

**Pour l'obtention du diplôme
de Statisticien Mention Actuariat
et l'admission à l'Institut des Actuaires
le 09/01/2023**

Par : **Clément GOUNEL**

Titre: **Optimisation du processus de pilotage du risque de spread par des méthodes de Machine Learning**

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

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

Membres présents du jury de la filière

*Entreprise : Deloitte **D.***

Nom :

Signature :

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

Directeur de mémoire en entreprise :

Nom : Limbert Guillaume, Michel Loïc

Signature :



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

Signature du responsable entreprise

CYRIL CHALIN



Secrétariat :

Signature du candidat

Bibliothèque :



Remerciements

C'est avec un grand plaisir que je m'apprête à adresser mes remerciements dans les lignes qui vont suivre. Ils me permettent de témoigner ma reconnaissance aux personnes qui m'ont accompagné tout au long de ce stage-alternance et de ma scolarité à l'ISUP et l'ESILV.

Je remercie en premier lieu l'ensemble du groupe Deloitte et plus particulièrement sa branche Risk Advisory pour son accueil et son intégration chaleureuse.

Je tenais à remercier plus particulièrement Guillaume LIMBERT, Loic MICHEL et Aurélien SURAY, mes tuteurs et encadrants de ce stage-alternance pour leur patience, leur écoute ainsi que leur disponibilité qui m'ont permis de compléter les missions qui m'étaient attribuées.

Je remercie aussi les associés Claude CHASSAIN, Cyril CHALAIN et Baptiste BRECHOT ainsi que la directrice Imène SEGHOUANI de m'avoir donné la chance d'intégrer leurs équipes.

Mes remerciements s'adressent aussi à l'ensemble de l'équipe du département Risk Advisory et plus particulièrement au département Financial Risk dont l'ambiance chaleureuse et humaine m'a permis de m'épanouir dans les missions qui m'ont été confiées.

Je souhaite exprimer ma reconnaissance à Olivier LOPEZ le directeur de l'ISUP et tuteur école de ce mémoire pour ses conseils avisés, Laurent IMBERT et Aline ELLUL les responsables des départements actuariat et informatique de l'ESILV. Je tenais aussi à remercier tout le personnel de l'ISUP et de l'ESILV m'ayant prodigué dans un environnement de travail optimal tous les prés requis et conseils nécessaires au bon déroulement de ce stage.

Enfin je souhaite remercier ma famille, mes proches et toutes les personnes qui m'ont soutenu et/ou aidé pour l'écriture de ce mémoire.

Table des matières

Résumé	4
Abstract.....	5
Synthèse.....	6
Executive summary	12
Introduction	18
1 Cadre de l'étude.....	21
1.1 Assurance vie et contrats euro	21
1.2 Solvabilité 2	22
1.3 Le calcul du SCR	24
1.4 Le fonctionnement d'un modèle de simulation actuariel et des GSE	30
2 Outils statistiques et environnement informatique	36
2.1 Outils statistiques	36
2.2 Généralité sur la datascience et description des algorithmes utilisés dans l'étude	42
3 Implémentation d'un réseau de neurone générateur de chocs obligataires	50
3.1 Création du générateur des chocs.....	50
3.2 Fonctionnalités du générateur de densité et calibration des chocs.....	57
3.3 Construction des bases de données de Machine Learning et présentation du cadre expérimental	66
4 Analyse des résultats	72
4.1 Présentation des résultats sur les Bases de données avec les chocs obligataires uniquement activés	72
4.2 Présentation des résultats sur des bases de données dans des configurations plus complexes	86
4.3 Application sur le SCR de spread obligataire et de taux	105
4.4 Les limites du projet	109
5 Conclusion	110
6 Annexes	112
6.1 Acronymes.....	112
6.2 Glossaire.....	114
6.3 Glossaire numérique.....	118
6.4 Table des figures.....	120
6.5 Hyperparamétrage XGboost et la descente de gradient	122
6.6 Hyperparamétrage des réseaux de neurones développés.....	124
6.7 Courbe de densité des coefficients de RN par univers pour 100 portefeuilles répliqués	127
6.8 Tableau des univers de création des densités grecques et françaises	129
6.9 Mémoire de fin d'étude d'Aurélien Suray, ESILV promo 2015-2020.....	130
6.10 Annexes Base de données chocs obligataires avec un pool de TMG	131
6.11 Annexes Base de données chocs obligataires avec 3 pools de TMG	133

6.12	Analyse des corrélations des variables explicatives ajoutés par la création de l’GSE variant	137
6.13	Annexe Base de données chocs obligataires et GSE variant avec un pool de TMG	148
6.14	Annexe base de données chocs obligataires, GSE variant et trois pools de TMG	151
6.15	Annexe Base de données GSE variant avec un pool de TMG	154
6.16	Annexe de la Base de données GSE variant avec 3 pools de TMG	160
6.17	Annexe pour la Base de données GSE variant et indicateurs obligataires avec un pool de TMG	166
6.18	Annexe Base de données GSE variant et indicateurs obligataires avec trois pools de TMG.....	170
6.19	Annexe Base de données GSE variant, chocs obligataires et réallocation dynamique avec un pool de TMG	173
6.20	Annexe Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG	176
7	Bibliographie.....	181

Résumé

Mots clefs : ALM, S2, BEL, SCR, GSE, Optimisation des processus de production, IA, Statistiques Machine Learning

Avec plus de 2 425,5 milliards d'euro de placement en France¹, le secteur de l'assurance est l'investissement préféré des français, les assureurs font donc face à de forts enjeux économiques et sociaux. Afin de répondre à leurs engagements envers les bénéficiaires, les organismes d'assurance doivent respecter de nombreuses normes assurantielles : Solvabilité 2 (S2), IFRS17, French GAAP, les critères Environnementaux, Sociaux et de Gouvernance (ESG)...

Le calcul de ces engagements nécessite l'appel à des outils complexes tel que le générateur de scénarios économiques (GSE). Un scénario économique correspond à un univers économique fictif dans lequel les actifs et le passif de l'assureur vont évoluer dans le cadre de simulation déterministe ou stochastique. La complexité de la création de GSE ainsi que le calcul des projections peuvent s'avérer chronophage.

Dans un environnement économique fortement volatile (5,2% d'inflation en mai 2022² accompagné d'une hausse des taux et d'incertitudes sur les marchés financiers³), les compagnies d'assurance souhaitent optimiser leurs processus de pilotage.

Une optimisation du processus de pilotage accorderait plus de temps aux services ALM pour des analyses plus approfondies des mouvements des actifs par rapport au passif.

Ce temps d'analyse supplémentaire permettrait une meilleure anticipation des changements de stratégie de pilotage à effectuer en cas de mouvements brutaux des marchés, ou de changements de comportement des assurés.

L'évolution de plus en plus rapide des systèmes d'informations, ainsi que la capacité à stocker et analyser les données ont permis l'utilisation et l'accès à de nouveaux outils statistiques basés sur l'Intelligence Artificielle (IA).

Le recours à ces nouveaux outils peut permettre d'optimiser le temps de pilotage ALM d'une compagnie.

Le but de ce mémoire est de mettre en place un processus de prédiction d'un BEL applicable au Solvency Capital Requirement (SCR) de spread obligatoire ainsi qu'une ouverture sur le SCR de taux et leur agrégation en fonction des variations de l'actif sous-jacent, des conditions de marchés et du passif par des algorithmes de Machine Learning.

Le processus de prédiction sera optimisé par l'implémentation d'un outil générateur de portefeuilles obligataires basé sur des méthodes statistiques inspirées de réseaux de neurones.

¹ France assurance (2021). *L'assurance Française, Données clefs 2020*

² INSEE (2022). *Indice des prix à la consommation - résultats provisoires (IPC) - mai 2022*

³ Voir Le Monde (2022). *Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux* et ECB (2022). *Statistical Data Warehouse*

Abstract

Keywords: ALM, S2, BEL, SCR, ESG, Production process optimization, AI, Machine Learning statistics

With more than 2,425.5 billion ⁴euros invested in France, the insurance policies is the preferred investment vehicle of French people, insurers are therefore facing strong economic and social challenges. In order to meet their commitments to beneficiaries, insurance companies must comply with numerous insurance standards: Solvency 2 (S2), IFRS17, French GAAP, Environmental, Social and Governance (ESG)...

The calculation of these commitments requires the use of complex tools such as the economic scenario generator (ESG). An economic scenario corresponds to a fictitious economic universe in which the insurer's assets and liabilities will evolve within the framework of deterministic or stochastic simulation.

The complexity of creating an ESG and calculating projections can be time consuming.

In a highly volatile economic environment (5.2% inflation in May 2022⁵ followed by rising interest rates and uncertainty in the financial markets ⁶), insurance companies want to optimize their management steering processes.

Optimizing the steering process would allow more time to perform more in-depth analyses of gap analysis. This additional analysis time would allow better anticipation of changes in driving strategy in case of sudden market movements or changes in policyholder behavior.

The increasingly rapid evolution of information systems, as well as the ability to store and analyze data, has enabled the use of and access to new intelligence-based statistical tools.

The use of these new statistical tools would allow to optimize the production time necessary for the ALM steering process of a company.

The goal of this thesis is to implement a process for predicting a BEL applicable to the Solvency Capital Requirement (SCR) of a bond spread and an opening on the SCR of interest and their aggregation as a function of variations in the underlying asset, market conditions and liabilities by Machine Learning algorithms.

The prediction process will be optimized by the implementation of a bond portfolio generator tool based on statistical methods inspired by neuronal networks.

⁴ See France assurance (2021). *L'assurance Française, Données clefs 2020*

⁵ See INSEE (2022). *Indice des prix à la consommation - résultats provisoires (IPC) - mai 2022*

⁶ See Le Monde (2022). *Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux* et ECB (2022). *Statistical Data Warehouse*

Synthèse

Dans un contexte économique fortement volatile (inflation de 5,2% en mai 2022⁷) il est compliqué pour les assureurs d'anticiper leurs engagements (forte variation des marchés boursier et montée des taux d'intérêts⁸). Face à ces incertitudes, les assureurs cherchent à estimer leurs engagements le plus rapidement possible.

Afin de réduire le coût temporel lié aux importants calculs à effectuer, des algorithmes basés sur différentes méthodes statistiques peuvent être utilisés. Une fois bien calibrés, ils seraient capables de déterminer la valeur du BEL et du SCR de spread obligataire, de taux et leur agrégation avec une faible erreur d'approximation.

Cette nouvelle solution permettrait d'économiser un temps important dans un environnement économique volatile où le pilotage est de plus en plus complexe.

Dans le but d'avoir des prédictions rapides et précises des résultats S2, ce mémoire s'intéresse aux possibilités d'utiliser des outils basés sur l'intelligence artificielle.

Une utilisation efficace de ces algorithmes permettrait de faciliter au maximum le temps de calcul des résultats S2.

Pour représenter les interactions entre les actifs et le passif, les assureurs doivent étudier le comportement des bilans de leurs compagnies. Il existe deux types de bilans nommés bilan comptable et économique, ayant deux buts et procédé de construction distincte.

La version originelle du bilan est celle du bilan comptable, ce bilan fait intervenir des actifs et du passif avec une valeur nette comptable « figée » (valeur d'achat – amortissement) amortie dans le temps.

Le bilan économique a pour but de renvoyer l'image la plus « réaliste » de l'actif et du passif de l'assureur⁹.

Pour effectuer un gain de temps sur la construction de la base de données, un générateur de densité de chocs obligataires a été codé sous Python. Ces algorithmes permettent de générer des chocs obligataires inspirés de méthodes d'apprentissage automatique et d'apprentissage profond tels que le perceptron informatique ou le Réseau de Neurones de Convolution (RNC)¹⁰.

Une fois les chocs créés, ils sont stockés puis connectés à Prophet via une table de chocs. Des variables explicatives pour le Machine Learning sont ensuite créées.

⁷ INSEE (2022). *Indice des prix à la consommation - résultats provisoires (IPC) - mai 2022*

⁸ Voir Le Monde (2022). *Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux* et ECB (2022). *Statistical Data Warehouse*

⁹ Voir partie 1, sous partie 1.2.2

¹⁰ Voir partie 2, chapitres 2.2.3.6 et 2.2.3.7

Le réseau de neurone est présenté par le schéma suivant :

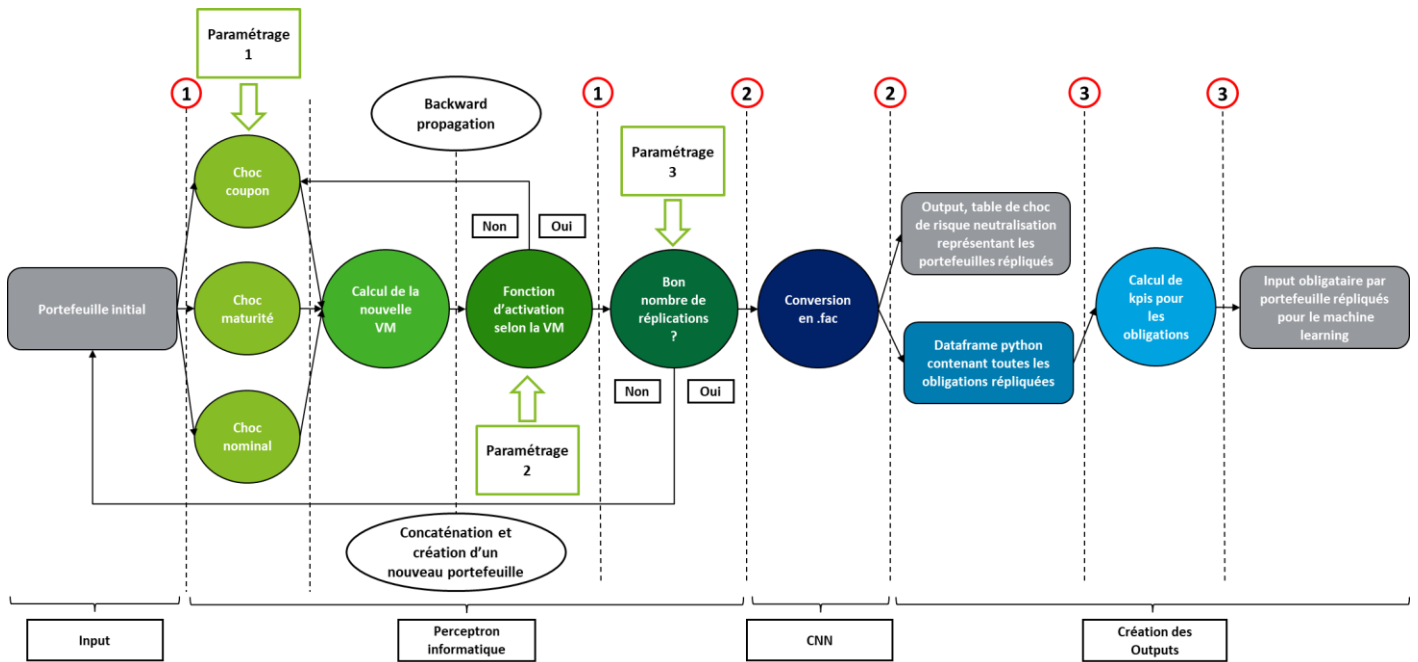


Figure 1 schéma du réseau de neurone pour la génération des inputs obligataires

La réplication des portefeuilles se fait à l'aide d'un perceptron informatique (partie 1 du schéma). Pour effectuer une connexion des résultats entre Python et Prophet, les nouvelles Valeurs de Marchés (VM) des obligations créées sont converties en coefficients de choc¹¹, à l'aide d'un convertisseur inspiré d'une méthode d'apprentissage profond (partie 2 du schéma). Les portefeuilles répliqués sont ensuite convertis en fichier Excel pour le retraitement des inputs de Machine Learning (ML), partie 3 du schéma.

Un convertisseur de fichier inspiré d'un RNC a été implémenté afin de connecter les coefficients de choc générés à Prophet :

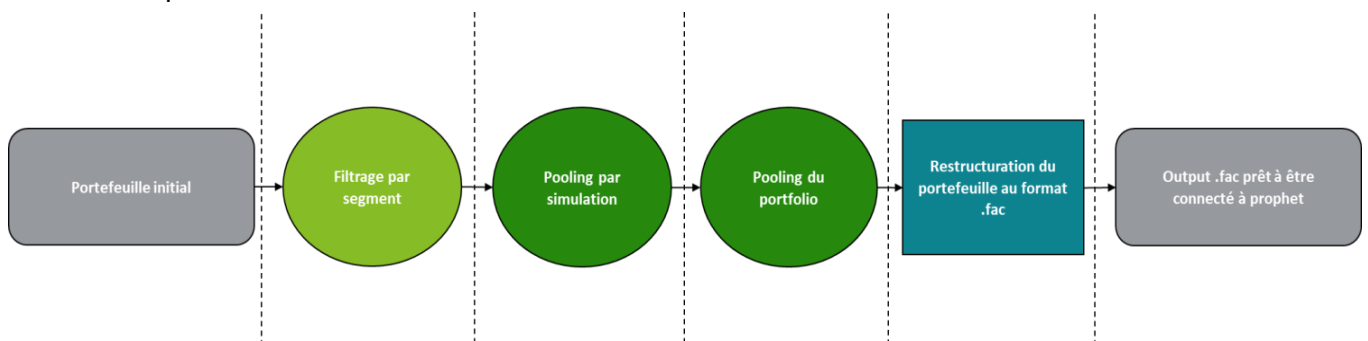


Figure 2 schéma du convertisseur inspiré d'un RNC

La mise en place des scénarios obligataires a été effectuée sur une analyse des obligations françaises puis des obligations grecques, pour vérifier l'évolution de la capacité de prédiction des algorithmes lors d'un changement de base de données.

¹¹ $Coeff_{choc} = \frac{VM_{nouvelle}}{VM_{initiale}}$

Les algorithmes ont appris sur les scénarios grecs étant donné leur plus forte diversité par rapport aux scénarios français.

L'analyse des obligations grecques fait apparaître un scénario de densité avec un début de distribution avec des densités réduites, et beaucoup plus marquée sur la queue de distribution¹².

Une fois calibré, l'outil génère les coefficients de chocs suivants :

Quantile	Valeur
0%	0.67
10%	0.91
20%	0.94
30%	0.95
40%	0.96
50%	0.97
60%	0.98
70%	0.99
80%	1.00
90%	1.02
100%	1.03

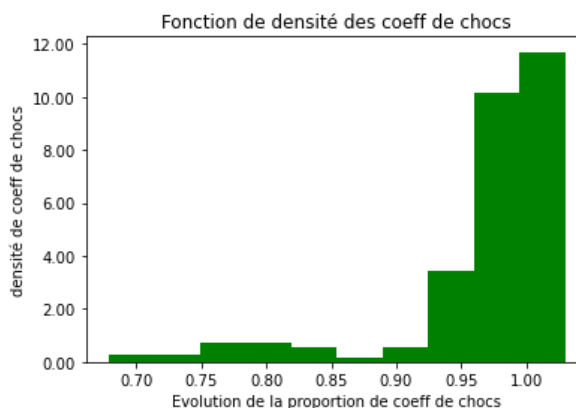


Figure 3 densité des chocs obligataires grecs

Afin de vérifier la robustesse des prédictions, le générateur de spread obligataire a été recalibré après une analyse des obligations françaises 10 ans.

Une fois calibré, l'outil génère la densité des coefficients de chocs suivants¹³:

Quantile	Valeur
0%	0.90
10%	0.92
20%	0.93
30%	0.95
40%	0.97
50%	0.98
60%	0.99
70%	1.00
80%	1.01
90%	1.01
100%	1.02

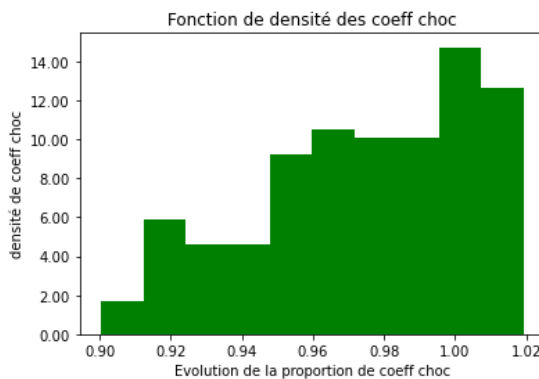


Figure 4 densité des chocs obligataires françaises

¹² Voir partie 3, chapitre 3.2.2.1

¹³ Voir partie 3, chapitre 3.2.2.2

La déclaration de l'actif est présentée par le graphique ci-dessous :

Répartition des actifs du portefeuille initial

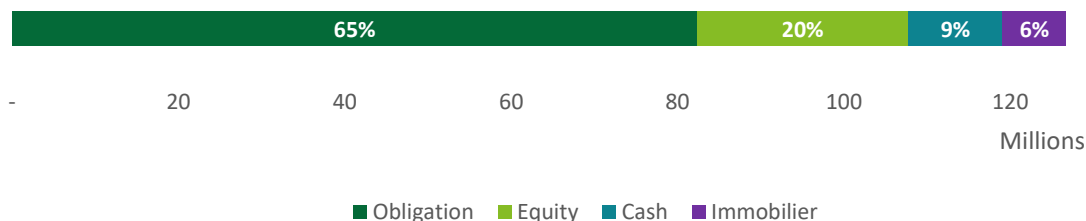


Figure 5 répartition des actifs des Bases De Données (BDD) pour le ML

Cette répartition s'inspire de la composition du portefeuille d'un fonds euro « moyen » français ¹⁴¹⁵.

Afin de tester la sensibilité des algorithmes à la complexification du passif lié à l'interaction actif-passif, deux types de portefeuilles de passif sont utilisés, dont l'un avec uniquement des TMG à 0% et un second avec trois groupes de TMG.

Afin d'effectuer des prédictions sur un maximum de contextes économiques et de vérifier l'efficacité du générateur de chocs obligataires, cinq BDD ont été implémentés ¹⁶.

- Une Base de données avec des chocs obligataires et un GSE continu
- Une Base de données chocs obligataires et un GSE variant
- Une Base de données avec un GSE variant
- Une Base de données avec un GSE variant et des inputs obligataires
- Une Base de données avec un GSE variant, chocs obligataires et une réallocation d'actif dynamique

Les BELs ont d'abord été prédits avec un passif composé d'un unique TMG, puis avec un passif composé de trois TMG différents pour chaque base de données.

Aucune sélection de variable¹⁷ n'a été effectuée sur les sept inputs obligataires.

Le dataset des variables de marché produit par XSG a été séparé en 6 parties issues des mêmes GSE stochastiques

La data visualisation des variables s'est réalisée par analyse des corrélations linéaires et de rangs. La sélection des variables a réduit le nombre de variables explicatives de 2304 à 130 inputs.

Modèle pour évaluer le BEL

¹⁴ France assurance (2021). *L'assurance Française, Données clefs 2020*

¹⁵ Voir partie 3, chapitre 3.3.2.7

¹⁶ Voir partie 3, chapitre 3.3.2.9

¹⁷ Voir parties 4 et annexe 6, chapitres 4.1.2.3 et les sept chapitres du 6.12

Afin d'effectuer la cross-validation des résultats, les bases de données ont été découpées en deux parties le dataset train et test sur un découpage de 80/20%.

La cross validation a été effectuée selon une méthode k-fold avec une réduction optimale de la moyenne des écarts absolus par quantiles de BE.

Tel que :

$$\text{Moyenne des écarts} = \frac{1}{nb_{quantile}} * \sum_{i=0}^n |Ecart_{quantile_i}|$$

Avec :

$$Ecart_{quantile_i} = \frac{Val_{modèle_quantile_i} - Prediction_{quantile_i}}{Val_{modèle_quantile_i}}$$

Pour éviter tout sous-apprentissage ou un sur-apprentissage¹⁸, la moyenne des écarts a été calculée sur 20 quantiles de prédiction.

Les résultats par quantile pour les datasets test et français pour la Base de données à GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG sont présentés ci-dessous.

Dataset test

Modèle Ecarts aux quantiles ¹⁹	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	> 0,1%	> 0,1%	0,3%	-0,6%
19%	-0,1%	0,0%	-0,1%	-0,2%
29%	> 0,1%	> 0,1%	> 0,1%	-0,1%
39%	> 0,1%	> 0,1%	> 0,1%	-0,1%
49%	0,1%	> 0,1%	-0,1%	-0,1%
59%	-0,1%	> 0,1%	-0,1%	-0,4%
69%	0,1%	> 0,1%	0,1%	-0,3%
79%	> 0,1%	-0,1%	-0,4%	0,6%
89%	-0,1%	0,2%	> 0,1%	1,9%
99%	0,7%	-1,3%	0,3%	-4,2%

Figure 6 Présentation des résultats par quantiles pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG, dataset test

Dataset Français

Modèle Ecarts aux quantiles	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	-0,5%	-0,5%	-0,4%	-0,8%
19%	-0,2%	-0,3%	-0,3%	-0,4%
29%	-0,2%	-0,2%	-0,2%	-0,2%
39%	0,1%	0,1%	0,1%	-0,1%
49%	0,3%	0,3%	0,3%	-0,2%
59%	0,2%	0,1%	0,1%	-0,1%
69%	> 0,1%	> 0,1%	-0,1%	-0,1%
79%	-0,1%	-0,2%	-0,3%	0,6%
89%	0,4%	0,3%	0,1%	2,0%
99%	4,9%	0,5%	1,4%	-4,5%

¹⁸ Voir annexe partie 6, chapitre 6.2.2

¹⁹ Pour les tableaux ayant des cases en gras, celles-ci représentent les écarts au BE des prédictions inférieurs à environ 1%.

Figure 7 Présentation des résultats par quantiles pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG, dataset test

Hors l’algorithme de régression linéaire, les trois autres algorithmes prédictifs possèdent des écarts dans la target fixée à environ 1% maximum. Deux d’entre eux (XGboost et Gradient Boosting) ont gardé de très bons résultats lors du changement de base de données.

Ces résultats étant représentatifs des autres bases de données²⁰, les outils implémentés sont capables de prédire de bons BE avec des portefeuilles d’actifs et de passif relativement simples.

L’étude a montré que la prédiction des BE donne des résultats concluants pour des prédictions aux quantiles entre 10 et 90%.

Une étude supplémentaire sur les SCR de spread, ainsi qu’une ouverture sur le SCR de taux ainsi que son agrégation avec le SCR de spread a été effectuée²¹.

Le résultat d’un SCR de spread obligataire pour une BDD chocs obligataires, GSE variant et réallocation d’actif avec un pool de TMG donne le résultat suivant :

Modèle	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR de spread	3 398	3 303	2 801	2 474	3 374
Deviation	0,00%	-2,80%	-17,57%	-27,19%	-0,70%

Figure 8 SCR de spread obligataire pour une BDD chocs obligataires, GSE variant et réallocation d’actif avec 1 pool de TMG

Cette étude a montré que les prédictions ligne à ligne peuvent être améliorés et que malgré de bons résultats sur des bases simples, le modèle nécessite quelques complexifications pour avoir de bonnes prédictions pour des modèles plus difficiles.

L’étude initiale et son ouverture peuvent être améliorées par l’apprentissage de scénarios basés sur des datasets avec des chocs de taux à la hausse ou à la baisse.

Le bilan de cette étude est positif avec une prédiction des BE et des SCR de spread satisfaisante pour la majorité des BDD abordées. Pour améliorer les prédictions du SCR de taux, il faudra faire apprendre des datasets avec des chocs de taux aux algorithmes de ML.

²⁰ Voir partie 4, sous parties 4.1 et 4.2

²¹ Voir partie 4, sous partie 4.3

Executive summary

In a highly volatile economic context (inflation of 5.2% in May 2022²²), it is complicated for insurers to anticipate their commitments (strong variation of the stock markets and rising interest rates²³). Faced with these uncertainties, insurers seek to estimate their commitments as quickly as possible.

To reduce the time cost related to the important calculations to be performed, algorithms based on different AI can be used. Once properly calibrated, they would be able to determine the value of the BEL with a low approximation error.

This new solution would save a lot of time in a multi-standard environment (IFRS17 comes into force in 2023) where ALM management is becoming more and more complex.

To have fast and accurate predictions of S2 results, this thesis focuses on the possibilities of using tools based on artificial intelligence.

An efficient use of these algorithms would facilitate the computation time of the S2 results as much as possible.

To represent the interactions between assets and liabilities, insurers need to study the behavior of their companies' balance sheets. There are two types of balance sheets, called accounting and economic balance sheets, with two distinct purposes and construction processes.

The original version of the balance sheet is the accounting balance sheet, this balance sheet was created by the monk Luca Pacioli in 1494. This balance sheet involves assets and liabilities with a "fixed" net book value (purchase value - depreciation) amortized over time.

The purpose of the economic balance sheet is to provide the most "realistic" picture of the insurer's assets and liabilities²⁴.

To save time on the construction of the database, a spread's shock density generator has been coded in Python. This generator allows to generate a replicating portfolio tool inspired by Machine Learning and deep learning methods such as the computer perceptron or the Convolutional Neuronal Network (CNN).²⁵

Once the portfolios replicated, this method allows the conversion of a bond shock table to be connected to Prophet and the creation of bond inputs.

The neural network is presented in the following diagram:

²² INSEE (2022). *Indice des prix à la consommation - résultats provisoires (IPC) - mai 2022*

²³ Voir Le Monde (2022). *Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux* et ECB (2022). *Statistical Data Warehouse*

²⁴ See part 1, subpart 1.2.2

²⁵ See part 2, sections 2.2.3.6 et 2.2.3.7

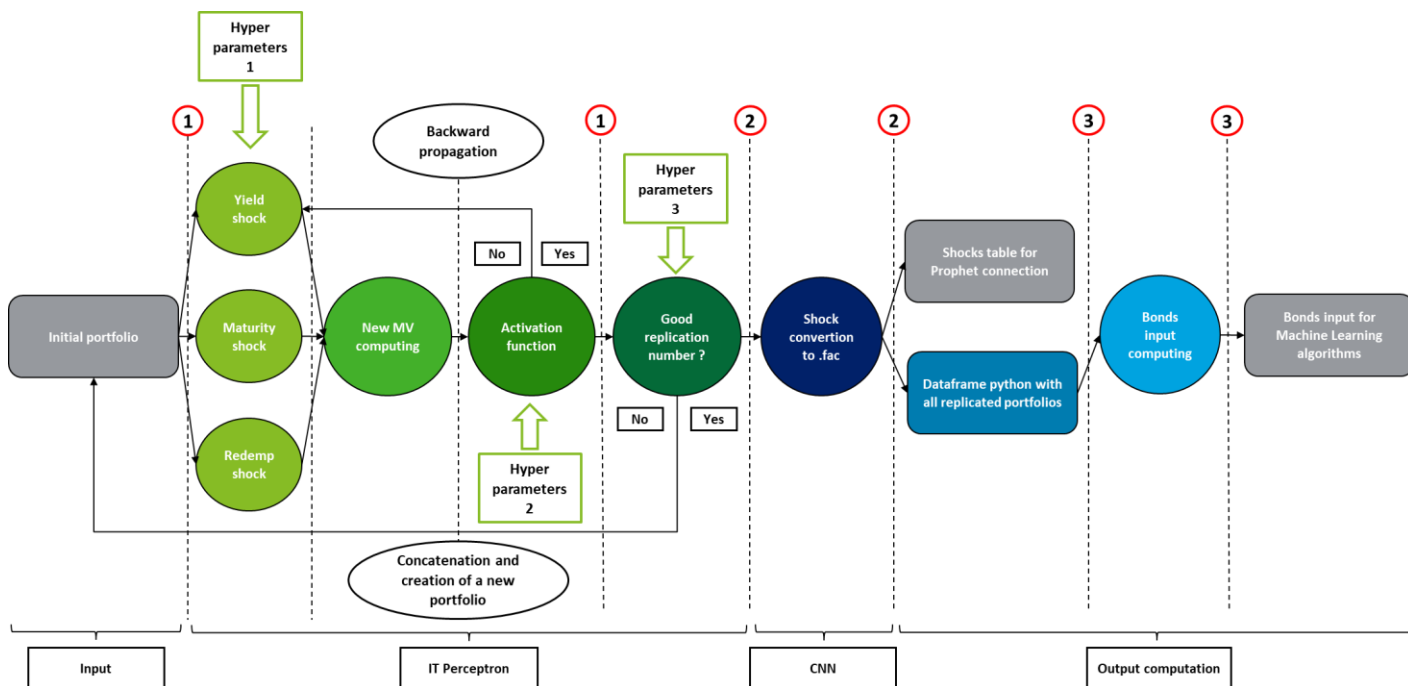


Figure 9 Neuronal bond's generator chart

The replication of the portfolios is done using a computer perceptron (part 1 of the chart). To connect the results between Python and Prophet, the new VMs of the created bonds will be converted into shock coefficients²⁶, using a converter based on a deep learning method (part 2 of the chart). The replicated portfolios will be converted into an Excel file for reprocessing ML inputs (part 3 of the chart).

A converter inspired by a CNN was implemented to connect the generated shock coefficients to Prophet:

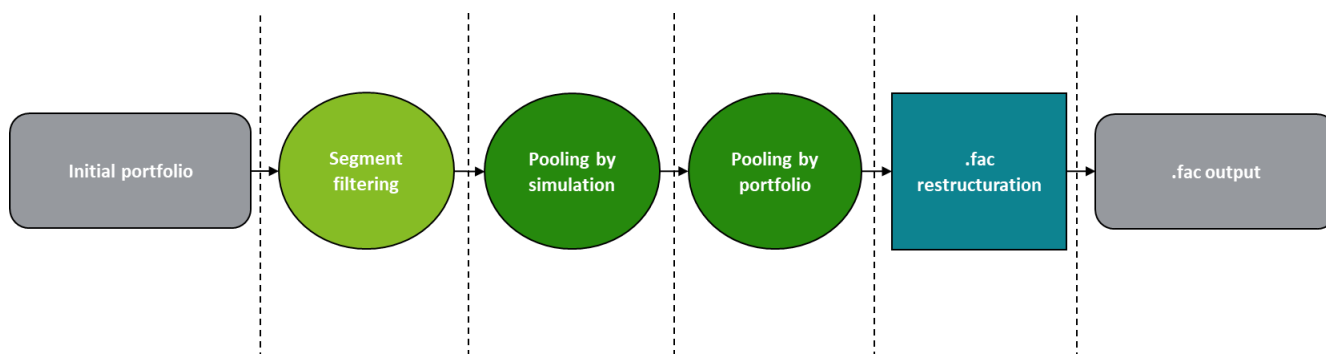


Figure 10 CNN chart

The implementation of the bond scenarios was carried out on an analysis of French bonds and Greek bonds, to check the evolution of the prediction capacity of the algorithms when the database is changed. The algorithms learned on the Greek scenarios given their higher diversity compared to the French scenarios.

The analysis of Greek bonds reveals a density scenario with a slight tail of distribution on the left, and much more marked on the right.²⁷

²⁶ $Coef_{shock} = \frac{VM_{new}}{VM_{initial}}$
²⁷ See part 3, section 3.2.2.1

Once calibrated, the tool will output the following shocks coefficients for the replication of 100 portfolios:

Quantile	Value
0%	0.67
10%	0.92
20%	0.94
30%	0.95
40%	0.96
50%	0.97
60%	0.98
70%	0.99
80%	1.00
90%	1.01
100%	1.03

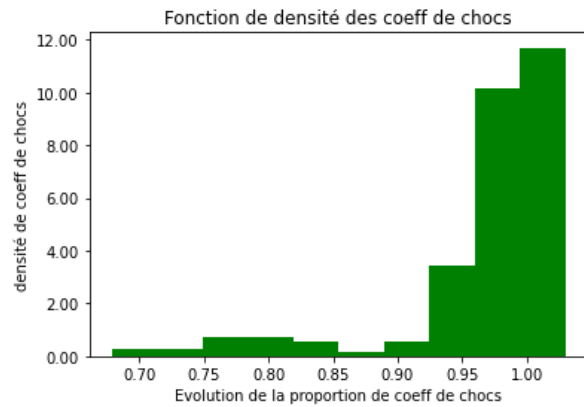


Figure 11 Density of Greek bonds shock

The density curve then created follows a similar shape to the desired curve (light tail on the left and heavy tail on the right).

to check the robustness of the algorithms, the tool was recalibrated after an analysis of French 10-years bonds since the adoption of the euro by France in 2002.²⁸

Once calibrated, the tool will output the following shock coefficients for the replication of 100 portfolios:

Quantile	Value
0%	0.90
10%	0.92
20%	0.93
30%	0.95
40%	0.97
50%	0.98
60%	0.99
70%	1.00
80%	1.01
90%	1.01
100%	1.02

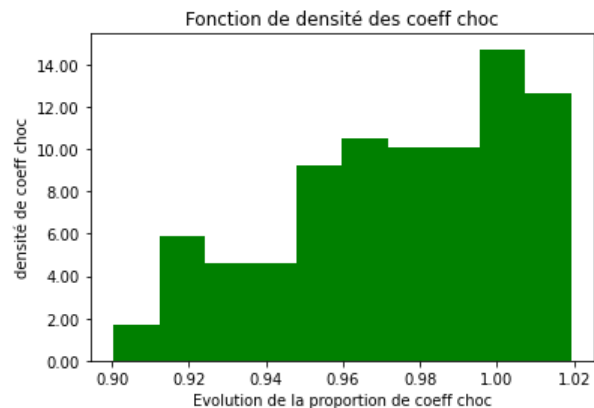


Figure 12 Density of Français bonds shock

The density curve then created follows a similar shape to the desired curve (very "heavy" distribution tail on the right).

The asset statement is presented by the graph below:

²⁸ See part 3, section 3.2.2.2

Initial asset allocation

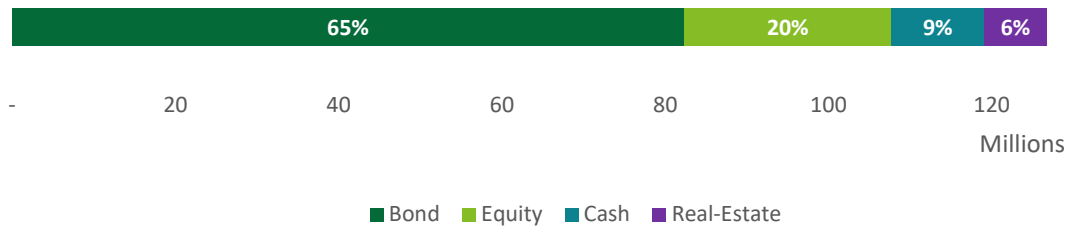


Figure 13 initial asset allocation

This allocation is based on the portfolio composition of an "average" French euro fund.²⁹³⁰

To test the sensitivity of the algorithms to the complexity of the liabilities, two types of liabilities are used, one with a single pool of 0% TMG and the other with three pools of Minimum Rate Pools (MRP).

To make predictions on a maximum of economic contexts and to check the efficiency of the bond shock generator and the asset reallocation tool coded previously, five DataBases (DBs) have been implemented.³¹ Among the DBs implemented there are:

- A Database with bond shocks and continuous ESG
- A Database with bond shocks and varying ESG
- A Database with varying ESG
- A Database with varying ESG and bond inputs
- A Database with varying ESG, bond shocks and dynamic asset reallocation

The BELs were first predicted with a liability to one pool of MRPs and then with three pools of TMGs for each DB.

For the exploration and selection of variables, no variable selection was performed on the seven bond inputs.

The dataset of market variables produced by XSG was separated into 6 parts from the same stochastic ESG. The data visualization of the variables was done by linear and rank correlation analysis³².

The selection of variables reduced the number of explanatory variables from 2304 to 130 inputs.

Model to evaluate the BEL:

To perform cross-validation of our results, the Dataset was split into two parts: the test dataset and the train dataset on an 80/20% split.

The cross-validation was performed using a k-fold method with an optimal reduction of the mean of the absolute deviations by BE quantiles.

²⁹ See France assurance (2021). *L'assurance Française, Données clefs 2020*

³⁰ See part 3, section 3.3.2.7

³¹ See part 3, section 3.3.2.9

³² See part 4 and annex part 6, chapters 4.1.2.3 and subpart 6.12

Such as:

$$\text{Mean absolute average} = \frac{1}{nb_{\text{quantile}}} * \sum_{i=0}^n |Ecart_quantile_i|$$

with:

$$Quantile_deviation_i = \frac{Quantile_model_i - Quantile_prediction_i}{Quantile_model_i}$$

To avoid under-learning or over-learning, the mean of the deviations was calculated on 20 quantiles of prediction.

The results per quantile for the test and French datasets for the Database with varying ESG, bond shocks and dynamic reallocation with three TMG pools are presented below:

Test dataset

Modeling Relative Deviation	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	> 0,1%	> 0,1%	0,3%	-0,6%
19%	-0,1%	0,0%	-0,1%	-0,2%
29%	> 0,1%	> 0,1%	> 0,1%	-0,1%
39%	> 0,1%	> 0,1%	> 0,1%	-0,1%
49%	0,1%	> 0,1%	-0,1%	-0,1%
59%	-0,1%	> 0,1%	-0,1%	-0,4%
69%	0,1%	> 0,1%	0,1%	-0,3%
79%	> 0,1%	-0,1%	-0,4%	0,6%
89%	-0,1%	0,2%	> 0,1%	1,9%
99%	0,7%	-1,3%	0,3%	-4,2%

Figure 14 Results by quantile for the DB5 with 3 pools of minimum rate, test dataset

French dataset

Modeling Relative Deviation	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	-0,5%	-0,5%	-0,4%	-0,8%
19%	-0,2%	-0,3%	-0,3%	-0,4%
29%	-0,2%	-0,2%	-0,2%	-0,2%
39%	0,1%	0,1%	0,1%	-0,1%
49%	0,3%	0,3%	0,3%	-0,2%
59%	0,2%	0,1%	0,1%	-0,1%
69%	> 0,1%	> 0,1%	-0,1%	-0,1%
79%	-0,1%	-0,2%	-0,3%	0,6%
89%	0,4%	0,3%	0,1%	2,0%
99%	4,9%	0,5%	1,4%	-4,5%

Figure 15 Results for the DB5 with 3 pools of minimum rate, French dataset

Apart from the linear regression algorithm, the three other predictive algorithms have deviations in the target set at about 1% maximum. Two of them (XGboost and Gradient Boosting) kept very good results when changing database.

These results being representative of the other results presented for the other databases³³, the implemented tools predict good results for predicting BEs with relatively simple asset and liability portfolios.

³³ See part 4, sub parts 4.1 and 4.2

An additional study on interest rate and spread SCRs and their aggregation was performed³⁴.

The result of a bond spread SCR for a DB with bond shocks, varying ESG and asset reallocation with a TMG pool gives the following result:

Modeling SCR of spread	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	3 398	3 303	2 801	2 474	3 374
Deviation	0,00%	-2,80%	-17,57%	-27,19%	-0,70%

Figure 16 Prediction of the spread SCR for a DB with varying ESG, moving asset allocation, bonds shocks and 1 pool of MR

This study showed that the line-by-line predictions could be improved and that despite good results on simple bases, the model requires some complexities to have good predictions for more complex models.

The initial and additional study can be improved by learning scenarios based on datasets with upward or downward rate shocks.

The outcome of this study is positive with satisfactory BE and spread SCR prediction for the majority of the DBs discussed. To improve the predictions of the interest SCR, it will be necessary to make the ML algorithms learn datasets with rate shocks.

³⁴ See part 4, subpart 4.3

Introduction

Par son cycle de production inversé, les assureurs se doivent d'utiliser des outils statistiques et informatiques pour approximer la meilleure estimation de leurs engagements aussi appelée Best Estimate (BE).

Le secteur de l'assurance étant à forts enjeux stratégiques dans l'univers économique actuel, de nombreuses réglementations ont été mises en application afin de limiter toutes ruines des compagnies.

Les premier et troisième piliers³⁵ de la réglementation S2 imposent des fortes exigences quantitatives et de reporting pour les organismes d'assurances.

Cette réglementation impose la mention du bilan économique des compagnies d'assurance dans ses exigences quantitatives et de reporting.

Le pilotage d'un bilan économique nécessite de longs délais de calcul et le recours à de nombreuses simulations (supérieur à 1000 trajectoires dans le cadre d'un scénario stochastique).

Dans un contexte de fortes incertitudes économiques (Recul de 20% de l'indice S&P500 entre Janvier et Juin 2022³⁶), les assureurs souhaitent réduire leur temps de calcul pour mieux piloter leur BEL S2. Ce gain de temps permettrait de mieux anticiper les réallocations d'actifs à effectuer en cas de mouvements brutaux sur les marchés, ou en cas de changements de comportement des assurés.

Les exigences quantitatives S2

Le principe quantitatif de la norme S2 va être de simuler la projection de l'actif et du passif sur 50 ans avec les hypothèses suivantes :

- Run non central (différents chocs effectués sur les actifs et le passif durant les simulations).
- Condition de risque neutre (tous les actifs rapportent le taux sans risque fourni par l'EIOPA).
- Caractérisation par « Run off » : l'évolution liée aux contrats actuels uniquement (pas d'ajout de contrats futurs) est suivie
- Scénarios économiques générés sur plusieurs milliers de trajectoires qui convergent vers l'UFR (Ultimate Forward Rate).
- Projection sur 50 ans.

Le Best Estimate Liability (valeur actualisée des engagements futurs)³⁷ est calculé comme indiqué ci-dessous :

- Prestations (+)
- Primes (-)
- Frais (+)
- Rétro-commissions (-)
- Commissions (+)
- Autres (+)

³⁵ Voir partie 1, chapitre 1.2.1

³⁶ Le Monde (2022). *Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux*

³⁷ Voir partie 1, chapitre 1.2.2

Les sensibilités calculées par les simulations stochastiques en comparaison à la trajectoire déterministe du modèle permettent d'évaluer le coût des options de l'assureur à son passif plus communément appelés TVOG.

Pour calculer ces résultats, les assureurs utilisent des logiciels de simulations coûteux en temps. Afin d'optimiser leur processus de pilotage, les assureurs souhaitent surfer sur la révolution du digital et une utilisation standardisée du big data.

En 1956, lors d'une conférence réunissant une vingtaine de statisticiens, mathématiciens et d'informaticiens à l'université de Darmouth naquit le concept d'Intelligence Artificiel (IA).

A l'instar du perceptron informatique³⁸, base du machine et deep learning inventé par le Psychologue Frank Rosenblatt en 1957, de nombreux concepts d'intelligence artificielle ont été mathématiquement démontrés de longue date.

La puissance des calculateurs ne suivant pas les avancées mathématiques, il faut attendre les années 2010 pour voir une standardisation du big data et de l'intelligence artificielle.

Cette avancée s'est faite grâce à la montée exponentielle de la puissance de calcul des ordinateurs vendus sur le marché, ainsi que l'abondance des données à disposition des entreprises.

Ce mémoire a pour but d'utiliser les méthodes de deep learning, de Machine Learning et d'optimisation³⁹ pour raccourcir le temps de production d'un BEL sous la norme S2.

L'objectif est d'estimer le BE et le SCR de spread d'un bilan S2 avec un écart maximum d'environ 1%. L'étude a ensuite été étendue à la prédiction d'un SCR de taux et de son agrégation avec le SCR de spread.

Dans un premier temps, le cadre de l'étude va être posé avec la présentation de la norme solvabilité II

Dans un deuxième temps, une présentation des outils statistiques et informatiques à dispositions sera faite.

Dans un troisième temps, l'implémentation d'un générateur de densité de chocs obligataires inspiré d'algorithmes d'IA, ainsi que son paramétrage retenu pour la construction des bases de données pour les prédictions Machine Learning seront présentés.

Dans un quatrième temps, une analyse des différentes bases de données sélectionnées ainsi que la vérification de la robustesse des algorithmes prédictifs aux changements de dataset seront effectuées.

³⁸ Voir partie 2, chapitre 2.2.3.6

³⁹ Voir parties 2 à 4

Chapitre 1

1 Cadre de l'étude

Dans ce chapitre, le cadre du mémoire va être présenté. Dans un premier temps, ce chapitre présentera brièvement le fonctionnement d'un contrat euro et le contexte économique de l'assurance vie. Dans un deuxième temps, cette partie se focalisera sur la norme Solvabilité II et ses exigences quantitatives avec un focus sur le calcul du SCR obligataire. Dans un troisième temps, cette partie s'intéressera à la modélisation d'environnements économiques à l'aide d'une étude du fonctionnement des GSE.

1.1 Assurance vie et contrats euro

Lorsqu'un assuré souscrit un contrat euro, celui-ci s'engage à confier un montant à l'assureur sous forme de prime unique ou périodique. En échange l'assureur va garantir des engagements périodiques (le plus souvent annuels) envers le bénéficiaire du contrat. L'assureur va investir les primes de ses assurés sur les marchés pour répondre à ses engagements, sans pour autant chercher à battre le marché (« gestion en bon père de famille »).

Le service ALM a pour but de veiller à une saine gestion de l'actif et du passif en accord avec leurs différents mouvements. Pour veiller au pilotage de l'équilibre de l'actif sur le passif qui peut s'avérer compliqué, le service ALM doit produire des résultats issus de scénarios fixés par le régulateur.

Les scénarios imposés par le régulateur présentent une exhaustivité maximale des interactions entre les actifs et le passif, nécessitant la mise en place d'un processus de production chronophage.

Les résultats sortis par les processus de productions sont dépendant des hypothèses économiques initiales des marchés, des comportements des assurés, de la mortalité des assurés...

Dans un contexte économique fortement volatile, tant sur les marchés financiers ⁴⁰, qu'au niveau du retour de l'inflation ⁴¹, après une longue période de taux bas voire négatifs⁴² ; il est de plus en plus compliqué pour les assureurs de définir une stratégie de pilotage ALM pour répondre à leurs engagements.

Lors du pilotage des actifs et du passif, les assureurs doivent répondre aux exigences fixées par la norme solvabilité 2.

⁴⁰ Voir article du Monde « Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux »

⁴¹ Indice des prix à la consommation - résultats provisoires (IPC) - mai 2022

⁴² Voir ECB (2022). *Statistical Data Warehouse*

1.2 Solvabilité 2

Cette sous partie décrit les principaux éléments de la norme S2 avec un focus sur le bilan économique et le BEL qui correspond à la principale variable à prédire.

1.2.1 Généralité

La Directive Solvabilité II a été adoptée en 2009 et mise en œuvre en 2016 pour tous les assureurs présents sur la zone de l'Union Européenne.

Cette norme se présente sous la forme de 3 piliers, le pilier quantitatif, le pilier qualitatif et le pilier d'exigence de reporting.

Ci-dessous un schéma des trois piliers :

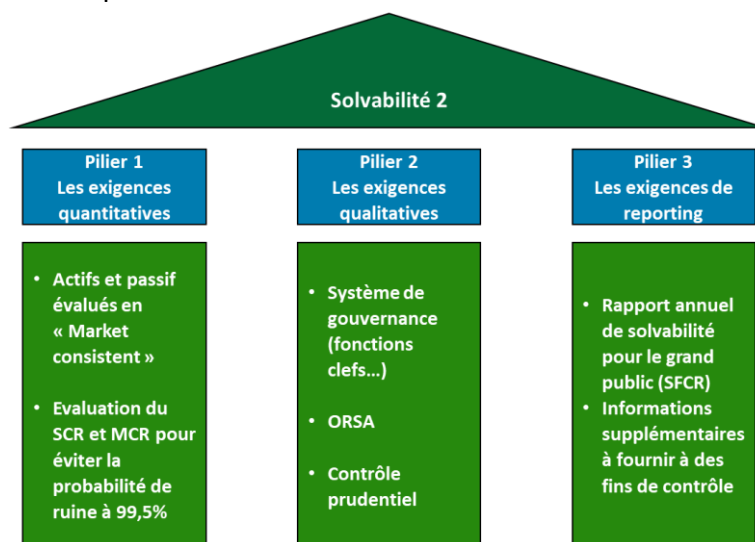
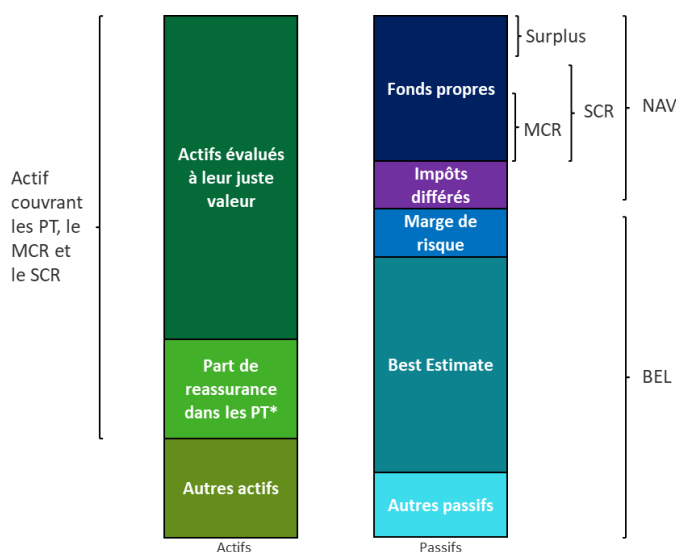


Figure 17 Schéma représentant les trois piliers de S2 (source réavis de S2 de la FFSA)

La suite de l'étude s'intéressera au pilier quantitatif de SII.

1.2.2 Bilan économique Solvabilité II

Cette norme s'appuie sur une vision économique du bilan. Le schéma ci-dessous représente le bilan économique sous Solvabilité II :



43

⁴³ Voir acronymes, annexe partie 6, chapitre 6.1

1.2.2.1 Les actifs

Les actifs sont exprimés en valeur de marché (aussi appelée juste valeur ou Fair Value en anglais). La part de réassurance dans les provisions techniques est mise à l'actif :

- Actifs exprimés en valeur de marché (ou juste valeur)
- Part de réassurance dans les PT (Provisions Techniques)
- Autres actifs

1.2.2.2 Le passif

Le passif se décompose en 5 grandes parties :

- Les fonds propres
- Les Impôts différés
- La marge de risque
- Le Best Estimate qui correspond à l'estimation des provisions techniques sous S2
- Autres types de passif

Le Best Estimate Liability

Le BEL correspond à la projection des flux liés aux contractions du portefeuille ainsi qu'à une actualisation des flux au passif à une courbe risque neutre à choisir parmi les options citées ci-dessous :

- La méthode EIOPA : Valeur Actuelle Probable des engagements simulés en « run off » (pas de choc de taux) en scénario risque neutre.
- La méthode indirecte : qui consiste à calculer d'autres éléments du passif et en déduire le BEL.

La simulation en scénario risque neutre permet de simplifier la modélisation de certains risques complexes tels que les risques de défaut et de crédit en :

- Simulant les rendements chaque année au taux sans risque
- En indexant le taux d'actualisation sur la courbe du taux sans risque différente pour chaque simulation.

1.2.2.3 Les fonds propres

Afin d'éviter la ruine (d'être en non capacité de respecter ses engagements à hauteur de 99,5% ou plus), les assureurs doivent décomposer leurs fonds propres de la manière suivante :

- Le MCR (Minimum Capital Requirement) calculé de manière simple et claire, il correspond à la value at risk (estimation des pertes qui ne devrait pas être dépassées sauf événement extrême) des fonds propres à un seuil compris entre 80 et 90% à horizon 1 an.
- Le SCR calculé à l'aide de chocs exercés sur les actifs qui vont être répercutés sur le BE au passif. Le SCR va correspondre à la différence de la NAV (net asset value) choquée et non choquée. Les scénarios de chocs sont paramétrés de telle sorte à ce que cette différence de NAV appelée SCR permette de provisionner un seuil de FP permettant d'éviter une ruine de l'assureur à 99,5% à minima.

- Un possible surplus FP qui permet de payer des « dividendes » aux actionnaire (assurance) ou une meilleure rémunération des contrats des adhérents (mutuelle).
- Il est aussi possible de déduire les fonds propres à l'aide de la formule suivante : $FP = \text{Actif} - \text{Engagements assureurs et surplus réglementaires (BEL+RM+Impôts différés)}$.

1.3 Le calcul du SCR

Dans cette sous partie, les étapes du calcul du SCR sont décrites, avec un focus sur le SCR de spread prédit en fin d'étude.

1.3.1 Les différentes étapes du calcul du SCR

Le SCR se calcule à l'aide d'un logiciel de simulation actuariel (Prophet dans le cadre de ce stage) selon les étapes suivantes :

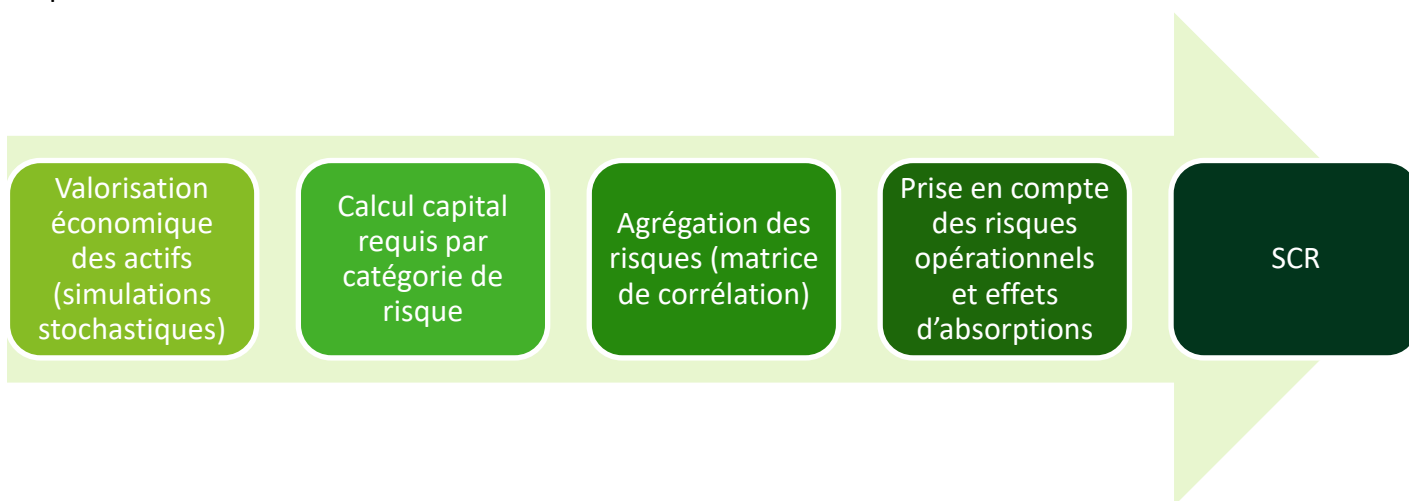


Figure 19 Schéma du calcul du BSCR (Source Deloitte)

1.3.2 Les enjeux du calcul de la TVOG et des BELs par des scénarios stochastiques

Les contrats d'assurances vies peuvent être la plupart du temps comparés à des instruments financiers, par exemple un contrat épargne rapportant un TMG de x% pourra être comparé à une option de vente⁴⁴(put), un produit dérivé très répandu sur les marchés financiers.

Les assureurs doivent évaluer le coût de ces options dans différentes conditions de marchés, leurs risques pouvant être comparés à des volatilités⁴⁵.

En finance de marché, la volatilité est une variance empirique de Pearson des rendements des actifs sur un portefeuille détenu.

