

Mémoire présenté devant l'ENSAE ParisTech
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 08/11/2017

Par : Adel Messoussi

Titre : Application d'algorithmes de *machine learning*
pour l'estimation du ratio de couverture
d'un assureur-vie détenteur d'un produit épargne

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

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

Membres présents du jury de la filière

Entreprise :

Nom : Actuaris

Signature :

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

Directeur du mémoire en entreprise :

Nom : Marine NIEDZWIEDZ

Signature :

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

Signature du responsable entreprise

Secrétariat :

Signature du candidat

Bibliothèque :

Résumé

Dans le cadre du pilier 2 de la directive Solvabilité 2, l'ORSA ou évaluation interne des risques et de la solvabilité (Own Risk and Solvency Assessment), se caractérise par la nécessité pour une compagnie d'assurance d'être en mesure de s'assurer en permanence de la couverture de son capital réglementaire. Il s'agit d'un processus d'évaluation prospectif des risques, appréhendé sur la durée du business plan de l'entreprise et peut s'étendre de 3 à 8 ans.

Ainsi, cette évaluation permanente doit conduire les organismes à simuler dans le temps l'évolution de leur capital requis (ou *SCR*, *Solvency Capital Requirement*) et de leurs fonds propres économiques (*FP*) sur la base de scénarii d'évolution de leur facteur de risque. Plusieurs possibilités d'environnements économiques futurs seront considérées pour effectuer ces calculs (tels que des changements de la situation économique, des modification subites des portefeuilles d'encours, changement règlementaire etc). L'ensemble de ces scénarii seront appelés des « possibles ».

L'exploration d'un nombre très élevé de « possibles », via des approches stochastiques, rend les calculs complexes, à la fois en termes de temps de calculs mais aussi d'un point de vue mathématiques. Cela incite les compagnies d'assurance à envisager plusieurs approches de simplification afin d'évaluer leurs engagements. La mise en œuvre de méthodes simplifiées doit cependant tenir compte du profil de risque de la compagnie tout en arbitrant entre coût de mise en œuvre, précision attendue, maîtrise dans la justification des résultats et sensibilité aux hypothèses.

L'objectif de ce mémoire consiste donc en l'estimation d'une fonction dite de proxy permettant d'évaluer le *SCR* et les fonds propres via des algorithmes de *machine learning* afin de rendre possible le calcul du ratio de couverture sur l'horizon de business plan sans avoir à relancer un grand nombre de calcul.

La construction d'une telle fonction se fera en deux étapes. Dans un premier temps, nous déterminerons les variables pertinentes au calcul du *SCR* et des fonds propres économiques via une régression *LASSO* avant de calibrer cette fonction via des *réseaux de neurones* et des *Support Vector Machine*.

Mots clés : Solvabilité 2, ORSA, Ratio de couverture, Capital économique, Fonds propres économiques, LASSO, Réseau de neurones, SVM, Machine learning

Abstract

Within the framework of the pillar 2 of Solvency 2, ORSA, the intern evaluation of risks and solvency (Own Risk and Solvency Assessment) characterizes itself with the necessity of being able to insure at any time the covering of the regulatory capital of a given insurance firm. It is an evaluation of risks based on a prospective vision within three to eight years which can correspond to the business plan of the company.

Thus, this permanent evaluation is supposed to lead organizations into making simulations through time portraying the evolution of their SCR (Solvency Capital Requirement) and of their eligible own fund (EOF) on the basis of scenarios of their risk factor. Different futur scenario's possibilities will be considered to make this regulatory computation (macroeconomic changes, regulatory and so on). These changes are qualified as « possible ».

Exploring the great number of « possible » states via stochastic approaches tends to make calculation complex be it regarding the time that calculations take but also on a mathematic point of view. This is why insurance companies are pushed to consider simplifications in their calculus approach. Setting up these simplified methods must however take into account the risk profiles while still juggling with the cost of implementation, the awaited precision of results, the control over the justification of the results and the sensitivity to the hypothesis.

The aim of this master thesis is to compute a parametric function based on the SCR and EOF through machine learning techniques so as to forecast the coverage ratio.

Computing such a function will be made in two steps. On the one hand, we will select the variables contributing the most to the variation of the SCR and the eligible own fund using *LASSO* which will then lead us to calibrate the function through *neural networks* and *Support Vector Machine*.

Key words : Solvency 2, ORSA, Coverage ratio, SCR, Eligible own fund, LASSO, Neural networks, SVM, Machine learning

Remerciements

Je tiens à remercier Marine NIEDZWIEDZ, ma tutrice lors de ce stage chez ACTUARIS pour m'avoir permis de réaliser ce mémoire ainsi que pour sa disponibilité et ses nombreux conseils.

Je souhaite également remercier l'ensemble du pôle Vie d'ACTUARIS pour son accueil et pour l'aide reçus sur la totalité de ma période de stage. Je pense en particulier à Delphine MORGANA, Anne-Claire MARTIAL, François UZAN et David MARIUZZA.

Je remercie Virak NOU, associé du pôle Vie chez Actuaris, de m'avoir permis de réaliser mon mémoire au sein de son équipe et d'avoir partagé les avancées du pôle de manière régulière au même titre que les autres consultants.

Je tenais à remercier très chaleureusement les autres consultants du cabinet pour leurs nombreuses relectures, leurs encouragements et leurs soutiens inconditionnels, en particulier Yasmine ESSAHEL, Manon VALDIGUIE, Celia COLLIAUX, Ines MANAI, Ghita CHERGOU, Maroi BEN ARIBI et Ghita BOUCHTA.

Pour finir, je remercie Ismaël MESSOUSSI ainsi que Badr-Eddine CHERIEF-ABDELLATIF pour leur soutien vital.

Sommaire

Introduction	9
I Calcul du capital économique prospectif sous Solvabilité 2	12
1 Cadre réglementaire	13
1.1 De Solvabilité 1 à Solvabilité 2	13
1.1.1 Evaluation prudente des engagements	13
1.1.2 Représentation des engagements	14
1.1.3 Marge de solvabilité	14
1.1.4 Limites de Solvabilité 1	14
1.2 Les 3 piliers de Solvabilité 2	15
1.2.1 Pilier 1	15
1.2.2 Pilier 2	16
1.2.3 Pilier 3	16
1.3 Calcul du capital économique sous Solvabilité 2	17
1.3.1 Définition	17
1.3.2 Calcul du capital économique en formule standard	17
1.3.3 Calcul du capital économique en modèle interne	19
2 Focus pilier 2 : le processus ORSA	21
2.1 Approche générale	21
2.2 Dimension prospective du capital économique	22
3 Alternative au calcul du capital économique	23
3.1 Méthode se basant sur un proxy	23
3.1.1 Curve-fitting	24
3.1.2 Least Square Monte-Carlo	25
3.2 Calcul d'un proxy par <i>machine learning</i>	26
II Sélection des facteurs de risques pertinents au calcul du capital économique	29
4 Description de la base	30
4.1 Hypothèses liées à l'actif	30
4.2 Hypothèses liées au passif	33
5 Construction d'une base d'apprentissage	34
5.1 Hypercube latin	34
5.2 Construction des courbes des taux	37
5.2.1 Générateur de scénarii économiques	37
5.2.2 Courbes en inputs	38
5.2.3 Projections des autres indicateurs	39
5.3 Etude des corrélations entre nos variables	40
5.3.1 Coefficient de corrélation de Pearson	40
5.3.2 Coefficient de corrélation de Kendall	42

6	Sélection des variables	43
6.1	Présentation du modèle	43
6.2	Convexité du modèle	44
6.3	Calcul des estimateurs	45
6.4	Validation-croisée des paramètres	47
6.4.1	Sélection de variables pour l'estimation du BE	48
6.4.2	Sélection de variable pour l'estimation des FP	51
 III Utilisation de techniques de <i>machine learning</i> pour la mise en place d'un proxy		53
7	Les Support Vector Machines	54
7.1	Contexte	54
7.2	Formulation du problème	54
7.2.1	Cas linéaire	54
7.2.2	Cas non-linéaire	58
7.3	Choix des paramètres	60
8	Les réseaux de neurones	61
8.1	Contexte	61
8.2	Neurone formel	61
8.3	Perceptron multicouche	64
8.4	Ajustement du réseau de neurones	66
8.4.1	Rétropropagation de l'erreur	66
8.4.2	Mise en place de l'algorithme	67
8.4.3	Résolution du problème	68
 IV Application du proxy pour l'estimation du capital économique		69
9	Application des chocs	70
9.1	Apprentissage sur le SCR	70
9.2	Apprentissage sur les sous-modules	73
9.2.1	SCR de marché brut	73
9.2.1.1	Choc de taux	74
9.2.1.2	Choc de spread	75
9.2.1.3	Choc actions	76
9.2.1.4	Choc immobilier	76
9.2.1.5	Agrégation des chocs	76
9.2.2	Calcul du SCR de marché net	77
9.2.3	SCR de souscription vie	77
9.2.3.1	Agrégation des chocs	78
9.2.4	SCR	78
10	Résultats d'estimations en t=0	80
10.1	Résultats obtenus pour le BE	82
10.2	Résultats obtenus pour le SCR	83
10.2.1	Module marché	83
10.2.2	Module vie	83
10.2.3	SCR	84
10.2.4	RC	84

V	Application de la méthode de proxy dans un cadre ORSA	85
11	Présentation du modèle ORSA	86
12	Mise en place du proxy en $t > 0$	88
12.1	Présentation d'un modèle ALM	88
12.2	Calcul du coefficient ALM	89
12.3	Projection des inputs	91
12.4	Résultats d'estimations	92
12.4.1	Module marché	92
12.4.2	Module de souscription-vie	93
12.4.3	RC	94
13	Limites des méthodes	95
VI	Pour aller plus loin	96
14	Hypothèses prospectives	97
14.1	Scénario de baisse des taux	97
14.2	Résultats d'estimations	98
	Conclusion générale	99
	Bibliographie	100
	Note de synthèse	102
	Summary	107
	Table des figures	112
	Liste des tableaux	114

Introduction

L'entrée en vigueur en 2016 de la réglementation Solvabilité 2 modifie les exigences quantitatives et qualitatives pour les compagnies d'assurances. Cette nouvelle réglementation comporte trois piliers :

- Le premier pilier définit les exigences quantitatives, notamment en matière de fonds propres et de calculs des provisions techniques
- Le deuxième pilier définit les exigences en matière de gouvernance des risques
- Le troisième pilier définit les exigences en termes d'informations prudentielles et de publication

Dans le cadre du pilier 2 de la directive Solvabilité 2, l'ORSA ou évaluation interne des risques et de la solvabilité (Own Risk and Solvency Assessment) se caractérise par la nécessité d'être en mesure de s'assurer en permanence de la couverture de son exigence de capital réglementaire. L'ORSA est défini à l'article 45 de la directive Solvabilité 2. Il s'agit d'un processus d'évaluation prospectif des risques et de la solvabilité permettant d'intégrer la dimension risque dans le pilotage de l'entreprise. Il est appréhendé sur la durée du business plan de l'entreprise pouvant s'étendre de 3 à 8 ans.

Ainsi, cette évaluation, permanente doit conduire les organismes à simuler dans le temps l'évolution de leur capital requis ou *SCR* (Solvency Capital Requirement) sur la base de scénarii d'évolution de facteur de risque. En effet, cette évaluation des risques doit tenir compte de l'ensemble des états de la nature dans le futur comme, des changements macroéconomiques (hausse ou baisse des taux d'intérêts) ou encore un changement règlementaire (évolution du pourcentage de redistribution des bénéfices) que nous appellerons des « possibles ».

L'exploration d'un nombre très élevé de « possibles » via des approches stochastiques rend les calculs complexes, à la fois en termes de temps de calculs mais aussi d'un point de vu mathématique, ce qui incite les compagnies d'assurance à envisager plusieurs approches de simplification afin d'évaluer leurs engagements. La mise en œuvre de méthodes simplifiées doit cependant tenir compte du profil de risque de la compagnie tout en arbitrant entre coût de mise en œuvre, précision attendue, maîtrise dans la justification des résultats et sensibilité aux hypothèses.

L'enjeu de ce mémoire consiste donc en l'estimation d'une fonction paramétrique à l'aide de techniques dite de « *machine learning* », permettant de projeter le ratio de couverture ou *RC* sur la durée du business plan, tout en ayant pour objectif la réduction du temps de calcul. De par sa définition, le *RC* est obtenu comme suit :

$$RC_t = \frac{FP_t}{SCR_t}$$

Ainsi, l'estimation du *RC* nécessite au préalable de pouvoir estimer les fonds propres *FP* ainsi que le *SCR* de notre compagnie d'assurance à chaque pas de projection *t*, variant de 0 à 6. Chacune de ces deux quantités étant dépendantes d'hypothèses économiques, nous les estimerons séparément afin d'en calculer le *RC*.

Pour ce faire, nous allons, dans le cadre de ce mémoire, mettre en place des fonctions dites « proxies » permettant d'estimer les indicateurs suivants :

- Les Best Estimate (BE)
- Les FP
- Les sous-modules de risques de la formule standard

Ces trois proxies seront calculées de la même manière, c'est-à-dire via des méthodes d'apprentissage statistique.

Nous distinguerons la fonction de proxy implémentée en $t = 0$ de celle qui sera mise en place en $t > 0$. En effet, entre t et $t + 1$, nos indicateurs de risque vont évoluer en même temps que l'environnement économique. Pour prendre en compte cette évolution, comme une diminution de la valeur de marché, nous estimerons un dernier coefficient, dit « coefficient ALM » que nous noterons δ dont le but sera de capter l'effet de la variation de la valeur de nos indicateurs économiques dans le futur. Le calcul de ce coefficient sera explicité dans la chapitre 5 lorsque nous présenterons le modèle ORSA.

Les chapitres 1 et 2 présentent le cadre d'étude du mémoire. Après avoir situé le cadre réglementaire et présenté les données utilisées dans notre étude, nous présenterons les méthodes d'apprentissages statistiques utilisées pour calibrer notre forme paramétrique dans le chapitre 3. Pour finir, le chapitre 4 s'attachera à présenter les résultats obtenus dans le cadre d'une estimation du *SCR* en vision pilier 1 tandis que le chapitre 5 proposera une application permettant d'estimer le *RC* dans le cadre prospectif de l'ORSA

Définition des mots-clés

SCR (Solvency Capital Requirement) : Exigence en capital de solvabilité nécessaire à la couverture des différents risques encourus par les compagnies d'assurance. Cette exigence en fonds propres correspond à une VaR à 1 an au seuil de 99.5%.

Valeur nette comptable (ou Prix de revient) : La valeur nette comptable est la valeur brute d'un actif minorée du montant des amortissements et/ou des provisions.

Valeur de marché : La valeur de marché de tout titre financier est le prix auquel il peut être acheté ou vendu, en particulier sur le marché secondaire.

Plus-ou-Moins-value latente :

- Les **moins-values latentes** sont des pertes potentielles et provisoires qui apparaissent lors de la valorisation des actifs qu'ils soient monétaires, financiers ou autres. Les moins-values latentes se déterminent en retirant du prix constaté sur un marché, à un instant donné, la valeur comptable de l'actif.
- Les **plus-values latentes** sont des profits potentiels donc non encore réalisés. Les plus-values latentes sont donc provisoires et restent théoriques. Elles apparaissent lors de la valorisation des actifs qu'ils soient monétaires, financiers ou autres. Les plus-values latentes se calculent en retirant du prix constaté sur un marché, à un instant donné, la valeur comptable de l'actif.

TMG (taux minimum garanti) : Le TMG correspond au rendement minimum servi par un contrat d'assurance-vie. C'est un taux contractuel que l'assureur s'engage à verser à l'assuré sur une durée déterminée dans le contrat.

Chargement sur encours : Le chargement est un montant prélevé par un assureur prenant la forme de prélèvement lors du versement de fonds par un assuré sur un contrat d'assurance (auto, habitation, vie,...). Le taux de chargement est le rapport entre les charges financières, ou les frais, avec les fonds collectés. Le chargement est donc un élément de rémunération essentiel pour les compagnies d'assurances.

Produits R332-20 : L'article R.332-2 du Code des assurances énumère les actifs pouvant venir en représentation des engagements des sociétés d'assurance. Parmi ceux-ci, le code distingue deux grandes catégories :

- Les obligations (art. R.332-19)
- Les autres actifs (art. R.332-20) comme les actions, immeubles, valeurs mobilières et prêts.

Best estimate (BE) : Le BE se traduit comme l'espérance des flux futurs sortants diminués des flux futurs entrants actualisés avec la courbe des taux sans risque à la date d'évaluation.

Première partie

Calcul du capital économique prospectif sous Solvabilité 2

Cadre réglementaire

Ce mémoire a pour but de proposer une méthode permettant de calculer le capital économique sous Solvabilité 2. Avant d'entrer plus en détail dans le cadre de cette étude, il est important de comprendre l'ancienne réglementation et les faiblesses de cette dernière ayant conduit à la réglementation actuelle.

1.1 De Solvabilité 1 à Solvabilité 2

La directive Solvabilité 1 datant de 1970 avait pour but de démontrer la capacité des assureurs à faire face à leurs engagements. Cette réglementation reposait sur 3 principes :

- 1 : évaluer prudemment les engagements
- 2 : représenter de ces engagements avec des actifs de bonne qualité
- 3 : avoir une marge de solvabilité

Nous proposons dans les parties suivantes d'effectuer un détail succinct sur chacun des principes.

1.1.1 Evaluation prudente des engagements

L'évaluation des engagements est représentée au niveau du bilan de la compagnie d'assurance. En effet, le bilan est une photographie du patrimoine de l'entreprise à une date donnée où nous pouvons différencier l'actif, qui représente ce que l'entreprise détient en portefeuille et le passif, qui représente les engagements de l'entreprise vis-à-vis de ses assurés.

Les actifs de la compagnie d'assurance sont comptabilisés selon le principe de coût historique, ce qui signifie à leur coût d'acquisition. La valeur des actifs n'a pas vocation à fluctuer dans le temps.

Au sein du passif, nous pouvons distinguer les fonds propres et les provisions techniques. Les fonds propres correspondent au capital social apporté par les actionnaires augmentés des bénéfices mis en réserve au fil des années tandis que les provisions techniques correspondent aux engagements envers les assurés.

La figure ci-dessous permet d'illustrer un bilan sous Solvabilité 1,

Actif	Passif
Actif en valeur comptable	Excédent de marge
	Exigence de marge de solvabilité
	Provisions techniques

} Fonds propres

FIGURE 1.1 – Bilan économique sous S1

1.1.2 Représentation des engagements

Le deuxième principe prévoit une représentation de ces engagements avec des actifs de bonne qualité. Ici, la notion de bonne qualité fait référence à des actifs sûrs, liquides et rentables. La réglementation prévoit une limite à détenir par classe d'actifs :

- 65% pour les actions
- 40% pour l'immobilier
- 10% pour les prêts

Il n'y a pas de plancher pour ces actifs.

1.1.3 Marge de solvabilité

L'article R334 du code des assureurs définit la marge de solvabilité comme le capital que doivent posséder les compagnies d'assurance pour faire face à toutes dépenses supplémentaires dues à des événements inattendus. Le calcul de la marge de solvabilité en assurance-vie se définit comme la somme de :

- 4% des provisions mathématiques (ou 1% si aucun risque de placement n'est supporté) x ratio de réassurance
- 0.1% à 0.3% des capitaux décès sous risque

où le ratio de réassurance se définit comme :

- PM nettes / PM brutes
- plancher à 85%

1.1.4 Limites de Solvabilité 1

Cependant, ce régime a été remis en question pour plusieurs raisons :

- Ce système prudentiel imposait aux assureurs de constituer des provisions techniques dites « suffisantes » pour faire face à leurs engagements sans tenir compte des risques cédés
- L'approche Solvabilité 1 était trop statique, reposant sur une vision comptable des engagements ce qui ne permettait pas de prendre en compte l'évolution du profil de risque de la compagnie d'assurance au cours du temps.
- Les interactions entre l'actif et le passif ne sont pas prises en compte.

Dans le but de corriger ces faiblesses, la réforme Solvabilité 2, entrée en vigueur au 1^{er} janvier 2016 s'est donnée pour objectif de mieux prendre en compte les risques inhérents à une compagnie d'assurance. Ainsi, Solvabilité 2 reposera sur une approche plus dynamique en se fondant sur des éléments prospectifs d'appréciation des risques : prise en compte de la valeur des actifs non plus de manière comptable mais en valeur de marché (valeur actuelle) ou encore l'introduction de la notion de « meilleur estimation » des provisions techniques. Cette bonne gestion permanente de la solvabilité et des risques de l'assureur se retrouvent au-travers des trois piliers de cette nouvelle réglementation que nous présentons par la suite.

1.2 Les 3 piliers de Solvabilité 2

1.2.1 Pilier 1

L'objectif du pilier 1 de la réforme Solvabilité 2 consiste à calculer les exigences quantitatives de la compagnie d'assurance. Sous Solvabilité 2, le bilan de la compagnie d'assurance est évalué selon le principe de « Juste Valeur » (Fair Value), permettant ainsi de refléter la richesse réelle de l'entreprise. Cela nécessite de valoriser les actifs en valeur de marché et les passifs en « Meilleure estimation ».

La figure ci-dessous présente un bilan économique sous Solvabilité 2.

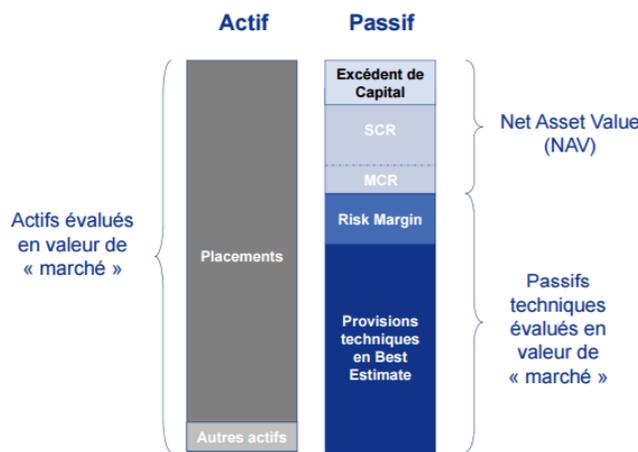


FIGURE 1.2 – Bilan économique sous S2

Les **provisions techniques** de ce bilan sont composées :

- Du **Best Estimate** : L'espérance des flux futurs sortants diminués des flux entrants calculé sous la probabilité risque-neutre¹. Cela revient à actualiser les flux avec la courbe des taux sans risque. Le BE se calcule comme suit :

$$BE = \mathbb{E}^Q \left(\sum_i \frac{F_i}{(1 + r_i)^i} \right)$$

Avec les notations suivantes :

- \mathbb{Q} représente la probabilité risque-neutre
- F_i représente le flux de passif à la date i (les engagements de l'entreprise vis à vis des assurés)
- r_i représente le taux sans risque
- De la **Marge pour Risque** : Il s'agit d'une marge complémentaire au BE calculée sur la base du SCR. Elle est destinée à garantir que la valeur des provisions techniques est équivalente au montant dont les entreprises auraient besoin pour reprendre et honorer les engagements.

1. On appelle probabilité risque-neutre toute probabilité \mathbb{Q} équivalente à \mathbb{P} qui rende martingales nos flux.

Le Best Estimate tient compte des prestations discrétionnaires futures versées aux assurés. Ainsi, les engagements de l'assureur peuvent être scindés en deux parties :

- La partie garantie : **BEG** (Best Estimate Garanti)
- La partie afférente aux prestations discrétionnaires futures : **FDB** (Future Discretionary Benefits)

La somme de ces deux éléments est notée **BEL** (Best Estimate Liabilities) et représente la meilleure estimation des engagements de l'assureur (hors marge de risque).

Quant à l'exigence en capital, elle se définit par :

- Le **MCR (Minimum Capital Requirement)** : représente le niveau de fonds propres en-dessous duquel les intérêts des assurés se verraient menacés si l'entreprise poursuivait son activité. Le MCR se situe entre 25% et 45% du SCR. En dessous de ce seuil, l'intervention du régulateur est immédiate.
- Le **SCR (Solvency Capital Requirement)** : capital nécessaire pour ne pas être en ruine à horizon 1 an avec une probabilité de 99.5%.

1.2.2 Pilier 2

A la différence du pilier 1, le pilier 2 de Solvabilité 2 a pour but de se concentrer davantage sur les exigences qualitatives. Cela consiste en la mise en place d'un processus de gestion des risques permettant à la compagnie de gérer et de définir l'ensemble des risques qu'elle est prête à prendre, assumer et gérer sur un horizon de temps donné. Ce pilier encourage les entreprises à instaurer une démarche dite ERM (Entreprise Risk Management) dont l'ORSA (Own Risk and Solvency Assessment) fait partie prenante. Nous détaillerons ce pilier dans la partie suivante.

1.2.3 Pilier 3

Enfin, le pilier 3 de Solvabilité 2 concerne les exigences de reporting à destination du public et du régulateur. Ce pilier a pour but d'améliorer, d'une part la transparence du secteur de l'assurance vis-à-vis du public en assurant un niveau d'informations suffisant aux assurés et aux acteurs financiers et d'autre part, de garantir un niveau de reporting suffisamment détaillé afin de permettre l'évaluation de la bonne gestion des risques de la part des compagnies d'assurance par les régulateurs.

La figure ci-dessous permet de résumer les trois piliers présentés précédemment.

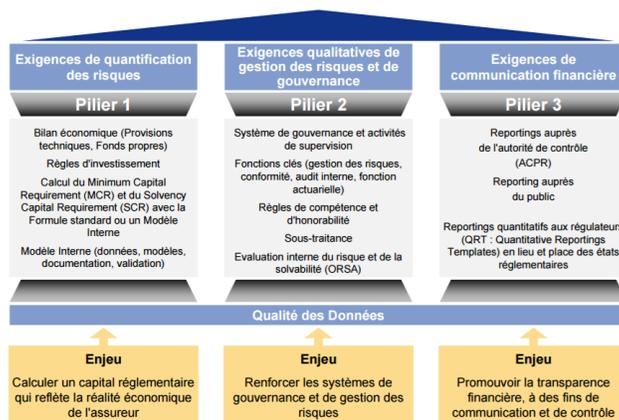


FIGURE 1.3 – Les piliers de Solvabilité 2

1.3 Calcul du capital économique sous Solvabilité 2

1.3.1 Définition

Sous Solvabilité 2, le capital économique est défini comme le montant de fonds propres (FP) dont un assureur doit disposer pour faire face à une ruine économique à 1 an au seuil de 99.5%. Autrement dit, l'assureur doit pouvoir faire face à un risque bicentenaire. D'un point de vue mathématique, cela se traduit à tout instant t comme suit :

$$\mathbb{P}(FP_t \geq 0) \geq 0.995$$

Le montant de capital économique peut être approximé par la différence entre les fonds propres en $t = 0$ et le quantile à 0.5% de la distribution des fonds propres en $t = 1$. Cette approximation est présentée dans l'article de LOISEL.S et DEVINEAU.L [5] et se réécrit de la façon suivante :

$$SCR = FP_0 - q_{0.5\%} \left(FP_1 \times e^{-\int_0^1 r(u) du} \right)$$

où r représente le taux de rendement de l'actif.

En supposant que l'actif évolue au taux zéro-coupon, la relation précédente se simplifie :

$$SCR \approx FP_0 - B(0, 1) \times q_{0.5\%} (FP_1)$$

avec $B(0, 1)$ la valeur actuelle d'une obligation zéro-coupon de maturité 1 an.

Pour calculer ce montant de FP , l'EIOPA propose deux approches :

- Une formule standard
- Un modèle interne

Nous allons présenter dans un premier temps la méthode de calcul du capital économique par formule standard avant de présenter la méthode par modèle interne.

1.3.2 Calcul du capital économique en formule standard

La formule standard a été développée par le régulateur afin de proposer une méthode de calcul qui soit, à la fois simple pour pouvoir s'adapter à tous les assureurs, et suffisamment complète pour prendre en compte l'ensemble des risques.

La formule standard est dite modulaire puisque chaque grand risque est divisé en module (risque de marché, risque de souscription vie etc) et chacun de ces modules est divisé en sous-module afin d'avoir une plus grande granularité et compréhension du risque auquel un assureur est exposé.

La figure ci-dessous présente l'architecture de la formule standard permettant d'obtenir le montant de capital économique.

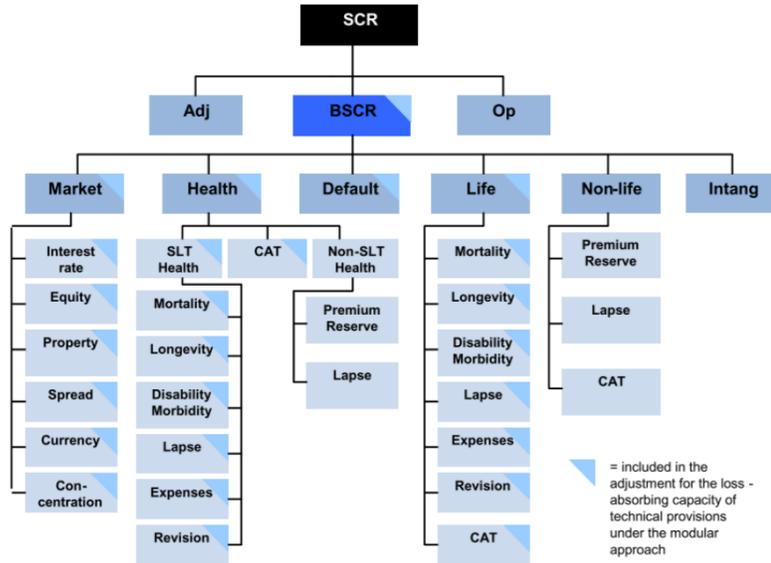


FIGURE 1.4 – Structure de la formule standard

Comme nous pouvons le constater sur la figure précédent, le capital économique sera calculé pour chaque sous-module de risque et sera obtenu comme suit :

$$C_R = FP_0^{central} - FP_0^{choc}$$

où :

- $FP_0^{central}$, correspondant au fond propre Solvabilité 2 initial de l'assureur
- FP_0^{choc} , correspondant au nouveau montant de fonds propres de solvabilité suite à l'application d'un choc

Un choc peut correspondre à une baisse des taux d'intérêts, une baisse du cours des actions ou encore un choc de mortalité instantané.

Une fois le besoin en capital au titre de chaque sous-module défini, ces derniers sont agrégés via une matrice de corrélation réglementaire afin de prendre en compte la dépendance entre chaque sous-risque (par exemple la dépendance entre le risque de taux d'intérêts et le risque actions) comme suit :

$$SCR_m = \sqrt{\sum_{i,j} \rho_{(i,j)}^{R_m} C_i C_j}$$

où :

- SCR_m correspond au capital associé au module m
- C_i correspond au capital associé au risque i
- ρ^{R_m} correspond à la matrice de corrélation des sous-modules de risque considérés

Enfin, une agrégation inter-modulaire est effectuée afin de prendre en compte la corrélation et l'effet de diversification entre les différents modules de la formule standard et d'obtenir le BSCR (Basic Solvency Capital Requirement) :

$$BSCR = \sqrt{\sum_{i,j} \rho_{(i,j)}^M SCR_i SCR_j}$$

Cette formule présente l'avantage d'être simple à implémenter mais fait de grosses hypothèses concernant l'agrégation des risques. Premièrement, la matrice de corrélation est déterministe, elle ne permet donc pas de refléter le montant réel de corrélation entre les risques d'un assureur donné. Deuxièmement, cette corrélation est linéaire ce qui suppose que le SCR se diffuse linéairement avec le risque. Troisièmement, cette formule suppose que les facteurs de risques sont elliptiques, ce qui traduit une dépendance en queue très faible et ne permet donc pas de mesurer les risques extrêmes.

