

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

Par : Laura BASCHUNG

Titre : Estimation de la FDB : mise en place d'un proxy sur la base d'indicateurs financiers

Confidentialité : NON (Durée : 1 an 2 ans)

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

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

..... François Xavier Négri

.....

.....

*Membres présents du jury de
l'ISFA*

..... Esterina Masiello

.....

.....

Entreprise :

Nom : MALAKOFF HUMANIS

Signature :

*Directeur de mémoire en entre-
prise :*

Nom : Imen SAID

Signature : said

Invité :

Nom :

Signature :

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

Signature du responsable entreprise

said

Signature du candidat

Baschung

Remerciements

Je tiens à remercier Papa Maguette GUEYE de m'avoir permis de faire cette alternance ainsi que de m'avoir donné ce sujet de mémoire. Je remercie également l'équipe Mesure De la Solvabilité pour leur accueil et leurs conseils.

Je remercie plus particulièrement ma tutrice en entreprise Imen SAID, manager de l'équipe, d'avoir accepté de reprendre mon suivi, de m'avoir fait bénéficier de son expérience, ainsi que de toute son aide qu'elle m'a apportée pour ce mémoire.

Je remercie également l'équipe Gouvernance Des Modèles, pour toute l'aide et les conseils qu'ils m'ont apportés.

Ma reconnaissance s'adresse également à ma tutrice académique Esterina MASIELLO pour son encadrement, sa disponibilité et ses remarques constructives.

Enfin, je remercie grandement ma famille pour leurs soutiens tout au long de mes études ainsi que pour la relecture de ce mémoire.

Liste des abréviations

ALE : Accumulated Local Effects plots
AOA : Absence d'Opportunité d'Arbitrage
BSDV : *Black-Scholes with Deterministic Volatility*
BEG : *Best Estimate Garantie*
BEL : *Best Estimate Liability*
CF : Coefficients de flexing
DDLMM : *Displaced Diffusion Libor Market Model*
ERE : Épargne-Retraite
FDB : *Future Discretionary Benefits*
GRU : *Gate Recurrent Unit*
GSE : Générateur de Scénario Économique
IA : Intelligence Artificielle
IT : Intérêt Technique
LSTM : *Long-Short Term Memory*
MAPE : Erreur Moyenne Absolue en Pourcentage
MSE : Erreur Quadratique Moyenne
PB : Participation aux Bénéfices
PDP : Partial Dependence Plots
PE : Provision d'Égalisation
PFI : Permutation Future Importance
PM : Provision Mathématique
PPE : Provision pour Participation aux Excédents
PS : Prévoyance-Santé
RMSE : Racine de l'Erreur Quadratique Moyenne
RNN : Réseau de Neurones Récurrents
SCR : Capital de Solvabilité Requis
TME : Taux Moyen d'emprunt D'État
TMG : Taux Minimum Garanti
TRC : Taux de Rendement Comptable
VM : Valeur de Marché
ZC : Zéro-Coupon

Résumé

La directive Solvabilité II impose aux assureurs le calcul de leurs provisions techniques, notamment composés par le *Best Estimate*. La détermination de ce dernier nécessite un modèle ALM, stochastique de par l'utilisation d'un générateur de scénarios économiques risque-neutre. L'une des raisons de cet usage est la participation aux bénéfiques. Cette dernière est une partie du résultat financier reversée aux assurés, et comporte un minimum réglementaire. Compte tenu de toutes les projections qu'il effectue, le recours à ce type de modèle est chronophage.

Ce mémoire a pour but de construire un modèle qui estimera la partie stochastique du *Best Estimate*. Diverses études pourront alors être menées suivant les différents contextes économiques, entraînant ainsi l'élaboration de stratégies de manières plus efficaces. On utilisera le *machine learning*, qui est une méthode de plus en plus développée dans ce domaine, afin de mettre en place ce proxy.

Pour cela, une courte présentation de la solvabilité ainsi qu'une analyse du modèle ALM seront d'abord menées afin de comprendre qu'elles variables seront essentielles pour créer le proxy. La méthode utilisée sera un réseau de neurones, dont les paramètres seront optimisés. Ce modèle sera ensuite interprété par l'intermédiaire de différentes méthodes, pour enfin tester les impacts et limites de ce proxy sur certains sous-modules de SCR.

Mots clés : Solvabilité II, *Best Estimate*, modèle ALM, GSE, risque-neutre, participation aux bénéfiques, *machine learning*, stratégies, proxy, SCR

Abstract

The Solvency II Directive imposes to insurance companies the calculation of their technicals provisions, especially composed by the Best Estimate. The latter determination requires an ALM model, which is stochastic through the use of a risk-neutral economic scenario generator. One of the reasons for this use is profit-sharing. It is the part of financial outcomes given to the policyholders, and includes a regulatory minimum. Using this type of model is time consuming, in view of all the projections it makes.

The goal of this paper is to build a model which estimates the stochastic part of the Best Estimate. Various studies can be carried out according to the different economic environments, leading to an early strategy development. We will use machine learning, which is more and more developed in this field, to set up the proxy.

For this purpose, a small Solvency II presentation along with the ALM model analysis will be broached in order to understand the primordial variables to create the proxy. The method used will be a neuronal network, with optimized factors. Thanks to several methods, it will afterward be interpreted, to finally test the limits and impacts on some under-units SCR.

Key words : Solvency II, Best Estimate, ALM model, ESG, risk-neutral, profit-sharing, machine learning, strategy, proxy, SCR

Sommaire

Remerciements	1
Liste des abréviations	2
Résumé	3
Abstract	4
Introduction	7
1 Cadre de l'étude	9
1.1 Contexte	9
1.1.1 L'épargne retraite	9
1.1.2 La prévoyance santé	10
1.2 Solvabilité II	11
1.2.1 Généralités	11
1.2.2 Le pilier 1	12
1.2.3 Le pilier 2	13
1.2.4 Le pilier 3	13
1.3 Le <i>Best Estimate</i>	14
1.3.1 Définition	14
1.3.2 L'épargne retraite	15
1.3.3 La prévoyance santé	15
1.4 Le SCR	16
1.4.1 La Formule Standard	16
1.4.2 Le modèle interne	18
2 Modélisation de l'actif et du passif	19
2.1 Présentation de l'entité étudiée	19
2.2 Le modèle passif	19
2.2.1 Les hypothèses	19
2.2.2 Les flux du passif	20
2.3 La modélisation des actifs	22
2.3.1 Structure des actifs	22
2.3.2 Les hypothèses	23
2.3.3 Projection des actifs	23
2.4 Le modèle ALM	29
2.4.1 Les hypothèses	29
2.4.2 Le fonctionnement	29
2.4.3 Le Générateur de Scénarios Économiques	30
2.4.4 Le <i>Flexing</i>	33
2.4.5 Les flux de BEL	34
3 Décomposition de la FDB	36
3.1 La partie épargne-retraite	36
3.1.1 Définition	36

3.1.2	L'algorithme de participation aux bénéfices	36
3.2	La partie prévoyance santé	40
3.2.1	Définition	40
3.2.2	La participation aux résultats	40
3.2.3	L'absorption par la PE	41
3.3	Comparaison des variations	42
4	Mise en place du Proxy	43
4.1	Recherche des variables explicatives	43
4.1.1	Analyse sur deux scénarios extrêmes	43
4.1.2	La base de données	47
4.2	Implémentation d'un réseau de neurones	49
4.2.1	Principe	49
4.2.2	Construction des modèles	51
4.2.3	Résultats	55
4.3	Interprétations	58
4.3.1	Au niveau Locale	58
4.3.2	Au niveau Global	64
5	Impacts et limites	71
5.1	Impacts sur les nSCR	71
5.1.1	Longévité	71
5.1.2	Taux d'intérêt	73
5.1.3	Catastrophe	74
5.2	Les limites du modèle	75
5.2.1	Application sur un trimestre	75
5.2.2	Test sur une entité similaire	76
5.2.3	Test sur une entité disparate	76
	Conclusion	78
	Tables des figures	79
	Liste des tables	80
	Bibliographie	81
	Annexe	83
	Annexe I : Tables corrélations SCR	83
	Annexe II : Exemple implémentation réseau de neurones	84
	Annexe III : Résultats SHAP détaillés	85
	Annexe IV : Résultats PFI détaillés	86
	Annexe V : Table chocs Solvabilité 2 taux d'intérêt	87

Introduction

La directive Solvabilité II est une norme européenne du monde de l'assurance. Son but est d'harmoniser la réglementation des assurances et réassurances dans l'Union Européenne. Ce régime prudentiel permet de se prémunir des risques systémiques en régulant les fonds propres exigés, ce qui va permettre de mieux protéger les assurés, ainsi que les assureurs, pour qu'ils puissent honorer leurs engagements. C'est pourquoi il est obligatoire pour les compagnies d'assurance de calculer le *Solvency Capital Requirement* (SCR) qui est le capital dont on a besoin pour se prémunir contre tous les risques que la branche assurantielle peut subir.

Le bilan de l'entreprise est également impacté par cette directive, qui passe désormais d'une vision comptable à une vision économique. En effet, à l'actif, les placements sont considérés en valeur de marché, tandis qu'au passif le *Best Estimate Liability* (BEL) brut est calculé sous une probabilité risque neutre.

Le *Best Estimate* peut être décomposé en deux parties. D'un côté il y aura la partie garantie et de l'autre la partie discrétionnaire, estimées au minimum chaque année. Cette dernière est appelée la *Future Discretionary Benefit* (FDB), qui est la participation aux bénéfices que l'on va verser aux assurés dans le futur, en plus de celle garantie. Elle est déterminée par différence entre le *BE* et sa partie garantie, nommée le *Best Estimate Garantie* (BEG). Cependant, l'estimation de cette quantité est chronophage, car elle résulte d'un modèle stochastique opulent. La calibration, le traitement des données en entrées ainsi que le modèle en lui-même prennent un temps considérable, ce qui ne permet pas de faire des tests approfondis.

L'objectif de ce mémoire est de trouver un proxy, par l'intermédiaire de différents indicateurs, afin d'estimer la FDB sans avoir recours à ces nombreuses simulations. On pourra ainsi étudier l'impact de cette estimation sur le capital de solvabilité requis ainsi que les limites de l'utilisation de ce procédé.

Des travaux ont déjà été menés par Florian GACH et Simon HOCHGERNER en 2022, intitulé « *Estimation of Future Discretionary Benefits in traditional life insurance* », pour essayer d'estimer la FDB en la bornant. Ils estiment la borne supérieure et la borne inférieure de la FDB, ensuite utilisées dans le calcul de cette dernière. Pour notre part, nous allons adopter une approche plus *Datascience* en se servant du *Machine Learning* pour éviter le recours du modèle ALM à partir du logiciel *Addactis Modeling* utilisé par l'entreprise.

L'intelligence artificielle (IA) se déploie de plus en plus dans le monde de l'assurance. Elle fournit des méthodes innovatrices plus performantes dans l'utilisation des données. Les assureurs doivent apprendre à utiliser ces méthodes et les incorporer au mieux dans leur organisation pour en tirer le meilleur. La surveillance de la fiabilité et de la véracité des modèles ainsi que des données générées par une IA est nécessaire, car une mauvaise utilisation peut entraîner des

décisions causant des dommages importants. Le *Deep Learning* est une méthode spécifique de *Machine Learning*. Plusieurs recherches ont porté sur l'implémentation de ce type de méthode afin de déterminer un modèle de mortalité, ou encore la projection de séries temporelles, que l'on peut retrouver dans le livre de *F. Planchet* : « *Insurance data analytics [2020]* ». Une approche de provisionnement individuelle a également été proposée par *K. Kuo* en 2020, qui utilise un modèle de prévision de paiements à partir de réseaux de neurones. Le proxy mis en place dans ce mémoire se basera sur une technique de *Deep learning*, à savoir un réseau de neurones récurrents.

L'inconvénient du *Deep Learning* est l'interprétabilité du modèle. En effet, bien que ces modèles soient très performants, il est très complexe, voir impossible, d'expliquer directement leurs fonctionnements, et d'interpréter comment les valeurs explicatives sont utilisées pour arriver aux résultats. C'est pourquoi des méthodes d'interprétations pour des modèles complexes, dits « *black-box* », ont été instaurées. Cela concerne à la fois des méthodes d'interprétation locales et globales des modèles. Ces méthodes montrent de plus en plus leur importance avec le développement du *Machine Learning*. Chaque approche a une utilité précise, avec son lot d'avantages et d'inconvénients, comme peuvent le montrer les travaux menés par *Delcaillau D., Ly A., Papp A., et Vermet F. [2022]*, sur l'application de quelques unes d'entre elles.

J'utiliserai les données d'une entité du groupe Malakoff Humanis pour déterminer ce proxy. Cette entité dispose de garanties épargne-retraite, de santé complémentaire et de prévoyance. La recherche du proxy sera scindée en deux, une partie épargne-retraite et une partie prévoyance-santé. Les calculs des flux de BEG et de BEL seront nécessaires pour pouvoir en déduire les flux de FDB.

Dans un premier temps, je vais comparer les flux de FDB avec deux scénarios extrêmes, ainsi que le comportement de différentes variables qui lui sont liées, telles que les intérêts techniques ou la courbe des taux. Ceci permettra dans un second temps de déterminer les variables explicatives qui permettront de mettre en place un proxy. Il se construira par une méthode de *Deep Learning*, qui sera plus précisément un réseau de neurones récurrents. Par la suite, un test sera mené sur un modèle, dans le but d'analyser l'écart entre les résultats obtenus et ceux produits par le lancement du modèle ALM.

Une fois ceci achevé, il y aura d'abord une première interprétation locale du modèle, pour analyser le comportement des variables sur certaines observations. Deux techniques différentes seront appliquées et comparées. Puis on pourra passer à une interprétation globale en utilisant trois méthodes différentes, qui chacune apporte des informations complémentaires.

Enfin on pourra analyser les impacts de ces résultats sur trois sous-modules de SCR ainsi que l'application sur une autre entité afin de constater les limites et conséquences que peut avoir l'utilisation d'un tel proxy.

1 Cadre de l'étude

1.1 Contexte

1.1.1 L'épargne retraite

L'épargne-retraite désigne tous les produits qui permettent de se constituer un capital jusqu'à l'arrivée de la retraite. Le capital ne pourra être sorti ou utilisé que quand l'épargnant aura atteint la retraite, à l'exception près de certains événements devant être spécifiés dans le contrat.

Il existe deux types de supports d'investissement en assurance vie, et plus particulièrement en épargne-retraite. Il y a ceux avec des contrats en euro, et ceux avec des contrats en unités de compte (UC).

- Les contrats en euro désignent l'épargne placée en fond euros, majoritairement des obligations, qui sera capitalisée chaque année selon un taux défini à la signature du contrat. La capital placé au départ est garanti.
- Les contrats en UC sont des portefeuilles d'actifs financiers qui évoluent selon le marché. Ces contrats sont à capital variable.

Les unités de compte sont plus risqués que les fonds euros, car l'investissement de départ n'est pas protégé, mais ils peuvent avoir de meilleurs rendements.

Les contrats Euro font intervenir un Taux Minimum Garantie (TMG). Les assureurs ont l'obligation de revaloriser à minima l'épargne des assurés au TMG. Cependant, il existe une clause de participation minimale défini dans l'*Article A132-11* du « *Code des Assurances* » qui précise que les assureurs doivent reverser au minimum 85% de leur résultat financier et 90% de leur résultat technique. Ainsi, les assurés auront leurs contrats revalorisés au maximum entre leur TMG contractuel et la limite réglementaire.

De plus, les compagnies d'assurances peuvent choisir de verser des participations aux bénéficiaires (PB) en plus, lorsqu'une année produit de grand rendements. Ce procédé permet de fidéliser les clients et de faire face à la concurrence. La distribution de ces participations aux bénéficiaires discrétionnaires doit être modélisée et nécessite l'utilisation de modèle stochastique afin de prendre en compte divers scénarios. Ces flux auront un impact sur la partie épargne-retraite de la FDB.

Il existe donc trois niveaux de participations aux bénéficiaires :

- La participation contractuelle.
- La participation réglementaire.
- La participation discrétionnaire.

1.1.2 La prévoyance santé

La prévoyance-santé permet au contraire, au souscripteur ou autres bénéficiaires, de se prémunir contre des conséquences financières liées à un problème de santé. Cela peut provenir d'un accident ou d'une maladie.

Contrairement à l'épargne-retraite, la garantie première n'est pas la constitution de capital. Si le contrat arrive à maturité et que le risque ne s'est pas réalisé, l'assuré ne récupère pas la somme versée. On peut dire que ce contrat est à « fonds perdus ». La prévoyance-santé complémentaire permet de bénéficier d'autres garanties, en plus du régime de base obligatoire qu'est la Sécurité Sociale. De plus, si le contrat est collectif, le principe de mutualisation permet de bénéficier de meilleures garanties avec des cotisations réduites.

Les principaux risques en prévoyance-santé sont :

- le risque décès, avec les sous garanties :
 - Capital décès
 - Rente conjoint
 - Rente éducation
- le risque arrêt de travail
 - Incapacité
 - Invalidité
- le risque santé

Il est fréquent, dans ce type de produit, qu'il soit prévu dans le contrat des mécanismes de participation aux bénéfices afin de faire face à la volatilité naturelle des sinistres pour les opérations collectives.

1.2 Solvabilité II

1.2.1 Généralités

Entrée en vigueur le 1^{er} janvier 2016, Solvabilité 2 est une norme s'appliquant à toutes les entreprises d'assurance et de réassurance européennes. Elle a pour enjeu d'harmoniser le marché de l'assurance européen, de garantir le respect des engagements des entreprises, et de protéger les bénéficiaires de polices d'assurances.

Cette norme succède à Solvabilité I, instaurée dans les années 1970 puis révisée en 2002, qui adopte une vision comptable. Elle repose sur trois principes de bases :

- **Les provisions techniques** : elles doivent être constituées de manière suffisante et prudente afin de respecter l'engagement vis-à-vis des assurés.
- **Des actifs de qualités** : il doit y avoir suffisamment d'actif afin de rembourser ses dettes auprès des assurés, du personnel et de l'Etat. De plus les actifs doivent être équivalents aux engagements, diversifiés entre différentes classes d'actifs pour plus de sûreté, et liquide.
- **La marge de solvabilité** : le niveau des fonds propres doit être supérieur à un montant minimal définit dans la réglementation, c'est l'exigence de marge de solvabilité.

La mise en place de ce système est relativement simple, cependant il entraîne certaines limites. En effet, la vision comptable amène une évaluation de l'actif et du passif non cohérente avec le marché. L'allocation des actifs n'est pas prise en compte dans le calcul de la marge de solvabilité, ainsi deux compagnies ayant une allocation totalement différente peuvent avoir le même ratio de solvabilité. De plus, certains risques ne sont pas pris en compte, tels que le risque de souscription ou le risque de défaut, alors qu'ils ont pourtant un impact sur la solvabilité de l'entreprise. Enfin, en plus du manque d'uniformité au niveau européen de par l'hétérogénéité des méthodes de calculs des provisions, aucune surveillance sur le contrôle interne n'est imposée dans Solvabilité I.

C'est pourquoi, afin de combler ces faiblesses, Solvabilité II a été mise en place. Cette nouvelle directive va requérir à un bilan économique en remplacement du bilan comptable. Voici, en Figure 1, les principales différences entre Solvabilité I et II au niveau du bilan :

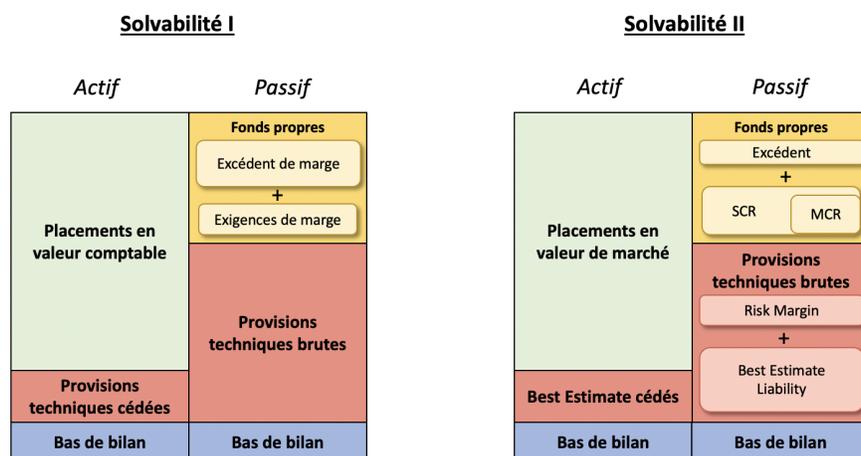


FIGURE 1 – Passage de SI à SII au bilan.

Cette réglementation se construit en trois piliers : un pilier quantitatif, un pilier qualitatif et enfin un pilier narratif.

1.2.2 Le pilier 1

Ce pilier définit toutes les exigences quantitatives, à savoir :

- Les modes de calculs des provisions techniques.
- Les règles de valorisation des actifs et des passifs.
- Les exigences de détention des fonds propres réglementaires.

Sous Solvabilité II, les provisions techniques sont la somme du *Best Estimate* et de la marge pour risque (*Risk Margin*). Elles servent à contrôler les risques liés aux sinistres qui influenceront par la suite les exigences de solvabilité. Le *Best Estimate* et la *Risk margin* doivent généralement être évalués séparément. La marge pour risque sert à prendre en compte les risques non répliquables et non mutualisables, car ces risques ne sont pas intégrés dans le calcul du *Best Estimate*.

Comme le montre en Figure 2, les actifs et passifs sont évalués en juste valeur, c'est-à-dire valorisés en valeur de marché.

Concernant les fonds propres, Solvabilité 2 va instaurer deux niveaux d'exigences :

- Le *Minimum Capital Requirement* (MCR), il correspond au niveau minimal de fond propre en dessous duquel, le risque de ne pas pouvoir honorer ses engagements est inacceptable. Il est calculé au moins une fois par trimestre. Si le niveau de fond propre est inférieur à ce seuil, les autorités de contrôles peuvent retirer l'agrément à la compagnie d'assurance.

- Le *Solvency Capital Requirement* (SCR), représente le niveau minimum de fond propre que la compagnie doit détenir afin d'assurer, dans 99,5% des cas, de ne pas être en faillite sur un horizon d'un an.

Dans ce pilier, les calculs se font sous la probabilité risque-neutre afin d'être *Market Consistent*, c'est-à-dire être en cohérence avec les prix du marché.

1.2.3 Le pilier 2

Ce second pilier consiste d'une part à définir les exigences qualitatives, telles que les règles de gouvernances et de gestion des risques, afin d'assurer une gestion prudente de l'entreprise et de ses risques. D'autre part, il établit les différentes règles de contrôles. L'objectif de ce pilier est double. Il a pour but d'assurer que la compagnie est bien gérée et en mesure de calculer et maîtriser ses risques, et aussi d'assurer qu'elle est bien capitalisée.

L'une des principales difficultés de ce pilier est qu'il se base sur le principe de proportionnalité. Ce principe signifie que le système de gouvernance doit être adapté à la nature, à la taille et à la complexité de l'entreprise. Il se divise en 4 fonctions clés : audit interne, gestion des risques, actuarielle, et vérification de la conformité, dont chacune est dirigée par un responsable unique.

Il met également en place un processus d'évaluation interne des risques et de la solvabilité, aussi appelé *ORSA* (*Own Risk and Solvency Assessment*). Ce processus réalise, entre autres, trois évaluations distinctes, dont certaines reprennent les exigences du pilier 1 sur un horizon plus élevé. L'*ORSA* doit être fait au moins une fois par an, et les résultats doivent être communiqués à l'ACPR (Autorité de Contrôle Prudentielle et Réglementaire).

1.2.4 Le pilier 3

L'objectif de ce dernier pilier est de définir les différents niveaux d'information à fournir au public et aux autorités de contrôle. Ainsi, la compagnie se doit d'être transparente, vis-à-vis des assurés et des investisseurs, quant aux informations fournies afin d'avoir une vision nette du marché de l'assurance. Enfin, l'entreprise doit également livrer à l'ACPR différents rapports, dans un but de surveillance et de contrôle.

Ce mémoire portera principalement sur les calculs du premier pilier.

1.3 Le *Best Estimate*

1.3.1 Définition

Par définition de l'Article 77 de la Directive Solvabilité II, le *Best Estimate of Liabilities*, ou encore la meilleure estimation, est : « la moyenne pondérée par leur probabilité des flux de trésorerie futurs, compte tenu de la valeur temporelle de l'argent, estimée sur la base de la courbe des taux sans risque pertinents. ».

Les flux concernés sont ceux liés à l'engagement pris vis-à-vis des assurés, ce sont donc : les frais, les primes, les commissions ainsi que les sinistres. Ces flux seront ensuite actualisés avec une courbe de taux sans-risque pertinente. L'EIOPA (European Insurance and Occupational Pensions Authority) fournit cette courbe de taux, mais les entreprises peuvent l'ajuster ou la redresser sous certaines conditions.

Dans le référentiel Solvabilité 2, les évaluations se font sous les hypothèses : d'Absence d'Opportunité d'Arbitrage (AOA), d'une probabilité risque-neutre et d'un univers *market consistent*. En outre, l'évaluation des provisions doit être au plus proche de la réalité. Cela concerne principalement les hypothèses telles que le rachat ou la sinistralité. Afin d'avoir la meilleure estimation la plus réaliste, on va avoir recours à des données historiques. Le *Best Estimate* est donc l'espérance des flux futurs actualisés sous une probabilité croisée risque-neutre et historique.

