortuné" et "Patrimonial"), le rachat semble être abordé différemment :

- Les assurés les plus jeunes rachètent leur contrat selon la gamme de produit de leur contrat ;
 - Les assurés plus âgés ont un comportement différent selon qu'ils soient Aisés, Fortunés ou Patrimoniaux : les assurés Fortunés et Patrimoniaux (Segments de Clientèle) rachètent légèrement moins que les assurés aisés ;

- ✘ Les assurés classés Grand Public (modalité du segment de clientèle) rachètent selon l'ancienneté de leurs contrats :

- Les plus jeunes rachètent souvent soit après 8 ans d'ancienneté soit, avant 1 an d'ancienneté ("a" et "b" réfèrent aux anciennetés 0 et 1 ; "i" et "j" à, respectivement, 8 ans d'ancienneté et à plus de 8 ans). Lorsque le rachat s'effectue entre 2 et 8 ans d'ancienneté, c'est l'option du TMG qui conditionne le rachat ;
 - Les plus âgés (>70 ans) rachètent leur contrat lorsque celui-ci a une ancienneté entre 0 et 1 an, de 4 ans, ou de 8 ans et plus.

Enfin, nous évaluons l'erreur de prévision sur cet arbre (Cf figure 9.25). Cette erreur est moins importante que celle évaluée sur les données 2005 à 2010 (10,4%).

```
> ar.predict<-predict(tree.reg, newdata = data_Val)
> s=sum((ar.predict-data_Val$Taux_R)^2)/nrow(data_Val)
> s
[1] 0.05260475
```

FIGURE 9.25 – Arbre de régression - Erreur de prédiction sur le portefeuille 2013

L'arbre de régression n'en demeure pas moins instable, en particulier pour la phase de prévision. Nous utiliserons dans la section V des méthodes d'agrégation pour en augmenter la robustesse.

Réseau de neurones

Nous appliquons aux données 2013 le modèle retenu par les données 2005-2010, à savoir celui d'un Perceptron Multi-Couches (PCM) à taux d'apprentissage de 0,25, à 2 couches intermédiaires de 3 neurones chacune et à fonction d'activation sigmoïde (tangente hyperbolique). La fonction de combinaison de ce réseau est une combinaison linéaire des poids et des entrées des neurones.

Rappelons que les réseaux de neurones n'offrent pas de rendu visuel intéressant à analyser, nous nous reposons sur l'erreur de prédiction obtenue par le modèle sélectionné. Celle-ci vaut 3,67 % dans le cas de ce modèle. Elle sera comparée à celles obtenues par les autres méthodes en conclusion de ce chapitre.

Agrégation de modèles

Forêt aléatoire

L'évolution du taux d'erreur par rapport au nombre d'arbres de la forêt est un peu plus erratique que celle observée sur les données 2005 à 2010. Le taux d'erreur ne parvient à se stabiliser qu'à partir de 450 arbres (contre 300 pour les données 2005 à 2010). Nous augmentons ce nombre d'une marge de prudence d'une cinquantaine d'arbres pour construire notre modèle.

L'intérêt est ensuite porté à l'importance des variables explicatives (Figure 9.27). Nous présentons les valeurs équivalentes obtenues pour la période 2005 à 2010 à des fins de comparaison.

La décroissance de la pureté des nœuds (*IncNodePurity*), vue en 2013, est sensible aux mêmes variables qu'entre 2005 et 2010. En revanche, la précision du modèle (*IncMSE*) est légèrement différente sur les données 2013 :

- Le type de rachat et le segment de clientèle correspondent toujours aux variables auxquels le modèle est le plus sensible ;
- Les variables Age et Gamme sont permutées et de même pour les variables de catégorie fiscale et d'ancienneté.

Nous rappelons que ce graphique représente la décroissance de la précision. La permutation observée indique donc que la précision du modèle est plus sensible à la variable Gamme (resp. à la variable Ancienneté) qu'à la variable Âge (resp. à la variable Catégorie Fiscale).

Le modèle nous indique ainsi que :

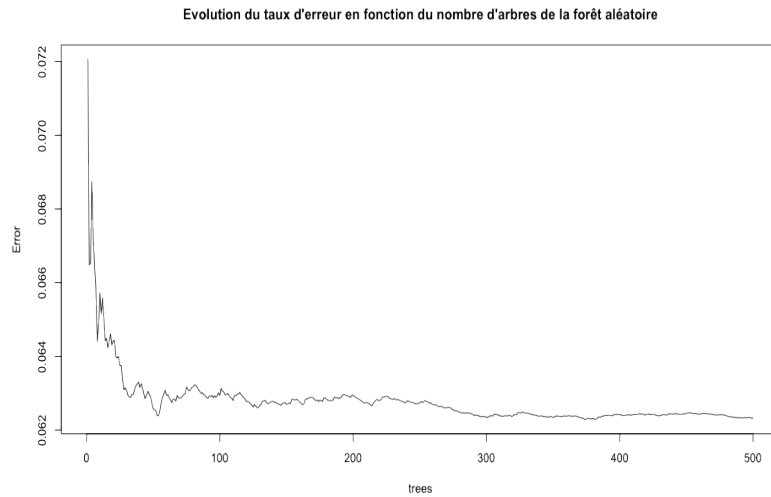


FIGURE 9.26 – Forêt Aléatoire - Taux d'erreur et nombre d'arbres - Portefeuille 2013

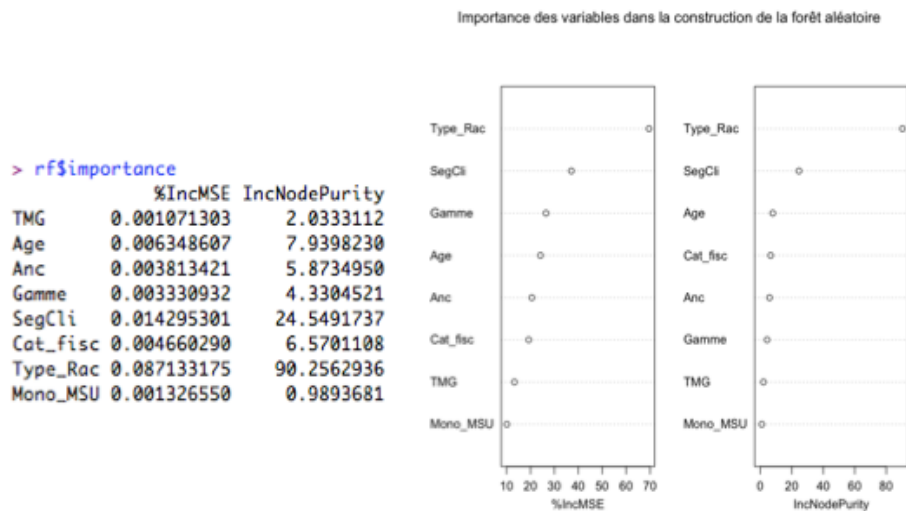


FIGURE 9.27 – Forêt Aléatoire - Importance des variables - Portefeuille 2013

- Le taux de rachat est particulièrement attaché au type de rachat et au segment de clientèle puis, bien que relativement moins, à la gamme de produit ;
- Le taux de rachat, en 2013, devient sensible à l'âge et à l'ancienneté indépendamment de la catégorie fiscale qui les lie.

Nous évaluons enfin l'erreur de prévision sur ce modèle. Elle est égale à 5,65% (contre 5,1% dans le cas des données 2005 - 2010) (figure 9.28). Le gain est considérable comparé à celui des arbres de régression (11,7% pour un seul arbre).

```

> sum((rf.predict-data_Val$Taux_R)^2)/nrow(data_Val)
[1] 0.05659251691

```

FIGURE 9.28 – Forêt Aléatoire - Erreur de prédiction sur échantillon test

Boosting d'arbres de régression

Les paramètres du boosting d'arbres de régression sont identiques à ceux utilisés dans le cas des données 2005 à 2010 : le nombre initial d'arbres est fixé à 50 et évolue par pas de 50 jusqu'à atteindre 500 arbres, la profondeur de la forêt varie entre 5 niveaux différents et le taux d'apprentissage est égal à 0,001.

Les résultats obtenus sont les suivants :

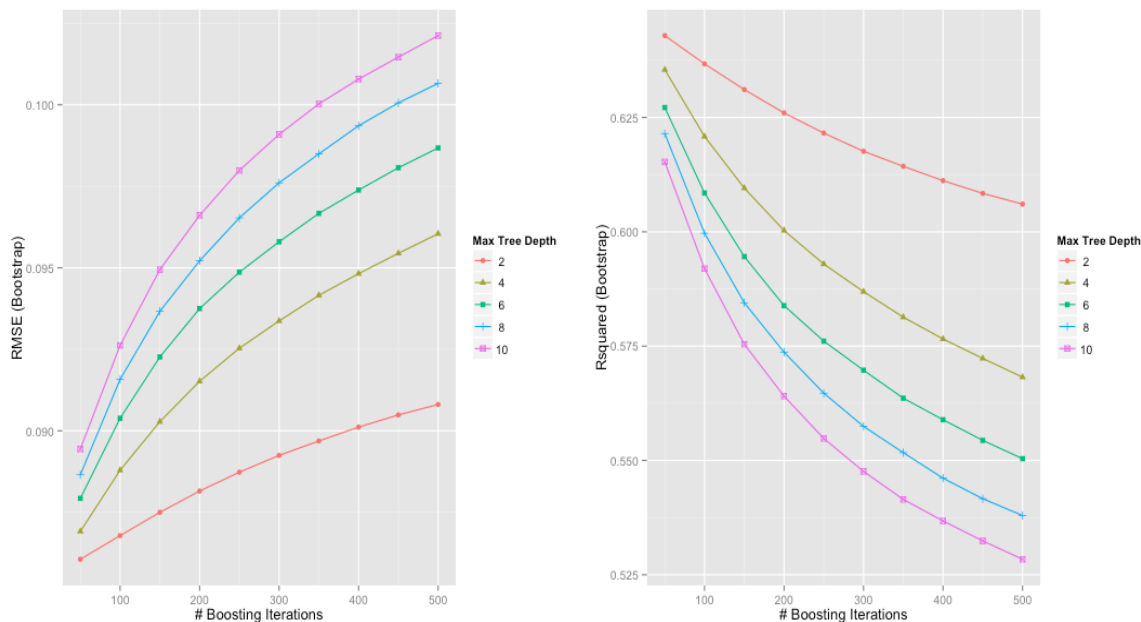


FIGURE 9.29 – Boosting Arbres de régression - RMSE et R^2 d'un Boosting d'arbres de régression - Portefeuille 2013

Dans le cas présent, le RMSE est compris entre 0,086 et 0,104 et est, comme dans le cas des précédentes données, plus sensible à la complexité de la forêt qu'à son nombre d'arbres.