Par exemple, la volatilité d'un portefeuille X contenant n actifs de rentabilité R_i peut-être exprimée de la sorte :

$$\sigma(R(X)) = \frac{1}{n} \sum_{i=1}^n (R(X_i) - E(R(X_i)))^2$$

Avec :

⁴⁴ Voir annexe partie 6, chapitre 6.2.3

⁴⁵ Les principes financiers cités dans ce mémoire (volatilité, instruments financiers, choc de spread...) sont décrits dans le mémoire d'Hélène LASSARTESES (2013). *Intérêt de l'investissement en actions protégées dans un contrat d'épargne*. ESSEC

$$E(R(X_i)) = \frac{1}{n} \sum_{i=1}^n R(X_i)$$

En ALM, lors de la simulation du passif, le calcul de la Time Value Option Garantie (TVOG) correspond au calcul de la volatilité du passif.

La TVOG est égale à la différence entre le BEL déterministe et l'espérance empirique des BEL stochastiques :

$$TVOG = BEL_{déter} - E(BEL_{sto})$$

La TVOG correspond à une quantification du risque produit par le passif de l'assureur.

Les BELs seront calculés à l'aide d'une méthode d'évaluation du passif qui prendra notamment en compte :

- Les mouvements du passif et des actifs pour chaque trajectoire
- Les hypothèses de calculs des différents chocs SCR
- La valeur des stress tests effectués au préalable

1.3.3 Description des différentes étapes du calcul du SCR

La première étape du calcul du SCR consiste à générer un scénario stochastique des actifs et du passif. Il s'agit d'un scénario dans lequel de nombreuses trajectoires d'actifs (plus de 1000 trajectoires en général) seront simulés selon différents scénarios économiques. Ces trajectoires permettront d'évaluer les actifs en valeur de marché dite « juste », et d'ajouter une provision proportionnelle à la volatilité des actifs et du passif (TVOG).

La TVOG est schématisé ci-dessous :



Figure 20 Représentation de la TVOG (Source Deloitte)

$$TVOG = BEL_{déter} - E(BEL_{sto})$$

La deuxième étape du calcul du SCR consiste à calculer les SCR des actifs et du passif selon des critères fixés pour chaque catégorie d'actif et de passif.

Les principaux modules de calcul du SCR sont mentionnés dans le schéma ci-dessous :

Risques de marché	Risques souscription vie	Risques de souscription non-vie	Risques de souscription santé
Actions	Mortalité	Prime et réserve	Mortalité
Taux	Longévité	Rachat	Longévité
Change	Dépense	Catastrophe	Invalidité
Immobilier	Révision		Dépense
Spread	Rachat		Révision
Concentration	Catastrophe		Prime et Réserve
Prime d'illiquidité			Rachat
			Catastrophe

Figure 21 Sous modules du calcul du SCR (source Deloitte)

Afin de calculer le SCR à hauteur de 99,5% à l'horizon de 1 an, la norme solvabilité 2 prévoit de faire varier la valeur de marché des actifs en appliquant des chocs à la baisse ou à la hausse. Par exemple pour un actif à composante action, un choc de 39% pour des actifs côtés OCDE, ou de 49% sinon est simulé.

Le schéma suivant illustre l'obtention d'un SCR marché pour une action côté OCDE :

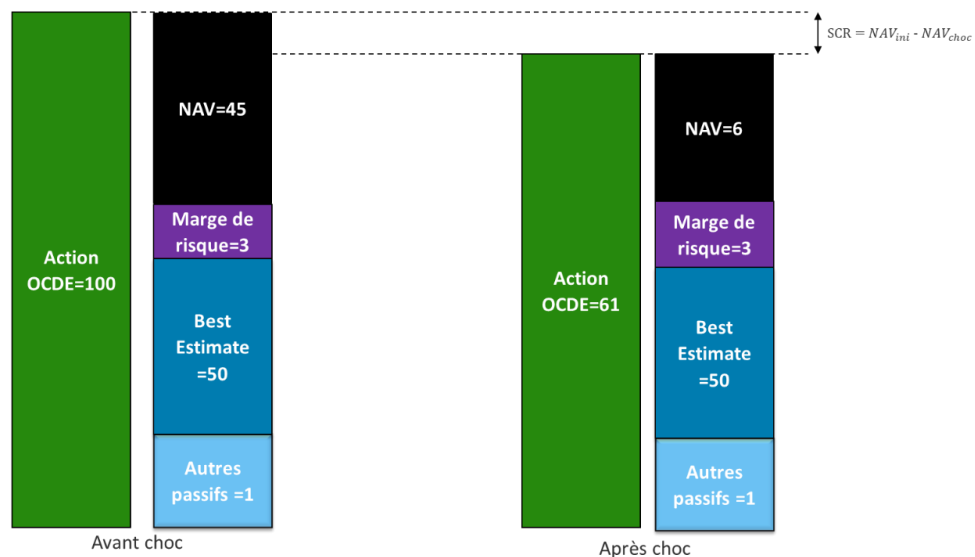


Figure 22 Représentation du choc action (Source Deloitte)

La troisième étape consiste en l'agrégation des différents SCR de chaque catégorie et sous-catégorie à l'aide de matrice de corrélation. Des produits matriciels sont effectués entre une matrice de corrélation intra module et les SCR des sous modules. Par exemple, lors de l'étape d'agrégation le SCR_{Mkt} (SCR de marché) sera agrégé de la manière suivante :

$$SCR_{Mkt} = \sqrt{(SCR_{rate}, SCR_{equity}, SCR_{Immobilier}, SCR_{Spread}) \cdot Corr_{Mkt}^T \cdot (SCR_{rate}, SCR_{equity}, SCR_{Immobilier}, SCR_{Spread})}$$

Avec :

SCR_{rate} : Le SCR de taux

SCR_{equity} : Le SCR action

$SCR_{Immobilier}$: Le SCR Immobilier

SCR_{Spread} : Le SCR spread

$Corr_{Mkt}^T$: La transposée de la matrice d'agrégation de marché

Avec la matrice d'agrégation de marché présentée ci-dessous

CrrMkt	Interest	Equity	Property	Spread	Currency	Concentration
Interest	100%	0%*	0%*	0%*	25%	0%
Equity	0%*	100%	75%	75%	25%	0%
Property	0%*	75%	100%	50%	25%	0%
Spread	0%*	75%	50%	100%	25%	0%
Currency	25%	25%	25%	25%	100%	0%
Concentration	0%	0%	0%	0%	0%	100%

Figure 23 Matrice de corrélation du SCR marché

(*) Dans le cas où le taux est à la hausse 0 sinon 0.5

La quatrième étape consiste à effectuer des ajustements sur les SCR prévus dans les étapes précédentes.

Elle se décompose en 3 grandes parties :

- Le calcul du BSCR (Brut Solvency Capital Required)
- La prise en compte des effets d'absorptions
- La prise en compte des risques d'absorptions

Où le BSCR est calculé à l'aide de matrice de corrélation intergroupe comme suit :

$$BSCR = \sqrt{\sum_{i,k} (Corr_{i,k} * SCR_i * SCR_k)} + SCR_{int}$$

Où :

- $Corr_{i,k}$ représente le coefficient de corrélation pour les éléments i et k de la liste d'agrégation ;
- SCR_i et SCR_k représentent les exigences de capital pour les éléments i et k de la liste d'agrégation, respectivement.
- SCR_{int} représente l'exigence de capital pour risque lié aux immobilisations incorporelles.

Le calcul des effets d'absorptions permet dans certains cas de revoir la valeur du BEL à la baisse ou à la hausse. Dans le cas d'une baisse conséquente de la production financière les effets d'absorptions sont revus à la baisse à cause d'une variation décroissante de la PPB.

Les exigences réglementaires imposées par la redistribution de 85% de la production financière aux assurés impliquent une décomposition du BEL en deux parties :

- Une partie fixée à la distribution des TMG
- Une partie dépendante de la production financière de l'assureur souvent nommée Future Discretionary Benefits (FDB).

La BEL peut alors être revue comme suit :

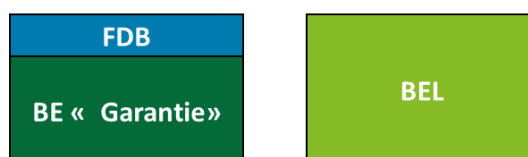


Figure 24 Représentation du FDB (Source Deloitte)

Avec la FDB, les Future Discretionary Benefits qui représentent les cash-flows induits par la PPB/PPE. Pour contraster des pertes ou des gains induits par une forte volatilité des marchés sur des actifs long terme, un paramètre nommé « effet Dampener » a été mis en place pour réduire les besoins SCR en cas d'anticipations des marchés à la hausse.

1.3.4 La méthode market consistency et le SCR obligataire

Afin de répliquer un actif de type obligataire en situation de risque neutre, une risque neutralisation de cet actif s'impose. Une risque neutralisation correspond à une simulation des rendements d'un actif en situation de risque neutre en appliquant la méthode de « Market Consistency ».

Il existe différentes méthodes pour effectuer la Market consistency :

- Faire varier le montant des coupons théoriques
- Faire varier le taux sans risque S2
- Mélanger ces deux manœuvres

Le but de cette méthode est que $VM_{theorique_RN} = VM_{réelle}$

La valeur « réelle » d'une obligation correspond à son prix sur le marché et la valeur théorique d'une obligation se calcule de la manière suivante :

$$VM_{theorique} = \sum_i \frac{CF_i}{(1 + r_i)^i}^{46}$$

Dans la pratique la risque neutralisation peut s'effectuer par :

- L'ajout d'une prime d'illiquidité à la courbe des taux
- L'ajout d'un coefficient constant à tous les cashflows
- L'ajout d'un coefficient constant appliqué uniquement aux coupons
- L'ajout d'un coefficient décroissant appliqué à tous les cashflows

A noter que les lignes générées par la méthode market consistency doivent être une martingale⁴⁷.

1.3.5 Modélisation du risque de spread

Le risque de spread fait intervenir des notions financières obligataires, de prime de liquidité et de crédit.

Pour calculer le SCR de spread, il faut calculer trois SCR faisant intervenir les notions suivantes :

$$SCR_{spread} = SCR_{Obligataire} + SCR_{struct} + SCR_{cd}$$

Avec :

- SCR_{struct} qui représente le SCR des produits structurés de crédit
- $SCR_{Obligataire}$ qui représente le SCR des obligations
- SCR_{cd} qui représente le SCR des propres dérivés

Pour calculer un SCR obligataire, un choc de duration et de taux doivent être appliqué selon la nature et la notation de l'obligation.

⁴⁶ Notations exprimées en annexe partie 6, chapitre 6.3, Glossaire numérique

⁴⁷ Voir partie 1, chapitre 1.4.7

1.3.6 Le SCR bond ou obligataire

Les SCR_{struct} et SCR_{cd} étant compliqués à modéliser, ils ne sont pas considérés par la suite.

La modélisation du SCR bond ou obligation en français sera effectuée par le calcul d'un facteur de risque $Stress_i$ dépendant de la durée modifiée de l'obligation exprimé en années (Dur_i) et de son échelon de crédit, Dur_i n'est jamais inférieure à 1.

Les obligations pour lesquels une évaluation par une agence de notation est disponible, se voient attribuer un facteur de risque $Stress_i$ en fonction de leur échelon de qualité de crédit.

Le risque de spread correspond au risque de variation, à la hausse, des spreads de crédit par rapport à la courbe des taux d'intérêts sans risque, correspondant à une hausse de la probabilité de défaut du portefeuille obligataire. Ces variations viennent impacter les valorisations des obligations.

De ce fait, le SCR de spread dépend de la notation et de la durée de l'obligation.

Le SCR lié au risque de spread s'écrit alors :

$$SCR_{spread} = \max(NAV_{Centrale} - NAV_{Choc\ spread}; 0)$$

Avec :

NAV : Net-Asset-Value qui constituent un indicateur de l'exposition au risque d'un assureur tel que :

$$NAV = VM_{actifs\ simulés} - (BE + RM)$$

Avec :

$VM_{actifs\ simulés}$: Valeur de marché SII en fair value

BE: Best Estimate

RM : Risk Margin

Ci-dessous le tableau des échelons de crédits permettant de calculer les valeurs de « stress » selon la formule standard :

Echelon qualité de crédit		0		1		2		3		4		5 et 6	
Duration	Stress i	Ai	bi	ai	bi	ai	bi	ai	bi	ai	bi	ai	bi
$dur_i < 5$	$b_i * dur_i$	0,0%	0,9%	0,0%	1,1%	0,0%	1,4%	0,0%	2,5%	0,0%	4,5%	0,0%	7,5%
$5 < dur_i < 10$	$a_i + b_i * (dur_i - 5)$	4,5%	0,5%	5,5%	0,6%	7,0%	0,7%	12,5%	1,5%	22,5%	2,5%	37,5%	4,2%
$10 < dur_i < 15$	$a_i + b_i * (dur_i - 10)$	7,0%	0,5%	8,4%	0,5%	10,5%	0,5%	20,0%	1,0%	35,0%	1,8%	58,5%	0,5%
$15 < dur_i < 20$	$a_i + b_i * (dur_i - 15)$	9,5%	0,5%	10,9%	0,5%	13,0%	0,5%	25,0%	1,0%	44,0%	0,5%	61,0%	0,5%
$20 < dur_i$	$\text{Min}(a_i + b_i * (dur_i - 20); 1)$	12,0%	0,5%	13,4%	0,5%	15,5%	0,5%	30,0%	0,5%	46,5%	0,5%	63,5%	0,5%

Figure 25 Chocs de spread obligataire selon la formule standard S2 (Source Deloitte)

Ci-dessous le Tableau de conversion des ratings aux échelons de crédit :

S&P and Fitch	Moody's	Echelon de crédits (CQS)
AAA	Aaa	0
AA	Aa	1
A	A	2
BBB	Bbb	3
BB	Bb	4
B	B	5
<B	<B	6

Toute obligation non notée est passée à l'échelon 9 de crédit.

L'implémentation d'un générateur de chocs obligataires inspiré du principe du choc de spread obligataire sera exprimé dans la suite de l'étude en partie 3.

1.4 Le fonctionnement d'un modèle de simulation actuariel et des GSE

Un logiciel de simulation permet de simuler de nombreuses trajectoires stochastiques à l'aide d'hypothèses économiques (comportement des assurés, type d'écoulement des contrats etc...); et des paramètres financiers mis en inputs (portefeuilles d'actifs, frais de prélèvements sur encours ...). Une fois toutes ces informations renseignées, le logiciel de simulation va calculer pour chaque trajectoire les engagements de l'assureur envers ses clients (BEL) et son SCR associé.

Pour calculer des trajectoires stochastiques et leurs TVOG associées, il est nécessaire d'explicitier des hypothèses économiques supplémentaires.

Ces hypothèses supplémentaires vont être créées à l'aide d'un générateur de scénarios économiques (GSE).

Ces GSE vont être connectés dans un dossier d'inputs séparé du dossier initial.

Dans le cadre de ce mémoire, les GSE vont être produits à l'aide de différentes méthodes stochastiques dont le LMM-DDSV, le modèle de Hull-White à un facteur, le modèle de Heston et le modèle de Black and Sholes.

Une fois les inputs connectés au logiciel de simulation actuariel, le modèle va calculer les engagements de l'assureur par un ensemble de calculs. Les résultats des simulations sont sortis en output.

Le déroulement du calcul d'un logiciel de simulation actuariel peut-être représenté ci-dessous :

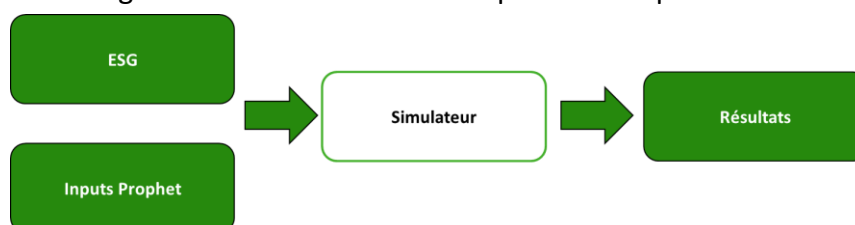


Figure 27 Schéma des inputs et outputs d'un logiciel de simulation actuariel

1.4.1 La différence entre le taux spot et le taux forward

Le taux spot représente un taux d'actualisation à court terme, dans la suite du mémoire la courbe des taux spots fournie par l'EIOPA⁴⁸ est utilisée.

Le taux forward représente un taux plus long terme sur plusieurs années (ou s'écoulant entre deux périodes définies), les taux forward sont utilisés lors de la construction des courbes ZCB d'un GSE.

⁴⁸ Les taux sont exprimés annuellement

1.4.2 Description de quelques modèles GSE

La description des modèles stochastiques s'est appuyée sur le document d'orientation des GSE publié par l'institut des actuaires.

1.4.3 Le modèle de Black and Sholes

L'équation différentiel du modèle de Black and Sholes se pose :⁴⁹

$$dF_t = \mu_t F_t dt + \sigma_t^2 F_t dB_t$$

Ce modèle va représenter l'évolution d'un actif selon l'espérance de ses rendements multipliés par le risque neutre pour représenter la partie « non risqué » de l'actif dans le temps. Sa variance (ou volatilité en langage financier) suit un mouvement brownien représentant la « prime de risque de l'actif ».

En considérant les conditions d'application pour appliquer la formule d'Itô comme respectées (F_t est de classe C^2) alors en découpant l'équation différentiel par morceaux suivante :

$$dF_t = \frac{\partial F_t}{\partial t}(t, B_t)dt + \frac{\partial F_t}{\partial B}(t, B_t)dB_t + \frac{1}{2} \frac{\partial^2 F_t}{\partial t^2} \sigma_t^2 dt^2 = \mu_t dt + \sigma_t^2 dB_t$$

L'équation du sous-jacent (le rendement immobilier) va être obtenue :

$$S_t = S_0 e^{(\sigma_t B_t + (\mu_t - \frac{1}{2}\sigma_t^2)t)}$$

Ce modèle possède les limites suivantes :

- Absence d'opportunités d'arbitrage
- Il est possible d'effectuer des ventes à découvert (Short selling)
- Pas de coûts de transactions
- Connaissance du taux sans risque
- Les sous-jacents sont parfaitement divisibles

Par son absence d'opportunité d'arbitrage, le modèle de Black and Sholes permet de représenter des actifs ayant des comportements de produits dérivés européens avec une exécution à échéance ; tels que les rendements immobiliers (bien qu'on émette l'hypothèse que les parts immobilières sont un actif liquide).

1.4.4 Le modèle de Heston

Le modèle de Heston s'appuie sur un modèle de volatilité stochastique implicite. A l'instar du modèle de Black and Sholes, le modèle simule le rendement d'un actif comme une partie non risquée (espérance des rendements de l'actif rapportant le taux sans risque), et d'une partie risquée reposant sur la volatilité des rendements. Au lieu d'être constante, la volatilité des rendements évolue dans le temps.

Le modèle de Heston est représenté de la manière suivante :

$$dS_t = \mu_t S_t dt + \sqrt{\vartheta_t} S_t dW_t$$

Avec :

$$d\vartheta_t = k(\theta - \vartheta_t)dt + \varepsilon\sqrt{\vartheta_t}dB_t$$

L'introduction d'une volatilité dans le temps va permettre de mieux modéliser les actifs de types actions dont la valeur est beaucoup plus volatile que des produits immobiliers.

⁴⁹ Les notations sont exprimées dans l'annexe partie 6, chapitre 6.3.2

1.4.5 Le modèle de Hull-White

Le modèle de Hull-White à un facteur s'appuie sur des modèles de type Heath, Jarrow et Morton (HJM). Ces modèles concernent les taux forward, c'est-à-dire les taux longue durée (supérieur à un an dans le cadre du mémoire).

Afin d'approximer ces taux longs, le modèle va toujours considérer une partie qui suit la continuité de l'évolution de l'environnement de taux « partie structurelle » et une partie plus volatile, plus sensible aux mouvements soudains de l'environnement de taux « partie conjoncturelle ».

Cette interprétation donne la structure d'évolution du taux suivante pour une maturité T et un incrément de temps t donné :

$$df_t = \alpha(t, T)dt + \sigma(t, T)dW_t$$

Avec W_t un mouvement Brownien.

Si les obligations 0 coupons sont considérées comme les actifs les moins risqués du marché, dans un modèle de temps discret et en absence d'opportunités d'arbitrage, la valeur d'une obligation de ce type peut être posée de la manière suivante :

$$VM_{t,T} = \sum_{i=t}^T CF_i * e^{-r_t * i}$$

Si un passage d'un intervalle de temps discret à un intervalle de temps continu est admis, alors par linéarité de la somme en intégrale, l'équation suivante est obtenue :

$$VM_{t,T} = \int_t^T CF_u * e^{-r_t * u} du$$

Cette intégrale peut être exprimée comme le type de fonction suivante (car intégrale d'une fonction exponentielle) :

$$VM_{t,T} = A_{t,T} * e^{-r_t * B_{t,T}}$$

Avec $A_{t,T}$ et $B_{t,T}$ deux fonctions déterministes dépendante du temps

De ce constat va être posée l'équation suivante :

$$dr_t = (\theta_t - \alpha * r_t)dt + \sigma_{t,T}dW_t$$

Par reconnaissance de fonction vont être obtenues $A_{t,T}$ et $B_{t,T}$:

$$A_{t,T} = (\theta_t - \alpha * r_t)dt \text{ et } B_{t,T} = \sigma_{t,T}dW_t$$

Avec l'hypothèse d'une volatilité évoluant dans le temps :

$$\sigma_{t,T} = \sigma e^{-\alpha * (T-t)}$$

θ_t Un paramètre dépendant de la courbe des taux forwards initiale.

α un paramètre de vitesse de retour à la moyenne.

Il faut faire attention à ne pas trop faire dépendre α et σ en fonction du temps car cela peut entraîner la construction d'une structure de volatilité pas adaptée aux travaux actuariels souhaités.⁵⁰

L'inflation faisant entrer en jeu une fonction structurelle et conjoncturelle d'un univers économique local, continental ou mondial évoluant a priori sans opportunité d'arbitrage, le modèle de Hull-White à un facteur a été retenu pour représenter les phénomènes d'inflation (variable ILZCB).

⁵⁰ Institut des ACTUAIRES (2018) Document d'orientation, les générateurs de scénarios économiques, éléments d'analyse et exemple de pratiques actuarielles applicables au marché français

1.4.6 Le modèle LMM

Le modèle LMM a pour but de décrire le comportement des taux forward sous la mesure risque neutre. Pour démontrer ce modèle, les conditions suivantes doivent-êre posées :

- Le prix d'une obligation Zéro coupon entre T sa maturité et t son pas de temps est noté $VM_{t,T}$
- L'opérateur $\tau_k = T_k - T_{k-1}$ avec T_k et T_{k-1} les pas de temps d'une série croissante définie de la manière suivante :
- $T_0 < T_1 < \dots < T_{k-1} < T_k$ avec k comme dernier incrément.

Si on pose les équations des valeurs de marché respectivement avec les taux forwards T_{k-1} et T_k :

$$VM_{t,T_{k-1}} = \frac{Nominal}{1+T_{k-1}} \text{ car l'obligation ne rapporte pas de coupon}$$

$$VM_{t,T_k} = \frac{Nominal}{1+T_k}$$

Alors :

$$\frac{VM_{t,T_{k-1}}}{VM_{t,T_k}} = \frac{1+T_k}{1+T_{k-1}}$$

$$\Leftrightarrow \frac{VM_{t,T_{k-1}}}{VM_{t,T_k}} - 1 = \frac{T_k - T_{k-1}}{1+T_{k-1}}$$

$$\Leftrightarrow \frac{1}{\tau_k} * \frac{VM_{t,T_{k-1}}}{VM_{t,T_k}} - 1 = \frac{1}{T_k - T_{k-1}} * \frac{T_k - T_{k-1}}{1+T_{k-1}} = \frac{1}{1+T_{k-1}} \sim T_k$$

De ce constat la fonction représentant le taux forward en t, $F_k(t)$ est posée de la manière suivante :

$$F_k(t) = \frac{1}{\tau_k} * \frac{VM_{t,T_{k-1}}}{VM_{t,T_k}} - 1$$

Si les taux forward sont considérés comme volatiles dans le temps selon une volatilité $\sigma_k(t)$, et que cette volatilité est simulée selon un mouvement Brownien multidimensionnel $Z_k(t)$, l'équation différentielle représentant l'évolution du taux sera :

$$dF_k(t) = F_k(t) * \sigma_k(t) * dZ_k(t)$$

Pour départager les effets structurels des effets conjoncturels de l'évolution des taux, la mesure Libor dans la dimension d va être posée :

$$\beta_d(t) = \frac{VM_{t,T_{\beta(t)-1}}}{\prod_{j=0}^{\beta(t)-1} VM_{T_{j-1},T_j}}$$

Avec $\beta_d(t)$ la plus grande valeur entre 1 et k tel que $t \leq T_{j-1}$ avec $\beta_0 = 1$ sous la mesure Libor spot en dimension d.

Le passage à cette mesure va permettre de découper $dZ_k(t)$ en deux parties :

$$dZ_k(t) = \sum_{j=\beta(t)}^k \frac{\tau_j * \rho_{j,k} * \sigma_j(t) * F_j(t)}{1 + \tau_j * F_j(t)} dt + dZ_k^d$$

Avec une variable qui correspond à une « corrélation entre les mouvements Browniens » :

$$\rho_{j,k} dt = dZ_j(t) * dZ_k(t)$$

Dans ce modèle il faut calibrer la fonction de volatilité $\sigma_k(t)$, posée de la manière suivante :

$$\sigma_k(t) = \varphi_k((a * (T_{k-1} - t) + d) * e^{-b(T_{k-1}-t)} + c)$$

Les principaux hyperparamètres seront a, b, c et d beaucoup plus de paramètres possibles à régler.

Afin de réduire certains biais inhérents à ce modèle, il peut être ajouté un shift de diffusion δ , sera donc obtenue la formule d'un modèle de type LMM-DDSV suivante :

$$dF_k(t) = (F_k(t) + \delta) * \sigma_k(t) * \left(\sum_{j=\beta(t)}^k \frac{\tau_j * \rho_{j,k} * \sigma_j(t) * F_j(t)}{1 + \tau_j * F_j(t)} dt + dZ_k^d \right)$$

Ce modèle complexe va être utilisé pour modéliser les taux forwards long terme correspondant au déflateur des Zeros Coupons Bonds (ZCB).

1.4.7 Le test de martingalité

Soit M_n un processus statistique définie comme une suite de nombre.

M_n est une martingale par rapport à une filtration F_n si pour tout entier naturel n si :

- M_n est adaptée à la filtration F_n
- M_n intégrable $\forall n \in \mathbb{N}$
- $E(M_{n+1}|F_n) = M_n$

Si M_n respecte les deux premières conditions et que $E(M_{n+1}|F_n) < M_n$, le processus étudié est une sur-martingale, et si au contraire $E(M_{n+1}|F_n) > M_n$, le processus est une sous-martingale.

Le test de martingalité repose sur le fait qu'en t=0, le sous-jacent d'un actif actualisé $S_t = VM_{actif}$ notée S_0 . Soit N le nombre de trajectoires générées et respectivement $D_{n,t}$ et $S_{n,t}$ les valeurs respectives du déflateur et du sous-jacent de l'actif pour la trajectoire n et l'incrément de temps t sont considérés comme une martingale si elles vérifient :

$$S_t = \frac{1}{N} * \sum_{n=1}^N D_{n,t} * S_{n,t} = S_0$$

En pratique pour qu'un test de martingalité soit vérifié, il faut que la moyenne des variables économiques pour chaque année soit égale à 1 avec +/-5% d'écart⁵¹.

Un test de martingalité est présenté ci-dessous pour les indices actions :

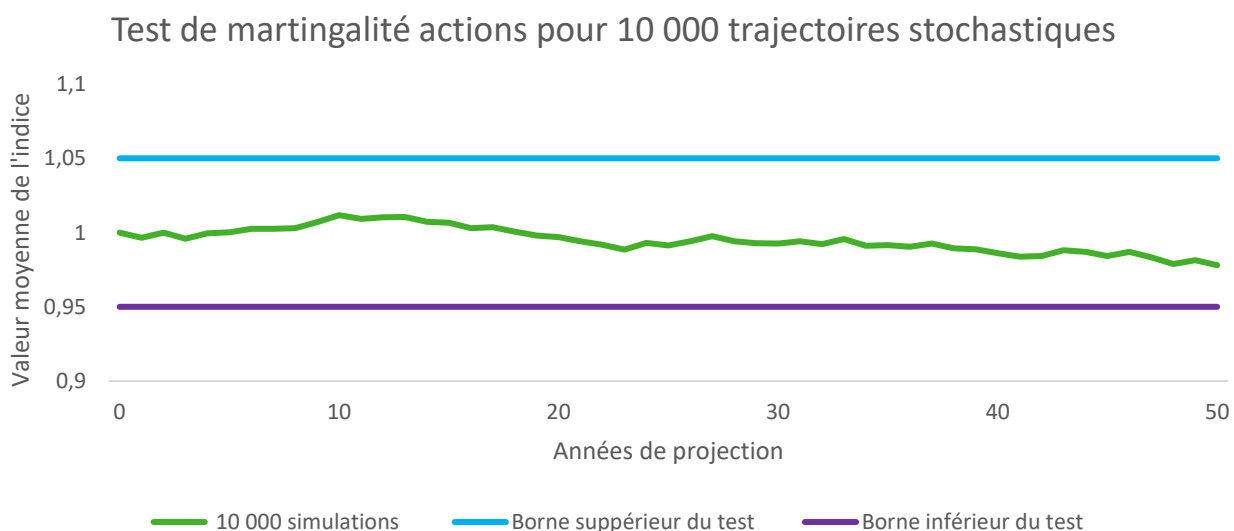


Figure 28 Graphique d'un test de martingalité action

⁵¹ Voir mémoire de Valentin SAVIDAN

Le test de martingalité est bien vérifié car la moyenne des indices sont compris dans les bornes 0,95 et 1,05 pour tous les pas de temps.

La partie suivante va s'intéresser aux outils statistiques et informatiques utilisés lors de l'étude.

Chapitre 2

2 Outils statistiques et environnement informatique

Dans un contexte de digitalisation massive des entreprises ainsi qu'avec l'émergence de nouveaux outils de traitements des données en masse (big data) de plus en plus puissants, les problématiques liées à l'analyse des données ainsi que leur traitement deviennent un enjeu majeur pour les compagnies d'assurance.

Dans ce contexte, de nombreux assureurs souhaitent tirer profit de leurs données à l'aide d'algorithmes basés sur l'intelligence artificielle. L'utilisation de ces algorithmes permettrait aux compagnies de gagner du temps dans leur pilotage ALM.

Le temps gagné grâce à ces nouvelles technologies permettrait aux assureurs d'être plus compétitifs. Déjà 31% des firmes d'assurance européennes utilisent l'intelligence artificielle.⁵²

Pour répondre à ces défis, ce mémoire fait l'objet d'un usecase concret de ces nouvelles technologies afin de concevoir un outil permettant de prévoir le BEL et le SCR de spread avec un écart de prédiction relatif d'environ 1% maximum.

2.1 Outils statistiques

Pour la réalisation de ce mémoire, un générateur de densité de chocs obligataires a été développé et implémenté sous Python. Les différentes lois statistiques utilisés par le générateur de densité de choc implémenté pour l'étude sont présentées dans cette partie⁵³.

Pour l'écriture de cette partie, les distributions ont toutes été normalisées à l'aide d'une normalisation min-max pour laquelle a été ajouté un facteur de dissymétrie $-0,5$ afin d'uniformiser les graphiques des densités et des répartitions.

La méthode min-max consiste à remettre à une échelle $[0,1]$ un échantillon de variable comme présenté ci-dessous :

$$X_{normalise} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

L'application du facteur de dissymétrie se fait de la manière suivante :

$$X_{normalise} = \frac{X - X_{min}}{X_{max} - X_{min}} - \text{facteur}_{dyssimétrie}$$

2.1.1 Présentation des principales lois statistiques utilisées pour la construction du générateur de spread aléatoire

Dans cette sous partie, les lois statistiques qui sont utilisées dans la construction des fonctions de distribution de chocs obligataires vont être présentées.

⁵² EIOPA (2019). ARTIFICIAL INTELLIGENCE GOVERNANCE PRINCIPLES: TOWARDS ETHICAL AND TRUSTWORTHY ARTIFICIAL INTELLIGENCE IN THE EUROPEAN INSURANCE SECTOR

⁵³ Les notations sont exprimées dans l'annexe partie 6, chapitre 6.5.3

2.1.1.1 La loi Uniforme

Le but de la loi uniforme est de simuler des événements aléatoires entre un intervalle donné, sa fonction de densité est représentée par :

$$U(a, b) \sim f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}$$

La fonction de densité de cette loi simulée sous Python pour un échantillon de 10 000 simulations sur un intervalle de simulation compris entre -0,5 et 0,5 a la forme suivante :

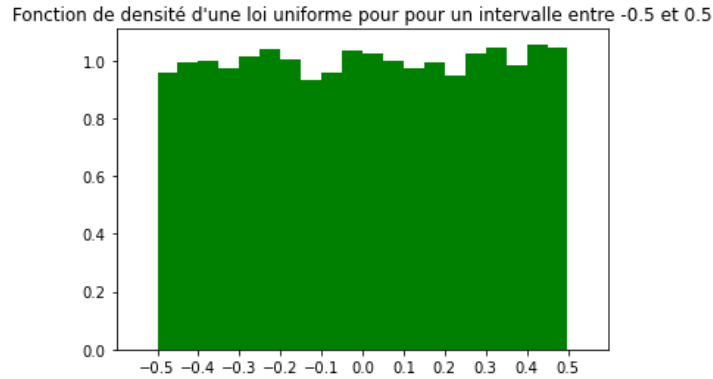


Figure 29 Densité d'une loi uniforme générée sous Python

Python faisant intervenir une notion de pseudo aléatoire et non pas de « parfaitement aléatoire », la fonction de densité représentée n'est pas exactement la fonction de densité d'une loi uniforme qui doit être un carré entre l'intervalle a et b.

La fonction de répartition de cette fonction prend la forme :

$$F(x) = \frac{x-a}{b-a} \mathbb{1}_{[a,b]}$$

2.1.1.2 La loi Normale

Le but de la loi normale est d'effectuer des simulations concentrées autour d'une moyenne et réparties selon un écart type. Cette loi admettra la fonction de densité suivante :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2} * \frac{(x-\mu)^2}{\sigma^2}}$$

La fonction de densité d'une loi normale $N(0,1)$ simulée sous Python pour un échantillon de 10 000 simulations :

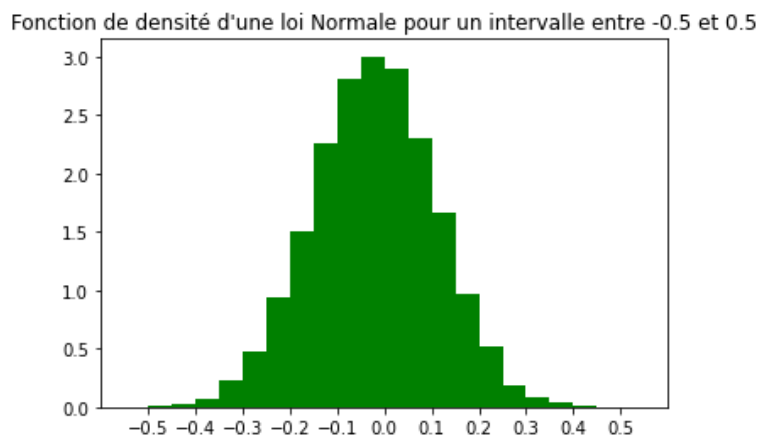


Figure 30 Densité d'une loi normale générée sous Python

La fonction de répartition d'une loi normale est obtenue en approximant l'intégrale de la fonction de densité de la loi :

$$F(x) = \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{1}{2} * \frac{(x-\mu)^2}{\sigma^2}} dx$$

2.1.1.3 La loi de Pareto

Le but d'une loi de Pareto est d'effectuer des simulations réparties de manière asymétrique à gauche d'une « moyenne ».

La fonction de densité d'une loi de Pareto s'écrit :

$$f(x)_{k,m} = \frac{k * m^k}{x^{k+1}}$$

La fonction de densité d'une loi de Pareto $P(10,1)$ simulée sous Python pour un échantillon de 10 000 simulations :

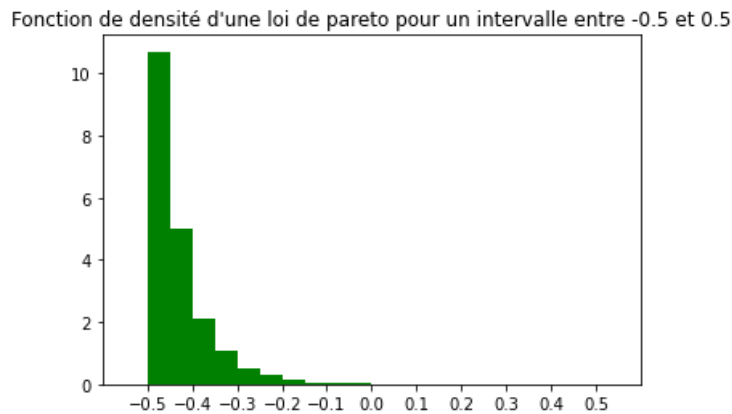


Figure 31 Densité d'une loi de Pareto générée sous Python

La fonction de répartition d'une loi de Pareto s'écrit :

$$F(x) = 1 - \frac{m^k}{x^k}$$

2.1.1.4 Loi de Weibull

Le but d'une loi de Weibull est d'effectuer des simulations réparties de manière asymétriquement à droite d'une « moyenne ».

La fonction de densité d'une loi de Weibull s'écrit :

$$f(x)_{\gamma,k} = \frac{k}{\gamma} * \frac{x^{k-1}}{\gamma^{k-1}} * e^{-\frac{x^k}{\gamma^k}}$$

Pour la simulation de 10 000 variables d'une loi de Weibull $W(1,10)$ sous Python, la densité suivante est obtenue :

Fonction de densité d'une loi de weibull pour des chocs entre -50 et 50%

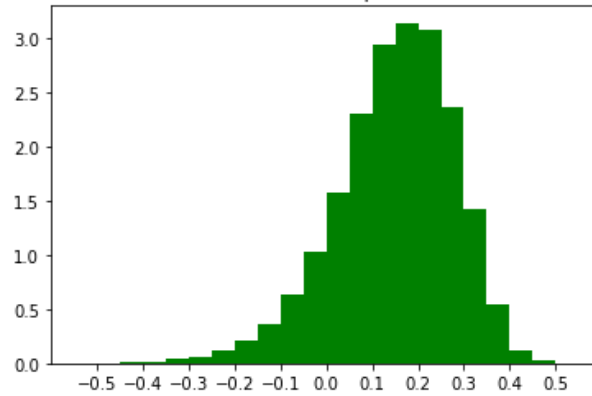


Figure 32 Densité d'une loi de Weibull générée sous Python

La fonction de répartition d'une loi de Weibull s'écrit :

$$F(x) = 1 - e^{-\frac{x^k}{\gamma^k}}$$

En utilisant la propriété sur les lois uniformes [0,1] tel que $X=G(U)$ alors, $F(X)=U$ le développement suivant est obtenu :

$$\begin{aligned} F(X) &= U \\ \Leftrightarrow 1 - e^{-\frac{x^k}{\gamma^k}} &= U \\ \Leftrightarrow -e^{-\frac{x^k}{\gamma^k}} &= 1 - U \end{aligned}$$

Par symétrie en 0.5 de la loi uniforme [0,1] :

$$1 - U = U$$

Donc :

$$\begin{aligned} \Leftrightarrow -e^{-\frac{x^k}{\gamma^k}} &= U \\ \Leftrightarrow -\frac{x^k}{\gamma^k} &= \ln(U) \\ \Leftrightarrow X &= \gamma * \ln(U)^{\frac{1}{k}} \\ X &= \gamma * \ln(U)^{\frac{1}{k}} \end{aligned}$$

La formule obtenue sera la formule utilisée dans la librairie Python pour effectuer les chocs aléatoires selon une loi de Weibull.

Cette formule plus simple à développer et à implémenter pour la machine se base sur une loi uniforme imparfaite (pseudo aléatoire) pouvant amener à quelques erreurs d'estimations de la loi.

2.1.1.5 Les mélanges de lois

Si plusieurs lois sont simulées sur le même intervalle de définition, alors il est possible de les mélanger, par exemple si les lois de Weibull et Pareto sont mélangées dans un échantillon de 20 000 simulations, la représentation graphique suivante de la densité sera obtenue :

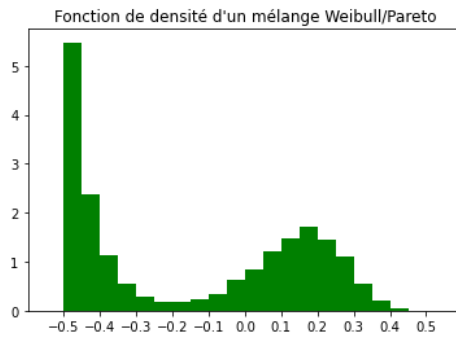


Figure 33 Densité d'un mélange d'une loi de Pareto/Weibull générée sous Python

Dans la suite de ce mémoire, ces lois et mélanges de lois seront implémentés dans un algorithme effectuant des chocs sur des paramètres obligatoires.

2.1.2 Présentation des autres outils et test d'adéquation statistiques utilisés pour la construction du générateur de densité et de la vérification des résultats générés

Dans cette sous partie, les méthodes d'échantillonnages et tests statistiques qui ont permis la construction et la vérification du bon fonctionnement du générateur des données vont être présentés.

2.1.2.1 La méthode bootstrap

La méthode bootstrap consiste à reproduire un ou plusieurs échantillons de données à partir d'un échantillon unique, cette méthode va être utilisée lors de l'étape de réattribution des poids des chocs (back propagation) dans l'algorithme prochainement présenté.⁵⁴

Le principe de fonctionnement de la méthode d'échantillonnage bootstrap consiste à effectuer n tirages aléatoires avec remise sur l'échantillon de départ.

Ci-dessous, La fonction de densité d'une loi uniforme bootstrapée sous Python pour un échantillon de 5 000 simulations sur un intervalle de simulation compris entre -0,5 et 0,5 :

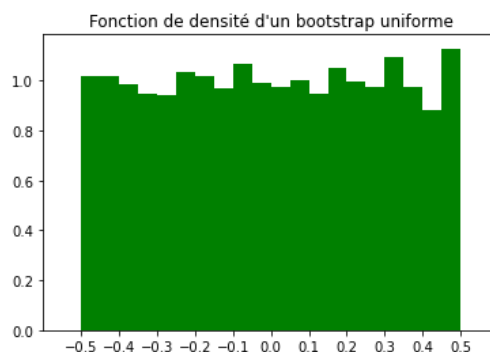


Figure 34 Densité d'un échantillon de loi uniforme généré par bootstrap

Cette courbe de densité correspond à vu d'œil à la fonction de densité d'une loi uniforme $U(-0,5 ; 0,5)$.

⁵⁴ Voir partie 3, sous partie 3.1.2

2.1.2.2 Le test de Kolmogorov-Smirnov⁵⁵

Un test d'adéquation statistique permet de vérifier si un échantillon suit une loi ciblée ou non.

Le test de Kolmogorov Smirnov permet de comparer deux lois et de vérifier si ces deux lois sont similaires ou non.

Ce test s'appuie sur deux hypothèses :

H0 : « Les deux lois ont une répartition différente »

H1 : « Les deux lois ont une répartition similaire »

Soit une fonction de répartition A prise comme référence qui va permettre de tester si la courbe B suit la même loi.

Le test de Kolmogorov-Smirnov va compter le nombre d'éléments moyens dans la courbe B inférieurs à un x fixé dans l'échantillon, et les diviser par le nombre d'éléments dans l'échantillon. Si le problème mathématique est posé, alors la statistique de test suivante sera posée :

$$\forall x \in \mathbb{R}, \omega \in \Omega, F_n(x, \omega) = \frac{\text{nombre d'élément} \leq x}{n} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i(\omega) \leq x}$$

En pratique, lorsqu'un test est effectué sous Python, il suffira d'observer la p-value du test.

La p-value correspond à la probabilité de rejeter l'hypothèse nulle (H0 décrite précédemment) pour un intervalle de confiance donné.

En général l'intervalle de confiance choisi est l'intervalle à 95%, cela signifie que si la p-value sortie par le test est inférieur à 0,05 alors, l'hypothèse H0 sera acceptée, elle sera rejetée sinon.

2.1.2.3 Le diagramme quantile par quantile⁵⁶

Le diagramme quantile par quantile permet de vérifier si deux distributions statistiques données sont similaires. Par exemple, si un échantillon A suit une loi normale est découpé en quantile dont les valeurs sont positionnées sur les abscisses. Dans le cas où une courbe est tracée à partir des quantiles d'un échantillon B et que cette courbe est linéaire, alors ces deux lois suivent une distribution normale.

Ci-dessous l'exemple d'un QQ-plot de deux lois normales et d'une loi normale et uniforme :

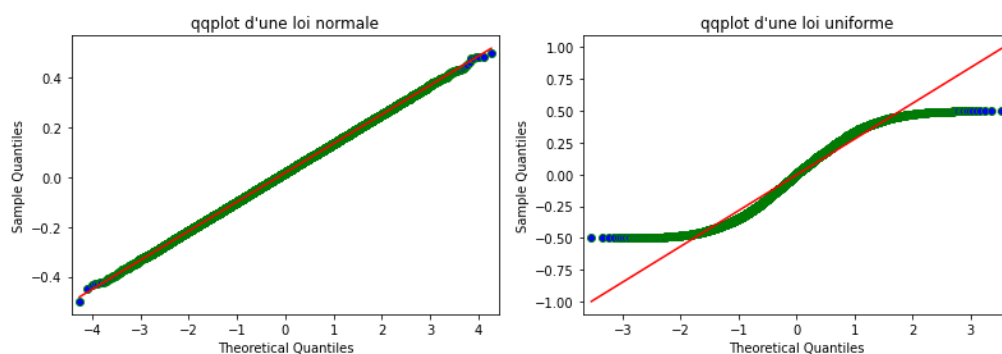


Figure 35 Qqplot d'une loi normale VS uniforme

⁵⁵ Les tests statistique et d'autres exemple d'utilisation sont cités dans le mémoire de Vanda ROUSSEAU (2011). *Le Best Estimate appliqué aux contrats d'épargne et de retraite selon les règles de Solvabilité II*. Mémoire d'actuariat, ISFA

⁵⁶ Voir un autre exemple d'utilisation du QQplot dans le mémoire de Jeremy LESNE (2012). *Calculation of the required capital for interest rates and equity risks for variable annuity products*. EURIA

2.1.3 Les outils ayant permis la sélection des variables

Dans cette sous partie, les outils statistiques qui ont permis la sélection des variables lors de la phase de construction des algorithmes prédictifs sont présentés.

2.1.3.1 La matrice de corrélation de Pearson basée sur la corrélation linéaire⁵⁷

L'analyse des dépendances entre variables débute souvent avec une analyse des corrélations linéaires. La corrélation linéaire est simple de compréhension et de calcul, mais possède néanmoins certains inconvénients. La corrélation linéaire est un estimateur pouvant manquer de robustesse sur des grandes bases de données, et ne permet pas de capter les autres types de dépendances (quadratiques...).

La corrélation linéaire est définie par l'équation suivante :

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Avec :

$$Cov(X,Y) = E(X * Y) - E(X) * E(Y)$$
$$\sigma_X = \sqrt{E(X^2) - E^2(X)}$$

2.1.3.2 La matrice de corrélation de Kendall basée sur les corrélations de rang

Afin d'analyser les corrélations non linéaires, l'analyse d'une matrice de corrélation de rang va permettre de capter d'autres types de dépendances entre les variables.

Le Tau de Kendall entre deux variables se définit comme suit :

$$\tau = \frac{(S - D)}{C_n^2}$$

Avec :

S : Nombre d'inégalités similaires par exemple si $X_i > X_j$ et $Y_i > Y_j$

D : Nombre d'inégalités discordantes par exemple si $X_i > X_j$ et $Y_i < Y_j$

n : Nombre de paires X,Y

C_n^2 combinaison de 2 parmi n, obtenue grâce à la formule suivante :

$$C_n^2 = \frac{n!}{(n-2)! * 2!}$$

Avec :

$$n! = n * (n - 1) * (n - 2) * ... * 1$$

Une analyse croisée de ces deux types de corrélation permet de déceler la grande majorité des dépendances.

2.2 Généralité sur la datascience et description des algorithmes utilisés dans l'étude

Dans cette sous partie, la démarche suivie pour la construction des modèles de ML, ainsi que les différents algorithmes prédictifs utilisés vont être présentés.

2.2.1 Description de la démarche du datascientist

⁵⁷ Les deux méthodes d'analyses des corrélations et d'autres méthodes ainsi que leur utilisation dans le cadre d'une analyse ERM sont présentées dans le mémoire d'Amine DRISSI BOUTAYBI (2014). *Technique d'agrégation des risques*. EURIA

Afin d'optimiser les performances des algorithmes et de garantir des prédictions pertinentes, le data scientist doit suivre la démarche suivante :

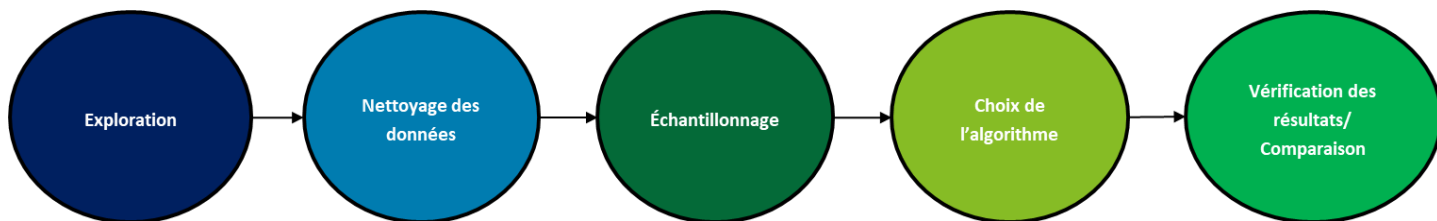


Figure 36 La démarche du data scientist

58

La phase d'exploration consiste à jeter un bref regard au set de données (voir si ces données sont lisibles, quelles sont les variables à analyser etc...)

La phase de Nettoyage des données consiste en la correction des défauts du dataset, c'est-à-dire corriger les valeurs manquantes, supprimer certaines valeurs aberrantes, ou transformer certaines variables explicatives pour pouvoir appliquer à ces données un algorithme voulu (exemple : transformation d'une variable qualitative en variable quantitative si on veut faire une régression linéaire) etc...

La phase d'échantillonnage (ou cross validation) fait intervenir des processus de découpage du data set en plusieurs sous-datasets (dont un dataset test et un ou plusieurs dataset(s) trains) permettant un calibrage des algorithmes sur ces datasets, la méthode de cross-validation k-fold est la plus utilisée.

La méthode de cross-validation k-fold consiste à découper la base train en k sous échantillons. Les sous échantillons vont servir de base d'apprentissage puis de test les uns après les autres pour le réglage des hyperparamètres des algorithmes.

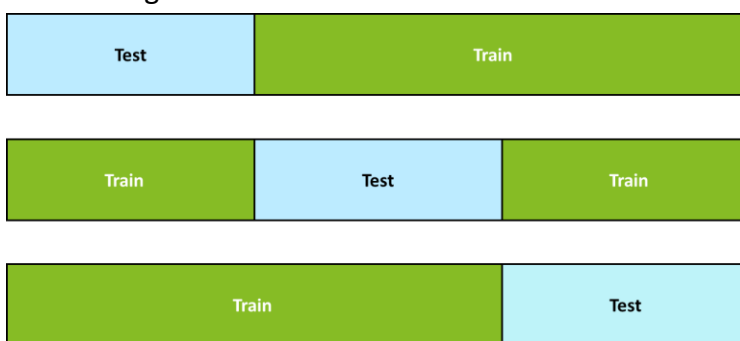


Figure 37 Découpage d'un dataset pour effectuer une 3-fold cross validation

Après cette phase de cross validation les algorithmes sont supposés être bien calibrés et de produire des résultats optimaux.

La sensibilité d'un algorithme à une variation des inputs peut se vérifier sur un autre dataset test, en analysant les variations des indicateurs de mesure de l'erreur du modèle choisis par l'utilisateur.

⁵⁸ Schéma tiré du cours de Machine Learning pour la tarification d'un produit d'assurance de M. Grondin, ESILV S8 année 2020

Avant de parler du fonctionnement des différents algorithmes composant l'intelligence artificielle, un récapitulatif sur les différentes catégories d'algorithmes composant l'intelligence artificielle et leur classification va être effectué.

2.2.2 Classification des algorithmes composants de l'IA

Schéma sur la classification des algorithmes composant l'IA

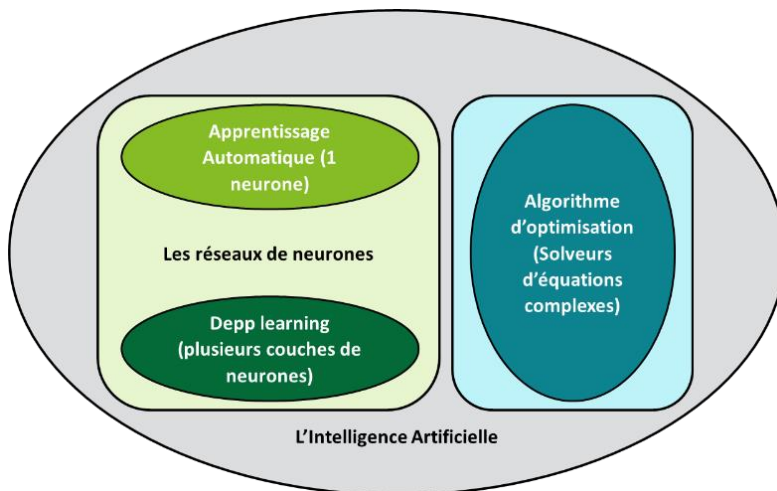


Figure 38 Schéma sur les différentes composantes de l'IA

Parmi les réseaux de neurones, il existe une classification exprimée dans le tableau ci-dessous avec des exemples pratiques pour illustrer leur usage :

Les réseaux de neurones		
Apprentissage supervisé	Apprentissage non supervisé	Deep learning
<ul style="list-style-type: none"> • La régression linéaire, logistique et l'arbre décisionnel en sont des exemples • Utilise des données connues et étiquetées en entrée • Des calculs pouvant être lourds et longs • Utilisé pour de la classification et de la régression • Produit des résultats précis • Le nombre de classes est connu à l'avance 	<ul style="list-style-type: none"> • L'ACP et le clustering en sont des exemples • Données inconnues en entrée • Plus « friendly user » que les apprentissages supervisés • Utilise l'analyse en temps réel des données • Exploitation des règles de clusterings et d'association • Moins précis que les méthodes supervisées • Le nombre de classe est a priori inconnu 	<ul style="list-style-type: none"> • Les réseaux neuronaux de convolution en sont des exemples • Données inconnues en entrée • Besoin d'adapter « l'architecture de l'algorithme » selon la tâche voulue • Utilise l'analyse en temps réel des données • Exploitation et interprétation de règles de clusterings et d'associations combinées • Utilisé dans la reconnaissance d'image, chatbot etc... • Peut créer du contenu à partir d'un fichier vide

Figure 39 Les différentes composantes des réseaux de neurones illustrées de usecases

Dans le cadre de l'étude menée, afin de produire des résultats précis et travaillant sur des données connues, différents algorithmes provenant des méthodes supervisées ont été choisis.

Le fonctionnement de ces algorithmes sont exprimés dans la partie suivante.

2.2.3 Descriptions du fonctionnement des algorithmes d'optimisation, de Machine Learning et de deep learning utilisés dans le mémoire⁵⁹⁶⁰⁶¹⁶²

Dans cette sous partie différents algorithmes et concepts d'intelligence artificielle utilisés lors de l'étude seront décrits.

2.2.3.1 Le SVM

Le SVM est un algorithme d'apprentissage supervisé basé sur des méthodes statistiques, dont le principe est de trouver un hyperplan séparateur entre deux groupes d'individus différents.

2.2.3.2 L'arbre de décision

Un arbre de décision est une méthode supervisée représentant un ensemble de choix sous une forme graphique appelée arbre de décision. L'algorithme pose des questions (chiffrées ou non) au dataset ce qui permet de trier chaque individu en groupes et sous-groupes selon leurs caractéristiques.

L'arbre de décision peut se schématiser de la manière suivante :

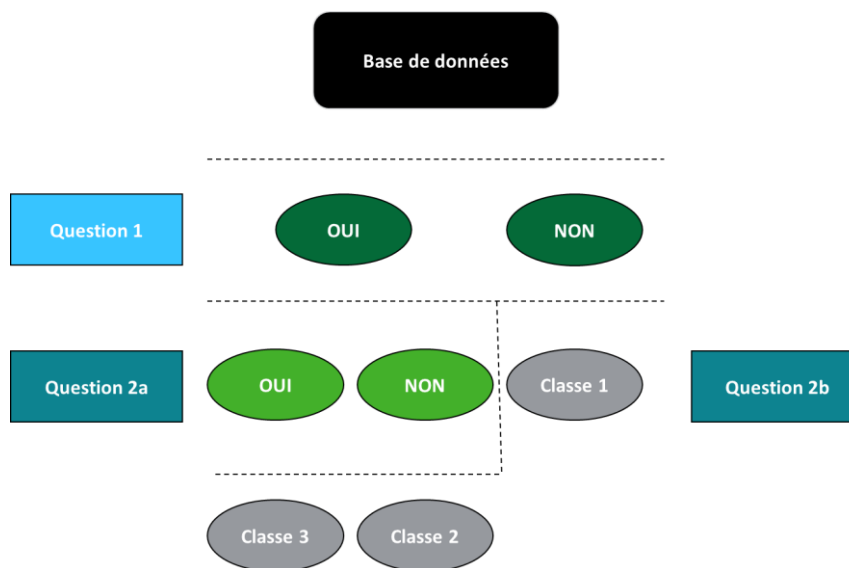


Figure 40 L'arbre de décision

2.2.3.3 Le Random Forest

Le Random Forest est un algorithme d'apprentissage supervisé faisant intervenir les deux concepts suivants :

- Le concept de bagging
- Le concept de « weak-learner » pouvant être aussi considéré comme un arbre de décision « simple ».

⁵⁹ Les schémas des algorithmes de Machine Learning sont inspirés du cours Mme. Boyer, cours de Machine Learning, ISUP CS3 2022

⁶⁰ Les schémas sur le perceptron informatique et sur le RNC sont inspirés du cours de M. Lopez, cours de réseaux de neurones, ISUP CS3 2022

⁶¹ Description des hyperparamètres du XGboost tirés du cours de M. Marcotorchino, cours d'IA et Big-data, ISUP CS3 2022

⁶² Pour plus de détail sur les principes mathématiques des algorithmes, voir la thèse de Ruocong ZHANG (2014). *Apprentissage statistique en gestion de portefeuille*. Télécom Paris

Le bagging est un mot pour qualifier le découpage de la base de données d'apprentissage de manière aléatoire.

Ce découpage de la base de données s'effectue à l'aide de méthodes de sous échantillonnages de la base de données (Undersampling) consistant en différents tirages avec remises pouvant être hyperparamétrés.

Une fois le bagging effectué, chaque sous échantillons va être appris par un « weak learner ». Un weak learner est un arbre de décision dit « naïf » c'est-à-dire avec une faible profondeur (peu de nœuds) pouvant être paramétrée.