Ainsi le *Best Estimate Liability* s'écrit :

$$BEL = \mathbb{E}^{\mathbb{Q} \otimes \mathbb{P}} \left(\sum_{t \geq 0} \frac{CF_t^{\text{assurés}}}{(1 + i_t)^t} \right) \quad (1)$$

Avec :

- \mathbb{Q} : la probabilité risque-neutre
- \mathbb{P} : la probabilité historique
- $CF_t^{\text{assurés}}$: les cash-flows des assurés à l'instant t
- i_t : le taux sans risque de l'année t

Par ailleurs, le BEL tient compte des prestations discrétionnaires futures versées aux assurés. On peut décomposer le BEL en deux parties :

- La partie garantie : appelée le *Best Estimate Garantie* (BEG). C'est la projection des flux au taux minimum garanti. C'est ce que les assureurs doivent obligatoirement aux assurés, suite à l'accord contractuel.
- La partie discrétionnaire : ce sont les *Future Discretionary Benefits* (FDB), soit les participations aux bénéfices futures que l'assureur décide de reverser suivant les résultats techniques et financiers futurs.

Pour résumer :

$$BEL = BEG + FDB \quad (2)$$

Dans notre cas, la FDB aura une partie épargne-retraite (ERE) et une partie prévoyance-santé (PS).

1.3.2 L'épargne retraite

Dans le cadre de l'assurance vie, et principalement lorsqu'il y a un dispositif de participation aux bénéfices, il est nécessaire d'avoir recours à un modèle ALM pour prendre en compte les interactions actif-passif, et qui sera un modèle stochastique. Pour ce faire, on utilisera un générateur de scénarios économiques risque-neutre, qui projettera les flux d'un modèle ALM dans différents scénarios.

Puis par Monte-Carlo, on pourra estimer le BEL, avec N scénarios :

$$BEL \approx \frac{1}{N} \sum_{k=1}^N \sum_{i \geq 0} \frac{CF_{t,k}^{assurés}}{(1 + i_{t,k})^t} \quad (3)$$

Dans le contexte de Solvabilité 2, les projections se font en *run-off*, c'est-à-dire que les primes futures pour les contrats épargne-retraite ne sont pas prises en compte.

C'est la partie FDB qui exige l'utilisation d'un modèle stochastique et la prise en compte des interactions actif-passif. En effet, la participation aux bénéfices entraîne la revalorisation des contrats, qui elle-même dépendra de la performance financière de l'actif. Il est alors important de prendre en compte divers scénarios possible.

1.3.3 La prévoyance santé

Dans ce cadre-ci, il n'est pas essentiel d'utiliser un modèle stochastique. En effet, la participation aux bénéfices est contractuelle, et est gérée contrat par contrat de façon indépendante. Cependant, elle peut tout de même être modélisée dans un modèle ALM, ce qui sera notre cas.

De plus, on prend en compte les cotisations dans le calcul du BEL. Un BE prime et un BE sinistre va être calculé, et le BE total sera la somme du sinistre avec les primes. Les frais sont directement intégrés dans ces deux BEL.

1.4 Le SCR

Comme il a été spécifié, le SCR est l'une des exigences du pilier 1 de Solvabilité II. Il représente le capital requis afin de pouvoir absorber un choc dû à un événement extrême survenu dans l'année. Il doit être calculé au moins une fois par an.

Une méthode standard est proposée par le régulateur, cependant une compagnie peut passer à un modèle interne si elle estime qu'elle est plus en adéquation avec son profil de risque. Chacune des méthodes sont spécifiées dans le règlement délégué ainsi que dans la directive S2. Dans tous les cas, comme il est précisé dans l'article *R352-2 du Code des Assurances*, le SCR doit couvrir au minimum :

- Le risque de souscription vie, non-vie et santé ;
- Le risque de marché ;
- Le risque de crédit ;
- Le risque opérationnel.

1.4.1 La Formule Standard

La formule standard est plus simple que le modèle interne, mais elle est aussi moins spécifique à l'entreprise. Cette méthode utilise le profil de risque moyen des entreprises d'assurance, calibré de manière prudente. Ainsi le calcul du SCR se décompose selon la structure suivante (Figure 2) :

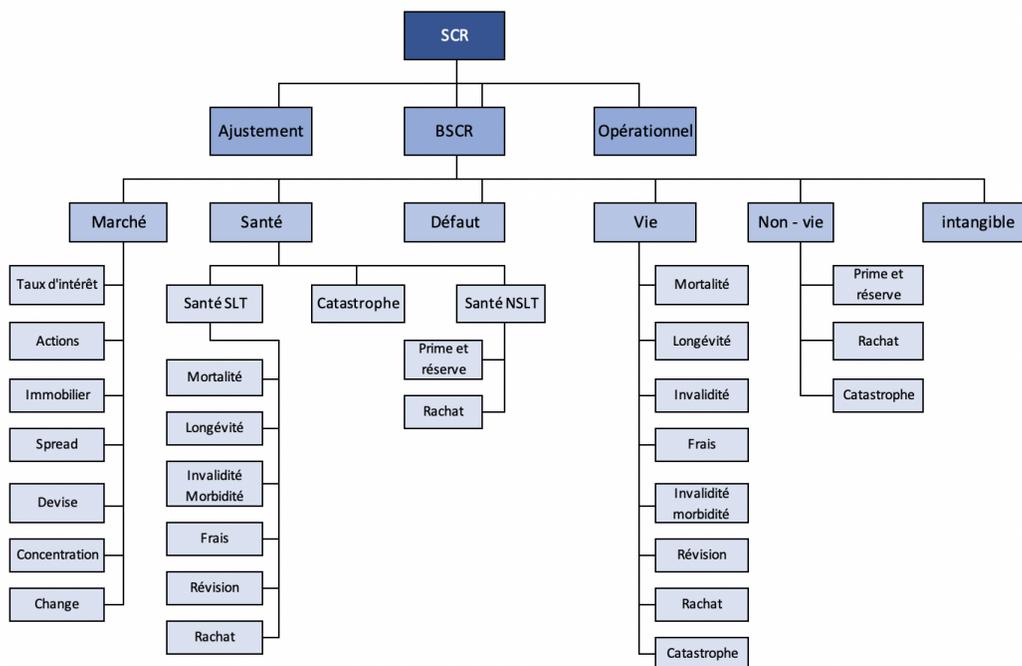


FIGURE 2 – Structure du SCR en Formule Standard

On peut tout d'abord décomposer le SCR de la manière suivante :

$$SCR = BSCR + SCR_{opérationnel} + Adj \quad (4)$$

Avec :

- Le *BSCR* est le capital de solvabilité requis de base, qui comprend différents modules de risque.
- Le *SCR opérationnel* représente les risques non pris en compte dans le BSCR.
- L'*Adj* représente l'*adjustement* et prend en compte la capacité d'absorption de pertes des provisions techniques et des impôts différés.

Le BSCR

Le BSCR est l'agrégation des six différents modules de risque de SCR, eux même décomposés en sous-modules de risques. Il se calcule généralement selon les étapes suivantes :

Étape 1 : Calcul des SCR sous-modulaires

On calcul les SCR de chaque sous-modules de risque. Pour cela, il y a deux approches de calculs selon le sous-module. Soit on va appliquer des chocs, dont leurs spécificités sont précisés dans la réglementation, soit on va utiliser une formule fermée.

L'approche par application de chocs consiste à calculer la *Net Asset Value* (NAV) pour un scénario central, puis pour un scénario choc, et d'avoir notre SCR sous-modulaire pas différence des deux.

$$SCR_{sous-module} = NAV_{centrale} - NAV_{choquée}$$

Étape 2 : Calcul des SCR modulaires

La déduction du SCR modulaire se fait par l'agrégation de tous les modules qui le composent, avec l'aide de matrices de corrélations propres à chaque module définies par la réglementation.

Étape 3 : Calcul du BSCR

Comme précisé précédemment, le BSCR s'obtient par l'agrégation des six SCR des modules avec des niveaux de corrélation définis dans la réglementation.

$$BSCR = \sqrt{\sum_{i,j} Corr_{ij} \times SCR_i \times SCR_j} + SCR_{intangible}$$

Avec :

- $Corr_{ij}$: le coefficient de corrélation entre le module i et j
- SCR_i : le SCR du module de risque i

Les coefficients de corrélation permettent de prendre en compte la diversité des activités afin de réduire l'exposition aux risques.

Le SCR opérationnel

Il permet de prendre en compte les risques opérationnels qui incluent : le risque occasionnant des pertes liées aux erreurs de procédures internes, de membres du personnel, de systèmes défaillants ou encore d'événements externes. Ce module de risque est indépendant des autres modules.

L'Ajustement

L'ajustement pour capacité d'absorption des pertes permet de diminuer le SCR de base par le biais de participation aux bénéfices et d'impôts différés. On se focalisera sur l'ajustement des pertes par les provisions techniques, aussi appelé l'ajustement FDB. Ainsi le lien entre la FDB et le SCR se fait, en partie, par cet ajustement.

De par son caractère général, la formule standard ne prend pas en compte divers paramètres qui ont pourtant un impact sur le montant du SCR. C'est pourquoi, certaines compagnies vont implémenter un modèle interne, afin d'avoir un SCR plus représentatif de leur profil de risque.

1.4.2 Le modèle interne

Certaines entreprises vont préférer mettre en place un modèle interne pour le calcul du SCR, afin que celui-ci soit en adéquation avec leur profils de risque. Ce modèle doit toutefois être validé par les autorités de contrôle. La méthodologie mise en place va utiliser des paramètres qui seront calibrés à partir de données spécifiques à l'entreprise. Cependant cette méthode est longue et complexe, c'est pourquoi beaucoup d'entreprises vont préférer adopter la formule standard.

Dans la suite de ce mémoire, les SCR seront calculés par la méthode de la formule standard.

2 Modélisation de l'actif et du passif

2.1 Présentation de l'entité étudiée

L'entité choisie est l'une des plus importantes du Groupe Malakoff Humanis. Elle est spécialisée dans l'épargne-retraite, la complémentaire santé et la prévoyance.

Elle est composée d'un portefeuille d'actifs variés et propose de nombreux produits. Parmi ces produits, on retrouve des produits d'épargne-retraite collectifs tels que le contrat « *Article 39* », « *Article 83* », « *Madelin* »..., avec des contraintes de Taux Minimum Garantie.

À ceci s'ajoute des produits de prévoyance et de santé complémentaire, tels que pour l'incapacité, l'invalidité, le décès

Pour mettre en place le proxy qui déterminera la FDB, il est important de comprendre comment fonctionne le modèle.

2.2 Le modèle passif

Un premier modèle va être utilisé, sans prendre en compte les interactions actif-passif. Cette modélisation déterministe permettra d'obtenir le BEG. L'épargne-retraite et la prévoyance-santé seront modélisées séparément.

2.2.1 Les hypothèses

Dans ce modèle, une ligne correspond à un assuré. Toutes les informations relatives aux contrats, telles que leurs âges, leurs TMG ..., seront mises en entrée du modèle. On ne prend pas en compte les nouveaux assurés futurs dans la projection.

Les tables de mortalité utilisées pour le calcul des rentes sont les tables réglementaires *TGF05* et *TGH05*.

On utilisera la courbe des taux sans risque de l'EIOPA pour déterminer le coefficient d'actualisation de l'année k :

$$Coef\ actu_k = \frac{1}{(1 + R_k)^k}$$

Avec R_k le taux zéro-coupon de l'année k en $t = 0$.

Enfin, seul le rachat structurel sera pris en compte pour l'instant, autrement dit les conjectures économique n'auront pas d'impact sur le rachat. La loi de rachat structurel est faite par rapport à l'historique des rachats passés selon l'ancienneté du contrat. Ainsi nous obtenons un rachat structurel moyen, pour tous les contrats, suivant (Figure 3) :

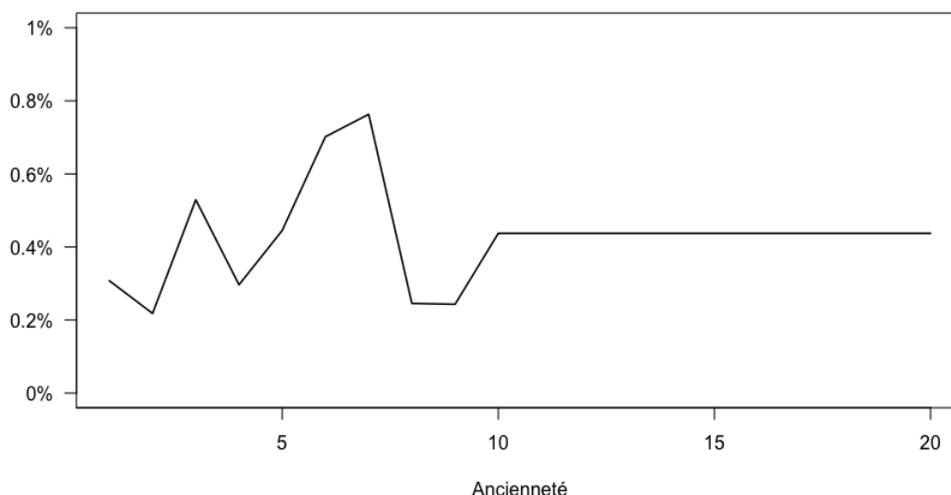


FIGURE 3 – Rachats structurels

À partir de dix ans d’ancienneté du contrat, la probabilité de rachat reste constante.

2.2.2 Les flux du passif

Après avoir posé les hypothèses dont nous avons besoin, les différents flux du passif vont pouvoir être projetés, et de ce fait, on obtiendra les flux du *Best Estimate Garantie*.

Voici un récapitulatif des principaux flux projetés dans le modèle :

	Prestations	Frais	Autres
Epargne	Décès, Rachats, Autres	Frais sur encours, Frais sur prestations	Provisions mathématiques, Intérêts techniques
Rente	Sortie de capital, Rentes versées, Rachats,	Frais sur encours, Frais sur prestations	Provisions mathématiques, Intérêts techniques
Prév-santé	Incapacité, Arrêt de travail, Décès ...	Frais sur prestations, Frais de gestion	Primes, Provision pour égalisation

Les provisions mathématiques sont estimées dans l’optique de pouvoir honorer les engagements de l’assureur, envers ses assurés, en ne tenant compte que des TMG et des taux techniques. Les différentes prestations et frais futurs probables sont également considérés.

La somme de ces flux actualisés donnera alors le BEG. De manière générale on aura :

$$BEG = Prestations - Primes + Frais - Chargements sur encours$$

Pour les contrats d'épargne, les primes ne seront pas prises en compte puisqu'elles ne contiennent pas de garanties financières pour les versements futurs. À la dernière année de projection, on met dans le BEG ce qu'il reste de provisions mathématiques.

Le BEG qui nous intéresse est celui net de réassurance. Il faut donc lui retirer les prestations et cotisations cédées ainsi que les commissions.

Les flux de BEG net de réassurance seront alors :

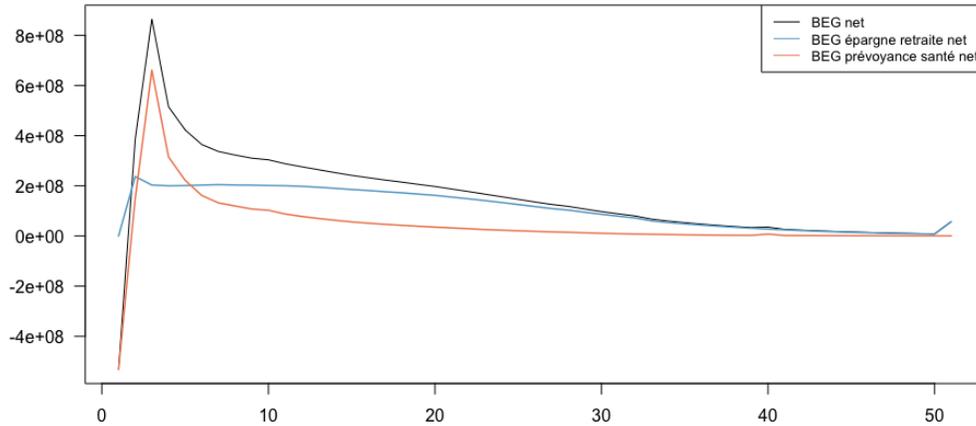


FIGURE 4 – Flux de BEG

Le montant négatif de l'année de projection pour la partie prévoyance-santé correspond aux primes acquises non émises, qui sont des primes afférentes à la période écoulée mais n'ayant pas été émises. Puis à partir de la première année de projection, les flux de BEG décroissent pour tendre vers 0. En effet, d'après les hypothèses les nouveaux assurés ne sont pas comptabilisés.

Le léger saut que l'on distingue à la dernière année de projection sur la courbe du BEG épargne-retraite, correspond aux provisions mathématiques restantes.

2.3 La modélisation des actifs

La modélisation des actifs jusqu'à leurs maturités, ainsi que l'application des stratégies d'investissement vont s'effectuer directement dans le modèle ALM. La projection des actifs permettra d'avoir les produits financiers futurs probables, ce qui aura un impact sur la participation aux bénéfices et donc sur la FDB.

2.3.1 Structure des actifs

Les actifs sont projetés par classe et selon leurs caractéristiques, qui seront préalablement renseignées en entrée du modèle. Les informations fournies se résument principalement par :

- Le libellé
- Le code CIC
- La notation
- La devise
- Le type d'actif
- Le prix d'acquisition
- La valeur de marché
- Les taux (remboursement, indexation, coupon ...)
- Le nominal
- Les différentes dates (acquisition, maturité, versement coupon, ...)
- Le spread
- La probabilité de défaut

Chaque classe d'actif aura un poids plus ou moins important dans le portefeuille. Cette répartition va jouer un rôle important pour les rendements futurs, et donc sur la participation aux bénéfices.

L'entité est composée de différents types d'actifs, dont l'allocation est la suivante (Table 1) :

Actions	3%
Obligations	58 %
OPCVM	21 %
Immobilier	12 %
Trésorerie	6 %

TABLE 1 – Allocation de l'actif

Parmi les obligations, on retrouve une grande majorité d'obligations à taux fixe, et une petite partie d'obligations à taux variable ainsi que des obligations indexées sur l'inflation.

2.3.2 Les hypothèses

Comme il a été précisé, la projection est effectuée en univers risque-neutre pour rester en cohérence avec les valeurs du marché et être en Absence d’Opportunité d’Arbitrage. Les actifs rapportent le taux sans-risque. On utilisera donc la courbe des taux sans-risque fournie par l’EIOPA.

L’allocation des actifs est constante tout au long de la projection. Par conséquent, la stratégie d’allocation devra prendre en compte les futurs réinvestissements à effectuer afin de maintenir l’allocation. Par exemple, lorsqu’une obligation arrive à maturité, les fonds devront être réinvestis dans de nouvelles obligations.

2.3.3 Projection des actifs

Pour chaque années, les valeurs de marché de toutes les classes d’actif seront calculées, avant la prise en compte de l’interaction actif-passif.

Actions

La valeur de marché des actions pour une année k , est la capitalisation de la valeur de marché, après interaction actif-passif, de l’année $k - 1$ au rendement de l’indice action.

Soit :

$$VM_{avant-inter}(k) = VM_{après-inter}(k - 1) \times (1 + Tx_{k,indice\ action})$$

Avec $Tx_{k,indice\ action}$ le taux de l’indice action de l’année k .

Immobilier

La valeur de marché pour les actifs immobilier est calculée de la même manière que les actions, seulement la capitalisation portera sur l’indice immobilier.

OPCVM

Il faut regarder la nature du fond de l’OPCVM. Dans notre cas, on modélise des OPCVM diversifiées. Ce qui signifie que le portefeuille des OPCVM est composé de différents types de fond. Il faut donc séparer le calcul des valeurs de marché des parts de taux, d’action et d’immobilier. Pour les parts action et immobilier, les calculs sont identiques à ceux expliqués précédemment.

Pour la partie taux, on va utiliser le taux *forward* 1 an de l’année k . Ces taux sont obtenus à partir de la courbe des taux zéro-coupon.

En effet, un taux zéro-coupon, également appelé taux *spot*, noté $R(t, T)$, est le taux de rendement actuariel d’une obligation zéro-coupon de maturité T acheté en t . On rappelle que le taux actuariel est le taux d’intérêt qui permet d’égaliser le prix d’un actif financier et ses flux futurs actualisés.

Le taux *forward*, noté $F(t, T_1, T_2)$, est le taux d'intérêt fixé en t pour les emprunts sur la période $[T_1, T_2]$. Ce taux peut être illustré de la manière suivante (Figure 5) :

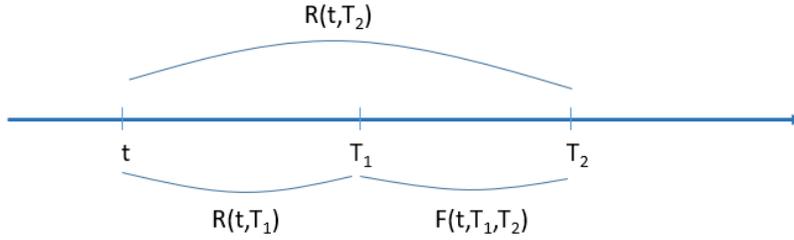


FIGURE 5 – Illustration d'un taux forward

On se place en convention composée, on a alors :

$$(1 + R(t, T_2))^{T_2-t} = (1 + R(t, T_1))^{T_1-t} \times (1 + F(t, T_1, T_2))^{T_2-T_1}$$

D'où

$$F(t, T_1, T_2) = \left(\frac{(1 + R(t, T_2))^{T_2-t}}{(1 + R(t, T_1))^{T_1-t}} \right)^{\frac{1}{T_2-T_1}} - 1$$

Or dans notre cas, on se place en $t = 0$, et $T_{i+1} - T_i = 1$

On a donc :

$$F(k, k + 1) = \frac{(1 + R_{k+1})^{k+1}}{(1 + R_k)^k} - 1 \quad (5)$$

Avec :

- $F(k, k + 1)$ le taux forward d'un an de l'année k ;
- R_k le taux zéro-coupon de maturité k .

La valeur de marché pour les OPCVM de produits de taux sera :

$$VM_{avant-inter}(k) = VM_{après-inter}(k - 1) \times (1 + F(k - 1, k))$$

Obligations

Le principe de calcul des valeurs de marché pour les trois types d'obligations est le même, cependant il y aura quelques variables à prendre en compte pour les OATi et les OTV en complément.

La valeur de marché pour ces obligations se calcule sur ce principe :

$$VM(t) = \sum_{i=1}^T \frac{coupon}{(1+r_i)^i} + \frac{Nominal}{(1+r_T)^T}$$

Avec :

- T la maturité de l'obligation
 - r_i le taux *forward* de l'année i
-
- Les OTF : la valeur de marché de chaque obligation est calculée avec l'actualisation des flux futurs, qui sont le nominal et les coupons à taux fixe. De plus il faut prendre en compte le facteur de défaut.
 - Les OTV : la méthode reste la même, cependant les taux de coupons ne sont pas fixes. Ils évoluent en fonction de la courbe des taux et sont additionnés d'un *spread*.
 - Les OATi : également identique aux OTF hormis la gestion de l'inflation qui s'ajoute. En effet, les nominaux sont indexés sur l'inflation, et les coupons seront déterminés sur ces nominaux indexés.

Suite à l'interaction entre l'actif et le passif, une valeur cible sera calculée. Pour une classe d'actif, la valeur de marché cible dépendra de la somme des VM totales pour toutes les classes d'actifs, des cash-flows totaux et de son allocation. Ainsi on aura :

$$VM_cible_{classe_actif} = (VM_{totale} + Cash_flows_{totaux}) \times allocation_{classe_actif} \quad (6)$$

Dans la valeur de marché totale, on prend en compte les valeurs de marché des achats/ventes prévues l'année concernée. En fonction de la valeur de marché par rapport à la cible, on va investir ou désinvestir dans la classe d'actif. Les désinvestissements vont être transformés en plus-values. Les principales plus-values viendront :

- Des plus-values réalisées ;
- Des coupons ;
- Des intérêts ;
- Des loyers.

Comme on réinvestit à chaque pas de temps les fonds qui sont arrivés à maturité, les rendements ne devraient pas converger vers 0 dans le temps.

