

Mémoire présenté pour l'obtention du DUAS et l'admission à l'Institut des Actuaire

Le 6 décembre 2024

Par : Sébastien MEYER

Titre : Évaluation des engagements en assurance vie à l'aide de méthodes de *machine learning*

Confidentialité : NON OUI Durée : 1 an 2 ans 3 ans 4 ans 5 ans

Membres du jury de l'IA :

Gregory BOUTIER

Alexandre YOU

Membres du jury de l'Unistra :

J. BERARD

~~E. BIRMELE~~

~~A. COUSIN~~

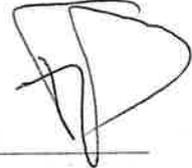
P.-O. GOFFARD

~~M. MAUMY-BERTRAND~~

Entreprise : **PRIM'ACT**

Directeur de mémoire (entreprise) :

Nom : Frédéric PLANCHET



Signature du responsable entreprise



Secrétariat : Mme Stéphanie Richard

Signature du candidat



Remerciements

«Il n’y a pas d’excès au monde aussi louable
que la gratitude excessive.»

LA BRUYÈRE

Je tiens en premier lieu à remercier tous ceux qui m’ont encouragé à reprendre des études en actuariat – famille et amis – premiers instigateurs de ce présent mémoire. Dans une même optique, je souhaite exprimer ma gratitude envers l’ensemble du corps enseignant de l’UNIVERSITÉ DE STRASBOURG ainsi qu’envers mes camarades de promotion. Ils m’ont donné l’occasion de revivre le plaisir de se sentir étudiant.

Merci à l’ensemble du cabinet PRIM’ACT de m’avoir permis de travailler sur ce sujet bien entouré. En particulier, je remercie Quentin GUIBERT et Frédéric PLANCHET. Finalement, j’adresse mes remerciements à mes encadrants, Pierre-Olivier GOFFARD et Sugiban RATNASOTHY pour leur suivi.

À mon regretté grand-père, à qui je dois mes valeurs dont je suis le plus fier.

Table des matières

Remerciements	i
Note de Synthèse	ix
Executive Summary	xv
Acronymes utilisés	xxi
<hr/>	
Introduction	1
1 Cadre d'étude	3
1 Le référentiel Solvabilité II	3
2 Le produit modélisé : l'assurance vie épargne euro	4
3 Les indicateurs actuariels clés	5
4 L'évaluation usuelle des engagements en assurance vie	6
<hr/>	
I Calcul des indicateurs par modèles GSE et ALM	9
2 Générateur de scénarios économiques et ALM	11
1 Le générateur de scénarios économiques	11
2 La gestion actif-passif : le modèle ALM	13
3 Création d'une base de données d'entraînement	21
1 Une vision paramétrique des inputs	21
2 La génération de paramètres synthétiques	23
3 Une analyse descriptive de la base obtenue	23

II	Proxy de calcul des engagements en <i>machine learning</i>	29
4	Mise en place d'un modèle de <i>machine learning</i> «distributionnel»	31
1	Les modèles linéaires	31
2	La régression étendue aux facteurs de forme : le modèle GAMLSS	33
3	Les réseaux de neurones et l'apprentissage distributionnel	34
4	Architecture finale du modèle	44
5	Résultats et discussion	47
1	Les performances des modèles	47
2	Comparaisons	52
3	Quelques interprétations	53
	Conclusion	55

Résumé

L'objectif de ce mémoire est d'explorer des alternatives aux méthodes usuelles exhaustives utilisant successivement un générateur de scénarios économiques (GSE) et un modèle de gestion actif-passif (ALM) pour le calcul du Best Estimate of Liabilities (BEL) des contrats d'épargne en euros, en utilisant comme proxy des techniques de machine learning. Nous cherchons à modéliser non seulement la valeur du BEL, mais aussi la distribution de l'incertitude liée aux différents scénarii économiques autour de cet estimateur, ce qui constitue un enjeu central dans la gestion des risques pour les assureurs.

Plusieurs modèles ont été testés, allant du GLM aux réseaux de neurones, en passant par des modèles comme le GAMLSS. Bien que les résultats obtenus soient globalement intéressants, ils ne performant pas significativement mieux que des modèles plus simples. Le réseau de neurones, malgré une plus grande flexibilité, ne permet pas d'obtenir une amélioration substantielle en termes de prédiction du BEL.

La principale contribution de ce mémoire réside dans la démonstration de la faisabilité d'une modélisation de la distribution du BEL à travers une telle approche, tout en soulignant les défis inhérents à l'amélioration de la précision de tels modèles dans un cadre actuariel.

Mots-clés : Générateur de Scénarios Économiques (GSE) ; Gestion Actif-Passif (ALM) ; Best Estimate (BE) ; Modélisation de distributions ; Machine Learning ; Réseaux de neurones ; GLM ; GAMLSS ; Épargne en euros ; Risque actuariel.

Abstract

The aim of this thesis is to explore alternatives to the use of economic scenario generator (ESG) and asset-liability management tools (ALM) for calculating the Best Estimate of Liabilities (BEL) of euro-denominated savings contracts, by using machine learning techniques as a proxy. We seek to model not only the BEL value but also the distribution of the uncertainty around this estimator induced by the different scenarios from the ESG, which is a key challenge in risk management for insurers.

Several models were tested, ranging from GLM to neural networks, including models such as GAMLSS. While the results obtained are generally satisfactory, they do not significantly outperform simpler models like GLM. Neural networks, despite their greater flexibility, do not provide substantial improvement in terms of BEL prediction.

The main contribution of this thesis is to demonstrate the feasibility of modeling the distribution of the BEL using machine learning methods, while highlighting the challenges inherent in improving the accuracy of these models in an actuarial context.

Keywords : Economic Scenario Generator (ESG) ; Asset-Liability Management (ALM) ; Best Estimate (BE) ; Distribution modeling ; Machine Learning ; Neural Networks ; GLM ; GAMLSS ; Euro-denominated savings ; Actuarial risk.

Note de Synthèse

Introduction

Dans un contexte réglementaire rigoureux, les assureurs vie doivent évaluer leurs engagements financiers avec précision, notamment pour le Best Estimate of Liabilities (BEL), un indicateur central du dispositif Solvabilité II. Ce mémoire s'inscrit dans la recherche d'alternatives aux méthodes classiques, en explorant l'apport des techniques de machine learning (ML) pour modéliser non seulement la valeur du BEL, mais aussi la distribution des incertitudes économiques autour de cet estimateur. L'objectif est d'établir un modèle plus efficace et transparent pour la valorisation des passifs, permettant une modélisation rapide et des analyses de sensibilité approfondies, tout en maintenant la rigueur exigée par les normes actuelles. Le mémoire s'articule autour de la construction d'un modèle proxy de calcul d'engagements en assurance vie épargne euro contournant l'utilisation d'un GSE puis d'un modèle ALM.

Cadre d'étude

Le cadre d'étude repose sur Solvabilité II, une réglementation européenne visant à protéger les assurés en imposant des exigences strictes de solvabilité aux entreprises d'assurance. Solvabilité II est structuré en trois piliers : les exigences quantitatives (Pilier 1), les exigences qualitatives pour la gestion des risques (Pilier 2), et la transparence à travers un reporting réglementé (Pilier 3). Le BEL constitue une estimation centrale dans le calcul des passifs d'assurance. Traditionnellement, il est calculé à l'aide d'un Générateur de Scénarios Économiques (GSE) pour projeter des trajectoires financières, couplé à un modèle de gestion actif-passif (ALM) pour estimer les flux de trésorerie futurs.

Cette approche traditionnelle, bien que fiable, est limitée par la complexité des calculs et la difficulté de modéliser des relations non linéaires entre variables économiques et engagements financiers. En réponse à ces contraintes, l'application des techniques de ML pourrait permettre de capturer de manière plus flexible ces relations complexes, tout en réduisant le temps de calcul et en améliorant la transparence des projections.

Générateur de Scénarios Économiques

Le Générateur de Scénarios Économiques (GSE) est une composante essentielle de l'évaluation des engagements. Il permet de simuler les trajectoires des indices de marché et des taux d'intérêt pour divers actifs, sur lesquels reposent les engagements de l'assureur. Dans cette étude, le GSE repose sur des modèles classiques de diffusion : le modèle de Black-Scholes pour les actions et les rendements immobiliers, et le modèle de Hull-White (ou sa simplification en modèle de Vasicek) pour les taux d'intérêt. Ces modèles permettent de générer un ensemble de trajectoires stochastiques en prenant en compte les risques inhérents aux actifs sous-jacents.

Les simulations stochastiques sont indispensables pour évaluer des engagements garantis, comme les produits d'assurance vie en euros, en raison de la présence d'options intégrées dans ces contrats (comme le Taux Minimum Garanti). Toutefois, cette approche reste complexe et lourde à mettre en œuvre, d'où la motivation d'explorer des modèles ML qui puissent non seulement simplifier ce processus mais aussi fournir des résultats comparables en termes de précision.

Gestion actif-passif

Le modèle ALM permet de projeter les flux de trésorerie en fonction des actifs et passifs, tout en intégrant des paramètres réglementaires et économiques, essentiels dans un cadre de gestion des risques. En simulant les actifs sous-jacents (actions, obligations, immobilier, etc.) et leurs caractéristiques (volatilité, rendement, etc.), le modèle ALM permet de calculer les engagements futurs. Les contraintes réglementaires imposées par Solvabilité II sont également intégrées, avec des scénarios de chocs pour évaluer la résilience des passifs face aux fluctuations économiques.

L'évaluation des passifs s'appuie traditionnellement sur des simulations imbriquées coûteuses en calcul, mais nécessaires pour évaluer des scénarios complexes et la répartition des flux financiers. Cependant, ces projections demandent une grande quantité de ressources en temps et en puissance de calcul. L'utilisation du ML pourrait représenter un proxy efficace pour simplifier ces projections, tout en conservant la capacité d'analyse des scénarios économiques essentiels.

Création d'une base de données

La construction d'une base de données d'entraînement constitue un point fondamental pour tester et comparer les différents modèles de ML proposés. Dans ce mémoire, la base de données a été générée de manière synthétique, à partir de distributions paramétriques pour les variables clés, alignées sur les paramètres de Solvabilité II. La base inclut des informations relatives aux actifs (par exemple, actions, obligations et immobilier), aux engagements (provisions mathématiques, rachats, mortalité) et aux chocs économiques simulés.

Chaque trajectoire est calibrée pour représenter des conditions économiques réalistes, en se basant sur des corrélations empiriques entre les différentes classes d'actifs. La normalisation par la provision mathématique (PM) a été utilisée pour permettre des comparaisons cohérentes entre les différents contrats et modèles d'engagements. Cette approche permet de tester la capacité des modèles ML à prédire des engagements diversifiés et à généraliser sur de nouveaux scénarios économiques.

Modélisation des engagements

Cette section présente les modèles utilisés pour modéliser les engagements et décrire l'incertitude des estimations en assurance. Trois modèles principaux sont étudiés : le modèle linéaire généralisé (GLM), le modèle GAMLSS (Generalized Additive Model for Location, Scale, and Shape), et un réseau de neurones.

Modèle GLM

Le GLM, ou modèle linéaire généralisé, constitue une méthode de base dans cette analyse. Il est construit en utilisant une sélection de variables basée sur le critère AIC, avec une approche de sélection *backward*. Le GLM se concentre sur la relation linéaire entre les variables explicatives et l'engagement, ce qui en fait une référence initiale pour comparer la performance de modèles plus complexes. Cependant, il peut s'avérer limité lorsqu'il s'agit de capturer des relations non linéaires ou des distributions asymétriques dans les données.

Modèle GAMLSS

Le modèle GAMLSS étend le GLM en permettant de modéliser plusieurs paramètres de distribution, incluant non seulement la moyenne, mais également l'échelle et la forme. Ces paramètres permettent de capturer des asymétries ou des queues lourdes dans la distribution, aspects fréquents dans les données d'assurance. La structure du GAMLSS est définie par un système d'équations où chaque paramètre de la distribution (position μ , échelle σ , asymétrie ν , et kurtosis τ) est lié à un ensemble de covariables via une fonction de lien spécifique :

$$\begin{cases} g_1(\mu) = X_1\beta_1 + \sum_{j=1}^{J_1} Z_{j1} \cdot \gamma_{j1}, \\ g_2(\sigma) = X_2\beta_2 + \sum_{j=1}^{J_2} Z_{j2} \cdot \gamma_{j2}, \\ g_3(\nu) = X_3\beta_3 + \sum_{j=1}^{J_3} Z_{j3} \cdot \gamma_{j3}, \\ g_4(\tau) = X_4\beta_4 + \sum_{j=1}^{J_4} Z_{j4} \cdot \gamma_{j4}. \end{cases}$$

Cette flexibilité permet de mieux ajuster la distribution aux données observées en tenant compte de l'hétérogénéité de l'échantillon et des potentiels déséquilibres dus aux garanties ou options incluses dans les contrats d'assurance.

Réseaux de neurones et régression distributionnelle

Nous proposons ici un cadre d'étude qui s'éloigne de la tâche classique de régression d'une quantité d'intérêt en fonction de covariables potentiellement explicatives $Y = \phi(X)$. Dans notre cas, le BEL à partir des données et paramètres initiaux lors de son calcul. On propose de modéliser plutôt la **distribution conditionnelle aux variables explicatives qui a engendré les observations** de la variable aléatoire Y . En effet, au vu de nos données qui ne sont pas des observations uniques mais bien un vecteur de plusieurs (ici 200) scenarii pour chaque condition initiale, nous choisissons d'intégrer toute l'information apportée par ces observations sur la distribution latente plutôt que de la résumer par une statistique telle que la moyenne empirique qui nous donnerai le BEL.

Plus formellement on pose l'hypothèse suivante : $Y|X \sim \mathcal{D}_{\theta=\phi(X)}$ avec \mathcal{D} une distribution et θ sa paramétrisation. En d'autres termes, la distribution des flux d'engagements futurs issus des trajectoires aléatoires du GSE étant donné un environnement économique dépend fonctionnellement de ces mêmes hypothèses de départ au travers d'un lien que l'on cherche à mettre en lumière.

On cherche ϕ la fonction (qu'il faut ici voir comme un réseau de neurones entraîné) qui minimise la divergence de Kullback-Leibler entre les distributions issues de la paramétrisation inférée $\hat{\theta}$ (l'estimateur de maximum de vraisemblance dans la famille de loi considérée inféré sur les observations, ici des scenarii) et de celle prédite $\hat{\phi}(X)$ par notre modèle de *machine learning*.

$$\hat{\phi} = \operatorname{argmin}_{\phi \in \Phi(\text{NN})} D_{\text{KL}}(f_{\mathcal{D}_{\phi(X)}} \parallel f_{\mathcal{D}_{\hat{\theta}}}) \quad (1)$$

$\Phi(\text{NN})$ est ici l'espace des fonctions représentables par un réseau de neurone NN d'une architecture donnée. $f_{\mathcal{D}_{\hat{\theta}}}$ la densité associée à la distribution ajustée sur les observations de Y et $D_{\text{KL}}(\cdot \parallel \cdot)$ la divergence de Kullback-Leibler.

La divergence de Kullback-Leibler comme fonction de perte

Lorsque l'on souhaite comparer des distributions, il vient assez vite l'envie d'étudier des distances entre distributions. La notion de divergence de Kullback-Leibler tombe alors sous le sens. Nous allons tenter de nous convaincre de sa pertinence et son interprétabilité dans un cadre simplifié.

Définition formelle dans le cadre des distributions continues :

$$D_{\text{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (2)$$

Cette divergence correspond à l'espérance du logarithme de la statistique du ratio de vraisemblance. Elle est aussi appelée gain d'information en ML en cela qu'elle renseigne par le calcul d'entropie le manque à gagner informationnel entre P et Q. De nombreuses interprétations existent, notamment bayésiennes. Nous allons nous concentrer sur le calcul de cette divergence pour la loi normale – car elle permet une formule fermée simple de sa divergence – afin de dériver une meilleure fonction objectif pour nos modèles. On retrouvera un lien avec le modèle *Weighted Least Squares*.

Résultats

Les résultats montrent une performance variable selon les modèles, mettant en lumière les avantages et limites de chacun. Le GLM offre une base de référence solide et interprétable, mais se révèle insuffisant pour capturer des relations complexes entre les variables.

Modèle	R^2 (Train)	R^2 (Test)	Temps d'entraînement
GLM	0.76	0.59	Faible
WLS	0.72	0.53	Faible
GAMLSS	0.75	0.60	Modéré
Réseau de Neurones	0.70	0.65	Plus élevé

TABLE 1 : Comparaison des performances des modèles

Les réseaux de neurones fournissent les meilleures performances de généralisation sur les données de test non vues, bien que leur interprétation reste plus complexe.

Variable	Plage des Coefficients
theta_alpha	[-4.034, -3.534]
alpha	[-0.0239, -0.0198]
sigma_taux	[-1.8744, -1.3880]
r	[0.7757, 0.9844]
sigma_action	[-0.4357, -0.1863]
marge_mar	[-0.1258, -0.0813]
rach_tot_from	[1.4710, 1.5575]

TABLE 2 : Variables consensuelles et plages de coefficients associées

Conclusion et perspectives

En conclusion, cette étude démontre la faisabilité et la pertinence d'utiliser des techniques de machine learning pour l'évaluation des engagements en assurance vie. Bien que les modèles traditionnels (GLM, GAMLSS) offrent une interprétation plus directe des résultats, les méthodes de ML, et en particulier les réseaux de neurones, présentent un potentiel intéressant pour améliorer la flexibilité et la précision des prédictions dans des contextes complexes. Cependant ces méthodes nécessitent une quantité de données conséquente qu'il est difficile d'avoir dans ces contextes de par les outils chronophages utilisés pour leur génération.

Executive Summary

Introduction

In a strict regulatory context, life insurers must assess their financial commitments accurately, particularly for the Best Estimate of Liabilities (BEL), a central indicator within the Solvency II framework. This thesis explores alternatives to traditional methods by examining the contribution of machine learning (ML) techniques to model not only the value of the BEL but also the distribution of economic uncertainties surrounding this estimator. The aim is to establish a more efficient and transparent model for liability valuation, allowing for rapid modeling and in-depth sensitivity analyses while maintaining the rigor required by current standards. The thesis is organized around constructing a proxy model for calculating liabilities in euro-denominated life insurance savings, bypassing the use of an ESG and then an ALM model.

Study Framework

The study framework is based on Solvency II, a European regulation aimed at protecting policyholders by imposing strict solvency requirements on insurance companies. Solvency II is structured into three pillars: quantitative requirements (Pillar 1), qualitative requirements for risk management (Pillar 2), and transparency through regulated reporting (Pillar 3). The BEL is a central estimate in the calculation of insurance liabilities. Traditionally, it is calculated using an Economic Scenario Generator (ESG) to project financial paths, coupled with an asset-liability management (ALM) model to estimate future cash flows.

While this traditional approach is reliable, it is limited by the complexity of calculations and the difficulty in modeling non-linear relationships between economic variables and financial commitments. In response to these constraints, ML techniques could enable a more flexible capture of these complex relationships, reducing computation time and enhancing projection transparency.

Economic Scenario Generator

The Economic Scenario Generator (ESG) is a key component in liability assessment. It allows for simulating the trajectories of market indices and interest rates for various assets on which the insurer's liabilities depend. In this study, the ESG is based on classic diffusion models: the Black-Scholes model for stocks and real estate returns, and the Hull-White model (or its simplification into the Vasicek model) for interest rates. These models generate a set of stochastic paths by accounting for the inherent risks in the underlying assets.

Stochastic simulations are essential to assess guaranteed commitments, such as euro-denominated life insurance products, due to the presence of embedded options in these contracts (such as the Guaranteed Minimum Rate). However, this approach remains complex and cumbersome to implement, motivating the exploration of ML models that could simplify this process while providing comparable results in terms of accuracy.

