

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

Par : Anissa Bendounan

Titre : Versements libres : analyse et projection du comportement des assurés d'un portefeuille d'épargne individuelle

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

---

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

*Membres présents du jury de l'Institut  
des Actuaires :*

*Membres présents du jury de l'ISFA :*

*Entreprise :*  
Nom : SeaBird Conseil  
Signature :

*Directeur de Mémoire en entreprise :*  
Nom : Pierre Théron  
Signature :

*Pierre Théron*  
Pierre Théron (Apr 29, 2022 15:15 GMT+2)

---

*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*

*Pierre Théron*  
Pierre Théron (Apr 29, 2022 15:15 GMT+2)

*Signature du candidat*





## Résumé

---

La question de l'anticipation des comportements assurés dans le contexte de contrats d'assurance vie préoccupe depuis longtemps les assureurs, en quête d'une gestion optimale de leurs portefeuilles. L'émergence de la norme IFRS 17 conduit les assureurs au développement de modèles de projection des versements libres, comportements encore peu abordés dans la littérature au profit de ceux de rachat ou d'arbitrage. C'est à ce titre que nous abordons dans ce mémoire les problématiques de segmentation et de modélisation des versements libres, dans le but de mieux connaître et prendre en compte l'ensemble des facteurs-clés qui jouent sur les décisions des assurés. Nous remarquons une forte variabilité temporelle des taux de versements libres et identifions par conséquent dans ces comportements l'intervention d'une composante conjoncturelle, à l'image de ce qui est constaté sur les comportements de rachats. L'introduction de variables conjoncturelles dans la modélisation de ces comportements permet alors d'améliorer de façon considérable les projections, en comparaison de modèles basés uniquement sur les caractéristiques assurées. La méthodologie implémentée permet la construction de lois de projection pertinentes et facilement intégrables dans les outils de projection des flux futurs de trésorerie.

---

*Mots-clés : Versements libres, Comportements assurés, Régression logistique, Segmentation, IFRS 17, Frontière des contrats.*

## Abstract

---

The modelling of policyholder behaviour in the context of life insurance has been of interest for insurers for a very long time as they intend to optimise the management of their portfolios. The new IFRS 17 accounting standard is leading them to develop models for projecting voluntary premium payments, behaviours that are still little discussed in the literature unlike those of withdrawals or fund switches. In this context this paper addresses the challenges of segmentation and modelling of voluntary premium payments, with the aim of better understanding and taking into account all the key factors that influence policyholders' decisions. We note a strong variability in the premium payment rates over the time, and as a consequence identify in these behaviours an economic component, similar to what is observed in the withdrawals behaviour. The introduction of economic variables in the modelling of these behaviours allows us to considerably improve the projected cash flows, in comparison to models based only on insured characteristics. The methodology implemented allows the construction of relevant projection rules that can be easily integrated into future cash flow projection tools.

---

*Keywords : Top-ups; Policyholder behaviours, Logistic regression, Partitioning, IFRS 17, Contract boundary.*





# Note de synthèse

À l'approche de l'entrée en application de la norme IFRS 17, ce mémoire se propose de fournir une méthodologie détaillée de la construction de lois de projection des versements libres. Cette étude est motivée par la nouveauté sous IFRS 17 de l'inclusion des versements libres dans la frontière des contrats, absente sous le cadre Solvabilité II, qui amène les assureurs vie à se questionner sur la mise en place de telles lois. Par ailleurs, et outre ces nouvelles considérations réglementaires, la bonne connaissance des comportements des assurés est un enjeu majeur pour les assureurs en quête perpétuelle d'une gestion optimale de leurs portefeuilles.

Pour cela, nous basons notre étude sur l'utilisation de données relatives à des contrats multisupports d'épargne individuelle sur un historique allant de 2006 à 2020. Ces données comprennent des variables caractéristiques des assurés ainsi que des variables relatives à l'encours annuel du contrat et aux versements enregistrés sur l'année.

Au préalable de la construction des lois de projection, ce mémoire se propose de fournir une étude des comportements de versements libres, peu étudiés dans la littérature. L'objectif de cette première étape est alors d'identifier les tenants et les aboutissants de ces comportements. Pour cela, nous distinguons les deux notions inhérentes à l'observation d'un versement libre : son occurrence et son montant. Nous cherchons alors à identifier les effets sur chacune de ces deux variables des informations à disposition : les caractéristiques des assurés.

À l'aide d'analyses statistiques classiques et de méthodes de classification, nous soulignons le pouvoir discriminant de certaines variables sur les comportements de versements libres. Ainsi l'ancienneté du contrat montre une importance particulière sur la décision de versements libres des assurés du portefeuille. Quant à l'étude des montants de ces versements, nos analyses nous permettent d'établir le fort lien de cette variable avec le niveau d'encours et le type de produit souscrit par l'assuré.

Nous observons par ailleurs une forte variabilité temporelle des taux de versements libres observés, contrairement aux montants moyens qui leur sont associés. Nous identifions

alors l'intervention d'une composante conjoncturelle dans la décision de versement libre, à l'image de ce qui est constaté sur les comportements de rachats. Nous établissons alors de fortes corrélations entre la variable décrivant l'évolution des taux de versements annuels et des variables conjoncturelles comme celle décrivant les évolutions des cotations moyennes du CAC 40. Nous considérons alors, sur la base de ces corrélations, un panel de variables conjoncturelles à prendre en compte dans la modélisation des taux de versements libres.

Sur la base des enseignements de cette première étape de l'étude, nous procédons ensuite à la modélisation des comportements de versements libres, en distinguant la modélisation de leurs occurrences de celle des montants.

Afin de juger de l'apport de l'introduction de variables conjoncturelles dans la qualité d'ajustement de la modélisation, nous ne les introduisons que dans un second temps de la modélisation des taux de versements libres. Nous adoptons ainsi deux approches de modélisation de cette variable : une approche statique considérant les versements observés sur le portefeuille comme réalisés au cours d'une même période ainsi qu'une approche dynamique considérant les évolutions du contexte économique des périodes d'observations. L'implémentation de méthodes de régression logistique selon les deux approches décrites révèle l'insuffisance des variables structurelles du portefeuille à l'explication des variations temporelles des taux de versements observés sur le portefeuille. La modélisation basée sur la deuxième approche considérée, permet quant à elle de fournir des résultats satisfaisants.

Pour la modélisation des montants associés aux versements libres, nous choisissons de miser sur la profondeur de l'historique à disposition et sur la stabilité observée de leurs valeurs moyennes. Nous construisons alors des tables de prédiction des montants moyens de versements libres que nous lisons ensuite à l'aide d'une méthode de régression locale, à l'image des pratiques de construction de lois d'expérience.

L'association des deux lois de projections ainsi construites fournit des estimations adaptées aux observations réalisées sur le portefeuille. Nous observons cependant une forte volatilité des estimations, en raison du couplage des volatilités des deux estimations. Ces volatilités sont à prendre en considération en raison de leurs potentiels impacts sur les grandeurs du passif prenant en compte ces projections. Néanmoins, cette étude permet la construction de lois comportementales pertinentes et facilement intégrables dans les outils ALM de projection des flux de trésorerie associés aux contrats.

# Synthesis note

In the run-up to IFRS 17, this paper provides a detailed methodology for the modelling of future top-up premiums to individual saving products. This analysis stems from the new inclusion under IFRS 17 of top-ups within the contract boundary. It leads insurers to consider the implementation of such models for future premiums not falling within the Solvency II contract boundary. Note that in addition to these new accounting considerations, in-depth knowledge of policyholder behaviour has been a major challenge for insurers in their constant seek for optimal management of their portfolios.

To do so, we have based our study on historical data relating to additional premiums paid to individual multi-fund saving policies (combining unit-linked and with-profit funds) over a period ranging from 2006 to 2020. This data includes policyholder details as well as the annual saving amount under each policy and premiums paid during the year.

Prior to building a methodology about future premiums, this paper presents an analysis about top-ups behaviours, rarely studied in the literature. This first step aims to get an in-depth understanding of these behaviours. To do this, we distinguish the two aspects of a voluntary payment: its occurrence and its amount. We then seek to identify the effects (on each of these two variables) of the available information regarding the policyholder characteristics.

Using traditional statistical analyses and classification methods, we highlight the discriminating power of some variables on top-ups behaviours. It turns out that the length of time the policy has been in force is a key decision factor. As for the payment amounts, our analyses establish a strong correlation with the level of saving amounts (in the policy) and the type of saving product.

We also observe a high variability in the top-ups rates over time, unlike for the average amounts. Therefore we identify an economic component in the decision to top up saving policies, similar to what is observed for withdrawals behaviours. We then establish strong correlations between the variable describing the evolution of annual premium payment rates and market variables such as the one describing the fluctuation of the average CAC

40 equity index. Then, based on these correlations, we consider a range of economic variables to be considered in the modelling of top-ups rates.

Based on the conclusions drawn from this first stage, we proceed to the modelling of top-ups behaviours, splitting the modelling of their occurrences from that of their amounts.

In order to assess the contribution of economic variables to the quality of the modelling, we introduce them only in a second stage. We adopt two approaches to modelling this variable: a static approach that considers the premiums paid in the portfolio as having been made during the same period, and a dynamic approach that considers changes in the economic context. The implementation of logistic regression methods according to the two approaches shows that there is no evidence of any structural variables associated with the portfolio to explain the variations of the premium payment rates over time. However the modelling based on the second approach gives satisfactory results.

To model the additional premium amounts, we rely on the available historical data and on the observed stability of their average values. Then we build prediction tables for the average top-up amounts, which we smooth using a local regression method, similar to what is used to build experience tables.

The combination of the two modelling approaches provides estimates based on the observations made regarding the portfolio. However, we observe a high volatility of the estimates resulting from the combination of the volatilities of the two estimates. These volatilities should be taken into consideration because of their potential impact on the liability components that take into account these projections. Nevertheless, this analysis allows the building of relevant behavioral rules that can be easily integrated into ALM tools projecting future cash flows attached to saving policies.

# Remerciements

Je tiens tout d'abord à remercier Pierre Théron et que Baptiste Dieltiens pour leur implication ainsi que pour leurs précieux conseils dans la réalisation de mon mémoire. De manière générale, je remercie aussi mes collègues de SeaBird pour leur accueil et leur bonne humeur qui m'ont permis de réaliser ce mémoire dans les meilleures conditions.

Je remercie également mon tuteur académique, Yahia Salhi, de m'avoir guidée dans l'orientation de ce mémoire.

Je remercie finalement l'ensemble de mes proches pour leur soutien dans ces derniers moments qui ont été tout sauf faciles ainsi que pour leurs nombreuses relectures.

# Table des matières

<b>Note de synthèse</b>	<b>5</b>
<b>Synthesis note</b>	<b>7</b>
<b>Remerciements</b>	<b>9</b>
<b>Introduction</b>	<b>11</b>
<b>1 Cadre de l'étude, enjeux et réglementation</b>	<b>14</b>
1.1 Présentation générale de l'assurance vie et chiffres clés . . . . .	14
1.2 Motivations et enjeux de l'étude . . . . .	19
1.3 Présentation des données et retraitements . . . . .	25
<b>2 Analyse et segmentation des comportements de versements libres</b>	<b>34</b>
2.1 Analyse des caractéristiques des assurés en portefeuille et de leur impact sur les comportements de versements libres . . . . .	36
2.2 Segmentation des comportements de versements libres . . . . .	43
2.3 Étude de la corrélation du taux de versements libres avec la conjoncture économique . . . . .	52
<b>3 Construction des lois de projection des versements libres</b>	<b>59</b>
3.1 Modélisation des probabilités de versements libres . . . . .	60
3.2 Modélisation des montants moyens de versements libres . . . . .	74
3.3 Agrégation et analyse des résultats des modèles . . . . .	81
<b>Conclusion</b>	<b>85</b>
<b>Bibliographie</b>	<b>87</b>

# Introduction

L'assurance vie constitue depuis plusieurs années le placement préféré des Français. À fin 2021, la Fédération Française de l'Assurance observe plus de 46 millions de contrats en cours, dont 2,3 millions de nouvelles souscriptions dans l'année et un encours s'élevant à 1 876 milliards d'euros, en progression de 4,4 % par rapport à l'année précédente. Ce produit continue de séduire les épargnants en répondant à leurs besoins de rendement et de liquidité accompagnés de la sécurité financière d'une compagnie d'assurance. Il affiche une collecte nette positive de 23,7 milliards d'euros à fin 2021. On distingue trois types de dépôt d'argent sur un contrat d'assurance vie : le versement initial, réalisé à l'ouverture du contrat ; les versements programmés, permettant la planification de son épargne ; et enfin les versements libres, donnant la possibilité à l'assuré de pouvoir augmenter le capital de son contrat à tout moment. C'est ce dernier type de versement que se propose d'étudier ce mémoire.

En effet, dans le cadre de l'application de la norme IFRS 17, les assureurs vie sont maintenant invités à intégrer la modélisation des versements libres dans leurs projections des flux de trésorerie futurs. Les versements libres n'étant pas pris en compte sous Solvabilité II, cette nouvelle exigence amène alors les assureurs vie à se questionner sur la mise en place d'une fonction de projection de versements libres. Outre ces nouvelles exigences réglementaires, la bonne compréhension du comportement des assurés est un enjeu majeur pour les assureurs en quête d'une gestion optimale de leurs portefeuilles. En effet, une meilleure connaissance des assurés du portefeuille pourrait par exemple permettre l'identification des profils les plus enclins à effectuer un versement libre et ainsi servir à des fins marketing.

La littérature sur l'étude des comportements assurés s'est largement développée ces dernières années et plus particulièrement sur le sujet de la modélisation des rachats, en raison de leurs risques significatifs pour l'assureur. Si le versement libre est, par essence, opposé au rachat, il serait prématuré de considérer que les facteurs déclencheurs des rachats produisent un effet inverse quant à la décision d'effectuer un versement libre. Néanmoins, nous nous sommes intéressés aux conclusions de ces études et aux tech-

niques employées pour leur modélisation. Ainsi, de nombreuses études sur les rachats distinguent ce phénomène en une partie structurelle (causée par les caractéristiques propres aux assurés) et en une partie conjoncturelle (causée par les fluctuations du contexte économique). [Milhaud \(2011\)](#) identifie même dans son étude une corrélation parmi les décisions de rachats de certains groupes d'assurés qui se distinguent par leurs réactions aux variations du climat économique. Les versements libres ne faisant pas l'objet d'un risque aussi significatif que celui apporté par le rachat (du moins lorsqu'ils ne sont pas rémunérés au taux contractuellement garanti par l'assureur, ce qui est la majorité des cas en pratique) leur étude n'a donné lieu qu'à peu de travaux. [Assaraf \(2020\)](#) s'est intéressée à la modélisation de ce phénomène, en excluant la prise en compte du contexte économique du fait du faible historique d'observations disponible à l'étude. Son étude, basée sur l'utilisation de techniques de machine learning, fournit des résultats concluants, mais peu utilisables en pratique au vu des contraintes imposées par les modèles de projection couramment utilisés.

La revue de la littérature sur les comportements assurés nous a permis d'élaborer une liste de questions auxquelles nous nous proposons de répondre. Quelles caractéristiques différencient un assuré réalisant un versement libre d'un autre n'en effectuant pas? Le montant du capital épargné a-t-il un impact sur le comportement de l'assuré? Verse-t-on moins à partir d'un certain niveau d'encours atteint? Observons-nous la même volonté d'épargne sur toute la durée de vie des contrats? La première phase du travail sera alors d'analyser le comportement des assurés afin de répondre à cette liste (non exhaustive) de questions, puis nous considérerons des effets moins évidents tels que ceux relatifs à la tenue des marchés financiers. Là encore, nous nous poserons la question de quels effets économiques pourraient avoir un impact sur le comportement assuré? Et quelles variables macroéconomiques prendre en compte le cas échéant? Le défi de cette deuxième phase de travail sera alors double, car demandera l'identification des variables reflétant l'éventuel phénomène économique impactant la décision des assurés. Nous pouvons cependant en intuitionner certaines comme le Taux Moyen des emprunts de l'État, guidant la rémunération servie sur les fonds euros, ou encore la cotation du CAC 40, corrélée au rendement moyen des unités de compte. Nous nous baserons alors sur les multiples études réalisées ces dernières années en temps de crise afin d'établir une présélection des potentiels effets de la conjoncture économique sur les comportements assurés. Nous pensons notamment à la récente crise sanitaire provoquée par la pandémie de Covid-19, où beaucoup de Français, par crainte de l'avenir, se sont constitués une épargne supplémentaire. Cependant, si les études ([ACPR \(2021a\)](#)) ont montré une hausse de l'épargne des ménages en 2020, ce sont vers des produits d'épargne bancaire plus liquides comme le Livret A que les cotisations des Français se sont dirigées au détriment de l'assurance vie qui a connu cette même année une baisse nette de ses cotisations. On retrouve alors deux aspects dans les comportements observés en 2020 : une volonté en temps de crise de se constituer une épargne de précaution,



accompagnée d'une préférence pour des produits d'épargne offrant un rendement plus faible, mais présentant l'avantage d'être plus liquides.

L'objectif de ce mémoire est donc de proposer une étude du comportement des assurés sur ce plan conduisant ensuite à la construction d'un modèle prédictif des versements libres sur la base des enseignements de la première étude. Préalablement au développement de ces deux axes, nous exposerons le cadre dans lequel s'inscrit cette étude. La fréquence et le montant de versements libres dépendant intuitivement de l'objectif d'épargne des assurés et donc du produit étudié, nous avons choisi de concentrer le périmètre de notre étude aux produits d'assurance vie à adhésion individuelle.

# Chapitre 1

## Cadre de l'étude, enjeux et réglementation

Dans ce chapitre, nous présenterons les principales caractéristiques du marché de l'assurance vie ainsi que le cadre réglementaire auquel ce secteur est soumis. Nous détaillerons les diverses options de gestion afférentes aux contrats d'assurance vie, dont les versements libres font partie. Nous soulignerons ensuite les enjeux de cette étude en exposant les différences existantes parmi les normes constituant le cadre réglementaire de l'assurance vie. Nous compléterons finalement la définition du cadre dans lequel s'inscrit ce mémoire en fournissant une description des données étudiées.

### 1.1 Présentation générale de l'assurance vie et chiffres clés

Un contrat d'assurance vie est un placement financier permettant à son souscripteur de constituer et de valoriser un capital. Cet outil d'épargne se distingue des autres placements par la fiscalité avantageuse qu'il propose, nous reviendrons sur ce point en détail dans une autre partie, mais aussi par sa volonté d'être un outil de transmission de patrimoine. En effet, le souscripteur a la possibilité, au moment de l'ouverture de son contrat, de désigner un ou plusieurs bénéficiaires qui se verront transmettre le capital constitué au moment de son décès.

On distingue ainsi communément deux types de contrats d'assurance vie : les contrats d'assurance en cas de vie, outil d'épargne pouvant servir à la réalisation d'un projet ou à

compléter sa retraite par exemple, et les contrats d'assurance en cas de décès, véhicules de transmission de patrimoine.

Ces deux avantages, en plus d'une rémunération attractive du capital, font de l'assurance vie le placement préféré des Français et lui offrent une place importante dans l'économie française avec un encours s'élevant à 1876 milliards d'euros à la fin de l'année 2021, en progression de 4,4 % par rapport à l'année précédente (Fédération Française de l'Assurance, FFA).

### 1.1.1 Différents supports d'investissement

Lors de l'ouverture d'un contrat d'assurance vie, le souscripteur a la possibilité d'investir son capital sur un seul support d'investissement, on parle alors de contrat monosupport, ou de répartir son investissement sur différents supports, on parle alors de contrat multisupport. Parmi ces différents supports d'investissement, on retrouve notamment les supports euros et les unités de compte.

#### Supports Euro

Lors de la signature d'un contrat d'assurance vie sur un support Euro, l'assureur garantit au souscripteur le capital déposé et s'engage à le revaloriser annuellement au Taux Minimum Garanti (TMG), inscrit dans le contrat. Les deux parties s'accordent ainsi sur le contrat (minimum) en euros que percevra le souscripteur (ou ses bénéficiaires) à l'issue du contrat, d'où l'appellation de ce support.

Ainsi, pour un contrat monosupport Euro, l'intégralité du risque de marché est supportée par l'assureur qui investit le capital sur des actifs peu risqués, essentiellement des obligations.

Par mesure de prudence, le TMG proposé par l'assureur est encadré par le Code des Assurances, d'après l'article A132-1 celui-ci doit être « *au plus égal à 75 % du taux moyen des emprunts de l'État français calculé sur une base semestrielle sans pouvoir dépasser, au-delà de huit ans, le plus bas des deux taux suivants : 3,5 % ou 60 % du taux moyen indiqué ci-dessus* ». Ces dernières années, les TMG moyens proposés par les assureurs vie ont vu leur valeur progressivement baisser, parallèlement à la chute des taux obligataires, pour atteindre une valeur moyenne de 0,39 % en 2020 (ACPR (2021b)). Le taux auquel l'assureur s'engage à revaloriser le contrat chaque année est exprimé brut de chargements. Ainsi, à la revalorisation fournie par l'assureur se déduisent les frais de gestion appliqués au contrat. En 2020, l'ACPR décrit un taux de chargement moyen pratiqué par les assureurs de 0,62 %, soit une différence moyenne de 23 points de pourcentage avec le taux technique moyen appliqué et donc un taux de revalorisation net de chargements négatif.

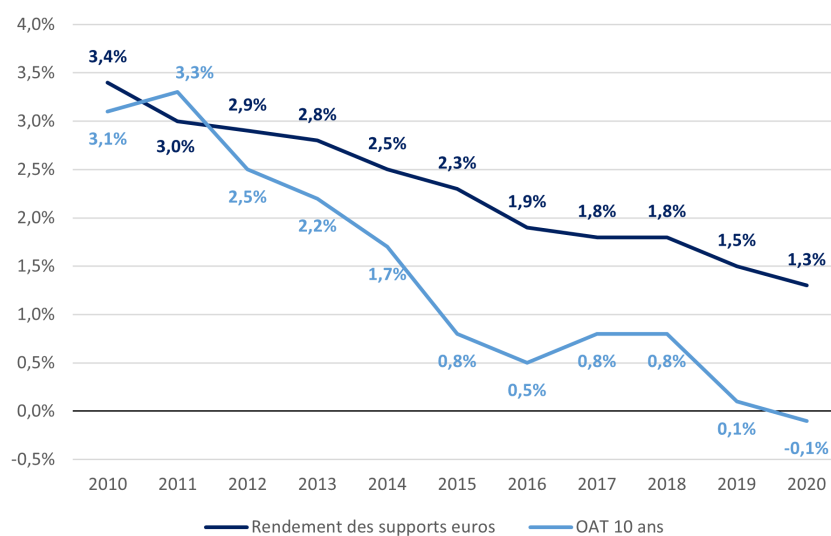


FIGURE 1.1 – Rendements<sup>1</sup> des supports euros (Source : FFA)

Par ailleurs, l'assuré peut se voir attribuer chaque année une revalorisation plus importante que celle prévue par son contrat. En effet, l'assureur est soumis par le Code des Assurances à reverser à ses assurés une partie de son résultat, appelée Participation aux Bénéfices. Son montant minimal se calcule comme la somme de 90 % du résultat technique de l'assureur ainsi que de 85 % de son résultat financier. L'assureur a ensuite huit ans pour reverser cette participation aux bénéfices selon les critères qu'il définit (selon les produits par exemple). La participation aux bénéfices est alors utilisée par les assureurs à des fins de pilotage de leur activité, leur permettant ainsi d'améliorer la compétitivité de leurs offres vis-à-vis de la concurrence ou alors de compenser la perte engendrée par d'anciens contrats toujours en portefeuille, proposant des TMG élevés.

### Supports en unités de compte (UC)

En cas de placement de son capital sur des supports en unités de compte, c'est l'assuré qui supporte le risque de marché. Au moment de la signature de son contrat, il se voit attribuer un nombre de parts d'UC en échange de son capital. Ce capital est alors investi sur différents actifs financiers comme des actions, des obligations ou des actifs immobiliers.

Ce nombre de parts lui est garanti jusqu'à la fin de son contrat, mais la valeur de ces parts, elle, fluctue. Ce support d'investissement, bien que plus risqué que le support euro, permet en contrepartie à l'assuré de prétendre à des rendements plus élevés.

Les assureurs ne sont soumis à aucune obligation de revalorisation vis-à-vis de ces supports,

1. Nets de chargements de gestion et bruts de prélèvements sociaux et fiscaux

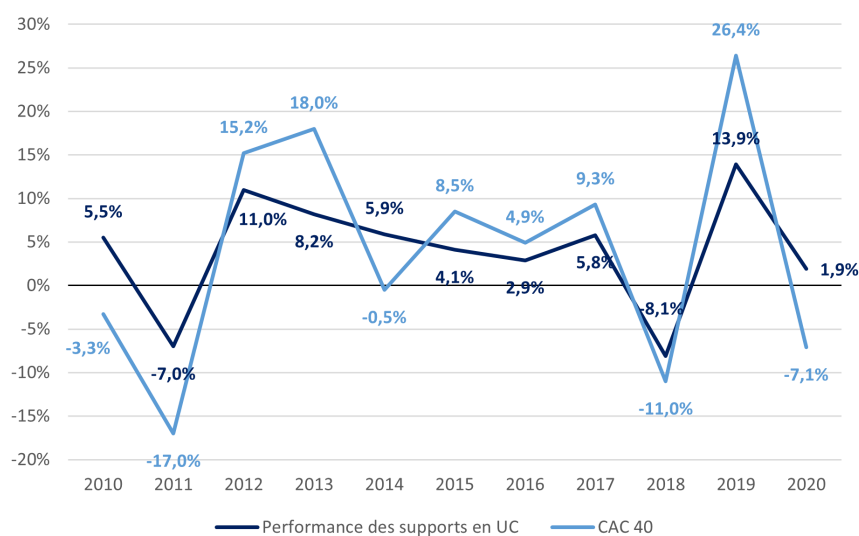


FIGURE 1.2 – Performance<sup>1</sup> des supports en unités de compte (Source : FFA)

mais proposent pour la plupart la souscription d'options payantes appelées garanties planchers, permettant à l'assuré de retrouver au moins la valeur de son capital investi en cas de chute de valeur de ses parts d'UC.

### 1.1.2 Fiscalité d'un contrat d'assurance vie

Comme tous produits financiers, les intérêts issus des contrats d'assurance vie sont soumis aux cotisations sociales et à l'imposition.