Pour un scénario central déterministe, les taux de rendements projetés, par classe d'actif, sont les suivants (Figure 6) :

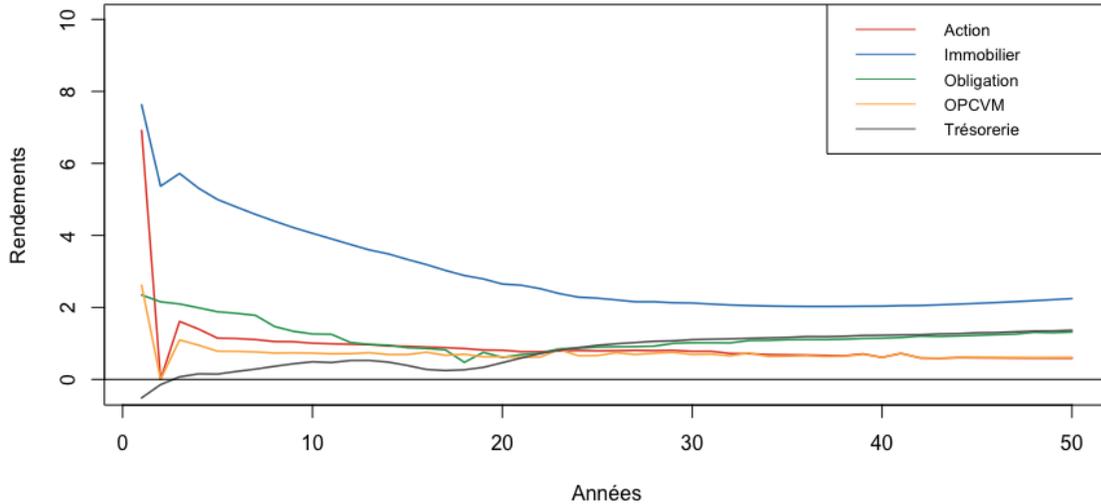


FIGURE 6 – Rendements par actifs

On remarque que l'immobilier est la classe d'actif qui a le plus haut taux de rendement. C'est aussi la classe la plus volatile, suivie par celle des actions. On observe également un pic descendant sur la deuxième année, principalement visible pour les classe d'actifs action, immobilier et OPCVM. Aussi, les courbes de taux de rendements pour ces trois classes ont la même allure, de part les calculs de leurs rendements assez similaires. Les obligations à taux fixe ont généralement une volatilité faible, cependant comme vu précédemment, certaines obligations ont des nominaux indexés sur l'inflation, ce qui ajoutera de la volatilité en plus de la courbe des taux sans risque. Enfin, pour ce qui est de la trésorerie, comme le rendement correspond au montant capitalisé par la courbe des taux, on y retrouve la forme de la courbe présente sur le schéma suivant.

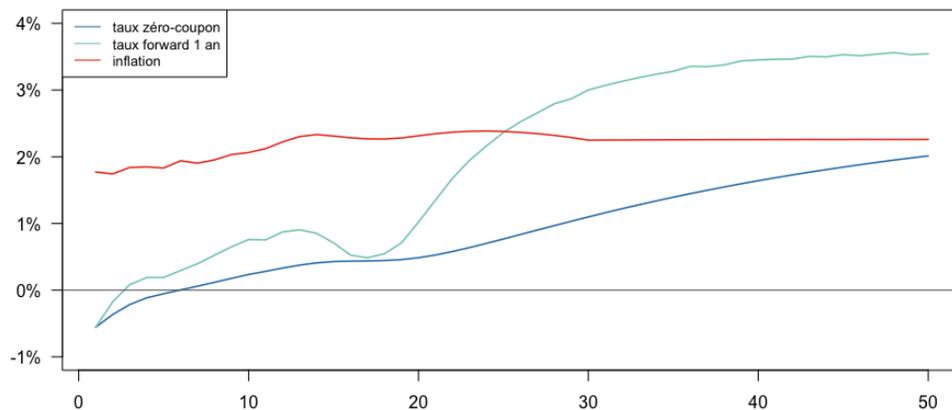


FIGURE 7 – Indicateurs des actifs

Comme montré en Figure 7, vers la 20-ième année de projection, il y a une rehausse des deux taux, dont on voit l’impact principalement sur les obligations.

L’indicateur pour les produits financiers totaux obtenus, qui correspondent aux rendements totaux, sera le Taux de Rendement Comptable (TRC). Il reflète le rendement attendu sur l’investissement (Figure 8).

On le calcul par :

$$TRC_t = \frac{\text{Produits financier net}}{\frac{1}{2}(\text{Montant_actifs}_t + \text{Montant_actifs}_{t-1})} \quad (7)$$

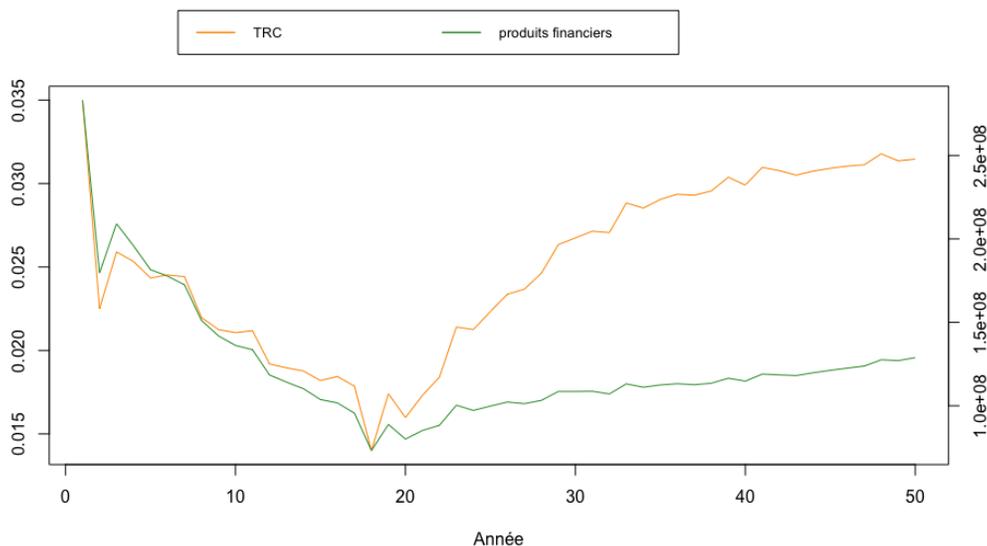


FIGURE 8 – Évolution du TRC et des produits financiers

Après l'obtention du montant final des produits financiers net de l'année modélisée, ils seront répartis entre l'épargne-retraite, la prévoyance-santé et les fonds propres. La partie qui nous intéressera principalement sera celle attribuée à l'épargne-retraite.

Une partie de ce montant sera les produits financiers destinés à la participations aux bénéfices. C'est pourquoi, il est important d'analyser les différents facteurs qui peuvent influencer le montant des produits financiers pour la mise en place du proxy.

2.4 Le modèle ALM

2.4.1 Les hypothèses

On prend en compte divers scénarios, toujours en univers-risque neutre, ainsi que la participation aux bénéfices discrétionnaires.

En addition du rachat structurel, le rachat dynamique (ou rachat conjoncturel) sera modélisé et dépendra cette fois de l'évolution du marché. On utilisera l'approche proposée dans les Orientations Nationales Complémentaires de 2013 proposé par l'ACPR. On modélisera le rachat dynamique par une fonction qui dépendra de l'écart entre le taux servi et le taux de rendement espéré par l'assuré. Ce taux de rendement espéré par l'assuré dépend du TME.

On aura alors la fonction de rachat dynamique suivante :

$$RD = \begin{cases} RD_{max} & \text{si } tx_{servi} - tx_{attendu} < \alpha \\ RD_{max} \frac{tx_{servi} - tx_{attendu} - \beta}{\alpha - \beta} & \text{si } \alpha < tx_{servi} - tx_{attendu} < \beta \\ 0 & \text{si } \beta < tx_{servi} - tx_{attendu} < \gamma \\ RD_{min} \frac{tx_{servi} - tx_{attendu} - \gamma}{\delta - \gamma} & \text{si } \gamma < tx_{servi} - tx_{attendu} < \delta \\ RD_{min} & \text{si } tx_{servi} - tx_{attendu} > \delta \end{cases} \quad (8)$$

Avec les paramètres :

α	β	γ	δ	RD_{min}	RD_{max}
-6%	-2%	1%	2%	-6%	20%

Dans notre cas, avec ces paramètres là, on pourra négliger le rachat dynamique.

2.4.2 Le fonctionnement

Comme expliqué dans la partie 2.2, le modèle va en premier lieu calculer les cash-flows des passifs avec uniquement la revalorisation garantie. Puis dans un second temps, il va récupérer ces cash-flows déterministes et, en comptabilisant les interactions actif-passif, va les revaloriser en utilisant la méthode du *Flexing* présentée dans la partie 2.4.4, pour chaque scénario économique. Le fonctionnement du modèle est illustré en Figure 9.

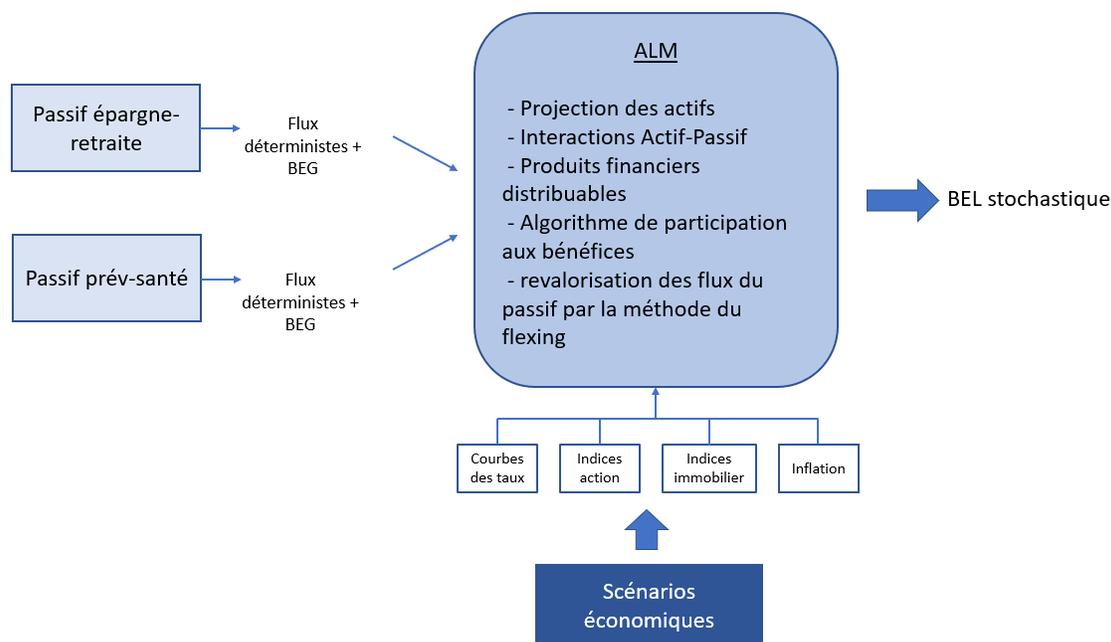


FIGURE 9 – Illustration du fonctionnement du modèle

2.4.3 Le Générateur de Scénarios Économiques

Le recours au générateur de scénarios économiques, que l'on notera GSE, en assurance vie est primordial pour pouvoir implémenter un modèle stochastique. Le GSE va projeter différents facteurs de risque, tels que le taux d'intérêt ou le taux d'inflation, dans divers scénarios, afin d'alimenter les calculs de modélisation du passif et de l'actif. On aura alors une estimation des rendements futurs des actifs et des valeurs futures du passif.

Le GSE utilisé dans le modèle étudié génère 2000 scénarios antithétiques en univers risque-neutre. Les indicateurs économiques projetés seront : le taux d'intérêt, l'indice actions, l'indice immobilier et l'inflation. On explicitera principalement le modèle de taux et celui de l'indice action.

Le modèle de taux

Le modèle de taux utilisé est le DDLMM (*Displaced Diffusion Libor Market Model*). La particularité de ce modèle est qu'il permet de générer des taux négatifs, ce qui est nécessaire dans le contexte économique actuel, puisque la modélisation doit refléter la réalité. Ce modèle est construit sur des taux *forwards*.

Sa dynamique à p facteurs, sous la probabilité *forward* neutre, est la suivante :

$$dF_k(t) = (F_k(t) + \delta) \sum_{q=1}^p \xi_k^q(t) \times dZ_{k+1}^q(t) \quad (9)$$

Avec :

- $F_k(t)$ le taux forward sur la période $[T_k, T_{k+1}]$;
- δ le coefficient de déplacement, appelé *shift*, qui permet la génération de taux négatifs ;
- ξ_k^q la volatilité du k -ième taux *forward* du facteur q ;
- dZ_{k+1}^q un mouvement Brownien sous la mesure *forward* neutre associé à la maturité T_{k+1} .

La volatilité permet de tenir compte de la corrélation entre les taux de différentes maturité. Elle est déterministe et se décompose en trois fonctions, dont deux fonctions paramétriques.

$$\xi_k^q(t) = f(t)g(T_k - t)b^q(T_k - t)$$

La fonction ***d'échelle***, notée f , dépend du temps et a deux paramètres :

$$f(t) = \theta + (1 - \theta)e^{-kt}$$

La fonction de ***forme***, notée g , dépend du temps qu'il reste à maturité et a quatre paramètres :

$$g(t) = (a + bt)e^{-ct} + d$$

La fonction de ***structure de corrélation interforward***, notée b^q , dépend également du temps restant à maturité. Elle est déterminée selon des données historiques.

Le modèle action

La modélisation des actions se fait à partir d'une extension du modèle Black & Scholes, qui sera le modèle BSDV (*Black-Scholes with Deterministic Volatility*). Le modèle Black & Scholes standard est utilisé pour projeter le cours des actions en temps continu. Le rendement des actions suit un mouvement Brownien géométrique avec une tendance et une volatilité constante. Le modèle BSDV utilise le modèle de taux stochastique et a une volatilité déterministe et dépendante du temps.

L'indice action $S(t)$ au temps t ($t \in [0, T]$), suit un processus log-normal et est donné par :

$$\frac{dS(t)}{S(t)} = r(t)dt + \sigma_{loc}(t)dW_t^{\mathbb{Q}} \quad (10)$$

Où :

- $r(t)$ est le taux court nominal obtenu avec le modèle de taux ;
- $\sigma_{loc}(t)$ est la volatilité déterministe dépendante du temps ;
- $dW_t^{\mathbb{Q}}$ est un mouvement Brownien sous la probabilité risque-neutre \mathbb{Q} .

Il n'y aura pas de dividende pour la modélisation des actions, pour moins de complexité.

Les tests de Martingalité

Pour vérifier la cohérence des scénarios et la justesse de la calibration des modèles, des tests sont mis en place. Les tests de martingalité permettent de vérifier la risque neutralité des scénarios économiques.

Pour les déflateurs, on vérifie qu'ils sont bien centrés sur la chroniques des zéro-coupons initiaux. Autrement dit, on actualise en moyenne au taux sans risque, soit :

$$\mathbb{E}^{\mathbb{Q}}(D(t)) = P_n(0, t)$$

Avec :

- $D(t)$ le déflateur à l'instant t ;
- $P_n(0, t)$ le prix du zéro coupon à l'année t .

En faisant ce test, on obtient :

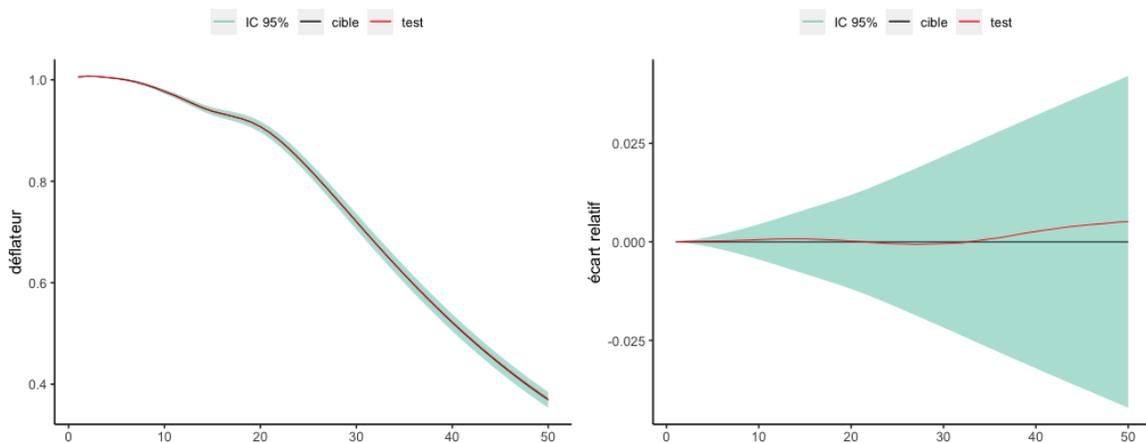


FIGURE 10 – Test de martingalité sur les déflateurs

La cible est la chronique des zéro-coupons. Nos données se trouvent à l'intérieur de l'intervalle de confiance, elles sont donc fiables à 95%.

Pour le test sur les actions, on vérifie qu'en moyenne elles ont rapporté le taux sans risque, soit :

$$\mathbb{E}^{\mathbb{Q}}(D(t)S(t)) = S(0)$$

Les résultats sont présentés en Figure 11 :

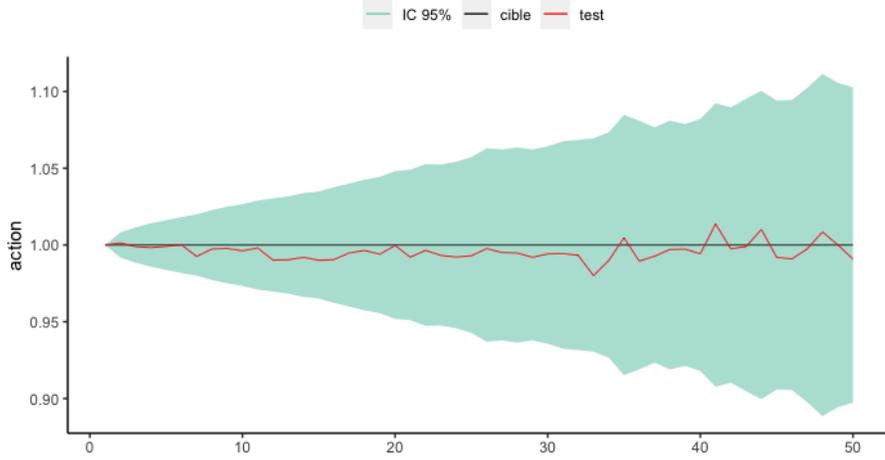


FIGURE 11 – Test de martingalité sur les actions

Là encore, on se retrouve à l'intérieur de l'intervalle de confiance, ce qui valide notre scénario économique. Si l'on veut vérifier le modèle pour l'immobilier, on aura recours à un test identique.

2.4.4 Le *Flexing*

Le *Flexing* consiste à utiliser les flux déterministes projetés au TMG pour les ajuster après la projection de l'actif en tenant compte de la revalorisation. Tous les éléments déterministes seront projetés tête par tête, puis agrégés en *Models Points*. Ces flux seront ensuite utilisés dans le modèle ALM, et seront ajustés à chaque pas de temps à l'aide de « coefficient de flexing », afin de tenir compte de la participation aux bénéfices. C'est le principal mécanisme qui joue sur l'interaction Actif-Passif. Les coefficients de flexing reposeront donc sur la proportionnalité des *cash-flows*.

Les coefficients de flexing se construisent de manière récurrente. On aura la provision mathématique (PM) revalorisée qui sera :

$$PM_{ALM}(t) = PM_{ALM}(t - 1) + IT_{ALM}(t) + PB_{servie}(t) - Chargements_{ALM}(t) \quad (11)$$

par conséquent,

$$Coeff_flexing(t) = \frac{PM_{ALM}(t)}{PM_{passif}(t)} \quad (12)$$

ou encore,

$$Coeff_flexing(t) = tx_IT(t) + tx_PB_servie(t) - tx_Chargement(t)$$

Avec :

$$IT_{ALM}(t) = IT_{ALM}(t - 1) \times Coeff_flexing(t - 1) \quad (13)$$

$$Chargements_{ALM}(t) = Chargements_{ALM}(t - 1) \times Coeff_flexing(t - 1)$$

Les avantages :

Le *Flexing* permet de gagner en temps de calcul. Les cash-flows ne sont pas simulés à chaque fois mais flexés. Tout ce qui est proportionnel n'est calculé qu'une seule fois, pour avoir des calculs plus rapides.

Aussi, la projection au TMG se fait police par police, ce qui permet de faire des analyses précises à ce niveau là.

Les inconvénients :

Bien que cette méthode soit plus rapide, elle reste plus complexe et moins abordable à mettre en place. Les mécanismes ALM sont masqués par la proportionnalité, ce qui rend la lecture plus délicate. Enfin, les agrégations en entrée du modèle ne permettent pas l'analyse d'une police donnée dans le modèle ALM.

2.4.5 Les flux de BEL

À la sortie du modèle ALM, on obtient alors le BEL dont on calcule les flux de la même façon vu pour le BEG.

On peut observer le BEL pour un scénario central déterministe en Figure 12 :

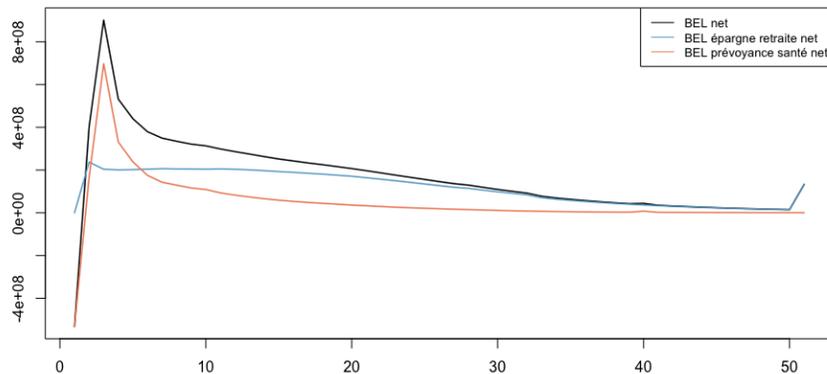


FIGURE 12 – Flux de BEL pour un scénario central

On constate que les flux sont sensiblement proches du BEG, que l'on observe dans la figure 4, et ont la même tendance. Si l'on regarde attentivement, c'est la partie du BEL épargne-retraite qui varie le plus, expliqué par l'incertitude d'un plus grand nombre de paramètres.

On peut les comparer avec les flux d'un BEL stochastique (Figure 13) :

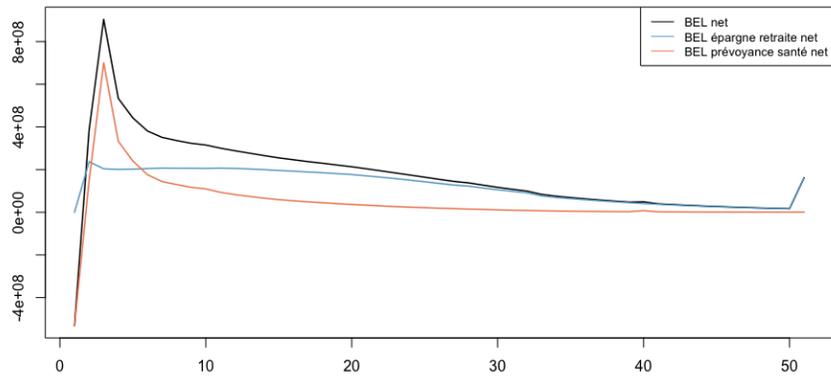


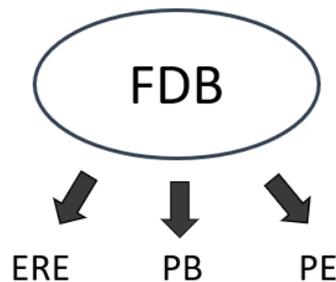
FIGURE 13 – Flux de BEL stochastique

Le BEL épargne-retraite est plus élevé et le montant des provisions ajoutées en dernière année est plus important. La forme des flux reste cependant la même.

3 Décomposition de la FDB

La FDB dépend des gains financiers et techniques. Une part de ces gains va être utilisée pour l'épargne-retraite, une autre pour la prévoyance-santé et le reste est placé en fond propre. Ce sont les deux premières parts qui nous intéressent, car ce sont celles qui vont impacter notre variable de recherche.

Comme nous l'avons vu en section 1.3.1, la FDB épargne-retraite se calcule séparément et indépendamment de celle de la prévoyance-santé. Pour notre proxy, on décomposera la FDB en trois parties, comme ceci :



3.1 La partie épargne-retraite

3.1.1 Définition

On définit cette FDB par la somme actualisée des participations aux bénéfices discrétionnaires futurs projetées. Il est important de définir la politique de distribution de la participation aux bénéfices qui va revaloriser les contrats. Cette politique de distribution peut être considérée comme un processus d'optimisation du résultat, sous des contraintes réglementaires et concurrentielles.

3.1.2 L'algorithme de participation aux bénéfices

Il existe plusieurs méthodes de distribution de la participation aux bénéfices, qui vont avec différentes méthodes de gestion de la PPE (Provision pour Participation aux Excédents). Les deux méthodes de gestion de la PPE les plus utilisées sont :

1. On donne la participation aux bénéfices générée l'année considérée, et on dote en PPE la différence entre la cible et ce qui est servi en cas de surplus, ou on reprend ce qu'il manque de la PPE ;
2. L'autre méthode consiste à mettre toute la participation aux bénéfices de l'année, et on reprend le montant de la cible dans la PPE déjà existante. Cette méthode a la particularité d'atténuer les reprises fiscales et de lisser la PPE pour plus de sécurité les années moins performantes.

Comme le montre l'exemple de l'étude menée par Gerondeau E. [2017], dans tous les cas on fini avec une PPE généralement vide les dernières années de projection. De plus, l'utilisation des deux méthodes ne montre pas de différence significatives sur le ratio de solvabilité.