Asset-Liability Management

The ALM model projects cash flows based on assets and liabilities while integrating regulatory and economic parameters essential in a risk management context. By simulating the underlying assets (stocks, bonds, real estate, etc.) and their characteristics (volatility, return, etc.), the ALM model calculates future liabilities. The regulatory constraints imposed by Solvency II are also integrated, with shock scenarios to assess the resilience of liabilities in the face of economic fluctuations.

Traditionally, liability assessment relies on nested simulations, which are computationally expensive but necessary for evaluating complex scenarios and the distribution of financial flows. However, these projections require a significant amount of time and computational resources. ML could provide an efficient proxy for simplifying these projections while preserving the capacity for analyzing essential economic scenarios.

Database Creation

The construction of a training database is fundamental for testing and comparing the proposed ML models. In this thesis, the database was generated synthetically from parametric distributions for key variables aligned with Solvency II parameters. The database includes information related to assets (e.g., stocks, bonds, and real estate), liabilities (mathematical provisions, surrenders, mortality), and simulated economic shocks.

Each path is calibrated to represent realistic economic conditions, based on empirical correlations between the different asset classes. Normalization by the mathematical provision (MP) was used to allow consistent comparisons between different contracts and liability models. This approach allows for testing the ability of ML models to predict diversified liabilities and generalize to new economic scenarios.

Liability Modeling

This section presents the models used to model liabilities and describe the uncertainty of estimates in insurance. Three main models are studied: the generalized linear model (GLM), the Generalized Additive Model for Location, Scale, and Shape (GAMLSS), and a neural network.

GLM Model

The GLM, or generalized linear model, is a basic method in this analysis. It is built using variable selection based on the AIC criterion, with a backward selection approach. The GLM focuses on the linear relationship between explanatory variables and liability, making it an initial reference for comparing the performance of more complex models. However, it may be limited in capturing non-linear relationships or asymmetric distributions in the data.

GAMLSS Model

The GAMLSS model extends the GLM by allowing for modeling multiple distribution parameters, including not only the mean but also scale and shape. These parameters capture asymmetries or heavy tails in the distribution, aspects frequently found in insurance data. The GAMLSS structure is defined by a system of equations where each distribution parameter (location μ , scale σ , skewness ν , and kurtosis τ) is linked to a set of covariates via a specific link function:

$$\begin{cases} g_1(\mu) = X_1\beta_1 + \sum_{j=1}^{J_1} Z_{j1} \cdot \gamma_{j1}, \\ g_2(\sigma) = X_2\beta_2 + \sum_{j=1}^{J_2} Z_{j2} \cdot \gamma_{j2}, \\ g_3(\nu) = X_3\beta_3 + \sum_{j=1}^{J_3} Z_{j3} \cdot \gamma_{j3}, \\ g_4(\tau) = X_4\beta_4 + \sum_{j=1}^{J_4} Z_{j4} \cdot \gamma_{j4}. \end{cases}$$

This flexibility allows for a better fit to observed data by accounting for sample heterogeneity and potential imbalances due to guarantees or options included in insurance contracts.

Neural Networks and Distributional Regression

Here, we propose a study framework that diverges from the classic task of regressing an interest quantity based on potentially explanatory covariates $Y = \phi(X)$, focusing instead on the **conditional distribution given the explanatory variables that generated the observations** of the random variable Y . Our data consists of not single observations but rather a vector of several (here, 200) scenarios for each initial condition, allowing us to incorporate the information these observations provide on the latent distribution rather than summarizing it with a statistic like the empirical mean, which would yield the BEL.

More formally, we assume: $Y|X \sim \mathcal{D}_{\theta=\phi(X)}$ with \mathcal{D} a distribution and θ its parameterization. In other words, the distribution of future liability cash flows from ESG-driven random paths given an economic environment depends functionally on these same initial assumptions through a link we seek to uncover.

We aim to find ϕ , the function (to be seen here as a trained neural network) that minimizes the Kullback-Leibler divergence between the parameterized inferred distribution $\hat{\theta}$ (the maximum likelihood estimator within the considered distribution family inferred from the observations, here scenarios) and that predicted $\hat{\phi}(X)$ by our *machine learning* model.

$$\hat{\phi} = \operatorname{argmin}_{\phi \in \Phi(\text{NN})} D_{\text{KL}}(f_{\mathcal{D}_{\phi(X)}} \parallel f_{\mathcal{D}_{\hat{\theta}}}) \quad (3)$$

$\Phi(\text{NN})$ here is the space of functions representable by a neural network (NN) of a given architecture. $f_{\mathcal{D}_{\hat{\theta}}}$ is the density associated with the distribution fitted to the observations of Y , and $D_{\text{KL}}(\cdot \parallel \cdot)$ is the Kullback-Leibler divergence.

The Kullback-Leibler Divergence as a Loss Function

When comparing distributions, it is intuitive to examine distances between distributions. The notion of Kullback-Leibler divergence naturally arises. We will aim to understand its relevance and interpretability in a simplified context.

Formal Definition in the Context of Continuous Distributions:

$$D_{\text{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (4)$$

This divergence corresponds to the expectation of the log likelihood ratio. It is also called information gain in ML as it informs about the informational loss between P and Q by entropy calculation. Numerous interpretations exist, including Bayesian ones. We will focus on calculating this divergence for the normal distribution – as it allows for a simple closed-form formula – to derive a better objective function for our models. This links with the *Weighted Least Squares* model.

Results

The results show varying performance across models, highlighting the strengths and limitations of each. The GLM offers a solid and interpretable baseline but proves insufficient to capture complex relationships between variables.

Model	R^2 (Train)	R^2 (Test)	Training Time
GLM	0.76	0.59	Low
WLS	0.72	0.53	Low
GAMLSS	0.75	0.60	Moderate
Neural Network	0.70	0.65	Higher

Table 3: Model Performance Comparison

Neural networks provide the best generalization performance on unseen test data, though their interpretation remains more complex.

Variable	Coefficient Range
theta_alpha	[-4.034, -3.534]
alpha	[-0.0239, -0.0198]
sigma_rate	[-1.8744, -1.3880]
r	[0.7757, 0.9844]
sigma_stock	[-0.4357, -0.1863]
mar_margin	[-0.1258, -0.0813]
total_surrender_from	[1.4710, 1.5575]

Table 4: Consensus Variables and Associated Coefficient Ranges

Conclusion and Perspectives

In conclusion, this study demonstrates the feasibility and relevance of using machine learning techniques for the evaluation of life insurance liabilities. While traditional models (GLM, GAMLSS) offer more direct interpretability of results, ML methods, particularly neural networks, present interesting potential for improving flexibility and prediction accuracy in complex contexts. However, these methods require a substantial amount of data, which is challenging to obtain due to the time-consuming tools used for their generation.

Acronymes utilisés

ALM	<i>Assets & Liabilities Management</i>
AOA	Absence d'Opportunités d'Arbitrage
BEG	<i>Best Estimate Garanti</i>
BEL	<i>Best Estimate of Liabilities</i>
EIOPA	<i>European Insurance and Occupational Pensions Authority</i>
FDB	<i>Future Discretionary Benefits</i>
GAM	<i>Generalized Additive Model</i>
GAMLSS	<i>General Additive Model for Location Scale and Shape</i>
GLM	<i>General Linear Model</i>
GSE	Générateur de Scénarios Économiques
ITS	<i>Implementation Technical Standards</i>
LSMC	<i>Least Squares Monte Carlo</i>
MCR	<i>Minimum Capital Requirement</i>
ML	<i>Machine Learning</i>
ORSA	<i>Own Risk Solvency Assessment</i>
PB	Participation au(x) Bénéfice(s)
PM	Provision Mathématique
PMVL	Plus-ou-Moins-Values Latentes
PPB	Provision pour Participation au(x) Bénéfice(s)
SCR	<i>Solvency Capital Requirement</i>
SdS	Simulations dans les Simulations
TMG	Taux Minimum Garanti
TVOG	<i>Time Value of Options & Guarantees</i>
UE	Union Européenne
ZC	Zéro-Coupon

Introduction

«Trouver n'est rien,
c'est le plan qui est difficile.»

Fiodor DOSTOÏEVSKI

Contexte

En qualité de membre de l'UE¹, la France dispose d'un cadre réglementaire très uniformisé qui se traduit pour la pratique actuarielle par la conformité aux directives européennes. Solvabilité II est au cœur de beaucoup de travaux actuariels depuis son entrée en application début 2016. Les indicateurs de santé financière des assureurs de la place sont strictement encadrés par ce dispositif, tant dans le calcul de ces derniers que dans leur communication aux parties prenantes. La juste évaluation des engagements que ces entreprises contractent envers leurs assurés et la capacité de les honorer constituent de fait des informations cruciales à l'ensemble du marché.

Problématique

Habituellement, le calcul de ces engagements s'effectue par l'utilisation successive de deux outils puissants mais chronophages, à savoir un GSE² et un modèle de gestion actif-passif dit ALM³. Cette composition de modèles rend quelque peu indirectes et donc opaques les relations existantes entre les indicateurs calculés et l'environnement économique et structurel dans lequel ils l'ont été.

L'adoption d'une approche de calcul plus directe, répondant au double enjeu de *(i)* : éclaircir les impacts structurels et conjoncturels sur la distribution des engagements et *(ii)* : s'affranchir du frein que représente le temps de calcul, représente le cœur de notre sujet. Une telle modélisation rendrait possibles des analyses de sensibilité très rapides pouvant servir dans un cadre de pilotage stratégique ou dans l'élaboration d'un ORSA⁴. Des enjeux similaires sont examinés dans l'article BONNIN et al. (2014) publié dans l'*European Actuarial Journal* ; les auteurs allant même jusqu'à chercher une formule fermée.

1. Union Européenne
2. Générateur de Scénarios Économiques
3. *Assets & Liabilities Management*
4. *Own Risk Solvency Assessment*

Approche

Dans un monde où la technologie occupe une place prépondérante et l'intelligence artificielle devient presque un standard, beaucoup se sont déjà questionnés sur l'éventuelle pertinence de méthodes plus modernes de modélisation. Il existe en effet une part non négligeable de mémoires traitant de méthodes d'apprentissage automatique – plus communément mentionnées sous la dénomination ML⁵ – et de leur apport à la pratique actuarielle. Une recherche bibliographique rapide sur l'état de l'art autour de l'utilisation du ML dans les mémoires d'actuariat nous apprend que non loin de 250 d'entre eux «mentionnent» ces techniques dans des applications très diverses, avec une montée en tendance du sujet. Néanmoins, au moment où j'écris ces lignes, seuls cinq d'entre eux traitent selon moi d'une thématique que je qualifierai de proche. Ce sont les mémoires suivants : ROUÉ (2015), GAUVILLE (2017), ZURFLUH (2019), CERISIER (2021) et BEN YOUSSEF (2023).

Il apparaît alors nécessaire de souligner quelques importantes distinctions avec ces travaux :

- Tout d'abord, les indicateurs prédits se limitent au BEL⁶ pour CERISIER (2021) et BEN YOUSSEF (2023), à savoir la moyenne de la distribution des engagements. ROUÉ (2015), GAUVILLE (2017) et ZURFLUH (2019) y intègrent le SCR⁷ afin de projeter le ratio de solvabilité, prenant en compte en un sens les quantiles de cette dernière. **La présente étude se donne pour objectif de modéliser la distribution des engagements dans son intégralité et non le simple estimateur de sa moyenne.**
- Ensuite, les modèles employés se limitent assez souvent aux arbres de décision, forêts aléatoires et méthodes dites de *Boosting*. Le terme «*machine learning*» renvoie aujourd'hui, dans l'imaginaire collectif, plutôt à des architectures plus denses, en réseaux de neurones dont la complexité et le nombre de couches peut amener vers de l'apprentissage profond (*deep learning*) qui ne sont malheureusement pas ou peu explorés. Seul ROUÉ (2015) étudie un modèle d'un tel format. **Il sera ici question d'exploiter le potentiel d'architectures en réseaux de neurones telles que nous voyons utilisées dans des domaines transverses en gardant en toile de fond une volonté d'interprétabilité.**
- Enfin, la flexibilité de paramétrisation initiale et *a fortiori* la répliquabilité semble à première vue assez réduite dans l'ensemble de ces travaux. **L'approche utilisée ici se veut le plus répliquable possible et d'une grande flexibilité autant sur les paramètres conjoncturels du GSE que structurels du modèle ALM, avec une représentation certes simplifiée mais souple et permissive quant à la structure du portefeuille modélisable.**

Après un court chapitre 1 introductif posant le cadre de notre étude, des travaux préliminaires de construction d'une base de données seront présentés au chapitre 3 comme une première partie de mémoire. Une seconde partie permettra de développer le cœur du sujet, à savoir la modélisation d'un proxy de calcul des engagements en ML au chapitre 4.

Bonne lecture,

5. *Machine Learning*

6. *Best Estimate of Liabilities*

7. *Solvency Capital Requirement*

Chapitre 1

Cadre d'étude

1 Le référentiel Solvabilité II

Le référentiel Solvabilité II est structuré en piliers et est régi par un ensemble de textes hiérarchisés en 3 niveaux. Comme son nom l'indique, ce dispositif répond à une problématique de solvabilité des organismes d'assurance au niveau européen. En premier lieu, il s'agit de protéger les assurés en leur garantissant que les acteurs présents sur le marché sont en capacité de faire face à leurs engagements et ce avec un très haut niveau de confiance.

1.1 Les textes encadrant le référentiel :

- Niveau 1 – Les principes généraux : LES DIRECTIVES – Solvabilité II et sa révision Omnibus II (PARLEMENT EUROPÉEN & CONSEIL DE L'UNION EUROPÉENNE, 2009, 2014) : La directive Solvabilité II établit le cadre réglementaire visant à assurer la solvabilité des entreprises d'assurance et à protéger les assurés. La directive Omnibus II introduit les prérogatives de l'EIOPA¹ et apporte des ajustements aux règles du pilier 1, notamment en ce qui concerne les aspects quantitatifs liés à la courbe des taux. Ils sont traduits en droit français pas des ordonnances, arrêtés et décrets que l'on retrouve dans le Code des Assurances.
- Niveau 2 – Les conditions d'applications : LE RÈGLEMENT DÉLÉGUÉ (COMMISSION EUROPÉENNE, 2015) : En tant que règlement européen, le règlement délégué est directement applicable et ce de manière uniforme dans tous les États membres de l'UE, sans nécessiter de mesures supplémentaires de transposition au niveau national.
- Niveau 3 – Les exemples concrets : LES NORMES TECHNIQUES D'EXÉCUTION & LES ORIENTATIONS : La directive Solvabilité II et le règlement délégué sont complétés par deux types de textes de niveau 3 : les normes techniques d'exécution (ou ITS²) et les orientations (ou *guidelines*). Les ITS, proposées par l'EIOPA et adoptées par la Commission européenne ont une application directe. Les orientations, adoptées par l'EIOPA, sont mises en œuvre au niveau national selon la procédure dite « *comply or explain*³ ».

1. *European Insurance and Occupational Pensions Authority*

2. *Implementation Technical Standards*

3. Mettre en œuvre les orientations proposées ou fournir une explication justifiant du contraire.

1.2 Les 3 piliers du référentiel Solvabilité II :

- Pilier 1 – Les exigences quantitatives : Ce pilier comprend les règles de valorisation des actifs et des passifs ainsi que le calcul des exigences de capital, à savoir le MCR⁴ et le SCR. Le SCR peut être calculé à l'aide de la formule standard ou d'un modèle interne.
- Pilier 2 – Les exigences qualitatives : Ce pilier concerne l'évaluation et la gestion des risques au niveau de l'entreprise. Il exige que les compagnies d'assurance identifient, évaluent et gèrent de manière adéquate leurs risques. Cela comprend la mise en place de politiques, procédures et contrôles internes pour atténuer les risques identifiés. L'objectif est d'assurer une solvabilité et une stabilité financière durable tout en renforçant la confiance des parties prenantes dans le secteur de l'assurance. Les deux pendants de ce pilier sont l'ORSA et l'enjeu de gouvernance.
- Pilier 3 – Le reporting et la communication : Ce pilier concerne la transparence et la communication des informations financières par les compagnies d'assurance. Son objectif est de fournir des rapports clairs et complets aux parties prenantes afin d'évaluer la solvabilité et la performance des compagnies d'assurance.

2 Le produit modélisé : l'assurance vie épargne euro

2.1 Généralités

L'assurance vie occupe une place prépondérante dans le paysage financier français en tant que second produit d'épargne préféré des français derrière le livret A avec un encours de 1 842 milliards d'euros recensés fin 2022 dans un communiqué de presse (FRANCE ASSUREURS, 2023).

Parmi les avantages à l'origine d'une telle popularité, nous retrouvons une fiscalité avantageuse, notamment successorale, ainsi qu'un accès simplifié à des supports d'investissement diversifiés. À cet effet, plusieurs supports sont proposés, avec deux grandes catégories de produits :

- les produits «monosupports» : unités de compte ou fonds en euros ;
- les produits «multisupports» : unités de compte et fonds en euros.

2.2 Le fonds en euros

Les fonds en euros possèdent une garantie en capital. La stabilité de ces fonds provient de sa composition principalement obligataire⁵. Ils disposent de deux caractéristiques propres, l'effet cliquet et le taux plancher garantie.

Effet cliquet :

L'effet cliquet correspond au mécanisme de consolidation des rendements par composition des intérêts acquis réintégrés.

4. *Minimum Capital Requirement*

5. (Obligation & trésorerie : 84,7 % ; Action : 7,7 % ; Immobilier : 7,6 %) selon l'AFER au 30/06/2023

Taux plancher garanti :

Le taux plancher garanti est aussi appelé taux minimum garanti et est régi aux articles A132-1 et A132-3 du Code des Assurances. Par la suite, il sera mentionné sous sa dénomination courante TMG⁶.

Par soucis de simplification, et de par sa popularité auprès des français, nous allons nous concentrer sur la modélisation de ce produit. Nous allons modéliser un produit assurance vie monosupport fonds en euros.

3 Les indicateurs actuariels clés

3.1 *Best Estimate of Liabilities*

Cet indicateur correspond à «*la meilleure estimation correspond à la moyenne pondérée par leur probabilité des flux de trésorerie futurs compte tenu de la valeur temporelle de l'argent estimée sur la base de la courbe des taux sans risque pertinente, soit la valeur actuelle attendue des flux de trésorerie futurs.*»⁷

Concrètement, il se calcule comme suit :

$$\text{BEL} = \mathbb{E}^{\mathbb{P} \otimes \mathbb{Q}} \left(\sum_{t \geq 0} \frac{F_t}{(1 + R_t)^t} \right);$$

avec \mathbb{P} la mesure historique amenant à l'évaluation de l'aléa d'origine structurelle des *cashflows* F_t (comme les tables de mortalité et de rachats) et \mathbb{Q} la mesure risque-neutre avec laquelle sont évaluées les trajectoires économiques de taux R_t et la partie financière conjoncturelle des *cashflows* F_t émanant de l'actif.

3.2 *Solvency Capital Requirement*

Nous allons capitaliser sur notre modélisation du BEL afin de retrouver le niveau de SCR à l'aide de chocs sur les conditions de calcul en se servant de la formule standard de Solvabilité II.

Formule standard :

Le SCR représente la quantité de fonds propres dont doit disposer l'entreprise pour être dite «solvable», à savoir être capable de faire face à ses engagements à horizon 1 an dans 99,5 % des cas. Il est calculé par un agrégat de modules et sous-modules selon la formule suivante :

$$\text{SCR} = \sqrt{\sum_{i,j} \rho_{i,j} \cdot \text{SCR}_i \cdot \text{SCR}_j} + \text{SCR}_{op} + \text{SCR}_{intang} + \text{Adj}.$$

Les modules i, j étant le risque de marché, de souscription en vie et en non-vie, en santé et en risque de contrepartie. $\rho_{i,j}$ étant la corrélation entre ces risques, auquel s'ajoute le risque opérationnel, les intangibles et un ajustement. Une représentation schématique de la décomposition du SCR en sous-modules (souvent appelée «pieuvre») est présentée ci-dessous comme figure 1.1.