Les cotisations sociales ont la particularité de ne pas être calculées de la même façon selon les supports d'investissement :

- dans le cas d'un support euro, les intérêts produits chaque année sont abattus de 17,2 % avant d'être comptabilisés dans la provision mathématique du contrat ;
- dans le cas d'un support en UC, la provision mathématique est déduite, à l'issue du contrat, de 17,2 % de la somme des intérêts produits chaque année.

À la clôture de son contrat ou lors de la récupération d'une partie de son capital revalorisé, l'assuré se voit offrir deux possibilités d'imposition sur les intérêts perçus :

- les intégrer à sa déclaration de revenus ;
- les déduire d'un Prélèvement Forfaitaire Libératoire (PFL) qui s'articule comme suit :

1. Bruts de chargements de gestion et de prélèvements sociaux et fiscaux

	Primes versées avant le 29/07/2017			Primes versées à partir du 27/09/2017		
Ancienneté du contrat	< 4 ans	4 – 8 ans	≥ 8 ans	< 8 ans	≥ 8 ans	
Montant des primes versées	Toutes primes confondues			< 150 000 €	≥ 150 000 €	
Abattement annuel de 4 600 € ou de 9 200 €	Non	Non	Oui	Non	Oui	Oui
Imposition lors du rachat (PFL)	35 %	15 %	7.5 %	12.8 %	7.5 %	12.8 %

FIGURE 1.3 – Valeur du PFL en fonction de l'ancienneté du contrat (Source : [impots.gouv.fr](http://impots.gouv.fr))

Ainsi, plus l'ancienneté du contrat augmente, plus l'assurance vie promet une fiscalité avantageuse sur les plus-values, cela a une influence sur la durée de vie des contrats. En effet, les nombreuses études réalisées sur le risque de rachat soulignent souvent l'observation de pics des taux de rachat après la 4e et 8e année de la durée de vie du contrat.

### 1.1.3 Options de gestion d'un contrat d'assurance vie

Tout au long de la durée de vie de son contrat, un assuré a la possibilité de gérer son épargne en en récupérant une partie, en y ajoutant du capital ou encore en modifiant ses supports d'investissements. Ces trois opérations se décrivent comme suit :

#### Les rachats

On parle de rachat total ou partiel lorsque l'assuré demande à récupérer tout ou une partie de son capital investi.

La possibilité d'effectuer cette opération à tout moment confère à l'assuré l'avantage d'une liquidité de son épargne, mais peut en contrepartie avoir un coût pour l'assureur. En effet, il semble évident que dans le cas d'un contrat en UC, l'assuré ne décidera pas de racheter son contrat si la valeur de ses UC est inférieure au capital investi initialement, auquel cas, il se retrouverait en situation de perte. Cependant, dans le cas d'un contrat en euro, c'est l'assureur qui peut se retrouver en situation de perte, car c'est lui qui supporte le risque de l'investissement : il se peut par exemple que l'actif sur lequel était investi ce capital ait diminué de valeur alors que l'assureur est obligé de le revendre pour pouvoir verser le rachat demandé par l'assuré. Face à cette situation, l'assureur peut décider d'appliquer des pénalités en cas de rachat.

## Les versements

On distingue trois types de versements sur un contrat d'assurance vie :

- les versements initiaux : montant versé à l'ouverture du contrat ;
- les versements programmés : montant défini versé à une périodicité définie ;
- les versements libres : montant libre versé à la volonté de l'assuré.

À noter que certains contrats exigent un montant minimal de versements dans l'année, sans distinction de leur type, et que la réalisation de versements programmés n'exclut pas la réalisation de versements libres. La possibilité d'effectuer un versement libre confère à l'assuré la possibilité d'augmenter son épargne à tout moment. Comme pour les rachats, cette option peut représenter un risque pour l'assureur, notamment si le TMG du contrat s'applique aussi aux versements libres et qu'il est élevé, alors que les taux d'intérêt proposés par le marché sont faibles. L'assureur se doit alors, dans ce cas, d'anticiper au mieux ces versements libres.

## Les arbitrages

Cette 3e opération permet à l'assuré de modifier la répartition de l'investissement de son épargne dans le cas d'un contrat multisupport.

L'assuré peut aussi choisir de confier la gestion de la stratégie d'investissement de son épargne à l'assureur qui décidera des arbitrages à effectuer sur le contrat en fonction des rendements des différents supports choisis par l'assuré et de son aversion au risque. On parle alors de gestion profilée du contrat.

## 1.2 Motivations et enjeux de l'étude

Dans le cadre de l'application de la norme IFRS 17, les assureurs vie sont maintenant invités à intégrer la modélisation des versements libres dans la projection des flux financiers associés aux contrats rentrant dans son champ d'application. Les versements libres n'étant pas pris en compte sous Solvabilité II, cette nouvelle exigence amène donc les assureurs vie à se questionner sur la mise en place d'une fonction de projection de versements libres. Nous introduisons dans cette partie une rapide présentation de ces deux normes ainsi que de leurs différences de définition des frontières des contrats constituant la motivation de cette étude.

### 1.2.1 La norme IFRS 17

L'IASC (*International Accounting Standard Committee*), dirigé par l'*IASC Foundation*, naît en 1973 en Angleterre. Dès 1982, il devient le normalisateur comptable international de référence. Son rôle est alors d'établir des standards de comptabilité visant à offrir une meilleure comparabilité des états financiers des entreprises cotées.

En 2001, l'IASC devient l'IASB (*International Accounting Standard Board*) et reprend le projet entrepris par l'IASC de s'intéresser à la comptabilisation des contrats d'assurance. Elle publie alors la norme IFRS 4 en deux phases : une première, à titre de mesure transitoire en 2004, complétée de la deuxième en 2017, dont le nom sera remplacé par IFRS 17.

IFRS 4 Phase 1 impose alors une nouvelle comptabilisation de l'actif à sa juste valeur tandis qu'aucune règle n'est imposée pour la comptabilisation du passif. La norme engendre ainsi une différence comptable entre l'évaluation de l'actif et celle du passif des assurances. Cette différence, bien que partiellement corrigée par l'application de la comptabilité de reflet, a la conséquence de complexifier la comparaison des états financiers des compagnies d'assurances, alors même que cela faisait partie des objectifs de l'IASB.

En 2017, l'IASB publie la norme IFRS 17 (anciennement IFRS 4 Phase 2), cette norme vient corriger les lacunes présentées par la précédente en imposant une comptabilisation du passif d'assurance à sa juste valeur.

#### Champ d'application de la norme

Selon la norme IFRS 17, une entreprise de (ré)assurance doit appliquer ses directives aux :

- contrats de (ré)assurance qu'elle émet ;
- contrats de réassurance qu'elle détient ;
- contrats d'investissement à participation discrétionnaires qu'elle émet.

Un contrat de (ré)assurance, au sens de la norme IFRS 17, se définit de la façon suivante :

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

(IFRS 17 *Insurance Contracts*, Annexe A – Définitions)

La norme définit le risque d'assurance comme tout autre risque que le risque financier qui est transféré à l'émetteur du contrat par le titulaire. Le risque financier étant le risque découlant de la variation d'un sous-jacent financier comme la variation de la valeur d'un actif ou des taux d'intérêt. Le risque de rachat par exemple n'est alors pas considéré comme



un risque d'assurance.

L'aspect significatif du risque d'assurance s'apprécie, quant à lui, de la façon suivante :

*"Le risque d'assurance est significatif si et seulement si un événement assuré peut obliger un assureur à payer des prestations complémentaires significatives dans n'importe quel scénario, à l'exclusion des scénarios qui manquent de substance commerciale"*

(IFRS 17 *Insurance Contracts*, Annexe B - paragraphe 18)

*"Les prestations complémentaires décrites au paragraphe B18 font référence aux montants qui excèdent ceux qui seraient payables si aucun événement assuré ne se produisait (à l'exclusion des scénarios qui manquent de substance commerciale)"*

(IFRS 17 *Insurance Contracts*, Annexe B - paragraphe 20)

La norme indique de plus que le caractère significatif du risque d'assurance doit être apprécié à la maille du contrat et non du portefeuille de contrats. Ainsi, le risque d'assurance relatif à un contrat d'assurance peut être considéré significatif, tandis que le risque mutualisé au sein d'un portefeuille de contrats est lui non significatif.

Les contrats d'investissements avec participation discrétionnaires se définissent, quant à eux, comme suit :

*"Un instrument financier qui fournit à un investisseur particulier le droit contractuel de recevoir, en complément d'une somme non soumis à la discrétion de l'émetteur, des montants supplémentaires :*

- 1. dont on s'attend à ce qu'ils représentent une part importante du total des avantages contractuels ;*
- 2. dont le calendrier ou le montant sont contractuellement à la discrétion de l'émetteur ; et*
- 3. qui sont contractuellement basés sur :*
  - (a) les rendements d'un groupe de contrats ou d'un type de contrat spécifique type de contrat spécifié,*
  - (b) les rendements d'investissement réalisés et/ou non réalisés sur un ensemble déterminé d'actifs détenus par l'émetteur, ou*
  - (c) le bénéfice ou la perte de l'entité ou du fonds qui délivre le contrat."*

(IFRS 17 *Insurance Contracts*, Annexe B - paragraphe 20)

La participation discrétionnaire est alors par exemple la part des bénéfices distribuée au-delà des minima légaux dans le cadre d'un contrat d'assurance vie en euros.

## **De nouvelles règles de comptabilisation des passifs de (ré)assurance**

La norme IFRS17 introduit une nouvelle approche par blocs pour la comptabilisation des passifs de (ré)assurance. De manière générale, on y retrouve les trois suivants :

- le *Best Estimate* (BE) représentant la provision constituée de la meilleure estimation des flux de trésorerie futurs actualisés des contrats entrant dans le champ d'application de la norme. Sa définition est très proche de la définition du Best Estimate sous Solvabilité II ;
- l'ajustement pour risque (*Risk Margin*, RA) qui correspond à la provision constituée pour se prémunir du risque non financier porté par l'estimation du BE. Le risque non-financier correspond au risque de mauvaise estimation des flux de trésorerie futurs pour le calcul du BE ;
- la marge contractuelle de service (*Contractual Service Margin*, CSM) représentant les bénéfices non acquis pour services futurs que l'entité comptabilisera au fur et à mesure qu'elle fournira des services dans le cadre de l'exécution des contrats.

La CSM est la grande nouveauté d'IFRS 17. Elle est introduite dans le but de constituer une provision des profits futurs associés aux contrats afin de lisser dans le temps les résultats des entreprises de (ré)assurance.

### 1.2.2 La directive Solvabilité II

C'est lors des années 1970 que la Commission Européenne, dans un contexte d'augmentation des échanges et d'accélération de la mondialisation, décide d'encadrer la solvabilité des entreprises de (ré)assurance et publie alors la directive Solvabilité I. L'application de cette directive impose alors l'immobilisation en fonds propres d'une marge de solvabilité, calculée en proportion des primes et sinistres provisionnés par l'entreprise de (ré)assurance.

Dans un mouvement de redéfinition des marges de solvabilité des banques et des entreprises de (ré)assurance, la Commission Européenne publie ensuite en 2009 la fameuse directive Solvabilité II. Cette directive prévoit une complexification de la définition de la marge de solvabilité (qui devient le capital de solvabilité requis) pour permettre la prise en compte, dans son calcul, de tous les risques inhérents aux activités de (ré)assurance.

La directive européenne a ainsi pour objectif de superviser la solvabilité des (ré)assureurs dans une volonté de protection de ses consommateurs. Ce contrôle est donné à la responsabilité de chacun des États membres de l'Union Européenne. Cette directive est transposée en droit français par l'ordonnance du 2 juin 2015. En France, c'est l'ACPR qui a la charge de veiller à l'application de la directive et de superviser la solvabilité des (ré)assureurs.

#### Champ d'application

Le champ d'application de Solvabilité II est dicté par les articles 1 à 12 de la directive 2009/138/CE. L'application de ses dispositions est imposée à toute entreprise de (ré)assurance de l'Union Européenne. On retrouve néanmoins des conditions d'exclusion de ce champ d'application reposant principalement sur la taille de l'entreprise.

## Une réglementation articulée autour de trois piliers

Solvabilité II s'articule autour de trois piliers, chaque pilier ayant pour objectif de cadrer un aspect différent de la solvabilité des entités. Les piliers sont :

- le pilier 1 relatif aux exigences quantitatives (exigences de fonds propres) établies par le calcul notamment du SCR qui représente une mesure des fonds propres nécessaires à la couverture à un niveau de probabilité de 99,5 %, des risques auxquels l'assureur est soumis ;
- le pilier 2 relatif aux exigences qualitatives de gestion de l'activité de (ré)assurance (gestion interne des risques, gouvernance) ;
- le pilier 3 relatif aux exigences de communication financière des entreprises de (ré)assurance à destination du public et du contrôleur (harmonisation européenne de la communication financière).

### 1.2.3 Quid de l'inclusion des versements libres dans l'estimation des flux de trésorerie futurs

Nous en venons maintenant aux enjeux de notre étude qui prennent appui sur la différence de définition de la frontière des contrats au sein des deux normes que nous venons de présenter.

La frontière des contrats est une notion définissant les limites de prise en compte des flux futurs des contrats pour le calcul des engagements des assureurs.

Pour la norme Solvabilité II, la limite de l'engagement de l'assureur est définie par l'article 18 du Règlement délégué 2015/35/CE :

*"Les engagements qui ne se rapportent pas aux primes déjà payées ne font pas partie d'un contrat d'assurance ou de réassurance si toutes les conditions suivantes sont remplies :*

- 1. le contrat ne prévoit pas d'indemnisation en cas de réalisation d'un événement incertain spécifié affectant la personne assurée ;*
- 2. le contrat ne prévoit pas de garantie financière des prestations ;*
- 3. l'entreprise ne peut pas contraindre le preneur à payer la prime future pour ces engagements."*

(Règlement délégué 2015/35/CE, Article 18 - paragraphe 5)

Les versements libres ne peuvent ainsi pas être pris en compte dans l'estimation des flux futurs relatifs aux contrats. Les engagements se rapportant à ces primes ne sont alors pas prévus dans le calcul du Best Estimate défini par le référentiel Solvabilité II.

Concernant la norme IFRS 17, les limites des flux à prendre en compte sont définies par

l'article 34, dont les stipulations sont précisées par les articles 33 et B61 à B71. L'article B65 détaille les types de flux entrant dans la frontière des contrats. On retrouve alors dans le paragraphe a) de cet article, la mention des primes et des flux de trésorerie supplémentaires qui en découlent. Si cette mention précise l'inclusion des primes programmées dans le périmètre des contrats, elle n'exclut pas explicitement l'inclusion des primes volontaires relatives à un contrat d'assurance vie.

Par ailleurs, l'article 33 stipule dans son paragraphe a) que l'estimation de ces flux doit :

*"intégrer, de manière impartiale, toutes les informations raisonnables et justifiables disponibles sans coût ni effort excessif concernant le montant, l'échéance et l'incertitude de ces flux de trésorerie futurs. Pour ce faire, une entité doit estimer la valeur attendue (c'est-à-dire la moyenne pondérée par les probabilités) de toute la gamme des résultats possibles."*

(IFRS 17 *Insurance Contracts*, Article 33 - paragraphe a)

La prise en compte sous IFRS 17 des versements libres dans les projections des flux futurs est par ailleurs cohérente avec les objectifs de la norme visant à fournir une vision fidèle des profits des entreprises d'assurance.

Cette position incertaine de la réglementation quant à la projection des versements libres à suscité les interrogations des assureurs. Les travaux réalisés en 2018 par un groupe de travail de l'Institut des Actuaire à ce sujet révèlent le début d'un consensus pour leur intégration. Selon le rapport rédigé par ce groupe ([S. Fallos et al. \(2018\)](#)), 3 ou 4 assureurs de la place envisageaient alors à cette époque le développement d'une loi de projection de ces comportements. En effet, la plupart des assureurs n'étaient alors pas en possession de telles lois, inutiles sous le cadre de Solvabilité II.

Cependant, il est alors nécessaire pour les assureurs de développer des modèles simples de projection des versements libres et donc aptes à être audités. Il est important de plus de fournir une modélisation robuste de ces comportements afin de limiter la volatilité des estimations de la CSM. En effet, si l'inclusion des versements libres permet d'affiner l'estimation de la CSM en offrant une vision plus fidèle de la rentabilité des contrats, un point d'attention doit être porté à la volatilité de leurs estimations afin d'éviter des écarts d'expérience importants.

## 1.3 Présentation des données et retraitements

### 1.3.1 Présentation des bases de données mises à disposition de l'étude

La fenêtre de notre étude s'étend de 2006 à 2020, soit sur 15 années. Les données mises à notre disposition se décomposent de la sorte : un portefeuille décrivant l'entièreté des contrats signés chez l'assureur ainsi que les caractéristiques des assurés associés aux contrats, 16 tables décrivant l'encours vu à la fin de chaque année par contrat et par support investi ainsi que 15 tables décrivant l'ensemble des versements enregistrés par année.

Les tableaux suivants présentent la composition de 2006 à 2020 des 3 types de tables en notre disposition ainsi qu'un résumé des valeurs prises - pour les variables quantitatives - et des modalités observées - pour les variables qualitatives.

Variable	Résumé des valeurs/modalités prises de 2006 à 2020
Numéro contrat	294 070 valeurs distinctes
Date d'effet	Min : 15/05/1985, Médiane : 04/11/2004, Max : 31/12/2020
Date de sortie	Min : 01/01/2006, Médiane : 26/07/2016, Max : 09/07/2020
Motif de sortie	Vide (67 %), Rachat (25 %), Décès (8 %)
Type de personne	Personne physique (94.5 %), Personne morale (5.5 %)
Civilité	Mr (50 %), Mme (43 %), Mr-Mme (1.5 %), Autre (5.5 %)
Date de naissance	Min : 01/01/1900, Médiane : 06/12/1955, Max : 22/06/2020
Sexe	Homme (50%), Femme (43%), Personne morale (5.5%), Indivision (1.5%)
Code postal	7599 valeurs distinctes (95% d'assurés résidents en France métropolitaine et 5% dans les DOM-TOM ou à l'étranger)
Produit	1115 produits différents
Partenaire	406 partenaires différents

TABLE 1.1 – Description de la table du portefeuille de contrats

Nous disposons ensuite de 16 tables d'encours, décrivant les encours au 31/12 de chaque contrat de notre portefeuille, et cela, des années 2005 à 2020. Le portefeuille est ainsi associé à un volume d'encours de 24 milliards d'euros en moyenne par année de 2006 à 2020, réparti sur un nombre moyen de 178 744 contrats. Si le volume d'encours a progressivement augmenté depuis 2005, en passant de 15 milliards d'euros à plus de 30 milliards en 2020, le nombre de contrats en portefeuille a lui connu de plus fortes variations. En effet, nous

observons une forte croissance de la population assurée jusqu'en 2015, qui atteint son pic de 191 618 contrats, suivie d'un déclin jusqu'en 2020 où elle atteint son minimum sur 15 ans de 161 366 contrats.

Variable	Résumé des valeurs/modalités prises de 2006 à 2020
Numéro contrat	292 564 valeurs distinctes
Date	31/12 de chaque année
Nature support	Euro (65 %), UC (35 %)
Montant	Min : -129 986 €, Médiane : 28 602 €, Max : 70 848 975 €

TABLE 1.2 – Description des tables d'encours

Enfin, 15 tables de versements ont été mises à notre disposition, décrivant pour les années 2006 à 2020 les types de versements dont chaque contrat a pu bénéficier ainsi que leurs montants. En 15 ans d'observations, les assurés effectuant des versements libres représentent en moyenne 4.5 % du portefeuille pour une récolte moyenne de 650 millions d'euros par an.

Variable	Résumé des valeurs/modalités prises de 2006 à 2020
Numéro contrat	294 070 valeurs distinctes
Code opération	Versement programmé (86%), Versement libre (7%), Versement initial (7%)
Date d'effet	Min : 01/01/2006, Médiane : 03/01/2013, Max : 31/12/2020
Nature support	Euro (66%), UC (34%)
Montant net	Min : 2.5€, Médiane = 146€, Max : 50 000 000€
Taux de chargement	Min : 0%, Médiane : 2.5 %, Max : 3%

TABLE 1.3 – Description des tables de versement

Afin de pouvoir travailler à partir d'une unique base de données, les tables présentées ont été réconciliées puis retraitées selon la démarche expliquée dans la partie suivante.

### 1.3.2 Retraitements effectués

#### Gestion des données manquantes

La création de ces bases de données par an a été compliquée par le manque de certaines données dans le portefeuille. En effet, certains contrats ne bénéficiaient pas de Date de

Sortie alors qu'un Motif de Sortie apparaissait dans le portefeuille. Afin d'extraire les contrats en cours pour une année N, nous avons premièrement filtré la table de portefeuille en conservant les contrats dont les dates d'effet et de sortie correspondaient à un contrat en cours lors de l'année N, ainsi que les contrats débutants avant l'année N et n'ayant pas de date de sortie. Nous avons ensuite effectué une jointure entre ce portefeuille filtré et les tables d'encours de l'année en question et de la précédente, afin d'associer à chaque contrat son encours de début et de fin d'année.

Finalement, nous joignons à la table précédemment créée, la table de versements de l'année N. Certains contrats du portefeuille affichent une durée d'effet inférieure à 1 an, ces contrats ne sont ainsi pas présents dans les tables d'encours, mais apparaissent parfois dans les tables de versements. Par conséquent, et afin de conserver l'entièreté des versements, les contrats sur lesquels n'apparaissent aucun mouvement et n'ayant ni d'encours en début ou fin d'année sont supprimés de la table. La démarche de construction des bases de données par an est résumée dans le graphique 1.4.

Nous retirons finalement les contrats "oubliés" avec des assurés de plus de 100 ans et aucun mouvement sur leurs contrats.

## **Périmètre d'étude**

L'objet de cette étude étant d'étudier le comportement individuel des assurés, nous restreignons le périmètre de l'étude à des contrats d'épargne individuelle. Le portefeuille mis à notre disposition présentant des contrats d'épargne collective ou encore de retraite, nous éliminons ces contrats. En effet, il est intuitivement possible d'observer des comportements de versements différents sur ces divers types de contrats en raison d'une part de la clientèle visée sur ces types de produits, mais aussi de par les conditions fiscales différentes qui leur sont associées.

## **Pas d'observation**

Nous choisissons d'étudier le comportement annuel des assurés de ce portefeuille. Ainsi, un assuré ayant par exemple effectué 2 versements libres de 1000 € et de 1500 € lors de l'année 2017, est identifié comme ayant effectué un versement libre de 2500 €. Ce choix est motivé par le manque d'observations d'autres "mouvements" (rachats, arbitrages) sur les contrats que les versements effectués. En effet, nous ne connaissons l'encours des contrats qu'à un pas annuel.

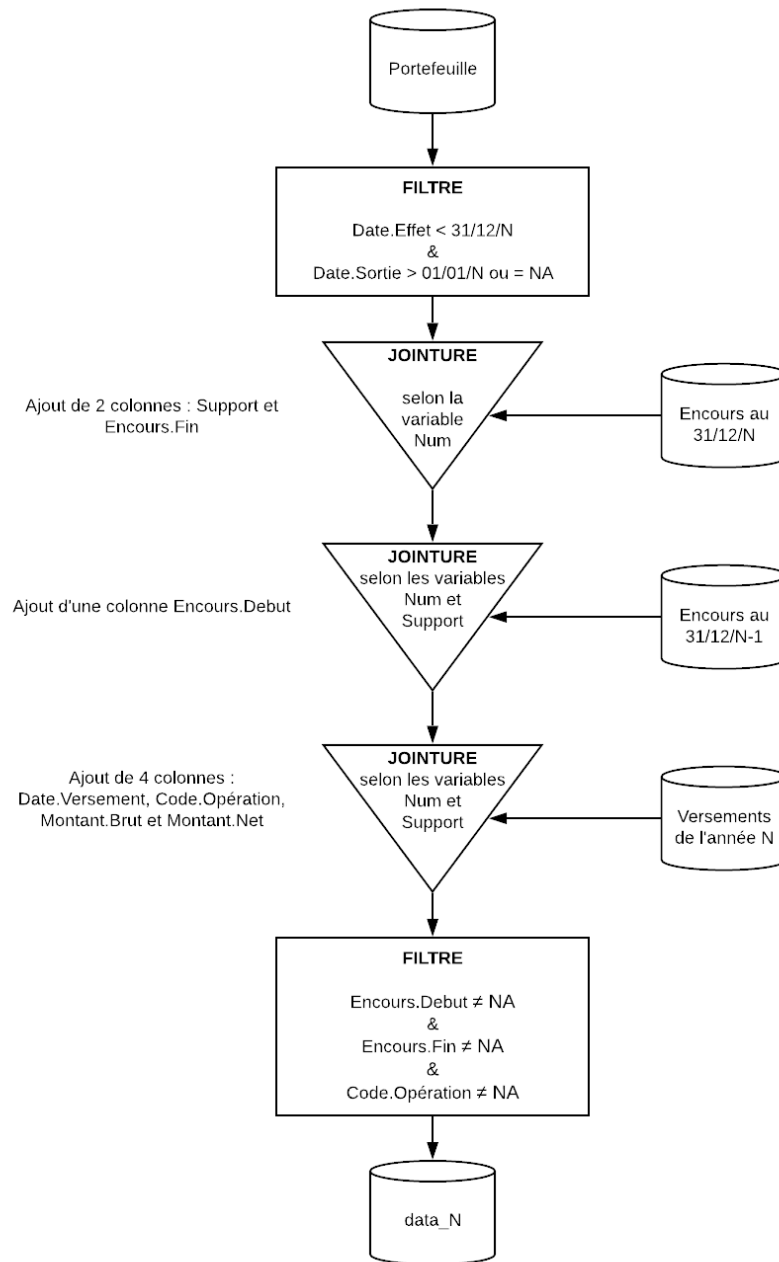


FIGURE 1.4 – Démarche de construction de la base de données utilisée

### Constitution des bases de données pour l'étude et suppression des valeurs aberrantes

Sur la base des retraitements effectués, nous créons deux bases de données : une pour l'étude de la décision de versement libre et une autre pour l'étude de leurs montants.

La distinction de ces 2 bases est motivée par le fait qu'un assuré ayant effectué un versement libre d'un montant réparti sur l'euro et l'UC constitue, par construction de la



base d'origine, deux lignes. Pour l'étude de la décision de versement libre, nous agrégeons les deux lignes du précédent exemple pour n'obtenir qu'une ligne par assuré et par an. Pour l'étude des montants, nous conservons les deux lignes du précédent exemple afin de conserver l'information de la répartition du montant sur les différents supports et nous retirons simplement les lignes relatives aux assurés n'ayant pas effectué de versement libre.