Le modèle ALM étudié dans ce mémoire utilise la première méthode décrite ci-dessus.

Dans tous les cas, la distribution de participation aux bénéfices de notre modèle suit un processus à plusieurs étapes :

Étape 1 : Détermination des produits financiers disponibles pour la PB

On détermine les produits financiers qui découlent des actifs, en leur retraitant les frais. Puis on alloue un pourcentage de ce résultat pour la participation aux bénéfices. Elle doit prendre en compte divers scénarios économique pour faire face à toutes les situations et ne pas se retrouver en perte. Le montant obtenu suite à cette étape servira de base pour la suite du processus.

Étape 2 : Respect du TMG et du niveau réglementaire

On vérifie que le montant de produit financier de l'étape 1 satisfait le maximum entre le TMG et le taux minimum réglementaire. En général, le TMG proposé est supérieur au taux minimum imposé.

Si ce n'est pas le cas, on aura recours à 3 niveaux de leviers :

- Le premier levier consiste à aller puiser dans la Provision pour Participation aux Excédents (PPE). Cette provision est destinée exclusivement à la participation aux bénéfices. On la dote lorsque les produits financiers sont supérieurs aux taux cibles pour se prémunir contre les années où les produits financiers sont inférieurs aux TMG ou à la cible fixée. Cette provision suit deux règles :
 - On ne peut reprendre plus de 50% de son montant par année ;
 - Le plafond de PPE ne peut dépasser 10% de PM attribuées aux PB.
- Le second levier est de réaliser des plus ou moins values latentes sur les actifs.
- Le troisième levier est l'abandon de sa marge sur le résultat.

Étape 3 : Le taux cible

Ce taux représente un indice de satisfaction. C'est ce qu'attend l'assuré. S'il est trop faible, cela incitera les assurés à partir, et s'il est trop élevé il y aura un risque de réaliser trop de pertes. Parfois il est déterminé selon le taux de la concurrence, qui est déterminé sur leur historique. Il peut être calculé comme un intervalle. Il est généralement supérieur au TMG ou au taux minimum réglementaire.

Notre taux cible est défini comme environ :

$$Taux_cible(t) = 80\% TME(t) \quad (14)$$

On peut alors définir notre montant de participation aux bénéfices cible :

$$PB_cible(t) = (PM_PB(t) \times tx_cible(t) - IT(t)) + Chargements(t) \quad (15)$$

Ce montant de PB cible est en addition du montant de PB obligatoire vu précédemment.

Étape 4 : Le taux réellement servi

Après la détermination du taux cible, on peut calculer le taux que l'on servira réellement à nos assurés.

Si le montant de produit financier restant est inférieur au montant cible, alors on utilise les deux premiers leviers. Certains assureurs ont même recours au 3^{ème} levier d'abandon de marge pour atteindre le taux cible.

Si le montant de produit financier restant est supérieur au montant cible, alors on place la différence dans la PPE. On aura ensuite 8 ans maximum pour écouler ce qui a été doté l'année considérée.

Le montant réellement servi aux assurés sera le montant minimal entre la cible et le montant atteint après les différents leviers, si besoin, soustrait des chargements.

Cet algorithme est schématisé en Figure 14 :

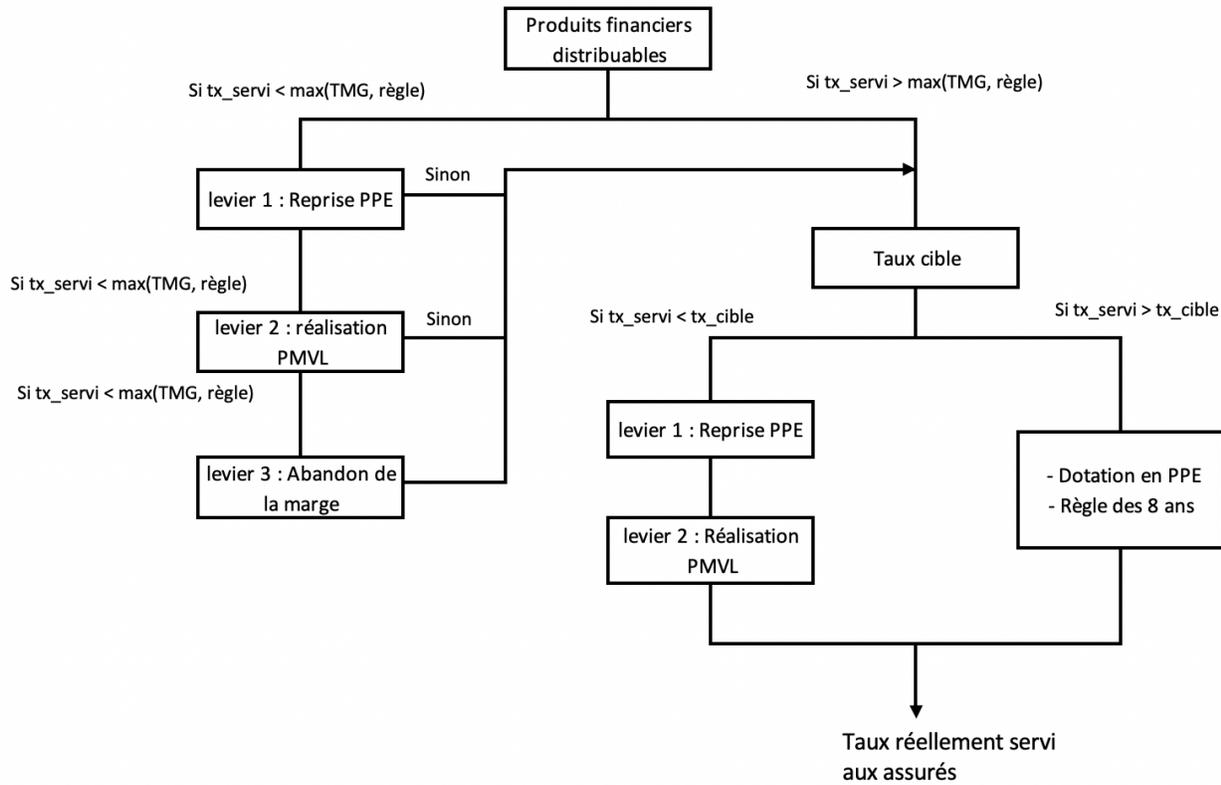


FIGURE 14 – Résumé de l'algorithme de PB

3.2 La partie prévoyance santé

3.2.1 Définition

La FDB en prévoyance-santé se compose de participation aux résultats et de l'absorption de provision d'égalisation. Il n'y a pas d'algorithme de distribution, et elle est moins dépendante des scénarios économique. L'utilisation d'un modèle stochastique pour déterminer le BEL n'est pas forcément nécessaire, cependant il est quand même incorporé dans notre modèle.

3.2.2 La participation aux résultats

Le mécanisme de participation aux bénéfices est complètement différent de celui en épargne-retraite. Il n'y a pas de montant minimum obligatoire mentionné dans la réglementation. La PB est également défini contractuellement, et approvisionnée identiquement à l'épargne-retraite, c'est-à-dire par les résultats techniques et financiers.

La provision principale associée à cette participation est la Provision d'Égalisation (PE). Elle ne peut être utilisée uniquement dans ce cadre, afin de faire face aux fluctuations des sinistralités rattachées aux risques concernés. Cette provision à l'avantage de pouvoir bénéficier de déduction fiscale, sous certaines conditions à respecter. La PE est individuelle à chaque contrat, autrement dit, un contrat possède sa propre PE et ne peut pas être utilisé pour autre chose.

Cette provision d'égalisation est gérée en amont du modèle ALM et va faire partie des flux déterministes. Pour chaque pas de projection, on va lui ajouter les intérêts qu'elle aura générés. Puis, si le résultat technique est positif, on va lui ajouter un pourcentage de celui-ci selon un taux défini préalablement dans le contrat.

Il faut par ailleurs tenir compte d'une limite de dotation qui est un pourcentage des cotisations acquises l'année de l'exercice. Si au contraire, il est négatif, on lui prélèvera le montant du résultat technique.

En résumé cela donne (Figure 15) :

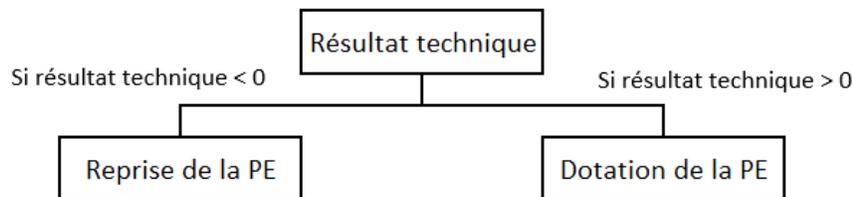


FIGURE 15 – Gestion de la provision d'égalisation

La détermination du montant de la participation va ensuite se faire dans le modèle ALM. Un pourcentage des produits financiers de l'année va être attribué à la prévoyance-santé. Le montant de la participation sera alors :

$$PB(t) = PM(t) (Tx_{PE} \times TRC(t-1) - Tx_{IT} - Tx_{revalo}(t))^+ + PE_{NET}(t-1) \times Tx_{PE} \times (TRC(t-1))^+ \quad (16)$$

Où :

- $P(t)t$: Provisions Mathématiques de l'année t ;
- $Tx_i(t)$ le taux de la variable i au temps t ;
- PE_{NET} est le montant net de la provision d'égalisation ;
- $(TRC(t-1))^+$ est le maximum entre la valeur du TRC de l'année t-1 et 0.

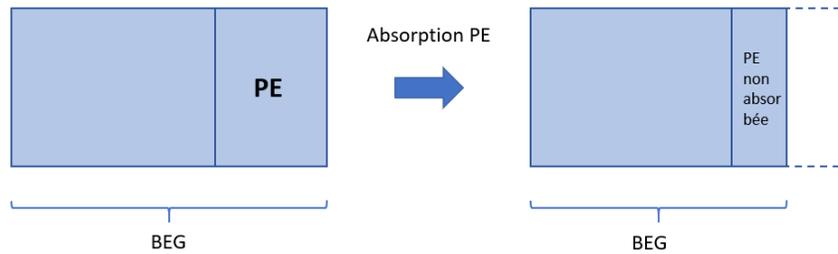
Si après l'attribution des PB, le montant des produits financiers n'est pas totalement consommé, le surplus est placé en fonds propres.

3.2.3 L'absorption par la PE

La FDB résultant de l'absorption par la PE s'obtient à posteriori du modèle ALM. Un premier SCR va être calculé sans prendre en compte l'absorption. Puis pour chaque compte de PE, une partie du SCR sera absorbée selon le montant de PE disponible. On aura un SCR après absorption :

$$SCR_{après\ absorption} = \sum_i \max(SCR_i - PE_i; 0) \quad (17)$$

Cette absorption va atténuer les chocs, en baissant les engagements. En effet, la PE est intégrée dans le BEG.



Cette méthode impactera seulement le montant de SCR net et l'ajustement. Les BEL en scénarios central ne changeront pas, mais ceux en scénario choqué seront impactés.

Cette PE absorbée est considérée comme de la FDB.

$$FDB_{PE} = SCR_{avant\ absorption} - SCR_{après\ absorption} \quad (18)$$

Cette partie de FDB est déterministe, on ne l'intégrera donc pas dans le proxy.

3.3 Comparaison des variations

Il est intéressant de voir comment varie les différentes FDB selon un scénario. Comme la FDB_{ERE} est celle qui requiert l'implémentation d'un modèle stochastique, et qu'elle est obtenue à partir de plusieurs variables dépendantes des hypothèses économiques, on s'attend à ce qu'elle varie plus que la FDB_{PS} , qui a seulement le TRC qui est stochastique.

On va donc comparer la variation des FDB, pour quatre scénarios différents par rapport au scénario déterministe central. Pour cela, on va calculer la part qu'elles prennent dans l'écart entre la FDB déterministe et les scénarios. Ainsi on aura, en pourcentage :

$$part\ ecart_{ERE} = \frac{|FDB\ centrale_{ERE} - FDB\ autre_{ERE}|}{|FDB\ centrale_{totale} - FDB\ autre_{totale}|} \times 100$$

Les quatre scénarios sont choisis en prenant : une FDB élevée et faible, un BEL élevé et faible. On obtient (Table 2) :

FDB	FDB min	FDB max	BEL min	BEL max
ERE	84%	92%	91%	81%
PS	16%	8%	9%	19%

TABLE 2 – Part des écarts avec le scénario central

Comme attendu, c'est la partie épargne-retraite qui fait varier en grande partie notre variable de recherche. Dans tous les cas, elle est responsable d'au moins 81% de l'écart de la FDB par rapport au scénario central. Il faudra déterminer un proxy différent pour chaque FDB, car leur fonctionnement sont distincts et indépendant. Le proxy de l'ERE sera plus complexe à mettre en place, car elle comporte différents facteurs stochastiques.

En effet, pour la partie prévoyance-santé, dans notre modèle, seul le TRC est stochastique. Dans certain modèle, on utilise même une projection de TRC unique, qui ne rentre pas en compte dans le modèle stochastique. Dans notre cas, on utilisera les projections de TRC de l'épargne-retraite, pour la prévoyance-santé.

Pour cette partie épargne-retraite, en plus de ce TRC, on aura d'autres variables qui dépendront de l'univers économique. C'est pourquoi on devra utiliser plusieurs modèles pour construire la FDB. Ces variables seront déterminés par la suite, et un modèle sera mis en place pour les estimer.

4 Mise en place du Proxy

4.1 Recherche des variables explicatives

4.1.1 Analyse sur deux scénarios extrêmes

En vue de comprendre comment évolue la FDB en fonction des variables dépendantes des conjectures économiques, on va analyser les variations des flux d'un scénario central avec un scénario extrême positif et un scénario extrême négatif.

Le choix de ces scénarios se fait à partir des montants de la FDB totale. Pour le scénario extrême positif, on va prendre celui qui donne une FDB nette maximale, et minimale pour le scénario négatif. Ainsi on aura les montants suivants (Table 3) :

FDB \ Scénario	Négatif	Central	Positif
ERE	31 600 400	433 196 945	2 494 290 032
PS	89 300 460	168 200 509	350 504 890
Total	120 900 860	601 397 454	2 844 794 922

TABLE 3 – FDB des scénarios choisis

En accord avec la section précédente, c'est le secteur épargne-retraite qui est le plus impacté par l'environnement économique. On peut faire le constat suivant : bien que les deux parties soient calculées distinctement, elles fluctuent selon une tendance identique. Par exemple, une baisse de la FDB totale provient d'une baisse des deux FDB. Ce qui semble logique puisque pour la partie prévoyance-santé, elle est en fonction du TRC, et pour l'épargne-retraite elle dépend en grande partie des produits financiers, dont le TRC est un indicateur. De plus, elles se basent sur le même TRC. Cependant, un petit nombre de scénarios montre un cas où l'une baisse et l'autre augmente. On étudiera cette éventualité ultérieurement.

La prévoyance-santé

Seul le TRC varie en fonction du scénario. On peut regarder la tendance des flux de FDB_{PS} par rapport à celle des TRC en Figure 16.

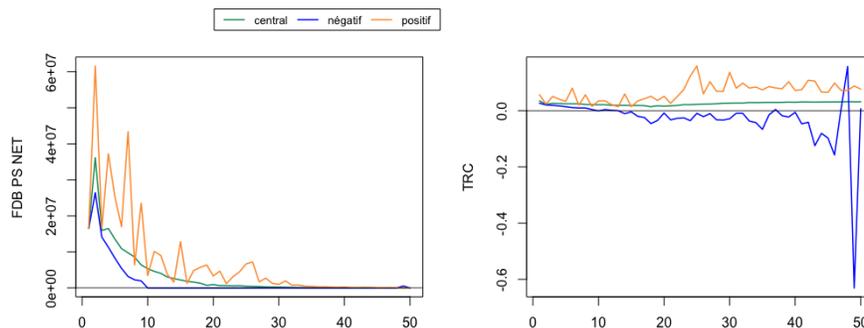


FIGURE 16 – Variations des FDB et des TRC

Les courbes se positionnent dans le même ordre pour les deux variables. Celles du scénario négatif sont les plus faibles, et celles pour le scénario positif sont les plus hautes, tandis que les centrales se retrouvent en général entre les deux autres. Il ne faut pas oublier que le TRC utilisé pour le calcul de la $FDB_{PS}(t)$ est celui du pas $t - 1$. Il y a une année de décalage, que l'on retrouve bien dans les graphes.

De manière générale, le stock global de PE diminue au cours du temps. on aura donc de moins en moins de participation qui se traduit par des flux de FDB décroissants. Plus la FDB diminue, et plus l'influence du TRC s'amenuise.

Les tendances concordent entre les deux graphes, avec un scénario positif en dents de scie, ainsi que ses entrecroisements avec le central. Lorsque le TRC devient négatif, on a une FDB nulle.

L'épargne-retraite

On va observer les principales variables qui déterminent la FDB_{ERE} . On sélectionnera donc, en plus du TRC ci-dessus, les intérêts techniques, le taux cible, la PPE et les chargements sur encours, comme illustré en Figure 17.

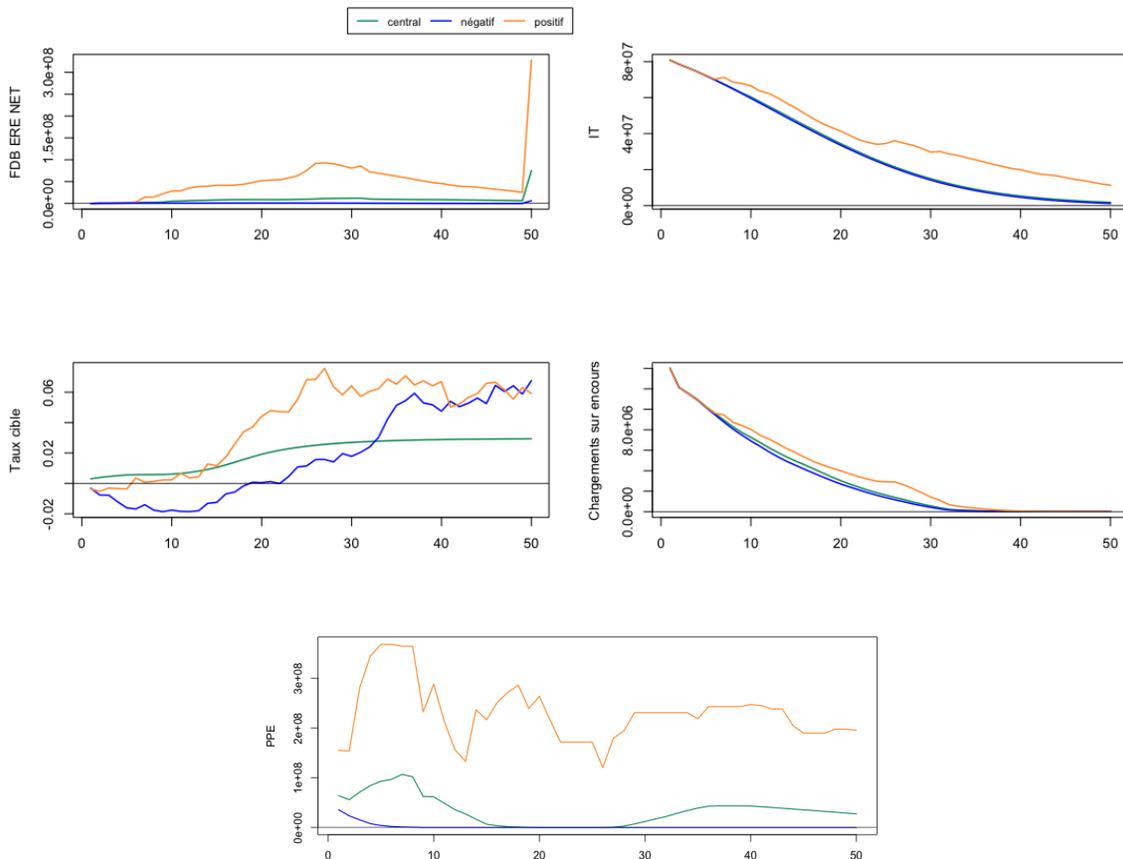


FIGURE 17 – Comparaisons des indicateurs

Tout d'abord, on remarque que la FDB est plus impactée par le scénario choisi. Les flux pour le scénario négatif sont extrêmement faibles par rapport à ceux du positif. En outre, les courbes des intérêts techniques et chargements sur encours sont relativement similaires. En effet, ils sont tous deux en proportion des provisions mathématiques. Leur variation par rapport aux scénarios reste assez faible compte tenu des autres variables, surtout si l'on se focalise sur le central et le négatif. Il est intéressant de regarder le taux cible. On remarque plusieurs enchevêtrements des trois courbes. Notamment celui du scénario positif en dessous du central en début de projection, puis de la montée du négatif en fin de projection dont la courbe rejoint celle du scénario positif. En ne se basant que sur cette variable, on pourrait penser que les trois FDB sont proches. Cependant la capacité à honorer ce taux cible est très opposée. On peut voir ça avec la PPE.

scénario positif : le taux cible est négatif en début de projection, on va donc placer tous les produits financiers dont on dispose dans la PPE. Puis quand ce taux augmente fortement, avec les produits financiers générés l'année exercée ainsi que le grand stock de PPE créé au début, on va pouvoir respecter le taux cible fixé. C'est la raison pour laquelle la hausse du taux cible est suivie de la hausse de la FDB_{ERE} .

scénario négatif : il est possible de se constituer un faible stock de PPE. Cependant, la période durant laquelle les taux cibles sont négatifs est trop longue pour conserver cette provision. De plus, le TRC devient assez rapidement négatif, ce qui va entraîner une insuffisance des produits financiers pour ne serait-ce qu'honorer ce qui est garanti. C'est pourquoi la FDB est aussi faible.

scénario central : le taux cible n'est jamais négatif. Il est faible au départ, ce qui donne l'occasion de placer en PPE. La hausse de ce taux qui ne s'accompagne pas d'une hausse de TRC vient épuiser le stock.

Variations opposées

On va maintenant étudier les cas où les variations des deux FDB sont contraires. Ils représentent environ 11% des scénarios. Les trois cas étudiés seront donc : le scénario central, le cas où la FDB_{ERE} augmente (ERE +) et la FDB_{PS} diminue (PS -) et le cas contraire.

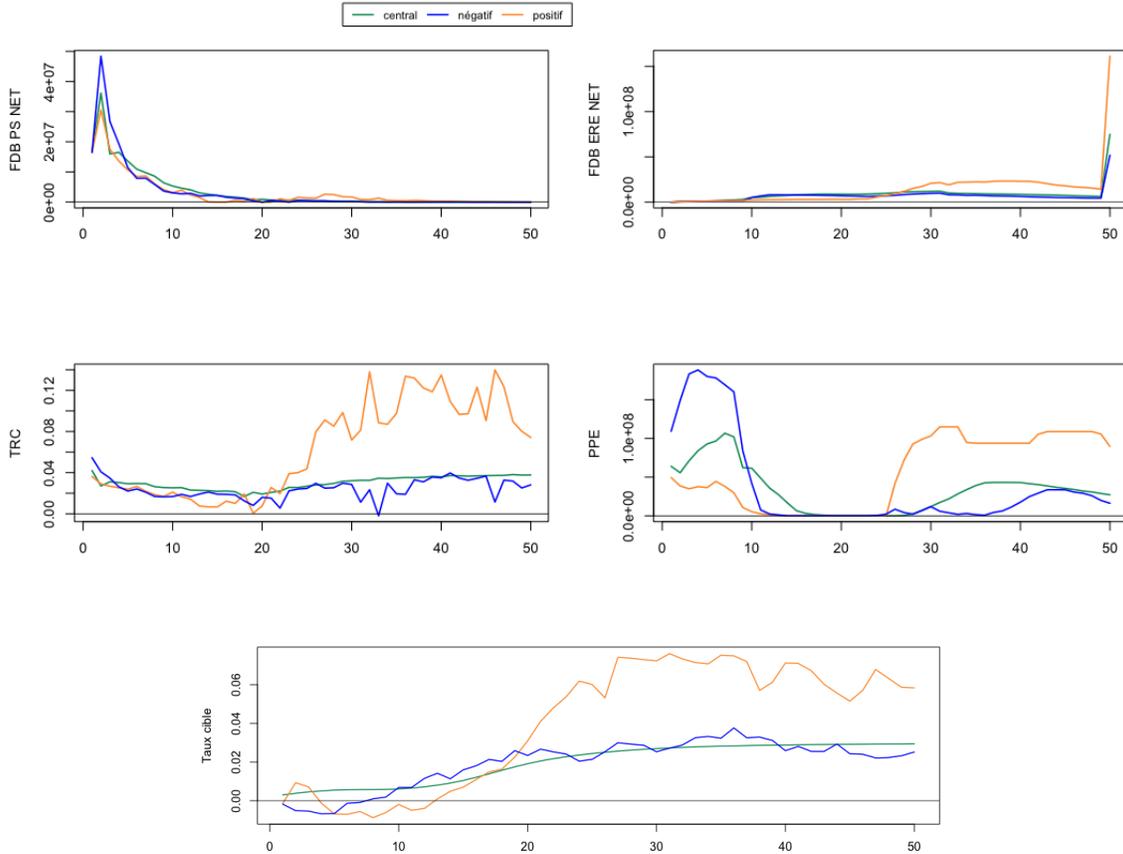


FIGURE 18 – Comparaisons des indicateurs pour des variations inverses

Cas ERE + et PS - : Le TRC ne commence à être nettement supérieur que vers la 20-ième année. Ce qui a épuisé la provision d'égalisation. C'est pourquoi la FDB_{PS} diminue plus fortement. Alors que pour l'ERE, la hausse du TRC permet de reconstituer la PPE et de pouvoir servir le taux cible fixé.

