

Mémoire présenté le :

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

Par : KONE ZIE SOULEYMANE

Titre OPTIMISATION DE LA CONSTRUCTION DE MODEL POINTS  
SUR LES DONNEES DU PASSIF

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

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

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

signature

*Entreprise :*

Nom : GPM

Signature : Yannick NANDJOU 

*Directeur de mémoire en entreprise :*

Nom : YANNICK NANDJOU

Signature :

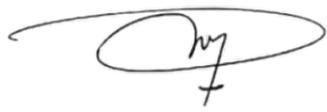
Invité :

Nom :

Signature :

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

Signature du responsable entreprise



Signature du candidat



## REMERCIEMENTS

Je souhaite remercier en tout premier lieu mon tuteur de mémoire Yannick Nandjou, Responsable du département actuariat au sein du Groupe Pasteur Mutualité, de m'avoir permis d'effectuer cette alternance et guidé vers ce sujet de mémoire intéressant. Son encadrement, son regard critique et son dynamisme m'a été précieux tout au long de mes recherches et de la rédaction de mon mémoire.

Dans ces remerciements, j'accorde une mention particulière à Omar ZAOUÏ et Yahia Salhi, respectivement actuaire responsable de la branche GPMA assurance vie et tuteur académique maître de conférence à l'ISFA, qui m'ont accompagné tout au long de ce mémoire en me prodiguant des conseils de la plus grande qualité.

Un merci particulier aussi à Charles CAZIER directeur du pôle actuariat pour ses conseils avisés, sa confiance lors de la proposition du sujet.

Un merci également aux membres de l'équipe actuariat de GPM pour leur accueil, leur sympathie et leur aide pendant cette alternance. Je citerai, Chantal Malika traoré, Amine Felah, Paul Roulland pour leurs conseils et soutiens lors de cette expérience enrichissante.

Enfin, je tiens aussi à exprimer ma gratitude à ma famille, mes amis et en particulier à mes parents, leur soutien et leurs encouragements m'auront été précieux tout au long de la réalisation de ce mémoire.

# RÉSUMÉ

**Mots clés :** Model Points - GSE - Solvabilité II - Classification Ascendante Hiérarchique (Cah) - K mean - Best Estimate (BE) - ALM - Déterministe - Stochastique - Clustering

La réglementation Solvabilité II a entraîné de nombreuses mutations au sein des compagnies d'assurance. S'articulant autour de 3 piliers, elle régule le processus de valorisation du bilan des compagnies tout en imposant des indicateurs essentiels pour la valorisation et la solvabilité des entreprises. Pour être en accord avec la réglementation, les assureurs développent des modèles qui interviennent dans tout le processus interne.

Les modèles en ALM sont ainsi devenus une préoccupation première de toute compagnie d'assurance (vie). Ils permettent la projection des actifs et des passifs à travers un générateur de scénarios économique sur plusieurs années avec plusieurs hypothèses. Ces modèles bien que souvent très précis se heurtent au problème de stockage et de temps de calcul.

Les assureurs ont ainsi recours à l'agrégation de donnée en groupe homogène appelé « Model Point ». Le but de ce mémoire est d'étudier le processus de constructions de ces Model Points, sur les données du passif de l'entreprise, pour proposer une méthode optimale.

La plupart du temps l'agrégation se fait par statistiques comme la moyenne ou la moyenne pondérée sur des variables discriminantes préalablement choisies (âge, ancienneté...). Cette méthode dite « traditionnelle » bien que simple peut manquer d'efficacité lors des lancements de modèles. Des méthodes de classifications algorithmiques permettent de répondre à ce genre de problématique. Elle apporte également une vision d'ensemble et des principes mathématiques qui permettent de mieux cerner le problème.

Toutes ces méthodes présentent des avantages et des inconvénients. La solution évoquée dans ce mémoire permet la combinaison des avantages cumulatifs de ces modèles sur notre portefeuille étudié.

Pour réaliser cette étude, un cadre a été fixé pour décrire le portefeuille d'étude ainsi que le modèle utilisé pour déterminer nos indicateurs solvabilité 2. Cette présentation a permis de circonscrire la problématique en présentant les sorties du modèle ALM pour soulever les critères de validité de la construction des Model Points.

Il a finalement été possible de produire des résultats convergents et précis qui sont cohérents avec les critères de validité établis. L'étude a montré aussi que les modèles réagissaient de façon différente en fonction du type de produit et les indicateurs cibles.

Le backtesting réalisé sur les données 2020 ont enfin validé notre méthode de construction en soulevant quelques limites. En effet, les évolutions de marché comme le rendement des actifs peuvent avoir un impact sur la précision des model points.

# ABSTRACT

**Key Words** : Model Points - ESG - Solvability II - Hierarchical Ascending Classification (HAC) - K mean - Best Estimate (BE) - ALM - deterministic - Stochastic - Clustering

The Solvency II regulation has led to many changes within insurance companies. Built around three pillars, it regulates the process of valuing companies' balance sheets while imposing key indicators for companies' valuation and solvency. In order to comply with the regulations, insurers develop models that intervene in the entire internal process.

The ALM models have thus become a primary concern of any (life) insurance company. They allow the projection of assets and liabilities through a Generator of economic scenarios over several years with several hypotheses. These models, although often very precise, face the problem of storage and calculation time.

Insurers use the homogeneous group data aggregation known as the Model Point. The purpose of this dissertation is to study the construction process of these Model Points, on the company's liability data, to propose an optimal method.

Most of the time the aggregation is done by statistics such as the average or weighted average on discriminating variables chosen beforehand (age, seniority, etc.). This so-called "traditional" but simple method may not be effective when launching models. Algorithmic classification methods are used to address this kind of problem. It also provides an overall vision and mathematical principles to better understand the problem.

All these methods have advantages and disadvantages. The solution mentioned in this brief allows the combination of the cumulative advantages of these models on our studied portfolio.

To conduct this study, a framework has been established to describe the study portfolio as well as the model used to determine our solvency indicators 2. This presentation made it possible to define the problem by presenting the outputs of the ALM model to raise the validity criteria of the Construction of Model Points.

It was ultimately possible to produce consistent and accurate results that are consistent with established validity criteria. The study also showed that the models reacted differently depending on the type of product and the target indicators.

The backtesting carried out on the 2020 data finally validated our construction method by raising some limits. Indeed, market developments such as the return on assets can have an impact on model point accuracy.

# SOMMAIRE

Table des figures	8
Table des Tableaux	9
INTRODUCTION	11
<b>I REGLEMENTATION ET CONTEXTE</b>	<b>13</b>
1 PRÉSENTATION CADRE D'ETUDES	14
1.1 Présentation GPM	14
1.2 Contexte et cadre d'étude	15
1.2.1 SOUSCRIPTION ET RENONCIATION	16
1.2.2 COTISATIONS	16
1.2.3 LES DIFFÉRENTS TYPES DE SUPPORT	16
1.2.4 LA REVALORISATION DU CONTRAT	17
1.2.5 LES POSSIBILITÉS DE RACHAT	18
1.3 Présentation des produits	18
1.3.1 ÉPARGNE	18
1.3.2 RENTE	19
2 RÈGLEMENTATION : SOLVABILITE 2	20
2.1 Présentation de Solvabilité II	20
2.2 Best Estimate	21
2.3 La marge pour risque (RM)	22
2.4 Solvency Capital Requirement (SCR)	23
3 ALM	25
3.1 La nécessité d'une modélisation ALM	25
3.2 Le modèle ALM	26
4 GSE	27
4.1 Généralités	27
4.2 Litterature sur les GSE	28
4.3 Présentation du GSE risque neutre GPM	30
4.4 Règlementation	32
5 MODEL POINT	34
5.1 Exigences règlementaires	34
5.2 Qualité de données	35
5.3 Méthode de constitution des model points actifs	36
5.4 Méthode de constitution des model points passifs	36
5.5 Tester la validité des model points	37

<b>II</b>	<b>ALM MODEL POINT</b>	<b>38</b>
1	CADRE D'ÉTUDE POUR LE LANCEMENT . . . . .	39
1.1	CALFITEC : modèle de lancement ALM . . . . .	39
1.1.1	Indicateurs solvabilité II sous CALFITEC . . . . .	39
1.1.2	Préparation des Inputs . . . . .	40
1.1.3	Principaux OUTPUTS de CALFITEC . . . . .	40
1.1.4	Modélisation CALFITEC . . . . .	41
1.2	Présentation des données brutes . . . . .	42
1.2.1	ÉPARGNE . . . . .	42
1.2.2	RENTE . . . . .	46
1.2.3	LANCEMENT INITIAL . . . . .	48
2	MODEL POINT . . . . .	49
2.1	Approche générale . . . . .	49
2.2	Introduction au clustering . . . . .	50
2.3	Approche non supervisée : Cah . . . . .	51
2.4	Approche non supervisée : Kmean . . . . .	52
2.5	Approche supervisée : CART . . . . .	54
3	CONSTRUCTION DES MP . . . . .	56
3.1	méthode générale (standard) . . . . .	56
3.2	Procédure de construction . . . . .	57
3.3	Épargne . . . . .	57
3.3.1	CART . . . . .	58
3.3.2	Cah/Kmean . . . . .	59
3.4	Rente . . . . .	60
3.4.1	Cart . . . . .	62
3.4.2	Cah . . . . .	62
3.4.3	Kmean . . . . .	62
4	ANALYSE ET COMPARAISON DETERMINISTE . . . . .	64
4.1	Épargne . . . . .	64
4.1.1	Lancement principal . . . . .	64
4.1.2	Lancements des chocs . . . . .	64
4.1.3	Analyse de trajectoires . . . . .	66
4.1.4	Conclusion . . . . .	66
4.2	Rente . . . . .	66
4.2.1	Lancement principal . . . . .	66
4.2.2	Lancements des chocs . . . . .	67
4.2.3	Analyse . . . . .	68
4.2.4	Conclusion . . . . .	68
5	BE STOCHASTIQUE . . . . .	68
5.1	Nécessité de la projection stochastique . . . . .	68
5.2	Estimation du BE stochastique . . . . .	70
<b>III</b>	<b>BACKTESTING</b>	<b>71</b>
1	Définition et statistique . . . . .	72
2	Épargne . . . . .	72
2.1	Comparaison déterministe . . . . .	72
2.2	Analyse . . . . .	73
3	Rente . . . . .	74
3.1	Comparaison déterministe . . . . .	74
3.2	Analyse des trajectoires . . . . .	74
4	Lancement stochastique . . . . .	75

**ANNEXE**

1	Annexe A : Notions importantes de clustering . . . . .	80
1.1	Relation et classes d'équivalence . . . . .	80
1.1.1	Similarité, dissimilarité et distance . . . . .	80
1.2	Hierarchies, partitions... . . . .	81
2	Annexe B : Classification hiérarchique . . . . .	83
2.1	Distances et dendogrammes . . . . .	83
2.2	Distances et variance . . . . .	83

# Table des figures

1	Bilan Solvabilité II . . . . .	20
2	Synthèse de l'analyse des risques . . . . .	23
3	Structure modulaire du SCR selon le QIS 5 . . . . .	24
4	Décomposition du SCR . . . . .	25
5	Bilan GPM . . . . .	26
6	Schema du modèle ALM. . . . .	27
7	Schéma du modèle de Wilkie . . . . .	29
8	Modèle de Brennan et Xia . . . . .	30
9	Schéma du modèle Ahlgrim et AL . . . . .	30
10	Processus ALM Calfitec . . . . .	39
11	Répartition des Adhérents par Age . . . . .	44
12	Répartition de la PM par Age . . . . .	44
13	Répartition de la PM pondérée par Ancienneté . . . . .	45
14	Répartition des Adhérents par sexe . . . . .	45
15	Répartition du montant de rente et de la PM par type de cotisation . . . . .	47
16	Répartition du montant de rente . . . . .	47
17	Répartition des Adhérents par Age . . . . .	48
18	Répartition des Adhérents par sexe . . . . .	48
19	Illustration du dendrogramme obtenu en appliquant la classification ascendante hiérarchique sur 10 individus indexés par $i_{1 \leq i \leq 10}$ . . . . .	52
20	Illustration de la méthode des Kmean . . . . .	53
21	Exemple de classification par la méthode CART . . . . .	54
22	Quel est l'arbre optimal ? . . . . .	55
23	Étape de construction des Model Points . . . . .	57
24	élagage par plotcp . . . . .	58
25	Arbre de décision de la classe 1 (sexe=H) . . . . .	59
26	Critère de décision . . . . .	60
27	partitionnement avec la méthode Kmean en 5 clusters . . . . .	60
28	Première partition Rente . . . . .	61
29	modèle CART (classe 2) . . . . .	62
30	modèle Cah (classe 2) . . . . .	62
31	modèle Kmean (classe 2) . . . . .	63
32	Analyse de la précision des méthodes sur le BEG . . . . .	65
33	Analyse de la précision des méthodes sur le BE . . . . .	65
34	Étude des projections (épargne) . . . . .	66
35	Analyse de la précision des méthodes sur le BEG (rente) . . . . .	67
36	Analyse de la précision des méthodes sur le BE (rente) . . . . .	67
37	Étude des projections (rente) . . . . .	68
38	densité du BE stochastique sur 1000 scénarios . . . . .	70
39	Étude des trajectoires épargne 2020 . . . . .	73
40	Étude des trajectoires rente 2020 . . . . .	75
41	densité du BE stochastique sur 1000 scénarios du produit épargne en 2020 . . . . .	75
42	Représentation graphique d'un dendrogramme . . . . .	82

# Table des tableaux

1	Activités GPM. . . . .	14
2	Description des variables de la CTAD . . . . .	42
3	Récapitulatif des tests de cohérence sur le produit d'épargne . . . . .	43
4	Récapitulatif des dates clés (épargne) . . . . .	43
5	Récapitulatif des tests de cohérence sur le produit rente . . . . .	46
6	Récapitulatif des dates clés (Rente) . . . . .	46
7	Résultat de lancement . . . . .	49
8	Première partition . . . . .	58
9	Récapitulatif des lancements déterministes (épargne) . . . . .	64
10	Récapitulatif des lancements déterministes (rente) . . . . .	67
11	Récapitulatif des lancements stochastiques . . . . .	70
12	description statistique des données 2020 . . . . .	72
13	Récapitulatif des lancements déterministe épargne 2020 . . . . .	73
14	Récapitulatif des lancements déterministes de rente 2020 . . . . .	74
15	Récapitulatif des avantages et inconvénients des méthodes . . . . .	77

## Liste des sigles et acronymes

**ALM** : Asset and Liability Management

**BE** : Best Estimate

**Cah** : Classification Ascendante Hiérarchique

**EIOPA** : European Insurance and Occupational Pensions Authority

**GPM** : Groupe Pasteur Mutualité

**GSE** : Générateur de Scénarios Economiques

**IARD** : Incidents Accident Risques Divers

**MP** : Model Point

**PB** : Participation aux bénéfices

**PM** : Provision Mathématique

**PT** : Provisions techniques

**QIS** : Quantitative Impact Studies

**RM** : Risk Margin

**SCR** : Solvency Capital Requirement

**TMG** : Taux Minimum Garanti

**UC** : Unité de Compte

## Liste des packages de R utilisés

**FactoClass** [[PD07](#)]

**fastcluster** [[M13](#)]

**factoextra** [[KM20](#)]

**ggdendro** [[dR22](#)]

**ggpubr** [[Kas23](#)]

**rpart** [[TA22](#)]

**rpart.plot** [[Mil22](#)]

# INTRODUCTION

Le Groupe Pasteur Mutualité (GPM) est un groupe mutualiste spécialisé dans le secteur de la santé. La mutuelle est actuellement composée principalement de 3 entités.

- AGMF qui propose des garanties santé et prévoyance individuelles ou collectives ;
- GPMA qui propose des garanties en assurance vie ;
- PANACEA qui propose des garanties IARD.

Ce mémoire porte sur l'entité GPMA en assurance VIE qui représente environ 20% de l'activité de l'entreprise. Ainsi, une partie du processus de gestion ALM est externalisée et effectuée par un prestataire externe. L'entreprise, étant dans une perspective de croissance, cherche désormais à internaliser tous les processus et méthodes de valorisation dans ce domaine.

En effet, l'ALM est un processus qui permet de modéliser l'interaction entre l'actif et le passif (Participation aux Bénéfices, Rachats...) à travers la projection du bilan de l'entreprise. Il est donc d'une importance capitale dans la gestion d'une société d'assurance vie. Les compagnies développent dans cette mesure des modèles internes pour la valorisation de leurs engagements et respecter les normes règlementaires Solvabilité 2.

La robustesse de tels modèles internes repose sur le principe de Monte-Carlo (loi des grands nombres). Le modèle projette les actifs et passifs de la société en prenant en compte plusieurs paramètres (rachats, décès, taux minimum garantie, revalorisation...). Dès lors, deux problèmes se posent, le temps de calcul et le stockage. Pour répondre à cette problématique, les assureurs ont recours le plus souvent soit à une méthode ALM moins directe (le flexing) soit en passant par la constitution de Model Points (MP) qui sera utilisé dans ce mémoire.

La méthode des « Model Points » consiste à regrouper les polices d'un portefeuille selon un certain nombre de critères. En pratique, les assureurs réalisent les MP au niveau de l'actif et du passif. Mais, dans cette étude, on se focalisera que sur ceux du passif.

L'objectif de ce mémoire est donc d'étudier le portefeuille pour construire des MP au niveau du passif tout en respectant les contraintes de solvabilité 2. Le model point optimal doit respecter trois principaux critères :

- Temps de calcul
- Stockage de donnée
- Estimation en solvabilité 2 : BE Central, BEG...

Ce mémoire étudie par ailleurs différents types de produit d'assurance vie. L'impact de nos MP sera relevé sur des produits d'épargne et de rente.

Pour la construction de MP, ce mémoire se base sur les principales méthodes de clustering dans la littérature. Le clustering est une méthode de regroupement de données qui consiste à grouper les éléments en classe homogènes. De ce fait, dans une même classe, l'objectif est d'obtenir des individus les plus similaires possibles, c'est-à-dire ayant les

mêmes caractéristiques.

Avant de faire le clustering, nous allons recueillir les résultats avec les données brutes pour voir la précision de nos modèles.

Les 3 méthodes retenues en fonction de nos données et notre objectif sont :

- CART : simple et intuitif,
- Cah : simple, mais difficulté de mise en œuvre sur beaucoup de donnée,
- Kmean : rapide, flexible et convient à une grande masse de donnée.

Ces méthodes seront challengées sur les différents produits à notre disposition en fonction des critères évoqués plus haut. À ce stade, pour qu'un MP soit considéré comme fiable, l'écart d'estimation par rapport à la projection des données brutes ne doit pas être significatif. La résolution de la problématique de ce mémoire s'articule ainsi en trois grandes parties :

## I. RÈGLEMENTATION ET CONTEXTE

Cette partie permet de fixer le cadre réglementaire de cette étude. Elle cherche à décrire l'entreprise, le contexte ainsi que la réglementation autour de laquelle tourne ce mémoire. On présentera également le modèle ALM dans sa généralité pour familiariser le lecteur au contexte du mémoire ainsi que la réglementation qui l'englobe.

## II. CONSTRUCTION DE MODEL POINT

Cette partie présente, analyse et synthétise la méthode de construction de MP. Les travaux menant à la méthode de construction des MP optimal sont décrits dans cette partie. Le lecteur pourra découvrir ou redécouvrir les techniques de clustering et comment l'intégrer dans le processus ALM d'une entreprise.

## III. BACKTESTING

La dernière partie valide l'utilisation du modèle précédent pour une étude ALM. En effet, elle cherche à tester la validité et la robustesse des MP construits.

## Part I

# REGLEMENTATION ET CONTEXTE

# 1 PRÉSENTATION CADRE D'ETUDES

## 1.1 Présentation GPM

Le Groupe Pasteur Mutualité est un groupe mutualiste qui a évolué au fil des années et contient un portefeuille diversifié d'adhérents des métiers de la santé. Il offre une gamme de couverture complète en santé, prévoyance, responsabilité civile professionnelle, épargne-retraite, garantie autonomie, assurances de prêt, habitation, automobile, à destination des professionnels de santé. Fort de plus de 130 000 adhérents, 170 000 personnes protégées, et un chiffre d'affaires prévisionnel de plus de 250 millions d'euros, GPM affiche une grosse solidité financière et est l'une des premières mutuelles dans le secteur de la santé en France.

Étant un groupe français de premier plan administré par des professionnels de santé, le Groupe Pasteur Mutualité propose des solutions conçues par et pour les professionnels de santé. Parmi les adhérents de GPM, nous pouvons compter :

- Médecins libéraux,
- Praticiens hospitaliers,
- Pharmaciens,
- Infirmiers, Chirurgiens-Dentistes
- Sage-femmes,
- Vétérinaires,
- professions libérales,
- Autres professions de santé.

L'offre offerte aux adhérents est ainsi complète et variée. On retrouve la répartition des activités par entités dans le tableau suivant :

AGMF	Complémentaire santé	Compléter le remboursement des frais de santé
	Arrêt travail invalidité	Maintenir les revenus en cas d'arrêt maladie ou d'invalidité
	Garantie en couverture emprunteur	Assurer les emprunts immobiliers en cas de décès ou perte d'autonomie
GPMA	Epargne	Placer de l'argent de côté et le récupérer avec des intérêts
	Retraite	Compléter les revenus de la retraite
	Rentes	Recevoir une somme fixée à l'avance, périodiquement
PANACEA	Responsabilité Civile Professionnelle	Se protéger des litiges dans le milieu professionnel
	Protection Juridique	Assurer l'indemnisation d'un préjudice causé par un tiers
	Perte Pécuniaires Diverses	Assurer la perte d'un associé par le versement d'un capital

Table 1: Activités GPM.

Le cadre de ce mémoire se restreint à l'étude du portefeuille d'adhérent de GPMA. On

présentera par la suite une généralité sur l'assurance vie pour mieux décrire les produits utilisés plus tard.

## 1.2 Contexte et cadre d'étude

GPMA est l'entité spécialisée en assurance vie. On distingue donc trois principales couvertures, à savoir l'épargne, la retraite et les rentes.

L'assurance vie est un placement à long terme qui répond à différents besoins :

- Protéger les siens ;
- Constituer un complément de retraite ;
- Bénéficier d'un avantage fiscal ;
- faire un placement intéressant ;
- Constituer une épargne de précaution (en cas d'invalidité, incapacité, décès) ;
- Préparer sa succession.

Cette définition du mémoire [Qur13] montre que l'assurance est un moyen très important pour les objectifs économiques du citoyen.

C'est pour cela qu'elle est considérée traditionnellement comme le placement préféré des Français. Mais ces dernières années les crises ainsi que la chute des taux d'intérêt fragilise le domaine.

Avant de poursuivre, il est important de faire la distinction entre l'assurance en cas de décès, dite "assurance décès" qui verse le capital ou la rente en cas de décès, et l'assurance en cas de vie, qui verse un capital ou une rente à la survie de l'adhérent à la date d'échéance du contrat.

Pour un contrat d'assurance vie souscrit pour GPMA, Il y a principalement trois parties qui entrent en jeu :

- L'assureur : GPMA qui rédige le contrat et s'engage à offrir des prestations
- L'adhérent : qui signe le contrat, verse les primes et désigne le bénéficiaire
- le(s) bénéficiaire(s) : perçoit les fonds en cas de décès (rente de conjoint, rente éducation)

L'assurance vie bénéficie aussi en général d'une fiscalité favorable. En effet, lors d'un rachat partiel ou total, seule la part constituée des plus-values est imposée. Bien que soumises aux charges sociales (CSG, CRDS) en vigueur, la part versée des plus values lors d'un rachat partiel ou total bénéficie d'avantages fiscaux non négligeables (selon l'ancienneté du contrat).

Nous détaillons ci-dessous les caractéristiques d'un contrat d'assurance vie.

### 1.2.1 SOUSCRIPTION ET RENONCIATION

Les contrats d'assurance-vie ou de capitalisation sont des produits financiers à long terme qui permettent de constituer une épargne et de bénéficier d'avantages fiscaux. La durée moyenne de ces contrats est souvent de 8 ans, car c'est le délai à partir duquel les gains sont exonérés d'impôt sur le revenu.

La prorogation tacite des contrats signifie que si l'assuré ne souhaite pas mettre fin au contrat à l'échéance, le contrat sera automatiquement renouvelé d'année en année. Cela évite à l'assuré de devoir renouveler manuellement le contrat chaque année. Toutefois, il est important de noter que les conditions du contrat peuvent être modifiées à chaque renouvellement, il est donc recommandé de vérifier régulièrement les clauses du contrat pour s'assurer que les conditions sont toujours en adéquation avec les objectifs de l'adhérent.

ce dernier peut ainsi renoncer légalement à sa souscription dans les 30 jours à compter de la date de remise de la notice d'information du contrat.

### 1.2.2 COTISATIONS

Les versements sur contrat peuvent être effectués sous différentes formes :

- **Cotisations périodiques:** les montants des versements (mensuels, trimestriels, annuels...) sont prédéfinis. Mais l'adhérent peut interrompre ou redéfinir à tout moment ces versements;
- **Cotisations uniques:** à la souscription, l'adhérent doit verser un certain montant;
- **Cotisations libres:** l'adhérent peut verser quand il le souhaite un montant, mais généralement, il est soumis à un versement minimal à la souscription du contrat.

Dans le cas où l'adhérent ne verse pas de cotisations, le contrat peut soit être réduit, soit être résilié. Chaque versement effectué par l'adhérent est investi après avoir été déduit des frais et chargements de gestion, qui sont précisés dans la demande de souscription. Ces frais permettent de rémunérer l'assureur et donnent accès à certains services pour l'adhérent.

### 1.2.3 LES DIFFÉRENTS TYPES DE SUPPORT

Les produits proposés en assurance vie se déclinent en deux familles de contrats :

- **Contrats monosupports en euros :** Il s'agit de contrats en euros qui fonctionnent comme un compte d'épargne où le capital investi est protégé. L'assureur garantit toujours un taux de rendement minimum (éventuellement nul) pour ses assurés. L'argent épargné est disponible à tout moment, mais des frais supplémentaires peuvent être appliqués en cas de retrait anticipé.
- **Contrats en unités de compte (UC) :** Ces contrats donnent aux souscripteurs une liberté de choix quant aux investissements dans lesquels ils souhaitent placer leur argent. Ils peuvent opter pour une garantie exprimée en unités de compte, telles que des obligations, des actions ou des parts d'OPCVM ou de SICAV, ce qui peut offrir une espérance de gain plus importante qu'un contrat classique de type obligataire. Toutefois, cela implique un risque accru, car la valeur des unités de compte dépend du comportement des marchés financiers.

Bien que ces contrats ne prévoient généralement pas de mécanisme de participation aux bénéfices, il est possible de bénéficier d'une garantie plancher en cas de décès, garantissant que le bénéficiaire recevra un capital d'au moins le montant investi net de charges. Cela peut rassurer les investisseurs soucieux de protéger leurs proches en cas d'imprévu.

En somme, ces contrats permettent aux souscripteurs de personnaliser leur investissement en fonction de leurs préférences et de leur niveau de tolérance au risque, tout en offrant une protection pour leur capital investi.

Les **contrats multisupports** sont formés par la juxtaposition d'un contrat en euro à taux garanti et d'un contrat en unité de compte.

#### 1.2.4 LA REVALORISATION DU CONTRAT

Les produits proposés en assurance vie ont deux mécanismes très importants impactant le comportement de l'assuré :

- **La participation aux bénéfices (PB):**

D'après le Code des Assurances, les entreprises d'assurance-vie doivent faire participer les assurés aux bénéfices techniques et financiers qu'elles réalisent. Elles doivent verser au minimum 85% des bénéfices financiers et 90% des bénéfices techniques de l'année : c'est la PB réglementaire.

La PB peut aussi avoir un aspect contractuel (clause de PB à intégrer au contrat) et un aspect discrétionnaire (à des fins commerciales, jusqu'à 100% de PB).