Pour pallier ces défauts, les assureurs ont la possibilité de développer leur propre modèle interne afin de modéliser plus finement leur structure de risque.

1.3.3 Calcul du capital économique en modèle interne

La directive Solvabilité 2 prévoit que les assureurs puissent utiliser leur propre modèle interne mais celui-ci doit être au préalable validé par le régulateur. Ce modèle interne présente l'avantage d'avoir été construit de sorte à prendre en compte le profil de risque de l'assureur, ce qui le rend plus fidèle que la formule standard. A la différence de la formule standard, le modèle interne cherche à obtenir une distribution des fonds propres en $t = 1$ et, d'en calculer la probabilité que l'assureur se retrouver en ruine.

La méthode dite des « Simulations dans les simulations » (SdS) apparaît comme la méthode la plus conforme pour calculer le capital économique sous Solvabilité 2. La méthode SdS s'effectue en deux étapes :

- Projection des différents facteurs de risque de l'assureur (courbe des taux, action, immobilier etc) à l'horizon 1 an dans une vision monde réel, ce qui signifie que les facteurs de risques sont projetés conformément à ce qui a été observé les années précédentes afin d'être réaliste. Ces différentes projections permettent d'obtenir les bilans économiques observés dans 1 an pour chacun des scénarios, que nous appellerons « Simulation primaires ».
- En fin de première période, les bilans économiques sont calculés de manière « Market consistent », ce qui signifie qu'ils doivent être conformes aux observations de marché. Ainsi, nous obtenons les FP économiques pour chaque simulation primaire grâce à des simulations secondaires en univers risque neutre (les actifs sont martingales : ils évoluent au taux sans risque). Ces projections seront appelées « Simulation secondaires ».

La figure ci-dessous permet d'illustrer la méthode SdS.

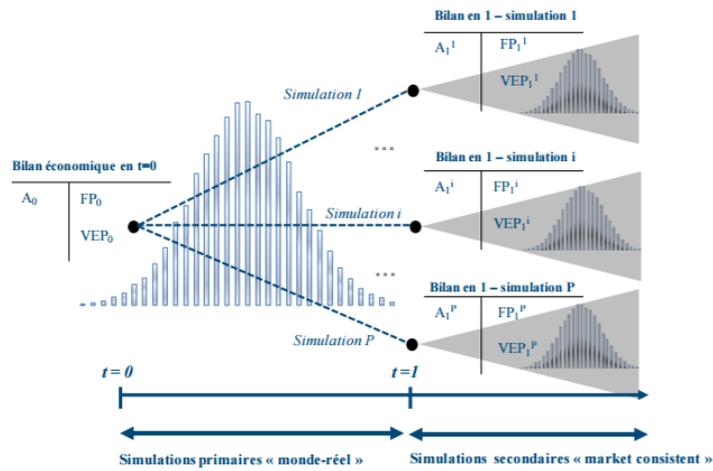


FIGURE 1.5 – Méthode de Simulation dans les Simulations

La précision des résultats est très sensible à la complexité du portefeuille d'assurance et aux nombres de simulations primaires (noté P) et secondaires (noté S). Le nombre de simulation total sera égal au produit des simulations primaires et secondaires. De ce fait, augmenter le nombre de simulation P ou S augmente considérablement le temps de calcul, ce qui rend son utilisation très compliquée en pratique.

Focus pilier 2 : le processus ORSA

2.1 Approche générale

L'ORSA ou évaluation interne des risques et de la solvabilité (Own Risk and Solvency Assessment) est défini à l'article 45 de la directive Solvabilité 2. Il s'agit d'un processus d'évaluation prospective des risques et de la solvabilité permettant d'intégrer la dimension risque dans le pilotage de l'entreprise. Il est appréhendé sur une durée de 3 ou 8 ans et peut correspondre à celle du business plan de l'entreprise. L'ORSA vise à évaluer les trois dimensions suivantes :

— **Besoin Global de Solvabilité (BGS) :**

Le BGS n'est pas défini par la réglementation et est propre à chaque entreprise. Il peut inclure l'ensemble des risques se trouvant dans la formule standard et/ou hors de la formule standard. Parmi les risques les plus fréquemment inclus dans le BGS, nous pouvons par exemple retrouver le risque souverain ou encore le risque de réputation. Le BGS peut inclure le capital de solvabilité requis (SCR), mais l'analyse des risques doit être complétée par d'autres évaluations. Le résultat du BGS doit se concrétiser par un niveau des moyens nécessaires pour faire face aux risques importants (capital ou autres) définis par la compagnie.

— **Exigence en capital :**

Au vu des risques portés par chaque assureur, la réglementation définit un montant en fonds propres réglementaires, qui doit être couvert à tout moment par des fonds propres reconnus conformes par la réglementation.

— **Adéquation des mesures de risque :**

Cette étape vise à cartographier l'ensemble des risques auxquels l'entreprise est exposée. Elle a pour but d'évaluer si son profil de risque s'écarte des hypothèses qui sous-tendent le calcul du SCR et si ces écarts sont significatifs.

L'ORSA prévoit également la définition de l'appétence aux risques (niveau maximum de risques qu'une entreprise est prête à prendre en vue d'atteindre ses objectifs) ainsi que sa déclinaison en tolérances aux risques (définition des différents indicateurs de performance) et limites opérationnelles (éventuelle limite de consommation en capital par branche d'activité).

Le processus ORSA peut se résumer à l'aide du tableau ci-dessous.

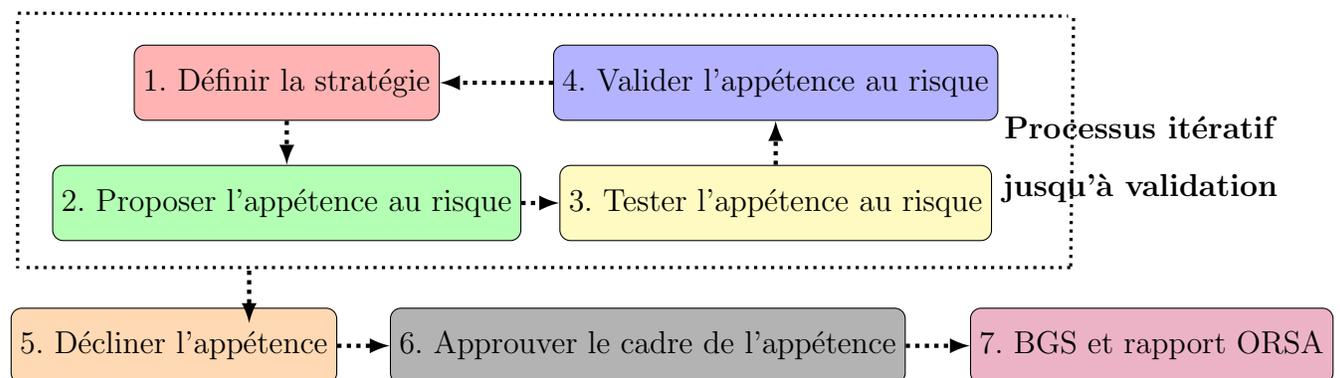


FIGURE 2.1 – Processus ORSA

2.2 Dimension prospective du capital économique

Cette évaluation permanente sur la durée du business plan doit conduire les organismes à simuler dans le temps l'évolution de leur SCR sur la base de scénarii d'évolution de facteurs de risque. En effet, cette évaluation des risques doit tenir compte de l'ensemble des états de la nature dans le futur comme, des changements macroéconomiques (hausse ou une baisse des taux d'intérêts) ou encore un changement réglementaire (changement du pourcentage de redistribution des bénéfices) que nous appellerons des « possibles ».

Ainsi, au terme de la démarche ORSA, le Besoin Global en Solvabilité est calculé afin de permettre à la compagnie de savoir si son niveau de fonds propres ou les actions qui seront entreprises pour lever les fonds nécessaires seront suffisants compte tenu de ses objectifs.

Par ailleurs, l'ORSA doit démontrer la capacité de l'entreprise à répondre aux exigences réglementaires sur toute la durée du business plan, par exemple, avoir un ratio de couverture au moins supérieur à 100%. En effet, le ratio de couverture de l'année t se définit comme :

$$RC_t = \frac{FP_t}{SCR_t}$$

Si le RC_t est trop faible, cela signifie que les fonds propres ne sont pas suffisants pour couvrir l'ensemble des risques de l'entreprise compte tenu de son profil de risque. Une des actions de l'entreprise pourrait consister en une modification de son allocation d'actifs afin d'en diminuer son exposition.

Une autre mesure permettant d'influer sur le ratio de couverture serait de modifier la politique de *management actions*. Les *management actions* sont des leviers d'actions de la part des assureurs qui, dans le cadre des calculs réglementaire de Solvabilité 2, permettent d'agir à la suite d'une situation défavorable conduisant à une dégradation de son profil de risque. Parmi les actions possibles, les assureurs peuvent envisager de :

- mettre en place un programme de réassurance
- modifier son algorithme de revalorisation de participations aux bénéfices discrétionnaires
- modifier les conditions de déclenchement de réalisation de plus ou moins-values latentes

Alternative au calcul du capital économique

Comme explicité en fin de première partie, la mise en place d'un SdS est très coûteuse en temps de calcul, c'est la raison pour laquelle, d'autres méthodes ont été développées afin de réduire le temps de calcul du capital économique.

La littérature existante recense deux grandes familles de méthodes :

- Le premier type de méthode vise à réduire le nombre de simulation primaires et secondaires des SdS. En effet, l'idée de cette méthode consiste à remarquer qu'il n'est pas nécessaire de connaître l'ensemble de la distribution des fonds propres à un an mais uniquement le quantile à 0.5% de cette distribution. Il s'agit donc de déterminer ce quantile en utilisant le moins de simulations possibles. Parmi ces méthodes, nous pouvons citer la méthode connue sous le nom « d'accélérateur SdS », développé par LOISEL.S et DEVINEAU.L [5].
- Le second type de méthode consiste à estimer une fonction de proxy des FP .

La méthode dite « d'accélérateur SdS » ne sera pas présentée dans le cadre de ce mémoire puisque nous cherchons à estimer une fonction de proxy du BE et de nos sous-modules de risque. Le cadre de ce mémoire s'inscrit dans les méthodes de deuxième type. Pour plus de détail sur la première méthode, le lecteur peut se référer à l'article « Construction d'un algorithme d'accélération de la méthode des simulations dans les simulations pour le calcul du capital économique Solvabilité II » de LOISEL.S et DEVINEAU.L [5].

3.1 Méthode se basant sur un proxy

Nous allons dans cette partie présenter les méthodes se basant sur une fonction de proxy permettant de calculer la distribution des FP en $t=1$. Le but des méthodes paramétriques est de pouvoir inférer la valeur des FP économiques à l'aide de fonctions de base ne dépendant que des facteurs de risques élémentaires représentant les aléas auxquels ils sont exposés. Cela se traduit de la façon suivante :

$$FP_1 = f(\epsilon_1, \dots, \epsilon_n)$$

où :

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \epsilon &\longmapsto p(\epsilon) \end{aligned}$$

avec :

$p(\epsilon)$ une fonction polynomiale
 ϵ_i le i -ème facteur de risque

Parmi les méthodes existantes permettant de calibrer notre fonction f , nous pouvons citer les deux suivantes :

- Le Curve-fitting : la forme paramétrique est calibrée sur les calculs de fonds propres obtenus en $t=1$ sur un nombre restreint de simulation primaire via la méthode SdS
- Le Least-Square Monte Carlo : la forme paramétrique est calibrée sur les calculs de fonds propres obtenus en $t=1$ sur un grand nombre restreint de simulation primaire et une seule simulation secondaire via la méthode SdS

Dans la suite, nous présenterons plus en détail ces deux méthodes.

3.1.1 Curve-fitting

La méthode curve-fitting consiste à calibrer une fonction paramétrique des fonds propres économiques en $t = 1$ en fonction des différents facteurs de risque. Le calibrage de cette fonction se fait en plusieurs étapes :

- **Etape 1** : On effectue un nombre restreint de simulations primaires \tilde{P} avant d'utiliser l'algorithme SdS sur ces scenarii.
- **Etape 2** : On calibre une forme paramétrique sur la valeur des fonds propres obtenus à l'étape 1 par minimisation de l'écart quadratique entre la valeur du proxy et la valeur des \tilde{P} fonds propres obtenus.
- **Etape 3** : Le reste de la distribution est calculé à l'aide de la forme paramétrique à partir des facteurs de risques des simulations primaires non utilisées précédemment.
- **Etape 4** : Calcul des FP et du quantile à 0.5%

Le mémoire de HAGUET.E [7] permet de comparer la précision des résultats d'une telle méthode. Cela permet de réduire considérablement le nombre de simulation tout en ayant des résultats très acceptables : le quantile est estimé avec une erreur inférieure à 1%.

La figure ci-dessous permet d'illustrer la méthode curve-fitting.

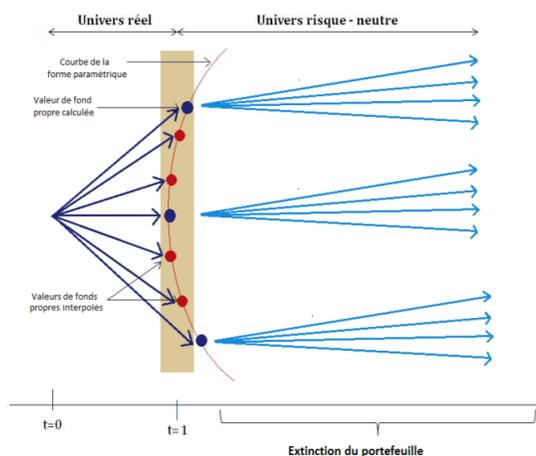


FIGURE 3.1 – Méthode curve-fitting

3.1.2 Least Square Monte-Carlo

A la différence de la méthode curve-fitting, la méthode Least Square Monte-Carlo permet d'obtenir un proxy en calibrant notre forme paramétrique non plus sur les fonds propres mais sur les VAN (valeur actuelle nette) de marge. En effet, nous pouvons remarquer une redondance d'information dans nos simulations. Il convient de remarquer que, pour des simulations primaires proches, les simulations secondaires attribuées posséderont des lois proches, c'est la raison pour laquelle, une seule simulation secondaire sera effectuée pour chaque simulation primaire et ce, pour un grand nombre de simulations primaire. Le calibrage se fera suivant les étapes ci-dessous :

- **Etape 1 :** Un grand nombre de scénarii primaires \tilde{P} est généré et se verra attribuer une unique simulation secondaire. Pour chaque simulation primaire, les VAN de marge seront calculées.
- **Etape 2 :** Une forme paramétrique sera calibrée par minimisation de l'écart quadratique entre la valeur du proxy et la valeur des VAN de marge.
- **Etape 3 :** Calcul des FP et du quantile à 0.5%.

A nouveau, le mémoire de HAGUET.E [7] permet de comparer la précision des résultats d'une telle méthode. Elle arrive à la conclusion que cette méthode permet de réduire considérablement le temps de calcul puisque le nombre de scénarii secondaires a été drastiquement réduit tout en gardant des résultats acceptables : le quantile est estimé avec une erreur inférieure à 1%.

La figure ci-dessous permet d'illustrer la méthode Least Square Monte-Carlo.

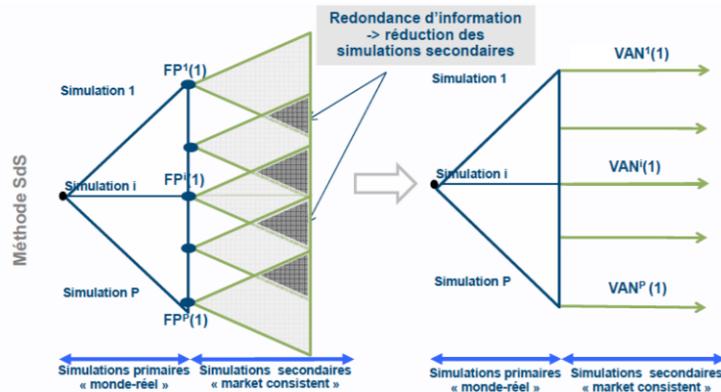


FIGURE 3.2 – Méthode Least Square Monte-Carlo

3.2 Calcul d'un proxy par *machine learning*

Nous venons de présenter les méthodes existantes se fondant sur une méthode de proxy. A ce jour, ces méthodes tentent d'approximer la distribution des fonds propres afin d'en inférer la valeur du *SCR*.

Nous allons dans le cadre de ce mémoire mettre en place une nouvelle fonction de proxy, non plus sur la distribution des fonds propres mais sur les indicateurs suivants :

- Les BE
- Les FP
- Les sous-modules de risques de la formule standard

Ces trois fonctions de proxy seront calculées de la même manière, c'est à dire via des méthodes dites de *machine learning* et auront la forme suivante :

$$y_t = f_t(\epsilon_1, \dots, \epsilon_n)$$

où :

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \epsilon &\longmapsto p(\epsilon) \end{aligned}$$

avec :

y la variable d'intérêt
 $p(\epsilon)$ une fonction polynomiale
 ϵ_i le i -ème facteur de risque

Cette approximation repose uniquement sur des facteurs de risques élémentaires représentant les aléas auxquels les assureurs sont soumis et susceptibles d'impacter les *BE*, les *FP* ainsi que les sous-modules de risques de la formule standard (changement d'environnement macroéconomique, réglementaire etc).

Contrairement aux différentes méthodes que nous venons de présenter, l'apprentissage statistique va nous permettre d'extraire, à partir des informations économiques de nos bilans, des relations entre nos facteurs de risques afin d'en inférer une valeur d'engagement.

La figure ci-dessous présente la structure simplifiée d'un bilan d'assurance sous Solvabilité 2.

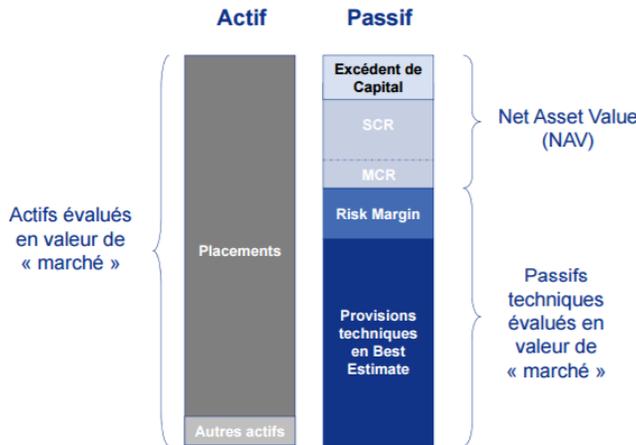


FIGURE 3.3 – Bilan économique sous S2

Comme nous pouvons le constater, dans le cadre d'un bilan simplifié nous pouvons réécrire :

$$FP = VM - BE$$

Or, nous savons que le SCR correspond à la perte de fonds propres à la suite d'un choc instantané, ce qui nous permet de le définir comme :

$$\begin{aligned} SCR_t^{choc} &= FP_t^{central} - FP_t^{choc} \\ &= (VM_t^{central} - BE_t^{central}) - (VM_t^{choc} - BE_t^{choc}) \\ &= (VM_t^{central} - VM_t^{choc}) - (BE_t^{central} - BE_t^{choc}) \\ &= \Delta NAV^{VM} - \Delta NAV^{BE} \end{aligned}$$

La partie actif de notre bilan peut se calculer aisément à l'aide de formules fermées. En effet, la variation de valeur de marché d'une classe d'actifs correspond tout simplement à une variation de valeur de marché à la suite du choc considéré (cette partie sera présentée plus en détail au chapitre 4). C'est sur cette simplification que les méthodes existantes reposent.

Pour rappel, l'enjeu de ce mémoire consiste en l'estimation d'une fonction de proxy par apprentissage statistique, dans le but de pouvoir projeter le RC sur la durée du business plan de la compagnie d'assurance. De part sa définition, le RC est obtenu comme suit :

$$RC_t = \frac{FP_t}{SCR_t}$$

Ainsi, l'estimation du RC nécessite au préalable de pouvoir estimer les FP ainsi que le SCR de notre compagnie d'assurance. Chacune de ses deux quantités étant dépendantes d'hypothèses économiques, nous les estimerons séparément afin d'en calculer le RC . La façon dont seront calibrées ces fonctions sera présentée par la suite.

Cependant, pour construire de telles fonctions, des données sont nécessaires afin de pouvoir, d'une part calibrer notre fonction et d'autre part prédire de nouveaux SCR et FP suite à un changement d'environnement économique. La partie suivante s'attache à présenter la façon dont nous avons construit notre base de données.

Nous distinguerons la fonction de proxy implémentée en $t = 0$ et celle qui sera mise en place en $t > 0$. En effet, entre t et $t + 1$, nos indicateurs de risques vont évoluer en même temps que l'environnement économique. Pour prendre en compte cette évolution, comme par exemple une diminution de la valeur de marché, nous estimerons un dernier coefficient, dit « coefficient ALM » que nous noterons δ dont le but est de capter l'effet de la variation de la valeur de nos indicateurs économiques dans le futur. Le calcul de ce coefficient sera explicité dans la chapitre 5 lorsque nous présenterons le modèle ORSA.

Finalement, nous obtiendrons la relation suivante :

$$\begin{aligned} \tilde{S\tilde{C}R}_t &= f_t(\epsilon_1, \dots, \epsilon_n) \mathbb{1}\{t = 0\} + f_{t-1}(\epsilon_1, \dots, \epsilon_n) \delta_{t-1}(\epsilon_1, \dots, \epsilon_n) (1 - \mathbb{1}\{t = 0\}) \\ \tilde{F\tilde{P}}_t &= f'_t(\epsilon_1, \dots, \epsilon_n) \mathbb{1}\{t = 0\} + f'_{t-1}(\epsilon_1, \dots, \epsilon_n) \delta'_{t-1}(\epsilon_1, \dots, \epsilon_n) (1 - \mathbb{1}\{t = 0\}) \end{aligned}$$

où ϵ_i correspond aux facteurs de risques et $f_t(\cdot)$, $f'_t(\cdot)$, $\delta_t(\cdot)$ et $\delta'_t(\cdot)$ correspondent à nos fonctions de proxies obtenues par apprentissage statistique.

La relation précédente fait apparaître l'importance de bien pouvoir calibrer la fonction de proxy en $t = 0$ puisque, à l'instant $t > 0$, nos indicateurs représenteront simplement une proportion du $\tilde{S\tilde{C}R}_0$ et des $\tilde{F\tilde{P}}_0$.

Finalement, nous obtiendrons le ratio de couverture :

$$\tilde{R\tilde{C}}_t = \frac{\tilde{S\tilde{C}R}_t}{\tilde{F\tilde{P}}_t}$$

Deuxième partie

Sélection des facteurs de risques pertinents au calcul du capital économique

Description de la base

Les algorithmes dit de *machine learning* sont de plus en plus utilisés dans le but de trouver des relations entre des variables d'intérêts et d'en inférer une valeur de sortie. Cependant, ces méthodes ne peuvent donner de bons résultats dans des situations qui n'ont jamais été expérimentées. En effet, ils apprennent sur un ensemble de données, trouvent des relations entre les variables et, suite à l'apprentissage qui a été fait de ces relations, arrivent à prédire la variable d'intérêt. La constitution de la base qui servira d'apprentissage à ces méthodes est donc un enjeu crucial, c'est ce que nous présenterons dans cette partie, au travers des hypothèses de notre portefeuille, et plus précisément la composition de notre bilan.

Le portefeuille utilisé dans ce mémoire représente une compagnie d'assurance fictive possédant un unique contrat d'épargne vie, représenté par un unique model point. On entend par model point, un outil simplifié permettant de réaliser rapidement des calculs portant, dans notre cas, sur un portefeuille « moyen ». Le choix des valeurs que peuvent prendre l'ensemble des variables d'actifs et de passifs ont été fait de manière cohérente avec nos observations de plusieurs portefeuilles d'assurance-vie sur le marché. En effet, nous avons choisi d'être le plus généraliste possible de sorte à représenter les tendances du marché mais aussi de sorte à capter les situations extrêmes.

4.1 Hypothèses liées à l'actif

Le portefeuille d'actif a été construit de telle sorte à être représentatif d'une société d'assurance vie de taille moyenne sur le marché. Ainsi, il aura une valeur de marché de 2 milliards d'euros et sera réparti comme suit :

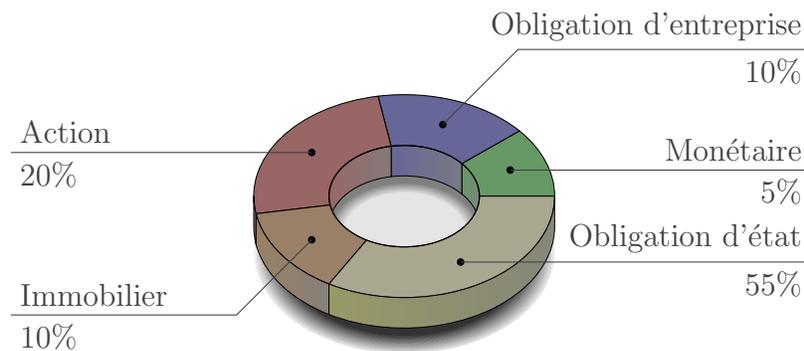


FIGURE 4.1 – Décomposition du portefeuille d'actif

Actif	Valeur de marché	Valeur comptable	% PMVL
Action	400	348	3%
Immobilier	200	174	1%
Monétaire	100	100	0%
OTF	975	945	3%
OAT	325	307	7%
Total	2 000	1 874	14%

TABLE 4.1 – Composition de l'actif en million d'€

L'objectif étant de construire une base de données comprenant les variables auxquelles le capital économique est le plus sensible, nous allons dans un premier temps présenter les caractéristiques des variables composant la partie actif de notre bilan ainsi que les bornes dans lesquelles ces variables peuvent varier. Par hypothèse, les plus ou moins-values latentes concernent uniquement les actifs nommés dans le code des assurances par articles R332-20 (action et immobilier). Comme nous pouvons le voir dans le tableau ci-dessous, la part de chacun des actifs en portefeuille se situe entre les bornes minimum et maximum. Par exemple, nos portefeuilles d'actifs seront composés à minima de 5% et d'au plus 20% d'actions.

Actif en % de la VM totale	Minimum	Maximum
% Actions	5%	20%
% Immobilier	5%	20%
% Monétaire	0%	10%
% Obligataire	60%	85%
Taux 1 ans	0.055%	0.55%
Taux 8 ans	0.61%	2.99%
Taux 20 ans	1.14%	5.79%
% R332-20	11%	38%

TABLE 4.2 – Variables d'actifs

D'autre part, nous avons décidé de considérer dans nos variables d'intérêt trois maturités de taux d'intérêts pour modéliser le changement d'environnement macroéconomique :

- le taux 1 an
- le taux 8 ans
- le taux 20 ans

Le choix de vouloir considérer trois horizons de taux d'intérêts pour expliquer la variation du *SCR*, des *FP* et *in fine* du *RC* répond à notre besoin de pouvoir estimer, quel que soit la situation économique, notre *RC* à horizon donné.

En effet, le taux 1 an permet de prendre en compte les chocs de court-terme, tandis que le taux 8 ans correspond à la durée de l'actif.

La durée d'un instrument financier est la durée de vie moyenne (durée entre la date actuelle et la date de paiement) des flux financiers, pondérée par leur valeur probable actualisée. La durée sert de mesure de risque : plus la durée est élevée et plus le risque est important. La durée de l'actif est donc une variable pertinente dans le cadre d'une politique de gestion des risques telle que prévue par l'ORSA.

Pour finir, le taux 20 ans permet de prendre en compte les risques lointains. L'année 20 correspond au dernier point liquide (Last Liquid Point) de la courbe des taux. Par convention, l'EIOPA retient la courbe des taux « type zéro coupon » construite à partir des taux swaps comme référence pour actualiser les flux futurs. Au-delà de 20 ans, la courbe des taux est extrapolée via la méthode de Smith-Wilson pour converger vers le niveau d'UFR.

Enfin, les PMVL sur les articles R 332-20 constituent une variable clé dans la variation du *SCR*. Comme nous l'avons dit précédemment, cette variable constitue un outil de *management action* qui, lors de contexte économique défavorable, s'avère utile dans l'atteinte du taux d'intérêt à servir aux assurés. En effet, il peut arriver que le taux de revalorisation servi soit inférieur au taux attendu par les assurés. Dans ce cas, la compagnie d'assurance peut réaliser des plus-values latentes dans le but d'honorer ses engagements vers ses assurés et redistribuer plus qu'elle n'aurait pu le faire par ailleurs.

Nous terminons la présentation de notre base d'actif en présentant dans le tableau ci-dessous, les caractéristiques des obligations (OAT et OTF).

Type de l'obligation	Nominal	Valeur de marché	Taux de coupon	Spread	Duration	Maturité	Notation
OAT	71 488 271	381 781 150	3.5%	0%	1.50	1.97	1
OAT	67 767 206	381 781 150	3.5%	0%	2.49	3.81	1
OAT	65 153 398	381 781 150	3.5%	0%	3.47	5.56	1
OAT	63 371 657	381 781 150	3.5%	0%	4.43	7.20	1
OAT	62 127 565	381 781 150	3.5%	0%	5.37	8.74	1
OTF	223 369 653	1 145 343 450	3.0%	1.60%	1.50	1.97	4
OTF	214 330 200	1 145 343 450	3.5%	1.40%	2.48	3.81	3
OTF	203 482 919	1 145 343 450	4.0%	1.20%	3.44	5.49	2
OTF	191 338 606	1 145 343 450	4.5%	1.00%	4.38	6.99	2
OTF	178 188 538	1 145 343 450	5.0%	0.80%	5.31	8.33	1

TABLE 4.3 – Caractéristiques des obligations

4.2 Hypothèses liées au passif

L'assurance vie se caractérise par la possibilité d'offrir deux types de contrats aux assurés :

- **Contrat d'assurance-vie** : Ces contrats ont pour but la constitution d'une épargne et le versement d'un capital ou d'une rente si l'assuré est toujours en vie au terme du contrat
- **Contrat d'assurance décès** : Ces contrats ont pour but de verser un capital ou une rente à un bénéficiaire désigné dans le contrat en cas de décès de l'assuré

Notre passif sera constitué d'un seul contrat d'épargne-vie. Un contrat d'épargne vie est un contrat qui garantit à un bénéficiaire, en contrepartie d'une prime, le versement d'un capital ou d'une rente en cas de décès de l'assuré. Ces contrats sont très répandus en France du fait d'une fiscalité et d'un rendement avantageux. Le capital perçu par le bénéficiaire correspond à l'encours versé, revalorisé par des intérêts techniques et des participations aux bénéfices (PB), diminué des frais liés à la gestion de ce contrat (taux de chargement sur encours, frais d'acquisitions et de gestion).

L'Article A331-3 du code des assurances impose à l'assureur de reverser aux assurés au minimum 90% des résultats techniques et 85% des résultats financiers. Le résultat financier de l'assureur reste majoritairement supérieur aux résultats techniques et correspond aux bénéfices réalisés sur les placements de l'assureur. Ce montant doit être reversé sous 8 ans.

De la même façon que pour l'actif, nous avons identifié les variables suivantes comme étant les plus caractéristiques de la variation du passif :

Variable	Minimum	Maximum
Age de l'assuré	31	51
TMG	1%	1.5%
Taux de chargement	0.6%	1%
TMG-Taux 8 ans	-0.11%	0.88%

TABLE 4.4 – Variables de passif

Nous avons choisi de créer une variable supplémentaire pour notre apprentissage : la variable mesurant la différence entre le TMG et le taux 8 ans nommée **TMG-Taux8ans**. Cette variable est censée capturer l'effet potentiel de rachat des contrats de la part des assurés. En effet, si le taux servi par les entreprises est inférieur au taux de rendement que les assurés peuvent obtenir par ailleurs, ils auront tendance à racheter les contrats obligeant ainsi les entreprises à venir piocher dans leurs fonds propres. Cette variable devrait avoir un effet significatif sur la variation du *SCR* en cas de changement d'environnement économique.