Cas ERE - et PS + : Le TRC du départ est plus élevé, et permet de ne pas puiser dans la PE directement et de servir plus de participation. Cependant elle reste très vite inférieure au scénario central tout au long de la projection, ce qui va écouler la PPE plus rapidement.

Pour conclure, la chronique des TRC est réellement importante car elle va définir en grande partie la FDB des deux parties. Cependant, elle aura une importance particulière pour la partie ERE, car elle va permettre d'alimenter ou non la PPE. Cette dernière ayant tout un algorithme de gestion qui servira à avoir une FDB_{ERE} plus élevée.

4.1.2 La base de données

Notre base de données se compose des 2000 scénarios antithétiques générés à la sortie du modèle ALM. Comme on l'a vu antérieurement, les variables primordiales pour évaluer la FDB sont : les produits financiers, les intérêts techniques, les chargements sur encours et le taux cible.

Les produits financiers nets seront représentés par le TRC. Les intérêts techniques ainsi que les chargements sur encours sont tous les deux un pourcentage des PM. Ces PM seront déterminées par l'intermédiaire des coefficients de flexing et des PM comptables calculées dans le modèle passif. La distribution de la FDB dépendra également du taux cible. Enfin, les taux ZC qui déterminent la courbe d'actualisation sont également importants pour notre modèle. Ainsi les variables qui nous serviront à établir notre proxy seront :

- les TRC ;
- les coefficients de flexing ;
- les taux cibles ;
- les taux ZC.

Ces variables seront à prendre sur les 50 années de projections.

Cependant notre proxy à pour but de ne pas utiliser le modèle ALM, car il prend beaucoup de temps. En outre, il faudra élaborer des modèles pour déterminer les TRC et les coefficients de flexing pour les utiliser dans le modèle final. Tous les modèles se construiront sur des données qui s'obtiennent en dehors de l'ALM, c'est à dire des GSE (Générateur de Scénario Economique).

Pour construire le modèle des TRC, on utilisera la courbe des taux ainsi que les variables qui influencent les actifs, qui semblent être les indices action et immobilier ainsi que l'inflation. Pour les coefficients de flexing, on utilisera les TRC ainsi que les taux ZC. Enfin le modèle final utilisera les données sorties des précédents modèles ainsi que la courbe des taux.

Notre modèle se construira donc comme ceci :

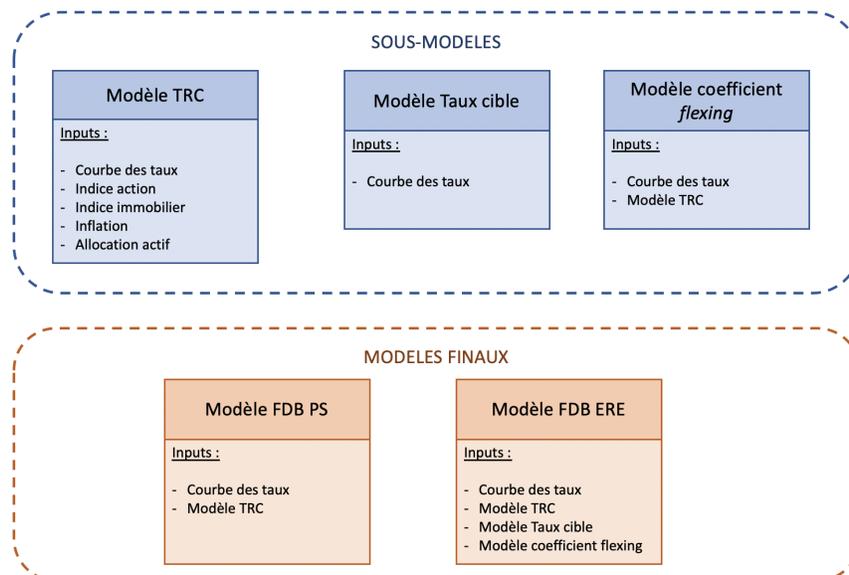


FIGURE 19 – Structure du proxy

Pour chaque modèle, il faudra diviser nos données en deux parties : train et test. La répartition sera 80% de train et 20% de test. Ainsi on aura 1600 lignes pour paramétrer nos modèles et les tester sur 400 prédictions.

4.2 Implémentation d'un réseau de neurones

4.2.1 Principe

Un réseau de neurones est une méthode d'apprentissage en profondeur (*Deep Learning*). Son fonctionnement imite celui d'un neurone du cerveau humain. Ils sont souvent utilisés pour la reconnaissance faciale, les prédictions de flux, ou encore des traitements de langage.

Un réseau de neurones se compose de plusieurs neurones organisés par couches. On retrouve trois sortes de couches (Figure 20) :

- Une couche d'entrée qui est la première et qui va recevoir les vecteurs de données initiales.
- Une ou plusieurs couches cachées, ce sont toutes les couches intermédiaires à celles d'entrées et de sorties.
- Une couche de sortie qui va produire les résultats recherchés.

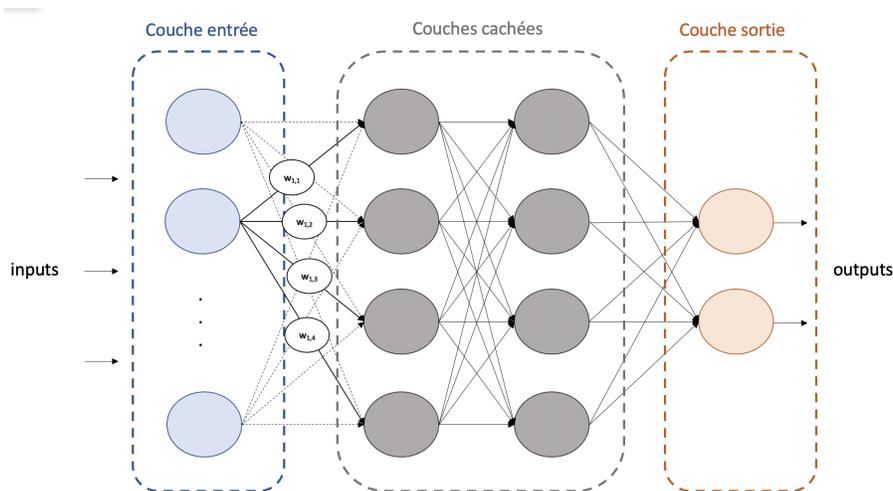


FIGURE 20 – Un réseau de neurones

Chaque neurone qui se connecte à un autre va lui être associé un poids ($w_{i,j}$) pour déterminer l'importance de chaque variable. Les poids vont s'ajuster au fur et à mesure de l'apprentissage du modèle. À chaque couche est attribuée une fonction d'activation. Elle permet de lisser et de normaliser les données d'entrée pour ensuite les transmettre à la couche suivante.

Ainsi, un neurone d'une couche va recevoir les données de la couche précédente, qui vont être sommées en pondération de leurs poids, puis la fonction d'activation va être appliquée pour renvoyer les données à la couche suivante.

Les poids vont s'adapter afin de minimiser l'erreur entre les données prédites du modèle et celles réellement attendues. Pour cela on utilise la méthode de descente de gradient. Elle consiste à choisir des poids aléatoires et suivre l'inverse du gradient, multiplié par le taux d'apprentissage, pour chaque itération. Cet algorithme est appelé **rétropropagation du gradient**.

Il existe trois grandes catégories de réseaux de neurones :

- Les réseaux de neurones *feed forwarded*
- Les réseaux de neurones récurrents (RNN)
- Les réseaux de neurones convolutifs

Le réseau de neurones qui s'adapte le mieux à notre sujet et que nous allons utiliser est de type récurrent, car il est parfaitement adaptée pour les données séquentielles.

Réseaux de neurones récurrents

Un réseau de neurones récurrents se caractérise par son traitement de l'information en cycle. Les sorties d'une couche vont être réinjectées dans celle-ci ou dans des couches précédentes. Cela permet de prendre en compte le traitement d'une même information en le mémorisant. Ces couches récurrentes sont souvent combinées avec des couches classiques. Une couche classique étant une couche de neurones sans propriétés particulières, c'est-à-dire que chaque neurones va uniquement appliquer sa fonction d'activation puis repasser l'information aux suivants. Schématiquement, cela se représenterait comme le montre la Figure 21 :

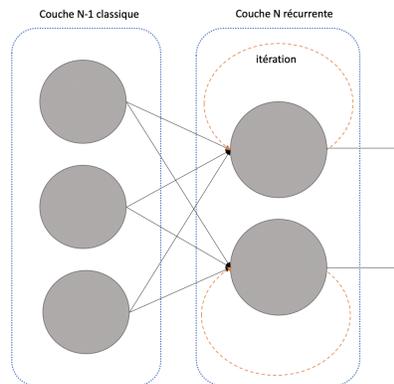


FIGURE 21 – Exemple d'une couche récurrente

La fonction d'activation d'une cellule récurrente est la tangente hyperbolique (\tanh) défini par :

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (19)$$

Néanmoins, dans un réseau de neurones récurrents classique, la mémoire reste à court-terme. Cela signifie que si on lui impose un trop grand nombre d'itérations ou trop complexes, il commence à oublier ce qu'il s'est passé avant. La performance du réseau baisse et le gradient va alors se dissiper ou bien augmenter de manière exponentielle. Pour pallier à ce problème, des réseaux de neurones spécifiques ont été créés en élargissant le concept du RNN, qui sont le *Long Short-Term Memory* (LSTM) et le *Gate Recurrent Unit* (GRU).

L'architecture d'un LSTM est identique à un RNN. C'est la gestion de l'état caché qui va être revue. La récurrence sera assurée par une cellule à mémoire. Le noeud central contient la

mémoire interne de la cellule. L'idée est de diviser ce qui est important à court terme, et ce qui est important à long terme. Elle est composée de trois portes : une porte d'entrée, une porte d'oubli et une porte de sortie. Ces portes vont permettre de réguler la mémoire et la production des sorties.

Une couche LSTM est beaucoup plus « lourde » qu'une couche récurrente simple. Elle prend généralement plus de temps d'exécution. Une cellule GRU a la même structure qu'une LSTM avec la porte d'oubli en moins, ce qui réduit le nombre de paramètres et donc le temps d'exécution. Leurs performances sont en général presque similaires.

Une étude menée par *Shudong YANG, Ying ZHOU et Xueying YU* en 2020 sur la reconnaissance de commentaire positif ou négatif, a montré que dans certains cas le GRU générait de meilleures performances que le LSTM et inversement pour les autres cas. Une autre étude, faite par *Gumaei A., Abd-el Ghany S., Ali A., Al-Rakhami M., Omran N. et Saleh H.* en 2021, sur la prédiction de séries temporelles pour faire des prévisions sur le nombre de décès du Covid-19 montrait que selon le pays, le LSTM ou le GRU obtenait de meilleurs résultats. Cela dépend des cas et des données que l'on étudie. Ainsi on comparera les deux modèles pour nos données.

4.2.2 Construction des modèles

Pour implémenter un réseau de neurones récurrents, on va utiliser la librairie *Keras* de la bibliothèque *Tensorflow* sur Python. Le premier modèle à construire est celui des TRC, car il va alimenter les autres.

Tous nos modèles comporteront une première couche récurrente et une dernière couche dense, qui est une couche neuronale classique. Le nombre de couches additionnelles sera à tester afin de voir quels sont les meilleurs résultats. Le nombre de neurones dans une couche, la fonction d'activation et le taux d'apprentissage seront des hyperparamètres à *tuner*.

Tuning des hyperparamètres

Un hyperparamètre est un paramètre qui est défini avant le processus d'apprentissage et va influencer sur sa performance, sa vitesse et sa qualité. Ainsi le choix de ses hyperparamètres est important. Trouver les hyperparamètres qui obtiennent le meilleur modèle est compliqué manuellement. Il faut tester de nombreuses combinaisons afin de trouver celle optimale. On utilisera l'indicateur de l'erreur quadratique moyenne (MSE) pour choisir le meilleur modèle, qui sera celui qui a la plus faible MSE.

Le réglage de ces hyperparamètres est appelé *tuning*. Pour automatiser leurs recherches, on va utiliser la librairie *Keras-Tuner*. On va alors choisir ceux que l'on va régler et sur quelles valeurs les tester.

- Le nombre de neurones dans une couche : Pour chaque couche, on va tester entre trois et quatre valeurs. Ces valeurs sont choisies selon la taille des inputs de départ. On testera entre 100, 200, 300 et 400 neurones par couches.
- Le *Dropout* : il évite l'*overfitting* (ou sur-apprentissage) qui est un problème que l'on

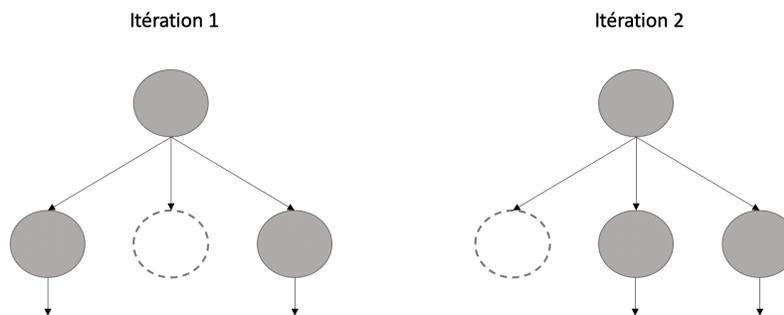
expliquera juste après. Nous allons le tester sur trois taux : 0.1, 0.2 et 0.5.

- La fonction d'activation : définie pour chaque couche, afin d'introduire de la complexité non-linéaire au modèle. Trois fonctions d'activation seront testées, en général ce seront presque toujours les mêmes qui seront dans la liste de choix, selon les résultats de petits tests préalables, à savoir : 'Selu', 'Relu' et 'Linear'.
- Le taux d'apprentissage : indique la vitesse à laquelle les coefficients, pour la descente de gradient, évoluent, que l'on testera pour 0.01, 0.001 ou 0.0001.

L'Overfitting

Ce problème survient lorsque le modèle essaye de trop s'adapter aux données d'apprentissages, et conduit à un manque de généralisation. La marge d'erreur est alors augmentée. Pour surmonter ce problème on utilise des méthodes de régularisation.

Celle que nous utilisons est la méthode du *Dropout*. Cela consiste à « supprimer » temporairement des neurones d'une couche. La désactivation d'un neurone est aléatoire selon une probabilité qui sera notre hyperparamètre recherché. Ceci est illustré dans la figure ci-dessous :



Ainsi à chaque itération, il y aura une configuration de neurones différente, ce qui viendra perturber les caractéristiques apprises par le modèle pour exploiter chaque neurone individuellement.

La fonction d'activation

Les fonctions d'activation testées seront donc :

La fonction d'*Unité Linéaire Exponentielle mise à l'Echelle* (SELU) définie par :

$$f(x) = \begin{cases} \xi \times x & \text{si } x > 0 \\ \xi \times \alpha \times (e^x - 1) & \text{si } x < 0 \end{cases} \quad (20)$$

Avec ξ et α des constantes prédéfinies. Elle évite les problèmes d'explosion et de disparition de gradients. L'explosion de gradients est lorsque l'augmentation des valeurs des gradients est trop rapide, et dépasse la capacité de représentation des nombres, arrêtant ainsi l'apprentissage. À l'inverse, la disparition de gradient consiste à une diminution accélérée des gradients, ce qui entraîne l'annulation du gradient et par conséquent l'arrêt de l'apprentissage.

La fonction d'Unité Linéaire Rectifiée (RELU), qui est :

$$f(x) = \max(0, x) \quad (21)$$

Elle a un temps de calcul rapide de par sa simplicité.

La fonction *linéaire* qui ne va pas modifier les données et sera simplement : $f(x) = x$

Le taux d'apprentissage

Les taux d'apprentissages sont choisis en fonction de ce qui est souvent utilisé. Ces taux sont associés à la méthode d'optimisation « *Adam* », qui est la plus utilisée, grâce à son efficacité et simplicité et permet d'effectuer la descente de gradient. Le taux d'apprentissage va permettre de déterminer la vitesse à laquelle les coefficients de la descente du gradient évoluent. Si ce taux est trop faible, alors le processus sera plus long, pouvant aller jusqu'à se bloquer. Tandis que s'il est trop élevé, le processus d'apprentissage sera instable.

Avec la méthode choisie, ce taux s'adapte aux fils du temps. Pour reprendre les termes déjà utilisés précédemment, le taux d'apprentissage va contrôler l'ajustement des poids entre chaque neurones.

Les hyperparamètres non *tunés*

Certains hyperparamètres ne font pas partie de nos réglages automatiques, soit parce qu'ils s'avèrent être trop compliqués à mettre en place dans notre fonction soit parce qu'il n'est pas vraiment utile de les mettre.

L'*epoch* correspond au nombre d'itération pour la descente de gradient, autrement dit le nombre de fois où l'on va entraîner notre réseau. Un nombre d'*epoch* trop grand amène à l'*overfitting*, alors qu'un nombre trop faible amène à un sous-apprentissage avec un modèle insuffisamment entraîné.

Pour ce qui est du nombre de couches, qui n'est pas vraiment un hyperparamètre, ceci a été déterminé manuellement à partir de plusieurs tests. Pour chaque modèle, on part d'une couche d'entrée et d'une couche de sortie. Puis on *tune* les hyperparamètres et on regarde les résultats. Puis on rajoute une autre couche et on recommence. On compare les différents résultats obtenus pour plusieurs couches et on prend celui optimal. Il est inutile de tester sur un trop grand nombre de couches, compte tenu de la taille de nos données.

Le modèle TRC

Après la détermination des hyperparamètres, le modèle TRC sera composé de trois couches dont les deux premières possèdent un *Dropout*. La première couche sera forcément des cellules LSTM ou GRU, puis les suivantes seront des couches normales *Dense*.

	LSTM	GRU
Couche 1	400	200
<i>Dropout</i> 1	0,5	0,1
Couche 2	100 , relu	80, linear
<i>Dropout</i> 2	0,3	0,5
Couche 3	50 , linear	50, relu
Taux apprentissage	0,001	0,01

Le modèle Coefficient de flexing

Ce modèle se compose de trois couches avec un *Dropout* sur la première couche récurrente.

	LSTM	GRU
Couche 1	100	100
<i>Dropout</i> 1	0,5	0,3
Couche 2	150 , selu	150, selu
Couche 3	50 , selu	50, relu
Taux apprentissage	0,001	0,001

Le modèle Taux cible

Le taux cible aura juste une couche entrée et une couche sortie avec un *Dropout* sur la couche d'entrée.

	LSTM	GRU
Couche 1	300	300
<i>Dropout</i>	0,1	0,3
Couche 2	50 , selu	50, relu
Taux apprentissage	0,001	0,001

Le modèle FDB PS

Il n'y a pas besoin d'un réseau de neurones car on a une formule déjà définie en section 3.2.2. On reprend cette formule (16) en utilisant les TRC sortis du modèle précédent. Puis on fait la moyenne par *Monte Carlo* pour avoir notre FDB de la partie prévoyance-Santé.

Le modèle FDB ERE

Le modèle final possède quatre couches dont les trois premières ont été affectées d'un *Dropout*. Ainsi la structure est la suivante :

	LSTM	GRU
Couche 1	400	300
Dropout 1	0,3	0,3
Couche 2	150 , selu	200, relu
Dropout 2	0,5	0,5
Couche 3	100 , relu	200, relu
Dropout 3	0,2	0,2
Couche 4	1 , selu	1, linear
Taux apprentissage	0,001	0,001

Les taux d'apprentissage sont optimaux à 0,001 pour presque tous les modèles, excepté pour celui du TRC pour le réseau de neurones du type GRU.

4.2.3 Résultats

Les métriques utilisées pour évaluer les modèles sont :

- **La Racine de l'Erreur Quadratique Moyenne** (RMSE), qui permet de quantifier les erreurs d'un modèle.

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{y}_i - y_i)^2}{n}}$$

Elle se lit dans la même unité que la variable prédite, et peut être lue en pourcentage en la normalisant.

- **L'Erreur Moyenne Absolue en Pourcentage** (MAPE), qui utilise les erreurs relatives :

$$MAPE = \frac{1}{n} \sum_i^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Cela nous donne une précision sur les prédictions.

- Le **Biais**, qui permet de savoir en plus si le modèle à tendance à surestimer ou sous-estimer les prédictions.

$$Biais = \frac{\sum_i^n \hat{y}_i - y_i}{n}$$

Pour les sous-modèles, les indicateurs se calculent sur la globalité des vecteurs pour chaque prédictions. On cherche à minimiser le plus possible ces indicateurs. On obtient comme métriques optimales (Table 4) :

Modèles	Layer	RMSE	MAPE	Biais
Modèle TRC	LSTM	1,8%	1,1%	0,18%
	GRU	1,9%	1,3%	0,8%
Modèle CF	LSTM	1,8	0,4%	-0,11
	GRU	1,9	0,9%	-0,18
Modèle Taux cible	LSTM	0,7%	1,1%	-0,03%
	GRU	1%	1,1%	0,04%
Modèle Final FDB	LSTM	41 205 112	3,6%	-6 032 354
	GRU	61 193 612	5,4%	-63 597 521

TABLE 4 – Résultats des métriques

On peut observer que dans chacun des modèles, ce sont les LSTM qui ont les meilleurs résultats. On va donc opter pour ces modèles pour la suite. De ce fait, on constate que l'on a tendance à sous-estimer les coefficients de flexing, et donc les provisions mathématiques, les taux cibles ainsi que la FDB, et à surestimer les TRC. Pour le modèle final, le RMSE paraît grand, mais compte tenu de la variabilité de la FDB il reste acceptable.

On peut regarder, dans la Figure 22, la moyenne des prédictions par rapport à la moyenne cible pour chaque modèle :

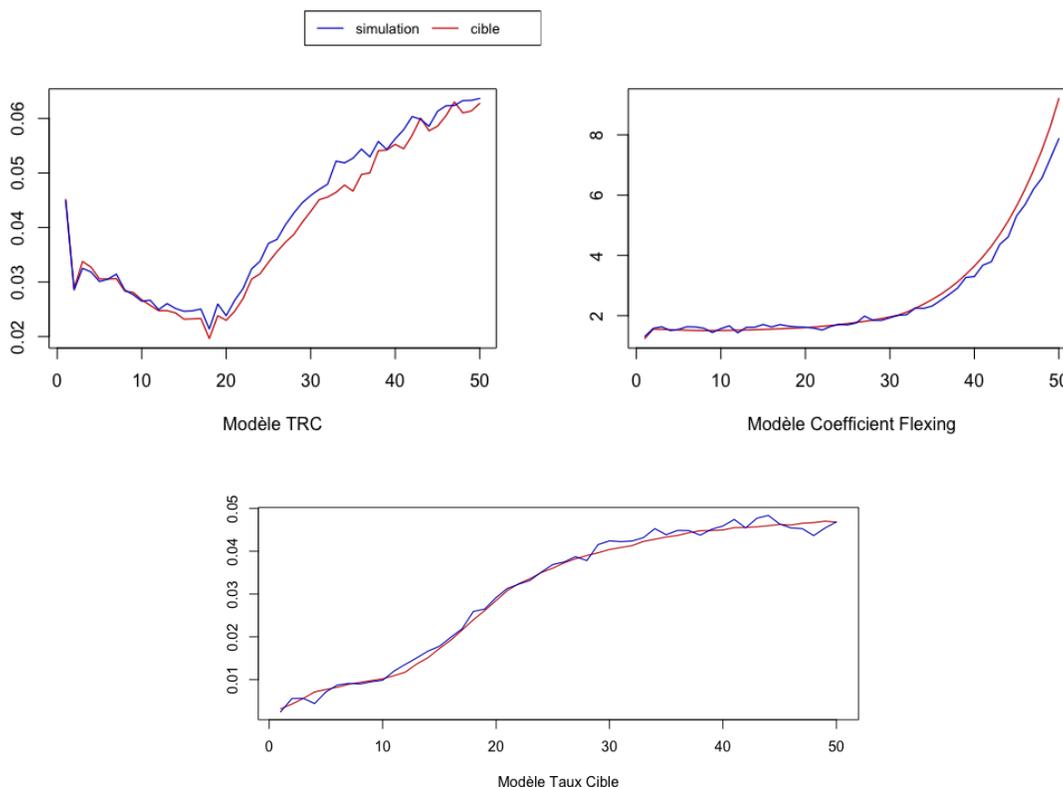


FIGURE 22 – Résultats des prédictions tests

Ces résultats semblent convenables. Si on prend une simulation précise, il peut y avoir des

écarts plus importants, alors qu'en prenant la moyenne de nos scénarios tests simulés à partir de nos modèles, on se rapproche fortement de la cible. Notre FDB totale sera, par la méthode Monte-Carlo, la moyenne de nos simulations, on peut ainsi espérer retrouver de bons résultats pour notre modèle final.