- **Le Taux Minimum Garanti (TMG) :**

Lorsqu'on souscrit un contrat d'épargne, l'assureur peut proposer un taux Minimum Garanti (TMG) garanti pour toute la durée du contrat. Cependant, l'assurance vie étant un engagement à long terme, l'assureur doit garantir ce taux sur une période étendue, même si le rendement de son portefeuille d'actifs est soumis aux aléas des marchés financiers.

Le Code des Assurances encadre le montant du TMG et établit un taux maximal pour assurer la solvabilité de l'assureur. Pour les contrats à prime unique ou à versements libres de moins de 8 ans, ce taux ne peut pas dépasser 75% du Taux Moyen de l'Emprunt d'Etat (TME). Pour les autres contrats, le taux maximal est égal à  $\min(60\% \text{ du TME}; 3,5\%)$ .

Bien que ces limites visent à protéger l'assureur, le marché de l'épargne est très concurrentiel et les compagnies doivent proposer des taux attractifs pour rester compétitives. Le taux effectif proposé par l'assureur est généralement supérieur au TMG, car les assurés sont très sensibles aux écarts entre les taux proposés par différentes compagnies et par rapport aux niveaux du marché financier. Les assurés peuvent ainsi décider de racheter leur contrat en fonction de ces écarts. Dans la suite, il est supposé que  $\text{TMG}=0\%$ .

### 1.2.5 LES POSSIBILITÉS DE RACHAT

Le rachat d'un contrat d'assurance vie permet à l'assuré de retirer une partie ou la totalité de son épargne avant la fin prévue du contrat. Cette option peut être exercée à tout moment par l'assuré, mais représente un risque important pour l'assureur car il ne peut pas prévoir comment l'assuré réagira à l'évolution du marché.

Il existe deux types de rachats : le rachat total, qui permet au souscripteur de mettre fin au contrat et de récupérer la provision mathématique (PM) correspondant à l'épargne atteinte, et le rachat partiel, qui permet de retirer une partie de l'épargne avant la date d'échéance prévue.

Les pénalités de rachat varient selon les contrats et les compagnies, mais sont limitées à 5% des PM selon l'article R. 331-5 du Code des Assurances, et deviennent nulles après 10 ans.

Les avances en cours de contrat permettent à l'assuré d'obtenir une somme d'argent sans affecter son épargne, sous réserve de rembourser le prêt à un taux d'intérêt fixé par l'assureur.

## 1.3 Présentation des produits

Pour notre étude, nous travaillerons avec 2 produits d'assurance vie (épargne et rente).

### 1.3.1 ÉPARGNE

Le produit d'épargne est un contrat d'assurance collective sur la vie à adhésion facultative en euros. Il s'agit d'un produit d'assurance vie mono-support en euros fermé à la commercialisation (run-off) au 01/01/1994 qui a servi une performance de 1% en 2021 net de frais de gestion et avant prélèvements sociaux et fiscaux.

Ce produit comporte une garantie en cas de vie qui prévoit le versement d'un capital différé pouvant être transformé en rente à la demande de l'adhérent et une garantie en cas de décès (contre-assurance en cas de décès) qui prévoit le versement d'un capital au(x) bénéficiaire(s) désigné(s).

À noter que :

- A l'échéance de l'adhésion, l'assureur verse à l'adhérent qui ne souhaite pas proroger son adhésion un capital égal au montant du capital acquis à la date de la demande.
- L'adhérent peut demander le service d'une rente viagère payable trimestriellement à terme échu.
- Adhérent peut demander le rachat de tout ou partie de son capital acquis
- Les valeurs de rachat sont calculées au terme de chacune des huit premières années en tenant compte d'un taux d'intérêt minimum de 0% brut de "frais de gestion financière" (ces frais correspondent au prélèvement sur encours au taux de 0,50% de l'encours géré), garanti sur 8 ans (et brut de prélèvements sociaux et fiscaux).
- La gestion administrative est financée par un prélèvement, sur chaque versement, égal aux taux suivants :

– 3% sur la partie du versement comprise entre 0 et 7 500 €;

- 2% sur la partie du versement comprise entre 7 501 € et 15 000 €;
- 1% sur la partie du versement comprise entre 15 001 € et 75 000 €;
- 0,5% sur la partie du versement supérieure à 75 000 €.

### 1.3.2 RENTE

Le produit de rente a été créé à l'origine sous la forme d'un régime L.441, qui assure un paiement d'une rente viagère au plus tôt à compter de la liquidation de la pension de l'adhérent dans un régime d'assurance vieillesse ou de l'âge fixé en application de l'article L. 351-1 du Code de la Sécurité Sociale, par conversion de l'épargne acquise.

Le contrat ne comporte pas de garantie en capital au moins égal aux sommes versées, nettes de frais. En cas de décès en cours de contrat, versement d'une rente aux bénéficiaires désignés. Les rentes sont revalorisables et disposent des options suivantes.

Les rentes sont payables trimestriellement à terme échu. Elle peut être stipulée réversible sur la tête du bénéficiaire à concurrence de 60% ou de 100% de son montant selon le choix opéré par l'adhérent lors de la demande de liquidation. Et pour finir, une option de rente certaine qui garantit 100% de la rente au bénéficiaire jusqu'au 75 anniversaire de l'adhérent.

Ce contrat présente une clause de participation aux bénéfices de 90% des résultats techniques et 85 % des résultats financiers sous déduction des pertes techniques éventuelles. À noter que :

- le produit de Rente bénéficie de la faculté de transfert sur les contrats de même nature et de rachat exceptionnelle.
- Les capitaux inférieurs à 120 €/trimestre sont versés sous forme de capital à la liquidation.
- Le montant de l'indemnité de transfert pour le produit RENTE est fixé à 0,50% de la valeur de transfert.
- Les rentes versées soient des rentes viagères immédiate ou différée, ou soit sous forme de rente temporaire d'éducation versée jusqu'au 25 anniversaire.

Ainsi s'achève une brève présentation de l'environnement de notre étude, dans la section suivante, nous allons décrire la réglementation autour de L'ALM.

## 2 RÈGLEMENTATION : SOLVABILITE 2

### 2.1 Présentation de Solvabilité II

La directive Solvabilité II est une approche fondée sur des principes (principle-based) qui vise une meilleure adaptation des exigences de capital et des pratiques de contrôle au profil de risque (risk-based). Concrètement, elle énonce des principes larges (plutôt que sur des règles (rule-based)) et laisse aux organismes assureurs une grande part de liberté dans l'application de ces principes. L'objectif de cette directive est de renforcer la protection des adhérents et l'intégration du marché européen, d'améliorer la gestion des risques des assureurs européens et de favoriser leur compétitivité sur le marché international via l'harmonisation des normes. Dans le schéma suivant, nous présentons le bilan selon Solvabilité II tiré du mémoire [Qur13] :

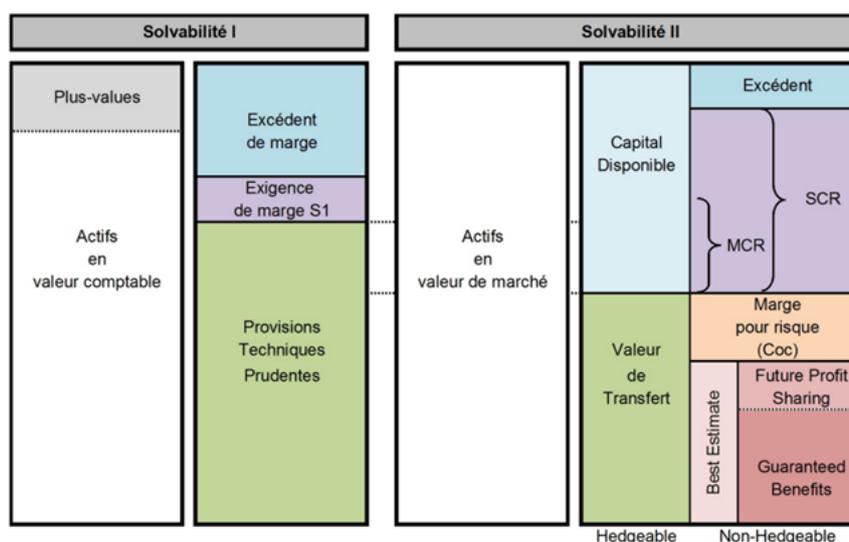


Figure 1: Bilan Solvabilité II

Ce projet repose sur trois piliers :

- **Pilier 1 : Exigences quantitatives**

Le premier pilier a pour objectif de définir des seuils d'exigence en capital pour les fonds propres et une méthodologie de calcul des provisions techniques. Deux exigences de capital sont définies :

- Le Minimum Capital Requirement (MCR) qui correspond au niveau minimum de fonds propres en dessous duquel l'intervention de l'autorité de contrôle sera automatique.
- Le Solvency Capital Requirement (SCR) qui représente le capital cible nécessaire pour absorber le choc provoqué par une sinistralité exceptionnelle.

- **Pilier 2 : Exigences qualitatives et activités de contrôle**

Le deuxième pilier de Solvabilité II complète les exigences quantitatives de capitaux par des exigences qualitatives et d'un système global de gestion des risques. La réforme prévoit des mesures en matière de gouvernance, de contrôle et d'audit interne afin d'assurer une pratique de gestion saine et prudente des assureurs. Le processus d'évaluation interne des risques et de solvabilité, connu sous le nom de « Own Risk and Solvency Assessment » (ORSA), qui comporte à la fois un volet qualitatif et un volet quantitatif, est la pièce maîtresse de ce plan. Ce pilier définit également le cadre d'intervention de l'autorité de contrôle.

- **Pilier 3 : Exigences en matière d'informations prudentielles et de publication**

Le troisième pilier a pour objectif de définir l'ensemble des informations nécessaires aux autorités de contrôle pour exercer leur pouvoir de surveillance. Afin de mesurer concrètement leurs conséquences, les directives européennes ont fait l'objet d'études quantitatives d'impact (Quantitative Impact Studies - QIS). Le QIS5 [dCP11] précise les modalités techniques applicables aux compagnies d'assurance et le calcul du SCR y est en particulier détaillé.

Deux catégories de passifs sont présentées (QIS5 TP.4.1) :

- **les passifs répliquables** : le passif est considéré comme répliquable si celui-ci peut admettre une couverture sur les marchés financiers par un produit suffisamment liquide répliquant parfaitement ses flux. La valeur des provisions techniques est donc estimée selon la valeur de marché de son portefeuille répliquant. En pratique, peu de passifs présentent cette caractéristique. Les marchés des passifs d'assurance sont incomplets pour l'essentiel dans le sens où ils sont peu liquides et il est impossible de répliquer les risques.
- **les passifs non répliquables** : le passif n'est pas parfaitement répliquable sur les marchés financiers. Dans ce cas, les provisions techniques sont estimées comme étant la somme du Best Estimate et d'une marge pour risque.

## 2.2 Best Estimate

Le Best Estimate ou meilleure estimation ci-dessous, est défini par l'extrait suivant de l'article 77 de la directive 2009/138/CE :

### Article 77-Calcul des provisions techniques

2. La meilleure estimation correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent (valeur actuelle attendue des flux de trésorerie futurs), estimée sur la base de la courbe des taux sans risque pertinents.

Le calcul de la meilleure estimation est fondé sur des informations actualisées et crédibles et des hypothèses réalistes et il fait appel à des méthodes actuarielles et statistiques adéquates, applicables et pertinentes.

La projection en matière de flux de trésorerie utilisée dans le calcul de la meilleure estimation tient compte de toutes les entrées et sorties de trésorerie nécessaires pour faire face aux engagements d'assurance et de réassurance pendant toute la durée de ceux-ci.

La meilleure estimation est calculée brute, sans déduction des créances découlant des contrats de réassurance et des véhicules de titrisation. Ces montants sont calculés séparément, conformément à l'article 81.

Dès lors, le calcul du Best Estimate doit tenir compte de toutes les garanties et options offertes par le contrat sous des hypothèses de comportement des adhérents (ex : exercice d'option). De l'ensemble des versements discrétionnaires futurs (ex : PB futures) sous des hypothèses de « future management actions ». Le calcul du Best Estimate s'effectue sous les hypothèses suivantes :

- **l'horizon de projection:** les spécifications techniques du QIS5 insistent sur le fait que l'horizon de projection doit être suffisamment long pour que la différence entre les provisions calculées avec cet horizon et celles calculées jusqu'à extinction totale du portefeuille soit négligeable;
- **les flux sortants:** ils comprennent les prestations et tous types de frais et dépenses générés au titre des contrats d'assurances;
- **l'actualisation:** les flux sont actualisés à l'aide de la courbe des taux sans risque fournie par l'EIOPA ou de la courbe des taux swaps du marché ;
- **les primes:** une partie des primes futures est désormais prise en compte dans le calcul;
- **la réassurance:** les provisions en Best Estimate sont calculées brutes de réassurance.

Le Best Estimate (BE) correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs actualisés.

$$BE = E \left( \sum_{i=1}^T \frac{F_t}{(1 + r_t)^t} \right)$$

où :

- $F_t$  : flux probable de l'année t
- $r_t$  : taux sans risque de maturité t
- $T$  : le nombre d'années de projection

La valeur des flux probables est estimée par la méthode de Monte-Carlo. Celle-ci nécessite la simulation d'un nombre important de scénarios. Les scénarios sont indépendants et identiquement distribués.

Le *Best Estimate* est évalué comme suit :

$$BE = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^T \frac{F_t^n}{(1 + r_t)^t}$$

où :

- $F_t^n$  : flux probable de l'année t pour le scénario n
- $N$  : le nombre de scénarios utilisés

## 2.3 La marge pour risque (RM)

Le tableau présenté ci dessous provient du cours de *Planchet et KAMAL Armel* [AP23] et permet de comprendre la définition du **Risk Margin**:

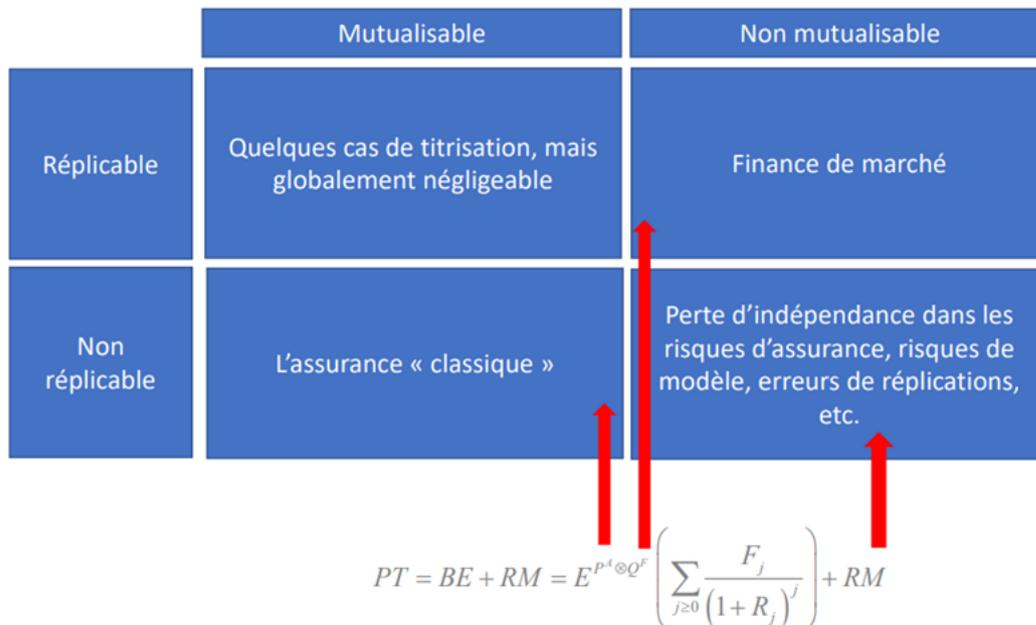


Figure 2: Synthèse de l'analyse des risques

La marge pour risque correspond donc aux risques non mutualisables et non réplicables. Elle peut être considérée aussi comme le montant supplémentaire requis par rapport au *best estimate* pour que les engagements puissent être transférés à un autre unique assureur.

Dans le QIS 5, le calcul de la risk margin se fait par la méthode "coût du capital". Pour plus de détails, le lecteur peut se référer aux spécifications techniques du QIS 5 (TP5.2-TP5.74).

## 2.4 Solvency Capital Requirement (SCR)

Il doit permettre à l'entreprise d'absorber les sinistres imprévus et ainsi d'honorer ses engagements auprès des assurés.

L'EIOPA propose deux possibilités de calcul aux assureurs :

- **par formule standard** : la formule évalue le capital par un calcul reposant sur des données comptables et des indicateurs connus.
- **soit par modèle interne** : méthode d'évaluation qui permet à l'assureur de calculer son SCR selon son profil de risque.

La structure modulaire de la formule standard selon le QIS 5 est représentée sur la figure suivante :

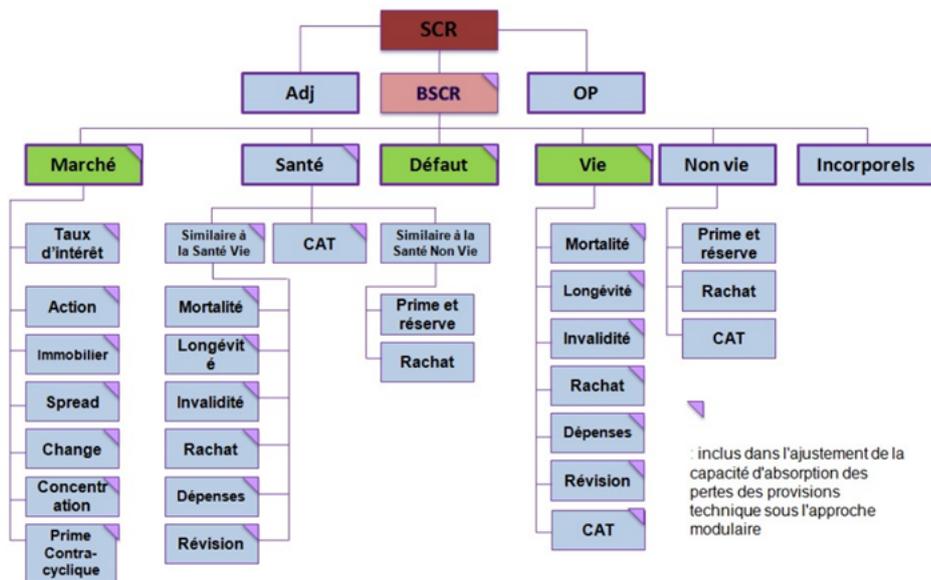


Figure 3: Structure modulaire du SCR selon le QIS 5

En effet, La formule standard de calcul du SCR utilise une approche modulaire. L'approche modulaire consiste en une agrégation d'un calcul de capital effectué individuellement pour chaque risque. Chaque module de risque est calibré de façon à ce que la probabilité de ruine à horizon d'un an soit égale à 0,5%. Le calcul du SCR est fait au moins une fois par an, mais suivi de manière permanente.

L'organisme doit calculer la perte subie en cas d'événement défavorable lié à une trentaine de facteurs de risque. Pour tenir compte de la probabilité faible de réalisation simultanée de tous ces événements, la formule standard introduit des corrélations entre ces facteurs de risque et permet ainsi à l'organisme de constater des bénéfices de diversification. Enfin, pour refléter au mieux la réalité économique, la perte brute observée par l'organisme (BSCR ou SCR de base) peut encore être atténuée par deux effets avant d'obtenir le SCR final :

- D'une part, la prise en compte de la capacité de l'organisme à transférer une partie de sa perte aux assurés via une moindre participation aux bénéfices que celle qu'il avait escomptée avant le choc.
- D'autre part, l'imputation de la perte au résultat fiscal qui conduira, dans la plupart des cas, à payer moins d'impôts dans le futur qui avait été comptabilisé au bilan initial.

Le SCR se décompose selon le graphique suivant du mémoire [Qur13]:

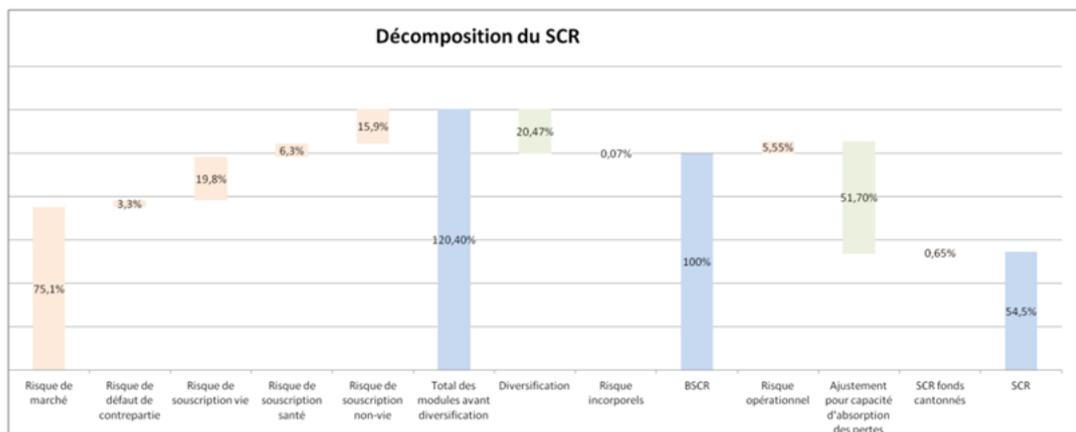


Figure 4: Décomposition du SCR

Le BSCR (*Basic SCR*) est composé d'éléments exprimés en pourcentage de ce dernier. Pour l'obtenir, les charges de chaque risque sont additionnées, puis l'effet de diversification est retranché. Il est important de noter que les graphiques doivent être lus de gauche à droite. Ensuite, la charge pour risque opérationnel est ajoutée au BSCR, suivi de la prise en compte négative de l'ajustement pour capacité d'absorption des pertes et de l'intégration du SCR des fonds cantonnés. Tout cela conduit à l'obtention du SCR final.

## 3 ALM

### 3.1 La nécessité d'une modélisation ALM

Avant de poursuivre, il est important de souligner l'importance pour une compagnie d'assurance vie d'un **modèle ALM**.

Par définition, le Best Estimate est un indicateur économique : il est calculé à partir des flux futurs sous-jacents (Cash-Flows) aux engagements d'une compagnie d'assurance à l'égard de ses assurés. Il est alors nécessaire de projeter l'ensemble des éléments comptables d'une société d'assurance vie dès lors qu'ils présentent un impact sur la valeur actuelle des prestations et des frais futurs.

En second lieu, dans le cas de l'assurance vie, les rendements financiers constatés à l'actif ont une influence directe sur l'évolution du passif d'une compagnie. Cela peut se justifier à travers la revalorisation des contrats d'épargne, par le biais de la participation aux bénéfices ainsi que les rachats, qui sont directement liés à la performance financière. On peut évoquer le cas des versements libres où l'assureur ne peut anticiper tous ces flux.

Du fait de ces contraintes de modélisation, le choix d'un modèle Actif-Passif est nécessaire dans le cadre d'un calcul de Best Estimate en assurance vie. Un tel modèle permet en effet de prendre en compte les différentes interactions existant entre l'actif et le passif d'une compagnie d'assurance.

On peut donc résumer l'ALM par un ensemble de techniques et de procédés qui permettent à l'assureur de prévoir, d'ajuster et de contrôler la composition et l'adéquation de tous ses actifs et passifs. Il doit permettre :

- Une optimisation des fonds propres de l'organisation (BE, SCR) ;
- Une minimisation du risque de taux : respecter ces garanties (TMG) ;
- Une adaptation permanente aux conditions du marché et au comportement des assurés.

### 3.2 Le modèle ALM

l'ALM repose sur la simulation de l'évolution du compte de résultat et du bilan pour un horizon donné. Ainsi, le modèle ALM sert à répondre aux exigences de solvabilité 2 en projetant les flux de trésorerie (Cash-Flows). On retrouve les différentes parties du Bilan projetées dans la figure ci-dessous :

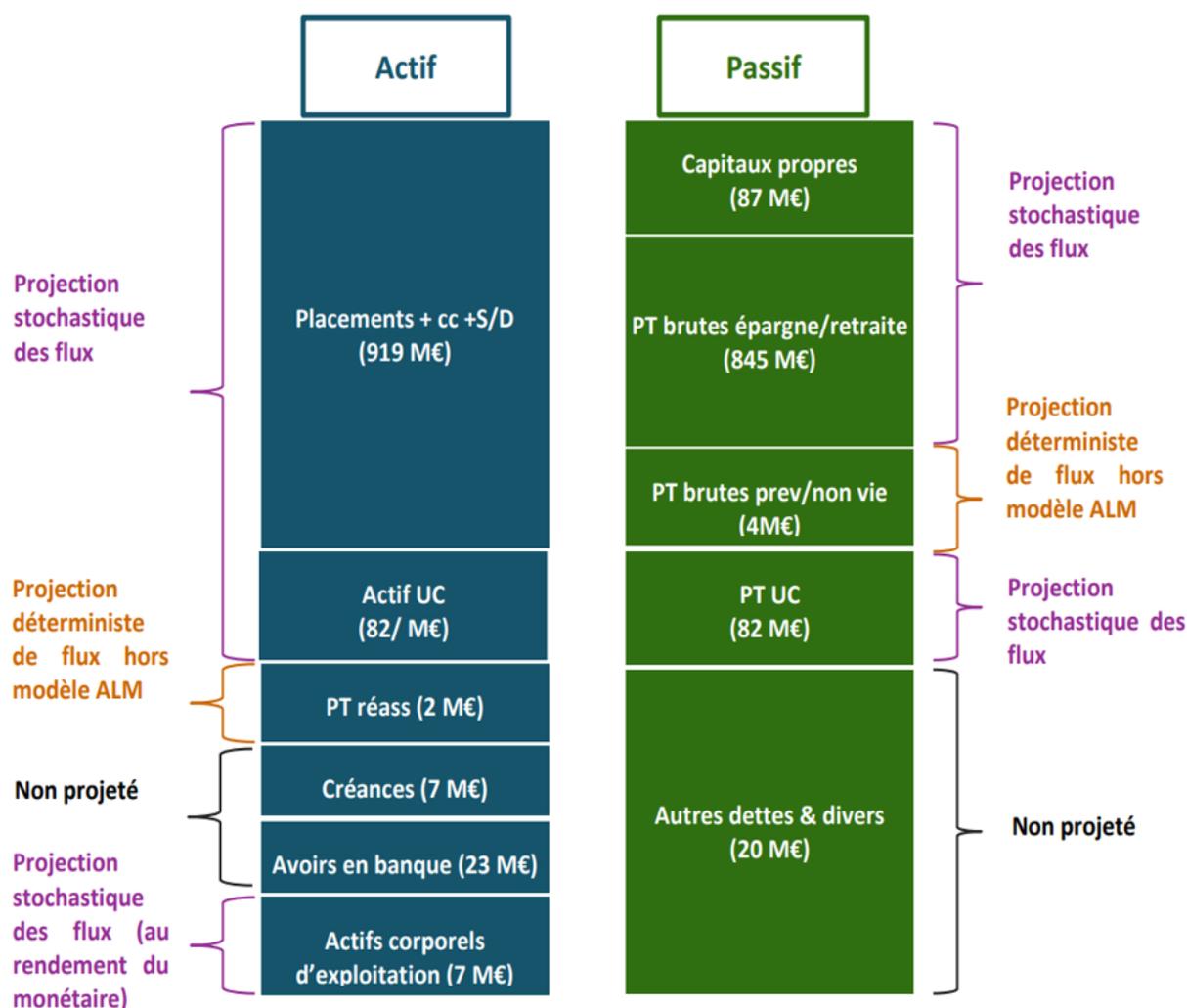


Figure 5: Bilan GPM

Pour ce mémoire, Les MP produits interviennent lors de la projection des PT brutes épargne/retraite.