Construction d'une base d'apprentissage

Comme évoqué dans la partie précédente, nous cherchons à déterminer une fonction de proxy afin de pouvoir estimer le RC . Pour ce faire, nous devons créer une base de données permettant de refléter l'ensemble des événements possibles.

5.1 Hypercube latin

Dans cette partie, nous allons traiter de la construction de la base de données qui sera par la suite utilisée pour déterminer les variables jugées significatives ou contribuant le plus à la variation du BE , des FP et des sous-modules de risques. Les variables que nous allons tester sont les suivantes :

- Le taux de chargement sur encours
- le TMG
- L'âge de l'assuré
- La proportion d'OAT
- La proportion d'OTF
- La proportion d'actions
- La proportion d'immobilier
- La proportion de monétaire
- Les taux d'intérêt 1,8 et 20 ans
- Le pourcentage d'articles R332-20
- La différence entre le TMG et le taux 8 ans
- La différence entre le TMG et le taux 8 ans maximum

Les variables précédentes ont été choisies car elles sont celles qui ont été jugées les plus susceptibles d'impacter le capital économique d'une société d'assurance vie compte tenu du contexte actuel. En effet, les caractéristiques des assurés vont déterminer le montant de capital nécessaire pour être solvable.

Un assuré âgé aura une probabilité plus forte de décéder, ce qui engendrera des flux sortant de la part de la compagnie d'assurance. De même, à TMG fixé, une hausse des taux d'intérêts va augmenter la probabilité de rachat des contrats des assurés dans le portefeuille de la compagnie, ce qui nécessitera à nouveau une sortie d'argent. En effet, si les taux servis sur le marché sont plus intéressants que le rendement que le contrat procure, les assurés vont être incités à racheter leur contrat afin de placer leur argent sur d'autres contrats du marché proposant un meilleur rendement.

Les résultats de l'analyse qui sera menée par la suite à l'aide de différentes techniques d'apprentissages statistiques sont très sensibles aux choix des données en entrée, c'est la raison pour laquelle, il est important de construire une base représentant le plus la variabilité de nos données. En effet, les algorithmes de *machine learning* ne peuvent prédire/donner de bons résultats sur des situations qui n'ont jamais été expérimentées. Ils vont apprendre sur un échantillon d'apprentissage et tenter de prédire la valeur du BE , des FP ou des sous-modules de risques basés sur une situation similaire qui a déjà été rencontrée.

Nous souhaitons donc trouver une base de dimension réduite avec nos 14 variables mais qui soit la plus répartie possible dans l'espace tout en optimisant le temps de calcul. La méthode de Monte-Carlo fut la méthode la plus utilisée dans le cadre d'échantillonnage de données. Considérons un espace de dimension d et N variables et notons par x la matrice associée à ces données de la façon suivante :

$$x = \begin{pmatrix} x_{11} & \dots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{N1} & \dots & x_{Nd} \end{pmatrix}$$

Plusieurs alternatives à la méthode de Monte Carlo ont été développées, non seulement pour accélérer la convergence des estimateurs mais aussi pour s'assurer que les domaines d'entrée des paramètres soient bien échantillonnés. En effet, choisir arbitrairement quelques points au hasard dans un espace de grande dimension peut conduire à sous représenter certaines dimensions. La méthode qui sera présentée ici pour s'assurer de la couverture de l'ensemble de l'espace est la méthode des Hypercubes Latins (Latin Hypercube Sampling ou LHS) introduite par McKay et al [14].

Cette méthode consiste à évaluer numériquement les intégrales multiples en s'assurant de la non-redondance de l'information au travers d'une projection sur les axes factoriels. Pour ce faire, l'intervalle de chaque entrée x_i est découpé en N segments de probabilité égale à $\frac{1}{N}$ ce qui nous donne le découpage suivant : $\{[0, \frac{1}{N}], [\frac{1}{N}, \frac{2}{N}], \dots, [\frac{N-1}{N}, 1]\}$.

Par la suite, une valeur est tirée aléatoirement dans chacun des N segments. Une fois qu'un point a été tiré aléatoirement dans un des segments de x_i , aucun nouveau point ne peut y être placé, assurant ainsi que chaque segment soit représenté une seule fois. Les N valeurs ainsi obtenues pour x_1 sont combinées aléatoirement aux N valeurs obtenues pour x_2 et forment ainsi une matrice de taille $N \times 2$. De la même manière, les N valeurs obtenues pour x_3 sont accolées à cette matrice et ainsi de suite jusqu'à la d^e variable. Cet échantillonnage sera effectué dans une espace de dimension d .

La définition mathématique d'un hypercube latin est la suivante :

Definition Un hypercube latin à M points sur $[a, b]^d$ est défini sur l'ensemble de points x^i tel que :

$$x_i^j = \frac{\pi_j^i + U_j^i}{M}, \quad i \in [1, n], \quad j \in [1, d]$$

où π_j est une permutation de $\{1, \dots, M\}$ et $U_j^i \sim U[a, b]$ est une valeur aléatoire de distribution uniforme sur $[a, b]$

Ainsi, $\pi_j(i)$ représente la cellule dans laquelle se trouve le point x^i et U_j^i désigne l'endroit où se trouve le point dans cette cellule. L'hypercube latin obtenu correspond donc à la matrice de N lignes et d colonnes à coefficients x_j^i .

Cette méthode garantit une couverture uniforme du domaine de chaque entrée évitant de sous échantillonner certains segments et à l'inverse d'en sur-échantillonner d'autres. Dans un souci de visualisation, la figure ci-dessous présente la méthode des LHS pour le TMG et le taux de chargement sur encours dans un espace de dimension $d = 2$ et $N = 100$. Comme nous pouvons le constater sur la figure ci-dessous, la méthode LHS permet une couverture de l'espace des données tandis qu'un tirage aléatoire aura tendance à sous-représenter certaines parties du cube.

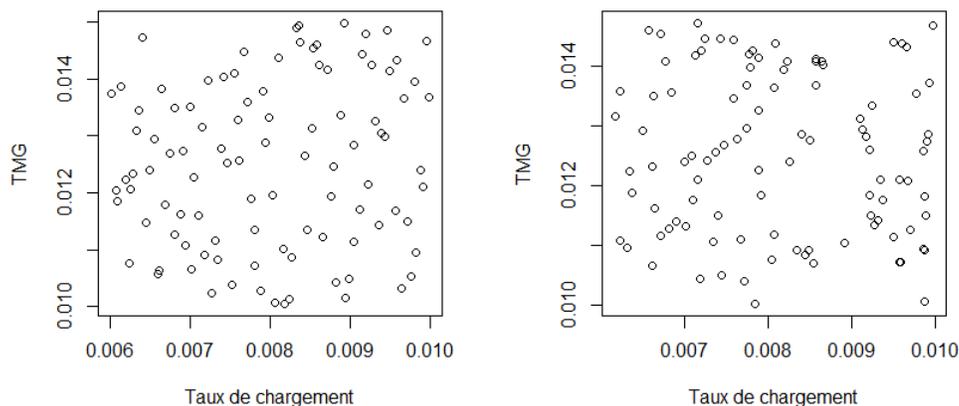


FIGURE 5.1 – Hypercube Latin (à gauche) contre tirage aléatoire (à droite)

Nous avons ensuite eu recours au LHS sur l'ensemble de nos 14 variables afin d'obtenir un nuage de point le plus uniformément réparti possible.

Il est à noter que le taux 8 ans a été échantillonné séparément. En effet, nous ne disposons pas de l'ensemble des courbes des taux (qui soient risque neutre) pour n'importe quelle valeur. Nous avons reconstruit 13 courbes des taux (correspondant à des scénarii générés via un générateur de scénario économique) pour 13 valeurs différentes de taux 8 ans.

Ces courbes des taux à la hausse et à la baisse permettent de prendre en compte des situations économiques s'éloignant plus ou moins fortement du scénario central, à savoir la courbe des taux fournit par l'EIOPA au 31/12/2015. Nous avons tiré aléatoirement l'une de ces 13 courbes pour rester cohérent avec la méthode LHS, ce qui nous a permis d'obtenir notre base de données pour $d = 14$ et $N = 100$.

La figure en cascade ci-dessous permet de résumer l'ensemble de variation de nos variables d'intérêts.

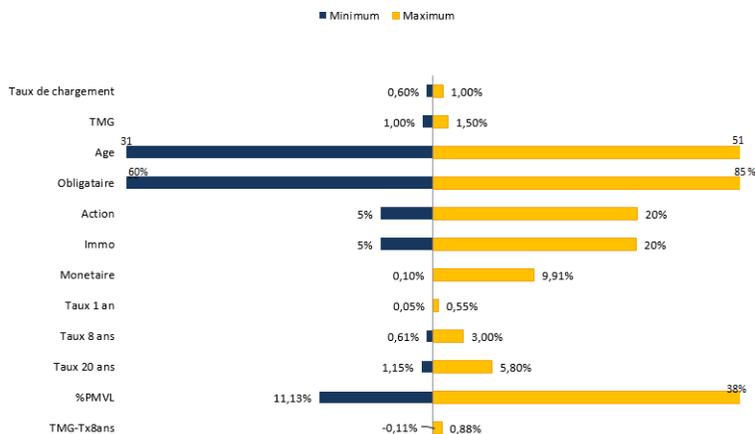


FIGURE 5.2 – Répartition LHS

5.2 Construction des courbes des taux

Dans cette section, nous allons décrire la façon dont les courbes des taux d'intérêts ont été construites. Comme cela a été expliqué précédemment, pour pouvoir construire la base la plus représentative possible, nous avons besoin de pouvoir évaluer le *SCR* pour différents niveaux de taux d'intérêts. Pour ce faire, nous avons eu recours à un générateur de scénarii économiques (GSE). Nous ne ferons qu'une brève description de cette méthode, pour plus de détails, se référer au mémoire de CHATAIGNIER.B [8].

5.2.1 Générateur de scénarii économiques

Un générateur de scénarii économique permet de générer des trajectoires de grandeurs financières comme les taux d'intérêt, les rendements actions, immobiliers et inflation, sur un horizon de temps défini. Pour calibrer nos scénarii économiques, nous devons distinguer deux univers :

- **Monde-réel** : La projection de la trajectoire d'un actif doit refléter son comportement historique. Les données historiques seront donc utilisées pour calibrer nos données.
- **Risque-neutre** : Les actifs modélisés ont pour rendement le taux sans risque afin d'obtenir une modélisation « Market-consistent », c'est-à-dire cohérente avec l'environnement actuel des prix observés sur le marché.

Nous allons nous servir du GSE dans un cadre risque-neutre.

La projection de ces grandeurs économiques se fait par le biais de modèles de diffusion comme le modèle Hull White à un facteur dans le cas de la courbe des taux d'intérêts. Ce modèle va prendre en input la courbe des taux initiale (celle fournie par l'EIOPA par exemple). C'est un modèle de taux court se définissant comme suit :

$$dr_t = (\theta_t - ar_t) dt + \sigma dW_t$$

La fonction θ_t est définie de sorte à ce que la structure des taux à l'instant $t = 0$ corresponde à celle des taux observés sur les marchés et mis en entrée du modèle. Cette adéquation avec les valeurs de marché se retrouve par la formule ci-dessous :

$$\theta_t = \frac{\partial f}{\partial T}(0, T) + a \times f(0, T) + \frac{\sigma^2}{2a} (1 - e^{-2at})$$

où $f(0, T)$ est le taux forward instantané associé au prix zéro coupon de maturité t noté $B(0, T)$.

Une fois les paramètres a et σ calibrés, nous pouvons en déduire le prix des obligations zéro-coupon via l'équation différentielle suivante :

$$dr_t = \left(\frac{\partial f}{\partial T}(0, T) + a \times f(0, T) + \frac{\sigma^2}{2a} (1 - e^{-2at}) r_t \right) dt + \sigma dW_t$$

Le terme $\frac{\partial f}{\partial t}(0, t)$ de l'expression précédente étant inconnu et dans un soucis de précision, nous allons approximer le taux par différence de processus suivante : $X_t = r_t - f(0, t)$ dont la dynamique est décrite par :

$$dX_t = \left(\frac{\sigma^2}{2a} (1 - e^{-2at}) - aX_t \right) dt + \sigma dW_t$$

L'équation précédente se résout en intégrant entre t et s et donne la formule permettant de simuler X_t à partir de X_s :

$$X_t = e^{-a(t-s)} X_s + \frac{\sigma^2}{a^2} e^{-at} (ch(at) - ch(as)) + \sigma \sqrt{\frac{1 - e^{-2a(t-s)}}{2a}} N(0, 1)$$

Les taux zéro-coupons forward sont alors calculés à partir des valeurs de $(X_t)_t$ en appliquant les formules suivantes :

$$P(t, T) = \frac{1}{(1 + R(t, T))^{(T-t)}}$$

où :

- $R(t, T)$ correspond aux taux zéro-coupon forward
- $P(t, T)$ correspond aux prix zéro-coupon

avec :

$$P(t, T) = \frac{P^{input}(0, T)}{P^{input}(0, t)} e^{-\frac{1}{2}\phi(t)B(t, T)^2 - X_t B(t, T)}$$

où :

- $P^{input}(0, T)$ correspond au prix des obligations zéro-coupon observés sur le marché
- $\phi(t) = \frac{\sigma^2}{2a} (1 - e^{-2at})$
- $B(t, T) = \frac{1 - e^{-a(T-t)}}{a}$

5.2.2 Courbes en inputs

Le modèle de Hull White à un facteur prend en input une courbe des taux initiale. Pour générer les différentes courbes des taux correspondant à différentes situations économiques espérées (faible hausse des taux, baisse des taux ou forte hausse des taux) nous avons appliqué un choc sur les 20 premières maturités de la courbe des taux initiale fournie par l'EIOPA. En effet, comme nous l'avons expliqué dans la partie précédente, au-delà de 20 ans, l'EIOPA a choisi d'extrapoler la courbe des taux.

$$Taux_n^\alpha = (1 + \alpha) \times Taux_n^{EIOPA}$$

α , représente le choc en pourcentage

Une fois ce choc mis en place et comme le prévoit la réglementation, nous avons appliqué la méthode de Smith-Wilson afin d'extrapoler nos courbes passé 20 ans et de les faire converger vers l'UFR (Ultimate Forward Rate), soit 4.2%. Il faut faire attention à la lecture de ces courbes ; elles correspondent à des taux spot tandis que le réglementation exige que les taux forward 1 an convergent vers ce niveau d'UFR. Or, il existe une relation entre les taux spots et forward 1 an nous permettant de vérifier que cette hypothèse de convergence est bien respectée :

$$f_n = \frac{(1 + s_n)^n}{(1 + s_{n-1})^{n-1}} - 1$$

avec : s_n et s_{n-1} , les taux spot aux dates n et $n-1$

La figure ci-dessous présente l'ensemble des courbes de taux choquées que nous allons donner en inputs à notre GSE. Sur ces différentes courbes, nous retrouvons l'ensemble des taux 8 ans présents dans notre base de données.

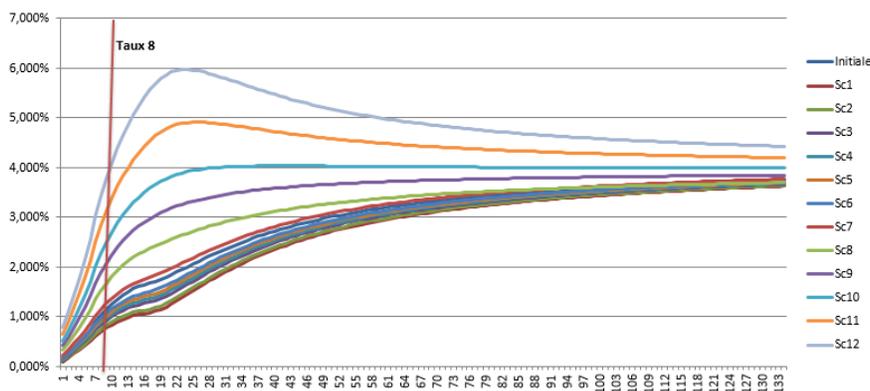


FIGURE 5.3 – Courbes des taux choquées en inputs

Une fois ces différentes courbes des taux obtenues, nous avons relancé notre GSE pour chacune de ces courbes afin d’obtenir 1 000 scénarii de taux RN sur l’horizon de projection et pour les maturités (durées) considérées.

5.2.3 Projections des autres indicateurs

Nous venons de présenter la méthodologie pour construire nos différentes courbes des taux sans risque. L’ESG a aussi permis d’obtenir des scénarii stochastiques pour les autres grandeurs financières comme l’inflation, l’indice immobilier et l’indice action. Il est important de remarquer qu’en univers risque-neutre, le rendement doit correspondre au taux 1 an de la courbe des taux associée. La méthodologie de calcul de ces différents indicateurs ne sera pas présentée, nous allons juste donner les modèles permettant de les projeter sur l’horizon considéré.

Le tableau ci-dessous permet de lister les modèles utilisés dans notre GSE :

Risque	Modèle	Diffusion
Action	Black-Scholes	$dS_t = (\mu - q)S_t dt + \sigma S_t dW_t$
Immobilier	Black-Scholes	$dS_t = (\mu - q)S_t dt + \sigma S_t dW_t$
Inflation	Gadmer	$i_t = \alpha + \beta \times r_t + \epsilon_t$
Taux	Hull-White	$dr_t = (\theta_t - ar_t) dt + \sigma dW_t$

TABLE 5.1 – Projection des indicateurs du GSE

5.3 Etude des corrélations entre nos variables

Avant d'étudier plus en profondeur les variables les plus significatives ou contribuant le plus à la construction de nos indicateurs d'intérêts, nous allons étudier les liaisons entre nos variables explicatives. Cette première étape va nous permettre d'éliminer les variables étant parfaitement corrélées. En effet, si deux variables sont parfaitement (positivement ou négativement) corrélées, cela signifie que l'information est redondante d'où la suppression de l'une de ces deux variables. Cette étape de suppression des données permet aussi de ne pas augmenter son poids dans la prise de décision de l'algorithme.

5.3.1 Coefficient de corrélation de Pearson

Soit x la matrice de variables explicatives et x_i l'une des variables explicatives, $\forall i \in [1 : p]$. Le coefficient de corrélation de Pearson permet de détecter et quantifier la présence ou l'absence de relation linéaire entre deux variables quantitatives et se calcule de la façon suivante :

$$\rho_{(x_i, x_j)} = \frac{\sigma_{x_i, x_j}}{\sigma_{x_i} \sigma_{x_j}}, \forall i, j \in [1 : p]$$

où :

$$\sigma_{x_i, x_j} = \text{Cov}((x_i, x_j)) = E[(x - E(x))(Y - E(Y))]$$

$$\sigma_{x_i} = \sqrt{E[(x - E[x])^2]} = \sqrt{E[x^2] - E[x]^2}$$

La figure ci-dessous présente la matrice de corrélations de Pearson entre nos variables. Plus la corrélation entre nos variables est forte et plus la case sera foncée, tendant vers le rouge lorsque cette corrélation est positive et vers le bleu lorsqu'elle est négative.

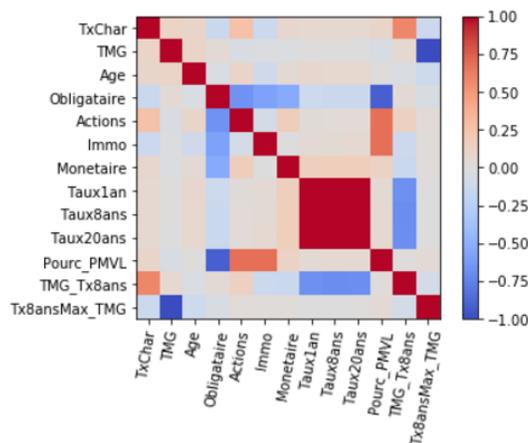


FIGURE 5.4 – Matrice de corrélation de Pearson

Comme nous pouvons le constater, la variable obligataire est parfaitement négativement corrélée à la variable % *PMVL*. La variable % *PMVL* se calcul comme suit :

$$PMVL = \frac{(VM^{actions} - VC^{actions}) + (VM^{immobiliers} - VC^{immobiliers})}{VM^{actifs} - VC^{actifs}}$$

Avec :

VM correspond à la valeur de marché

VC correspond à la valeur comptable

Or, nous savons par construction que :

$$\%action + \%immobilier + \%obligataire + \%monétaire = 1$$

La variable *PMVL* dépend donc par construction de l'obligataire.

De même, la variable « Taux 8 ans Max -*TMG* » est parfaitement positivement corrélée à la variable *TMG* ce qui était attendu par construction.

Cependant, la variable *TMG-Taux8ans* est fortement négativement corrélée mais pas parfaitement corrélée au Taux 8 ans ou au *TMG* ce qui traduit un effet que le coefficient de Pearson n'arrive pas à capter comme, l'impact d'une autre variable qui viendrait affecter cette dernière. Nous allons donc introduire une nouvelle mesure : le coefficient de corrélation de Kendall.

5.3.2 Coefficient de corrélation de Kendall

Précédemment, nous avons supposé qu'une relation linéaire liait l'ensemble de nos variables ce qui peut ne pas être le cas. Ainsi, nous allons avoir recours au coefficient de corrélation de Kendall.

Le tau de Kendall permet de mesurer l'association entre deux variables en considérant les rangs de chacune d'elle. Ici, nous ne considérons plus les valeurs que prennent les variables mais leur classement ou rang par rapport à l'ordre observé sur l'autre variable. Ce coefficient permet de mesurer des liaisons monotones entre nos variables (logarithmique, linéaire, exponentielle ...) ce qui en fait une statistique plus robuste que celle de Pearson.

Pour calculer le tau de Kendall, la notion de concordance et de discordance joue un rôle crucial. La concordance entre deux variables se traduit par le fait qu'elles bougent dans le même sens et discordante si elles varient en sens opposé. La statistique se calcule donc comme suit :

$$\tau = \frac{P - D}{\binom{n}{2}} = \frac{P - D}{\frac{n(n-1)}{2}}$$

avec P et D respectivement le nombre de paires concordantes et discordantes définies comme suit :

$$\begin{aligned} &\forall k, l \in [1 : n], \forall i, j \in [1 : p] \\ P &= \# \{x_{k,i} - x_{l,i} > 0\} \cap \{x_{k,j} - x_{l,j} > 0\} \cup \{x_{k,i} - x_{l,i} < 0\} \cap \{x_{k,j} - x_{l,j} < 0\} \\ D &= \# \{x_{k,i} - x_{l,i} < 0\} \cap \{x_{k,j} - x_{l,j} > 0\} \cup \{x_{k,i} - x_{l,i} > 0\} \cap \{x_{k,j} - x_{l,j} < 0\} \end{aligned}$$

Comme nous pouvons le voir avec la matrice corrélation de Kendall ci-dessous, l'association entre la part d'obligataire et le %PMVL est forte mais n'est pas parfaite. En ce qui concerne la variable « TMG-Taux8ans », nous constatons qu'elle est moins fortement associée à la variables TMG comparé au cas de Pearson. Raisonner sur les rangs plutôt que sur les montants permet donc d'obtenir une corrélation plus robuste dans le sens où cette dernière nécessite moins d'hypothèses sur la distribution sous-jacente de nos données.

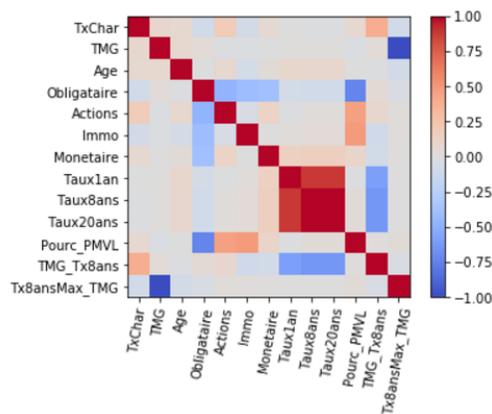


FIGURE 5.5 – Matrice de corrélation de Kendall

Au vu des corrélations obtenues au travers de ces deux méthodes, nous allons éliminer la variable ***Tx8ansMax - TMG***. Une nouvelle étude des corrélations confirme que sans cette variable, nous n'avons plus de variables parfaitement corrélées.

Sélection des variables

Dans cette section, nous allons nous intéresser au choix des variables à intégrer dans nos algorithmes de *machine learning*. Pour ce faire, nous allons étudier la relation liant le *BE*, les *FP* et les sous-modules de risque à l'ensemble de nos variables explicatives. Comme nous avons pu le voir précédemment, nos variables sont plus ou moins corrélées, une approche par Moindres carrés ordinaire (MCO) n'est donc plus pertinente. En effet, en cas de multicollinéarité, la matrice de variables explicatives n'est plus de rang plein : elle n'est donc plus inversible et l'estimateur des MCO ne peut être défini. En supprimant l'une des variables dans la partie précédente, nous avons éliminé ce cas.

Cependant, une forte corrélation subsiste dans nos données. Même si cette dernière n'est plus parfaite, l'estimation par MCO aura pour conséquence de biaiser nos estimateurs (forte imprécision) ce qui impactera les tests usuels de significativité des coefficients (T de student ou F de Fisher). Pour contourner ce problème, nous allons recourir à une régression de type Least Absolute Shrinkage and Selection Operator (LASSO) qui présente l'avantage de pouvoir gérer les variables fortement corrélées et de faire de la sélection de modèle par ailleurs.

6.1 Présentation du modèle

La méthode de régression LASSO a été introduite par Robert Tibshirani en 1996 pour prendre en compte le cas des données présentant une forte colinéarité parmi les variables explicatives. Dans le cadre du modèle linéaire standard, nous estimons le modèle suivant par minimisation de la somme des carrés des résidus :

$$\begin{aligned} Y &= X\beta + \epsilon \\ X &= \{(X_{i,j}), i \in [1, n], j \in [1, p]\} \\ \epsilon &\sim N(0, \sigma^2) \end{aligned}$$

Les variables $X_{i,p}$ n'étant pas toutes pertinentes dans la prédiction du *BE*, des *FP* et des sous-modules de risque, l'objectif est donc d'éliminer les variables inutiles. Le LASSO ne va donc pas faire une régression linéaire classique mais une régression linéaire augmentée d'un paramètre de régularisation ou pénalité afin de rendre nul certains coefficients dans l'estimation de β .

Le problème d'optimisation est donc le suivant :

$$\arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

où pour $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ et $y = (y_1, \dots, y_p) \in \mathbb{R}^p$:

$$\|x\|_2^2 = \sum_{i=1}^n x_i^2 \text{ et } \|y\|_1 = \sum_{i=1}^p |y_i|$$

Le paramètre $\lambda \geq 0$ contrôle la complexité du modèle. En effet, si $\lambda = 0$, nous nous retrouvons dans le cas classique de la régression par MCO. Si au contraire λ tend vers l'infini, on annule tous les coefficients $\hat{\beta}_j$ ce qui aboutit à ne retenir aucune variable. La variable λ permet donc de faire un choix parcimonieux des variables.

6.2 Convexité du modèle

Le problème précédent peut se décomposer comme la somme de deux fonctions convexes. En effet, en nommant h la fonction à minimiser :

$$\begin{aligned} h(\beta) &= f(\beta) + g(\beta) \\ f(\beta) &= \frac{1}{2} \|Y - X\beta\|_2^2 \\ g(\beta) &= \lambda \|\beta\|_1 \end{aligned}$$

Pour montrer que la fonction f est convexe, il suffit de montrer que sa matrice hessienne en tout point est semi-définie positive (f étant deux fois différentiable).

$$\begin{aligned} \nabla f(\beta) &= \frac{\partial}{\partial \beta} \left(\frac{1}{2} \langle Y, Y \rangle - \langle Y, X\beta \rangle + \frac{1}{2} \langle X\beta, X\beta \rangle \right) \\ &= \frac{\partial}{\partial \beta} \left(\frac{1}{2} \|X\beta\|_2^2 - \langle Y, X\beta \rangle \right) \\ &= X^T X\beta - X^T Y \\ \nabla^2 f(\beta) &= \frac{\partial^2}{\partial \beta^2} (X^T X\beta - X^T Y) \\ &= X^T X \end{aligned}$$

Or, $\forall z \ z^T X^T X z = (Xz)^T X z = \|Xz\|_2^2 \geq 0$ ce qui implique que la matrice hessienne $\nabla^2 f(\beta) = X^T X$ est semi-définie positive. Le caractère semi-défini positif de la matrice hessienne pour tout $\beta \in \mathbb{R}^p$ est une condition suffisante pour conclure que f est convexe.

Montrons désormais que la fonction g est également convexe. On ne peut toutefois pas raisonner comme précédemment. En effet, la norme 1 n'étant pas différentiable en 0, nous ne pourrions pas utiliser la même méthode. La convexité de g se montre comme suit :

$\forall \beta_1, \beta_2 \in \mathbb{R}^p, \forall \alpha \in [0, 1]$

$$\begin{aligned} g(\alpha\beta_1 + (1-\alpha)\beta_2) &= \lambda \|\alpha\beta_1 + (1-\alpha)\beta_2\|_1 \\ &\leq \lambda \|\alpha\beta_1\|_1 + \lambda \|(1-\alpha)\beta_2\|_1 \\ &= \lambda \alpha \|\beta_1\|_1 + \lambda (1-\alpha) \|\beta_2\|_1 \\ &= \alpha g(\beta_1) + (1-\alpha)g(\beta_2) \end{aligned}$$

Ainsi, g est convexe.

Par suite, h est convexe comme somme de deux fonctions convexes.

Notre problème $\arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1$ est donc un problème de minimisation sans contrainte d'une fonction convexe. Nous avons donc un problème convexe.

Comme h est une fonction convexe, si nous avons pour un β donné $\nabla h(\beta) = 0$, alors β est un minimiseur global de h .

6.3 Calcul des estimateurs

Nous venons de montrer que h était convexe, ainsi, $\tilde{\beta}$ est un minimiseur global de $h = f + g$ si et seulement si $0 \in \partial(\tilde{\beta})$ où $\partial(\tilde{\beta})$ est la sous-différentielle de g . Ce résultat découle des propriétés d'une sous-différentielle suivante :

Definition : sous-gradient

Soit $f : \mathbb{E} \rightarrow \mathbb{R} \cup \{+\infty\}$ une fonction convexe et propre. On dit que $s \in \mathbb{E}$ est un sous-gradient de f en $x \in \text{dom } f$ si l'une des propriétés équivalentes suivantes est vérifiée :

- $\forall d \in \mathbb{E}, f'(x; d) \geq \langle s, d \rangle$, où f' est la dérivée au sens de Gâteaux
- $\forall y \in \mathbb{E}, f(y) \geq f(x) + \langle s, y - x \rangle$
- x minimise $y \in \mathbb{E} \mapsto f(y) - \langle s, y \rangle$
- $f^*(s) + f(x) \leq \langle s, x \rangle$
- $f^*(s) + f(x) = \langle s, x \rangle$

Définition : sous-différentiel

L'ensemble des sous-gradients de f en x est appelé le **sous-différentiel** de f en x . Il est noté $\partial f(x)$. On dit que f est **sous-différentiable** en x si $\partial f(x) \neq \emptyset$. Par convention, $\partial f(x) = \emptyset$ si $x \notin \text{dom } f$.

Propriété : condition d'optimalité

Soit $f \in \text{Conv}(\mathbb{E})$. Un point $\bar{x} \in \mathbb{E}$ minimise f sur \mathbb{E} si, et seulement si, $0 \in \partial f(\bar{x})$.