6. Taux Minimum Garanti

7. Article R351-2 du Code des Assurances transposé de l'article 77 de la directive Solvabilité II

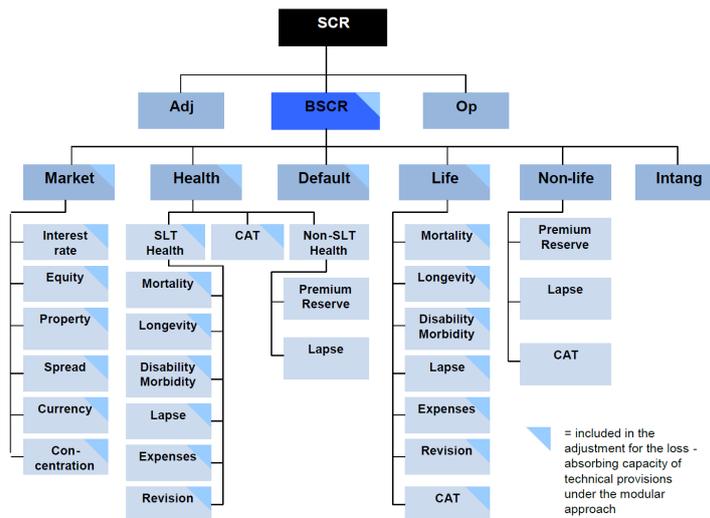


FIGURE 1.1 : Décomposition des composantes risques du SCR (SOURCE : ACPR)

4 L'évaluation usuelle des engagements en assurance vie

4.1 La méthode des simulations dans les simulations

Pour le calcul du SCR, au delà de l'approche par modules de la formule standard, on utilise parfois également une valorisation par un modèle interne. Dans ce contexte, une approche de simulations imbriquées est couramment employée, également connue sous le nom de SdS⁸ ou parfois *nested simulations*. Le processus commence par une première couche de simulations qui concerne l'évaluation de l'espérance sous la mesure empirique \mathbb{P} de la situation comptable à horizon d'un an. Cette situation comptable est composée de passifs calculés comme une espérance risque-neutre sous \mathbb{Q} des flux du portefeuille en *run-off*. Ce processus implique donc un Monte Carlo imbriqué, qui est généralement très chrono-phage. Dans ce contexte, il est rare d'avoir une formule fermée disponible en raison des mécanismes de PB⁹ – qui sont discrétionnaires – ainsi que des dynamiques de rachats conjoncturels. Ces facteurs rendent complexe l'utilisation de méthodes analytiques, ce qui explique le recours aux simulations imbriquées pour obtenir des estimations réalistes des risques en assurance. La table 1.1 ainsi que la figure 1.2 représentent respectivement un bilan minimaliste ainsi que la méthode des SdS projetant ce dernier.

ACTIF	PASSIF
A_t	FP_t
	VEP_t

A_t représente la valeur de marché de l'actif en t
 VEP_t représente la valeur économique du passif
 FP_t le montant des fonds propres disponibles

TABLE 1.1 : Bilan économique minimaliste

8. Simulations dans les Simulations

9. Participation au(x) Bénéfice(s)

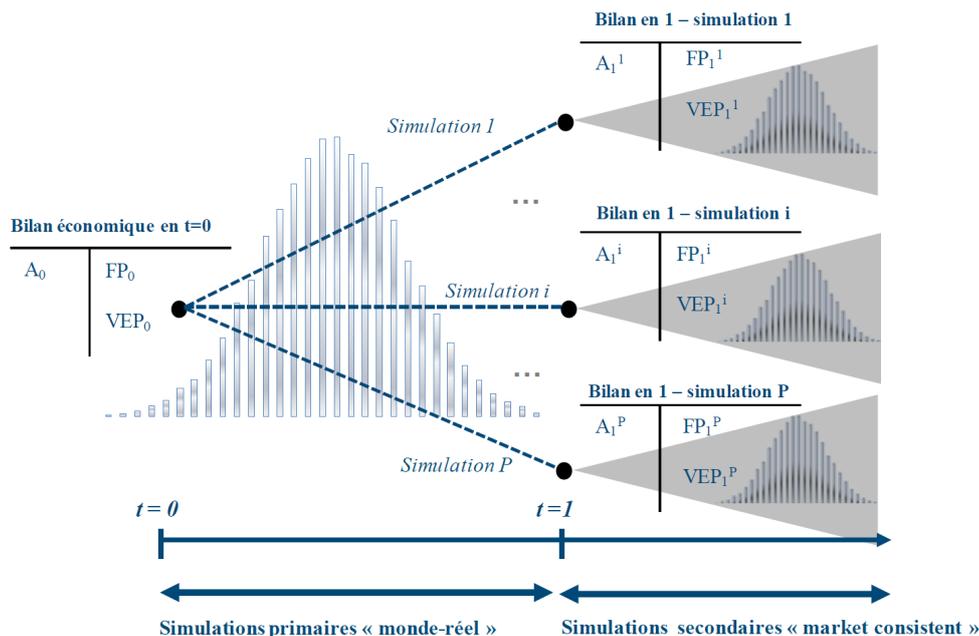


FIGURE 1.2 : Obtention de la distribution de fonds propres économiques par la méthode des simulations dans les simulations (SOURCE : DEVINEAU ET LOISEL (2009))

4.2 Le portefeuille répliquant

La méthode du portefeuille répliquant ou *replicating portfolio* agit également en *proxy* des passifs de l'assureur. L'idée sous-jacente consiste à construire un portefeuille d'actifs simples à modéliser qui réplique les flux de trésorerie attendus des engagements de ce dernier. Pour atteindre cet objectif, le *replicating portfolio* est créé en utilisant une combinaison de produits financiers divers pour lesquels des évaluations en formule fermée sont possibles. Par exemple, les mécanismes asymétriques de TMG et PB peuvent se répliquer par des obligations ZC¹⁰ et des options *calls*. En d'autres termes, en choisissant astucieusement les actifs financiers appropriés, il devient possible de modéliser les flux de trésorerie futurs des passifs de l'assurance et de s'assurer que ces actifs répliquent parfaitement les caractéristiques financières attendues des passifs sous-jacents. Si ce portefeuille réplique tous les flux, par hypothèse d'AOA¹¹, il doit être valorisé à l'identique. Cela permet habilement de se départir des simulations par des formules mathématiques directes de chaque produit simple le composant. Il reste que la calibration et la détermination de ce portefeuille demeure un point délicat et loin d'être trivial.

4.3 Les moindres carrés Monte Carlo

L'approche LSMC¹² est elle aussi utilisée en actuariat comme *proxy* des passifs, mais cette fois-ci en utilisant la régression. Dans cette approche, on remplace l'espérance conditionnelle des flux de trésorerie futurs liés aux engagements de l'assurance par une régression sur une base de fonctions. Pour ce faire, on utilise un grand nombre de scénarios primaires pour tenir compte des différentes évolutions possibles des marchés financiers. Une hypothèse importante de cette méthode est l'indépendance des

10. Zéro-Coupon

11. Absence d'Opportunités d'Arbitrage

12. *Least Squares Monte Carlo*

erreurs en deuxième couche dans cet espace. On remplace alors les moyennes empiriques des scénarios secondaires de passifs par une simple observation grossière primaire, ce qui nous permet d'obtenir un nuage de points représentant les valeurs des actifs et des passifs pour chaque scénario. En ajustant une régression sur cette base de données, on peut estimer un lien fonctionnel entre les flux de trésorerie futurs de manière efficace et ainsi mieux gérer les risques associés aux engagements de l'assurance. Souhaitant développer une approche différente, nous n'élaborerons pas la description exhaustive de cette méthode, bien que très intéressante. Cependant, le lecteur intéressé par un approfondissement théorique de ce sujet trouvera les références suivantes utiles : LY (2015) et BAUER et al. (2010).

Discussion

Limites des approches précédentes

Le temps de calcul inhérent aux scénarios primaires et secondaires est le point bloquant principal pour la plupart des approches sus-mentionnées. La méthode des SdS est la plus chronophage d'entre elles. Le LSMC s'affranchit au moins de faire beaucoup de simulations secondaires en lissant l'incertitude de second niveau par un ajustement paramétrique. Le *curve fitting* à l'inverse se concentre sur les simulations de second ordre. Quant au *replicating portfolio*, sa construction demeure une tâche complexe et ne permet pas toujours de capturer avec une assez grande flexibilité certaines options ou garanties complexes ou bien encore certains risques hors marché.

Perspectives étudiées

Une approche basée sur le ML représenterait potentiellement un *proxy* plus immédiat que d'autres méthodes, tout en offrant une souplesse bien plus grande que des approches de régression. Une structure en réseau parviendrait à capturer des mécanismes bien plus complexes grâce à de multiples couches de résultats intermédiaires. De plus, la possibilité d'interpréter les prédictions, notamment par le biais de techniques telles que la rétro-propagation ou l'analyse de l'importance des variables est un point intéressant. Cependant, cette interprétabilité dépend souvent de la supervision de la forme du réseau.

En bref...

- Solvabilité II impose un cadre de calcul des engagements ;
- Ce mémoire portera sur le produit assurance vie épargne euro ;
- Les indicateurs calculables ne se limitent pas au BEL et au SCR ;
- La distribution complète et paramétrique des engagements futurs sera modélisée.

PREMIÈRE PARTIE

Calcul des indicateurs par modèles GSE et ALM

Chapitre 2

Générateur de scénarios économiques et ALM

Il est difficile de prévoir, surtout lorsqu'il s'agit de l'avenir.

Proverbe danois

1 Le générateur de scénarios économiques

1.1 La nécessité d'une approche stochastique

L'utilisation de trajectoires déterministes pourrait sembler a priori une bonne idée. Seulement, la présence d'options et garanties (le taux plancher du TMG, l'obligation de partage du bénéfice si il y en a par la PB...) dans les contrats d'épargne euro impose l'utilisation d'une évaluation stochastique car l'application de ces options et garanties introduit une asymétrie. Cette asymétrie est assimilable à la valeur temporelle d'une option *call*. Comme l'exercice ou non d'une telle option correspond à une fonction convexe, d'après l'inégalité de Jensen, la moyenne des trajectoires est supérieure à la trajectoire moyenne. La quantité qui matérialise le coût de ces options et garanties est la TVOG¹.

TVOG : La valeur temporelle des options et garanties financières se quantifie en calculant la différence entre la moyenne des valeurs actuelles des bénéfices distribuables à travers un ensemble de scénarios stochastiques et la valeur actuelle des bénéfices distribuables dans un scénario déterministe central, appelé scénario de meilleure estimation. Tous les scénarios doivent être construits en appliquant les mêmes hypothèses de meilleure estimation et une méthodologie cohérente. Les scénarios stochastiques explorent diverses projections des rendements des actifs, tandis que le scénario déterministe repose sur une projection de rendement moyen, généralement alignée sur l'hypothèse de rendement utilisée pour l'évaluation de la valeur intrinsèque.

Formellement, la TVOG peut être exprimée comme :

$$\text{TVOG} = \mathbb{E} [\text{PV}(B)]_{\text{scénarios stochastiques}} - \text{PV}(B)_{\text{scénario déterministe}}$$

1. *Time Value of Options & Guarantees*

où $PV(B)$ représente la valeur actuelle des bénéfices distribuables.

Afin d'intégrer correctement ce coût des options, une approche stochastique s'impose. Pour ce faire, il est d'usage de s'appuyer sur des trajectoires d'indices pour les différentes classes d'actifs ainsi que pour les taux. Ces trajectoires proviennent de modèles de diffusions que l'on regroupe sous la dénomination GSE. Nous allons maintenant présenter deux modèles parmi les plus simples pour les trajectoires d'actions et de taux.

1.2 Le modèle action : Black-Scholes-Merton

La loi du prix d'un actif est supposé dans ce modèle être une loi log-normale. Elle peut s'écrire comme l'équation différentielle stochastique suivante, avec une dérive r correspondant au taux sans risque et σ une volatilité ici constante et adossée à un mouvement brownien B_t :

$$\frac{dS_t}{S_t} = r dt + \sigma dB_t. \quad (2.1)$$

Il devient alors facile de générer des accroissements du logarithme du prix de l'actif par un schéma d'Euler en posant :

$$\ln(S_{t+\delta}) = \ln(S_t) + \left(r - \frac{\sigma^2}{2}\right) \delta + \sigma \mathcal{N}(0, \delta). \quad (2.2)$$

Bien que très simple dans sa formulation et sa paramétrisation, ce modèle reste très largement utilisé en pratique. Aussi, le choix a été fait de considérer les trajectoire de rendement immobilier comme étant des actions dont les dividendes seraient des loyers.

1.3 Le modèle de taux : De Hull & White à Vasicek

Afin de modéliser les taux, essentiels à la fois pour les facteurs d'actualisation des flux et pour les produits sensibles au taux tel que les obligations, nous allons utiliser la version à un facteur du modèle de HULL et WHITE (1993) dont l'équation de diffusion est la suivante :

$$dr_t = (\theta(t) - \alpha(t) \cdot r_t)dt + \sigma(t)dW_t \quad (2.3)$$

Dans le but de simplifier les calculs et d'éviter une sur-paramétrisation de notre modèle final, nous allons considérer la paramétrisation avec $\theta(t), \alpha(t), \sigma(t)$ constantes afin de rester dans une forme paramétrique simple de notre base finale avec des trajectoires dont la distribution est connue et en formule fermée relativement simple. Cela nous ramène au modèle de VASICEK (1977).

$$r_t \sim \mathcal{N}\left(r_0 \cdot e^{-\alpha t} + \frac{\theta}{\alpha} (1 - e^{-\alpha t}), \frac{\sigma^2}{2\alpha} (1 - e^{-2\alpha t})\right). \quad (2.4)$$

Preuve :

Le lemme d'Îto nous donne alors :

$$d(e^{\alpha u}) = e^{\alpha u} (\alpha r_u du + dr_u) = \theta e^{\alpha u} du + \sigma e^{\alpha u} dW_u$$

Par identification d'une intégrale de Wiener, et puisque $\int_0^t f(u) dW_u \sim \mathcal{N}(0, \int_0^t f(u)^2 du)$ on obtient le résultat voulu.

1.4 La structure de corrélation

En pratique, les trajectoires des différentes classes d'actif ne se comportent pas au niveau macro de manière totalement aléatoire. Une certaine cohérence structurelle induit des corrélations entre les différents indices. Nous devons donc répliquer cette structure observée.

Les browniens à l'origine de nos différentes diffusions doivent donc être identiquement corrélés. La décomposition de Cholesky nous permet de transformer un vecteur de variables aléatoires indépendantes en un vecteur de variables aléatoires corrélées, conformément à la matrice de covariance spécifiée. Nous aurions pu utiliser des copules, mais la simplicité de cette approche qui fait apparaître directement les coefficients de corrélation entre les actifs nous convient pour cette étude.

Considérons une matrice de covariance symétrique définie positive Σ . La décomposition de Cholesky de cette matrice nous donne $\Sigma = \mathbf{L}\mathbf{L}^\top$, où \mathbf{L} est une matrice triangulaire inférieure. Si \mathbf{z} est un vecteur de variables aléatoires indépendantes suivant une loi normale standard, alors le vecteur $\mathbf{x} = \mathbf{L}\mathbf{z}$ suivra une loi normale multivariée avec une matrice de covariance Σ .

Preuve :

Soit $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, où \mathbf{I} est la matrice identité.

$$\mathbb{E}[\mathbf{x}] = \mathbb{E}[\mathbf{L}\mathbf{z}] = \mathbf{L}\mathbb{E}[\mathbf{z}] = \mathbf{0}.$$

Donc,

$$\text{Cov}(\mathbf{x}) = \mathbb{E}[\mathbf{x}\mathbf{x}^\top] = \mathbb{E}[\mathbf{L}\mathbf{z}\mathbf{z}^\top\mathbf{L}^\top] = \mathbf{L}\mathbb{E}[\mathbf{z}\mathbf{z}^\top]\mathbf{L}^\top = \mathbf{L}\mathbf{I}\mathbf{L}^\top = \mathbf{L}\mathbf{L}^\top = \Sigma.$$

Ainsi, $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \Sigma)$.

2 La gestion actif-passif : le modèle ALM

L'outil dont nous allons tenter de répliquer les résultats est un *package* **R** open source de calcul de «*best estimates*» pour des contrats d'assurance vie épargne en euros individuels nommé **SimBEL**. Nous

allons parcourir dans les grandes lignes son fonctionnement.

2.1 Les données et paramètres d'entrée

Afin de calculer les engagements dans un contexte réaliste, l'environnement économique est répliqué au travers de paramètres financiers et l'entité d'assurance au travers de son portefeuille et de ses choix de gestion. Tout cela est renseigné au travers de données et paramètres que nous allons passer en revue.

Données :

- **Actif** - Les actifs sont représentés par des lignes de produits détenus dans le portefeuille de l'entité modélisée. Ils se répartissent en quatre grandes classes : les actions, les obligations, l'immobilier et la trésorerie. Chaque titre détenu se voit attribuer un triplet de valorisation comprenant la valeur de marché, la valeur nette comptable et la valeur d'achat. Des informations supplémentaires telles que la cessibilité des actifs, la durée de détention et une éventuelle provision pour dépréciation durable sont également renseignées. La devise du produit financier et le taux de change par rapport à l'euro sont pris en compte. Pour une caractérisation plus fine des actifs, les actions sont distinguées en type 1 ou 2 selon qu'elles soient cotées ou non, et les obligations selon qu'elles soient émises par un État ou une entreprise, avec mention de leur notation. Les loyers immobiliers, les dividendes des actions et les taux de coupons des obligations sont naturellement quantifiés, offrant ainsi une vision complète et détaillée de chaque actif au sein du portefeuille.

- **Passif** - Le passif, quant à lui, est modélisé par des *Model Points*, c'est-à-dire des groupes de contrats représentatifs et homogènes en fonction de leurs caractéristiques clés.

Ces caractéristiques incluent l'âge du souscripteur, le sexe (mentionné à travers la table de mortalité utilisée), ainsi que d'éventuels chargements sur encours. Le montant de la provision mathématique actuelle, l'ancienneté du contrat, un taux technique et un TMG et leurs termes sont également pris en compte. Enfin, des tables de rachats spécifiques sont utilisées pour chaque individu.

Model Points : Le temps de calcul étant un des principaux obstacles à la modélisation ALM, le regroupement de contrats en un nombre restreint de *model points* est un sujet à part entière. Réduire drastiquement les temps de calcul sans perdre d'information sur le portefeuille complet modélisé est un enjeu majeur pour toute compagnie d'assurance disposant d'une grande part de marché. Ici nous avons fait le choix de modéliser un calcul de BEL à partir de 10 *model points*. L'utilisateur souhaitant utiliser notre paramétrisation de modèle pourra trouver des méthodes très prometteuses d'une telle agrégation dans GOFFARD et GUERRAULT (2015), qui nous rassurent également sur la précision obtenue à partir de «seulement» 10 *model points*.

Paramètres :

- **Allocation d'actifs** - L'allocation cible, définie comme la proportion d'actifs détenus dans chacune des quatre classes est spécifiée en amont comme des pourcentages d'actions, d'obligations, d'immobilier et de trésorerie. Pour atteindre cette allocation cible, un mécanisme d'achats et de ventes est mis en place entre le portefeuille actuel et un portefeuille d'investissement de référence. Ce processus permet de réaligner continuellement le portefeuille détenu avec l'allocation stratégique souhaitée. Grâce à cette approche dynamique, l'entité peut maintenir une allocation d'actifs équilibrée et adaptée aux objectifs financiers à long terme.