Lorsque chaque arbre de décision a « appris sa partition », ces arbres vont être agrégés afin que l'algorithme puisse prédire le résultat « global » de la forêt d'arbres aléatoires.

Le random-forest peut-être schématisé de la manière suivante :

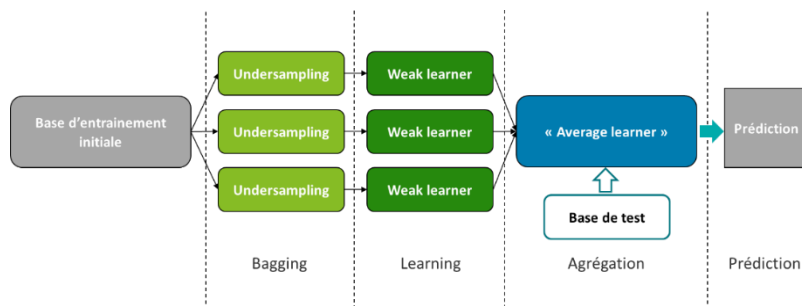


Figure 41 Fonctionnement d'un Random Forest

2.2.3.4 Le XGboost et sa principale différence avec un Random Forest

La principale différence entre le XGboost et le Random Forest se fait lors de l'étape de bagging et de learning. Pour le XGboost, cette étape va faire intervenir un principe de « Backward propagation », c'est-à-dire que les weak learners vont améliorer leurs résultats en revenant en arrière. Pour ce faire ils vont améliorer leurs performances en allant sélectionner des données supplémentaires dans d'autres sous échantillons que celui qui leur été initialement attribué. Ce retour en arrière va améliorer leurs performances individuelles pour améliorer la performance collective une fois agrégés.

Le fonctionnement du XGboost peut être schématisé de la manière suivante :

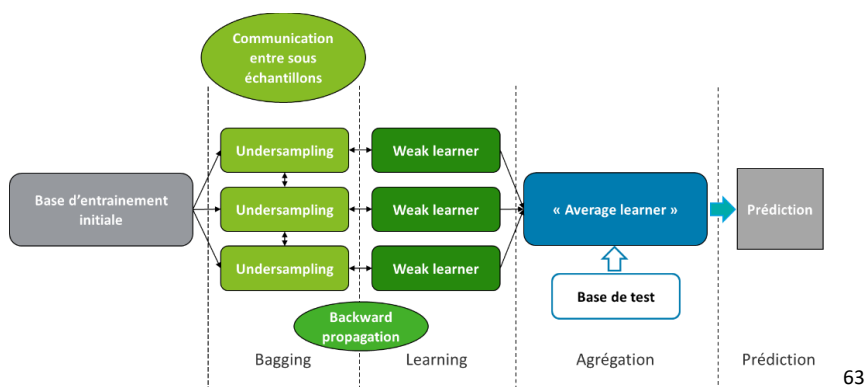


Figure 42 Fonctionnement d'un XGboost

⁶³ Description de l'hyperparamétrage du XGboost de la descente de gradient en annexe partie 6, chapitre 6.5

2.2.3.5 La méthode Nelder-mead, un algorithme d'optimisation utilisé lors du calcul de l'input de spread pondéré⁶⁴

La méthode Nelder-mead est utilisée dans l'étude pour calculer la variable explicative de spread pondérée⁶⁵. La méthode de Nelder-Mead est un algorithme d'optimisation qui permet de trouver le minimum d'une fonction pouvant être non linéaire ou non quadratique à l'aide de la méthode du simplexe.

La méthode du simplexe permet d'approcher une solution en prenant N points et un pas de résolution à son initialisation.

Ces N points forment alors un objet géométrique nommé simplexe possédant un centre de gravité x_0 . Cet algorithme se construit sur les trois règles suivantes :

- Supprimer l'expérience pour laquelle la réponse est la plus mauvaise et la remplacer par une expérience située à l'opposé du centre de gravité x_0 formée par les points.
- Si le simplexe vacille entre plusieurs points identiques alors, le deuxième plus mauvais point est abandonné
- Si après l'itération n+1, le même sommet est conservé, alors le point de ce sommet est considéré comme la solution maximale, ou l'algorithme peut être réinitialisé à ce point avec un pas graduel plus petit.

La méthode Nelder-Mead va reprendre ce principe en y appliquant des coefficients supplémentaires pour l'étape 3.

L'algorithme peut être représenté de la manière suivante :

1. Choix de $N+1$ points de l'espace des inconnues de dimension N . Cela forme un simplexe : $\{x_1, \dots, x_{N+1}\}$
2. Calcul des valeurs de la fonction f en ces points, tri des points de façon à avoir $f(x_1) \leq \dots \leq f(x_n) \leq f(x_{n+1})$.
3. Calcul de x_0 , centre de gravité de tous les points sauf x_{N+1} .
4. Calcul de $x_r = x_0 + \alpha * (x_0 - x_{n+1})$.
5. Soit $f(x_1) < f(x_r) < f(x_N)$, remplacement de x_{n+1} par x_r et retour à l'étape 2.
6. Soit $f(x_r) < f(x_1)$, calcul de $x_e = x_0 + \gamma * (x_r - x_0)$ (expansion du simplexe). Si $f(x_e) \leq f(x_r)$, remplacement de x_{n+1} par x_r et retour à l'étape 2.
7. Soit, calcul de $f(x_r) \geq f(x_N)$ calcul de $x_c = x_0 + \rho * (x_{N+1} - x_0)$. Si $f(x_c) < f(x_{N+1})$, remplacement de x_{N+1} par x_c sinon aller à l'étape 8.
8. Homothétie de rapport de σ et de centre x_1 : remplacement de x_i par $x_1 + \sigma * (x_i - x_1)$ et retour à l'étape 2.

Avec α, γ, ρ et σ des coefficients réels tels que $\alpha > 0, \gamma > 1$ et $0 < \rho \leq 0,5$.

2.2.3.6 Le perceptron informatique, la base du machine et deep learning⁶⁶

⁶⁴ Voir description plus détaillée dans la thèse de Alaa DAHER (2019). *Optimisation des réseaux cellulaires pour les communications professionnelles critiques*. Télécom Paris

⁶⁵ Voir partie 3, chapitre 3.3.2.5

⁶⁶ Schéma et description inspiré du mémoire de Silvia BUCCI (2021). *Etude et implémentation de techniques d'analyse de sensibilité dans les modèles de tarification Non-Vie*. ENSAE

L'architecture du perceptron informatique a inspiré la partie qui produit les chocs obligataires du générateur de données⁶⁷.

Le perceptron informatique peut se schématiser de la manière suivante :

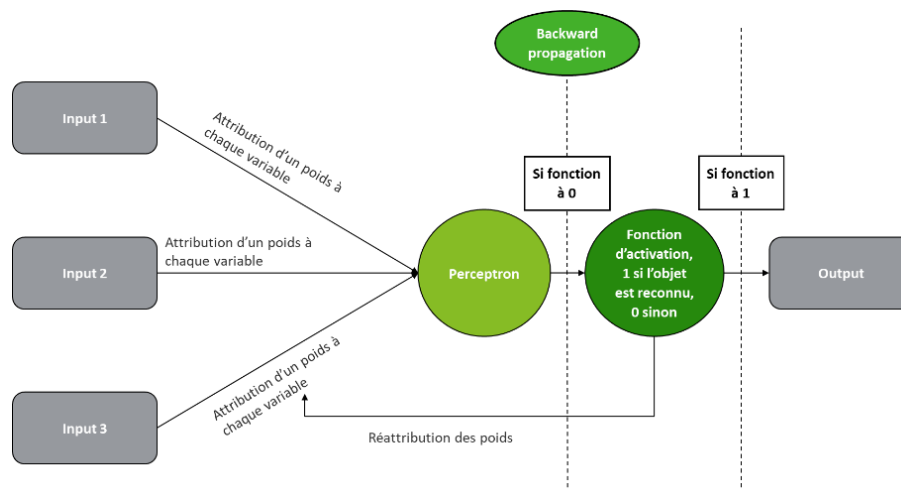


Figure 43 Fonctionnement d'un perceptron informatique

Le perceptron informatique prend en entrée une série d'inputs.

Chaque input va se voir dans un premier temps attribué un poids (entre 0 et 1) qui va permettre la reconnaissance de l'objet rentré selon un classement « d'importance entre les inputs ».

Les poids initiaux attribués aux variables sont différents pour chaque input et différents de 0.

Si le perceptron ne reconnaît pas un objet (la fonction d'activation ne s'active pas) par l'attribution des poids initiaux, alors le réseau va réattribuer les poids des inputs différemment (backpropagation), jusqu'à ce que la fonction d'activation passe à 1.

Une fois l'objet reconnu (fonction d'activation à 1), alors l'objet prédit sort en output.

Pour activer la fonction d'activation, le perceptron va renvoyer une liste de probabilités des différents objets, si une probabilité de « vraisemblance » d'un objet atteint un certain seuil, alors la fonction d'activation passe à 1.

2.2.3.7 Le réseau de neurone de convolution faisant parti des apprentissages profonds ⁶⁸

Les RNC sont une famille de réseau de neurones de type deep-learning très utilisés dans la reconnaissance et le traitement de l'image. Cette structure est utilisée pour la conversion en table de choc des portefeuilles obligataires répliqués.

Ces réseaux de neurones fonctionnent sur des étapes de filtrage (réduction de la base de données) et de pooling (rassemblement de la base de données). Une fois toutes ces étapes implémentées, l'algorithme est capable de reconnaître des data placées en input.

Le schéma de reconnaissance d'une image suivant peut être pris à titre d'exemple :

⁶⁷ Voir sous partie 3, sous partie 3.1

⁶⁸ Voir thèse de Zong Xiangyu (2021). Machine Learning in stock indices trading and pair trading. University of Glasgow

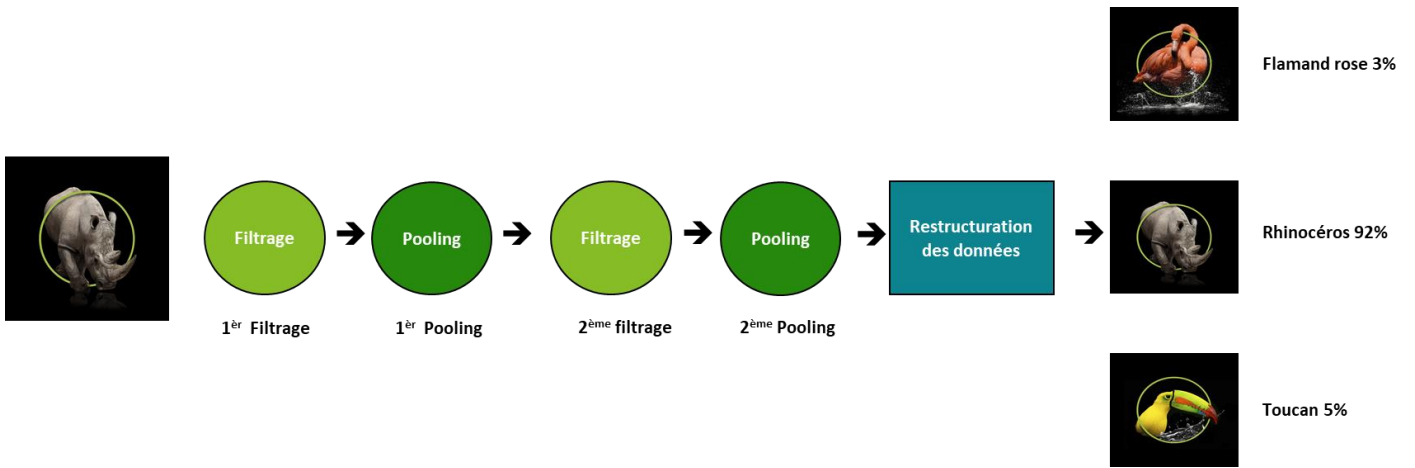


Figure 44 Le réseau de neurone de convolution (images de source Deloitte)

Chapitre 3

3 Implémentation d'un réseau de neurone générateur de chocs obligataires ⁶⁹

Cette partie s'intéresse à la création, puis au calibrage du créateur d'inputs obligataires pour l'étude des algorithmes de ML pouvant aussi être réutilisé comme GSE obligataire.

La création de cet outil permet un gain de temps et un large choix de paramétrage pour la création des bases de données et des inputs obligataires pour l'étude Machine Learning.

3.1 Création du générateur des chocs

Dans cette partie, les différentes étapes menant à la création de l'outil de choc obligataire sont décrites.

L'outil a la capacité d'être utilisé pour :

- Effectuer des chocs obligataires et pouvoir être utilisé comme un GSE de spread obligataire
- Pouvoir répliquer un portefeuille obligataire de manière réaliste
- Créer les inputs obligataires pour une utilisation ML sur un portefeuille répliqué ou initial
- Avoir une transparence maximale dans les résultats sortis

3.1.1 Implémentation d'un générateur de table de chocs uniformes

Pour avoir des prédictions cohérentes les algorithmes de Machine Learning nécessitent des bases de données relativement importantes d'au moins 10 000 lignes ⁷⁰.

Pour ce faire deux solutions se présentent à nous :

- Générer 10 000 trajectoires stochastiques à l'aide du logiciel Prophet
- Générer 10 000 trajectoires déterministe en modifiant les lignes d'actifs du portefeuille, en faisant varier chaque variable de manière aléatoire entre différents plafonds donnés jusqu'à l'obtention d'une base de données de 10 000 lignes ou plus.

La première méthode présente une facilité d'implémentation car déjà existante. Cependant cette méthode présente l'inconvénient d'effectuer des calculs longs avec une tendance boîte noire au vu de la complexité de l'implémentation des calculs du SCR de spread. En cas de changement brutaux des marchés ou des courbes de taux, cette méthode implique une actualisation lente de la base de données.

Le principal inconvénient de la deuxième méthode est son implémentation car non existante, cependant une fois implémentée, cette méthode permettrait d'actualiser en temps réel la BDD. Cela permettrait d'offrir des résultats compétitifs en temps réel.

⁶⁹ Une étude de portefeuilles répliquants au passif est proposée dans le mémoire de Styven PONNUSAMY (2020). *Méthodes de sélection d'instruments financiers pour le calibrage de portefeuilles répliquants*. ISUP

⁷⁰ Voir le mémoire de Valentin CERISIER (2021). *Application de méthodes de Machine Learning dans le calcul de la solvabilité infra-annuelle*. Dauphine

3.1.1.1 Le générateur de spreads aléatoires

Afin de générer des chocs de spreads sur la base de données initiale, un générateur de spread aléatoire a dans un premier temps été développé. Ce générateur va répliquer les actifs du portefeuille déjà existant en y ajoutant des bruits.

Ce générateur s'inspire de la méthode de choc de spread S2 standard ⁷¹ à une maille plus fine, c'est-à-dire qu'un choc va être généré par maturité année par année et selon l'échelon de crédit de l'obligation.

Deux listes de plafonds pour les chocs ont été établies (la liste basse et la liste haute).

La liste basse est constituée de la valeur 0 pour le premier choc puis de la valeur de tous les chocs précédents D_i .

La liste haute est constituée d'indices augmentant linéairement selon la longueur des maturités et du nombre d'échelons. Soit Le coefficient haut U_i définit de la manière suivante :

$$U_i = \frac{100}{\sum_{i=1}^{nb_annees} \sum_{j=1}^{nb_echelons} ((1))} \cdot$$

Un tirage aléatoire uniforme pour chaque choc entre la liste haute et basse des valeurs (D_i, U_i) est ensuite effectué

Cette fonction génère une table de choc de telle sorte à ce qu'elle puisse être utilisée sous Prophet, afin d'appliquer (soustraction ou addition) et/ou de modifier les chocs voulus directement sous Prophet.

Voilà comment s'organise l'output :

Maturité résiduelle	Echelons de spread	Chocs minimaux	Type des lois	Tirages
1	0	0	uniforme	0.0<0.08<0.25
2	0	1,71	uniforme	1.71<2.08<2.25
...
50	0	97,58	uniforme	97.58<97.74<98.25
1	1	0,080	uniforme	0.08<0.34<0.50
...
50	9	99,32	uniforme	99.32<99.43<100.00

Figure 45 Table sortie par le générateur de spread aléatoire uniforme initialement codé

Le générateur de spread aléatoire générant des chocs trop significatifs et étant pensé pour fonctionner de manière uniquement uniforme, un générateur de spread obligataire plus sophistiqué a été implémenté. Ce générateur est inspiré d'un empilement de neurone informatique.

3.1.2 L'implémentation du générateur de portefeuilles aléatoires et sa connexion à Prophet

Dans cette sous partie va être présentée la construction du générateur de spread.

3.1.2.1 Présentation du générateur d'un point de vu global

Afin d'éviter le lancement d'un nombre important de runs Prophet synonyme d'un cout temporel important, il a été décidé d'implémenter une méthode de génération de portefeuilles aléatoires pouvant être appelée un grand nombre de fois.

L'algorithme initialement pensé pour effectuer une méthode de portefeuille répliquant a été amélioré afin de pouvoir aussi définir un GSE de spread obligataire.

⁷¹ Voir partie 1, chapitre 1.3.6

Pour calibrer l'outil selon différents scénarios, il est possible de définir les bornes des coefficients de chocs obligataires à renvoyer sous Prophet.

Ci-dessous est schématisée la méthode de portefeuille répliquant inspirée d'un réseau de neurone auquel a été ajoutés des hyperparamètres pour mieux créer/expliciter les résultats sorties.

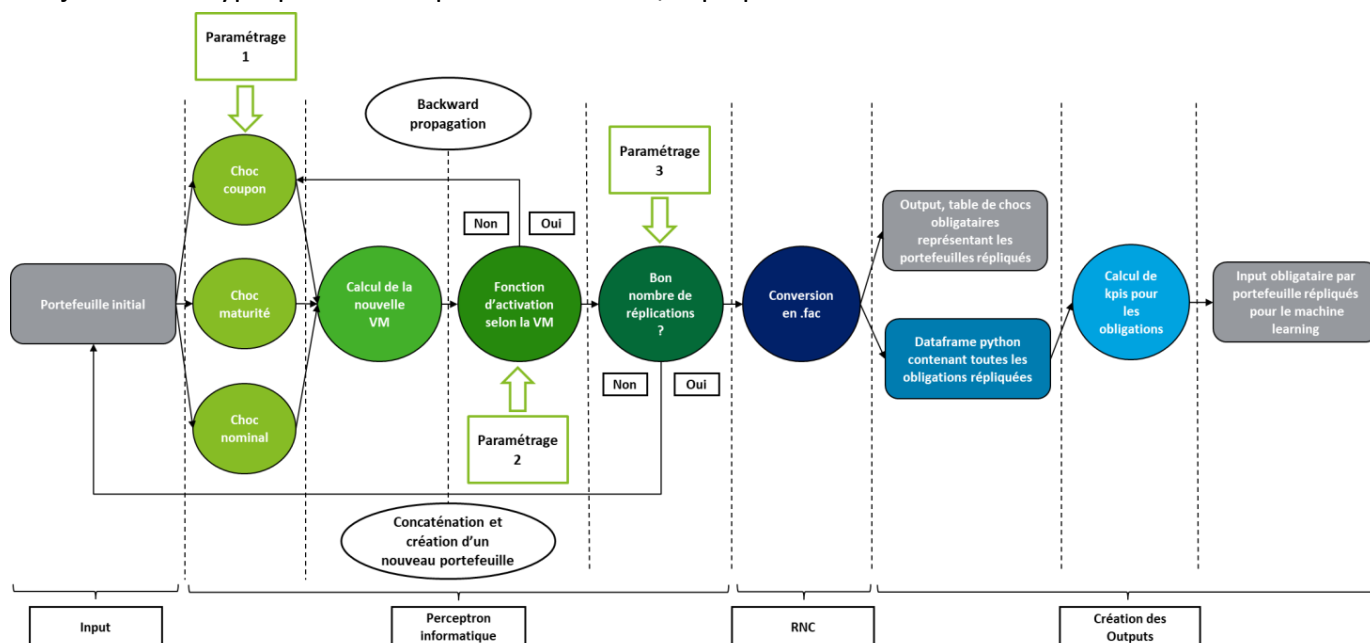


Figure 46 Schéma du réseau de neurones générateurs des chocs obligataires

Une option afin de choisir la fonction de densité des chocs effectués sur la maturité, le nominal et les coupons (première couche de neurones) a été ajoutée sur ce réseau.

Le but de cette première couche (présentée ci-dessous) est de tirer aléatoirement des échantillons selon des lois statistiques rentrées par les utilisateurs.

3.1.2.2 Zoom sur les générateurs de chocs

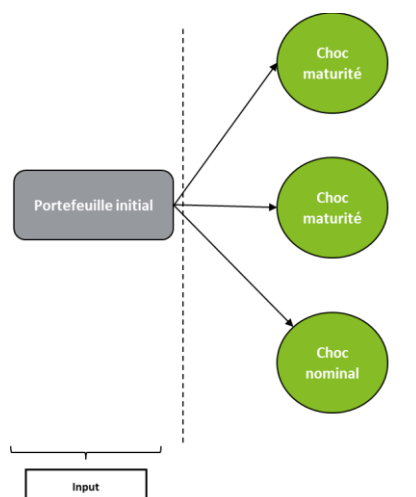


Figure 47 Première couche de neurone du générateur de données obligataires

Le choix des lois se fait parmi les quatre lois suivantes :

- Pareto
- Gaussienne
- Uniforme
- Weibull

Ces quatre lois statistiques ont été choisies pour leur variété de fonction de répartition et de densité ⁷²
 Une fois la loi choisie, les vecteurs Pythons créés par cette loi sont normalisés selon la méthode min-max⁷³.
 Pour une loi normale $N(0,1)$ la fonction de densité suivante est obtenue :

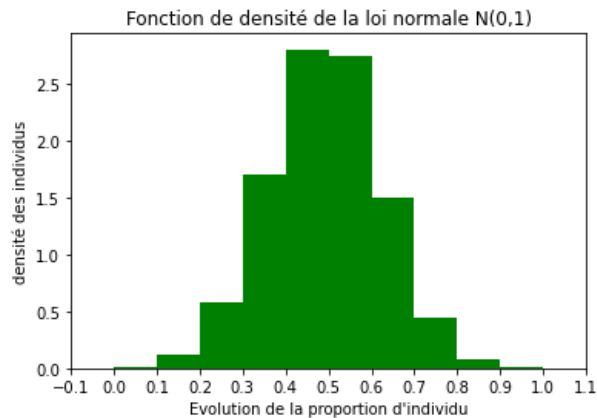


Figure 48 Fonction de densité d'une loi de choc normale après une transformation min-max

Afin de pouvoir personnaliser la loi à sa guise l'utilisateur pourra utiliser un facteur de dissymétrie intérieur permettant d'appliquer un biais sur la fonction de répartition normalisée min-max comme suit :

$$X_{custom} = X_{normalise} - Dissymétrie_{int}$$

Si une dissymétrie de $-0,5$ est appliquée, la densité suivante sera obtenue :

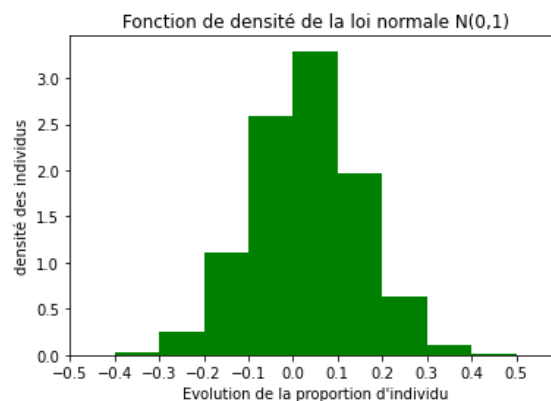


Figure 49 Fonction de densité d'une loi normale après une transformation min-max et l'ajout d'un biais

Un facteur d'intensité peut-être utilisé afin de régler « la variance » de la loi appliquée :

$$X_{custom} = (X_{normalise} - Dissymétrie_{int}) * Facteur_{intensité}$$

Si un facteur de 10 est sélectionné, appairé à un facteur de dissymétrie int à $-0,5$, la courbe de densité suivante sera obtenue :

⁷² Voir partie 2, sous partie 2.1.1
⁷³ Voir partie 2, introduction de la sous partie 2.1

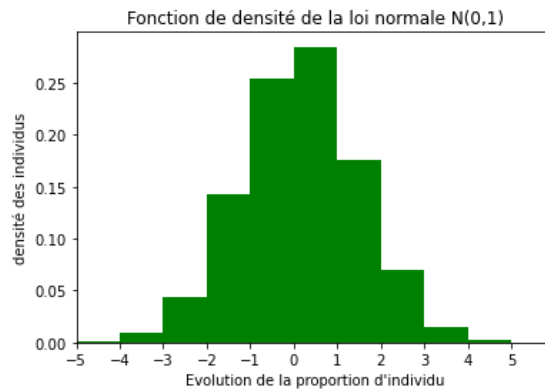


Figure 50 Fonction de densité d'une loi normale après une transformation min-max, l'ajout d'un biais et d'un facteur de variance multiplicatif

Un facteur de dissymétrie externe a été ajouté au calcul pour affiner la fonction de densité si l'utilisateur le souhaite, l'ajout de ce facteur est présenté ci-dessous :

$$X_{custom} = (X_{normalise} - Dissymétrie_{int}) * Facteur_{intensité} + Dissymétrie_{ext}$$

En reprenant le paramétrage de la loi ci-dessus et en appliquant un choc de dissymétrie extérieur de 1, la courbe de densité suivante sera obtenue :

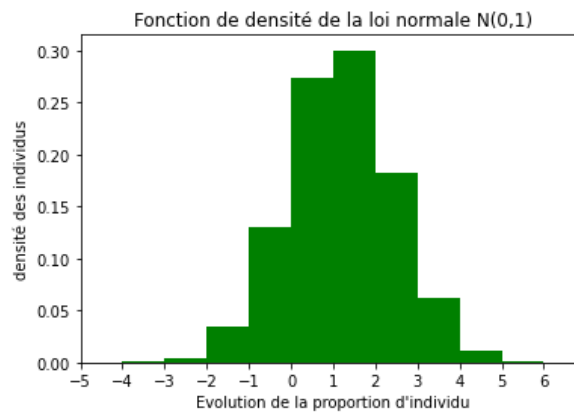


Figure 51 Fonction de densité d'une loi normale après une transformation min-max, l'ajout d'un biais, d'un facteur de variance multiplicatif et d'un biais supplémentaire

Une fois les chocs effectués sur les coupons, nominaux et maturité, un neurone calculera la $VM_{nouvelle}$ égale à la somme des VM des obligations répliquées par portefeuille.

Une fois cette nouvelle VM calculée, ce réseau calculera les coefficients de chocs à appliquer dans Prophet de la manière suivante :

$$Coeff_{choc} = \frac{VM_{new}}{VM_{initiale}}$$

L'ajout d'une fonction d'activation au réseau permet de vérifier la cohérence des coefficients sortis.

La « cohérence » des coefficients sera jugée par l'utilisateur qui rentrera une borne de coefficients de chocs voulus $Coeff_{down}$ et $Coeff_{up}$ (fonction d'activation).

Si l'algorithme constate que les coefficients ne sont pas dans la borne donnée, le réseau effectuera une backpropagation.

Par soucis d'explicabilité des résultats, cette backpropagation s'effectuera à l'aide d'une méthode bootstrap⁷⁴ qui permet de retirer de manière uniforme les coefficients de choc pour les maturités, les coupons et les nominaux.

L'utilisateur a également le choix ou non d'activer la fonction d'activation dans les hyperparamètres.⁷⁵

3.1.2.3 Possibilité d'utilisation de plusieurs générateurs de chocs d'affilés

Il est possible de créer une couche avec plusieurs générateurs de chocs afin de pouvoir proposer une plus grande possibilité de densité de chocs comme schématisé ci-dessous :

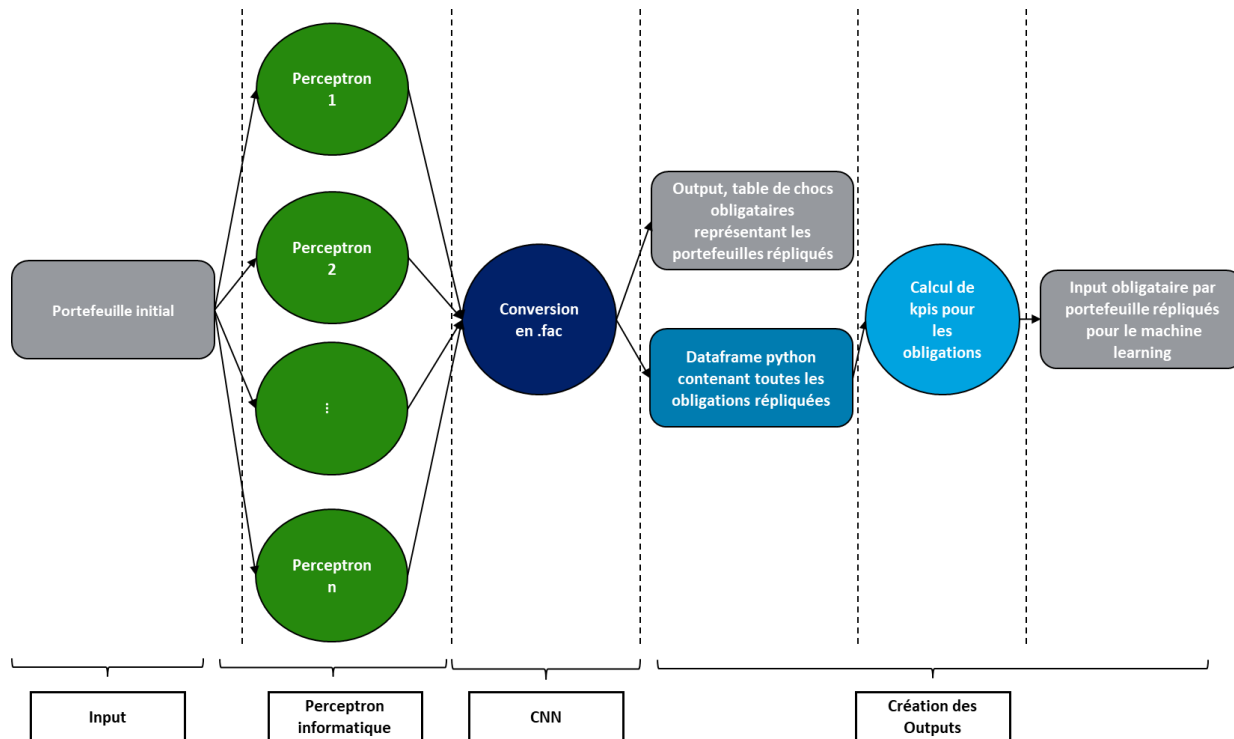


Figure 52 Illustration de la possibilité d'ajouter plusieurs neurones à la couche du générateur des données

La possibilité de pouvoir employer plusieurs couches génératrices de chocs permet de répliquer des fonctions de densité de chocs très complexes, voir graphique ci-dessous :

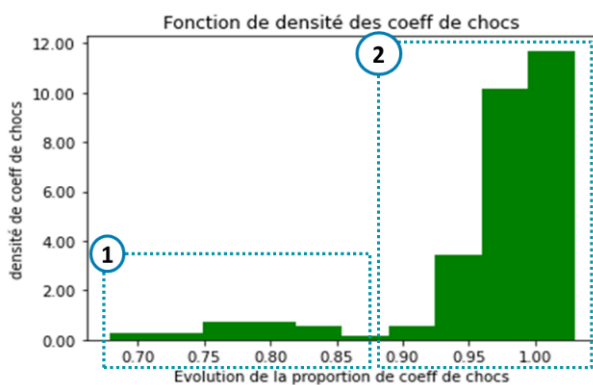


Figure 53 Exemple de fonction de densité produite par un réseau à deux neurones générateurs de chocs⁷⁶

⁷⁴ Voir partie 2, chapitre 2.1.2.1

⁷⁵ Voir annexe partie 6, chapitre 6.11.1

⁷⁶ Fonction de densité des chocs calibrés sur les obligations grecques, voir partie 3, chapitre 3.2.2.2

Cette possibilité peut permettre de créer un GSE de spread par échelon de rating en superposant la génération de sept distributions de chocs différentes puis de les regrouper lors de la conversion en format Prophet.

3.1.2.4 Connexion à Prophet inspirée par une méthode deep-learning de type RNC

La conversion de ce portefeuille en table de choc pour la connexion à Prophet, se fait à l'aide d'un algorithme inspiré d'un réseau de neurone de convolution pensé et développé pour cet usage :

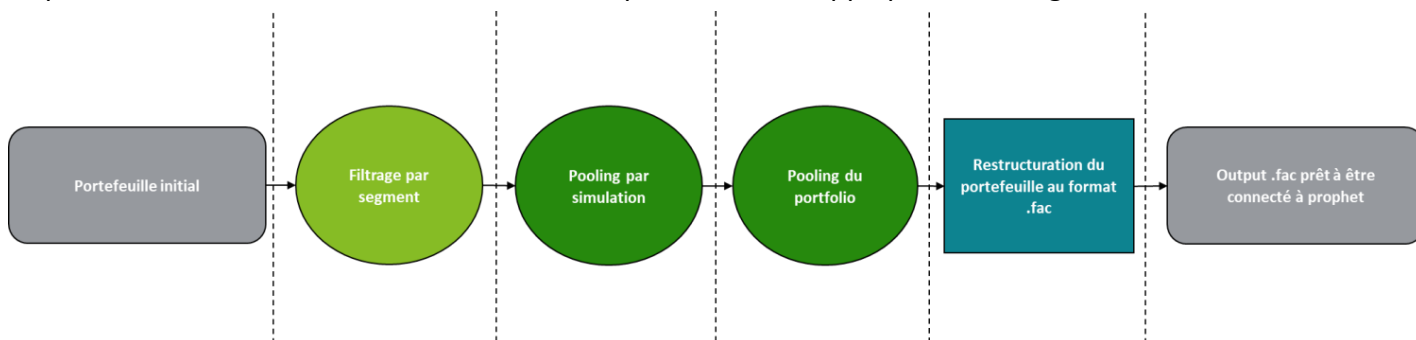


Figure 54 Convertisseur inspiré d'un RNC

Ce convertisseur génère une table structurée de la manière suivante :

SEGMENT_NO	SIMULATION	Maturité résiduelle et échelon de crédit	Chocs
1	1	Maturité 6 ans, échelon 1	0,88
1	2	Maturité 6 ans, échelon 1	0,87
1	3	Maturité 6 ans, échelon 1	0,85
1	4	Maturité 6 ans, échelon 1	0,82
2	1	Maturité 11 ans, échelon 1	0,87
2	2	Maturité 11 ans, échelon 1	0,86
2	3	Maturité 11 ans, échelon 1	0,81
2	4	Maturité 11 ans, échelon 1	0,82

Figure 55 Table de chocs générée par le convertisseur

Pour la structuration de cette table, l'utilisateur peut déclarer le nombre de simulations stochastiques et la dimension de la table générée, le nombre de segments étant automatiquement détectés par l'algorithme selon le nombre de simulations rentrées.

Un développement Prophet a ensuite été effectué afin d'appeler cette table lors du calcul des simulations choquées.

La connexion Prophet peut se schématiser comme suit :

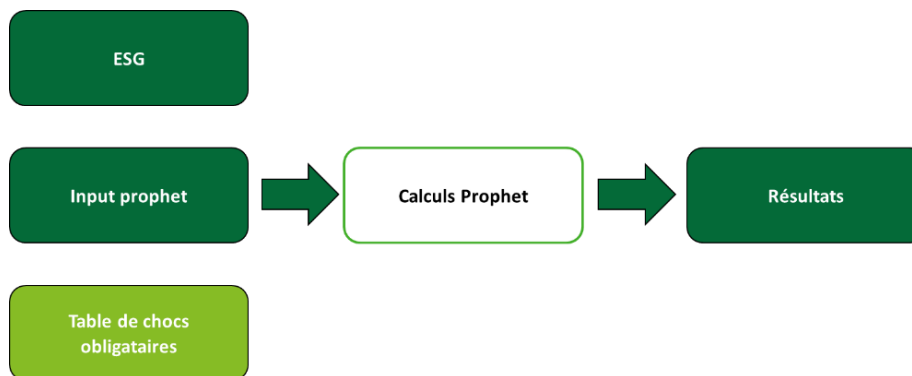


Figure 56 Schéma des nouveaux inputs branchés au simulateur actuariel

3.2 Fonctionnalités du générateur de densité et calibration des chocs

Les résultats générés avec différents paramétrages de l'outil développé sont présentés et commentés. La méthode de portefeuilles répliquants est testée et vérifiée.

Un calibrage sur la densité des chocs de l'algorithme sortis selon deux scénarios est ensuite effectué. Enfin, la construction de différentes bases de données pour les prédictions des algorithmes de ML est présentée.

3.2.1 Analyse des coefficients de chocs sortis par le perceptron informatique

Dans cette partie, une analyse de la répartition des coefficients de chocs sortis par le générateur selon différentes combinaisons d'hyperparamètres est proposée.

Cette partie s'intéresse notamment au comportement de la distribution selon le choix des lois rentrées par l'utilisateur.

Les hyperparamètres vont être fixés pour la suite de cette partie.

Paramétrage

La backward propagation est activée

- Le nombre de répétition avant laquelle la backward est arrêtée est fixé à 1000
- La borne basse de la risque neutralisation paramétrée pour la backward propagation est fixée à 0,7
- La borne haute de la risque neutralisation paramétrée pour la backward propagation est fixée à 1,3
- 19 obligations dans le portefeuille initial
- 100 portefeuilles vont être répliqués

Hors neurone générateur de choc de maturité, Les paramètres suivants restent fixes :

- Une dissymétrie de choc intérieur de -0.4
- Un biais supplémentaire de 0
- Pour les neurones de choc nominal et coupon le paramètre suivant va rester fixe :
- Un facteur d'intensité de choc à 1

Pour le neurone de choc de maturité, le paramètre suivant est fixé :

- Un facteur d'intensité à 10

La loi des chocs va être amenés à varier.

Les options de loi de chocs sont présentées ci-dessous :

- Option 1 : Loi de Pareto
- Option 2 : Loi Normale
- Option 3 : Loi Uniforme
- Option 4 : Loi de Weibull, param1= γ , param2= k

L'équation génératrice de choc pour ces réseaux est :

$$X_{custom} = (X_{normalise} - Dissymétrie_{int}) * Facteur_{intensité} + Dissymétrie_{ext}$$

3.2.1.1 Univers de chocs de Pareto

Loi de Pareto de paramètre 20 pour tous les chocs

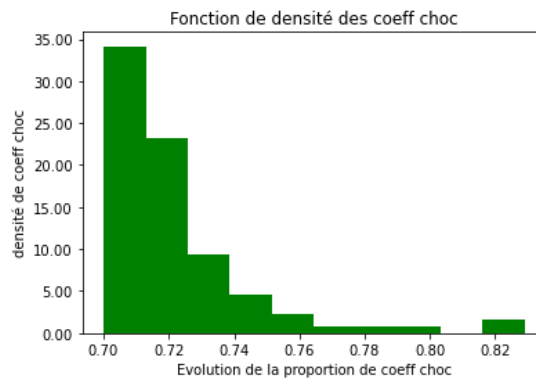


Figure 57 Densité des coefficients de chocs obtenus pour l'univers de Pareto

Après avoir effectué un test d'adéquation de Kolmogorov-Smirnov⁷⁷ entre cet échantillon, et un échantillon représentatif d'une loi de Pareto normalisée à l'aide de la méthode min-max, la p-value de 0,995 est obtenue, confirmant la similitude de la courbe avec une loi de Pareto⁷⁸.

3.2.1.2 Univers de chocs Normaux

Loi Normale $N(0,1)$ pour tous les chocs

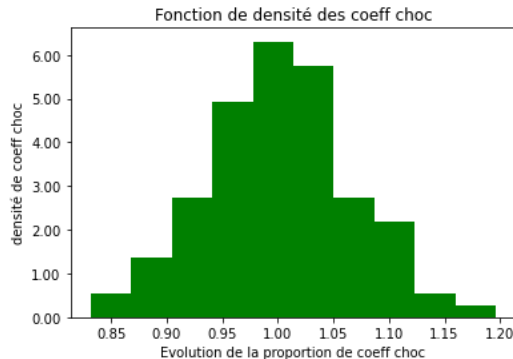


Figure 58 Densité des coefficients de chocs obtenus pour l'univers Normal

Après avoir effectué un test d'adéquation de Kolmogorov-Smirnov entre cet échantillon et un échantillon produit par une loi normale⁷⁹ normalisée selon la méthode min-max, une p-value de $3e - 16$ est obtenue, ne confirmant pas la similitude de la courbe avec une loi normale.

⁷⁷ Voir partie 2, chapitre 2.1.2.2

⁷⁸ Voir partie 2, chapitre 2.1.1.3

⁷⁹ Voir partie 2, chapitre 2.1.1.2

Pour avoir la confirmation de ces résultats, un QQplot⁸⁰ de la courbe va être effectué :

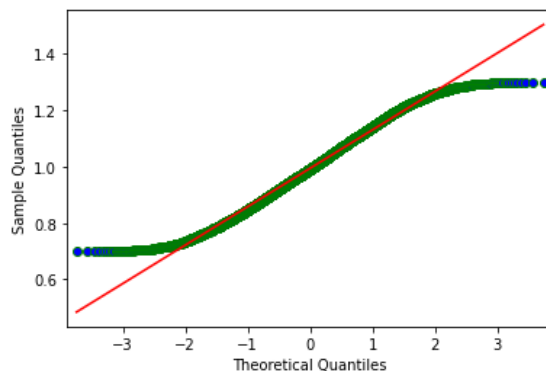


Figure 59 Qqplot de la densité des coefficients de chocs obtenus pour l'univers de choc normal

Si le réseau de neurone est uniquement paramétré avec des lois normales pour chaque choc, avec des bornes de backward propagation restreintes, alors le réseau produira une loi « pseudo normale » ne suivant pas une loi normale aux extrémités de sa courbe.

3.2.1.3 Univers de chocs Uniformes

Loi Uniforme ⁸¹ $U(-0.5,0.5)$ pour tous les chocs pour les générateurs de chocs de coupons et nominaux et $U(-5,5)$ pour les générateurs de chocs de maturité.

Ci-dessous le graphique de densité :

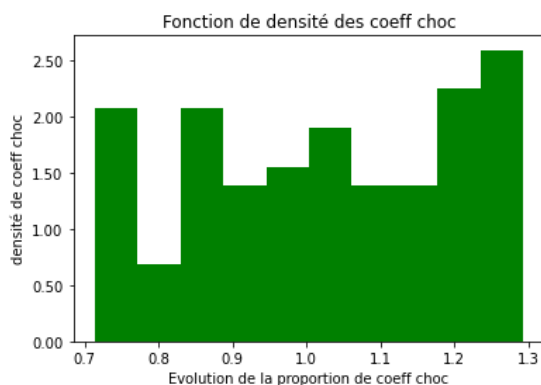


Figure 60 Densité des coefficients de chocs obtenus pour l'univers Uniforme

Après avoir effectué un test d'adéquation de Kolmogorov-Smirnov entre cet échantillon et un échantillon de choc produit par une loi uniforme aux bornes $[0.7,1.3]$ la p-value de 0,89 est obtenue, confirmant la similitude de la courbe avec une loi uniforme.

3.2.1.4 Univers de chocs de Weibull

Loi de Weibull pour tous les chocs

⁸⁰ Voir partie 2, chapitre 2.1.2.3

⁸¹ Voir partie 2, chapitre 2.1.1.3

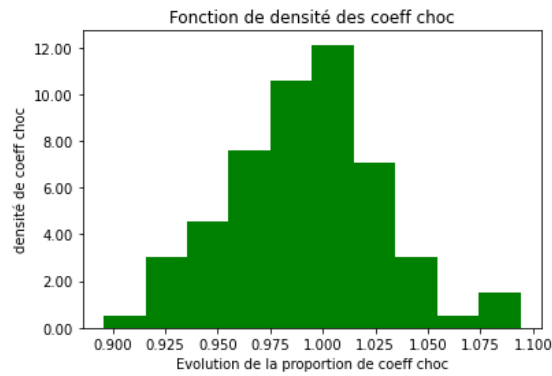


Figure 61 Densité des coefficients de chocs obtenus pour l'univers de Weibull

Après avoir effectué un test d'adéquation de Kolmogorov-Smirnov entre cet échantillon et un échantillon avec une loi de Weibull ⁸²normalisée avec la méthode min-max, la p-value de 0.4 est obtenue, confirmant la similitude de la courbe avec une loi de Weibull.

3.2.1.5 Univers de chocs de Pareto pour les coupons et le nominal et chocs Normaux

Loi de Pareto pour les chocs coupons et nominal, loi normale pour les chocs de maturité.

Ci-dessous le graphique de densité :

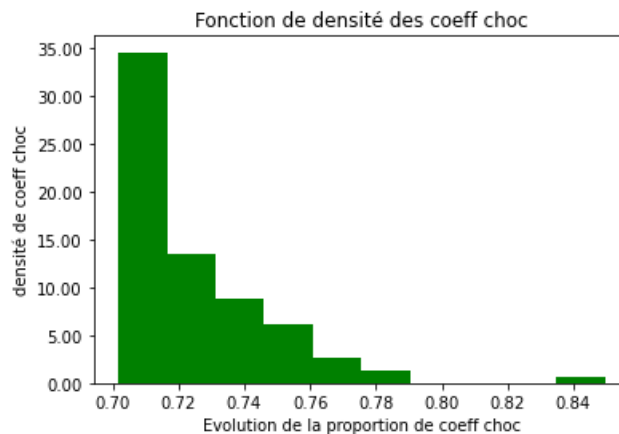


Figure 62 Densité des coefficients de chocs obtenus pour l'univers Pareto pour les coupons et Normaux pour le reste

Après avoir effectué un test d'adéquation de Kolmogorov-Smirnov entre cet échantillon et un échantillon avec une loi de Pareto normalisé avec la méthode min-max, la p-value de 0,995 est obtenue, confirmant la similitude de la courbe avec une loi de Pareto.

Ce résultat peut laisser supposer que le choix de la loi de maturité importe peu sur la répartition des chocs, un test d'influence des paramètres va être effectués sur les coupons.

3.2.1.6 Univers de chocs de Pareto pour les coupons et chocs Normaux pour les maturités et les nominaux

Loi de Pareto pour les chocs coupons, loi normale pour les chocs de maturité et nominal.

⁸² Voir partie 2, chapitre 2.1.1.4

Ci-dessous le graphique de densité :

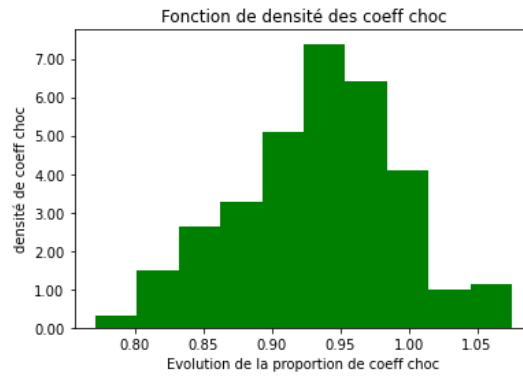


Figure 63 Densité des coefficients de chocs obtenus pour l'univers de choc 6

Après avoir effectué un test d'adéquation de Kolmogorov-Smirnov entre cet échantillon et un échantillon avec une loi de Weibull normalisé avec la méthode min-max, la p-value de 0,855 est obtenue, confirmant la similitude de la courbe avec une loi de Weibull.

Le choix de la loi statistique des coupons et du nominal semble fortement corrélé à l'allure de la répartition des chocs sortis par l'algorithme.

3.2.1.7 Test de l'option de backpropagation à l'aide de chocs infinitésimaux

Le graphique de densité des VM sans modification est le suivant :

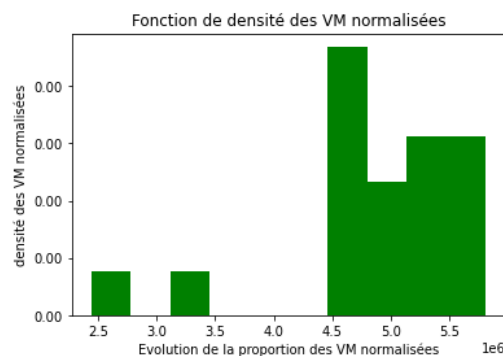


Figure 64 Densité des VM obligataires du portefeuille initial

La durée pondérée de ce portefeuille a été calculée de la manière suivante :

$$Duration_{pondérée} = \frac{\sum_{i=1}^{nb_oblig} VM_i * Duration_i}{\sum_{i=1}^{nb_oblig} VM_i}$$

Avec :

$$Duration = \sum_{i=1}^{maturité_résiduelle_oblig} \frac{i * CF_i}{(1 + rf_i)^i} \frac{CF_i}{(1 + rf_i)}$$

La durée pondérée du portefeuille initial est de 9,34.

Une comparaison de la répartition du portefeuille présenté avec celle d'un portefeuille répliqué est proposée.

Afin de vérifier la pertinence de la backpropagation, un portefeuille a été répliqué avec des chocs infinitésimaux.

La fonction de densité des VM suivante est obtenue :

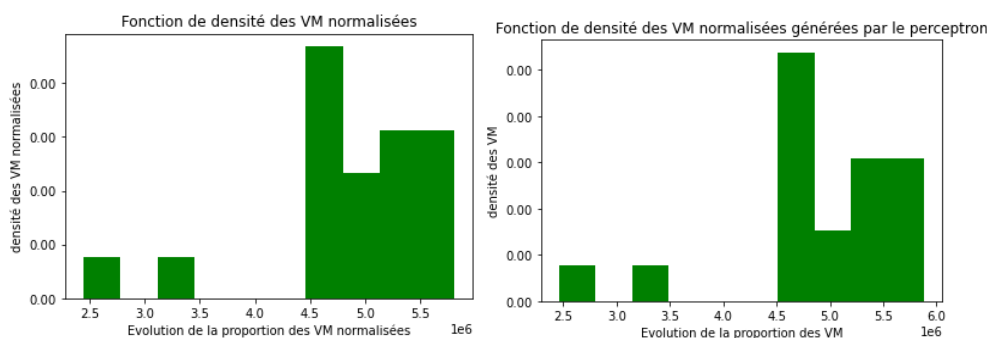


Figure 65 Vérification du bon fonctionnement du réseau de neurone par la réplique exacte du portefeuille initial

Les deux courbes sont identiques.⁸³

L'algorithme développé est capable de répliquer à la perfection un portefeuille donné pour des chocs infinitésimaux.

3.2.1.8 Récapitulatif des différents univers de chocs testés

Pour mieux appréhender la fonction de densité des portefeuilles répliqués, différents scénarios de chocs avec des bornes élargies sont testés.

La suite de la partie se concentre sur des combinaisons de lois qui semblent intéressantes en termes de modulation des chocs. L'allure des densités des coefficients de chocs ainsi que le temps d'implémentation des méthodes sont analysées.

Les sept univers de chocs suivants vont être testés⁸⁴:

Univers ⁸⁵	Pareto	Normal	Uniforme	Weibull	Pareto /Normal	Normal /Uniforme	Normal /Weibull
Type de loi Coupons	Pareto	Normale	Uniforme	Weibull	Pareto	Normale	Normale
Type de loi Nominal	Pareto	Normale	Uniforme	Weibull	Normale	Uniforme	Weibull
Type de loi Maturité	Pareto	Normale	Uniforme	Weibull	Normale	Normale	Normale
Temps de réplique ⁸⁶	53.85	19.24	23	23.44	21.64	21,9	47,42
Type de Symétrie	Centrée à gauche	Centrale	Aucune	Centrale à droite	Centrale à droite	Aucune	Centrale

Figure 66 Présentation d'un échantillon d'univers choisis pour l'utilisation du réseau de neurone générateur de choc

Par sa densité centrée autour de la moyenne, la loi normale est la plus optimale sur les chocs de type « centraux ». Par ces mêmes caractéristiques, elle est la loi la plus simple à optimiser pour chaque type de choc. Cependant dans certaines situations asymétriques, d'autres combinaisons de lois pourront être utilisées telles que la loi de Weibull ou la loi normale mélangée à une loi de Pareto offrant des belles dissymétries des coefficients de chocs produits.

⁸³ L'annexe partie 6, chapitre 6.6.3 présente la possibilité de répliquer exactement les mêmes densités obligataires d'un portefeuille avec des VM ayant perdu 30% de leurs valeurs

⁸⁴ Les courbes de densité obtenues lors de cette étude sont disponibles annexe partie 6, sous partie 6.7

⁸⁵ Une version plus approfondie du tableau est présentée en annexe partie 6, chapitre 6.6.1

⁸⁶ Exprimés en secondes pour la génération de 100 chocs

3.2.1.9 Analyse de la fonction de densité d'un portefeuille obligataire répliqué dans univers de choc Normal

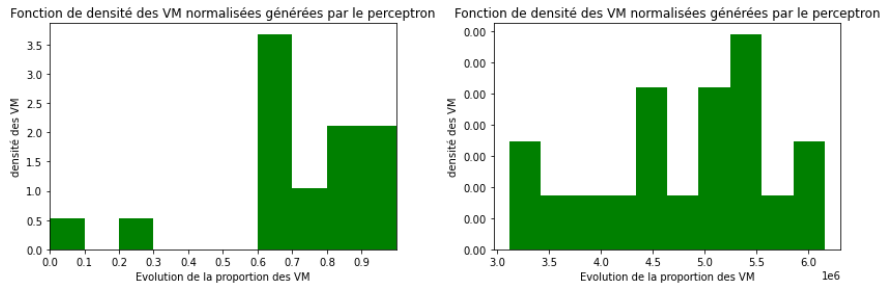


Figure 67 Comparaison de la densité des VM d'un portefeuille répliqué avec l'algorithme de portefeuilles répliquants et le portefeuille initial

La répartition du portefeuille a sensiblement changé.

Les deux analyses menées précédemment permettent de démontrer que la variabilité des portefeuilles créés dépend des bornes et des lois des chocs choisies par l'utilisateur.

3.2.2 Choix de l'allure de répartition des chocs pour la construction de la BDD de ML

Dans cette sous partie, les calibrages du réseau de neurone pour la création des BDD de ML sont présentés.

3.2.2.1 Reproduction de la densité des chocs sur les obligation françaises 10 ans

Avec le retour de l'inflation ce semestre ainsi qu'une remontée des taux obligataires, il a été choisi de répliquer les portefeuilles selon deux scénarios pessimistes.

Utilisant la version française de Prophet, une première calibration de l'algorithme a été effectuée sur les obligations française de maturité 10 ans.

La date de maturité 10 ans fait référence à la durée de vie moyenne d'un portefeuille épargne.

Ci-dessous le graphique des variations des coupons des obligations française 10 ans sur 35 ans



Figure 68 graphique sur l'évolution des coupons obligataires français (source : trading economics)

Le taux de coupons maximum représenté sur cette courbe atteint les 12% lors de la période inflationniste du début des années 1990. Le taux minimum est proche de -1% et le taux de coupon actuel est d'environ 2%.

Le taux de coupon moyen d'une obligation Française est de 4% depuis l'entrée en vigueur de l'euro le 1^{er} Janvier 2002.

Le prix de l'obligation est considéré comme proportionnelle à son taux de coupon possible.

Si les taux directeurs montent (tendance économique de la zone euro actuelle visible sur la courbe), alors les obligations précédemment émises perdront en valeur et vice versa en cas de baisse des taux.

Dans le contexte de baisse des taux et au regard du coupon maximum de l'obligation, la borne inférieure de réplication est définie comme suit :

$$100\% - 12\% + 2\% = 90\%$$

La borne supérieure est définie par :

$$100\% + 2\% - 1\% = 102\%$$

La valeur moyenne des taux de la courbe étant de 4%, un biais de répartition de la courbe est ajouté :

$$2\% - 4\% = -2\%$$

La courbe a une allure quasi linéaire, il sera choisi l'univers de choc Normal dont l'allure de la courbe sera modifiée avec l'ajout des bornes plus restreintes entre 0.9 et 1.02.

Les coefficients de risque neutralisation suivants sont obtenus pour la réplication de 100 portefeuilles :

Quantile	Valeur
0%	0.90
10%	0.92
20%	0.93
30%	0.95
40%	0.97
50%	0.98
60%	0.99
70%	1.00
80%	1.01
90%	1.01
100%	1.02

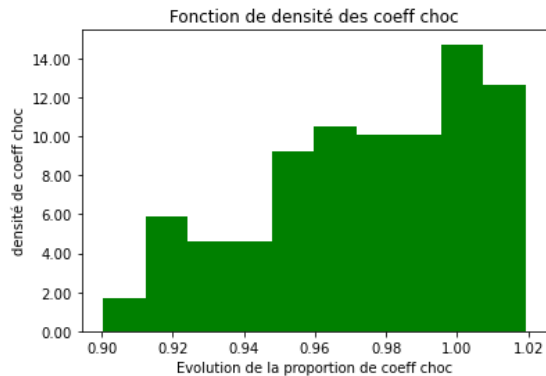


Figure 69 densité des chocs obligataires français

La répartition de la courbe des coupons obligataires français étant relativement centrée et souhaitant effectuer l'étude pour une projection la plus pessimiste maximum, l'historique de la valorisation des coupons des obligations Grecques a été étudiée.

3.2.2.2 Reproduction de la densité des chocs sur les obligations grecques 10 ans

L'étude des obligations grecques s'est centrée sur une analyse de l'évolution des taux de coupons obligataires sur une période de 10 ans.

Ci-dessous le graphique de cette évolution :



Figure 70 graphique sur l'évolution des coupons obligataires grecques (source : trading economics)

Le taux de coupons maximum représenté sur cette courbe atteint les 36% en pleine crise de la dette Grecque (2011).

Le taux minimum est proche de 0% et le taux de coupon actuel est d'environ 3%.

Le taux de coupon moyen d'une obligation Grecques est de 6%.

Le prix de l'obligation est considéré comme proportionnel à son taux de coupon.

Si les taux directeurs montent (tendance économique de la zone euro actuelle visible sur la courbe), alors les obligations précédemment émises perdront en valeur et vice versa en cas de baisse des taux.

Dans le contexte de baisse des taux et au regard du coupon maximum de l'obligation, la borne inférieure de répliation est définie comme suit :

$$100\% - 36\% + 3\% = 67\%$$

La borne supérieure est définie par :

$$100\% + 3\% - 0\% = 103\%$$

La valeur moyenne des taux de la courbe étant de 6%, le biais de répartition de la courbe suivant est ajouté :

$$3\% - 6\% = -3\%$$

La courbe ayant une allure fortement dissymétrique, avec une queue de distribution très légère à gauche et beaucoup plus marquée à droite, un neurone avec une loi uniforme est implémenté pour effectuer une queue légère à droite et un autre neurone avec une loi normale est utilisé pour le reste des simulations.