Nous avons que :

$$\left\{ \begin{array}{l} h = f + g \\ f \text{ est différentiable : } \partial f(\beta) = \{\nabla f(\beta)\} \forall \beta \end{array} \right. \quad \text{donc } \forall \beta \partial h(\beta) = \nabla f(\beta) + \partial g(\beta)$$

$$\begin{aligned} \text{Donc } \tilde{\beta} \text{ minimiseur global de } h &\Leftrightarrow 0 \in \partial h(\tilde{\beta}) \\ &\Leftrightarrow 0 \in \nabla f(\tilde{\beta}) + \partial g(\tilde{\beta}) \\ &\Leftrightarrow -\nabla f(\tilde{\beta}) \in \partial g(\tilde{\beta}) \end{aligned}$$

Or :

$$\begin{aligned} g(\tilde{\beta}) &= \partial \|\lambda \cdot \|_1(\tilde{\beta}) \\ &= \left\{ \gamma \mid \forall \beta \lambda h(\beta) \geq \lambda h(\tilde{\beta}) + \langle \gamma, \beta - \tilde{\beta} \rangle \right\} \\ &= \left\{ \gamma \mid \forall \beta h(\beta) \geq h(\tilde{\beta}) + \langle \frac{\gamma}{\lambda}, \beta - \tilde{\beta} \rangle \right\} \\ &= \left\{ \lambda \gamma' \mid \forall \beta h(\beta) \geq h(\tilde{\beta}) + \langle \gamma', \beta - \tilde{\beta} \rangle \right\} \\ &= \lambda \left\{ \gamma' \mid \forall \beta h(\beta) \geq h(\tilde{\beta}) + \langle \gamma', \beta - \tilde{\beta} \rangle \right\} \\ &= \lambda \partial \|\cdot\|_1 \\ &= \lambda \partial h(\tilde{\beta}) \end{aligned}$$

D'où $\tilde{\beta}$ minimiseur global de $h \Leftrightarrow X^T(Y - X\tilde{\beta}) \in \lambda\partial h(\tilde{\beta})$
 $\Leftrightarrow \exists \nu$ tel que $X^T(Y - X\tilde{\beta}) = \lambda\nu$

où

$$\nu_i = \begin{cases} \text{sign}(\tilde{\beta}_i) & \text{si } \tilde{\beta}_i \neq 0 \\ \in [-1, 1] & \text{si } \tilde{\beta}_i = 0 \end{cases}$$

Ainsi, ceci est équivalent à avoir ν tel que :

$$\tilde{\beta} = (X^T X)^{-1} (\lambda\nu + X^T Y)$$

Ce qui correspond à $\forall i \in [1, p]$:

$$\tilde{\beta} = \begin{cases} |X^T(Y - X\tilde{\beta})| \leq \lambda & \text{si } \tilde{\beta}_i = 0 \\ X^T(Y - X\tilde{\beta}) = \lambda \text{sign}(\tilde{\beta}_i) & \text{si } \tilde{\beta}_i \neq 0 \end{cases}$$

Graphiquement, la solution du problème du LASSO peut se visualiser ci-après. La zone bleue représente la zone de pénalité du LASSO, autrement dit la norme L^1 tandis que les ellipses rouges correspondent aux contours de la fonction d'erreur des moindres carrés.

Le $\hat{\beta}$ correspond à la solution des moindres carrés. Sur le graphique de gauche se trouve la solution du LASSO tandis qu'à droite se trouve la solution du problème Ridge. La différence entre les deux problèmes réside dans le choix de la norme pénalisante (L^2 pour le ridge). La solution du problème se trouve au croisement entre les ellipses rouges et la zone bleue. Comme nous pouvons le constater, la solution Lasso se situe là où l'une des composantes a été annulée tandis que le ridge n'en annule pas. La norme 1 semble donc plus adaptée à la réduction de dimensionnalité.

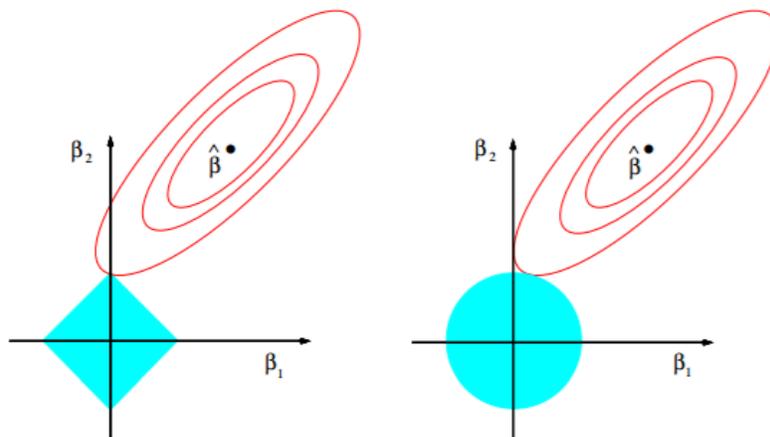


FIGURE 6.1 – Problème du LASSO-Ridge

L'estimateur du LASSO ne présentant pas de forme explicite, il est en pratique calculé de manière itérative afin d'en garantir sa stabilité, notamment dans le cas de matrice sparse. La méthode suivante s'inspire de l'article de Alquier [22] et a été introduit par Tibshirani, Saunders, Rosset, Zhu, et Knight [23] sous le nom de « fused LASSO ». Nous ne présentons ici que la forme du nouveau problème d'optimisation. Pour plus de détails sur les résultats théoriques, se référer aux articles précédents.

Le problème se réécrit donc de la façon suivante :

$$\hat{\beta}_{F-LASSO}(\lambda, \mu) = \arg \min_{\beta \in \mathbb{R}^p} \frac{1}{2} \|Y - X\beta\|_2^2 + \lambda \sum_{i=1}^p |\beta_i| + \mu \sum_{i=2}^p |\beta_i - \beta_{i-1}|$$

L'algorithme permettant de calculer le problème précédent procède de manière itérative en initialisant le vecteur $\beta^{(0)} = (0, \dots, 0)$ et en mettant à jour un groupe de coordonnées à chaque étape. Ainsi, pour un $\hat{\beta}^{(m)}$ donné, $\hat{\beta}^{(m+1)}$ sera défini.

Pour finir, pour tout groupe de données $j \in [1, p]$ et $k \in [1, p - j + 1]$ où k désigne la longueur du groupe de coordonnées et j sa position.

$$\hat{\beta}^m = \arg \max_{j,k} \|X \left(\tilde{\beta}^{(m+1),j,k} - \tilde{\beta}^m \right)\|_2^2$$

6.4 Validation-croisée des paramètres

Nous avons estimé notre modèle par validation-croisée (cross-validation ou CV) afin de choisir le paramètre λ de sorte à minimiser l'erreur de prédiction. Cette validation-croisée sera effectuée K-fois. On parle communément de K-fold cross-validation.

Considérons un échantillon de données $(x_i, y_i)_{i=1, \dots, n}$ et un estimateur \hat{f}_λ qui dépend d'un paramètre λ à « régler » (tuning parameter). Le principe de la validation-croisée K-fois est le suivant :

- On divise aléatoirement l'échantillon $\{1, \dots, n\}$ en K sous-échantillon de même taille F_1, \dots, F_k
- Pour $k = 1, \dots, K$:
 - On considère l'échantillon d'apprentissage (x_i, y_i) , $i \notin F_k$ et on valide sur l'échantillon (x_i, y_i) , $i \in F_k$
 - Pour chaque valeur du paramètre à « régler » $\lambda \in \{\lambda_1, \dots, \lambda_n\}$, on estime \hat{f}_λ^{-k} sur la base d'apprentissage et on calcule l'erreur sur l'échantillon test : $err_k(\lambda) = \sum_{i \in F_k} \left(y_i - \hat{f}_\lambda^{-k}(x_i) \right)^2$
- Pour chaque paramètre de réglage λ , on calcule l'erreur moyenne sur tous les sous échantillons : $CV(\lambda) = \frac{1}{n} \sum_{k=1}^K err_k(\lambda) = \frac{1}{n} \sum_{k=1}^K \sum_{i \in F_k} \left(y_i - \hat{f}_\lambda^{-k}(x_i) \right)^2$

6.4.1 Sélection de variables pour l'estimation du BE

Dans cette sous-section, nous allons présenter les résultats d'estimations de la sélection de variables sur le BE .

La figure ci-dessous présente la variation de l'erreur quadratique moyenne en fonction du paramètre de pénalisation de la régression LASSO sur chacun des K sous-échantillons.

Ainsi, à l'issue de cette étape, nous allons retenir le $\hat{\lambda}$ qui minimise l'erreur quadratique ou autrement dit :

$$\hat{\lambda} = \arg \min_{\lambda \in \{\lambda_1, \dots, \lambda_n\}} CV(\lambda)$$

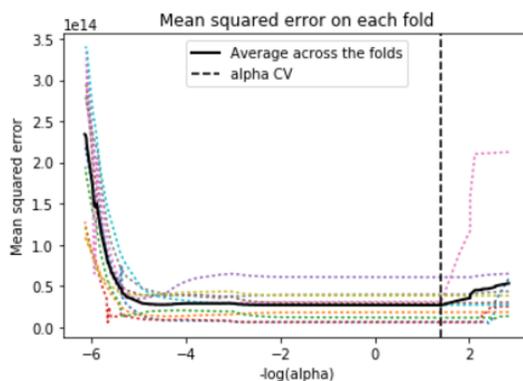


FIGURE 6.2 – Cross-Validation

La figure ci-dessous représente le chemin de régularisation pris par chacune des variables. Comme évoqué précédemment, certains coefficients ont une valeur nulle, indiquant qu'ils ne sont pas significatifs ou bien qu'ils n'apportent pas d'informations pertinentes lors de la détermination du BE . Ces variables ne seront donc pas retenues par la suite lorsque nous allons utiliser des méthodes d'apprentissage statistique pour prédire le BE . La même étude sera faite sur chacune des variables que nous chercherons à prédire.

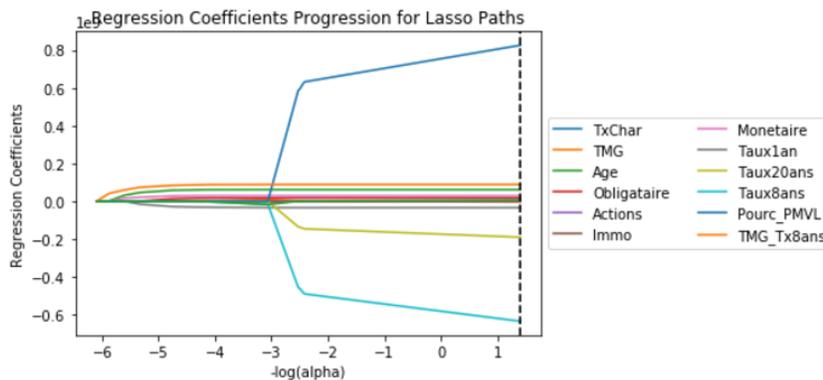


FIGURE 6.3 – Variables LASSO

L'estimation des coefficients via validation-croisée est la suivante :

Taux de chargt	TMG	Age	Obligataire	Action	Immo	Monétaire	Taux 1 an	Taux 8 ans	Taux 20 ans	% PMVL	TMG-Tx 8 ans
-3×10^9	4×10^9	3×10^5	0	0	7×10^8	-1×10^8	-1×10^9	1×10^9	-5×10^9	1×10^8	3×10^8

Comme nous pouvons le remarquer, le BE de notre compagnie d'assurance est principalement déterminé par des variables de passif mais aussi par des variables d'actifs comme les taux, l'immobilier et le monétaire. En effet, le BE se définit comme :

$$\mathbb{E}^Q \left(\sum_i \frac{F_i}{(1+r_i)^i} \right)$$

Q représente la probabilité risque-neutre

F_i représente le flux de passif à la date i (les engagements de l'entreprise vis-à-vis des assurés)

r_i représente le taux sans risque

De par sa définition, le BE représente les engagements de l'assureur envers les assurés, évalué en valeur de marché. De ce fait, le BE sera sensible aux variables suivantes :

- Taux de chargement sur encours
- TMG
- Age de l'assuré
- Courbe des taux d'intérêts

Le taux de chargement sur encours représente les frais qui seront prélevés à l'assuré pour garantir la bonne gestion de son contrat. Quant au TMG, il représente le taux de rendement du contrat d'assurance, qui est le taux que l'assureur s'engage contractuellement à verser aux assurés. Dès lors, il apparaît que les engagements de l'assureur (les flux de passifs) dépendent directement de la différence entre les sommes qui seront prélevées par l'assureur et les sommes qui seront reversées aux assurés.

Par ailleurs, l'âge de l'assuré va avoir un impact significatif sur les flux de passif de l'assureur. Prenons l'exemple de deux personnes entrant au même moment dans un contrat d'assurance épargne mais n'ayant pas le même âge. Nous supposons de plus que ces personnes sortiront du contrat à 61 ans. La première personne sera prélevée sur une période de 30 ans tandis que la deuxième sur une période de 20 ans. A taux de chargement et TMG fixés, la compagnie d'assurance gagnera plus d'argent sur le premier contrat puisqu'elle chargera l'assuré sur une période plus longue. D'autre part, elle devra redistribuer plus d'argent. Rappelons que les assureurs sont tenus légalement de redistribuer au souscripteur d'un contrat d'assurance vie en euros au moins 85 % des bénéfices financiers (gains réalisés en plaçant l'épargne du titulaire du contrat) et 90 % des bénéfices techniques (résultant de la différence entre les frais prélevés par l'assureur et les frais réels).

Sur le deuxième contrat, elle gagnera donc moins d'argent via le taux de chargement sur encours (car sera prélevée sur une plus courte période) mais redistribuera moins d'argent via les mécanismes de participation aux bénéfices et de bénéfices techniques. D'autre part, le taux de décès augmente avec l'âge ce qui aura pour conséquence de réduire la provision mathématique (PM).

Enfin, comme nous pouvons le constater via la formule précédente, le BE dépend directement des taux via le facteur d'actualisation. Il est important de remarquer qu'il existe un lien très fort entre le TMG et les taux. En effet, la différence entre le taux que l'assureur s'engage à verser aux assurés et le taux servi sur le marché aura un impact sur les taux de rachats des contrats.

Le rachat est une option qui permet à l'assuré de récupérer à n'importe quel moment, tout ou partie de son encours. Il existe deux types de rachats en assurance :

- les rachats **structurels** : ils correspondent aux rachats observés de manière récurrente indépendamment de l'environnement économique. Ce type de rachat peut s'expliquer par le besoin de la part des assurés de transmettre un patrimoine ou de faire face à des dépenses imprévues. Ils dépendent en grande partie de la fiscalité comme nous pouvons le constater sur le tableau ci-dessous.

Ancienneté du contrat	Imposition des intérêts
Moins de 4 ans	35% ou impôt sur le revenu
De 4 à 8 ans	15% ou impôt sur le revenu
Plus de 8 ans	0% jusqu'à: - 4 600€ pour une personne seule - 9 200€ pour un couple partageant la même feuille fiscale 7,5% au-delà

FIGURE 6.4 – Fiscalité des rachats structurels

Ainsi, les compagnies d'assurance observent généralement une vague de rachat à partir de la 8^{ème} année après l'ouverture du contrat.

- les rachats **dynamiques** : ils correspondent aux rachats conjoncturels et modélisent l'arbitrage que les assurés vont faire entre rendement du contrat et rendement qu'il pourrait obtenir par ailleurs. Si les assurés se rendent compte qu'ils peuvent obtenir un meilleur rendement en plaçant leur argent sur un autre produit financier qu'en le laissant dans le contrat d'assurance-vie actuel, ils auront tendance à racheter leur contrat pour investir l'argent ailleurs.

6.4.2 Sélection de variable pour l'estimation des FP

Dans cette sous-section, nous allons présenter les résultats d'estimations de la sélection de variables sur les *FP*.

La figure ci-dessous nous donne la courbe de validation-croisée :

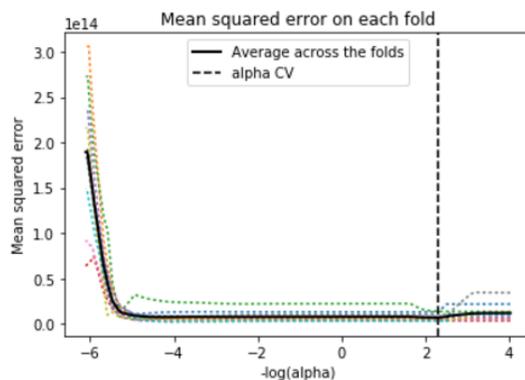


FIGURE 6.5 – Cross-Validation

De la même façon que dans l'exemple précédent, la figure ci-dessous présente le chemin de régulation pris par chacune des variables expliquant la variation des *FP* en ayant choisi le λ optimal.

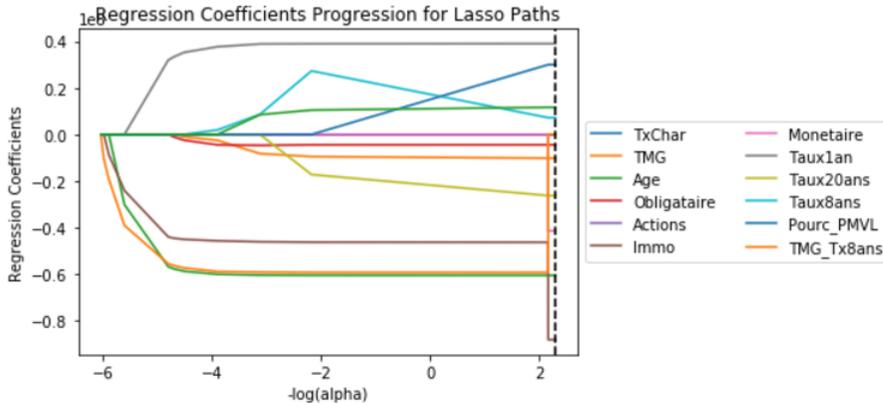


FIGURE 6.6 – Variables LASSO

L'estimation des coefficients via validation-croisée est la suivante :

Taux de chargt	TMG	Age	Obligataire	Action	Immo	Monétaire	Taux 1 an	Taux 8 ans	Taux 20 ans	% PMVL	TMG-Tx 8 ans
-9×10^9	-4×10^9	-8×10^5	0	-2×10^8	-1×10^8	1×10^8	-1×10^9	4×10^9	5×10^9	6×10^{-8}	6×10^8

Comme nous pouvons le constater, les FP de notre compagnie d'assurance sont déterminés par :

- l'actif
- le passif

La partie actif de notre bilan est modélisée par les variables :

- Action
- Immobilier
- Monétaire
- Obligataire
- % PMVL
- Courbe des taux

Tandis que la partie passif est déterminée par :

- Taux de chargement
- TMG
- Age de l'assuré
- TMG-Tx 8 ans

Sur la partie actif, seule la variable obligataire converge vers 0 du fait d'une multicolinéarité avec les autres variables d'investissement. En effet, le proportion d'obligataire correspond tout simplement à :

$$\%Obligataire = 1 - \%Actions - \%Immo - \%Monétaire$$

La connaissance des trois autres variables capte donc de l'information quant à la part d'obligataire en portefeuille.

L'interaction entre l'actif et le passif de notre bilan d'assurance apparaît au travers des variables TMG et courbes des taux. Comme cela a été dit auparavant, en fonction de l'environnement économique, les assureurs seront soumis à des rachats dynamiques. Pour faire face à ses rachats, l'assureur sera amené à vendre des actifs afin de respecter ses engagements. Nous voyons donc que la variable « Pourcentage de plus-ou-moins-value latente » va jouer un rôle important, permettant aux assureurs de piocher dans cette réserve pour faire face à ses obligations.

La même analyse sera faite sur les sous-modules de risques afin de déterminer les variables les plus contributives mais nous ne les présenterons pas.

Troisième partie

Utilisation de techniques de *machine learning* pour la mise en place d'un proxy

Les Support Vector Machines

7.1 Contexte

Dans cette partie, nous allons présenter l'une des méthodes d'apprentissage statistique les plus utilisées, les Support Vector Machines (SVM). Les SVM sont principalement utilisés dans des problèmes de classification, lorsque l'on veut prédire la classe d'appartenance d'une variable. Cependant, lorsque les données ne sont pas binaires mais quantitatives, les SVM peuvent être utilisés pour faire de la régression : Support Vector Machines Régression (SVR). Dans notre cas, les SVM pour la régression peuvent servir à prédire la valeur du BE , des FP ou des sous-modules de risques en fonction des différentes variables jugées pertinentes. Nous cherchons donc à déterminer une fonction de n variables qui ajustent au mieux nos données, dans le sens où l'erreur quadratique est minimisée.

Ainsi, les SVM peuvent nous servir à calculer le capital économique dans un cadre ORSA puisque cette méthode va nous permettre de prédire le BE , les FP ainsi que les sous-modules de risques sur l'ensemble du plan stratégique de la compagnie d'assurance. Cette méthode est donc censée réduire le temps de calcul du capital économique puisqu'elle permet d'éviter d'avoir recours à des simulations stochastiques qui sont coûteuses à implémenter.

7.2 Formulation du problème

Considérons l'ensemble d'apprentissage $E = \{(x_i, y_i), \forall i = 1 : N\}$, $x \in \mathbb{R}^n, y \in \mathbb{R}$. Le problème consiste en l'estimation d'une fonction f qui minimise la distance entre les y_i et $f(x_i)$. D'après l'article de Vapnik [19], le but est de trouver une fonction $f(x)$ qui a au plus une déviation ϵ de la vraie valeur y_i pour toute la base d'échantillon et qui soit la plus lisse possible (constante possible).

7.2.1 Cas linéaire

Nous allons d'abord considérer le cas où f est linéaire :

$$f(x) = x^T \beta + \beta_0$$

Pour que la fonction soit la plus lisse possible, il faut que β soit faible, ce qui revient à minimiser sa norme euclidienne. Le problème s'écrit donc comme suit :

$$\min_{\beta} \frac{1}{2} \|\beta\|^2$$
$$\text{s.c.} \begin{cases} y_i - f(x_i) \leq \epsilon \\ f(x_i) - y_i \leq \epsilon \end{cases}$$

Le problème précédent suppose que cette optimisation convexe est réalisable. Dans le cas où ce problème n'est pas réalisable, et de manière similaire aux « marges souples » des fonctions de pertes (Cortes and Vapnik, [24]), nous pouvons introduire des variables d'écarts ξ_i, ξ_i^* pour satisfaire les

autres contraintes. Ainsi, nous obtenons le problème suivant :

$$\min_{\beta, \xi, \xi^*} \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*)$$

$$\text{s.c} \begin{cases} y_i - f(x_i) \leq \epsilon \\ f(x_i) - y_i \leq \epsilon \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

La constante $C > 0$ représente l'arbitrage entre la constance de f et le montant de déviation ϵ toléré.

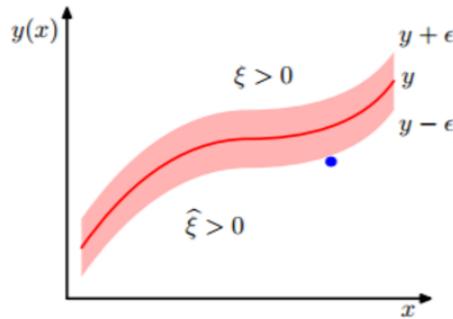


FIGURE 7.1 – Représentation des Support Vector Regression

La partie rouge ombragée dans la figure ci-dessus représente le tube ϵ -intensif où ϵ est une variable de seuil qui sert à encadrer le niveau de variation des variables prédites. En d'autres termes, les variables prédites doivent se trouver dans cet intervalle pour que la prédiction soit jugée acceptable. Les variables d'écart ξ servent à prendre en compte les erreurs d'approximations et permettent la réalisation du problème dans le cas où ce dernier ne l'était pas à l'origine.

La variable ξ représente la fonction de perte intensive qui peut-être définie comme suit :

$$\xi_\epsilon = \begin{cases} 0 & \text{si } |\xi| < \epsilon \\ |\xi| - \epsilon & \text{sinon} \end{cases}$$

Le problème précédent se résout sous sa forme duale comme suit à l'aide d'un Lagrangien :

$$L(\beta, \xi_i, \xi_i^*, \epsilon, \alpha_i, \alpha_i^*, \eta_i, \eta_i^*) = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) - \sum_{i=1}^N \alpha_i (\epsilon + \xi_i - y_i + f(x_i))$$

$$- \sum_{i=1}^N \alpha_i^* (\epsilon + \xi_i^* + y_i - f(x_i)) - \sum_{i=1}^N (\eta_i \xi_i + \eta_i^* \xi_i^*), \tag{7.1}$$

Puis, en appliquant les conditions du premier ordre :

$$\partial_{\beta_0} L = \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0 \quad (7.2)$$

$$\partial_{\beta} L = \beta - \sum_{i=1}^N (\alpha_i^* - \alpha_i) x_i = 0 \quad (7.3)$$

$$\partial_{\eta_i^*} L = C - \alpha_i^* - \eta_i^* = 0 \quad (7.4)$$

$$\partial_{\eta_i} L = C - \alpha_i - \eta_i = 0 \quad (7.5)$$

En injectant les équations (7.2), (7.3), (7.4) et (7.5) dans le problème dual (7.1), nous nous retrouvons avec le nouveau problème d'optimisation suivant :

$$\begin{aligned} \max_{\alpha, \alpha^*} & -\frac{1}{2} \sum_{i,j=1}^N (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \epsilon \sum_{i=1}^N (\alpha_i - \alpha_i^*) + \sum_{i=1}^N y_i (\alpha_i - \alpha_i^*) \\ \text{s.c} & \begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases} \end{aligned} \quad (7.6)$$

Les variables duales η_i, η_i^* ont été éliminées au travers de l'équation (7.4) puisqu'elles n'apparaissent plus dans la fonction objectif mais dans la condition de faisabilité. L'équation (7.3) peut se réécrire de la façon suivante :

$$\beta = \sum_{i=1}^N (\alpha_i - \alpha_i^*) x_i \quad (7.7)$$

Ce qui nous donne :

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) \langle x, x_i \rangle + \beta_0 \quad (7.8)$$

Comme nous pouvons le constater dans l'équation (7.7), β peut être complètement décrit comme une combinaison linéaire de la base d'apprentissage x_i . Ceci signifie que la complexité d'une fonction représentée par les Support Vectors (SV) ne dépend pas de la dimension de l'espace des données en entrée mais dépend seulement du nombre de SV. Les SV correspondent à certains points de la base de données qui contribuent à la détermination de la marge ϵ .

En ce qui concerne le terme β_0 , il peut être exprimé explicitement à l'aide des conditions de Karush-Kuhn-Tucker (KKT). En effet, à l'optimum, les variables duales sont orthogonales ce qui se traduit dans notre cas par :

$$\begin{aligned} \alpha_i (\epsilon + \xi_i - y_i + f(x_i)) &= 0 \\ \alpha_i^* (\epsilon + \xi_i^* + y_i - f(x_i)) &= 0 \\ (C - \alpha_i) \xi_i &= 0 \\ (C - \alpha_i^*) \xi_i^* &= 0 \end{aligned}$$

A partir de ces quatre équations, nous obtenons plusieurs résultats intéressants en résolvant les équations de KKT.

Premièrement, si $\epsilon + \xi_i - y_i + f(x_i) = 0$ alors $\alpha_i \neq 0$ et nous avons deux possibilités :

- $C - \alpha_i = 0$ et $\xi_i > 0$
- $C - \alpha_i > 0$ et $\xi_i = 0$

La première possibilité correspond au cas où $C = \alpha_i$ et $\xi_i > 0$ ce qui implique que les points se trouvent au-dessus de la limite supérieure tandis que la seconde possibilité implique que les points se trouvent sur la limite supérieure du tube ϵ ($\xi_i = 0$ et $C > \alpha_i$). De manière symétrique sur $\epsilon + \xi_i^* + y_i - f(x_i) = 0$ alors $\alpha_i^* \neq 0$ ce qui implique que l'on se trouve en dessous ou sur la limite inférieure du tube ϵ . Dans les deux cas les autres contraintes sont satisfaites.

Deuxièmement, les deux contraintes $\epsilon + \xi_i - y_i + f(x_i) = 0$ et $\epsilon + \xi_i^* + y_i - f(x_i) = 0$ sont incompatibles. En effet, en sommant les deux contraintes nous obtenons : $\epsilon = -\frac{1}{2}(\xi_i + \xi_i^*)$. La résolution du Lagrangien associé à notre problème d'optimisation suppose que $\xi_i, \xi_i^* \geq 0$. Ceci nous donne à nouveau deux possibilités :

- $\epsilon = \xi = 0$ et $C \geq \alpha_i$
- $\epsilon \leq 0, \xi \geq 0$ et $C = \alpha_i$

Or, ϵ doit être strictement positif et donc nous devons avoir nécessairement $\alpha_i, \alpha_i^* \in [0, C]$.

Les Support Vectors sont les points qui contribuent à la prédiction donnée par l'équation (7.8), en d'autres termes, ceux pour lesquels $\alpha_i^* \neq 0$ ou $\alpha_i \neq 0$. Ce sont les points qui se trouvent sur la frontière du tube ϵ intensif ou à l'extérieur du tube. Tous les points à l'intérieur du tube correspondent au cas où $\alpha_i = \alpha_i^* = 0$. Dans ce cas, nous obtenons à nouveau une solution explicite : $f(x) = \beta_0$.

Enfin, pour α_i^* (respectivement α_i) $\in [0, C]$, nous avons $\xi_i^* = 0$ (respectivement $\xi_i = 0$) ce qui nous permet d'obtenir β_0 en vérifiant les deux premières contraintes de KKT :

$$\begin{aligned}\beta_0 &= y_i - \langle \beta, x_i \rangle - \epsilon \text{ for } \alpha_i \in [0, C] \\ \beta_0 &= y_i - \langle \beta, x_i \rangle + \epsilon \text{ for } \alpha_i^* \in [0, C]\end{aligned}$$

7.2.2 Cas non-linéaire

La précision de l'estimation dépend en grande partie du choix de la fonction noyau appliquée à nos données. Dans le cas précédent, nous avons considéré un noyau de la forme $K(x, x_i) = \langle x, x_i \rangle$ qui correspond à un noyau linéaire (Linear Kernel), c'est-à-dire à une approximation des données par une fonction linéaire.

Cependant, comme nous pouvons le remarquer ci-dessous, les données ne peuvent pas s'approxi-mer directement par une fonction linéaire, c'est la raison pour laquelle elles seront transformées via une fonction noyau.

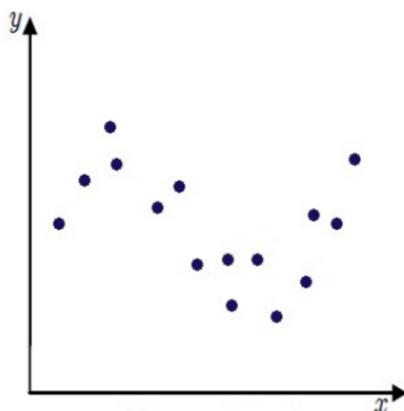


FIGURE 7.2 – Données non-linéairement séparables

Les données n'étant pas toujours linéairement séparables, l'astuce consiste à les transformer via une fonction noyau dans un espace de dimension supérieure.

Définition : Noyau défini positif

Une fonction symétrique $K : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ est un noyau défini positif sur \mathcal{X} si

$$\sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0$$

pour tout $n \in \mathbb{N}$, $x_1, \dots, x_n \in \mathcal{X}$, $c_1, \dots, c_n \in \mathbb{R}$.