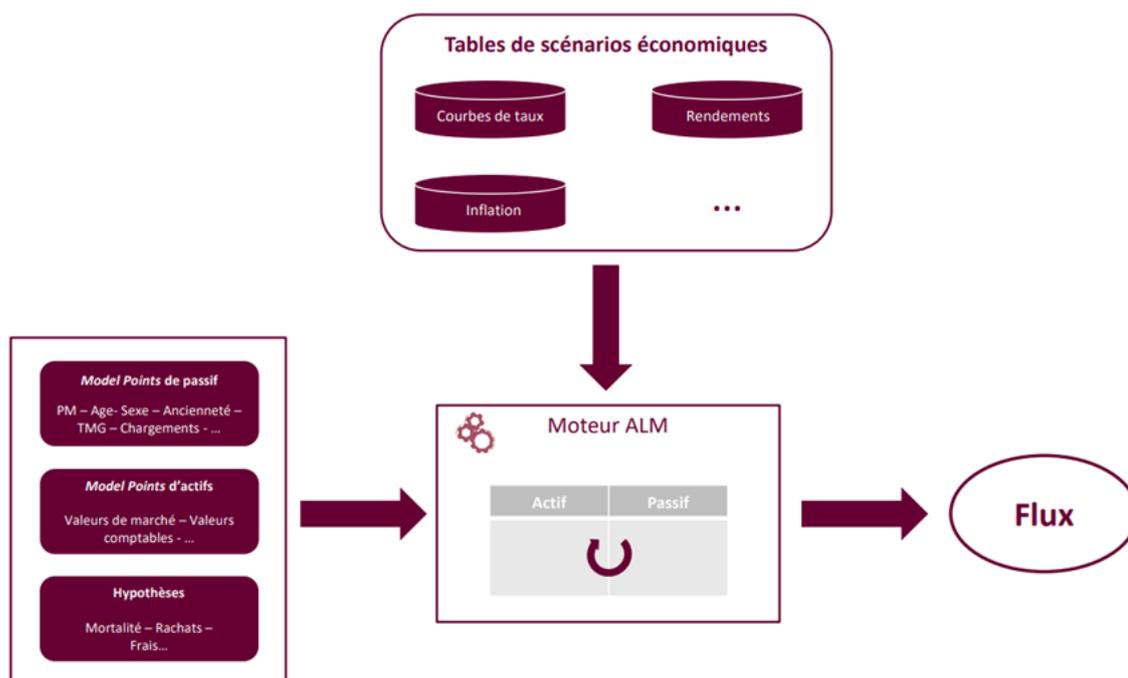


Figure 6: Schema du modèle ALM.

La figure 6 présente le principe d'un modèle ALM. À partir d'un portefeuille initial et de scénarios économiques, il projette le bilan pour estimer les flux futurs (puis le best estimate).

Le modèle ALM projette les flux de l'entreprise en prenant en compte plusieurs hypothèses (financières, comportementales et économiques).

## 4 GSE

### 4.1 Généralités

Le générateur de scénarios économiques (GSE) est un outil de simulation stochastique qui permet de projeter dans le temps des trajectoires réalistes de grandeurs financières telles que les taux d'intérêt, l'indice d'inflation, les rendements d'actions, les rendements obligataires et les rendements immobiliers.

Toutes les compagnies d'assurance sont exposées à deux types de risques : les risques assurantiels liés à leurs engagements envers les assurés et les risques financiers liés à leur portefeuille d'actifs. Le GSE, combiné à un modèle d'assurance prospectif, permet de projeter l'environnement économique de l'assureur, ainsi que les interactions entre l'actif et le passif, tout en prenant en compte le comportement des assurés en réponse à la situation financière de l'assureur (comme l'illustre le schéma ci-dessus). En outre, le GSE permet de quantifier et d'analyser l'impact sur l'activité de l'assureur d'une variation imprévue de l'environnement financier, conformément à la réglementation européenne en vigueur.

L'implémentation et la validation d'un GSE s'effectue classiquement en cinq étapes, énumérées ci-dessous :

1. Choix des modèles de projection ;

2. Choix des inputs ;
3. Calibrage des modèles ;
4. Génération des scénarios ;
5. Validation des scénarios générés

Dans le secteur de l'assurance, on distingue deux catégories de GSE, selon leur univers de modélisation :

- **Monde réel** : un GSE monde réel permet de déterminer l'évolution attendue du bilan économique à l'aide de scénarios stochastiques. Il est utilisé pour générer des variables pour tous les principaux facteurs de risque. Ces scénarios reflètent l'évolution future de l'économie du point de vue de la compagnie d'assurance. Le rendement des actifs sous ce GSE inclut souvent une prime de risque liée à l'actif. Ces modèles sont calibrés en utilisant un historique de données représentatif de l'économie au moment du calcul.
- **Monde risque-neutre** : Les primes de risque sont souvent très complexes pour certains actifs financiers dans le monde réel, alors que sous l'univers risque-neutre, elles sont nulles, car tous les agents économiques sont neutres au risque. Cet univers nous évite la tâche délicate de les déterminer. En monde risque neutre (RN), les modèles doivent être calibrés sur les prix de marché. Un GSE risque neutre est utile pour les valorisations market consistent préconisées par Solvabilité II. C'est cette hypothèse qui est utilisée dans le modèle ALM de ce mémoire.

## 4.2 Litterature sur les GSE

Outre l'univers retenu pour le GSE, la structure de dépendance entre les différents facteurs de risque est un élément important. Ainsi, les modèles d'actifs peuvent être répartis en deux catégories :

- les modèles intégrés : description structurée de plusieurs classes d'actifs à partir d'une variable explicative de référence ;
- les modèles composites : description ad-hoc de chaque classe d'actifs avant de les agréger pour proposer une description globale de l'actif.

La littérature sur les GSE intégrés est abondante. Les plus connus au sein de la communauté actuarielle sont ceux de Wilkie (1986), d'Ahlgrim et al. (2005) ou encore de Brennan et Xia (2000).

Les modèles intégrés proposent une description structurée des classes d'actifs à partir d'une variable explicative de référence. Dans le cadre de l'assurance, l'inflation a une grande importance au sein d'une compagnie d'assurance et ces modèles traitent l'inflation en tant que variable de référence, ce qui n'est pas le cas dans les modèles composites.

Dans les modèles intégrés, on considère généralement que le modèle d'actifs intègre comme supports d'investissement les obligations, les actions, l'immobilier et le monétaire.

L'objectif de cette partie est de les présenter très sommairement. Les méthodes sont développées dans la thèse de (cf. **FALEH** [Fal11]) que nous citons en bibliographie. Le lecteur pourra s'y reporter pour davantage de détails.

### a) Modèle de Wilkie

Le modèle de Wilkie se singularise par une modélisation des prix des actions à partir des taux et des montants de dividendes. Ce modèle inclut par ailleurs le taux d'inflation et les taux d'intérêt à long terme.

Dans ce premier modèle intégré, le taux d'inflation joue un rôle majeur :

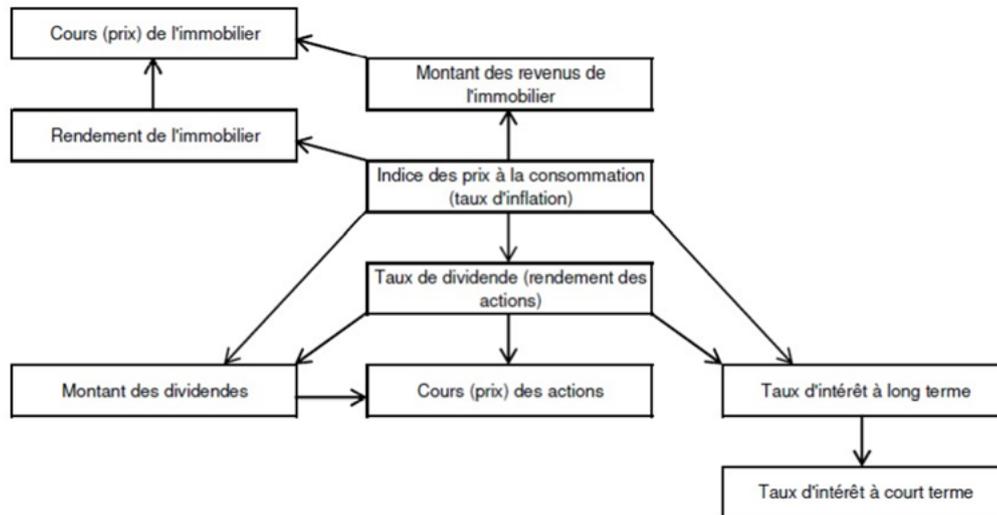


Figure 7: Schéma du modèle de Wilkie

Il peut toutefois apparaître relativement sommaire dans la modélisation interne de chaque type d'actif.

### b) Modèle de Brennan et Xia

Le modèle de Brennan et Xia utilise des modèles stochastiques de taux d'intérêt réels, de l'inflation et des actions en vue d'en déduire un indice des prix à la consommation stochastique et un facteur d'actualisation stochastique réel.

L'une des caractéristiques de ce modèle, développé en 2000, est qu'il considère que l'inflation réalisée n'est pas parfaitement corrélée à l'inflation attendue. L'inflation est obtenue à partir de l'indice des prix à la consommation, et dépend du bruit propre au modèle sur l'inflation réalisée d'une part, et des bruits des modèles sur les actions, les taux d'intérêt et l'inflation attendue d'autre part.

Ce modèle est plus restrictif que le modèle de Wilkie puisqu'il n'intègre que les taux d'intérêt réels, l'inflation et les actions.

Le modèle de Brennan et Xia peut être représenté par les relations suivantes :

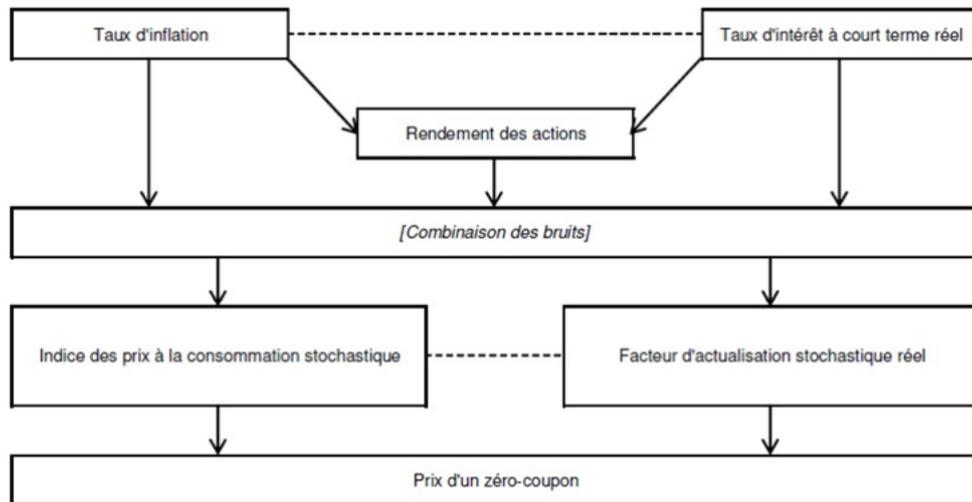


Figure 8: Modèle de Brennan et Xia

### c) Modèle de Ahlgrim et Al

Le modèle d'Ahlgrim et al, constitue une alternative intéressante (aux deux modèles précédents) en combinant une approche intégrée s'appuyant sur l'inflation avec la possibilité d'utiliser des modèles élaborés à l'intérieur de chaque classe d'actif. Au surplus, ce modèle est développé sous l'impulsion de la CASACT 1, ce qui lui confère une certaine légitimité.

Ce modèle couvre, entre autres, les taux d'intérêt, les rendements des actions, le taux d'inflation et les revenus de l'immobilier. Dans le modèle *Ahlgrim et Al*, la projection des taux court et long dépendent de celle de l'inflation. Il est donc plus complexe à mettre en œuvre.

Comme la modélisation prend en compte l'ensemble des dynamiques, il s'agit d'un modèle cohérent. Le modèle d'*Ahlgrim et Al* s'élabore de la façon suivante :

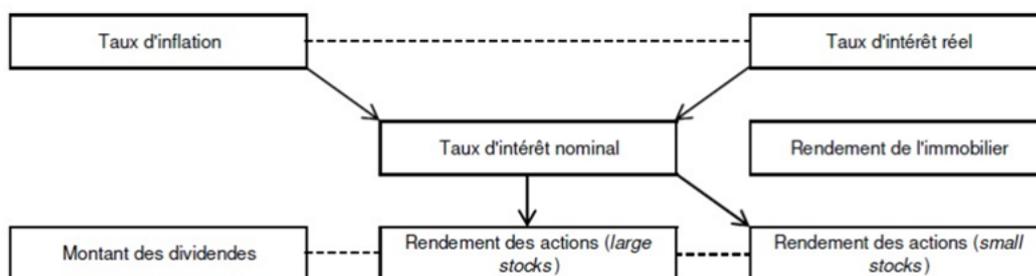


Figure 9: Schéma du modèle Ahlgrim et AL

## 4.3 Présentation du GSE risque neutre GPM

Le GSE utilisé pour l'évaluation s'appuie sur un modèle stochastique basé sur une approche risque-neutre.

Il permet d'obtenir 1 000 trajectoires économiques risque neutre pour le prix de zéro coupon (maturité 1 à 30), l'indice action, l'indice immobilier, l'indice monétaire, les

déflateurs. Pour chacun de ces indices, un modèle de projection est choisi.

**Modèle de taux** : le modèle de taux risque neutre est basé sur un modèle de Hull et White à un facteur à volatilité déterministe. Ce modèle projette des taux négatifs.

$$dr_t = (b(t) - r_t)dt + \sigma dW_t \quad (1)$$

où :

- $r_t$  : le taux d'intérêt
- $b$  : une fonction de  $t$  représentant le retour à la moyenne
- $\sigma$  : Volatilité instantanée
- $W_t$  : Mouvement Brownien

Le calibrage est effectué sur la courbe à la date de valorisation et sur les prix de marché des swaptions sur Euribor à la monnaie. Les prix sont issus de Bloomberg. Il est important de signaler qu'il n'y a pas diffusion stochastique du risque de crédit dans la modélisation. La diffusion des spreads de crédit est déterministe (en l'occurrence le spread des obligations, calibré à  $t=0$  sur la valeur de marché est projeté constant).

Sur plusieurs jeux de simulations lancés avec des graines différentes, le choix du jeu de simulations s'effectue en fonction des critères suivants évalués après lancement du modèle ALM sur les données les plus récentes à disposition (en l'occurrence celles des calculs d'atterrissage à fin 2021):

- Les écarts de fuite du modèle ALM doivent être le plus faible possible
- Des scénarios économiques prudents (c'est-à-dire un taux de couverture pas parmi les plus élevés).
- La qualité de réplication des prix de marché
- La bonne martingalité des scénarios

**Modèle actions** : Le modèle utilisé pour diffuser les actions et l'immobilier est de type log-normal avec une volatilité constante. Il s'agit du modèle Black-Scholes dérivant des travaux de Black, Scholes et Merton (1973) sur l'évaluation des options européennes.

Soit  $S_t$  : le prix de l'actif sous-jacent à la date  $t$ ,  $0 \leq t \leq T$  de maturité  $T$ . Ce prix suit la dynamique suivante :

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \quad (2)$$

où :

- $\mu$  : l'espérance de rentabilité de l'actif (le drift), constante
- $\sigma$  : sa volatilité, constante
- $W_t$  : un mouvement brownien standard (processus de Wiener), tel que les accroissements disjoints  $W_t - W_s \sim N(0, t - s)$  sont indépendants et stationnaires,  $0 \leq s \leq t$

On déduit de cette équation différentielle stochastique, grâce au lemme d'Itô que :

$$S_t = S_0 \exp \left[ \left( \mu - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right],$$

C'est-à-dire que le rendement de l'actif sous-jacent est log-normal car on a :

$$\ln \left( \frac{S_t}{S_s} \right) = \ln(S_t) - \ln(S_s) = \left( \mu - \frac{1}{2} \sigma^2 \right) (t - s) + \sigma (W_t - W_s)$$

Ce résultat sert à la calibration de toutes les options financières. Ainsi, La méthode consiste à calibrer la volatilité des actions type 1 sur la base des prix du call et du put, à la monnaie, sur EuroStoxx50. Cela correspond au call et au put de strike identique de maturité la plus longue qui demeurent assez liquides qui ont été trouvés sur Bloomberg. Tout comme pour les taux, des tests de martingalité, ainsi que de réplcation des prix des options (call et put ayant servi au calibrage) ont été réalisés pour vérifier la market-consistency des scénarios actions générés. Les méthodes de calculs et hypothèses utilisées pour le calibrage de la volatilité des actions sont les suivantes :

- Caractéristiques du call et du put sur EuroStoxx50 utilisés (Maturité : 17/12/2021, taux sans risque : taux 1 an EIOPA hors VA, strike : 3600, cours EuroStoxx50 : 3572, Prix call : 196, Prix put : 325)
- Calibrage du dividende implicite à partir de la parité call-put
- Calibrage de la volatilité implicite à partir du prix du call, de ses caractéristiques et du dividende calibré et avec un modèle de Black-Scholes pour le sous-jacent. Elle est ensuite constante dans le temps.
- Liquidité (en volume de transaction call : 3650, put : 500)

En annexe, un exemple de résolution de ce type d'équation est présenté.

## 4.4 Règlements

La directive Solvabilité II impose plusieurs exigences sur les modèles de GSE et leurs calibrages. Ceux-ci doivent respecter les exigences définies dans l'article 22(3) du règlement Délégué de la commission européenne :

Règlement Délégué article 22(3) – Dispositions générales

Les entreprises d'assurance et de réassurance fixent des hypothèses concernant les paramètres et scénarios futurs des marchés financiers qui sont appropriés et conformes à l'article 75 de la directive 2009/138/CE. Lorsque l'entreprise d'assurance ou de réassurance recourt à un modèle pour produire des projections de paramètres et scénarios futurs des marchés financiers, celui-ci respecte l'ensemble des exigences suivantes :

- (a) il génère des prix d'actifs cohérents avec les prix des actifs observés sur les marchés financiers;
- (b) il ne suppose aucune opportunité d'arbitrage;
- (c) le calibrage des paramètres et des scénarios est cohérent avec la courbe des taux sans risque pertinents utilisée pour calculer la meilleure estimation visée à l'article 77, paragraphe 2, de la directive 2009/138/CE.

Tout d'abord, il faut que les prix d'actifs générés soient market consistent (a), c'est-à-dire compatibles avec un marché financier profond, liquide et sans aucune opportunité d'arbitrage.

Plus précisément :

- un marché liquide où les actifs peuvent être facilement achetés et vendus sans impacter significativement les mouvements de prix ;
- un marché profond tel qu'un grand nombre d'actifs peuvent être traité sans significativement modifier les prix des instruments financiers ;
- un marché transparent dans le sens où les informations sur les prix sont facilement accessibles au public.

Les exigences en ce qui concerne la calibration des modèles d'actifs en monde risque neutre sont :

- Les modèles doivent être calibrés sur des prix de marché observés sur des marchés suffisamment « profonds, liquides et transparents » et cette procédure devrait être basée sur des instruments qui en quelque sorte reflètent la nature et la durée des engagements ;
- Les modèles doivent vérifier le principe d'absence d'opportunité d'arbitrage et doivent permettre de retrouver les prix de marché qui ont été utilisés pour le calibrage (notamment les prix des options) ;
- Pour le calibrage des volatilités, le QIS 5 laisse le choix entre 2 approches :
  1. À partir des volatilités implicites issues des prix de marché des options (inconvenients : disponibilité des options, sur-estimation en période de crise, sous-estimation en période calme, etc.) ;
  2. À partir des volatilités historiques calculées sur des historiques de cours (inconvenients : disponibilité de l'historique, etc.).
- Le calibrage peut aussi se fonder sur des analyses statistiques et actuarielles, tant que les résultats sont market consistent ;

Il existe par ailleurs une liste de validation et de vérifications pour le GSE en univers risque neutre décrits dans l'orientation n°58 des orientations sur la valorisation des provisions techniques (EIOPA-BoS-14/166 FR) :

- Il doit conduire à ce que les rendements moyens de tous les actifs soient égaux au taux sans risque ;
- Le GSE doit reproduire la structure par terme des taux sans risque actuelle utilisé pour actualiser les flux de trésorerie et doit permettre de retrouver les prix de marché qui ont été utilisés pour le calibrage, notamment les prix des options ;
- Il doit permettre de reproduire tous les paramètres de calibrage, notamment les volatilités implicites ;

Le but derrière une évaluation market consistent du passif pour une compagnie d'assurance est de reproduire le « prix de marché » de son passif comme si celui-ci était négocié sur un marché.

## 5 MODEL POINT

### 5.1 Exigences règlementaires

Le 30 janvier 2015 un règlement délégué a été publié par la commission européenne. Ce règlement délégué (UE) 2015/850 de la Commission du 30 janvier 2015 modifiant le règlement délégué (UE) n° 241/2014 complétant le règlement (UE) n° 575/2013 du Parlement européen et du Conseil par des normes techniques de réglementation concernant les exigences de fonds propres applicables aux établissements (Texte présentant de l'intérêt pour l'EEE).

Le Règlement Délégué précise à l'article 35 que « les projections des flux de trésorerie [...] sont effectuées séparément pour chaque contrat ».

Cependant, ce règlement suggère un regroupement de polices lorsque celles-ci et les risques sous-jacents présente des caractéristiques similaires tant que cette constitution de model points ne fausse pas les résultats de projections.

Les guidelines Solvabilité 2 portant sur la constitution des model points sont concentrés dans le Règlement Délégué (l'Article 35) et l'Orientation EIOPA sur la valorisation des provisions techniques. Le groupe de travail de l'institut des actuaires [da16] donne des extraits de cette réglementation ainsi que les contraintes de constitution de model points.

Règlement Délégué Article 35 - Groupes de risques homogènes d'engagements d'assurance vie

Les projections des flux de trésorerie utilisées dans le calcul des meilleures estimations concernant les engagements d'assurance vie sont effectuées séparément pour chaque contrat. Lorsqu'un calcul distinct pour chaque contrat représente une charge induite pour l'entreprise d'assurance ou de réassurance, celle-ci peut regrouper les contrats pour effectuer ces projections, à condition de satisfaire à l'ensemble des exigences suivantes :

1. il n'y a pas de différences significatives dans la nature et la complexité des risques sous-jacents aux contrats faisant partie d'un même groupe;
2. le regroupement de contrats ne crée pas une image faussée des risques sous-jacents aux contrats et n'introduit pas de biais dans leurs coûts;
3. il est probable que le regroupement de contrats produise approximativement les mêmes résultats pour le calcul de la meilleure estimation qu'un calcul contrat par contrat, en particulier pour ce qui concerne les garanties financières et les options contractuelles prévues par les contrats.

Orientations sur la valorisation des provisions techniques, orientation 19 - Déterminer et évaluer l'adéquation d'un groupe de risques homogènes

1.49. Les entreprises d'assurance et de réassurance devraient calculer les provisions techniques en utilisant des groupes de risques homogènes afin d'établir des hypothèses.

1.50. Un groupe de risques homogènes comprend un ensemble de polices présentant des caractéristiques de risque similaires. En sélectionnant un groupe de risques homogènes, les entreprises devraient parvenir à un équilibre approprié entre la crédibilité des données disponibles, permettant de réaliser des analyses statistiques fiables, et l'homogénéité des caractéristiques de risque au sein du groupe. Les entreprises devraient définir des groupes de risques homogènes pour lesquels il est prévu qu'ils demeureront raisonnablement stables au fil du temps.

1.51. Le cas échéant, aux fins de définir les risques, les entreprises devraient tenir compte, entre autres, des éléments suivants :

1. politique de souscription ;
2. modèles de règlement des sinistres ;
3. profil de risque des preneurs d'assurance ;
4. caractéristiques des produits, notamment garanties ;
5. décisions futures de gestion.

## 5.2 Qualité de données

Les textes relatifs à Solvabilité 2 insistent sur la qualité des données en définissant trois critères - exactitude, exhaustivité et caractère approprié - auxquels s'ajoute une exigence de traçabilité. La fonction actuarielle, dans le cadre de son rapport actuariel, doit s'exprimer sur l'impact de la qualité des données sur la fiabilité des provisions techniques.

Ces exigences portent sur les données utilisées pour le calcul des provisions techniques : données de passif, données d'actif, lois et hypothèses. Par extension, toutes les données sources à l'origine de ces données sont également concernées.

Pour répondre à ces exigences, le dispositif suivant peut être envisagé :

- Une gouvernance spécifique qui peut intégrer la mise en place d'un comité Qualité des Données et/ou d'un gestionnaire de la qualité des données (Data Quality Manager). Ceci ayant pour objectif de piloter le dispositif relatif à la qualité des données, dont la collecte des résultats des contrôles et la définition d'un plan d'amélioration continue.
- Une politique relative à la qualité des données et un dictionnaire des données, et plus largement une base documentaire relative à la qualité des données.
- La réalisation de contrôles tout au long de la vie de la donnée : chaque intervenant sur les données sera responsable d'un certain nombre de contrôles qui devront être documentés.

Ce dispositif est mis en place au sein de GPM avec un outil comme QlickView. Le traitement et l'amélioration de la qualité est une problématique gérée régulièrement pour les travaux solvabilité 2. Des tests de cohérence et des échanges ont donc été réalisés pour s'assurer de la fiabilité des données de projection.

### 5.3 Méthode de constitution des model points actifs

Le calcul de la valeur et du rendement des actifs devrait idéalement résulter des projections de chaque actif de manière indépendante. Cependant, en pratique, le très grand nombre d'actifs détenus, le recours à des gestionnaires d'actifs externes et l'investissement sur des fonds communs de placement (OPCVM) ne rendent pas toujours possible une projection ligne à ligne de l'actif.

Pour regrouper les différents actifs détenus, les critères à considérer peuvent être :

- La nature de l'actif (exemples : obligations d'État, obligations d'entreprises dites obligations « corporate », actions, immobilier, swap, etc.) ;
- La maturité ;
- Le type de taux (exemple : fixe, variable, indexé) ;
- L'allocation des actifs (exemple : actifs mis en représentation des fonds propres, du fonds général ou d'un fonds cantonné) ;
- La nature homogène par rapport aux stress de marché dans le cadre du calcul du SCR (par exemple écart entre le strike d'un call ou put et la valeur de marché du titre de référence) ;

La granularité doit également être cohérente avec les classes d'actifs modélisées. Par exemple, lorsque les chocs SCR marché sont appliqués en amont du regroupement des données d'actif, il pourrait être envisagé de ne pas distinguer les obligations d'État et corporate lors du regroupement des données dès lors que ces actifs sont modélisés de façon identique.

Disposer de données d'actif ligne à ligne permet par exemple une projection des provisions pour dépréciation actif par actif.

Par ailleurs, le réinvestissement des actifs au cours de la projection peut mener à la création de nouveaux model points. Leur granularité devra être suffisante pour refléter la politique de réinvestissement future.

De manière générale l'application du principe de proportionnalité devrait guider la granularité retenue pour les classes d'actifs et model points d'actifs du modèle.

### 5.4 Méthode de constitution des model points passifs

Le Règlement Délégué publié par la commission européenne en janvier 2015 précise à l'article 35 que « les projections des flux de trésorerie [...] sont effectuées séparément pour chaque contrat ».

Cependant, le Règlement Délégué indique que des regroupements de polices sont possibles dès lors que les polices considérées et les risques sous-jacents sont similaires, et que l'utilisation de ces regroupements de polices (model points) ne fausse pas les résultats des

projections.

Aucune technique n'est prescrite par la réglementation Solvabilité 2 pour la création de ces regroupements, qui reste donc à définir par chaque compagnie. Le processus de création de model points doit répondre à plusieurs enjeux qui coïncident avec ceux de ce mémoire :

- Réduire le nombre de lignes de données projetées afin d'optimiser les temps de calcul et de faciliter l'analyse ;
- Garantir la qualité des regroupements créés : homogénéité, granularité permettant de refléter correctement les options et garanties et/ou la modélisation des engagements par typologie d'assurés ou de contrats (taux de rachats par canal de distribution/ancienneté/âge/montant de provision, taux de mortalité par âge/sexe, etc.) en conditions « normales » et « stressées »;
- Permettre une analyse pertinente à un niveau de maille utile pour le pilotage de l'entreprise (type de produit, canal de distribution, niveau de garanties, etc.).