Les coefficients de chocs suivants sont obtenus pour la réplication de 100 portefeuilles :

Quantile	Valeur
0%	0.67
10%	0.91
20%	0.94
30%	0.95
40%	0.96
50%	0.97
60%	0.98
70%	0.99
80%	1.00
90%	1.02
100%	1.03

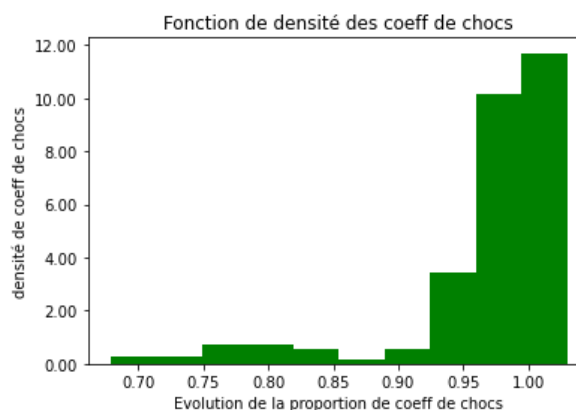


Figure 71 densité des chocs obligataires grecques

Les chocs suivent la densité souhaitée.

3.3 Construction des bases de données de Machine Learning et présentation du cadre expérimental

Dans cette partie, les paramétrages utilisés pour la production des BEL à prédire sont présentés. Dans un premier temps les paramétrages pour générer les databases de ML sont donnés. Dans un second temps les différents inputs obligataires retenus pour l'étude sont dévoilés.

3.3.1 Récapitulatifs des différents chocs implémentés

Les trois scénarios de taux exprimés ci-dessous ont été implémentés :

- Un scénario central
- Deux scénarios de choc de la courbe des taux de +/-20 bps

Afin de garder de l'homogénéité dans les résultats et de capter les variations propres aux changements de la courbe des taux, seules les courbes de taux EIOPA changeront entre les scénarios +/- 20bps.

L'hyperparamétrage du neurone change seulement pour les scénarios Français et Grecs.

Deux générateurs réglés différemment ont produits la densité de choc Grecs. Un générateur pour la queue de distribution à gauche simulée pour 10% des variables du dataset et un autre pour le reste des 90% des chocs restants.

Les univers suivants ont été implémentés :

Univers ⁸⁷	Français	Grec queue distribution à gauche	Grec autres parties de la distribution
Type de loi Coupons	Normale	Uniforme	Gaussienne
Type de loi Nominal	Normale	Uniforme	Gaussienne
Type de loi Maturité	Normale	Uniforme	Normale

Figure 72 Paramètres ayant permis la construction des densités de chocs obligataires des bases grecques et français

⁸⁷ Un tableau plus détaillé du paramétrage est disponible en annexe partie 6, chapitre 6.8

3.3.2 Les variables d'inputs créées par les algorithmes

Afin de garantir les meilleures prédictions possibles, sept variables d'entrées obligatoires ont été implémentées.

Parmi les sept variables explicatives développées, deux sont indépendantes de la maturité des obligations, deux sont à la fois centrées sur la maturité et les valeurs quantitatives des obligations (coupons et nominaux), enfin les trois derniers indices vont se concentrer majoritairement sur la maturité résiduelle des portefeuilles.

Les deux indicateurs quantitatifs choisies sont la somme des coupons et nominaux par portefeuilles créés.

3.3.2.1 La variable explicative de somme des coupons par portefeuille

$$Somme_{coupons} = \sum_{i=1}^{nb_obligations} Coupons_i$$

3.3.2.2 La variable explicative de somme des nominaux par portefeuille

$$Somme_{nominaux} = \sum_{i=1}^{nb_obligations} Nominaux_i$$

Ces variables explicatives étant indépendantes de la valeur temps, des indices mettant à égalité les variables quantitatives et temporelles ont été implémentées.

3.3.2.3 La variable explicative de valeur de marché par portefeuille

$$VM_{portefeuille_répliqué} = \sum_{i=1}^{nb_oblig} VM_répliquée_i$$

Avec :

$$VM_répliqué_i = \sum_{i=1}^{maturite_résiduelle_oblig} \frac{CF_répliqués_i}{(1 + rf_i)^i}$$

Afin de compléter cet input avec la valeur de marché initiale du portefeuille répliqué (qui est toujours la même), deux variables prenant en compte le coefficient pondéré de chocs ainsi que le spread obligataire ont été implémentées sur les portefeuilles répliqués.

3.3.2.4 La variable explicative de coefficient pondéré de chocs par portefeuille

$$Coeff_{choc} = \frac{VM_{portefeuille_répliqué}}{VM_{portefeuille_initial}}$$

Avec :

$$VM_{portefeuille_initial} = \sum_{i=1}^{maturite_résiduelle_oblig} \frac{CF_initiaux_i}{(1 + rf_i)^i}$$

3.3.2.5 La variable explicative de coefficient de spread pondéré par portefeuille

$$Spread_{pondéré} = \frac{\sum_{i=1}^{nb_oblig} Spread_i * VM_répliquée_i}{\sum_{i=1}^{nb_oblig} VM_répliquée_i}$$

Avec le calcul du spread s obligatoire solution de l'équation :

$$VM_{portefeuille_initial} = \sum_i \frac{CF_{portefeuille_répliqué_i}}{(1 + rf_i + Spread)^i}$$

Cette équation a été résolue à l'aide de l'algorithme d'optimisation Nelder-Mead⁸⁸ disponible dans la bibliothèque Scipy de Python.

Une réflexion sur des input à dominante temporelle a été effectuée.

Des variables explicatives de maturité résiduelle pondérée et de duration pondérée ont été développées.

3.3.2.6 La variable explicative de maturité résiduelle pondérée

$$Maturité_{pondérée} = \frac{\sum_{i=1}^{nb_oblig} VM_répliquée_i * Maturité_répliquée_i}{\sum_{i=1}^{nb_oblig} VM_répliquée_i}$$

Avec :

$$VM_répliqué_i = \sum_{i=1}^{maturité_résiduelle_oblig} \frac{CF_répliqué_i}{(1 + rf_i)^i}$$

3.3.2.7 La variable explicative de duration pondérée

Cet input ne faisant pas entièrement rentrer en compte la sensibilité à une variation de taux des obligations, un coefficient de duration pondérée a été développé :

$$Duration_{pondérée} = \frac{\sum_{i=1}^{nb_oblig} VM_répliquée_i * Duration_répliquée_i}{\sum_{i=1}^{nb_oblig} VM_répliquée_i}$$

Avec :

$$Duration_répliquée = \sum_{i=1}^{maturité_résiduelle_oblig} \frac{i * CF_répliquée_i}{(1 + rf_i)^i} \frac{CF_répliquée_i}{(1 + rf_i)}$$

La création de la fonction de calcul de la duration pour chaque obligation et sa pondération par portefeuille a permis le calcul de la duration pondéré du portefeuille de départ, et de créer les différents univers cités en début de partie.

3.3.2.8 Répartition de l'actif

Un portefeuille d'assurance vie se compose en moyenne d'une très forte proportion d'obligations (65% environ). Les actions représentent 20% des parts restantes, l'immobilier 6% et 9% pour les autres liquidités qui seront considérées comme du cash.⁸⁹

Pour effectuer les estimations des BEL, il a été choisi de se placer sur un portefeuille épargne ayant un encours d'un peu plus de 120 millions d'euros.

Le taux de coupon du portefeuille obligataire a été fixé à 1,08%⁹⁰ et Les frais à 0,65%.

Par simplification du modèle, les produits dérivés tels que les puts, floors, swaptions... n'ont pas été pris en comptes.

⁸⁸ Voir partie 2, chapitre 2.2.3.5

⁸⁹ France assurance (2021). *L'assurance Française, Données clefs 2020*

⁹⁰ D'après le rapport sur les résultats annuels de CNP Assurances 2021, les nouvelles productions financières au 31 décembre 2021 sur les produits à taux fixe brut de frais

Les actifs étudiés ont été considérés comme directement détenus par l'assureur et non « par un intermédiaire » comme il est le cas pour les Organismes de Placements Commun à Valeurs Mobilières (OPCVM) ⁹¹par exemple.

Le portefeuille qui permettra l'adossement du passif sera composé de la manière suivante :

Répartition des actifs du portefeuille initial

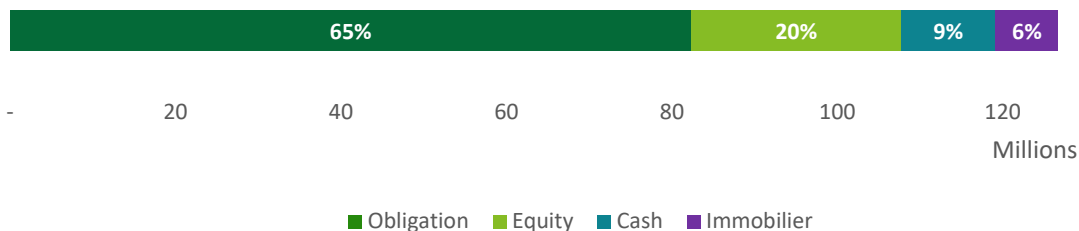


Figure 73 répartition des actifs des BDD pour le ML

Le portefeuille obligataire initial (qui va ensuite être répliqué) est initialisé avec des échelons de crédits à 1, c'est-à-dire que les obligations le composant disposent d'une notation AA⁹².

3.3.2.9 Extraction des données

Une fois le portefeuille créé, les simulations Python effectuées et les runs Prophet lancés, une extraction des BEL à l'aide d'une maquette de reporting Prophet est effectuée.

La variable à prédire extraite de Prophet est le BEL et sa NAV associée.

Pour l'étude du SCR, la NAV se calcule de la manière suivante :

$$NAV = PVFP + PMVL$$

Une fois l'extraction terminée, la base d'input Prophet a l'allure suivante :

Index ⁹³	Coupon	Nominal	VM Oblig Python	Choc obligataire	Spread pondéré	Maturité pondérée	Duration pondérée	NAV	VM Prophet	BE
0	800	68 325	77 143	84%	3%	10,37	9,39	5 180	130 846	136 026
1	692	63 405	70 734	77%	5%	10,10	9,23	-33 433	124 788	158 221
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10 000	702	67 816	74 981	81%	3%	9,84	9,06	8 891	129 052	137 943

Figure 74 Inputs ML obligataires

Afin de garantir une étude optimale des algorithmes, différentes bases de données avec différents passifs et univers économiques sont prédites.

3.3.2.10 Présentation des différentes BDD pour les travaux de prédiction de l'étude

Pour cette étude dix bases de données ont été retenues. Ces dix bases de données peuvent être divisées en 5 grands groupes auxquels ont été ajoutés des sous variations comme décrit dans le tableau ci-dessous :

⁹¹ Voir annexe partie 6, chapitre 6.2.3

⁹² Voir partie 1, chapitre 1.3.6

⁹³ Les valeurs Coupon, Nominal, VM oblig, NAV, MV Prophet et BE sont exprimées en k€

Type de base de données	Caractéristiques communes	Spécificités	Nom de la base
Base de données chocs obligataires	Chocs obligataires GSE EIOPA similaires pour chaque simulation	1 TMG unique	Base de données chocs obligataires
		3 groupes de TMG	Base de données chocs obligataires TMG
Base de données chocs obligataires et GSE variant	GSE variant pour chaque simulation sans variation d'allocation d'actif Chocs obligataires	1 TMG unique	Base de données chocs obligataires et GSE variant
		3 groupes de TMG	Base de données chocs obligataires et GSE variant TMG
Base de données GSE variant	GSE variant pour chaque simulation sans variation d'allocation d'actif Pas de chocs obligataires	1 TMG unique	Base de données GSE variant
		3 groupes de TMG	Base de données GSE variant TMG
Base de données GSE variant et indicateurs obligataires	GSE variant pour chaque simulation sans variation d'allocation d'actif Ajout d'indicateurs obligataires	1 TMG unique	Base de données GSE variant et indicateurs obligataires
		3 groupes de TMG	Base de données GSE variant et indicateurs obligataires TMG
Base de données GSE variant, chocs obligataires et réallocation dynamique	Chocs obligataires GSE variant pour chaque simulation avec variation d'allocation d'actif	1 TMG unique	Base de données GSE variant, chocs obligataires et réallocation dynamique
		3 groupes de TMG	Base de données GSE variant, chocs obligataires et réallocation dynamique TMG

Figure 75 Présentation des BDDs

Un TMG à 0% a été défini pour les simulations à TMG unique.

Afin de challenger les algorithmes et de voir les variations de prédictions dues à l'ajout de complexité dans le portefeuille, des simulations avec trois groupes de TMG ont été ajoutées, comme présenté ci-dessous :

TMG	Proportion
0%	50%
0,50%	25%
1%	25%

Figure 76 Répartition des TMG pour les databases à trois pools

Les algorithmes ont été développés selon une logique orientée objet, il est possible de créer des indicateurs obligataires basés sur le déflateur de chaque simulation stochastique.

Cette modification a permis la création et l'ajout de variables explicatives obligatoires à la Base de données à GSE variant. Cette modification va permettre de comparer les résultats obtenus par les algorithmes de ML avec et sans l'ajout de ces inputs.

Dans la Base de données GSE variant, chocs obligataires et réallocation dynamique, les chocs d'allocations d'actifs correspondent à une modification de la répartition de l'actif variant entre différentes bornes fixées par l'utilisateur.

Dans les chocs simulés, l'allocation des obligations varie entre 50 et 80%, l'allocation des actions entre 8 et 30%, l'allocation de l'immobilier entre 3 et 10% et l'allocation de cash entre 5 et 15%.

Le choc attribué aux valeurs de marchés obligataires pour changer la part d'allocation des actifs est compris entre 60 et 130%, le choc attribué pour les actions est compris entre 30 et 120% et le choc attribué pour l'immobilier est compris entre 30 et 120%.

L'étude va d'abord se focaliser sur le BE et s'étendre aux SCR obligataires et de taux et leur agrégation.

Chapitre 4

4 Analyse des résultats

Une fois les tables de chocs obligataires connectées à Prophet et le perceptron calibré de telle sorte à avoir une réplique des portefeuilles optimaux, il est désormais possible de lancer les algorithmes prédictifs.

Afin de réduire les écarts de prédictions aux quantiles extrêmes, le but de la prochaine partie va être d'utiliser le générateur obligataire résumé dans les parties précédentes, afin d'enrichir les BDD pour la prédiction du BE.

Cette amélioration se fera avec l'aide d'une série de portefeuilles obligataires simulés de manière réalistes. L'ajout d'inputs obligataires pertinents devraient sensiblement améliorer les prédictions.

Une méthode de sélection des variables et de cross-validation sont implémentées.

La target d'erreur fixée est d'environ 1% d'écart maximum pour tous les quantiles de BE situés aux quantiles entre 10% et 90%, ainsi que pour le SCR obligataire. Le SCR de taux et son agrégation avec le SCR obligataire sont aussi prédits dans cette partie.

L'analyse de la Base de données GSE variant ayant déjà effectuée, les résultats obtenus par les algorithmes nouvellement calibrés sont présentés en annexe.⁹⁴

4.1 Présentation des résultats sur les Bases de données avec les chocs obligataires uniquement activés

Dans cette partie le processus de production des résultats prédits par les algorithmes de Machine Learning, la construction et l'analyse des résultats prédictifs sont décrits.

La démarche suivie dans cette partie est présentée ci-dessous :



Figure 77 Processus de production des résultats des algorithmes de ML

4.1.1 Présentation des résultats de la Base de données avec chocs obligataires et GSE unique avec un et trois pools de TMG

Dans cette partie, les résultats pour la Base de données chocs obligataires sont présentés.

Les algorithmes de Machine Learning retenus pour cette base de données sont les suivants :

⁹⁴ Voir annexe partie 6, sous parties 6.20 et 6.21

- Une régression linéaire multiple avec constante
- Un Random Forest
- Un Gradient Boosting
- Une régression XGboost
- Un SVM par la possibilité de sélectionner une base d'apprentissage et de test de petites dimensions (3 à 4 variables).

4.1.1.1 Préambule

Pour l'hyperparamétrage des algorithmes une cross validation k-fold ⁹⁵est implémentée.

Le critère à minimiser pour cette cross validation est la moyenne des quantiles absolus des écarts en % entre les prédictions et les valeurs du modèle.

Ce critère a été calculé de la manière suivante :

$$Moyenne_{\text{écart}} = \frac{1}{nb_{\text{quantile}}} * \sum_{i=0}^n |Ecart_quantile_i|$$

Avec :

$$Ecart_quantile_i = \frac{Val_modèle_quantile_i - Prediction_quantile_i}{Val_modèle_quantile_i}$$

Un pas de quantile de 5% a été choisi pour la Cross-Validation.

Une fois les résultats sur la base train et test obtenus, d'autres résultats basés sur une base de données ont été prédits⁹⁶.

Les chocs représentatifs des OAT françaises étant compris dans la borne des chocs obligataires Grecs, les résultats attendus devraient être proche des résultats de test si les algorithmes de Machine Learning sont bien calibrés ou/et les variables d'apprentissages pertinentes.

La présentation des résultats se fait en deux parties pour les BDD chocs obligataires avec un et trois pools de TMG. La première partie de l'étude se consacre à l'analyse du contexte économique dans lequel évoluent les variables.

La deuxième partie se focalise sur l'analyse des résultats sorties par les algorithmes.

Les résultats sont présentés pour le dataset test avec un commentaire sur le dataset français simulées selon les conditions de la Base de données chocs obligataires avec un et trois pools de TMG.

L'analyse du contexte économique fait intervenir les éléments suivants :

- Un graphique sur l'évolution de la production financière et de l'engagement assureur⁹⁷
- Un graphique sur l'évolution des BE en fonction des VM par percentiles
- Les matrices de corrélation des données

L'analyse des résultats fait intervenir les éléments suivants :

⁹⁵ Partie 2, chapitre 2.2.1

⁹⁶ Partie 3, sous partie 3.2.2

⁹⁷ Avec engagement assureur=IT + dotation PPB - reprise PPB

- Un tableau présentant les prédictions des algorithmes sur la database test pour le BE minimum, médian, moyen et maximum
- Un graphique présentant l'allure des BE prédits sur le dataset test par percentile
- Deux graphiques présentant les écarts des prédictions aux quantiles 10 et 90% du BE sur le dataset test
- Un graphique présentant les RMSE après normalisation des données par algorithmes sur le dataset test.

4.1.1.2 Analyse de l'univers économique Base de données chocs obligataires avec un pool de TMG à 0%

La Base de données chocs obligataires avec un pool de TMG à 0% résulte d'un GSE constant pour chaque trajectoire choquée, avec des chocs obligataires activés et un pool de TMG à 0%.

Analyse du BE en fonction de la VM

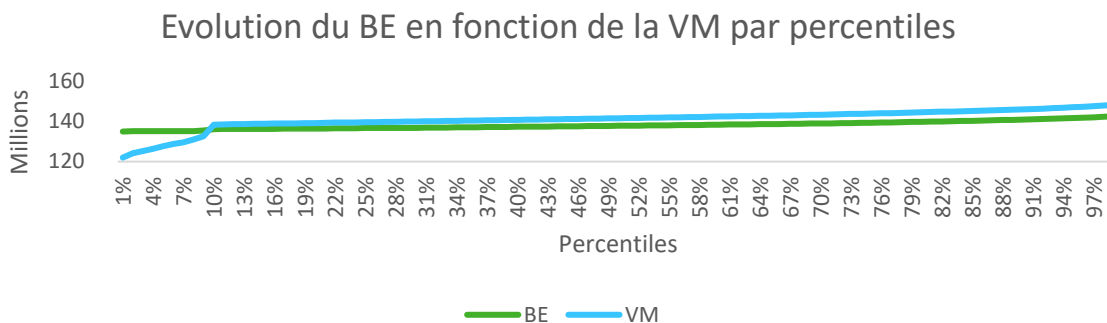


Figure 78 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires avec 1 pool de TMG

La cassure au quantile des 10% des VM les plus extrêmes peut s'expliquer par la répartition des coefficients de choc⁹⁸ qui ont été produits par une couche impliquant deux générateurs différents.

Le BE croît lorsque la VM lui est supérieur, ce qui semble logique au vu de l'asymétrie des engagements assureurs qui doit reverser à minima 85% de sa production financière.

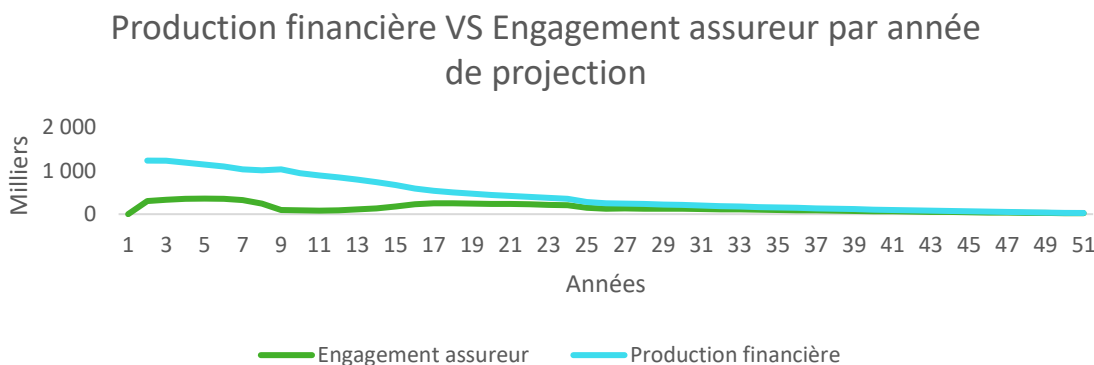


Figure 79 Production financière VS l'engagement de l'assureur Base de données chocs obligataires avec 1 pool de TMG

⁹⁸ Voir partie 3, chapitre 3.2.2.2

La production financière étant toujours supérieure aux engagements assureurs, le BE devrait être linéairement corrélé avec la VM.

L'analyse des matrices des corrélations de Pearson et Kendall ⁹⁹ confirment les hypothèses définies précédemment et montrent une forte corrélation de toutes les variables avec le BE, hormis les variables « temps » (Maturité et Duration pondérée).

N'ayant que sept inputs obligatoires pour l'étude de la Base de données chocs obligataires, il n'y a pas de sélection d'inputs pour les algorithmes de Machine Learning sauf, pour le SVM efficace dans des bases de données à très faibles dimensions (3 à 4 variables explicatives maximum).

Pour la sélection des variables, il a été choisi de croiser une analyse des corrélations de Pearson (corrélation linéaire) et de Kendall (corrélation de rang).

Sur les deux matrices de corrélation ¹⁰⁰, les variables coupon, maturité et duration pondérée sont très décorréliées de la variable cible (BE).

Malgré une forte corrélation entre elles, les variables Nominal, VM obligataire python, Coefficient de choc et Spread pondéré ont été retenus pour l'application d'un SVM

4.1.1.3 Présentation des résultats pour la BDD à un pool de TMG

Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données chocs obligataires avec un pool de TMG est le suivant :

Modèle ¹⁰¹ 102		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	135	131	3,2%
	Moy	138	138	> 0,1%
	Med	138	138	-0,3%
	Max	143	141	1,0%
Random Forest	Min	135	135	> 0,1%
	Moy	138	138	> 0,1%
	Med	138	138	> 0,1%
	Max	143	143	0,1%
Gradient Boosting	Min	135	135	-0,1%
	Moy	138	138	> 0,1%
	Med	138	138	> 0,1%
	Max	143	142	0,3%
XGboost	Min	135	135	> 0,1%
	Moy	138	138	> 0,1%
	Med	138	138	> 0,1%
	Max	143	143	0,2%
SVM	Min	135	135	-0,3%
	Moy	138	138	-0,1%
	Med	138	138	> 0,1%
	Max	143	143	0,1%

Figure 80 Tableau récapitulatif des performances des algorithmes de ML Base de données chocs obligataires avec 1 pool de TMG

⁹⁹ Voir partie 2, chapitres 2.1.3.1 et 2.1.3.2

¹⁰⁰ Voir annexe, partie 6, chapitre 6.10.2

¹⁰¹ Tous les chiffres issus de ce type de tableau dans la partie 5 sont exprimés en M€

¹⁰² Tous les algorithmes avec des lignes mises en gras dans ce type de tableau prédisent des résultats dans la target fixée

La régression linéaire avec constante semble avoir le pouvoir prédictif le plus faible sur cette base de données, vient ensuite le SVM.

La courbe des BE du dataset test prédits par percentile est la suivante :

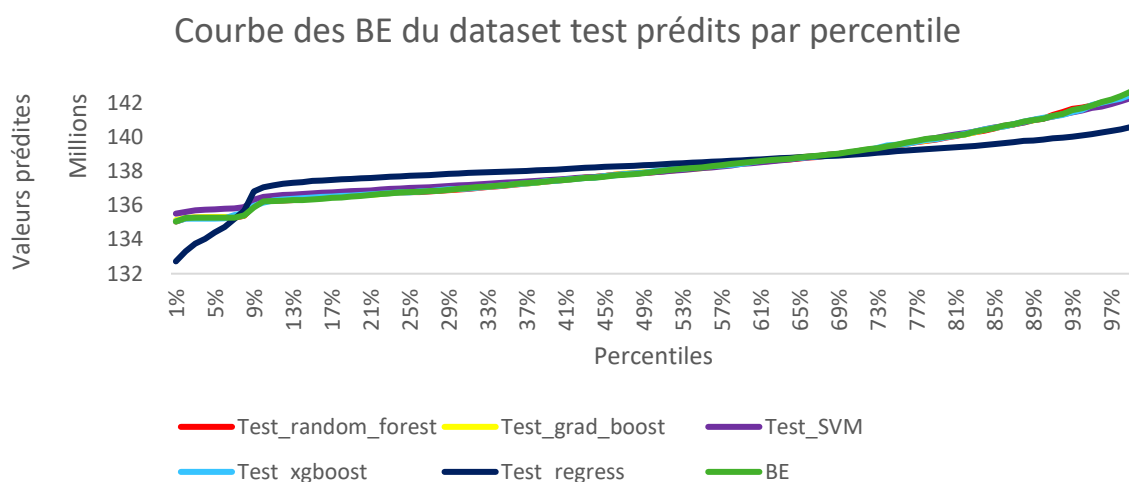


Figure 81 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires avec 1 pool de TMG

Cette courbe permet d'affirmer les premières hypothèses émises lors de l'analyse du tableau précédent. Les courbes violettes et bleues foncées (SVM et régression linéaire multiple) se détachent de la courbe du BE. Pour les autres algorithmes, ils prédisent quasi parfaitement le BE.

La courbe des écarts aux quantiles 10 et 90% confirme ces observations :

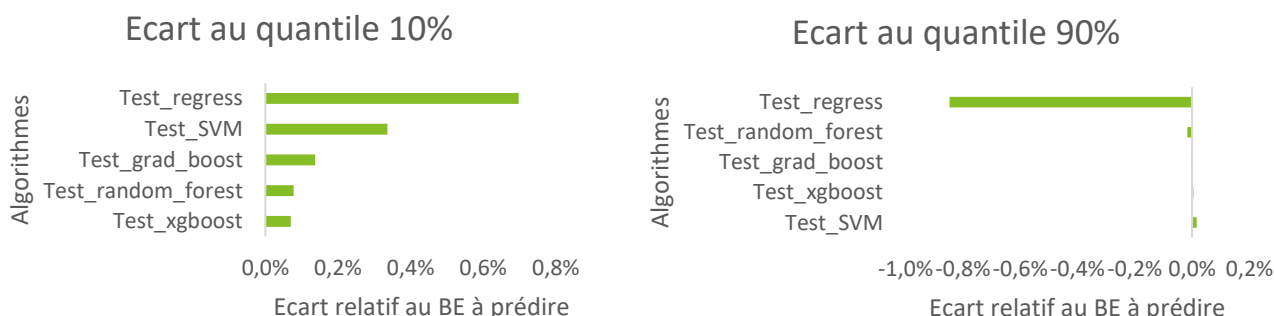


Figure 82 Ecart aux quantiles 10% et 90% par algorithmes Base de données chocs obligataires avec 1 pool de TMG

Le graphique des RMSE est présenté afin de comparer les performances des algorithmes entre eux.

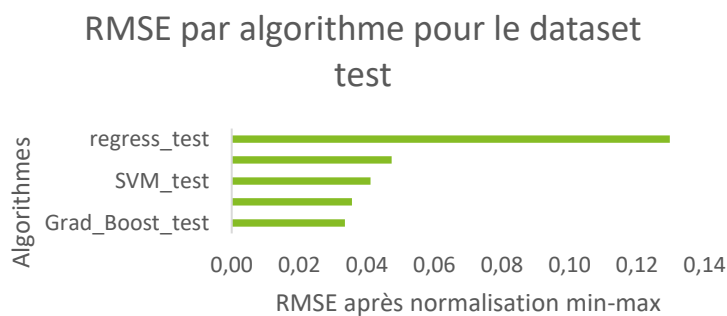


Figure 83 RMSE par algorithme Base de données chocs obligataires avec 1 pool de TMG

Le tableau des RMSE confirme que les algorithmes à retenir pour l'estimation du BEL dans ce cas sont le Gradient Boosting, le XGboost et le Random Forest.

Les temps d'apprentissage et de prédiction sont tous inférieurs à 2 secondes.¹⁰³

Sur la Base de données françaises¹⁰⁴, les algorithmes de Random Forest, Gradient Boosting et XGboost continuent de prédire les BE avec des écarts à environ 1%.

Cela signifie que les algorithmes ont été bien calibrés car ils obtiennent de bonnes prédictions pour une base de données avec une densité de VM obligataires fortement modifiées.

4.1.1.4 Analyse de l'univers économique Base de données chocs obligataires avec trois pools de TMG

Analyse du BE

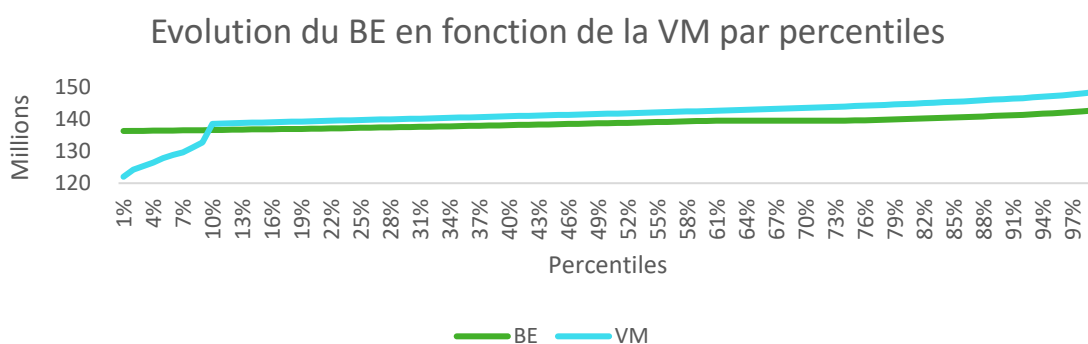


Figure 84 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires avec 3 pools de TMG

Le VM suit la même trajectoire que le graphique de la Base de données chocs obligataires avec un pool de TMG.

Le BE suit une trajectoire très proche du BE avec un pool de TMG à 0% (voir graph un peu plus haut). Les deux seules différences avec l'autre courbe des BE sont la « linéarisation » de la cassure qui était présente au quantile 10% et une position plus haut dans le graphique de la courbe des BE par l'ajout des TMG.

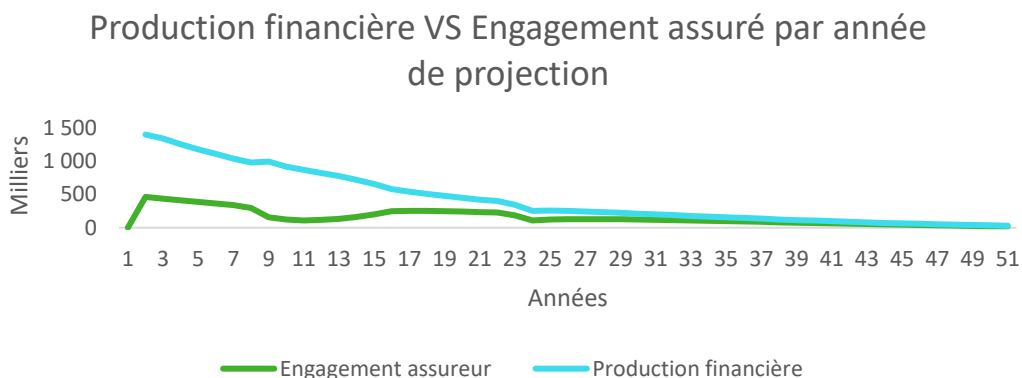


Figure 85 Production financière VS l'engagement de l'assureur Base de données chocs obligataires avec 3 pools de TMG

¹⁰³ Voir annexe partie 6, chapitre 6.10.1

¹⁰⁴ Voir annexe partie 6, chapitre 6.10.4

La production financière reste largement suffisante pour répondre au TMG bien que l'engagement assureur a été sensiblement augmenté par rapport à l'univers avec un pool de TMG.

En raison de l'augmentation des engagements assureurs par l'ajout de TMG, des corrélations moins fortes entre le BE et les valeurs de marchés sont attendues.

Les corrélations¹⁰⁵, sont moins fortes par rapport à la Base de données chocs obligataires avec un pool de TMG, cependant les corrélations non linéaires restent fortes.

La tendance des corrélations entre variables reste la même par rapport à la Base de données chocs obligataires, donc la sélection des variables est la même pour le SVM.

4.1.1.5 Présentation des résultats

Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données chocs obligataires avec trois pools de TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	136	136	0,3%
	Moy	139	139	> 0,1%
	Med	139	139	> 0,1%
	Max	143	140	1,7%
Random Forest	Min	136	136	> 0,1%
	Moy	139	139	> 0,1%
	Med	139	139	> 0,1%
	Max	143	143	> 0,1%
Gradient Boosting	Min	136	136	-0,2%
	Moy	139	139	> 0,1%
	Med	139	139	0,1%
	Max	143	142	0,3%
XGboost	Min	136	137	-0,4%
	Moy	139	139	> 0,1%
	Med	139	139	0,1%
	Max	143	142	0,3%
SVM	Min	136	137	-0,4%
	Moy	139	139	> 0,1%
	Med	139	139	> 0,1%
	Max	143	143	-0,1%

Figure 86 Tableau récapitulatif des performances générales des algorithmes de ML Base de données chocs obligataires avec 3 pools de TMG

La régression linéaire avec constante semble avoir le pouvoir prédictif le plus faible sur cette Base de données.

La courbe des BE du dataset test prédits par percentile est la suivante :

¹⁰⁵ Voir annexe partie 6, chapitre 6.11.2

Courbe des BE du dataset test prédits

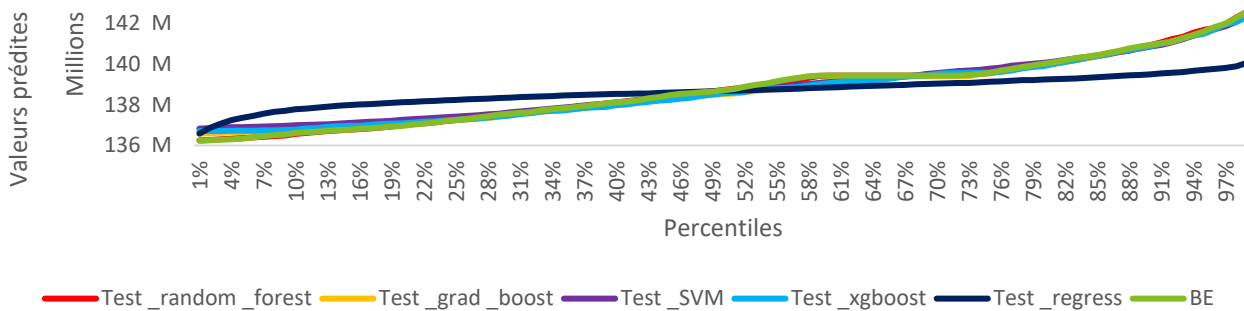


Figure 87 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires avec 3 pools de TMG

A l'instar de la Base de données chocs obligataires avec un pool de TMG, la régression linéaire multiple a un très faible pouvoir prédictif par rapport aux autres algorithmes (courbe bleu foncé).

Les autres algorithmes ont un fort pouvoir de prédiction.

Bien que moins marquée par rapport à la Base de données chocs obligataires avec un pool de TMG, le SVM reste fragile par rapport aux autres algorithmes pour la prédiction aux quantiles 10 et 90%¹⁰⁶, son amélioration de prédiction au quantile 10% pour cette base de données est probablement dû au « lissage » du BE aux quantiles inférieurs à 10% par l'ajout d'un TMG non nul.

L'analyse du graphique des RMSE¹⁰⁷ permet d'aboutir aux mêmes constats, les algorithmes XGboost, Gradient boost et du Random Forest obtiennent des meilleurs résultats que les autres.

Les temps d'apprentissage¹⁰⁸ et de prédictions sont tous inférieurs à 2 secondes.

Sur la Base de données français, les algorithmes de Random Forest, Gradient Boosting et XGboost continuent de prédire les BE avec des écarts bien inférieurs à 1%.

Cela signifie que les algorithmes ont été bien calibrés car ils obtiennent de bonnes prédictions pour une base de données avec une densité de VM obligataires fortement modifiées.

L'analyse de la Base de données chocs obligataires donne des résultats très encourageant quant à la production des chocs et à la création des indicateurs obligataires.

L'étude se complexifie avec un changement dans les GSE du modèle.

4.1.2 Présentation des résultats de la Base de données avec chocs obligataires et GSE variant avec un et trois pools de TMG

Dans cette partie, les résultats pour la BDD chocs obligataires et GSE variant sont présentés.

La Base de données chocs obligataires et GSE variant possède les mêmes caractéristiques que la Base de données chocs obligataires à l'exception de son GSE qui est désormais variant pour chaque simulation.

¹⁰⁶ Voir annexe partie 6, chapitre 6.11.3

¹⁰⁷ Voir annexe partie 6, chapitre 6.11.3

¹⁰⁸ Voir annexe partie 6, chapitre 6.11.1

A l'exception du SVM, les mêmes algorithmes de Machine Learning ont été retenus que pour la BDD précédente. L'élimination du SVM fait suite à l'ajout de nombreuses variables explicatives induit par l'utilisation de l'GSE variant.

4.1.2.1 Préambule

La technique de cross-validation est la même que celle utilisée pour l'analyse de la Base de données chocs obligataires.

A l'exception de la sélection des variables explicatives, l'acheminement aux résultats reste le même que pour la Base de données chocs obligataires.

La présentation des résultats est faite de manière similaire à la BDD précédente.

4.1.2.2 Analyse de l'univers économique Base de données chocs obligataires et GSE variant avec un pool de TMG à 0%

La Base de données chocs obligataires et GSE variant avec un pool de TMG à 0% résulte d'un GSE variant pour chaque trajectoire choquée, avec des chocs obligataires activés et un pool de TMG à 0%.

Analyse du BE en fonction de la VM

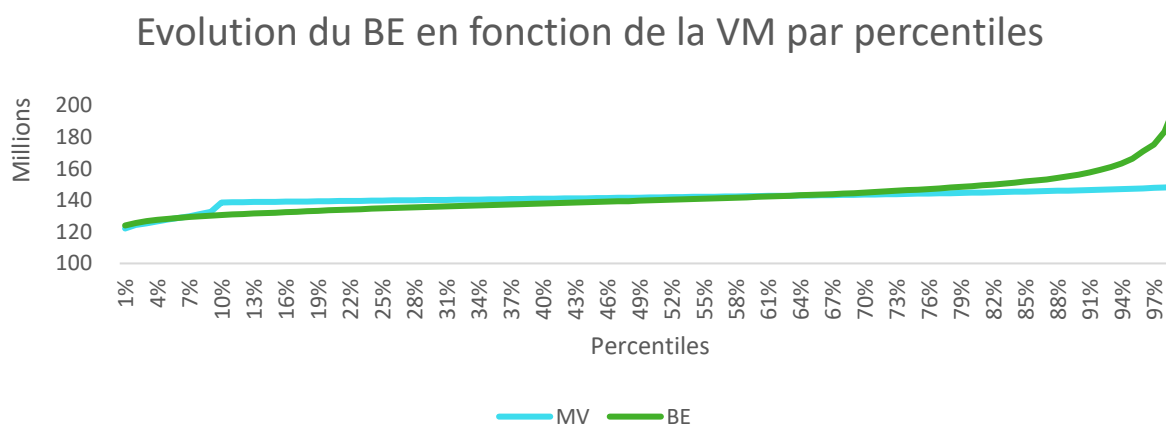


Figure 88 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires et GSE variant avec 1 pool de TMG

Le BE croît avec lorsque la VM lui est supérieur, ce qui semble logique au vu de l'asymétrie des engagements assureurs qui doit reverser à minima 85% de sa production financière.

Cependant, l'ajout d'un GSE variant, basé sur des méthodes stochastiques a introduit certaines trajectoires économiques « extrêmes » qui font décoller les BE du dernier quantile.

La production financière par années de projection VS les engagements assurés a été tracée pour la première trajectoire économique du GSE

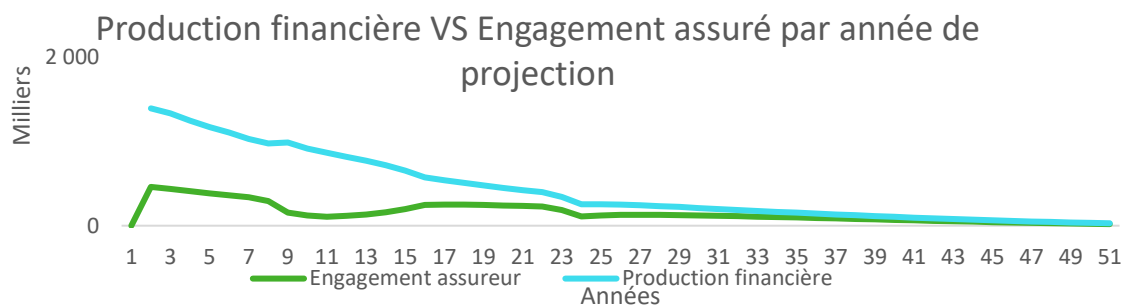


Figure 89 Production financière VS l'engagement de l'assureur Base de données chocs obligataires et GSE variant avec 1 pool de TMG

Une forte variation à la hausse de l'engagement assureur est visible par rapport au graphique de la Base de données chocs obligataires avec un pool de TMG.

Il est possible d'imaginer que cette différence est d'autant plus marquée pour certaines trajectoires économiques ce qui expliquerait la trajectoire du BE sur le graphique.

Le GSE de la Base de données chocs obligataires étant basé sur la courbe des taux EIOPA de décembre 2021, il était prévisible d'avoir des VM plus basses pour la Base de données chocs obligataires et GSE variant dont la production est basée sur une courbe avec des taux supérieurs.

4.1.2.3 Traitement de la base de données des variables explicatives ajoutées par le GSE¹⁰⁹

L'ajout du GSE variant pour chaque trajectoire introduit les quatre types de variables de marché suivant :

- Les rendements actions : « Actions_Y »
- Les taux d'inflation : « Inflation_Y »
- Les rendements immobiliers : « Immobilier_Y »
- Les taux zéros coupons : « ZCB_1_Y », « ZCB_2_Y », ..., « ZCB_30_Y »

La base de données contenant toutes ces variables étant lourde à traiter et ralentissant fortement le temps d'apprentissage, une sélection des variables a été effectuée.

La sélection de ces variables s'est basée sur les cinq analyses des corrélations suivantes :

- L'analyse des corrélations sur les rendements actions
- L'analyse des corrélations sur les taux d'inflation
- L'analyse des corrélations sur les rendements immobiliers
- L'analyse des corrélations inter ZCB
- L'analyse des corrélations intra ZCB

Cette sélection des variables a permis de supprimer :

- 32 variables actions
- 30 variables d'inflation
- 36 variables d'immobilier
- 1450 variables ZCB

¹⁰⁹ Le traitement des variables s'est inspiré du mémoire de Valentin CERISIER (2021). *Application de méthodes de Machine Learning dans le calcul de la solvabilité infra-annuelle*. Dauphine

Douze inputs de Machine Learning ont été créés en moyennant des variables issues d'un même GSE par groupe de corrélation entre variables.¹¹⁰

Les autres bases de données étant basées sur le même GSE variant, la sélection des variables explicatives sur ce dataset reste inchangée.

Comme le montre les graphiques représentant l'importance des variables pour les algorithmes XGboost, Random Forest et Gradient Boosting, l'introduction des moyennes de certaines variables de marché ont amélioré la capacité de prédiction des algorithmes.¹¹¹

La sélection des variables a permis de garder un total de 130 VS 266 variables explicatives sélectionnées lors du mémoire d'Aurélien Suray¹¹² (étude précédente).

Analyse des matrices des inputs obligataires

Comme vu lors de l'analyse du graph il existe une très faible corrélation linéaire¹¹³ entre les inputs obligataires et le BE, cependant une corrélation de rang non négligeable existe entre les inputs Nominal, VM obligataire python, Coefficient de choc, Spread pondéré et le BE, signe que la production financière des actifs a permis d'au moins rémunérer les TMG sur au moins une bonne partie des simulations.

4.1.2.4 Présentation des résultats

Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données chocs obligataires et GSE variant avec un pool de TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	131	129	1,3%
	Moy	143	143	0,1%
	Med	141	141	0,3%
	Max	278	278	20,3%
Random Forest	Min	131	132	-0,9%
	Moy	143	143	> 0,1%
	Med	141	141	> 0,1%
	Max	278	226	18,9%
Gradient Boosting	Min	131	132	-1,1%
	Moy	143	143	> 0,1%
	Med	141	141	> 0,1%
	Max	278	222	20,2%
XGboost	Min	131	133	-1,5%
	Moy	143	143	> 0,1%
	Med	141	141	> 0,1%
	Max	278	241	13,4%

Figure 90 Tableau récapitulatif des performances générales des algorithmes de ML Base de données chocs obligataires et GSE variant avec 1 pool de TMG

Les algorithmes semblent prédire de manière correcte (prédictions dans la target des 1%) les valeurs les plus faibles du BE ainsi que les valeurs médianes et moyennes.

Cependant les algorithmes ont des difficultés à prédire la valeur maximale du BE.

Pour avoir une vision plus globale, la courbe des BE du dataset test prédits par percentile est la suivante :

¹¹⁰ Voir annexe partie 6, sous partie 6.12

¹¹¹ Voir annexe partie 6, chapitre 6.12.7

¹¹² Voir annexe partie 6, chapitre 6.9.1

¹¹³ Voir annexe partie 6, chapitre 6.13.2

Courbe des BE du dataset test prédits par percentiles

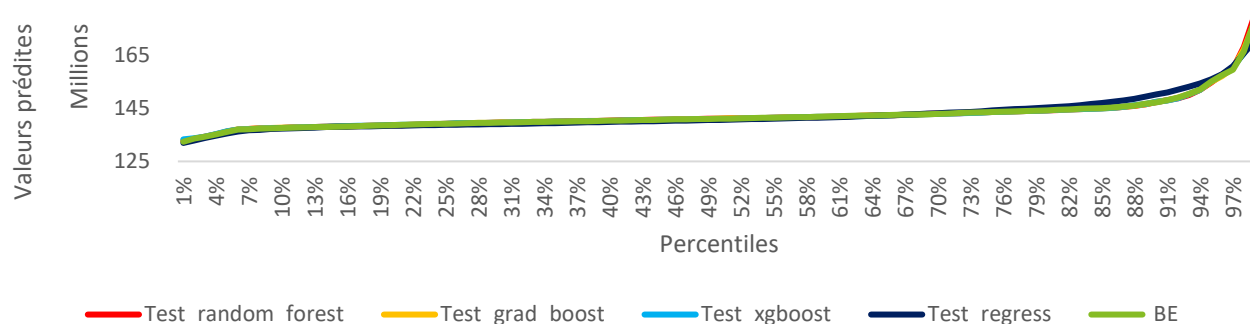


Figure 91 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires et GSE variant avec 1 pool de TMG

Cette courbe permet d'affirmer que seul le percentile allant de 99 à 100% est mal prédit, un tableau des écarts relatifs est présenté ci-dessous :

Modèle ¹¹⁴	Random Forest	Gradient Boosting	XGboost	Régression Multiple
Ecart relatif				
9%	0,05%	0,11%	0,07%	-0,19%
19%	-0,05%	-0,07%	-0,01%	-0,19%
29%	-0,01%	0,03%	-0,02%	-0,37%
39%	0,03%	-0,03%	-0,04%	-0,32%
49%	0,04%	0,03%	-0,07%	-0,28%
59%	-0,06%	-0,03%	-0,05%	-0,24%
69%	0,06%	-0,02%	0,03%	0,14%
79%	0,01%	0,05%	0,01%	0,69%
89%	-0,17%	-0,23%	-0,08%	1,76%
99%	1,08%	-0,40%	-1,95%	-4,54%

Figure 92 Tableau d'écarts aux quantiles de BE, BDD chocs obligataires et GSE variant avec 1 pool de TMG

Comme pour la Base de données chocs obligataires, la régression linéaire a un pouvoir prédictif bien inférieure aux autres algorithmes.

Les graphiques des écarts aux quantiles 10 et 90% ainsi que le tableau des RMSE confirment que le Gradient Boosting et le XGboost ont les meilleurs pouvoir prédictifs.¹¹⁵

Les temps d'apprentissage et de prédiction sont tous inférieur à 20 secondes.¹¹⁶

Sur la base de données française, les algorithmes ont un meilleur pouvoir prédictif que sur la base de données test¹¹⁷. Ce meilleur pouvoir prédictif peut s'expliquer par la fonction de densité des chocs de la base de données française moins « complexe » que les chocs inspirés des obligations Grecques.

Si le dernier percentile des valeurs prédites est omis, les algorithmes XGboost et Random Forest peuvent être utiliser pour prédire les BE dans le contexte économique fixé, une analyse avec une complexification du passif suit.

¹¹⁴ Tous les algorithmes avec des colonnes mises en gras dans ce type de tableau prédisent des résultats dans la target fixée

¹¹⁵ Voir annexe partie 6, chapitre 6.13.3

¹¹⁶ Voir annexe partie 6, chapitre 6.13.1

¹¹⁷ Voir annexe partie 6, chapitre 6.13.4

4.1.2.5 Analyse de l'univers économique Base de données chocs obligataires et GSE variant avec trois pools de TMG

Analyse du BE en fonction de la VM

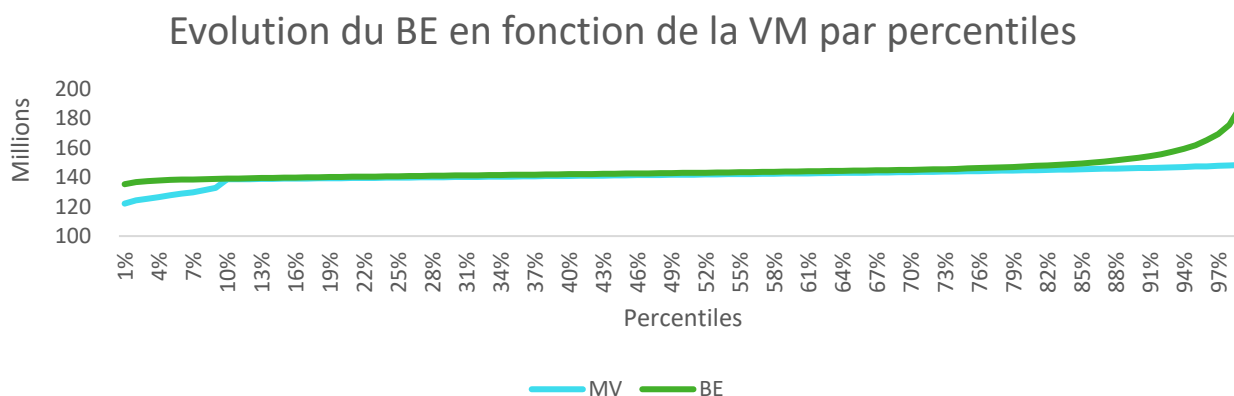


Figure 93 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires et GSE variant avec 3 pools de TMG

Par rapport au graphique de la Base de données chocs obligataires et GSE variant, les BE ont sensiblement augmenté à la suite de l'ajout des TMG non nuls. La VM est désormais quasi toujours en dessous du BE. Ce résultat peut s'expliquer par les chocs obligataires pour la majorité à la baisse et le rendement obligataire fixé à 1,08% relativement faible pour rembourser des TMG moyen à 0,375% hors définition du taux cible.

La production financière par années de projection VS les engagements assurés a été tracée pour la première trajectoire économique de l'GSE :

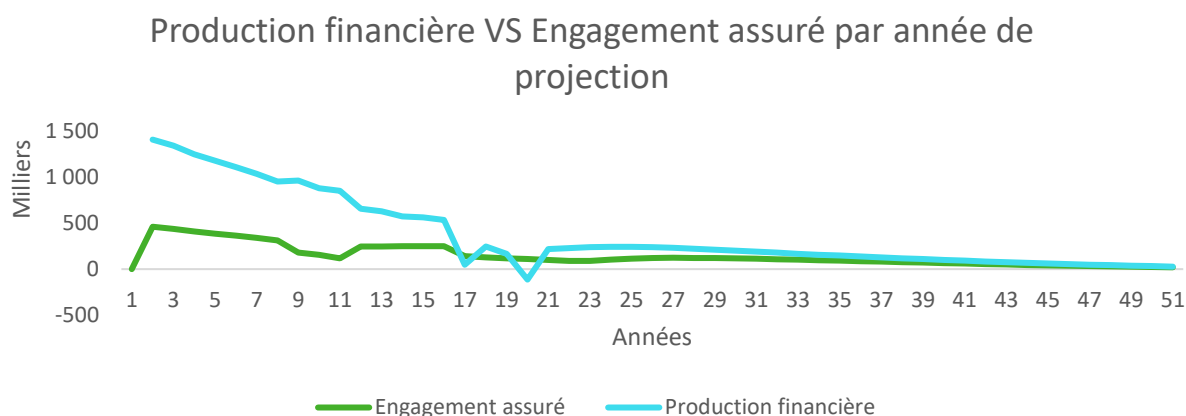


Figure 94 Production financière VS l'engagement de l'assureur Base de données chocs obligataires et GSE variant avec 3 pools de TMG

Comme attendu à la lecture du graphique des BE en fonction des VM, la production financière est insuffisante sur certaines périodes pour payer les TMG.

Cela présage un fort nombre de NAV négative d'une part et d'une certaine décorrélacion des inputs obligataires avec le BE à cause d'un grand nombre de scénarios cout fonds propres¹¹⁸.

¹¹⁸ Voir annexe partie 6, chapitre 6.2.3

Comme discuté lors de l'analyse de la production financière VS l'engagement assureur, les corrélations linéaires comme non linéaires ¹¹⁹entre le BE et les VM sont faibles bien que non nuls.

4.1.2.6 Présentation des résultats

Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données chocs obligataires et GSE variant avec trois pools de TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	131	130	0,8%
	Moy	143	143	>0,1
	Med	141	141	0,3%
	Max	234	234	16,5%
Random Forest	Min	131	132	-0,4%
	Moy	143	143	>0,1
	Med	141	141	>0,1
	Max	234	227	3,0%
Gradient Boosting	Min	131	132	-0,5%
	Moy	143	143	>0,1
	Med	141	141	-0,1%
	Max	234	233	0,5%
XGboost	Min	131	135	-2,8%
	Moy	143	143	>0,1
	Med	141	141	0,3%
	Max	234	222	5,2%

Figure 95 Tableau récapitulatif des performances générales des algorithmes de ML Base de données chocs obligataires et GSE variant avec 3 pools de TMG

La régression linéaire avec constante semble avoir le pouvoir prédictif le plus faible sur cette base de données.

Le Gradient Boosting semble pouvoir être utilisé pour les prédictions sur cette base de données avec toutes les valeurs qui semblent dans la target des 1% d'écart.

La courbe des BE du dataset test prédits par percentile est la suivante :

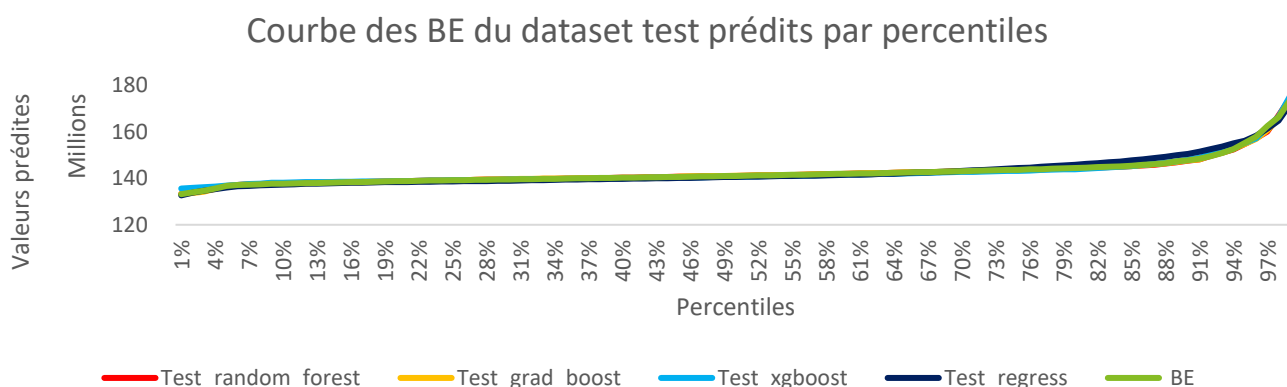


Figure 96 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires et GSE variant avec 3 pools de TMG

Les résultats semblent meilleurs sur cette database que sur la Base de données chocs obligataires et GSE variant avec un pool de TMG.

¹¹⁹ Voir annexe partie 6, chapitre 6.14.2

La courbe des écarts aux quantiles 10 et 90% et le tableau des RMSE confirment ces observations.¹²⁰

L'analyse précédemment effectuée pour la Base de données chocs obligataires et GSE variant à un pool de TMG et cette analyse montrent des meilleurs résultats pour la base de données avec trois pools de TMG qu'avec un pool de TMG, il est possible qu'un léger surapprentissage a été effectué pour les algorithmes de la Base de données chocs obligataires et GSE variant avec un pool de TMG.

Les algorithmes de prédiction de Gradient Boosting et Random Forest semblent les plus adaptés à ces bases de données.

Les temps d'apprentissage et de prédiction pour la Base de données chocs obligataires et GSE variant avec trois pools de TMG sont tous inférieurs à 10 secondes contre 20 secondes pour la Base de données chocs obligataires et GSE variant avec un pool de TMG¹²¹, confirmant l'hypothèse d'un éventuel sur-apprentissage des algorithmes.

Sur la base de données française¹²², les algorithmes de Random Forest, Gradient Boosting et XGboost continuent de prédire les BE avec des écarts bien inférieurs à 1%.

Cela signifie que les algorithmes ont été assez bien calibrés car ils obtiennent de bonnes prédictions pour une base de données avec une densité de VM obligataires fortement modifiées.

L'analyse de la Base de données chocs obligataires et GSE variant donne des résultats globaux dans la target souhaitée.

Dans la partie suivante, des prédictions sur des BDD plus complexes sont réalisées.

4.2 Présentation des résultats sur des bases de données dans des configurations plus complexes

La base de données avec GSE variant sans choc obligataire ayant déjà été traitée dans le mémoire de fin d'étude d'Aurélien Suray¹²³, les résultats propres à cette base ont été mis en annexe¹²⁴.

Ils vont néanmoins servir de base de comparaison aux résultats obtenus avec la base de données où des indicateurs obligataires ont été ajoutés.

Cette étude finit par l'analyse d'une base de données où l'allocation des actifs est changée de manière aléatoire et où les obligations sont choquées.