- **Chocs Solvabilité II** - En plus du scénario central, la norme Solvabilité II impose l'application de divers chocs pour évaluer la résilience du portefeuille. Parmi ces chocs, on trouve les chocs de frais, de mortalité et de longévité, ainsi que les chocs de rachats à la hausse et à la baisse. Les chocs sur les classes d'actifs incluent les chocs sur les actions et l'immobilier. Des chocs de taux d'intérêt, aussi bien à la hausse qu'à la baisse, ainsi que des chocs de *spread* et des chocs de devise dans les deux directions, sont également pris en compte. Ces différents chocs permettent de simuler une gamme étendue de conditions de marché et de risques spécifiques, garantissant ainsi une évaluation complète et robuste de la solvabilité de l'entité.

- **Historique de PPB** - L'historique de la PPB² pour les huit dernières années est également renseigné. La PPB permet aux assureurs de lisser les rendements servis aux assurés au fil du temps. Elle consiste à mettre en réserve une partie des bénéfices réalisés par l'assureur, afin de pouvoir les redistribuer lors des années moins favorables. Selon la réglementation, les bénéfices mis en provision doivent être redistribués aux assurés dans un délai maximum de huit ans.

- **Taux et frais** - Le taux de PB représentant la part des bénéfices redistribuée aux assurés est donné en entrée du modèle pour la part du résultat technique et financier. Les taux de chargement, prélevés sur les primes versées, ainsi que les frais fixes et variables sur encours, qui influencent directement la rentabilité des contrats sont renseignés également. Les frais sur encours, calculés annuellement en pourcentage de l'épargne accumulée, et les frais sur prestations, appliqués lors du dénouement du contrat ou des rachats partiels sont aussi intégrés.

- **Mortalité** - La mortalité est représentée au travers de tables réglementaires pour les hommes et les femmes comme classiquement en actuariat par âge et par génération, au choix par la représentation de population écoulee l_x ou bien en taux de mortalité q_x .

- **Rachats** - Les rachats peuvent être structurels ou conjoncturels tout comme ils peuvent être partiels ou totaux. Les rachats conjoncturels répondent à une loi paramétrique alors que les rachats structurels prennent la forme d'une table sur l'âge du bénéficiaire et l'ancienneté du contrat. Les rachats conjoncturels (ou dynamiques) sont représentés par une fonction de réaction dépendant du taux servi et du taux cible.

2. Provision pour Participation au(x) Bénéfice(s)

Pour des paramètres $\alpha < \beta < \gamma < \delta$ et $RC_{min} < RC_{max}$ à fixer, cette fonction peut s'écrire comme suit :

$$q(r_{cib}, r_{rev}) = \begin{cases} RC_{max} & \text{si } r_{rev} - r_{cib} \leq \alpha \\ RC_{max} \frac{r_{rev} - r_{cib} - \beta}{\alpha - \beta} & \text{si } \alpha < r_{rev} - r_{cib} \leq \beta \\ 0 & \text{si } \beta < r_{rev} - r_{cib} \leq \gamma \\ RC_{min} \frac{r_{rev} - r_{cib} - \gamma}{\delta - \gamma} & \text{si } \gamma < r_{rev} - r_{cib} \leq \delta \\ RC_{min} & \text{sinon.} \end{cases} \quad (2.5)$$

- Priorités des sorties - L'ordre des sorties est supposé être le suivant : les rachats totaux interviennent avant les décès puis enfin les rachats partiels ont lieu.

$$q^{tot}(t) = 1 - \frac{l_{x,anc}^{tot}}{l_{x-1,anc-1}^{tot}} + q^{tot,dyn}(r_{cib}(t), r_{rev}(t)).$$

$$q^{dc}(t) = \left(1 - q^{tot}(t)\right) \left(1 - \frac{l_{x,gen}^{dc}}{l_{x-1,gen}^{dc}}\right).$$

$$q^{part}(t) = \left(1 - q^{tot}(t)\right) \frac{l_{x,gen}^{dc}}{l_{x-1,gen}^{dc}} q_{x,anc}^{part} + q^{part,dyn}(r_{cib}(t), r_{rev}(t)).$$

Finalement les flux liés aux rachats, aux décès et les flux à l'échéance sont respectivement :

$$pm(t-1) \left(q^{tot}(t) + q^{part}(t)\right) \mathbb{1}_{t \leq T}.$$

$$pm(t-1) q^{dc}(t) \mathbb{1}_{t \leq T}.$$

$$pm(t-1) \mathbb{1}_{t > T}.$$

2.2 Les Calculs

Hypothèses diverses :

- Le pas de discrétisation est annuel ;
- Les flux financiers interviennent en milieu d'année ;
- L'épargne est modélisée par la PM^3 .

Scénarii issus du GSE :

Avant de commencer, un ensemble de scénarios sous la mesure dite «risque-neutre» est créé par un GSE. Ces scénarios nous donnent différentes trajectoires possibles pour les actifs et les taux d'actualisation. Ils sont essentiels pour modéliser les incertitudes et évaluer les risques financiers auxquels l'assureur pourrait être confronté.

Première évaluation en norme S1 :

Nous évaluons d'abord les flux de trésorerie entrants et sortants sans prendre en compte les éventuelles revalorisations. Un premier résultat financier est calculé et projeté en suivant la politique ALM. Ensuite, nous déterminons les niveaux de provisions techniques.

Application des règles de partage du bénéfice :

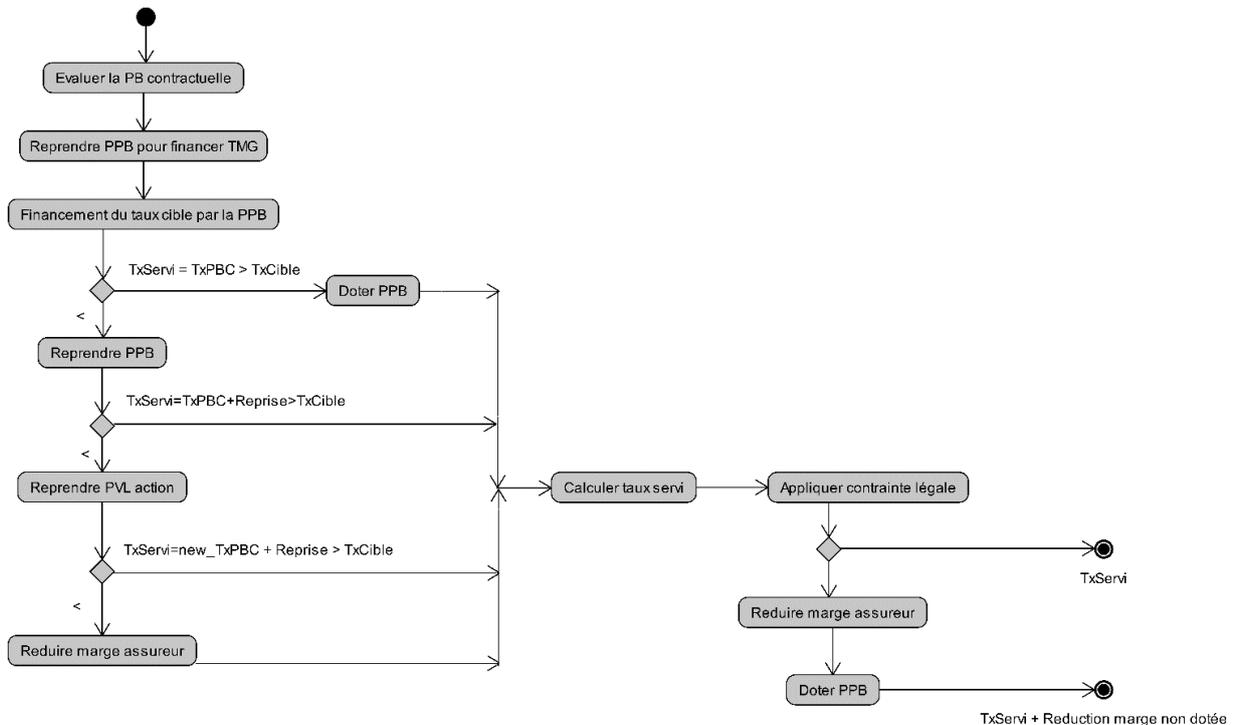


FIGURE 2.1 : Algorithme de revalorisation

Selon la revalorisation prévue et la marge cible de l'assureur, nous ajustons la richesse disponible via la PPB et les PMVL⁴ selon les règles de l'algorithme présenté figure 2.1. Cela assure une répartition équitable des bénéfices entre assureur et assurés.

Revalorisations :

Les contrats d'assurance sont ensuite revalorisés et les bénéfices sont distribués. Nous réajustons les actifs en fonction de l'allocation cible par des investissements stratégiques et des cessions d'actifs. Les flux financiers sont différenciés entre ceux qui sont discrétionnaires et ceux qui sont garantis.

La part garantie des revalorisation des prestations est définie comme :

$$rev_{prest}(t) = prest(t) \left[(1 + r_{min}(t))^{1/2} - 1 \right], \text{ et}$$

$$r_{min}(t) = \max(r_{tech}, r_{tmg} \mathbb{1}_{t \leq T_{tmg}}).$$

Le taux r_{min} est défini comme le meilleur taux entre le taux technique et le taux minimum garanti lorsque le terme de ce dernier n'est pas expiré.

Itérations sur la double-somme de l'estimateur Monte Carlo :

Ce processus est répété pour chaque scénario généré par le GSE, ainsi que pour différentes durées de projection.

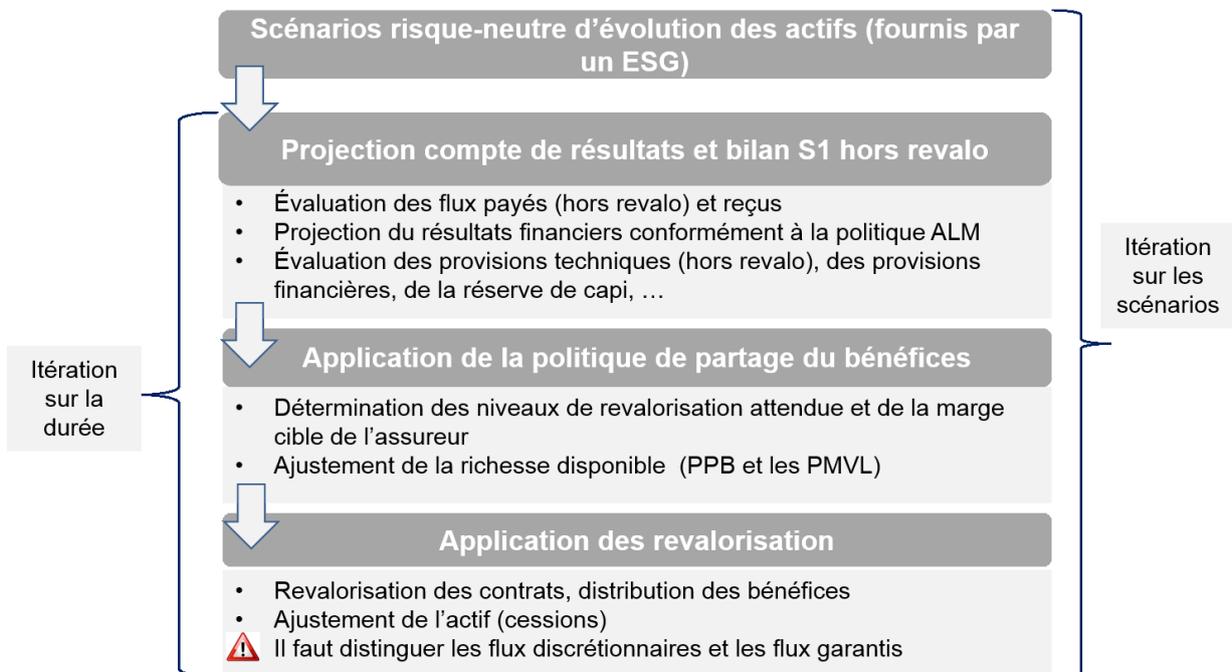


FIGURE 2.2 : Logique générale d'évaluation (selon LAURENT et al. (2016))

4. Plus-ou-Moins-Values Latentes

2.3 Les résultats en sortie

L'outil calcule les flux constituant le BEL. Ceci est fait pour les différents chocs de la formule standard de Solvabilité II. Il est donc possible de calculer un SCR pour les sous-modules souscription vie et risque de marché. Une séparation entre FDB⁵ et BEG⁶ est également proposé.

SCR

En reprenant la formule présentée en section 3.2 du chapitre 1, nous n'avons qu'à calculer les différents SCR de chaque sous-module correspondant à chaque choc. Concrètement, pour chaque choc, il s'agit de quantifier la différence en fonds propres économiques engrangée par l'application du choc en question.

$$\text{SCR}_{\text{choc}} = (\text{FP}_{\text{initiaux}} - \text{FP}_{\text{choqués}})_+$$

L'agrégation de ces chocs, sous-modules et modules est décrite à la section sus-mentionnée.

BEG

Le BEG correspond à la valeur actualisée des garanties offertes aux assurés dans le cadre d'un contrat d'assurance vie en épargne euro. Ces garanties incluent, par exemple, les taux d'intérêt garantis sur le capital ou les prestations minimales garanties en cas de décès ou à l'échéance du contrat. Le calcul du BEG repose sur des hypothèses réalistes et objectives, telles que les probabilités de mortalité, les taux de rachat, et les scénarios économiques futurs. Le BEG reflète ainsi la charge future associée à ces garanties, en tenant compte des engagements actuels et futurs de l'assureur.

FDB

Les FDB représentent les prestations futures discrétionnaires que l'assureur peut attribuer aux assurés dans le cadre d'un contrat d'assurance vie en épargne euro. Ces prestations incluent principalement la participation aux bénéfices discrétionnaire, qui dépend de la performance financière des actifs sous-jacents et des choix de gestion de l'assureur. Contrairement aux garanties représentées par le BEG, les FDB ne sont pas des engagements fermes, mais des engagements potentiels qui restent à la discrétion de l'assureur. Cette composante permet de compléter l'estimation globale des engagements de l'assureur, contribuant ainsi au calcul du BEL.

5. *Future Discretionary Benefits*

6. *Best Estimate Garanti*

En bref...

- Le **générateur de scénarios économiques (GSE)** est essentiel pour simuler les rendements futurs des actifs dans un cadre stochastique et évaluer la valeur temporelle des options et garanties (TVOG).
- Les modèles de **Black-Scholes-Merton** (pour les actions) et de **Hull & White** (pour les taux d'intérêt) sont utilisés pour générer les trajectoires des prix des actifs et des taux à partir de mouvements browniens corrélés.
- La décomposition de **Cholesky** permet de générer des mouvements browniens corrélés afin de construire les trajectoires des différents actifs simulés.
- Le modèle **ALM** (gestion actif-passif) prend en compte les actifs, les passifs, les paramètres économiques et les contraintes réglementaires pour projeter les flux financiers futurs.
- Les entrées clés du modèle incluent : la composition du portefeuille d'actifs (actions, obligations, immobilier, trésorerie), les engagements d'assurance (modélisés par des *Model Points*), l'allocation d'actifs cible, les taux de participation aux bénéfices et les chocs réglementaires de Solvabilité II.
- Les résultats permettent d'estimer les engagements d'assurance (BEL) et d'évaluer les impacts des chocs réglementaires sur la solvabilité de l'assureur.

Chapitre 3

Création d'une base de données d'entraînement

1 Une vision paramétrique des inputs

1.1 GSE

L'ensemble des paramètres des modèles de diffusion pour les trajectoires d'action, d'immobilier et de taux ainsi que les corrélations entre ces dernières sert d'input au modèle. En effet, puisque nous utilisons des équations de diffusion stochastiques, nul besoin de calibrer sur des données de marché lors de la génération d'environnements synthétiques.

1.2 Rachats

Ici nous faisons un choix de paramétrisation arbitraire de loi décrivant les rachats structurels. Les rachats augmentent linéairement avec l'âge entre un taux minimum vers un taux maximum, avec des sauts à certains seuils d'ancienneté (8 ans) et d'âge (60 ans).

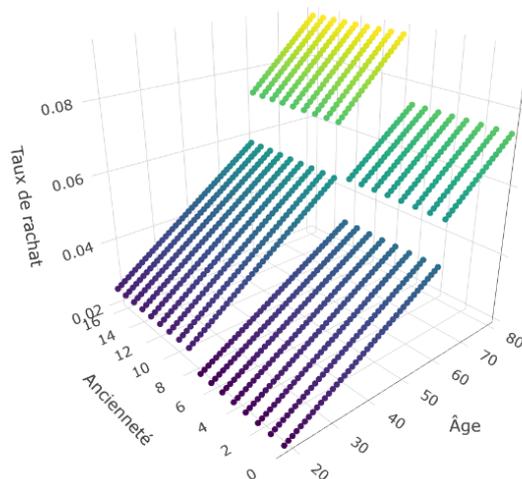


FIGURE 3.1 : Représentation 3D d'un exemple de taux de rachat partiel structurel

Les rachats dynamiques quant à eux utilisent les paramètres présentés dans l'équation 2.5.

1.3 Mortalité

Afin de permettre au modèle de capter les dérives de mortalité à l'avenir ou de modéliser une mortalité non standard, nous choisissons de rendre la mortalité paramétrique également. Nous utilisons une hypothèse de simplification qui considère la mortalité comme suivant une relation de type KANNISTO (1992) et non une table empirique. Nous établissons une régression linéaire affine sur les logit des q_x . Nous distinguons cependant la table masculine et féminine.

$$\text{logit}(q_x^h) = a_h + b_h \cdot x$$

$$\text{logit}(q_x^f) = a_f + b_f \cdot x$$

1.4 Model Points

Comme discuté précédemment, une dizaine de *model points* semble suffisant pour obtenir une précision acceptable. Un choix de 5 *model points* masculins et 5 *model points* féminins est fait à la base. Chaque individu parangon est caractérisé par un âge et une ancienneté dans le portefeuille, un taux de chargement sur encours, une PM et un taux technique que l'on suppose égal au TMG ici.

1.5 Autres paramètres

Actif :

Une valorisation est donnée à l'ensemble des actions détenues. Les obligations *corporate* et souveraines y voient également associés un taux de coupon et un *rating*. L'immobilier est caractérisé par sa valeur ainsi que le taux de loyer. La trésorerie est vue comme un simple montant. Un taux de chargement financier global ainsi qu'un taux de PB est renseigné.

Taux légaux :

Les taux de prélèvements sociaux, d'imposition ainsi que les parts de résultats financier et technique à intégrer dans la PB réglementaire sont pris en compte et modifiables.

2 La génération de paramètres synthétiques

2.1 Plages de valeurs admissibles, observations empiriques et bayésiannisme

Prenons pour exemple les corrélations entre les trajectoires action, immobilier et de taux. Nous savons qu'une corrélation est une quantité entre -1 et 1. Nous savons aussi empiriquement que les trajectoires de taux sont légèrement corrélées positivement avec celles d'action et d'immobilier. Cependant les trajectoires action et immobilier sont corrélées plutôt négativement. Générer une valeur admissible de $\rho_{a,i}$ pourrait se faire par une loi a priori non informative sur l'intervalle $[-1, 1]$.

$$\rho_{a,i}^{\hat{}} \sim 2 \cdot \mathcal{U} - 1.$$

Seulement, l'on sait que la corrélation est plutôt négative et que l'on se trouve autour de -0.5 en moyenne. La vision bayésienne nous encourage à pondérer plus fortement les plages de valeurs observées empiriquement et la mise à jour d'un a priori uniforme pour un paramètre nous amène à utiliser une loi a posteriori Beta.

Exemple :

Considérons une pièce pour laquelle nous voulons déterminer la probabilité p de faire face (resp. pile). A priori, sans information, nous modélisons ce paramètre par une loi uniforme sur l'intervalle $[0, 1]$, soit $p \sim \mathcal{U}(0, 1)$. Supposons maintenant que nous effectuons n lancers de la pièce, obtenant k fois face (resp. pile). Selon le théorème de Bayes, la loi a posteriori de p étant donné ces observations est proportionnelle à la vraisemblance multipliée par la loi a priori :

$$P(p | k, n) \propto P(k | p, n) \cdot P(p).$$

La vraisemblance $P(k | p, n)$ suit une distribution binomiale, donnée par $P(k | p, n) = \binom{n}{k} p^k (1-p)^{n-k}$. En substituant cette expression dans l'équation précédente, et sachant que la loi a priori est uniforme (constante), nous obtenons :

$$P(p | k, n) \propto p^k (1-p)^{n-k}.$$

Cela correspond à une distribution Beta($k + 1, n - k + 1$). Ainsi, après avoir observé les résultats des lancers, la probabilité a posteriori de faire face est modélisée par une loi Beta qui s'affine avec chaque nouvelle observation. Nous utilisons ceci afin d'intégrer un «jugement d'expert» dans le choix de nos paramètres à générer.