On s'accordera une erreur de 5% par rapport à la valeur cible pour le résultat de FDB. Ainsi, à partir de nos données tests on obtient :

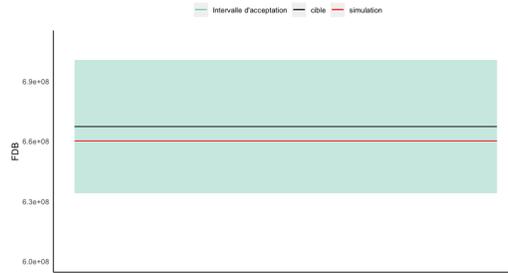


FIGURE 23 – Résultat de la FDB ERE sur les données tests

Ce résultat est correct, on peut donc appliquer tous nos modèles à la suite, en gardant la même marge d'erreur, pour parvenir aux résultats suivants :

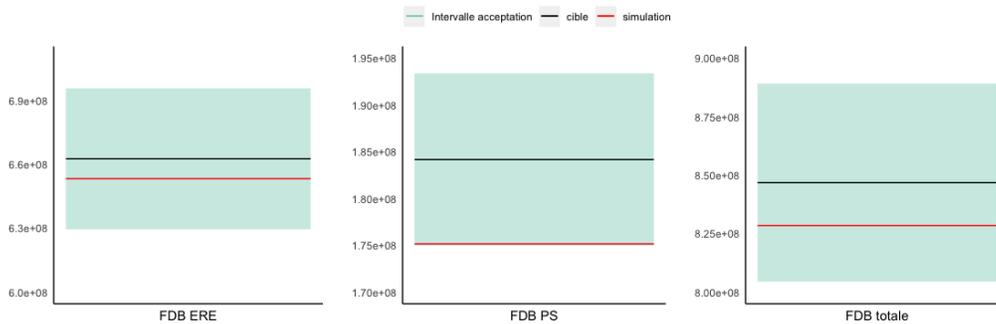


FIGURE 24 – Résultats finaux du modèle

La FDB totale est assez éloignée de la cible, mais elle se retrouve tout de même dans l'intervalle d'acceptation que l'on s'est fixé. Ainsi notre modèle, pour l'entité considérée, estime dans une marge d'erreur inférieure à 5% la FDB totale.

4.3 Interprétations

Le *Deep Learning*, et par conséquent les réseaux de neurones, sont des modèles dit « *Black-Box* ». Cela signifie que les traitements qui ont lieu entre l'entrée des données et la sortie des résultats ne sont pas totalement observables ou bien trop complexes pour être compréhensibles. Or, en assurance, lorsque l'on construit un modèle, il vient un moment où se pose la question de savoir s'il est éthique ou non. C'est à dire, si notre modèle peut se révéler discriminatoire selon les variables que l'on a prit en compte. L'interprétabilité du modèle est alors important, surtout en Solvabilité II, où chaque modèle se doit de pouvoir être justifié à chaque instant, avec une réglementation précise. C'est pourquoi notre modèle ne peut être utilisé qu'à titre indicatif.

Des méthodes ont été mises en place afin de pouvoir interpréter les modèles *Black-Box*. On retrouve deux niveaux d'interprétations qui sont : Locale ou Globale.

4.3.1 Au niveau Locale

A cette échelle, on cherche à expliquer le modèle pour une seule prédiction. On analyse chaque réponse séparément. L'utilisation de l'interprétation locale peut se faire lorsque l'on cherche à analyser le comportement d'un assuré en particulier, par exemple.

Nous allons utiliser deux méthodes locales pour interpréter notre modèle. La première sera la méthode de *Lime* et la seconde sera celle de *Shap*.

Lime

La méthode de *Lime* (*Local Interpretable Model-agnostic Explanation*) a été introduite par *Ribeiro, Samer and Guestrin [2006]* et va utiliser ce qu'on appelle des « *Surrogate model* », qui sont des modèles simples, construit sur la base de prédictions issues de modèles complexes. Cette méthode est donc indépendante du modèle choisi et n'est pas spécifique aux *Deep Learning*, comme le sont d'autres méthodes.

Pour construire ces modèles simplifiés, *Lime* va prendre une observation, et en générer d'autres fictives proches d'elle, pondérées selon leurs proximités. Plus l'observation introduite sera voisine de l'originale, plus elle sera importante. Il va ensuite appliquer le modèle initial sur les observations de substitutions pour avoir leurs valeurs cibles.

Il va sélectionner les caractéristiques du modèle les plus décisifs, avec des techniques de sélections telles que *Lasso* ou encore *Forward*. Un modèle de régression sera ensuite construit à partir de ces caractéristiques restantes.

On aura :

$$\mathbb{E}(Y) = \beta_0 + \sum_j \beta_j X_j$$

Avec les bêtas qui seront les explications de *Lime*, valables uniquement pour le voisinage de Y.

Cette méthode est plus souvent expliquée, et plus facilement compréhensible, avec les schémas suivant (Figure 25) :

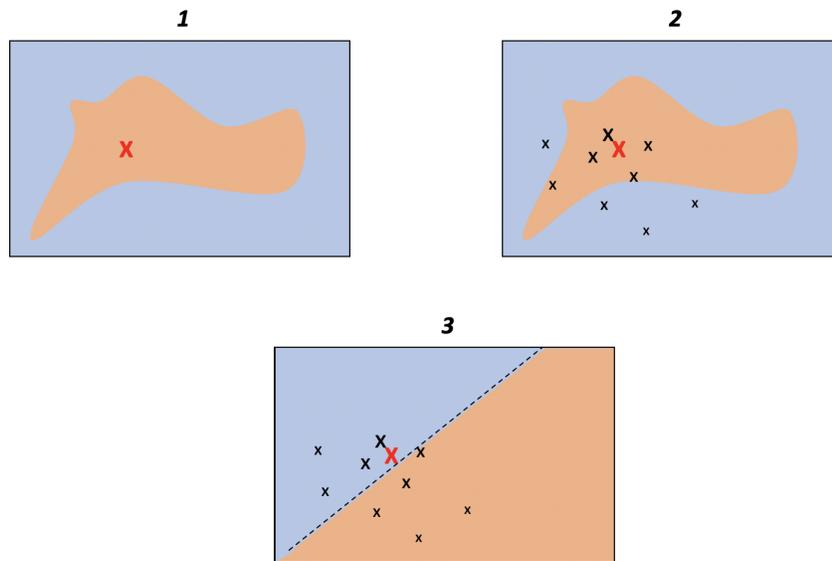


FIGURE 25 – Explication LIME

La croix rouge représente l'observation choisit dans le jeu de données. L'arrière plan correspond aux prédictions du modèle complexe que l'on cherche à expliquer. Dans cet exemple, la prédiction se fait sur une variable binaire, où le fond orange correspond à une valeur et le fond bleu à une autre. *Lime* va ensuite générer d'autres observations plus ou moins proches, représentées par les autres croix. Cela permettra de construire notre régression linéaire et d'avoir notre explication locale.

Pour notre utilisation de *Lime*, il va falloir choisir divers paramètres. Il faut tout d'abord, choisir le module de *Lime* qui convient à notre jeu de données. En effet, il existe trois objets explicatifs pour expliquer les modèles, qui sont : « *lime_tabular* », « *lime_text* », « *lime_image* ». Les deux derniers, comme leurs noms l'indiquent, sont utilisés pour les données textuelles et des images. Nous allons dans notre cas utiliser celui pour les données tabulaires.

Il faudra ensuite spécifier l'utilisation d'une régression pour expliquer les résultats, et non une classification. Enfin, la sélection des caractéristiques se fera par la technique « *auto* », qui consiste à prendre les poids les plus élevés par régression par crête mis à l'échelle. On prendra ici les dix caractéristiques les plus importantes, et on comparera les résultats sur deux observations.

Ainsi nous obtenons les résultats suivants :

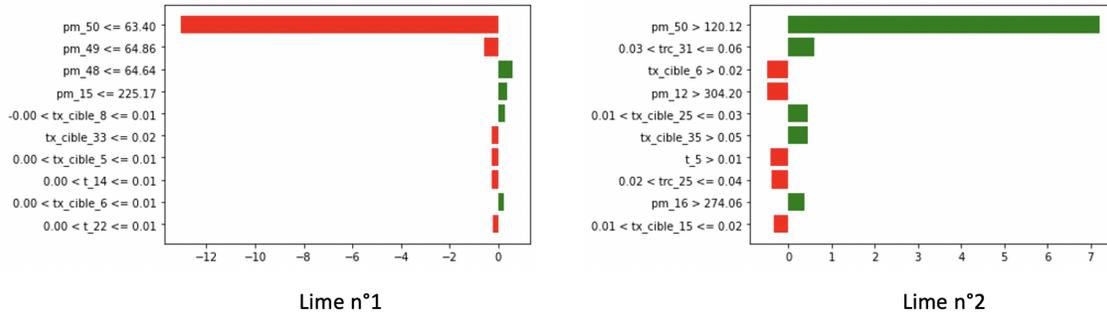


FIGURE 26 – Comparaison *Lime* sur deux observations

Les variables choisies par *Lime* pour expliquer le modèle ne sont pas totalement les mêmes pour les deux observations. On remarque cependant que la PM de la dernière année a une importance particulière dans nos deux cas. Elle a un poids décisif dans le modèle de prédiction de la FDB. Cela reste compréhensible car dans le calcul du *Best Estimate*, on ajoute les PM restantes à la fin de la projection. De plus, dans l'observation 1, elles sont assez faibles, ce qui entraîne un poids négatif dans le modèle. Tandis que dans l'observation 2, elles sont plus élevées et apportent un poids positif au modèle. En effet, plus il reste de PM à la dernière année, plus on peut en mettre dans le BEL, et donc la FDB de 50^{ième} année sera élevée, ce qui aura un fort impact sur le montant de notre valeur de recherche finale. Donc cela paraît assez cohérent que cette variable soit aussi importante.

Toutefois, le fait qu'elle pèse autant dans le modèle peut être dangereux, et moins représentatif de la réalité. Si par exemple, pour une année il y a peu de PM restante en fin de projection, alors la FDB prédite laisse à penser qu'elle sera très faible, compte tenu des résultats que l'on observe. Mais ce n'est pas forcément le cas, et notre prédiction sera alors trompeuse.

Pour le reste des variables, on remarque que dans le premier schéma ce sont les PM et taux cibles qui sont choisis, tandis que pour le deuxième il y a le TRC en plus qui reste important. Ce qui reste cohérent. On pourrait même s'attendre à ce qu'il soit plus présent, avec la corrélation des produits financiers que l'on pourra distribuer qui impactera directement le montant de FDB.

On peut comparer les prédictions que donne *Lime*, par rapport aux prédictions issues du modèle *Deep Learning*, afin de voir si l'explication de ces variables ont un sens, selon la proximité des résultats (Table 5) :

Graphique	Réseaux de neurones	Lime
n°1	492 679 863	493 934 378
n°2	672 876 663	650 776 619

TABLE 5 – Comparaison réseaux de neurones et Lime

Les prédictions sont relativement proches, ce qui montre l'importance des variables obtenues. L'explication locale que nous apporte *Lime* reste correcte quand aux similitudes des prédictions.

Cela reste utile si on veut étudier une situation économique précise.

On peut maintenant se poser la question de l'importance de toutes les variables du modèle. C'est ce qui va être analysé pour une observation.

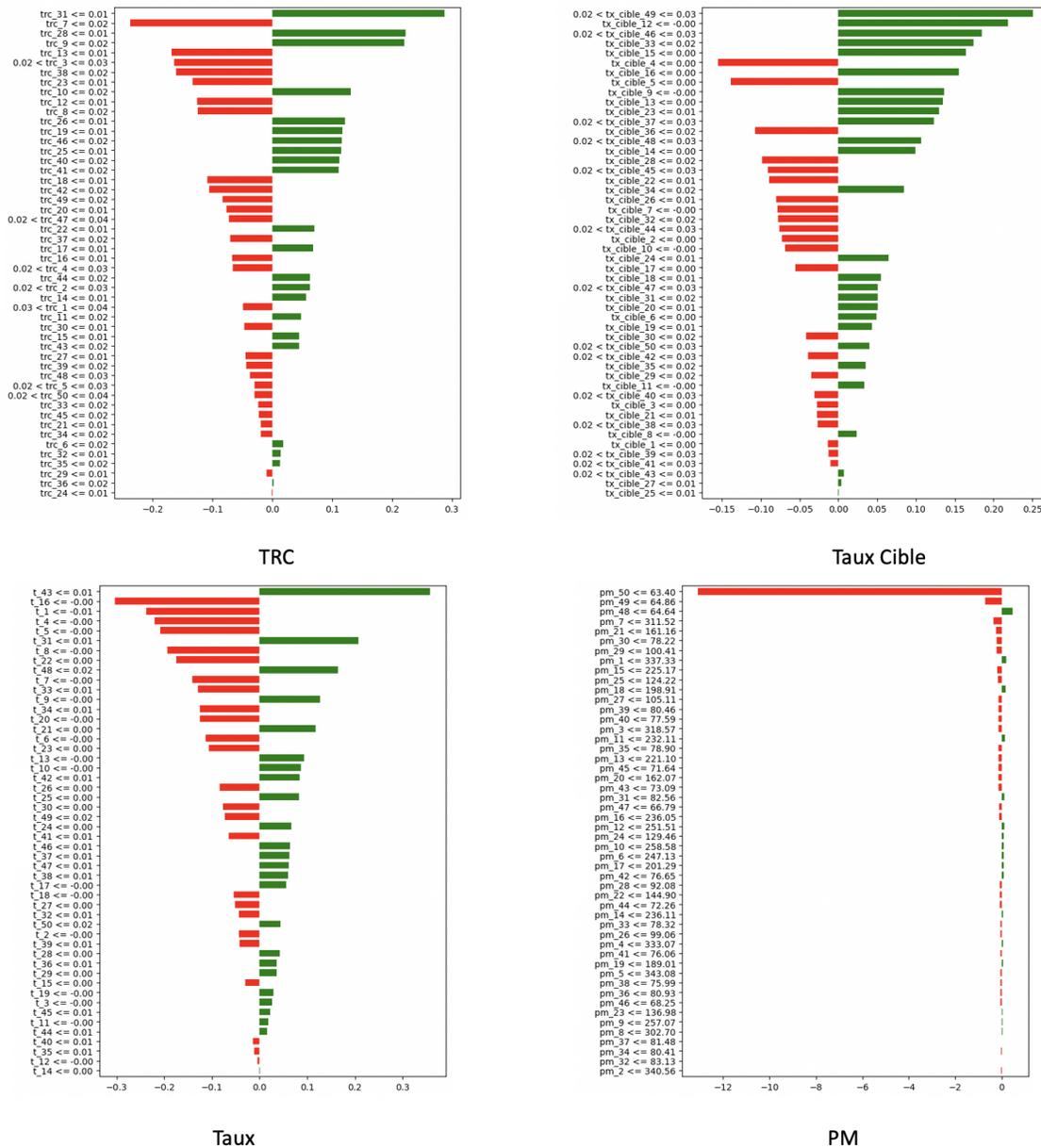


FIGURE 27 – Poids variables *Lime*

Si l'on additionne tous ces poids, avec l'*intercept* en plus, on retrouve le montant de FDB prédit par *Lime*. On retrouve encore une fois la dominance du poids des PM de la dernière année de projection. Mise à par celle-ci, les poids des autres variables restent dans le même ordre de grandeur. De manière générale, le TRC, les PM et les taux ont un poids négatif dans le modèle pour la FDB, contrairement au taux cible qui a un poids positif.

L'inconvénient de cette méthode est la définition du voisinage, surtout pour des données tabulaires. De plus, les explications peuvent être instables. Il est possible de se retrouver avec des explications complètement différentes alors que les observations choisies sont très proches l'une de l'autre.

On peut pour la suite comparer ces résultats avec la méthodes de *Shapley*, pour les mêmes observations

Shap

La méthode de *Shap* (*SHapley Additive Explanations*), donnée par Lundberg et Lee en 2017, peut être utilisée pour les interprétations locales ou globales. Comme *Lime*, c'est une méthode agnostique, qui peut être appliquée sur n'importe quel modèle de *Machine Learning*.

Cette méthode se base sur la valeur de Shapley, beaucoup utilisée dans la théorie des jeux pour répartir équitablement la contribution de chaque variables à un résultat collectif, qui est ici la FDB. Ainsi, dans notre cas, les valeurs de Shapley vont calculer les contributions de toutes nos variables explicatives, utilisées dans le modèle final, pour parvenir à notre prédiction de FDB.

On commence par approximer les valeurs de Shapley ϕ_j par la méthode Monte-Carlo :

$$\hat{\phi}_j = \frac{1}{M} \sum_{m=1}^M \left(\hat{f}(x_{+j}^m) - \hat{f}(x_{-j}^m) \right)$$

Avec \hat{f} le modèle de prédiction, x qui est l'observation que l'on cherche à interpréter, j la variable dont on cherche la contribution par rapport à la prédiction, M la taille de l'échantillon et enfin x_{+j}^m et x_{-j}^m deux instances permutées venant de l'ensemble des données.

Le point commun qu'a la méthode *Shap* avec *Lime*, est qu'elle peut être représentée comme une attribution additive des effets des variables, en plus de la valeur de base qui est la moyenne de toutes les prédictions, qui donnera le modèle d'explication.

Plusieurs approches existent dans cette méthode. Celle que l'on utilisera et qui convient le mieux à notre étude est celle du *Kernel Shap*. Elle se décompose en cinq étapes :

1. Créer une coalition c_k , qui indique si une caractéristique est dans la coalition ($c_{s,j} = 1$ si X_j est dans la coalition, $c_{s,j} = 0$ dans le cas contraire).
2. Application du modèle de prédiction sur les coalitions, en utilisant en plus une fonction h_x , qui permet d'avoir en résultat :

$$z_i = \begin{cases} x_j^{(i)} & \text{si } c_{s,j} = 1 \\ x_j^{(k)} & \text{si } c_{s,j} = 0 \end{cases}$$

avec $x_j^{(k)}$ qui est un autre observation de la variable j prit aléatoirement dans les données.

3. Calcul des poids de tous les z_i , $w_{c_s}^{(i)} = \frac{p-1}{\binom{p}{|c_s|} |c_s|^{(p-|c_s|)}}$
4. Calibrage d'un modèle linéaire pondéré
5. Renvoyer les valeurs de Shapley ϕ_j , qui seront les coefficients du modèle linéaire

En utilisant cette méthode, on obtient pour les deux mêmes observations que *Lime* :

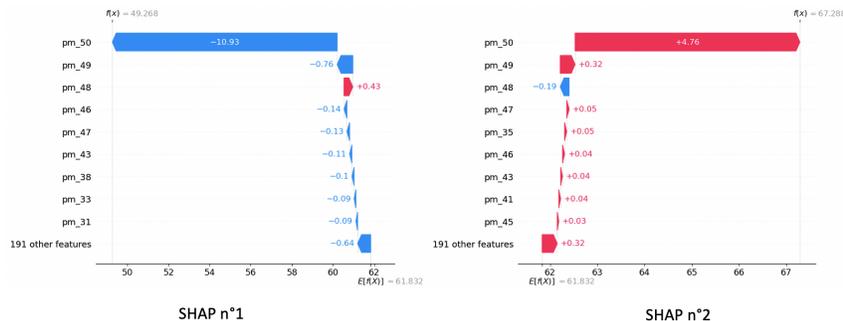


FIGURE 28 – Comparaison *SHAP* sur deux observations

On constate des similitudes par rapport aux résultats de *Lime*. Tout d'abord, la démarcation de la dernière année de PM est la même, avec un effet qui va dans le même sens pour les deux observations. En effet, celle de dernière année a un effet négatif sur la première observation, tandis qu'elle a un effet positif sur la deuxième. La dissemblance avec *Lime* se trouve au niveau des poids pour les autres variables. Ici, les autres poids les plus élevés sont pour les autres années de PM, alors que précédemment, d'autres variables se démarquaient.

En sommant les poids des diverses variables et en ajoutant la valeur de base, on obtient la FDB du modèle d'interprétation. Comparé aux FDB prédites par le modèle complexe (Table 6) :

Graphique	Réseaux de neurones	Shap
n°1	492 679 863	492 679 825
n°2	672 876 663	672 876 587

TABLE 6 – Comparaison réseaux de neurones et Shap

Les résultats sont beaucoup plus proches des prédictions que ceux de *Lime*, ce qui paraît cohérent car ici il n'y a pas d'observations générées aléatoirement. Cela ne veut pas forcément dire que les explications de cette méthode sont meilleures pour ce modèle complexe.

Afin de voir les poids d'une variable pour toutes ses années de projection, on va sommer tous les poids obtenus. Ainsi par exemple pour la variable TRC, on va sommer les poids des 50 années de projection. Les poids des quatre variables du modèle final pour ces deux observations sont résumés ainsi :

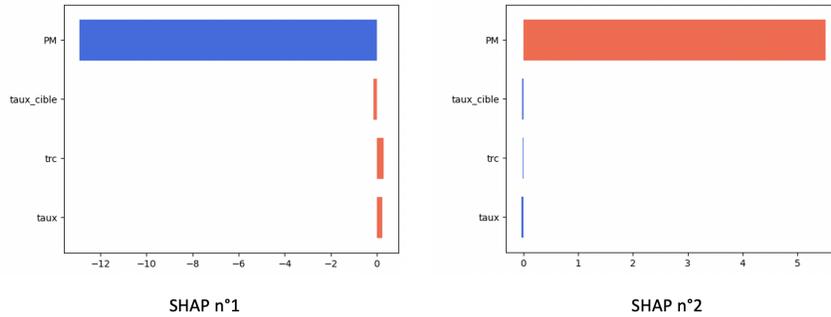


FIGURE 29 – Résumé *SHAP* des deux observations

La PM reste, en général, la variable qui a le plus de poids dans le modèle. Pour les autres, cela dépend des observations. Dans le SHAP 1, c'est le TRC qui a le deuxième plus grand poids, alors que pour le SHAP 2 c'est le taux. Comme ces méthodes sont locales, on ne peut pas déduire quelles variables sont les plus importantes entre le TRC, le taux et le taux cible. On peut cependant s'attendre à ce qu'à un niveau global, la PM soit la plus décisive du modèle. On remarque par ailleurs que lorsque la PM a un poids négatif, les trois autres en ont un positif et inversement.

4.3.2 Au niveau Global

Ce qui nous intéresse beaucoup dans cette étude est de pouvoir interpréter notre modèle au niveau global, et d'avoir les comportements généraux de chaque variable sur le modèle.

Avec la méthode *Lime*, le seul moyen d'avoir une représentation au niveau global serait de sélectionner uniquement les observations qui ont des caractéristiques différentes, afin qu'elles représentent l'entièreté des données sans qu'il n'y ait d'observation similaires. Pour ensuite faire une moyenne des résultats *Lime* obtenus. Cependant cette méthode est très peu utilisée car elle n'est pas fiable et qu'un modèle linéaire local ne peut pas être valable pour tous les cas.

L'interprétation globale du modèle peut être comprise de différentes manières. On peut tout d'abord regarder l'importance plus ou moins forte de chaque variable dans le modèle. Il est également possible d'analyser les effets marginaux des variables sur le modèle de prédiction. Nous allons utiliser différentes méthodes pour avoir ces différentes interprétations du modèle jusqu'alors construit.

Permutation Futures Importance (PFI)

Cette méthode, dont l'idée provient de *Fisher, Rudin et Dominici [2018]*, permet de découvrir les variables qui sont les plus utiles dans le modèle. Le principe de cette méthode est de permuter les valeurs d'une variable et voir l'impact que cela a sur le modèle de prédiction. Si cela débouche sur un fort impact avec une grande augmentation de l'erreur, alors cette valeur a beaucoup d'importance. Au contraire si le résultat varie très peu, avec une erreur qui ne change presque pas, alors elle sera considérée comme moins importante. On va donc calculer l'erreur originale puis l'erreur après les permutations, et l'importance de la variable sera mesurée par le quotient,

ou la différence, de l'erreur permutée avec l'erreur originale. On ne permute uniquement la variable dont on veut calculer l'importance, et plusieurs permutations sont nécessaires.

Ainsi on considère notre modèle f , X une matrice des différentes variables en entrée et Y les valeurs de FDB cible. De plus, l'erreur sera mesurée par $Err(Y, f(X))$ et σ une permutation. L'importance de la variable x_i , que l'on notera FI_i , sera donc :

$$FI_i = \frac{Err(Y, X_{perm})}{Err(Y, X_{init})} \quad \text{ou bien} \quad FI_i = Err(Y, X_{perm}) - Err(Y, X_{init})$$

On pourra alors déterminer quelles variables contribuent le plus à l'obtention de la valeur prédite. On va utiliser la bibliothèque « *mlxtend* » de Python pour cette méthode. Cette méthode va utiliser la différence des erreurs pour mesurer l'importance des variables.

Les résultats obtenus pour toutes les variables détaillées selon l'année de projection seront mis en annexe. On peut noter que l'on obtient le même constat que pour les méthodes locales, à savoir une importance plus forte sur les PM de dernière année.