4.2.2 Présentation des résultats de la base de données avec inputs obligataires et GSE variant avec un et trois pools de TMG et comparaison à la base de données sans input obligataire et GSE variant.

¹²⁰ Voir annexe partie 6, chapitre 6.13.3 et 6.14.3

¹²¹ Voir annexe partie 6, chapitres 6.13.1 et 6.14.1

¹²² Voir annexe partie 6, chapitre 6.13.4 et 6.14.4

¹²³ Voir annexe partie 6, sous partie 6.9

¹²⁴ Voir annexe partie 6, sous parties 6.15 et 6.16

Dans cette partie, les résultats pour la Base de données GSE variant et indicateurs obligataires sont présentés.

La Base de données GSE variant et indicateurs obligataires se différencie de la Base de données chocs obligataires et GSE variant par la suppression des chocs obligataires, elle se différencie de la Base de données GSE variant par l'ajout de certains indicateurs obligataires calculés sous Python pour chaque trajectoire économique.

Des comparaisons avec les résultats de la Base de données GSE variant sont réalisés.

Les algorithmes de Machine Learning retenus pour cette base de données sont les mêmes que pour la BDD précédente.

4.2.2.1 Préambule

La méthode de cross validation reste inchangée par rapport à la Base de données chocs obligataires et GSE variant.

Une fois les résultats sur la base train et test obtenus, d'autres résultats basés sur des bases de données simulées avec d'autres GSE créés à l'aide de la courbe des taux initiale choquée de +/-20 bps ont été calculés (database U20/D20).

Le but de cette analyse est d'une part de vérifier si les résultats sont en adéquation avec la target fixée et si l'ajout d'indicateurs obligataires permet d'améliorer les prédictions de la Base de données GSE variant sur la base test et/ou les bases U20 et D20.

Pour les deux databases qui vont être présentés, les indicateurs suivants¹²⁵ ont été rajoutés :

- VM Obligataire calculée sous Python
- Les Coefficients de choc
- Spread pondéré
- La maturité pondérée
- La duration pondérée

La présentation des résultats est faite de manière similaire aux BDD précédentes.

4.2.2.2 Analyse de l'univers économique Base de données GSE variant et indicateurs obligataires avec un pool de TMG à 0%

La Base de données chocs obligataires avec un pool de TMG à 0% résulte d'un GSE constant pour chaque trajectoire choquée, avec des chocs obligataires activés et un pool de TMG à 0%.

Analyse du BE en fonction de la VM par quantile

¹²⁵ Voir partie 3, chapitre 3.3.2

Evolution du BE en fonction de la VM par percentiles

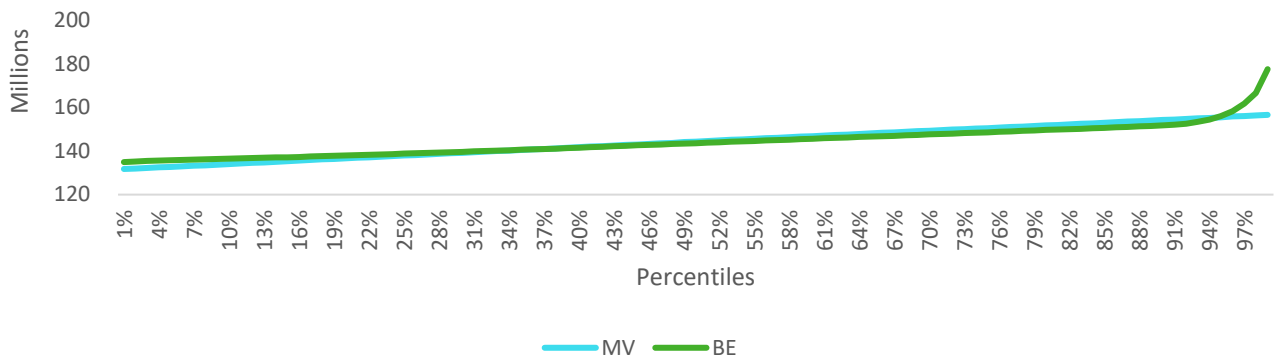


Figure 97 Evolution du BE en fonction de la VM par percentile Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG

Sans les chocs obligataires, le GSE fait varier la VM de manière quasi linéaire.

La position du BE par rapport à la VM laisse penser que le BE est très corrélé linéairement à la VM jusqu'aux quantiles 90%. De plus cette position signale une situation soit de cout fonds propres ou de pertes de revenus pour l'assureur.

La production financière par années de projection VS les engagements assurés a été tracée pour la première trajectoire économique du GSE :

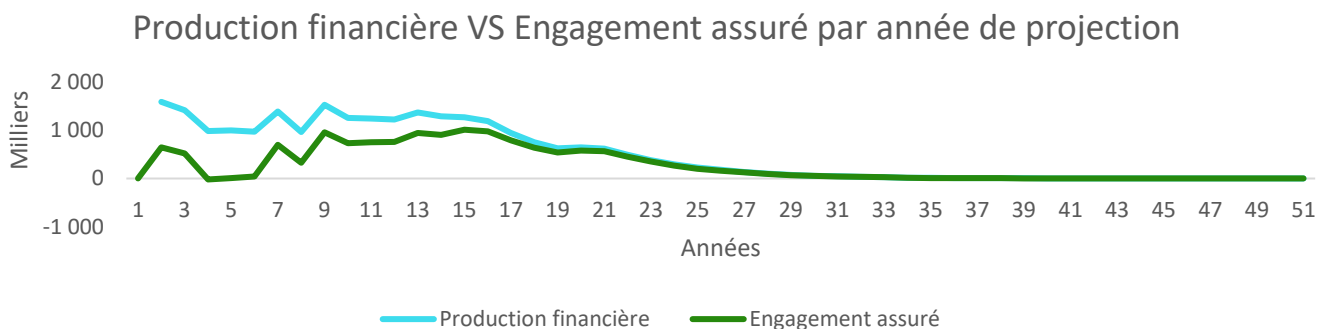


Figure 98 Production financière VS l'engagement de l'assureur Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG

Comme pour la Base de données chocs obligataires et GSE variant, une forte variation à la hausse de l'engagement assureur est visible par rapport au graphique de la Base de données chocs obligataires avec un pool de TMG.

Il est possible d'imaginer que cette différence est d'autant plus marquée pour certaines des trajectoires économiques ce qui expliquerait la position du BE par rapport à la VM sur le graphique.

Analyse des corrélations des inputs obligataires

Il existe de fortes corrélations linéaires entre le BE et les autres variables obligataires avec les variables temps y comprises.¹²⁶

4.2.2.3 Présentation des résultats pour la BDD à un pool de TMG

¹²⁶ Voir annexe partie 6, chapitre 6.17.1

Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données chocs obligataires sans TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	134	135	-0,4%
	Moy	145	145	>0,1
	Med	144	143	0,6%
	Max	245	248	-1,1%
Random Forest	Min	134	133	0,6%
	Moy	145	145	-0,1%
	Med	144	144	-0,3%
	Max	245	258	-5,2%
Gradient Boosting	Min	134	123	8,2%
	Moy	145	145	-0,1%
	Med	144	144	-0,2%
	Max	245	225	8,3%
XGboost	Min	134	128	4,8%
	Moy	145	145	>0,1
	Med	144	143	0,1%
	Max	245	239	2,5%

Figure 99 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG

La régression linéaire semble avoir le meilleur pouvoir prédictif sur le BE dans cette configuration économique, ce qui peut sembler logique à la vue du graphique de l'évolution du BE en fonction de la VM. Cependant ces résultats sont vérifiés sur un tableau des quantiles qui va aussi servir de comparaison à la Base de données avec GSE variant sans indicateur obligataire.

Comparaison du tableau des écarts aux quantiles de la Base de données GSE variant et Base de données GSE variant et indicateurs obligataires :

Modèle Ecart relatif ¹²⁷	Random Forest avec indicateurs obligataires	Gradient Boosting avec indicateurs obligataires	XGboost avec indicateurs obligataires	Régression Multiple avec indicateurs obligataires	Random Forest sans indicateurs obligataires	Gradient Boosting sans indicateurs obligataires	XGboost sans indicateurs obligataires	Régression Multiple sans indicateurs obligataires
9%	0,16%	-0,68%	1,27%	3,46%	0,36%	-0,69%	2,10%	2,89%
19%	0,22%	0,40%	1,63%	2,75%	0,57%	0,43%	2,29%	2,28%
29%	0,35%	0,73%	1,40%	1,86%	0,41%	0,77%	1,72%	1,40%
39%	0,40%	0,61%	0,82%	0,78%	0,09%	0,32%	0,79%	0,26%
49%	0,32%	0,29%	-0,05%	-0,51%	-0,11%	0,20%	-0,12%	-0,60%
59%	0,14%	-0,04%	-0,73%	-1,60%	-0,08%	-0,19%	-1,11%	-1,46%
69%	-0,02%	-0,17%	-1,27%	-2,51%	-0,03%	-0,23%	-1,74%	-1,84%
79%	> 0,01%	0,04%	-1,30%	-2,77%	-0,24%	0,06%	-2,12%	-1,84%
89%	-0,34%	0,72%	-1,06%	-1,77%	-0,18%	0,90%	-1,19%	-0,50%
99%	-0,86%	0,32%	-0,93%	-0,06%	-1,53%	0,14%	1,18%	-3,25%

Figure 100 Comparaison des écarts aux quantiles entre la Base de données GSE variant et la Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG

¹²⁷ Pour les tableaux ayant des cases en gras, celles-ci représentent les écarts au BE des prédictions inférieures à environ 1%.

Ce tableau permet de montrer que l'ajout des indicateurs obligatoires permet d'obtenir deux algorithmes dans le benchmark prédictif fixé contre un sans les indicateurs, excepté la régression linéaire qui perd en précision, les autres algorithmes ont gagné en pouvoir prédictif après cet ajout.

Les forts pouvoirs prédictifs du Random Forest et du Gradient Boosting sont vérifiés à l'aide du graphique représentant la courbe des BE du dataset test prédits par percentile :

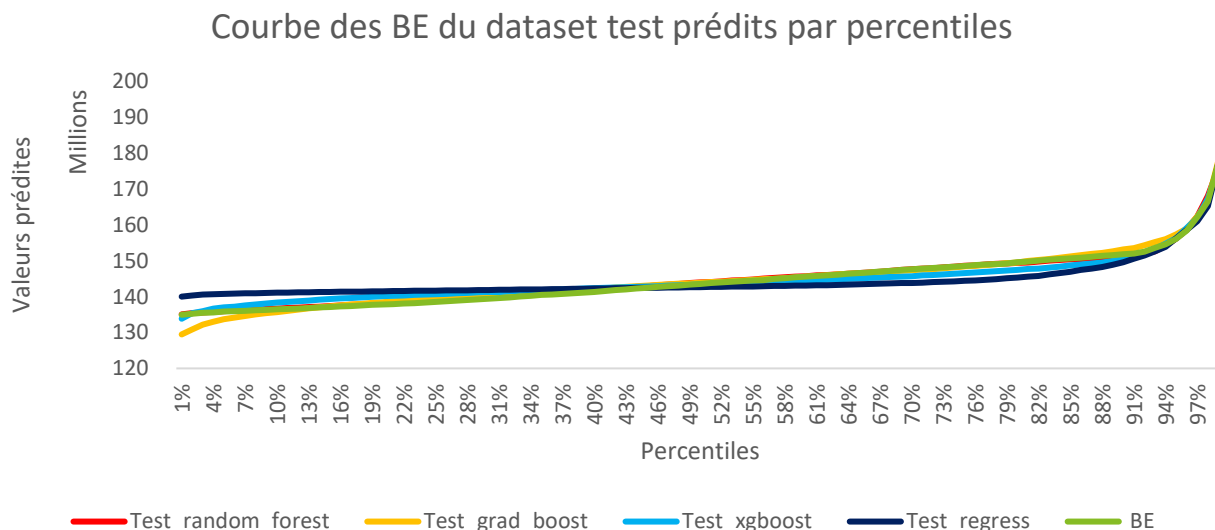


Figure 101 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant et indicateurs obligatoires avec 1 pool de TMG

Cette courbe permet d'affirmer que malgré des mauvaises prévisions sur le dernier percentile, le Random Forest est capable de prédire avec une précision inférieure à 1% le BE comme souhaité.

La courbe des écarts aux quantiles 10 et 90% confirme ces observations.¹²⁸

Le tableau des RMSE n'apporte pas d'indications supplémentaires à l'analyse¹²⁹. Malgré de très mauvaises prédictions pour le dernier percentile, le Random Forest est l'algorithme préféré pour la globalité de ses prédictions dans la target d'écart fixée.

Les tableaux suivants montrent une comparaison des résultats avant et après l'ajout des indicateurs obligatoires pour les datasets U20 et D20.

¹²⁸ Voir annexe partie 6, chapitre 6.17.3

¹²⁹ Voir annexe partie 6, chapitre 6.17.3

Database U20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple avec constante	Régression Multiple
Amélioration par quantile					
9%	0,29%	0,63%	-1,74%	-14,60%	-1,44%
19%	0,54%	0,05%	-2,05%	-13,51%	-0,99%
29%	0,54%	-0,17%	-1,55%	-12,30%	-0,01%
39%	0,23%	-0,11%	-0,47%	-11,13%	1,00%
49%	-0,19%	0,16%	0,74%	-9,92%	2,01%
59%	-0,38%	0,48%	1,92%	-8,87%	2,44%
69%	-0,67%	0,47%	2,69%	-8,50%	2,19%
79%	-0,60%	0,27%	3,08%	-8,46%	1,68%
89%	0,37%	-1,03%	1,96%	-9,73%	0,46%
99%	0,99%	1,58%	0,71%	-3,02%	0,90%

Figure 102 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 1 pool de TMG sur le dataset U20

Database D20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple avec constante	Régression Multiple
Amélioration par quantile					
9%	0,14%	-1,56%	0,07%	-18,54%	-5,71%
19%	0,16%	-1,00%	-0,26%	-16,65%	-5,05%
29%	0,16%	-0,51%	-0,27%	-15,45%	-4,00%
39%	0,23%	-0,20%	0,09%	-14,40%	-2,84%
49%	0,07%	0,03%	0,39%	-13,46%	-1,76%
59%	> 0,01%	0,28%	0,73%	-12,67%	-0,92%
69%	-0,08%	0,50%	0,94%	-12,08%	-0,68%
79%	-0,22%	0,74%	0,99%	-11,06%	-0,85%
89%	0,02%	0,77%	0,64%	-9,60%	-2,48%
99%	0,09%	1,01%	-0,55%	-1,52%	3,66%

Figure 103 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 1 pool de TMG sur le dataset D20

Malgré une nette amélioration des résultats (hors régressions linéaires) sur les database D20 et U20, les résultats prédits sur ces bases de données restent médiocres¹³⁰. Un apprentissage supplémentaire pour ce type de BDD est nécessaire.

4.2.2.4 Analyse de l'univers économique Base de données GSE variant et indicateurs obligataires avec trois pools de TMG

Analyse du BE en fonction de la VM

¹³⁰ Voir annexe partie 6, chapitres 6.17.4 et 6.17.5

Evolution du BE en fonction de la VM par percentiles

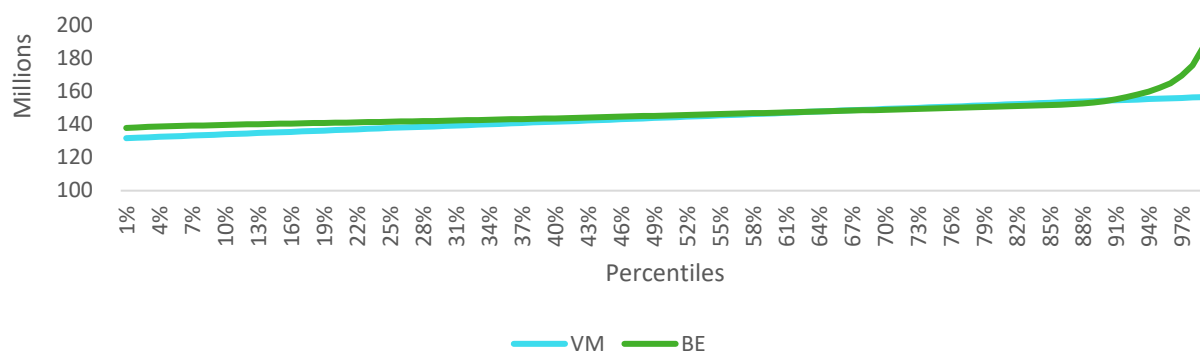


Figure 104 Evolution du BE en fonction de la VM par percentile Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG

La VM suit la même trajectoire que le graphique de la Base de données GSE variant avec un pool de TMG. Le BE suit une trajectoire très proche du BE avec un pool de TMG à 0% (voir graph un peu plus haut).

La seule différence avec l'autre courbe des BE, est la position de la courbe du BE plus haute par rapports aux VM, cette différence peut s'expliquer par l'ajout de TMG non nul.

La production financière par années de projection VS les engagements assurés a été tracée pour la première trajectoire économique de l'GSE

Production financière VS Engagement assuré par année de projection

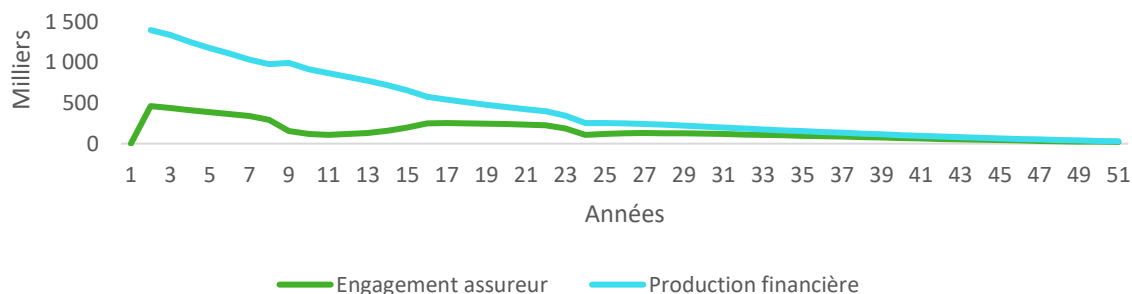


Figure 105 Production financière VS l'engagement de l'assureur Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG

La production financière reste suffisante pour répondre au TMG bien que l'engagement assureur a été sensiblement augmenté par rapport à l'univers avec un pool de TMG.

Du fait de la conservation de la relation linéaire entre le BE et la VM, des fortes corrélations linéaires entre les indicateurs obligataires et le BE sont attendues.

A cause de l'ajout de complexité dans la répartition des TMG, il est possible que les corrélations non linéaires entre les indicateurs de VM et le BE augmentent sensiblement.

La gestion différente des actifs par rapport à l'ajout des TMG permettrait d'expliquer la différence entre la production financière de la database avec un et trois pools de TMG.

Analyse des corrélations des indicateurs obligataires

Comme dit ci-dessus de fortes corrélations linéaires sont à enregistrer ainsi qu'une hausse sensible des corrélations non linéaires entre les inputs obligataires et le BE.¹³¹

4.2.2.5 Présentation des résultats

Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données GSE variant et indicateurs obligataires avec trois pools de TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	136	136	-0,3%
	Moy	148	148	0,1%
	Med	146	144	0,9%
	Max	254	254	9,5%
Random Forest	Min	136	137	-1,0%
	Moy	148	148	> 0,1%
	Med	146	146	> 0,1%
	Max	254	268	-5,6%
Gradient Boosting	Min	136	136	-0,3%
	Moy	148	148	> 0,1%
	Med	146	145	0,4%
	Max	254	240	5,6%
XGboost	Min	136	145	-6,7%
	Moy	148	148	-0,2%
	Med	146	145	0,6%
	Max	254	239	6,0%

Figure 106 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG

Ces résultats semblent sensiblement moins bons aux quantiles de valeurs les plus fortes que ceux de la base de données sans les indicateurs.

Cependant une nette amélioration est à constater sur les prédictions minimales.

¹³¹ Voir annexe partie 6, chapitre 6.18.2

Comparaison du tableau des écarts aux quantiles de la Base de données GSE variant avec et sans indicateurs obligatoires :

Modèle	Random Forest avec indicateurs obligatoires	Gradient Boosting avec indicateurs obligatoires	XGboost avec indicateurs obligatoires	Régression Multiple avec indicateurs obligatoires	Random Forest sans indicateurs obligatoires	Gradient Boosting sans indicateurs obligatoires	XGboost sans indicateurs obligatoires	Régression Multiple sans indicateurs obligatoires
Ecart relatif¹³²								
9%	0,20%	1,53%	3,62%	1,91%	-0,01%	1,41%	2,04%	1,04%
19%	0,27%	1,29%	2,55%	1,31%	-0,11%	1,15%	1,55%	0,58%
29%	0,33%	1,02%	1,66%	0,73%	-0,02%	0,92%	1,06%	0,26%
39%	0,36%	0,55%	0,69%	0,11%	-0,05%	0,51%	0,38%	-0,14%
49%	-0,01%	-0,30%	-0,51%	-0,87%	0,02%	-0,07%	-0,41%	-0,59%
59%	-0,11%	-0,94%	-1,44%	-1,42%	-0,04%	-0,79%	-1,24%	-1,07%
69%	-0,27%	-1,46%	-2,11%	-1,74%	-0,18%	-1,31%	-1,84%	-0,94%
79%	-0,33%	-1,41%	-2,13%	-1,37%	-0,36%	-1,79%	-2,15%	-0,45%
89%	0,14%	-0,03%	-0,65%	0,77%	-0,20%	-0,19%	-0,24%	1,52%
99%	-2,96%	-2,42%	-4,44%	-2,06%	-3,34%	-1,74%	-1,85%	-5,00%

Figure 107 Comparaison des écarts aux quantiles entre la Base de données GSE variant et la Base de données GSE variant et indicateurs obligatoires avec 3 pools de TMG

Dans le tableau ci-dessus, seule la fonction Random Forest donne des résultats dans la marge d'erreur bien qu'elle sous-estime les résultats pour les quantiles au-delà de 90%.

Le temps d'entraînement des algorithmes inférieur à 40 secondes pour la Base de données GSE variant et indicateurs obligatoires VS inférieur à 1 minute ¹³³ pour la Base de données GSE variant, dénote la possibilité que la Base de données GSE variant et indicateurs obligatoires a été légèrement sous-apprise.

La courbe des BE du dataset test prédits par percentile est la suivante :

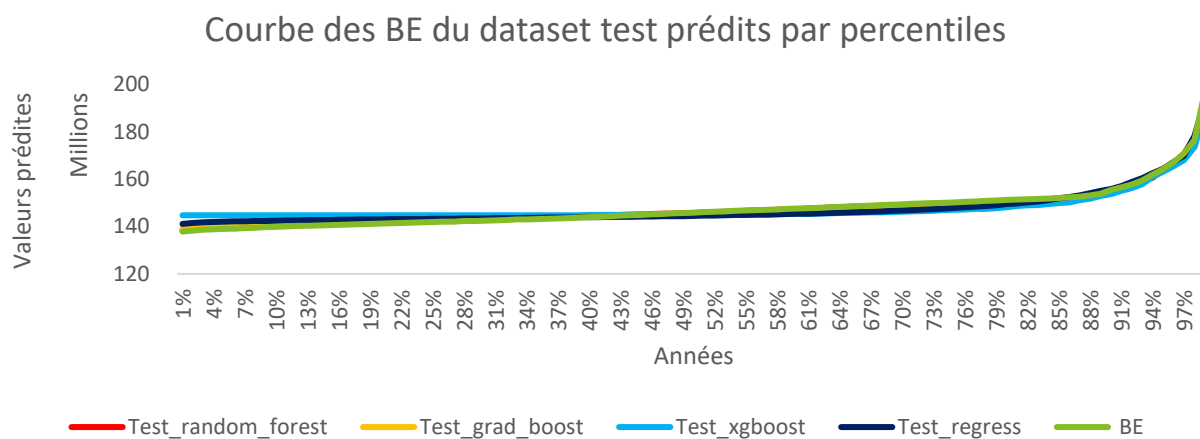


Figure 108 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant et indicateurs obligatoires avec 3 pools de TMG

Le Random Forest devrait bien prédire le BE aux quantiles les plus faible, entre 0 et 10% :

¹³² Pour les tableaux ayant des cases en gras, celles-ci représentent les écarts au BE des prédictions inférieures à environ 1%.

¹³³ Voir annexe partie 6, chapitres 6.17.1 et 6.18.1

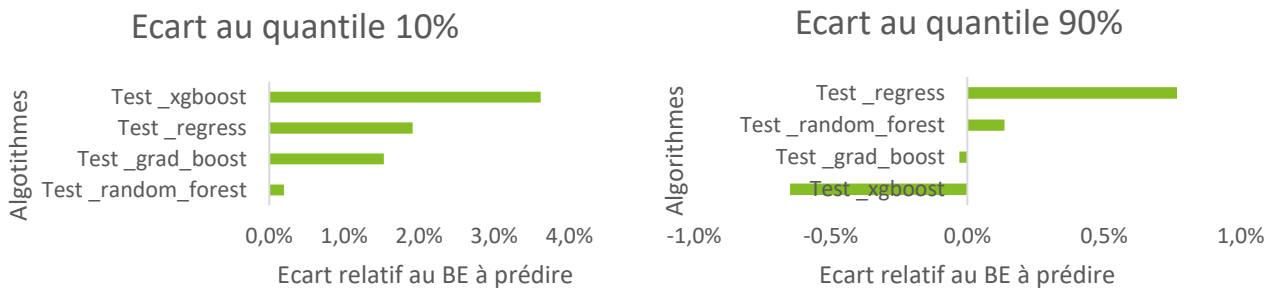


Figure 109 Ecart aux quantiles 10% et 90% par algorithmes Base de données GSE variant et indicateurs obligatoires avec 3 pools de TMG

La complexification de l'univers économique par l'ajout du TMG a nettement dégradé les résultats. Cependant l'algorithme de Random Forest garde des prédictions dans la target pour au moins 90% des données et ce malgré un léger sous apprentissage supposé.

Le tableau des RMSE ne permet pas de meilleure comparaison des algorithmes.¹³⁴

Par son fort pouvoir prédictif sur une très grande partie du jeu de données, le Random Forest est l'algorithme préféré pour cette situation.

L'ajout des indicateurs obligatoires permet de sensiblement améliorer les prédictions des algorithmes (hors régression linéaire) sur le dataset D20 comme le montre les tableaux comparatifs suivants.

Pour la database U20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple avec constante	Régression Multiple
Amélioration par quantile					
9%	-0,12%	-2,24%	-1,30%	-11,17%	1,36%
19%	-0,11%	-1,21%	-0,93%	-11,51%	0,80%
29%	-0,13%	-0,51%	-0,69%	-11,75%	0,35%
39%	-0,08%	0,06%	-0,46%	-11,89%	-0,02%
49%	-0,04%	0,61%	-0,28%	-11,90%	-0,27%
59%	-0,06%	1,08%	-0,18%	-11,60%	-0,47%
69%	0,04%	0,63%	-0,06%	-11,31%	-0,27%
79%	0,02%	-0,04%	0,16%	-10,84%	-0,04%
89%	-0,07%	0,74%	-0,26%	-9,51%	1,07%
99%	-0,29%	0,22%	0,20%	-6,38%	-2,10%

Figure 110 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 3 pools de TMG sur le dataset U20

Pour la database D20

¹³⁴ Voir annexes partie 6, chapitres 6.17.3

Modèle Amélioration par quantile	Random Forest	Gradient Boosting	XGboost	Régression Multiple avec constante	Régression Multiple
9%	4,20%	4,74%	5,71%	-12,72%	-0,11%
19%	2,23%	3,61%	3,91%	-13,70%	-2,59%
29%	1,30%	2,09%	2,74%	-14,48%	-4,28%
39%	0,47%	0,42%	0,47%	-14,28%	-5,46%
49%	-0,02%	-1,34%	-1,08%	-12,57%	-4,40%
59%	0,05%	-0,89%	-0,14%	-11,17%	-3,11%
69%	1,04%	0,15%	1,18%	-9,70%	-2,38%
79%	2,93%	2,34%	2,61%	-7,02%	-1,26%
89%	5,91%	5,49%	5,58%	-2,56%	0,65%
99%	12,74%	13,59%	11,13%	12,33%	16,72%

Figure 111 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 3 pools de TMG sur le dataset D20

L'ajout d'indicateurs obligataires améliore très nettement les résultats (hors régression linéaire) de la database D20 et n'a aucun impact sur la database U20.

Cela suppose qu'une hausse de 20 bps, fait passer l'assureur d'une situation de perte de revenus à une situation de cout fonds propres non apprise par les algorithmes.

Dans l'analyse suivante, une réallocation des actifs et des chocs obligataires sont effectués en même temps.

4.2.3 Présentation des résultats de la base de données avec chocs obligataires, GSE variant et réallocation d'actif.

Dans cette partie, les résultats pour la Base de données GSE variant, chocs obligataires et réallocation dynamique sont présentés.

La Base de données GSE variant, chocs obligataires et réallocation dynamique se différencie de la Base de données GSE variant et indicateurs obligataires par l'ajout de chocs obligataires ainsi qu'une variation dans l'allocation de l'actif.

4.2.3.1 Préambule

La méthode de cross validation reste inchangée par rapport à la Base de données GSE variant et indicateurs obligataires.

Une fois les résultats sur la base train et test obtenus, d'autres résultats basés sur des bases de données simulées avec d'autres GSE créés à l'aide de la courbe de taux initiale choquée de +/-20 bps ont été calculés (database U20/D20).

Des prédictions sur le dataset français sont aussi effectuées.

Le but de cette analyse est de vérifier si les résultats sont en adéquation avec la target fixée et si les algorithmes sont robustes à une variation du contexte économique.

Pour la Base de données GSE variant, chocs obligataires et réallocation dynamique les indicateurs suivants sont rajoutés :

- Indicateur de la réallocation des actions
- Indicateur de la réallocation Immobilière

4.2.3.2 Analyse de l'univers économique Base de données GSE variant, chocs obligataires et réallocation dynamique avec un pool de TMG à 0%.

La Base de données GSE variant, chocs obligataires et réallocation dynamique avec un pool de TMG à 0% résulte d'un GSE variant pour chaque trajectoire, avec des chocs obligataires et une réallocation des actifs.

Analyse du BE en fonction de la VM

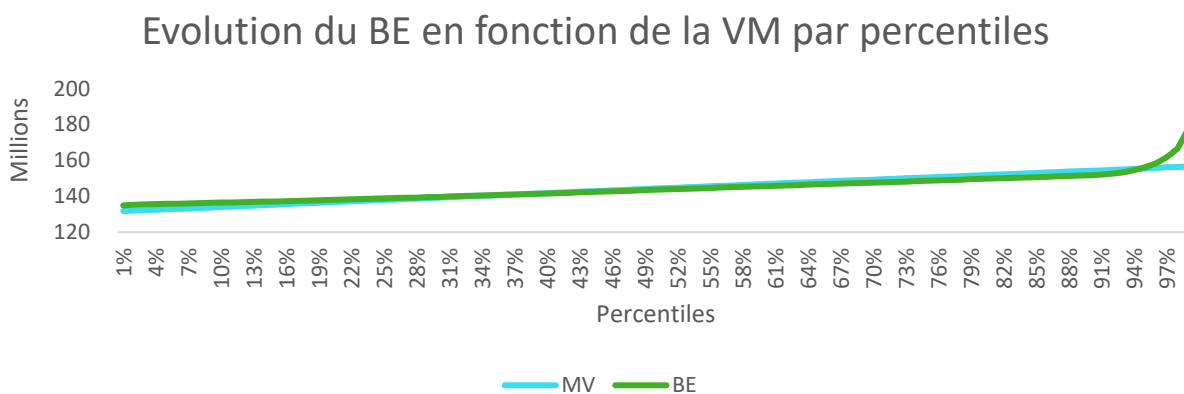


Figure 112 Evolution du BE en fonction de la VM par percentile Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

Le BE paraît corrélé linéairement à la VM jusqu'aux quantiles 90%. La position du BE par rapport à la VM laisse présager une situation de pertes de revenus ou de cout fonds propres pour l'assureur.

La production financière par années de projection VS les engagements assurés a été tracée pour la première trajectoire économique de l'GSE

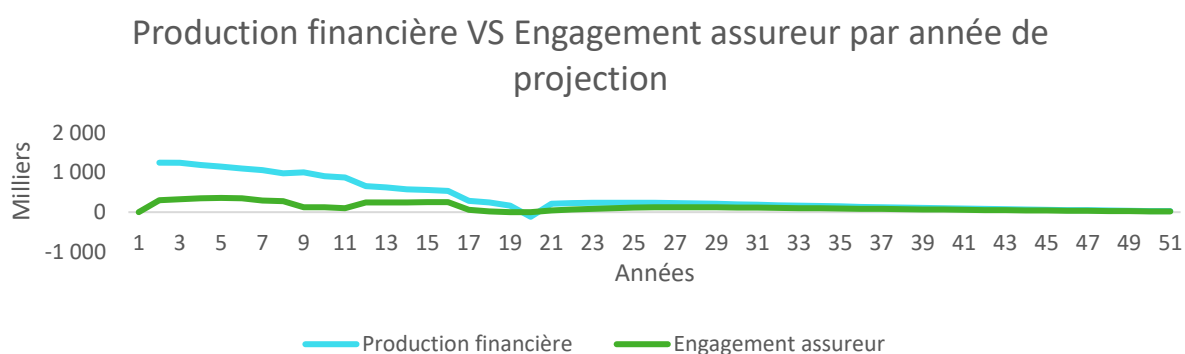


Figure 113 Production financière VS l'engagement de l'assureur Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

La production financière passant sous l'engagement de l'assureur pour certaines années laisse à présager un BE fortement décorrélé des inputs obligataires.

Analyse des corrélations des inputs obligataires

Le BE est fortement décorrélé inputs obligataires dû au manque de production financière dans les simulations. ¹³⁵

4.2.3.3 Présentation des résultats

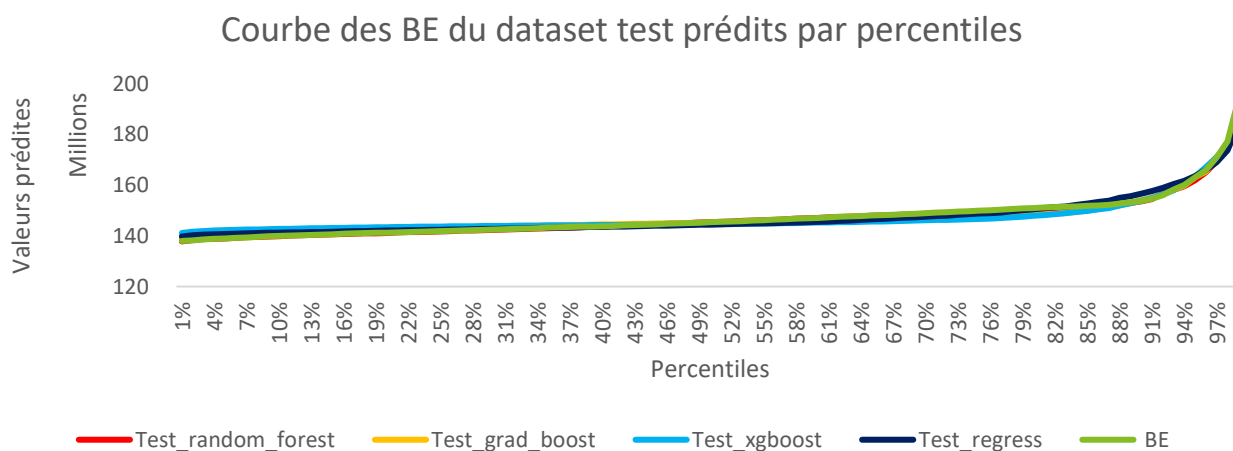
Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec un pool de TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	132	130	1,6%
	Moy	143	143	-0,1%
	Med	142	143	-0,2%
	Max	209	190	9,1%
Random Forest	Min	132	132	0,1%
	Moy	143	143	> 0,1%
	Med	142	142	0,1%
	Max	209	214	-2,7%
Gradient Boosting	Min	132	133	-0,7%
	Moy	143	143	> 0,1%
	Med	142	142	> 0,1%
	Max	209	208	0,1%
XGboost	Min	132	133	-0,9%
	Moy	143	143	> 0,1%
	Med	142	142	-0,1%
	Max	209	209	-0,2%

Figure 114 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

Le Gradient Boosting et le XGboost semblent les candidats parfaits pour la prédiction dans cet environnement économique.

Le pouvoir prédictif du Random Forest et du XGboost vont être vérifiés à l'aide du graphique représentant la courbe des BE du dataset test prédits par percentile :



¹³⁵ Voir annexe partie 6, chapitre 6.19.2

Figure 115 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

Cette courbe permet d'appuyer le bon pouvoir prédictif sur ces bases de données

Les courbe des écarts aux quantiles 10 et 90%¹³⁶ confirment le bon pouvoir prédictif du Random Forest.

Le graphique des RMSE afin de comparer les algorithmes

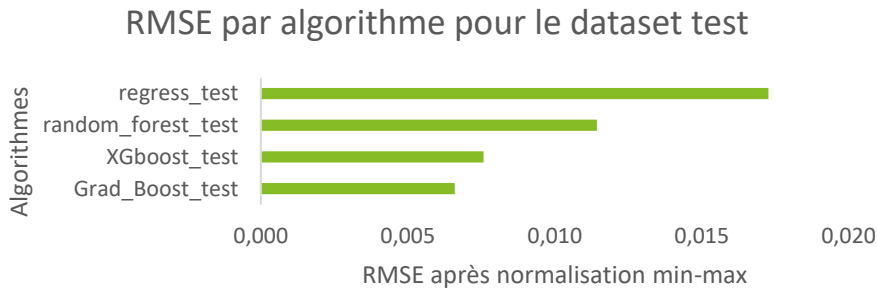


Figure 116 RMSE par algorithme Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

Ce tableau indique en effet que le XGboost et le Gradient Boosting ont un pouvoir prédictif équivalent à la médiane, mais que le Random Forest aura tendance à avoir un plus gros écart à la médiane.

La robustesse des algorithmes est testée sur les datasets D20, U20 et français dont les écarts aux quantiles sont affichés ci-dessous :

Database U20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	1,36%	1,35%	1,35%	-0,13%
19%	1,45%	1,47%	1,38%	0,66%
29%	1,77%	1,71%	1,71%	1,27%
39%	2,01%	2,00%	1,98%	1,55%
49%	2,27%	2,29%	2,35%	1,68%
59%	2,46%	2,44%	2,49%	1,69%
69%	2,54%	2,60%	2,57%	1,63%
79%	2,32%	2,33%	2,26%	1,47%
89%	1,75%	1,70%	1,65%	2,07%
99%	2,41%	2,23%	2,18%	0,54%

Figure 117 Ecarts aux quantiles sur le dataset U20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

¹³⁶ Voir partie 6, chapitre 6.19.3

Database D20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	-1,14%	-1,11%	-1,14%	-2,51%
19%	-0,94%	-0,86%	-0,94%	-1,61%
29%	-0,62%	-0,65%	-0,64%	-0,96%
39%	-0,27%	-0,21%	-0,22%	-0,62%
49%	0,02%	0,01%	0,14%	-0,48%
59%	0,13%	0,19%	0,17%	-0,42%
69%	0,21%	0,28%	0,19%	-0,38%
79%	0,02%	0,07%	0,03%	-0,01%
89%	-0,37%	-0,54%	-0,52%	1,06%
99%	1,48%	0,67%	0,40%	-2,45%

Figure 118 Ecarts aux quantiles sur le dataset D20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

Database française

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	-0,22%	-0,15%	-0,17%	-0,58%
19%	-0,16%	-0,05%	-0,26%	-0,13%
29%	-0,04%	-0,01%	-0,02%	0,25%
39%	0,20%	0,12%	0,06%	0,33%
49%	0,08%	-0,06%	0,14%	0,09%
59%	-0,05%	0,08%	0,14%	-0,02%
69%	-0,03%	0,07%	0,02%	-0,06%
79%	0,01%	-0,03%	-0,10%	-0,07%
89%	-0,37%	-0,20%	-0,33%	1,52%
99%	3,00%	1,73%	1,17%	-3,37%

Figure 119 Ecarts aux quantiles sur le dataset français, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG

Les prédictions effectuées sur les datasets français et D20 sont satisfaisantes avec un écart maximum au BE de l'ordre de 1% pour certains algorithmes.

L'écart relatif plus conséquent pour le dataset U20 peut s'expliquer par le passage d'une situation à pertes de revenus à une situation coûts fonds propres.

4.2.3.4 Analyse de l'univers économique Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG

Analyse du BE en fonction de la VM

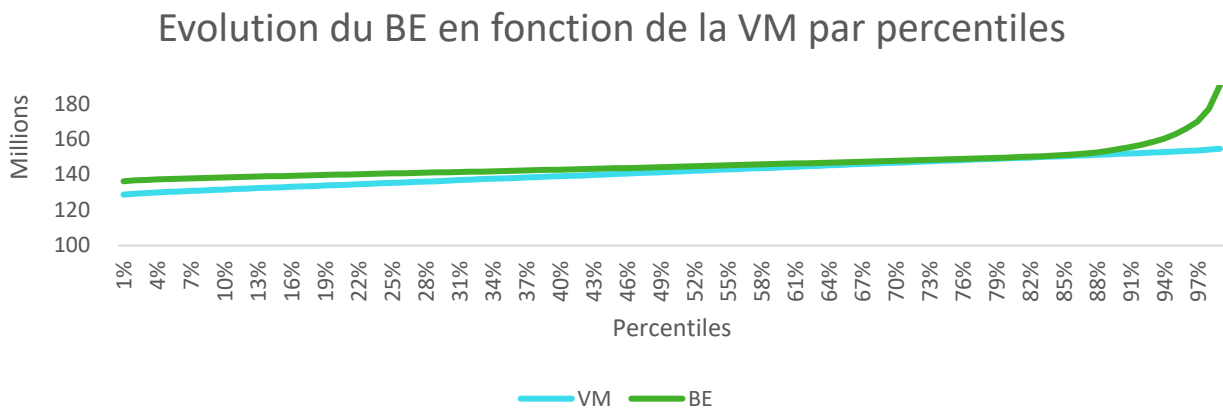


Figure 120 Evolution du BE en fonction de la VM par percentile Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

La VM suit la même trajectoire que le graphique de la Base de données GSE variant avec un pool de TMG. Le BE suit une trajectoire très proche du BE avec un pool de TMG à 0% (voir graph un peu plus haut).

La seule différence avec l'autre courbe des BE est la position de la courbe du BE plus haute par rapport aux VM, cette différence peut s'expliquer par l'ajout de TMG non nul.

A noter que le BE est désormais constamment en dessous de la courbe des VM, l'assureur est constamment en situation de coût fonds propres.

La production financière par années de projection VS les engagements assurés a été tracée pour la première trajectoire économique de l'GSE

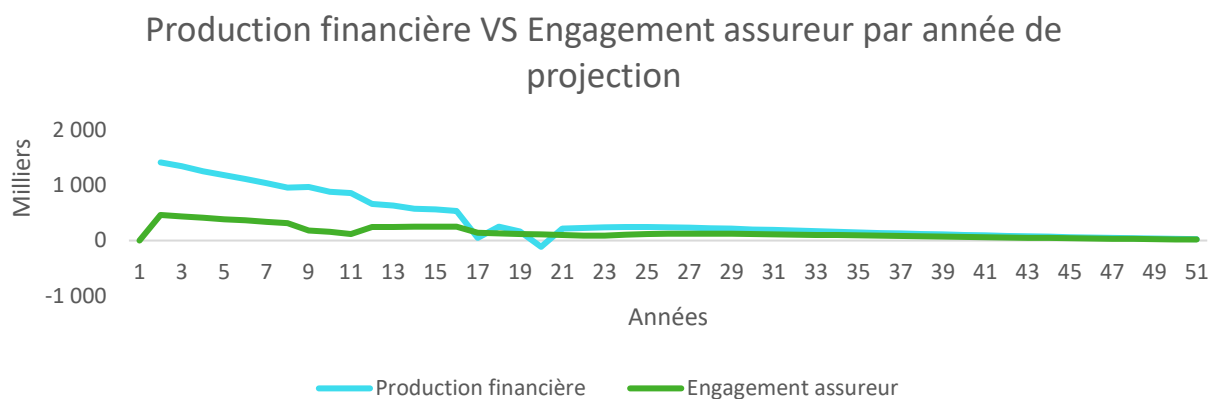


Figure 121 Production financière VS l'engagement de l'assureur Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

Comme attendu vis à vis du graphique précédent, l'assureur n'est pas capable de répondre à son engagement.

Les corrélations entre les inputs obligataires et le BE sont probablement très faibles voire nulles.

Analyse des corrélations des indicateurs obligataires

Comme dit ci-dessus les corrélations linéaires entre le BE et les inputs obligataires sont nulles.¹³⁷

La décorrélation entre les VM obligataires calculées sous Python et la VM Prophet peut s'expliquer par une réallocation d'actif fortement différente de l'allocation initiale.

4.2.3.5 Présentation des résultats

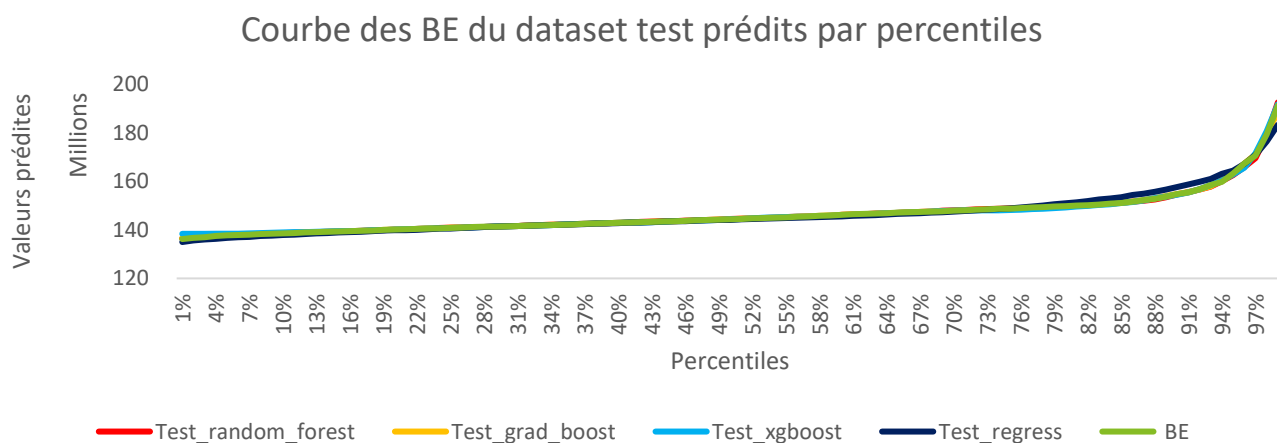
Après cross validation, le tableau des prédictions pour le dataset test pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG est le suivant :

Modèle		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	133	131	1,6%
	Moy	146	146	0,1%
	Med	144	144	0,2%
	Max	261	261	15,1%
Random Forest	Min	133	135	-1,1%
	Moy	146	146	> 0,1%
	Med	144	144	> 0,1%
	Max	261	269	-2,9%
Gradient Boosting	Min	133	135	-1,8%
	Moy	146	146	> 0,1%
	Med	144	144	> 0,1%
	Max	261	245	6,2%
XGboost	Min	133	138	-4,0%
	Moy	146	146	> 0,1%
	Med	144	144	> 0,1%
	Max	261	218	16,4%

Figure 122 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

Ces résultats semblent sensiblement moins bons aux quantiles de valeurs les plus fortes que ceux de la base de données avec un pool de TMG, la complexification de cette base a sensiblement fait varier les résultats.

La courbe des BE du dataset test prédits par percentile est la suivante :



¹³⁷ Voir annexe partie 6, chapitre 6.20.2

Figure 123 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

Sur cette courbe, seule la régression linéaire semble avoir un pouvoir prédictif moindre.

Des moins bonnes prédictions aux quantiles extrêmes (en dessous de 1% et au-delà de 99%) sont relevées.

Ci-dessous les graphiques des écarts aux quantiles :

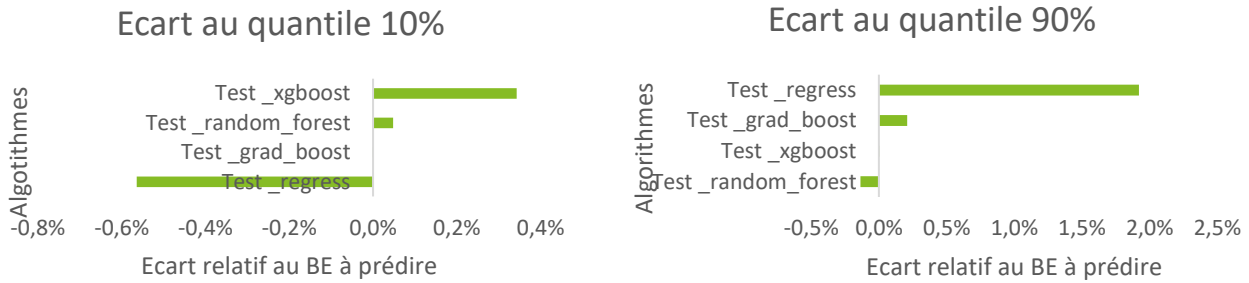


Figure 124 Ecart aux quantiles 10% et 90% par algorithmes Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

La complexification de l'univers économique par la complexification du TMG a sensiblement dégradé les résultats de l'étude aux quantiles extrêmes, l'écart de prédiction reste néanmoins inférieur à 1% pour les quantiles entre 10 et 90%.

La RMSE¹³⁸ des XGboost, Random Forest et du Gradient Boosting ne permet pas de juger si un algorithme est meilleur que l'autre.

Cependant la régression linéaire possède un pouvoir prédictif bien moindre que les autres algorithmes.

Le tableau des écarts au BE des différents quantiles par algorithmes¹³⁹ permet de mettre en évidence l'efficacité des algorithmes dont les écarts au BE sont compris dans la target aux environs des 1% fixés hors régression linéaire.

La complexification des TMG a sensiblement fait augmenter le temps d'apprentissage des algorithmes¹⁴⁰ qui est inférieur à 25 secondes pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec un pool de TMG VS inférieur à 45 secondes pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG.

¹³⁸ Voir annexe partie 6, chapitre 6.20.3

¹³⁹ Voir annexe partie 6, chapitres 6.20.3

¹⁴⁰ Voir annexe partie 6, chapitres 6.19.1 et 6.20.1

Présentation des tableaux des écarts relatifs par quantile de BE pour les datasets U20, D20 et Français :

Pour le dataset U20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	8,16%	8,13%	8,15%	6,03%
19%	8,39%	8,37%	8,36%	6,76%
29%	8,44%	8,45%	8,42%	7,03%
39%	8,41%	8,46%	8,40%	7,14%
49%	8,39%	8,34%	8,35%	7,08%
59%	8,39%	8,33%	8,28%	6,88%
69%	7,96%	7,94%	7,87%	6,58%
79%	7,44%	7,45%	7,38%	6,83%
89%	7,08%	7,03%	7,01%	8,34%
99%	21,26%	20,80%	19,79%	18,27%

Figure 125 Ecarts aux quantiles sur le dataset U20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

Pour le dataset D20

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	-1,05%	-1,05%	-1,01%	-2,82%
19%	-0,85%	-0,89%	-0,91%	-2,25%
29%	-0,68%	-0,61%	-0,68%	-1,83%
39%	-0,52%	-0,46%	-0,46%	-1,57%
49%	-0,25%	-0,26%	-0,27%	-1,30%
59%	-0,12%	-0,13%	-0,17%	-1,03%
69%	-0,18%	-0,16%	-0,18%	-0,47%
79%	-0,46%	-0,53%	-0,52%	0,47%
89%	-0,27%	-0,21%	-0,14%	1,71%
99%	-1,17%	-0,86%	-0,20%	-3,49%

Figure 126 Ecarts aux quantiles sur le dataset D20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

Pour le dataset Français

Modèle	Random Forest	Gradient Boosting	XGboost	Régression Multiple
9%	-0,50%	-0,46%	-0,44%	-0,78%
19%	-0,20%	-0,26%	-0,28%	-0,37%
29%	-0,16%	-0,17%	-0,17%	-0,18%
39%	0,07%	0,14%	0,12%	-0,12%
49%	0,35%	0,32%	0,27%	-0,17%
59%	0,19%	0,11%	0,09%	-0,11%
69%	-0,03%	-0,03%	-0,09%	-0,14%
79%	-0,12%	-0,25%	-0,25%	0,64%
89%	0,40%	0,31%	0,12%	1,97%
99%	4,87%	0,46%	1,38%	-4,53%

Figure 127 Ecarts aux quantiles sur le dataset français, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG

Les prédictions effectuées sur les datasets français et D20 sont satisfaisantes avec un écart maximum au BE de l'ordre de 1% pour certains algorithmes.

L'écart relatif plus conséquent pour la database U20 peut s'expliquer par la baisse plus rapide de la VM que du BE par effet d'actualisation et de l'incapacité de l'assureur à « payer » son TMG.

Cela entraine une stagnation de l'engagement assureur à son engagement minimal pour chaque trajectoire et implique un changement de modèle de prédictif.

4.3 Application sur le SCR de spread obligatoire et de taux

La duration du portefeuille initiale étant d'environ 9 ans et son échelon de rating à 1, le choc de spread selon la formule standard sera de 7,9% soit un choc obligatoire de 0,921.

Le SCR de choc de spread a été calculé de la manière suivante :

$$SCR_{oblig} = NAV_{ini} - NAV_{0,939}$$

Avec :

$$NAV = VM_{actifs} - BE$$

Le $SCR_{oblig_prédit}$ a été calculé de la manière suivante :

$$SCR_{prédit} = NAV_{ini} - NAV_{prédite0,939}$$

Le SCR de taux a été calculé à l'aide des courbes choquées à +/-20bps.

Le SCR_{taux} a été calculé comme suit :

$$SCR_{taux} = NAV_{ini} - NAV_{+/-20bps}$$

Avec :

$$NAV_{+/-20bps} = VM_{actifs_{moyenne+/-20bps}} - BE_{moyen+/-20bps}$$

La base de données n'ayant pas été adaptée pour le calcul du SCR de taux¹⁴¹, les prédictions des SCR de taux et leur agrégation avec le SCR obligatoire ne devraient pas être optimales.

4.3.1 Application du SCR de spread pour la base de données avec GSE continu et chocs obligatoires avec un et trois pools de TMG

Le tableau des SCR obligatoires et leurs écarts de prédiction pour la BDD à un pool de TMG est présenté dans le tableau ci-dessous :

Modèle \ SCR de spread ¹⁴²	Prophet	SVM	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	10 418	10 150	10 272	10 085	8 647	10 339¹⁴³
Ecart	0,00%	-2,57%	-1,40%	-3,20%	-17,00%	-0,75%

Figure 128 Prédiction du spread obligatoire pour la BDD GSE continu, chocs obligatoires et 1 pool de TMG

L'algorithme de XGboost est capable de prédire un SCR avec une erreur inférieure à 1%.

¹⁴¹ Les inputs des scénarios +20/-20 bps n'ont pas été ajoutés à la base d'apprentissage

¹⁴² Chiffres exprimés en k€ pour ce type de tableau

¹⁴³ Tous les algorithmes avec des colonnes mises en gras dans ce type de tableau prédisent des résultats dans la target fixée

Le tableau des SCR obligataires et leurs écarts de prédiction pour la BDD à trois pools de TMG est présenté dans le tableau ci-dessous :

Modèle SCR de spread	Prophet	SVM	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	7 272	7 202	7 160	7 216	4 249	7 176
Ecart	0,00%	-0,96%	-1,53%	-0,77%	-41,57%	-1,32%

Figure 129 Prédiction du spread obligataire pour la BDD GSE continu, chocs obligataires et 3 pools de TMG

L'algorithme du SVM et du Random Forest sont capables de prédire un SCR avec une erreur inférieure à 1%.

4.3.2 Application du SCR de spread et de taux pour la base de données avec GSE variant, chocs obligataires et réallocation d'actif avec un et trois pools de TMG

Dans cette partie, le SCR de spread puis le SCR de taux seront présentés pour la BDD avec GSE variant, chocs obligataires et réallocation d'actif avec un et trois pools de TMG

4.3.2.1 SCR de spread

Le tableau des SCR obligataires et leurs écarts de prédiction pour la BDD à un pool de TMG est présenté dans le tableau ci-dessous :

Modèle SCR de spread	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	3 398	3 303	2 801	2 474	3 374
Ecart	0,00%	-2,80%	-17,57%	-27,19%	-0,70%

Figure 130 Prédiction du spread obligataire pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG

Le XGboost à un écart de prédiction inférieur à 1%

Le tableau des SCR obligataires et leurs écarts de prédiction pour la BDD à trois pools de TMG est présenté dans le tableau ci-dessous :

Modèle SCR de spread	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	3 247	3 413	3 834	5 988	3 533
Ecart	0,00%	5,12%	18,08%	84,41%	8,80%

Figure 131 Prédiction du spread obligataire pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG

La complexification du passif à fortement amplifiée les écarts de prédictions, si l'option de réallocation d'actif avait été désactivée, les résultats auraient probablement été meilleurs, par une augmentation de l'impact prédictif des inputs obligataires.

4.3.2.2 SCR de taux avec un choc à la baisse

Le tableau des SCR de taux à la baisse et leurs écarts de prédiction pour la BDD à un pool de TMG est présenté dans le tableau ci-dessous :

Modèle	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR de taux à la baisse					
SCR	2 300	2 691	2 686	3 362	957
Ecart	0,00%	16,98%	16,80%	46,16%	-58,41%

Figure 132 Prédiction du spread de taux à la baisse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG

Les prédictions sont bien au-delà de la target fixée, cela peut s'expliquer par un non-apprentissage de la courbe D20 et un changement de comportement du passif lors de la baisse des taux.

Le tableau des SCR de taux à la baisse et leurs écarts de prédiction pour la BDD à trois pools de TMG est présenté dans le tableau ci-dessous :

Modèle	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR de taux à la baisse					
SCR	6 987	6 233	6 235	6 224	5 283
Ecart	0,00%	-10,78%	-10,75%	-10,92%	-24,38%

Figure 133 Prédiction du spread de taux à la baisse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG

Même constat que pour le tableau précédent.

4.3.2.3 SCR de taux avec un choc à la hausse

Le tableau des SCR de taux à la hausse et leurs écarts de prédiction pour la BDD à un pool de TMG est présenté dans le tableau ci-dessous :

Modèle	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR de taux à la hausse					
SCR	3 371	582	621	597	1 583
Ecart	0,00%	-82,74%	-81,56%	-82,28%	-53,02%

Figure 134 Prédiction du spread de taux à la hausse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG

Le fait que l'algorithme n'est pas appris sur des scénarios à +20bps, et les changements de comportements des variables induits, rendent les prédictions aberrantes.

Le tableau des SCR de taux à la hausse et leurs écarts de prédiction pour la BDD à trois pools de TMG est présenté dans le tableau ci-dessous :

Modèle	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR de taux à la hausse					
SCR	5 608	5 902	5 824	5 819	4 329
Ecart	0,00%	5,24%	3,85%	3,77%	-22,81%

Figure 135 Prédiction du spread de taux à la hausse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG

Le scénario à +20bps et -20bps sont mieux prédits pour la database avec trois pools de TMG du fait que peu importe le choc, le BE est rémunéré par des coûts fonds propre (scénario très pessimiste).