Théorème

Une fonction $K : \mathcal{X} \times \mathcal{X}$ est un noyau défini positif s'il existe un ensemble discret T et une application $\phi : \mathcal{X} \rightarrow l^2(T)$, l'ensemble des valeurs réels $\{u_t, t \in T\}$ tel que $\sum_{t \in T} |u_t|^2 < \infty$ où :

$$\forall (x, y) \in \mathcal{X} \times \mathcal{X}, K(x, y) = \sum_{t \in T} \phi(x)_t \phi(y)_t = \langle \phi(x), \phi(y) \rangle_{l^2(\mathcal{X})}$$

Une application ϕ de cette forme est appelé une « feature map » pour K .

Dans le cas non-linéaire, le problème reste le même à la différence que la quantité $\langle x_i, x_j \rangle$ est remplacée par une fonction noyau : $K(x_i, x_j)$ et la valeur prédire s'obtiendra comme suit :

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x, x_i) + \beta_0$$

La figure ci-dessous présente l'utilisation d'un noyau pour transformer les données de sorte à pouvoir les séparer linéairement.

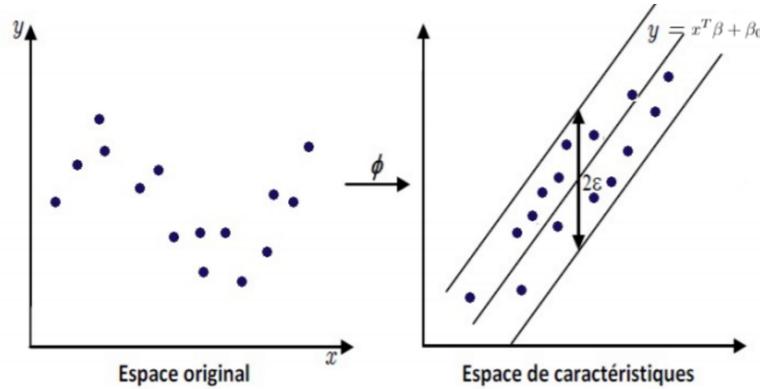


FIGURE 7.3 – Transformation des données via un noyau

Une fois les données estimées dans ce nouvel espace, il est possible d'utiliser l'image inverse du noyau pour les renvoyer dans son espace de départ. La figure ci-dessous résume le procédé de transformation des données via un noyau. Finalement, cette astuce nous aura permis d'approximer linéairement des données qui ne le sont pas afin d'en déduire les engagements de la compagnie d'assurance.

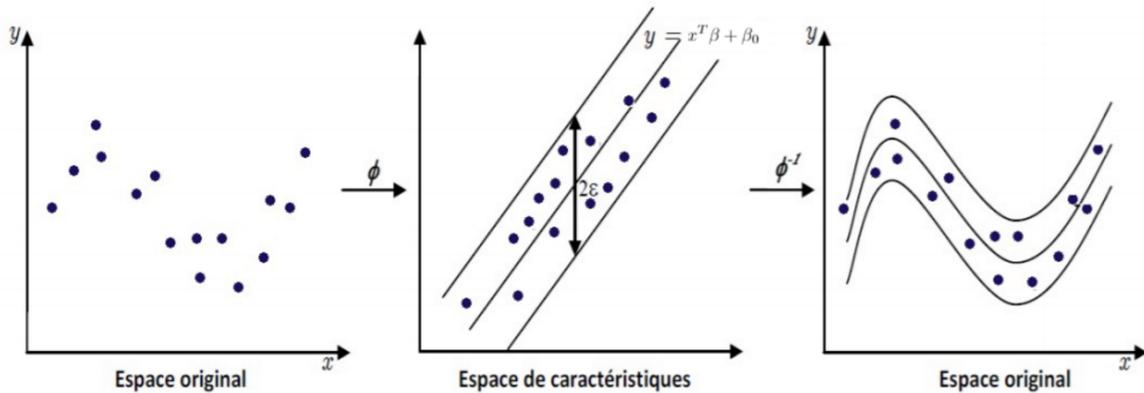


FIGURE 7.4 – Transformation de l'espace des données via un Kernel

Voici quelques exemples des noyaux les plus utilisés :

- **Linéaire** : Lorsque les données sont linéairement séparables, l'espace initial n'a pas besoin d'être changé :

$$K(x_i, x_j) = \langle x_i, x_j \rangle$$

- **Polynomial** : Le noyau polynomial sera utilisé lorsque l'on souhaite non seulement tenir compte des similitudes dans notre échantillon d'apprentissage mais aussi des combinaisons entre nos variables :

$$K(x_i, x_j) = (\langle x_i, x_j \rangle + c)^d$$

- **Gaussian Radial Basis Function (GRBF)** : Le noyau gaussien permet de calculer de similarités dans un espace de dimension infinie. Le paramètre σ permet de calibrer la distance/similarité entre deux points. Plus σ sera grand et plus les points seront « proches ».

$$\begin{aligned} K(x_i, x_j) &= \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \\ &= \exp(-\gamma\|x_i - x_j\|^2) \end{aligned}$$

7.3 Choix des paramètres

Dans cette section, nous allons présenter les paramètres qui ont été retenus lors de l'estimation de notre SVR. Dans ce mémoire, nous avons décidé d'utiliser un noyau gaussien. L'utilisation du noyau gaussien nécessite la connaissance de deux paramètres :

- **C** : un paramètre dit de régularisation permettant d'arbitrer entre la constance de notre fonction et le niveau de déviation autorisé.
- **σ** : un paramètre permettant de calculer la distance entre deux points.

Il n'existe pas de règle précise pour calculer ces paramètres. En général, une validation croisée est utilisée afin de choisir les paramètres optimaux permettant d'obtenir l'erreur d'approximation la plus faible sur la base d'apprentissage.

γ^*	σ^*
0.1	10

TABLE 7.1 – Résultat de la validation croisée

Les réseaux de neurones

8.1 Contexte

Dans cette partie, nous allons présenter les réseaux de neurones qui constituent, avec les SVM, l'une des techniques de *machine learning* les plus utilisées. Les réseaux de neurones se sont largement répandus d'une part, grâce à leur capacité à pouvoir traiter à la fois des problèmes de classification et de régression et d'autre part, pour leur capacité à obtenir de bonnes performances sur la plupart des problèmes traités. Cependant, cette technique peine à se développer dans le secteur de l'assurance du fait de son aspect « boîte noire » qui rend son utilisation et son interprétation parfois compliquées.

Cette méthode d'apprentissage s'inspire de la biologie. Le but d'un réseau de neurones est de synthétiser l'information contenue dans nos variables explicatives afin de prédire une variable à expliquer. Chaque composante du réseau est représentée par un noeud ou neurone lié aux autres au travers d'un poids.

8.2 Neurone formel

L'appellation *réseaux de neurones* fait référence à son analogue en biologie et peut-être visualisé ci-dessous.

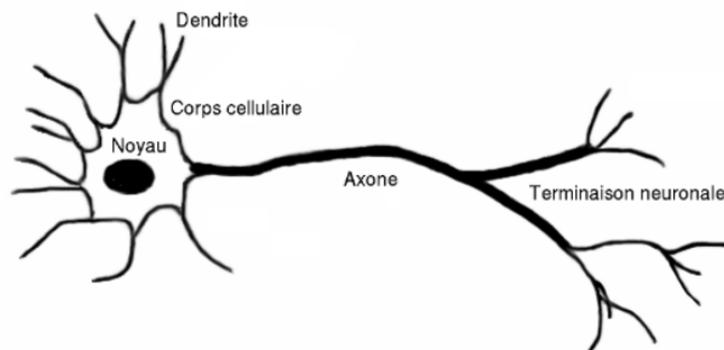


FIGURE 8.1 – Neurone en biologie

Chaque cellule nerveuse dispose d'un canal pour émettre et recevoir des signaux sous forme d'impulsions électriques. Les neurones reçoivent des signaux par des extensions très ramifiées de leur corps cellulaire (les dendrites) et envoient l'information par de longs prolongements (les axones).

Les synapses permettent d'entrer en contact entre deux neurones, reliant les dendrites d'un neurone à l'axone d'un autre neurone. C'est la connexion synaptique qui permet le passage de l'information (excitateur ou inhibiteur).

Un neurone correspond donc à une unité élémentaire de traitement d'un réseau de neurones qui, à une entrée lui associe une information de sortie :

- Les **entrées** du neurone notées $x_1, \dots, x_n \in \mathbb{R}$
Les neurones de la couche d'entrée reçoivent comme information les variables explicatives. Dans notre cas, nous avons normalisé nos données afin d'éviter tout phénomène de valeur extrême. Les données seront donc comprises entre 0 et 1.
- Les **poids** notés $w_i \in \mathbb{R} \forall i \in \{1, \dots, n\}$
Les poids sont estimés au cours de la phase d'apprentissages. La $i^{\text{ème}}$ information parvenant au neurone sera donc $w_i x_i$. Ces poids permettent d'accentuer ou d'inhiber l'information entrante de sorte à ce que certaines informations aient plus de poids dans la détermination de la valeur de sortie y_i .
- La fonction **d'activation** notée $f : \mathbb{R} \rightarrow \mathbb{R}$
Cette fonction va effectuer une transformation d'une combinaison affine des vecteurs de poids et des entrées afin d'obtenir la valeur de sortie. En fonction du choix de la fonction d'activation que nous ferons, les données seront normalisées différemment afin que l'intervalle de variation des données corresponde à l'ensemble de définition de la fonction d'activation.
- La **sortie** y_i correspond à la combinaison linéaire entre les poids et les entrées du réseau de neurone

Le neurone artificiel va effectuer la somme pondérée des données en entrées plutôt que de les considérer séparément. Nous définissons donc la fonction h permettant d'effectuer cette opération comme :

$$h : \mathbb{R} \rightarrow \mathbb{R}$$

$$x \mapsto wx$$

Comme décrit précédemment, le neurone va traiter la quantité précédente via une fonction d'activation f . La fonction que nous cherchons à déterminer est donc de la forme :

$$y = f(h(x))$$

$$= f(w^t x)$$

$$= f(w_0 + \sum_{i=1}^p w_i x_i)$$

où :

w_0 est le biais du neurone
 f une fonction d'activation

Finalement, le réseau de neurones formel peut être décrit à l'aide du graphique suivant :

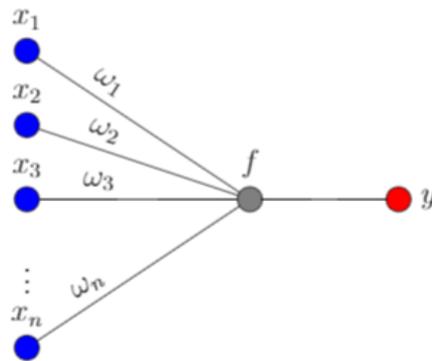
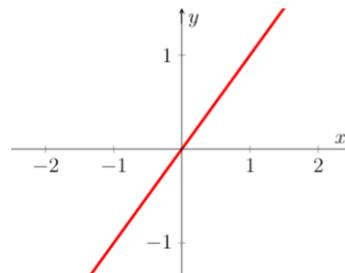


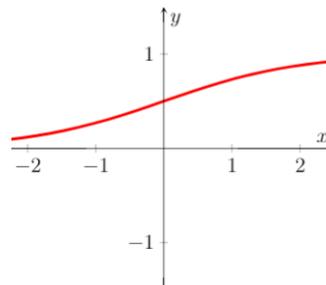
FIGURE 8.2 – Réseau de neurones formel

Comme nous pouvons le constater sur la figure précédente, le choix de la fonction d'activation va déterminer en grande partie la valeur prédite y . Les fonctions d'activation les plus connues sont les suivantes :

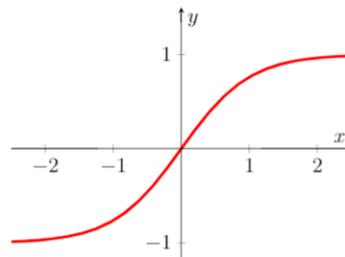
— Linéaire : $f(x) = x$



— Sigmoïde : $f(x) = \frac{1}{1+e^{-x}}$



- Tangente hyperbolique : $f(x) = \tanh(x)$



Les modèles linéaires et sigmoïdaux (Tangente hyperbolique et sigmoïde) sont les plus utilisés et les mieux adaptés aux algorithmes d'apprentissage utilisant une rétro-propagation du gradient car leur fonction d'activation est infiniment différentiable.

8.3 Perceptron multicouche

Le perceptron, inventé en 1957 par Frank Rosenblatt est le réseau de neurone le plus simple. Il s'agit d'un classifieur linéaire ne contenant aucun cycle : cela signifie que le perceptron est monocouche et ne possède qu'une seule sortie à laquelle toutes les entrées sont connectées.

Cette notion de perceptron a été généralisée dans un cadre multicouche. A la différence du cadre précédent, le perceptron multicouche (PMC) contient une ou plusieurs couches cachées et se caractérise par :

- Les **entrées** du neurone notées $x_1, \dots, x_n \in \mathbb{R}$
Les neurones de la couche d'entrée reçoivent comme information les variables explicatives. Dans notre cas, nous avons normalisé nos données afin d'éviter tout phénomène de valeur extrême. Les données seront donc comprises entre 0 et 1.
- Les **poids** notés $w_{ij} \in \mathbb{R} \forall i, j \in \{1, \dots, n\}$
Dans le cas du PMC, le poids w_{ij} correspond à l'importance donnée à une entrée du neurone i d'une couche avec le neurone j de la couche suivante. Nous chercherons les poids optimaux grâce à notre algorithme d'apprentissage afin de minimiser l'erreur commise entre la valeur prédite et la vraie valeur.
- Une ou plusieurs **couches cachées**
Ces couches relient la couche d'entrée à la couche de sortie via une fonction d'activation. L'ajout d'une ou de plusieurs couches vise à augmenter le pouvoir prédictif du modèle. En effet, plus la complexité du modèle augmente, c'est à dire que plus le nombre de couches augmentent et plus notre modèle apprendra de ses erreurs. Cependant, il est préférable de ne pas augmenter trop fortement la complexité de notre modèle afin d'éviter tout phénomène de « surapprentissage » où notre réseau reproduit exactement ce qu'il observe sur l'échantillon d'apprentissage.
- la couche de **sortie** y_i
Cette couche fournit la valeur prédite par notre modèle en fonction des différents inputs d'entrée.

La figure ci-dessous permet d'illustrer le perceptron multicouche (PMC) dans le cas d'une seule couche cachée :

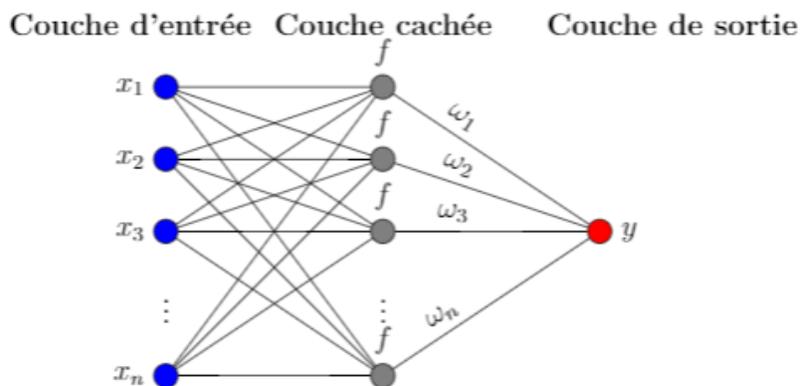


FIGURE 8.3 – Perceptron multi-couche

Le théorème de Cybenko [15], connu sous le nom de *théorème d'approximation universelle* justifie l'utilisation du PMC à au moins une couche cachée pour approcher des fonctions continues.

Theorem 8.3.1. *théorème d'approximation universelle*

Soit $\varphi(\cdot)$ une fonction continue, non-constante, bornée et croissante.

Notons par I_m l'hypercube unitaire de dimension m $[0, 1]^m$.

L'espace des fonctions continues sur I_m est noté $C(I_m)$.

Ainsi, pour $\forall f \in C(I_m)$ et $\varepsilon > 0$, $\exists N, v_i, b_i \in \mathbb{R}$ et $w_i \in \mathbb{R}^m$, où $i = 1, \dots, N$ tel que :

$$F(x) = \sum_{i=1}^N v_i \varphi(w_i^T x + b_i)$$

F est une approximation de la fonction f , où f est indépendants de φ , ce qui signifie que :

$$|F(x) - f(x)| < \varepsilon, \forall x \in I_m$$

En d'autres termes, les fonctions de la forme $F(x)$ sont denses dans $C(I_m)$.

Cela reste vrai lorsque nous remplaçons I_m par n'importe quel sous espace compact de \mathbb{R}^m .

Ce théorème permet de justifier la possibilité d'approcher tout type de fonctions continues par un réseau de neurones. Par ailleurs, un résultat intéressant de ce théorème stipule qu'un simple réseau de neurones à une couche est suffisant pour approcher tout type de fonctions avec une précision donnée.

Cependant, ce théorème ne donne aucune information sur la vitesse de convergence des suites de fonctions F_n vers la fonction f .

8.4 Ajustement du réseau de neurones

Comme vu précédemment, le perceptron multicouche réalise une transformation des variables d'entrées : $Y = \phi(X_1, \dots, X_p; w)$ où w est un vecteur contenant chacun des paramètres w_{jkl} de la $j^{\text{ième}}$ entrée du $k^{\text{ième}}$ neurone de la $l^{\text{ième}}$ couche.

De façon générale, la dernière couche est constituée d'un seul neurone muni d'une fonction d'activation identité tandis que les couches cachées sont munies d'autres fonctions d'activation (sigmoïde ou tangente hyperbolique en général).

Prenons le cas le plus simple d'une base d'entraînement $(x_1, y_1), \dots, (x_n, y_n)$, $x_i, y_i \in \mathbb{R}^n$, avec une couche cachée, un neurone caché et posons :

$$\begin{aligned} \phi(x; w) &= w_2 f(w_1^t x + b_1) + b_2 \\ &= w_2 f(h(x)) \end{aligned}$$

où :

$w_1 \in \mathbb{R}^m$ représente les poids de la couche d'entrée.

$w_2, b_1, b_2 \in \mathbb{R}$ et représentent respectivement le poids de la couche cachée et les biais ajoutés aux couches d'entrée et cachée.

Nous allons voir dans la partie suivante la façon dont ces paramètres seront calculés.

8.4.1 Rétropropagation de l'erreur

Dans cette partie, nous allons présenter l'algorithme utilisé afin de calculer les paramètres optimaux de notre réseau de neurones : l'algorithme de rétro-propagation du gradient.

Comme nous pouvons le constater sur la figure ci-dessous, à l'issue de l'estimation du réseau de neurones, seule la sortie associée à un input donné est observable et donc, l'erreur qui a été commise pour cette estimation. Partant de cette erreur, nous aimerions remonter notre réseau afin d'ajuster les paramètres ayant le plus contribué à cette erreur. L'idée de la rétro-propagation de l'erreur consiste donc à utiliser l'erreur commise, de la dernière jusqu'à la première couche, afin de modifier les poids pour la diminuer. Cette technique revient à corriger les erreurs en réajustant de manière significative les poids synaptiques qui ont contribué à engendrer une erreur importante.

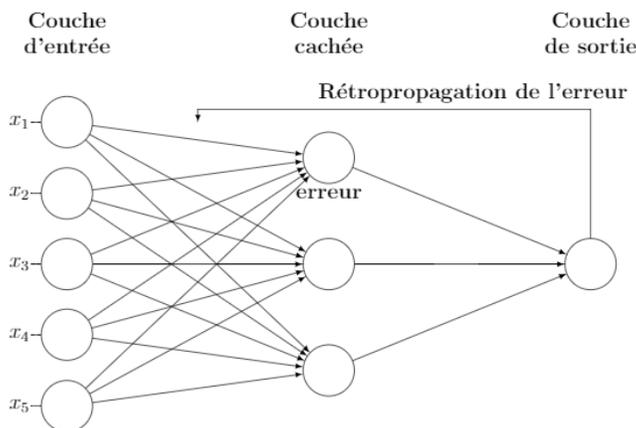


FIGURE 8.4 – Rétropropagation de l'erreur

Nous allons maintenant expliciter le calcul de cet algorithme.

8.4.2 Mise en place de l'algorithme

Une fois les paramètres du modèle identifiés, nous devons fixer un critère de décision afin d'optimiser notre problème. Ce critère de décision ou fonction de perte correspond à une mesure de la différence entre la vraie valeur des données et celle prédite par notre modèle. Lors de l'ajustement du modèle, nous allons chercher à minimiser ce critère.

Nous présentons ci-dessous quelques-unes des fonctions de perte les plus utilisées :

Fonction	Définition
Hinge	$\max(0, 1 - y\hat{y})$
Quadratique	$\ \hat{y} - y\ _2^2$
Logistique	$\ln(1 - e^{y\hat{y}})$

TABLE 8.1 – Fonctions de perte

Dans le cadre de ce mémoire nous avons eu recours à la fonction de perte quadratique. Notons par L la fonction de perte quadratique que nous allons chercher à minimiser :

$$L(\hat{y}, y, w, b) = \frac{1}{2} \|\hat{y} - y\|_2^2 - \alpha \|w\|_2^2$$

où $\alpha \|w\|_2^2$ est une régularisation de type L^2 permettant de pénaliser les modèles trop complexes et $\alpha > 0$ est un hyperparamètre contrôlant la magnitude de la pénalité.

La résolution du programme précédent nous donnera les paramètres optimaux servant au calibrage de notre modèle. Une des techniques les plus utilisées en apprentissage statistique pour résoudre ce type de problèmes consiste à recourir à la méthode de *descente de gradient*. Cette classe regroupe des méthodes itératives qui permettent de converger vers un optimum. Le défaut principal de cette méthode réside dans le fait que l'optimum trouvé peut être local et non global.

La figure ci-dessous permet d'illustrer la descente de gradient :

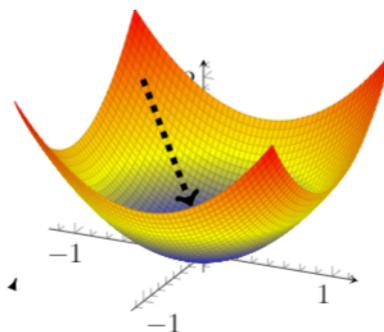


FIGURE 8.5 – Descente de gradient

Comme nous pouvons le voir sur la figure précédent, nous commençons par initialiser aléatoirement nos paramètres afin de donner un point de départ à notre algorithme puis, au fur et à mesure que les itérations augmentent, les paramètres se modifient de sorte à minimiser le risque et converger vers la solution.

8.4.3 Résolution du problème

Dans cette partie, nous allons présenter la façon dont nous résolvons le problème précédent par descente du gradient. Pour ce faire, nous devons évaluer les dérivées partielles de la fonction de perte :

$$\begin{aligned}\nabla_w L(\hat{y}, y, w, b) &= \frac{\partial}{\partial w_{jk}} L(\hat{y}, y, w, b) \\ \nabla_b L(\hat{y}, y, w, b) &= \frac{\partial}{\partial b_j} L(\hat{y}, y, w, b)\end{aligned}$$

Sachant évaluer les gradients, nous pouvons calculer une solution itérative de notre problème. En tout point de l'espace des paramètres, le vecteur gradient de L pointe dans la direction de l'erreur croissante. Pour faire décroître L , il suffit de se déplacer en sens contraire en modifiant les poids à chaque étape. L'algorithme du gradient cherche donc à converger de manière itérative vers les paramètres optimaux :

$$\begin{aligned}w^{(i+1)} &= w^{(i)} - \alpha \nabla_w L^{(i)}(\hat{y}, y, w, b) \\ b^{(i+1)} &= b^{(i)} - \alpha \nabla_b L^{(i)}(\hat{y}, y, w, b)\end{aligned}$$

L'algorithme est résolu comme suit :

- Initialiser les poids w_{jkl} par tirage aléatoire selon une loi uniforme $[0,1]$
- Tant que $L < \text{errmax}$ ou $\text{niter} < \text{itermax}$ faire :
 - Ranger la base d'apprentissage dans un nouvel ordre aléatoire
 - Pour chaque élément de la base faire :
 - $\epsilon(i) = y_i - \phi(x_i^1, \dots, x_i^p; w(i-1))$. Les entrées sont propagées vers l'avant. L'erreur est « rétro-propagée » dans les différentes couches afin d'affecter à chaque entrée une responsabilité dans l'erreur globale.
 - Mise à jour des poids : $w_{jkl}(i) = w_{jkl}(i-1) + \Delta w_{jkl}(i)$
 - fin pour
- fin tant que

L'algorithme précédent est une méthode dite de premier ordre et nécessite de fixer le paramètre α représentant le taux d'apprentissage.

Dans le cadre de ce mémoire, nous avons eu recours à l'algorithme L-BFGS (Limited Memory Broyden–Fletcher–Goldfarb–Shanno) qui consiste à faire une approximation au second ordre. Le nouvel ajustement des poids devient donc :

$$\begin{aligned}w^{(i+1)} &= w^{(i)} - [H(L(\hat{y}, y, w))]^{-1} \nabla_w L^{(i)}(\hat{y}, y, w) \\ b^{(i+1)} &= b^{(i)} - [H(L(\hat{y}, y, w))]^{-1} \nabla_b L^{(i)}(\hat{y}, y, w)\end{aligned}$$

où $H(L(\hat{y}, y, w))$ désigne la matrice hessienne de $L(\hat{y}, y, w)$.

Cette méthode permet de résoudre des problèmes d'optimisation non-linéaire tout en limitant le coût de stockage des données, ce qui la rend particulièrement utile en *machine learning*. En effet, elle ne cherche pas à calculer directement la matrice Hessienne mais l'approxime. A la différence des techniques de premier ordre, elle permet d'obtenir des ajustements de poids plus précis via le calcul de la matrice Hessienne, décrivant la courbure locale de la fonction de perte.

Quatrième partie

Application du proxy pour l'estimation du capital économique

Application des chocs

Cette partie s'attache à présenter les différentes fonctions de proxy qui ont été mises en place afin de prédire le RC . Nous nous intéresserons dans un premier temps à la présentation du proxy mis en place sur le SCR avant de présenter brièvement la fonction de proxy sur les FP .

Comme nous avons pu le voir durant cette première partie et au travers de la formule standard, le SCR de la compagnie d'assurance est complexe à calculer car, dépend de beaucoup de paramètres (aspect modulaire de la formule standard), c'est la raison pour laquelle, deux approches ont été envisagées pour déterminer ce proxy sur le besoin en capital :

- Approximer directement le SCR global
- Approximer le BE et les sous-modules de risque de la formule standard avant réagrégation

9.1 Apprentissage sur le SCR

Le but de ce mémoire est de pouvoir fournir un proxy du RC en fonction de paramètres économiques observés et attendus sur le marché. Le ratio de couverture étant défini comme :

$$RC_t = \frac{SCR_t}{FP_t}$$

nous avons décidé de l'approximer en passant par le SCR et les FP en ajustant une fonction de la forme :

$$\begin{aligned} \tilde{S}CR_t &= f_t(\epsilon_1, \dots, \epsilon_n) \\ \tilde{F}P_t &= f'_t(\epsilon_1, \dots, \epsilon_n) \end{aligned}$$

Nous permettant ainsi d'obtenir le ratio de couverture attendu comme :

$$\tilde{R}C_t = \frac{\tilde{S}CR_t}{\tilde{F}P_t}$$

Cependant, pour pouvoir approximer le SCR en fonction de n variables (14 dans notre cas) suppose de pouvoir ajuster une telle fonction à nos données comme par exemple, un polynôme de degré n . Pour un n suffisamment grand, il est impossible de trouver une telle fonction à moins de supposer que notre SCR varie linéairement avec nos variables explicatives.

Les graphiques ci-dessous présentent, pour chacun de nos 100 portefeuilles, la valeur du capital économique correspondant à chacune des situations économiques construites précédemment (graphique de gauche) et d'autre part, le taux 8 ans en fonction de la valeur du capital économique.

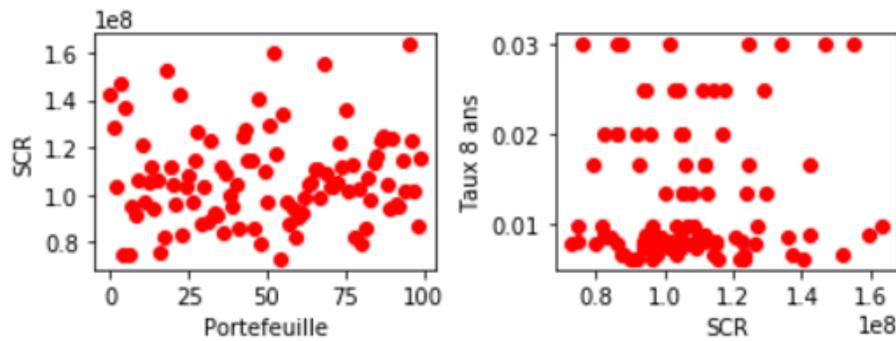


FIGURE 9.1 – Non linéarité du SCR

Comme nous pouvons le constater au travers de ces deux graphiques, le *SCR* présente une forte non-linéarité. En effet, lorsque nous voulons expliquer la variation du capital économique en fonction du taux 8 ans, nous remarquons qu'il n'est pas possible de calibrer une fonction linéaire permettant d'ajuster nos données. Ceci peut s'expliquer par le fait qu'il existe une forte interdépendance entre nos variables comme nous avons pu l'expliquer précédemment.

Par exemple, pour un taux 8 ans donné, le *SCR* ne sera pas le même en fonction du *TMG* puisque lorsque la différence entre le taux servi par l'assureur et celui servi sur le marché devient trop grand, cela déclenche des mécanismes de rachats dynamiques des contrats d'assurances. Ce mécanisme oblige la compagnie d'assurance à piocher dans ses liquidités, se faisant, vendant des actifs à une date donnée qui, peuvent aboutir à des moins-values. Approximer directement le *SCR* ne permet donc pas de capturer les effets sous-jacents propres à chaque module de risque.

La figure ci-dessous permet d'illustrer la méthode que nous avons voulu mettre en place en réalisant un apprentissage direct.

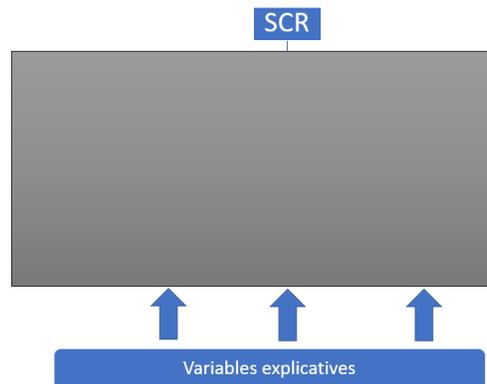


FIGURE 9.2 – Méthode « boîte-noire »

Comme l'illustre la figure et l'exemple précédent, ce type d'approximation s'apparente à une méthode dite de « boîte-noire » où l'assureur serait incapable d'expliquer les variations induites par chacun des changements d'hypothèses économiques. Par exemple, si nous nous attendons à une forte remontée des taux, toutes choses égales par ailleurs, nous obtiendrons seulement une valeur terminale mais nous ne serons pas capables d'expliquer le module ayant été impacté le plus significativement par cette remontée des taux. D'autre part, ce manque de transparence nous empêche de vérifier que nos intuitions en termes de variation par sous-modules est vérifiée, ce qui constitue un premier indicateur permettant de tester l'exactitude de notre modèle.

Pour les raisons mentionnées précédemment, nous avons décidé de réaliser un apprentissage sur chacun des sous-modules. Cela permet de mieux capter les interactions entre nos variables et pouvoir, à chaque niveau, expliquer et mieux comprendre la variation d'un paramètre comme, par exemple, l'impact du *TMG* sur le besoin en capital réglementaire. Il paraît plus logique de considérer que sur un sous-ensemble, nos variables bougent de manière linéaire plutôt que de faire l'hypothèse sur tout l'espace. Nous effectuons donc des estimations sur une partition de notre espace total, à savoir sur chacun des sous-modules de la formule standard.