Nous remarquons une très forte hétérogénéité des montants de versements libres dans la base ainsi créée. En effet, nous observons des montants très faibles avec un minimum de 0,45 € ainsi que des montants extrêmement élevés avec un maximum de 35 millions d'euros versés. Cette hétérogénéité des montants, marquée par les montants extrêmes observés, est visualisable sur le graphique suivant qui en décrit les déciles :

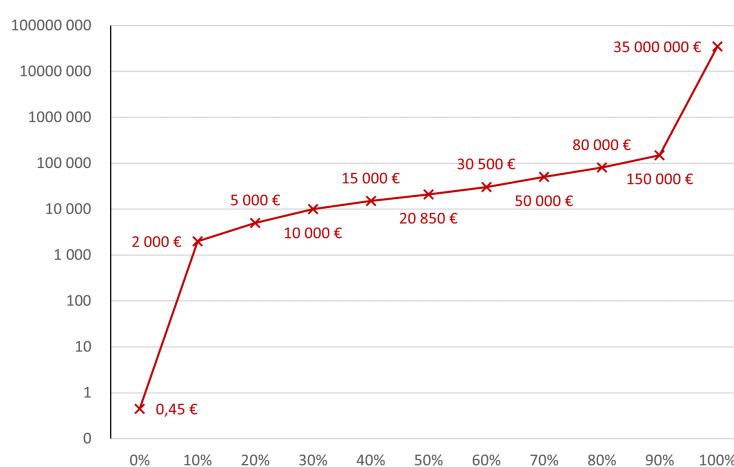


FIGURE 1.5 – Déciles des montants de versements libres observés (échelle logarithmique)

La forte volatilité de ces montants en queue de distribution impacte leur moyenne sur les 15 ans d'observation en la faisant augmenter significativement vers les hautes valeurs extrêmes à une valeur de plus de 750 000 €. Nous identifions alors le risque, porté par ces montants extrêmes, d'induire un biais important dans leur modélisation et de complexifier le calibrage des modèles, nous choisissons alors d'écarter de l'étude les 0,5 % des montants les plus élevés ainsi que les plus faibles pour ainsi se ramener à une moyenne de montant de 80 000 €.

Ce choix est néanmoins lourd de conséquences sur l'étude, mais s'inscrit dans une démarche d'établir une loi "prudente" des versements libres. Nous pouvons cependant argumenter ce choix par le caractère aberrant que revêt l'observation de montants si faibles et alors penser qu'ils sont la cause d'erreurs de saisie. La suppression des très hauts montants est quant à elle plus problématique dans le sens où ces montants représentent en moyenne près de 10 % du chiffre d'affaires annuel apporté par les versements libres. L'objectif de ce mémoire étant de fournir une étude générale des comportements de versements libres, nous choisissons de nous concentrer sur l'étude des versements apportant 90 % de ce chiffre d'affaires. L'étude de ces montants extrêmes ne sera ainsi pas abordée dans ce mémoire

et constitue alors une première ouverture. Nous pouvons imaginer la création d'une "loi des extrêmes" pour la projection des versements libres, à l'image de ce qui est fait dans d'autres domaines de l'assurance pour la modélisation de sinistres.

### 1.3.3 Composition du portefeuille étudié

Le portefeuille ainsi retraité conserve les mêmes tendances que celles détaillées précédemment.

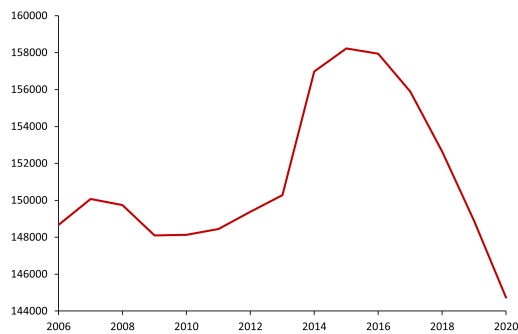


FIGURE 1.6 – Évolution du nombre de contrats en portefeuille de 2006 à 2020

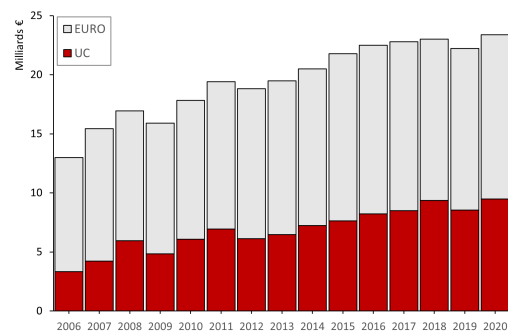


FIGURE 1.7 – Évolution des encours au 1er janvier de 2006 à 2020

Ainsi, la population assurée observe une croissance jusqu'en 2015, atteignant son pic de 158 223 contrats, suivie d'un déclin significatif les cinq années suivantes. De même, le volume d'encours conserve son augmentation progressive en 15 ans, en passant de 13 milliards d'euros à plus de 23 milliards.

Le portefeuille étudié se compose alors d'une moyenne de 151 208 contrats par an, conservant ainsi 85 % des contrats initialement en portefeuille, associés à un encours moyen de 19.5 milliards d'euros, le retraitement imposant ainsi une perte de moins de 20 % du volume d'encours moyen initial. Le pourcentage d'encours investi sur des unités de compte croît progressivement depuis 2006 et passe de 26 % à 41 % au 1er janvier 2020.

Le tableau suivant détaille les profils moyens des assurés du portefeuille pour chaque année observée. Le profil type, sur 15 ans, d'un assuré de ce portefeuille est celui d'un homme de 60 ans habitant à Paris, possédant un encours de 50 000€ dont 31 % investi sur des unités de compte et ayant une ancienneté de contrat de 10 ans.

Année	Age moyen	Ancienneté moyenne	Encours moyen	Pourcentage d'encours investi sur des UC
2006	57	6	70 422 €	26 %
2007	57	7	79 416 €	27 %
2008	58	7	85 875 €	35 %
2009	59	8	81 998 €	30 %
2010	59	9	91 889 €	34 %
2011	59	9	100 205 €	36 %
2012	59	10	97 460 €	33 %
2013	60	10	100 657 €	33 %
2014	60	10	100 806 €	35 %
2015	60	10	104 237 €	35 %
2016	61	11	106 600 €	37 %
2017	61	12	108 010 €	37 %
2018	62	12	110 192 €	41 %
2019	62	13	107 703 €	38 %
2020	63	13	115 818 €	41 %

TABLE 1.4 – Profil moyen des assurés du portefeuille par an

### Zoom sur les produits associés aux contrats en portefeuille

Les contrats en portefeuille sont associés à 790 produits différents regroupés en fonction des caractéristiques des supports offerts. On observe ainsi 3 groupes de produits : les produits monosupport euro, ainsi que les produits multisupports avec ou sans option d'orientation personnalisée. Cette option d'orientation personnalisée offerte pour certains produits multisupport permet à l'assuré de déléguer la gestion de son épargne aux soins d'un expert des marchés financiers.

Les produits monosupport euro, n'étant plus disponibles à la vente en raison de la baisse des taux obligataires, réunissent une minorité des contrats en portefeuille : 10 % en moyenne sur les 15 années observées. On observe alors logiquement une décroissance progressive du nombre de contrats associés à ces produits : ils réunissaient en 2006 un peu plus de 26 000 contrats et plus que la moitié de ce nombre en 2020. Le volume d'encours associé à ces produits observe en conséquence la même décroissance. Tandis que les produits monosupport euro regroupaient environ 1 milliard d'euros d'encours en 2006 (soit 9 % du total d'encours sur cette année), on observe une perte du tiers de ce volume d'encours sur 15 ans.

Les produits multisupports sans option d'orientation personnalisée réunissent quant à eux la majorité des contrats en portefeuille, avec une proportion moyenne de 75 % sur les 15

années observées. Si le nombre de contrats associés à ce type de produits décroît légèrement et affiche une baisse de 10 % en 15 ans, son niveau d'encours est quant à lui stable sur la période d'observation avec une moyenne de 10 milliards d'euros.

Les produits multisupports avec option d'orientation personnalisée, bien qu'encore minoritaires en 2006, représentent en 2020, 17 % des contrats en portefeuille. L'encours totalisé sur ces produits affiche une croissance fulgurante de près de 10 milliards d'euros en 15 ans pour atterrir à un montant de 13 milliards d'euros en 2020. Les contrats associés à ce type de produits portent ainsi la hausse du volume d'encours observée sur la totalité du portefeuille. Les contrats associés à ces produits appartiennent à une clientèle fortunée avec une moyenne de 250 000 € épargnés par contrat, soit près de quatre fois plus que les montants moyens observés pour les deux autres types de produits.

L'étude des conditions générales des cinq produits rassemblant le plus de contrats en portefeuille et les plus gros volumes d'encours soulève les points suivants :

- certains produits visent une clientèle aisée avec un montant minimum de versement à la souscription allant jusqu'à 600 000 €, les autres sont destinés à une clientèle plus large avec des versements initiaux imposés à 5 000 € au minimum ;
- les frais de gestion sont homogènes parmi les produits avec un taux moyen de 0,8 % pour le fond euro et de 0,95 % pour les unités de compte ;
- chacun des produits affiche un taux maximal de chargement sur versement libre compris entre 4 et 4,5 %. En pratique, nous observons une moyenne de 1,2 % de prélèvements sur les versements libres du portefeuille. On observe par ailleurs une décroissance à partir de 2006 des taux de chargements appliqués aux versements libres, comme l'illustre le graphique suivant.

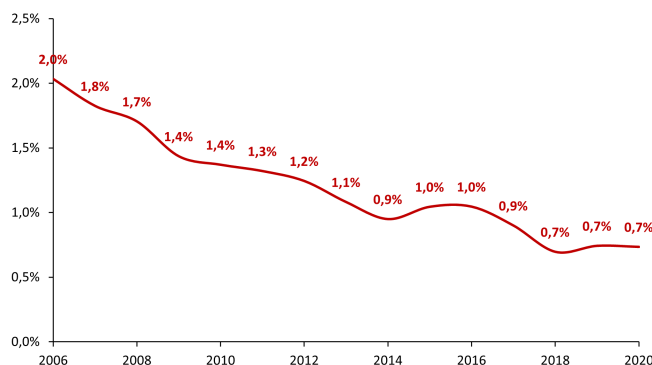


FIGURE 1.8 – Évolution des taux de chargements moyens observés sur les montants de versements libres versés

### **Zoom sur les partenaires commerciaux associés aux contrats en portefeuille**

Les contrats en portefeuille sont associés à 384 partenaires différents que nous classifions selon la segmentation utilisée par l'assureur :

- Réseau propriétaire : 3 partenaires associés à 3 % des contrats en portefeuille ;
- Banques de réseau : 29 partenaires associés à 34 % des contrats en portefeuille ;
- Courtiers : 121 partenaires associés à 25 % des contrats en portefeuille ;
- Gestionnaires de fortune : 54 partenaires associés à 31 % des contrats en portefeuille ;
- Sociétés de Gestion : 75 partenaires associés à 5 % des contrats en portefeuille ;
- Autres : 102 partenaires associés à 2 % des contrats en portefeuille.

## Chapitre 2

# Analyse et segmentation des comportements de versements libres

Dans ce chapitre, nous exposons les résultats de nos analyses des comportements de versements libres observés sur le portefeuille. Les variables étudiées dans cette partie sont les suivantes :

- la proportion d'assurés effectuant un versement libre dans l'année ;
- le montant moyen des versements libres observés dans l'année.

Par souci de simplification de notre rédaction, nous désignons par "verseurs" les assurés effectuant un versement libre dans l'année et par "non-verseurs" ceux n'en effectuant pas. Nous étudions ainsi l'influence de chacune des variables mises à notre disposition, de manière parallèle, sur l'évolution de la proportion de verseurs ainsi que sur l'évolution des montants moyens de versements libres.

## Proportion de verseurs

D'un point de vue global, sans regarder les comportements individuels de chacun, il est notable de constater les fortes variations en 15 ans des proportions de verseurs illustrées par les baisses significatives en 2008 ou 2012, années tourmentées par les crises des subprimes et des dettes souveraines. La composition du portefeuille n'évoluant pas de façon significative durant ces années, on peut à première vue supposer que les caractéristiques des assurés ne seront pas essentielles à l'explication de la décision d'effectuer un versement libre.

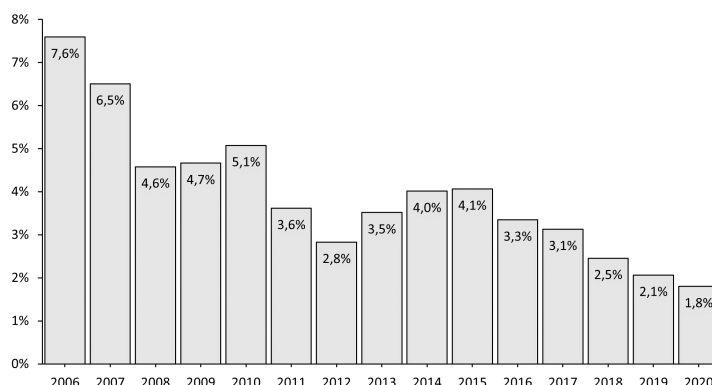


FIGURE 2.1 – Évolution annuelle de la proportion de verseurs observée

## Montant moyen des versements libres

Le montant moyen de versement libre perçu par an connaît lui des évolutions moins nettes. Le montant moyen des versements libres observés sur 15 ans a progressivement augmenté, en doublant depuis 2006, pour se stabiliser au montant de 90 000 € à partir de 2012. Les montants moyens de versements libres observent une évolution en 15 ans très similaire à celle des encours observés sur le portefeuille, et nous laissent ainsi à penser à une forte influence de l'encours du contrat sur les montants de versements libres.

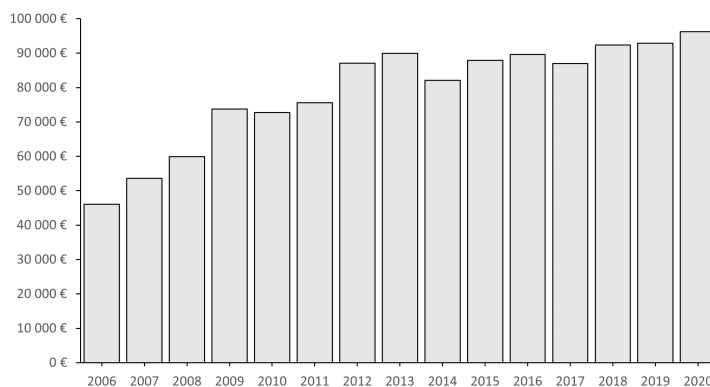


FIGURE 2.2 – Évolution annuelle des montants moyens de versements libres observés

## 2.1 Analyse des caractéristiques des assurés en portefeuille et de leur impact sur les comportements de versements libres

Dans cette première sous-partie, nous nous intéressons à l'étude des variables structurales du portefeuille : les caractéristiques des contrats observés.

### 2.1.1 Variable Age

Nous nous sommes questionnés ici sur l'influence de l'âge de l'assuré sur sa décision d'effectuer un versement libre.

À première vue, nous n'observons pas de différence d'âge significative entre les assurés effectuant un versement libre dans l'année et ceux n'en effectuant pas. En effet, nous observons chaque année une différence d'âge moyen de 3 ans entre ces deux profils, avec une moyenne d'âge sur 15 ans de 57 ans chez les verseurs et de 60 ans chez les non-verseurs.

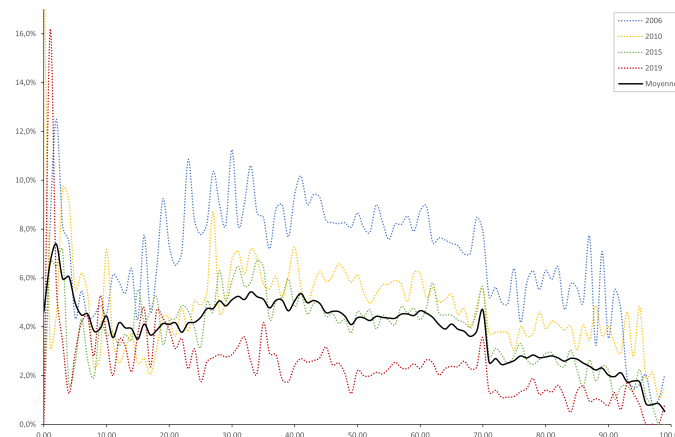


FIGURE 2.3 – *Évolution de la proportion de verseurs observée en fonction de l'âge*

Cependant, l'âge de l'assuré semble avoir un grand impact sur sa probabilité d'effectuer un versement libre. En réalité, l'observation de l'évolution des taux de versements libres en fonction des âges, illustrée par le graphique ci-dessus, permet d'identifier 4 "plateaux" :

- un premier entre 0 et 5 ans, où on observe un pic des taux de versements. Cette catégorie d'âge est celle présentant le plus haut taux de versements libres moyen (7 %), et cela, pour toutes les 15 années d'observations ;
- un deuxième entre 5 et 25 ans, associé à un taux de versement moyen de 4 % ;
- un troisième entre 25 et 70 ans, correspondant à la tranche active de la population et affichant un taux moyen de versements libres de 5 % ;



- un quatrième correspondant aux âges au-delà de 70 ans, période de retraite des assurés, qui affiche le taux le plus bas de versements libres (3 %);

Les assurés ayant un âge compris entre 25 et 70 ans représentent la majorité des assurés en portefeuille (plus de 75 %), ainsi la probabilité d'effectuer un versement libre quasiment constante sur cette tranche d'âge explique la différence d'âge non significative que l'on observe entre les profils de verseurs et de non-verseurs. Nous retrouvons par ailleurs sur ce graphique, la forte variabilité temporelle de la décision d'effectuer un versement libre observée en introduction à cette partie.

Le montant moyen des versements libres connaît quant à lui une nette croissance avec l'âge de l'assuré, et ce, jusqu'à l'âge de 70 ans. Ainsi, si les assurés ayant un âge compris entre 0 et 5 ans sont ceux qui effectuent le plus de versements libres sur leurs contrats, ils sont aussi ceux qui versent les montants les plus faibles. À partir de l'âge de 70 ans, on observe une stabilisation des montants moyens versés sur les contrats associés.

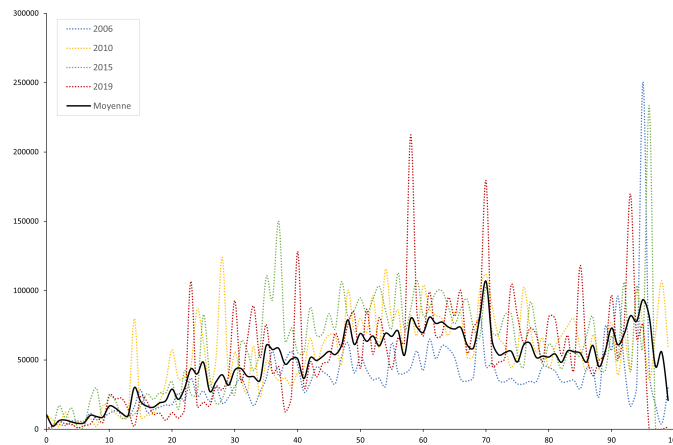


FIGURE 2.4 – Évolution des montants moyens de versement libre en fonction de l'âge

### 2.1.2 Variable Ancienneté

L'ancienneté semble avoir une influence encore plus forte sur la probabilité d'effectuer un versement libre. En effet, nous observons une différence significative de cinq ans en moyenne sur l'ancienneté fiscale des contrats verseurs et non-verseurs : les premiers étant associés à une ancienneté moyenne de 5 ans, tandis que les autres sont associés à une ancienneté moyenne de 10 ans.

L'étude du graphique représentant l'évolution de la proportion de verseurs en fonction de l'ancienneté du contrat confirme ce constat. Nous observons ainsi une décroissance quasi-linéaire de la proportion de verseurs avec l'ancienneté. Nous constatons, par ailleurs, que les assurés sont plus enclins à effectuer un versement libre lors de la deuxième année de leur contrat.

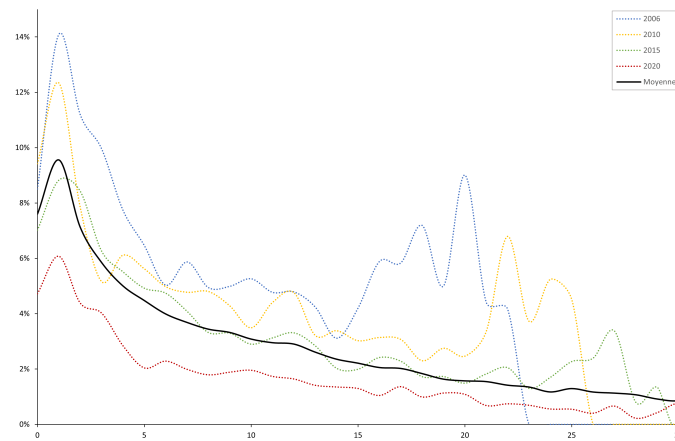


FIGURE 2.5 – Évolution de la proportion de verseurs observée fonction de l'ancienneté des contrats

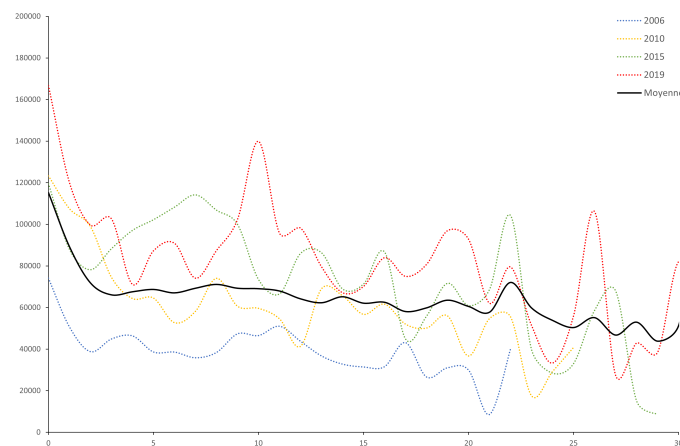


FIGURE 2.6 – Évolution des montants moyens de versement libre en fonction de l'ancienneté des contrats

Concernant le montant moyen associé aux versements libres observés, l'ancienneté du contrat ne semble pas avoir de réelle influence. En effet, nous ne distinguons pas de tendance d'évolution de ce montant avec l'ancienneté.

### 2.1.3 Variable Encours

Nous nous intéressons ici à l'influence du montant d'encours investi par l'assuré sur son comportement de versement libre. Si les verseurs présentent en moyenne le même encours que les non-verseurs, nous souhaitons cependant observer les différences de comportement entre les assurés possédant les plus faibles encours et ceux possédant les plus élevés. Pour cela, nous étudions les encours associés aux contrats en portefeuille, que nous découpons selon leurs quartiles afin de constituer quatre classes équilibrées.

Nous observons ainsi que les assurés possédant les plus faibles encours sont ceux qui

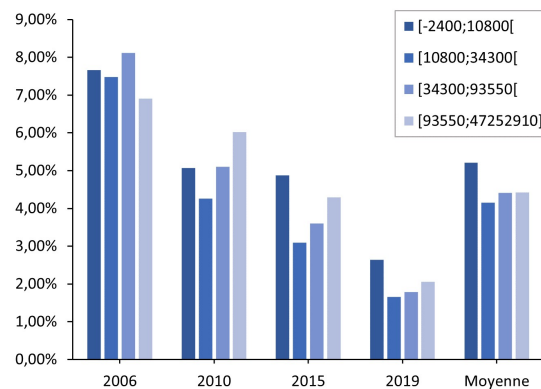


FIGURE 2.7 – Évolution de la proportion de verseurs en fonction des niveaux d'encours

représentent, en moyenne, le groupe pour lequel on enregistre le plus de versements libres. Pour les autres classes, la proportion de verseurs croît avec le volume d'encours associé à la classe. Si ces tendances observées parmi les classes d'encours ainsi construites sont vraies en moyenne, elles dépendent cependant de l'année d'observation.

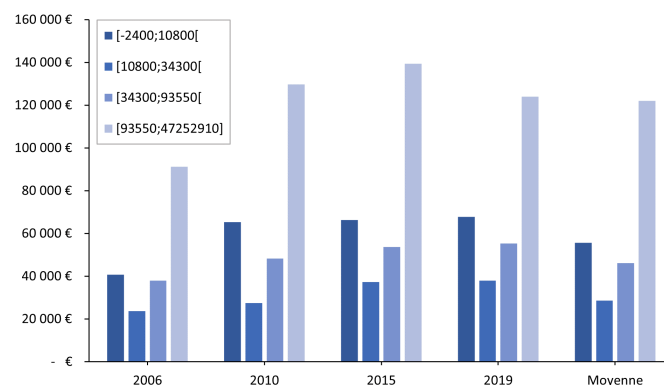


FIGURE 2.8 – Évolution des montants moyens de versement libre en fonction des niveaux d'encours

Concernant les montants moyens associés à ces versements libres, le niveau d'encours semble y apporter une nette influence. Si le montant moyen versé par les assurés associés à la première classe d'encours reste quasiment constant dans le temps à 8000 €, les montants moyens associés aux autres classes augmentent de façon exponentielle avec le niveau d'encours. Les assurés associés au quart des encours les plus élevés versent en moyenne plus de 120 000 € dans l'année sur leurs contrats d'épargne, soit deux fois plus que ceux associés aux encours les plus faibles et quatre fois plus que les assurés associés à la 2e classe d'encours. Cette tendance reste la même, quelle que soit l'année observée.

### 2.1.4 Variable Produit

Nous nous intéressons ici à l'influence des produits, regroupés en fonction des caractéristiques présentées dans la partie précédente des supports offerts, sur le comportement de versements libres des assurés.

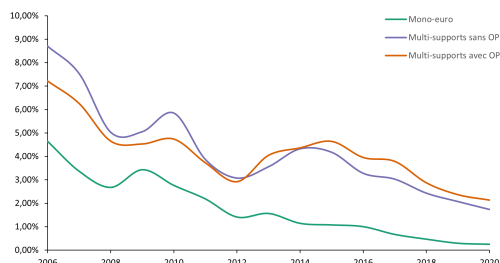


FIGURE 2.9 – Évolution de la proportion de verseurs observée en fonction du type de produit

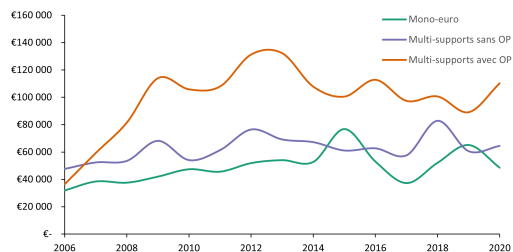


FIGURE 2.10 – Évolution des montants moyens de versements libres en fonction du type de produit

Nous remarquons une proportion de verseurs semblable entre les produits multisupports. Les produits mono-euro, eux, affichent une proportion de verseurs plus faible, en moyenne deux fois moins élevée que pour les autres produits.