Nous nous prononçons sur son efficacité lors de la comparaison des modèles.

Agrégation de réseaux de neurones

Les paramètres sont encore une fois ceux des données 2005 à 2010, à savoir : 200 itérations bootstrap, 0,9 de Decay (le paramètre de pénalisation des poids synaptiques) et 6 neurones cachés.

Les résultats obtenus sont les suivants :

L'intervalle de valeurs prises par le RMSE est moins important que dans le cas des données 2005 à 2010. Cela peut être dû au fait que le modèle assimile mieux les comportements des nouvelles données mais peut également être lié à un effet volume : les données 2013 étant naturellement moins denses que les données 2005 à 2010.

Nous nous prononçons quant à l'efficacité de ces méthodes dans la dernière partie de ce chapitre.

```

> summary(resamps)

Call:
summary.resamples(object = resamps)

Models: bag_nn, boost_nn
Number of resamples: 200

RMSE
      Min. 1st Qu.  Median    Mean 3rd Qu.  Max. NA's
bag_nn  0.08557 0.09261 0.09516 0.09497 0.09724 0.1035  0
boost_nn 0.08659 0.09227 0.09451 0.09447 0.09665 0.1027  0

Rsquared
      Min. 1st Qu.  Median    Mean 3rd Qu.  Max. NA's
bag_nn  0.5517  0.5940 0.6130 0.6128  0.6301 0.6932  0
boost_nn 0.5373  0.5934 0.6091 0.6103  0.6303 0.6833  0

```

FIGURE 9.30 – Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Portefeuille 2013

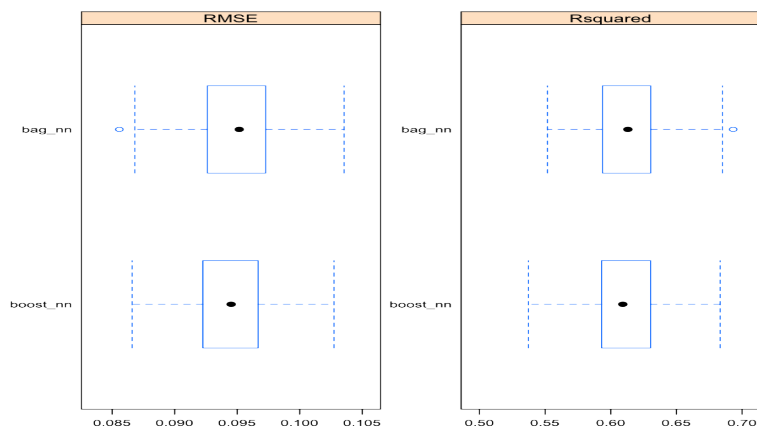


FIGURE 9.31 – Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Portefeuille 2013 - Boxplot RMSE et R^2

Conclusion

Nous comparons les 4 agrégations de modèles grâce à leurs erreurs de prévisions respectives en rappelant les valeurs analogues dans les modèles simples :

Arbres de régression		
<u>Modèle Simple</u>	<u>Bagging</u>	<u>Boosting</u>
5,26%	5,65%	2,32%

Réseaux de neurones		
<u>Modèle Simple</u>	<u>Bagging</u>	<u>Boosting</u>
3,67%	3,02%	4,33%

De ces 4 modèles, l'erreur la moins importante est celle du boosting d'arbres de régressions. Opitz et Maclin [25] l'avaient remarqué : ce type d'agrégation de modèles peut très bien s'adapter à certaines données et permettre d'obtenir des résultats intéressants.

Nous sélectionnons donc ce modèle et observons l'importance des variables (figure 9.32) :

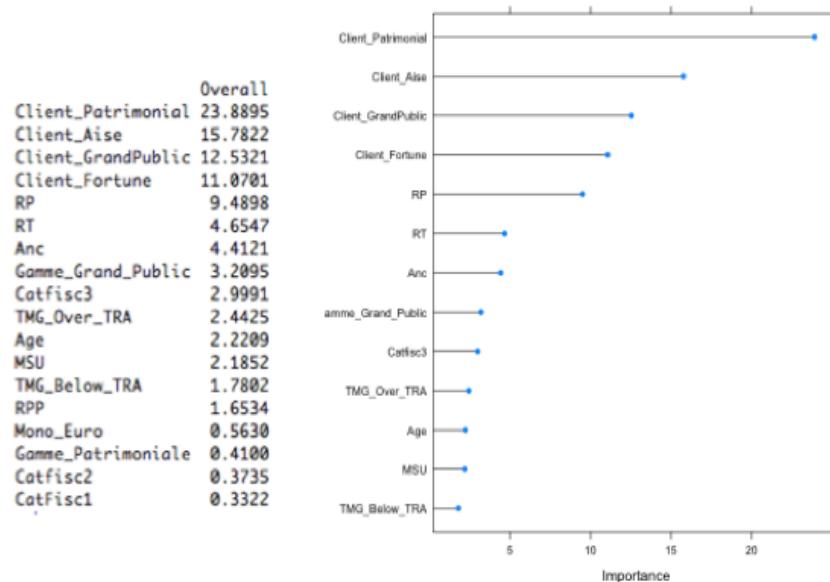


FIGURE 9.32 – Boosting d’arbres de régressions - Importance des variables

Nous retrouvons un constat fait au niveau du bagging de réseaux de neurones appliqué aux données 2005 à 2010 : la variable Segment de clientèle prend de l’importance avec le temps, c’est d’autant plus imposant en 2013. Ses quatre modalités sont celles qui remontent le plus dans le modèle retenu.

Le type de rachat intervient ensuite. Le rachat total prend le pas sur le rachat partiel programmé, qui apparaissait en 4^{ème} position dans le cas des données 2005 à 2010. Rappelons-nous que, dans la répartition des modalités (Section V), le rachat total avait pris une position différente dans l’espace des modalités, en comparaison avec celle qu’elle prenait au sein des données 2005 - 2010.

Le boosting d’arbres de régressions détecte donc ce changement de tendance.

L’ancienneté des contrats et les autres variables interviennent enfin avec une importance assez proche.

Globalement, les résultats obtenus sur les données 2013 rejoignent ceux établis entre 2005 et 2010. Les principales différences sont principalement dues à des décisions commerciales ou au passage du temps.

Nous avons ainsi constaté que :

- Le segment de clientèle prend, avec le temps, de plus en plus d’importance dans l’explication du taux de rachat ;
- L’ancienneté du contrat est toujours attachée au TMG : le TMG n’est présent que sur les contrats les plus anciens. Nous l’observons dans l’analyse factorielle : le vecteur portant le TMG est plus court que sur des données anciennes, signe des sorties des contrats portant un TMG. Avec le temps, cette variable est de moins en moins pertinente. Par ailleurs, il est important de remarquer l’absence de la variable TMG comme critère incitant au rachat pour les assurés les plus aisés ;

- L'âge conditionne le type du rachat, cela s'entend dans la mesure où, à chaque tranche d'âges correspondent des besoins différents ;
- Le segment de clientèle et le type de rachat sont les critères les plus pondérants dans la détermination du taux de rachat, l'âge les talonnant de près ;
- Pour les assurés les plus aisés, c'est d'abord la gamme de produits qui influence la décision du rachat. Les assurés les moins nantis sont davantage conditionnés par l'ancienneté de leurs contrats ;

Sixième partie

Conclusion

Dans ce mémoire, nous avons souhaité mettre en lumière l'un des risques les plus pondérants dans le portefeuille d'un assureur, celui du **Rachat**. Il était incontournable de rappeler les enjeux qui gravitent autour de ce risque. Les besoins des agents économiques évoluent, consciemment ou non. L'assureur doit alors anticiper les comportements de ses assurés, comprendre leurs motivations, détecter des périodes critiques,... Ces informations peuvent être extraites à partir des médias, de recherches économiques, mais peuvent surtout se trouver dans les données dont dispose l'assureur et qu'il lui faut simplement exploiter.

Pour étudier le risque de rachat, l'approche linéaire n'était pas suffisante car pouvait passer outre des liens complexes entre les variables. Nous avons donc fait appel à des méthodes de *machine learning* réputées pour leur puissance statistique. Les conclusions de ce travail ont mené à des réflexions sur la construction des Model Points définis au niveau de l'*ALM*, celle-ci pouvant alors être éprouvée via des études d'impact sur des indicateurs de rentabilité (Return-On-Equity (ROE),...).

Les méthodes appelées dans ce mémoire sont intéressantes d'un point de vue théorique, toutefois, des réserves doivent être émises sur les modèles construits et leur application dans le cadre d'un portefeuille d'Epargne. Souvenons-nous : les modèles issus d'approximateurs non-linéaires sont instables car sensibles aux minima locaux et au surapprentissage. La manipulation de certains paramètres (le moment ou le taux d'apprentissage par exemple) peuvent permettre de réduire ce risque mais il n'en demeure pas moins présent. Bien qu'existent des règles théoriques en facilitant le contrôle, ce sont généralement des tests empiriques (propres donc à chaque base de données) qui aboutissent à des paramètres satisfaisants. D'autre part, les réseaux de neurones, en particulier, aboutissent à des erreurs de prédiction très faibles et détectent des relations non-linéaires intéressantes, ce qui en fait un outil de modélisation et de prévision précieux. En revanche, et bien que la parcimonie soit l'un de leurs principaux atouts, ce sont des modèles dont la calibration initiale est délicate. Non seulement chaque paramètre agit sur les résultats du modèle mais certaines combinaisons de paramètres impactent tout autant les résultats. Le développement de tels modèles peut alors se faire dans le cadre de cellules de Recherche & Développement moyennant une grande patience. En effet, les temps d'exécution sont parfois extrêmement coûteux, en particulier pour les réseaux de neurones et les méthodes d'imputation des données mixtes. Limiter ce temps requiert le contrôle d'arguments (nombre de couches intermédiaires, présence ou non de biais, choix de la fonction d'activation,...) qui peuvent pénaliser le modèle. Ceci étant, ces sujets, en particulier celui de l'agrégation de modèles spécifiques (dans le cas de ce mémoire, les arbres de régression et les réseaux de neurones) et hormis dans le cas des forêts aléatoires, ne bénéficie pas d'une littérature abondante. Cela en fait un champ de recherches encore inexploité pouvant receler de promesses insoupçonnables.