Le scénario avec un pool de TMG à 0% est à perte de revenus (et non coût fonds propre). De ce fait avec un choc des taux à la baisse, le contrat passe dans la catégorie profitable, et avec choc à la hausse dans la catégorie onéreuse. Ces deux catégories n'ayant pas été apprises par les algorithmes, ils ne seront pas en mesure d'effectuer de bonnes prédictions sur les résultats.

4.3.2.4 Prédiction sur l'agrégation des deux modules

Pour l'agrégation des SCR, le SCR de taux le plus élevé pour les deux BDD est choisi. Pour la base de données avec un pool de TMG, le SCR de taux à la hausse est sélectionné et le SCR de taux à la baisse est pris pour la BDD à trois pools de TMG.

Conformément à la formule standard les SCR de taux et de spread sont agrégés à l'aide de l'équation suivante :

$$SCR_{agrégé} = \sqrt{SCR_{taux}^2 + SCR_{spread}^2 + 2 * Corr_{Interest,Spread} * SCR_{taux} * SCR_{spread}}$$

Avec :

$Corr_{Interest,Spread} = 0$ pour la BDD avec trois pools de TMG et 0,5 pour la BDD à un pool de TMG¹⁴⁴.

Les résultats suivants sont obtenus pour la BDD avec un pool de TMG :

Modèle \ SCR agrégé	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	5 862	3 629	3 158	2 820	4 385
Ecart	0,00%	-38,09%	-46,14%	-51,89%	-25,19%

Figure 136 Prédiction du spread de taux et de spread agrégés pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG

Le XGboost a tendance à mieux prédire le SCR de taux et de spread agrégés pour ce scénario.

Les résultats suivants sont obtenus pour la BDD avec trois pools de TMG :

Modèle \ SCR agrégé	Prophet	Gradient Boosting	Random Forest	Régression Multiple	XGboost
SCR	7 705	7 106	7 319	8 637	6 355
Ecart	0,00%	-7,77%	-5,00%	12,10%	-17,51%

Figure 137 Figure 138 Prédiction du spread de taux et de spread agrégés pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG

Malgré un écart un peu éloigné de la target à 1%, le Random Forest produit des résultats prometteurs pour la reprise de l'étude.

Pour le calcul de SCR de taux, il aurait fallu apprendre les scénarios avec les chocs de taux aux algorithmes.

¹⁴⁴ Voir partie 1, chapitre 1.3.3

Le SCR obligataire est dans la plupart des exemples bien prédit, bien qu'il soit sensible à une complexification du passif et à une baisse de l'impact des inputs obligataires.

4.4 Les limites du projet

Le projet présente les limites suivantes :

- Le GSE obligataire et la création des inputs obligataires est pensé pour des obligations à un coupon par an et ne prend pas en compte les coupons courus (adapter la table d'obligation à répliquer)
- Pas d'input de passif dans l'apprentissage, possibilité d'améliorer les résultats en ajoutant des inputs actions/immobiliers
- Les obligations étant déjà risque neutralisées, il n'est pas possible de vérifier si le spread de l'obligation répliqué est inférieur à 0 (bien faire attention aux bornes limites de choc en amont pour un cas pratique réel)
- Pas de calcul du SCR de marché complet
- Une amélioration de l'interprétation des résultats sortis par les algorithmes de Machine Learning peut être envisagée.
- Nécessité d'envisager un grand nombre de scénarios pour créer la base d'apprentissage de machine learning

Chapitre 5

5 Conclusion

Dans un premier temps, ce mémoire s'est consacré à l'implémentation d'un GSE de choc de spread obligataire, à l'aide de la création d'un outil inspiré de concept machine et deep learning.

Dans un deuxième temps, cet outil a été calibré puis utilisé pour créer des bases de Machine Learning. Le GSE de choc de spread s'est construit sur la base d'un réseau de neurone informatique composé d'un perceptron informatique générateur de choc appairé à un RNC pour la connexion à Prophet.

Dans un troisième temps, des algorithmes prédictifs ont été lancés afin d'estimer des BEL et des SCR de spread, de taux et leur agrégation.

Afin de tester la robustesse des algorithmes, deux types de passif ont été implémentés :

- Un passif avec un groupe de TMG à 0%
- Un passif avec trois groupes de TMG

Pour la sélection des variables, la data visualisation s'est effectuée à l'aide d'une étude des corrélations linéaires (Pearson) et de rang (Kendall).

Par leur faible proportion, aucune sélection des variables obligataires n'a été effectuée, l'élimination des variables de marché s'est faite par groupe de variables issues d'un même modèle stochastique.

Afin de synthétiser l'information des variables éliminées, des indicateurs moyens pour chaque groupe de variables issues du même modèle stochastique ont été créés.

Cette sélection des variables rigoureuse a permis l'élimination de 1548 et l'ajout de 12 variables explicatives. L'élimination des variables a permis une forte optimisation du temps d'apprentissage des algorithmes.

L'hyperparamétrage des algorithmes, s'est construit sur une cross validation k-fold, dont le critère à minimiser a été la moyenne des quantiles absolus des écarts en % entre les prédictions et les valeurs du modèle.

Après le lancement des algorithmes les prédictions au BE se situent dans la target fixée, avec un écart au BE maximum d'environ 1%.

Grâce à l'ajout des indicateurs moyens de marché, les prédictions ont été nettement améliorées. Avec un écart inférieur à environ 1% pour la Base de données GSE variant contre un écart >5% à la prise en main du projet.

L'étude des SCR a montré des résultats dans la target des 1% pour tous les SCR prédits à l'exception de la base de données avec chocs de spread, GSE variant, réallocation d'actif et trois TMG différents. Cet écart de prédiction par rapport aux autres jeux de données pouvant s'expliquer par une grande proportion de scénarios couts fonds propre.

Les prédictions des SCR de taux se sont montrées moins précises par le non-apprentissage des bases de données avec les chocs de taux, constituant une amélioration à l'étude.

L'étude a montré qu'il était possible de prédire le BE et des SCR de spread avec un écart de prédiction inférieur à environ 1% pour des scénarios relativement simples.

Cette étude pourrait être plus exhaustive en faisant intervenir un passif plus complexe en termes de TMG ainsi qu'un portefeuille d'actif plus complexe.

L'élaboration d'un processus d'utilisation du générateur de chocs développé sous python avec l'outil de réallocation d'actif sous Prophet pourrait permettre une amélioration des résultats.

Enfin, l'ajout d'indicateur actions et immobiliers permettraient d'améliorer les résultats des prédictions.

6 Annexes

6.1 Acronymes

ACPR : Autorité de Contrôle Prudentiel et de Résolution

ALM : Asset Liability Management

BDD : Base De Données

BE : Best Estimate

BEL : Best Estimate Liabilities

BOF : Basic Own Funds

BSCR : Brut Solvency Capital Requirement

CNN : Convolution Neuronal Network

DB/DBs : Database/ Databases

DLL : Dynamic Link Library

EIOPA : European Insurance and Occupational Pensions Authority

GSE : Economic Scenario Generator

FDB : Future Discretionary Benefits

FP : Fonds Propres

GANS : Generative Adversarial Neuronal networks

IA : Intelligence Artificielle

IDE : Internal Development environment

IFRS17 : International Financial Reporting Standard 17

MCEV : Market Consistent Embedded Value

MCR : Minimum Capital Requirement

ML : Machine Learning

NAV : Net Asset Value

OCDE : Organisation de Coopération et de Développement économique (majoritairement composée de pays d'Amérique du Nord et d'Europe)

OPCVM : Organisme de Placement Collectif en Valeur Mobilière

ORSA : Own Risk Solvency Assessment

PDD : Provision pour Déperdition Durable

PM : Provision Mathématique

PMVL : Plus ou Moins-Values Latentes

PT : Provisions techniques

PPB : Provison Participation aux bénéfices

PRE : Provision Risque d'Exigibilité

PVFP : Present Value of Future Profits

RC : Réserve de Capitalisation

RM : Risk Margin

RNC : Réseau de Neurone de Convolution

S2 : Solvabilité 2

SCR : Solvency Capital Requirement

SFCR : Solvency Financial Control Report

TMG : Taux Minimum Garanti

TVOG : Time Value of Options and Guarantee

UFR : Ultimate Forward Rate

VaR : Value at Risk

VM/MV : Valeur de Marché/Market Value

XGboost : Xtreme Gradient Boosting

YC : Yield Curve

ZCB : Zero Copounds bonds

6.2 Glossaire

Le glossaire est découpé en trois parties, la partie avec les notions d'actuariat utilisées dans le mémoire, une partie avec les notions statistiques et data-science et une partie orientée économie et finance.

6.2.1 Partie actuariat

ACPR : Régulateur français des assurances

Asset Liability Management : Gestion Actif-Passif. Pratique consistant à analyser la situation actuelle du bilan de l'entreprise, ainsi que son évolution possible sur un horizon de temps donné. Cette comparaison permet de déterminer si les engagements pris au passif sont cohérents avec les rendements des actifs.

Best Estimate of Liabilities : Il s'agit de « la meilleure estimation du passif ». Représente l'évaluation de tous les engagements de l'assureur au passif, bruts de réassurance. Les acronymes BE ou BEL pourra être utilisé pour désigner cette grandeur.

BSCR : Il s'agit du SCR considéré en brut de la capacité d'absorption des pertes par la participation aux bénéfices.

European Insurance and Occupational Pensions Authority : Autorité européenne des assurances et des pensions professionnelles. Comité regroupant l'ensemble des autorités de contrôle des pays membres de l'Union Européenne pour le domaine de l'assurance. Rôle consultatif auprès du parlement européen.

Minimum Capital Requirement : Capital minimum requis. Niveau de capital plancher en dessous duquel l'intervention du régulateur est automatique auprès de la société d'assurance. L'idée ici est que l'assureur ne fasse pas faillite dans l'année qui vient avec une probabilité de 85%.

Best Estimate of Liabilities : Il s'agit de « la meilleure estimation du passif ». Représente l'évaluation de tous les engagements de l'assureur au passif, bruts de réassurance. Les acronymes BE ou BEL pourra être utilisé pour désigner cette grandeur.

Générateur de Scénarios Économiques/GSE : Outil dont l'objectif est de projeter à un certain horizon de temps des indicateurs économiques de risques nécessaires à la simulation des actifs financiers.

Own Risk Solvency Assesment/ORSA : Évaluation interne du risque et de la solvabilité. Processus réglementaire par lequel une société d'assurance cartographie ses risques et anticipe leur évolution sur un horizon de temps au moins égal à son plan d'activité.

Participation aux Bénéfices/Provision pour Participation aux Excédents : Redistribution aux assurés par l'assureur d'une partie des bénéfices réalisés avec les fonds qui lui sont confiés. Cette fraction doit au moins représenter 85% du résultat financier net, 90% du résultat technique ainsi que 100% du solde de réassurance lorsque ceux-ci sont positifs. La PPB doit être redistribuée dans les huit années qui suivent le rattachement du bénéfice.

Pool de TMG : Représente un groupe de contrat au passif avec des assurés ayant le même TMG.

Prophet : Logiciel de simulation actuariel

Provision Mathématique : Représente l'engagement de l'assureur envers ses assurés. Dans le cadre d'un contrat en euros, les PM équivalent à l'ensemble des fonds confiés par les assurés et à leurs revalorisations accordées par l'assureur.

Ratio de solvabilité : Rapport entre les fonds propres économiques et le SCR. Le plancher légal est à 100%, mais un ratio supérieur est apprécié afin de montrer la solidité financière de l'entreprise qui le présente.

Risk Margin : « Marge de risque ». Constitue les provisions techniques avec le Best Estimate. Cette valeur augmente le BEL au niveau théorique nécessaire au transfert des engagements à un autre assureur. Cette valeur est obtenue en multipliant le coût du capital de l'assureur par le SCR nécessaire pour supporter le transfert des engagements de l'assureur, tout en tenant compte de l'actualisation.

Réserve de Capitalisation : Réserve constituée par les plus-values réalisées sur la cession d'obligations et les reprises effectuées dans le cas de moins-values sur ce type d'actif. Elle permet de lisser le résultat de l'assureur sur plusieurs années.

Ruine : Situation dans laquelle un assureur ne peut plus faire face à ses engagements.

Solvency Capital Requirement/SCR : Capital de solvabilité requis. Capital à détenir pour absorber un choc important. En théorie, le SCR représente le capital nécessaire pour que l'assureur ne fasse pas faillite dans l'année qui vient avec une probabilité de 99,5%.

Taux Minimum Garanti : Taux de revalorisation annuel appliqué à l'épargne investie sur le fonds du contrat euros. Plus précisément, il s'agit de l'engagement minimum pris par l'assureur envers les assurés.

XSG : logiciel utilisé comme GSE

6.2.2 Partie statistique et data-science

Biais : Un modèle Machine Learning est dit biaisé s'il prédit des valeurs avec le même écart de prédiction pour différents quantiles de prédiction.

Bootstrap : Méthode d'échantillonnage permettant de reproduire la distribution d'une loi statistique en effectuant des tirages aléatoires avec remise sur un vecteur suivant la loi statistique à échantillonner.

Fonction de coût : Estimateur statistique qu'un algorithme d'apprentissage automatique va chercher à minimiser lors de la construction de son modèle prédictif.

Filtrage : Terme issu du deep learning qui désigne l'application d'un filtre (nommée matrice de convolution) sur un objet informatique afin de faire ressortir certaines caractéristiques de cet objet.

L'Intelligence Artificielle/IA/Artificial Intelligence/AI :

L'intelligence artificielle est un domaine de la statistique et de l'informatique regroupant les algorithmes de Machine Learning, de deep learning et d'optimisation (dont la descente de gradient est un exemple).

Couche de neurone/Layer : Terme issu du deep-learning désignant une couche comprenant différents neurones reliés à une autre couche de neurone (voir Illustration de la possibilité d'ajouter plusieurs neurones à la couche du générateur des données P88)

Pooling : Terme issu du deep learning qui consiste à condenser les informations les plus importantes d'un objet informatique.

Sélection des variables : Etape effectuée lors d'un projet de ML où les variables les plus importantes d'une BDD vont être sélectionnées afin d'optimiser le temps de calcul et/ou d'améliorer les résultats de certains algorithmes.

Sur-apprentissage : Désigne un modèle informatique autoapprenant étant trop complexe pour la tâche qui lui est confiée. Un modèle qui a surappris va généralement être très sensible à une variation de sa base de données à prédire.

Sous-apprentissage : Désigne un modèle informatique autoapprenant n'étant pas assez complexe pour la tâche qui lui est confiée. Un modèle qui a sous appris va généralement produire des mauvais résultats sur sa base de données initiale mais va être peu sensible à une variation de sa base de données.

Sous-échantillonnage/Undersampling/Bagging : Méthode d'échantillonnage statistique consistant à reconstruire un dataset d'apprentissage et/ou de test en construisant un dataset réduit par rapport au dataset initial. La construction du dataset réduit se fait à l'aide de tirages effectués sur le dataset initial.

Sur-échantillonnage/Oversampling : Méthode d'échantillonnage statistique consistant à reproduire des données issues d'un dataset initial afin de l'enrichir et/ou de l'agrandir.

Target : Variable à prédire pour les algorithmes de ML ou seuil d'erreur de prédiction maximum fixé.

Weak-learner : neurone « simple », se comportant le plus souvent comme un arbre de décision (voir le schéma intitulé L'arbre de décision). Assemblés en grand nombre ces neurones donnent de très bons résultats de classification ou de prédiction.

6.2.3 Partie économique et financière

Agence de notation économique/ de rating : Organisme qui va attribuer une note à un acteur économique (entreprise, état, mutuelle ...) indiquant la solidité financière et économique de cet acteur. Les trois plus grandes agences de notation sont Moodies, Standard and Poors et Fitch

Emprunt in-fine : Emprunt dont les intérêts sont remboursés chaque année et dont le remboursement du montant initial se fait entièrement la dernière année

Les effets structurels économiques : Répercussions économiques moyen-long terme

Les effets conjoncturels économiques : Répercussions économiques court terme

La probabilité de défaut : Souvent calculée sur le spread d'un instrument financier, elle représente les chances qu'une entreprise ne soit plus en disposition de répondre à ses obligations financières envers ses créanciers, actionnaires...

OPCVM : Instrument financier structuré par différentes classes d'actifs (action, obligation...) le plus souvent mis en circulation par un fonds d'investissement.

Put : Option financière qui donne le droit à son détenteur de vendre un actif. Selon l'origine du put, il est possible de l'exercer à tout instant avant son expiration ou à la date de son exécution.

Scénario à perte de revenu : Trajectoire déterministe ou stochastique dans laquelle l'assureur ne réalise pas un profit suffisant pour se rémunérer et/ou rémunérer ses actionnaires

Scénario à cout fonds propres : Trajectoire déterministe ou stochastique dans laquelle l'assureur ne peut répondre à ses engagements envers les assurés.

Sous-jacent d'un actif : Ce terme désigne le prix contractuel d'un instrument financier sur lequel une option peut être exercée (un achat ou une vente à un instant t par exemple). Ce prix contractuel peut-être estimé sur des instruments financiers tels que les actions, les obligations etc...

Spread : Ecart de valeur entre la valeur « théorique » et la valeur « réelle » d'un instrument financier sur le marché. Plus le spread est élevé plus les marchés et/ou la compagnie délivrant les titres financiers sont en mauvaise posture financière.

Swaption : Option qui permet d'entrer dans un swap de taux, qui est un contrat définissant à une date donnée l'échange d'un taux fixe contre un taux variable observé sur une période.

Taux sans risque spot : Taux de rendement du risque neutre à court terme, son déflateur se calcule à l'aide de la formule suivante :

$$\text{Déflateur}_{spot_n} = \frac{1}{(1 + \text{taux}_{spot})^n}$$

Taux sans risque forward : Taux de rendement risque neutre entre deux périodes fixées, son déflateur entre l'année 0 et n se calcule à l'aide de la formule suivante :

$$\text{Déflateur}_{forward_n} = \frac{1}{(1 + \text{taux}_{forward})^n}$$

Volatilité : Terme financier désignant la variance des rendements d'un actif. Plus cette mesure est élevée, plus l'investisseur prend des risques.

Value at Risk : Définie par rapport à un horizon de temps T pour un seuil α , correspond au quantile de niveau α de la distribution des pertes en T .

Zero Copounds bonds : Désigne une obligation zéro coupon.

6.3 Glossaire numérique

Le glossaire numérique présente les notations mathématiques et financières utilisées dans le mémoire pour les démonstrations et les descriptions scientifiques des sujets abordés.

6.3.1 Pour l'ensemble du mémoire

- CF_t ou CF_i cashflows d'un actif à l'incrément de temps i ou t
- r_f risk free rate ou taux sans risque
- μ espérance mathématique d'une loi statistique
- σ écart type d'une loi statistique
- X vecteur représentant les valeurs d'une loi statistique
- X_{min} valeur minimum du vecteur X représentant une loi statistique
- X_{max} valeur maximum du vecteur X représentant une loi statistique
- X_{Custom} vecteur de valeur dont l'allure de la fonction de densité a été choisie par l'utilisateur
- \hat{y} valeur prédite
- \bar{y} moyenne empirique des valeurs d'un dataset

6.3.2 Partie 1

Variables pour le modèle de Black and Sholes (immobilier)

- dF_t représente la différentielle de l'évolution du sous jacent dans le temps
- F_t et S_t représente les sous jacents d'un actif
- μ_t représente le rendement d'un actif non risqué
- σ_t représente la prime de risque dans les rendements d'un actif
- S_0 représente le prix de départ d'un actif
- dB_t représente la différentielle d'un mouvement brownien dans le temps
- B_t représente un mouvement brownien, pouvant être traduit par une loi normale centrée $N(0, \sqrt{t})$

Variables pour le modèle de Heston (action)

- dS_t représente la différentielle de l'évolution du sous jacent dans le temps
- S_t représente le sous jacents d'un actif selon Black and Sholes
- B_t et W_t représente des mouvements Browniens distincts
- ϑ_t représentent le sous jacents d'un actif ajouté à la formule de Black and Sholes pour créer une volatilité implicite supplémentaire
- dB_t et dW_t représentent les différentielles de mouvements browniens distincts dans le temps
- k, θ, ε constantes

Variables pour le modèle de Heston (action)

- $\alpha(t, T), A_{t,T}$ fonction représentant la partie structurelle de l'évolution des taux entre deux intervalles de temps T et t
- $\sigma(t, T), B_{t,T}$ fonction représentant la partie conjoncturelle de l'évolution des taux entre deux intervalles de temps T et t .
- $VM_{T,t}$ valeur de marché d'une obligation zéro coupon entre deux intervalles de temps
- θ_t Un paramètre dépendant des taux forwards
- r_t rendement d'un actif non risqué
- B_t et W_t représentent des mouvements Browniens distincts
- dB_t et dW_t représentent les différentielles de mouvements browniens distincts dans le temps

Variables pour le modèle LMM

Les variables contenues dans le modèle LMM sont présentés de manière claire dans la partie résumant ce modèle stochastique.

6.3.3 Partie 2

- a, b, k et m sont des constantes
- U représente un vecteur de valeurs suivant une loi uniforme $U(0,1)$
- $X_{i(\omega)}$ Élément d'un vecteur issu d'une distribution statistique
- σ_X ou σ_Y deux écarts types associés aux vecteur X ou

6.4 Table des figures

Figure 1 schéma du réseau de neurone pour la génération des inputs obligataires	7
Figure 2 schéma du convertisseur inspiré d'un RNC	7
Figure 3 densité des chocs obligataires grecs	8
Figure 4 densité des chocs obligataires françaises	8
Figure 5 répartition des actifs des Bases De Données (BDD) pour le ML	9
Figure 6 Présentation des résultats par quantiles pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG, dataset test	10
Figure 7 Présentation des résultats par quantiles pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG, dataset test	11
Figure 8 SCR de spread obligataire pour une BDD chocs obligataires, GSE variant et réallocation d'actif avec 1 pool de TMG	11
Figure 9 Neuronal bond's generator chart	13
Figure 10 CNN chart	13
Figure 11 Density of Greek bonds shock	14
Figure 12 Density of Français bonds shock	14
Figure 13 initial asset allocation	15
Figure 14 Results by quantile for the DB5 with 3 pools of minimum rate, test dataset	16
Figure 15 Results for the DB5 with 3 pools of minimum rate, French dataset	16
Figure 16 Prediction of the spread SCR for a DB with varying ESG, moving asset allocation, bonds shocks and 1 pool of MR	17
Figure 17 Schéma représentant les trois piliers de S2 (source réavie de S2 de la FFSa)	22
Figure 18 Bilan Solvabilité II (Source Deloitte)	23
Figure 19 Schéma du calcul du BSCR (Source Deloitte)	24
Figure 20 Représentation de la TVOG (Source Deloitte)	25
Figure 21 Sous modules du calcul du SCR (source Deloitte)	26
Figure 22 Représentation du choc action (Source Deloitte)	26
Figure 23 Matrice de corrélation du SCR marché.....	27
Figure 24 Représentation du FDB (Source Deloitte).....	27
Figure 25 Chocs de spread obligataire selon la formule standard S2 (Source Deloitte)	29
Figure 26 Tableau de conversion des notations obligataires en échelon de crédit (Source Deloitte).....	30
Figure 27 Schéma des inputs et outputs d'un logiciel de simulation actuariel.....	30
Figure 28 Graphique d'un test de martingalité action.....	34
Figure 29 Densité d'une loi uniforme générée sous Python	37
Figure 30 Densité d'une loi normale générée sous Python	37
Figure 31 Densité d'une loi de Pareto générée sous Python	38
Figure 32 Densité d'une loi de Weibull générée sous Python.....	39
Figure 33 Densité d'un mélange d'une loi de Pareto/Weibull générée sous Python.....	40
Figure 34 Densité d'un échantillon de loi uniforme généré par bootstrap	40
Figure 35 Qqplot d'une loi normale VS uniforme.....	41
Figure 36 La démarche du datascientist	43
Figure 37 Découpage d'un dataset pour effectuer une 3-fold cross validation.....	43
Figure 38 Schéma sur les différentes composantes de l'IA.....	44
Figure 39 Les différentes composantes des réseaux de neurones illustrées de usecases	44
Figure 40 L'arbre de décision.....	45
Figure 41 Fonctionnement d'un Random Forest	46
Figure 42 Fonctionnement d'un XGboost	46
Figure 43 Fonctionnement d'un perceptron informatique.....	48
Figure 44 Le réseau de neurone de convolution (images de source Deloitte)	49
Figure 45 Table sortie par le générateur de spread aléatoire uniforme initialement codé	51
Figure 46 Schéma du réseau de neurones générateurs des chocs obligataires	52
Figure 47 Première couche de neurone du générateur de données obligataires	52
Figure 48 Fonction de densité d'une loi de choc normale après une transformation min-max	53
Figure 49 Fonction de densité d'une loi normale après une transformation min-max et l'ajout d'un biais.....	53
Figure 50 Fonction de densité d'une loi normale après une transformation min-max, l'ajout d'un biais et d'un facteur de variance multiplicatif	54
Figure 51 Fonction de densité d'une loi normale après une transformation min-max, l'ajout d'un biais, d'un facteur de variance multiplicatif et d'un biais supplémentaire	54
Figure 52 Illustration de la possibilité d'ajouter plusieurs neurones à la couche du générateur des données	55
Figure 53 Exemple de fonction de densité produite par un réseau à deux neurones générateurs de chocs	55
Figure 54 Convertisseur inspiré d'un RNC	56
Figure 55 Table de chocs générée par le convertisseur.....	56
Figure 56 Schéma des nouveaux inputs branchés au simulateur actuariel.....	56
Figure 57 Densité des coefficients de chocs obtenus pour l'univers de Pareto	58
Figure 58 Densité des coefficients de chocs obtenus pour l'univers Normal	58
Figure 59 Qqplot de la densité des coefficients de chocs obtenus pour l'univers de choc normal.....	59
Figure 60 Densité des coefficients de chocs obtenus pour l'univers Uniforme	59
Figure 61 Densité des coefficients de chocs obtenus pour l'univers de Weibull.....	60
Figure 62 Densité des coefficients de chocs obtenus pour l'univers Pareto pour les coupons et Normaux pour le reste.....	60
Figure 63 Densité des coefficients de chocs obtenus pour l'univers de choc 6	61
Figure 64 Densité des VM obligataires du portefeuille initial.....	61
Figure 65 Vérification du bon fonctionnement du réseau de neurone par la réplique exacte du portefeuille initial.....	62
Figure 66 Présentation d'un échantillon d'univers choisis pour l'utilisation du réseau de neurone générateur de choc.....	62
Figure 67 Comparaison de la densité des VM d'un portefeuille répliqué avec l'algorithme de portefeuilles répliquants et le portefeuille initial	63
Figure 68 graphique sur l'évolution des coupons obligataires français (source : trading economics)	63
Figure 69 densité des chocs obligataires français	64
Figure 70 graphique sur l'évolution des coupons obligataires grecques (source : trading economics)	65
Figure 71 densité des chocs obligataires grecques	66
Figure 72 Paramètres ayant permis la construction des densités de chocs obligataires des bases grecques et françaises.....	66
Figure 73 répartition des actifs des BDD pour le ML.....	69
Figure 74 Inputs ML obligataires	69
Figure 75 Présentation des BDDs.....	70
Figure 76 Répartition des TMG pour les databases à trois pools.....	70
Figure 77 Processus de production des résultats des algorithmes de ML.....	72
Figure 78 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires avec 1 pool de TMG	74
Figure 79 Production financière VS l'engagement de l'assureur Base de données chocs obligataires avec 1 pool de TMG	74
Figure 80 Tableau récapitulatif des performances des algorithmes de ML Base de données chocs obligataires avec 1 pool de TMG	75
Figure 81 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires avec 1 pool de TMG.....	76

Figure 82 Ecart aux quantiles 10% et 90% par algorithmes Base de données chocs obligataires avec 1 pool de TMG	76
Figure 83 RMSE par algorithme Base de données chocs obligataires avec 1 pool de TMG	76
Figure 84 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires avec 3 pools de TMG	77
Figure 85 Production financière VS l'engagement de l'assureur Base de données chocs obligataires avec 3 pools de TMG	77
Figure 86 Tableau récapitulatif des performances générales des algorithmes de ML Base de données chocs obligataires avec 3 pools de TMG	78
Figure 87 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires avec 3 pools de TMG	79
Figure 88 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires et GSE variant avec 1 pool de TMG	80
Figure 89 Production financière VS l'engagement de l'assureur Base de données chocs obligataires et GSE variant avec 1 pool de TMG.....	81
Figure 90 Tableau récapitulatif des performances générales des algorithmes de ML Base de données chocs obligataires et GSE variant avec 1 pool de TMG.....	82
Figure 91 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires et GSE variant avec 1 pool de TMG	83
Figure 92 Tableau d'écart aux quantiles de BE, BDD chocs obligataires et GSE variant avec 1 pool de TMG	83
Figure 93 Evolution du BE en fonction de la VM par percentile Base de données chocs obligataires et GSE variant avec 3 pools de TMG	84
Figure 94 Production financière VS l'engagement de l'assureur Base de données chocs obligataires et GSE variant avec 3 pools de TMG	84
Figure 95 Tableau récapitulatif des performances générales des algorithmes de ML Base de données chocs obligataires et GSE variant avec 3 pools de TMG	85
Figure 96 Courbe des BE du dataset test prédits par percentile pour la Base de données chocs obligataires et GSE variant avec 3 pools de TMG	85
Figure 97 Evolution du BE en fonction de la VM par percentile Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG	88
Figure 98 Production financière VS l'engagement de l'assureur Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG	88
Figure 99 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG	89
Figure 100 Comparaison des écarts aux quantiles entre la Base de données GSE variant et la Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG	89
Figure 101 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant et indicateurs obligataires avec 1 pool de TMG	90
Figure 102 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 1 pool de TMG sur le dataset U20	91
Figure 103 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 1 pool de TMG sur le dataset D20	91
Figure 104 Evolution du BE en fonction de la VM par percentile Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG	92
Figure 105 Production financière VS l'engagement de l'assureur Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG	92
Figure 106 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG.....	93
Figure 107 Comparaison des écarts aux quantiles entre la Base de données GSE variant et la Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG	94
Figure 108 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG	94
Figure 109 Ecart aux quantiles 10% et 90% par algorithmes Base de données GSE variant et indicateurs obligataires avec 3 pools de TMG	95
Figure 110 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 3 pools de TMG sur le dataset U20	95
Figure 111 Amélioration des résultats entre les prédictions de la Base de données GSE variant et 4 avec 3 pools de TMG sur le dataset D20	96
Figure 112 Evolution du BE en fonction de la VM par percentile Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	97
Figure 113 Production financière VS l'engagement de l'assureur Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	97
Figure 114 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	98
Figure 115 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	99
Figure 116 RMSE par algorithme Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	99
Figure 117 Ecart aux quantiles sur le dataset U20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	99
Figure 118 Ecart aux quantiles sur le dataset D20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	100
Figure 119 Ecart aux quantiles sur le dataset français, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 1 pool de TMG	100
Figure 120 Evolution du BE en fonction de la VM par percentile Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG	101
Figure 121 Production financière VS l'engagement de l'assureur Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG	101
Figure 122 Tableau récapitulatif des performances générales des algorithmes de ML Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG	102
Figure 123 Courbe des BE du dataset test prédits par percentile pour la Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG	103
Figure 124 Ecart aux quantiles 10% et 90% par algorithmes Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG.....	103
Figure 125 Ecart aux quantiles sur le dataset U20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG.....	104
Figure 126 Ecart aux quantiles sur le dataset D20, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG.....	104
Figure 127 Ecart aux quantiles sur le dataset français, Base de données GSE variant, chocs obligataires et réallocation dynamique avec 3 pools de TMG.....	104
Figure 128 Prédiction du spread obligataire pour la BDD GSE continu, chocs obligataires et 1 pool de TMG	105
Figure 129 Prédiction du spread obligataire pour la BDD GSE continu, chocs obligataires et 3 pools de TMG.....	106
Figure 130 Prédiction du spread obligataire pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG.....	106
Figure 131 Prédiction du spread obligataire pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG	106
Figure 132 Prédiction du spread de taux à la baisse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG	107
Figure 133 Prédiction du spread de taux à la baisse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG	107
Figure 134 Prédiction du spread de taux à la hausse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG.....	107
Figure 135 Prédiction du spread de taux à la hausse pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG	107
Figure 136 Prédiction du spread de taux et de spread agrégés pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 1 pool de TMG	108
Figure 137 Figure 138 Prédiction du spread de taux et de spread agrégés pour la BDD GSE variant, chocs obligataires, réallocation d'actifs et 3 pools de TMG.....	108

6.5 Hyperparamétrage XGboost et la descente de gradient

Lors de l'utilisation de nos algorithmes de XGboost, les hyperparamètres suivants ont été utilisés :

- **La fonction objectif** : Est l'un des hyperparamètres les plus importants lorsque des algorithmes de Machine Learning sont implémentés. Il s'agira d'un indicateur à maximiser ou minimiser lors de l'étape d'apprentissage de l'algorithme. En général un estimateur de l'espérance (erreur des moindres carrés) ou de la vraisemblance (cross-entropy) sont choisis. L'estimateur convergera vers un minimum ou un maximum à l'aide d'un algorithme de descente de gradient.
- **Le paramètre n_estimators** : permet de choisir le nombre de weak-learner que va comporter le Random Forest ou XGboost, plus ce nombre est élevé, plus le modèle sera complexe.
- **Le paramètre max_depth** permet de contrôler la profondeur des arbres de décisions utilisés pour l'implémentation de l'algorithme, moins ces weak learners seront profonds, moins le modèle une fois agrégé sera complexe.
- **Le paramètre de learning rate** permet d'ajuster la descente de gradient stochastique de l'algorithme lors de l'étape de maximisation/minimisation de la fonction objectif, plus le learning rate est petit, plus la convergence vers l'estimateur de la fonction objectif est supposée précise mais la compilation de l'algorithme sera lente.
- **Le paramètre Colsample_bytree** permet de régler la taille des sous échantillons tirés pour effectuer l'apprentissage des weak-learners.
- **Le coefficient de pénalisation L^1 (Lasso)** permet de réduire/régler le biais d'un modèle.
La descente de gradient, un algorithme d'optimisation utilisé pour maximiser

La descente de gradient qui permet de minimiser les fonctions objectifs du XGboost et du Random Forest

La descente de gradient est un algorithme d'optimisation dont l'objectif est de rechercher un estimateur d'une variable quantitative. Elle intervient lors de la minimisation ou la maximisation de la fonction objectif pendant l'étape d'apprentissage d'un algorithme de ML.

Algorithme :

1. Initialiser avec x_0 (au hasard)
2. Répéter $x(t+1) = x(t) - \eta \times \nabla f(x(t))$
3. Jusqu'à convergence

Avec :

- $\nabla f(x(t))$ le gradient (dérivée multidimensionnelle) de la fonction du modèle
- η est un paramètre de correction de la convergence du modèle (convergence trop lente si trop petit, Oscillation du modèle si trop élevé).
- A noter que le nombre d'itérations est soit fixé au départ, ou se fait par la différence entre valeurs successives $x(t)$, ou lorsque $\nabla f(x(t))$ est très petit.

Exemple d'une descente de gradient

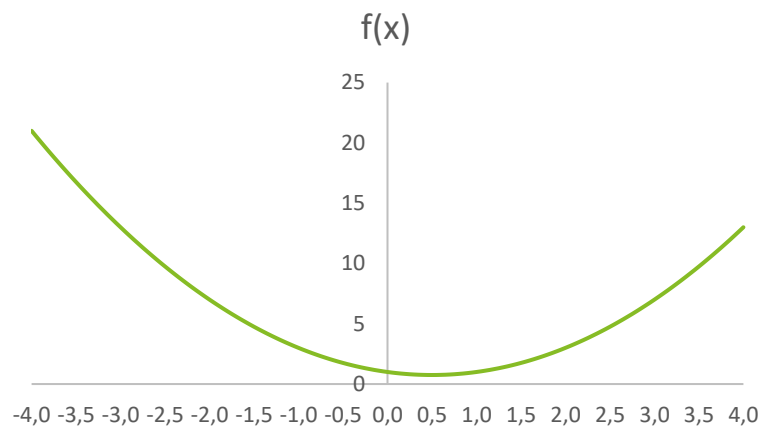
Fonction :

$$f(x) = x^2 - x + 1$$
$$\nabla f(x) = \partial f(x) / \partial x = f'(x) = 2x - 1$$

Algorithme :

x	f'(x)	f(x)	n
5,00		21,00	0,3
2,30	9,00	3,99	
1,22	3,60	1,27	
0,79	1,44	0,83	
0,62	0,58	0,76	
0,55	0,23	0,75	
0,52	0,09	0,75	
0,51	0,04	0,75	
0,50	0,01	0,75	

Courbe de la fonction :



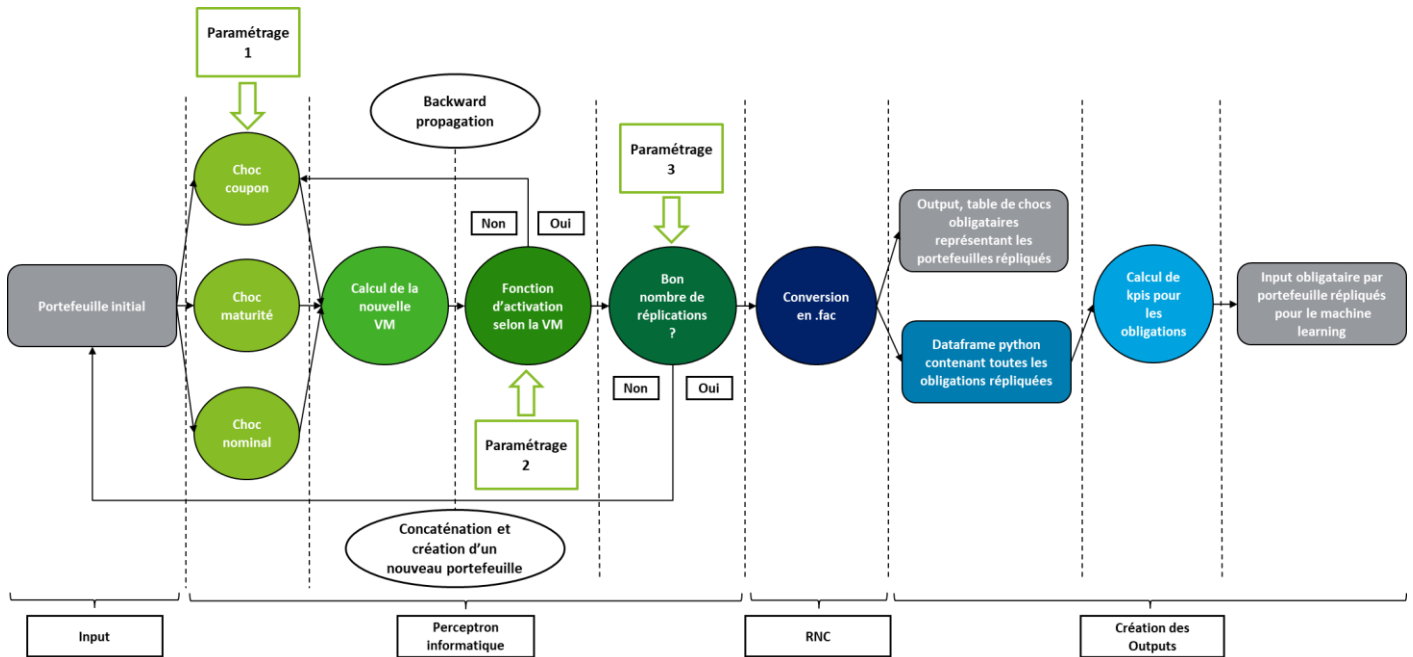
Conclusion : la fonction converge vers un minimum lorsque $x=0,5$

6.6 Hyperparamétrage des réseaux de neurones développés

Cette annexe se consacre au détail de l'hyperparamétrage du réseau de neurone développé et à l'hyperparamétrage de situations testées.

6.6.1 Hyperparamétrage du perceptron informatique

Pour rappel le réseau de neurone peut être représenté comme ci-dessous :



Hyperparamétrage déclaré une fois pour le réseau de neurone :

- Année du début de la simulation
- Portefeuille initial
- Taux EIOPA
- Nombre de réplification du portefeuille
- L'activation de la backward ou non
- Le nombre maximum d'itération effectuées par la backward que l'utilisateur peut rentrer
- La borne inférieure des coefficients de risque neutralisations souhaités par l'utilisateur en cas de backward propagation
- La borne supérieure des coefficients de risque neutralisations souhaités par l'utilisateur en cas de backward propagation
- Le numéro de portefeuille initialisé à 0 automatiquement pour signaler les obligations qui ont des coefficients de risques neutralisations hors des bornes fixés par l'utilisateur
- Pour chaque réseau de choc (*3) :
- L'option de la loi statistique à prendre :
- Option loi 1 : Pareto param1=coeff de Pareto, pas de param2
- Option loi 2 : Gaussienne param1= μ , param2= sigma
- Option loi 3 : Uniforme param1=a, param2=b

- Option loi 4 : Weibull param1=lambda, param2=k
- Un choc de dissymétrie à l'intérieur de la normalisation (pensé pour être compris entre -1 et 1)
- Un facteur d'intensité multiplicatif pour élargir la fonction de répartition de la courbe
- Un facteur de dissymétrie externe à la normalisation pour l'introduction d'un biais.

Rappel de la formule de choc :

$$X_{custom} = (X_{normalise} - Dissymétrie_{int}) * Facteur_{intensité} + Dissymétrie_{ext}$$

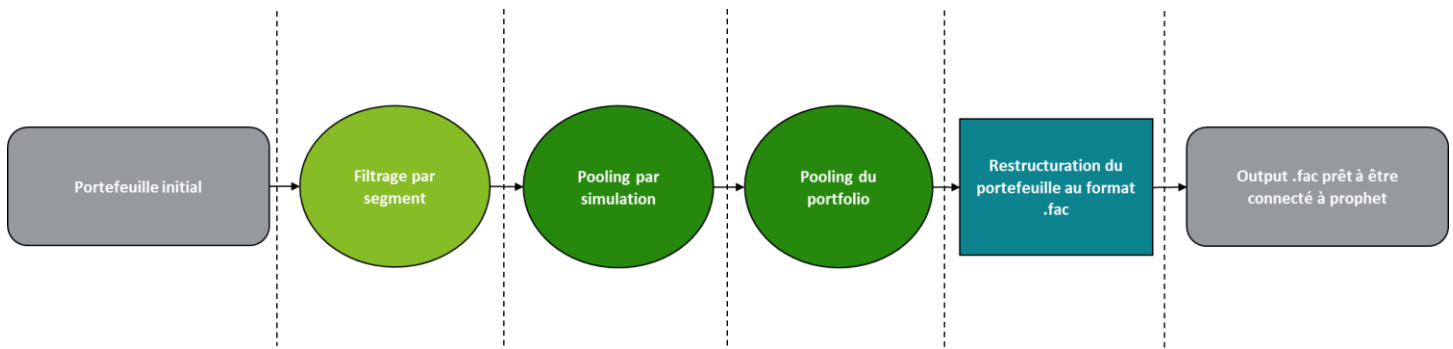
Tableau récapitulatif des hyperparamètres utilisés :

Univers	Pareto	Normal	Uniforme	Weibull	Pareto /Normal	Normal /Uniforme	Normal /Weibull
Type de loi Coupons	Pareto	Normale	Uniforme	Weibull	Pareto	Normale	Normale
Paramètre loi 1 Coupons	20	0	-0,5	4	20	0	0
Paramètre loi 2 Coupons	-	1	0,5	4	-	1	1
Facteur Dissymétrie int Coupons	-0,5	-0,5	-	-0,5	-0,5	-0,5	-0,5
Facteur Dissymétrie ext Coupons	-	-	-	-	-	-	-
Type de loi Nominal	Pareto	Normale	Uniforme	Weibull	Normale	Uniforme	Weibull
Paramètre loi 1 Nominal	20	0	-0,5	4	0	-0,5	4
Paramètre loi 2 Nominal	-	1	0,5	4	1	0,5	4
Facteur Dissymétrie int Nominal	-0,5	-0,5	-	-0,5	-0,5	-	-0,5
Facteur Dissymétrie ext Nominal	-	-	-	-	-	-	-
Type de loi Maturité	Pareto	Normale	Uniforme	Weibull	Normale	Normale	Normale
Paramètre loi 1 Maturité	20	0	-0,5	4	0	0	0
Paramètre loi 2 Maturité	-	1	0,5	4	1	1	1
Facteur Dissymétrie int Maturité	-0,5	-0,5	-	-0,5	-0,5	-0,5	-0,5
Facteur Dissymétrie ext Maturité	-	-	-	-	-	-	-
Temps de répliation ¹⁴⁵	53.85	19.24	23	23.44	21.64	21,9	47,42
Borne supérieur Coeff RN	1,3	1,3	1,3	1,3	1,3	1,3	1,3
Borne inférieur Coeff RN	0,7	0,7	0,7	0,7	0,7	0,7	0,7
Type de Symétrie	Centrée à gauche	Centrale	Aucune	Centrale à droite	Centrale à droite	Aucune	Centrale

¹⁴⁵ Exprimé en secondes pour la génération de 100 chocs

6.6.2 Hyperparamétrage du RNC

Pour rappel le RNC peut être représenté comme ci-dessous :



L'hyperparamétrage du réseau RNC se définit comme suit :

- Data frame à convertir
- Nombre de simulations
- Nombre de dimensions souhaitées pour la table Prophet
- La liste du nom des bucket du portefeuille initiale (indicateur visuel qui permet de vérifier le bon traitement des datas par le réseau).

6.6.3 Test du neurone dans un univers choqué non central

Analyse univers de choc Normal modifié avec des bornes réduites entre 0.69999 et 0.700001

Afin de vérifier la possibilité de répliquer exactement la même répartition de portefeuille obligataire tout en changeant la valeur des VM, la configuration suivante a été adoptée :

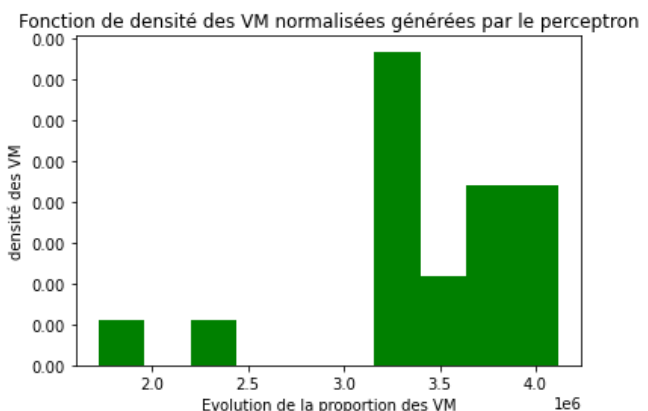
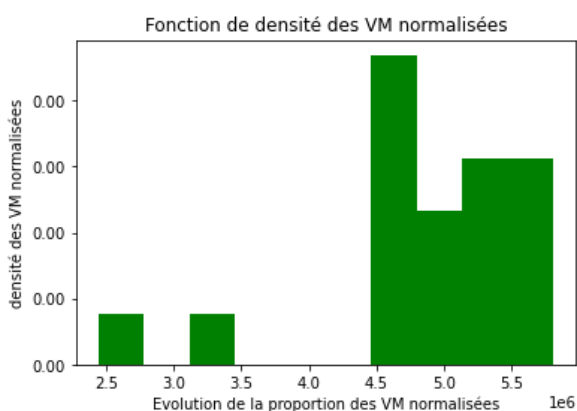
$$Facteur_{intensité} = 0$$

$$Dissymétrie_{ext} = -0,3$$

Dans cette configuration, les VM du portefeuille répliqué devrait avoir perdues 30% de leur valeur initiale.

$$X_{custom} = (X_{normalise} - Dissymétrie_{int}) * Facteur_{intensité} + Dissymétrie_{ext}$$

Comparaison des deux densités



Les deux densités sont identiques.

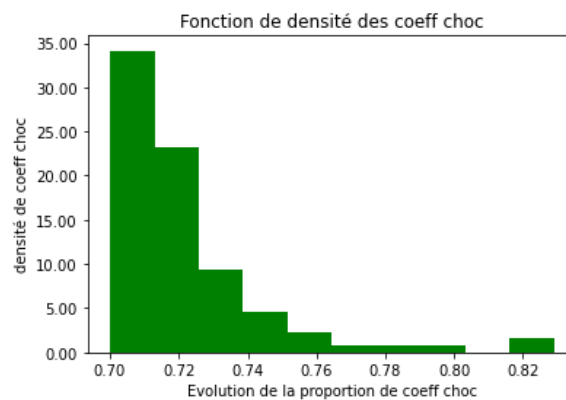
La VM maximal du portefeuille initial est de l'ordre de 6M€ contre une VM maximale de l'ordre 4M€ pour le portefeuille répliqué soit une différence d'environ 30%.

Cette d'analyse montre que le réseau de neurone codée permet de répliquer avec un fort réalisme la fonction de densité des VM du portefeuille obligataire et peut donc être utilisé comme une méthode de portefeuille répliquant obligataire.

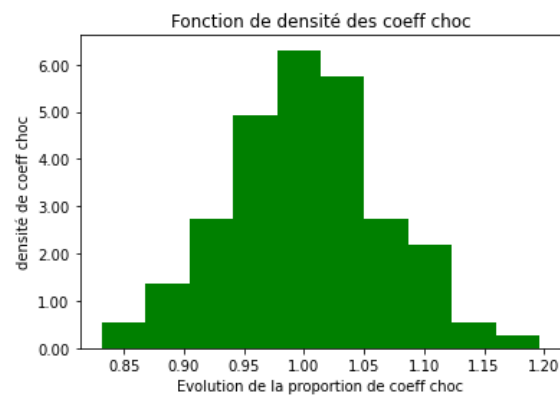
6.7 Courbe de densité des coefficients de RN par univers pour 100 portefeuilles répliqués

Cette annexe présente les différentes courbes de densité de chocs sorties par le réseau de neurone lors de son calibrage.

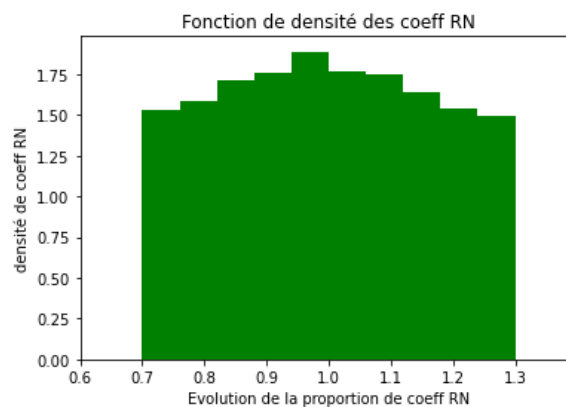
6.7.1 Univers de Pareto



6.7.2 Univers Normal

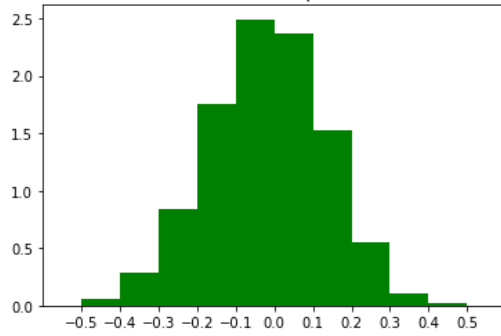


6.7.3 Univers Uniforme



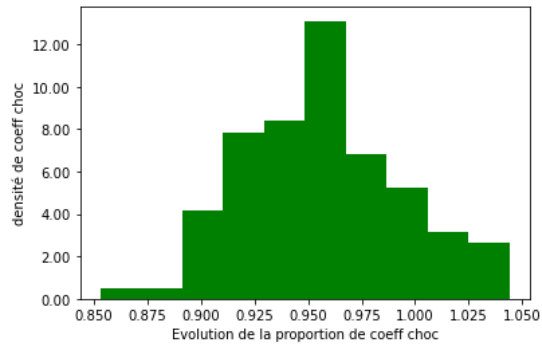
6.7.4 Univers de Weibull

Fonction de densité d'une loi de weibull pour des chocs entre -50 et 50%



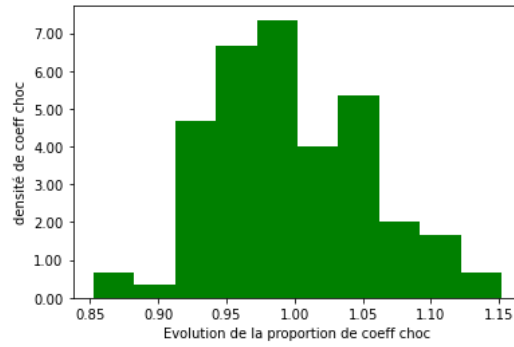
6.7.5 Univers Pareto/Normal

Fonction de densité des coeff choc



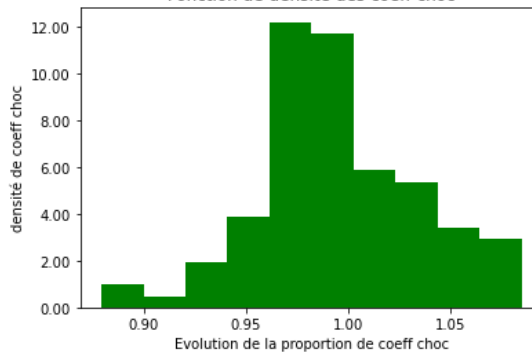
6.7.6 Univers Normal/Uniforme

Fonction de densité des coeff choc



6.7.7 Univers Normal/Weibull

Fonction de densité des coeff choc



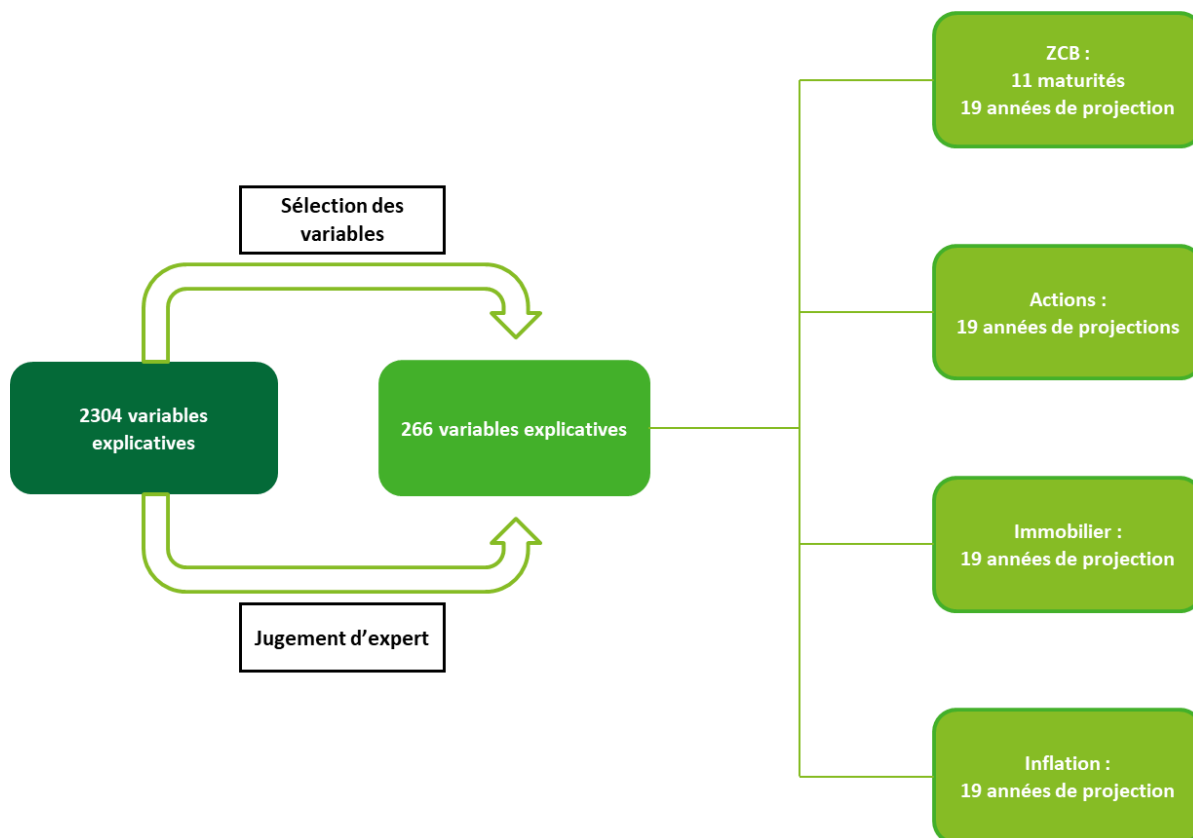
6.8 Tableau des univers de création des densités grecques et françaises

Univers	Français	Grecques queue distribution à gauche	Grecques autres parties de la distribution
Type de loi Coupons	Normale	Uniforme	Gaussienne
Paramètre loi 1 Coupons	0	-0,6	0
Paramètre loi 2 Coupons	1	0,2	1
Facteur Dissymétrie int Coupons	-0,5	0	-0,5
Facteur Dissymétrie ext Coupons	-0,01	0	0
Facteur intensité Coupons	1	1	1
Type de loi Nominal	Normale	Uniforme	Gaussienne
Paramètre loi 1 Nominal	0	-0,6	0
Paramètre loi 2 Nominal	1	0,2	1
Facteur Dissymétrie int Nominal	-0,5	0	-0,5
Facteur Dissymétrie ext Nominal	-0,01	0	0
Facteur intensité nominal	1,2	1	1
Type de loi Maturité	Normale	Uniforme	Normale
Paramètre loi 1 Maturité	0	-0,6	0
Paramètre loi 2 Maturité	1	0,2	1
Facteur Dissymétrie int Maturité	-0,5	0	-0,5
Facteur Dissymétrie ext Maturité	0	0	0
Facteur intensité maturité	10	1	10
Nombre de portefeuilles	10 000	10 000	10 000
Borne supérieur Coeff RN	1,02	0,92	1,03
Borne inférieur Coeff RN	0,90	0,67	0,92

6.9 Mémoire de fin d'étude d'Aurélien Suray, ESILV promo 2015-2020

Cette annexe présente les principaux résultats issus du mémoire d'Aurélien SURAY utilisés à titre comparatif dans la partie d'analyse de l'étude.