Les produits multisupports avec OP sont associés aux montants moyens de versements libres les plus élevés, avec une moyenne de 110 000 € contre 55 000 € pour les autres produits. On remarque une similitude des montants moyens associés à ces produits avec les montants moyens, présentés précédemment, versés par les assurés ayant les plus hauts encours. Parallèlement, les montants moyens versés sur les produits multisupports sans OP se rapprochent des montants moyens versés par les assurés ayant les plus faibles encours.

### 2.1.5 Variable Partenaire

Nous nous intéressons ici à l'influence des partenaires, regroupés selon la segmentation utilisée par l'assureur, sur le comportement de versements libres des assurés.

Le type de partenaire associé au contrat semble avoir une plus grande influence sur la proportion de verseurs que pour le montant moyen associé à ces versements libres. Exception faite des contrats associés aux gestionnaires de fortunes, dont les encours sont les plus élevés, qui rassemblent les montants de versements les plus élevés en moyenne comme observé précédemment.

Les contrats associés au réseau de l'assureur sont ceux présentant les plus fortes proportions de verseurs en 15 ans. Si ce réseau de partenaires est parvenu à réunir de plus en plus de versements libres dans le temps, on observe néanmoins une baisse en parallèle des montants moyens versés. À l'inverse, les gestionnaires de fortunes ont vu la proportion de verseurs

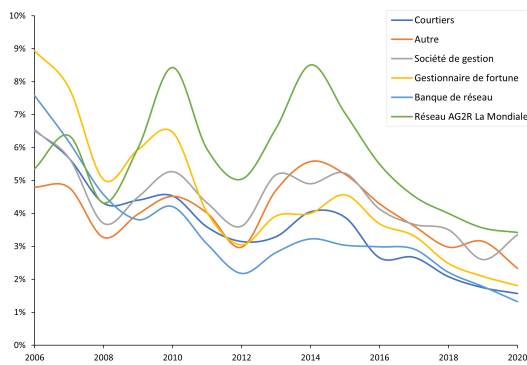


FIGURE 2.11 – Évolution de la proportion de verseurs observée en fonction du type de partenaire

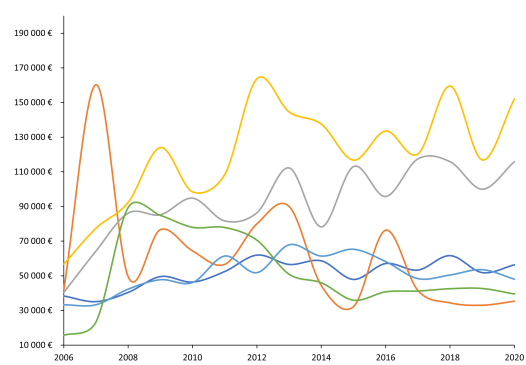


FIGURE 2.12 – Évolution des montants moyens de versements libres en fonction du type de partenaire

sur leur périmètre baisser avec le temps : alors qu'ils étaient en tête du classement des partenaires en 2006, ils redescendent dès 2012 et se positionnent à la moitié du classement. Parallèlement, ils observent une augmentation du montant moyen versé.

### 2.1.6 Analyses complémentaires

Nous n'observons pas de différences significatives des taux de verseurs parmi les hommes, les femmes ou les co-souscriptions. Cependant, le sexe de l'assuré affiche un impact sur le montant moyen versé :

- Les contrats associés à des co-souscriptions enregistrent des montants moyens de versements libres plus élevés, ce qui est logique. Cette tendance tend néanmoins à s'effacer avec le temps.
- Les hommes versent en moyenne 1,5 fois plus d'argent sur leurs contrats.

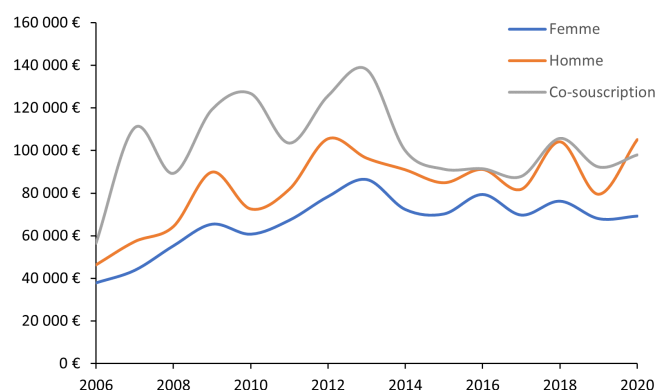


FIGURE 2.13 – Évolution des montants moyens versés en fonction du sexe de l'assuré

Les graphiques suivants illustrent les évolutions des taux de versement ainsi que des montants moyens versés parmi les assurés effectuant des versements programmés ou non, en

vert sont représentés ces évolutions pour les assurés effectuant des versements programmés, en rose les autres.



FIGURE 2.14 – *Évolution des proportions de verseurs en fonction de la présence de versements programmés*

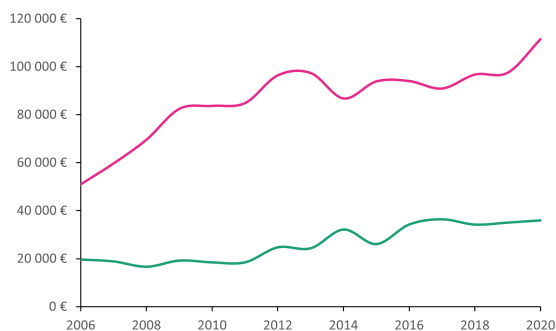


FIGURE 2.15 – *Évolution des montants moyens versés en fonction de la présence de versements programmés*

Les assurés effectuant des versements programmés dans l'année affichent une proportion de verseurs plus élevée que ceux n'en effectuant pas. Les assurés programmant la constitution de leur épargne sont ainsi plus enclins à effectuer un versement libre que les autres. Cet écart dans la proportion de verseurs parmi les deux profils d'assurés se réduit néanmoins dans le temps. Parallèlement, nous remarquons que si le fait d'effectuer des versements programmés semble impacter la décision de versement libre, il impacte encore plus leurs montants moyens. En effet, nous observons des montants moyens de versements libres en moyenne trois fois plus élevés chez les assurés n'effectuant pas de versements programmés.

### 2.1.7 Conclusions

Ces premières analyses nous ont permis d'obtenir un premier aperçu des variables les plus discriminantes aux comportements de versements libres.

Ainsi, nous avons pu observer que la décision d'effectuer un versement libre est majoritairement guidée par l'ancienneté du contrat, mais aussi par l'âge de l'assuré, tandis que le montant moyen du versement se rapporte plus à l'encours investi par l'assuré. Nous vérifions ces conclusions en complétant ces premières analyses univariées par une étude multivariée de notre portefeuille.

## 2.2 Segmentation des comportements de versements libres

Dans cette partie, nous exposons les résultats de l'étude multivariée réalisée sur le portefeuille. En effet, si nous avons pu précédemment identifier l'impact de chacune des caractéristiques des assurés sur leurs comportements de versements libres, ces liens sont peut-être différents en considération d'une autre variable. Nous souhaitons par ailleurs identifier les variables les plus discriminantes par rapport aux comportements assurés. Cette segmentation permettra alors, en complétant les premières analyses réalisées, de mieux comprendre les comportements assurés, mais aussi d'identifier les déclencheurs essentiels de versements libres.

Nous choisissons, pour ces fins, l'utilisation d'arbres de classification (CART) pour leur simplicité d'interprétation graphique. En effet, la représentation de la segmentation de notre portefeuille, établie par la méthode CART, sera illustrée à l'aide d'arbres nous renseignant sur les variables identifiées comme étant les plus discriminantes.

### 2.2.1 Arbres de régression (CART)

L'acronyme CART signifie Classification And Regression Trees. Il désigne une méthode non paramétrique de régression et de classification, développée par [Leo Breiman \(1984\)](#). Basée sur un algorithme à la fois itératif et récursif, la méthode CART permet la construction d'arbres de prédiction aussi simples que puissants.

Dans la pratique, l'algorithme CART procède en 3 étapes afin de déterminer l'arbre optimal de classification :

1. *Construction de l'arbre maximal :*

Au moyen de divisions binaires récursives, une suite de partitions de  $\chi$  est construite jusqu'à ce que chacune des partitions les plus fines construites ne corresponde qu'à un nombre minimal d'observations  $Y$  (choisi arbitrairement par l'utilisateur) ;

2. *Élagage de l'arbre maximal :*

De la suite d'arbres précédemment construite, on extrait une sous-suite à partir d'un critère permettant de faire un compromis entre la taille de l'arbre et l'information retenue ;

3. *Sélection finale :*

Cette dernière étape sélectionne le meilleur arbre dans la sous-suite construite.

### Construction de l'arbre de classification maximal

Considérons un échantillon de données  $D = (y_i, x_i^1, \dots, x_i^p)_{1 \leq i \leq N}$  constitué de  $N$  observations d'une variable à expliquer  $Y \in \mathcal{Y} = \{1, 2, \dots, K\}$  et de  $p$  variables explicatives  $X^j \in \chi_j$ . On note  $\chi = \prod_{j=1}^p \chi_j$ .

Le principe de construction de l'arbre de classification est de diviser  $\chi$  en  $q$  classes (où  $q$  n'est à priori pas connu à l'avance). Une suite de partitions de  $\chi$  est alors construite en appliquant une *règle de division binaire*. Chaque partition de  $\chi$  constitue alors un *nœud* de l'arbre de classification et les  $q$  partitions de  $\chi$  obtenues à la dernière étape de la méthode sont appelées *nœuds terminaux* ou *feuilles*. Nous notons  $\tilde{T}$  l'ensemble des feuilles de l'arbre et  $T^t$  l'ensemble des *descendants* de l'ancêtre  $t$ .

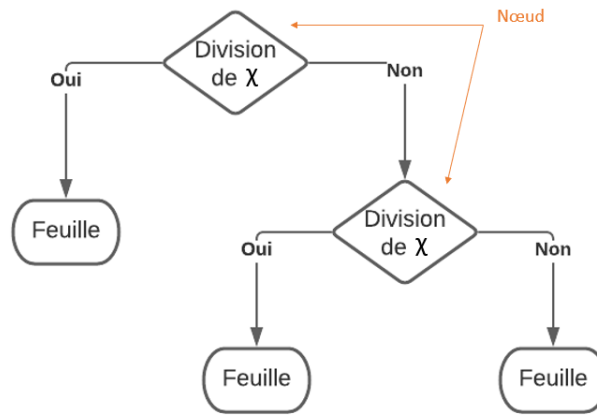


FIGURE 2.16 – Exemple d'arbre de classification

Le critère de division d'une partition de  $\chi$  en deux, repose sur une fonction d'*hétérogénéité*. L'objectif est de segmenter notre base de données en sous-groupes homogènes au sens de la variable à expliquer  $Y$ .

Nous notons  $g$  la fonction d'hétérogénéité. Cette fonction peut être définie de plusieurs façons, en voici quelques exemples :

- Dans le cas d'une variable  $Y$  qualitative, elle peut être définie à partir de la notion d'*entropie* ou à partir de la *concentration de Gini* ;
- Dans le cas d'une variable  $Y$  quantitative, elle peut être définie simplement comme la variance des observations  $y_i$  au sein de la partition.

Cette fonction  $g$  se doit d'être positive, nulle si les observations  $y_i$  au sein de la partition sont homogènes et maximales si elles sont très dispersées.

Dans le cas de la division d'un nœud  $t$  en deux nœuds  $t_1$  et  $t_2$ , l'algorithme retiendra la variable explicative ainsi que son critère de segmentation maximisant le plus le "gain"



en homogénéité induit par la création des deux nœuds. Cela revient à résoudre, à chaque étape de la construction de l'arbre, l'équation suivante :

$$\max_{\chi_i, 1 \leq i \leq p} g(t) - (g(t_1) + g(t_2))$$

Cette règle de division a néanmoins le désavantage de favoriser la sélection de variables explicatives avec beaucoup de modalités. En effet, ces variables offrent plus de souplesse dans la construction des sous groupes et risquent alors de favoriser un sur-apprentissage, il est alors préférable de réduire le nombre de modalités auparavant en les fusionnant.

Même si la méthode CART n'impose pas de règles d'arrêt de la construction de l'arbre, d'où la qualification de "maximal", il peut être nécessaire d'en imposer quelques-unes afin de limiter la complexité de l'algorithme. Les divisions s'arrêtent naturellement lorsque plus aucun "gain" en homogénéité n'est rendu possible par la création de sous-groupes, ce qui est le cas par exemple lorsque les observations des variables explicatives sont identiques dans une classe donnée. Pour autant, afin d'éviter un découpage inutilement trop fin, il est possible de définir arbitrairement un seuil minimal de "gain" en homogénéité ou un nombre minimal d'observations  $y_i$  qu'une feuille de l'arbre doit contenir.

Aucune règle d'arrêt supplémentaire des divisions ne doit être imposée à ce stade, car l'objectif est d'obtenir l'arbre le plus ramifié, afin de l'élaguer par la suite par le biais d'une procédure détaillée dans la prochaine partie de ce mémoire.

Enfin, il est nécessaire d'attribuer à chaque feuille  $t$  de notre arbre maximal, une ou plusieurs valeurs de  $Y$ , notée  $k(t)$ . Pour cela, nous introduisons les notations suivantes, pour tout nœud  $t$  et pour tout  $k \in \mathcal{Y}$  :

- $N(t) = \text{card}\{(y_i, x_i^1, \dots, x_i^p) \in D \mid (x_i^1, \dots, x_i^p) \in t\}$
- $N_k(t) = \text{card}\{(y_i, x_i^1, \dots, x_i^p) \in D \mid y_i = k, (x_i^1, \dots, x_i^p) \in t\}$
- $p(k, t) = \frac{N_k(t)}{N(t)}$ , la probabilité a posteriori d'appartenir à la classe.  $k$

Cette affectation doit si possible classer au mieux les données et avoir un pouvoir prédictif intéressant. Pour cela, elle est déterminée de façon à minimiser l'estimation de mauvaise classification  $r(t) = \sum_{k \neq k(t)} p(k, t) = 1 - \max_{k \in \mathcal{Y}} p(k, t)$  au sein d'une feuille de l'arbre.

La définition de  $k(t)$  est alors logiquement  $k(t) = k \in \mathcal{Y} \mid p(k, t)$

## Élagage de l'arbre maximal et sélection de l'arbre optimal

La démarche de construction précédente nous a permis de construire une suite d'arbres de plus en plus ramifiés avec  $A_{max}$  l'arbre de segmentation maximal. Cependant,  $A_{max}$  est souvent beaucoup trop complexe et conduirait donc à un phénomène de sur-apprentissage, car construit à partir de partitions de  $\chi$  inutilement trop fines. Il s'agit alors de trouver l'arbre optimal au sein de la suite d'arbres précédemment construite. Malheureusement, tous les sous-arbres sont admissibles et il est souvent impossible de tous les considérer, car leur nombre est de croissance exponentielle.

Afin de contourner ce problème, *Breiman et al.* proposent alors une démarche dite d'élagage de l'arbre maximal, consistant à construire une *suite emboîtée de sous-arbres* de l'arbre maximal afin de sélectionner l'arbre optimal seulement parmi cette "sous-suite" créée.

Pour un arbre  $A$  donné, on note  $\tilde{T}_A = \{1, \dots, T_A\}$  l'ensemble de ses feuilles, la valeur de  $T_A$  exprime la complexité de  $A$ . L'erreur d'ajustement d'un arbre  $A$  est mesurée par :

$$g(A) = \sum_{t=1}^{T_A} g(t)$$

où  $g$  est la fonction d'hétérogénéité vue précédemment.

La sélection de la sous-suite d'arbres repose alors sur un arbitrage ajustement/complexité quantifié par un paramètre  $\alpha$ . On crée alors une nouvelle mesure de l'erreur d'ajustement d'un arbre en introduisant un terme de coût de complexité à la mesure d'hétérogénéité. L'objectif sera alors de minimiser la quantité suivante :

$$C_\alpha(A) = g(A) + \alpha T_A$$

Pour  $\alpha = 0$  cette méthode, sélectionnera l'arbre maximal, tandis que pour  $\alpha \rightarrow \infty$  elle privilégiera l'arbre minimal. En résumé, toute division qui n'aura pas créé de "gain" en homogénéité au moins supérieur au paramètre  $\alpha$  sera considérée comme superflue.

**Procédure d'élagage :** On part de l'arbre de segmentation maximal  $A_{max}$ , auquel on retire progressivement les feuilles puis les nœuds qui ne font pas strictement décroître la fonction  $C$ . On obtient ainsi un sous-arbre  $B_1$  de  $A_{max}$ , tel que pour toute branche  $B_1^t$  issue du nœud  $t$  de  $B_1$ , l'inégalité  $g(t) > g(B_1^t)$  est toujours vérifiée.

1. Soit un nœud non terminal  $t$  de  $B_1$ , tant que  $C_{\alpha_t}(\{t\}) > C_{\alpha_t}(B_1^t)$  (ce qui est le cas avec  $\alpha = 0$ ) il est inutile d'élaguer la branche  $B_1^t$ . On cherche alors le seuil  $\alpha_t$  à

partir duquel  $C_{\alpha_t}(\{t\}) \leq C_{\alpha_t}(B_1^t)$  :

$$g(t) + \alpha_t = g(B_1^t) + \alpha_t T_{B_1^t} \iff \alpha_t = \frac{g(t) - g(B_1^t)}{T_{B_1^t} - 1}$$

À partir de  $\alpha = \alpha_t$  l'élagage de la branche  $B_1^t$  est intéressant, car la qualité d'ajustement est équivalente et l'arbre serait plus simple.

On répète cette étape en considérant tous les nœuds non terminaux de l'arbre  $B_1$  afin de trouver le nœud  $\hat{t}$  qui minimise  $\alpha_t$ . On pose alors  $\alpha_1 = \alpha_{\hat{t}}$

2. On élague ensuite l'arbre  $B_1$  à partir du nœud  $\hat{t}$ , et on obtient ainsi l'arbre  $B_2 = B_1 - B_1^{\hat{t}}$ . La procédure est ensuite à répéter à partir de la première étape, jusqu'à revenir à la racine de l'arbre maximal, de telle façon à obtenir une suite d'arbres élagués de tailles décroissantes  $(B_1, \dots, B_{root})$  et la suite croissante  $(\alpha_1, \dots, \alpha_{root})$  correspondante.

Une fois la suite d'arbres élagués créée, on sélectionne le meilleur d'entre eux en comparant leurs erreurs d'ajustement  $C_{\alpha_i}(B_i)$ . L'*arbre optimal de segmentation* est alors celui qui minimise le plus cette quantité.

### 2.2.2 Classification de la décision de versement libre

Nous présentons dans cette partie, l'application de la démarche détaillée ci-dessus à la classification de la décision d'effectuer un versement libre. Pour cela, nous cherchons à expliquer la variable binaire décrivant l'occurrence d'un versement libre à l'aide des caractéristiques des assurés du portefeuille.

Nous construisons l'arbre maximal pour ensuite l'élaguer afin de diminuer sa taille et ainsi de simplifier les résultats. Comme expliqué précédemment, la procédure d'élagage de l'arbre maximal nécessite un choix du coût de complexité à appliquer pour la sélection de l'arbre optimal. La librairie *rpart* du logiciel R, nous permet d'afficher l'erreur d'ajustement, ainsi que la taille de l'arbre associé, en fonction du coût de complexité. Le coût de complexité optimal est alors celui permettant de minimiser l'erreur d'ajustement de l'arbre final. Dans notre cas, le coût de complexité optimal est de  $\alpha = 1,58.10^{-5}$  pour un arbre composé de 26 feuilles. Notre but étant ici d'identifier les déclencheurs essentiels de versements libres et non de produire une classification parfaite de ceux-ci, nous privilégions la construction d'un arbre final moins complexe que l'arbre en théorie optimal. Nous choisissons alors  $\alpha = 3,7.10^{-5}$  pour une faible augmentation de l'erreur d'ajustement de l'arbre et une taille de 16 feuilles. L'arbre obtenu est illustré dans le graphique suivant :

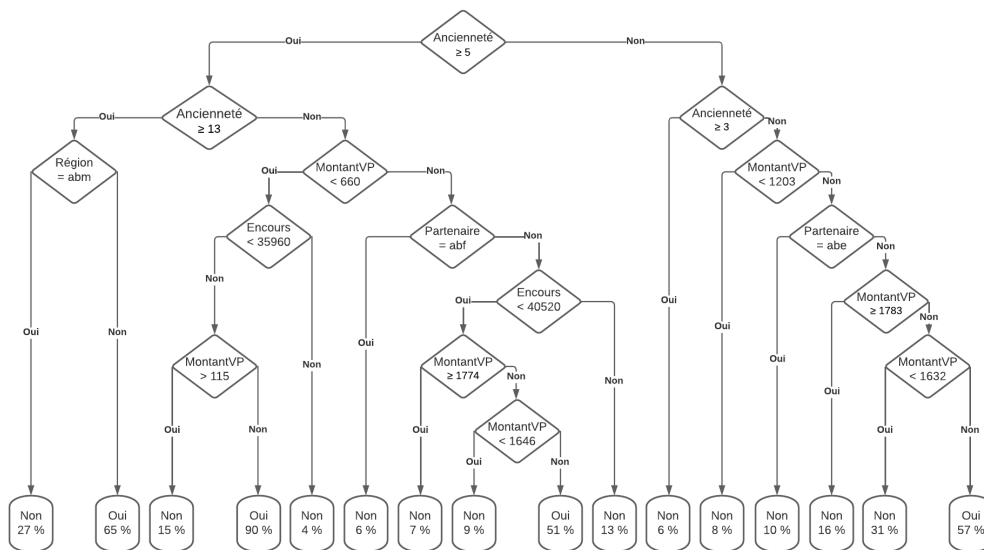


FIGURE 2.17 – Arbre final de classification de la variable binaire décrivant l'occurrence d'un versement libre

Les variables sélectionnées pour la segmentation sont l'ancienneté, le montant des versements programmés, le type de partenaire, l'encours et la région. Dans le cas de variables qualitatives, la règle de division affichée est exprimée à l'aide de lettres représentant les modalités prises par la variable. Une des règles de division est par exemple "Partenaire = abe", les lettres correspondent à la 1e, 2e et 5e modalité de la variable Partenaire. Sur les feuilles de l'arbre, nous observons la classe majoritaire parmi le regroupement de caractéristiques établi ainsi que le pourcentage de verseurs composant ce groupe. Ainsi, les assurés ayant une ancienneté de contrat supérieure à 13 ans sont considérés comme non-verseurs, car on n'observe que 2 % de verseurs sur l'historique du portefeuille parmi les assurés de ce profil.

Nous retrouvons l'observation faite dans la section précédente que l'ancienneté semble être la variable la plus discriminante à la décision d'effectuer un versement libre. En effet, cette variable est associée aux trois premières divisions de l'arbre qui, par construction, identifie des comportements différents sur les contrats ayant une ancienneté inférieure à 3 ans, comprise entre 3 et 5 ans, comprise entre 5 et 13 ans. Nous remarquons, de plus, l'absence des variables âge et sexe dans l'arbre final malgré l'identification, lors des analyses univariées, de leurs impacts sur la décision d'effectuer un versement libre.

Cependant, l'absence d'une variable dans l'arbre final est à interpréter avec précaution, car cela ne veut pas forcément dire que la variable en question n'apporte pas de pouvoir explicatif à la segmentation. En effet, lors de la création d'un nœud et donc d'une règle de division portant sur une variable explicative, il se peut qu'une autre division impliquant une autre variable ait apporté un gain en homogénéité similaire, mais inférieur. C'est

ce que Breiman et al. appellent les variables cachées. Ils introduisent alors la notion d'importance des variables pour améliorer l'interprétation des résultats obtenus par leur méthode, ce phénomène étant inobservable sur l'arbre de classification. Le graphique suivant détaille l'importance des variables explicatives dans la construction de l'arbre final de segmentation.

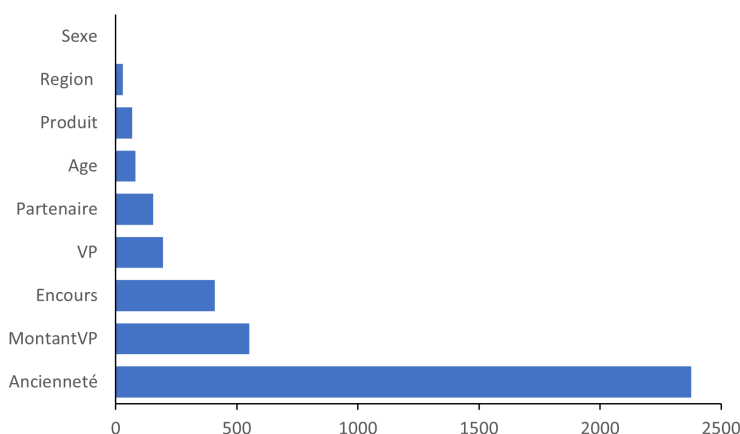


FIGURE 2.18 – Importance des variables explicatives dans la classification de la variable binaire décrivant l'occurrence d'un versement libre

L'observation de ce graphique confirme le fort pouvoir discriminant de la variable ancienneté sur la décision d'effectuer un versement libre. Par ailleurs, nous observons que la variable âge affecte, elle aussi, la décision des assurés comme nous avons pu l'observer dans la section précédente, son importance est même plus élevée que celle de la variable Région qui constitue pour autant une règle de division de l'arbre final. L'absence de la variable Sexe dans l'arbre, quant à elle, révèle bien l'absence d'impact de cette variable sur la décision de versement. En effet, nous avons constaté des taux de verseurs similaires parmi les hommes et les femmes, seules les co-souscriptions se différenciaient, mais pas de façon assez significative pour la méthode CART.

Ainsi, nous avons identifié les variables constituant les déclencheurs essentiels des versements libres : l'ancienneté, les versements programmés, l'encours, le partenaire et l'âge de l'assuré.