Quant au sujet des rachats structurels, il est extrêmement difficile de l'aborder en prenant compte de tous les paramètres véritablement impliqués dans la décision du rachat. Ainsi, certains facteurs indissociables des rachats n'ont pas été retenus dans cette étude pour des raisons inhérentes à la disponibilité des données (et au manque de temps et de ressources logistiques !) : les pénalités de rachat, les frais d'arbitrages (si l'assuré se sent lésé, il songera à racheter son épargne, à demander une avance ou un prêt mais peut également transférer son investissement vers un autre support si les frais d'arbitrage l'y incitent), la clause de participation aux bénéfices, l'implication de la conjoncture économique via un taux de satisfaction vis-à-vis du taux servi par exemple, les montants de versements, le rating de l'assureur,... Par ailleurs, il faut être réaliste : bien que l'étude soit focalisée sur les rachats structurels, les données ne peuvent être complètement indépendants de la conjoncture économique : nous captions donc malgré nous un effet conjoncturel (traité par ailleurs dans l'étude des rachats conjoncturels réalisée par le service *ALM*). Par ailleurs,

une profusion de données peut également ouvrir la voie à des méthodes aptes à gérer des volumes importants. Les techniques de Big Data doivent nécessairement intervenir alors.

Il est, également, un paramètre déterminant qui est sciemment écarté de cette étude, dans la mesure où la validation des modèles le mettrait en exergue : les facteurs du rachat ne sont pas statiques. Ils subissent l'effet du temps. Nous retrouvons bien là l'idée soulevée par X. Milhaud lorsqu'il a fait appel à des modèles GLM pour évoquer la crise des corrélations.

Cette étude repose donc sur des méthodes non nécessairement exploitées par le passé dans le cadre de l'Épargne. La calibration des modèles, tout en respectant les aspects théoriques de construction, s'est toujours aliénée à l'objectif poursuivi : la constitution de groupes homogènes de risque, traduisant des réalités économiques. Nous avons veillé à traduire immédiatement nos conclusions en les connectant au choix de nos paramètres.

Enfin, le travail entamé dans ce mémoire pourrait être amélioré à bien des égards. L'analyse factorielle pourrait être suivie d'une Classification Ascendante Hiérarchique (CAH) qui permettrait de réaliser une première segmentation des variables explicatives. Nous aurions également pu rester dans le thème des réseaux de neurones et réaliser une carte de Kohonen en première partie de ce mémoire. Il aurait été intéressant de modéliser les taux de rachat grâce à d'autres méthodes d'apprentissage, telles les Machines à Vecteurs de Support (SVM), réputées pour leur efficacité. Nous aurions également pu développer un algorithme d'agrégation entre la bagging et le boosting dans la mesure où les deux nous procuraient des taux d'erreur relativement proches. Il serait également intéressant de penser à un pendant de ce mémoire en lien avec l'étude récente menée sur le sujet par O. Lopez, X. Milhaud et P.-E. Thérond [46]. Dans cet article, les auteurs combinent l'utilisation d'arbres de décision CART et l'analyse de la probabilité inverse de la pondération de la censure (IPCW). En soulignant l'aliénation des données au temps, cette étude veille à ne pas occulter l'existence de données censurées, les imputant ainsi grâce à des méthodes de survie. Les auteurs mesurent ensuite la cohérence de cette approche en l'appliquant à des données assurantielles réelles (Prévoyance et Responsabilité Civile).

Un travail abouti comprendrait sans doute à la fois une approche macro-économique, une vision sectorielle et un développement scientifique. Les sujets abordés dans ce mémoire pourront certainement être développés dans ces cadres lors de prochains travaux.

Annexes

Juridiction autour du Rachat

Textes législatifs

Le souscripteur d'un contrat d'Assurance Vie peut à tout moment mettre fin à son contrat. Il perçoit alors – si la nature du contrat souscrit le permet – une somme appelée "valeur de rachat". Le versement de cette somme libère définitivement l'assureur de toutes ses obligations.

Cette opération est autorisée est mentionnée dans le code des Assurances, dans le code de la mutualité et dans le code de la Sécurité Sociale via les articles suivants, cités dans leur forme originelle :

Article L132-21 du Code des Assurances

Le contrat précise les modalités de calcul de la valeur de rachat ou de la valeur de transfert et, le cas échéant, de la valeur de réduction. Pour le calcul de la valeur de réduction, il ne peut être prévu d'imputer sur la provision mathématique du contrat une indemnité de réduction.

Dans la limite de la valeur de rachat du contrat, l'assureur peut consentir des avances au contractant.

En cas de demande de rachat du contrat par le contractant, l'entreprise d'assurance ou de capitalisation lui verse la valeur de rachat du contrat dans un délai qui ne peut excéder deux mois.

En cas de demande de transfert du contrat par l'adhérent, l'entreprise d'assurance verse à l'organisme d'assurance d'accueil la valeur de transfert du contrat dans un délai et selon des modalités fixées par décret.

Au-delà des délais mentionnés aux deux alinéas précédents, les sommes non versées produisent de plein droit intérêt au taux légal majoré de moitié durant deux mois, puis, à l'expiration de ce délai de deux mois, au double du taux légal.

Article L932-23 du Code de la Sécurité Sociale

A l'exception des articles L. 131-2, L. 131-3, L. 132-5-1, L. 132-6, L. 132-7, L. 132-10, L. 132-15, L. 132-17, et L. 132-19 et, pour les opérations collectives à adhésion obligatoire ou facultative, des articles L. 132-2, L. 132-8 et L. 132-9, les dispositions du chapitre Ier, de la section 1 du chapitre II du titre III et du chapitre IV du titre IV du livre Ier du code des assurances sont applicables aux règlements et contrats des institutions de prévoyance lorsqu'elles réalisent des opérations d'assurance sur la vie et de capitalisation.

Pour l'application du présent article, les mots : " assureurs " et " entreprises d'assurance " figurant dans ces dispositions du code des assurances sont remplacés par les mots : " institutions de prévoyance " ; le mot : " assuré " est remplacé par le mot :

" participant " ; le mot : " primes " est remplacé par le mot :

" cotisations " ; les mots : " police " et " contrat " sont remplacés par les mots : " bulletin d'adhésion à un règlement ou contrat " ; les mots : " participations bénéficiaires " sont remplacés par les mots :

" participation aux excédents " ; les mots : " contrats d'assurance de groupe " sont remplacés par les mots : " opérations collectives à adhésion obligatoire ou facultative ". Toutefois, les dispositions de l'article L. 132-20 ne s'appliquent qu'aux bulletins d'adhésion ou contrats comportant une valeur de rachat.

Par dérogation aux dispositions de l'article L. 132-3, le représentant légal d'un majeur en tutelle peut, pour les opérations collectives à adhésion obligatoire ne comportant pas une valeur de rachat, adhérer au nom de celui-ci à un règlement ou souscrire un contrat auprès d'une institution de prévoyance.

L'institution de prévoyance communique chaque année aux entreprises adhérant à son règlement ou à ses contrats le montant et les modalités de répartition pour l'année écoulée de la participation aux excédents.

Les dispositions de l'article L. 932-21-1 ne s'appliquent pas aux opérations dépendant de la durée de la vie humaine qui comportent une valeur de rachat.

Article L223-20 du Code de la Mutualité

Le bulletin d'adhésion ou le contrat précise les modalités de calcul de la valeur de rachat ou de la valeur de transfert et, le cas échéant, de la valeur de réduction. Pour le calcul de la valeur de réduction, il ne peut être prévu d'imputer sur la provision mathématique du contrat une indemnité de réduction.

Dans la limite de la valeur de rachat, la mutuelle ou l'union peut consentir des avances au membre participant.

En cas de demande de rachat du contrat par un membre participant, la mutuelle ou l'union lui verse la valeur de rachat du contrat dans un délai qui ne peut excéder deux mois.

En cas de demande de transfert du contrat par un membre participant, la mutuelle ou l'union verse à l'organisme d'assurance gestionnaire du plan d'accueil la valeur de transfert du plan d'épargne retraite populaire dans un délai et selon des modalités fixées par décret.

Au-delà des délais mentionnés aux deux alinéas précédents, les sommes non versées produisent de plein droit intérêt au taux légal majoré de moitié durant deux mois, puis, à l'expiration de ce délai de deux mois, au double du taux légal.

Par ailleurs, l'article L. 331-2 du Code des Assurances pose le principe de calcul de la valeur de rachat : « Pour tout contrat d'assurance sur la vie comportant une valeur de rachat ou de transfert et pour tout contrat de capitalisation, la valeur de rachat ou, le cas échéant, de transfert, est égale à la provision mathématique dans la limite, pour la valeur de rachat des contrats d'assurance sur la vie, du montant assuré en cas de décès ». Notons enfin que l'article L 331-2 prévoit que « la valeur de rachat ou de transfert peut être diminuée d'une indemnité »

Contrats rachetables et contrats non rachetables

En faisant une telle distinction, la loi pointe une taxation spécifique qui existe pour certains contrats. Cette scission est particulièrement pertinente dans le calcul de l'assiette de l'Impôt de Solidarité sur la Fortune (ISF).

Contrats non rachetables

Au regard de la loi^{vi}, les contrats non rachetables sont les suivants :

- Assurance temporaire en cas de décès ;
- Rentes viagères immédiates ou en cours de service ;
- Assurance de capitaux de survie et de rente de survie ;
- Assurance en cas de vie sans contre-assurance ;
- Rentes viagères différées sans contre-assurance.

En général, les primes versées par le détenteur d'un contrat non rachetable ne contribuent pas au patrimoine taxable de l'ISF. Il y a toutefois une règle précise dans le cadre des contrats non rachetables :

- Pour ceux souscrits après le 20 novembre 1991, toute prime versée après les 70 ans de l'assurée doit être ajoutée au patrimoine du souscripteur du contrat ;
- Si le contrat reçoit des cotisations avant les 70 ans du souscripteur, celles-ci sont exonérées de l'ISF.