Une première technique de construction consiste à regrouper les polices qui ont des caractéristiques semblables :

- Le produit, le réseau et la garantie souscrite;
- L'âge et le sexe de l'assuré;
- L'ancienneté du contrat (épargne);
- Le taux garanti (contrats en euro, prévoyance), le support utilisé (UC/multisupports),  
...

Cette technique permet de construire les model points en regroupant les produits qui se ressemblent à partir de critères prédéfinis. Il s'agit de l'approche la plus répandue dénommée "approche traditionnelle" dans la suite. Une autre approche (moins répandue) consiste à utiliser des méthodes statistiques d'analyse de données. Ces méthodes de construction seront bien décrites dans la suite.

## 5.5 Tester la validité des model points

Afin de valider que les model points n'entraînent pas une erreur significative dans la projection des flux par rapport au calcul police par police, il est nécessaire de comparer l'approche police par police et l'approche par regroupement sur un échantillon représentatif, voire sur la totalité du portefeuille à une fréquence régulière.

Il est nécessaire d'établir un compromis entre le gain de temps de calcul et la perte de précision afin de déterminer le nombre optimal de model points à constituer.

L'absence de déformation matérielle des résultats du fait du niveau de granularité des model points devrait être vérifiée aussi bien sur le BE que sur le BE après choc (car les mêmes model points sont également utilisés pour le calcul des SCR).

Cette procédure de backtesting est développée dans la dernière partie de ce mémoire.

## Part II

# ALM MODEL POINT

# 1 CADRE D'ÉTUDE POUR LE LANCEMENT

## 1.1 CALFITEC : modèle de lancement ALM

CALFITEC est un outil informatique développé par une entreprise de conseil, qui permet d'effectuer des calculs actuariels clés pour les modèles S1 et S2 Vie et non-Vie. Grâce à ses fonctionnalités avancées, cet outil fournit une documentation automatique et une traçabilité complète des traitements, ainsi qu'une bibliothèque pré-paramétrée pour faciliter les opérations. Bien qu'il diffère du modèle interne utilisé par le prestataire, CALFITEC a été préféré pour la modélisation ALM dans ce mémoire en raison de sa grande flexibilité d'utilisation.

Ainsi, l'outil CALFITEC permet lorsqu'il est bien alimenté de produire plusieurs sorties modélisant les interactions actifs passifs. Il est relié au logiciel SQL à travers les entrées (INPUTS) et sorties (OUTPUTS).

Le processus ALM respecte les étapes suivantes :

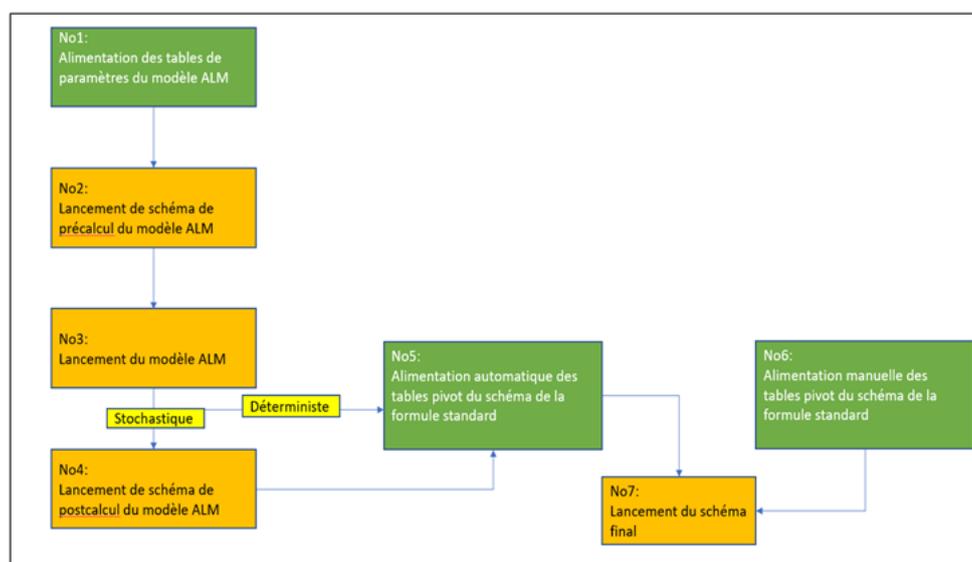


Figure 10: Processus ALM Calfittec

### 1.1.1 Indicateurs solvabilité II sous CALFITEC

L'objectif de ce mémoire est d'évaluer la stabilité et l'impact des model points sur le Best Estimate (BE). Le BE total correspond à l'ensemble des engagements contractuels souscrits par l'assureur, qui peuvent être classés en différentes catégories :

- les provisions destinées à couvrir l'engagement contractuel de taux minimum garanti (TMG) de revalorisation de l'épargne ou de la rente (BEG),
- les provisions destinées au respect de la contrainte réglementaire de participation aux bénéfices (PB),
- les provisions destinées au respect d'une clause de PB contractuelle,
- les provisions destinées à la participation aux bénéfices purement discrétionnaire au-delà des autres seuils,
- la provision pour participation aux bénéfices (PPB) déjà constituée à la date d'inventaire.

Conformément aux dispositions des spécifications techniques de base (point TP.2.102), ces engagements doivent être ventilés en 2 parties une partie garantie du BE appelée BEG et une partie incluant les participations aux bénéfices futurs (appelés Future Discretionary Benefits ou FDB).

$$BE = BEG + FDB$$

Ainsi le modèle ALM nous permet de déterminer le BE total en prenant en compte toutes les hypothèses et le BE garantie (BEG) en excluant les participations aux bénéfices futurs. Ces deux indicateurs seront d'une utilité importante pour mesurer la précision de notre construction de model point.

Des chocs sur les hypothèses retenues pour le modèle sont effectués (chocs intervenant dans le calcul du SCR). Ces chocs (action, immobilier, taux d'intérêt ...) créent de nouveaux indicateurs du BEG et du BE qui vont nous servir pour s'assurer de la stabilité des MP.

### 1.1.2 Préparation des Inputs

Pour construire un bon modèle ALM, il faut bien préparer les données. Ces inputs sont les différents éléments du bilan ainsi que certaines hypothèses économiques :

- **ACAP**: Table d'allocation des différentes classes d'actifs dans le portefeuille (PM);
- **CTAD**: Table de MP du passif et des différentes caractéristiques des adhérents;
- **GSEP**: Table de description du GSE (courbe EIOPA) ;
- **LOSP**: Table des rachats;
- **MSPP**: Choc en fonction de la notation ;
- **PARP**: Frais par canton;
- **PBHD**: Dotation et participation aux bénéfices ;
- **PFAD**: Description et rentabilité des classes d'actifs (VM);
- **PROD**: Réserve de capitalisation.

Notre mémoire porte sur l'optimisation des MP au niveau de la CTAD qui sera étudiée dans la suite.

### 1.1.3 Principaux OUTPUTS de CALFITEC

Grâce au modèle CALFITEC, nous avons une traçabilité précise de chaque étape de la modélisation. Cette fonctionnalité nous permet de tirer parti de plusieurs résultats et d'effectuer des analyses approfondies pour comparer différentes méthodes. Plusieurs tables sont générées par le modèle, mais les principaux outputs qui seront utilisés dans ce mémoire sont les suivants :

- **Sur le passif** : La CTARIS est un outil de projection qui regroupe toutes les évolutions individuelles de la CTAD sur une période de 50 ans. Grâce à cet outil, il est possible de suivre l'évolution complète du portefeuille en termes de PM ou de rachat/décès. Les résultats obtenus à partir de la CTARIS sont utilisés pour construire la table CTARS, qui permet d'estimer le Best Estimate (BE) par point de modélisation. Cette table offre une vision détaillée du calcul du BE en prenant en compte les particularités de chaque model point.

- **Sur l'actif** : La PFARIS regroupe toute la projection tête par tête de la PFAD au bout de 50 ans. Ainsi, il permet d'avoir toute l'évolution du portefeuille en termes de VM ou d'allocation d'actif.
- **Sur l'ensemble du bilan** : Les résultats de la PFARIS sont combinés avec la CTARIS pour construire la table RAPRIS représentant le résumé par année de projection de tous les calculs. Elle donne le détail de l'estimation des cash-flows. Il s'agit de l'évolution du bilan. Ces données sont concaténées pour construire une table RCARS présentant le BE par hypothèse (scenario Central, chocs, BEG...) pour chaque produit.

#### 1.1.4 Modélisation CALFITEC

La modélisation sous CALFITEC se fait en deux principales étapes :

##### a) Pré-calculs

Cette étape consiste à définir les principaux paramètres nécessaires à l'alimentation du modèle ALM. Elle commence par attribuer un code de lancement unique au modèle et initialiser une table pour les projections du passif avec la CTAD. Cette table, nommée  $CTAW_0$ , contient toutes les informations sur les assurés, y compris la PM qui évoluera avec le temps.

Après cette première phase dite d'initialisation, le modèle est configuré en utilisant diverses variables telles que le taux de rachat, le taux de décès, les taux d'actualisation, etc. Cette étape est cruciale, car elle calcule tous les paramètres de marché, tels que les taux forward, les taux choqués, les facteurs d'actualisation associés, les prix unitaires pour le calcul des valeurs de marché des obligations, les probabilités nécessaires au calcul des provisions mathématiques.

En somme, le pré-calcul permet de mettre en place toutes les tables qui auront une incidence sur la projection du modèle (comportementales et économiques). À ce stade, des hypothèses clés, telles que le type de projection (déterministe/stochastique), sont également établies pour assurer une modélisation optimale des données.

##### b) Modélisation

Cette étape concerne le processus de projection des Cash-Flows, qui constitue la modélisation proprement dite. Elle commence par l'initialisation et l'alimentation des tables d'outputs. Cette étape nécessite plusieurs calculs de PM et de VM, notamment pour les rentes à l'itération 1.

Ensuite, le modèle est paramétré en intégrant la table de décès et de rachat pour modéliser l'évolution de la PM et de la VM pour chaque classe d'actifs. Il détermine ainsi une PM à l'ouverture et met en place une politique de revalorisation de la PM à chaque itération. Des calculs intermédiaires sont également effectués, tels que la politique de réassurance ou l'investissement/désinvestissement dans chaque classe d'actifs. À partir de tous ces paramètres, le modèle peut projeter le bilan sur plusieurs années d'itération (50 ans dans ce cas). À la fin de cette étape, les tables des outputs CTARIS (Passifs) et PFARIS (Actifs) sont mises à jour.

Enfin, la dernière étape consiste à déterminer les cash-flows en fonction des hypothèses initiales établies lors du paramétrage. Le modèle calcule les cash-flows dans une table

appelée RAPRIS. Le modèle détermine le BE final en actualisant les cash-flows.

## 1.2 Présentation des données brutes

Les données brutes du passif sont consignées dans la CTAD. Chaque ligne représente un assuré et ses principales caractéristiques. Ce tableau nous donne la description des principales variables de la CTAD.

Variables	Description
DT_IMAGE	Date de la période clôture
FAMPDT	Type de produit (Épargne ou Rente)
SEXE_ASSURE	Sexe de l'adhérent
SEXE_BENEF	Sexe du bénéficiaire
DT_NAISS_ASSURE	Date de naissance de l'assuré
DT_EFFECT	Date de début de contrat (permet de déterminer l'ancienneté)
DT_NAISS_BENEF	Date de naissance du bénéficiaire
NB_CONTRATS	Nombre de contrats
TX_TAUX_REVALO	Taux Minimum Garanti (TMG)
MT_ARRERAGE_RENTES	Arrérage de rente pour les contrats de rente
MT_PM_INIT	Montant de la PM
IND_REGIME_PTS	Coefficient de majoration légal "O" si oui ou "N" non
TYPE_COTIS	Type de cotisation (1 : annuelle, 2 : semestrielle, 4 : trimestrielle)
ETAT_CONTRAT	État du contrat pour les rentes
PARAM_CAD	si l'assuré a un capital décès (1 si oui 0 sinon).

Table 2: Description des variables de la CTAD

### 1.2.1 ÉPARGNE

La base d'épargne initiale comprend un total de **10 949 lignes**. Pour notre étude, nous avons sélectionné la variable qualitative "sexe" ainsi que les variables quantitatives "date d'effet" et "date de naissance", jugées pertinentes pour notre analyse. Afin de garantir la fiabilité des résultats, nous avons procédé à un retraitement visant à détecter toute anomalie éventuelle, telles que des dates de naissance ou d'effet non renseignées, un montant PM nul ou une incohérence dans la cohérence du sexe ou de la date entre N et N-1. Les résultats des tests mentionnés ci-dessus sont présentés dans le tableau suivant.

Test de cohérence et de stabilité	Nombre de lignes concernées	Pourcentage de lignes concernées
Date de naissance non renseignée	0	0%
Date d'effet non renseignée	0	0%
Montant PM nul	0	0%
cohérence sexe N vs N-1	4	0,04%
date de naissance qui change entre N et N-1	7	0,06%
date d'effet qui change entre N et N-1	0	0%

Table 3: Récapitulatif des tests de cohérence sur le produit d'épargne

Dans le cadre de notre étude, nous avons utilisé les données de l'année 2020 et 2021 pour effectuer des tests visant à évaluer la stabilité de notre portefeuille. Les résultats obtenus ont montré une certaine stabilité, avec un faible pourcentage d'anomalies détectées. Toutefois, nous avons identifié quelques anomalies qui peuvent être expliquées par la gestion et ne nécessitent donc ni traitement ni suppression de lignes. Nous avons donc pu conclure que notre portefeuille était globalement fiable.

Par ailleurs, pour garantir la qualité de nos données, nous avons utilisé des statistiques descriptives afin de vérifier l'absence de valeurs aberrantes.

- **Minimum/ Maximum des Dates :**

	Minimum	Maximum
Date de naissance	16/08/1912	30/03/2021
Date d'effet	15/11/1990	14/12/2021

Table 4: Récapitulatif des dates clés (épargne)

- Répartition des Adhérents et de la PM par Age :

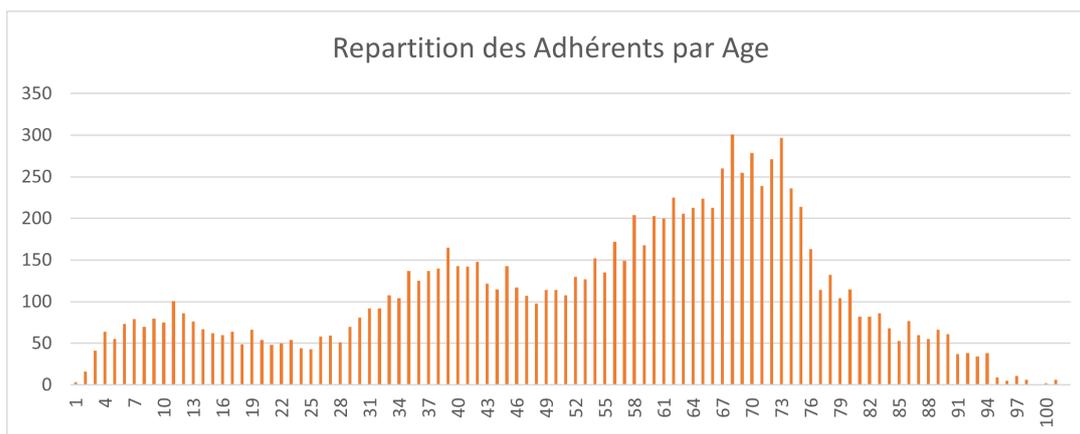


Figure 11: Répartition des Adhérents par Age

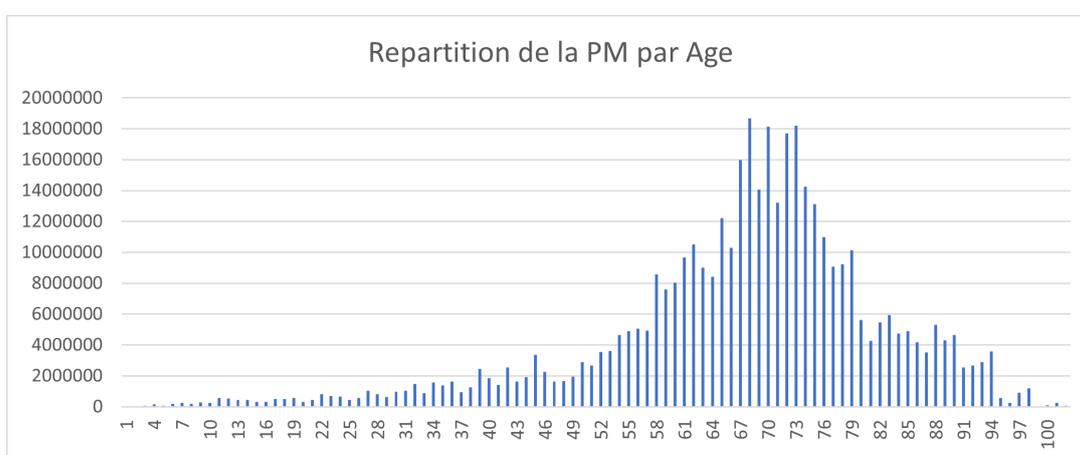


Figure 12: Répartition de la PM par Age

L'âge des assurés est un critère clé pour déterminer la prime d'assurance et donc la PM. Notre analyse a mis en évidence que l'âge moyen des assurés dans notre portefeuille est de 54 ans. Pour étudier l'impact de l'âge sur la PM, nous avons utilisé deux graphiques qui montrent une corrélation significative entre ces deux variables.

Le premier graphique montre une augmentation de la PM jusqu'à l'âge de 70 ans, où elle atteint son pic avant de diminuer légèrement. Le deuxième graphique quant à lui, illustre une concentration importante de la PM et du nombre d'adhérents entre 58 et 80 ans. Cette observation peut s'expliquer par le fait que les assurés de cette tranche d'âge sont généralement plus enclins à souscrire à des contrats d'assurance.

- Répartition de la PM par Ancienneté :

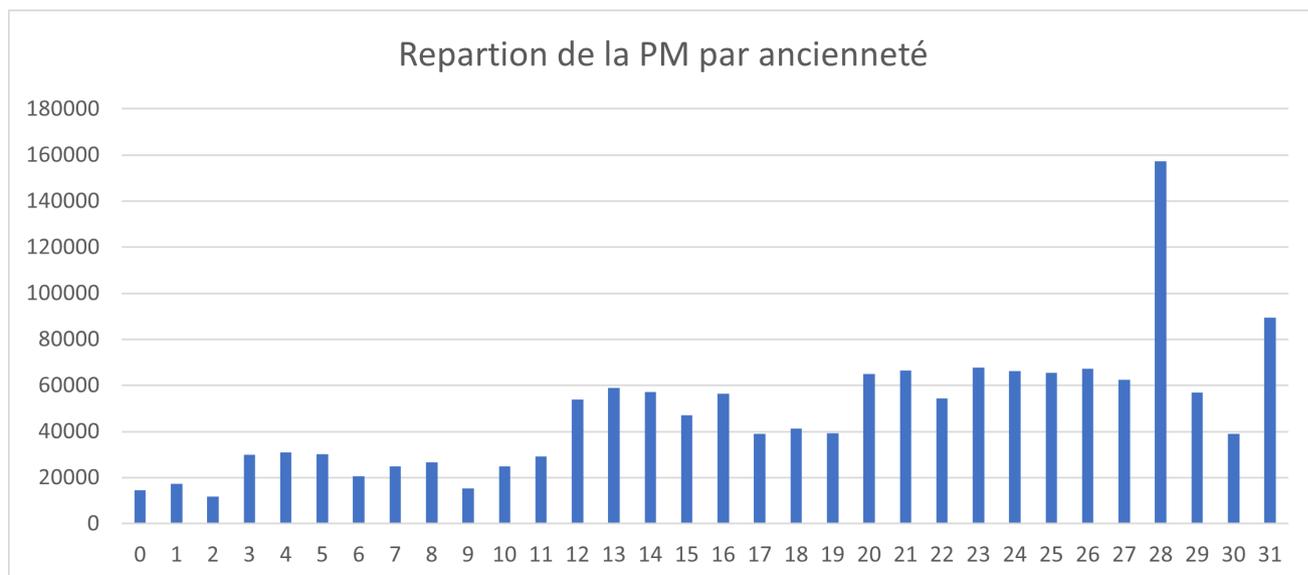


Figure 13: Répartition de la PM pondérée par Ancienneté

L'analyse de l'impact de l'ancienneté sur la PM révèle des résultats intéressants. En effet, la variable d'ancienneté semble avoir un impact assez homogène sur la PM. Toutefois, une étude plus fine de la distribution de la PM selon l'ancienneté a mis en évidence des pics aux niveaux des anciennetés les plus élevées, en particulier à 28 ans.

Cette observation peut s'expliquer par le fait que les assurés ayant une longue ancienneté dans le portefeuille ont généralement souscrit des contrats plus avantageux en termes de garanties et de taux de rentabilité, ce qui se traduit par une PM plus élevée.

- Répartition des Adhérents par sexe :

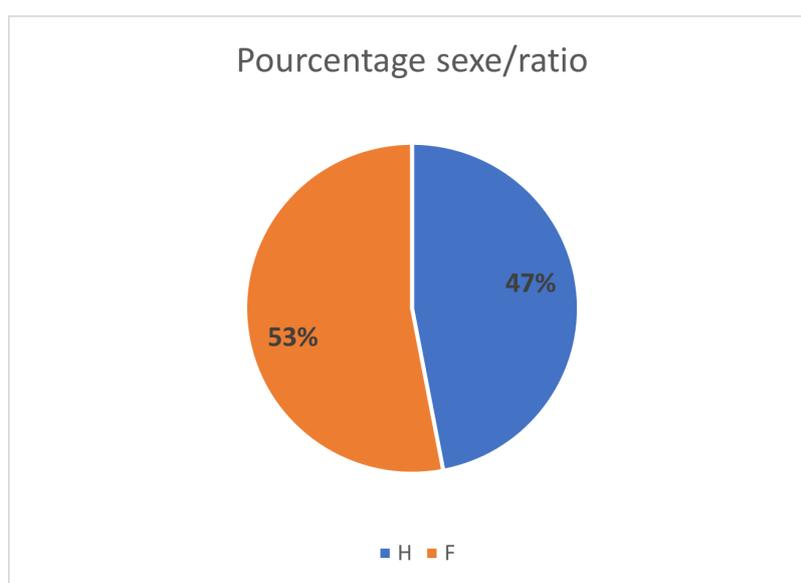


Figure 14: Répartition des Adhérents par sexe

La table CTAD est assez équilibrée avec une légère dominance féminine. Les femmes représentent 53% de notre base contre 47% des hommes.

### 1.2.2 RENTE

La base initiale contient **6 562 lignes**. Notre étude a pris en compte un certain nombre de facteurs pour permettre une analyse approfondie. Parmi eux, nous avons choisi de retenir cinq variables qualitatives ainsi que six variables quantitatives, toutes jugées pertinentes pour notre objectif d'analyse. Les tests nécessaires ont ensuite été appliqués à l'ensemble de cette sélection afin d'obtenir des résultats fiables et significatifs.

Test de cohérence et de stabilité	Nombre de lignes concernées	Pourcentage de lignes concernées
Date de naissance non renseignée	0	0%
Date d'effet non renseignée	0	0%
Montant d'arrérages de rente nul	135	1,94%
Montant PM nul	581	8,36%
cohérence sexe N vs N-1	0	0%
date de naissance qui change entre N et N-1	1	0,01%
date d'effet qui change entre N et N-1	2	0,03%

Table 5: Récapitulatif des tests de cohérence sur le produit rente

Pour évaluer la qualité de nos données, nous avons effectué des tests sur les données relatives à l'année 2020 et 2021. Nous avons constaté que pour ce produit de rente, un certain nombre de valeurs manquantes ou nulles étaient présentes sur nos variables quantitatives essentielles. Pour que notre modèle puisse fonctionner correctement, nous avons donc décidé de supprimer 581 lignes pour lesquelles la PM et le montant d'arrérages de rente étaient nuls.

- **Minimum/ Maximum des Dates :**

	Minimum	Maximum
Date de naissance	01/01/1914	23/09/1976
Date d'effet	27/12/1964	30/12/2004
Date de naissance bénéficiaire	22/06/1911	19/01/1979

Table 6: Récapitulatif des dates clés (Rente)

- Répartition du montant de rente et de la PM par type de cotisation :

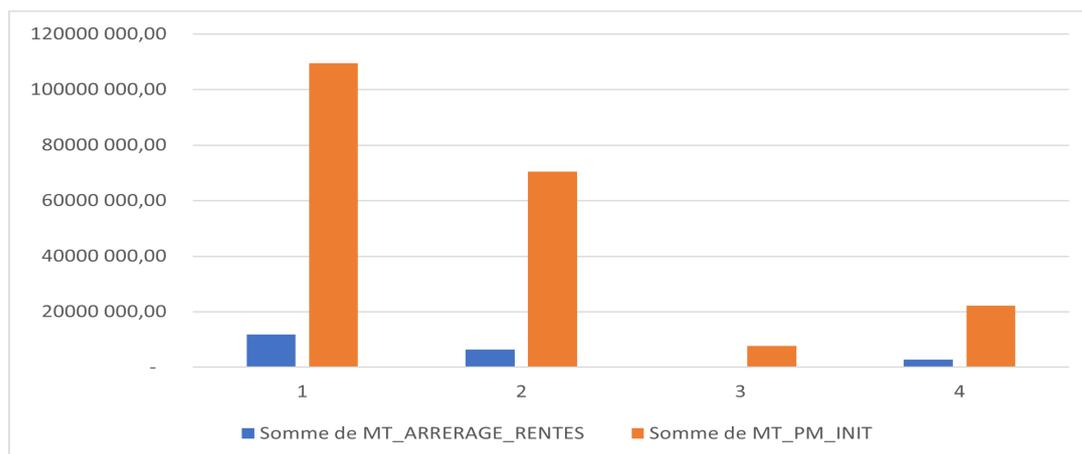


Figure 15: Répartition du montant de rente et de la PM par type de cotisation

La cotisation annuelle est la plus souscrite dans notre portefeuille et regroupe l'essentielle de la PM (rente).