Afin d'avoir une vision de l'importance des variables, j'ai rassemblé les résultats des différentes années de projection en une seule pour avoir l'impact global de la variable. On obtient les résultats suivants (Figure 30) :

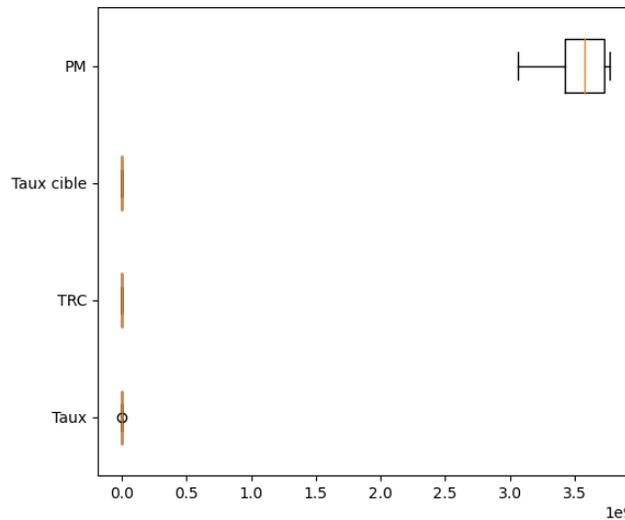


FIGURE 30 – Résultats Permutation Features Importance

La prédominance des PM est bien visible. C'est cette variable qui est la plus importante pour prédire la FDB dans notre modèle. Même en prenant la valeur de l'importance la plus basse des PM, elle reste excessivement éloignée des autres variables. L'analyse des autres caractéristiques est difficile avec ce graphe, à cause de la différence d'échelle notable. On peut zoomer sur ces trois variables pour plus de précision.

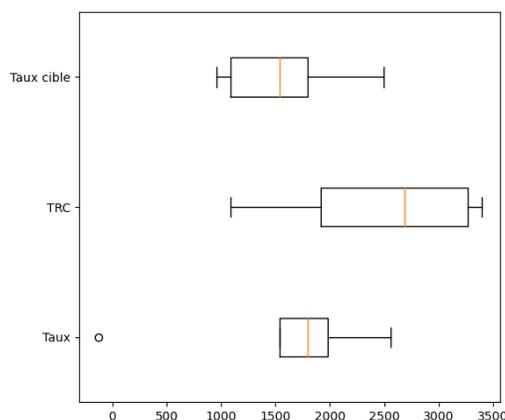


FIGURE 31 – Focalisation sur les autres variables en excluant les PM

"La deuxième variable la plus importante est le TRC, ce qui est cohérent. Effectivement, le TRC va déterminer la capacité de ce que l'on pourra distribuer de PB, dans le but de pouvoir atteindre le taux cible fixée. C'est une variables importante dans la détermination de la FDB. Les deux autres variables ont un niveau d'importance plutôt proche. On remarque également un cas particulier pour le taux, avec une valeur d'importance très éloignée, presque nulle.

L'avantage de cette méthode est qu'elle est simple à comprendre et les résultats graphiques sont facilement interprétables. De plus, grâce aux permutations, on retire les interactions possibles initiales avec les autres variables, ce qui permet de prendre en compte les effets qui existent entre les variables. Aussi, comme l'importance est basée sur l'erreur, on peut l'utiliser pour comparer plusieurs modèles.

Cependant, les résultats dépendent des permutations effectuées, qui se font de manière aléatoire. Il faut utiliser plusieurs permutations et en prendre la moyenne pour stabiliser les résultats. Aussi, si deux variables sont fortement corrélées, avec les permutations, cela peut donner lieu à des situations irréalistes et improbables dans la réalité, qui seront utilisées pour le calcul de l'importance de la variable. C'est sûrement ce que l'on retrouve avec la valeur particulière de l'importance du taux, qui a dû faire face à une situation improbable.

Pour la suite, deux méthodes qui détermineront les effets marginaux des variables vont être utilisées. La première sera le *Partial Dependence Plots* (PDP) et le second sera l'*Accumulated Local Effects Plots* (ALE) qui, comme on le verra, pallie à certains défauts du PDP.

Partial Dependence Plots

C'est une méthode graphique, apportée par *J.H Friedman [2001]*, qui permet de visualiser les effets marginaux des variables explicatives du modèle de prédiction. Il sert à découvrir quelle type de relation il y a entre la cible et une variable, que ce soit linéaire, monotone ou bien plus complexe. Le principe est d'estimer les marges en utilisant leurs équivalents empiriques, calculés à partir de \hat{f} , le modèle de réseau de neurones. On aura alors la fonction de dépendance partielle définie par :

$$\hat{f}_S(x_S) = \mathbb{E}_{X_C} \left[\hat{f}(x_S, X_C) \right] = \int \hat{f}(x_S, X_C) d\mathbb{P}(X_C)$$

Avec x_S qui est la variable dont on veut l'explication et X_C qui sont les autres variables explicatives. La sortie est marginalisée par rapport aux variables X_C , et démontre la relation qu'il existe entre le résultat du modèle et x_S .

Néanmoins, la densité $d\mathbb{P}(X_C)$ est en réalité inconnue. A l'aide de la méthode Monte-Carlo, on approxime \hat{f} par :

$$\hat{f}_S(x_S) \approx \frac{1}{N} \sum_{i=1}^N \hat{f}(x_S, x_C^{(i)})$$

Comme on regarde l'effet marginal d'une variable, il faut un PDP par variable, ce qui n'est pas possible ici compte tenu de toutes les variables. Il est possible d'avoir un graphe avec l'effet de deux variables, mais le problème reste le même. D'après les résultats de la méthode PFI, il est intéressant de regarder le PDP de l'année la plus utile du modèle pour chaque variable. En mettant en place cette méthode, on acquiert les graphes suivants (Figure 32) :

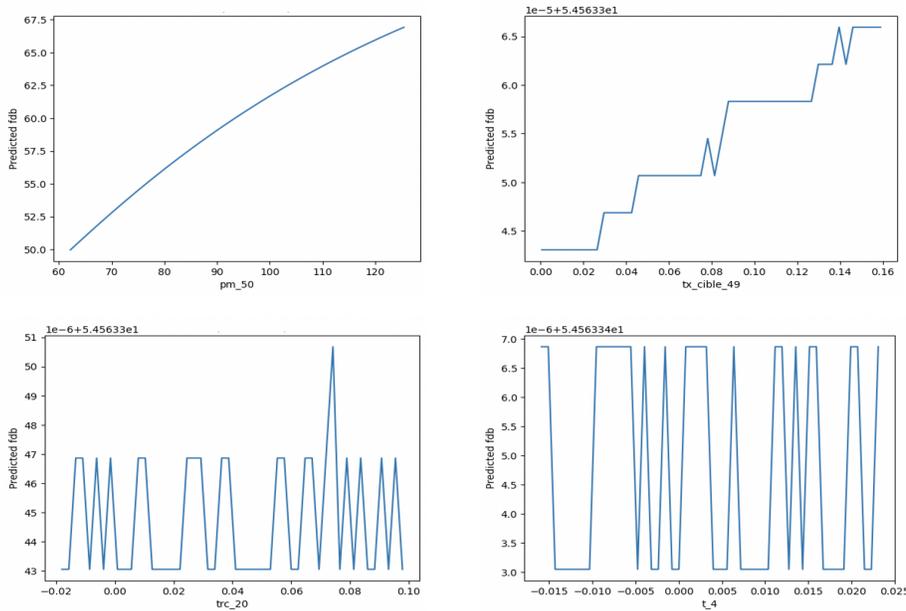


FIGURE 32 – PDP des variables les plus utiles

La relation des PM de la dernière année avec la valeur de FDB prédite ressemble presque à une relation linéaire. Le montant de FDB prédit croît avec la PM de dernière année de projection. Pour le taux cible, la courbe est en escalier, avec quelques pics occasionnels. Identiquement à la PM, plus le taux cible est important, plus le montant de FDB prédit sera élevé. Enfin pour les deux dernières variables, il y a une relation plus complexe avec la prédiction de la FDB. Les courbes sont en dents de scies, sans périodicité apparente.

L'avantage de cette méthode est sa rapidité ainsi que sa facilité de mise en place, avec un calcul pour obtenir l'effet marginale intuitif. On discerne clairement la relation que l'on cherche entre une variable et la valeur à prédire.

L'un des désavantages de cette méthode, et qu'on retrouve également dans le PFI, est l'exclusion des possibles corrélations entre les variables. S'il en existe une entre les variables de l'ensemble S avec l'ensemble C, alors des calculs de moyenne seront improbables et mèneront à des explications biaisées. Afin de pallier à ce problème, une nouvelle méthode a été introduite par *D.W. Apley [2016]* que nous allons voir juste après, et qui utilise la distribution conditionnelle au lieu de celle marginale.

Un autre point négatif important est qu'on ne peut prendre que deux variables maximum pour faire un PDP, ce qui est problématique lorsque l'on a beaucoup de variables, ce qui est notre cas.

Enfin, le fait que cette méthode se repose sur la moyenne des prédictions cache les effets hétérogènes. Il se peut qu'une variable impacte la valeur de la prédiction à travers une autre variable par sa distribution conditionnelle. Cependant sur le PDP, elle sera considérée comme non impactante sur la variable prédite.

Accumulated Local Effects Plots

La méthode ALE, de *Apley et Jingyu Zhu [2016]*, a le même but que le PDP. Afin de combler le défaut du PDP qui ne prend pas en compte les corrélations pouvant exister entre les différentes variables, on va utiliser la différence des distributions conditionnelles au lieu des effets marginaux. Si à la place de prendre la différence, on prenait la moyenne, alors cela mélangerait les effets de corrélations de toutes les variables, c'est la méthode M-plot. En prenant la différence on ne garde que l'effet recherché.

Ainsi on aura les valeurs de ALE qui seront :

$$\begin{aligned}\hat{f}_S(x_S) &= \int_{x_{min,S}}^{x_S} \mathbb{E}_{X_C|X_S} \left[\frac{\partial \hat{f}(X_S, X_C)}{\partial x_S} \mid X_S = z_S \right] dz_S - \text{constante} \\ &= \int_{x_{min,S}}^{x_S} \int_{x_C} \frac{\partial \hat{f}(z_S, X_C)}{\partial z_S} d\mathbb{P}(X_C \mid X_S = z_S) dz_S - \text{constante}\end{aligned}$$

La constante permet de centrer les résultats en mettant l'effet moyen à 0. La particularité de ALE est la présence du gradient. Toutefois, tous les modèles n'ont pas de gradient, et il est également difficile à mettre en place d'un point de vue algorithmique. On va donc utiliser une estimation de ALE qui est la suivante :

$$\hat{g}_j(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i:x_j^{(i)} \in N_j(k)} \left[\hat{f}(z_{k,j}, x_{-j}^{(i)}) - \hat{f}(z_{k-1,j}, x_{-j}^{(i)}) \right]$$

Avec :

- $n_j(k)$ le nombre d'observations de la variable j dans le k^{ieme} intervalle ;
- $z_{k,j}$ les bornes des intervalles de la variable j ;
- $x_{-j}^{(i)}$ l'observation i où l'on a retiré la variable explicative j .

Les caractéristiques sont divisées en plusieurs intervalles, puis on calcule la différence des prédictions entre chaque intervalle pour en prendre la moyenne, pour avoir la différence moyenne des prédictions pour cet intervalle. Puis les effets moyens sont cumulés pour tous les intervalles. Afin de retrouver l'effet moyen centré en 0, on obtient au final :

$$\hat{f}_j(x) = \hat{g}_j(x) - \frac{1}{n} \sum_{i=1}^n \hat{g}_j(x_j^{(i)})$$

Ainsi on peut considérer la valeur \hat{f} de ALE comme l'effet moyen d'une variable donnée à une certaine valeur, sur la prédiction moyenne. La méthode ALE sera appliquée sur les mêmes variables que le PDP afin de comparer les résultats.

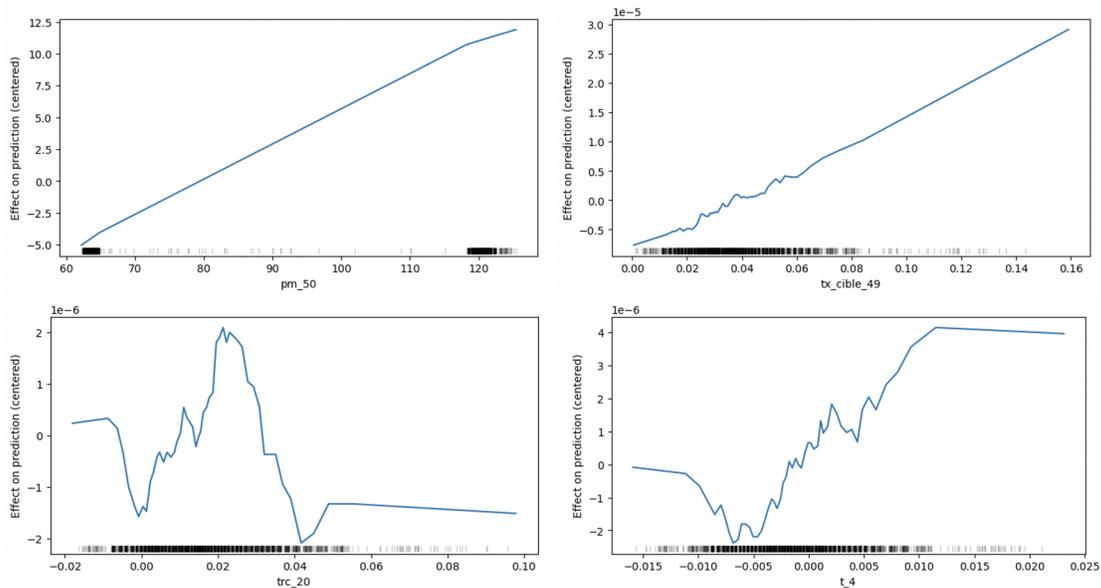


FIGURE 33 – ALE plot des variables les plus utiles

On remarque que la relation entre la FDB et la PM de la dernière année est plus linéaire par rapport aux résultats du PDP. On retrouve tout de même le montant de FDB prédit moyen qui augmente en même temps que la valeur de la PM. Lorsqu'elle est inférieure à 80 millions, cela a un effet négatif sur le montant prédit moyen. Pour le taux cible, la relation a également une forme plus linéaire avec des variations plus importantes lorsque le taux cible est compris entre 0,018 et 0,06. Si le taux est inférieur à 0,03 environ, cela a un effet négatif sur la prédiction moyenne.

Pour les deux autres variables, l'effet est plus volatile. Le TRC a tendance à avoir un effet négatif sur la prédiction moyenne, excepté principalement pour des valeurs comprises entre 0,01 et 0,03. Alors logiquement l'effet devrait être positif lorsque le TRC est élevé puisque cela entraîne plus de rendement et donc plus de participation aux bénéfices à distribuer. Lorsque le taux est supérieur à 0 environ, alors cela a un effet positif sur la prédiction moyenne de FDB. Enfin, on peut remarquer que globalement les effets ont le même ordre de grandeur, sauf encore une fois pour la PM qui a un effet nettement plus important.

L'avantage de cette méthode est qu'il n'y a pas de biais lorsqu'il y a des caractéristiques corrélées, et évite les situations improbables. De plus, le tracé d'une ALE est au moins huit fois plus rapide que pour obtenir celui d'un PDP. Il est également simple à comprendre graphiquement. Comme on le voit dans la formule ci-dessus, la valeur ALE finale repose sur une somme de fonction ALE, on peut donc le décomposer facilement.

A l'inverse, comme pour le PDP, il a le désavantage de ne prendre qu'une, voir deux variables au maximum par graphique. Ce qui est problématique lorsqu'il y a un grand nombre de variables explicatives.

Grâce à ces interprétations à la fois locales et globales, on a pu mieux comprendre le fonctionnement de notre modèle de prédiction. En effet, au global les PM et les TRC sont les plus utiles dans le modèle. Les TRC et les taux sont ceux qui ont les liens plus complexes avec notre variable de prédiction. On retrouve à chaque fois une démarcation de la PM de dernière année de projection par rapport aux autres variables. Il faut faire attention à cet effet de la PM largement plus important, qui peut non seulement engendrer des valeurs de prédictions aberrantes, non conforme à la réalité, mais également enlever les effets des autres variables qui sont tout aussi importantes dans le monde réel.

5 Impacts et limites

La FDB sert également à déterminer d'autres variables importantes en Solvabilité 2, comme les nSCR sous-modulaires, c'est-à-dire les SCR net de capacité d'absorption. Il est intéressant de vérifier les impacts que le modèle peut avoir sur ces variables selon les résultats de FDB que l'on a obtenu. Cela permettrait également de voir jusqu'où le proxy peut être utilisé pour faire des sensibilités, et s'il peut servir à estimer d'autres variables par l'intermédiaire des résultats prédis.

De plus, tester le modèle sur d'autres entités, d'allocation et d'ampleur différentes que celles étudiées va permettre de mesurer les limites du réseau de neurones implémenté.

5.1 Impacts sur les nSCR

La FDB a un impact direct sur le SCR à travers la capacité d'absorption des pertes par les provisions techniques. En effet, si l'on regarde la formule de l'ajustement de la capacité d'absorption, on a :

$$Adj = -\max(\min(BSCR - nBSCR); FDB); 0)$$

La FDB a donc un impact direct sur cette variable, qui est l'un des composants du SCR brut. Autre que l'impact à travers cette formule, il va y avoir un effet sur le calcul des SCR sous-modulaires. En effet, lorsque l'on va les déterminer, on va les calculer à la fois bruts et nets de capacité d'absorption de pertes, afin de pouvoir avoir le nBSCR. Le nBSCR va également être impacté, c'est pourquoi il va être compliqué de calculer l'ajustement avec le modèle puisqu'il faudra calculer chaque sous-module de nSCR.

La FDB va donc jouer un rôle à la fois sur l'ajustement et également sur le BSCR. Le mécanisme de distribution de la participation aux bénéfices va affecter les modules de SCR suivants :

- Risque de marché
- Risque de souscription Vie et Santé
- Risque de défaut des contreparties

Dans notre cas, et compte tenu de l'entité étudiée, il va être plus intéressant de mesurer l'impact de trois sous-modules de nSCR que sont : le nSCR longévité (risque de souscription vie), le nSCR taux d'intérêt (risque de marché) et enfin le nSCR catastrophe (risque de souscription vie).

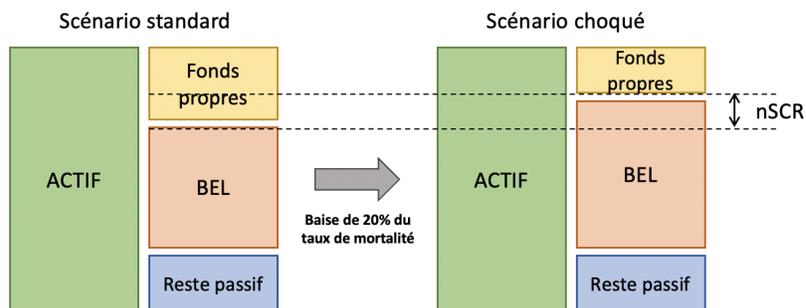
5.1.1 Longévité

Le risque de longévité est le risque de surestimation de la mortalité des assurés. Il provient de l'incertitude présente lors de la définition du taux de mortalité dans le calcul des provisions techniques.

Le nSCR de longévité correspond à la variation du *Best Estimate* de base avec l'application d'un choc prédéfini. Ainsi on a :

$$nSCR_{longevite} = BEL_{choque} - BEL_{base}$$

En Solvabilité 2, le choc à appliquer est une baisse permanente de 20% des taux de mortalité utilisés dans le calculs des provisions techniques.



On va récupérer les provisions du modèle passif qui mesure le risque de longévité, de base et choqué, qu'on utilisera ensuite dans le modèle de détermination de la FDB. Il ne faudra prendre en compte uniquement les *Model Points* de contrat pour lesquels une chute du taux de mortalité entraîne une hausse des provisions.

En appliquant les bonnes données d'input au modèle, les résultats sont les suivants (Table 7) :

	Standard	choqué
fdb ere	654 760 174	690 060 700
fdb ps	49 661 558	50 038 124
BEG	6 854 056 273	7 168 618 014
BEL	7 558 478 005	7 908 716 838

nSCR	350 238 833
-------------	--------------------

TABLE 7 – Résultats du choc longévité

La différence entre le BEL choqué et le standard donne un nSCR de 350 238 833. Si l'on compare avec le nSCR obtenu par les modèles ALM, qui s'élève à 367 304 741, on obtient un écart d'environ 4,6% , ce qui est un écart acceptable.

Cela montre que le proxy prend bien en compte le risque de longévité dans le module de la souscription vie, avec une marge d'erreur inférieure à 5%. Il est donc possible de faire des sensibilités de la FDB sur ce risque, ou sur ce sous-module de SCR.

5.1.2 Taux d'intérêt

Le risque de marché est la surestimation ou sous-estimation de la courbe des taux d'intérêt sans risque. Il est issu de l'incertitude de ce taux utilisé dans la valorisation des actifs et des passifs.

Le nSCR de taux est l'exigence en capital minimal liée au risque de taux à la hausse ou à la baisse. Il est calculé pour les deux chocs, à la hausse et à la baisse. Le nSCR final sera le plus élevé des deux.

$$nSCR_{taux\ interet} = \max(nSCR_{hausse}; nSCR_{baisse}; 0)$$

En solvabilité 2, les taux de choc à appliquer, à la hausse ou à la baisse, sont définis dans une table selon la maturité du taux sans risque. On aura comme taux choqué :

- A la baisse :

$$R_{choque} = R_{standard} \times (1 + choc_{baisse} \cdot \mathbb{1}_{R_{standard} \geq 0})$$

Si le taux d'intérêt sans risque standard est négatif, alors il n'y a pas de choc appliqué.

- A la hausse :

$$R_{choque} = R_{standard} + \max(1\%; R_{standard} \times choc_{hausse})$$

Un choc de minimum 1% doit être appliqué.

Le calcul du nSCR sera le même que pour celui de longévité, à savoir la différence entre le BEL choqué et le BEL standard. Dans notre cas ce sera le choc à la baisse qui prédominera. On applique donc la courbe des taux choqué à la baisse au modèle de prédiction. Les résultats sont en Table 8 :

	Standard	choqué
fdb ere	662 636 207	621 978 893
fdb ps	190 832 670	192 548 723
BEG	8 028 411 776	8 255 898 620
BEL	8 881 880 653	9 070 426 236

nSCR	188 545 583
-------------	--------------------

TABLE 8 – Résultat du choc taux d'intérêt

Le nSCR calculé à partir des résultats du modèle ALM étant de 180 361 019, on arrive à un écart de 4,5% par rapport à celui obtenu ci-dessus. Le choc de taux est donc assez bien représenté par le proxy, avec une erreur qui reste relativement faible. Le fait que la courbe des taux fasse partie des inputs de tous les sous-modèles et du modèle final construit en est sûrement l'une des raisons. Cela montre qu'il est possible de mesurer les variations de la courbe des taux sur la FDB et sur ce sous-module de SCR du risque de marché, à partir de ce proxy.

5.1.3 Catastrophe

En assurance vie, le risque de catastrophe est le risque de perte sur la survenance de phénomènes extrêmes et exceptionnels, dont l'incertitude des hypothèses de tarification et de provisionnement est importante.

Le principe de calcul du nSCR reste le même que pour les deux précédents. Le choc appliqué sera une hausse de 0,15 point sur le taux de mortalité du calcul des provisions, pour la mortalité des 12 prochains mois. Par ailleurs, comme pour le choc longévité, on ne prend en compte que les *Models Points* où la baisse du taux de mortalité se suit par une augmentation des provisions techniques.

En appliquant ce choc, on obtient en Table 9 :

	Standard	choqué
fdb ere	654 760 174	656 751 822
fdb ps	49 661 558	56 978 502
BEG	6 854 056 273	6 886 906 206
BEL	7 558 478 005	7 600 636 530

nSCR	42 158 525
-------------	-------------------

TABLE 9 – Résultats du choc Catastrophe

En comparant avec le nSCR initial, il y a un écart de 20%, ce qui est important. Le choc appliqué pour le risque catastrophe n'est pas bien comptabilisé dans le modèle de réseau de neurones. Les variations enregistrées par ce risque ne seront pas conformes à la réalité en utilisant ce modèle.

Au final, le modèle mesure bien les impacts des chocs de longévité et de taux d'intérêt de l'entité, mais représente mal le choc de catastrophe. Si on étend ces résultats pour tous les sous-modules de SCR, alors on peut s'attendre à ce que d'autres chocs entraînent trop d'écart par le modèle, et donc entraînent un montant de SCR total assez éloigné. Il faudrait appliquer le modèle pour tous les chocs imposés par Solvabilité 2, nécessaire pour calculer le SCR et utiliser la matrice de corrélation présentée en première partie, pour obtenir le véritable impact du SCR de cette entité.

5.2 Les limites du modèle

Le proxy a été construit sur le modèle annuel, tout comme les données. Il est aussi utile de pouvoir avoir une idée de la FDB pour les périodes trimestrielles, ou même sur un moment précis. Le modèle sera donc testé sur un trimestre de l'entité. Enfin, deux autres tests seront effectués, dont l'un portera sur une autre entité avec une importance et une composition assez proche de l'entité étudiée, et l'autre sur une entité plus modeste avec une composition plus éloignée.