### 2.2.3 Classification des montants de versements libres

Nous présentons maintenant l'application de la même démarche que celle détaillée précédemment, appliquée cette fois aux montants des versements libres. De la même façon que pour la variable binaire étudiée précédemment, nous cherchons à expliquer les montants de versements libres à l'aide des caractéristiques des assurés du portefeuille, et cela, afin d'identifier les plus influentes. Cette seconde classification se distingue tout de même de la première du fait du type de la variable étudiée. Dans cette analyse, l'application de la méthode CART permet non pas d'attribuer aux profils créés une classe d'appartenance, puisque nous considérons une variable continue, mais un montant moyen de versement libre.

Les étapes de construction de l'arbre final ayant déjà été expliquées précédemment de manière générale, puis de manière appliquée dans le cadre de la première classification présentée, nous passerons les détails de sa construction dans cette partie. L'arbre final se compose de 17 feuilles et est illustré dans le graphique suivant :

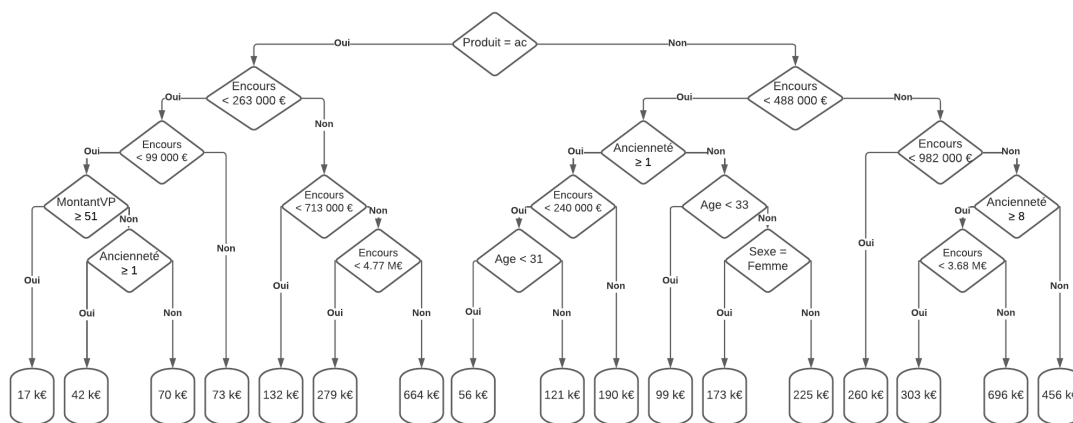


FIGURE 2.19 – Arbre final de classification des montants de versements libres

La première règle de division de cet arbre porte sur la variable *Produit* et distingue les produits "multisupports avec OP" des autres. La majorité des nœuds suivants correspondent à des règles de division portant sur la variable d'encours. Les montants moyens associés aux profils ainsi créés sont cohérents avec les résultats des analyses univariées présentés précédemment. Si l'étude des montants moyens de versements libres en fonction de l'ancienneté ne laissait pas apparaître de relation nette entre les deux variables, nous pouvions en effet remarquer un montant plus élevé en moyenne lors de la première année du contrat. Une ancienneté supérieure à une année constitue ainsi une règle de division de l'arbre final de classification.

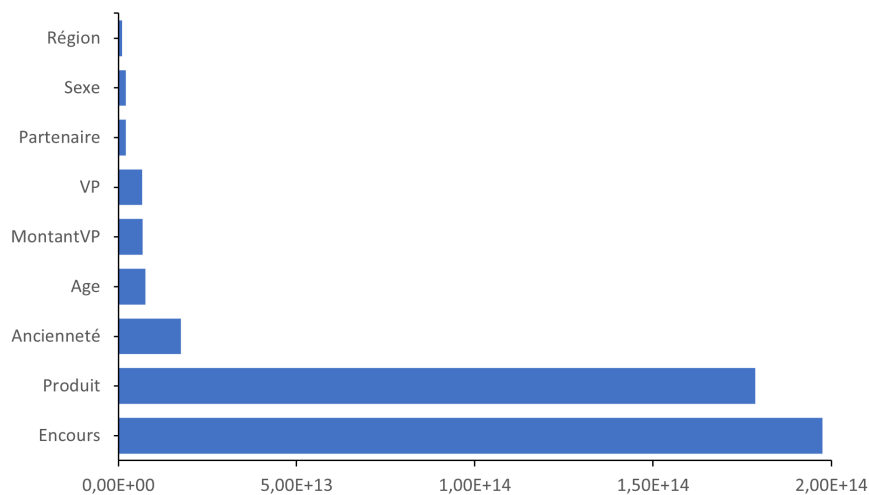


FIGURE 2.20 – *Importance des variables explicatives dans la classification des montants de versements libres*

Par ailleurs, l'analyse de l'importance des variables pour la classification des montants de versements libres révèle le fort pouvoir discriminant du niveau d'encours et du type de produit associé au contrat. Ainsi, la variable produit est significativement plus influente sur ces montants que la variable décrivant la présence de versements programmés sur le contrat, contredisant ainsi les conclusions que nous avons pu établir des analyses précédentes.

## 2.3 Étude de la corrélation du taux de versements libres avec la conjoncture économique

Si les caractéristiques des assurés semblent avoir un impact notable sur leur comportement de versement libre, les observations réalisées dans la partie précédente nous amènent à penser que les informations fournies par le portefeuille étudié ne constituent pas les seuls facteurs explicatifs de ce comportement. Prenons l'exemple du graphique décrivant, en moyenne et pour quelques années, la répartition des taux de versements libres en fonction de l'âge de l'assuré. Force est de constater que bien que ces courbes observent toutes la même tendance, elles restent néanmoins très différentes les unes des autres. Pour autant, comme nous l'avons détaillé avant, le portefeuille étudié n'évolue pas de façon assez significative pour pouvoir expliquer de telles fluctuations du comportement assuré dans le temps. Par ailleurs, l'évolution en 15 ans de la proportion de versements libres enregistrés affiche d'importants "creux" à des périodes tourmentées des marchés financiers.

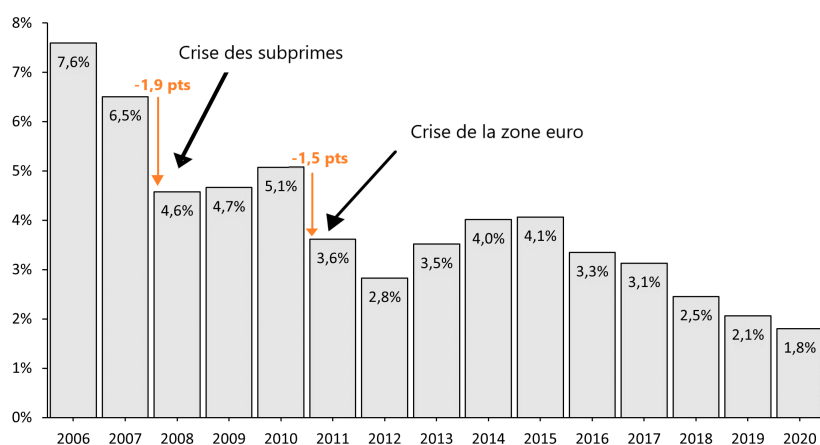


FIGURE 2.21 – Évolution des taux de versements annuels

On observe en effet deux importantes baisses de la proportion de verseurs en 15 ans : une baisse de près de 2 points en 2008, année de la crise des subprimes, ainsi qu'une baisse d'un point et demi en 2011, année de déclaration de la crise de la dette de la zone euro. Ces observations plaident en faveur de l'existence d'un impact de l'environnement économique sur le comportement de versement libre. Nous cherchons alors dans cette partie à identifier des indicateurs économiques permettant d'expliquer les fluctuations dans le temps des comportements observés.

Au-delà de l'observation que nous avons pu faire de l'impact notable des crises financières sur la décision de versement, nous nous intéressons aussi aux motivations de cette opération. En effet, la décision de versement étant étroitement liée à la volonté d'épargne de l'assuré, nous considérerons, de plus, des indicateurs "sociétaux" décrivant les évolutions de cette volonté ou encore les causes de celle-ci.



## Le CAC 40

Le premier indicateur étudié est celui du CAC 40, indicateur par excellence de la santé des marchés financiers français. Si l'ACPR (2016) relève une corrélation entre la collecte brute sur les supports UC et le cours du CAC 40, nous souhaitons vérifier si cette observation se conforme à l'échelle de notre portefeuille. Nous étudions ainsi l'éventuel lien entre la cotation moyenne du CAC 40 et la proportion de versements libres observés sur l'année.

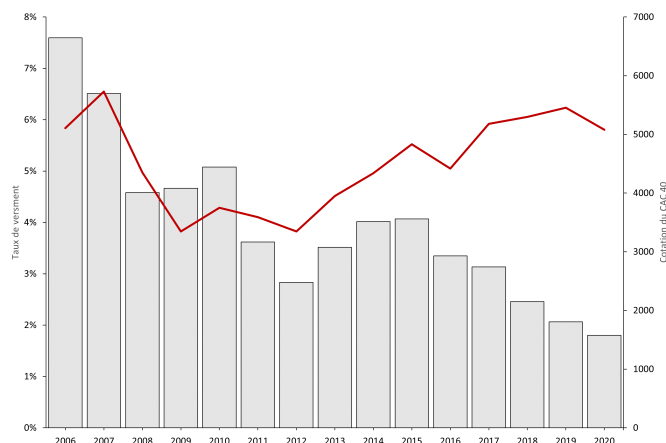


FIGURE 2.22 – Évolution des taux de versements en comparaison à celle du CAC40

Si nous observons une symétrie des évolutions de la cotation du CAC 40 et des taux de versements libres sur le portefeuille jusqu'à l'année 2017, les tendances des deux courbes s'inversent ensuite. On observe ainsi une corrélation de 40 % entre 2006 et 2017, mais de -18 % en considérant les 15 années d'observations. Nous introduirons aussi dans nos modèles la volatilité du cours du CAC 40 afin de ne pas biaiser les résultats de cette analyse par le "large" pas annuel de l'étude. En effet, il est intuitivement possible d'observer, pour une même cotation de l'indice, des comportements différents en fonction de la volatilité de celui-ci.

En plus du CAC 40, nous considérons les variables macroéconomiques suivantes :

### L'Indice de Confiance des Ménages (ICM)

Établi par l'INSEE, cet indicateur permet de refléter la confiance des ménages envers l'environnement économique. Évalué mensuellement, il synthétise les réponses d'un pool significatif de personnes sur des questions s'intéressant aussi bien à leur situation financière présente et future, qu'à leurs capacités d'épargne à venir, ou encore à leur crainte du chômage. Il se présente alors comme une source d'informations clé à la prédiction du comportement d'individus sur des sujets d'épargne.

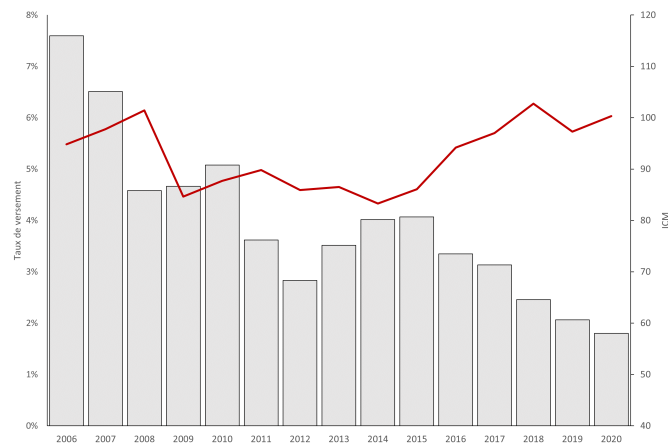


FIGURE 2.23 – Évolution des taux de versements en comparaison à celle de l'ICM

En raison de sa qualité d'indicateur des aspirations futures des ménages, nous remarquons une corrélation plus élevée de l'évolution des taux de versement avec celle de l'ICM pris en compte avec une année de décalage. De la même façon que pour le CAC 40, nous observons jusqu'en 2015 une forte corrélation des taux de versements et de l'ICM. À partir de 2016, l'ICM adopte une tendance inverse à celle observée sur les taux de versements du portefeuille, amoindrissant ainsi la corrélation de ces deux variables sur l'historique observé.

### Le taux d'inflation

Cet indicateur de la santé économique du pays, publié par l'[INSEE](#), se comporte comme un baromètre de la hausse des prix à la consommation et traduit une baisse du pouvoir d'achat de la monnaie. Le taux d'inflation impacte la valeur de l'épargne : effectuer un placement avec un rendement inférieur à ce taux ne rapporte plus rien et coûte même de l'argent à l'agent.

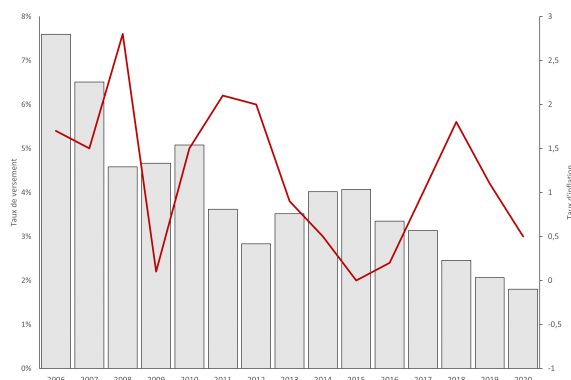


FIGURE 2.24 – Évolution des taux de versements en comparaison à celle de l'inflation

En diminuant le pouvoir d'achat, une augmentation de l'inflation a ainsi l'effet de diminuer par conséquent la capacité d'épargne des consommateurs. Pour ces raisons, nous nous attendons à observer une corrélation négative de l'inflation et des taux de versements, nous retrouvons cette observation par l'étude graphique de l'évolution de ces deux variables sur l'historique étudié.

### Le taux de chômage

Établi par l'INSEE sur la base de l'Enquête Emploi en Continu, le taux de chômage représente la proportion de chômeurs au sens du Bureau International du Travail (BIT) parmi la population de personnes actives. Les études de l'INSEE (Antonin (2019)) révèlent l'apparition d'un mécanisme d'épargne de précaution en climat incertain du fait de l'augmentation du chômage et donc de l'augmentation de la probabilité d'être au chômage.

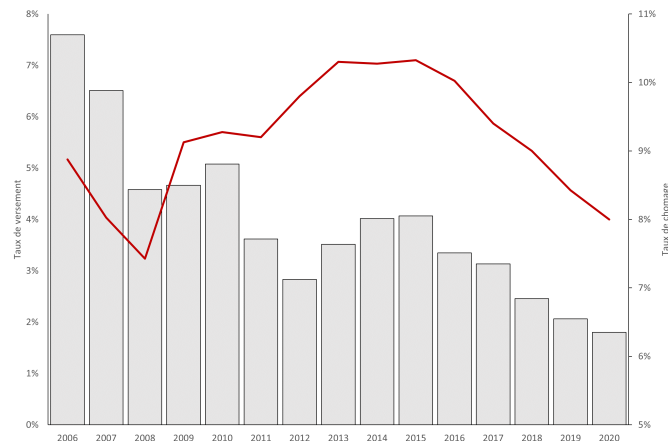


FIGURE 2.25 – Évolution des taux de versements en comparaison à celle du taux de chômage

La réaction de l'agent est donc de se constituer une épargne supplémentaire au regard de l'incertitude sur ses revenus futurs. À l'échelle de notre portefeuille, nous retrouvons bien une corrélation positive entre le taux de chômage et les taux de versements.

### Le Taux Moyen d'emprunt d'État (TME)

Fourni par la Banque de France, ce taux représente la moyenne arithmétique des rendements moyens hebdomadaires des rendements des emprunts d'État. La baisse ces dernières années des rendements obligataires et notamment du TME est liée à la chute des rendements des supports euros. Les assurés se voient alors contraints de diversifier leur épargne pour pouvoir obtenir de meilleures performances sur leurs contrats.

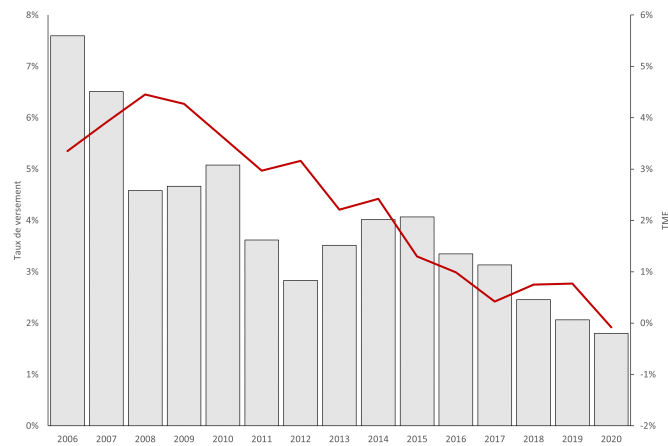


FIGURE 2.26 – Évolution des taux de versements en comparaison à celle du TME

Le TME a connu une baisse continue ces dernières années, et cela à partir de 2008, cette baisse pourrait, pour les raisons évoquées précédemment, expliquer en partie la baisse globale du nombre de versements libres observés à l'échelle de notre portefeuille sur l'historique considéré.

### La différence des rendements servis par la concurrence et par l'assureur

Étudier l'impact du taux de PB sur la probabilité d'effectuer un versement libre reviendrait à utiliser l'information du TME. Cependant, la possession de cette information est précieuse, car elle nous permet d'étudier l'impact de la rémunération de l'assureur en comparaison de celle de ses concurrents. En effet, un assuré pourrait être en possession de plusieurs contrats d'assurance vie chez différents assureurs ou encore être amené à en souscrire à un nouveau, il est alors légitime de penser qu'en cas de volonté d'effectuer un versement libre, son choix se portera sur le contrat promettant le meilleur rendement. Ainsi, nous choisissons de considérer la différence des taux servis sur le portefeuille avec les taux moyens servis sur le marché.

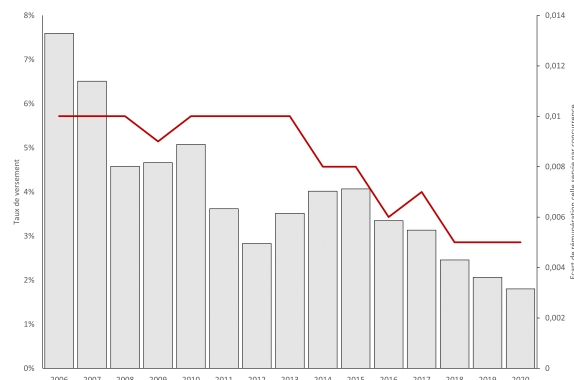


FIGURE 2.27 – Évolution des taux de versements en comparaison à celle de l'écart de rendement des supports Euros du portefeuille et ceux observés sur le marché

Pour cela, nous recueillons les rendements moyens des supports Euros observés sur le marché et fournis par la FFA. L'écart du rendement offert sur le portefeuille est stable à + 1 point de pourcentage jusqu'en 2012, mais chute par la suite pour atteindre un écart réduit à un demi-point de pourcentage. Cette baisse observée pourrait permettre d'expliquer la baisse du nombre de versements libres observée à partir de 2016, qui semble peu explicable par les indicateurs précédemment décrits.

### La différence des rendements servis sur le Livret A et par l'assureur

Après s'être intéressés à l'environnement concurrentiel de l'assureur, nous nous intéressons aux concurrents de l'assurance vie et plus particulièrement aux produits d'épargne bancaire. Nous choisissons ainsi d'étudier l'impact, sur le taux de versements, de la comparaison des rendements servis sur le portefeuille et des rendements offerts sur les produits d'épargne bancaire. Le choix de considérer cette variable est motivé par la lecture de diverses études sur le comportement d'épargne des Français.

L'ACPR (2021a) révèle par exemple une préférence des Français pour l'épargne bancaire en temps de crise et plus particulièrement pendant la crise sanitaire, pour l'avantage d'une meilleure liquidité qu'ils procurent, et cela, au détriment de l'assurance vie qui a observé une collecte nette négative en 2020. Cette observation est par ailleurs appuyée par Le Cercle des Épargnants (2021) qui, par le biais de son baromètre 2021 sur l'épargne et la retraite en France, décrit une attraction grandissante des Français pour des produits d'épargne bancaire tels que le Livret A.

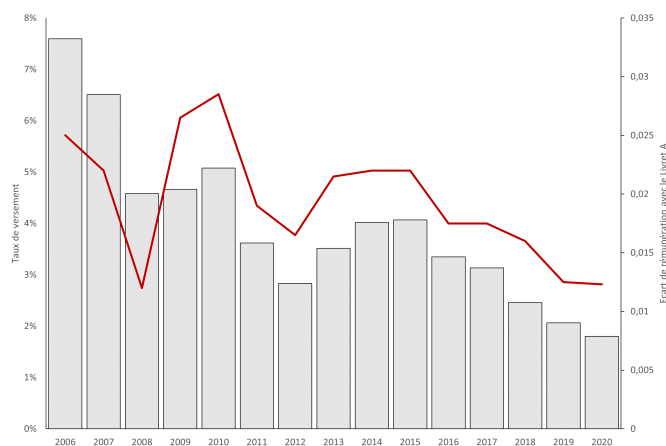


FIGURE 2.28 – Évolution des taux de versements en comparaison à celle de l'écart de rendement des supports Euros du portefeuille et de celui du Livret A

Le Livret A constituant le plus gros produit d'épargne bancaire, nous choisissons d'étudier l'impact, sur les taux de versements du portefeuille, de la différence des rendements servis par l'assureur avec ceux proposés sur ce support d'épargne. Ainsi, il serait possible d'observer, à partir d'un certain écart dans la rémunération du Livret A et celle servie par

l'assureur, une baisse du nombre de versements au profit de supports d'épargne bancaire. Nous remarquons une forte corrélation (positive) entre les taux de versements observés sur le portefeuille et la variable ainsi créée.

### Les périodes électorales

Il est usuel de voir à chaque campagne présidentielle, le sujet de la fiscalité de l'assurance vie remis sur la table. En effet, nombre de candidats à cette élection prévoient dans leurs programmes une révision de cette fiscalité. Néanmoins, les assureurs appuyés par des associations d'épargnants telles que l'Afer parviennent à obtenir des candidats des retours sur leurs propositions. Cependant, nous retrouvons dans plusieurs articles de presse l'observation d'un attentisme des Français à l'approche des périodes électorales, provoqué par l'incertitude du devenir de la fiscalité de l'assurance vie ([Le Figaro \(2017\)](#)). Cet attentisme se traduit par l'observation d'une augmentation des rachats associée à une baisse des cotisations ([ACPR \(2016\)](#)).

Nous retrouvons ces observations à l'échelle de notre portefeuille, où les taux de versements observés affichent une baisse en 2007, 2012 et 2017. Nous considérons alors une variable binaire, décrivant les années de périodes électorales, dans la construction de nos modèles.

### Conclusion

Nous avons ainsi identifié un large panel de variables conjoncturelles et ceci dans le but de sélectionner par la suite les plus significatives sur le taux de versement libre. Avant de mesurer la significativité de ces variables pour l'explication de la décision de versement libre, nous étudions leurs corrélations. En effet, certaines de ces variables, comme l'ICM et le taux de chômage ou encore les variables relatives au rendement des supports euros, recourent des informations similaires. Nous observons ainsi de fortes corrélations, d'un coefficient calculé supérieur à 0.8, entre la cotation moyenne annuelle du CAC40 et l'ICM ainsi qu'entre les variables décrivant les écarts de rendements des supports euros du portefeuille et ceux observés sur le marché ou celui du Livret A.

Par ailleurs, afin d'établir des liens entre conjoncture et taux de versement, il est nécessaire de se questionner sur le décalage à prendre en compte sur l'observation du contexte économique. En effet, nos observations étant réalisées à un pas annuel, nous pouvons alors nous demander si les comportements de versements sont impactés par la situation économique de l'année, de l'année précédente ou bien des deux? Nous avons déjà pu constater que dans le cas de l'ICM, la piste d'un décalage d'une année serait à privilégier. Nous répondons à ces questions dans la suite de ce mémoire, lors de la présentation des résultats des modèles.

## Chapitre 3

# Construction des lois de projection des versements libres

Nous développons dans ce chapitre une méthodologie de construction de lois de projection des comportements de versements libres. Pour cela nous distinguons, de la même façon que dans nos analyses, la modélisation de l'occurrence d'un versement libre de celle de son montant. Les méthodes implémentées sont basées sur des techniques courantes et présentant l'avantage d'être simples d'interprétation. Nous répondons ainsi aux contraintes opérationnelles imposées à l'utilisation de telles lois. Ce chapitre présente alors dans un premier temps, la modélisation des probabilités de versements libres, suivie dans un second temps de celle des montants moyens associés pour conclure sur l'agrégation des projections et leur analyse.

### 3.1 Modélisation des probabilités de versements libres

Nous nous proposons dans cette section de modéliser la décision des assurés d'effectuer un versement libre ou non, par le biais de régressions logistiques.

Afin d'évaluer l'apport de la prise en compte de la conjoncture économique dans la modélisation de ce phénomène, nous effectuons pour chacune de ces méthodes :

- une approche statique de l'occurrence des versements libres : nous cherchons à expliquer ce phénomène uniquement à partir de l'historique fourni par nos données. Nous considérons ainsi que les comportements sont homogènes dans le temps, l'impact de chacune des variables associées aux contrats sera alors apprécié en moyenne sur l'historique. Les résultats des études précédentes nous laissent déjà intuire la mauvaise qualité d'ajustement de cette approche.
- une approche dynamique de l'occurrence de phénomène : nous prenons en compte la variabilité temporelle des observations faites sur notre portefeuille, pour cela nous introduisons les variables macroéconomiques étudiées précédemment. L'introduction de ces variables permet alors de différencier les divers environnements économiques rencontrés sur l'historique d'observation du portefeuille.

Nous considérons dans cette section et pour chacune de nos modélisations un échantillon d'apprentissage constitué des observations du portefeuille réalisées entre 2006 et 2016. Nous conservons ainsi un historique amplement raisonnable de 10 ans - riche de toutes les crises économiques rencontrées sur cette période - pour l'apprentissage des modèles, et nous réservons les années de 2017 à 2020 pour l'appréciation de la prédiction de nos modèles.

#### 3.1.1 Régression logistique

La régression logistique appartient à la famille des modèles linéaires généralisés, connus sous le nom de GLM. Cette méthode semi-paramétrique permet de modéliser la probabilité d'occurrence d'un événement à partir de variables explicatives. L'utilisation de cette méthode est donc parfaitement adaptée à notre étude de la probabilité d'effectuer un versement libre. Dans le contexte de cette partie, rappelons que l'objectif est de segmenter nos variables afin de déterminer les profils d'assurés les plus enclins à effectuer des versements libres. Cette procédure étant assez populaire, nous l'expliquerons donc brièvement dans ce mémoire.