Contrats rachetables

Tous les contrats d'Assurance-Vie ne figurant pas dans la liste ci-dessus sont rachetables.

En phase d'épargne, tout contrat rachetable doit figurer dans le patrimoine de son détenteur (quel que soit son âge) à la valeur de rachat évaluée au 1^{er} janvier de l'année d'imposition.

Sortie des contrats

Les contrats peuvent connaître une sortie en capital ou en rente :

- Lorsqu'une sortie en capital se produit, tout le patrimoine de l'assuré est sujet à taxation.
- S'il est question d'une sortie en rentes, que le contrat soit rachetable ou non, le montant capitalisé total de la rente contribue au calcul de l'assiette de l'ISF.

Les deux types de contrats sont donc traités de la même manière, quel que soit le type de sortie.

Principes de calcul

Soit ${}_tV$ la provision mathématique d'un contrat, t années après la souscription. Cette provision représente le montant de la valeur de rachat, modulo l'application d'une indemnité couvrant le risque de réalisation d'actifs en moins-value.

En France, cette indemnité est contrôlée par l'article R 212-27 du Code de la mutualité : autorisée pendant les 10 premières années, elle ne doit pas dépasser 5% de la provision mathématique. Depuis 1993, les sociétés françaises ont la possibilité de prévoir des garanties de « fidélité » payables au terme, mais dont la provision mathématique est exclue du calcul de la valeur de rachat. Signalons enfin que lors d'un rachat, une plus-value comptable correspondant à la différence entre la provision mathématique d'un bilan et la valeur de rachat est systématiquement dégagée.

vi. Code des assurances, art. L. 132-23 ; Code de la mutualité, art. 223-22 ; Code de la Sécurité Sociale, art. L. 932-23

Abattement Fiscal

Les cotisations versées, pour celles dépassant un certain seuil, subissent une fiscalité différente selon l'âge auquel le souscripteur du contrat les a versées, sans que les intérêts capitalisés ne soient imposés :

- ✘ Pour des cotisations de plus de 30 500€, si les versements ont été effectués après le 70^{ème} anniversaire de l'assuré et que le contrat a été souscrit après le 20 novembre 1991, les droits de succession sont dus pour le montant dépassant le seuil de 30 500€. L'imposition est alors fonction du degré de parenté entre l'assuré et le bénéficiaire ;
- ✘ Si des cotisations de plus de 152 500€ sont relevées dans le cas d'un contrat souscrit avant les 70 ans de l'assuré, alors, **pour la partie dépassant le seuil de 152 500 €** :
 - Si l'assuré décède avant le 31 juillet 2011 ; seul est appliqué un taux de prélèvement de 20% ;
 - Si l'assuré décède entre le 31 juillet 2011 et le 1^{er} juillet 2014 ;
 - Un taux de prélèvement de 20% s'applique à la part de chaque bénéficiaire inférieure ou égale 700 000€ ;
 - Si la part de chaque bénéficiaire excède 700 000€, le taux de prélèvement s'élève à 31,25%.
 - Si l'assuré décède après le 1^{er} juillet 2014 ;
 - Un taux de prélèvement de 20% s'applique à la part de chaque bénéficiaire inférieure ou égale 902 838€. Cela concerne les revenus de l'année 2012 ;
 - Toujours pour les revenus de 2012, si la part de chaque bénéficiaire excède 902 838€, le taux de prélèvement s'élève à 25%.

Compléments historiques sur le Rachat

Notion transverse aux approches du Rachat

Bien que les approches présentées au chapitre 2.1 soient différentes par essence, des notions transverses les rapprochent. L'une des plus importantes est la suivante :

Rationalité et risquophilie des investisseurs

Un assuré est considéré comme rationnel s'il rachète son contrat au moment où les taux du marché sont supérieurs aux taux de rendement des supports d'investissement de son épargne. La rationalité, en ce sens, peut être liée au comportement des agents économiques face au risque : devant un panel de produits différents, un agent va essayer, par un choix raisonné des produits, de retirer une utilité maximale tout en prenant le moins de risques possible.

Pour comprendre ce lien, il faut se ramener à la définition d'une fonction d'utilité :

Soit $\Omega = \{w_1, \dots, w_n\}$ l'ensemble d'états du monde.

Une fonction $u : \Omega \rightarrow \mathbb{R}$ est une fonction d'utilité si et seulement si

$$\forall i \neq j, u(w_i) \geq u(w_j) \iff w_i \succeq w_j^{\text{vii}}$$

Une telle fonction est concave et est strictement croissante en chacun de ses arguments.

La théorie de Von Neumann et Morgenstern(1944) [51] est fondée sur la maximisation de la fonction d'utilité des agents. Seulement, cette fonction n'a pas forcément une forme simple. Dans le cas des contrats multisupports en unités de compte par exemple, il est plutôt d'usage de recourir aux fonctions CRRA (voir Fauvel et Le Pevedic(2011) [28]). Les fonctions d'utilité type CRRA (Constant Relative Risk Aversion) se mettent sous la forme d'une puissance :

$$u(x) = \frac{x^{1-\alpha}}{1-\alpha}, \alpha \neq 1$$

et sont caractérisées par un indice d'aversion relative au risque^{viii} constant.

Le lien entre rationalité et risquophilie est ici immédiat.

En admettant que l'assuré n'agit jamais de la manière la plus optimale qui soit vis-à-vis de son contrat, cette modélisation offre, du point de vue de l'assureur, le "pire" scénario de rachat possible.

Au demeurant, un assuré avisé quant aux mouvements des marchés financiers est généralement considéré comme rationnel et neutre au risque. Les décisions de rachat des

vii. Le symbole \succeq désigne la relation de préférence large.

viii. Cet indice s'écrit : $A_r(x) = x \times A_a(x) = -x \frac{u''(x)}{u'(x)}$

investisseurs financiers sont donc nécessairement convergentes. Ceci justifie une approche individuelle des contrats. Ce raisonnement a d'ailleurs été mis à profit dans l'étude des rachats dynamiques où le rachat est assimilé à une option américaine type Put : c'est l'approche financière du rachat qui a été présentée au début du chapitre 2.

Considérons l'intervalle de temps $[0, T]$ où T représente la date de maturité d'un contrat donné. Cet intervalle est divisé en n périodes : $0 = t_0 < t_1 < \dots < t_n = T$

Soit Ω l'ensemble des états du monde, autrement dit l'ensemble des trajectoires possibles pour la valeur du contrat :

$$\Omega = \{(w_1, \dots, w_n), w_i \in \{w_i^u; w_i^d\}, \forall i = 1, \dots, n\}$$

- En partant de $t = 0$: il suffit de déployer l'arbre en utilisant les paramètres u (up) et d (down) pour représenter les variations de la valeur du contrat selon les valeurs de rachat et la maturité de la police ;
- A partir de $t = T - 1$, la valeur initiale du contrat s'obtient grâce à :

$$V_t = \max(w^u V_{t+1}^u + w^d V_{t+1}^d; R_t)$$

Modélisation des taux de rachat

La construction des taux de rachats s'est popularisée grâce à deux méthodes :

- La première, défendue par Kolkiewicz et Tan (2006) [40], Tsai et al. (2009) [74], De Giovanni (2010) [33], Li et Szimayer (2010) [44] et par Le Courtois et Nakagawa (2011) [12] s'appuie sur un paramètre modélisant l'intensité des rachats ;
- La seconde, développée par l'American Academy of Actuaries (2005) [54], Morgan et Kent (2008) [50], Ledlie et al. (2008) [43] et Kling et al. (2011) [39] utilise la notion de multiplicateur de rachat dynamique.

Dans les lignes suivantes, n désigne le nombre de facteurs de rachat et $\Theta_{t,i}$ les facteurs de rachat, $\forall i \in [1, n]$.

Modèle à Décroissance Multiple

Lorsqu'une entreprise ne peut honorer les engagements qu'elle a vis-à-vis de ses créanciers, il y a risque de défaut. Un tel événement peut se produire dans le cas d'un risque de rachat mal évalué. Le facteur d'intensité constitue alors une mesure de la rationalité des investisseurs et de l'influence de facteurs exogènes sur leur décision de rachat.

Le **paramètre d'intensité** peut être décomposé en une constante (associée au rachat structurel) et en une composante stochastique (assimilée au rachat dynamique). Bernard, Le Courtois et Quittard-Pinon(2006) [6] ont en effet démontré que la non-prise en compte du caractère aléatoire des taux de rachat pouvait entraîner une sous-estimation du risque de rachat de l'entreprise. De fait, le rachat dit rationnel est lié à l'évolution des marchés, donc à un univers instable.

Ce modèle peut être formalisé en notant :

- $\mu(t, \Theta_{t,1}, \dots, \Theta_{t,n})$ le taux de hasard stochastique ;
- $\mu_{det}(t)$ la fonction de hasard liée à la partie déterministe du paramètre de rachat.

Ainsi, sur l'intervalle $[t - 1, t]$, le taux de rachat global est donné par :

$$l_{t-1,t} = 1 - \exp\left(-\int_{t-1}^t \mu_{det}(s) + \mu(s, \Theta_{s,1}, \dots, \Theta_{s,n}) ds\right)$$

Multiplicateurs de Rachat Dynamique (Dynamic Lapse Multiplier)

Cette notion a été introduite par l’American Academy of Actuaries pour étudier le rachat dans le cas des Variable Annuities, type GMDB (une extension a, par la suite, été faite pour les contrats type GMAB).

Ces multiplicateurs reflètent le risque d’anti-sélection lorsque la garantie est dans-la-monnaie : l’intervention de ces facteurs dans la construction des taux de rachat reflète le fait que ces derniers doivent baisser dès que la garantie présente des valeurs importantes [54].

Un multiplicateur de Rachat Dynamique est une fonction non décroissante du sous-jacent qui va opérer un ajustement à un instant t du taux de rachat à l’aide du ratio $\frac{\text{Garantie du contrat}}{\text{Valeur du fond d’investissement}}$

Cela signifie que plus ce ration augmente, plus le souscripteur d’un contrat type GMDB/GMAB a de chances de racheter son contrat.