5.2.1 Application sur un trimestre

Le *Best Estimate* doit être calculé tous les trimestres, en plus de l'annuel en fin d'année. Il est donc intéressant de voir comment agit le modèle sur des données d'une autre échelle, qui est ici trimestrielle. Ainsi des stratégies pourront être mises en place à tout moment. Cela permettrait de tester si le modèle est fiable ou s'il ne fonctionne uniquement sur une échelle annuelle.

Le principe reste le même, sauf que l'on prend les scénarios économiques de la période estimée, et on change également les données du passif ainsi que l'allocation d'actifs au moment de projection. En changeant uniquement ces paramètres et en prenant la même entité, il serait logique que le modèle prédise une FDB qui reste dans une marge d'erreur acceptable, que l'on s'est fixé à 5%.

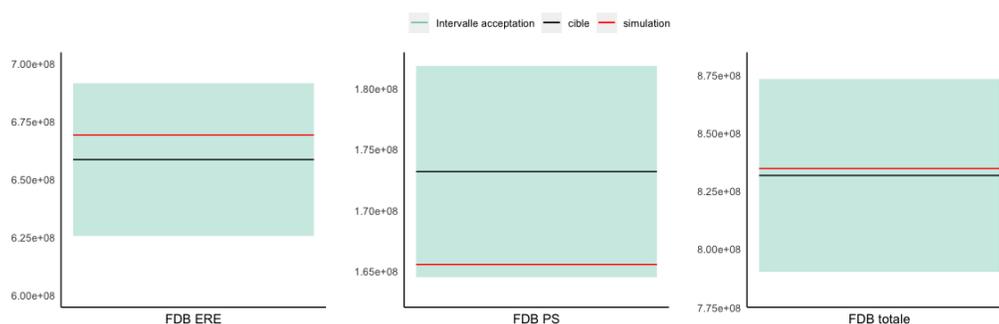


FIGURE 34 – Résultats proxy sur un trimestre

En Figure 34, on voit que la FDB épargne retraite est relativement proche de la valeur réelle, ce qui confirme la supposition initiale qui correspond aux résultats que l'on a pu avoir jusqu'ici. En ce qui concerne la FDB prévoyance-santé, l'écart est plus important mais reste dans une marge d'erreur faible, de 5%, dont on retrouve la même situation que pour les résultats finaux du Graphe 24.

Avec la surestimation de la partie Epargne-Retraite, et la sous-estimation de celle en prévoyance-santé, on obtient un FDB totale qui est proche de la valeur réelle. Le plus important reste les résultats obtenus pour l'Epargne-Retraite, car c'est celui qui nécessite d'être calculé de manière stochastique. Les résultats restent donc globalement satisfaisants pour l'utilisation du modèle à un niveau trimestriel.

5.2.2 Test sur une entité similaire

Arrive le moment de tester les limites du modèle sur d'autres entités. En effet, chaque entité est composée de produits différents, avec une allocation d'actifs différente, et d'ordre de grandeur de provision hétérogène. Dans cette partie, le modèle sera testé sur une entité d'ampleur similaire, avec cependant plus de produits en prévoyance-santé que d'épargne-retraite. On devrait donc trouver une FDB ERE plus faible.

Les scénarios économiques ne changent pas car on reste sur une même période. En prenant les données du passif de la nouvelle entité considérée, et l'allocation d'actifs basée sur plus d'actions et d'OPCVM, avec moins d'immobilier, on obtient alors :

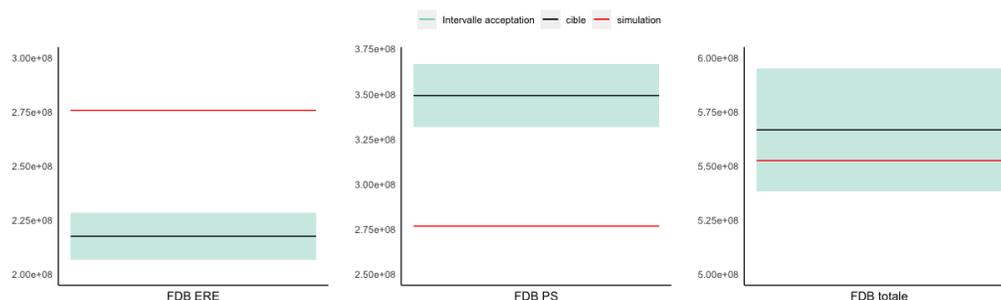


FIGURE 35 – résultats proxy sur une entité différente

Dans les deux parties de FDB estimée, on se retrouve en dehors de l'intervalle d'erreur fixée. Même si au final, avec la large surestimation de la partie ERE et la large sous-estimation de celle de Prévoyance-Santé, on tombe au final sur une FDB totale qui se trouve dans l'intervalle d'acceptation, cela ne peut être valide. En effet, le but de ce mémoire était de trouver un proxy sur la FDB_{ERE} , ce qui ici ne fonctionne pas avec le modèle.

Cela reste cohérent dans le sens où la composition de l'entité est différente, et les taux appliqués ainsi que la distribution de participation aux bénéfices ne sont pas les mêmes. Pour que le modèle fonctionne, il faudrait recalibrer les réseaux de neurones, des sous modèles et du modèle final, avec les données de cette entité pour avoir un modèle fonctionnel.

5.2.3 Test sur une entité disparate

Enfin, le modèle peut être challengé sur une entité avec un montant de provisions beaucoup plus modeste. Pour ce cas là, on ne regardera le résultat uniquement pour la FDB_{ERE} , car la partie prévoyance-santé est négligeable.

On peut s'attendre à avoir le même constat que pour la partie précédente, à savoir de devoir recalibrer le modèle à partir de l'entité prise en compte.

En appliquant le réseau de neurones sur les données de l'entité, avec une allocation d'actifs comprenant principalement des obligations et des OPCVM (ce qui se rapproche plus de l'allocation de la première entité), le résultat est le suivant :

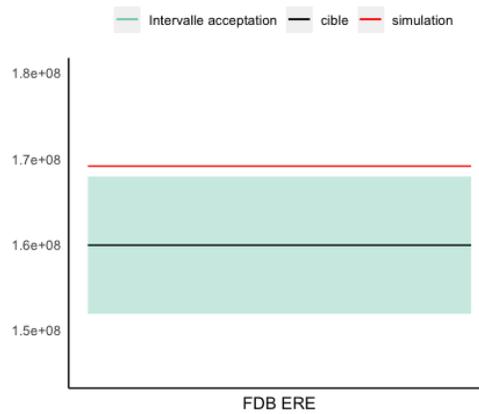


FIGURE 36 – résultats proxy sur une entité plus petite

La prédiction est aussi en dehors de la marge d'erreur de 5%, mais n'en est pas non plus très éloignée. Cela est peut être dû au fait que l'allocation d'actifs se rapproche de l'entité du Groupe étudiée tout au long de ce mémoire.

Ainsi on a pu tester les limites de ce modèle, en remarquant qu'il peut être utilisé à partir de données trimestrielles et donc calculé à tout moment pour cette entité. Cependant pour pouvoir le mettre en place pour d'autres entités du Groupe, il faudrait repartir du calibrage des hyperparamètres de chaque réseaux de neurones en utilisant les données propres à chaque entité. On a pu cependant remarquer que plus l'allocation d'actifs est proche de l'entité initiale, plus les prédictions sont proches de la cible, même si elles ne sont pas de la même ampleur.

Conclusion

Dans ce mémoire, nous avons vu le cheminement du calcul de la FDB dans les deux branches que sont : l'épargne-retraite et la prévoyance-santé. Cette variable dépend grandement de l'environnement économique. Dans le contexte économique actuel, il est intéressant pour l'entreprise d'estimer ce que l'on pourra servir aux assurés le plus rapidement possible, selon des indicateurs financiers.

L'étude de différents scénarios a permis d'analyser les variables impactant notre cible, et leurs liens avec les indicateurs financiers. Ainsi nous avons pu construire les premiers modèles qui permettent de simuler ces variables à partir des indicateurs, pour ensuite utiliser les résultats dans un modèle final qui estimera la FDB. Ces indicateurs financiers découlent de générateurs de scénarios économiques préalablement testés afin de vérifier leurs adéquations avec les valeurs de marchés.

Les modèles mis en place sont des réseaux de neurones récurrents, qui suite à de nombreux travaux ont montré qu'ils fournissent de bonnes prédictions pour des séries temporelles. Nous avons comparé les résultats entre un réseau de neurones de type LSTM et un de type GRU afin d'avoir les meilleures prédictions possibles. Nous opterons pour un réseau de neurones LSTM car il fournit de meilleurs résultats. Les prédictions alors obtenues par le modèle construit rentrent dans l'intervalle d'acceptation de 5% que l'on s'est fixé, en particulier pour la partie FDB_{ERE} qui est la raison du proxy construit.

L'interprétation de ce modèle *black-box* a pu être possible grâce à diverses méthodes, à la fois locales et globales. Ainsi on a pu mieux comprendre le modèle, et en déduire que les PM jouent un rôle des plus importants dans celui-ci. Cela entraîne une vigilance particulière à avoir lors de son utilisation, lorsque les inputs sont construits. Les résultats obtenus ne peuvent être utilisés qu'à titre indicatif, et en aucun cas pour un but de résultat officiel.

Enfin on a pu voir que le modèle prenait en compte les sensibilités sur la longévité, ainsi que sur les taux d'intérêt, mais pas celles des risques de catastrophes. Cela peut laisser à penser que d'autres sensibilités ne peuvent être mesurées par le réseau de neurones. Une autre limite de ce modèle est son périmètre d'utilisation sur d'autres entités, en effet les résultats sur les deux autres entités choisies ne sont pas satisfaisants, avec une erreur trop importante. La recalibration du modèle sera alors nécessaire. Chaque entité doit avoir son modèle spécifique, sauf si l'on essaye de construire à partir d'autres variables en plus.

Dans ce mémoire, pour construire le modèle, on s'est placé entre le modèle passif et le modèle ALM. La construction de l'input nécessite des travaux préalables pour fiabiliser les données. Cela peut également prendre du temps. Il peut alors être intéressant d'essayer de construire un modèle qui part d'avant le modèle passif, à partir de données que l'on possède déjà. Ce modèle pourrait prendre en entrée des informations par rapport aux assurés telles que leurs âges, leurs taux ..., afin de pouvoir appliquer le modèle sans avoir à faire de traitement préalable et gagner d'avantage de temps, peut être même en précision et en domaine d'utilisation.

Table des figures

1	Passage de SI à SII au bilan	12
2	Structure du SCR en Formule Standard	16
3	Rachat structurel	20
4	Flux de BEG	21
5	Illustration d'un taux forward	24
6	Rendements par actifs	26
7	Indicateurs des actifs	27
8	Évolution du TRC et des produits financiers	27
9	Illustration du fonctionnement du modèle	30
10	Test de martingalité sur les déflateurs	32
11	Test de martingalité sur les actions	33
12	Flux de BEL pour un scénario central	34
13	Flux de BEL stochastique	35
14	Résumé de l'algorithme de PB	39
15	Gestion de la provision d'égalisation	40
16	Variations des FDB et des TRC	43
17	Comparaisons des indicateurs	44
18	Comparaisons des indicateurs pour des variations inverses	46
19	Structure du proxy	47
20	Un réseau de neurones	49
21	Exemple d'une couche récurrente	50
22	Résultats des prédictions tests	56
23	Résultat de la FDB ERE sur les données tests	57
24	Résultats finaux du modèle	57
25	Explication LIME	59
26	Comparaison <i>Lime</i> sur deux observations	60
27	Poids variables <i>Lime</i>	61
28	Comparaison <i>SHAP</i> sur deux observations	63
29	Résumé <i>SHAP</i> des deux observations	64
30	Résultats Permutation Features Importance	65
31	Focalisation sur les autres variables en excluant les PM	66
32	PDP des variables les plus utiles	67
33	ALE plot des variables les plus utiles	69
34	Résultats proxy sur un trimestre	75
35	résultats proxy sur une entité différente	76
36	résultats proxy sur une entité plus petite	77

Liste des tableaux

1	Allocation de l'actif	22
2	Part des écarts avec le scénario central	42
3	FDB des scénarios choisis	43
4	Résultats des métriques	56
5	Comparaison réseaux de neurones et Lime	60
6	Comparaison réseaux de neurones et Shap	63
7	Résultats du choc longévité	72
8	Résultat du choc taux d'intérêt	73
9	Résultats du choc Catastrophe	74

Bibliographie

- Armel K., Planchet F. : *L'évaluation économique des engagements en assurance vie : écueils, bonnes pratiques et préconisations pour une mise en œuvre pertinente*. Laboratoire des sciences actuarielles et financière, 2020.
- Apley D. W., Zhu J. : *Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models*, arXiv e-prints, arXiv-1612, 2016.
- Choquer A. : *Méthodes de classification pour l'optimisation des calculs Best Estimate en assurance-vie*, 2019.
- Commission Européenne : *Règlement Délégué (UE) 2015/35*, 2014.
- Delcaillau D., Ly A., Papp A., Vermet F. : *Model Transparency and Interpretability : Survey and Application to the Insurance Industry*, European Actuarial Journal, 2022.
- Fisher A., Rudin C., Dominici F. : *All Models are Wrong, but Many are Useful : Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously*, arXiv preprint arXiv :1801.01489, 2018.
- Friedman J. H. : *Greedy Function Approximation : A Gradient Boosting Machine*, Annals of statistics, 2001.
- Gach F., Hochgerner S. : *Estimation of future discretionary benefits in traditional life insurance*, ASTIN Bulletin : The Journal of the IAA, 2022.
- Gemelgo P., Housseau S. : *L'assurance retraite-supplémentaire : de Solvabilité I à Solvabilité II*, 2014.
- Gerondeau E. : *Ratio de couverture de Solvabilité 2 d'un contrat d'épargne en euros, quels leviers de pilotage pour l'assureur ?*, 2017.
- Gumaei A., Abd-el Ghany S., Ali A., Al-Rakhami M., Omran N. Saleh H. : *Applying deep learning methods on time-series data for forecasting COVID-19 in Egypt, Kuwait, and Saudi Arabia*. Complexity, 2021.
- Hammouaadi S. : *Mise en place d'un proxy pour le calcul de la capacité d'absorption par le passif en assurance vie*, 2020
- Institut des Actuaires : *Groupe de travail « Best estimate Liabilities Vie »*, 2016.
- Kuo K. : *Individual claims forecasting with Bayesian mixture density networks*, arXiv preprint arXiv :2003.02453, 2020.
- Lundberg SM., Lee SI. : *Consistent feature attribution for tree ensembles*, arXiv preprint arXiv :1706.06060, 2017.
- Planchet F., Robert C. : *Insurance data analytics*, Economica, 2020.
- Ribeiro M., Samer S., Guestrin C. : *"Why should I trust you ?" : Explaining the predictions of any classifier*, 2016.

Rihani S. : *Estimation du coût des options et garanties des contrats épargne-retraite collectif*, 2019.

Yang S., Yu X. Zhou Y. : *Lstm and gru neural network performance comparison study : Taking yelp review dataset as an example*. 2020 International workshop on electronic communication and artificial intelligence (IWECAI), 2020.

Annexe

Annexe I : Tables corrélations SCR

Niveau de corrélation	Marché	Défaut	Vie	Santé	Non-vie
Marché	1				
Défaut	0,25	1			
Vie	0,25	0,25	1		
Santé	0,25	0,25	0,25	1	
Non-vie	0,25	0,50	0	0	1

Matrice corrélation SCR modulaires

Niveau de corrélation (CorrL)	Mortalité	Longévité	Invalité/morbidité	Frais	Rachat	Révision	CAT
Mortalité	1						
Longévité	-0,25	1					
Invalité/morbidité	0,25	0	1				
Frais	0,25	0,25	0,50	1			
Révision	0	0,25	0	0,50	1		
Rachat	0	0,25	0	0,50	0	1	
CAT	0,25	0	0,25	0,25	0	0,25	1

Matrice corrélation SCR sous-modulaires Vie

Niveau de corrélation (CorrM)	Taux	Actions	Immobilier	Spread	Concentration	Change
Taux	1					
Actions	0 ou 0,5 *	1				
Immobilier	0 ou 0,5 *	0,75	1			
Spread	0 ou 0,5 *	0,75	0,5	1		
Concentration	0	0	0	0	1	
Change	0,25	0,25	0,25	0,25	0,25	1

Matrice corrélation SCR sous-modulaires Marché

Annexe II : Exemple implémentation réseau de neurones

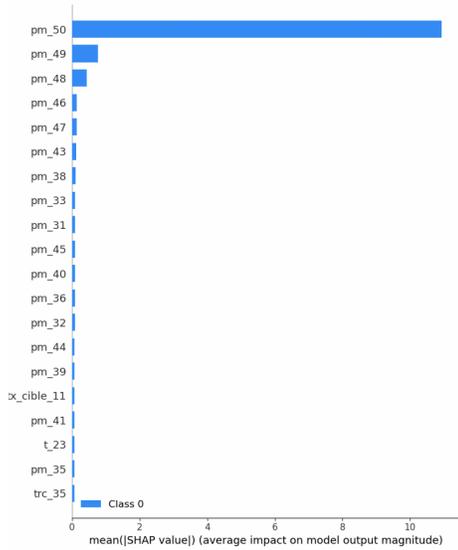
On définit la fonction avec les différentes valeurs à tester pour les hyperparamètres :

```
1 def build_model(params):
2     model = keras.Sequential()
3     params_units1 = params.Choice('units1', values=[200, 300, 400])
4     model.add(LSTM(params_units1, input_shape=(200,1), activation='tanh',
5                   recurrent_activation='sigmoid'))
6     params_drop = params.Choice('dropout', values=[0.3, 0.5])
7     model.add(Dropout(params_drop))
8     params_units2 = params.Choice('units2', values=[100, 150, 200])
9     params_activation1 = params.Choice('activation1', values=['linear', 'relu', '
10    selu'])
11    model.add(Dense(params_units2, activation = params_activation1))
12    params_drop2 = params.Choice('dropout2', values=[0.3, 0.5])
13    model.add(Dropout(params_drop2))
14    params_units3 = params.Choice('units3', values=[100, 150, 200])
15    params_activation2 = params.Choice('activation2', values=['linear', 'relu', '
16    selu'])
17    model.add(Dense(params_units3, activation = params_activation2))
18    params_drop3 = params.Choice('dropout3', values=[0.2, 0.4])
19    model.add(Dropout(params_drop3))
20    params_activation3 = params.Choice('activation3', values=['linear', 'relu', '
21    selu'])
22    model.add(Dense(1, activation=params_activation3))
23    params_lr = params.Choice('learning_rate', values=[1e-2, 1e-3, 1e-4])
24    model.compile(loss='mse', optimizer=keras.optimizers.Adam(params_lr), metrics
25                  =['mean_squared_error'])
26    return model
```

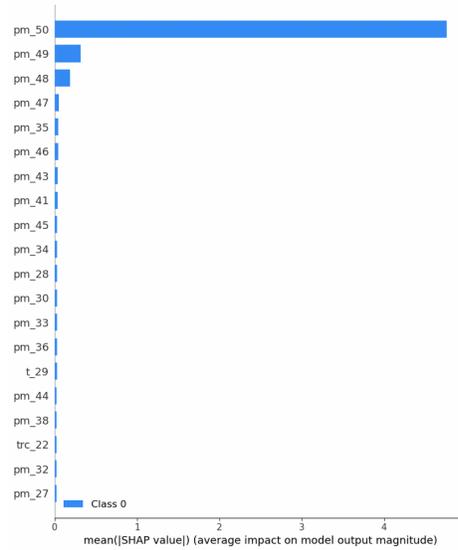
Puis on regarde les résultats de ces modèles pour avoir les paramètres optimaux :

```
1 tuning_fdb = keras_tuner.RandomSearch(build_model,
2                                       objective="val_mean_squared_error",
3                                       max_trials=25,
4                                       directory='C:\\Users\\M21II03.SI2M\\OneDrive - SI2M\\
5                                       Documents\\Models saved',
6                                       project_name='fdb',
7                                       overwrite=True)
8 tuning_fdb.search(train_fdb_1, train_cible_fdb, epochs=15, validation_data = (
9     test_fdb_1, test_cible_fdb ))
```

Annexe III : Résultats SHAP détaillés

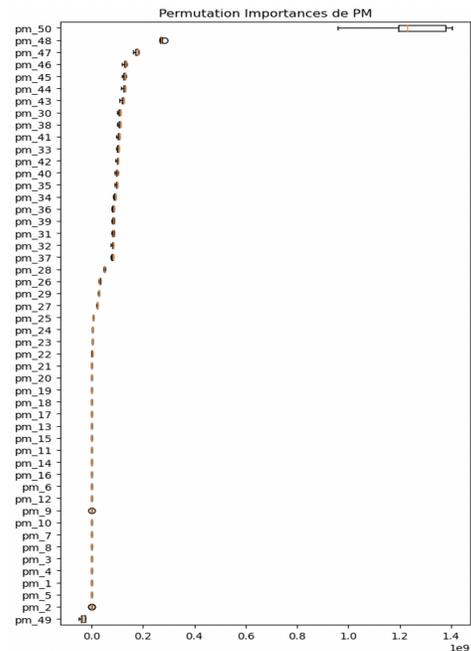
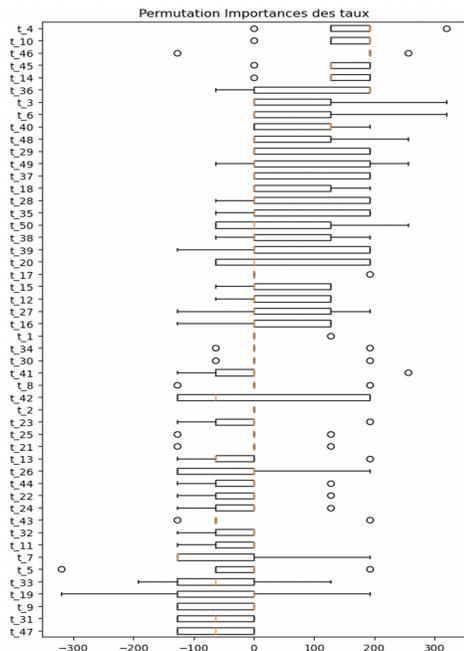
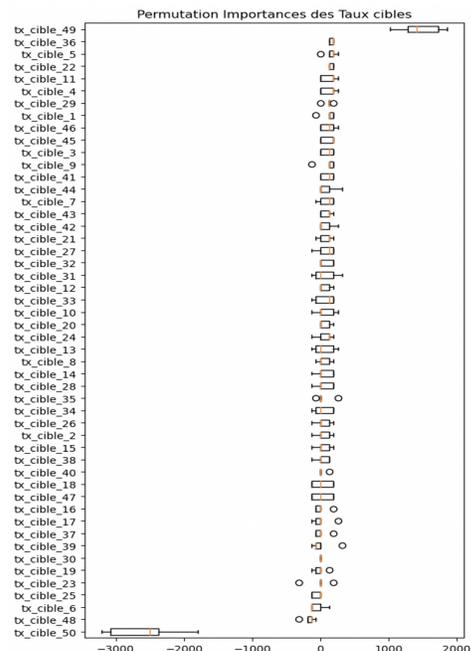
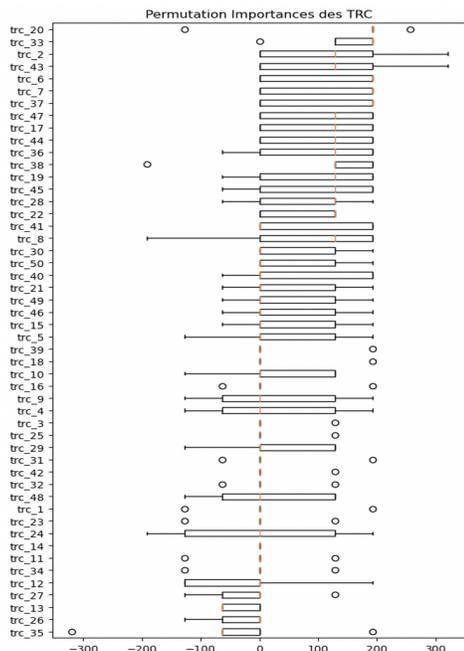


SHAP n°1



SHAP n°2

Annexe IV : Résultats PFI détaillés



Annexe V : Table chocs Solvabilité 2 taux d'intérêt

Échéance (en années)	Choc à la baisse ($Shock_{down}$)	Choc à la hausse ($Shock_{up}$)
≤ 1	- 75%	+ 70%
2	- 65%	+ 70%
3	- 56%	+ 64%
4	- 50%	+ 59%
5	- 46%	+ 55%
6	- 42%	+ 52%
7	- 39%	+ 49%
8	- 36%	+ 47%
9	- 33%	+ 44%
10	- 31%	+ 42%
11	- 30%	+ 39%
12	- 29%	+ 37%
13	- 28%	+ 35%
14	- 28%	+ 34%
15	- 27%	+ 33%
16	- 28%	+ 31%
17	- 28%	+ 30%
18	- 28%	+ 29%
19	- 29%	+ 27%
20	- 29%	+ 26%
21 à 89	Interpolation linéaire	Interpolation linéaire
≥ 90	- 20%	+ 20%