6.9.1 Schéma sur la sélection des variables explicatives



6.9.2 Schéma sur les pouvoirs prédictifs des algorithmes

	Random Forest	XGboost	GBR
Temps d'apprentissage	15 minutes	6 minutes	8 minutes
Temps de prédiction sur la base de test	1 seconde	1 seconde	1 seconde
-A la moyenne	-0,06%	0,02%	-0,12%
-Quantile à 90%	[-7,3%-7,2%]	[-4,8%-3,4%]	[-7,0%-6,5%]
Temps de prédiction sur un nouveau scénario	1 minute 30	1 minute 30	1 minute 30
-Prédiction moyenne	-0,2%	-0,20%	-0,04%
-Quantile à 90%	[-6,9%-5,7%]	[-2,8%-1,8%]	[-6,2%-5,9%]

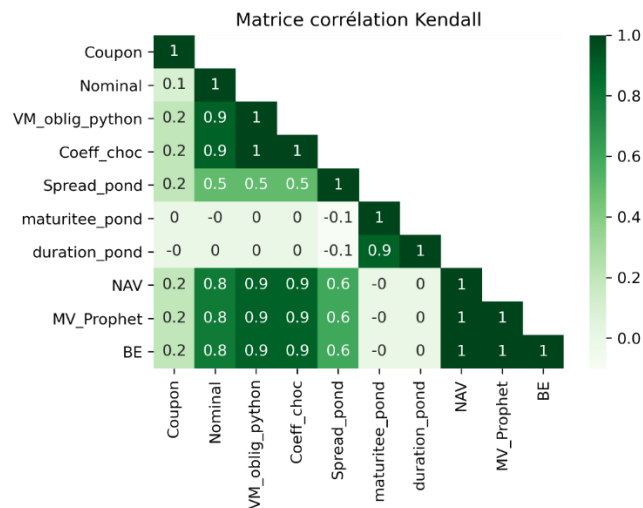
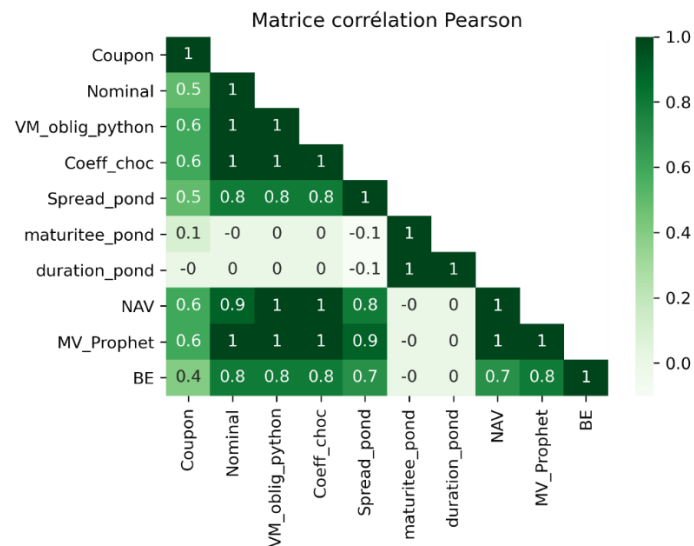
6.10 Annexes Base de données chocs obligataires avec un pool de TMG

Les annexes qui suivent présentent des graphiques et tableaux supplémentaires sur les résultats obtenus lors des prédictions de ML sur les différentes BDD utilisées.

6.10.1 Tableau des temps d'apprentissage des algorithmes

Algorithme	temps_apprentissage
Regression_linéaire_const	0,01
Regression_linéaire_min	0,01
grad_boost	0,07
SVM	0,17
XGboost	0,37
rd_forest	1,29

6.10.2 Les matrices de corrélations



6.10.3 Le dataset test

Ecart BE par percentile :

Percentile	Test_random_forest	Test_grad_boost	Test_SVM	Test_XGboost	Test_regress
9%	0,1%	0,1%	0,3%	0,1%	0,7%
19%	> 0,1%	0,1%	0,2%	> 0,1%	0,8%
29%	> 0,1%	> 0,1%	0,2%	> 0,1%	0,7%
39%	> 0,1%	> 0,1%	0,1%	> 0,1%	0,5%
49%	> 0,1%	> 0,1%	> 0,1%	> 0,1%	0,3%
59%	> 0,1%	> 0,1%	> 0,1%	> 0,1%	0,1%
69%	> 0,1%	> 0,1%	> 0,1%	-0,1%	-0,1%
79%	-0,1%	> 0,1%	> 0,1%	> 0,1%	-0,4%
89%	> 0,1%	> 0,1%	> 0,1%	> 0,1%	-0,8%
99%	-0,1%	-0,2%	-0,3%	-0,1%	-1,5%

6.10.4 Le dataset français

Graphique de la répartition des BE par percentile

Courbe des BE du dataset français prédits par percentiles

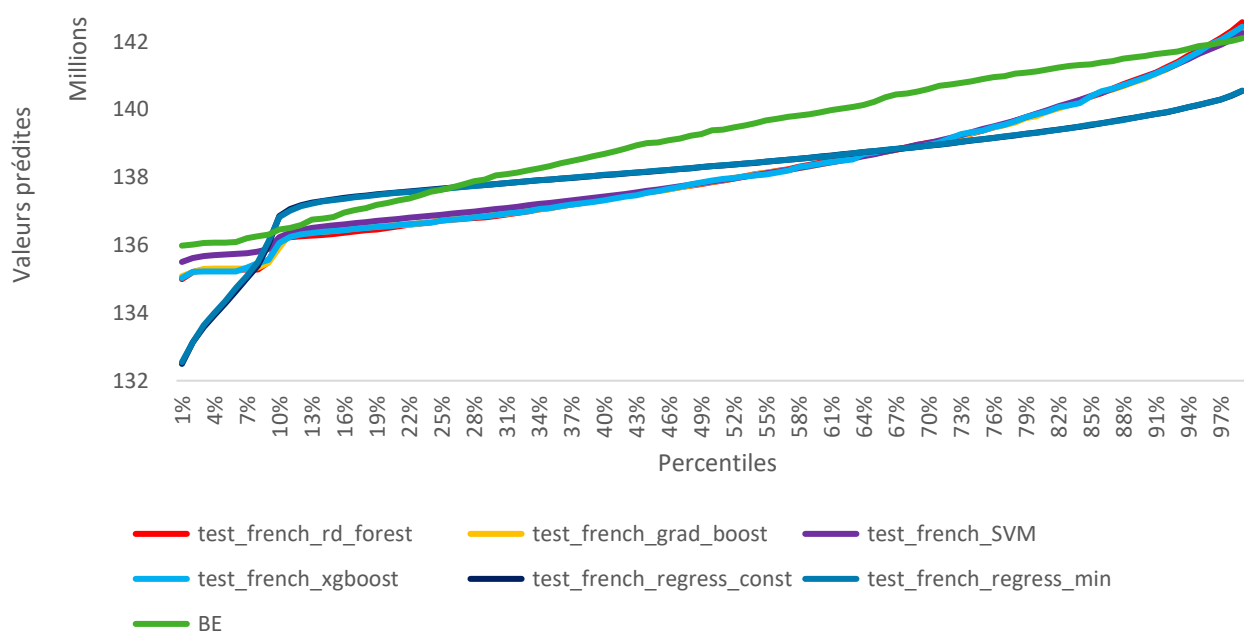


Tableau des écarts relatifs par percentiles

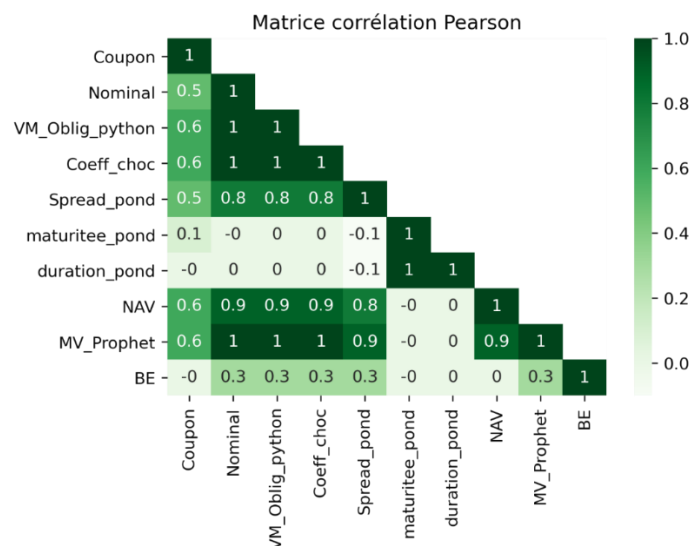
Percentile	test_français_rd_forest	test_français_grad_boost	test_français_SVM	test_français_XGboost	test_français_regress_const	test_français_regress_min
9%	-0,6%	-0,6%	-0,3%	-0,5%	-0,2%	-0,2%
19%	-0,5%	-0,5%	-0,3%	-0,5%	0,2%	0,2%
29%	-0,8%	-0,8%	-0,7%	-0,8%	-0,1%	-0,1%
39%	-1,0%	-1,0%	-0,9%	-1,0%	-0,4%	-0,4%
49%	-1,1%	-1,0%	-1,0%	-1,0%	-0,7%	-0,7%
59%	-1,1%	-1,1%	-1,1%	-1,1%	-0,9%	-0,9%
69%	-1,1%	-1,2%	-1,1%	-1,2%	-1,2%	-1,2%
79%	-0,9%	-0,9%	-0,9%	-0,9%	-1,3%	-1,3%
89%	-0,5%	-0,5%	-0,5%	-0,5%	-1,3%	-1,3%
99%	0,3%	0,2%	0,1%	0,2%	-1,1%	-1,1%

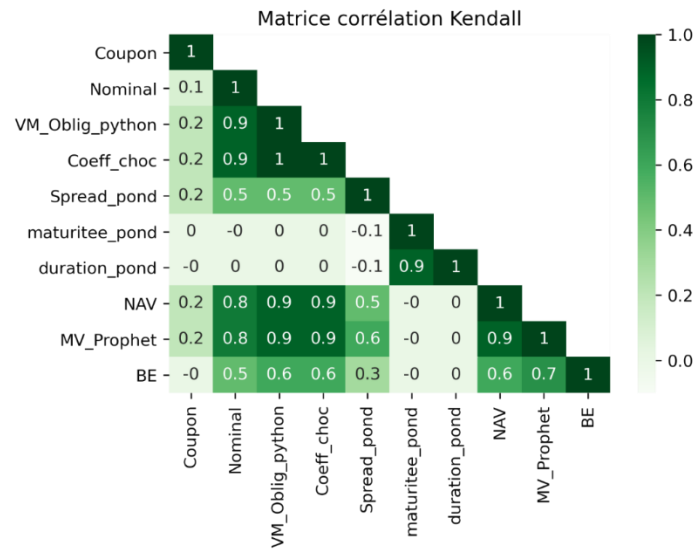
6.11 Annexes Base de données chocs obligataires avec 3 pools de TMG

6.11.1 Tableau des temps d'apprentissage des algorithmes

Algorithme	Temps de test en secondes
regress_test	>0,01
XGboost_test	0,01
Grad_Boost_test	0,01
random_forest_test	0,01
SVM_test	0,30

6.11.2 Les matrices de corrélations



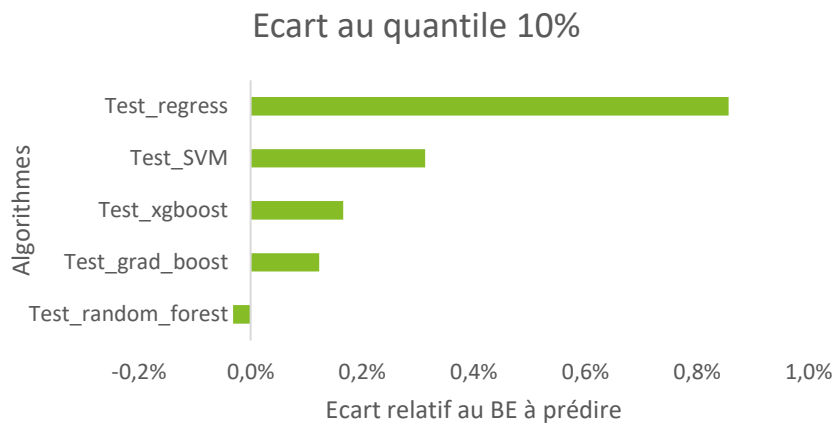


6.11.3 Le dataset test

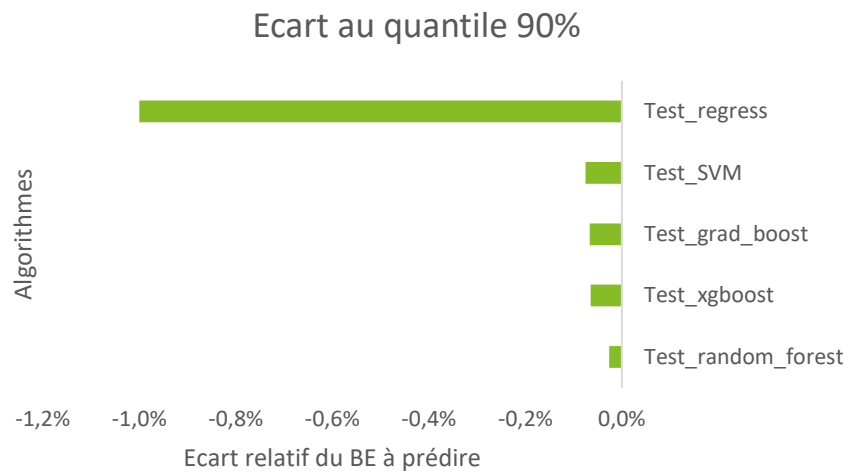
Tableau des écarts relatifs au BE

Ecarts relatifs	Test_random_forest	Test_grad_boost	Test_SVM	Test_XGboost	Test_regress
9%	-0,03%	0,12%	0,31%	0,17%	0,86%
19%	> 0,01%	0,12%	0,20%	0,11%	0,84%
29%	-0,01%	-0,08%	0,04%	-0,05%	0,60%
39%	0,00%	-0,09%	-0,01%	-0,12%	0,33%
49%	0,01%	-0,12%	-0,04%	-0,11%	0,01%
59%	-0,01%	-0,25%	-0,25%	-0,30%	-0,43%
69%	> 0,01%	-0,01%	0,06%	0,01%	-0,29%
79%	-0,02%	-0,01%	0,06%	-0,05%	-0,51%
89%	-0,03%	-0,07%	-0,08%	-0,06%	-1,00%
99%	0,01%	-0,14%	-0,10%	-0,16%	-1,74%

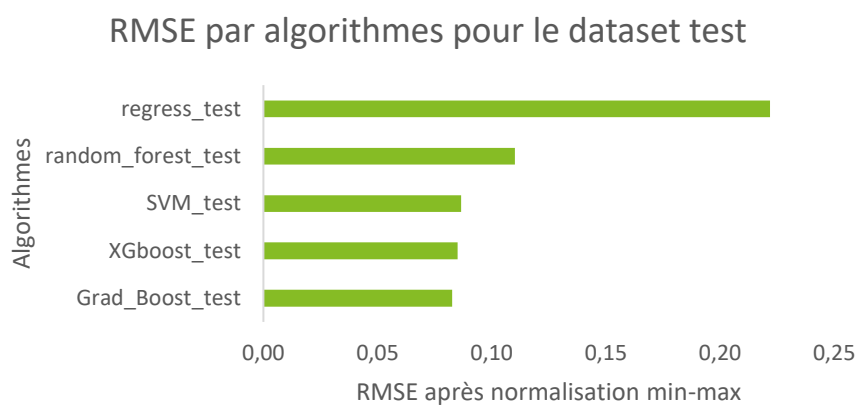
Graphique des écarts relatifs au quantile 10%



Graphique des écarts relatifs au quantile 90%



Graphique RMSE



6.11.4 Le dataset français

Graphique de l'évolution des BE par percentiles

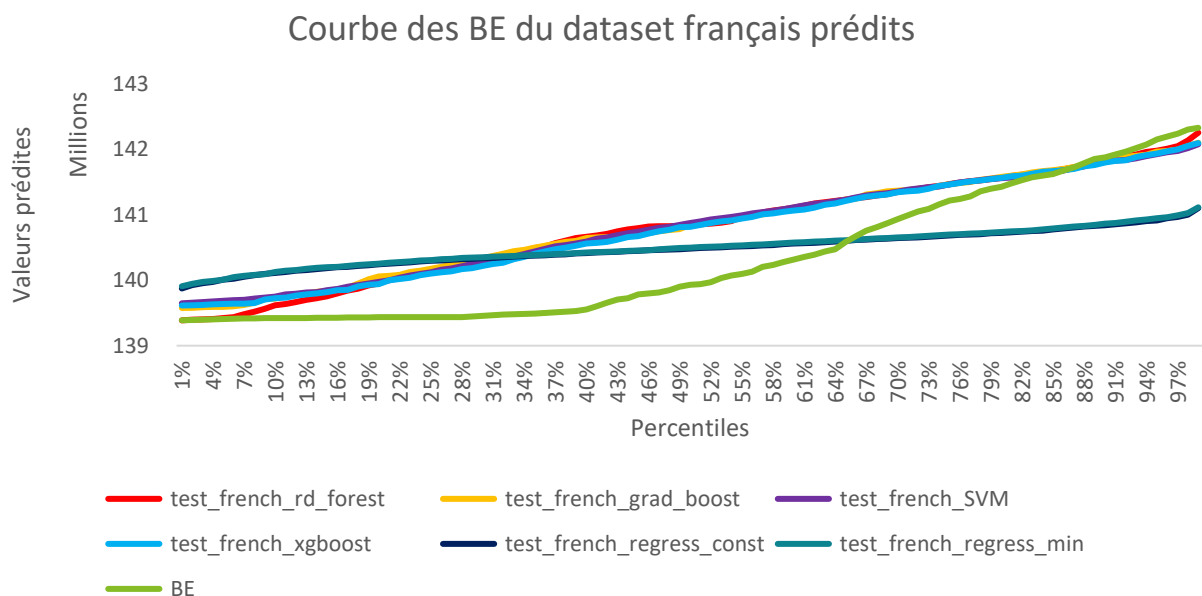
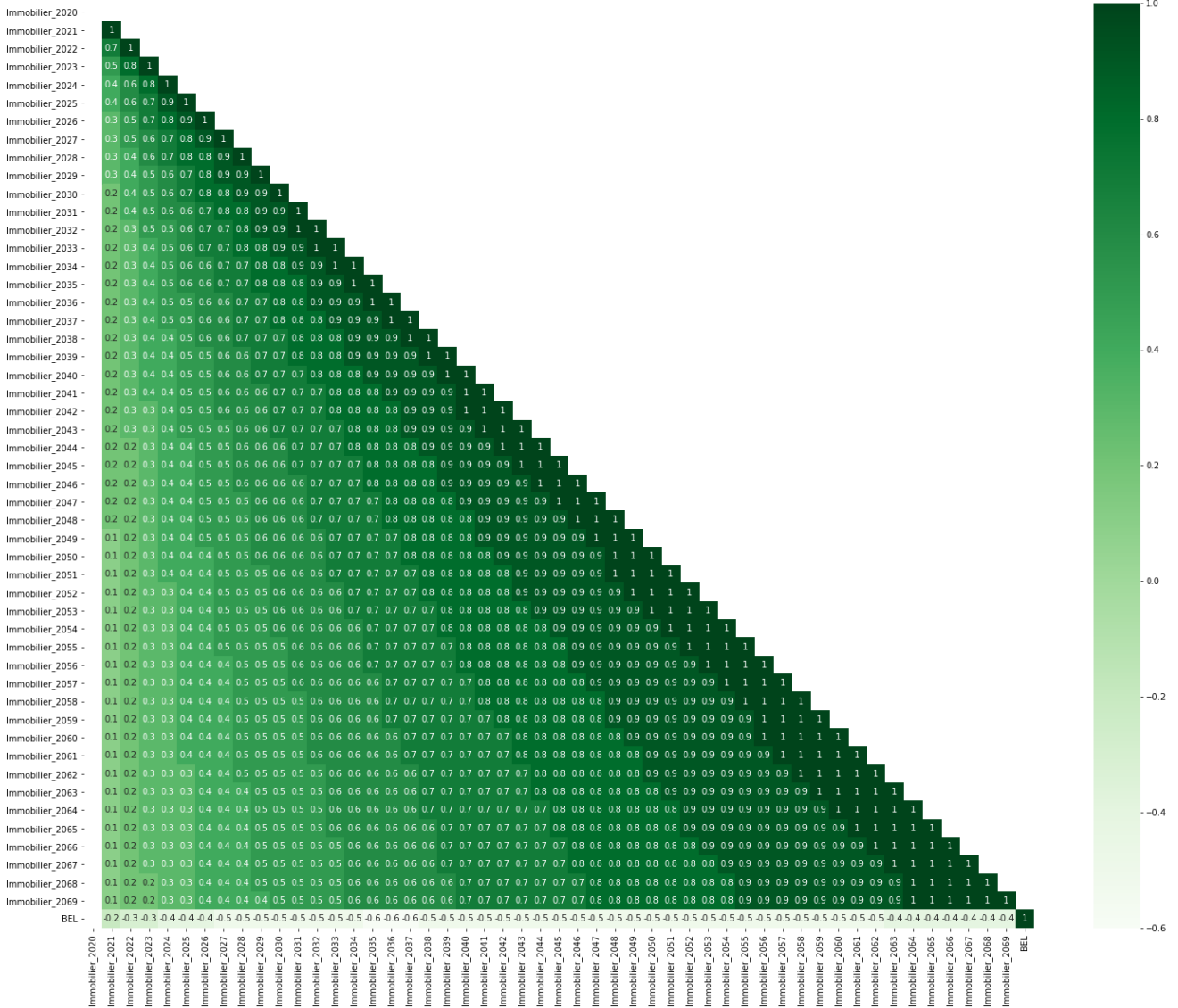


Tableau des écarts relatif des BE prédits

Ecart relatif ¹⁴⁶	test_français _rd_forest	test_français _grad_boost	test_français _SVM	test_français _XGboost	test_français _regress_const	test_français _regress_min
9%	0,10%	0,21%	0,22%	0,20%	0,48%	0,48%
19%	0,35%	0,41%	0,37%	0,36%	0,57%	0,58%
29%	0,59%	0,60%	0,57%	0,53%	0,63%	0,64%
39%	0,80%	0,77%	0,74%	0,71%	0,63%	0,63%
49%	0,66%	0,63%	0,68%	0,65%	0,41%	0,42%
59%	0,57%	0,56%	0,57%	0,55%	0,19%	0,20%
69%	0,32%	0,35%	0,33%	0,31%	-0,17%	-0,16%
79%	0,10%	0,11%	0,11%	0,10%	-0,48%	-0,47%
89%	-0,04%	-0,06%	-0,05%	-0,07%	-0,72%	-0,71%
99%	-0,05%	-0,16%	-0,18%	-0,16%	-0,86%	-0,85%

¹⁴⁶ Les BE sont exprimés en M€

Matrice corrélation Kendall



Les taux immobiliers des années 2056 à 2069, 2047 à 2055, 2040 à 2046 ainsi que de 2034 à 2039 sont très fortement corrélés linéairement.

L'analyse des corrélations non linéaires fait apparaître les mêmes tendances que l'analyse des corrélations linéaires.

Les mesures suivantes vont être appliquées :

- Suppression des taux 2020
- Création d'un input moyen 2056-2069 et suppression des taux 2057-2069
- Création d'un input moyen 2047-2055 et suppression des taux 2048-2055
- Création d'un input moyen 2040-2046 et suppression des taux 2041-2046
- Création d'un input moyen 2034-2039 et suppression des taux 2035-2039

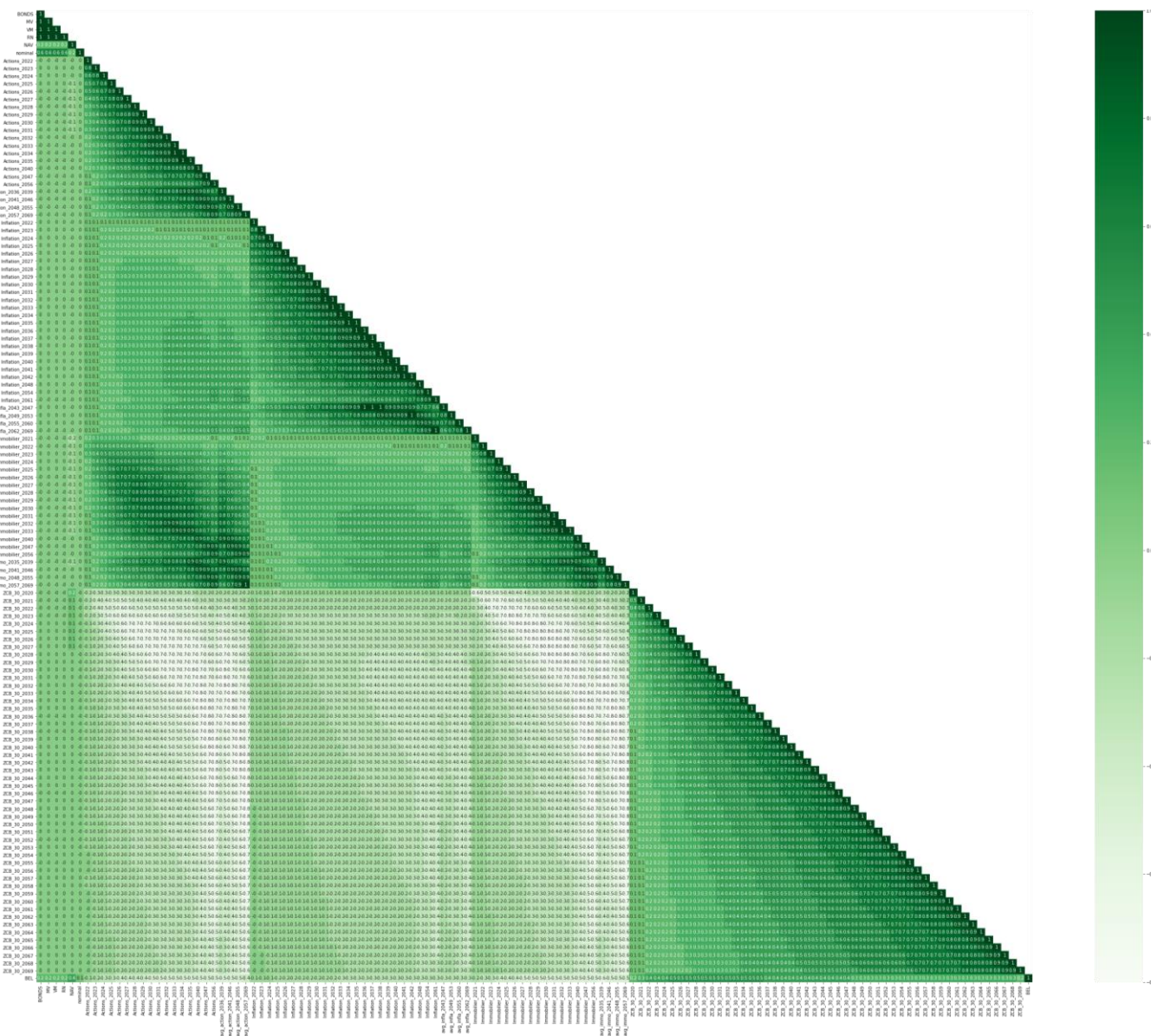
L'analyse des corrélations linéaires est sans équivoque, toutes les ZCB sont quasi intégralement corrélées positivement, les corrélations non linéaires suivent cette même tendance.

Pour la base d'apprentissage des algorithmes va être choisi d'implémenter la variable ZCB_30 sur tous les pas de temps.

Les corrélations de rang ont été plus atténuées que les corrélations linéaires.

Cette atténuation peut s'expliquer que les variables ont toutes été normalisées par une normalisation men-max, mettant sur la "même dimension" toutes les variables explicatives et à prédire.

6.12.6 Analyse des corrélations sur la base finale



La matrice des corrélations linéaires entre variables indique par sa couleur (la matrice étant trop grande pour être visualisée avec précision sur un pdf) que les principaux groupes de variables (Obligataires, immobiliers, inflation et ZCB) sont relativement décorrélés entre eux (voir couleur bas du triangle).

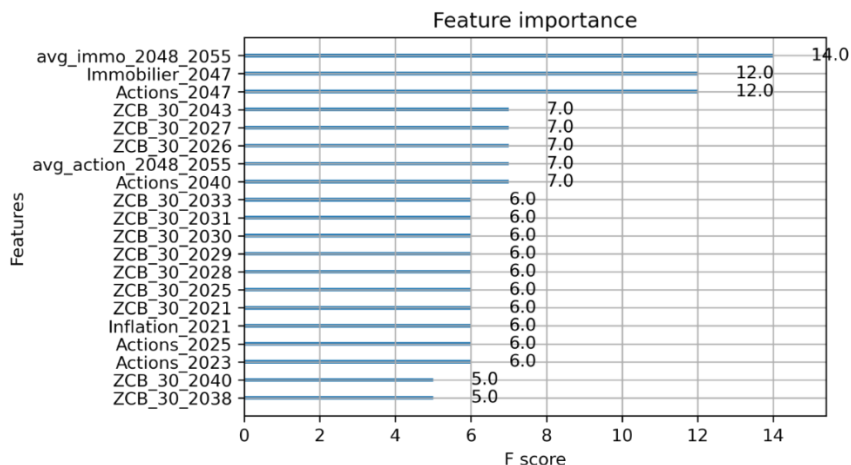
A noter une forte corrélation entre certains coefficients obligataires et les ZCB entre eux par choix de vouloir garder une diversité de ces variables qui sont fortement corrélées au BEL.

L'analyse des corrélations de Kendall suit les mêmes tendances.

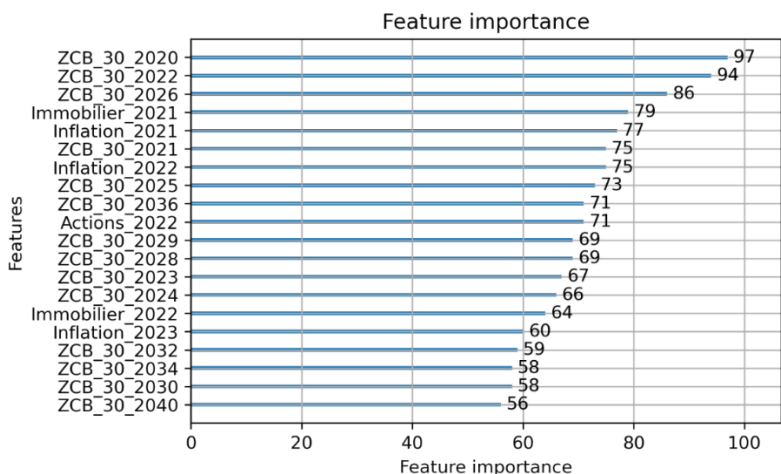
6.12.7 Graphique d'importance des variables par algorithmes pour la BDD GSE variant et chocs obligataires

Importance des variables par algorithme (hors régression linéaire)

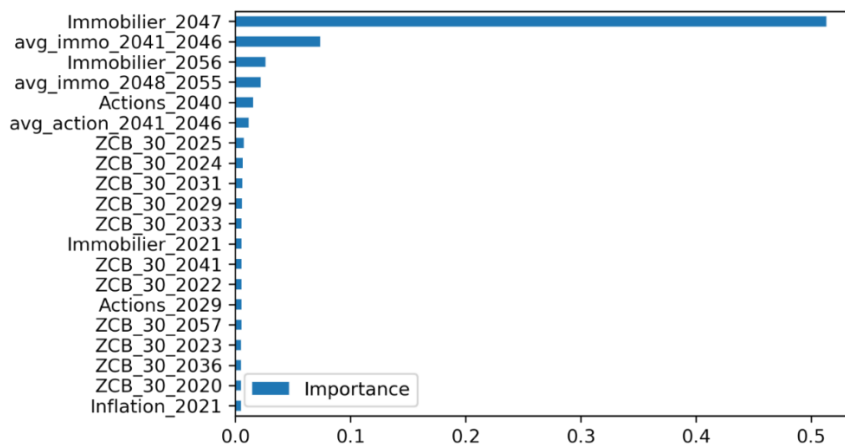
Le XGboost



Le Gradient Boosting



Le Random Forest

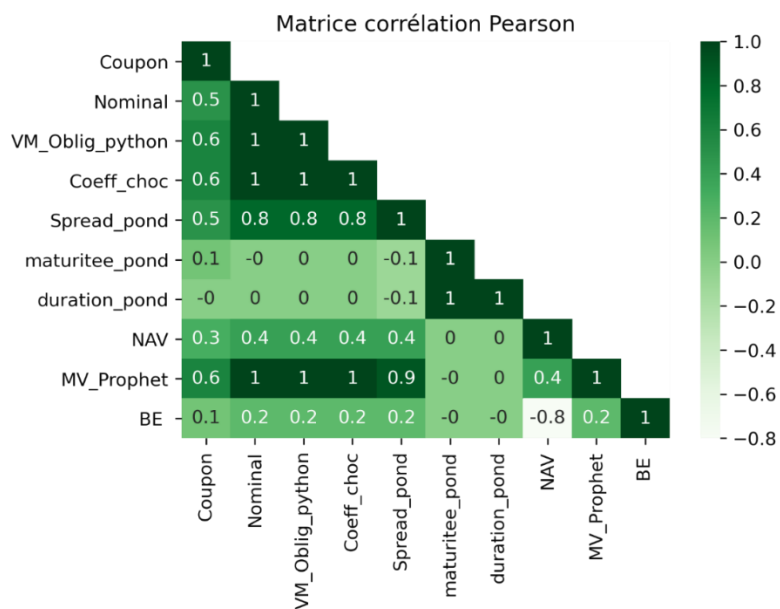


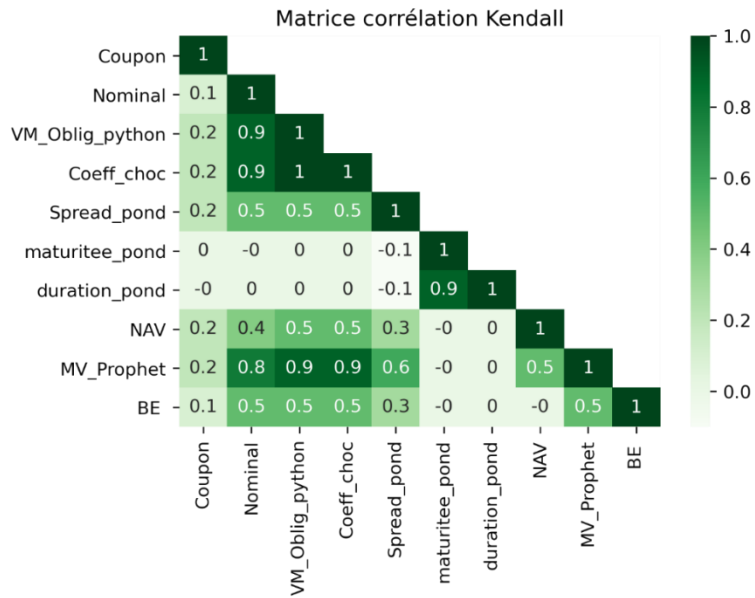
6.13 Annexe Base de données chocs obligataires et GSE variant avec un pool de TMG

6.13.1 Tableau des temps d'apprentissage

Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,14
Regression_linéaire_min	0,19
grad_boost	0,94
XGboost	4,14
rd_forest	19,83

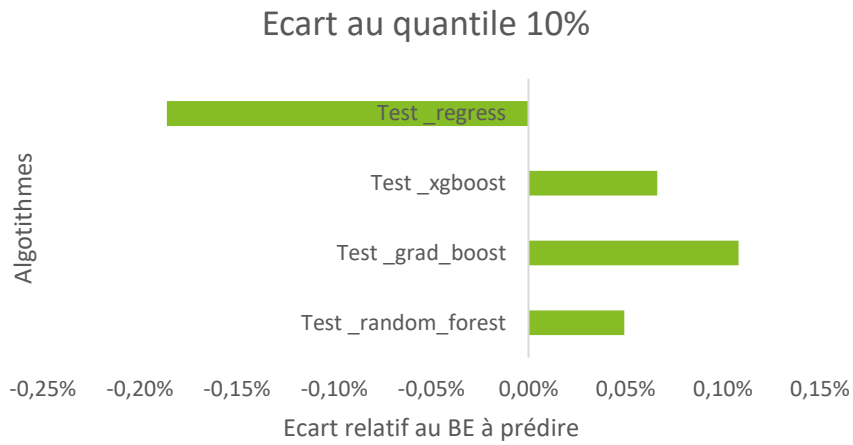
6.13.2 Les matrices de corrélation



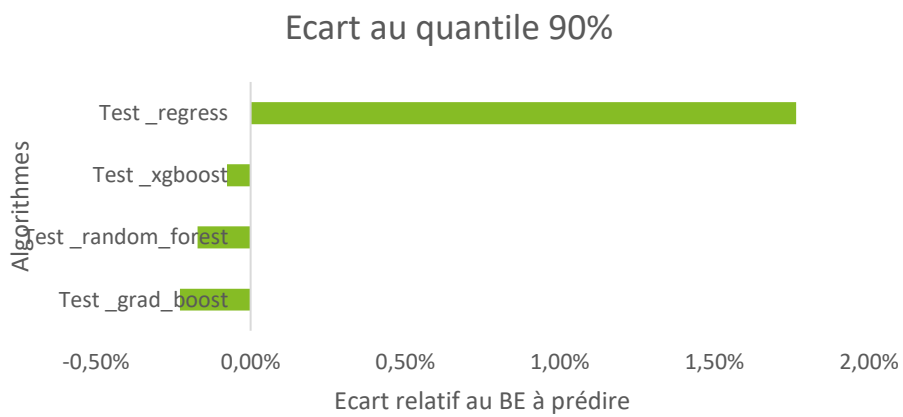


6.13.3 Le dataset test

Graphique des écarts relatifs au quantile 10%

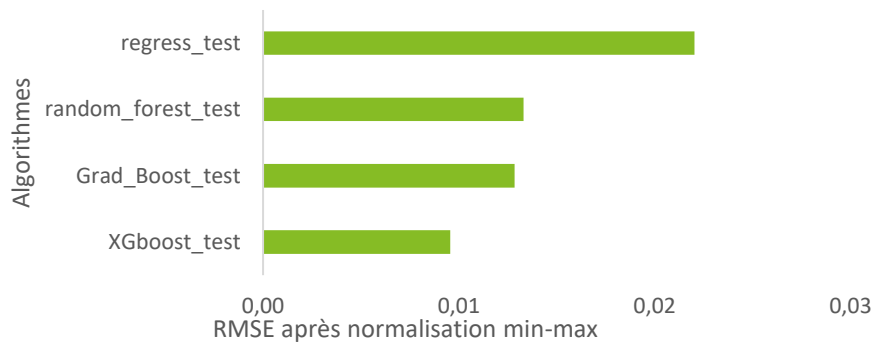


Graphique des écarts relatifs au quantile 90%



Graphique RMSE

RMSE par algorithme pour le dataset test



6.13.4 Le dataset français

Graphique de l'évolution des BE par percentiles

Courbe des BE du dataset français prédits par percentiles

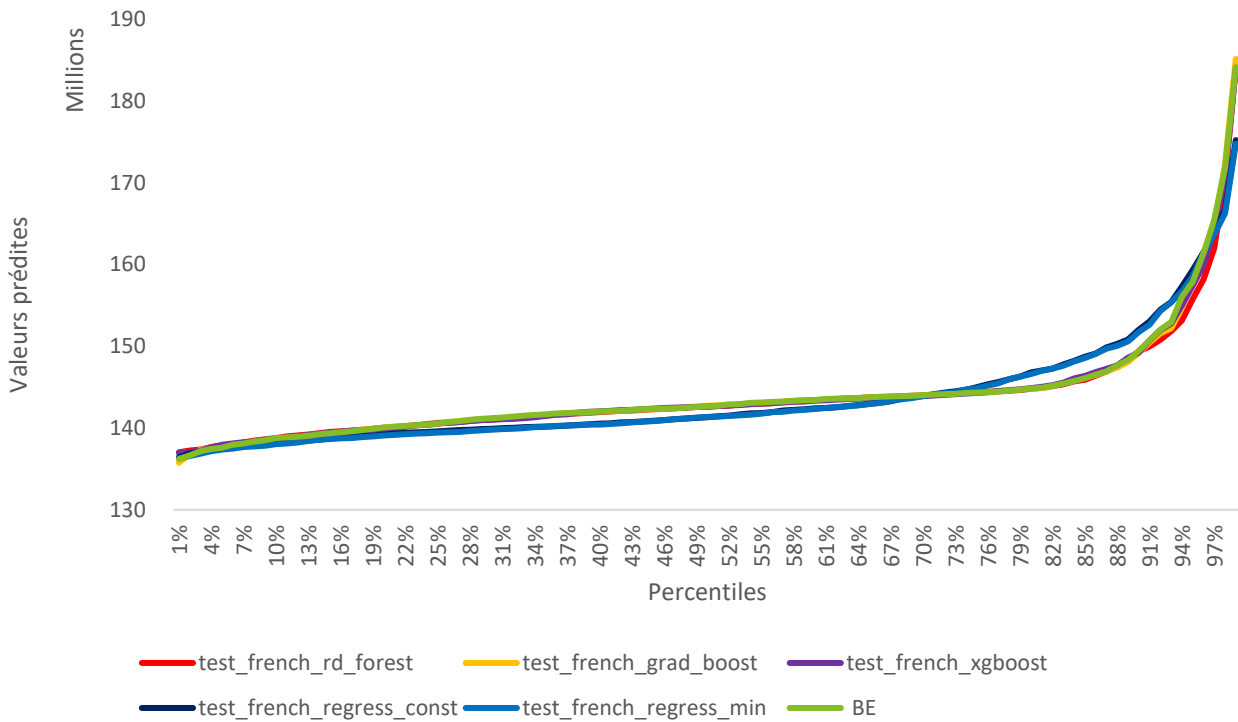


Tableau des écarts relatif des BE prédits

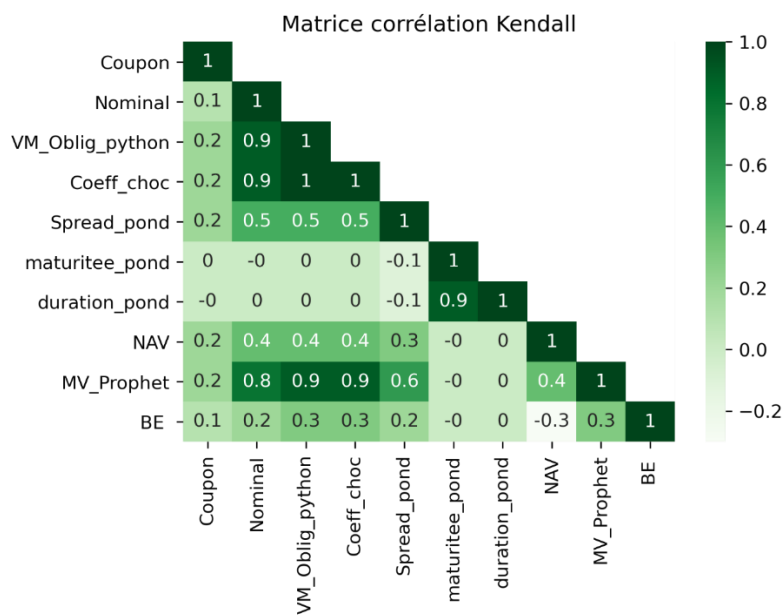
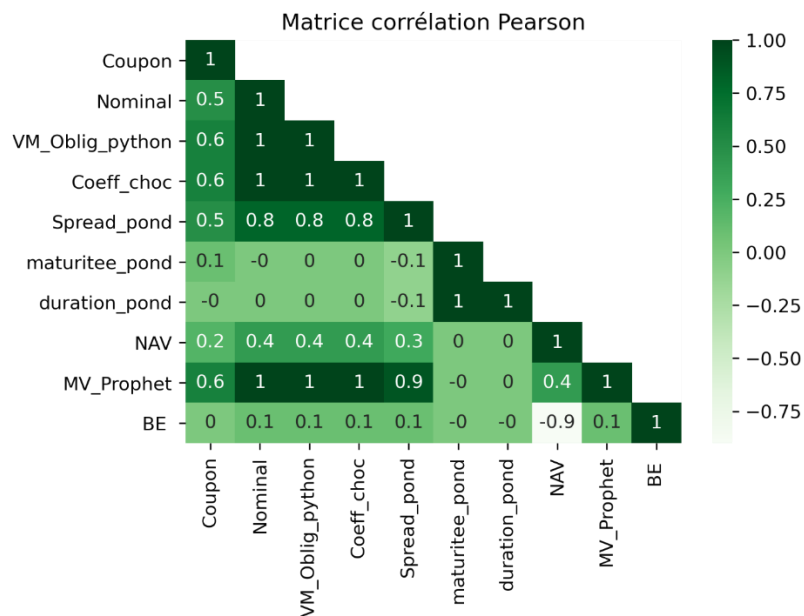
Ecart relatif	test_français_rd_forest	test_français_grad_boost	test_français_XGboost	test_français_regress_const	test_français_regress_min
9%	0,06%	0,03%	0,03%	-0,36%	-0,49%
19%	0,01%	-0,02%	-0,08%	-0,59%	-0,68%
29%	-0,10%	-0,04%	-0,12%	-0,87%	-0,98%
39%	-0,05%	-0,02%	0,02%	-1,06%	-1,09%
49%	> 0,01%	0,06%	0,02%	-0,93%	-0,91%
59%	-0,04%	-0,07%	-0,10%	-0,76%	-0,81%
69%	> 0,01%	-0,02%	-0,06%	-0,12%	-0,20%
79%	-0,02%	0,02%	0,03%	1,11%	1,10%
89%	0,02%	-0,15%	0,20%	1,70%	1,54%
99%	0,01%	0,54%	-0,09%	-4,85%	-5,10%

6.14 Annexe base de données chocs obligataires, GSE variant et trois pools de TMG

6.14.1 Tableau des temps d'apprentissage

Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,17
Regression_linéaire_min	0,22
grad_boost	0,94
XGboost	1,42
rd_forest	9,51

6.14.2 Les matrices de corrélations

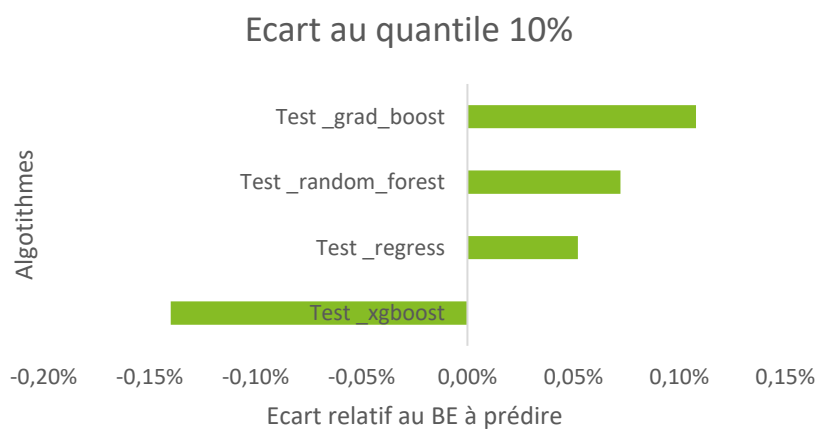


6.14.3 Le dataset test

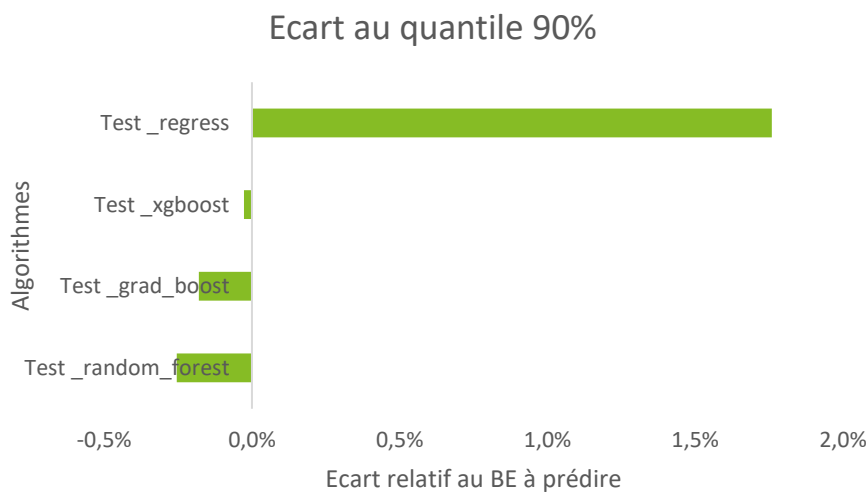
Tableau des écarts relatifs au BE

Ecart relatif	Test _random_forest	Test _grad_boost	Test _XGboost	Test _regress
9%	0,02%	0,02%	0,29%	-0,34%
19%	-0,03%	-0,07%	0,05%	-0,23%
29%	0,02%	0,06%	-0,07%	-0,39%
39%	-0,03%	-0,07%	-0,19%	-0,36%
49%	0,01%	0,05%	-0,28%	-0,32%
59%	-0,02%	-0,07%	-0,29%	-0,36%
69%	-0,03%	-0,02%	-0,22%	0,00%
79%	-0,03%	0,03%	-0,27%	0,87%
89%	-0,25%	-0,18%	-0,03%	1,76%
99%	0,45%	0,55%	0,96%	-1,05%

Graphique des écarts relatifs au quantile 10%

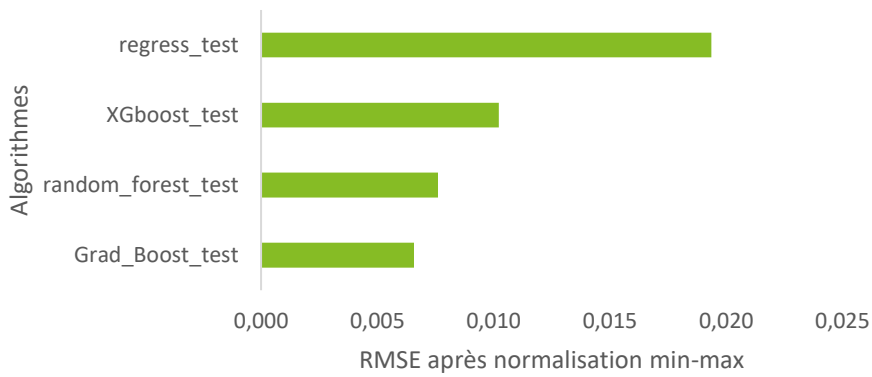


Graphique des écarts relatifs au quantile 90%



Graphique RMSE

RMSE par algorithme pour le dataset test



6.14.4 Le dataset français

Graphique de l'évolution des BE par percentiles

Courbe des BE du dataset français prédits par percentiles

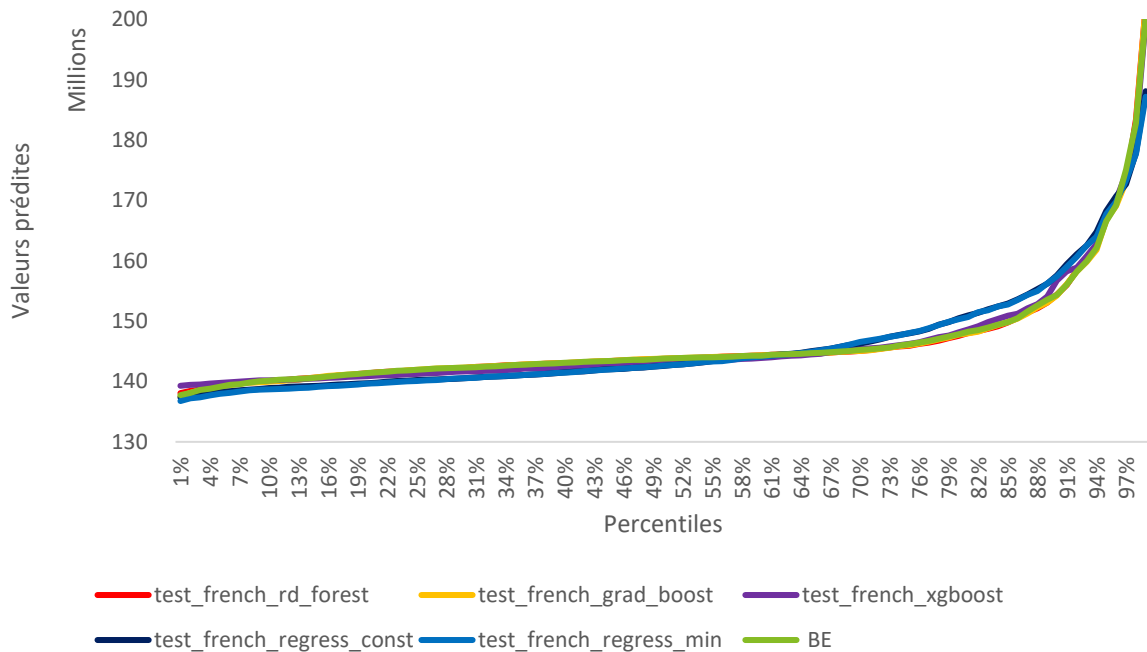


Tableau des écarts relatif des BE prédits

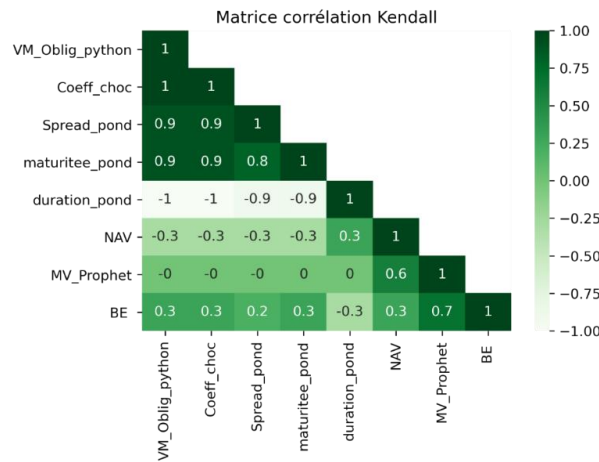
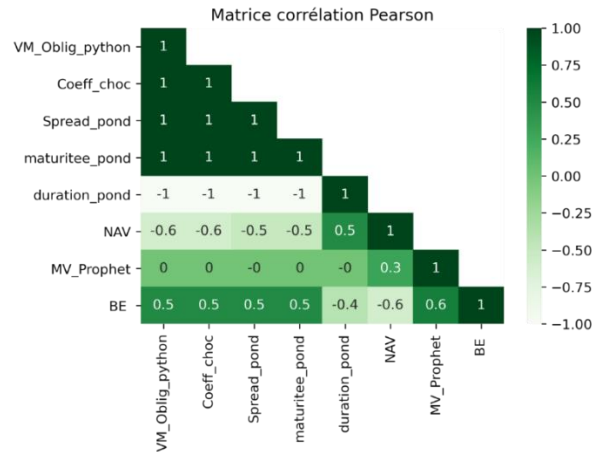
Ecart relatif	test_français_rd_forest	test_français_grad_boost	test_français_XGboost	test_français_regress_const	test_français_regress_min
9%	-0,06%	-0,10%	0,10%	-0,89%	-1,00%
19%	> 0,01%	-0,04%	-0,28%	-1,13%	-1,25%
29%	-0,04%	> 0,01%	-0,46%	-1,26%	-1,24%
39%	-0,02%	> 0,01%	-0,56%	-1,17%	-1,12%
49%	-0,02%	0,04%	-0,56%	-0,85%	-0,83%
59%	-0,01%	-0,01%	-0,32%	-0,13%	-0,18%
69%	-0,09%	-0,06%	0,07%	0,67%	0,72%
79%	-0,20%	-0,12%	0,09%	1,59%	1,63%
89%	-0,33%	-0,33%	0,34%	1,73%	1,77%
99%	1,04%	1,07%	-0,11%	-6,03%	-6,49%

6.15 Annexe Base de données GSE variant avec un pool de TMG

6.15.1 Tableau des temps d'apprentissage

Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,15
Regression_linéaire_min	0,25
grad_boost	0,86
XGboost	2,97
rd_forest	20,26

6.15.2 Les matrices de corrélations



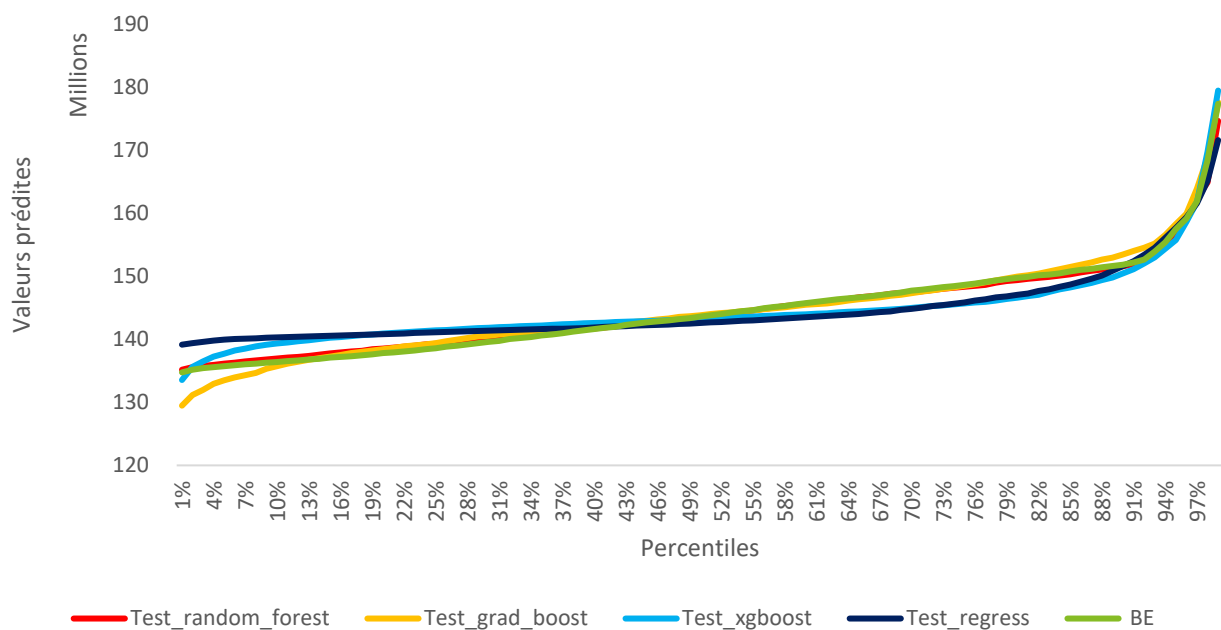
6.15.3 Le dataset test

Les prédictions pour la base de données test

Modèle ¹⁴⁷		BEL	BEL	Ecart
		Modèle	prédit	
Régression avec constante	Min	133	138	-3,2%
	Moy	145	145	> 0,1%
	Med	144	143	0,7%
	Max	245	245	8,9%
Random Forest	Min	133	135	-1,0%
	Moy	145	145	> 0,1%
	Med	144	143	0,2%
	Max	245	257	-4,8%
Gradient Boosting	Min	133	123	7,6%
	Moy	145	145	> 0,1%
	Med	144	144	-0,1%
	Max	245	224	8,5%
XGboost	Min	133	130	2,4%
	Moy	145	145	> 0,1%
	Med	144	143	0,3%
	Max	245	245	-0,2%