D'autre part, faire de l'apprentissage à un niveau plus granulaire permet d'éviter de se retrouver avec un outil de type « boîte noire » où nous obtenons seulement la valeur terminale sans maîtriser aucuns résultats.

La figure ci-dessous permet d'illustrer la seconde méthode, où nous reprenons la formule standard pour déterminer le *SCR*. Cette méthode permet une approximation plus granulaire de nos chocs et ainsi une meilleure compréhension des interactions entre nos variables.

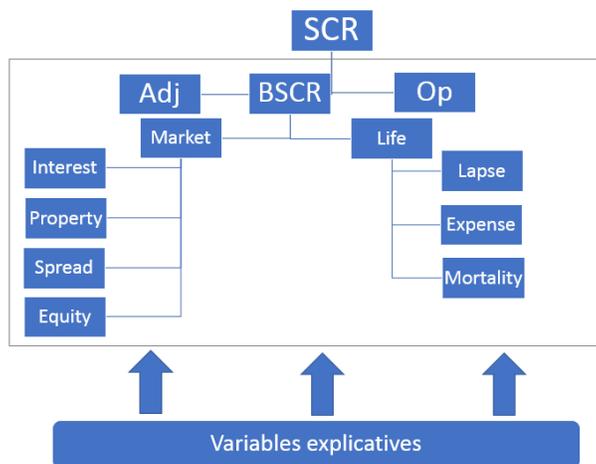


FIGURE 9.3 – Méthode proxy

9.2 Apprentissage sur les sous-modules

Comme présenté précédemment, nous nous intéressons à un portefeuille d'épargne vie, c'est la raison pour laquelle nous ne calculerons que des chocs sur les modules marché et vie. En effet, ces modules représentent la majeure partie du SCR global. Nous considérerons donc comme négligeable le risque de défaut de contrepartie et certains autres risques tels que le module de concentration, de longévité, d'incapacité, de révision ou encore de catastrophe.

9.2.1 SCR de marché brut

Pour calculer le *SCR* brut, l'article 83 des actes délégués précise :

- Lorsque le calcul d'un module ou d'un sous-module du capital de solvabilité requis de base se fonde sur l'impact d'un scénario sur les fonds propres de base de l'entreprise d'assurance ou de réassurance, toutes les hypothèses suivantes sont faites dans ce calcul :
 - (a) le scénario ne modifie pas le montant de la marge de risque incluse dans les provisions techniques ;
 - (b) le scénario ne modifie pas la valeur des actifs et des passifs d'impôts différés ;
 - (c) le scénario ne modifie pas le montant de la valeur des prestations discrétionnaires futures incluses dans les provisions techniques ;

D'un point de vue théorique certaines simplifications apparaissent et permettent de diminuer le nombre de lancement de modèle nécessaires à l'ensemble des calculs. En effet, les prestations discrétionnaires futures, par définition, restent identiques à celles du scénario central (articles 83 et 206 des actes délégués).

En effet, on sait que :

$$SCR_i^{brut} = (Actif_{scenario-central} - BEL_{scenario-central}) - (Actif_{scenario-choque} - BEL_{scenario-choque})$$

Comme cela a été décrit dans le chapitre 1, le *BEL* se décompose comme la somme d'une partie discrétionnaire et d'une partie garantie : *FDB* + *BEG*. La relation devient donc :

$$SCR_i^{brut} = (Actif_{scenario-central} - BEG_{scenario-central} - FDB_{scenario-central}) - (Actif_{scenario-choque} - BEG_{scenario-choque} - FDB_{scenario-choque})$$

Or, on sait que $FDB_{scenario-central} = FDB_{scenario-choque}$ d'après l'article 83 point (c). On obtient donc :

$$SCR_i^{brut} = (Actif_{scenario-central} - BEG_{scenario-central}) - (Actif_{scenario-choque} - BEG_{scenario-choque})$$

Le calcul du Best Estimate Garanti est effectué pour les scénarii suivants :

- Scénario Central
- Scénario choc de taux à la hausse/baisse
- Scénario choc de souscription (i.e. hausse et baisse des rachats, rachat massif, mortalité, longévité, CAT et hausse des frais)

Pour les autres sous-modules de risques (notamment choc action, immobilier, devise et spread), le Best Estimate Garanti choqué est supposé égal au Best Estimate Garanti central. La formule devient donc :

$$SCR_i^{Brut} = Actif_{scenario-central} - Actif_{scenario-choque}$$

Ainsi, en cas de choc de marché (hors risque de taux), le BEG étant inchangé, le SCR brut revient à un calcul de variation de valeur de marché de l'actif comme nous allons le voir dans les sous-sections suivantes.

9.2.1.1 Choc de taux

Dans cette partie, nous allons présenter la façon de déterminer le besoin en capital au titre du risque de taux d'intérêt. Le risque de taux d'intérêt concerne tous les actifs et passif donc la valeur est sensible à un changement du niveau des taux d'intérêt. Les actifs concernés par ce risque sont :

- Les investissements à revenu fixe : OAT, OTF et OPCVM obligataire
- Les produits dérivés de taux
- Les instruments de financement : emprunt

Notre portefeuille étant composé d'OAT et d'OTF, nous allons présenter les méthodes permettant de les valoriser suite à un choc de taux à la baisse.

Une Obligation Assimilable du Trésor (OAT) est un emprunt d'état émis pour une durée de 5 à 50 ans. Il s'agit d'un titre de dette détenu sur un état.

Quant à « Organised Trading Facility » (OTF), il englobe toutes les formes organisées d'exécution et de négociation ne correspondant pas aux spécifications réglementaires des plates-formes existantes (plates-formes d'exécution de swaps, systèmes de croisement d'ordres ou de gré à gré).

Nous présentons ci-dessous les méthodes de calcul permettant de valoriser une obligation suite à un choc.

$$\text{Flux}_i^{central} = \text{Coupon}_i \sum_{j=1}^{2i} \left(\frac{1}{1 + r_j^{central} + spread_j} \right)^{\frac{\text{Date}_j^{achat} - \text{Date}_j^{echeance}}{365.25}} + \left(\frac{1}{1 + r_{2i}^{central} + spread_{2i}} \right)^{\frac{\text{Date}_{2i}^{achat} - \text{Date}_{2i}^{echeance}}{365.25}}$$

$$\text{Flux}_i^{down} = \text{Coupon}_i \sum_{j=1}^{2i} \left(\frac{1}{1 + r_j^{down} + spread_j} \right)^{\frac{\text{Date}_j^{achat} - \text{Date}_j^{echeance}}{365.25}} + \left(\frac{1}{1 + r_{2i}^{down} + spread_{2i}} \right)^{\frac{\text{Date}_{2i}^{achat} - \text{Date}_{2i}^{echeance}}{365.25}}$$

où :

$\text{Flux}_i^{central}$ correspond aux flux de l'obligation valorisée avec la courbe des taux sans choc

Flux_i^{down} correspond aux flux de l'obligation valorisée avec une courbe des taux choqué à la baisse

$r_j^{central}$ le taux d'intérêt central de l'obligation j

r_j^{down} le taux d'intérêt choqué à la baisse de l'obligation j

La valeur de marché de l'obligation à la suite d'un choc correspond donc à la différence entre la valeur de marché en central et la valeur de marché calculé à partir d'une courbe des taux à la baisse, comme illustré par la formule ci-dessous :

$$\begin{aligned}
 VM_{obligation}^{choc} &= VM_{central} - VM_{choc} \\
 &= \sum_{i=1}^n \text{Nominal}_i \left(\text{Flux}_i^{central} - \text{Flux}_i^{down} \right)
 \end{aligned}$$

Le SCR taux brut s'obtient donc en prenant en compte la variation de valeur de marché et de BEG à la suite d'un changement du niveau des taux :

$$SCR_{taux}^{brut} = \Delta VM_{OAT} + \Delta VM_{OTF} - \Delta BEG$$

où :

$$\begin{aligned}
 \Delta BEG &= BE_t^{central} - BE_t^{choc} \\
 &= f_t(\Delta BEG)
 \end{aligned}$$

où la variable ΔBEG est approximée par nos méthodes d'apprentissages statistiques présentées dans les parties précédentes.

9.2.1.2 Choc de spread

En plus du besoin en capital au titre du risque de taux, la réglementation prévoit un besoin en capital concernant le risque de spread, devant prendre en compte la dégradation de la qualité de crédit de l'émetteur. Ce besoin en capital est fonction des paramètres suivant :

- la valeur de marché du titre
- la durée
- la notation de l'émetteur

Le tableau ci-dessous permet de donner les paramètres permettant de calculer le besoin en capital au titre du spread par formule fermée.

Échelon de qualité de crédit		0		1		2		3		4		5 et 6	
Duration (dur_i)	$stress_i$	a_i	b_i										
Jusqu'à 5 ans	$b_i \cdot dur_i$	—	0,9 %	—	1,1 %	—	1,4 %	—	2,5 %	—	4,5 %	—	7,5 %
Supérieure à 5 et inférieure ou égale à 10 ans	$a_i + b_i \cdot (dur_i - 5)$	4,5 %	0,5 %	5,5 %	0,6 %	7,0 %	0,7 %	12,5 %	1,5 %	22,5 %	2,5 %	37,5 %	4,2 %
Supérieure à 10 et inférieure ou égale à 15 ans	$a_i + b_i \cdot (dur_i - 10)$	7,0 %	0,5 %	8,5 %	0,5 %	10,5 %	0,5 %	20,0 %	1,0 %	35,0 %	1,8 %	58,5 %	0,5 %
Supérieure à 15 et inférieure ou égale à 20 ans	$a_i + b_i \cdot (dur_i - 15)$	9,5 %	0,5 %	11 %	0,5 %	13,0 %	0,5 %	25,0 %	1,0 %	44,0 %	0,5 %	61,0 %	0,5 %
Plus de 20 ans	$\min[a_i + b_i \cdot (dur_i - 20); 1]$	12,0 %	0,5 %	13,5 %	0,5 %	15,5 %	0,5 %	30,0 %	0,5 %	46,6 %	0,5 %	63,5 %	0,5 %

FIGURE 9.4 – Choc de spread

Par exemple, pour une OTF ayant une durée de 6 ans et une qualité de crédit de 1, le calcul sera :

$$\begin{aligned} SCR_{spread} &= VM_{OTF} \times \text{choc-spread} \\ &= VM_{OTF} \times (5.5\% + 0.6\% \times (6 - 5)) \end{aligned}$$

9.2.1.3 Choc actions

L'exigence de capital correspond à la variation de la NAV suite à une diminution de la valeur de marché des actions de 39%. Afin d'éviter les comportements pro-cycliques, qui remettraient en cause le rôle fondamental des assureurs dans le financement de l'économie, ce choc est corrigé par un mécanisme d'ajustement symétrique ou dampener. Le but du dampener est d'atténuer le choc action lorsque les marchés sont bas de cycle, et de l'augmenter lorsque les marchés sont en haut de cycle et que la probabilité de baisse est importante. Dans le cadre de cette étude, nous allons appliquer un choc action de 36.76%. Autrement dit, le choc sera calculé comme :

$$SCR_{actions}^{brut} = VM_{actions} \times 36.76\%.$$

9.2.1.4 Choc immobilier

L'exigence de capital correspond à la variation de la NAV suite à une diminution de la valeur de marché de l'immobilier de 25% et se calcul comme suit :

$$SCR_{immobilier}^{brut} = VM_{immobilier} \times 25\%.$$

9.2.1.5 Agrégation des chocs

La formule standard étant modulaire, pour obtenir le SCR de marché brut, nous avons d'abord dû calculer le besoin en capital sur chacun des sous-modules. Une fois ces sous-modules réagréés via une matrice de corrélation règlementaire, nous obtenons le besoin en capital au titre du risque de marché. En ce qui concerne le risque de marché, il existe deux matrices de corrélations :

- une matrice de corrélation up prenant en compte les corrélations au sein du module de marché suite à un risque de taux à la hausse
- une matrice de corrélation down prenant en compte les corrélations au sein du module de marché suite à un risque de taux à la baisse

Il est à noter que, dans le cadre de cette étude, nous considérons une matrice de corrélation de marché dans le cas d'un choc de taux à la baisse où les sous modules devise et concentration sont considérés comme nuls car négligeables par hypothèse.

$$SCR_{marche}^{brut} = \sqrt{\sum_{i,j} \rho_{i,j} SCR_i^{brut} SCR_j^{brut}}$$

avec :

$$\rho_{i,j} = \begin{pmatrix} & \textit{Interest} & \textit{Equity} & \textit{Property} & \textit{Spread} & \textit{Currency} & \textit{Concentration} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0.75 & 0.75 & 0.25 & 0 \\ 0 & 0.75 & 1 & 0.5 & 0.25 & 0 \\ 0 & 0.75 & 0.5 & 1 & 0.25 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{pmatrix}$$

$$SCR_{i,j} = \begin{pmatrix} SCR_{Taux} & SCR_{Actions} & SCR_{Immobilier} & SCR_{Spread} & 0 & 0 \end{pmatrix}$$

9.2.2 Calcul du SCR de marché net

Nous distinguons les calculs de SCR net et de SCR brut par le fait que le SCR brut ne tiennent pas compte de la capacité d'absorption des pertes par les provisions techniques (les prestations discrétionnaires futures sont identiques à celles versées dans le scénario central), alors que le SCR net tient compte de l'absorption grâce à une diminution éventuelle des prestations discrétionnaires futures résultant des pertes observées lors du choc.

$$\begin{aligned} SCR_{sous-module}^{net} &= (VM_{centrale}^{sous-module} - VM_{choc}^{sous-module}) - (BEL_{central} - BEL_{choc}) \\ &= (VM_{centrale}^{sous-module} - VM_{choc}^{sous-module}) - (BEG_{central} - BEG_{choc}) - (FDB_{central} - FDB_{choc}) \\ &= SCR_{sous-module}^{brut} - (FDB_{central} - FDB_{choc}) \\ &= SCR_{sous-module}^{brut} - \Delta FDB \\ &\approx SCR_{sous-module}^{brut} - f_t(\Delta FDB) \end{aligned}$$

où $f_t(\Delta FDB)$ correspond à l'effet d'absorption d'un sous module obtenu par apprentissage statistique en fonction de nos variables d'intérêts.

Une fois ce besoin en capital au titre de chaque sous-module obtenu, nous les agrégerons via la matrice de corrélation afin d'en obtenir le SCR de marché net :

$$SCR_{marche}^{net} = \sqrt{\sum_{i,j} \rho_{i,j} SCR_i^{net} SCR_j^{net}}$$

9.2.3 SCR de souscription vie

La méthodologie de calcul utilisée pour le module de souscription vie est analogue à ce qui a été présenté précédemment. Le module de risque de souscription vie regroupe l'ensemble des risques élémentaires provenant de la souscription des contrats d'assurance vie. Dans le cadre de cette étude d'un portefeuille d'épargne, nous avons considéré les risques de rachat, de frais et de mortalité :

- SCR Frais : Le risque de frais correspond au fait que l'assureur a pu mal évaluer ses dépenses futures, ce qui pourra entraîner une sous-estimation de ses décaissements.
- SCR Rachat : Le risque de rachat vise à prendre en compte la variation des taux de rachats, de résiliation et de réduction des couvertures des contrats souscrits par les assurés.
- SCR Mortalité : Le risque de mortalité vise à prendre en compte la sous/surestimation des provisions techniques engendrées par une mauvaise estimation de la mortalité des assurés en portefeuille.

Ces trois SCR brut ont été calculés de la même façon :

$$SCR_{sous-module}^{brut} = f_t(\epsilon_1, \dots, \epsilon_n)$$

où les ϵ_i représentent nos variables explicatives et $f_t(\cdot)$ représente notre fonction d'apprentissage statistique.

9.2.3.1 Agrégation des chocs

Comme pour le SCR de marché, nous allons avoir recours à une matrice de corrélation afin d'obtenir le SCR vie brut comme suit :

$$SCR_{vie}^{brut} = \sqrt{\sum_{i,j} \rho_{i,j} SCR_i^{brut} SCR_j^{brut}}$$

avec :

$$\rho_{i,j} = \begin{pmatrix} \text{Mortalité} & \text{Longévité} & \text{Incap/Inval} & \text{Rachats} & \text{Frais} & \text{Révisions} & \text{Cat} \\ \begin{pmatrix} 1 & -0.25 & 0.25 & 0 & 0.25 & 0 & 0.25 \\ 0.25 & 1 & 0 & 0.25 & 0.25 & 0.25 & 0 \\ 0.25 & 0 & 1 & 0 & 0.5 & 0 & 0.25 \\ 0 & 0.25 & 0 & 1 & 0.50 & 0 & 0.25 \\ 0.25 & 0.25 & 0.50 & 0.50 & 1 & 0.50 & 0.25 \\ 0 & 0.25 & 0 & 0 & 0.50 & 1 & 0 \\ 0.25 & 0 & 0.25 & 0.25 & 0.25 & 0 & 1 \end{pmatrix} \end{pmatrix}$$

Après calcul des effets d'absorptions comme précédemment, nous obtenons le SCR de vie net ci-dessous :

$$SCR_{vie}^{net} = \sqrt{\sum_{i,j} \rho_{i,j} SCR_i^{net} SCR_j^{net}}$$

9.2.4 SCR

Conformément à l'article 103 de la directive, le capital de solvabilité requis calculé selon la formule standard est :

$$SCR = BSCR + Adj + SCR_{Op}$$

- **BSCR** représente le SCR de base.
- **Adj** représente les ajustements dû à l'effet d'absorption des risques induits par les mécanismes de participations aux bénéficiaires et d'impôts différés.
- **SCR_{Op}** représente le besoin en capital au titre du risque opérationnel, ce qui correspond aux pertes résultantes de procédures internes, de membres du personnel ou de systèmes inadéquats ou défaillants, ou d'événements extérieurs. Il se calcule de la façon suivante :

$$SCR_{Ope} = \min(0.3 \times BSCR; Op) + 0,25Exp_{ul}$$

où :

- Op correspond au capital de base calculé en fonction des primes acquises au cours des 12 derniers mois et des provisions techniques vie
- Exp_{ul} correspond aux dépenses encourues au cours des 12 derniers mois pour les contrats d'assurance vie où le risque est porté par les assurés

où, conformément à l'article 104 de la directive, le capital de solvabilité requis de base ou le $BSCR$ est calculé par agrégation des différents modules de risques, en utilisant la matrice de corrélation définie par la directive :

$$BSCR = \sqrt{\sum_{i,j} \rho_{i,j} SCR_i^{brut} SCR_j^{brut}}$$

Résultats d'estimations en $t=0$

Dans cette section, nous allons présenter les résultats d'estimations de nos apprentissages au temps $t = 0$. Comme nous l'avons explicité au chapitre 1, la fonction de proxy qui sera utilisée lors du calcul de l'ORSA se base sur la fonction estimée à l'instant $t - 1$, déformée d'un certain montant, afin de prendre en compte la perte de valeur des actifs sur l'horizon du business plan. Ainsi, estimer avec précision nos différents indicateurs en $t = 0$ devient primordial, puisque de cette estimation va dépendre l'ensemble de nos *BE*, *SCR* et *FP*.

Les résultats d'estimations portent sur les trois méthodes d'apprentissage statistique suivantes :

- LASSO
- réseau de neurone
- SVM

Nous avons, dans un premier temps, découpé notre base de données en deux bases distinctes :

- une base d'apprentissage
- une base de test

La base d'apprentissage contient 80% de nos données, soit 80 portefeuilles et la base de test contient les 20% restantes. La base d'apprentissage a été utilisée dans le but de calibrer nos modèles, c'est à dire qu'ils ont appris des « situations » sur notre sous-ensemble de donnée afin d'en extraire des relations entre nos variables. Suite à l'apprentissage qui a été fait de ces relations, ils arrivent à prédire la variable d'intérêt. La base de test nous sert à challenger notre modèle, c'est à dire que cette base sert à s'assurer que nos modèles ont réussi à apprendre de nos portefeuilles sans pour autant faire du surapprentissage.

Le surapprentissage est un phénomène qui se produit en *machine learning* lorsqu'un algorithme ajuste parfaitement les données de la base d'apprentissage. Le modèle ayant appris « par coeur » des événements, il n'est pas capable de donner des estimations satisfaisantes lorsque les données servant à faire de la prédiction sont différentes des données ayant servis à calibrer le modèle.

Une des façons de détecter le surapprentissage consiste à comparer les erreurs de prévisions entre les deux bases :

- Si l'erreur de prévision de notre modèle diminue sur la base d'apprentissage tandis que l'erreur augmente significativement sur la base de test, alors le modèle est à questionner..
- Si les erreurs d'apprentissages sont du même ordre de grandeur entre les deux bases alors le modèle ajuste correctement.

La figure ci-dessous permet d'illustrer le phénomène de surapprentissage.

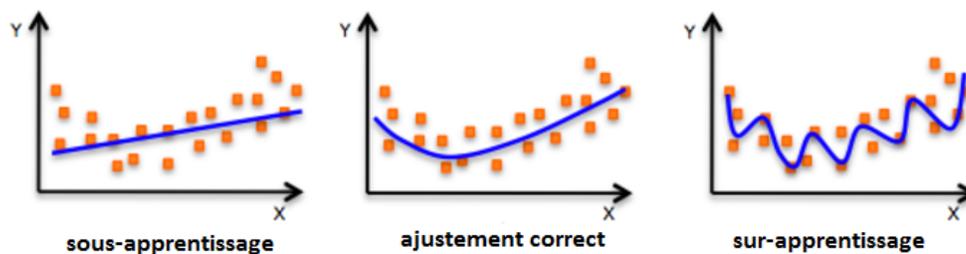


FIGURE 10.1 – Phénomène d'ajustement

Les bases ont été construites par tirage aléatoire successif, limitant le risque d'avoir des données trop proches en apprentissage.

Pour comparer la performance de nos différentes méthodes, nous auront recours aux métriques suivantes :

- MAE
- L'erreur minimum
- L'erreur maximum
- La variance

Le MAE ou Mean Absolute Error correspond à l'erreur moyenne en valeur absolue et est défini de la façon suivante :

$$\frac{\sum_{i=1}^n |y_i^{true} - y_i^{pred}|}{n}$$

Cette quantité correspond à une moyenne arithmétique des écarts en valeur absolue entre la valeur prédite et la vraie valeur. Contrairement à une moyenne arithmétique simple, le MAE permet d'éviter de compenser les erreurs d'approximations, empêchant d'avoir l'illusion que notre modèle performe trop bien.

La variance se calcule comme :

$$V(X) = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$$

avec m la moyenne de notre série

La variance permet de nous rendre compte de la variabilité de notre série, à savoir l'erreur de prévision dans notre cas. Une variance faible impliquerait donc une erreur maîtrisée puisque stable autour d'une valeur.

Enfin, l'erreur minimum et maximum nous permettront de réaliser une analyse plus poussée dans la compréhension de nos portefeuilles et des variables créant le plus d'erreurs. Une fois les portefeuilles identifiés, nous pourrons analyser leurs caractéristiques afin de pouvoir nuancer nos méthodes et de définir un cadre dans lequel notre modèle peut être jugé comme « performant » ou dans lequel nous devons prêter une attention particulière.

Dans la suite de l'analyse, une erreur de 10% sur la durée du business plan sera jugée comme acceptable. Ce seuil est subjectif et est propre à chaque compagnie d'assurance. Les résultats d'estimations seront ceux obtenus sur la base de test, soit les 20% des données restantes.

10.1 Résultats obtenus pour le BE

Le tableau ci-dessous nous donne les erreurs d'approximations en pourcentage entre la vraie valeur du BE, obtenue via par le modèle et celle prédite par nos méthodes de *machine learning*. A noter que l'outil de modélisation est celui utilisé par Actuaris en interne dénommé ADDACTIS Modeling.

Métrique	LASSO	Réseau de neurone	Support Vector Regressor
MAE	0.5%	0.1%	0.9%
Minimum	-0.9%	-0.7%	-2.8%
Maximum	1.6%	1.1%	3.3%
Variance	0.09%	0.06%	0.1%

TABLE 10.1 – Erreur d'approximation sur la base de test

Comme nous pouvons le constater, le réseau de neurone donne les résultats les plus précis et les plus robustes sur l'estimation du *BE*. En effet, l'erreur moyenne en valeur absolue est très proche de 0 tout comme la variance, ce qui traduit une volatilité quasiment nulle dans l'erreur d'estimation. Cette faible variance traduit donc une erreur maîtrisée. De même, le Lasso donne des résultats robustes. Enfin, en ce qui concerne les résultats d'estimations par SVR, l'erreur moyenne est de 1% avec une erreur allant de -2.9% à 3.4%. Le réseau de neurone est donc le meilleur modèle selon nos critères pour prédire le *BE* en fonction de nos variables d'intérêts.

Il est utile de rappeler qu'une erreur de 1% sur le *BE* équivaut à une erreur d'environ 10% sur les fonds propres. En effet, si nous supposons que le passif de notre compagnie d'assurance vaut 100 et que les *FP* et *BE* représentent respectivement 10% et 90% des passifs (comme c'est le cas pour notre compagnie d'assurance), alors, commettre une erreur de 1% sur les *BE* se traduira comme suit :

$$\begin{aligned}
 BE_{\text{estimé}} &= 89.1 \\
 FP_{\text{estimé}} &= \text{Valeur du passif} - BE_{\text{estimé}} \\
 &= 100 - 89.1 \\
 &= 10.9
 \end{aligned}$$

Ce qui équivaut à une erreur de 9% sur les *FP*. Cet exemple illustre la nécessité d'avoir une erreur sur les *BE* qui soit « maîtrisée ». Le tableau précédent nous montre que notre erreur d'estimation est en effet bien « maîtrisée » sur les *BE* puisque la moyenne et la variance sont proches de 0, attestant du caractère peu volatil de notre modèle.

10.2 Résultats obtenus pour le SCR

Dans cette partie, nous allons présenter les résultats d'estimation des proxies nous ayant permis d'estimer le *SCR* global à l'aide des sous-modules de risques.

10.2.1 Module marché

Le tableau ci-dessous permet de comparer les SCR de marché brut obtenu via notre modèle simplifié et ceux obtenus via le logiciel ADDACTIS Modeling.

Méthode	Métrique	Ecart BE	SCR _{Actions}	SCR _{Immobilier}	SCR _{Spread}	SCR _{Taux}	SCR _{marche}
NN	MAE	0.2%	0%	0%	0.2%	0.3%	0.1%
SVR	MAE	6%	0%	0%	2.3%	0.8%	0.9%
LASSO	MAE	0.4%	0%	0%	2.3%	0.1%	0.6%

TABLE 10.2 – Erreur d'approximation SCR brut marché

Comme nous pouvons le constater sur le tableau d'estimation du *SCR* de marché, l'erreur d'approximation en valeur absolue est la même selon les trois modèles sur les modules Actions, Immobilier et Spread. En effet, ces trois modules se calculant par formules fermées, les erreurs sont logiquement identiques, et égale à 0. En ce qui concerne le module de taux, le réseau de neurone est à nouveau le meilleur modèle puisque son erreur d'approximation est la plus faible. Le réseau de neurone semble bien capter l'effet de la variation du *BEG* tandis que le SVR n'y arrive pas.

Finalement, les erreurs d'approximation sur le *SCR* de marché restent correctes sur les trois méthodes, avec une erreur de prévision inférieure à 1% en valeur absolue. Le réseau de neurone produisant l'erreur la plus faible sur chacun des sous-modules de risques de marché, reste le modèle le plus robuste.

10.2.2 Module vie

Le tableau ci-dessous permet de comparer les SCR de vie brut obtenu via notre modèle simplifié et ceux obtenus via le logiciel ADDACTIS Modeling.

Méthode	Métrique	SCR _{Mortalité}	SCR _{Rachat}	SCR _{Frais}	SCR _{vie}
NN	MAE	2.4%	1.8%	3.8%	2%
SVR	MAE	80%	3.5%	9.3%	8.6%
LASSO	MAE	20.1%	1.5%	0.4%	19%

TABLE 10.3 – Erreur d'approximation SCR brut vie

En ce qui concerne le module de souscription-vie, nous pouvons remarquer que les erreurs d'apprentissages sont bien plus élevées que dans le cas du module de marché. Dans ce module, nous avons réalisé des approximations sur les trois sous-modules afin de capter l'effet d'un changement d'hypothèses économiques sur la variation du besoin en capital. Nous avons fait l'hypothèse implicite que les modules de mortalité et de rachat étaient principalement influencés par l'âge de l'assuré, âge censé capter l'ancienneté de l'assuré dans le contrat. Quant au *SCR* de frais, il sera principalement porté par les taux de chargement.

Le réseau de neurone semble à nouveau bien capter les effets non-linéaire des différents sous-modules de risques suite à un changement de nos hypothèses économiques, avec une erreur d'approximation sur le *SCR* brut de souscription-vie largement inférieur à nos deux autres méthodes.

10.2.3 SCR

Le tableau ci-dessous permet de comparer les valeurs des différents modules servant au calcul du *SCR* global de la compagnie d'assurance via notre modèle simplifié et ceux obtenus via le logiciel ADDACTIS Modeling.

Méthode	Métrique	SCR _{Défaut}	SCR _{Opérationnel}	BSCR	Adj TP	SCR
NN	MAE	0%	0.1%	0.3%	1.5%	1.1%
SVR	MAE	0%	1%	2.1%	23.3%	11.8%
LASSO	MAE	0%	0.2%	0.4%	2.5%	2.5%

TABLE 10.4 – Erreur d'approximation SCR global

Le tableau précédent présente les résultats d'estimation sur les métriques permettant d'arriver au *SCR* global de la compagnie d'assurance. Comme nous pouvons à nouveau le constater, le réseau de neurone donne les résultats les plus satisfaisants, avec une erreur d'approximation d'environ 1.1%.

Il est à noter que le *SCR* obtenu par réseau de neurone varie entre +5% et -5% pour l'ensemble des portefeuilles estimés.

10.2.4 RC

Le tableau ci-dessous permet de comparer les valeurs des différents modules servant au calcul du *RC* de la compagnie d'assurance via notre modèle simplifié et ceux obtenus via le logiciel ADDACTIS Modeling.

Méthode	Métrique	SCR	FP	RC
NN	MAE	1.1%	1.1%	2.1%
SVR	MAE	11.8%	4.5%	16.3%
LASSO	MAE	2.5%	1.1%	3.8%

TABLE 10.5 – Erreur d'approximation RC

Finalement, ce dernier tableau résume les résultats d'estimations permettant d'obtenir le *RC* en $t = 0$. Comme nous pouvons le constater, les *FP* sont assez bien estimés avec nos trois modèles.

Au vu de l'ensemble des résultats d'estimations, nous garderons le modèle de réseau de neurone afin de prédire le besoin en capital sur l'horizon du business plan de la compagnie d'assurance.

Cinquième partie

Application de la méthode de proxy dans un cadre ORSA

Présentation du modèle ORSA

Dans cette première partie, nous allons présenter la méthode permettant de projeter le *SCR* dans un cadre ORSA, méthode ayant été implémentée sous le logiciel ADDACTIS Modeling. De plus, les résultats de cet ORSA serviront de comparatif à nos méthodes de *machine learning* lorsque nous voudrions tester la robustesse et la fiabilité de nos méthodes d'apprentissage.

Le but de ce mémoire est de pouvoir fournir un *RC* sur la durée du plan stratégique de l'entreprise. Pour pouvoir fournir ce *RC*, la compagnie d'assurance doit pouvoir calculer un *SCR* prospectif, ce qui correspond à un montant de capital économique calculé à chaque pas de temps à l'aide de notre modèle dit ORSA.