## Principe

Considérons un échantillon de données  $D = (y_i, x_i^1, \dots, x_i^p)_{1 \leq i \leq N}$  constitué de  $N$  observations d'une variable à expliquer  $Y \in \mathcal{Y} = [0; 1]$  et de  $p$  variables explicatives  $X^j \in \mathcal{X}_j$ . On note  $\chi = \prod_{j=1}^p \mathcal{X}_j$ . Dans le cadre d'une régression logistique, l'exposition à un ensemble des variables explicatives est représentée par le prédicteur linéaire  $z$ , formalisé par l'équation de régression classique :

$$z = \beta_0 + \beta_1 X^1 + \dots + \beta_p X^p$$

où  $X = (X^j)_{1 \leq j \leq p}$  est la matrice des  $N$  observations des  $p$  variables explicatives du jeu de données. La suite  $\beta = (\beta_i)_{0 \leq i \leq p}$  est une suite de coefficients de régression associés à chacune des variables explicatives.

Le résultat de la régression est ensuite transformé en probabilité à l'aide d'une fonction de lien : la fonction logistique  $\Phi$  dans le cadre de notre régression logistique.

$$\forall z \in \mathbb{R}, \Phi(z) = \frac{1}{1 + e^{-z}} = \frac{e^z}{1 + e^z}$$

La fonction logistique permet de projeter tout antécédent réel dans l'espace  $[0, 1]$  et respecte de plus la propriété de non-décroissance d'une fonction de répartition classique.

L'estimation des coefficients de régression se fait ensuite par la méthode du maximum de vraisemblance. L'estimateur du maximum de vraisemblance  $\hat{\beta}$  satisfait  $\frac{\partial \ln(L)}{\partial \beta}(\hat{\beta}, X) = 0$  avec :

$$L(\beta, X) = \prod_{i=1}^N \Phi(X_i \beta')^{Y_i} (1 - \Phi(X_i \beta'))^{1 - Y_i}$$

Ce système est résolu à partir d'algorithmes d'optimisation tels que celui de Newton-Raphson.

La probabilité individuelle d'effectuer un versement libre est alors naturellement,

$$\forall i \in [1; N], \hat{p}_i = \Phi(X_i \hat{\beta}')$$

La validation du modèle se fait ensuite par le biais du test du ratio de vraisemblance qui permet d'évaluer la significativité globale du modèle. La validation de la significativité de chaque variable explicative se fait, quant à elle, à l'aide du test de Wald.

## Limites de la régression logistique

Les hypothèses nécessaires à la pratique de la régression logistique constituent les principales limites de cette technique de modélisation. En particulier, l'estimation des coefficients du modèle requiert la condition que les variables explicatives ne soient pas significativement corrélées les unes entre les autres. Cette hypothèse est satisfaite par les variables structurelles du portefeuille, mais nous avons précédemment vu qu'elle ne l'était pas dans le cas des variables macroéconomiques sélectionnées. L'utilisation de la régression logistique lors de l'analyse dynamique de la décision de versements libres impose alors une étape préliminaire de sélection des variables.

### 3.1.2 Analyse statique

Nous cherchons dans cette partie à expliquer l'occurrence d'un versement libre, soit la variable binaire associée, en fonction des caractéristiques associées au contrat, à savoir :

- l'âge de l'assuré ;
- le sexe de l'assuré ;
- le montant d'encours ;
- l'ancienneté du contrat ;
- le produit souscrit ;
- le partenaire commercial ;
- la présence de versements programmés (variable binaire) ;
- le montant de ces versements programmés ;
- le montant du versement initial s'il y en a eu un durant l'année.

### Sélection des variables explicatives

Si nous cherchons, dans un premier temps, à utiliser dans le modèle toutes les variables citées, nous cherchons cependant, dans un second temps, à éviter un sur-apprentissage de notre modèle. Un modèle trop complexe, et donc en situation de sur-apprentissage, présenterait l'écueil de décrire l'erreur aléatoire des données plutôt que de décrire les relations entre les variables explicatives et la variable à expliquer. En d'autres termes, le modèle "collerait" trop aux données d'apprentissage au lieu d'intégrer les informations essentielles à la prédiction. Un modèle avec peu de variables a ainsi plus de chances de fournir des prédictions robustes.

Il est alors nécessaire d'appliquer une démarche de sélection des variables explicatives

afin de trouver un équilibre entre qualité d'adéquation aux données d'apprentissage et complexité du modèle (nombre  $K$  de variables explicatives introduites). Pour cela, nous utilisons le *Critère d'Information d'Akaike*, dit critère AIC, s'écrivant comme suit :

$$AIC(\beta, X) = 2K - 2\ln(L(\beta, X))$$

Afin de sélectionner le modèle offrant le meilleur compromis complexité/qualité d'adéquation, nous partons du modèle initialement implémenté à partir de l'entièreté des variables explicatives présentées, nous identifions la variable la moins significative et nous appliquons la démarche suivante :

- dans le cas d'une variable quantitative : nous la supprimons du modèle ;
- dans le cas d'une variable qualitative : nous regroupons la modalité avec une autre ayant un coefficient estimé semblable à celui de la première.

Nous calculons le critère AIC associé à ce nouveau modèle puis nous réitérons la démarche jusqu'à ce que toutes les variables explicatives soient significatives. Nous créons ainsi une série de modèles imbriqués et sélectionnons le modèle associé au critère AIC le plus faible.

L'application de cette démarche à notre régression logistique indique :

- la suppression de la variable Encours, n'apportant pas de pouvoir explicatif à l'occurrence d'un versement libre.
- le regroupement de la modalité "Femme" avec la modalité "Homme". Ce regroupement est cohérent avec les analyses préliminaires présentées dans le chapitre précédent, les pourcentages de verseurs parmi les hommes et les femmes n'étant pas significativement différents.
- le regroupement des modalités "Autres", "Courtiers" et "Banque de réseau" de la variable partenaire, la distinction de ces trois types n'apportant aucun pouvoir explicatif significatif au modèle.

Le tableau suivant (Figure 3.1) détaille les coefficients de régression estimés par le modèle.

### Interprétation des coefficients du modèle

Le détail des coefficients estimés par le modèle est fourni par le tableau suivant. Si la valeur des coefficients d'une régression logistique ne peut être interprétée directement, sans calcul préalable, il est possible d'identifier la nature d'impact de chacune des modalités/variables du modèle sur la variable expliquée, en fonction du signe de ces coefficients. Ainsi le coefficient positif de 0.002 estimé pour la variable Age nous indique que le modèle considère la probabilité d'effectuer un versement libre comme augmentant avec l'âge.

Par ailleurs, les coefficients associés aux variables qualitatives sont à interpréter avec

	Modalités	Coefficient estimé	Ecart type	p-value du test de Wald
$\beta_0$		-2,276E+00	1,507E-02	< 2E-16 ***
Sexe	Indivision	-7,633E-02	6,929E-03	3,29E-13 ***
Region	Auvergne	-1,370E-02	1,176E-02	< 2E-16 ***
	Bourgogne	-2,029E-02	2,501E-02	4,91E-16 ***
	Bretagne	-2,135E-01	1,427E-02	< 2E-16 ***
	Centre	-1,409E-01	1,818E-02	< 2E-16 ***
	Corse	-3,090E-01	6,296E-02	9,17E-07 ***
	Etranger	-1,409E-01	1,861E-02	3,78E-14 ***
	Grand Est	-3,701E-01	1,769E-02	< 2E-16 ***
	Hauts de France	-3,445E-01	1,671E-02	< 2E-16 ***
	Normandie	-1,220E-01	1,889E-02	1,06E-10 ***
	Nouvelle-Aquitaine	-2,204E-02	1,240E-02	< 2E-16 ***
	Occitanie	-7,234E-02	1,560E-02	3,55E-06 ***
	Outre Mer	-1,841E-01	2,340E-02	3,61E-15 ***
	Pays de la Loire	-3,569E-01	2,046E-02	< 2E-16 ***
PACA	-1,901E-01	1,306E-02	< 2E-16 ***	
VP		6,938E-01	1,006E-02	< 2E-16 ***
MontantVP		2,554E-07	2,393E-08	< 2E-16 ***
MontantV1		8,354E-06	1,257E-02	< 2E-16 ***
Produit	Mono-euro	-1,561E-01	2,167E-02	< 2E-16 ***
	Multisupport avec OP	-1,003E-01	1,481E-02	9,71E-04 ***
Partenaire	Gestionnaire de fortune	2,517E-01	1,340E-02	< 2E-16 ***
	Réseau propriétaire	1,960E-01	1,943E-02	< 2E-16 ***
	Société de gestion	2,290E-01	9,302E-03	< 2E-16 ***
Age		1,740E-03	2,116E-04	< 2E-16 ***
Ancienneté		-6,722E-02	8,153E-04	< 2E-16 ***

FIGURE 3.1 – Résultats de l'estimation des coefficients de régression - Approche statique

précaution. En effet, si nous avons trois types de produits, seules deux modalités de la variable produit apparaissent dans le tableau des coefficients estimés. Les produits multisupports sans option d'orientation personnalisée sont en réalité considérés comme classe de référence de la variable Produit, car c'est la plus représentée dans l'échantillon d'apprentissage. Les coefficients affichés pour les deux classes sont alors à lire en référence à la classe de référence. Ainsi, le coefficient négatif estimé pour les produits monosupport euro nous indique que les assurés ayant souscrits à ces produits sont estimés comme ayant moins de chances d'effectuer un versement libre que ceux ayant souscrit aux produits de la classe de référence.

Ce rapport de chances, plus connu sous le nom d'*Odds-Ratio* (OR), est cependant quantifiable et se définit comme le rapport de probabilité  $\frac{p}{1-p}$  au sein d'un groupe. Ainsi, si l'on prend l'exemple d'un groupe d'assurés (ayant les mêmes caractéristiques) associés à une probabilité d'effectuer un versement libre estimée 0.7, on a alors sur un assuré de ce groupe  $0.7/0.3 = 2.33$  fois plus de chance d'observer un versement libre.

Cette idée se généralise ensuite au calcul d'un *Risque relatif* lié à un changement de caractéristiques. Ce risque relatif étant le rapport des OR associés aux deux caractéristiques que l'on souhaite comparer. Prenons comme exemple la variable Ancienneté, le risque relatif d'observer un versement libre chez un assuré d'ancienneté 5 ans en comparaison avec un assuré d'ancienneté 10 ans vaut :

$$\frac{OR(X_{anc} = 5)}{OR(X_{anc} = 10)} = \frac{\frac{\mathbb{P}[Y=1|X_{anc}=5]}{\mathbb{P}[Y=0|X_{anc}=5]}}{\frac{\mathbb{P}[Y=1|X_{anc}=10]}{\mathbb{P}[Y=0|X_{anc}=10]}} = \frac{e^{5\hat{\beta}_{anc}}}{e^{10\hat{\beta}_{anc}}}$$

Les Odds-Ratios des variables du modèle sont illustrés par le graphique suivant. Les cercles (pleins en cas de significativité de la variable et vides sinon) présentent la valeur de l'odds-ratio associé à la variable/modalité et sont accompagnés de barres de part et d'autre, illustrant leurs intervalles de confiance. Les résultats présentés sont à lire de la façon suivante :

- pour les variables quantitatives : l'OR décrit la variation de la probabilité d'effectuer un versement libre pour une augmentation d'une unité de la variable considérée. Ainsi, nous remarquons que le vieillissement d'une année d'un contrat engendre une diminution de 7 % sur sa probabilité d'effectuer un versement libre ;
- pour les variables qualitatives : l'OR décrit la variation de la probabilité d'effectuer un versement libre pour la modalité considérée, en comparaison avec la modalité de référence (celle qui n'apparaît pas dans le tableau). Ainsi, nous remarquons que la probabilité d'observer un versement libre sur un contrat associé au réseau de partenaire des gestionnaires de fortunes est par exemple 30 % plus élevé que celle d'en observer sur un contrat associé au groupe de partenaires composé des courtiers, des banques de réseau et des autres.

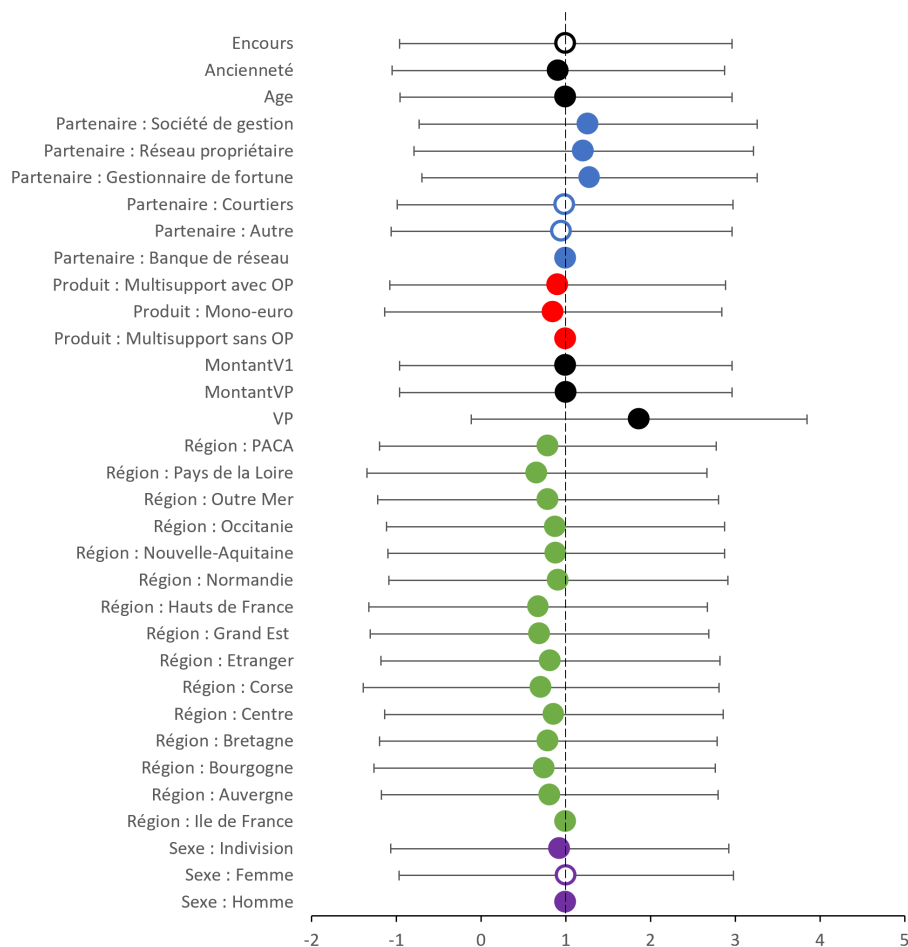


FIGURE 3.2 – Odds-Ratios - Approche statique

Les écueils de cette approche statique sont alors facilement identifiables en observant les odds-ratios établis par le modèle. Prenons l'exemple de l'impact estimé de la variable décrivant la présence de versements programmés sur le contrat : le modèle estime qu'il est deux fois plus probable qu'un assuré faisant des versements programmés, effectue un versement libre en comparaison avec un assuré n'en faisant pas. Cette conclusion du modèle n'est vraie qu'en moyenne : sur l'année 2010, nous observons 2,5 fois plus de verseurs parmi ces 2 profils d'assurés tandis qu'en 2016 nous n'en observions que 1,3 fois plus.

Ces écarts d'ajustements du modèle sont observables en considération de l'âge de l'assuré sur le graphique suivant.

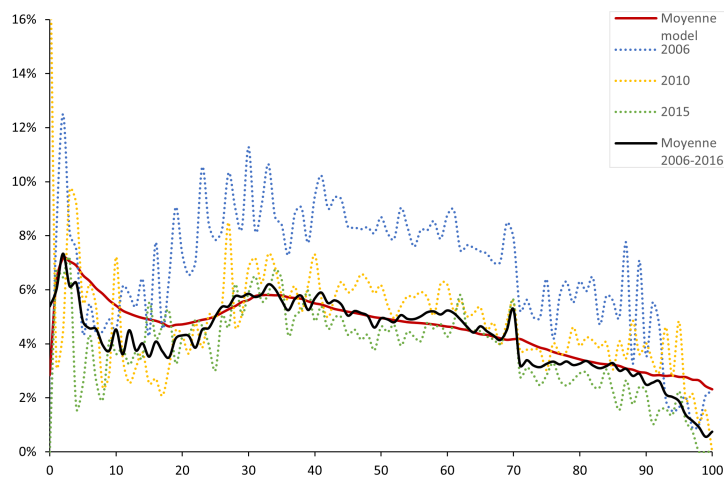


FIGURE 3.3 – *Taux de versements observés ajustés par le modèle en fonction de l'âge*

La courbe rouge représente la moyenne des probabilités d'effectuer un versement ajustées par le modèle sur l'ensemble de l'échantillon d'apprentissage, et cela, en fonction des âges du portefeuille.

La représentation des moyennes de ces probabilités ajusté en fonction de plus de l'année d'observation du portefeuille est inutile, car cela résulterait en multitude de courbes confondues. En effet, comme expliqué précédemment pour l'effet des versements programmés sur la variable d'intérêt, le modèle vise dans cette démarche statique à approcher la courbe noire du graphique ci-dessous, représentant les taux de versements par âge moyennés de 2006 à 2016.

### Qualité d'ajustement du modèle

L'interprétation des coefficients estimés nous amène alors à intuire la mauvaise qualité des résultats fournis par la régression logistique sous cette approche statique. Le modèle ne dispose d'aucune information permettant d'expliquer les différences de taux de versements

chaque année, il ne dispose que des caractéristiques des contrats. Ainsi, ces caractéristiques n'évoluant pas significativement dans le temps, les taux de versements ajustés par le modèle ne varient pas dans le temps. Le graphique ci-dessous illustre l'évolution des taux de versement prédits les échantillons d'apprentissage et de validation, distingués dans le graphique par la ligne verticale en pointillés. Le taux de versements sur l'année est obtenu à partir de la moyenne des scores ajustés par le modèle.

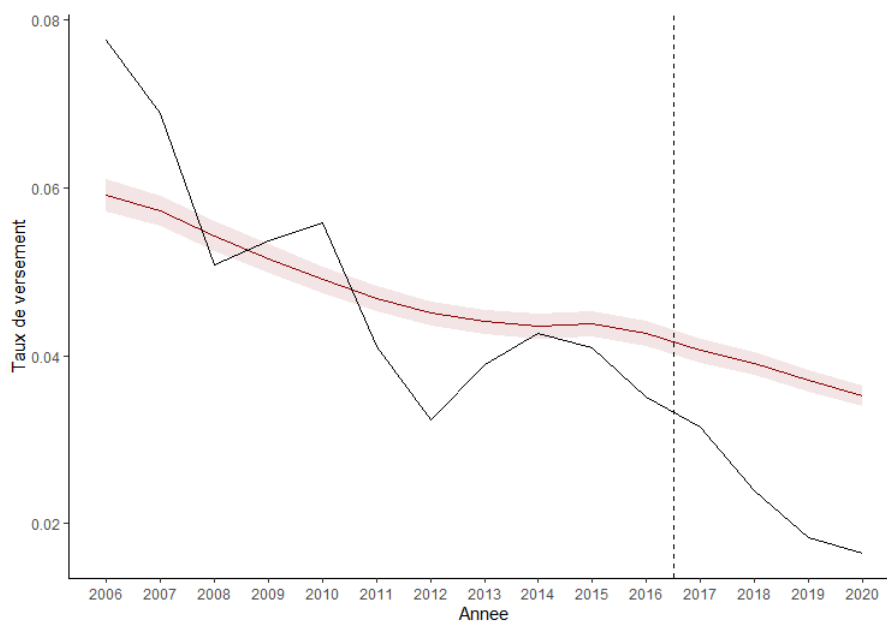


FIGURE 3.4 – *Taux de versements ajustés via une régression logistique - Approche statique*

Nous remarquons tout de même la tendance baissière des taux de versements ajustés par régression logistique. Cette baisse est explicable par le vieillissement du portefeuille : les âges et les anciennetés des contrats augmentent, faisant ainsi baisser la probabilité d'observer des versements libres.

Nous introduisons alors, dans une seconde partie, des variables explicatives supplémentaires dans ce modèle afin de prendre en compte la variabilité temporelle des impacts de chacune des variables structurelles.

### 3.1.3 Approche dynamique

Nous cherchons dans cette partie à expliquer l'occurrence d'un versement libre, soit la variable binaire associée, en fonction des caractéristiques associées au contrat introduites dans le précédent modèle ainsi que des variables économiques présélectionnées et décrites précédemment. Ces variables se composent de :

- l'indice de confiance des ménages avec et sans décalage d'une année ( $ICM_N$  et  $ICM_{N-1}$ );
- la cotation moyenne annuelle du CAC 40 avec et sans décalage d'une année ( $CAC_N$  et  $CAC_{N-1}$ );
- la volatilité annuelle des cours du CAC 40 avec et sans décalage d'une année ( $V\_CAC_N$  et  $V\_CAC_{N-1}$ );
- le taux de chômage;
- le taux d'inflation;
- le TME;
- la différence du taux de PB servis par l'assureur et des taux moyens servis par la concurrence ( $PB\_CONC$ );
- la différence du taux de PB servis par l'assureur et des taux servis sur le Livret A ( $PB\_LA$ );
- une variable indicatrice des périodes électorales.

Comme nous l'avons expliqué auparavant, la régression logistique impose la non-corrélation des explicatives introduites dans le modèle. Si nous avons pu échapper à cette étape de sélection des variables explicatives introduites dans le modèle dans notre première approche, nous devons cependant dans cette seconde approche procéder à cette sélection.

#### Sélection des variables explicatives

Les corrélations observées entre les variables présélectionnées choisies imposent un choix entre les suivantes :

- l'ICM de l'année N ( $ICM_N$ ) ou la cotation moyenne du CAC 40 de l'année N ( $CAC_N$ );
- l'ICM de l'année N-1 ( $ICM_{N-1}$ ) ou la cotation moyenne du CAC 40 de l'année N-1 ( $CAC_{N-1}$ );
- la différence du taux de PB et de la moyenne des taux appliqués par la concurrence ( $PB\_CONC$ ) ou la différence du taux de PB et du taux du Livret A ( $PB\_LA$ ).

Nous implémentons alors les modèles correspondant aux huit combinaisons de variables



explicatives citées, nous choisissons alors de conserver celui offrant le critère AIC le plus faible. La complexité des huit modèles étant la même, le critère AIC nous informe ici de la déviance des modèles. La combinaison des variables  $CAC_N$ ,  $ICM_{N-1}$  et  $PB\_LA$  ainsi que des autres variables économiques présélectionnées permettent à la régression logistique de mieux approcher la variable binaire caractéristique de la décision d'effectuer un versement libre.

L'introduction des variables économiques a un impact majeur sur les coefficients estimés des variables communes au premier modèle ainsi que sur leur significativité.

L'application de la démarche de sélection des variables explicatives décrite dans la section précédente nous amène à effectuer les mêmes regroupements/suppressions de variables que décrits précédemment. L'information du niveau d'encours de l'assuré et de son sexe, ou encore la distinction des partenaires "Autres", "Courtiers" et "Banque de réseau" n'apportent aucun pouvoir explicatif au modèle. Par ailleurs, l'effet de la modalité correspondant aux produits multisupports avec options apparaît ici comme n'ayant aucune différence avec celui de la modalité de référence : les produits multisupports sans options.

Quant aux variables économiques introduites, la volatilité des cours du CAC 40 décalée d'une année n'affiche pas d'effet significatif sur la réalisation d'un versement libre. Le modèle identifie néanmoins un pouvoir explicatif de cette variable prise en compte à l'année d'observation sans décalage. Cette variable reflétant l'incertitude observée sur les marchés financiers, l'exclusion de la variable  $V\_CAC_{N-1}$  a du sens. En effet, il est intuitif de penser qu'un agent basera sa décision d'effectuer un versement libre au regard des fluctuations observées sur le moment présent plutôt que sur celles observées l'année précédente. De même, bien que la prise en compte des périodes électorales ait été motivée par l'observation de chutes des taux de versements à ces périodes, le modèle n'identifie pas de pouvoir explicatif significatif pour cette variable.

Le détail des coefficients de régression estimés par le modèle est fourni en annexe. Les coefficients estimés pour les variables structurelles déjà introduites dans le premier modèle restent pour la plupart semblables, et donc les odds-ratios aussi. En effet, le premier modèle estime correctement l'impact moyen des caractéristiques des contrats, comme celui de l'âge de l'assuré, que nous avons étudié plus en détail précédemment. Cet impact moyen est alors ajusté dans la seconde régression par l'impact des variables économiques.

### Interprétation des coefficients

Les odds ratios des variables du modèle sont illustrés dans le graphique suivant, nous relevons les observations suivantes au sujet des variables économiques introduites :

- le taux d'inflation impacte négativement l'appétence des assurés à effectuer un versement libre, tandis que les autres variables influent positivement sur celle-ci ;

- *PB\_LA* est la variable économique ayant la plus grande influence sur les comportements de versements libre : pour une augmentation d'un point de l'écart de ces taux, les assurés en portefeuille auraient une probabilité 44 % plus élevée d'effectuer un versement libre.

Les odds-ratios estimés pour les variables structurelles du portefeuille restent pour la plupart du même ordre de grandeur que celles estimées dans le modèle précédent.