- Répartition du montant de rente par état du contrat :

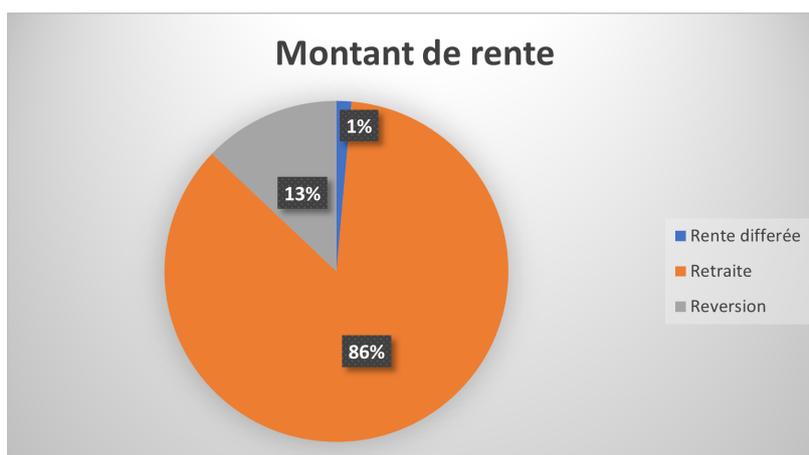


Figure 16: Répartition du montant de rente

- Répartition des Adhérents par Age :

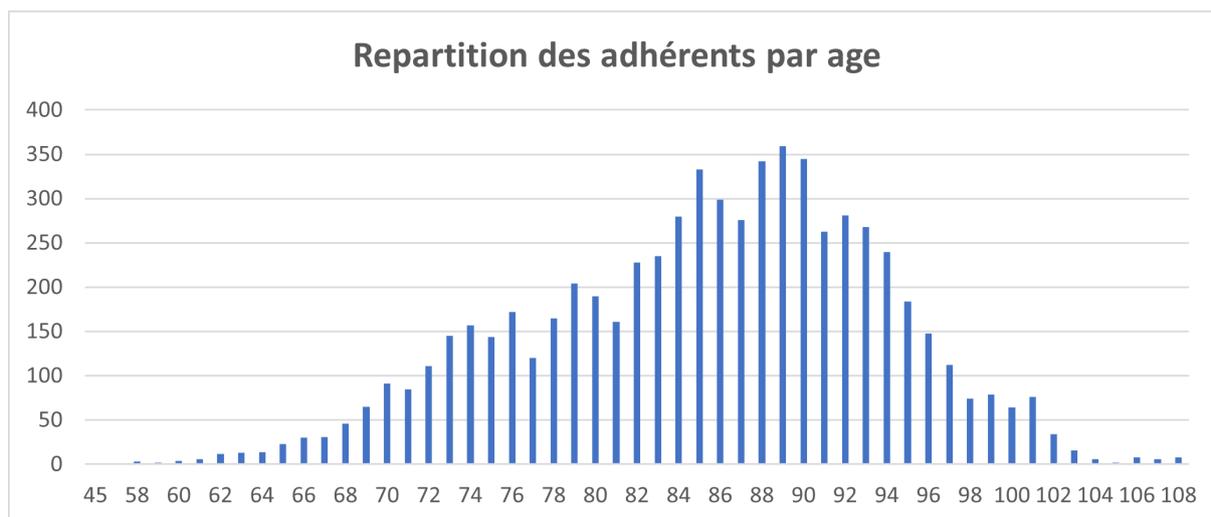


Figure 17: Répartition des Adhérents par Age

Étant donné que les adhérents de ce portefeuille sont principalement en situation de retraite, l'âge moyen est de 85 ans avec un âge minimum de 45 ans.

- Répartition des Adhérents par sexe :

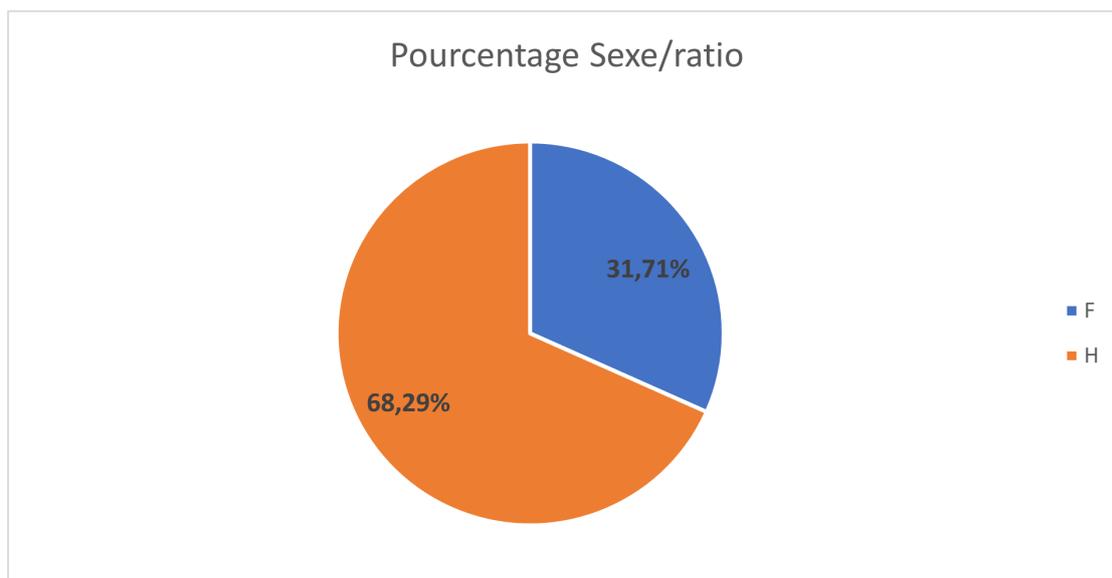


Figure 18: Répartition des Adhérents par sexe

La table CTAD est à dominance masculine contrairement à l'autre portefeuille d'épargne. Les hommes représentent 68% de notre base contre 32% des femmes.

### 1.2.3 LANCEMENT INITIAL

Nous avons utilisé les informations de la table du passif des deux produits en 2021 pour lancer une analyse des données brutes de manière déterministe et avons enregistré les résultats dans le tableau ci-dessous. Ces résultats seront utilisés comme base pour nos futurs modèles. Ces modèles devront être conçus de manière à réduire le nombre de contrats et le temps de calcul tout en minimisant l'écart par rapport à ces estimations.

	Épargne	Rente
Temps d'exécution	2 h 33 min 46 s	1 h 28 min 51 s
Nombre de contrats	10 949	5 982
BEG_CENTRAL	412 625 137,56 €	223 938 699,28 €
BE_CENTRAL	443 570 096,51 €	276 283 270,12 €

Table 7: Résultat de lancement

## 2 MODEL POINT

Cette section est consacrée à la mise en place des Model Points (MP). Comme mentionné dans l'introduction, les MP jouent un rôle crucial dans le processus ALM, en raison de la complexité des modèles internes. L'objectif est donc de regrouper les contrats en groupes homogènes de MP, en utilisant un ratio de compression qui peut varier selon les modèles, qu'ils soient lancés de manière stochastique ou déterministe.

Cependant, cette opération d'agrégation entraîne inévitablement une perte d'information, ce qui justifie l'attention particulière accordée à la précision des estimations dans ce mémoire. En plus de chercher à réduire le temps de calcul, il est donc important de trouver la méthode de regroupement optimale qui permettra d'obtenir le nombre minimal de MP tout en garantissant la représentation la plus précise possible de l'ensemble des polices.

Il existe plusieurs approches pour regrouper les contrats, relevant du domaine du clustering en data science. Dans le cadre de cette recherche, nous avons décidé de tester trois approches différentes afin d'identifier la plus appropriée pour notre modèle :

- **L'approche générale** : cette méthode consiste à regrouper en fonction de caractéristiques similaires et évidentes.
- **L'approche supervisée** : La méthode la plus connue est l'algorithme CART. Nous construisons des arbres de décision avec une variable cible
- **L'approche non supervisée** : la classification ascendante hiérarchique (Cah) et Kmean

### 2.1 Approche générale

Lors de l'établissement de Model Points, il est possible de réunir certains critères avant de procéder à la réduction du nombre de polices. Ainsi, le regroupement doit être effectué en utilisant des critères qui favorisent la création de groupes homogènes, qui peuvent être définis à partir de l'information brute (CTAD) ou déterminés a posteriori.

L'utilisation de l'approche générale peut nécessiter un regroupement préalable du nombre de tarifs, qui consiste à rassembler les tarifs dominants en termes de représentativité avec les tarifs similaires en termes de caractéristiques, mais dont le volume est peu important. À la fin de cette étape, le nombre total de Model Points doit rester inchangé, seul le nombre de tarifs étant réduit. Cette étape peut être nécessaire lorsque le modèle interne utilisé nécessite un Model Point distinct pour chaque tarif, mais que la complexité ne

permet pas de réaliser des calculs stochastiques pour chaque tarif.

Cependant, chaque tarif ne peut pas être regroupé avec n'importe quel autre tarif disponible. En effet, certains critères doivent toujours être pris en compte lors des calculs, en fonction du modèle interne utilisé et du cadre de conformité pouvant être utilisé. Par exemple, il est impossible de regrouper un produit en capital différé avec un produit en rente différée viagère.

Cela nécessite donc une étude approfondie de son portefeuille et les différentes statistiques des variables. Cette méthode fait beaucoup appel à notre vision (intuition) du portefeuille.

Les valeurs de variables ne correspondant pas à un critère de regroupement seront définies de calcul de moyennes simples ou alors pondérées ou par une somme. L'ancienneté par exemple est obtenue avec la variable date d'effet par moyenne pondérée (de même l'âge avec la date de naissance).

Cependant, cette méthode présente des limites. En effet, il est difficile de contrôler le nombre de polices souhaitées après regroupements. Cela demande un ajustement itératif de la méthode jusqu'à ce qu'on trouve un nombre suffisant de polices et diminue la précision du modèle. La qualité de ce regroupement est vraiment aléatoire, car on ne considère pas l'évolution des contrats dans le temps, mais juste au moment de l'extraction de la base de données. Selon le modèle interne utilisé, les modèles points pourront réagir de manière différente lors des projections.

Par ailleurs, l'identification des critères de regroupement permettant d'obtenir de bons résultats n'est pas facile et cela peut entraîner une répétition du processus jusqu'à obtention de meilleurs regroupements. Cela est dû au caractère intuitif de la méthode qui demande beaucoup de vigilance et de connaissance du portefeuille lors du regroupement.

La thèse de Goffard [GG15] est un bon moyen pour le lecteur pour comprendre les réserves concernant cette méthode ainsi que l'utilisation d'une autre méthode.

Cette méthode est par ailleurs la plus utilisée par les assureurs par sa simplicité. Mais il existe d'autres méthodes, dites de clustering (classification), qui proposent des processus et méthode mathématiques pour comprendre, organiser et synthétiser l'information.

## 2.2 Introduction au clustering

L'un des objectifs majeurs de la data science est de parvenir à une classification efficace de données. C'est pourquoi la littérature sur les méthodes de clustering est abondante. La classification consiste en effet à regrouper des individus présentant des comportements similaires ou des caractéristiques communes.

Cette méthode est très utilisée en data science pour analyser et segmenter les données en groupes homogènes, afin de faciliter leur interprétation et de pouvoir appliquer des traitements adaptés à chaque groupe. Le principe de regroupement est donc fondamental pour permettre une analyse approfondie des données, en offrant une vision synthétique et claire de la structure sous-jacente. Il faut donc :

- Une grande similarité intra groupe afin d'assurer de l'homogénéité entre chaque classe ;

- Une faible similarité inter groupe afin d'assurer une grande différence entre chaque classe.

Cette similarité se déduit à travers de nombreux algorithmes reposant sur des notions de distance et d'inertie. Le clustering consiste en la segmentation d'un jeu de donnée en classes (hiérarchisées) selon certains critères.

En annexe, vous trouverez un cours de Denis Clot [1](#) expliquant bien cette notion de distance et inertie.

Pour résoudre un problème de classification, le choix de la méthode de segmentation est la plupart du temps complexe et doit tenir compte de nombreux points comme évoqué dans [\[COU13\]](#):

- Le type de données que l'algorithme peut traiter ;
- la dimension du jeu de données sur l'algorithme peut travailler ;
- l'hétérogénéité de l'ensemble des données et leurs dépendances ;
- le coût en termes de temps ;
- le nombre de paramètres initiaux à compléter ;
- la possibilité d'interpréter ou non les résultats.

### 2.3 Approche non supervisée : Cah

Les méthodes de classification non supervisée sont utilisées pour structurer une base de données d'individus caractérisés par des variables quantitatives en formant des groupes homogènes d'individus. En considérant une base de données  $\mathbf{X}$  de  $N$  individus, l'objectif est de les regrouper en  $K \leq N$  populations distinctes, sans connaître à l'avance les règles d'appartenance d'un individu  $i \in [1, N]$  à une population  $k \in [1, K]$ . Cette appartenance est établie en se basant sur la similarité entre les valeurs des variables quantitatives associées à chaque individu. Dans notre étude, les variables quantitatives utilisées sont les cashflows obtenus après lancement des données brutes.

La classification ascendante hiérarchique (CAH) permet de former des partitions emboîtées en réalisant des regroupements successifs des individus. Ce processus aboutit à la création de partitions de  $N, N - 1, N - 2, \dots, 2$  éléments. Le résultat de l'algorithme est représenté sous forme de dendrogramme, comme illustré dans la figure ci-dessous. Il est également possible de réaliser une coupure au niveau adéquat du dendrogramme, afin de déterminer une partition de  $K$  classes distinctes.

Pour obtenir cette classification, l'algorithme construit dans un premier temps un partitionnement à  $N$  singletons, à partir des  $N$  individus disponibles dans la base des données à partitionner. Ensuite, l'algorithme calcule la matrice des distances entre les individus. Finalement, les partitions à  $N - 1; N - 2; \dots; 2$  classes sont créées de proche en incluant dans chaque classe l'individu ayant la plus petite distance à la classe en question.

Ainsi, cette méthode requiert de définir non seulement une mesure de distance entre deux individus, mais aussi une mesure de distance entre une classe et un individu.

Plusieurs critères existent pour réaliser la fusion, générant ainsi différents types de classification ascendante hiérarchique. Cinq critères courants sont choisis : le saut minimum,

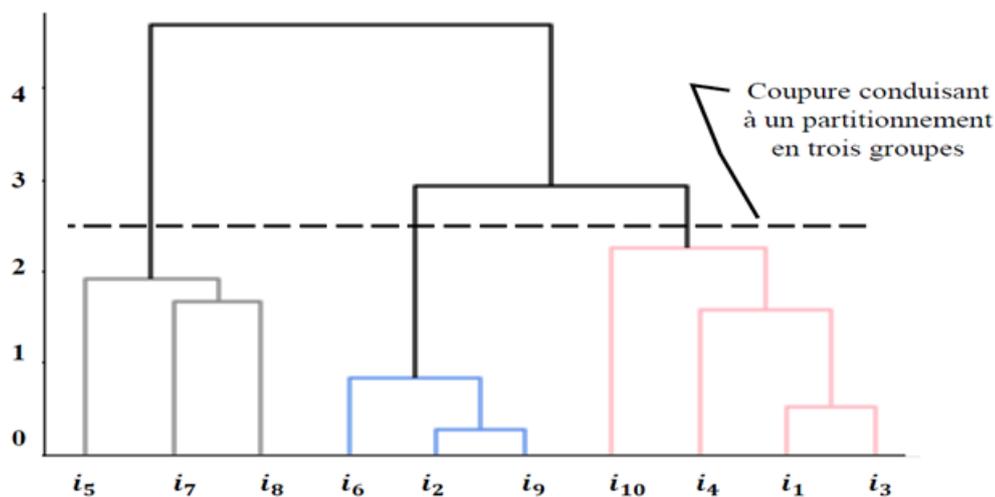


Figure 19: Illustration du dendrogramme obtenu en appliquant la classification ascendante hiérarchique sur 10 individus indexés par  $i_{1 \leq l \leq 10}$

le saut maximum, le saut moyen, le saut proportionnel et la méthode de Ward.

Le critère de Ward est celui retenu pour ce mémoire, les différents critères de mesure distance ainsi que le détail de l'algorithme du Cah sera présenter en **annexe 2**.

La classification ascendante hiérarchique (Cah) est une méthode de classification qui présente les avantages suivants :

- On travaille à partir des dissimilarités entre les objets que l'on veut regrouper. On peut donc choisir un type de dissimilarité adapté au sujet étudié et à la nature des données.
- L'un des résultats est le dendrogramme, qui permet de visualiser le regroupement progressif des données. On peut alors se faire une idée d'un nombre adéquat de classes dans lesquelles les données peuvent être regroupées.

Cependant, elle présente les inconvénients suivants :

- Les regroupements sont définitifs, ce qui ne permet pas d'optimisation postérieure au clustering.
- Lourdeur des calculs dès qu'on a un nombre de données important.

Il est important de souligner que la classification non supervisée est utile pour agréger et structurer une base de données, sans avoir besoin de connaître à l'avance les groupes d'appartenance de chaque individu. Cependant, elle ne fournit pas nécessairement de résultats précis ou interprétables, car elle est basée uniquement sur les caractéristiques quantitatives des individus.

## 2.4 Approche non supervisée : Kmean

L'algorithme Kmean permet d'aborder un autre type de problème de clustering : le nombre de classes à former étant fixé à K, on cherche à créer K groupes d'individus aussi homogènes que possible.

Étant donné une partition de départ en  $K$  groupes, on détermine pour chaque groupe son centre de gravité, puis on reforme les groupes en associant ensemble les points qui sont les plus proches d'un centre de gravité. La procédure est itérée jusqu'à satisfaction d'un critère d'arrêt (généralement la stabilisation des groupes). Il faut bien comprendre qu'à l'issue des itérations, on espère avoir une partition de bonne qualité, mais il n'y a aucune garantie d'optimalité globale (par rapport à l'inertie intra-classe).

En résumé, La méthode repose sur un algorithme itératif et sur une mesure de distance entre deux individus. Elle consiste à réitérer les étapes suivantes afin de minimiser un critère mesurant la qualité du partitionnement : le critère utilisé est l'inertie intra-classe.

- **Initialisation:**  $K$  éléments sont tirés aléatoirement dans la base de données à segmenter, ces derniers correspondent dans un premier temps aux centres des classes ;
- **Affectation:** l'algorithme parcourt l'ensemble des individus et les affecte à la classe dont le barycentre est le plus proche ;
- **Représentation :** le barycentre des classes est recalculé à chaque allocation d'un individu à une classe.

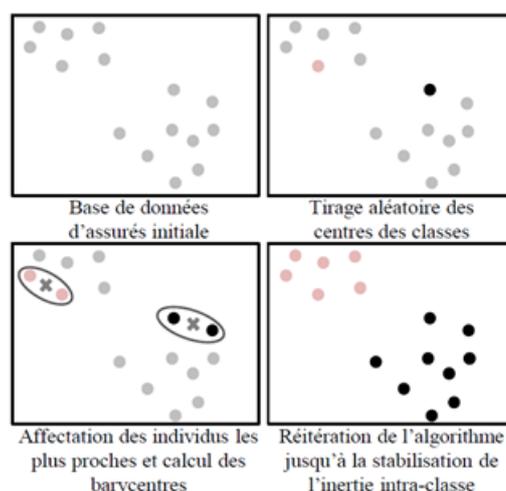


Figure 20: Illustration de la méthode des Kmean

La classification Kmean présente ainsi de nombreux avantages, notamment :

- Un objet peut être affecté à une classe au cours d'une itération puis changer de classe à l'itération suivante, ce qui n'est pas possible avec la classification ascendante hiérarchique pour laquelle une affectation est irréversible.
- En multipliant les points de départ et les répétitions, on peut explorer plusieurs solutions possibles.
- Applicable à des données de grandes tailles, et aussi à tout type de données (mêmes textuelles), en choisissant une bonne notion de distance.
- Sa simplicité conceptuelle et sa rapidité.

Cependant, elle a les inconvénients suivants :

- L'inconvénient de cette méthode est qu'elle ne permet pas de découvrir quel peut être un nombre cohérent de classes, ni de visualiser la proximité entre les classes ou les objets.
- Le résultat dépend de tirage initial des centres des classes.

## 2.5 Approche supervisée : CART

La méthode Classification And Regression Trees (CART) est une méthode introduite par Breiman, Friedman, Olshen et Stone en 1984 (cf. [BFOS17]). Cette méthode peut être utilisée aussi bien en classification qu'en régression.

C'est une méthode qui permet de construire des arbres de décision. Un arbre est généralement constitué de nœuds de décision (construits à partir des variables caractérisant les individus), et de feuilles.

Dans le but de comprendre le mode de construction des arbres de décision, nous considérons un exemple simple expliqué dans le mémoire [Oth] qui s'approche davantage de notre problématique. Supposons que l'on dispose d'une base contenant 26 assurés caractérisés par leur âge et leur ancienneté répartie en trois classes de couleurs différentes dans la figure 21.

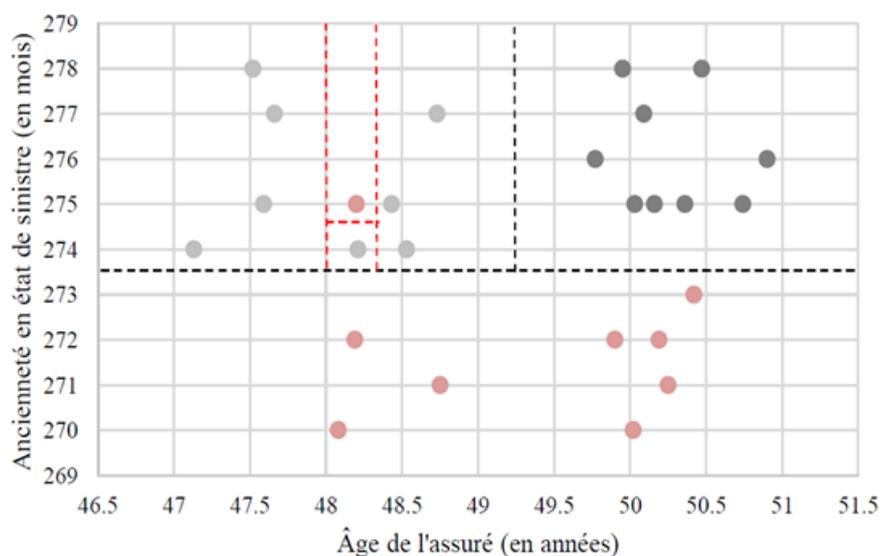


Figure 21: Exemple de classification par la méthode CART

Dans ce mémoire, la méthode CART sera notre indicateur pour déterminer le nombre de classes optimal. En effet, considérons pour chaque  $k \in 1, \dots, n$ , le “meilleur” arbre de taille  $k$ , autrement dit l'arbre à  $k$  feuilles qui minimise l'erreur de classification ou les résidus suivant le cadre d'étude. le modèle CART construit cette suite seulement implicitement, en fait, une sous-suite est astucieusement choisie. Ensuite, à l'aide d'un critère pénalisé, sont comparés ces différents arbres et on en sélectionne un, le critère pénalisé mis en œuvre servant à réaliser un compromis entre la taille des arbres et leur qualité d'ajustement. En effet, l'objectif est de trouver un compromis entre l'erreur d'apprentissage et l'erreur de généralisation. Or, un arbre de grande taille est trop fidèle aux données d'apprentissage, ce qui a pour conséquence d'obtenir une estimation de l'erreur de généralisation trop optimiste, car trop faible. À l'inverse, un arbre de petite taille engendre une perte de précision et une estimation de l'erreur de généralisation trop grande. Ainsi, l'arbre sélectionné a une taille ni trop grande ni trop petite.

Dans la pratique, afin de déterminer le “meilleur” arbre, le modèle procède en trois étapes que sont :

- étape 1 : Construction de l'arbre maximal  $T_{max}$ .

De manière récursive et dyadique, on construit une suite de partitions de plus en plus fines de l'espace des observations  $X$  jusqu'à ce que chacun des éléments de la partition ne contienne qu'une seule observation ou des observations de même réponse. Cela revient à construire un arbre en le développant au maximum, autrement dit jusqu'à ce que les nœuds terminaux ne contiennent que des observations de même réponse. À chacune de ces partitions est associé un arbre. L'arbre maximal, noté  $T_{max}$ , est celui associé à la partition la plus fine.

- étape 2 : Élagage de  $T_{max}$ .

De la suite d'arbres précédemment obtenue, on extrait une sous-suite par minimisation, pour  $\alpha \geq 0$ , du critère pénalisé

$$crit_{\alpha}(T) = \gamma_n(S_t) + \alpha \frac{|T|}{n},$$

où  $\gamma_n$  est l'erreur quadratique définie par

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

et  $|T|$  le nombre de feuilles de l'arbre T.

Ce critère permet de faire un compromis entre la taille de l'arbre et l'information retenue.

- étape 3 : Sélection finale.

Cette dernière phase sélectionne, à l'aide d'un échantillon témoin ou par validation croisée, le "meilleur" arbre dans la suite précédemment construite.

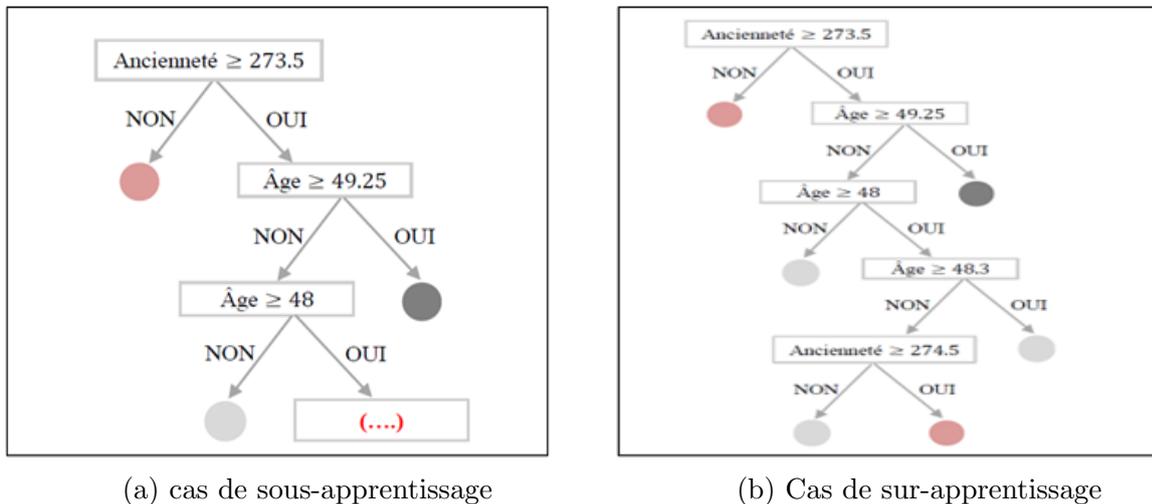


Figure 22: Quel est l'arbre optimal ?

Dans le premier scénario, la classification ne capture pas suffisamment d'informations et peut donc être améliorée. Dans le second cas, l'arbre de décision souffre d'un sur-apprentissage, ce qui le rend inexploitable. Il est important de rappeler que l'objectif final est de généraliser la classification à de nouveaux individus qui n'ont pas été inclus dans l'ensemble d'apprentissage initial.

Pour résoudre cette problématique, nous utilisons la technique de l'élagage de l'arbre qui permet de définir une profondeur optimale pour l'arrêt du développement de l'arbre.

Pour notre étude, nous créons des arbres de décision avec différentes profondeurs possibles, jusqu'à la profondeur maximale qui permet de s'ajuster aux données d'apprentissage. Ensuite, nous évaluons la précision de chaque arbre construit à l'aide de la méthode de validation croisée mentionnée précédemment. La profondeur optimale sera celle où la précision ne s'améliore plus de manière significative.

Bien que les arbres de décision présentent plusieurs avantages tels que leur interprétabilité, leur rapidité d'exécution et leur capacité à fonctionner sans hypothèse préalable sur les données, ils présentent également certaines limitations. L'instabilité est la principale limite des arbres de décision. En effet, l'arbre construit par l'algorithme décrit précédemment est très sensible aux données d'apprentissage. Un changement même minime dans les observations ou les variables utilisées peut conduire à la création d'un arbre de décision complètement différent, ce qui affecte la précision de la prédiction de l'algorithme lorsqu'il est utilisé pour prédire de nouvelles observations.

### 3 CONSTRUCTION DES MP

#### 3.1 méthode générale (standard)

Pour ce mémoire, l'approche générale utilisée est inspirée de celle mise en place par le prestataire en 2021. Elle a été revue par des recommandations de l'ACPR entre 2020 et 2021. Elle a été ajustée pour prendre en compte les évolutions survenues cette année.

Le travail d'agrégation des données du passif a été effectué afin de réduire les lignes de Model Points et ainsi optimiser les temps de calculs et de façon homogène en termes de caractéristiques déterminantes pour la projection afin de correctement modéliser les flux de passif.

L'approche générale présentée dans ce mémoire se base sur plusieurs critères de regroupement qui sont selon le type de produit :

- Age et ancienneté pondérés
- Le "Sexe pondéré" (épargne) : model point considéré comme entièrement des femmes si la part des PM des femmes est supérieure à celle des hommes et inversement pour les hommes.
- État Contrat et type de cotisation.