3 Une analyse descriptive de la base obtenue

3.1 Les covariables économiques

Black-Scholes :

Pour le modèle de diffusion des actions et de la part immobilière de l'actif, nous avons besoin de 3 paramètres. Ils sont calibrés sur les données de marché pour reproduire fidèlement les prix observés. Le taux sans risque est dérivé d'obligations d'État à long terme (OAT 10 ans par exemple).

- r représente le taux sans risque sur le marché qui correspond à la dérive en vision risque-neutre ;
- σ_{action} représente la volatilité incorporée aux trajectoires actions ;
- σ_{immo} représente la volatilité incorporée aux trajectoires de l'immobilier.

Hull-White – Vasicek :

Pour la reproduction des mouvements de taux, nous calibrons 3 paramètres également, de manière à reproduire au mieux la courbe des taux initiale que l'on projette ensuite par le modèle de Hull & White qui dans notre paramétrisation simplifiée est assimilable à un modèle de Vasicek.

- α est le paramètre qui contrôle la convergence vers le taux ultime espéré du modèle ;
- θ_{α} est le rapport $\frac{\theta}{\alpha}$ du modèle et correspond à un taux ultime ;
- σ_{taux} est le paramètre de volatilité s'y afférant.

Cholesky :

Afin d'obtenir les différents mouvements browniens corrélés qui structurent les différents scénarii de diffusion d'indices, nous utilisons une matrice de corrélation qui devient un jeu de paramètres de notre modèle.

- ρ_{a_i} est le coefficient de corrélation entre les trajectoires action et immobilier ;
- ρ_{a_t} celui entre les trajectoires d'action et de taux ;
- ρ_{i_t} celui entre l'immobilier et les taux.

Les valeurs prises par ces différents paramètres ainsi que leur corrélation avec le BEL sont répertoriées dans le tableau 3.1 ci-dessous. Le taux sans risque r et le taux ultime $\frac{\theta}{\alpha}$ semblent – ainsi que plus étonnement la corrélation des trajectoires de taux et de l'immobilier – corrélés au BEL.

	min	max	médiane	moyenne	écart-type	corrélacion	p-valeur (cor.test)
ρ_{a_i}	-0,774	-0,016	-0,503	-0,493	0,131	0,041	0,473
ρ_{a_t}	-0,050	0,282	0,096	0,097	0,067	0,078	0,166
ρ_{i_t}	-0,090	0,216	0,046	0,049	0,055	-0,118	0,037
r	0,002	0,086	0,027	0,031	0,018	0,139	0,013
σ_{action}	0,068	0,252	0,149	0,149	0,034	0,075	0,182
σ_{immo}	0,042	0,210	0,098	0,099	0,027	0,024	0,677
σ_{taux}	0,00004	0,048	0,007	0,009	0,009	-0,045	0,423
α	0,003	6,160	0,597	0,860	0,953	-0,081	0,153
θ_{α}	0,010	0,104	0,043	0,044	0,016	-0,156	0,006

TABLE 3.1 : Description sommaire des variables de contrôle relatives au GSE de notre base de données

3.2 Les portefeuilles modélisés

Actif :

Nous modélisons différentes allocations d'actifs pour chaque portefeuille fictif à l'origine d'un calcul de BEL. Nous gardons cependant une forte composante obligataire, typique des poches d'actifs mis en face de produits garantis tels que les fonds en euro. Les valeurs des actifs sont modélisées afin d'obtenir un portefeuille d'actif d'une valeur totale en espérance autour de 100 k€ avec une allocation à 80 % obligataire, avec 7,5 % d'actions, autant d'immobilier et 5 % de trésorerie. Les allocations présentes dans notre base de données sont synthétisées par la figure 3.2 ci-dessous.

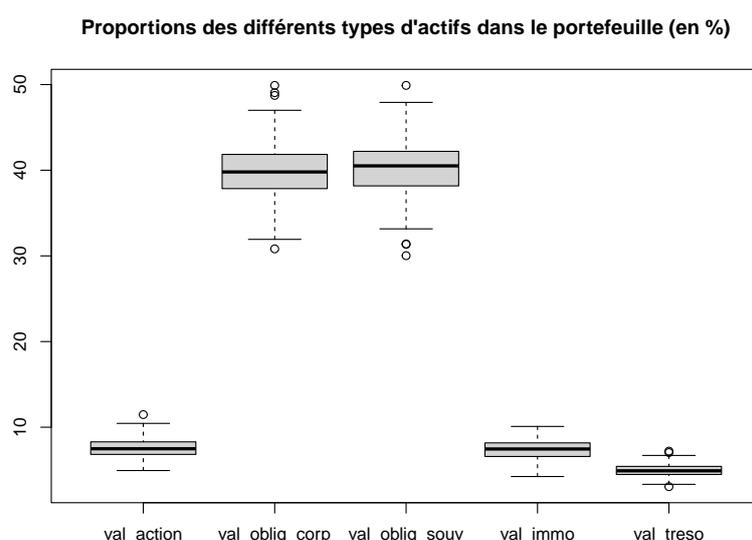


FIGURE 3.2 : Boîtes à moustaches des allocations d'actif dans la base de données

Nous retrouvons en moyenne les proportions souhaitées initialement comme le montre la table 3.2 suivante.

Actions	Obligations <i>corporate</i>	Obligations souveraines	Immobilier	Trésorerie
7,6 %	39,8 %	40,2 %	7,4 %	5 %

TABLE 3.2 : Allocations d'actifs dans la base de données

Passif :

Pour nos différents *model points* de passif, nos situations initiales synthétiques correspondent à un assuré median ayant 50 ans, depuis 10 ans dans le portefeuille, avec une PM de 50 000€ un TMG moyen de quasiment 2 % et un chargement sur encours d'un demi-pourcent. La répartition plus en détail des valeurs prises de ces variables dans notre base de données est consignée dans le tableau 3.3 suivant.

Variable	Min.	1er Quart.	Médiane	Moyenne	Max.
TMG moyen	0,005603	0,015057	0,018707	0,019097	0,036423
Âge moyen	38,50	47,85	50,00	50,19	62,00
Ancienneté moyenne	5,50	9,10	10,20	10,33	15,40
Chargement sur encours moyen	0,003164	0,004579	0,004977	0,005041	0,007564
Provision Mathématique totale	17805	39813	50008	50473	98752

TABLE 3.3 : Résumé des différents *model points* de notre base de données

3.3 Normalisation par la PM

La normalisation par la PM dans le calcul du BEL est essentielle pour rendre les résultats comparables entre les différents contrats d'assurance. En effet, les provisions mathématiques varient considérablement selon la taille, la durée ou les caractéristiques des contrats. La normalisation permet d'éliminer cet effet de taille et de se concentrer sur les différences relatives de BEL, facilitant ainsi l'analyse comparative et la mise en évidence de tendances ou de facteurs explicatifs. Elle assure également une meilleure stabilité des modèles et évite les biais liés à des valeurs extrêmes. Elle permet également de retrouver une grandeur étudiée dans l'article de BONNIN et al. (2014) au point 4, le $\rho(0, T)$.

$$\rho(0, T) = \frac{\text{BEL}(0, T)}{\text{PM}(0)}$$

En bref...

- **Modélisation paramétrique** : L'utilisation d'une modélisation stochastique paramétrique permet de générer des environnements synthétiques sans calibration sur des données de marché, simplifiant ainsi la création de la base de données.
- **Rachats et mortalité** : Les taux de rachats structurels sont paramétrés avec des seuils arbitraires en fonction de l'âge et de l'ancienneté. La mortalité est modélisée par une régression logistique sur les taux q_x selon l'âge, avec des distinctions entre hommes et femmes.
- **Synthèse des *model points*** : Une dizaine de *model points* permet une précision satisfaisante, avec une distinction selon l'âge, l'ancienneté, le taux de chargement et les provisions mathématiques.
- **Corrélations économiques** : La génération des paramètres économiques, comme les corrélations entre actions, immobilier et taux, s'appuie sur des lois bayésiennes pour intégrer des observations empiriques et expertises.
- **Modèles de diffusion** : Les modèles Black-Scholes pour les actions et Hull-White – Vasicek pour les taux sont utilisés, chacun avec leurs paramètres spécifiques, afin de simuler les trajectoires économiques.
- **Normalisation par la PM** : La normalisation des résultats par la Provision Mathématique (PM) permet de rendre les données comparables et d'éliminer les effets de taille entre différents contrats.

DEUXIÈME PARTIE

Proxy de calcul des engagements en *machine learning*

Chapitre 4

Mise en place d'un modèle de *machine learning* «distributionnel»

1 Les modèles linéaires

1.1 La définition du modèle *Ordinary Least Squares*

Lorsque nous avons un jeu de données qui correspond à une matrice de paramètres \mathbf{X} et un vecteur cible \mathbf{y} que l'on cherche à modéliser, le premier lien à tester de par sa simplicité est le lien linéaire. Comme point de départ de modélisation, cherchons la solution à l'équation :

$$\hat{\beta} = \arg \min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|.$$

Lorsque la matrice \mathbf{X} permet une solution, elle est la suivante :

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$$

Cet estimateur est efficient voire optimal sous certaines conditions, parmi lesquelles on retrouve l'indépendance des observations, l'homoscédasticité ou encore la normalité des résidus. Cherchons à quantifier la sensibilité du modèle à l'hypothèse d'homoscédasticité qui est la moins respectée dans notre cas. Nous avons la chance d'avoir accès à la distribution ayant engendré le BEL et ainsi pouvoir quantifier la variance en chaque observation.

1.2 Prise en compte de l'hétéroscédasticité avec les *Weighted Least Squares*

Prenons pour exemple un jeu de données fictif, construit pour répondre à toutes les hypothèses sauf celle d'égalité des variances. Nous pouvons observer une nette amélioration de la robustesse du modèle lorsque l'on pondère les observations par un paramètre `weights` inversement proportionnel aux variances attendues. L'estimateur devient, avec idéalement $W = M^{-1}$ et M la matrice de variance-covariance des observations :

$$\hat{\beta} = (\mathbf{X}^T W \mathbf{X})^{-1} \mathbf{X}^T W \mathbf{y}.$$

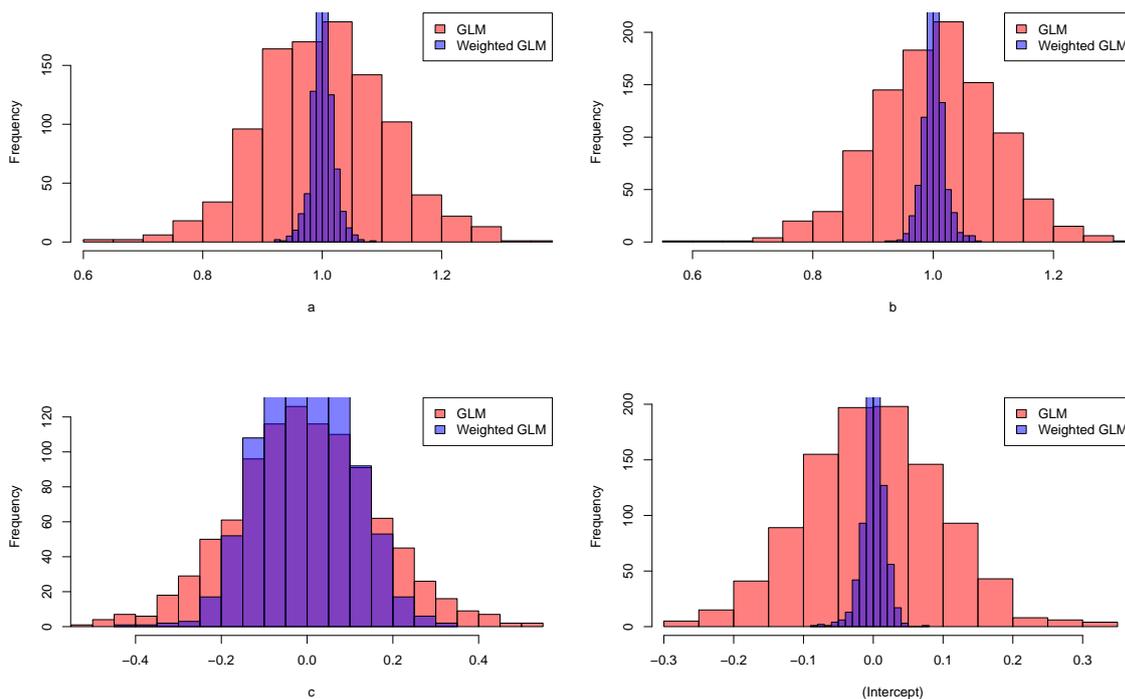


FIGURE 4.1 : Histogramme des coefficients inférés avec ou sans correction de l'hétéroscédasticité pour $Y|\{A = a, B = b, C = c\} \sim \mathcal{N}(a + b, c^2)$ avec $A, B, C \sim \mathcal{N}(0, 1)$

La figure 4.1 nous montre des coefficients bien plus rapprochés de leur valeur théorique. Le critère d'Akaike (AIC) nous donne un modèle beaucoup plus «vraisemblable» lorsque l'on utilise l'information de la dispersion inégale des observations :

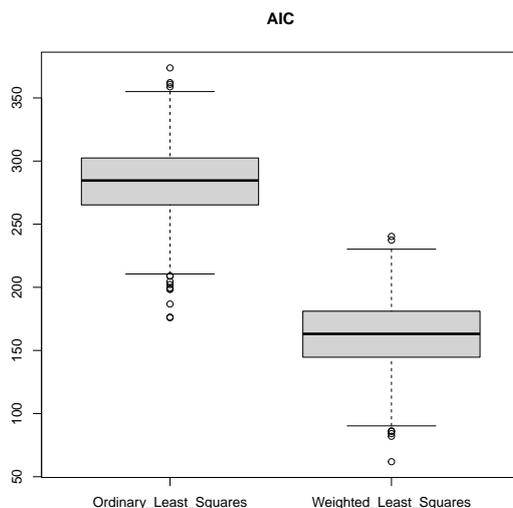


FIGURE 4.2 : Boîtes à moustaches des AIC avec et sans pondération

2 La régression étendue aux facteurs de forme : le modèle **GAMLSS**

2.1 La définition du modèle

Le modèle **GAMLSS**¹ étend le principe d'un GLM² ou d'un GAM³ à d'autres paramètres que la simple moyenne. Des paramètres d'échelle et de forme sont pris en compte. Par exemple, en notant μ le paramètre de position σ le paramètre d'échelle et (ν, τ) des paramètres de forme, on peut écrire le système d'équations suivant :

$$\begin{cases} g_1(\mu) = \mathbf{X}_1\beta_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1} \cdot \gamma_{j1}, \\ g_2(\sigma) = \mathbf{X}_2\beta_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2} \cdot \gamma_{j2}, \\ g_3(\nu) = \mathbf{X}_3\beta_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3} \cdot \gamma_{j3}, \\ g_4(\tau) = \mathbf{X}_4\beta_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4} \cdot \gamma_{j4}. \end{cases} \quad (4.1)$$

Une telle paramétrisation de distribution en dimension 4 offre une flexibilité permettant de régler l'asymétrie et le *kurtosis*⁴, et donc de gérer différents types de queues, tout en restant relativement simple à paramétrer.

2.2 La modélisation de distributions

La loi normale

Nous allons présenter la méthode de résolution proposée par défaut dans le modèle **GAMLSS** présenté dans la section précédente. La méthode est renseignée par le paramètre `method = RS()` de la fonction `gamlss(...)` sur **R** et implémente le modèle additif semi-paramétrique hétéroscédastique de RIGBY et STASIPOULOS (1996) qui reprend en grande partie les travaux de AITKIN (1987).

Le cadre de résolution simplifié qui reprend les 2 premières lignes du système d'équations 4.1 de la section précédente pour μ et σ nous donne la formule suivante pour la log-vraisemblance :

$$l = -\frac{1}{2} \left[\sum_{i=1}^n \frac{(y_i - \mu_i)^2}{\sigma_i^2} + \ln(2\pi\sigma_i^2) \right]. \quad (4.2)$$

On retrouve en partie – pour une dispersion fixée – l'enjeu de minimisation des écarts au carrés pondérés par l'inverse de leur dispersion. Ce qui est la démarche intuitive que l'on avait déjà pour corriger les problèmes de variances inégales dans le modèle *Weighted Least Squares* vu précédemment.

Modéliser asymétrie et *kurtosis* : la loi Sinh-Arcsinh

La distribution **Sinh-Arcsinh** (SHASH) est une extension flexible de la distribution normale, introduite pour modéliser des données univariées continues présentant des asymétries ou des queues plus lourdes ou plus légères que celles de la distribution normale classique. La transformation `sinh-arcsinh`, appliquée à une variable aléatoire normale standard, génère une famille de distributions paramétrée

1. *General Additive Model for Location Scale and Shape*
 2. *General Linear Model*
 3. *Generalized Additive Model*
 4. aplatissement

par quatre paramètres : ξ (position), η (échelle), ϵ (asymétrie), et δ (poids des queues). La distribution normale est un cas particulier de cette famille lorsque $\epsilon = 0$ et $\delta = 1$.

Le paramètre ϵ contrôle le degré d'asymétrie, avec une valeur positive induisant une asymétrie à droite et une valeur négative une asymétrie à gauche. Le paramètre δ , quant à lui, influence la lourdeur des queues : $\delta < 1$ entraîne des queues plus lourdes, tandis que $\delta > 1$ conduit à des queues plus légères que celles de la normale.

Dans notre cas, la distribution SHASH peut être particulièrement utile pour modéliser la distribution des passifs dans les produits d'épargne-vie en euros, où la distribution des sinistres peut présenter une potentielle asymétrie et des queues semi-lourdes, violant l'hypothèse de normalité. Cette flexibilité permet une estimation plus précise des potentiels engagements.

Formalisation

La distribution SHASH est définie par la transformation d'une variable aléatoire Z suivant une loi normale standard $\mathcal{N}(0, 1)$. La transformation est donnée par :

$$X = \mu + \sigma \cdot \sinh \left(\frac{\sinh^{-1}(Z) + \nu}{\tau} \right) \quad (4.3)$$

où :

- $\mu \in \mathbb{R}$ est un paramètre de position,
- $\sigma > 0$ est un paramètre d'échelle,
- $\nu \in \mathbb{R}$ contrôle l'asymétrie,
- $\tau > 0$ contrôle le poids des queues.

Intervalles de confiance

La transformation étant monotone et strictement croissante, si on pose $S_{\mu, \sigma, \nu, \tau}(t) = \mu + \sigma \cdot \sinh \left(\frac{\sinh^{-1}(t) + \nu}{\tau} \right)$, alors un intervalle à 95 % peut être construit immédiatement par :

$$IC_{0.95} = \left[S_{\mu, \sigma, \nu, \tau}(\Phi^{-1}(0.025)), S_{\mu, \sigma, \nu, \tau}(\Phi^{-1}(0.975)) \right]$$

3 Les réseaux de neurones et l'apprentissage distributionnel

3.1 Une formalisation des réseaux de neurones

Les réseaux de neurones ont le double avantage sur les modèles de régressions classiques de pouvoir opérer sur un vecteur de valeurs cibles et d'intégrer des transformations autres que affines. Ceci offre une flexibilité à double tranchant puisque le choix de l'architecture du modèle se fait parmi un panel très vaste et les méthodes d'optimisation qui en résultent sont parfois capricieuses à converger sur un espace des paramètres et hyper-paramètres difficile à explorer.