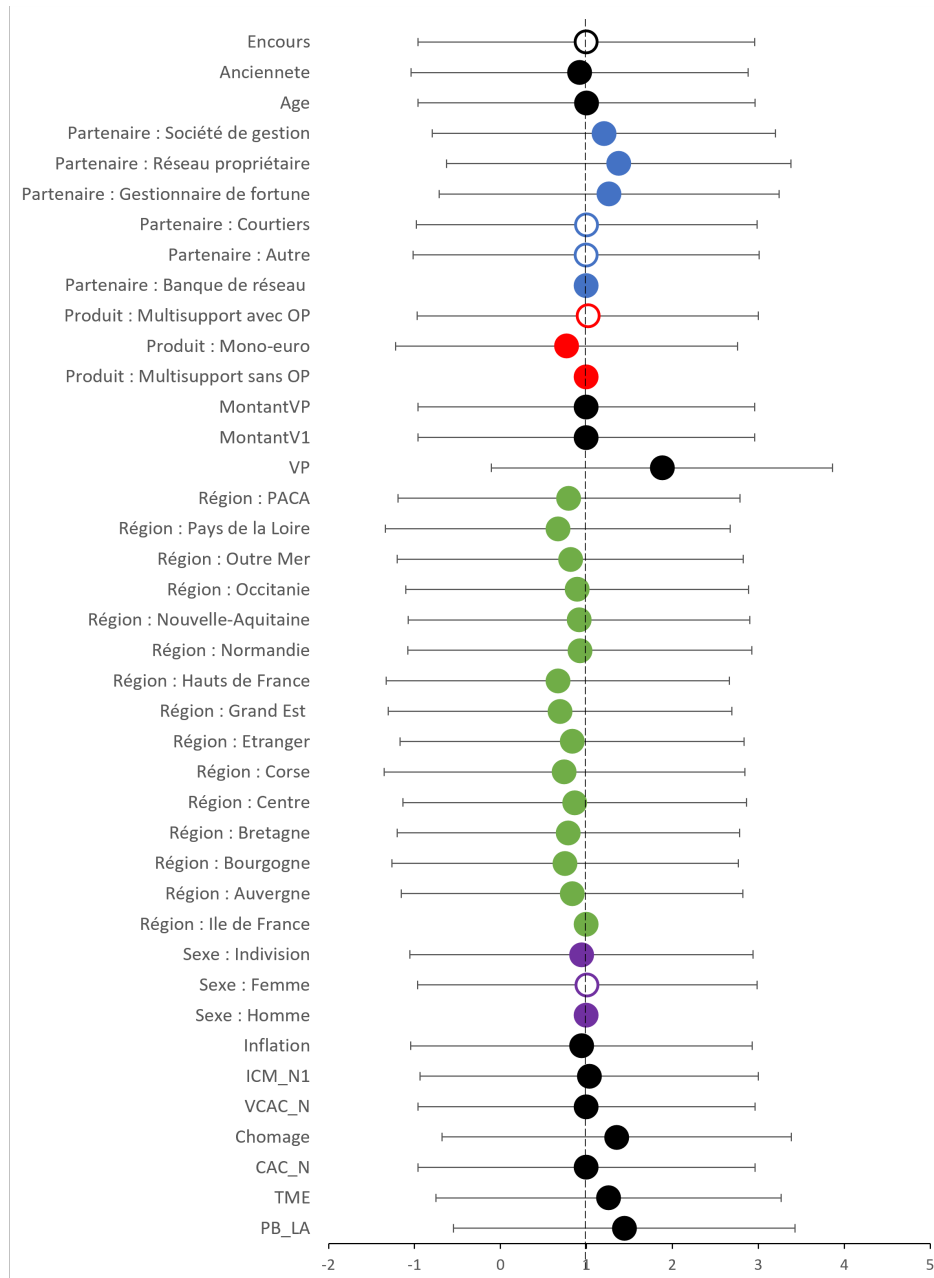


FIGURE 3.5 – Odds-Ratios - Approche dynamique

Les taux de versements approchés par ce second modèle devraient dépendre maintenant du contexte économique et donc de la période considérée. Reprenons l'exemple de l'évolution des taux de versements en fonction de l'âge des assurés, une des caractéristiques les plus influentes sur ce taux.

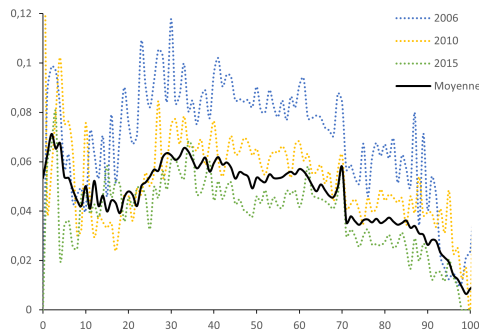


FIGURE 3.6 – Évolution des taux de versement observés de 2006 à 2016 en fonction des âges des assurés

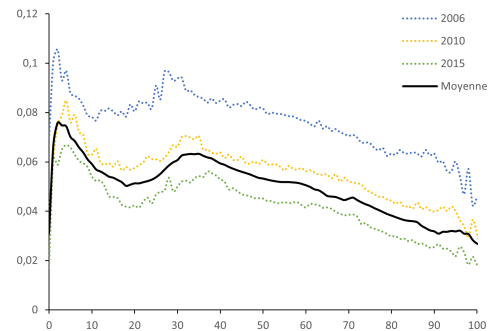


FIGURE 3.7 – Évolution des taux de versements, ajustés sur l'échantillon d'apprentissage, en fonction des âges des assurés

Le modèle implémenté précédemment estimait une distribution des taux de versement en fonction des âges qui était la même, quelle que soit l'année considérée. Ainsi, il ne considérait ainsi pas les variations temporelles de la décision d'effectuer un versement libre. Sous cette nouvelle approche dynamique, nous observons que le modèle parvient à capter les variations temporelles de la variable d'intérêt. Les graphiques suivants illustrent les évolutions en fonction des âges des assurés des taux de versements observés et ajustés par le modèle sur l'échantillon d'apprentissage.

Nous observons ainsi la similitude des moyennes de taux de versements ajustés par les modèles sous les deux approches. Cependant, l'avancée majeure que confère cette approche dynamique est celle de la prise en compte de la variabilité temporelle des taux de versements. Ainsi, même si les ajustements du modèle ne sont pas parfaits, nous retrouvons l'observation précédemment faite d'un taux de versement plus élevé sur l'année 2006 que sur l'année 2015 par exemple.

### Qualité d'ajustement du modèle

L'interprétation des coefficients estimés nous laisse alors penser que cette approche fournit de meilleurs résultats en prenant en compte les fluctuations de l'environnement économique pour expliquer la décision de versement libre. Le graphique ci-dessous illustre l'évolution des taux de versement ajustés par cette 2e régression logistique sur les mêmes échantillons d'apprentissages et de validation, distingués dans le graphique par la ligne verticale en pointillés.



FIGURE 3.8 – Taux de versements ajustés par régression logistique - Approche dynamique

Les résultats des prévisions de la régression logistique sous cette approche sont nettement meilleurs. Les taux ajustés par le modèle sont fidèles à la réalité des observations faites sur le portefeuille. Bien que les taux de versements ajustés ne soient pas exactement les mêmes que ceux observés, ils y correspondent au taux d'erreur de 95 %.

Nous observons cependant des écarts entre les prévisions sur l'échantillon de validation et les observations faites sur le portefeuille, et cela, notamment sur le taux prédit pour l'année 2018. Nous en concluons que les variables introduites dans notre modélisation ne parviennent pas totalement à expliquer la baisse du nombre de versements observée sur le portefeuille à partir de 2017. La variable *PB\_CONC* décrit à partir de 2017 une baisse des rendements servis par l'assureur détenteur du portefeuille étudié en comparaison ceux servis par ses pairs. Si cette variable n'est pas jugée significative dans la construction du modèle sur la base des informations relevées de 2006 à 2016, elle l'aurait sûrement été si nous avions choisi de calibrer notre modélisation sur l'ensemble de l'historique à disposition. À noter, par ailleurs, que le nombre de contrats en portefeuille observe une légère baisse à partir de 2017, nous faisant alors penser au lien entre rachats et versements libres. Il aurait alors été intéressant de pouvoir intégrer à notre modélisation l'ensemble des informations relatives aux opérations réalisées par les assurés, comme les rachats ou les arbitrages. Il est en effet intuitif de penser qu'un assuré venant ou prévoyant de racheter partiellement son contrat sera moins enclin à effectuer un versement libre.

Ces écarts pourraient aussi provenir du manque de précision dans la calibration des effets de la conjoncture sur la décision de versement des individus. L'utilisation de la régression logistique présente l'inconvénient d'estimer un impact global des variables explicatives introduites, sur les versements observés sur le portefeuille et non un impact propre à

chacun des individus observés. Cependant, il est possible qu'une variation d'une variable conjoncturelle, le CAC40 par exemple, n'impacte pas de la même façon la décision de versement chez les assurés du portefeuille étudié. Afin de pallier ce problème, [Milhaud \(2011\)](#) dans son étude des comportements de rachat, identifie des groupes d'individus ayant la même sensibilité à la conjoncture et présentant donc des comportements corrélés. Il utilise alors des méthodes régressions logistiques pour la modélisation des rachats, lui permettant de prendre en compte les différences de comportements parmi ces groupes. Les résultats de notre modélisation restant satisfaisants, nous n'explorerons pas cette piste. Néanmoins, il serait intéressant de mesurer l'apport de l'utilisation de telles techniques sur la modélisation de la décision de versement libre.

## 3.2 Modélisation des montants moyens de versements libres

Dans cette section, nous présentons une modélisation des montants moyens associés aux versements libres. Pour cela, nous faisons le choix de baser notre démarche sur les conclusions établies lors de l'étude de ces montants.

Du fait de la stabilité observée des montants de versements libres, nous décidons de ne pas introduire de variables conjoncturelles à la modélisation afin de la simplifier. Nous nous cantonnons ainsi aux variables représentant les caractéristiques des assurés du portefeuille, et plus particulièrement aux variables relatives au niveau d'encours et au type de produit associé au contrat. Ce choix est motivé par les résultats obtenus lors de la classification des montants moyens et présente de plus l'avantage de satisfaire ainsi les contraintes opérationnelles classiques en assurance des modèles de projection. En effet, nous avons pu observer que les deux variables sélectionnées constituent (de loin) les variables les plus discriminantes aux montants de versements libres. En misant sur cette conclusion, nous décidons de procéder simplement à la construction de tables de montants moyens de versements libres selon les valeurs des deux variables précédemment citées, pour ensuite les lisser de manière à réduire les erreurs d'échantillonnage.

De la même façon que pour les taux de versement, nous établissons nos estimations sur un échantillon d'apprentissage constitué des observations réalisées de 2006 à 2016 pour ensuite les comparer à des observations, réalisées de 2017 à 2020 et n'ayant pas servies à la construction de ces estimations.

Il est important de préciser que la démarche développée pour l'estimation du montant des versements libre se devrait d'être complétée de l'estimation de sa répartition sur les différents supports du contrat (Euro/UC). Nous avons fait le choix de ne pas nous intéresser à la projection de cette répartition, information pourtant essentielle en vue des projections des flux futurs de trésorerie, qui mériterait une étude approfondie. Nous pouvons par exemple intuitiver l'intervention d'une composante conjoncturelle dans la décision de répartition du montant d'un versement libre.

Néanmoins, nous pouvons, à l'aide de nos analyses, en établir une approximation. En effet, nous remarquons que les assurés appliquent à leurs versements libres la même répartition que celle de leur encours au moment du versement. Bien sûr, cette remarque n'est valable que pour les contrats affichant un encours non nul au 1<sup>er</sup> janvier de l'année considérée. À noter, par ailleurs, que sur une majorité de ses produits, l'assureur détenteur du portefeuille impose aux versements libres la condition d'être réparti au minimum à 30 % sur des supports UC. Nous testons alors l'approximation, pour les contrats présentant un encours nul en début d'année, d'un montant de versement placé à 70 % sur des supports euros et d'un montant de versement respectant la même répartition que l'encours pour les autres contrats. Le graphique suivant présente, par la courbe noire, l'évolution annuelle de la



FIGURE 3.9 – *Évolution annuelle des montants totaux de versements libres investis sur le support euro*

somme des versements sur le support euro. La courbe rouge quant à elle décrit la somme des versements sur le support euro calculés sur la base des montants totaux de versements et de l'approximation de leur répartition précédemment décrite.

### 3.2.1 Construction des tables

La construction de tables en fonction de l'encours des contrats est compliquée par le caractère continu de cette variable. En effet, si l'ancienneté ou l'âge sont techniquement des variables continues, leur expression en nombre d'années permet ainsi de discrétiser leurs valeurs. Le montant d'encours, quant à lui, peut par essence prendre une infinité de valeurs. Avant de procéder à la construction des tables de montants, il est alors nécessaire de discrétiser cette variable en la distinguant en intervalles.

Une alternative à la discrétisation de la variable d'encours, pour sa considération dans la construction des tables, aurait été de rapporter les montants de versements libres au niveau d'encours du contrat et ainsi estimer les proportions ainsi calculées. Cependant, cette approche serait difficilement applicable dans notre cas du fait de la présence d'affaires nouvelles dans le portefeuille étudié et donc de contrats associés à un encours nul au 1<sup>er</sup> janvier de leur première année de vie. Par ailleurs, il est intuitif de penser que les résultats d'une telle approche ne seraient fiables que dans le cas où nous observons une relation linéaire des montants de versements libres en fonction de l'encours. Dans notre cas, l'analyse de cette relation à l'échelle de notre portefeuille révèle une forme "parabolique" :

nous observons des montants moyens de versements libres élevés pour les contrats associés aux encours les plus faibles et les plus élevés, ainsi que des montants plus faibles pour les encours médians. Pour ces raisons, nous conservons la première approche détaillée.

Il existe deux méthodes "naturelles" pour la constitution de classes d'une variable continue : sa séparation en intervalles d'amplitudes égales ou en intervalles de populations égales. La première méthode présente deux principaux inconvénients. Le premier est sa sensibilité aux valeurs extrêmes des queues de distribution de la variable considérée, induisant ainsi le deuxième inconvénient en présentant la possibilité de créer des intervalles vides d'observations. La deuxième méthode, par essence, ne présente pas de tels inconvénients. Nous la choisissons pour le retraitement de la variable d'encours que nous discrétisons en 10 classes établies selon les déciles de sa distribution. En faisant l'analyse des montants moyens de versements libres associés à chacun de ces intervalles d'encours, nous retrouvons la forme parabolique de la relation entre ces deux variables.

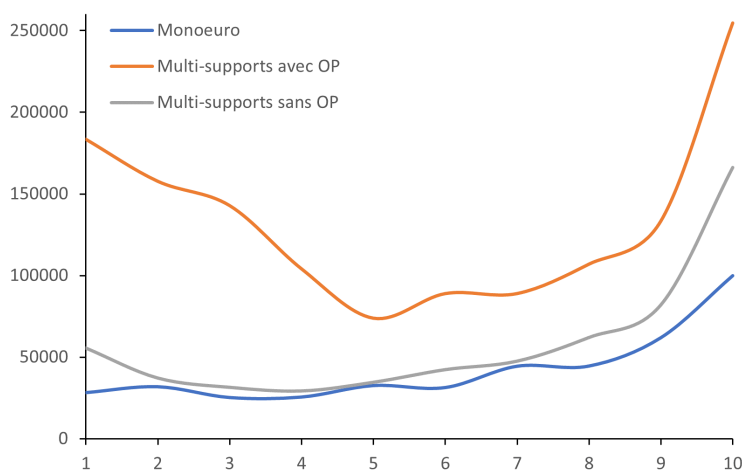


FIGURE 3.10 – Moyennes brutes des montants de versements libres en fonction du niveau d'encours et du type de produit (échantillon d'apprentissage)

Nous affinons ensuite la maille de moyennage de ces montants en incluant dans la table ainsi construite l'information du type de produit associé au contrat. Nous obtenons ainsi trois lois de projection des montants moyens de versement libres (toutes trois en fonction des classes d'encours créées), que nous lisons ensuite afin d'écartier les erreurs d'échantillonnage possiblement introduites par le calcul de ces moyennes.



### 3.2.2 Lissage des tables

La régression LOESS, développée par Cleveland (1981), est une méthode de régression non paramétrique tirant son nom de l'acronyme de "LOcally Estimated Scatterplot Smoothing". Cette méthode permet, par l'ajustement d'un nuage de points, de produire des courbes lissées de ces points. La construction de ces courbes repose sur l'étude des *k plus proches voisins* et permet ainsi le contrôle de la flexibilité de l'estimateur tout en écartant les potentielles erreurs d'échantillonnage induites par construction. Pour cela, l'ajustement de la courbe se fait localement : pour un point d'abscisse considéré, l'ajustement du point de la courbe sera établi à partir de tous les points associés au voisinage de ce point d'abscisse.

La prise en compte des points au voisinage du point considéré est pondérée par leurs distances à ce point. La taille du voisinage est quant à elle déterminée par un paramètre  $\alpha$  qui décrit la proportion de points du jeu de données qui y est inclus. L'ajustement de la courbe lissée consiste ensuite à l'ajustement local d'un polynôme de degré 1 ou 2.

Les courbes lissées des montants moyens de versement en fonction de l'encours sont représentées dans les graphiques suivants, pour les trois types de produits que présente le portefeuille :

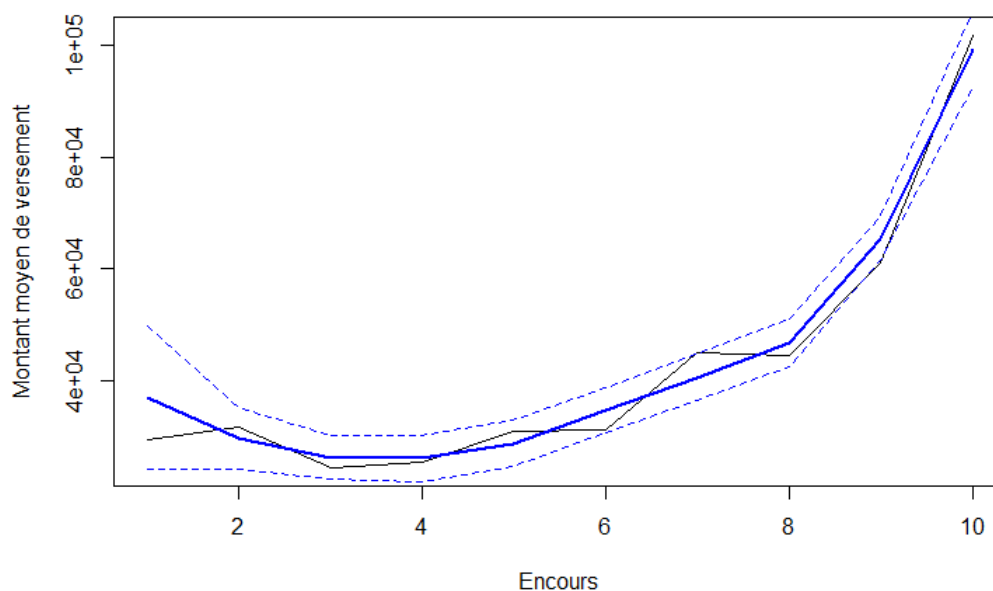


FIGURE 3.11 – Lissage des montants moyens de versements sur les produits monosupports euro, en fonction de l'encours

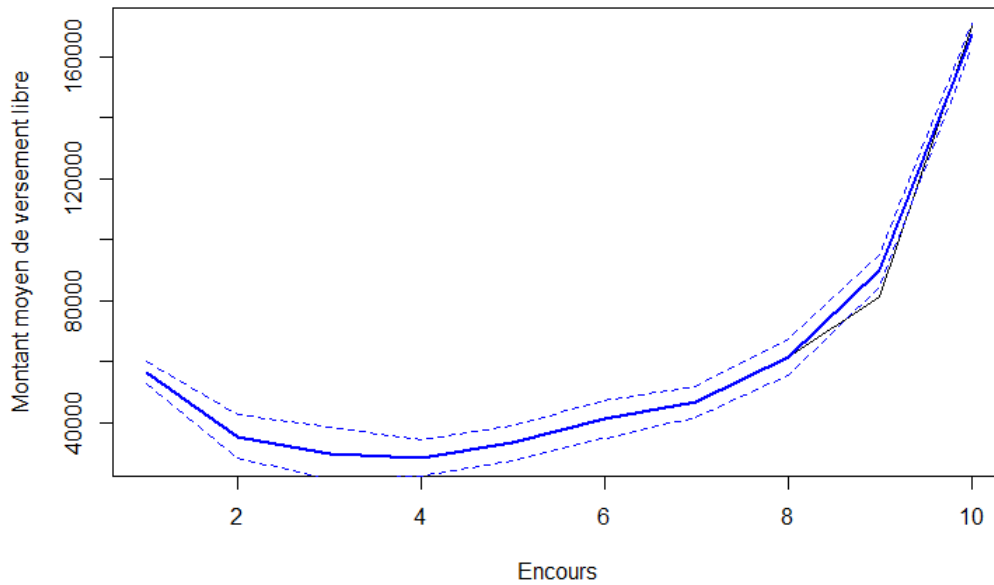


FIGURE 3.12 – Lissage des montants moyens de versements sur les produits multi-supports sans *OP*, en fonction de l'encours

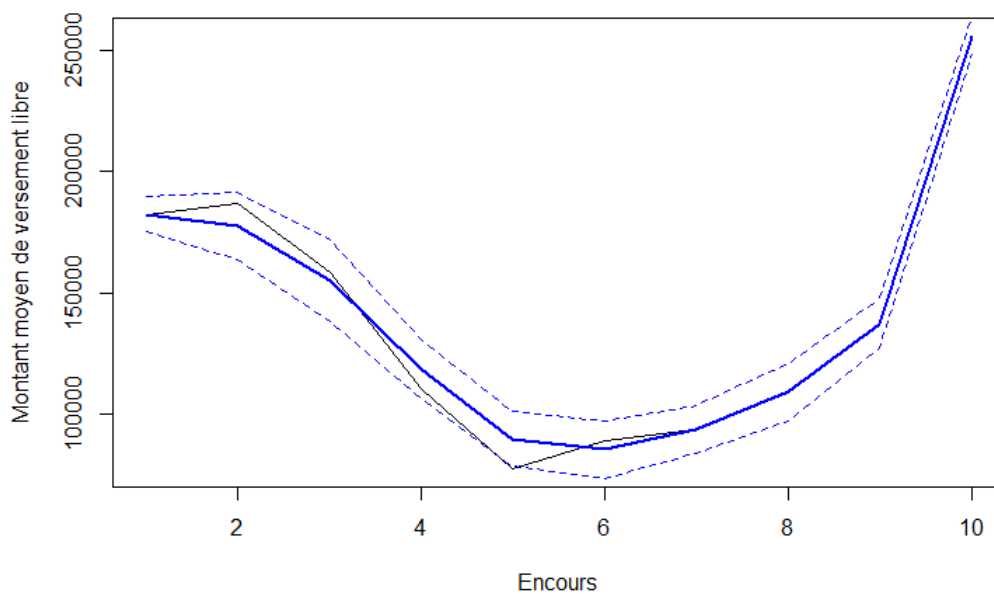


FIGURE 3.13 – Lissage des montants moyens de versements sur les produits multi-supports avec *OP*, en fonction de l'encours

Nous choisissons, pour le lissage de chacune de ces trois lois, un paramètre de lissage de 0.4 permettant à ce que la loi initialement construite se situe dans l'intervalle de confiance

à 95 % de la nouvelle loi. Nous observons des fluctuations d'échantillonnage relativement importantes pour la loi associée aux produits mono-euro. Ces fluctuations sont corrigées par la régression LOESS qui lisse ainsi la courbe afin d'éviter de futures erreurs lors de la généralisation de cette loi. Nous remarquons par ailleurs que la courbe générée pour les produits multisupports sans OP est quasiment confondue à celle initialement créée. Ce produit étant le plus représenté dans le portefeuille, la considération d'un plus grand nombre d'observations a ainsi permis d'établir une loi initiale assez adaptée à la généralisation de son utilisation.

Nous avons ainsi pu établir, à l'aide de régressions LOESS, des estimateurs des montants moyens associés aux versements libres. Cette méthode présente l'avantage d'une simplicité d'implémentation en exigeant de l'utilisateur uniquement le renseignement de l'amplitude du lissage désiré et du degré du polynôme considéré pour l'ajustement de la courbe. Néanmoins, son application demande un effort d'interprétation important, car elle ne fournit pas de formule fermée pour la construction de la courbe.

### 3.2.3 Validation du modèle

Nous évaluons ensuite la qualité des estimateurs en testant les valeurs prédites sur l'échantillon de validation constitué des observations relatives aux individus du portefeuille ayant effectué un versement libre entre 2017 et 2020. Pour cela, nous attribuons un montant à chacun de ces individus en fonction de la classe d'appartenance, de leur niveau d'encours et du type de produit auquel ils ont souscrit.

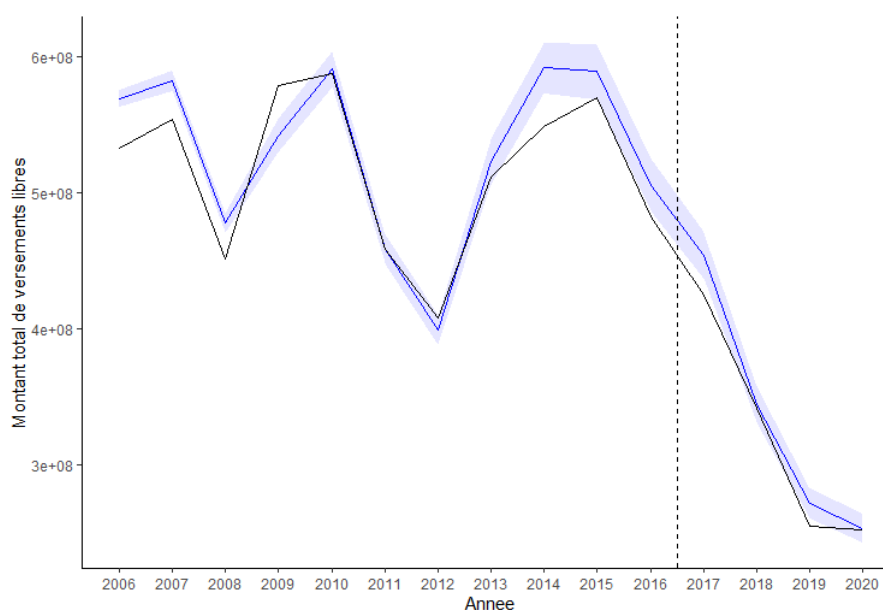


FIGURE 3.14 – Évolutions annuelles des montants totaux de versement libre et de leurs estimations

Le graphique 3.13 présente l'évolution annuelle des montants associés à l'ensemble des versements libres perçus sur le portefeuille. Comme pour la modélisation des taux de versement, la distinction des échantillons d'apprentissage et de validation est marquée par la ligne verticale en pointillée.

Les estimations établies approchent convenablement la somme des montants de versements libres observés sur le portefeuille. Nous constatons une erreur d'estimation tout à fait acceptable de 3 % en moyenne par année. Ainsi, la restriction des variables explicatives du modèle à celles de l'encours et du type de produit associé au contrat semble suffisante à la constitution d'un modèle fiable et performant. L'analyse des résultats confirme, par ailleurs, la capacité des estimateurs ainsi construits à expliquer la variabilité annuelle des montants de versements libres perçus. Sous réserve de la possession de l'information de l'existence d'un versement, l'introduction de composantes dynamiques se confirme alors non nécessaire pour l'étude des montants. Nous présentons alors, dans une ultime section, le résultat apporté par l'agrégation des modèles de prédiction des taux de versement et de leurs montants moyens.