Les model points ont été regroupés de la façon suivante :

Produit	Classe ancienneté	Classe d'âge	Sexe	Etat contrat	Type de cotisation	Nb de MP
Epargne	De 0 à 8 par pas de 1; ]8,12]; et >12	De 0 à 80 par pas de 10 ; >80				87
Rente	De 1 à 8 par pas de 1; ]8,12]; et >12	[0,20];[20,30];[30,40];[40,50];[50,60]; de 60 à 80 par pas de 1; >80	F & H	Rente différée Réversion Retraite	1; 2; 3 ou 4	118

Une fois les MP regroupés, les variables utilisées sont recalculées en effectuant des moyennes pondérées par la PM. Pour exemple si on considère  $x_1, \dots, x_n$  et  $PM_1, \dots, PM_n$  l'âge et la PM respective de n individus constituant un MP, l'âge pondéré x du MP se calcule :

$$x = \sum_{i=1}^n \frac{x_i PM_i}{PM} \quad (3)$$

Ainsi, Le choix d'agrégation des MP en fonction de l'ancienneté est discriminant dans la mesure où la loi de rachat est calibrée en fonction de l'ancienneté. Les MP ont été agrégés par TMG car le coût des options et garanties est lié au TMG.

### 3.2 Procédure de construction

Dans les conditions définies précédemment, on construit des MP pour chaque type de produit à partir des différentes méthodes selon une procédure bien définie.

En effet, des données brutes (non agrégées) sont mises à notre disposition dans une table rassemblant les informations des adhérents ainsi que leur PM initiale. Avant de commencer tout regroupement, un lancement déterministe est effectué pour constituer une base de clustering (données non agrégées et sorties déterministes). Cela est nécessaire pour disposer de variables quantitatives pour les méthodes de clustering.

Une fois la base construite, les étapes qui suivent doivent être respectées pour chaque méthode algorithmique :

- **Première partition** : la manière de calculé les engagements peut changer d'une catégorie de contrat à une autre. C'est pour cela qu'on réalise une première classification sur certaines variables qualitatives importantes pour une projection de flux ("retraite" ou "rente différée")
- **Détermination du nombre de classes** : le nombre de classes est le nombre de MP. Pour chaque première classe de la première partition, un élagage est effectué pour avoir le nombre optimal de classes minimisant l'erreur quadratique. Cette minimisation s'établit à l'aide du logiciel R.
- **Regroupement algorithmique** : les paramètres étant établis, les modèles de classification sont lancés et les données agrégées.
- **Lancements** : On procède au lancement stochastique ou déterministe pour avoir les indicateurs de solvabilité 2.

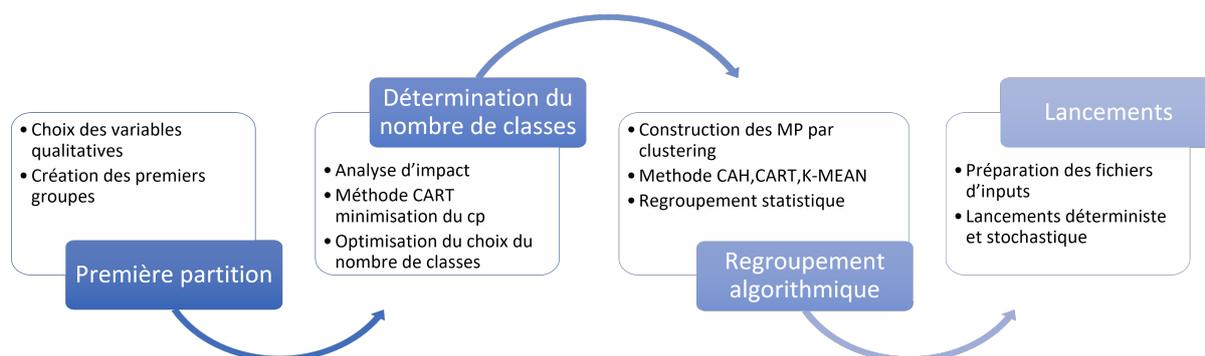


Figure 23: Étape de construction des Model Points

### 3.3 Épargne

Pour ce produit spécifique, il y a peu d'informations contractuelles disponibles pour élaborer des modèles complexes. Les variables clés exploitables sont l'âge et l'ancienneté qui proviennent de la table du passif (données non agrégées). Cependant, nous pouvons également utiliser les résultats de la projection initiale des données brutes pour exploiter

d'autres variables. Par conséquent, pour le modèle CART, nous incluons comme variables explicatives l'âge, l'ancienneté. La variable cible est le BE CENTRAL (sortie déterministe).

On réalise la première partition en 2 classes prenant en compte le sexe :

classe	caractéristique	nombre de contrats
1	H	5 774
2	F	5 175

Table 8: Première partition

### 3.3.1 CART

Le modèle CART est pertinent pour ce produit au vu du peu de variables et le nombre de contrats reste raisonnable. Ce modèle facile à mettre en place et à interpréter constitue la base dans la détermination du nombre de classes finales.

Ainsi, nous cherchons le nombre de classes optimal par tranche prédéfini en ajustant le paramètre `cp` du modèle CART. On réalise un élagage grâce au `plotcp` introduit par le package `rpart` en figure 24 de R [R C23]. Celle-ci est à cet effet essentiel pour évaluer la pertinence d'un arbre à travers la moyenne et l'écart type des erreurs dans la prédiction à validation croisée par rapport à chacune des moyennes géométriques. On cherche ainsi à choisir le `cp_optimal` qui minimise l'erreur d'estimation pour chaque tranche.

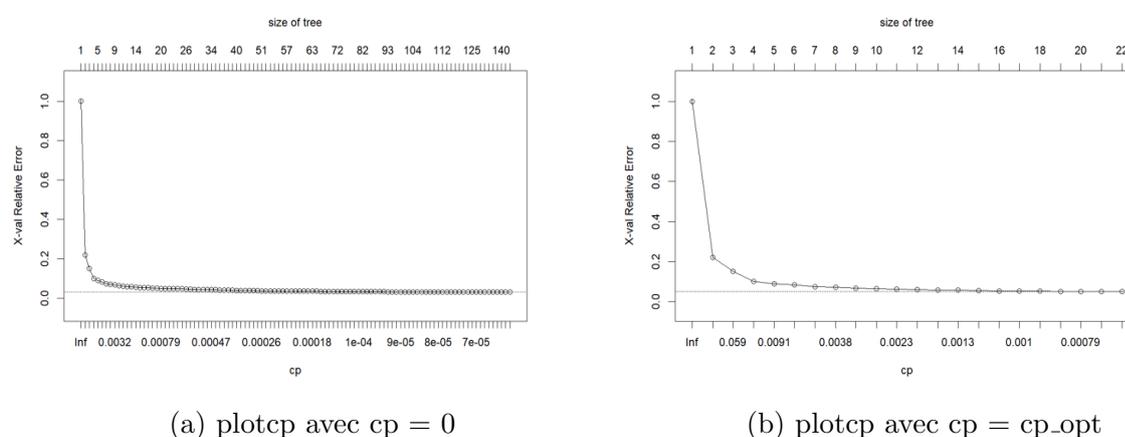


Figure 24: élagage par `plotcp`

La figure ci-dessus représente l'élagage réalisé sur la classe 1 celle des hommes. on part d'un arbre qui a pour paramètre par défaut 0 (arbre maximal) puis on essaie de minimiser le nombre de classes en conservant une erreur relativement nulle. Ce qui correspond à 22 classes pour la classe 1 (et 20 pour la classe 2).

Cet élagage est effectué manuellement pour avoir le `cp_opt`. En effet, il existe une manière d'automatiser le processus en utilisant le critère de minimisation algorithmique sous R (`cp_opt = fit$cpable[which.min(fit$cpable[, 4]), 1]`). Mais cela conduit à créer un grand nombre de classes (cas de surapprentissage).

En ajustant les CP, on retient finalement 42 MP qui réduit considérablement le nombre de contrats par rapport à la base initiale. l'arbre final construit pour la classe 1 est représentée ci-dessous.

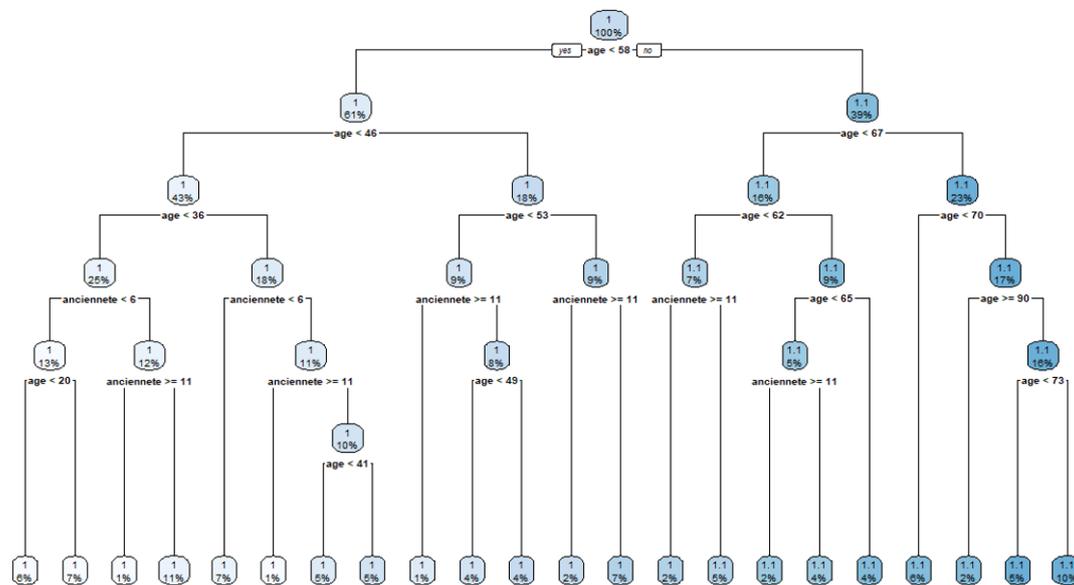


Figure 25: Arbre de décision de la classe 1 (sexe=H)

### 3.3.2 Cah/Kmean

Après la construction des arbres de décisions par l'algorithme supervisé du CART, on passe maintenant aux modèles non supervisés. Pour ce type de modèle, Un certain nombre de variables quantitatives est nécessaire pour sa mise en œuvre. C'est dans cette optique que les cash flows qui permet de construire une base de 16 variables quantitatives.

Pour conserver l'information sur le sexe, deux classes sont constituées comme première classification. Néanmoins, le choix du nombre de classes optimal reste un problème. Pour y remédier, le critère du coude et du coefficient de corrélation multiple  $R^2$  sont observés pour sélectionner des nombres de classes pertinents.

La suite de l'analyse consiste à explorer les modèles non supervisés, qui nécessitent un certain nombre de variables quantitatives pour être mis en œuvre. Pour cette raison, les cash flows ont été utilisés pour construire une base de 16 variables quantitatives.

Dans un premier temps, la classification des hommes a été effectuée en appliquant CAH. La figure 26 présente les critères de décision obtenus à partir de l'application du Cah sur la classe des hommes. Le critère du coude permet de voir les points d'inflexion, révélant ainsi les ruptures dans les données. Ainsi, on remarque 4 sauts d'inertie évidents à des seuils de 2, 6, 10 et 25. Le coefficient de corrélation multiple  $R^2$ , quant à lui, mesure le ratio de l'inertie intraclasse sur l'inertie totale (cf. 2). Cette grandeur qui juge de la qualité d'une partition doit être idéalement proche de 1 tout en correspondant à une partition moins fine que la partition discrète.

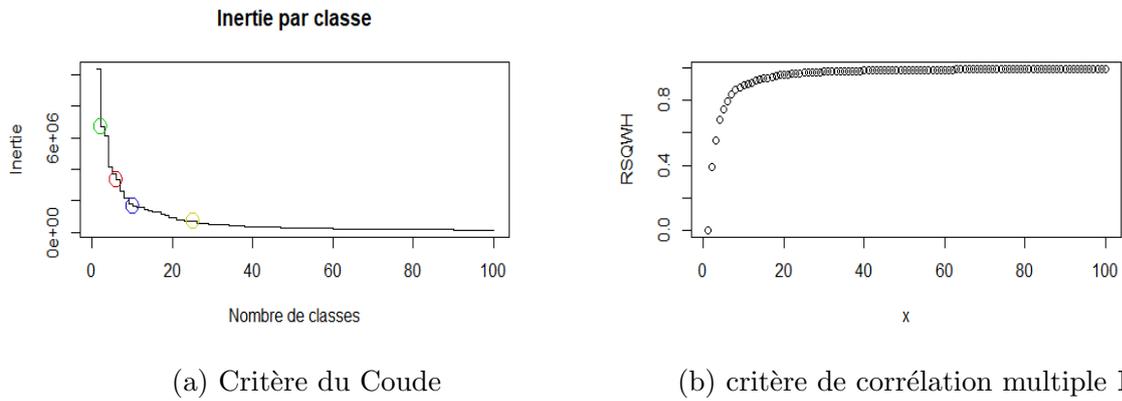


Figure 26: Critère de décision

À partir de cette analyse, on a retenu 22 classes comme nombre optimal pour la classification des hommes, car elle présente un  $R^2$  de 97% et respecte le critère du coude. Néanmoins, pour une exploration plus approfondie, nous avons également testé un nombre de classes de 40 avec un  $R^2$  de 98,30%.

La même analyse a été effectuée pour la classe des femmes, révélant un nombre optimal de classes compris entre 19 et 30.

Concernant la méthode de Kmeans, le nombre de centres a été fixé en fonction des analyses précédentes, et le paramètre *nstart* a été ajusté à 100 pour ajouter plus de complexité. La figure qui suit illustre les principaux clusters obtenus à partir de ces analyses.

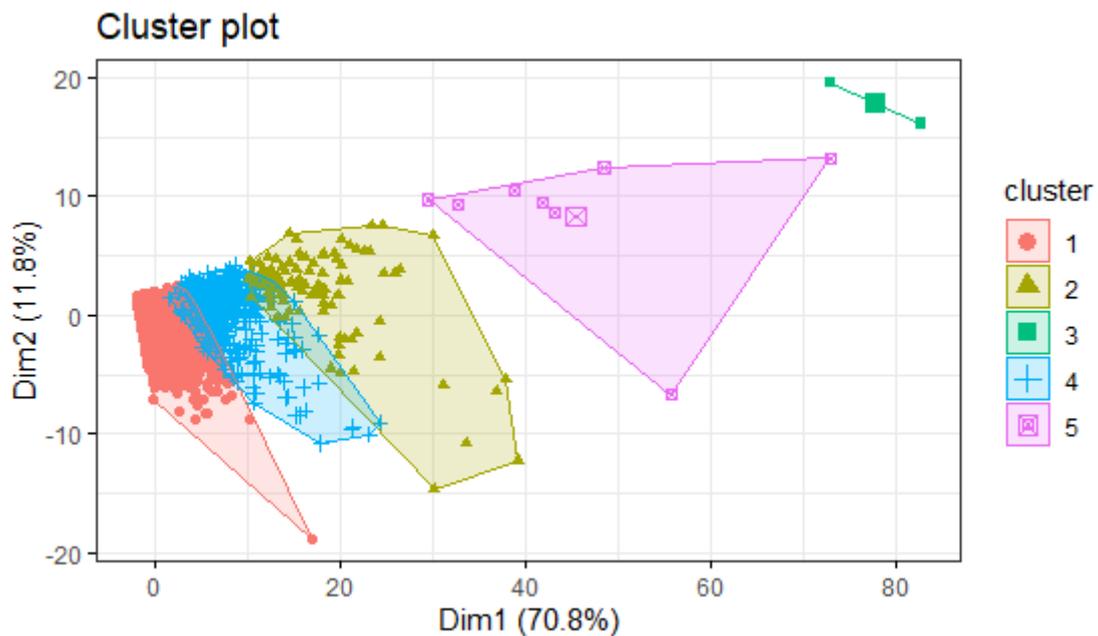


Figure 27: partitionnement avec la méthode Kmean en 5 clusters

### 3.4 Rente

Le produit de Rente sur lequel porte cette étude a beaucoup de variables qualitatives. Pour pouvoir exploiter au maximum ces informations impactant les résultats du modèle alm, une première partition sera effectuée comme avec le produit d'épargne. Le récapitulatif

des caractéristiques de cette classification est présentée dans la figure 28.

group_alm	sex_assure	sex_benef	param_cad	type_coti	regime_point	etat_contrat	nb_contrat	
1	H	F		0	4	N	Reversion	510
2	H	F		0	2	N	Retraite	784
3	F	F		0	2	N	Retraite	7
4	H	F		0	3	O	Rente différée	58
5	H	F		0	1	N	Retraite	908
6	F	F		0	4	N	Reversion	1
7	H	H		0	2	N	Retraite	12
8	H	H		0	2	O	Retraite	11
9	F	H		0	2	O	Retraite	57
10	H	F		1	3	N	Rente différée	73
11	F	H		0	4	N	Reversion	12
12	F	F		0	1	O	Retraite	652
13	F	F		1	3	N	Rente différée	53
14	F	F		0	3	N	Rente différée	4
15	F	F		0	4	O	Reversion	1
16	F	F		0	3	O	Rente différée	37
17	H	F		0	4	O	Reversion	472
18	F	F		0	2	O	Retraite	6
19	H	F		0	2	O	Retraite	626
20	H	F		0	1	O	Retraite	687
21	F	H		0	4	O	Reversion	11
22	F	H		0	2	N	Retraite	76
23	F	F		0	1	N	Retraite	924

Figure 28: Première partition Rente

Après avoir effectué une première partition pour le produit de Rente, nous avons procédé à la modélisation en utilisant les algorithmes CART, Cah et Kmean. Cette première classification nous a permis de faire ressortir des classes hétérogènes. Cependant, il était nécessaire de poursuivre l'analyse en utilisant des modèles plus sophistiqués pour mieux comprendre la structure des données et extraire des informations supplémentaires.

Ainsi, nous avons mis en place les modèles en nous basant sur les critères de décisions définis précédemment. Pour chaque modèle, nous avons choisi le nombre optimal de classes en utilisant des techniques spécifiques pour évaluer la pertinence de la partition. La classe 2 (group\_alm de la figure 28) est choisi pour illustrer la construction de MP pour chaque méthode.

### 3.4.1 Cart

L'image suivante montre l'élagage par plotcp et l'arbre de décision de la classe 2 :

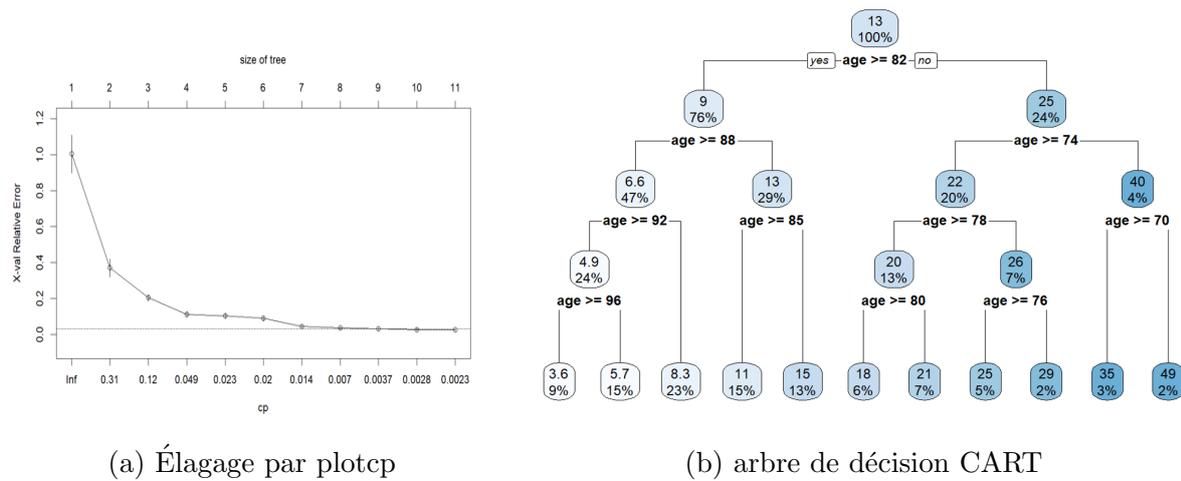


Figure 29: modèle CART (classe 2)

On remarque que le modèle ne prend pas en compte l'ancienneté lors de la classification. Cela est dû à la concentration de l'ancienneté des contrats. Le modèle peut exclure des variables qui ne lui semble pas autant pertinente pour un certain jeu de données (ce qu'on ne peut pas avoir avec l'approche générale a posteriori).

### 3.4.2 Cah

L'image suivante montre l'élagage par le critère du coude et le dendrogramme de la classe 2 :

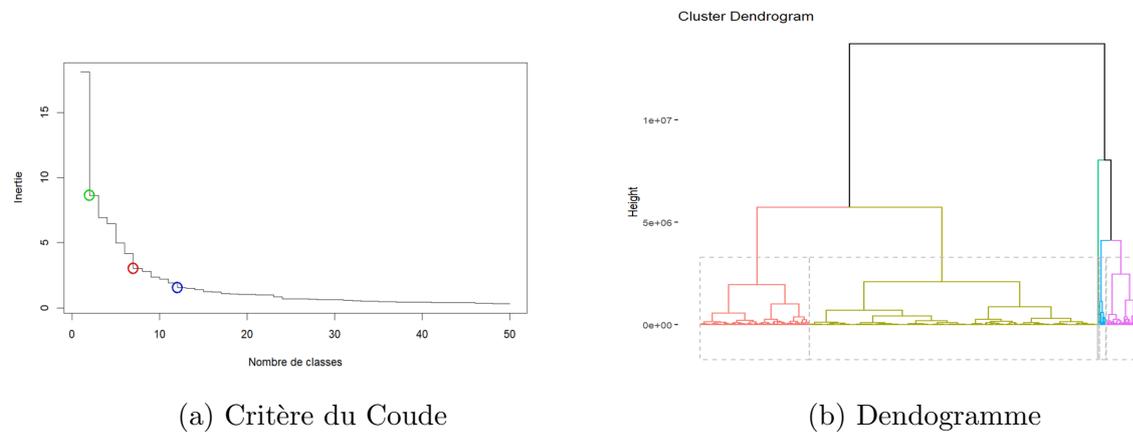


Figure 30: modèle Cah (classe 2)

La figure ci-dessus représente un partitionnement en 7 classes observables par code couleur sur le dendrogramme. Le critère du coude illustré confirme le choix de 7 classes comme avec le plotcp de la méthode CART en proposant d'autres alternatives comme 12 classes.

### 3.4.3 Kmean

L'image suivante montre l'élagage par le critère du coude et les principaux clusters de la classe 2 :

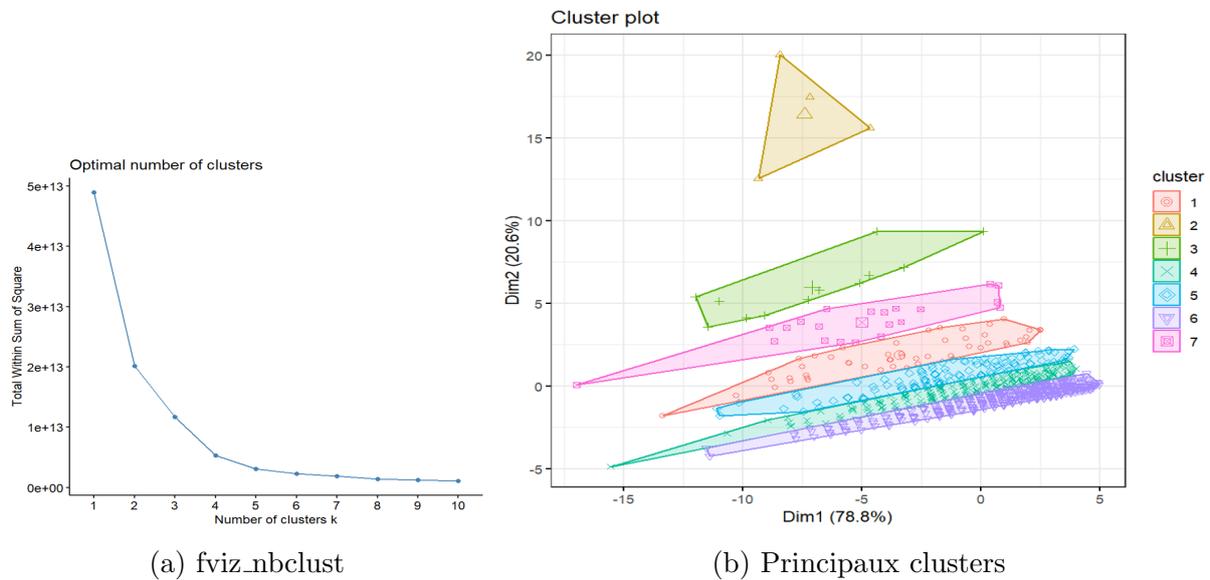


Figure 31: modèle Kmean (classe 2)

La figure ci-dessus représente un partitionnement en 7 classes observables par chaque cluster sur le graphique. On perçoit bien la notion de centre d'inertie et de regroupement.

Pour le modèle Kmean la fonction *fviz\_nbclust* de R a été utilisé pour observer le nombre optimal de clusters. Cette fonction permet de tracer le graphique de la variation du score d'inertie en fonction du nombre de clusters pour aider à déterminer le nombre optimal de clusters pour le modèle Kmean (graphique de gauche). Le score d'inertie est une mesure de la cohérence des clusters et représente la somme des distances carrées entre chaque point de données et le centre de son cluster. La variation du score d'inertie en fonction du nombre de clusters peut aider à déterminer le nombre optimal de clusters en identifiant un point où l'ajout de plus de clusters n'améliore pas de manière significative la cohérence des clusters qui est autour de 8 dans ce cas.

Ainsi à travers ces trois manières de déterminer le nombre de classes, on remarque une stabilité entre les méthodes. C'est pourquoi dans la suite, on gardera une seule<sup>1</sup> méthode pour déterminer le nombre de classes optimales pour chaque modèle.

<sup>1</sup>c à d : si le critère du plotcp avec Cart nous donne un nombre évident de clusters optimaux ce même nombre sera utilisé pour les méthodes Cah ou Kmean

## 4 ANALYSE ET COMPARAISON DETERMINISTE

### 4.1 Épargne

#### 4.1.1 Lancement principal

Le tableau qui suit révèle les différents résultats de lancement sur le produit d'épargne.

Méthode	nb de MP	temps	BE	écart BE
Approche générale	78	12 min 44 s	444 237 399,55 €	0,15%
CART	42	8 min 50 s	443 546 229,62 €	-0,005%
Cah	42	10 min 6 s	443 650 082,76 €	0,018%
Kmean	42	9 min 9 s	443 575 083,35 €	0,001%
<b>Brut</b>	<b>10 949</b>	<b>2 h 33 min 46 s</b>	<b>443 570 096,51 €</b>	<b>0%</b>

Table 9: Récapitulatif des lancements déterministes (épargne)

De manière générale, les résultats obtenus par toutes les méthodes permettent d'optimiser significativement le temps de calcul en réduisant celui-ci de plus de 92% sur CALFITEC. En ce qui concerne la précision, toutes les méthodes restent très pertinentes avec une erreur relative d'environ 0,02%. Cependant, il est important de noter que le modèle CART, bien qu'il soit précis, a tendance à sous-estimer le BE.

À ce stade de notre étude, toutes les méthodes peuvent être retenues pour un lancement en stochastique. Cela signifie que les regroupements statistiques obtenus grâce à ces méthodes pourront servir de base pour préparer les inputs de la CTAD et lancer le modèle ALM de manière stochastique. En effet, les résultats obtenus par ces méthodes fournissent des informations pertinentes pour la modélisation de l'ALM, ce qui permettra de mieux appréhender les risques liés au produit étudié.

#### 4.1.2 Lancements des chocs

Afin d'approfondir notre analyse et d'obtenir une vision plus complète des résultats obtenus avec les différentes méthodes de clustering utilisées, nous avons préparé un graphique comparatif qui met en évidence les différences entre les chocs générés par le lancement déterministe du Best Estimate (BE, BEG) et les résultats obtenus avec chacune des méthodes.