Les modèles séquentiels :

Présentons l'architecture la plus simple et la plus intuitive, à savoir l'enchaînement de couches élémentaires qui formeront un réseau. Chaque couche fonctionne comme une abstraction mathématique d'un neurone, d'où le nom de réseau de neurones. La réception par l'intermédiaire des synapses de l'information est agrégée et retransmise le long de l'axone. Ici mathématiquement, l'on pondère les entrées, puis applique une fonction d'activation.

Si l'on enchaîne d couches on obtient la formule suivante :

$$\hat{Y} = \phi_d(\dots(\phi_2(\phi_1(X \cdot M_1) \cdot M_2) \dots) \cdot M_d). \quad (4.4)$$

$\phi_1, \phi_2, \dots, \phi_d$ sont les fonctions d'activation qui s'appliquent élément par élément et M_1, M_2, \dots, M_d des matrices de dimension compatibles, afin de réduire la dimension en partant de X et en arrivant à la dimension du vecteur Y .

3.2 Rôle des fonctions d'activation et présentation de la fonction Tanhshrink

Dans un réseau de neurones, les fonctions d'activation jouent un rôle crucial dans l'apprentissage et la modélisation de relations non linéaires. Après que les données aient traversé une couche linéaire (généralement sous la forme d'une transformation matricielle du type $Wx + b$), elles sont souvent passées à travers une fonction d'activation. Celle-ci permet d'introduire des non-linéarités dans le modèle, rendant le réseau de neurones capable de capturer des relations complexes dans les données. Sans ces non-linéarités, le réseau serait équivalent à une simple régression linéaire, même avec plusieurs couches.

Quelques fonctions d'activation courantes

Il existe plusieurs fonctions d'activation, chacune avec des propriétés spécifiques :

- **ReLU (Rectified Linear Unit)** : La fonction ReLU est définie par $f(x) = \max(0, x)$. Elle est largement utilisée en raison de sa simplicité et de son efficacité dans les réseaux profonds.
- **Sigmoïde** : La fonction sigmoïde est définie par $f(x) = \frac{1}{1+e^{-x}}$. Elle est souvent utilisée dans les tâches de classification binaire.
- **Tanh (Tangente hyperbolique)** : La fonction Tanh est définie par $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$, et elle contraint les valeurs de sortie entre -1 et 1.

Chaque fonction d'activation présente des avantages et des inconvénients en fonction des caractéristiques de la tâche. Par exemple, la fonction ReLU permet d'éviter l'effet de saturation que l'on rencontre avec la fonction sigmoïde ou la fonction tanh, mais peut entraîner des neurones morts (Dead Neurons) lorsque les valeurs sont constamment négatives. (DOUGLAS & YU, 2018)

Présentation de la fonction Tanhshrink

La fonction **Tanhshrink**, moins connue et moins fréquemment utilisée, est définie par la formule suivante :

$$f(x) = x - \tanh(x)$$

Cette fonction retourne la différence entre l'entrée x et la valeur renvoyée par la fonction tangente hyperbolique $\tanh(x)$. Contrairement à la fonction **Tanh** classique, qui ramène les valeurs entre -1 et 1, la fonction **Tanhshrink** introduit une correction en soustrayant cette approximation non linéaire du signal d'entrée.

Propriétés de la fonction Tanhshrink :

- **Linéarité à l'infini** : Pour des valeurs absolues de x très élevées (i.e., $x \rightarrow +\infty$ ou $x \rightarrow -\infty$), $\tanh(x) \approx 1$ ou $\tanh(x) \approx -1$. La fonction **Tanhshrink** se comporte alors comme une fonction asymptotiquement affine.
- **Réduction des valeurs petites** : Pour des valeurs proches de zéro, $\tanh(x) \approx x$, et la fonction **Tanhshrink** réduit ces petites valeurs en les contractant vers zéro. Cela peut être utile pour limiter l'effet des petits bruits dans les données.

Cette fonction est souvent utilisée dans des contextes où l'on souhaite moduler ou lisser l'effet d'une activation non linéaire classique (comme **Tanh**) tout en conservant une composante linéaire, afin de bénéficier à la fois de la non-linéarité pour les grandes valeurs de x et d'une simplification pour les petites valeurs.

3.3 Rôle des optimiseurs et présentation de l'optimiseur AdamW

Dans les réseaux de neurones, les optimiseurs jouent un rôle crucial dans l'ajustement des paramètres du modèle afin de minimiser une fonction de coût. À chaque étape d'apprentissage, un optimiseur met à jour les poids du modèle en fonction de la direction et de la magnitude du gradient de la fonction de perte. Le but est de trouver les paramètres qui minimisent le plus possible l'erreur du modèle sur les données d'entraînement, tout en garantissant une bonne généralisation sur les données non vues.

Les différents types d'optimiseurs

Il existe plusieurs types d'optimiseurs utilisés dans les réseaux de neurones. Voici quelques exemples courants :

- **Descente de gradient stochastique (SGD)** : La méthode de base, qui met à jour les poids dans la direction opposée au gradient de la fonction de coût. Bien qu'efficace, la SGD peut parfois être lente à converger et sensible à l'échelle des gradients.
- **SGD avec *momentum*** : Cet optimiseur améliore la SGD en ajoutant un terme de «*momentum*», qui accumule une fraction des gradients passés pour atténuer les oscillations et accélérer la convergence.

- **RMSprop (*Root Mean Square Propagation*)** : Cet optimiseur ajuste dynamiquement la vitesse d'apprentissage pour chaque paramètre en fonction de la moyenne mobile des carrés des gradients passés, réduisant ainsi les oscillations et accélérant la convergence dans les directions où les gradients sont plus faibles.
- **Adam (*Adaptive Moment Estimation*)** : L'optimiseur Adam combine les avantages de la SGD avec *momentum* et de RMSprop en adaptant le taux d'apprentissage pour chaque paramètre en fonction de la moyenne mobile des gradients et de leur carré. C'est l'un des optimiseurs les plus populaires en apprentissage profond.

Présentation de l'optimiseur AdamW (Loshchilov & Hutter, 2019)

L'optimiseur **AdamW** est une variante de l'optimiseur Adam, introduite pour corriger un problème spécifique lié à la régularisation L2 (ou pénalisation de poids), appelée communément *weight decay*. Dans l'optimiseur Adam classique, la régularisation L2 est appliquée de manière incorrecte en ajustant non seulement les poids mais aussi les moments d'apprentissage, ce qui peut nuire à la convergence optimale. L'optimiseur AdamW dissocie correctement la régularisation du poids des étapes de mise à jour du gradient. Cela améliore la performance générale du modèle, en particulier sur des réseaux profonds, en assurant que la pénalisation L2 agit uniquement sur les poids, et non sur les moments ou les gradients, ce qui est essentiel pour une meilleure généralisation.

Formule de mise à jour d'AdamW : AdamW suit le schéma de mise à jour suivant :

$$\theta_{t+1} = \theta_t - \eta \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon) - \eta \cdot \lambda \cdot \theta_t$$

Où :

- θ_t représente les poids du modèle à l'étape t .
- η est le taux d'apprentissage.
- \hat{m}_t est la moyenne mobile des gradients (terme de *momentum*).
- \hat{v}_t est la moyenne mobile des carrés des gradients (adaptation du taux d'apprentissage).
- ϵ est un petit terme ajouté pour éviter la division par zéro.
- λ est le facteur de pénalisation de poids (*weight decay*).

La différence clé par rapport à Adam réside dans l'ajout du terme $-\eta \cdot \lambda \cdot \theta_t$, qui assure une pénalisation correcte des poids.

Avantages de l'optimiseur AdamW :

- **Meilleure généralisation** : AdamW tend à améliorer la capacité de généralisation du modèle en appliquant correctement la régularisation L2 sur les poids, évitant parfois les sur-ajustements (*overfitting*).

- **Convergence plus rapide** : Grâce à l'adaptation des taux d'apprentissage pour chaque paramètre et au découplage de la régularisation, AdamW converge souvent plus rapidement et de manière plus stable que les variantes de SGD classiques.
- **Efficace sur de grands modèles** : AdamW est particulièrement performant sur les modèles comportant un grand nombre de paramètres, comme les réseaux de neurones profonds ou les architectures complexes (e.g., les *transformers*).

3.4 Régression étendue à une distribution paramétrique

Nous proposons ici un cadre d'étude qui s'éloigne de la tâche classique de régression d'une quantité d'intérêt en fonction de covariables potentiellement explicatives $Y = \phi(X)$. Dans notre cas, le BEL à partir des données et paramètres initiaux lors de son calcul. On propose de modéliser plutôt la **distribution conditionnelle aux variables explicatives qui a engendré les observations** de la variable aléatoire Y . En effet, au vu de nos données qui ne sont pas des observations uniques mais bien un vecteur de plusieurs (ici 200) scenarii pour chaque condition initiale, nous choisissons d'intégrer toute l'information apportée par ces observations sur la distribution latente plutôt que de la résumer par une statistique telle que la moyenne empirique qui nous donnerai le BEL.

Plus formellement on pose l'hypothèse suivante : $Y|X \sim \mathcal{D}_{\theta=\phi(X)}$ avec \mathcal{D} une distribution et θ sa paramétrisation. En d'autres termes, la distribution des flux d'engagements futurs issus des trajectoires aléatoires du GSE étant donné un environnement économique dépend fonctionnellement de ces mêmes hypothèses de départ au travers d'un lien que l'on cherche à mettre en lumière.

On cherche ϕ la fonction (qu'il faut ici voir comme un réseau de neurones entraîné) qui minimise la divergence de Kullback-Leibler entre les distributions issues de la paramétrisation inférée $\hat{\theta}$ (l'estimateur de maximum de vraisemblance dans la famille de loi considérée inféré sur les observations, ici des scenarii) et de celle prédite $\hat{\phi}(X)$ par notre modèle de *machine learning*.

$$\hat{\phi} = \operatorname{argmin}_{\phi \in \Phi(\text{NN})} D_{\text{KL}}(f_{\mathcal{D}_{\phi(X)}} \parallel f_{\mathcal{D}_{\hat{\theta}}}) \quad (4.5)$$

$\Phi(\text{NN})$ est ici l'espace des fonctions représentables par un réseau de neurone NN d'une architecture donnée. $f_{\mathcal{D}_{\hat{\theta}}}$ la densité associée à la distribution ajustée sur les observations de Y et $D_{\text{KL}}(\cdot \parallel \cdot)$ la divergence de Kullback-Leibler.

La divergence de Kullback-Leibler comme fonction de perte

Lorsque l'on souhaite comparer des distributions, il vient assez vite l'envie d'étudier des distances entre distributions. La notion de divergence de Kullback-Leibler tombe alors sous le sens. Nous allons tenter de nous convaincre de sa pertinence et son interprétabilité dans un cadre simplifié.

Définition formelle dans le cadre des distributions continues :

$$D_{\text{KL}}(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \left(\frac{p(x)}{q(x)} \right) dx \quad (4.6)$$

Cette divergence correspond à l'espérance du logarithme de la statistique du ratio de vraisemblance. Elle est aussi appelée gain d'information en ML en cela qu'elle renseigne par le calcul d'entropie le manque à gagner informationnel entre P et Q . De nombreuses interprétations existent, notamment bayésiennes. Nous allons nous concentrer sur le calcul de cette divergence pour la loi normale – car elle permet une formule fermée simple de sa divergence – afin de dériver une meilleure fonction objectif pour nos modèles. On retrouvera un lien avec le modèle *Weighted Least Squares*.

Divergence d'une paramétrisation gaussienne : Posons le cadre d'étude suivant.

$$\begin{aligned} P &: X \sim \mathcal{N}(\mu_1, \sigma_1^2) \\ Q &: X \sim \mathcal{N}(\mu_2, \sigma_2^2). \end{aligned}$$

En reprenant l'équation 4.6, on obtient :

$$\begin{aligned} D_{\text{KL}}(P \parallel Q) &= \mathbb{E}^P \left[\ln \left(\frac{\frac{1}{\sqrt{2\pi}\sigma_1} \cdot \exp\left(-\frac{1}{2}\left(\frac{X-\mu_1}{\sigma_1}\right)^2\right)}{\frac{1}{\sqrt{2\pi}\sigma_2} \cdot \exp\left(-\frac{1}{2}\left(\frac{X-\mu_2}{\sigma_2}\right)^2\right)} \right) \right] = \ln \frac{\sigma_2}{\sigma_1} + \frac{1}{2} \cdot \mathbb{E}^P \left[\left(\frac{X-\mu_2}{\sigma_2} \right)^2 - \left(\frac{X-\mu_1}{\sigma_1} \right)^2 \right] \\ &= \frac{1}{2} \left[\ln \left(\left(\frac{\sigma_2}{\sigma_1} \right)^2 \right) + \frac{\mathbb{E}^P \left[(X-\mu_2)^2 \right]}{\sigma_2^2} - 1 \right] = \frac{1}{2} \left[\ln \frac{\sigma_2^2}{\sigma_1^2} + \frac{\mathbb{E}^P \left[(X-\mu_1 + \mu_1 - \mu_2)^2 \right]}{\sigma_2^2} - 1 \right] \\ &= \frac{1}{2} \left[\ln \frac{\sigma_2^2}{\sigma_1^2} + \frac{\mathbb{V}^P \left[X-\mu_1 \right] + (\mu_1 - \mu_2)^2}{\sigma_2^2} - 1 \right] = \frac{1}{2} \left[\ln \frac{\sigma_2^2}{\sigma_1^2} + \frac{\sigma_1^2}{\sigma_2^2} + \frac{(\mu_1 - \mu_2)^2}{\sigma_2^2} - 1 \right] \end{aligned}$$

On remarque que $\lim_{\sigma_1 \rightarrow \sigma_2} D_{\text{KL}}(P \parallel Q) \propto \frac{(\mu_1 - \mu_2)^2}{\sigma_2^2}$ ce qui revient asymptotiquement à faire du *Weighted Least Squares* avec les pondérations $w_i = \frac{1}{\sigma_2^2}$.

3.5 Modéliser l'asymétrie et la kurticité : la loi hyperbolique généralisée

Dans le but de capter une potentielle asymétrie issue du déséquilibre induit par la présence d'options et garanties, nous cherchons une distribution plus flexible. La loi hyperbolique généralisée représente à l'instar de la loi **Sinh-Arcsinh** évoquée pour le modèle GAMLSS une famille de distributions flexible qui permet de modéliser asymétrie et aplatissement. Elle est caractérisée par la fonction de densité suivante :

$$f(x) = \frac{\left(\sqrt{\alpha^2 - \beta^2}/\delta\right)^\lambda}{\sqrt{2\pi}K_\lambda\left(\delta\sqrt{\alpha^2 - \beta^2}\right)} \times \frac{K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)}{\left(\sqrt{\delta^2 + (x-\mu)^2}/\alpha\right)^{1/2-\lambda}} e^{\beta(x-\mu)}. \quad (4.7)$$

$K_\nu(x)$ représente la fonction modifiée de Bessel définie comme tel :

$$K_\nu(z) = \frac{2^\nu \Gamma(\nu + 1/2)}{\sqrt{\pi}} z^\nu \int_0^{+\infty} \frac{\cos x}{(z^2 + x^2)^{\nu+1/2}} dx.$$

Pour ne citer qu’elles, la distribution t de Student, la distribution de Laplace, la distribution hyperbolique, la distribution normale-inverse gaussienne et la distribution variance-gamma font partie de cas particuliers de la loi hyperbolique généralisée.

Inférence : L’ajustement d’une telle loi souffre de la flexibilité de sa paramétrisation. Afin de ne pas subir un «faux» minimum local lors de la maximisation de la vraisemblance, le choix d’un point de départ cohérent dans l’espace des paramètres est crucial. L’on remarque en posant les deux premiers moments que la paramétrisation $\alpha = 1, \beta = 0, \delta = \hat{V}(X), \mu = \hat{E}(X), \lambda = -1/2$ nous donne des estimateurs de l’espérance et de la variance qui coïncident. Nous utiliserons ce point de départ pour notre processus d’optimisation.

3.6 Acceptabilité d’une hypothèse de normalité et perte d’information

L’idée même d’utiliser une famille de loi plus permissive que la distribution gaussienne provient du postulat initial que l’hypothèse de normalité est violée par la présence d’une légère asymétrie. En effet, seule la moitié des jeux d’observations passe le test de Shapiro-Wilk. Cependant, il est intéressant de s’interroger sur la sensibilité de nos résultats à ces hypothèses. Au niveau de l’estimateur final du BEL, il reste identique quelque soit la distribution ajustée. La moyenne empirique reste l’estimateur cohérent de l’espérance. Néanmoins nous souhaitons modéliser l’incertitude autour de cette statistique à l’aide d’un intervalle de confiance plus adapté. Dans la suite, nous allons présenter le calcul de cet intervalle de confiance inféré sur une distribution modélisant asymétrie et *kurtosis* telle que la *generalized hyperbolic* et le comparerons à celui obtenu en admettant l’hypothèse de normalité. Une première inspection visuelle figure 4.3 de la différence d’ajustement sur la distribution observée la plus asymétrique parmi nos données nous donne un indice sur le caractère a priori négligeable du gain d’information obtenu avec une paramétrisation plus flexible, d’autant plus au regard du coût supplémentaire en complexité engendré par une telle approche.

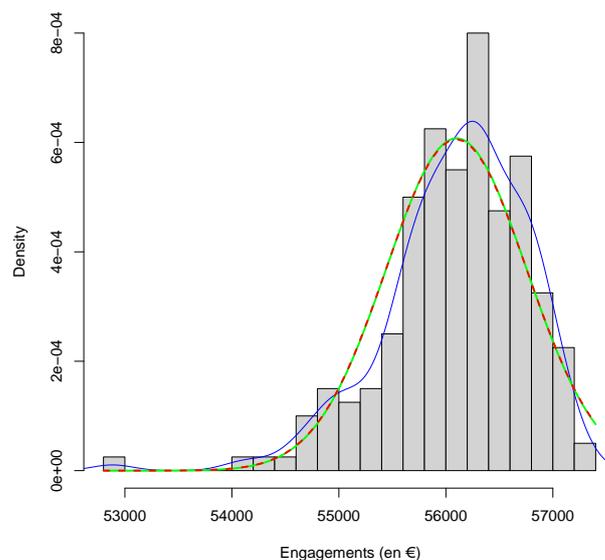


FIGURE 4.3 : Valeur actualisée des engagements futurs par simulation du GSE

La courbe bleue représente l'estimateur par noyau de la densité accompagnant l'histogramme des observations. La courbe verte est la densité ajustée gaussienne, et la courbe en pointillés rouge son alternative asymétrique, la distribution *generalized hyperbolic*. Il saute aux yeux que la différence est très subtile et probablement négligeable par la suite. Pour nous rassurer un peu plus, nous allons quantifier la similarité des deux approches. Pour ce faire, calculons des intervalles de confiance à 95 % autour de l'estimateur du BEL issus des deux distributions ajustées pour chaque environnement économique synthétique simulé à la création de notre base de données.

Ne connaissant pas explicitement la loi que suit une somme de variables aléatoires i.i.d. de loi *generalized hyperbolic*, nous utilisons une approche par simulation proche du bootstrap percentile. Il nous faut maintenant une métrique afin de quantifier la similarité des intervalles obtenus.

Introduisons la distance de JACCARD (1901).

$$J(I_1, I_2) = \frac{|I_1 \cap I_2|}{|I_1 \cup I_2|} \quad (4.8)$$

Cette formule présente l'avantage de représenter une quantité interprétable. Elle vaut 1 quand les intervalles se superposent parfaitement et 0 quand ils sont totalement disjoints. Dans notre cas, la figure 4.4 qui suit appuie notre intuition précédente avec une superposition quasi totale sur l'ensemble de notre base de données.