### 3.3 Agrégation et analyse des résultats des modèles

La régression logistique, adoptée pour notre première modélisation, est une méthode permettant de prédire la classe d'appartenance d'un individu (à un phénomène binaire), par le biais de l'estimation d'une probabilité qui lui est associée. On utilise usuellement le seuil de 0.5 appliqué à cette probabilité afin de déterminer la classe qu'attribue le modèle à un individu. Cependant, dans le cas de notre étude, les probabilités prédites par le modèle sont très faibles et aucunes ne dépassent ce seuil de 0.5. Cela est dû à la rareté de l'occurrence d'un versement libre sur le portefeuille étudié. En effet, la population de non-verseurs représente, en 15 ans d'observation, 95,5 % du portefeuille. Le modèle a donc tout intérêt de classer tous les individus comme non-verseurs afin de ressortir un taux d'erreur acceptable de 4,5 %.

Nous ne pouvons ainsi pas utiliser les probabilités prédites en tant que telles pour prédire les décisions individuelles des assurés. Néanmoins, nous considérons l'estimation de la probabilité faite par le modèle comme un indicateur de score d'appétence des individus. Les résultats des prédictions nous permettent plutôt d'obtenir une estimation de la quantité de versements à attendre sur un sous échantillon d'individus ayant le même profil. C'est en ce sens que nous utilisons les termes de modélisation des taux de versements plutôt que de celle de la décision individuelle de verser.

De nombreuses méthodes existent pour la résolution des obstacles induits par l'utilisation de données déséquilibrées. Cet obstacle, plus communément connu sous le nom d'*Imbalanced Data*, a longuement été traité dans la littérature, car intervenant dans la modélisation de tout phénomène rare. Cependant, nous n'avons pas exploré la piste de ces méthodes en raison du faible intérêt d'une prédiction à l'échelle individuelle pour le marché de l'assurance. En effet, du fait des contraintes opérationnelles imposées par les modèles de projection usuellement utilisés sur le marché, les différentes lois afférentes aux contrats d'assurance se doivent d'être établies à une large maille. Pour cela, ces lois sont couramment exprimées par regroupements des contrats selon les réalisations de quelques variables seulement.

C'est ainsi que l'étude des variables portant le plus d'impact sur les comportements de versements libres révèle tout son intérêt. Elle permet, en effet, d'établir les mailles auxquelles il serait optimal de présenter les lois construites.

Nous prédisons déjà les montants moyens de versements libres à partir des réalisations de deux variables : le niveau d'encours et le type de produit associés au contrat. Par conséquent, aucune agrégation n'est à faire sur l'expression de cette loi. Nous procédons alors à l'agrégation de nos lois de projection des taux de versements libres.

### 3.3.1 Agrégation des lois de projection des taux de versements libres

Les taux de versements issus de nos modélisations sont prédits à l'échelle individuelle de notre portefeuille. Cependant, la méthode utilisée pour leur modélisation présente l'avantage de fournir une formule fermée de ces taux et de se présenter (avant transformation) comme une somme de composantes afférentes aux diverses variables explicatives utilisées. On peut ainsi considérer que les coefficients de régression estimés ne sont pas amenés à varier dans le temps et que la seule raison de modification de cette loi serait celle d'un changement des hypothèses d'entrée qui mènerait donc à la réévaluation de la formule fermée fournie. Ces hypothèses sont celles de la valeur des variables conjoncturelles utilisées pour l'estimation de ces taux ainsi que celles relatives à la composition du portefeuille. L'utilisateur de cette loi se devra donc de la mettre à jour annuellement pour prendre en compte les évolutions des variables conjoncturelles et dans, à une fréquence moindre, en cas de changement significatif de la composition du portefeuille. Nous pouvons ainsi agréger les taux de versements prédits en les moyennant à la maille choisie, afin d'établir une loi agrégée de prédiction des taux de versements libres.

L'analyse de l'influence des caractéristiques des assurés sur leurs décisions de versement libre souligne le fort pouvoir discriminant de l'ancienneté du contrat. Nous choisissons d'établir nos lois selon les valeurs prises par cette variable et nous les construisons pour les années constituant notre échantillon de validation, soit de 2017 à 2020. Nous calculons ainsi pour chaque ancienneté la moyenne des taux de versements prédits selon nos deux approches de modélisation, et cela, pour chacune des années considérées. Nous les comparons ensuite aux taux de versements observés sur ces années à la même maille. Les graphiques suivants illustrent les projections de ces lois pour les années 2017 et 2019, en comparaison aux taux réellement observés sur ces années.

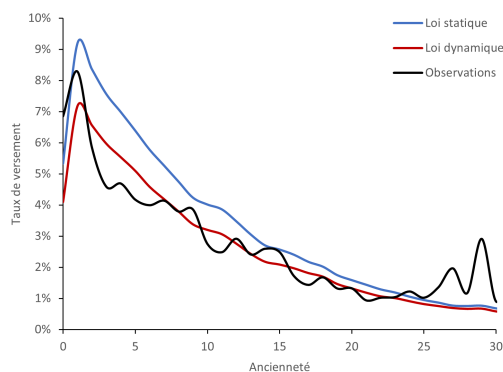


FIGURE 3.15 – *Lois de projection des taux de versements établies pour l'année 2017, en comparaison aux taux observés sur l'année*

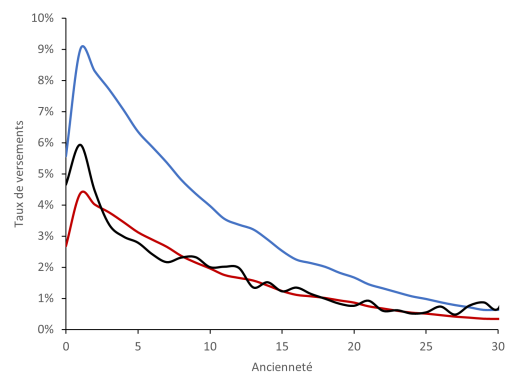


FIGURE 3.16 – *Lois de projection des taux de versements établies pour l'année 2019, en comparaison aux taux observés sur l'année*

Nous retrouvons la mauvaise qualité des projections de la loi établie selon l'approche statique. En effet, cette loi ne parvient pas à prévoir les variations temporelles des taux de versements et ainsi pas leur chute observée depuis 2016. Elle surestime alors largement le nombre de versements observé sur l'échantillon de validation et produit des erreurs allant jusqu'à 110 % pour l'année 2020.

Année	Nombre observé	Nombre estimé	Erreur
2017	4 997	6 415	+28 %
2018	3 742	6 025	+61 %
2019	2 824	5 565	+97 %
2020	2 448	5 134	+110 %

TABLE 3.1 – *Prédictions et erreurs de la loi de projection des taux de versements établie par approche statique*

Année	Nombre observé	Nombre estimé	Erreur
2017	4 997	5 115	+2 %
2018	3 742	4 445	+19 %
2019	2 824	2 746	-3 %
2020	2 448	2 563	+5 %

TABLE 3.2 – *Prédictions et erreurs de la loi de projection des taux de versements établie par approche dynamique*

La loi de projection des taux de versements établie par approche dynamique anticipe bien, quant à elle, cette chute du nombre de versements observée à partir de 2016, conformément à ce que nous avons pu constater dans les sections précédentes. C'est sans grande surprise que nous retrouvons alors l'erreur la plus élevée dans la prédiction relative à l'année 2018, qui s'établit à 19 %. Sans tenir compte de l'année 2018, l'erreur maximale d'estimation du nombre de versements s'établit à 5 % ce qui est tout à fait acceptable. Nous conservons alors cette loi de projection pour en coupler les résultats avec ceux des lois de projection des montants moyens de versement, afin d'obtenir les montants annuels des versements libres.

### 3.3.2 Projection des montants annuels de versements libres

Afin d'obtenir les projections des montants annuels de versements libres, nous associons les résultats de nos projections des taux et des montants de versements libres. Nous attribuons à chacun des individus de notre échantillon de validation (2017-2020), un taux de versement selon son ancienneté et l'année considérée. Nous leur associons de plus un montant moyen de versement selon leur niveau d'encours et le type de produit souscrit. De la même façon, nous leur associons les intervalles de confiances de chacune de ces deux estimations. Nous multiplions enfin chacune de ces deux valeurs attribuées, que nous sommions sur l'année pour obtenir les résultats suivants :

Année	Montant observé	Montant estimé	Erreur	Int. de conf. à 95 %
2017	425	441	4 %	[ 398 ; 485 ]
2018	342	395	16 %	[ 350 ; 443 ]
2019	255	247	-3 %	[ 205 ; 292 ]
2020	253	241	-5 %	[ 200 ; 285 ]

TABLE 3.3 – *Prédictions et erreurs des montants totaux de versements libres (exprimés en milliards d'euros)*

Le couplage des deux lois de projection fournit finalement des prédictions satisfaisantes. Nous retrouvons une erreur élevée de 16 % dans la projection de l'année 2018, en raison de la surestimation des taux de versements de cette année. Cependant, la plus haute erreur sur les trois années restantes s'établit à une sous estimation raisonnable de 5 % du montant total des versements libres.

Les montants annuels observés sont quasiment tous compris dans les intervalles de confiance à 95 % de nos projections, dénotant alors de la bonne qualité des prédictions de nos modèles. Nous constatons cependant la largeur de ces intervalles (environ 83 milliards d'euros) qui révèle la volatilité de nos prédictions.



# Conclusion

Ce mémoire développe une méthodologie détaillée de la construction de lois de projection des comportements de versements libres, distinguant la modélisation de son occurrence et des montants versés. La construction de ces lois repose sur l'utilisation conjointe de méthodes d'apprentissage statistique et d'estimateurs classiques de lois d'expérience. La démarche de modélisation entreprise s'appuie sur une étude, au préalable, de ces comportements peu étudiés dans la littérature jusqu'à présent.

Les premières analyses nous ont permis de constater la rareté d'occurrence des versements libres à l'échelle du portefeuille étudié. Par le biais d'analyses statistiques complétées par l'utilisation d'arbres de classification, nous avons pu identifier le marqueur essentiel à cette décision : l'ancienneté du contrat. De plus, nous avons pu confirmer nos intuitions en observant une forte corrélation des montants associés à ces versements libres avec les niveaux d'encours, mesure de richesse des assurés. Enfin, nous avons pu identifier le fort lien entre la conjoncture économique et les taux de versements libres observés sur le portefeuille. Nous relevons ainsi une composante conjoncturelle dans les observations du portefeuille, à l'image de ce que l'on retrouve dans les études des comportements de rachat.

La modélisation des taux de versements à partir uniquement des caractéristiques des assurés du portefeuille ne fournit alors pas de résultats concluants. En effet, le modèle ainsi implémenté ne parvient pas, à partir des informations à disposition, à expliquer la variabilité temporelle de ce phénomène. L'introduction de variables conjoncturelles, telles que le TME ou encore le taux de chômage, permet alors d'expliquer ces variations et fournit ainsi des projections plus adaptées aux observations. Néanmoins, la méthode employée ne permet d'estimer qu'un effet global au portefeuille de ces variables conjoncturelles. En réalité, les assurés pourraient avoir des aversions différentes au risque ou encore des cultures financières différentes, induisant ainsi des sensibilités financières différentes aux variations du climat économique. La modélisation proposée dans ce mémoire ne permet pas de tenir compte des différences d'effets des variables sur chacun et présente alors ses limites. Une piste pour leur prise en compte serait, par exemple, l'utilisation de mélanges de régression logistiques.

La modélisation des montants moyens de versements libres, quant à elle, permet de fournir des résultats concluants sur la base de méthodes simples. L'association finale des projections révèle finalement une forte volatilité des estimations du montant total de versements libres. Il est important de porter une attention particulière à cette volatilité, engendrant un risque sur le calcul des grandeurs du passif faisant intervenir ces projections. Il serait alors pertinent d'observer la robustesse d'autres modèles pour la projection des versements libres.

Pour les travaux futurs, il serait intéressant d'étudier les comportements de versements libres à une échelle mensuelle. Une étude à cette échelle pourrait permettre de mieux calibrer leurs composantes conjoncturelles. À voir, toutefois du fait de la rareté de ces phénomènes, si l'affinage de leur maille d'observation ne complexifierait pas leur modélisation. Nous avons pu présenter, par ailleurs, les limites de notre étude avec qui n'aborde pas la modélisation des montants "extrêmes" des versements libres (particularité, sans doute, de notre portefeuille) ou encore celle de leur répartition selon les supports d'investissement.

# Bibliographie

- ACPR (2016), ‘Les différentes composantes de l’assurance vie et leur évolution’, *Analyses et synthèses n°65*.
- ACPR (2021a), ‘Le marché de l’assurance vie pendant la crise sanitaire’, *Analyses et synthèses n°121*.
- ACPR (2021b), ‘Revalorisation 2020 des contrats d’assurance-vie et de capitalisation – engagements à dominante épargne et retraite individuelle’, *Analyses et synthèses n°126*.
- Antonin, C. (2019), ‘The links between saving rates, income and uncertainty : An analysis based on the 2011 household budget survey’, *Economie et Statistique / Economics and Statistics*, 513, 47–68.
- Assaraf, K. (2020), Modélisation de versements libres sous IFRS 17 par des méthodes de machine learning, Mémoire d’actuariat, Université Paris-Dauphine, Paris.
- Banque de France (2022), ‘Les indices obligataires’, <https://www.banque-france.fr/statistiques/taux-et-cours/les-indices-obligataires>.
- Benabdelkrim, F. Z. (2017), Modélisation des versements libres en assurance-vie : utilisation de méthodes de scoring, Mémoire d’actuariat, ISUP, Sorbonne Université, Paris.
- Brugirard, X. (2012), Les rachats dynamiques : comportement des assurés en période de crise, nouvelle modélisation et impact financier pour l’assureur, Mémoire d’actuariat, ISUP, Sorbonne Université, Paris.
- Cleveland, W. S. (1981), ‘Lowess : A program for smoothing scatterplots by robust locally weighted regression’, *Journal of the American Statistical Association*, col. 83, n°403.
- Commission Européenne (2014), ‘Règlement délégué 2015/35/ce’.
- de France, B. (2022), ‘Taux de rémunération des dépôts bancaires’, <https://www.banque-france.fr/statistiques/epargne-et-comptes-nationaux-financiers/lepargne-dans-les-banques/taux-de-remuneration-des-depots-bancaires>.
- FFA (2021), ‘Chiffres clés de l’assurance française 2020’, <https://www.franceassureurs.fr/wp-content/uploads/VF-Donnees-cles-2020.pdf>.
- FFA (2022), ‘2021, une année record pour l’assurance vie’, [https://www.franceassureurs.fr/wp-content/uploads/220128\\_FRANCE\\_ASSUREURS\\_CP-VIE\\_DECEMBRE.pdf](https://www.franceassureurs.fr/wp-content/uploads/220128_FRANCE_ASSUREURS_CP-VIE_DECEMBRE.pdf).

- Hosmer D.W., L. S. (2000), 'Applied logistic regression', *Wiley Series in Probability and Mathematical Statistics* .
- IASB (2017), 'IFRS 17 Insurance Contracts'.
- INSEE (2021a), 'Chômage au sens du bit et indicateurs sur le marché du travail (résultats de l'enquête emploi) - quatrième trimestre 2021', <https://www.insee.fr/fr/statistiques/6051733>.
- INSEE (2021b), 'Enquête mensuelle de conjoncture auprès des ménages', <https://www.insee.fr/fr/statistiques/6016218>.
- INSEE (2021c), 'Taux d'inflation, données annuelles de 1991 à 2021', <https://www.insee.fr/fr/statistiques/2122401>.
- KABA, S. I. (2017), IFRS 17 : Intégration des versements libres dans la frontière des contrats, Mémoire d'actuariat, ISFA, Université Claude Bernard Lyon 1, Lyon.
- Le Cercle des Épargnants (2021), 'Les français, l'épargne et la retraite - baromètre 2021'.
- Le Figaro (2017), 'Les français boudent toujours l'assurance vie', <https://www.lefigaro.fr/assurance/2017/05/24/05005-20170524ARTFIG00216-les-francais-boudent-toujours-l-assurance-vie.php>.
- Leo Breiman, Jerome H. Friedman, R. A. O. C. J. S. (1984), 'Classification and regression trees', *Wadsworth Books* .
- Milhaud, X. (2011), Segmentation et modélisation des comportements de rachat en assurance vie, Mémoire d'actuariat, ISFA, Université Claude Bernard Lyon 1, Lyon.
- Planchet, F. (2021), 'Statistique des modèles paramétriques et semi-paramétriques'.
- Rakotomalala, R. (2017), 'Pratique de la régression logistique'.
- S. Fallos et al. (2018), 'Ifrs 17- point d'actualités et présentation des travaux du gt ifrs 17', *Institut des Actuaïres* .
- Saugnier, G. (2017), Application de la future norme IFRS Contrats d'assurance à un portefeuille de contrats d'Epargne Euro, Mémoire d'actuariat, CNAM, Paris.
- Thérond, P. (2017), 'IFRS 17 : Principes et enjeux opérationnels', *Institut des Actuaïres* .

# Annexes

## Approche dynamique - Sélection des variables explicatives

	Modalités	Coefficient estimé	Ecart type	p-value du test de Wald
$\beta_0$		-1,04E+01	7,16E-01	< 2e-16 ***
PB_LA		3,64E-01	1,39E-02	< 2e-16 ***
TME		2,30E-01	2,37E-02	< 2e-16 ***
CAC_N		1,65E-04	9,82E-06	< 2e-16 ***
Chomage		3,02E-01	3,68E-02	2,48E-16 ***
VCAC_N		6,23E-04	5,82E-05	< 2e-16 ***
ICM_N1		3,23E-02	3,50E-03	< 2e-16 ***
Inflation		-6,02E-02	1,33E-02	5,62E-06 ***
Sexe	Femme	1,02E-02	7,53E-03	0,175679
	Indivision	-5,86E-02	1,79E-02	0,001088 **
Région	Auvergne	-1,83E-01	1,30E-02	< 2e-16 ***
	Bourgogne	-2,86E-01	2,78E-02	< 2e-16 ***
	Bretagne	-2,33E-01	1,61E-02	< 2e-16 ***
	Centre	-1,47E-01	2,02E-02	3,25E-13 ***
	Corse	-2,94E-01	6,79E-02	1,45E-05 ***
	Etranger	-1,80E-01	2,06E-02	< 2e-16 ***
	Grand Est	-3,65E-01	1,97E-02	< 2e-16 ***
	Hauts de France	-4,02E-01	1,86E-02	< 2e-16 ***
	Normandie	-7,94E-02	2,09E-02	0,000146 ***
	Nouvelle-Aquitaine	-9,21E-02	1,36E-02	1,42E-11 ***
	Occitanie	-1,16E-01	1,71E-02	1,45E-11 ***
	Outre Mer	-2,08E-01	2,60E-02	1,18E-15 ***
Pays de la Loire	-4,06E-01	2,29E-02	< 2e-16 ***	
PACA	-2,27E-01	1,45E-02	< 2e-16 ***	
VP		6,31E-01	1,17E-02	< 2e-16 ***
MontantV1		2,59E-07	2,39E-08	< 2e-16 ***
MontantVP		8,52E-06	1,19E-06	8,86E-13 ***
Produit	Mono-euro	-2,62E-01	1,54E-02	< 2e-16 ***
	Multisupport avec OP	1,65E-02	1,17E-02	0,159722
Partenaire	Autre	-2,62E-03	2,71E-02	0,923054
	Courtiers	4,54E-03	1,04E-02	0,663663
	Gestionnaire de fortune	2,35E-01	9,91E-03	< 2e-16 ***
	Réseau propriétaire	3,19E-01	2,23E-02	< 2e-16 ***
	Société de gestion	1,85E-01	1,74E-02	< 2e-16 ***
Age		1,77E-03	2,31E-04	1,58E-14 ***
Anciennete		-8,33E-02	8,46E-04	< 2e-16 ***
Encours		3,42E-09	8,13E-09	0,67355

FIGURE 17 – Résultats de l'estimation des coefficients de régression - Approche dynamique

# Liste des tableaux

1.1	<i>Description de la table du portefeuille de contrats . . . . .</i>	25
1.2	<i>Description des tables d'encours . . . . .</i>	26
1.3	<i>Description des tables de versement . . . . .</i>	26
1.4	<i>Profil moyen des assurés du portefeuille par an . . . . .</i>	31
3.1	<i>Prédictions et erreurs de la loi de projection des taux de versements établie par approche statique . . . . .</i>	83
3.2	<i>Prédictions et erreurs de la loi de projection des taux de versements établie par approche dynamique . . . . .</i>	83
3.3	<i>Prédictions et erreurs des montants totaux de versements libres (exprimés en milliards d'euros) . . . . .</i>	84

# Table des figures

1.1	<i>Rendements des supports euros (Source : FFA)</i>	16
1.2	<i>Performance des supports en unités de compte (Source : FFA)</i>	17
1.3	Valeur du PFL en fonction de l'ancienneté du contrat (Source : impots.gouv.fr)	18
1.4	<i>Démarche de construction de la base de données utilisée</i>	28
1.5	<i>Déciles des montants de versements libres observés (échelle logarithmique)</i>	29
1.6	<i>Évolution du nombre de contrats en portefeuille de 2006 à 2020</i>	30
1.7	<i>Évolution des encours au 1e janvier de 2006 à 2020</i>	30
1.8	<i>Évolution des taux de chargements moyens observés sur les montants de versements libres versés</i>	32
2.1	<i>Évolution annuelle de la proportion de verseurs observée</i>	35
2.2	<i>Évolution annuelle des montants moyens de versements libres observés</i>	35
2.3	<i>Évolution de la proportion de verseurs observée en fonction de l'âge</i>	36
2.4	<i>Évolution des montants moyens de versement libre en fonction de l'âge</i>	37
2.5	<i>Évolution de la proportion de verseurs observée fonction de l'ancienneté des contrats</i>	38
2.6	<i>Évolution des montants moyens de versement libre en fonction de l'ancienneté des contrats</i>	38
2.7	<i>Évolution de la proportion de verseurs en fonction des niveaux d'encours</i>	39
2.8	<i>Évolution des montants moyens de versement libre en fonction des niveaux d'encours</i>	39
2.9	<i>Évolution de la proportion de verseurs observée en fonction du type de produit</i>	40
2.10	<i>Évolution des montants moyens de versements libres en fonction du type de produit</i>	40
2.11	<i>Évolution de la proportion de verseurs observée en fonction du type de partenaire</i>	41

2.12	<i>Évolution des montants moyens de versements libres en fonction du type de partenaire</i> . . . . .	41
2.13	<i>Évolution des montants moyens versés en fonction du sexe de l'assuré</i> . . .	41
2.14	<i>Évolution des proportions de verseurs en fonction de la présence de versements programmés</i> . . . . .	42
2.15	<i>Évolution des montants moyens versés en fonction de la présence de versements programmés</i> . . . . .	42
2.16	<i>Exemple d'arbre de classification</i> . . . . .	44
2.17	<i>Arbre final de classification de la variable binaire décrivant l'occurrence d'un versement libre</i> . . . . .	48
2.18	<i>Importance des variables explicatives dans la classification de la variable binaire décrivant l'occurrence d'un versement libre</i> . . . . .	49
2.19	<i>Arbre final de classification des montants de versements libres</i> . . . . .	50
2.20	<i>Importance des variables explicatives dans la classification des montants de versements libres</i> . . . . .	51
2.21	<i>Évolution des taux de versements annuels</i> . . . . .	52
2.22	<i>Évolution des taux de versements en comparaison à celle du CAC40</i> . . . . .	53
2.23	<i>Évolution des taux de versements en comparaison à celle de l'ICM</i> . . . . .	54
2.24	<i>Évolution des taux de versements en comparaison à celle de l'inflation</i> . . .	54
2.25	<i>Évolution des taux de versements en comparaison à celle du taux de chômage</i>	55
2.26	<i>Évolution des taux de versements en comparaison à celle du TME</i> . . . . .	56
2.27	<i>Évolution des taux de versements en comparaison à celle de l'écart de rendement des supports Euros du portefeuille et ceux observés sur le marché</i>	56
2.28	<i>Évolution des taux de versements en comparaison à celle de l'écart de rendement des supports Euros du portefeuille et de celui du Livret A</i> . . . . .	57
3.1	<i>Résultats de l'estimation des coefficients de régression - Approche statique</i> .	64
3.2	<i>Odds-Ratios - Approche statique</i> . . . . .	65
3.3	<i>Taux de versements observés ajustés par le modèle en fonction de l'âge</i> . . .	66
3.4	<i>Taux de versements ajustés via une régression logistique - Approche statique</i>	67
3.5	<i>Odds-Ratios - Approche dynamique</i> . . . . .	70
3.6	<i>Évolution des taux de versement observés de 2006 à 2016 en fonction des âges des assurés</i> . . . . .	71
3.7	<i>Évolution des taux de versements, ajustés sur l'échantillon d'apprentissage, en fonction des âges des assurés</i> . . . . .	71



---

3.8	<i>Taux de versements ajustés par régression logistique - Approche dynamique</i>	72
3.9	<i>Évolution annuelle des montants totaux de versements libres investis sur le support euro . . . . .</i>	75
3.10	<i>Moyennes brutes des montants de versements libres en fonction du niveau d'encours et du type de produit (échantillon d'apprentissage) . . . . .</i>	76
3.11	<i>Lissage des montants moyens de versements sur les produits monosupports euro, en fonction de l'encours . . . . .</i>	77
3.12	<i>Lissage des montants moyens de versements sur les produits multi-supports sans OP, en fonction de l'encours . . . . .</i>	78
3.13	<i>Lissage des montants moyens de versements sur les produits multi-supports avec OP, en fonction de l'encours . . . . .</i>	78
3.14	<i>Évolutions annuelles des montants totaux de versement libre et de leurs estimations . . . . .</i>	79
3.15	<i>Lois de projection des taux de versements établies pour l'année 2017, en comparaison aux taux observés sur l'année . . . . .</i>	82
3.16	<i>Lois de projection des taux de versements établies pour l'année 2019, en comparaison aux taux observés sur l'année . . . . .</i>	82
17	<i>Résultats de l'estimation des coefficients de régression - Approche dynamique</i>	89