Il convient de souligner que ces chocs représentent des indicateurs clés en matière d'ALM, puisqu'ils permettent d'évaluer la sensibilité du Best Estimate (BE) aux différentes fluctuations et aléas du marché. Le BE représente l'estimation de la valeur actuelle de tous les flux monétaires futurs qui seront générés par les contrats d'assurance vie en portefeuille, tout en prenant en compte l'impact de différents scénarios possibles sur ces flux monétaires. En d'autres termes, il s'agit d'une mesure de la performance financière globale de l'assureur, qui prend en compte les risques associés à ses engagements.

Le "BE\_ACTION" représente quant à lui le BE avec application d'un choc sur les investissements en actions. En effet, il s'agit de l'estimation de la valeur actuelle de tous

les flux monétaires futurs générés par les investissements en actions, tout en prenant en compte l'impact de différents chocs ou scénarios possibles sur ces flux monétaires.

Les autres chocs réalisés dans notre étude portent sur les frais, le taux à la hausse (baisse), le spread et l'immobilier. En effectuant ces chocs sur les différents indicateurs, l'outil peut évaluer leur impact sur le BE et ainsi mesurer leur importance dans l'analyse des risques en ALM.

En résumé, il est important de bien analyser l'impact de notre étude sur ces indicateurs :

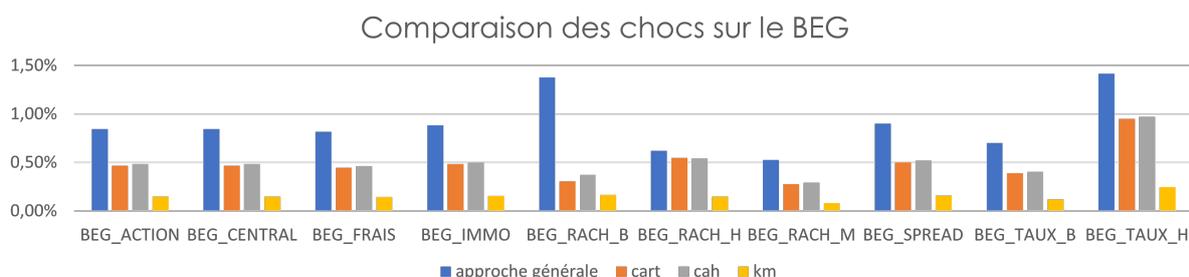


Figure 32: Analyse de la précision des méthodes sur le BEG

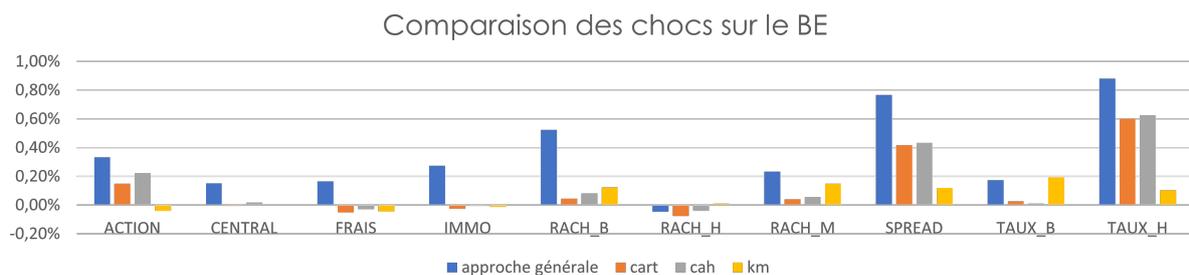


Figure 33: Analyse de la précision des méthodes sur le BE

En observant attentivement les deux graphiques, nous constatons que les méthodes de clustering sont plus performantes sur le lancement du BE que sur le BEG, avec une précision avoisinant les 0,2% pour le BE et 0,4% pour le BEG. Néanmoins, il convient de souligner que certaines méthodes de clustering, notamment Kmean, surpassent les autres en termes de précision. En effet, les méthodes de clustering telles que Kmean permettent généralement d'atteindre une précision inférieure à 0,5% pour chaque indicateur. De plus, ces méthodes sont considérées comme plus sûres et plus convergentes que l'approche générale, qui semble généralement moins performante que les autres méthodes évoquées dans ce mémoire.

Il est également important de noter que la précision n'est pas le seul indicateur à prendre en compte dans l'analyse de ces méthodes. D'autres critères tels que la complexité de mise en œuvre, la flexibilité et la capacité d'adaptation à différents types de données doivent être pris en compte dans le choix final d'une méthode de construction de MP pour l'ALM.

Enfin, il est intéressant de souligner que ces méthodes de clustering ont montré leur efficacité dans la gestion des chocs sur différents indicateurs tels que les frais, le taux à la hausse ou à la baisse, le spread et l'immobilier.

### 4.1.3 Analyse de trajectoires

Après une analyse de la précision des estimations, l'analyse peut encore être poussée pour étudier comment la construction des MP impact la projection des cash-flows.

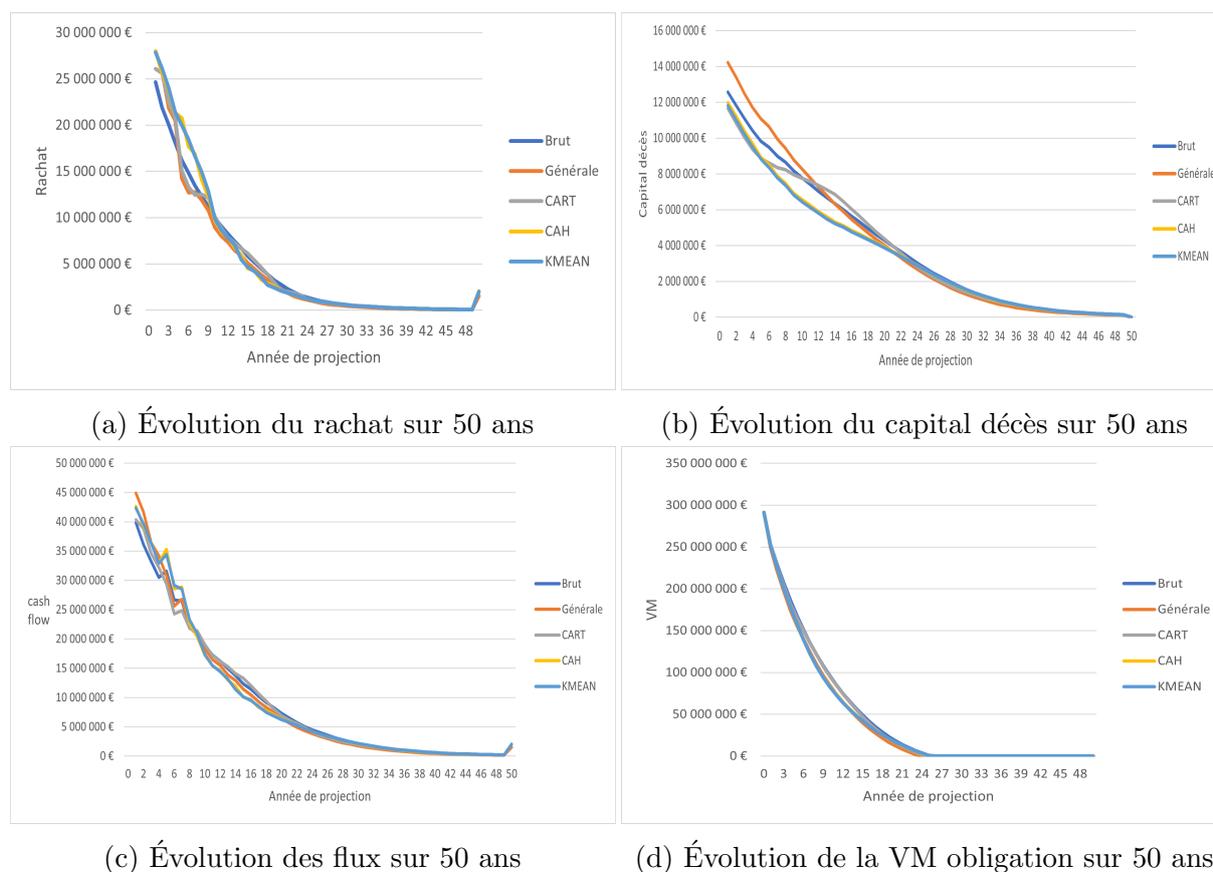


Figure 34: Étude des projections (épargne)

Cette étude de projection met en avant une fois de plus la crédibilité des modèles élaborés. En effet, toutes les méthodes utilisées ont réussi à reproduire fidèlement le résultat avec les données non agrégées en termes d'allure, ce qui atteste de leur performance. Par ailleurs, les approches simples, notamment le CART, semblent plus appropriées en ce qui concerne ce critère.

### 4.1.4 Conclusion

Pour conclure, l'analyse menée a permis de confirmer la fiabilité des modèles construits. Toutes les méthodes ont permis de réduire significativement le temps de calcul et de reproduire fidèlement la trajectoire des Cash-Flows, ce qui a abouti à des estimations précises du BE et du BEG. Toutefois, parmi toutes les méthodes étudiées, la méthode Kmean s'est avérée être la plus performante en combinant la meilleure précision et une grande viabilité pour cette construction.

## 4.2 Rente

### 4.2.1 Lancement principal

Le tableau qui suit révèle les différents résultats de lancement sur le produit de rente.

Dans l'ensemble, toutes les méthodes de clustering étudiées permettent d'optimiser considérablement le temps de calcul d'environ 90%, mais la précision obtenue reste relativement faible avec une erreur relative d'environ 1% et une sous-estimation du BE. Cette

Méthode	nb de MP	temps	BE	écart BE
approche générale	118	9 min 22 s	267 852 669,9€	-3,05%
Cart	171	9 min 13 s	274 984 043,95€	-0,47%
Cah	171	9 min 16 s	274 108 564,8€	-0,79%
Kmean	171	10 min 32 s	273 135 910,3€	-1,14%
<b>Brut</b>	5 982	1 h 28 min 51 s	276 283 270,12€	0%

Table 10: Récapitulatif des lancements déterministes (rente)

précision relativement faible peut s'expliquer par la complexité intrinsèque de ce type de produit, qui nécessite la prise en compte de nombreux paramètres.

Cependant, parmi les différentes méthodes étudiées, le modèle CART se distingue en produisant le meilleur résultat à ce stade de l'étude.

#### 4.2.2 Lancements des chocs

Comme pour le produit d'épargne un lancement des chocs est effectué sur le BE ou le BEG.

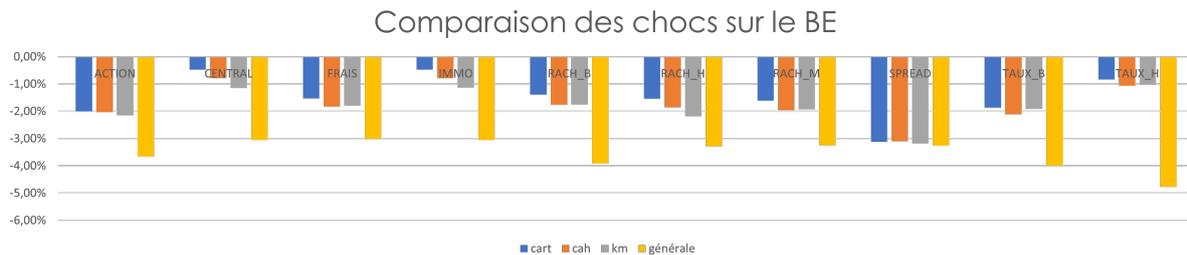


Figure 35: Analyse de la précision des méthodes sur le BEG (rente)

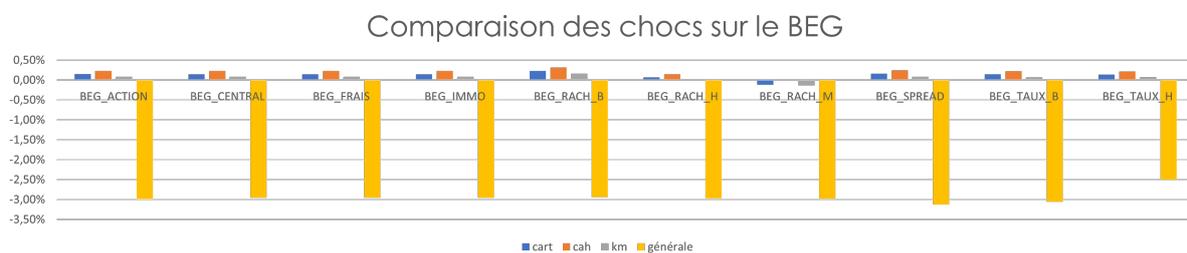


Figure 36: Analyse de la précision des méthodes sur le BE (rente)

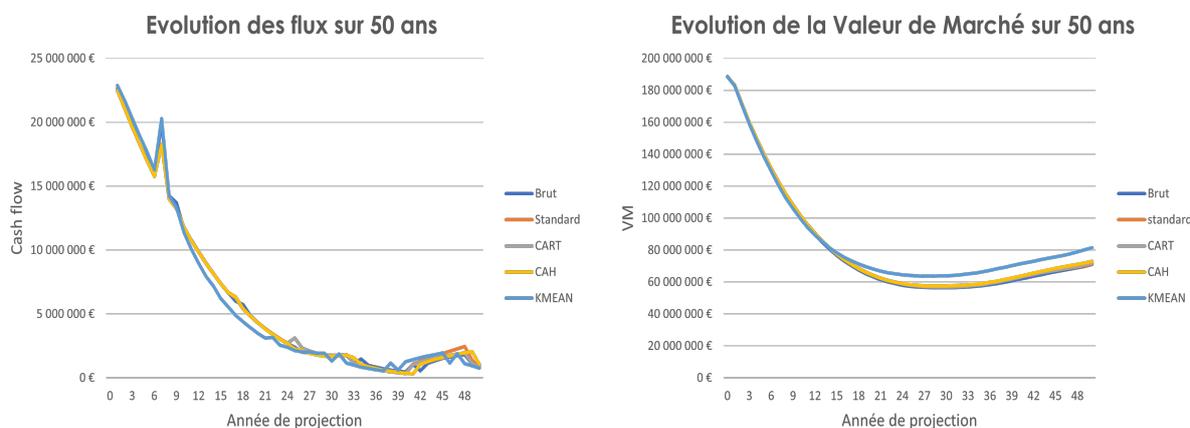
En observant les deux graphiques, il est évident que les méthodes sont plus performantes lors du lancement du BE, avec une précision d'environ 0,8% sur le BE tandis qu'il y a une sous-estimation d'environ 3% sur le BEG. En général, l'approche générale est moins performante que les autres méthodes évoquées dans ce mémoire (optimisation réussie) qui permettent généralement d'atteindre une précision inférieure à 0,5%.

Bien que les résultats de chaque méthode soient globalement assez proches, la méthode CART minimise fortement l'écart au niveau du BE central, ce qui en fait une méthode

plus sûre et plus convergente. Cependant, il est important de noter que la précision reste assez faible par rapport au produit d'épargne, ce qui peut être attribué à la complexité de ce type de produit où il faut prendre en compte de nombreux paramètres.

### 4.2.3 Analyse

Le graphique suivant sert à montrer l'évolution de certaines caractéristiques du portefeuille sur les 50 ans de projection par méthode de construction de MP.



(a) Évolution des flux sur 50 ans

(b) Évolution de la VM obligation sur 50 ans

Figure 37: Étude des projections (rente)

Le modèle CART est celui qui se rapproche le plus des tendances du modèle brut sur l'ensemble des graphiques dans les deux cas. le modèle Kmean est celui qui s'éloigne le plus de la courbe des données brutes.

### 4.2.4 Conclusion

En somme, l'ensemble des modèles construits ont permis de valider notre étude en réduisant significativement le temps de calcul, reproduisant la trajectoire des Cash-Flows, et produisant des estimations précises du BE et du BEG. Bien que toutes les méthodes aient montré une précision pertinente autour de 1% d'erreur relative en sous-estimant le BE, la méthode Kmean se démarque en combinant la meilleure précision avec la représentation de la majorité des trajectoires, la rendant ainsi la méthode la plus viable pour cette construction.

## 5 BE STOCHASTIQUE

### 5.1 Nécessité de la projection stochastique

Dans la section précédente, nous avons utilisé un lancement déterministe pour construire les model points, c'est-à-dire un scénario unique et cohérent avec le marché, utilisant la courbe de taux sans risque. Cependant, en pratique, l'évaluation en Best Estimate d'un produit d'épargne ou de rente nécessite l'utilisation d'un modèle stochastique, tel que décrit dans [HEL].

Bien que la méthode déterministe soit simple et facile à comprendre, elle ne permet pas de modéliser toutes les options financières garanties présentes dans les contrats d'assurance vie, ce qui rend l'utilisation de modèles stochastiques nécessaire pour une évaluation plus

précise et réaliste.

En effet, pour être représentatif de l'incertitude des flux de trésorerie, le BE doit prendre en compte la probabilité et la gravité des situations qui découlent de scénarios multiples combinant les facteurs de risque pertinents. Par exemple, dans le cas d'un contrat d'épargne à taux minimum garanti, la garantie ne sera pas revalorisée si le taux de rendement moyen des actifs utilisé lors de la projection reste toujours supérieur au taux minimum garanti. Cependant, si le taux de rendement des actifs devient inférieur à ce taux minimum, l'assureur subit une charge effective supplémentaire, ce qui diminue son résultat.

Le groupe de travail initié par l'IA [da16] suggère donc que "la valorisation de ces options et garanties peut être effectuée soit par une méthode stochastique, soit par utilisation de formules fermées (ex: formule de Black-Scholes)." Cependant, pour les contrats en euros français, les formules fermées pour l'évaluation d'options sont complexes à expliciter et ne sont pas utilisées en pratique en raison de la multitude d'options sous-jacentes au contrat et des interactions actifs/passifs, telles que les rachats dynamiques. Par conséquent, la méthode Monte Carlo basée sur des simulations stochastiques est souvent privilégiée pour modéliser les produits du marché français de l'assurance vie.

Les flux de trésorerie entrant dans le calcul du BE sont calculés en tenant compte, entre autres, d'hypothèses d'évolution sur l'horizon de projection, telles que :

- De l'environnement économique : niveau des marchés actions, niveau des taux sans risque, prix de l'immobilier, etc.
- Des risques biométriques : mortalité, morbidité, etc. La modélisation de l'incertitude dans l'évolution de ces types de variables peut être gérée par des modèles de diffusion stochastique conformément aux orientations sur la valorisation des provisions techniques (EIOPA-BoS-14/166 FR) n°25 (1.58) et n°54 (1.97).

Les évolutions des environnements financiers sont ainsi généralement basées sur des diffusions stochastiques simulées à l'aide de Générateurs de Scénarii Economiques (GSE).

L'utilisation de la méthode Monte-Carlo dans l'évaluation du BE est basée sur un jeu de scénarii stochastiques qui sert à évaluer les principales options et garanties financières sous-jacentes au contrat d'assurance, dans leur composante de valeur intrinsèque et valeur temps.

Pour évaluer une meilleure estimation par méthode Monte-Carlo, il est généralement procédé de la façon suivante :

1. Pour chaque scénario  $i$  du jeu de scénarii stochastiques, construire l'échéancier des flux de trésorerie;
2. Actualiser les flux de trésorerie à la courbe des taux sans risque du scénario  $i$  pour obtenir l'estimation  $i$ ;
3. La meilleure estimation est égale à la moyenne probabiliste (espérance) de l'ensemble des estimations  $i$ .

Le nombre de scénarii stochastiques financiers est à définir pour assurer une convergence suffisante de l'estimateur de la moyenne. Communément, le nombre de scénarii utilisés est le millier (1000) pour des produits d'épargne en euros, mais un nombre plus élevé

peut s'avérer nécessaire pour obtenir une précision compatible avec le seuil de matérialité défini par la société.

## 5.2 Estimation du BE stochastique

Les MP ayant été établi en déterministe, un lancement en stochastique sur 1000 scénarios est réalisé pour déterminer le BE stochastique final. Le tableau 10 récapitule les résultats de chaque produit.

Produit	méthode	BE sto	écart BE sto/dét	BEG sto	écart BEG sto/dét
Épargne	Kmean	465 708 500,9 €	3,17%	425 689 296,7 €	4,94%
Rente	CART	279 495 204,5 €	1,16%	233 350 981,4 €	4,2 %

Table 11: Récapitulatif des lancements stochastiques

Comme présenté plus haut, la méthode stochastique est privilégiée pour évaluer les produits d'assurance vie. Cela est dû à la nature incertaine des flux de trésorerie entrant dans le calcul du Best Estimate (BE). En effet, pour tenir compte de l'incertitude des flux de trésorerie, il est nécessaire de prendre en considération la probabilité et la gravité des situations issues de scénarios multiples combinant les facteurs de risque pertinents.

C'est dans ce contexte que le principe de Monte Carlo est utilisé pour la modélisation stochastique. Ainsi, le BE stochastique est la moyenne des BE estimés par scénario, représentant la convergence des scénarios économiques.

Le tableau 10 permet de visualiser les écarts de valorisation entre le BE stochastique et le BE déterministe. Ces écarts sont dus à la nature stochastique de la méthode Monte Carlo. En effet, les scénarios économiques générés par cette méthode peuvent conduire à des résultats différents de ceux obtenus avec la méthode déterministe. Les écarts observés permettent donc de mesurer l'impact de l'incertitude des flux de trésorerie sur la valorisation des produits d'assurance vie.

La figure suivante illustre bien cette estimation.

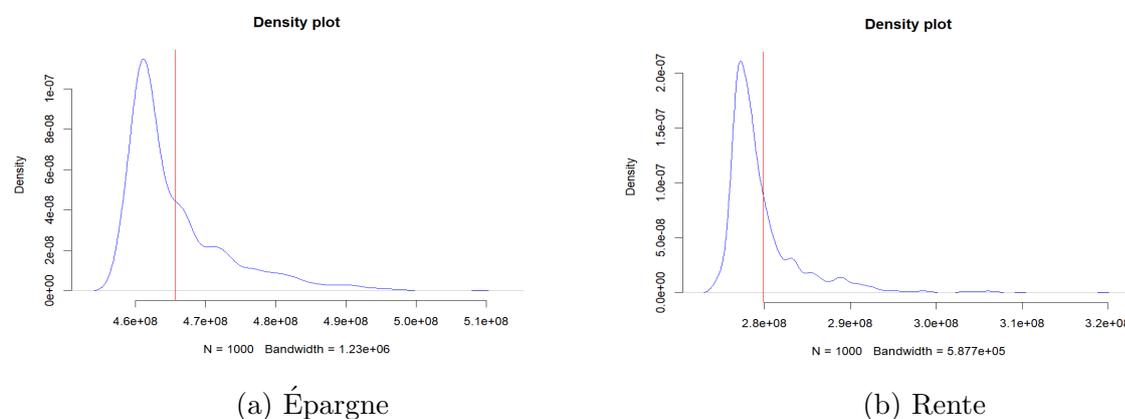


Figure 38: densité du BE stochastique sur 1000 scénarios

Cet écart de valorisation permet de respecter le principe de prudence en prenant en compte les variations du marché et le comportement des assurés.

**Part III**  
**BACKTESTING**

# 1 Définition et statistique

Le backtesting ou test rétroactif de validité est une étape cruciale dans la validation d'une modélisation ou d'une stratégie. Il permet de mesurer la pertinence et la fiabilité de ces dernières en les confrontant à des données historiques réelles sur une période donnée. Cette méthode nécessite une approche statistique rigoureuse pour une évaluation objective des performances des modèles.

Dans cette partie, nous allons appliquer le backtesting à notre construction de MP en utilisant les données de l'année 2020. Les modèles présentés précédemment seront ainsi testés et comparés pour évaluer leur capacité à prédire les résultats sur une période donnée.

Le tableau suivant résume les données de l'année 2020 qui seront utilisées pour cette analyse. Cette étape de validation est essentielle pour garantir la pertinence et la fiabilité de notre modélisation dans des conditions réelles et pour s'assurer de sa capacité à fournir des résultats cohérents sur le long terme.

	Épargne	Rente
sexe ratio (F)	52,5%	31,6%
PM	434 445 596,58 €	230 510 629,06 €
moyenne Age	53 ans	85 ans
moyenne Ancienneté	10 ans	47 ans
arrérage de rente	0 €	22 666 482,19 €

Table 12: description statistique des données 2020

Ce tableau nous donne des informations sur la population assurée en 2020, en distinguant les contrats d'épargne et de rente. Nous pouvons voir que la majorité des assurés est de sexe féminin pour les deux types de contrats. La prime moyenne est plus élevée pour les contrats d'épargne que pour les contrats de rente. L'âge moyen des assurés est significativement plus élevé pour les contrats de rente, reflétant la nature même de ce type de contrat qui est destiné aux personnes âgées. Enfin, le tableau indique que les assurés ayant souscrit à une rente ont un arrérage total de 22 666 482,19 €.

## 2 Épargne

### 2.1 Comparaison déterministe

La méthode mise en œuvre dans cette étude est la méthode K-mean. Pour mettre en œuvre cette méthode, nous avons suivi la même procédure décrite dans la partie 2 pour lancer les modèles et procéder à leurs évaluations.

Méthode	nb de MP	BE	écart BE	BEG	écart BEG
approche générale	78	465 012 794,41 €	0,627%	429 095 202 €	0,531%
Kmean	21	464 625 340,7 €	0,54%	427 620 971,94 €	0,185%
<b>Brut</b>	11 385	462 115 707,9 €	0%	426 829 249,1 €	0%

Table 13: Récapitulatif des lancements déterministe épargne 2020

Il est notable que le modèle a un impact significatif dans l'établissement des MP, permettant ainsi d'améliorer la précision de leur estimation du BE. Cela valide le backtesting réalisé, car il a permis de construire des MP plus performants que ceux établis avec l'approche générale de l'année précédente.

## 2.2 Analyse

On va maintenant voir l'impact de cette construction de modèle sur l'évolution des cash-flows. À cet effet, la trajectoire des cash-flows sur 50 ans est étudié pour comparer le comportement du lancement brut par rapport au comportement du lancement avec les MP.

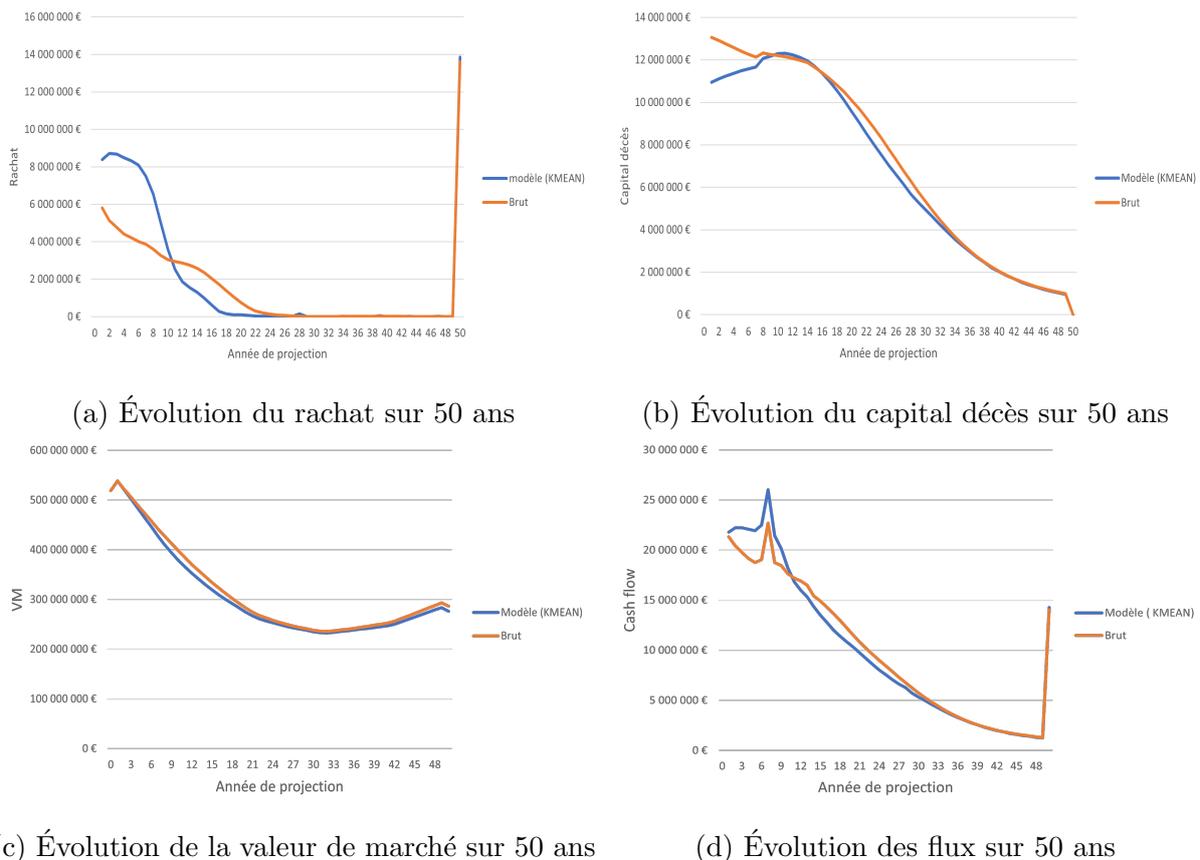


Figure 39: Étude des trajectoires épargne 2020

Le modèle reproduit de façon générale la trajectoire des cash-flows. Au niveau du rachat, du flux, il y a un certain décalage entre le modèle et le lancement brut qui se compense dans le temps. Ce décalage s'explique par l'âge et l'ancienneté qui impacte le rachat structurel ou les flux. On remarque aussi un décrochage fort à la dernière année pour le

rachat et les flux, cela est dû à l'hypothèse de distribution des flux totaux à la fin de la simulation (rachat total). Le paramétrage du modèle peut être mis en cause à ce niveau sur 2020. On voit que les modèles ne sont pas parfaits et évoluent en fonction des données et des hypothèses de marchés.