Formellement, un multiplicateur de rachat dynamique en un instant t peut être défini comme suit :

$$\lambda_t = \min \left(1; \max \left(0, 5; 1 - 1, 25 \times \left(\frac{\text{Valeur garantie du contrat en } t}{\text{Valeur en } t \text{ du fond d’investissement}} - 1, 1 \right) \right) \right)$$

Cette expression correspond à celle d’un facteur dynamique de rachat Θ_i . Elle peut être déclinée pour n facteurs en notant :

- $l_{t-1,t}^{det}$ la partie déterministe du taux de rachat ;
- λ_t le multiplicateur de rachat dynamique ;

L’expression du taux de rachat global sur l’intervalle $[t - 1, t]$ s’écrit :

$$l_{t-1,t} = \max \left(1, l_{t-1,t}^{det} \cdot \prod_{i=1}^n \lambda_{t,\Theta_{t,i}} \right)$$

Le principe de cette méthode consiste donc à utiliser un taux de rachat déterministe comme base du taux de rachat global et de l’ajuster grâce à des facteurs dynamiques qui greffent une tendance stochastique au taux de rachat global. Ce sont ces facteurs qui portent le nom de multiplicateurs de rachat dynamique.

Principaux travaux sur les Rachats dynamiques

Les études sur les rachats dynamiques ont mis en avant deux hypothèses fondamentales déclenchant le rachat :

L’hypothèse du fond d’urgence : EFH (Emergency Fund Hypothesis)

C’est Linton (1932) [45] qui, le premier, a proposé cette hypothèse qui repose sur le principe de détresse financière. Les ménages considèrent que leur épargne est avant tout une réserve en cas de besoin. Le taux de chômage est le principal levier de contrôle de cette hypothèse. Lorsque le taux de chômage augmente, le rachat dynamique se déclenche : les assurés rachètent ou demandent un prêt sur leur épargne.

L’hypothèse du taux d’intérêt : IRH (Interest Rate Hypothesis)

Cette hypothèse a été introduite par Schott (1971) [67] et développée par Pesando (1974) [58]. Elle est basée sur l'arbitrage qu'effectue un investisseur en cas de hausse des taux d'intérêt. Elle établit la corrélation positive qui existe entre le taux de rachat et les taux d'intérêt des marchés financiers et celle, négative, qui les lie aux taux de rendement de leur épargne (cas de la participation excédentaire par exemple).

Dans ce genre de scénario, la hausse des taux est accompagnée d'une augmentation des rachats alors que le taux de remboursement des actifs garde un niveau élevé. Il y a donc un *mismatch* entre le niveau minimum d'actif disponible et celui du passif exigible.

Marché secondaire de l'Assurance Vie

Pinquet définit le marché secondaire de l'Assurance Vie comme concernant "... les assurés vie entière ayant soit perdu leur motivation à léguer le capital assuré (marché du life settlement), soit une espérance de vie très réduite et des besoins financiers liés à leur état de santé (marché du *viatical settlement*)" [59]. Dans un contexte où l'assuré souhaiterait racheter son contrat d'assurance vie, son assureur peut lui proposer de lui racheter sa police à un montant supérieur à la valeur de rachat stipulée dans son contrat. Les entreprises versent ainsi les primes du contrat d'assurance racheté afin de toucher le capital au décès de l'assuré : elles supportent donc un risque de longévité (voir Bauwers(2008) [5]).

Gatzert et Schmeiser(2009) [32] alertent sur le risque de sélection aversive qui menace le marché du life settlement : c'est ce dernier qui concentre le plus de "mauvais risques" dans le portefeuille des assureurs.

Résultats fondamentaux

Les premiers travaux sur les rachats dynamiques ont été menés par Dar et Dodds (1989) [16] sur des produits d'assurance-mixte souscrits à partir de 1952 et clôturés au plus tard en 1985. Les auteurs, grâce à des tests économétriques, des séries temporelles et des régressions sur les motifs de rachat des ménages mettent en avant la sensibilité de l'épargne au taux de rachat : une corrélation positive a été constatée vis-à-vis du taux d'intérêt de la police, négative par rapport aux taux de rendement d'autres actifs financiers. Bien qu'ayant établi un lien direct entre les taux de rachat et le niveau national du chômage, ils ne concluent pas quant à une éventuelle relation entre les taux de rachat et les taux d'intérêt.

Outreville (1990) [55] analyse les rachats sur des contrats d'Assurance Décès. Il souligne la pertinence de l'EFH en utilisant la méthode des moindres carrés ordinaires et l'ajustement de Cochrane-Orcutt pour évaluer l'autocorrélation des erreurs d'ordre 1. Ses travaux mettent en exergue l'effet de variables macroéconomiques et de tendances du marché de l'Assurance Vie sur les rachats anticipés.

Afin d'analyser les éventuelles interactions à long terme entre les variables, Kuo et al. (2003) [41] ont, quant à eux, utilisé l'analyse de cointégration de Johansen et le modèle VECM (Vector Error Correction Model). Ils construisent un modèle empirique pour approcher les rachats dynamiques. Leur constat est le suivant : d'un point de vue économique, le taux d'intérêt est plus significatif que le taux de chômage dans l'explication de l'évolution du taux de rachat.

Les conclusions de Dar et Dodds, Outreville et Kuo et al. ne convergent pas du fait des contextes différents dans lesquels leurs études ont été réalisées. Toutefois, leurs travaux s'accordent sur certains points :

- L'analyse multivariée n'est généralement pas appropriée dans l'étude des comportements de rachat ;

- L'IRH est l'hypothèse la plus souvent retenue .

Outre la donnée économique que véhiculent l'IRH et l'EFH, la considération de caractéristiques propres à l'entreprise accorde une dimension intéressante à l'analyse des rachats : la possibilité d'utiliser des modèles linéaires généralisés.

Kim(2005) [38], Cox et Lin(2006) [13] et Kiesenbauer(2011) [37] mènent des études dans ce sens. En délimitant le même périmètre pour les facteurs exogènes incitant au rachat, ils aboutissent aux mêmes conclusions :

- Les rachats augmentent quand la croissance économique chute et que le chômage augmente ;
- Le spread entre le taux de rendement des marchés et le taux d'intérêt de la police influence la décision de rachat des assurés.

Kim (2005), dans un premier temps, s'appuie sur des données mensuelles, récoltées entre 1997 et 2000 et démontre, grâce à un modèle Logit, que le taux de chômage et le taux de rachat de l'époque sont positivement corrélés.

Cox et Lin (2006), en faisant appel à un modèle Tobit sur des taux de rachat annuels, affirment qu'en plus du taux de chômage et des taux d'intérêt, le rachat dépend de facteurs exogènes. Ils recommandent, en présence d'un volume important de données, de considérer les contrats individuellement afin de les approcher avec une loi discrète (loi de Poisson ou loi Binomiale Négative).

Analogie avec d'autres univers

En banque et en Assurance Non Vie, la conclusion volontaire d'un contrat par son souscripteur est un événement fréquent.

Crise des subprimes

La modélisation peut aider à anticiper d'éventuelles chutes vertigineuses des marchés. Le modèle construit par Diamond et Diabvig(1983) [19] prévient ainsi les paniques bancaires par deux mécanismes : la suspension de la conversion des actifs financiers et la mise en place d'une assurance pour les dépôts bancaires. Ils parviennent également aux constats suivants :

- Plus une banque consent à puiser dans les dépôts d'agents excédentaires pour répondre aux besoins d'investisseurs solvables, plus elle est influente au niveau des marchés financiers ;
- Lorsqu'un mouvement de panique naît entre les déposants, ce mouvement se propage rapidement et il en résulte un retrait général de l'argent des marchés ;
- La panique bancaire est une gangrène qui atteint les membres les plus sains du système bancaire : les banques les plus solides peuvent souffrir de la méfiance brutale des agents vis-à-vis des marchés et faire faillite.

Reconduction tacite en assurance non-vie

Les contrats d'Assurance Non Vie ont généralement une échéance d'un an avec une clause de tacite de reconduction. A l'approche de l'échéance, il y a un risque important de clôture du contrat par l'assuré. Cette clôture de contrat peut être abordée, d'un point de vue économique, comme le rachat d'un contrat d'Assurance Vie.

La seule étude à ce jour qui ait sondé la question de la reconduction tacite en Assurance Non Vie est celle de Pohl(2009) [60] qui a analysé un portefeuille de contrats d'Assurance Moto.

Extension et graphiques autour des réseaux de neurones

Principaux réseaux de neurones

Réseau à apprentissage supervisé	Réseau multicouches	Réseau linéaire	Réseau à apprentissage hebbien - Hebb (1949), Fausett (1994)
			Perceptron - Rosenblatt (1958), Minsky and Papert (1969/1988), Fausett (1994)
			Réseau Adaline - Widrow and Hoff (1960), Fausett (1994)
			Réseau à neurones d'ordre supérieur - Bishop (1995)
			Réseau à entrées fonctionnelles - Pao (1989)
		Perceptron multicouches	Réseau à rétropropagation - Rumelhart, Hinton, and Williams (1986)
			Réseau à architecture en cascade - Fahlman and Lebiere (1990), Fausett (1994)
	Réseau à fonctions de base radiales (ou RBF)	Réseau à propagation rapide - Fahlman (1989)	
	Réseau CMAC (Cerebellar Model Articulation Controller) - Albus (1975), Brown and Harris (1994)	Réseau à algorithmes des moindres carrés - Chen, Cowan and Grant (1991)	
	Réseau de classification	Modèle LVQ (Learning Vector Quantization) - Kohonen (1988), Fausett (1994)	
	Réseau de régression	Modèle PNN (Probabilistic Neural Network) - Specht (1990), Masters (1993), Hand (1982), Fausett (1994)	
	Réseau récurrent	Réseau BAM (Bidirectional Associative Memory) - Kosko (1992), Fausett (1994)	Modèle GNN (General Regression Neural Network) - Specht (1991), Nadaraya (1964), Watson (1964)
		Modèle de reconnaissance de série temporelle	Machine de Boltzman - Ackley et al. (1985), Fausett (1994)
Modèle d'Elman - Elman (1990)			
Modèle FIR (Finite Impulse Response) - Wan (1990)			
Modèle de Jordan - Jordan (1986)			
Réseau récurrent à pas de temps réel - Williams and Zipser (1989)			
Modèle récurrent à rétropropagation - Pineda (1989), Fausett (1994)			
Modèle TDNN (Time Delay NN) - Lang, Waibel and Hinton (1990)			
Réseau compétitif	Modèle ARTMAP - Carpenter, Grossberg and Reynolds (1991)		
	Réseau à contre-propagation - Hecht-Nielsen (1987; 1988; 1990), Fausett (1994)		
	Néocognitron - Fukushima, Miyake, and Ito (1983), Fukushima, (1988), Fausett (1994)		
Réseau à apprentissage non-supervisé	Réseau compétitif	Modèle à quantification vectorielle	Modèle de Grossberg - Grossberg (1976)
		Modèle à conscience - Desieno (1988)	
		Carte auto-organisatrice	Carte de Kohonen - Kohonen (1995), Fausett (1994)
			Modèle GTM - Bishop, Svensén and Williams (1997)
		Modèle à Théorie de la Résonance Adaptative (ART)	Modèle ART 1 - Carpenter and Grossberg (1987a), Moore (1988), Fausett (1994)
			Modèle ART 2 - Carpenter and Grossberg (1987b), Fausett (1994)
	Modèle ART 2-A - Carpenter, Grossberg and Rosen (1991a)		
	Modèle ART 3 - Carpenter and Grossberg (1990)		
	Modèle ART flou - Carpenter, Grossberg and Rosen (1991b)		
	Modèle à apprentissage compétitif différentiel - Kosko (1992)		
	Réseau réducteur de dimension	Modèle de Oja - Oja (1989)	
Modèle de Sanger - Sanger (1989)			
Modèle différentiel hebbien - Kosko (1992)			
Réseau auto-associatif	Modèle auto-associatif linéaire - Anderson et al. (1977), Fausett (1994)		
	Modèle BSB (Brain State in a Box) - Anderson et al. (1977), Fausett (1994)		
	Modèle de Hopfield - Hopfield (1982), Fausett (1994)		
Réseau sans apprentissage	Modèle de Hopfield - Hertz, Krogh, and Palmer (1991)		
	Modèle de Cichocki et Unbehauen (1993)		