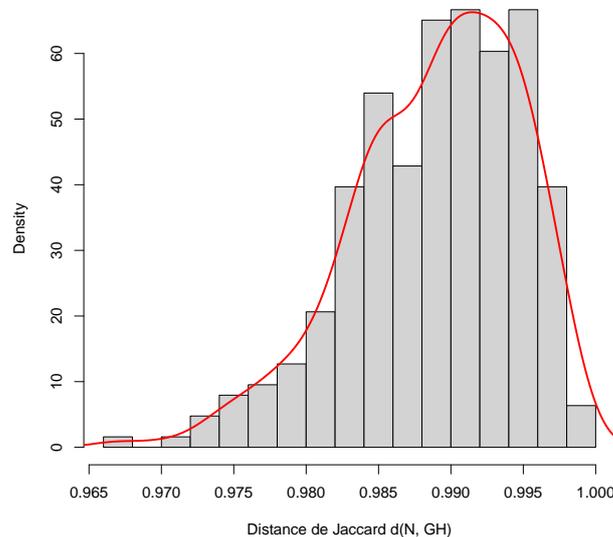


FIGURE 4.4 : Similarité de Jaccard des intervalles de confiance à 95 % d'une approximation gaussienne de la distribution *generalized hyperbolic*

La perte en taux de couverture de notre intervalle de confiance suite à cette simplification devrait donc être limitée à 3-4 % au maximum puisque la distance de Jaccard est un minorant de la proportion de l'intervalle de confiance initial contenu dans son approximation.

3.7 L'*overfitting* et la séparation des ensembles d'entraînement et de test

L'*overfitting*, ou sur-apprentissage, est un problème courant dans l'apprentissage automatique, y compris dans les réseaux de neurones. Il survient lorsqu'un modèle s'ajuste trop étroitement aux données d'entraînement, apprenant non seulement les tendances générales, mais aussi les fluctuations aléatoires ou le bruit. En conséquence, le modèle peut montrer des performances exceptionnelles sur les données d'entraînement mais échouer à généraliser sur des données nouvelles ou non vues.

Définition de l'*overfitting*

L'*overfitting* se produit lorsque le modèle est trop complexe par rapport à la structure sous-jacente des données. Cela peut être dû à plusieurs facteurs :

- **Complexité du modèle** : Un modèle avec trop de paramètres (par exemple, un réseau de neurones très profond) peut apprendre les moindres détails des données d'entraînement, y compris le bruit.
- **Trop peu de données** : Si le modèle est entraîné sur un ensemble de données trop petit, il risque de mémoriser les exemples spécifiques au lieu d'apprendre des *patterns* généralisables.
- **Absence de régularisation** : Sans techniques de régularisation, telles que le *dropout* ou la régularisation L2, les modèles peuvent avoir tendance à sur-apprendre les données d'entraînement.

Lorsque le modèle *overfit*, la fonction de coût sur les données d'entraînement continue de diminuer, tandis que la performance sur les données de validation ou de test se détériore. Ce phénomène est visible lorsque l'écart entre les erreurs sur les ensembles d'entraînement et de test devient significatif.

La séparation entre les ensembles d'entraînement et de test

Afin de prévenir l'*overfitting*, il est essentiel de diviser les données disponibles en au moins deux ensembles distincts : l'**ensemble d'entraînement** et l'**ensemble de test**. Cette séparation permet de mesurer la capacité du modèle à généraliser sur des données qu'il n'a jamais vues.

Ensemble d'entraînement : L'ensemble d'entraînement est utilisé pour ajuster les paramètres du modèle (les poids et biais dans un réseau de neurones). Le modèle est entraîné en minimisant la fonction de coût sur cet ensemble.

Ensemble de test : L'ensemble de test est un ensemble de données distinct, non utilisé pendant l'entraînement. Il permet d'évaluer la performance du modèle après son entraînement. En mesurant l'erreur sur l'ensemble de test, on peut estimer la capacité du modèle à généraliser à de nouvelles données.

Ensemble de validation (optionnel) : Un troisième ensemble, l'ensemble de validation, peut être utilisé pour affiner les hyperparamètres du modèle, tels que le taux d'apprentissage, la profondeur du réseau, ou encore la régularisation. Cela permet d'ajuster ces paramètres sans regarder les performances sur l'ensemble de test, ce qui évite d'optimiser le modèle directement sur cet ensemble.

Stratégies pour éviter l'*overfitting*

Il existe plusieurs stratégies pour réduire le risque d'*overfitting* :

- **Régularisation L2 (*Weight Decay*)** : Cette technique consiste à ajouter une pénalité dans la fonction de coût pour décourager les poids du modèle d'atteindre des valeurs trop élevées. Cela tend à simplifier le modèle en le contraignant à privilégier les poids plus faibles.
- ***Dropout*** : Le *dropout* est une technique où certains neurones sont aléatoirement désactivés pendant chaque étape d'entraînement, forçant le modèle à ne pas dépendre de neurones spécifiques. Cela introduit de la robustesse dans l'apprentissage et réduit la sur-adaptation.
- ***Cross-validation*** : La validation croisée (*cross-validation*) consiste à diviser les données en plusieurs sous-ensembles et à entraîner le modèle sur plusieurs combinaisons de ces ensembles. Cela permet de tester le modèle sur différentes partitions des données et de mieux évaluer sa capacité à généraliser.
- **Arrêt anticipé (*Early Stopping*)** : L'arrêt anticipé consiste à surveiller les performances du modèle sur l'ensemble de validation pendant l'entraînement. Si les performances se dégradent alors que l'erreur sur l'ensemble d'entraînement continue de diminuer, l'entraînement est stoppé pour éviter l'*overfitting*.
- **Augmentation des données** : Une autre méthode consiste à augmenter la taille de l'ensemble de données, soit en collectant davantage de données, soit en appliquant des transformations aux données existantes.

3.8 Valeurs de Shapley et interprétabilité

Les valeurs de SHAPLEY (1953) proviennent de la théorie des jeux coopératifs et sont utilisées pour attribuer la contribution de chaque caractéristique dans un modèle prédictif. Pour un modèle de machine learning, elles ont été démocratisées dans ce contexte par LUNDBERG et LEE (2017) et quantifient l'importance marginale de chaque caractéristique. La valeur de Shapley ϕ_i pour une caractéristique i est définie par :

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} [f(S \cup \{i\}) - f(S)]$$

où N représente l'ensemble des caractéristiques, S est un sous-ensemble des caractéristiques sans i , et $f(S)$ est la fonction de prédiction pour le sous-ensemble S . Cette valeur capture la différence de contribution entre la prédiction avec et sans la caractéristique i , en moyenne sur toutes les permutations possibles.

Dans le contexte des réseaux de neurones, calculer directement les valeurs de Shapley est intrac-table en raison du nombre exponentiel de permutations. Une méthode approximative, basée sur les gradients, est utilisée : le `GradientExplainer`, qui est une implémentation des ***Expected Gradients*** (ERION et al., 2019), une extension des ***Integrated Gradients*** (SUNDARARAJAN et al., 2017) visant à approcher les valeurs d'Aumann-Shapley. `GradientExplainer` fonctionne en intégrant les gradients

des interpolations entre l’échantillon cible (c’est-à-dire l’échantillon expliqué) et un échantillon de référence. Formellement, les **Integrated Gradients** pour une caractéristique i sont donnés par :

$$IG_i(x) = (x_i - \tilde{x}_i) \int_{\alpha=0}^1 \frac{\partial f(\tilde{x} + \alpha(x - \tilde{x}))}{\partial x_i} d\alpha$$

où \tilde{x} est un point de référence (background sample) et α est un coefficient d’interpolation. Cette approche permet d’obtenir une approximation des valeurs d’Aumann-Shapley en calculant les gradients sur le chemin reliant \tilde{x} à x .

Cependant, comme ce processus est coûteux en calcul, **GradientExplainer** utilise un échantillonnage aléatoire pour les interpolations et pour la sélection de l’échantillon de référence, ce qui rend l’algorithme non déterministe. En augmentant le nombre d’échantillons utilisés pour l’approximation, via le paramètre `nsamples`, les valeurs calculées convergeront progressivement vers une estimation plus précise des valeurs de Shapley.

4 Architecture finale du modèle

L’architecture finale du modèle repose sur un réseau de neurones simple avec une régularisation par *dropout*, une initialisation des poids basée sur un modèle de régression linéaire, et une fonction de perte spécifique pour les distributions gaussiennes. Ce modèle vise à prédire deux variables : μ et σ , dans le cadre d’une modélisation actuarielle, où μ représente la meilleure estimation BEL et σ l’écart-type de sa distribution sous-jacente imputable aux scénarii stochastiques à l’origine de ces engagements.

4.1 Structure générale du modèle

Le modèle est composé de deux couches linéaires successives, séparées par une activation non linéaire et un mécanisme de régularisation par *dropout*. La première couche utilise une initialisation des poids basée sur un modèle de régression linéaire préalablement ajusté pour améliorer la convergence initiale.

- **Entrée du modèle** : L’entrée du modèle est un vecteur de dimension 76, représentant les variables explicatives (*features*).
- **Première couche cachée** : Une couche linéaire connecte les 76 variables d’entrée à 16 neurones. Cette couche utilise une fonction d’activation `Tanhshrink` pour introduire une non-linéarité, et est suivie par un *dropout* avec un taux de 20% pour prévenir l’*overfitting*.
- **Deuxième couche cachée** : Une deuxième couche linéaire réduit le nombre de neurones de 16 à 2, fournissant les deux sorties finales, μ et σ .
- **Sorties du modèle** : Le modèle renvoie un vecteur de dimension 2. La première sortie correspond à μ , la meilleure estimation, et la seconde sortie σ est contrainte à être positive via la fonction `softplus`.

La régularisation par *dropout* est utilisée après la première couche pour éviter le sur-apprentissage, un problème courant dans les modèles complexes.

4.2 Fonction d'activation Tanhshrink et softplus

La fonction d'activation **Tanhshrink** est utilisée dans la première couche pour introduire une non-linéarité dans le modèle. Elle est définie par :

$$f(x) = x - \tanh(x)$$

Elle combine une composante linéaire avec une correction par \tanh , ce qui permet de conserver une certaine linéarité tout en lissant les extrêmes. La deuxième sortie, σ , est contrainte à être strictement positive en appliquant une fonction **softplus** sur la sortie brute du modèle. La fonction **softplus** est définie par :

$$\text{softplus}(x) = \log(1 + e^x)$$

4.3 Initialisation des poids

L'initialisation des poids dans la première couche du réseau est basée sur un modèle de régression linéaire ajusté pour prédire la variable μ . Ce modèle linéaire est entraîné séparément avant la mise en place du réseau de neurones, et ses poids sont utilisés comme initialisation pour la première couche afin de bénéficier de la simplicité des relations linéaires dans les premières étapes de l'entraînement.

4.4 Fonction de perte : divergence KL gaussienne

Le modèle est entraîné en minimisant une fonction de perte customisée correspondant à la divergence de Kullback-Leibler (KL) entre deux distributions gaussiennes. Cette fonction de perte est adaptée pour les tâches où les sorties du modèle sont des paramètres de distributions (comme μ et σ dans notre cas). Cette fonction de perte encourage le modèle à prédire non seulement la meilleure estimation μ de la variable cible, mais également une incertitude σ associée à cette prédiction.

4.5 Optimisation et régularisation

Le modèle utilise l'optimiseur **AdamW** présenté précédemment pour ajuster les poids. AdamW est une variante de l'optimiseur Adam qui introduit une régularisation explicite par *weight decay* (pénalisation L2) sur les poids, ce qui améliore la capacité de généralisation du modèle et réduit l'*overfitting*. Le taux d'apprentissage initial est fixé à 0,001, et la technique de **gradient clipping** (majoration du gradient) est utilisée pour éviter les problèmes de gradients explosifs lors de l'entraînement.

4.6 Résumé de l'architecture

- **Entrée** : Vecteur de dimension 76.
- **Couche 1** : Couche linéaire avec 16 neurones, activation `Tanhshrink`, suivie d'un *dropout* à 20%.
- **Couche 2** : Couche linéaire avec 2 neurones, fournissant μ et σ .
- **Sortie** : μ (sortie brute) et σ (après `softplus` pour garantir $\sigma > 0$).
- **Fonction de perte** : Divergence KL gaussienne.

Cette architecture simple, mais plutôt efficace, a été optimisée pour la tâche de prédiction des paramètres d'une distribution gaussienne, avec un accent particulier sur la régularisation pour améliorer la robustesse des prédictions.

En bref...

- Le modèle est composé de deux couches linéaires avec des fonctions d'activation `Tanhshrink` et `softplus`.
- Il est régularisé par *dropout* et une initialisation basée sur un modèle de régression linéaire.
- La fonction de perte utilisée est la divergence KL gaussienne, adaptée pour des prédictions μ et σ .
- L'optimisation est réalisée via AdamW avec régularisation *weight decay* et clipping des gradients.

Chapitre 5

Résultats et discussion

Dans ce chapitre, nous analysons et comparons les performances de divers modèles pour la prédiction des valeurs du BEL et de ses paramètres associés (ici la variance), à l'aide de différentes approches, allant de la régression linéaire aux réseaux de neurones. L'objectif est de comprendre la capacité de chaque modèle à généraliser les résultats et d'identifier les variables les plus influentes dans la prédiction des engagements. Les performances des modèles sont évaluées sur les ensembles d'entraînement et de test, afin de juger de leur capacité de généralisation.

1 Les performances des modèles

1.1 Le modèle GLM gaussien

Dans un premier temps, nous appliquons un modèle linéaire gaussien (GLM) comme méthode de base, avec sélection des variables par un processus *backward* basé sur le critère AIC. Ce modèle permet d'identifier les variables explicatives les plus pertinentes pour la prédiction du BEL, constituant ainsi une référence pour évaluer les autres méthodes.

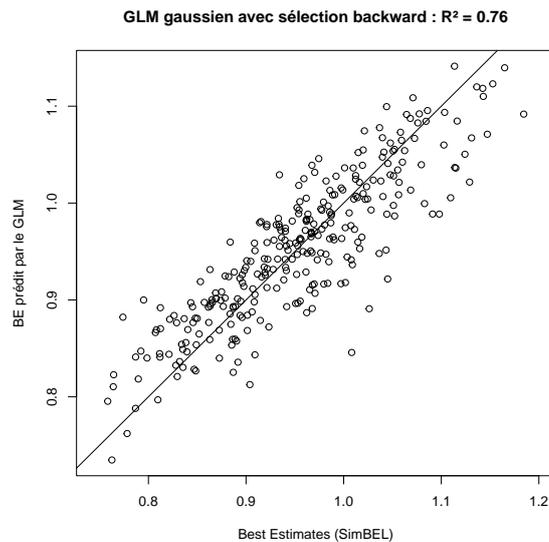


FIGURE 5.1 : Modèle GLM avec sélection *backward* des variables

Analyse des résultats : Comme montré à la figure 5.1, le modèle GLM offre une prédiction correcte sur l'ensemble d'entraînement, mais présente une légère baisse de performance sur le jeu de test (figure 5.1). Ce résultat est attendu pour un modèle linéaire, qui capture les relations directes mais peut manquer de flexibilité face à des interactions non linéaires entre variables.

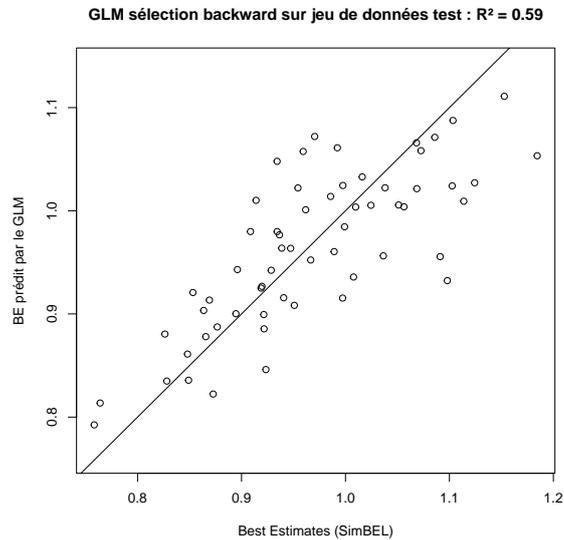


FIGURE 5.2 : Performance du GLM sur le jeu de test

Interprétation des coefficients : Le tableau 5.1 présente les coefficients estimés pour les 10 variables les plus importantes. Ces valeurs fournissent une première idée de l'impact de chaque variable, mais leur interprétation doit tenir compte des limites d'un modèle linéaire dans des situations potentiellement plus complexes.

Variable	Estimate	Std. Error	t value	Pr(> t)
theta_alpha	-3.7225	0.1827	-20.370	< 2e-16 ***
alpha	-0.0239	0.0030	-8.066	2.61e-14 ***
sigma_taux	-1.8671	0.3059	-6.103	3.73e-09 ***
r	0.9844	0.1562	6.303	1.23e-09 ***
tx_tech.8.	0.6387	0.1489	4.290	2.51e-05 ***
sigma_action	-0.2862	0.0842	-3.397	0.0008 ***
marge_mar	-0.1104	0.0304	-3.628	0.0003 ***
tx_tech.3.	0.4970	0.1571	3.164	0.0017 **
tx_tech.6.	0.5509	0.1702	3.237	0.0014 **
tx_tech.10.	0.6442	0.2011	3.204	0.0015 **

TABLE 5.1 : Les 10 variables les plus importantes du modèle GLM

1.2 Modèle *Weighted Least Squares*

Afin de mieux gérer l'hétéroscédasticité dans les données, un modèle de moindres carrés pondérés (WLS) est utilisé. Ce modèle attribue un poids variable aux observations, selon leur variance estimée, afin d'améliorer la robustesse des prédictions dans des contextes où la variance des erreurs fluctue.

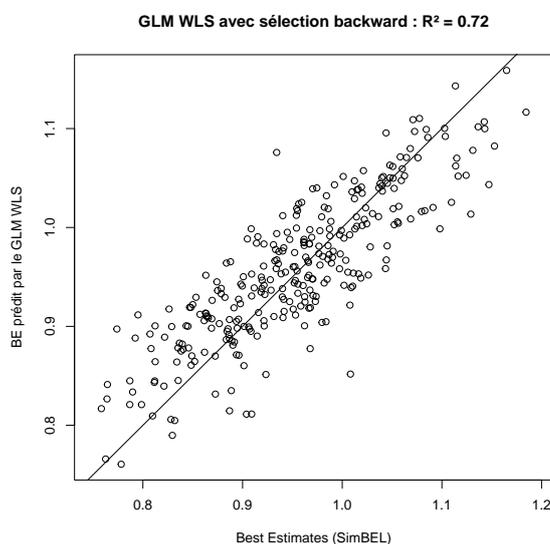


FIGURE 5.3 : Modèle WLS avec prise en compte de l'hétéroscédasticité

Analyse des résultats : Contrairement aux attentes, le modèle WLS n'apporte pas d'amélioration significative en termes de précision. Cependant, l'utilisation des poids rend les coefficients estimés plus robustes, en limitant l'influence des observations à forte variance. Le tableau 5.2 présente les coefficients obtenus pour les variables les plus influentes.

Variable	Estimate	Std. Error	t value	Pr(> t)
theta_alpha	-3.5340	0.1927	-18.337	< 2e-16 ***
alpha	-0.0232	0.0028	-8.379	3.42e-15 ***
sigma_taux	-1.8744	0.3876	-4.837	2.27e-06 ***
r	0.7757	0.1450	5.352	1.92e-07 ***
sigma_action	-0.4357	0.0841	-5.178	4.50e-07 ***
tx_tech.6.	0.7624	0.1586	4.808	2.58e-06 ***
marge_mar	-0.1258	0.0296	-4.249	3.00e-05 ***
tx_tech.3.	0.6347	0.1509	4.207	3.56e-05 ***
val_treso	0.5807	0.1592	3.649	0.0003 ***
rach_tot_from	1.5575	0.5199	2.996	0.0030 **

TABLE 5.2 : Les 10 variables les plus importantes du second modèle GLM

1.3 Le modèle GAMLSS

Le modèle GAMLSS permet une flexibilité supplémentaire en modélisant plusieurs paramètres de distribution (par exemple, la moyenne et la variance) en fonction des covariables. Ce modèle est particulièrement adapté aux données présentant des distributions variables en fonction des valeurs des variables explicatives.