Malgré cet écart de trajectoire, le modèle reste assez fiable en reproduisant la plupart des trajectoires.

## 3 Rente

### 3.1 Comparaison déterministe

La méthode utilisée est la combinaison de la méthode CART. Comme présenté plus haut cette méthode permet de prendre en compte toutes les variables qualitatives pour conserver le plus d'informations.

Nous avons suivi la même procédure que celle décrite plus haut dans la partie 2 pour effectuer le lancement des modèles.

Méthode	nb de MP	BE	écart BE	BEG	écart BEG
approche générale	60	328 556 246,59 €	1,832%	267 852 669,9 €	1,278%
CART	122	324 907 868 €	0,7%	263 334 284,72 €	-0,430%
<b>Brut</b>	6983	322 644 007,54 €	0%	264 471 477,25 €	0%

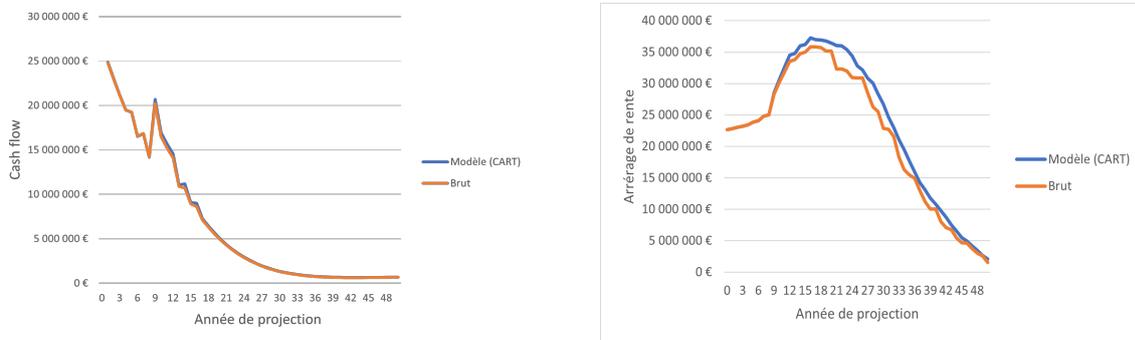
Table 14: Récapitulatif des lancements déterministes de rente 2020

On remarque bien que le modèle CART a un impact considérablement significatif dans l'établissement des MP. Il permet d'améliorer la précision des MP dans l'estimation du BE.

Le backtesting est validée à ce niveau puisqu'on arrive à construire des MP plus performants que ceux établis avec l'approche générale l'an dernier.

### 3.2 Analyse des trajectoires

On va maintenant voir l'impact de cette construction de modèle sur l'évolution des cash-flows. À cet effet, la trajectoire des cash-flows sur 50 ans est étudié pour comparer le comportement du lancement brut par rapport au comportement du lancement avec les MP.



(a) Évolution des flux sur 50 ans

(b) Évolution de l'arrage de rente sur 50 ans

Figure 40: Étude des trajectoires rente 2020

Le modèle reproduit de façon précise la trajectoire des cash-flows. Au niveau du BE, les deux courbes sont presque confondues ce qui évoque une bonne modélisation qui produit les mêmes cash-flows à chaque itération du modèle ALM (backtesting validé).

## 4 Lancement stochastique

Nous avons effectué comme dans l'analyse des données 2021 un lancement stochastique en prenant en compte les caractéristique du marché 2020 pour déterminer le BE sur le produit d'épargne.

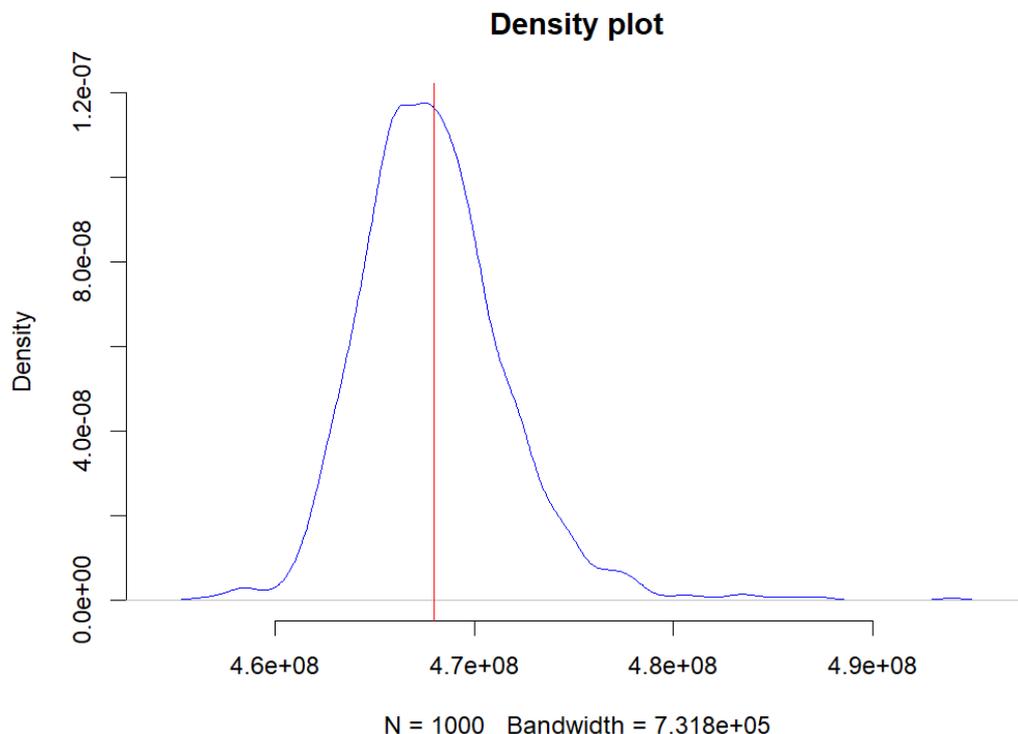


Figure 41: densité du BE stochastique sur 1000 scénarios du produit épargne en 2020

En étudiant les résultats obtenus, nous pouvons observer que le processus établi dans ce modèle a eu un impact significatif dans l'établissement des modèles prévisionnels. En effet, elle a permis d'améliorer considérablement la précision des prévisions et de l'estimation du Best Estimate (BE). Cette amélioration est due aux bénéfices des algo-

rithmes supervisés et non supervisés utilisés dans la méthode CART et Kmean.

De plus, les résultats de backtesting validés montrent que nous sommes parvenus à construire des modèles prévisionnels plus performants que ceux établis avec l'approche générale l'année précédente. Cela témoigne de l'efficacité de la procédure utilisée pour la construction des modèles prévisionnels.

Par ailleurs, pour évaluer l'impact de cette construction de modèle sur l'évolution des cash-flows, nous avons étudié la trajectoire des cash-flows sur 50 ans. Nous avons comparé le comportement du lancement brut avec celui du lancement avec les MP. Les résultats ont montré que l'utilisation des MP a permis une meilleure estimation des cash-flows et donc une meilleure évaluation de la solvabilité de l'entreprise.

Enfin, en ce qui concerne l'estimation du BE stochastique, nous avons obtenu une valeur de 467 976 091,3 €, avec un écart par rapport au déterministe de 1,27%. Nous avons observé que la densité de BE avait une allure normale un peu différente de la queue observée en 2021. Cette différence est due aux légères différences entre les deux modèles utilisés, qui sont liées à l'évolution du marché. Malgré ces différences, l'impact sur le ratio de couverture (SCR) reste viable.

# CONCLUSION GÉNÉRALE

Dans l'ensemble, notre travail s'inscrit dans un contexte plus large qui propose des outils théoriques et des approches pour étudier la structure de nos portefeuilles d'assurés dans un premier temps, puis les regrouper en utilisant des modèles de clustering adaptés dans une seconde phase. Notre recherche a nécessité une profonde réflexion et une recherche intensive en modélisation mathématique, ainsi qu'un investissement significatif dans la mise en œuvre des algorithmes de clustering.

Les contributions de notre travail sont multiples. Du point de vue théorique, il propose une méthode cohérente et générale pour construire des MP qui optimisent réellement le processus ALM. Cette méthode nous permet d'étudier le modèle ALM général de l'entreprise et de l'optimiser en utilisant le principe de clustering supervisé et non supervisé.

Plus précisément, le modèle de construction des MP examine les interactions entre les variables du passif, en particulier l'âge et l'ancienneté, pour évaluer leur impact sur la valeur cible, le BE. Il détermine ensuite un nombre optimal de classes en utilisant l'algorithme supervisé CART, puis construit le MP final en utilisant une méthode de classification pertinente.

Cette étude a permis également de montrer que les modèles supervisés (CART) sont plus efficaces sur le produit de rente (complexe) au vu du nombre de variables qualitatives à notre disposition. Le produit d'épargne quant à lui permet d'établir que les modèles de clustering non supervisés sont plus significatifs en situation de peu de données qualitatives. Une entreprise peut donc adapter la méthode de construction en fonction du portefeuille et du produit. Ces méthodes présentent des avantages et des inconvénients consignés dans le tableau suivant récapitulant nos conclusions.

CART	CAH	KMEAN
<p><b>Avantages:</b></p> <ul style="list-style-type: none"> <li>• Simple, interprétable</li> <li>• Rapidité d'exécution</li> <li>• Utilisation de variables inputs (qualitatives)</li> <li>• Visualisation de la classification et des variables utilisées</li> <li>• Reproduit bien les trajectoires des cash-flows</li> </ul> <p><b>Inconvénients</b></p> <ul style="list-style-type: none"> <li>• Instabilité (Réaction aux chocs)</li> <li>• Sous-estimation du BE</li> </ul>	<p><b>Avantages:</b></p> <ul style="list-style-type: none"> <li>• Visualisation du regroupement progressif des données (dendogramme).</li> <li>• choix un type de dissimilarité (Ward)</li> <li>• Stabilité sur les chocs</li> </ul> <p><b>Inconvénients</b></p> <ul style="list-style-type: none"> <li>• Regroupements sont définitifs</li> <li>• Lourdeur des calculs</li> <li>• Nécessité de variables quantitatives (lancement déterministe)</li> </ul>	<p><b>Avantages:</b></p> <ul style="list-style-type: none"> <li>• Concept simple, rapide</li> <li>• Plusieurs possibilités d'améliorations</li> <li>• Affectation réversible</li> <li>• Stabilité sur les chocs</li> </ul> <p><b>Inconvénients</b></p> <ul style="list-style-type: none"> <li>• Dépend du tirage initial</li> <li>• Pas de moyen de visualisation progressif</li> <li>• Nécessité de variables quantitatives (lancement déterministe)</li> <li>• Nécessite beaucoup de données</li> </ul>

Table 15: Récapitulatif des avantages et inconvénients des méthodes

Ces méthodes permettent ainsi d'effectuer un lancement déterministe ou stochastique pour observer le gain de précision par rapport à une approche classique. Les modèles retenus sont donc pertinents à la suite de ces analyses.

Dans le cadre de la validation, nous avons proposé un processus de validation croisée des données de l'année 2020 en nous appuyant sur le processus de construction des MP décrit précédemment pour évaluer l'évolution du BE entre les deux années.

En conclusion, notre étude permet de construire des Model points robustes et d'optimiser le processus de modélisation en ALM. Cependant, une des limites de notre étude est qu'on n'a considéré que deux produits d'assurance spécifiques (épargne et rente), ce qui pourrait limiter la portée de nos résultats. Des recherches futures pourraient étendre notre approche à d'autres types de produits d'assurance, tels que les produits d'assurance vie ou les produits d'assurance maladie.

Il est aussi important de noter que notre étude a été menée dans un contexte spécifique et sur des données spécifiques. Il est donc possible que les résultats obtenus ne soient pas généralisables à d'autres contextes ou à d'autres ensembles de données.

Enfin, des tests supplémentaires pourraient être introduits pour appréhender l'impact de cette construction sur le ratio de couverture (SCR) ou la sensibilité à d'autres méthodes.

En résumé, notre étude représente une contribution importante à la littérature sur l'ALM en proposant une méthode cohérente et générale pour construire des MP qui optimisent réellement le processus ALM. Bien que notre étude ait des limites, elle ouvre de nouvelles perspectives pour la construction de model points.

# ANNEXE

# 1 Annexe A : Notions importantes de clustering

## 1.1 Relation et classes d'équivalence

Certaines méthodes de classification ont recours au concept de classe d'équivalence pour définir les classes que l'on cherche à identifier. En effet, dans une classe d'équivalence, tout point peut représenter cette car équivalent à n'importe quel autre point de sa classe.

**Rappel 1 :** Une relation d'équivalence est une relation binaire  $R$  sur un ensemble  $E$  qui est réflexive, symétrique et transitive. Plus formellement, une relation  $R$  est une relation d'équivalence sur  $E$  si :

- $R \subseteq E \times E$  (relation binaire sur  $E$ )
- $\forall x \in E, xRx$  (réflexivité)
- $\forall x, y \in E, xRy, \text{ alors } yRx$  (symétrie)
- $\forall x, y, z \in E, \text{ si } xRy \text{ et } xRz, \text{ alors } xRz$  (transitivité).

**Rappel 2 :** Étant donné un élément de  $x$  de  $E$ , l'ensemble des points  $y$  qui sont équivalents à  $x$ , est appelé la classe d'équivalence de  $x$ .

### 1.1.1 Similarité, dissimilarité et distance

Le concept de similarité est un concept que l'on rencontre dans de nombreuses disciplines et il y a à la fois une terminologie assez variable selon les auteurs et une mise en correspondance entre les ensembles, des caractéristiques proches, mais non équivalents. Ci-dessous, nous considérons un ensemble noté  $E$ .

**Définition 1 :** Une similarité sur  $E$  est une application  $s : E \times E \mapsto R^+$  qui vérifie :

- $\forall x, y \in E, s(x, y) = s(y, x)$  (symétrie)
- $\forall x, y \in E, s(x, x) \geq s(x, y)$

**Définition 2 :** Un indice de dissimilarité sur  $E$  est une application  $d : E \times E \mapsto R^+$  qui vérifie :

- $\forall x, y \in E, d(x, y) = d(y, x)$  (symétrie)
- $\forall x \in E, d(x, x) = 0$

On parle également parfois plus simplement de dissimilarité.

**Définition 3 :** Un indice de distance sur  $E$  est un indice de dissimilarité qui vérifie :

- $\forall x, y \in E, d(x, y) = 0$  implique que  $x = y$

On trouve dans la littérature l'appellation de dissimilarité propre.

Il y a parfois un amalgame entre dissimilarité et indice de distance.

**Définition 4 :** Une distance sur  $E$  est une application  $d : E \times E \mapsto R^+$  qui vérifie :

- $\forall x, y \in E, d(x, y) = d(y, x)$  (symétrie)
- $\forall x, y \in E, d(x, y) = 0 \Leftrightarrow x = y$
- $\forall x, y, z \in E, d(x, z) \leq d(x, y) + d(y, z)$  (inégalité triangulaire)

## 1.2 Hiérarchies, partitions...

L'objectif principal des méthodes de classification automatique est de répartir les éléments d'un ensemble en groupes, c'est-à-dire d'établir une partition de cet ensemble. Différentes contraintes sont bien sûr imposées, chaque groupe devant être le plus homogène possible, et les groupes devant être les plus différents possibles entre eux.

On considère ici des ensembles finis, donc des collections d'objets au sens habituel.  $A$  est un **ensemble** :

$$A = a_1, a_2, \dots, a_n \Leftrightarrow a_j \in A \text{ pour } 1 \leq j \leq n$$

Une partie de  $A$  un sous-ensemble :

$$B = b_1, b_2, \dots, b_p \mapsto b_k \in A \text{ pour } 1 \leq k \leq p$$

Si on compte la partie vide et l'ensemble tout entier, il y a  $2^n$  parties dans  $A$ . L'ensemble de toutes les parties de  $A$  se note  $\Phi(A)$ . Si  $A$  est formé de  $a, b, c, d$ ,  $(A)$  compte 16 éléments.

Deux parties d'un ensemble sont :

- soit **chevauchantes** (non égales et d'intersection non nulle),
- soit **disjointes** (sans élément commun, d'intersection nulle),
- soit **incluses** l'une dans l'autre,
- soit **égales**.

Une partition est un sous-ensemble de parties deux à deux disjointes dont la réunion fait l'ensemble tout entier.

Ainsi, Un ensemble quelconque de parties est formé de parties chevauchantes, disjointes ou incluses. Un ensemble de parties formant une partition ne comporte que des parties disjointes recouvrant le tout. Entre ces deux classes, la première trop large pour être utile et la seconde trop étroite pour être nuancée, on trouve les hiérarchies de parties.

Une **hiérarchie** de parties de  $\mathbf{A}$  est un ensemble de parties ayant quatre propriétés :

1. La partie vide en fait partie
2. Les parties réduites à un seul élément en font partie.
3. L'ensemble total  $\mathbf{A}$  lui-même en fait partie.
4. Si  $\mathbf{X}$  et  $\mathbf{Y}$  en font partie, alors soit  $\mathbf{X}$  et  $\mathbf{Y}$  sont disjointes, soit  $\mathbf{X}$  contient  $\mathbf{Y}$ , soit  $\mathbf{Y}$  contient  $\mathbf{X}$ .

Un arbre est un graphe raciné :

- les feuilles sont les parties à un seul élément (qui sont toujours dans une hiérarchie),
- la racine est l'ensemble tout entier (qui est toujours dans la hiérarchie).

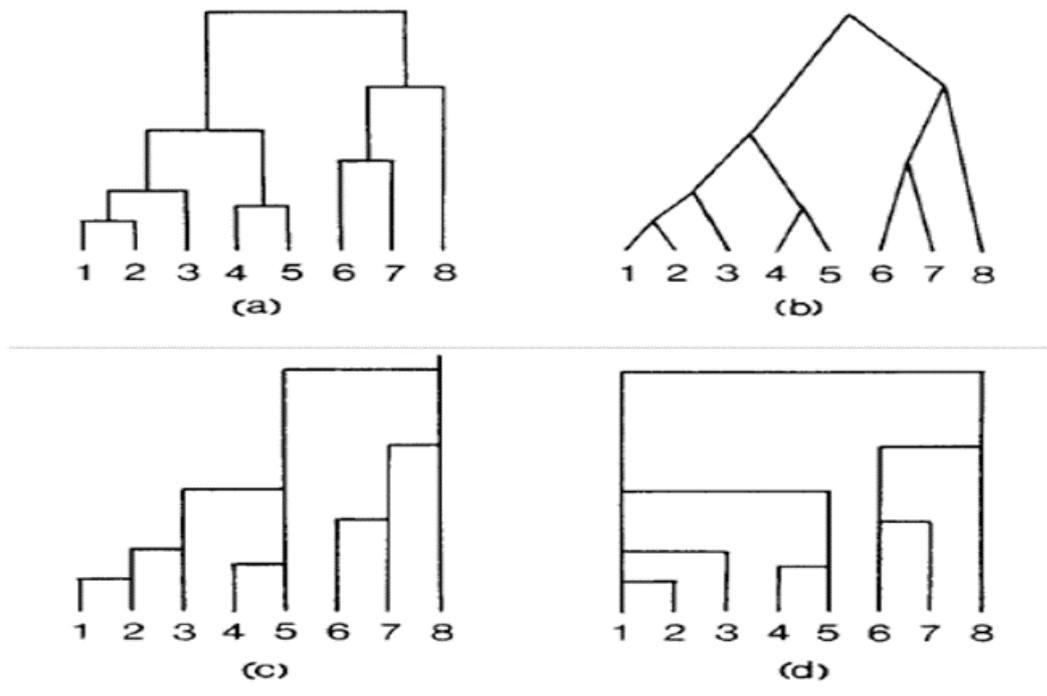


Figure 42: Représentation graphique d'un dendrogramme

Chaque partie n'a qu'un ancêtre, à l'exclusion de la racine qui n'en a pas. Si l'arbre est binaire, chaque partie a deux descendants, à l'exclusion des feuilles qui n'en ont pas. On dit alors que la hiérarchie est totalement résolue.

La hiérarchie est évaluée si à chaque partie, on peut associer une valeur numérique qui vérifie la définition :

$$X \mapsto Y \Leftrightarrow h(X) \leq h(Y)$$

où  $h$  est la fonction associant une valeur à la position d'un individu ou d'une classe d'individus dans la hiérarchie. La représentation graphique d'une hiérarchie évaluée s'appelle un dendrogramme. Il est essentiel de comprendre d'entrée que cette représentation est très peu contrainte. On a ci-dessous 4 représentations (parmi un très grand nombre possible) d'une hiérarchie évaluée.

## 2 Annexe B : Classification hiérarchique

En classification hiérarchique, on distingue les méthodes ascendantes et les méthodes descendantes. Les méthodes ascendantes créent une partie en regroupant deux parties existantes. Les méthodes descendantes divisent au contraire une partie existante pour en faire deux nouvelles.

Pour regrouper, il faut un critère. Au début, il est naturel de regrouper les deux individus les plus proches au sens de la dissimilarité de départ. Mais immédiatement après cette opération, on peut regrouper soit des individus, soit un individu et une classe, soit, un peu plus tard, deux classes. Plusieurs stratégies peuvent alors s'insérer dans le schéma général :

**Étape 1.** On dispose d'une matrice de dissimilarités entre  $n$  individus. Chaque individu donne une partie réduite à lui-même à laquelle on attribue la valeur 0. Prendre la plus petite valeur de cette matrice et faire avec le couple correspondant une partie à deux éléments. Attribuer à cette nouvelle partie une valeur positive. On a alors  $n-1$  parties.

**Étape 2.** À chaque pas, on a  $m$  parties et une valeur  $h(i)$  associée à chacune d'entre elles. Regrouper deux d'entre elles sur le critère  $M$  et attribuer à la réunion une valeur  $h$  supérieure ou égale à la valeur des deux composantes.

**Étape 3.** Recommencer jusqu'à ce qu'il ne reste que la classe regroupant le tout et lui attribuer une valeur supérieure à toutes les autres.

En conclusion, chaque procédé qui définit  $M$  et  $h$ , respectivement le choix pour le regroupement et la fonction de valuation, donne une classification hiérarchique particulière. Parmi les procédés les plus répandus figurent d'abord ceux qui sont basés sur les distances entre parties.

### 2.1 Distances et dendogrammes

La recherche d'une classification sur les individus dépend de différentes valeurs de  $h$  pour un même critère d'agrégation  $M$ . La fonction `hclust` de **R** en propose plusieurs et on va étudier les plus courantes :

- **Lien simple** : Saut minimum  $d(A, B) = \min(d(a, b))$
- **Lien complet**: Agrégation par le diamètre  $d(A, B) = \max(d(a, b))$
- **Lien moyen** :  $d(A, B) = \text{mean}(d(a, b))$
- **Critère de Ward** : une fonction de valuation particulière. Agrégation de Ward = Moment d'ordre 2 = Inertie minimale

### 2.2 Distances et variance

Le critère de **Ward** s'appuie sur la forte connexion entre les notions de distances et de variance. On a :

$$\text{var}_p(x) = \sum_{i=1}^n p_i(x_i - x)^2 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n p_i p_j (x_i - x_j)^2$$

$$var_p(x) = \frac{1}{n} \sum_{i=1}^n (x_i - x)^2 = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2$$

D'où la généralisation en terme d'inertie :

$$Iner(x) = \frac{1}{n} \sum_{i=1}^n \|x_i - x\|^2 = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \|x_i - x_j\|^2$$

D'où la généralisation en terme d'hétérogénéité :

$$Heter(\Omega) = \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n d_{ij}^2$$

avec  $\Omega$  une collection de  $n$  objets et  $d_{ij}^2$  le carré de la distance de l'objet  $i$  à l'objet  $j$ . On peut juger de l'hétérogénéité dans une partie (inertie intra-classe) ou entre parties (inertie inter-classe). On peut faire de la statistique avec des matrices de distances entre objets.

# Bibliography

- [AP23] Kamal ARMEL and Frédéric PLANCHET. Modèles financiers en assurance et analyses dynamiques. 2022-2023.
- [BFOS17] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.
- [COU13] A COULOUMY. Critères d’agrégations de jeu de données passif pour les calculs sous la formule standard de solvabilité 2. *actuarial master’s thesis, ISFA*, 2013.
- [da16] Institut des actuaires. Groupe de travail ”best estimate liabilities vie” exemples de pratiques actuarielles applicables au marché français. 2016.
- [dCP11] Autorité de Contrôle Prudentiel. Solvabilité 2: principaux enseignements de la cinquieme étude quantitative d’impact (qis5). *Banque de France*, 2011.
- [dR22] Andrie de Vries and Brian D. Ripley. *ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'*, 2022. R package version 0.1.23.
- [Fal11] Alaeddine Faleh. *Allocation stratégique d’actifs et ALM pour les régimes de retraite*. PhD thesis, Université Claude Bernard-Lyon I, 2011.
- [GG15] Pierre-Olivier Goffard and Xavier Guerrault. Est-il optimal de regrouper les souscripteurs par âge, sexe et ancienneté pour les calculs bel basés sur les model points ? *European Actuarial Journal*, 5(1):165–180, 2015.
- [HEL] Alexandre HELUIN. Titre: Solvency ii: Techniques de modélisation du best estimate en assurance-vie.
- [Kas23] Alboukadel Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*, 2023. R package version 0.6.0.
- [KM20] Alboukadel Kassambara and Fabian Mundt. *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*, 2020. R package version 1.0.7.
- [M13] Daniel Müllner. fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python. *Journal of Statistical Software*, 53(9):1–18, 2013.
- [Mil22] Stephen Milborrow. *rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'*, 2022. R package version 3.1.1.
- [Oth] Jabri Othman. Application des méthodes de data mining à l’agrégation du passif en prévoyance et santé.
- [PD07] C.E. Pardo and P.C. DelCampo. Combinacion de metodos factoriales y de analisis de conglomerados en r: el paquete factoclass. *Revista Colombiana de Estadística*, 30(2):231–245, 2007.

- [Qur13] T Qureshi. Scénarios économiques dans un contexte solvabilité ii. *Mémoire d'actuariat*, 2013.
- [R C23] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023.
- [TA22] Terry Therneau and Beth Atkinson. *rpart: Recursive Partitioning and Regression Trees*, 2022. R package version 4.1.19.