Tests de sensibilité sur un PCM à rétropropagation du gradient l'erreur

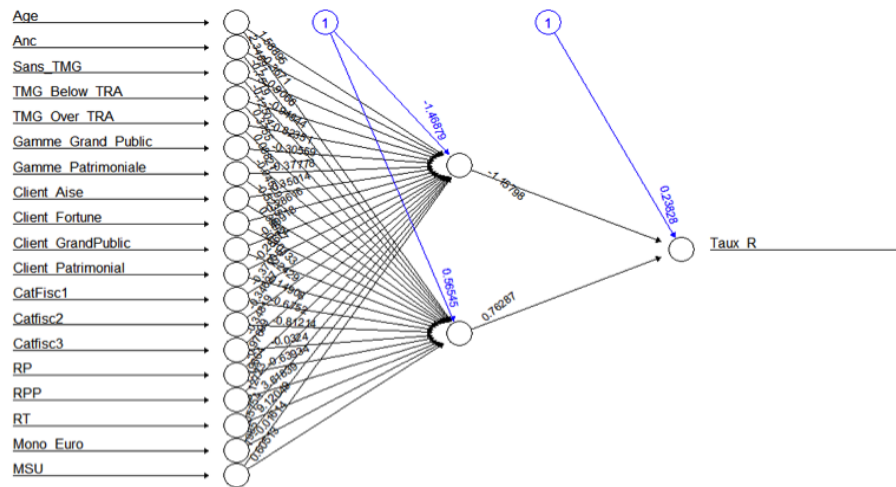


FIGURE 9.33 – Réseau de Neurons - PCM à une couche intermédiaire et à 2 neurones cachés

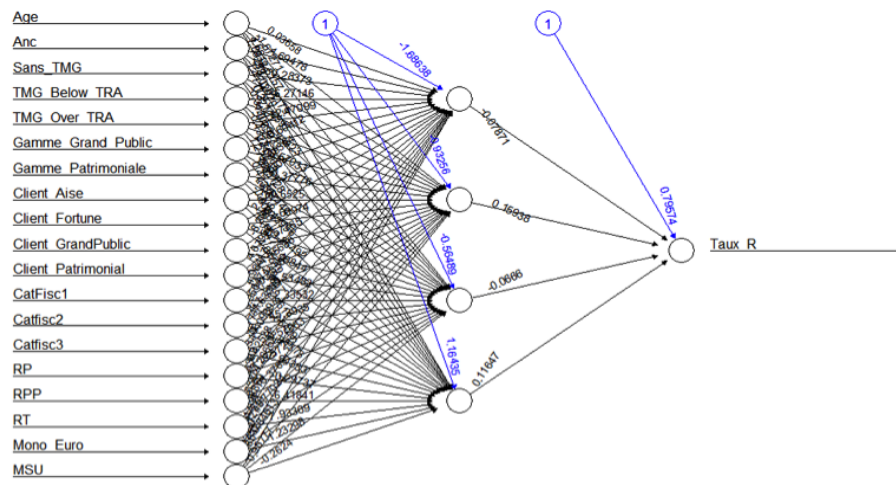


FIGURE 9.34 – Réseau de Neurons - PCM à une couche intermédiaire et à 4 neurones cachés

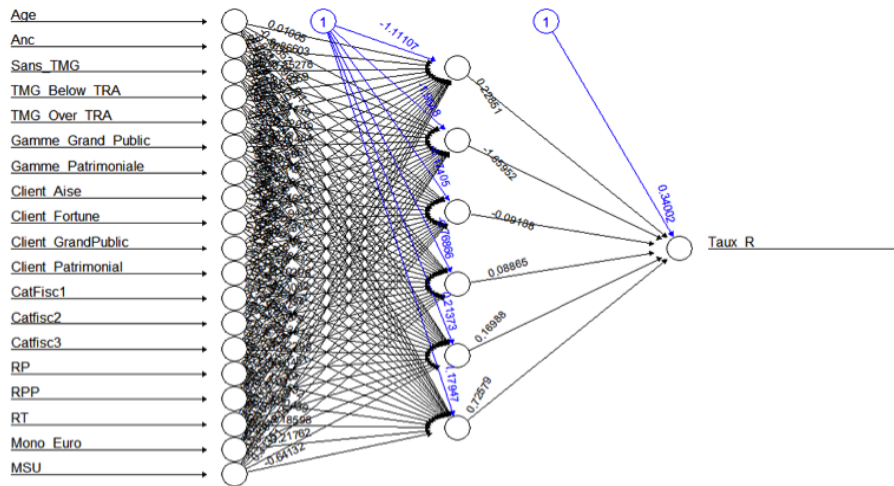


FIGURE 9.35 – Réseau de Neurons - PCM à une couche intermédiaire et à 6 neurones cachés

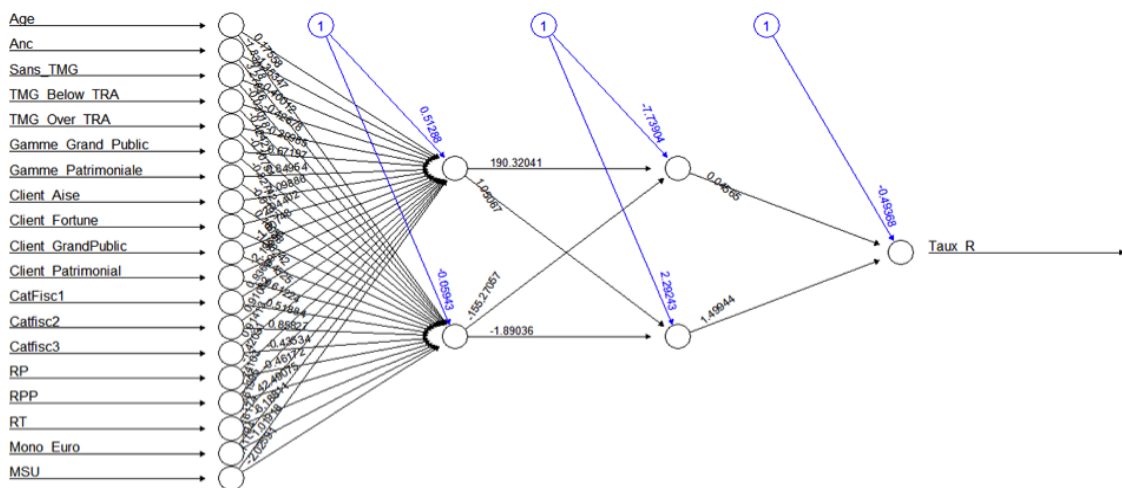


FIGURE 9.36 – Réseau de Neurons - PCM à 2 couches intermédiaires et à 6 neurones cachés

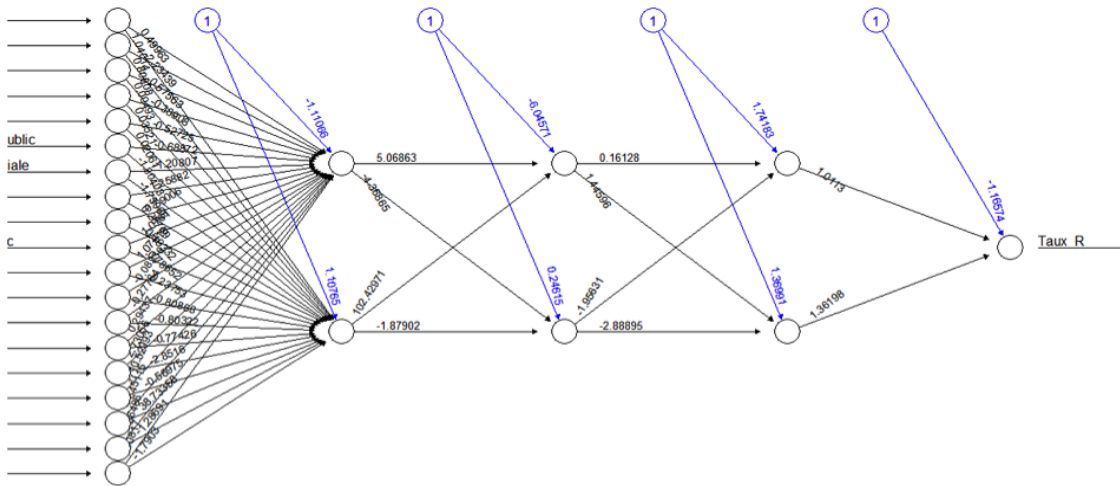


FIGURE 9.37 – Réseau de Neurons - PCM à 3 couches intermédiaires et à 6 neurones cachés

Boosting/Bagging - Graphiques

L'agrégation de modèles, dans le cadre du boosting d'arbres de régression notamment, offre des visuels graphiques différents. Nous présentons dans cette section, à titre d'illustration, ceux obtenus sur le portefeuille 2005 à 2010. Le premier représente l'évolution de l'erreur quadratique en fonction de la complexité de la forêt et du nombre d'arbres la composant, le second, la distribution de l'erreur quadratique moyenne et du R^2 .