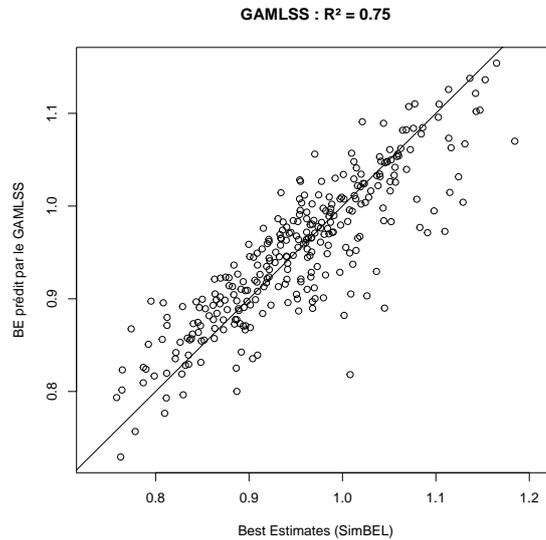


FIGURE 5.4 : Prédiction par le modèle GAMLSS

Analyse des résultats : Bien que le GAMLSS fournisse des résultats comparables au GLM, la flexibilité supplémentaire permet de capturer davantage de variation dans les données, en particulier pour des variables influencées par des distributions asymétriques.

Variable	Estimate	Std. Error	t value	Pr(> t)
alpha	-1.975e-02	2.323e-04	-85.040	< 2e-16 ***
theta_alpha	-4.034e+00	1.565e-02	-257.754	< 2e-16 ***
sigma_taux	-1.388e+00	2.863e-02	-48.470	< 2e-16 ***
sigma_immo	4.326e-01	8.607e-03	50.263	< 2e-16 ***
tx_chargement_fin	5.162e+00	8.797e-02	58.683	< 2e-16 ***
marge_mar	-8.129e-02	2.330e-03	-34.897	< 2e-16 ***
rho_i_t	-1.555e-01	3.841e-03	-40.480	< 2e-16 ***
r	9.019e-01	1.177e-02	76.655	< 2e-16 ***
rach_tot_from	1.471e+00	4.011e-02	36.681	< 2e-16 ***
sigma_action	-1.863e-01	6.685e-03	-27.867	< 2e-16 ***

TABLE 5.3 : Les 10 variables les plus importantes pour μ dans le modèle GAMLSS

1.4 Le réseau de neurones

Un réseau de neurones a été utilisé pour capturer des relations non linéaires complexes entre les variables explicatives et la valeur du BEL. Ce modèle est particulièrement adapté aux données avec un grand nombre de paramètres d'entrée et permet de modéliser des interactions difficiles à appréhender pour des modèles linéaires. Il permet aussi immédiatement la modélisation simultanée de plusieurs quantités d'intérêt, ici les deux premiers moments, μ et σ .

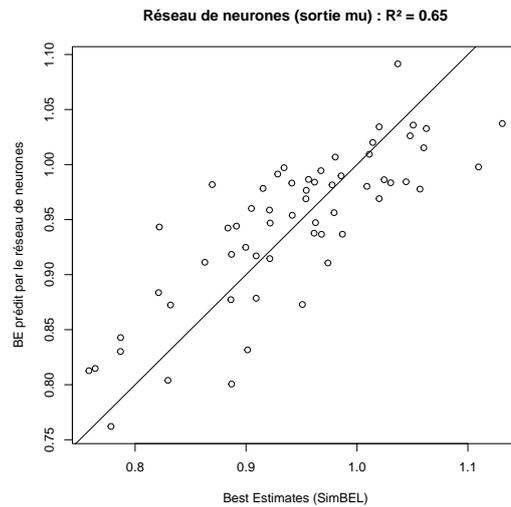


FIGURE 5.5 : Prédiction de μ par le réseau de neurones

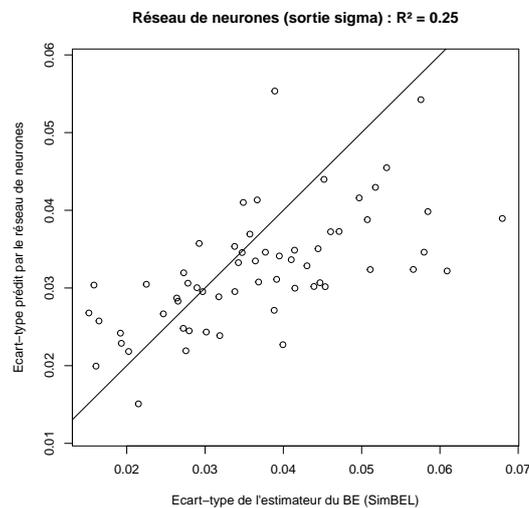


FIGURE 5.6 : Prédiction de σ par le réseau de neurones

Analyse des résultats : La figure 5.5 est à comparer à la figure 5.2 et nous montre de bien meilleurs résultats sur le jeu test. La figure 5.6 montre également que le réseau de neurones permet une estimation grossière de sa propre incertitude au travers de σ , permettant d'anticiper les situations plus complexes à modéliser de la sorte.

Pour évaluer l'importance des différentes variables explicatives dans la prédiction de μ le réseau de neurones, nous avons utilisé les valeurs de l'indicateur SHAP (SHapley Additive exPlanations) présenté à la section 3.8 du chapitre 4. Cet indicateur permet d'interpréter les contributions moyennes de chaque variable à la prédiction, offrant une mesure de leur importance relative. Le tableau ci-dessous présente les 10 variables ayant la plus forte valeur SHAP moyenne, ce qui reflète leur influence significative sur la prédiction de μ .

Variable	Valeur SHAP Moyenne pour Mu
theta_alpha	0.023010
alpha	0.009902
r	0.008529
sigma_taux	0.007943
marge_mar	0.007884
age.5.	0.007861
rho_i_t	0.006300
tx_tech.10.	0.005864
tx_tech.8.	0.005781
tx_tech.5.	0.005450

TABLE 5.4 : Les 10 variables les plus importantes selon les valeurs SHAP pour Mu

2 Comparaisons

Afin d'évaluer la performance des différents modèles, nous avons comparé leurs scores R^2 sur les jeux de données d'entraînement et de test, ainsi que le temps de calcul associé à chaque méthode. Le tableau 5.5 présente un résumé des résultats pour les GLM, les moindres carrés pondérés (WLS), le modèle généralisé GAMLSS, ainsi que le réseau de neurones. Nous observons que, bien que les modèles GLM et GAMLSS montrent de bonnes performances avec des scores R^2 proches sur le jeu de test, le réseau de neurones offre une meilleure généralisation avec un R^2 plus élevé sur le jeu de test, au prix d'un temps de calcul pour l'inférence plus important. La prédiction est ensuite par contre immédiate ce qui est le résultat espéré.

Modèle	R^2 (Train)	R^2 (Test)	Temps d'entraînement
GLM	0.76	0.59	Faible
WLS	0.72	0.53	Faible
GAMLSS	0.75	0.60	Modéré
Réseau de Neurones	0.70	0.65	Plus élevé

TABLE 5.5 : Comparaison des performances des modèles

3 Quelques interprétations

Le tableau 5.6 suivant présente les variables communes identifiées à travers les différents modèles (GLM, WLS, GAMLSS et réseau de neurones) et leur plage de valeurs de coefficients. Ces variables jouent un rôle clé dans la modélisation du BEL et ont un impact significatif sur la prédiction des engagements. Les coefficients des variables varient en fonction du modèle utilisé, mais leur présence constante démontre leur importance dans l'évaluation des risques financiers associés aux contrats d'assurance. Le tableau met en évidence la robustesse de ces variables à travers différentes approches, et leur interprétation permet de mieux comprendre les mécanismes sous-jacents influençant la valorisation des engagements. **Bien évidemment, si nous ne normalisons pas par la PM, elle ressortirait comme variable explicative principale, ce qui est normal et intuitif puisqu'il s'agit du volume brut matérialisant les quantités d'épargne engagées.**

Variable	Plage des Coefficients
theta_alpha	[-4.034, -3.534]
alpha	[-0.0239, -0.0198]
sigma_taux	[-1.8744, -1.3880]
r	[0.7757, 0.9844]
sigma_action	[-0.4357, -0.1863]
marge_mar	[-0.1258, -0.0813]
rach_tot_from	[1.4710, 1.5575]

TABLE 5.6 : Variables consensuelles et plages de coefficients associées

Commentaire sur les variables communes : Les variables communes identifiées dans tous les modèles soulignent des facteurs fondamentaux dans la détermination du BEL. Par exemple, les variables `theta_alpha` et `alpha` – respectivement le taux attendu à l'ultime et sa vitesse de convergence – associées à la dynamique des taux d'actualisation, ont un effet négatif sur le BEL, indiquant qu'une hausse des taux réduit la valeur actuelle des engagements. Les variables de volatilité, `sigma_taux` et `sigma_action` – respectivement les volatilités des dynamiques du taux et des actions – montrent également un impact négatif, signifiant que l'augmentation de la volatilité des indices de taux d'intérêt et des actions diminue la valeur des engagements futurs. On peut mieux le comprendre inversement, une volatilité très faible voire nulle pour notre expérience de pensée induira aucun rachat conjoncturel et les épargnants auront intérêt à garder leur épargne le plus longtemps possible soumise à un rendement stable et quasi certain et donc un engagement revalorisé. À l'inverse, la variable `r`, représentant le taux sans risque, a un effet positif, traduisant une augmentation du BEL par une augmentation des rachats conjoncturels faisant intervenir ces flux plus tôt et donc avec moins d'actualisation. La variable `marge_mar` montre qu'une diminution des marges d'exploitation de la concurrence conduit à une baisse de la rentabilité future, réduisant ainsi la valeur des engagements. Enfin, `rach_tot_from`, qui mesure le niveau des rachats totaux (conjoncturel + structurel), reflète une réduction des engagements à long terme, augmentant la valeur actuelle du BEL. Ensemble, ces variables démontrent l'importance des facteurs économiques et financiers dans la modélisation actuarielle des engagements d'assurance.

Conclusion

Synthèse

Dans ce mémoire, nous avons exploré diverses approches de modélisation pour l'évaluation des engagements en assurance vie, notamment dans le cadre de l'assurance épargne en euros. Nous avons comparé des méthodes traditionnelles telles que les modèles linéaires généralisés (GLM) et les moindres carrés pondérés (WLS) à des approches plus avancées comme le GAMLSS et les réseaux de neurones. Chaque modèle présente des avantages et des limites en termes de complexité, de précision et de temps de calcul. Il apparaît néanmoins que les réseaux de neurones offrent une capacité de généralisation particulièrement intéressante, bien que cela implique des exigences computationnelles plus élevées. La comparaison des modèles via leurs scores R^2 et leurs temps de calcul montre que les méthodes linéaires fournissent une performance appréciable en raison de leur simplicité. En revanche, le GAMLSS et les réseaux de neurones, grâce à leur capacité à modéliser des distributions plus flexibles et complètes, affichent des performances accrues sur des jeux de test non observés lors de l'entraînement.

Il est important de noter que si les performances peuvent sembler modestes comparées aux attentes générées par les méthodes de *machine learning* ou par certains résultats issus d'autres études, les conditions de nos simulations diffèrent souvent. Dans de nombreux travaux, l'environnement économique est fixé pour toutes les observations, alors que nous soulignons ici l'importance cruciale des paramètres économiques. En outre, les tables de mortalité et les lois de rachat sont souvent fixées, tandis que dans notre étude, presque tous les paramètres imaginables ont été variés, avec un volume de données limité pour compenser cette complexité. De plus, la normalisation par la PM, souvent omise dans d'autres études, impacte notablement la qualité des prédictions. Sans cette normalisation, les R^2 sont artificiellement élevés, une dérive que nous avons voulu éviter.

Perspectives

Pour approfondir ce travail, plusieurs pistes d'amélioration peuvent être envisagées. Tout d'abord, fixer certains paramètres et disposer d'une base de données plus vaste pourrait sans doute apporter des résultats plus robustes. Par ailleurs, explorer différentes architectures de modèles de machine learning, comme les modèles d'apprentissage profond ou les réseaux de neurones récurrents, pourrait permettre une meilleure capture des dynamiques temporelles des engagements. La mise en place d'algorithmes d'explicabilité plus avancés afin de mieux analyser l'importance des variables dans les réseaux de neurones offrirait également une interprétabilité accrue des résultats. Enfin, tester le modèle dans une projection type ORSA permettrait de le comparer à celui de l'étude de BONNIN et al. (2014). En parallèle, l'application de méthodes d'optimisation pour identifier des allocations d'actifs optimales constituerait une voie prometteuse.

Bibliographie

Articles de recherche

- AITKIN, M. (1987). Modelling Variance Heterogeneity in Normal Regression Using GLIM. *Applied Statistics*, 36(3), 332.
- BAUER, D., BERGMANN, D. R., & REUSS, A. (2010). Solvency II and Nested Simulations a Least-Squares Monte Carlo Approach.
- BONNIN, F., JUILLARD, M., & PLANCHET, F. (2014). Best estimate calculations of savings contracts by closed formulas : application to the ORSA. *European Actuarial Journal*, 4, 181-196.
- DEVINEAU, L., & LOISEL, S. (2009). Construction d'un algorithme d'accélération de la méthode des "simulations dans les simulations" pour le calcul du capital économique Solvabilité II. *Bulletin Français d'Actuariat*, 10(17), 188-221.
- DOUGLAS, S. C., & YU, J. (2018). Why RELU Units Sometimes Die : Analysis of Single-Unit Error Backpropagation in Neural Networks. *2018 52nd Asilomar Conference on Signals, Systems, and Computers*, 864-868.
- ERION, G. G., JANIZEK, J. D., STURMFELS, P., LUNDBERG, S. M., & LEE, S. (2019). Learning Explainable Models Using Attribution Priors. *CoRR*, abs/1906.10670.
- GOFFARD, P.-O., & GUERRAULT, X. (2015). Is it optimal to group policyholders by age, gender, and seniority for BEL computations based on model points? *European Actuarial Journal*, 5(1), 165-180.
- HULL, J., & WHITE, A. (1993). One-Factor Interest-Rate Models and the Valuation of Interest-Rate Derivative Securities. *The Journal of Financial and Quantitative Analysis*, 28(2), 235.
- JACCARD, P. (1901). Distribution de la flore alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société vaudoise des sciences naturelles*.
- KANNISTO, V. (1992). *Presentation at a Workshop on "Old Age Mortality"* [Odense University, Denmark, June 1992].
- LAURENT, J.-P., NORBERG, R., & PLANCHET, F. (Éd.). (2016). *Modelling in Life Insurance A Management Perspective*. Springer International Publishing.
- LOSHCHILOV, I., & HUTTER, F. (2019). Decoupled Weight Decay Regularization.
- LUNDBERG, S. M., & LEE, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30.
- RIGBY, R. A., & STASINOPOULOS, D. M. (1996). A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, 6(1), 57-65.
- SHAPLEY, L. S. (1953). A Value for n-Person Games. In H. W. KUHN & A. W. TUCKER (Éd.), *Contributions to the Theory of Games II* (p. 307-317). Princeton University Press.

SUNDARARAJAN, M., TALY, A., & YAN, Q. (2017). Axiomatic Attribution for Deep Networks. *CoRR*, *abs/1703.01365*.

VASICEK, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177-188.

Mémoires d'actuariat

BEN YOUSSEF, G. (2023). *Élaboration d'un proxy Machine Learning du BEL à partir de sa sensibilité aux aléas du marché* [Mém. d'actuariat]. ISUP.

CERISIER, V. (2021). *ORSA : Application de méthodes de machine learning dans le calcul de la solvabilité infra-annuelle* [Mém. d'actuariat]. Paris-Dauphine.

GAUVILLE, R. (2017). *Projection du ratio de solvabilité : des méthodes de machine learning pour contourner les contraintes opérationnelles de la méthode des SdS* [Mém. d'actuariat]. EURIA.

LY, A. (2015). *Étude de risques comportementaux et détermination du capital économique par la méthode des réseaux de neurones* [Mém. d'actuariat]. ENSAE.

ROUÉ, A. (2015). *Les réseaux bayésiens comme technique d'optimisation numérique dans le cadre de l'ORSA* [Mém. d'actuariat]. EURIA.

ZURFLUH, E. (2019). *Utilisation du Machine Learning dans l'estimation du ratio de solvabilité d'un assureur vie et application au Reverse Stress Tests* [Mém. d'actuariat]. ENSAE.

Textes légaux

COMMISSION EUROPÉENNE. (2015). *Règlement délégué (UE) 2015/35* (Règlement délégué). Journal officiel de l'Union européenne.

PARLEMENT EUROPÉEN & CONSEIL DE L'UNION EUROPÉENNE. (2009). *Directive 2009/138/CE* (Directive). Journal officiel de l'Union européenne.

PARLEMENT EUROPÉEN & CONSEIL DE L'UNION EUROPÉENNE. (2014). *Directive 2014/51/UE* (Directive). Journal officiel de l'Union européenne.

Webographie

FRANCE ASSUREURS. (2023, février). *Après une année 2021 de rattrapage, l'assurance vie a retrouvé en 2022 son niveau d'avant la crise sanitaire avec 144,4 milliards d'euros de cotisations* (Communiqué de presse) (Consulté le 30/01/2024). https://www.franceassureurs.fr/wp-content/uploads/20221201-france__assureurs_cp-vie_decembre.pdf

Table des figures

1.1	Décomposition des composantes risques du SCR (SOURCE : ACPR)	6
1.2	Obtention de la distribution de fonds propres économiques par la méthode des simulations dans les simulations (SOURCE : DEVINEAU ET LOISEL (2009))	7
2.1	Algorithme de revalorisation	17
2.2	Logique générale d'évaluation (selon LAURENT et al. (2016))	18
3.1	Représentation 3D d'un exemple de taux de rachat partiel structurel	21
3.2	Boîtes à moustaches des allocations d'actif dans la base de données	25
4.1	Histogramme des coefficients inférés avec ou sans correction de l'hétéroscédasticité pour $Y \{A = a, B = b, C = c\} \sim \mathcal{N}(a + b, c^2)$ avec $A, B, C \sim \mathcal{N}(0, 1)$	32
4.2	Boîtes à moustaches des AIC avec et sans pondération	32
4.3	Valeur actualisée des engagements futurs par simulation du GSE	40
4.4	Similarité de Jaccard des intervalles de confiance à 95 % d'une approximation gaussienne de la distribution <i>generalized hyperbolic</i>	41
5.1	Modèle GLM avec sélection <i>backward</i> des variables	47
5.2	Performance du GLM sur le jeu de test	48
5.3	Modèle WLS avec prise en compte de l'hétéroscédasticité	49
5.4	Prédiction par le modèle GAMLSS	50
5.5	Prédiction de mu par le réseau de neurones	51
5.6	Prédiction de sigma par le réseau de neurones	51



Liste des tableaux

1	Comparaison des performances des modèles	xiii
2	Variables consensuelles et plages de coefficients associées	xiii
3	Model Performance Comparison	xix
4	Consensus Variables and Associated Coefficient Ranges	xix
1.1	Bilan économique minimaliste	6
3.1	Description sommaire des variables de contrôle relatives au GSE de notre base de données	24
3.2	Allocations d'actifs dans la base de données	25
3.3	Résumé des différents <i>model points</i> de notre base de données	26
5.1	Les 10 variables les plus importantes du modèle GLM	48
5.2	Les 10 variables les plus importantes du second modèle GLM	49
5.3	Les 10 variables les plus importantes pour μ dans le modèle GAMLSS	50
5.4	Les 10 variables les plus importantes selon les valeurs SHAP pour Mu	52
5.5	Comparaison des performances des modèles	52
5.6	Variables consensuelles et plages de coefficients associées	53