Pour comprendre le fonctionnement du modèle ORSA, il convient de faire certains rappels. Pour pouvoir projeter notre capital économique sur n années, nous avons besoin de fournir un ensemble d'indicateurs, comme des jeux de scénario économiques, issus de notre Générateur de Scénario Economique (GSE ou ESG en anglais). Comme présenté précédemment, le GSE permet de fournir une projection d'un ensemble de grandeurs économiques comme les taux d'intérêts, les indices actions, immobiliers etc.

L'ESG va nous permettre d'obtenir une multitude de simulation pour chacune de ces grandeurs économiques, correspondant à différents états de la nature. Le but étant de pouvoir fournir une approximation du *RC* quel que soit l'état de la nature à horizon donné, nous avons besoin de faire apprendre à nos modèles différents scenarii.

Ainsi, l'ESG va nous fournir n simulations d'indicateurs financiers différentes, sur un horizon de temps et pour un pas de projection donné. Dans le cadre de notre étude, l'ESG va nous donner 1 000 simulations de courbes de taux différentes, sur un horizon de temps de 105 ans associé à un horizon de projection de 55 ans.

Le modèle ORSA, afin d'obtenir un *SCR* prospectif va projeter les bilans économiques selon deux univers :

- Un univers **Monde-réel**
- Un univers **Risque-neutre**

Dans l'univers monde-réel, la projection de la trajectoire d'un actif doit refléter son comportement statistique et historique. Des données historiques seront donc utilisées pour calibrer nos données afin de savoir comment « vieillit » notre portefeuille d'une année sur l'autre.

Dans l'univers risque-neutre, les actifs modélisés ont pour rendement le taux sans risque afin d'obtenir une modélisation « Market-consistent », c'est à dire cohérente avec l'environnement actuel des prix observés sur le marché.

La figure ci-dessous présente le fonctionnement d'un modèle ORSA.

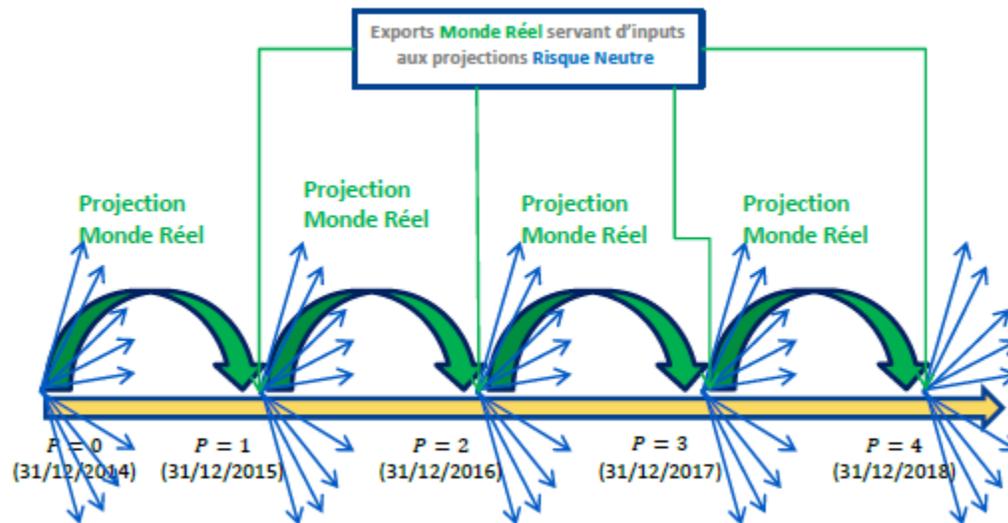


FIGURE 11.1 – Fonctionnement du modèle ALM

Dans le cadre du modèle ORSA qui a été mis en place, une unique simulation monde réel a été considérée dans le but de faire évoluer la situation économique de notre compagnie d'assurance. De même, un unique scénario risque-neutre a été considéré, le scénario théorique où nous supposons que les taux sont martingales afin de valoriser les engagements de l'assureur.

Mise en place du proxy en $t > 0$

Dans les parties précédentes, nous avons distingué deux types de proxy :

- un proxy à l'instant $t = 0$
- un proxy en $t > 0$

Ces deux proxies diffèrent d'un coefficient : le « coefficient ALM », que nous présenterons par la suite.

12.1 Présentation d'un modèle ALM

Dans les parties précédentes, nous avons mis en avant les interactions pouvant exister entre l'actif et le passif de notre compagnie d'assurance. En effet, certaines variables d'actifs comme l'allocation cible ou encore le montant à investir dans chaque classe d'actif est très fortement lié aux engagements pris par l'assureur. Un exemple concret serait de considérer une hausse des taux d'intérêts dans un contexte de *TMG* assez faible. Comme nous avons pu l'évoquer, cet effet peut engendrer des rachats dynamiques, que les assureurs vont pouvoir absorber en réalisant des plus-values latentes. De même, en cas de taux servi trop faible, ils vont pouvoir piocher dans leurs réserves afin de servir un taux plus important aux assurés.

Enfin, le modèle ALM permet de définir les règles comptables impactant le calcul de la participation aux bénéfices. Cette participation aux bénéfices dépend du rendement des investissements financiers de l'assureur et donc, de sa politique d'investissement ou de désinvestissement des actifs afin de satisfaire son allocation cible.

L'exemple précédent illustre bien la dépendance existant entre l'actif et le passif, dépendance devant être prise en compte lors de la valorisation des engagements de la compagnie d'assurance à chaque pas de projection, notamment lors du calcul d'indicateurs comme le *BE*.

La figure ci-dessous illustre le fonctionnement d'un modèle ALM :

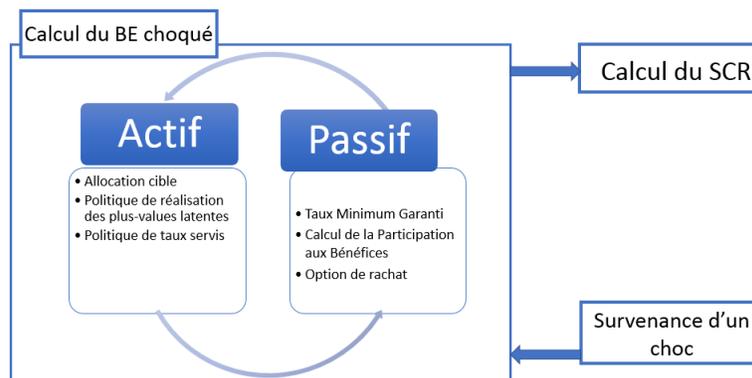


FIGURE 12.1 – Fonctionnement du modèle ALM

12.2 Calcul du coefficient ALM

Pour calculer le SCR de chacun des sous-modules de risques, nous avons besoin de savoir comment « vieillit » notre portefeuille d'une année sur l'autre afin d'être en mesure de recalculer les engagements de l'entreprise suite à une variation d'environnement économique.

Prenons l'exemple de la partie actif de notre portefeuille, nous savons qu'à l'instant $t = 0$, le besoin en capital au titre du risque immobilier se calculera comme :

$$SCR_{t=0}^{immobilier} = VM_{t=0}^{immobilier} * 0.25$$

Cependant, dans 1 an, nous percevrons un flux de loyer en moins avec des taux de rendements différents. Nous avons donc besoin d'évaluer la nouvelle VM de l'immobilier, donnant ainsi le nouveau besoin en capital comme :

$$\begin{aligned} SCR_{t=1}^{immobilier} &= VM_{t=1}^{immobilier} \times 0.25 \\ &= VM_{t=0}^{immobilier} \times \delta_t(\epsilon_1, \dots, \epsilon_n) \times 0.25 \\ &= SCR_{t=0}^{immobilier} * \delta_t(\epsilon_1, \dots, \epsilon_n) \end{aligned}$$

où $\delta_t(\epsilon_1, \dots, \epsilon_n)$ est un paramètre prenant en compte la variation de valeur de marché entre les deux années.

De même, si nous prenons la poche obligataire (OAT et OTF), entre $t = 0$ et $t = 1$, la VM des obligations va varier, d'une part du fait du changement de maturité résiduelle, engendrant le versement d'un coupon en moins, d'un changement de taux modifiant sa valeur actualisée et de l'arrivée à échéance de certaines obligations composant le portefeuille. Nous avons donc besoin d'être capable de réévaluer la valeur de l'obligation, pour un taux, une maturité résiduelle et un nombre de coupon donné. Ce calcul de valorisation étant compliqué à mettre en place, nous avons opté pour la mise en place d'une fonction paramétrique permettant de capter l'effet d'un changement d'hypothèses économiques sur la valeur de nos obligations.

La fonction paramétrique sera de la forme :

$$\frac{VM_{t+1}}{VM_t} = \delta_t(\epsilon_1, \dots, \epsilon_n)$$

où

VM_t représente la valeur de marché du module considéré

$\delta_t(\cdot)$ la fonction estimée par apprentissage statistique

Pour créer cette variable, nous avons projeté nos 100 portefeuilles sur 5 ans afin d'en récupérer les différents indicateurs comme les valeurs de marchés et de faire un apprentissage sur le rapport des VM entre deux années afin de capter la stratégie d'allocations d'actifs suite à un changement de variables d'actifs et de passifs. Ces différents modèles ont été calculés en ayant recours à un modèle ALM.

La figure ci-dessous permet d'illustrer la formule précédente permettant le calcul du coefficient $\delta_t(\cdot)$. Comme nous pouvons le constater, nous avons besoin des VM à l'instant $t + 1$ afin de calculer le coefficient d'ALM entre les dates t et $t + 1$. Notre business plan s'étendant sur 5 ans, nous sommes partis de l'année $t + 5$ pour remonter jusqu'à l'année 0.

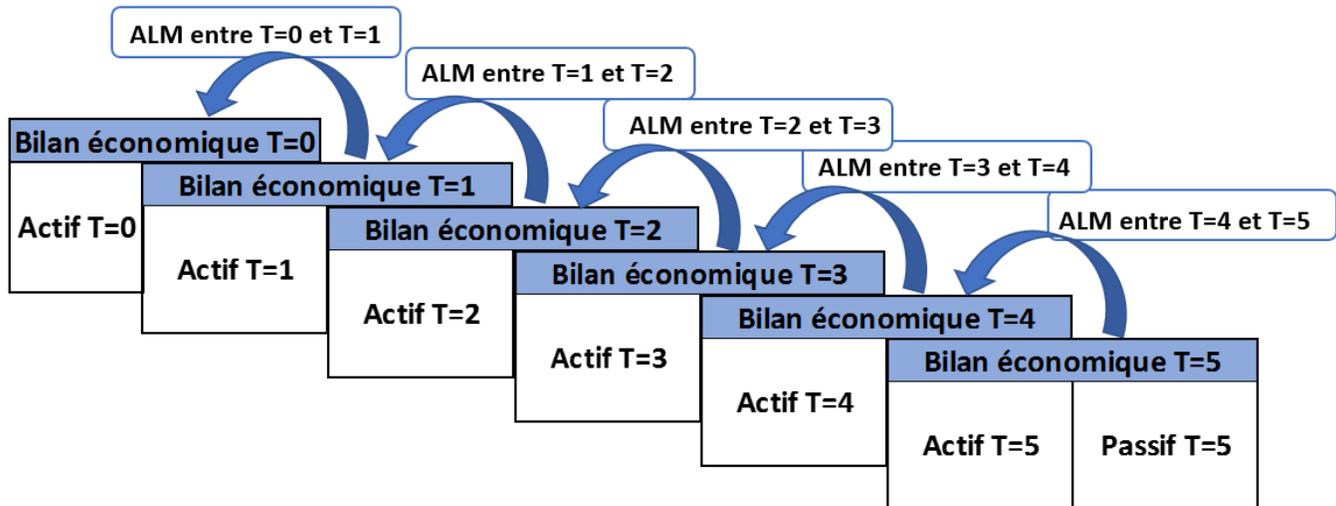


FIGURE 12.2 – Extraction des coefficients ALM

12.3 Projection des inputs

Dans la partie précédente, nous avons explicité le fait que la construction de nos « coefficients ALM » ont nécessité le lancement préalable de nos 100 portefeuilles sur 5 ans. Une fois ces 5 années de bilans économiques récupérées, nous avons pu faire apprendre à nos modèles la façon dont les actifs se déforment compte tenu du contexte macroéconomique et des interactions entre l'actif et le passif.

Une fois ces variables connues, nous sommes en mesure de faire vieillir notre portefeuille quel que soit le contexte économique. Dans la partie suivante, nous présenterons les erreurs d'estimations commises en ORSA. Avant cette étape, nous devons fournir à nos modèles de *machine learning* des inputs.

Afin d'éviter d'avoir l'illusion de bien prédire notre *RC* en fournissant directement les bilans économiques nous ayant permis de calibrer nos modèles sur les 5 années de projection, nous allons partir du bilan à l'année $t = 0$ et nous servir des « coefficients ALM » pour déformer notre bilan initial et ainsi obtenir un bilan prospectif en $t = 1$ vu à la date 0. A la date $t = 1$, nous allons procéder de la même façon afin d'obtenir le bilan choqué pour l'année $t = 2$ et ainsi de suite.

La figure ci-dessous résume la méthode permettant d'obtenir les bilans prospectifs. Comme nous pouvons le constater, seul le bilan en $t = 0$ correspond au vrai bilan de la compagnie d'assurance. En effet, à un instant donné, une compagnie d'assurance connaît les engagements qu'elle a, ainsi que les actifs en portefeuille au titre de l'année en cours. L'année suivante, elle ne peut que supposer la façon dont son bilan va évoluer, c'est ce que nous appelons le bilan prospectif. Ces bilans correspondent à une situation économique future basée sur les hypothèses que la compagnie a définies.

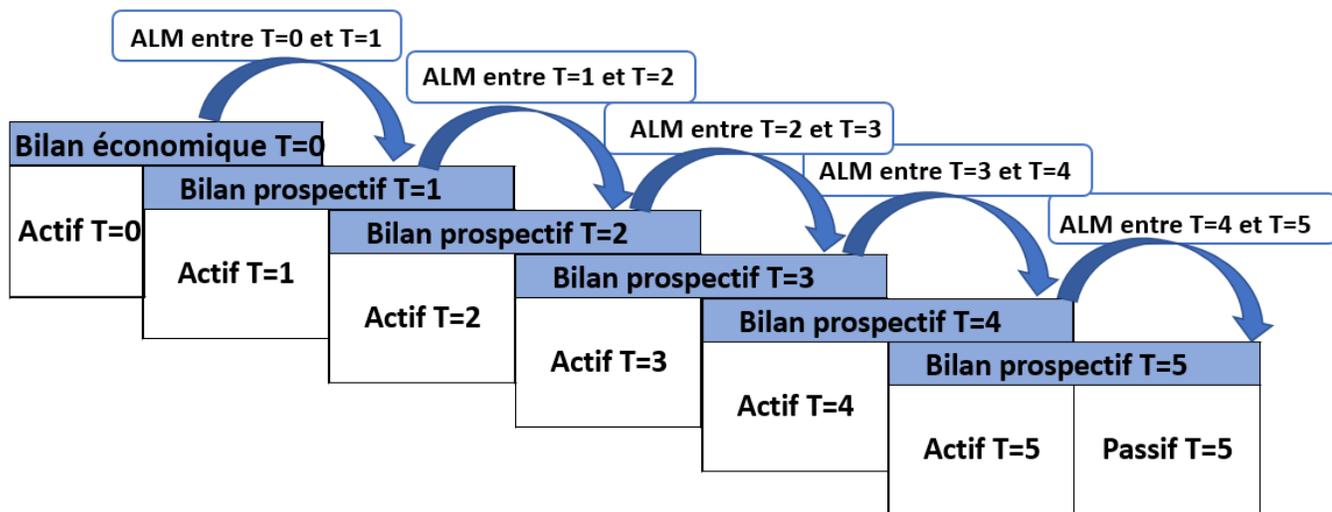


FIGURE 12.3 – Construction des bilans prospectifs

12.4 Résultats d'estimations

Comme dans la partie précédente, nous allons présenter les erreurs d'estimations commises sur les différents modules de la formule standard à chaque pas de projection. Ces erreurs concernent la base de test. Le but de cette étape est de pouvoir vérifier la capacité de notre fonction de proxy à prédire le capital économique sur l'horizon du business plan et sur les hypothèses prospectives que nous avons définies lors de l'étape de construction de la base de donnée.

Au vu des résultats d'estimation au temps $t = 0$, nous présenterons les résultats obtenus via les réseaux de neurones en ayant recours au *MAE*.

12.4.1 Module marché

Le tableau ci-dessous permet de comparer les SCR de marché brut obtenu via notre modèle simplifié et ceux obtenus via le logiciel ADDACTIS Modeling.

Pas	BE	Ecart BE	SCR _{Actions}	SCR _{Immobilier}	SCR _{Spread}	SCR _{Taux}	SCR _{marche}
t=0	0.1%	0.2%	0%	0%	0.2%	0.3%	0.1%
t=1	0.2%	0.6%	0.1%	0.1%	0.3%	0.6%	0.2%
t=2	0.2%	0.7%	0.1%	0.1%	0.3%	0.6%	0.3%
t=3	0.2%	0.8%	0.1%	0.1%	0.3%	0.6%	0.3%
t=4	0.2%	0.6%	0.1%	0.1%	0.3%	0.6%	0.3%
t=5	0.3%	0.6%	0.1%	0.1%	0.3%	0.6%	0.4%

TABLE 12.1 – Erreur d'approximation SCR brut marché

Comme nous pouvons le constater sur le tableau d'estimation du module de marché, l'erreur d'approximation en valeur absolue est inférieure à 5% et ce, sur la durée du business plan. Ce tableau met en avant le fait que notre modèle de *machine learning* est capable de comprendre les mécanismes de déformations de nos bilans en fonction du contexte économique changeant et, d'en inférer une valeur d'engagement cohérente.

Ce résultat d'estimation plus que correcte est à nuancer par le fait que le module de marché est principalement constitué de formules fermées à l'exception du module de taux qui nécessite de recalculer une variation de *BEG*.

Le *BE* hors choc est estimé de manière assez précise puisque l'erreur varie entre 0.15% et 0.3%. Comme nous avons pu le mentionner précédemment, ce résultat laisse entrevoir une estimation des *FP* ne dépassant pas les 3% d'erreurs ce qui n'aura pas pour conséquence de détériorer l'estimation du *RC*

12.4.2 Module de souscription-vie

Le tableau ci-dessous permet de comparer les SCR de souscription-vie bruts obtenus via notre modèle simplifié et ceux obtenus via le logiciel ADDACTIS Modeling.

Pas	SCR _{Mortalité}	SCR _{Rachat}	SCR _{Frais}	SCR _{vie}
t=0	2.4%	1.8%	3.8%	2%
t=1	11%	7.4%	4%	6.2%
t=2	11%	8.7%	4.1%	7%
t=3	11%	10%	4.1%	8.6%
t=4	11%	11%	4.2%	9.4%
t=5	11%	13%	4.3%	11.1%

TABLE 12.2 – Erreur d'approximation SCR brut marché

En ce qui concerne l'estimation du module de souscription-vie, nous remarquons que les erreurs sont bien supérieures à celle commises lors de l'estimation du module de marché. Cela s'explique par le fait que dans le cas du module de marché, nous avons des formules fermées dont seul la VM variait. D'une année sur l'autre, la valeur de marché des actifs financiers reste globalement stable ce qui permet d'estimer ces modules sans trop d'erreurs.

Dans ce cas précis, nous avons directement estimé le sous-module de risque comme suit :

$$SCR_{t+1}^{sous-module} = SCR_t^{sous-module} \times \delta_t(.)$$

Cette formule ne permet donc pas de capter avec précision les effets sous-jacents à la mortalité par exemple, rendant les estimations très volatiles, passant d'une erreur de 2% à 11% entre l'année $t = 0$ et $t = 1$.

12.4.3 RC

Le tableau ci-dessous permet de comparer les valeurs des différents modules servant au calcul du RC de la compagnie d'assurance via notre modèle simplifié et ceux obtenus via le logiciel AD-DACTIS Modeling.

Pas	$SCR_{Défaut}$	$SCR_{Opérationnel}$	BSCR	Adj TP	SCR	FP	RC
t=0	0%	0.1%	0.3%	1.5%	1.1%	1.1%	2.1%
t=1	0%	0.1%	1%	12.5%	5.6%	1.4%	6.3%
t=2	0%	0.1%	1.1%	13.4%	6.1%	1.4%	6.5%
t=3	0%	0.2%	1.1%	13.5%	6.1%	1.5%	6.5%
t=4	0%	0.2%	1.2%	13.5%	6.2%	1.6%	6.6%
t=5	0%	0.2%	1.6%	13.6%	6.8%	1.6%	6.8%

TABLE 12.3 – Erreur d'approximation RC

Nous finissons cette partie en comparant les erreurs d'estimations faite sur le RC de la compagnie d'assurance. Comme nous pouvons le remarquer, le SCR global de notre compagnie d'assurance est estimé avec une erreur inférieure à 7% sur la durée du business plan. Ce résultat très encourageant atteste de la robustesse de notre méthode sur le cadre d'estimation que nous avons considéré.

Malgré une erreur plus importante sur la partie souscription-vie, l'erreur globale est « maîtrisé » du fait d'une forte précision lors de l'estimation du capital économique au titre du risque de marché. Il est important de rappeler que le risque de marché est le plus gros risque pour les compagnies d'assurance-vie, risque que nous estimons avec grande précision dans le cas d'un SCR brut.

Par ailleurs, l'estimation du SCR global aurait pu être bien plus précise si nous n'avions pas commis une erreur aussi grande sur le module d'ajustement par les provisions techniques. Il apparaît, à ce stade, très compliqué de calibrer une fonction sur les mécanismes d'ajustement par les provisions techniques en considérant seulement le contexte économique changeant. En effet, cette variable représente les ajustements dû à l'effet d'absorption des risques induits par les mécanismes de participations aux bénéfices et d'impôts différés. Capter l'effet d'une telle variable reviendrait à mettre en place un proxy sur la politique de *management action* de la compagnie en question ou du moins, d'inclure le montant de participations aux bénéfices pour chacune des situations économiques.

Quant aux fonds propres, nous arrivons à les estimer avec une erreur inférieure à 2%, ce qui reste tout à fait acceptable. En somme, le ratio de couverture de notre compagnie d'assurance est estimé avec une erreur moyenne en valeur absolue de 2% la première année et de 6% les années suivantes. Comme nous venons de le dire, la variable d'ajustement par les provisions techniques contribue grandement à cette erreur.

Limites des méthodes

Les méthodes qui ont été mises en place dans la partie précédente nous permettent d'obtenir des résultats tout à fait acceptables sur l'horizon du business plan de la compagnie d'assurance. En effet, grâce à ces méthodes, nous pouvons prédire le *SCR*, les *FP* et le *RC* avec des erreurs d'estimations qui sont respectivement bornées à 7%, 2% et 7%.

Cependant, ces approximations restent fiables dans un cadre bien défini à savoir, le cadre d'estimation qui a été considéré. Dans la construction de la base de données, nous avons défini des bornes de variations pour chacune de nos variables. Le choix de ces bornes dépend en grande partie de la compagnie d'assurance considérée et de son portefeuille. Tant que nous souhaitons lancer des stress-test au sein de cette espace de variation, nos modèles resteront pertinent et fiables, nous pourrions interpoler une valeur d'engagement.

Cependant, en dehors de ce cadre d'étude, le modèle ne sera pas capable d'interpoler une valeur de capital économique puisqu'il n'aura jamais rencontré une telle situation. C'est en ce sens que notre modèle de *machine learning* présente une limite, il ne pourra pas extrapoler.

D'autre part, ayant fait l'hypothèse du caractère négligeable de certains sous-modules de risques, notre modèle d'apprentissage n'est pas capable de prendre en compte l'ensemble des risques d'une compagnie d'assurance-vie mais donne néanmoins de très bon résultats sur les plus gros risques.

Enfin, cette étude a été réalisée en ayant recours à un ORSA déterministe dans lequel une seule simulation risque-neutre a été considéré. Cette hypothèse présente des avantages et des limites. Cette hypothèse forte nous permet d'économiser en temps de calcul mais présente l'inconvénient de ne pas pouvoir évaluer le coût des options et garanties financières, que seul un modèle stochastique peut capter.

Pour finir, nous avons réalisé notre étude sur un seul model point, et nous nous sommes restreints à certaines classes d'actifs, ne prenant pas en compte les produits structurés, plus compliqués à modéliser.

Sixième partie

Pour aller plus loin

Hypothèses prospectives

Dans cette dernière partie, nous allons appliquer un ORSA à notre modèle en prenant des hypothèses ne faisant pas partie des inputs ayant initialement servis lors de la phase d'apprentissage ou du test de notre fonction. Cette ultime étape vise à évaluer les limites de notre modèle en testant sa robustesse.

Un scénario ORSA a été mis en place dans lequel nous supposons une baisse des taux. Pour ce scénario, nous avons décidé de ne pas faire varier le niveau du taux de chargement sur encours ainsi que du TMG sur la durée du business plan. Cette hypothèse restrictive pourra être relâché dans un second temps mais permet de rester cohérente avec le contexte actuel de taux bas et dans l'idée de ne pas provoquer de rachats de la part des assurés.

14.1 Scénario de baisse des taux

Dans ce scénario, nous considérons un contexte de taux bas avec une baisse plus marquée des taux. En effet, en $t = 0$, nous allons considérer la courbe des taux EIOPA au 31/12/2015 puis, celle fournie par l'EIOPA au 31/12/2016.

La figure ci-dessous permet de visualiser le niveau des courbes considéré :

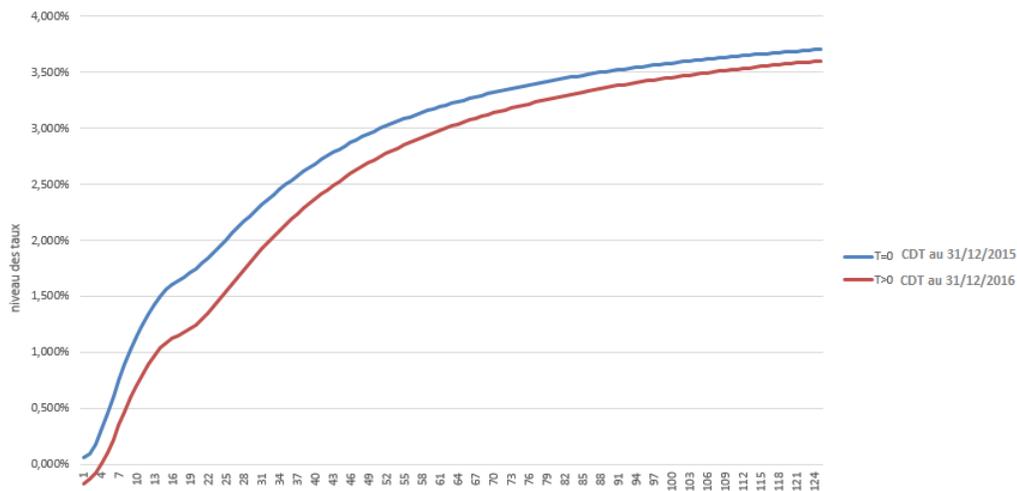


FIGURE 14.1 – Evolution des taux à la baisse

En ce qui concerne l'univers monde-réel, l'évolution des marchés actions et immobiliers est basée sur un historique prudent :

- évolution de 3% pour les actions
- évolution de 2% pour l'immobilier
- l'inflation a été fixée à 0.5%

Le tableau ci-dessous permet de résumer l'ensemble des hypothèses du scénario central

Horizon	Taux de chargt	TMG	Age	Obligataire	Action	Immo	Monétaire	Taux 1 an	Taux 8 ans	Taux 20 ans	% R332-20	TMG-Tx 8 ans	VM
t=0	0.80%	1.25%	40	65%	20%	10%	5%	0.063%	0.887%	1.747%	30%	0.363%	2 000 000 000
t=1	0.80%	1.25%	41	65%	20%	10%	5%	-0.172%	0.471%	1.247%	30%	0.779%	1 963 702 109
t=2	0.80%	1.25%	42	60%	23%	11%	6%	-0.172%	0.471%	1.247%	34%	0.779%	1 675 330 009
t=3	0.80%	1.25%	43	60%	23%	11%	6%	-0.172%	0.471%	1.247%	34%	0.779%	1 605 325 434
t=4	0.80%	1.25%	44	54%	26%	13%	7%	-0.172%	0.471%	1.247%	39%	0.779%	1 326 292 472
t=5	0.80%	1.25%	45	54%	26%	13%	7%	-0.172%	0.471%	1.247%	40%	0.779%	1 262 294 910

TABLE 14.1 – Hypothèses scénario baisse des taux

14.2 Résultats d'estimations

Le tableau ci-dessous permet de obtenir les erreurs d'approximations de nos différents indicateurs sur la durée du business plan.

Horizon	BE	SCR Marché	SCR Vie	BSCR	SCR Opé	SCR Adj TP	SCR	FP	RC
t=0	-0.8%	0.5%	3%	0.6%	0.1%	1.6%	1.3%	1.3%	2.4%
t=1	-4.2%	-1.9%	231.8%	35.1%	-2.6%	-17.5%	50.4%	18.5%	-21.2%
t=2	5.1%	18.1%	91.8%	29.1%	-2.6%	-11.9%	43.7%	26.7%	-11.8%
t=3	4.3%	23%	145%	41.1%	-2.6%	-8.4%	57.7%	44.1%	-8.6%
t=4	-6.1%	25.6%	162.6%	44.7%	-2.6%	3.9%	56.4%	64.9%	5.4%
t=5	-2.5%	47.9%	295.2%	90.9%	-2.6%	69.9%	93.5%	80.5%	-6.7%

TABLE 14.2 – Erreur d'approximation scénario de baisse des taux

Comme nous pouvons le constater, dans le cadre d'une baisse des taux, avec des taux négatifs comme cela a été le cas au 31/12/2016, notre modèle est incapable de prédire nos métriques avec précision. En effet, seule l'année $t = 0$ correspond à des inputs entrant dans le cadre de notre étude.

Ainsi, à la date de lancement, notre modèle est capable d'interpoler une valeur d'engagement tandis qu'aux dates $t > 0$, notre modèle doit extrapoler. L'étude préliminaire a servi à capter les effets d'interactions entre nos différentes variables or, dans ce cas précis, l'impact d'un taux négatif sur nos différents indicateurs n'a pas été appris par notre modèle, c'est la raison pour laquelle il ne sait pas comment se comporter dans une telle situation ; ce qui se ressent sur les estimations.

Finalement, ce dernier test permet de conclure sur l'incapacité de notre méthode à extrapoler d'où, la nécessité de bien définir le cadre de projection en amont.

Conclusion générale

Le but de ce mémoire était de proposer une méthodologie basée sur du *machine learning* dans le but de projeter le ratio de couverture dans un cadre ORSA. L'utilisation de ces méthodes d'apprentissage statistique répondait à un objectif double :

- la réduction du temps de calcul
- la simplification mathématiques

Ces objectifs répondent aux besoins de la part des compagnies d'assurance d'être en mesure de pouvoir évaluer en permanence et à moindre coût l'évolution de leur besoin en capital à la suite d'un changement d'hypothèses économiques (hausse ou une baisse des taux d'intérêts).