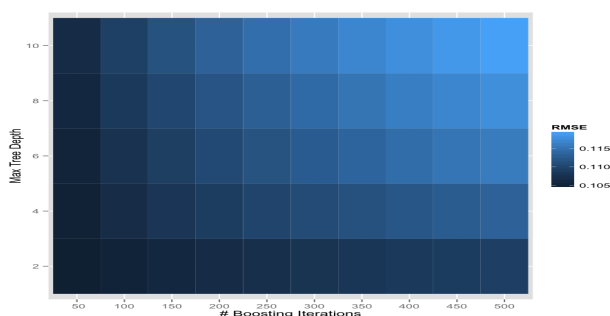


FIGURE 9.38 – Boosting Arbres de régression - Evolution de l'erreur quadratique moyenne en fonction de paramètres du modèle de Boosting d'arbres

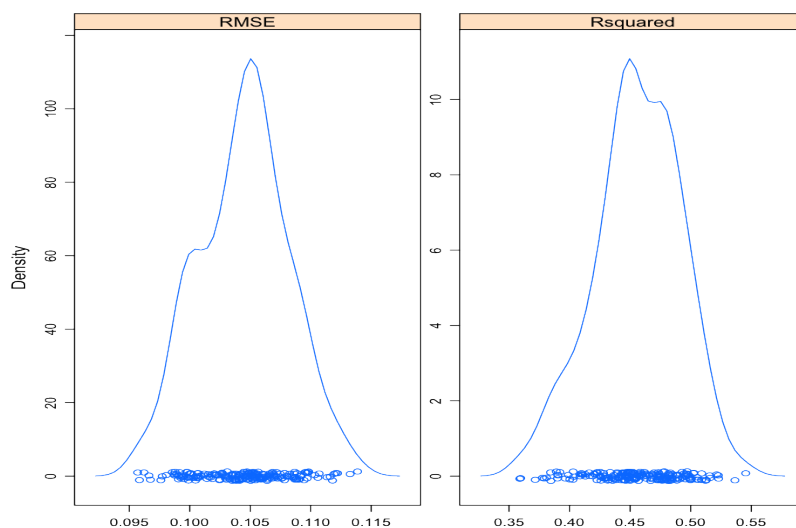


FIGURE 9.39 – Boosting Arbres de régression - Distribution de l'erreur quadratique moyenne et du R^2 dans le cadre du Boosting d'arbres de régression

Table des figures

1.1	Cartographie du SCR sous Solvabilité II	10
3.1	Définition de classes d'assurés à partir de leurs encours bancaires et assurantiels	24
3.2	Catégories fiscales	25
3.5	Lien entre données manquantes et données renseignées d'un jeu de données rudimentaire	29
3.6	Les données manquantes vues par Schafer et Graham	29
3.7	Comparaison de l'évolution du taux d'erreur entre 3 méthodes d'imputation de données manquantes mixtes	36
3.8	Comparaison du taux d'erreur entre les méthodes d'imputations : kNN, MICE et MissForest	37
3.9	Règle d'affectation d'un type de rachat à un contrat en cas de vagues multiples durant un exercice	38
3.10	Evaluation du taux de rachat partiel par variation de l'assiette de calcul - cas pratique (1/2)	39
3.11	Evaluation du taux de rachat partiel par variation de l'assiette de calcul - cas pratique (2/2)	39
4.1	Principe de l'AFDM	48
5.1	Répartition des variables quantitatives après rotation	51
5.2	Répartition des variables après rotation	52
5.3	Analyse croisée des variables	53
7.1	Composition d'un neurone biologique	61
7.2	Exemple d'un réseau de neurones élémentaire	62
7.3	Fonction de combinaison et d'activation dans un réseau de neurones	63
7.4	Illustration d'un réseau de neurones bouclé	64
9.1	Arbre de régression sur le Taux de rachat - Initialisation du coefficient de pénalisation	76
9.2	Coefficients de pénalité - Elagage de l'arbre de régression	76
9.3	Arbre de régression élagué	77
9.4	Taux de rachat par populations d'assurés	78
9.5	Arbre de régression élagué - version développée	78
9.6	Réseau de Neurones - Exemple d'un modèle à 3 couches intermédiaires de, successivement, 2, 4 et 6 neurones cachés	81
9.7	Réseau de Neurones - Sensibilité du PMC au nombre de couches intermédiaires	82
9.8	Réseau de Neurones - Sensibilité du PMC au nombre de neurones cachés . .	83
9.9	Réseau de Neurones - Sensibilité du PMC à la fonction d'activation	83
9.10	Réseau de Neurones - Sensibilité du PMC au taux d'apprentissage	83
9.11	Réseau de Neurones - Résultats d'apprentissage et de prévision sur un PMC {3 3} à fonction d'action sigmoïde et à taux d'apprentissage de 25%	84

9.12	Réseau de Neurones - Erreur de prévision sur un PMC {3 3}	84
9.13	Forêt Aléatoire - Premiers paramétrages	84
9.14	Forêt Aléatoire - Taux d'erreur et nombre d'arbres	85
9.15	Forêt Aléatoire - Importance des variables	86
9.16	Boosting Arbres de régression - RMSE et R^2 d'un Boosting d'arbres de régression	87
9.17	Agrégation de réseaux de neurones - Comparaison Bagging/Boosting	88
9.18	Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Boxplot RMSE et R^2	88
9.19	Bagging de Réseaux de Neurones - Importance des variables	89
9.20	Répartition des variables quantitatives après rotation - Portefeuille 2013	92
9.21	Répartition des modalités après rotation - Portefeuille 2013	93
9.22	Analyse croisée des variables - Portefeuille 2013	94
9.23	Arbre de régression sur le Taux de rachat - 2013	95
9.24	Taux de rachat par populations d'assurés	96
9.25	Arbre de régression - Erreur de prédiction sur le portefeuille 2013	97
9.26	Forêt Aléatoire - Taux d'erreur et nombre d'arbres - Portefeuille 2013	98
9.27	Forêt Aléatoire - Importance des variables - Portefeuille 2013	98
9.28	Forêt Aléatoire - Erreur de prédiction sur échantillon test	98
9.29	Boosting Arbres de régression - RMSE et R^2 d'un Boosting d'arbres de régression - Portefeuille 2013	99
9.30	Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Portefeuille 2013	100
9.31	Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Portefeuille 2013 - Boxplot RMSE et R^2	100
9.32	Boosting d'arbres de régressions - Importance des variables	101
9.33	Réseau de Neurones - PCM à une couche intermédiaire et à 2 neurones cachés	117
9.34	Réseau de Neurones - PCM à une couche intermédiaire et à 4 neurones cachés	117
9.35	Réseau de Neurones - PCM à une couche intermédiaire et à 6 neurones cachés	118
9.36	Réseau de Neurones - PCM à 2 couches intermédiaires et à 6 neurones cachés	118
9.37	Réseau de Neurones - PCM à 3 couches intermédiaires et à 6 neurones cachés	119
9.38	Boosting Arbres de régression - Evolution de l'erreur quadratique moyenne en fonction de paramètres du modèle de Boosting d'arbres	120
9.39	Boosting Arbres de régression - Distribution de l'erreur quadratique moyenne et du R^2 dans le cadre du Boosting d'arbres de régression	120

Bibliographie

- [1] L. Anzili and L. De Cesare. Valuation of the surrender option in unit-linked life insurance policies in a non-rational behavior framework. *Working Paper. Astin Colloquium Berlin*, 2007.
- [2] J.-M. Aouizerate. Alternative neuronale en tarification santé. Master's thesis, Master's thesis, CNAM, 2010.
- [3] A.R. Bacinello. Endogeneous model of surrender conditions in equity-linked life insurance. *Insurance Mathematics and Economics* 37, pages 270–296, 2005.
- [4] A.R. Barron. Universal approximation bounds for superpositions of a sigmoid function. *IEEE Transactions on Information Theory*, 39, pages 930–945, 1993.
- [5] V. Bauwers and J.-F. Valhin. *La titrisation du risque d'assurance*. Larcier, 2008.
- [6] C. Bernard, O. Le Courtois, and F. Quittard-Pinon. Development and pricing of a new participating contract. *North American Actuarial Journal*, 10, 2006.
- [7] O. Berne. Estimation des lois de rachat et de mortalité en assurance vie. Master's thesis, Master's thesis, ULP, 2013.
- [8] G. Biau. *Analysis of a random forests model*. 2010.
- [9] A. Blum. *Neural Network in C++*. NY : Wiley, 1992.
- [10] R.R. Cerchiara, M. Edwards, and A. Gambini. General linear models in life insurance : Decrements and risk factor analysis under solvency ii. working paper. afir colloquium rome. 2009.
- [11] A. Consiglio and D. De Giovanni. Pricing the option to surrender in incomplete markets. *Journal of Risk and Insurance*, pages 935–957, 2010.
- [12] O. Le Courtois and H. Nakagawa. On surrender and default risks. *Mathematical Finance, forthcoming*, 2011.
- [13] S.H. Cox and Y. Lin. Annuity lapse rate modeling : Tobit or not tobit? *Working paper. Society of Actuaries*, 2006.
- [14] L.A. Crosby and N. Stephens. Effects of relationship marketing on satisfaction, retention, and prices in the life insurance industry. *Journal of Marketing Research* 24, pages 404–411, 1987.
- [15] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems* 2, pages 303 – 314, 1989.
- [16] A. Dar and C. Dodds. Interest rates the emergency fund hypothesis and saving through endowment policies : Some empirical evidence for the u.k. *Journal of Risk and Insurance*, pages 415–433, 1989.
- [17] Autorité de Contrôle Prudentiel. Orientations nationales complémentaires aux spécifications techniques pour l'exercice 2013 de préparation à solvabilité ii. 2013.
- [18] J. Dhaene, D. Linders, W. Schoutens, and D. Vyncke. The herd behavior index : a new measure for systematic risk in financial markets. *Research Report. Feb, K.U. Leuven*, 2011.

- [19] D.W. Diamond and P.H. Dybvig. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy* 91, pages 401–419, 1983.
- [20] B. Escoffier. Traitement simultané de variables qualitatives et quantitatives en analyse factorielle. *Les Cahiers de l'analyse de données Vol. IV - 1979 - n°2*, pages 137–146.
- [21] B. Escoffier. Traitement simultané de variables quantitatives et qualitatives en analyse factorielle. *Les cahiers de l'analyse des données*, pages 137–146, 1979.
- [22] R.J.A. Little et D.B. Rubin. Statistical analysis with missing data. *New York : John Wiley*, 1987.
- [23] M. J. A. Berry et G. Linoff. *Data Mining Techniques*. NY : John Wiley Sons, 1997.
- [24] J. L. Schafer et J. W. Graham. Missing data : Our view of the state of the art. *Psychological Methods*, pages 144–177, 2002.
- [25] Opitz D. et Maclin R. Popular ensemble methods : An empirical study. *Journal of Artificial Intelligence Research* 11, pages 169–198, 1999.
- [26] D. J. Stekhoven et P. Bühlman. Missforest - nonparametric missing value imputation for mixed-type data. *Bioinformatics* 28, pages 113–118, 2011.
- [27] McCulloch et Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, pages 115–133, 1943.
- [28] S. Fauvel and M. Le Pévédic. Analyse des rachats d'un portefeuille vie individuelle : Approche théorique et application pratique. Master's thesis, Master's thesis, ENSAE, 2007.
- [29] A.J. Feelders. Handling missing data in trees : Surrogate splits or statistical imputation. *PKDD, Lecture Notes in Computer Science, t. 1704, Springer, 1999*, pages 144–177, 1999.
- [30] D. Fum, F. Del Missier, and S.A. The cognitive modeling of human behavior : Why a model is (sometimes) better than 10,000 words. *Cognitive Systems Research* 8, pages 135–142, 2007.
- [31] K. Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks* 2, pages 183 – 192, 1989.
- [32] N. Gatzert and H. Schmeiser. The impact of the secondary market on life insurer's surrender profits. *Journal of Risk and Insurance, Vol. 76*, pages 887–908, 2009.
- [33] D. De Giovanni. Lapse rate modeling : A rational expectation approach. *Scandinavian Actuarial Journal 2010*, pages 56–67, 2010.
- [34] M. Stinchcombe et H. White K. Hornik. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, pages 359–366, 1989.
- [35] S. Van Buuren et J. Van Rijkevorsel K. Oudshoorn, K. Flexible multiple imputation by chained equations of the avo-95. *TNO report PG/VGZ/99.045*, pages 520–525, 1999.
- [36] Y. Kagraoka. Quel avenir pour l'assurance dépendance ? leçons de l'expérience américaine. 2008.
- [37] D. Kiesenbauer. Main determinants of lapse in the german life insurance industry. *North American Actuarial Journal* 16, pages 52–73, 2011.
- [38] C. Kim. Modeling surrender and lapse rates with economic variables. *North American Actuarial Journal* 9, pages 56–70, 2005.
- [39] A. Kling, J. Ruß, and F. Ruez. The impact of policyholder behavior on pricing, hedging, and hedge efficiency of withdrawal benefit guarantees in variable annuities. *Working Paper. AFIR Colloquium Madrid.*, 2011.
- [40] A.W. Kolkiewicz and K.S. Tan. Unit-linked life insurance contracts with lapse rates dependent on economic factors. *Annals of Actuarial Science* 1, pages 49–78, 2006.

- [41] W. Kuo, C. Tsai, and W.-K. Chen. An empirical study on the lapse rate : The cointegration approach. *Journal of Risk and Insurance* 70, pages 489–508, 2003.
- [42] R. Olshen et C. Stone L. Breiman, J. Friedman. *Classification and regression trees*. Wadsworth Brooks, 1984.
- [43] AM.C. Ledlie, D.P. Corry, G.S. Finkelstein, A.J. Ritchie, K. Su, and D.C.E. Wilson. Variable annuities. *British Actuarial Journal* 14, pages 327–430, 2008.
- [44] J. Li and A. Szimayer. The effect of policyholders’ rationality on unit-linked life insurance contracts with surrender guarrantees. *Working paper*, 2010.
- [45] M.A. Linton. Panics and cash values. *Transactions of the Actuarial Society* 33, pages 365–394, 1932.
- [46] O. Lopez, X. Milhaud, and P.-E. Thérond. Tree-based censored regression with applications to insurance. 2015.
- [47] R. Luce. Behavioral assumptions for a class of utility theories : A program of experiments. *Journal of Risk and Uncertainty* 41, pages 19–37, 2010.
- [48] X. Milhaud. Segmentation et modélisation des comportements de rachat en assurance vie. Master’s thesis, Master’s thesis, ISFA, 2011.
- [49] X. Milhaud, S. Loisel, and V. Maume-Deschamps. Surrender triggers in life insurance : Classification and risk predictions. working paper. 2010.
- [50] E. Morgan and J. Kent. Dynamic policyholder behavior. *Working paper. Staple Inn Actuarial Society*, 2008.
- [51] J. Von Neumann and O. Morgenstern. *Theory of Games and economic Behavior*. Princeton University Press, 1947.
- [52] L. Personnaz G. Dreyfus et S. Marcos O. Nerrand, P. Roussel-Ragot. Neural networks and non-linear adaptive filtering : unifying concepts and new algorithms. *Neural Computation*, 5, pages 165–197, 1993.
- [53] G. Sherlock P. Brown T. Hastie R. Tibshirani D. Boltstein et R. Altman O. Troyanska, M. Cantor. Missing value estimation methods for dna microarrays. *Bioinformatics* 17(6), pages 520–525, 2001.
- [54] American Academy of Actuaries. Recommended approach for setting regulatory risk-based capital requirements for variable annuities and similar products. *Report. American Academy of Actuaries, Boston, MA*, 2005.
- [55] J.F. Outreville. Whole-life insurance lapse rates and the emergency fund hypothesis. *Insurance : Mathematics and Economics* 42, pages 249–255, 1990.
- [56] J. PAGES. Analyse factorielle de données mixtes. *Revue de statistique appliquée, tome* 52, n°4, pages 93–111, 2004.
- [57] X. Pan, C.S. Han, K. Dauber, and K.H. Law. Human and social behavior in computation modeling and analysis of egress. *Automation in Construction* 15, pages 448–461, 2006.
- [58] J.E. Pesando. The interest sensitivity of the flow of funds through life insurance companies : An econometric analysis. *Journal of Finance* 29, pages 1105–1121, 1974.
- [59] J. Pinquet. Quel avenir pour l’assurance dépendance ? leçons de l’expérience américaine. Technical report, 2008.
- [60] S. Pohl. Cancelling at the principal expiry date in motor insurance - discrete-time analysis of the risk rate with left-truncated data. *Eul Verlag*, 2009.
- [61] A.E. Renshaw and S. Haberman. Statistical analysis of life insurance lapses. *Journal of The Institute of Actuaries* 113, pages 459–497, 1986.
- [62] A.-C. Richard. La garantie en capital dans les produits d’épargne : de nouvelles solutions pour les assureurs. Master’s thesis, Master’s thesis, ISFA, 2013.

- [63] P. Royston. Multiple imputation for nonresponse in surveys. *New York : Wiley*, 2004.
- [64] Williams R.J. Rumelhart D.E., Hinton G.E. Learning internal representations by error backpropagation. *Parallel Distributed Processing : Explorations in the Microstructure of Cognition*, pages 318 – 362, 1986.
- [65] Freund Y. Bartlett P. Lee W. Schapire, R. Boosting the margin : A new explanation for the effectiveness of voting methods. *Proceedings of the Fourteenth International Conference on Machine Learning. Nashville, TN*, page 322–330, 1997.
- [66] R. Schapire. The strength of weak learnability. *Machine Learning*, 5(2), page 197–227, 1990.
- [67] F.H. Schott. Disintermediation through policy loans at life insurance companies. *Journal of Finance* 26, pages 719–729, 1971.
- [68] E. Scornet. Apprentissage et forêts aléatoires. Master’s thesis, Master’s thesis, ENS Ulm, 2011 - 2012.
- [69] D. Seror and D. Nkihoubonga Yengue. Modélisation des comportements de rachats dans un cadre de risque compétitif. Master’s thesis, Master’s thesis, ENSAE, 2013.
- [70] M. Shumrak, M. Greenbaum, V. Darley, and R. Axtell. Modeling annuity policyholder behavior using behavioral economics and complexity science. *Working Paper. Canadian Institute of Actuaries*, 1999.
- [71] A. Shun-Ichi. A theory of adaptive pattern classifiers. *Electronic Computers, IEEE Transactions on (Volume :EC-16 , Issue : 3)*, pages 299–307, 1967.
- [72] K. Swingler. *Applying Neural Networks : A Practical Guide*. London : Academic Press, 1996.
- [73] M. Tenenhaus. Analyse en composantes principales d’un ensemble de variables nominales ou numériques. *Revue de Statistique Appliquée*, XXV, pages 39–56, 1977.
- [74] C. Tsai, W. Kuo, D. Chiang, and M.-H. D. The distributions of policy reserves considering the policy-year structures of surrender rates and expense ratios. *Journal of Risk and Insurance* 76, pages 909–931, 2009.
- [75] E. Vannucci. The valuation of unit linked policies with minimal return guarantees under symmetric and asymmetric information hypotheses. *Proceedings of the 6th Spanish-Italian Meeting on Financial Mathematics, Vol II*, pages 569–582, 2003.
- [76] V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks, Vol.10, n°5*, 1999.
- [77] W.K. Viscusi, O. Phillips, and S. Kroll. Risky investment decisions : How are individuals influenced by their groups? *Vanderbilt Law and Economics Research Paper, N°11-33*, 2011.
- [78] P. Werbos. Beyond regression : New tools for prediction and analysis in the behavioral sciences. *Ph. D. dissertation, Committee on Appl. Math., Harvard Univ., Cambridge, MA*, 1974.