La démarche que nous avons mise en place tout au long de ce mémoire avait pour but l'atteinte des deux objectifs précédents. Cette méthode, à savoir le calcul d'une nouvelle fonction de proxy s'est faite en plusieurs étapes :

- définition du champ de l'étude
- variabilité de nos facteurs de risques
- choix des variables les plus pertinentes
- calibrage et test de la fonction

Le calcul de ce proxy par apprentissage a nécessité un travail préalable d'une semaine afin de pouvoir, d'une part définir le champ d'étude notre compagnie d'assurance et d'autre part, d'entraîner et de tester les performances nos modèles. Ainsi, à l'issu de ce temps de calcul d'une semaine, nous disposons d'une fonction de proxy robuste, permettant d'effectuer des stress-test en cours d'année de manière instantanée et avec une très grande précision. Cette fonction permet aussi de prédire sur la durée du business plan, le *SCR*, les *FP* ainsi que le *RC* avec des erreurs étant respectivement inférieures à 7%, 2% et 7%.

Ces résultats sont très satisfaisants dans la mesure où la base d'apprentissage contient seulement 100 scénarii de variations économiques. Il est probable qu'en augmentant le nombre de points et en affinant les bornes de variations de chacune de nos variables, les résultats soient bien plus précis et moins volatils mais dans ce cas, l'apprentissage serait plus coûteux en temps de calcul.

Cependant, lors de l'estimation sur un scénario type adverse (baisse des taux), notre modèle s'est avéré incapable de prédire le *RC* avec précision. Ce test a fait apparaître la limite des méthodes d'apprentissages statistiques, à savoir, que les modèles peuvent seulement interpoler sur des situations apprises. Pour être en mesure de répondre à ces scénarii adverses, il aurait fallu recalibrer nos modèles en prenant en compte des situations extrêmes. Dès lors, l'étape la plus importante reste la construction de la base de données et des hypothèses que nous souhaitons mettre dedans. Enfin, le modèle a été testé sur un unique model point avec des hypothèses simplifiées, il devra par la suite être testé sur un portefeuille réel.

A l'issue de cette étude, nous avons pu développer une méthodologie se basant sur le *machine learning* et permettant d'obtenir un proxy simple à utiliser pour prédire le capital économique et les fonds propres en $t = 0$ ou dans le cadre du business plan de la compagnie d'assurance. Ce proxy reste néanmoins propre à chaque compagnie d'assurance puisqu'il dépend de son profil de risque et de ses caractéristiques.

Bibliographie

- [1] ACPR, Notice « Solvabilité 2 » Évaluation interne des risques et de la solvabilité (ORSA).
- [2] Marc JUILLARD , Les aspect quantitatif de l'ORSA.
- [3] Actuaris, L'ORSA comme outil de pilotage, Université d'été de l'IA.
- [4] Institut des Actuaires, Groupe de travail ORSA (mars 2014)
- [5] LOISEL.S et DEVINEAU.L, Construction d'un algorithme d'accélération de la méthode des « simulations dans les simulations » pour le calcul du capital économique Solvabilité II, 2009
- [6] Matthieu Chauvigny Laurent Devineau, Replicating portfolios : techniques de calibrage pour le calcul du capital économique solvabilité 2, HAL - Archives ouvertes, hal-00508517, 2010.
- [7] Haguët.E, Mise en place d'indicateurs de suivi du risque dans un cadre d'ORSA Epargne
- [8] Châtaignier.B, Ajustements financiers dans un cadre ORSA
- [9] Anne-Claire Martial Jean-Baptiste Garnier, Optimisation du calcul du capital économique d'une compagnie d'assurance vie par la méthode des simulations dans les simulations, Ressources Actuarielles, 2013.
- [10] Marine Niedzwiedz, Utilisation des supports vecteurs machines pour l'accélération du calcul du capital économique, Ressources Actuarielles, 2014
- [11] Luo Yufei, Amélioration de la modélisation de sinistres graves à l'aide d'une approche d'apprentissage, Ressources Actuarielles 2016
- [12] Maroi Ben Aribi, Optimisation de la réassurance sous une vision ORSA en assurance de personnes, Ressources Actuarielles, 2016
- [13] Roué, A. 2015, « Les réseaux bayésiens comme technique d'optimisation numérique dans le cadre de l'ORSA », Mémoire d'actuariat, EURIA.
- [14] McKay, M.D., Beckman, R.J., Conover, W.J., 1979. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21, 239–245.
- [15] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems*, 2 :303–314, 1989.
- [16] Trevor Hastie, Robert Tibshirani, Jerome Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction*

- [17] V. Viallon, Régression pénalisée : le Lasso
- [18] Anisse Ismaili et Pierre Gaillard, Le Lasso, ou comment choisir parmi un grand nombre de variables à l'aide de peu d'observations, Juin 2009
- [19] Vapnik, V. N. 1998, Statistical learning theory, vol. 1, Wiley New York. 41
- [20] Rosenblatt, F. 1958, «The perceptron : a probabilistic model for information storage and organization in the brain.», Psychological review, vol. 65, no 6, p. 386. 3
- [21] Habermehl et Kesner. Programmation logique et IA - Réseaux de neurones.
- [22] P. Alquier, An Algorithm for Iterative Selection of Blocks of Features, Algorithmic Learning Theory (ALT), 2010, M. Hutter, F. Stephan, V. Vovk and T. Zeugmann Eds., Springer - Lecture Notes in Artificial Intelligence n. 6331, pp. 35-49. You can download a corrected version, also test the ISBF package for R.
- [23] Tibshirani, Saunders, Rosset, Zhu, et Knight, Sparsity and smoothness via the fused lasso
- [24] C. Cortes, V. Vapnik, Support-Vector Networks

Note de synthèse

Contexte

L'entrée en vigueur en 2016 de la réglementation Solvabilité 2 modifie les exigences quantitatives et qualitatives pour les compagnies d'assurances. Cette nouvelle réglementation comporte trois piliers :

- Le premier pilier définit les exigences quantitatives, notamment en matière de fonds propres et de calculs des provisions techniques
- Le deuxième pilier définit les exigences en matière de gouvernance des risques
- Le troisième pilier définit les exigences en terme d'informations prudentielles et de publication

Dans le cadre du pilier 2 de la directive Solvabilité 2, l'ORSA ou évaluation interne des risques et de la solvabilité (Own Risk and Solvency Assessment), se caractérise par la nécessité pour une compagnie d'assurance d'être en mesure de s'assurer en permanence de la couverture de son capital réglementaire. Il s'agit d'un processus d'évaluation prospectif des risques, appréhendé sur la durée du business plan de l'entreprise et peut s'étendre de 3 à 8 ans.

Ainsi, cette évaluation permanente doit conduire les organismes à simuler dans le temps l'évolution de leur capital requis (ou *SCR*, *Solvency Capital Requirement*) et de leurs fonds propres économiques (*FP*) sur la base de scénarii d'évolution de leur facteur de risque. Plusieurs possibilités d'environnements économiques futurs seront considérées pour effectuer ces calculs (tels que des changements de la situation économique, des modification subites des portefeuilles d'encours etc). L'ensemble de ces scénarii seront appelés des « possibles ».

L'exploration d'un nombre très élevé de « possibles », via des approches stochastiques, rend les calculs complexes, à la fois en termes de temps de calculs mais aussi d'un point de vue mathématiques. Cela incite les compagnies d'assurance à envisager plusieurs approches de simplification afin d'évaluer leurs engagements. La mise en œuvre de méthodes simplifiées doit cependant tenir compte du profil de risque de la compagnie tout en arbitrant entre coût de mise en œuvre, précision attendue, maîtrise dans la justification des résultats et sensibilité aux hypothèses.

L'objectif de ce mémoire consiste donc en l'estimation d'une fonction dite de proxy permettant d'évaluer le *SCR* et les fonds propres via des algorithmes de *machine learning* afin de rendre possible le calcul du ratio de couverture sur l'horizon de business plan sans avoir à relancer un grand nombre de calcul :

$$y_t = f_t(\epsilon_1, \dots, \epsilon_n)$$

où :

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \epsilon &\longmapsto p(\epsilon) \end{aligned}$$

avec :

$p(\epsilon)$ une fonction polynomiale
 ϵ_i le i -ème facteur de risque

Cette approximation repose uniquement sur des facteurs de risques élémentaires représentant les aléas auxquels les assureurs sont soumis et susceptibles d'impacter les BE , les FP ainsi que les sous-modules de risques de la formule standard. Contrairement aux différentes méthodes ayant été développées auparavant, l'apprentissage statistique va nous permettre d'extraire les informations économiques de nos bilans afin d'en inférer une valeur d'engagement.

Par la suite, cette approximation des sous-modules va nous permettre de « remonter » jusqu'au SCR de marché/vie afin de calculer les exigences de la compagnie d'assurance suite à un changement d'environnement économique et d'obtenir le SCR global de la compagnie d'assurance.

La schéma ci-dessous permet d'illustrer la méthode de proxy que nous allons mettre en place.

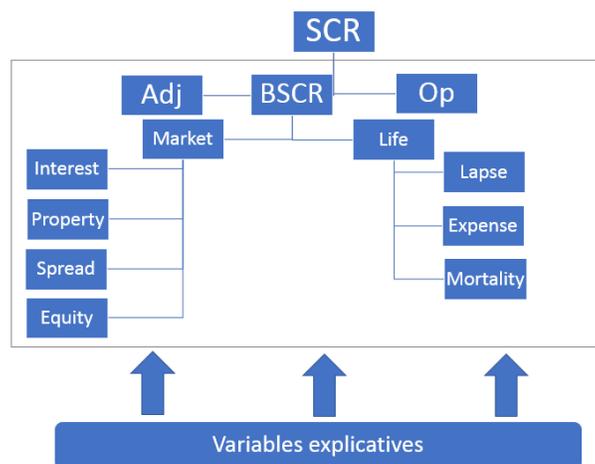


FIGURE 14.2 – Méthode proxy

Cependant, pour construire de telles fonctions, des données sont nécessaires afin de pouvoir, d'une part calibrer notre fonction et d'autre part prédire de nouveaux SCR et FP suite à un changement d'environnement économique.

Construction de la base de données

Pour être en mesure d'inférer des situations économiques à partir de nos bilans, nous avons dans un premier temps créé la base de données la plus représentative possible. En effet, les algorithmes de machine learning ne peuvent donner de bons résultats sur des situations qu'ils n'ont jamais rencontrés. Le but de cette première étape est donc de pouvoir balayer un panel de situations économiques le plus large possible. Nous avons retenus les variables suivantes :

- Le taux de chargement sur encours
- le TMG
- L'âge de l'assuré
- La proportion d'OAT
- La proportion d'OTF
- La proportion d'actions
- La proportion d'immobilier

- La proportion de monétaire
- Le taux d'intérêt 1,8 et 20 ans
- Le pourcentage d'articles R332-20
- La différence entre le TMG et le taux 8 ans
- La différence entre le TMG et le taux 8 ans maximum

Nous souhaitons donc trouver une base de dimension réduite avec nos 13 variables mais qui soit la plus répartie possible dans l'espace tout en optimisant le temps de calcul. Pour ce faire, nous avons opté pour la méthode des hypercubes latins (Latin Hypercube Sampling).

Cette méthode consiste à évaluer numériquement les intégrales multiples en s'assurant de la non-redondance de l'information au travers d'une projection sur les axes factoriels. Pour ce faire, l'intervalle de chaque entrée x_i est découpé en N segments de probabilité égale à $\frac{1}{N}$ ce qui nous donne le découpage suivant : $\{[0, \frac{1}{N}], [\frac{1}{N}, \frac{2}{N}], \dots, [\frac{N-1}{N}, 1]\}$.

Par la suite, une valeur est tirée aléatoirement dans chacun des N segments. Une fois qu'un point a été tiré aléatoirement dans un des segments de x_i , aucun nouveau point ne peut y être placé assurant ainsi que chaque point soit représenté une seule fois.

Cette méthode garantit une couverture uniforme du domaine de chaque entrée évitant de sous-échantillonner certains segments et à l'inverse d'en sur-échantillonner d'autres. Dans un soucis de visualisation, la figure ci-dessous présente la méthode des LHS pour le TMG et le taux de chargement sur encours dans un espace de dimension $d = 2$ et $N = 100$. Comme nous pouvons le constater sur la figure ci-dessous, la méthode LHS permet une couverture de l'espace des données tandis qu'un tirage aléatoire aura tendance à sous-représenter certaines parties du cube.

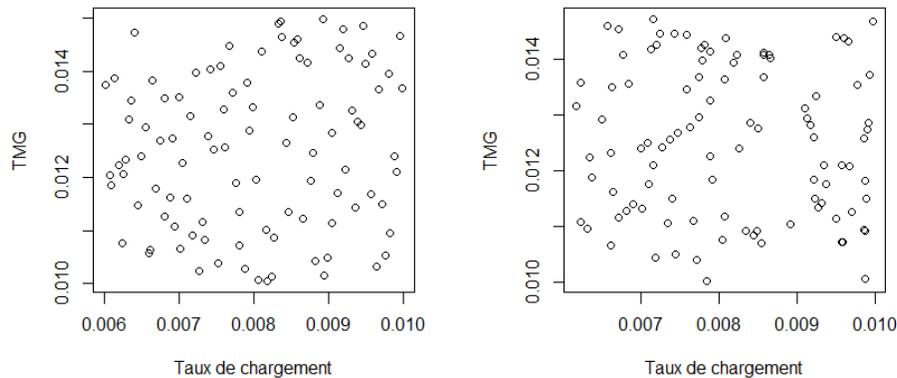


FIGURE 14.3 – Hypercube Latin contre tirage aléatoire

Mise en place du proxy

Nous distinguons la fonction de proxy implémentée en $t = 0$ et celle mise en place en $t > 0$. En effet, entre t et $t + 1$, nos indicateurs de risques vont évoluer en même temps que l'environnement économique. Pour prendre en compte cette évolution, comme une diminution de la valeur de marché, nous estimerons un dernier coefficient, dit « coefficient ALM » que nous noterons δ dont le but est de capter l'effet de la variation de la valeur de nos indicateurs économiques dans le futur.

Finalement, nous obtiendrons la relation suivante :

$$\begin{aligned} \tilde{S}C\tilde{R}_t &= f_t(\epsilon_1, \dots, \epsilon_n) \mathbb{1}\{t = 0\} + f_{t-1}(\epsilon_1, \dots, \epsilon_n) \delta_{t-1}(\epsilon_1, \dots, \epsilon_n) (1 - \mathbb{1}\{t = 0\}) \\ \tilde{F}P_t &= f'_t(\epsilon_1, \dots, \epsilon_n) \mathbb{1}\{t = 0\} + f'_{t-1}(\epsilon_1, \dots, \epsilon_n) \delta'_{t-1}(\epsilon_1, \dots, \epsilon_n) (1 - \mathbb{1}\{t = 0\}) \end{aligned}$$

La relation précédente fait apparaître l'importance de bien pouvoir calibrer la fonction de proxy en $t = 0$ puisque, à l'instant $t > 0$, nos indicateurs représenteront simplement une proportion du $\tilde{S}C\tilde{R}_0$ et des $\tilde{F}P_0$.

Finalement, nous obtiendrons le ratio de couverture :

$$R\tilde{C}_t = \frac{\tilde{S}C\tilde{R}_t}{\tilde{F}P_t}$$

où ϵ_i correspond aux facteurs de risques et $f_t(\cdot)$, $f'_t(\cdot)$, $\delta_t(\cdot)$ et $\delta'_t(\cdot)$ correspondent à nos fonctions de proxies obtenues par apprentissage statistique.

Résultats d'estimations

Notre apprentissage statistique a été réalisé sur un unique model point, sur lequel nous obtenons des résultats tout à fait convenable puisque les erreurs d'approximations pour le SCR , les FP et le RC sont respectivement inférieures à 7%, 2% et 7% en valeur absolue sur la durée du business plan. Ce premier résultat est tout à fait encourageant puisqu'il met en avant la capacité des méthodes d'apprentissages statistiques à répondre à des besoins concrets en assurance-vie.

Pas	$SCR_{Défaut}$	$SCR_{Opérationnel}$	BSCR	Adj TP	SCR	FP	RC
t=0	0%	0.11%	0.37%	1.57%	1.08%	1.1%	2.1%
t=1	0%	0.16%	0.98%	12.5%	5.6%	1.41%	6.3%
t=2	0%	0.18%	1.1%	13.4%	6.07%	1.47%	6.5%
t=3	0%	0.2%	1.17%	13.48%	6.09%	1.55%	6.56%
t=4	0%	0.21%	1.18%	13.5%	6.16%	1.58%	6.6%
t=5	0%	0.25%	1.6%	13.58%	6.77%	1.65%	6.8%

TABLE 14.3 – Erreur d'approximation RC

Conclusion

Le but de ce mémoire était de proposer une méthodologie basée sur du *machine learning* dans le but de projeter le ratio de couverture dans un cadre ORSA. L'utilisation de ces méthodes d'apprentissage statistique répondait à un objectif double :

- la réduction du temps de calcul
- la simplification mathématiques

Ces objectifs répondent aux besoins de la part des compagnies d'assurance d'être en mesure de pouvoir évaluer en permanence et à moindre coût l'évolution de leur besoin en capital à la suite d'un changement d'hypothèses économiques (hausse ou une baisse des taux d'intérêts).

Le calcul du proxy par apprentissage a nécessité un travail préalable d'une semaine afin de pouvoir, d'une part définir le champs d'étude notre compagnie d'assurance et d'autre part, d'entraîner et de tester les performances nos modèles. Ainsi, à l'issue de ce temps de calcul d'une semaine, nous disposons d'une fonction de proxy robuste, permettant d'effectuer des stress-test en cours d'année de manière instantanée et avec une très grande précision. Cette fonction permet aussi de prédire sur la durée du business plan, le *SCR*, les *FP* ainsi que le *RC* avec des erreurs étant respectivement inférieurs à 7%, 2% et 7%.

Cependant, lors des estimations sur un scénario type adverse (baisse des taux), notre modèle s'est avéré incapable de prédire le *RC* avec précision. Ce test a fait apparaître la limite des méthodes d'apprentissages statistiques, à savoir, que les modèles peuvent seulement interpoler sur des situations apprises. Pour être en mesure de répondre à ces scénarii adverses, il aurait fallut recalibrer nos modèles en prenant en compte des situations extrêmes. Dès lors, l'étape la plus importante reste la construction de la base de données et les hypothèses que nous souhaitons mettre dedans. Enfin, le modèle a été testé sur un unique model point avec des hypothèses simplifiés, il devra par la suite être testé sur un portefeuille réel.

A l'issue de cette étude, nous avons pu développer une méthodologie se basant sur le *machine learning* et permettant d'obtenir un proxy simple à utiliser pour prédire le capital économique en $t = 0$ ou dans le cadre du business plan de la compagnie d'assurance. Ce proxy reste propre à chaque compagnie d'assurance puisqu'il dépend de son profil de risque et de ses caractéristiques.

Summary

Context

The entry into force of Solvency 2 in 2016 change the quantitative and qualitative expectation for insurers companies. This new regulation framework is composed of three pillars :

- The first pillar define the quantitative expectations, especially in terms of eligible own fund and technical provision.
- The second pillar define the qualitative expectation in terms of risk management.
- The third pillar define the expectation in terms of disclosure and transparency requirement.

Within the framework of the pillar 2 of Solvency 2, ORSA, the intern evaluation of risks and solvency (Own Risk and Solvency Assessment) characterizes itself with the necessity of being able to insure at any time the covering of the regulatory capital of a given insurance firm. It is an evaluation of risks based on a prospective vision within three to eight years which can correspond to the business plan of the company.

Thus, this permanent evaluation is supposed to lead organizations into making simulations through time portraying the evolution of their SCR (Solvency Capital Requirement) and of their eligible own fund (EOF) on the basis of scenarios of their risk factor. Different futur scenario's possibilities will be considered to make this regulatory computation (macroeconomic changes, regulatory and so on). These changes are qualified as possible ».

Exploring the great number of « possible » states via stochastic approaches tends to make calculation complex be it regarding the time that calculations take but also on a mathematic point of view. This is why insurance companies are pushed to consider simplifications in their calculus approach. Setting up these simplified methods must however take into account the risk profiles while still juggling with the cost of implementation, the awaited precision of results, the control over the justification of the results and the sensitivity to the hypothesis.

The aim of this master thesis is to compute a parametric function based on the SCR and EOF through machine learning techniques so as to forecast the coverage ratio :

$$y_t = f_t(\epsilon_1, \dots, \epsilon_n)$$

where :

$$\begin{aligned} f : \mathbb{R}^n &\longrightarrow \mathbb{R} \\ \epsilon &\longmapsto p(\epsilon) \end{aligned}$$

with :

$$\begin{aligned} p(\epsilon) &\text{ a polynomial function} \\ \epsilon_i &\text{ the } i\text{-th risk factor} \end{aligned}$$

This approximation relies on elementary risk factors potentially affecting the best estimate, the eligible own fund as well as the sub risk modules of the standard formula. Unlike other methods that have been developed so far, the statistical learning will allow us to extract the economic information

from our balance sheet in order to infer an economic value.

Thereafter, this approximation of sub risk modules will allow us to compute the SCR Market/Life in order to compute the global SCR of the insurance company following a change in economic hypothesis.

The figure below illustrates the proxy methodology we have implemented

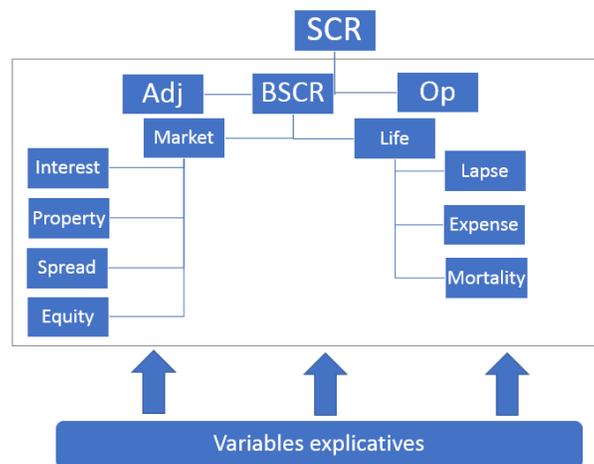


FIGURE 14.4 – Proxy methodology

However, to compute such a function, data are needed in order to fit our function and then, predict the value of the SCR and EOF following a change in economic hypothesis.

Construction of the data base

In order to be able to predict the amount of SCR based on different economic situations we need to create the most representative database possible. Indeed, the machine learning algorithms can't predict a situation they haven't experienced yet. The aim of this first step is then to define a broad range of economic situations. We have chosen the following variables :

- the rate of loading
- the minimum guarantee rate
- the age of the insured
- the proportion of government bond
- the proportion of corporate bond
- the proportion of equity
- the proportion property
- the proportion of cash
- the 1, 8 and 20 year interest rate
- the percentage of R332-20 products
- the difference between the minimum guarantee rate and the 8-year interest rate
- the difference between the minimum guarantee rate and the minimum 8-year interest rate

We want to construct a reduced database based on our 13 variables that might be the most spread out in our projection set while optimizing the computing time. To do so, we have been using the latin hypercube sampling.

The algorithm aims at numerically computing multiple integral ensuring the non-redundant information through a projection within the factorial axes. To do so, the range of each input x_i is divided into N segments with equal probability $\frac{1}{N}$ resulting in the following partition : $\{[0, \frac{1}{N}], [\frac{1}{N}, \frac{2}{N}], \dots, [\frac{N-1}{N}, 1]\}$.

We will then randomly draw a variable in each of the N segments. Once a point has been randomly drawn from one of the segment x_i , no other point can be put into this segment, ensuring that each and every point is placed one time.

This method guarantee a uniform representation of all the dimensions avoiding to sub-sample or over-sample some dimensions. The following graph displays the LHS method applied to the minimum guarantee rate and the rate of loading in a $d=2$ dimension space with $N=100$ points. As we can see, the LHS method is more efficient for covering all the space compared to the randomly draw.

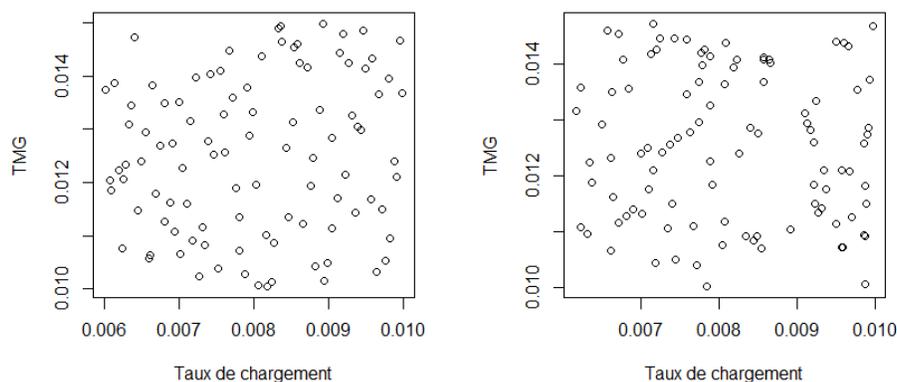


FIGURE 14.5 – LHS against random drawn

Defining the proxy

We will distinguish the proxy implemented at time $t = 0$ to the one implemented at time $t > 0$. Indeed, between t and $t + 1$, our risk indicator will evolve as well as the economic environment. To take into account this change, for example a decrease in the market value, we will compute a last coefficient called the « ALM coefficient » denoted by δ . The aims of this coefficient is to capture the change in the value of our risk indicator.

Eventually, we get the final relationship :

$$\begin{aligned}
 S\tilde{C}R_t &= f_t(\epsilon_1, \dots, \epsilon_n)\mathbb{1}\{t = 0\} + f_{t-1}(\epsilon_1, \dots, \epsilon_n)\delta_{t-1}(\epsilon_1, \dots, \epsilon_n)(1 - \mathbb{1}\{t = 0\}) \\
 F\tilde{P}_t &= f'_t(\epsilon_1, \dots, \epsilon_n)\mathbb{1}\{t = 0\} + f'_{t-1}(\epsilon_1, \dots, \epsilon_n)\delta'_{t-1}(\epsilon_1, \dots, \epsilon_n)(1 - \mathbb{1}\{t = 0\})
 \end{aligned}$$

The previous relationship exhibit the need to correctly fit the function at time $t = 0$ since for $t > 0$, our risk indicator are only a proportion of \tilde{SCR}_0 and $E\tilde{OF}_0$.

The coverage ratio is computing as :

$$RC_t = \frac{\tilde{SCR}_t}{\tilde{FP}_t}$$

where ϵ_i are the risks factors and $f_t(\cdot)$, $f'_t(\cdot)$, $\delta_t(\cdot)$ and $\delta'_t(\cdot)$ are our proxies computing by statistical learning.

Estimated results

Our statistical learning has been made on a unique model point on which we get very good result. Indeed, the errors of approximation on the SCR , the EOF and the coverage ratio are respectively of 7%, 2% and 7% in absolute value on the duration of the business plan. This first result is quite encouraging since it enlighten us on the ability of the statistical learning method to answer to insurance life needs.

Pas	$SCR_{Défaut}$	$SCR_{Opérationnel}$	BSCR	Adj TP	SCR	FP	RC
t=0	0%	0.11%	0.37%	1.57%	1.08%	1.1%	2.1%
t=1	0%	0.16%	0.98%	12.5%	5.6%	1.41%	6.3%
t=2	0%	0.18%	1.1%	13.4%	6.07%	1.47%	6.5%
t=3	0%	0.2%	1.17%	13.48%	6.09%	1.55%	6.56%
t=4	0%	0.21%	1.18%	13.5%	6.16%	1.58%	6.6%
t=5	0%	0.25%	1.6%	13.58%	6.77%	1.65%	6.8%

TABLE 14.4 – Error of approximation on the coverage ratio

Conclusion

The aim of this master thesis was to propose a methodology based on machine learning techniques in order to forecast the coverage ratio in an ORSA framework. The use of these statistical learning methods answer to two main goals :

- reducing the computation time
- mathematical simplification

These goals answer to the needs of insurance companies to be able to permanently evaluate and at a lower cost the capital requirement following a change in economic hypothesis (increase or decrease in interest rates).

Our models required a week of work to define the scope of study of our insurance company before training and testing the accuracy of our models. Thus, at the end of this computation time of one week, we have a robust function, enable us to perform stress-test instantaneously with great accurateness during the year. This function also enable us to forecast the *SCR*, the *EOF* and the coverage ratio on the duration of the business plan with errors being respectively below 7%, 2% and 7%.

However, when forecasting the coverage ratio on an adverse event (decrease in interest rate), our models poorly performed. This test shows the weaknesses of the statistical learning method namely that they can only predict on known situations. In order to be able to get accurate results on these adverses events, we should have calibrated our models on a new database which take into account extremes events. Hence, the most important step of our methodology is the construction of the database et the hypothesis that we want to define. At last, our model has been tested on a unique model point with simplified hypothesis and needs to be tested on a real portfolio.

At the end of this study, we have been able to develop a methodology based on machine learning. This has allowed us to get a proxy that is easy to use to predict the capital requirement in $t = 0$ or on the duration of the business plan of the insurance company. This proxy is specific to each insurance company since it's based on its own risk profile and characteristics.

Table des figures

1.1	Bilan économique sous S1	13
1.2	Bilan économique sous S2	15
1.3	Les piliers de Solvabilité 2	16
1.4	Structure de la formule standard	18
1.5	Méthode de Simulation dans les Simulations	20
2.1	Processus ORSA	21
3.1	Méthode curve-fitting	24
3.2	Méthode Least Square Monte-Carlo	25
3.3	Bilan économique sous S2	27
4.1	Décomposition du portefeuille d'actif	30
5.1	Hypercube Latin (à gauche) contre tirage aléatoire (à droite)	36
5.2	Répartition LHS	36
5.3	Courbes des taux choquées en inputs	39
5.4	Matrice de corrélation de Pearson	40
5.5	Matrice de corrélation de Kendall	42
6.1	Problème du LASSO-Ridge	46
6.2	Cross-Validation	48
6.3	Variabes LASSO	48
6.4	Fiscalité des rachats structurels	50
6.5	Cross-Validation	51
6.6	Variabes LASSO	51
7.1	Représentation des Support Vector Regression	55
7.2	Données non-linéairement séparables	58
7.3	Transformation des données via un noyau	59
7.4	Transformation de l'espace des données via un Kernel	59
8.1	Neurone en biologie	61
8.2	Réseau de neurones formel	63
8.3	Perceptron multi-couche	65
8.4	Rétropropagation de l'erreur	66
8.5	Descente de gradient	67
9.1	Non linéarité du SCR	71
9.2	Méthode « boîte-noire »	71
9.3	Méthode proxy	72
9.4	Choc de spread	75
10.1	Phénomène d'ajustement	81
11.1	Fonctionnement du modèle ALM	87
12.1	Fonctionnement du modèle ALM	88
12.2	Extraction des coefficients ALM	90

12.3	Construction des bilans prospectifs	91
14.1	Evolution des taux à la baisse	97
14.2	Méthode proxy	103
14.3	Hypercube Latin contre tirage aléatoire	104
14.4	Proxy methodology	108
14.5	LHS against random drawn	109

Liste des tableaux

4.1	Composition de l'actif en million d'€	31
4.2	Variables d'actifs	31
4.3	Caractéristiques des obligations	32
4.4	Variables de passif	33
5.1	Projection des indicateurs du GSE	39
7.1	Résultat de la validation croisée	60
8.1	Fonctions de perte	67
10.1	Erreur d'approximation sur la base de test	82
10.2	Erreur d'approximation SCR brut marché	83
10.3	Erreur d'approximation SCR brut vie	83
10.4	Erreur d'approximation SCR global	84
10.5	Erreur d'approximation RC	84
12.1	Erreur d'approximation SCR brut marché	92
12.2	Erreur d'approximation SCR brut marché	93
12.3	Erreur d'approximation RC	94
14.1	Hypothèses scénario baisse des taux	98
14.2	Erreur d'approximation scénario de baisse des taux	98
14.3	Erreur d'approximation RC	105
14.4	Error of approximation on the coverage ratio	110