Graphique des BE par percentile

Courbe des BE du dataset test prédits par percentiles

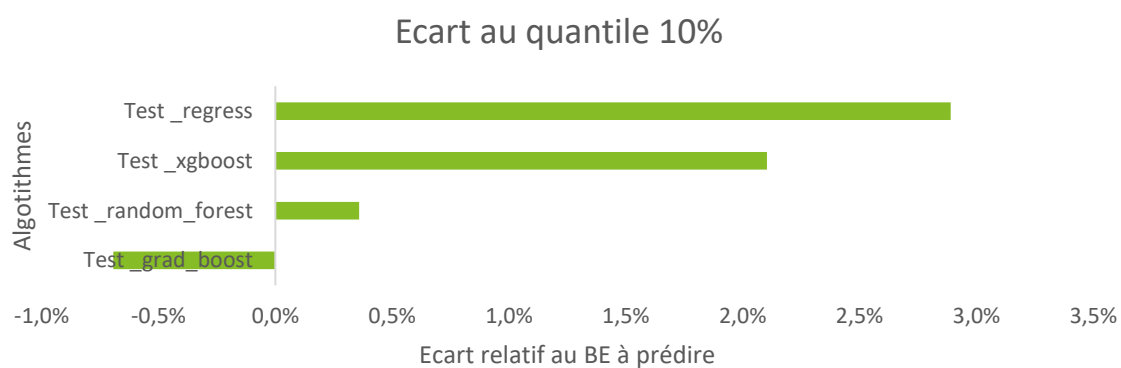


¹⁴⁷ Les BE sont exprimés en M€

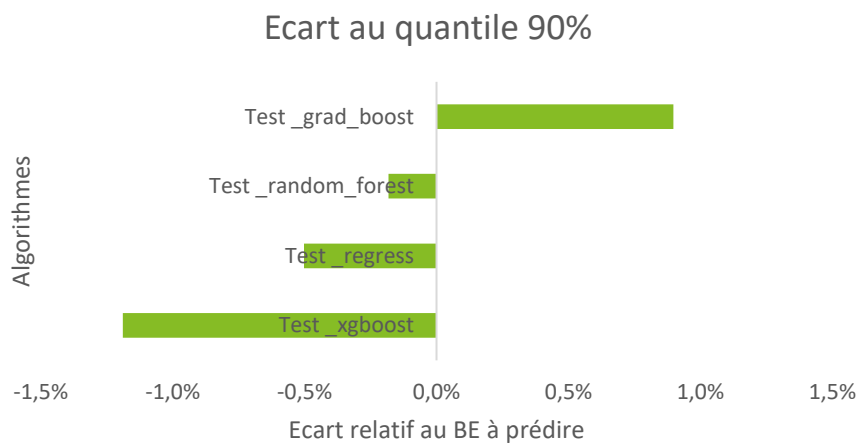
Tableau des écarts relatifs au BE

Ecart relative Base de données GSE variant	Test_random_forest	Test_grad_boost	Test_XGboost	Test_regress
9%	0,36%	-0,69%	2,10%	2,89%
19%	0,57%	0,43%	2,29%	2,28%
29%	0,41%	0,77%	1,72%	1,40%
39%	0,09%	0,32%	0,79%	0,26%
49%	-0,11%	0,20%	-0,12%	-0,60%
59%	-0,08%	-0,19%	-1,11%	-1,46%
69%	-0,03%	-0,23%	-1,74%	-1,84%
79%	-0,24%	0,06%	-2,12%	-1,84%
89%	-0,18%	0,90%	-1,19%	-0,50%
99%	-1,53%	0,14%	1,18%	-3,25%

Graphique des écarts relatifs au quantile 10%

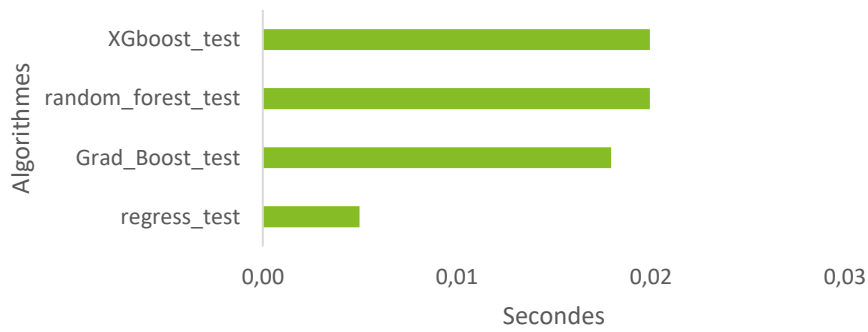


Graphique des écarts relatifs au quantile 90%



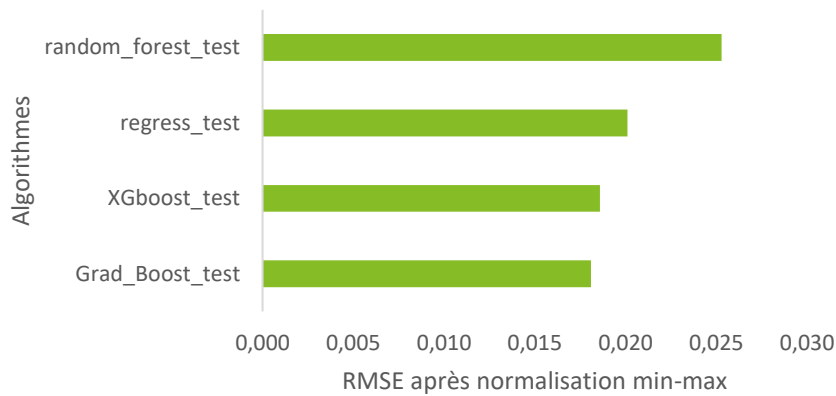
Graphique RMSE

Temps de test par algorithme



Graphique RMSE

RMSE par algorithme pour le dataset test



6.15.4 Le dataset D20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la hausse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

Courbe des BE du dataset D20 prédits par percentiles

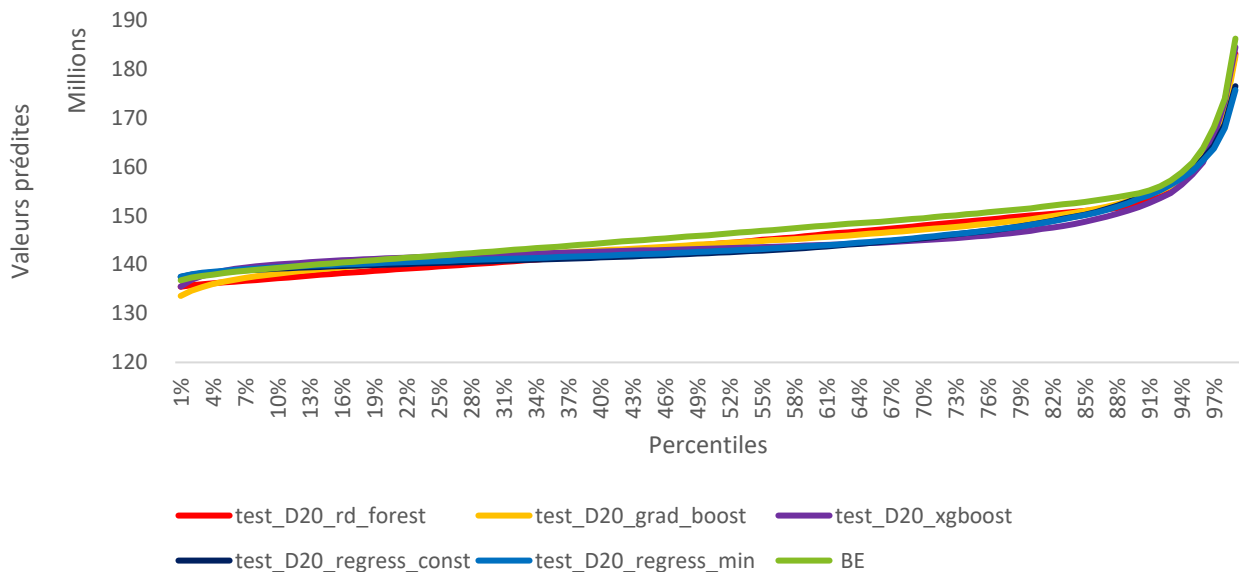


Tableau des écarts relatif des BE prédits

Ecart relatif	test_D20 _rd_forest	test_D20 _grad_boost	test_D20 _XGboost	test_D20 _regress_const	test_D20 _regress_min
9%	-1,50%	-0,90%	0,51%	-0,04%	0,15%
19%	-1,55%	-0,65%	0,19%	-0,70%	-0,45%
29%	-1,60%	-0,74%	-0,38%	-1,31%	-1,05%
39%	-1,56%	-0,97%	-1,12%	-1,94%	-1,68%
49%	-1,35%	-1,22%	-1,82%	-2,47%	-2,25%
59%	-1,22%	-1,51%	-2,52%	-2,85%	-2,63%
69%	-1,03%	-1,55%	-2,96%	-2,74%	-2,65%
79%	-0,97%	-1,51%	-3,09%	-2,30%	-2,33%
89%	-1,50%	-0,78%	-2,06%	-0,89%	-1,06%
99%	-1,67%	-1,92%	-0,94%	-5,22%	-5,64%

6.15.5 Le dataset U20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la baisse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

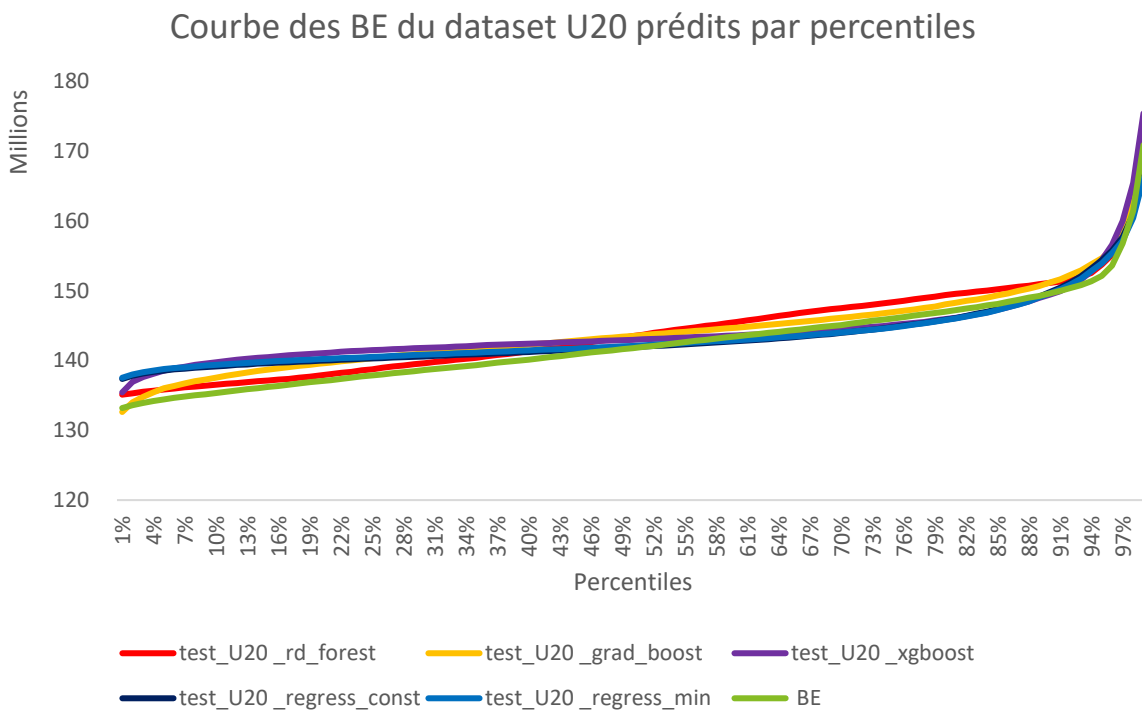


Tableau des écarts relatif des BE prédits

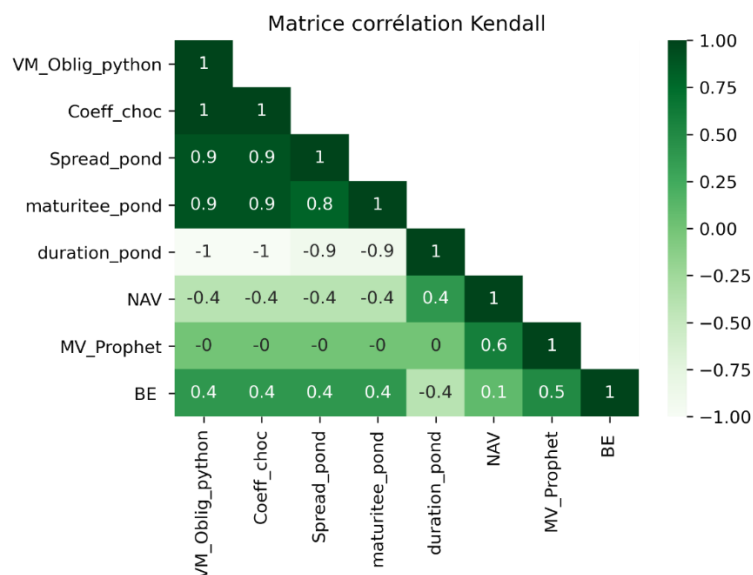
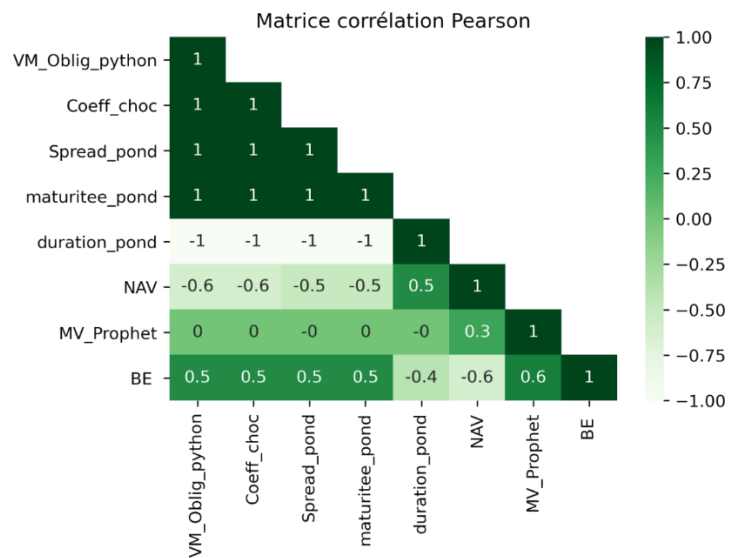
Ecart relatif	test_U20 _rd_forest	test_U20 _grad_boost	test_U20 _XGboost	test_U20 _regress_const	test_U20 _regress_min
9%	-1,50%	-0,90%	0,51%	-0,04%	0,15%
19%	-1,55%	-0,65%	0,19%	-0,70%	-0,45%
29%	-1,60%	-0,74%	-0,38%	-1,31%	-1,05%
39%	-1,56%	-0,97%	-1,12%	-1,94%	-1,68%
49%	-1,35%	-1,22%	-1,82%	-2,47%	-2,25%
59%	-1,22%	-1,51%	-2,52%	-2,85%	-2,63%
69%	-1,03%	-1,55%	-2,96%	-2,74%	-2,65%
79%	-0,97%	-1,51%	-3,09%	-2,30%	-2,33%
89%	-1,50%	-0,78%	-2,06%	-0,89%	-1,06%
99%	-1,67%	-1,92%	-0,94%	-5,22%	-5,64%

6.16 Annexe de la Base de données GSE variant avec 3 pools de TMG

6.16.1 Tableau des temps d'apprentissage

Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,21
Regression_linéaire_min	0,31
grad_boost	2,89
XGboost	12,51
rd_forest	36,38

6.16.2 Les matrices de corrélations



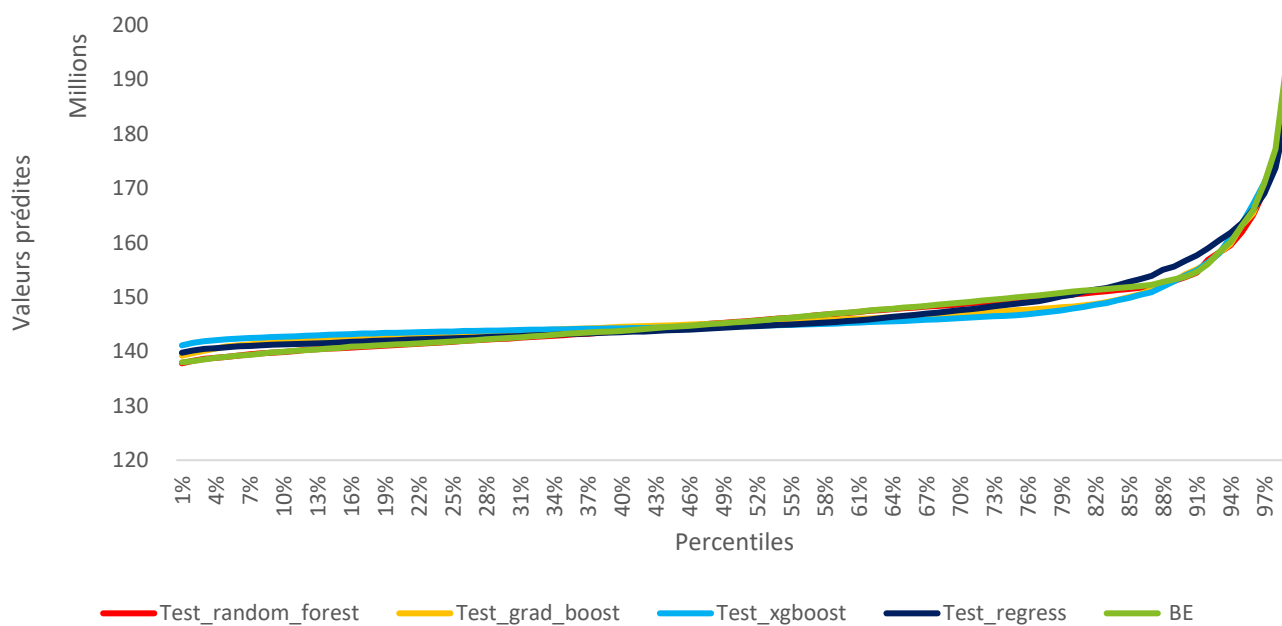
6.16.3 Le dataset test

Les prédictions pour la base de données test

Modèle ¹⁴⁸		BEL Modèle	BEL prédit	Ecart
Régression avec constante	Min	137	138	-0,9%
	Moy	147	147	0,1%
	Med	145	144	0,6%
	Max	261	261	8,9%
Random Forest	Min	137	137	0,2%
	Moy	147	147	0,1%
	Med	145	145	> 0,1%
	Max	261	261	0,3%
Gradient Boosting	Min	137	137	0,3%
	Moy	147	147	0,1%
	Med	145	145	0,1%
	Max	261	243	6,9%
XGboost	Min	137	139	-1,5%
	Moy	147	147	0,0%
	Med	145	145	0,5%
	Max	261	262	-0,2%

Graphique des BE par percentile

Courbe des BE du dataset test prédits par percentiles

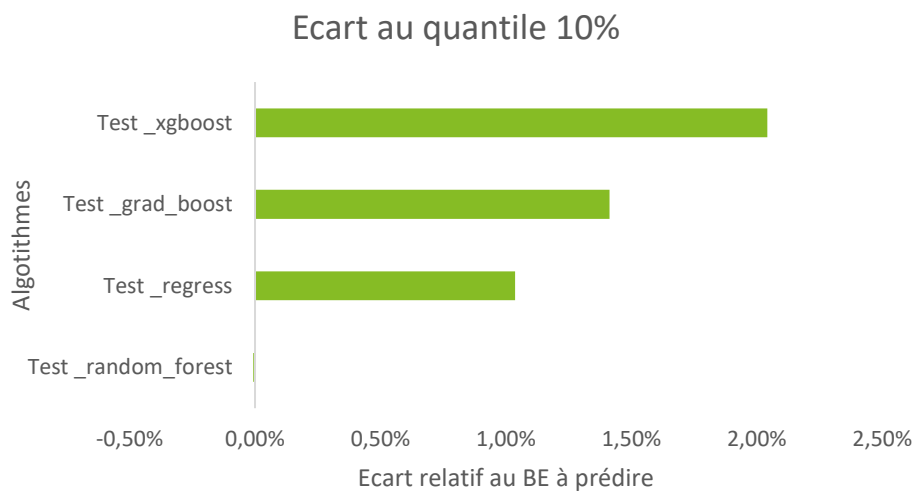


¹⁴⁸ Les BE sont exprimés en M€

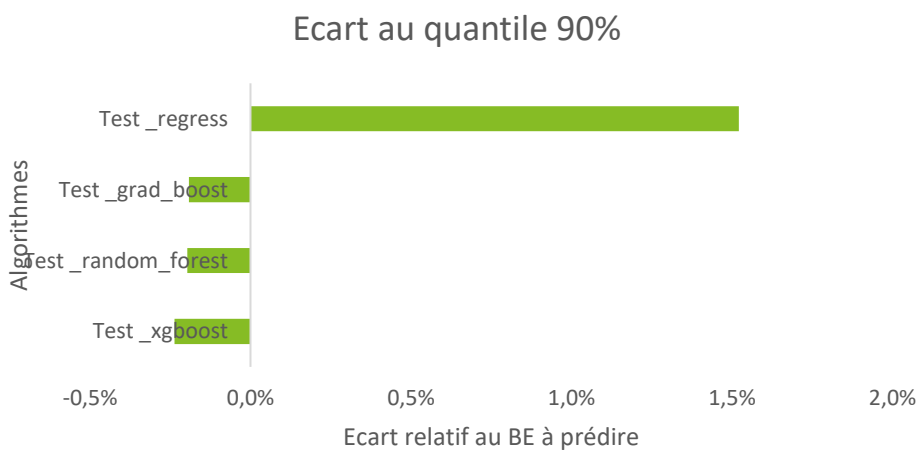
Tableau des écarts relatifs au BE

Ecart relatif	Test _random_forest	Test _grad_boost	Test _XGboost	Test _regress
9%	-0,01%	1,41%	2,04%	1,04%
19%	-0,11%	1,15%	1,55%	0,58%
29%	-0,02%	0,92%	1,06%	0,26%
39%	-0,05%	0,51%	0,38%	-0,14%
49%	0,02%	-0,07%	-0,41%	-0,59%
59%	-0,04%	-0,79%	-1,24%	-1,07%
69%	-0,18%	-1,31%	-1,84%	-0,94%
79%	-0,36%	-1,79%	-2,15%	-0,45%
89%	-0,20%	-0,19%	-0,24%	1,52%
99%	-3,34%	-1,74%	-1,85%	-5,00%

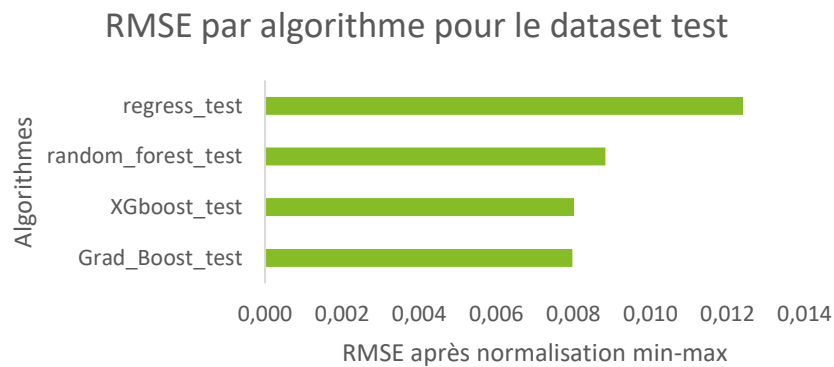
Graphique des écarts relatifs au quantile 10%



Graphique des écarts relatifs au quantile 90%



Graphique RMSE



6.16.4 Le dataset D20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la baisse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

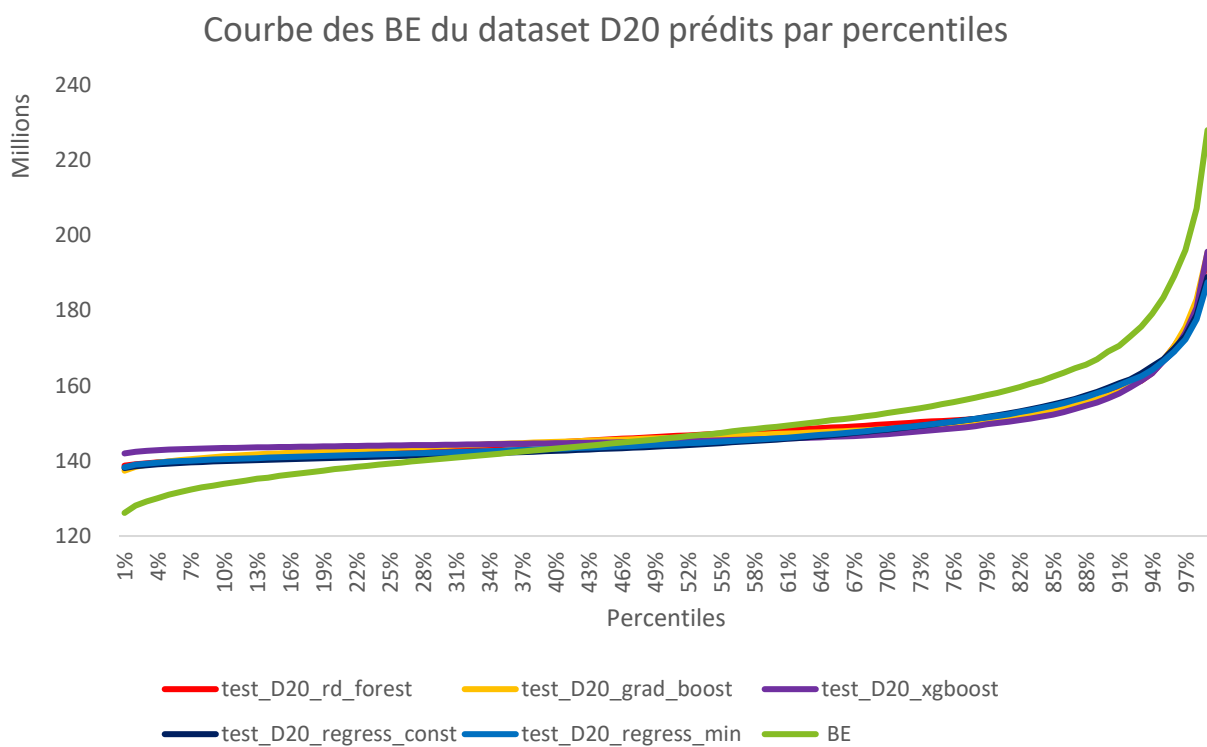


Tableau des écarts relatif des BE prédits

Ecart relatif	test_D20_rd_forest	test_D20_grad_boost	test_D20_XGboost	test_D20_regress_const	test_D20_regress_min
9%	5,28%	5,66%	7,46%	4,81%	5,21%
19%	3,27%	3,83%	4,71%	2,45%	2,85%
29%	2,16%	2,46%	2,75%	0,86%	1,26%
39%	1,19%	1,36%	1,14%	-0,32%	0,06%
49%	0,48%	0,22%	-0,38%	-1,31%	-0,96%
59%	-0,65%	-1,16%	-1,98%	-2,20%	-1,98%
69%	-1,74%	-2,56%	-3,47%	-2,74%	-2,62%
79%	-3,91%	-4,42%	-4,95%	-3,69%	-3,78%
89%	-6,53%	-6,33%	-6,84%	-5,17%	-5,40%
99%	-14,76%	-14,29%	-14,18%	-17,12%	-17,75%

6.16.5 Le dataset U20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l’aide de chocs de taux de 20bps à la hausse sur la courbe de taux EIOPA.

Graphique de l’évolution des BE par percentiles

Courbe des BE du dataset U20 prédits par percentiles

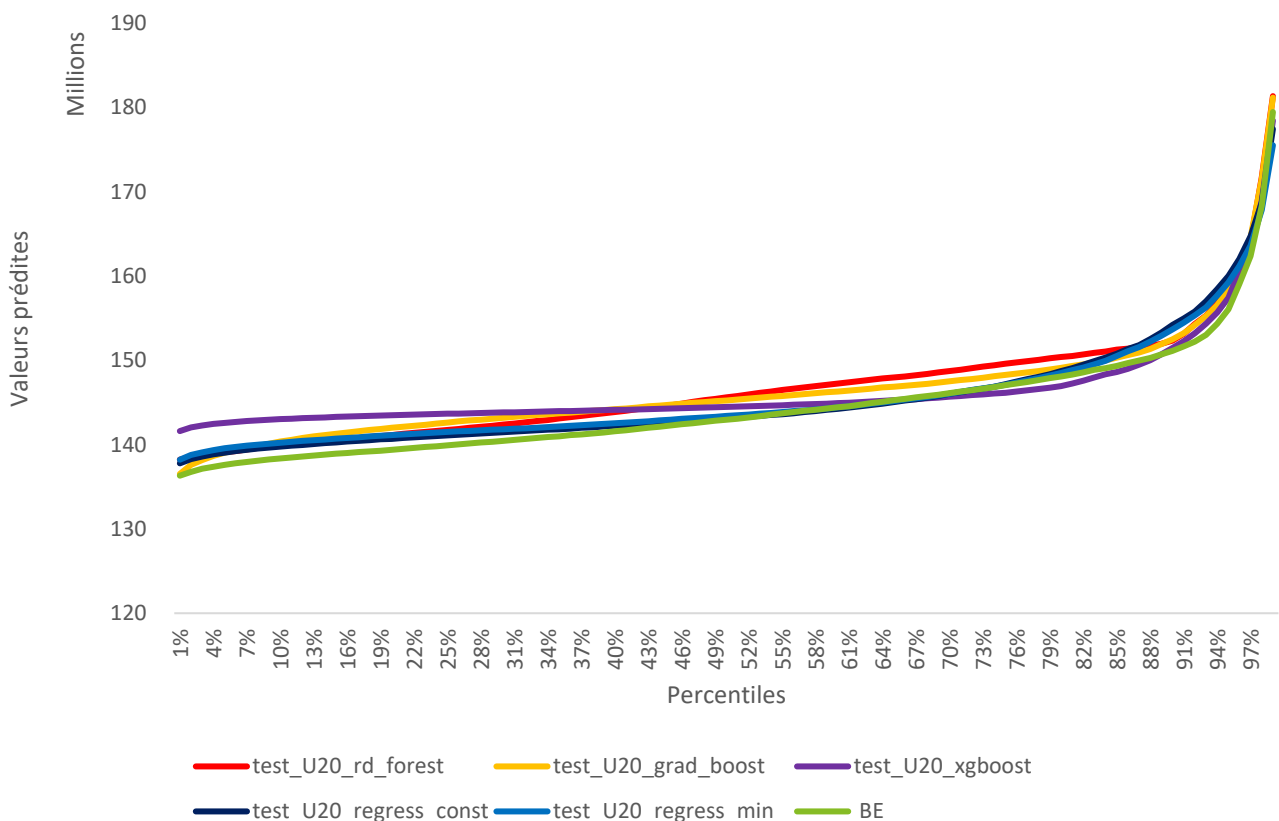


Tableau des écarts relatif des BE prédits

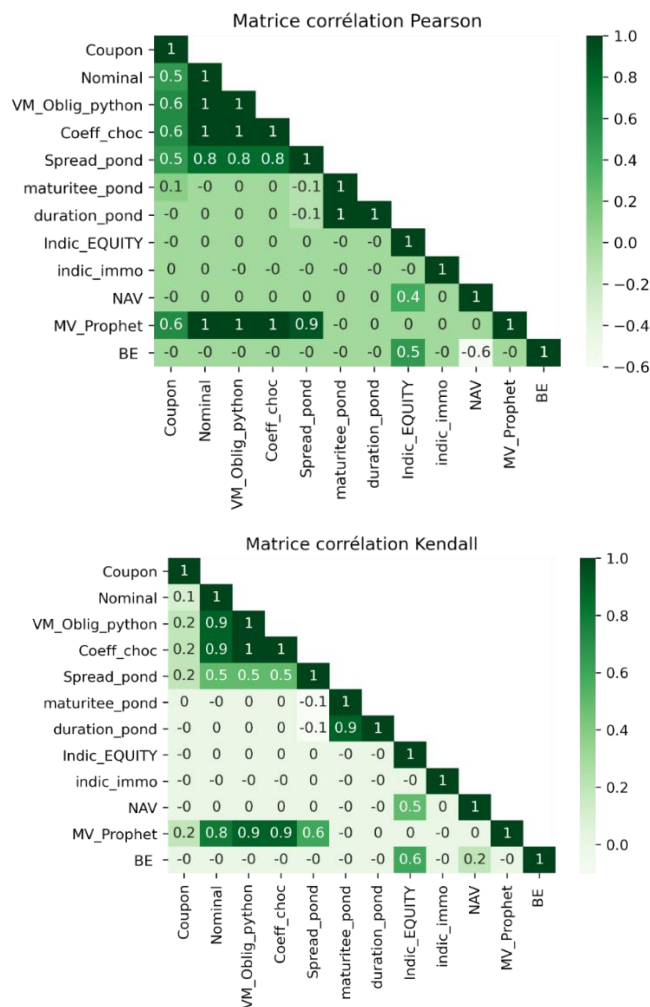
Ecart relatif	test_U20 _rd_forest	test_U20 _grad_boost	test_U20 _XGboost	test_U20 _regress_const	test_U20 _regress_min
9%	1,11%	1,37%	3,42%	1,03%	1,37%
19%	1,24%	1,85%	2,99%	0,96%	1,26%
29%	1,35%	1,94%	2,44%	0,75%	1,01%
39%	1,60%	1,87%	1,87%	0,49%	0,71%
49%	1,82%	1,63%	1,13%	0,14%	0,34%
59%	1,94%	1,33%	0,40%	-0,12%	0,00%
69%	1,83%	1,00%	-0,23%	-0,09%	-0,05%
79%	1,60%	0,70%	-0,77%	0,32%	0,21%
89%	0,83%	0,80%	-0,01%	1,70%	1,47%
99%	1,05%	0,95%	-0,57%	-1,12%	-2,20%

6.17 Annexe pour la Base de données GSE variant et indicateurs obligataires avec un pool de TMG

6.17.1 Tableau temps d'apprentissage

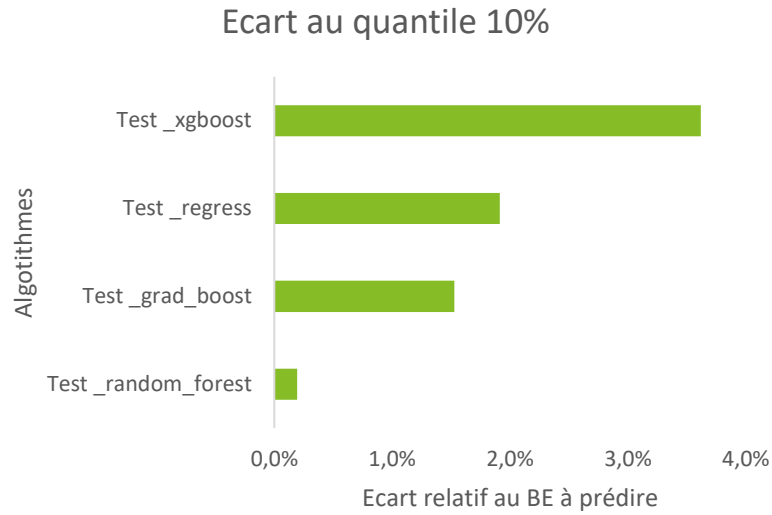
Algorithme	Temps d'apprentissage
Regression linéaire min	0,14
Regression linéaire const	0,25
XGboost	6,66
grad_boost	26,31
rd_forest	59,17

6.17.2 Les matrices de corrélations



6.17.3 Le dataset test

Graphique des écarts relatifs au quantile 10%



Graphique des écarts relatifs au quantile 90%

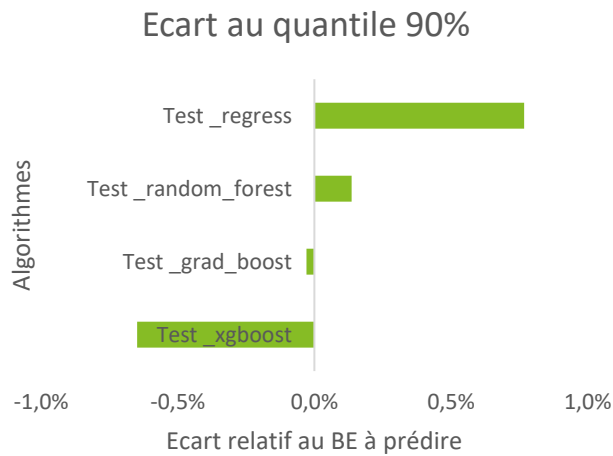
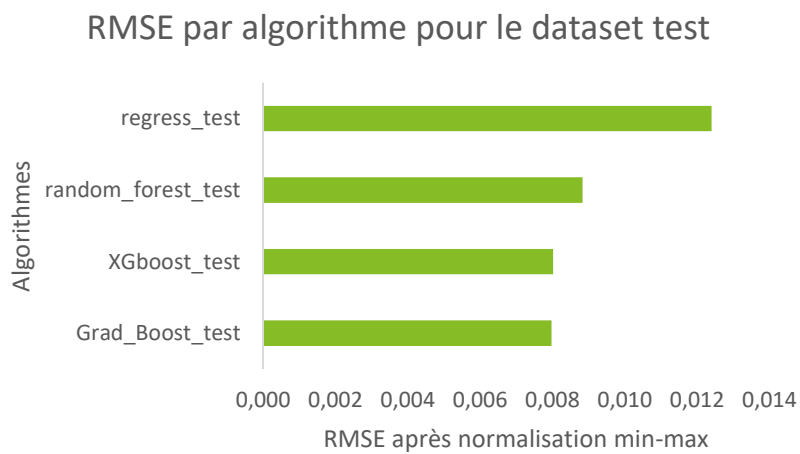


Tableau RMSE



6.17.4 Le dataset D20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la baisse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

Courbe des BE du dataset D20 prédits par percentiles

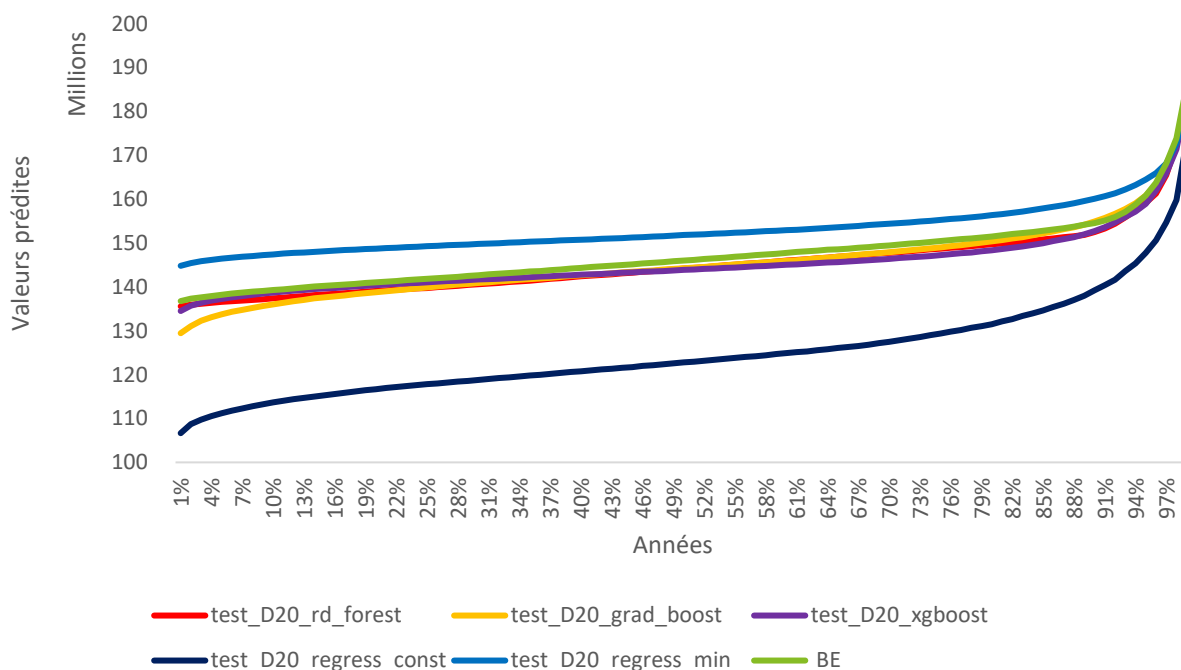


Tableau des écarts relatif des BE prédits

Ecart relatif	test_D20_rd_forest	test_D20_grad_boost	test_D20_XGboost	test_D20_regress_const	test_D20_regress_min
9%	-1,36%	-2,46%	-0,44%	-18,58%	5,86%
19%	-1,39%	-1,65%	-0,45%	-17,35%	5,50%
29%	-1,44%	-1,25%	-0,65%	-16,76%	5,05%
39%	-1,33%	-1,17%	-1,03%	-16,34%	4,52%
49%	-1,28%	-1,19%	-1,43%	-15,93%	4,01%
59%	-1,22%	-1,23%	-1,79%	-15,52%	3,55%
69%	-1,11%	-1,05%	-2,02%	-14,82%	3,33%
79%	-1,19%	-0,77%	-2,10%	-13,36%	3,18%
89%	-1,48%	0,01%	-1,42%	-10,49%	3,54%
99%	-1,58%	-0,91%	-1,49%	-6,74%	-1,98%

6.17.5 Le dataset U20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la hausse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

Courbe des BE du dataset U20 prédits par percentiles

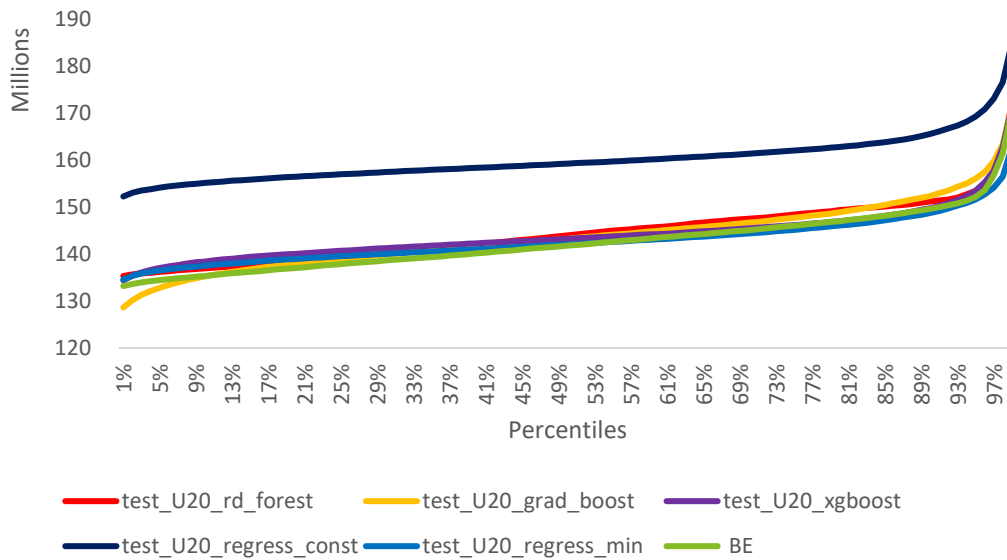


Tableau des écarts relatif des BE prédits

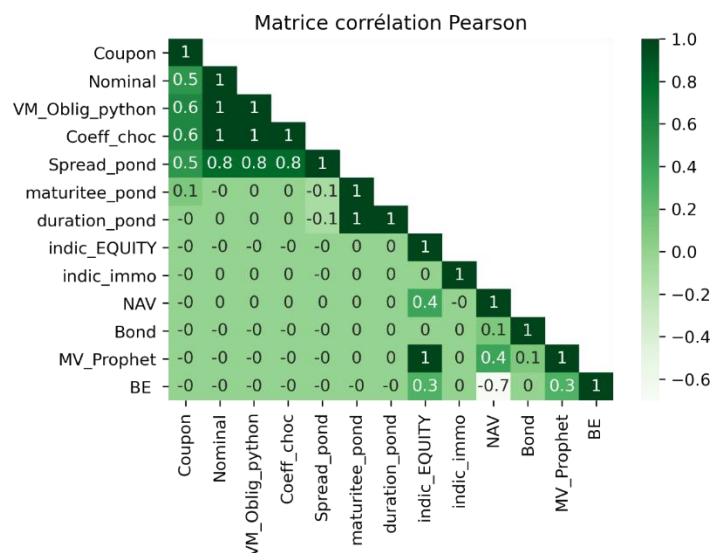
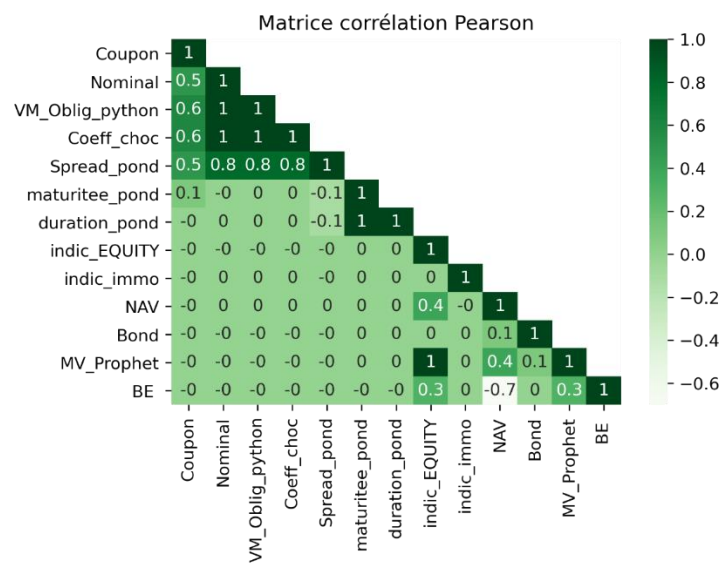
Ecart relatif	test_U20_rd_forest	test_U20_grad_boost	test_U20_XGboost	test_U20_regress_const	test_U20_regress_min
9%	1,21%	-0,27%	2,25%	14,64%	1,59%
19%	1,01%	0,60%	2,24%	14,21%	1,44%
29%	1,06%	0,91%	1,93%	13,61%	1,06%
39%	1,33%	1,08%	1,59%	13,07%	0,68%
49%	1,54%	1,06%	1,08%	12,39%	0,24%
59%	1,60%	1,03%	0,60%	11,72%	-0,19%
69%	1,70%	1,08%	0,27%	11,24%	-0,46%
79%	1,57%	1,24%	-0,01%	10,76%	-0,65%
89%	1,13%	1,81%	0,10%	10,62%	-0,60%
99%	0,68%	0,34%	0,23%	8,24%	-4,74%

6.18 Annexe Base de données GSE variant et indicateurs obligataires avec trois pools de TMG

6.18.1 Tableau temps d'apprentissage

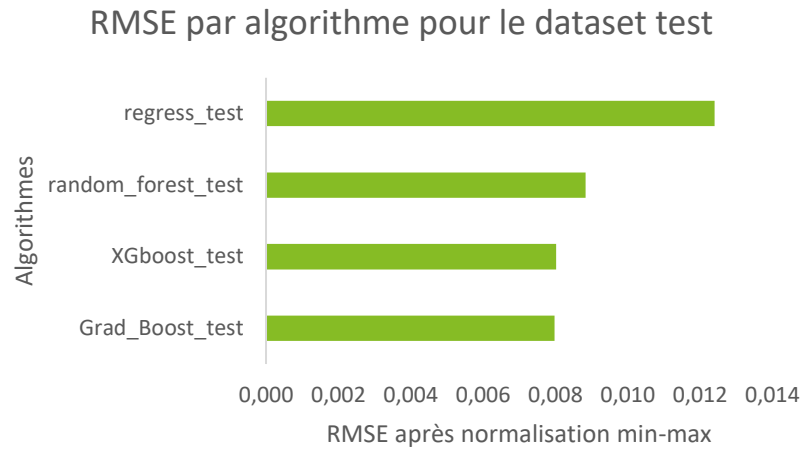
Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,21
Regression_linéaire_min	0,31
grad_boost	2,89
XGboost	12,51
rd_forest	36,38

6.18.2 Les matrices de corrélations



6.18.3 Le dataset test

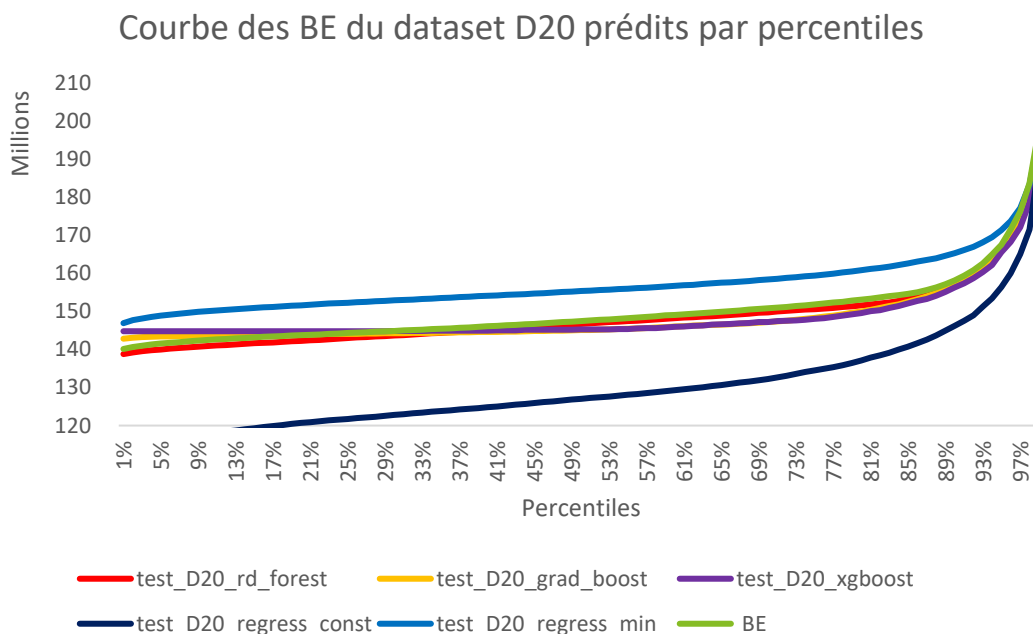
Graphique RMSE



6.18.4 Le dataset D20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la baisse sur la courbe de taux EIOPA.

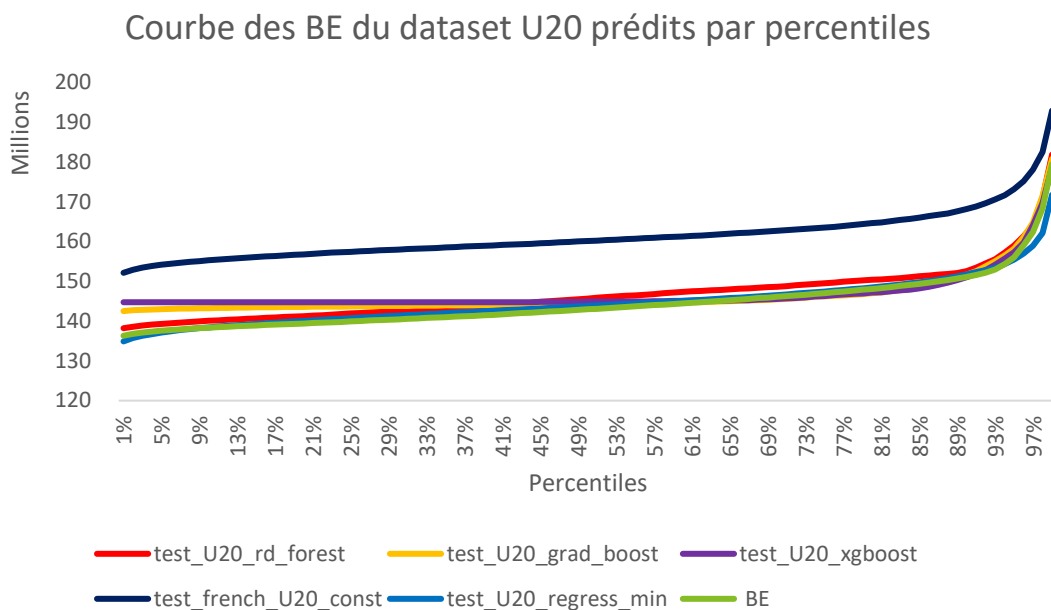
Graphique de l'évolution des BE par percentiles



6.18.5 Le dataset U20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la hausse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

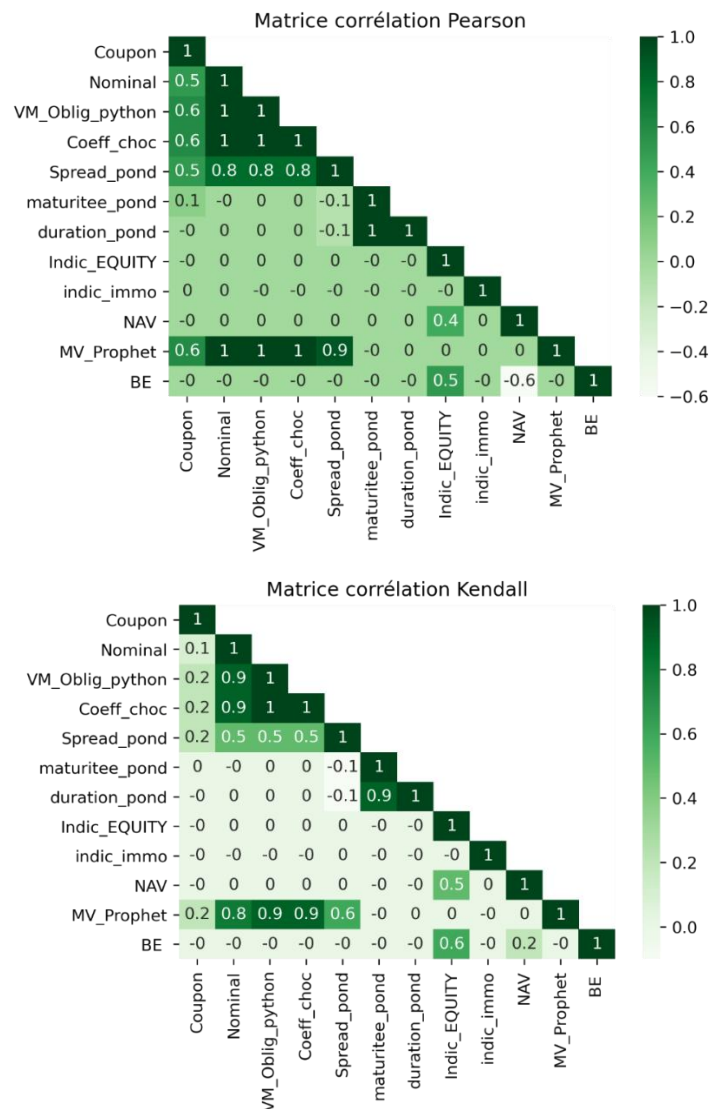


6.19 Annexe Base de données GSE variant, chocs obligataires et réallocation dynamique avec un pool de TMG

6.19.1 Tableau temps d'apprentissage

Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,15
Regression_linéaire_min	0,25
grad_boost	0,86
XGboost	2,97
rd_forest	20,26

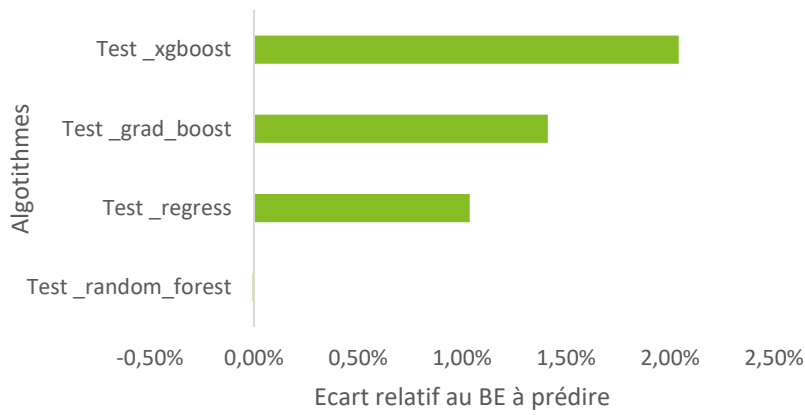
6.19.2 Les matrices de corrélations



6.19.3 Le dataset test

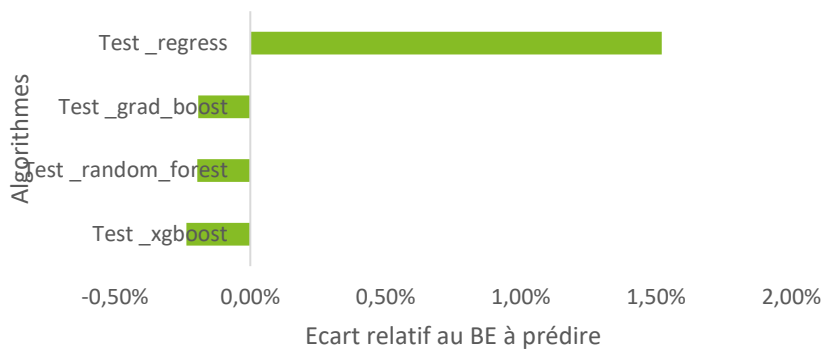
Graphique des écarts relatifs au quantile 10%

Ecart au quantile 10%



Graphique des écarts relatifs au quantile 90%

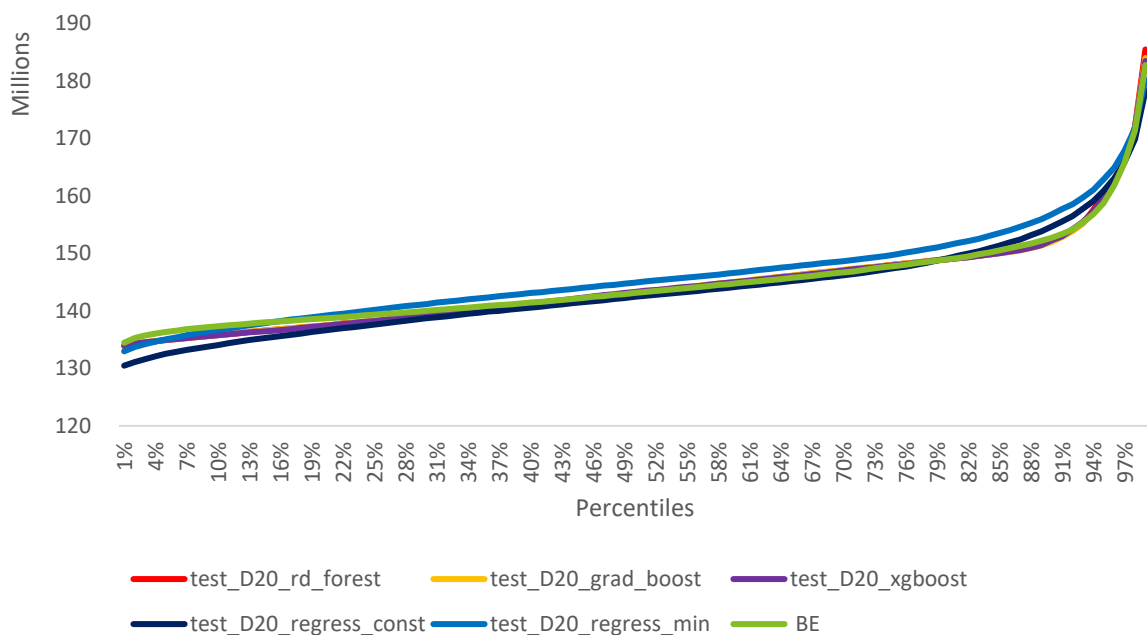
Ecart au quantile 90%



6.19.4 Le dataset D20

Graphique de l'évolution des BE par percentiles

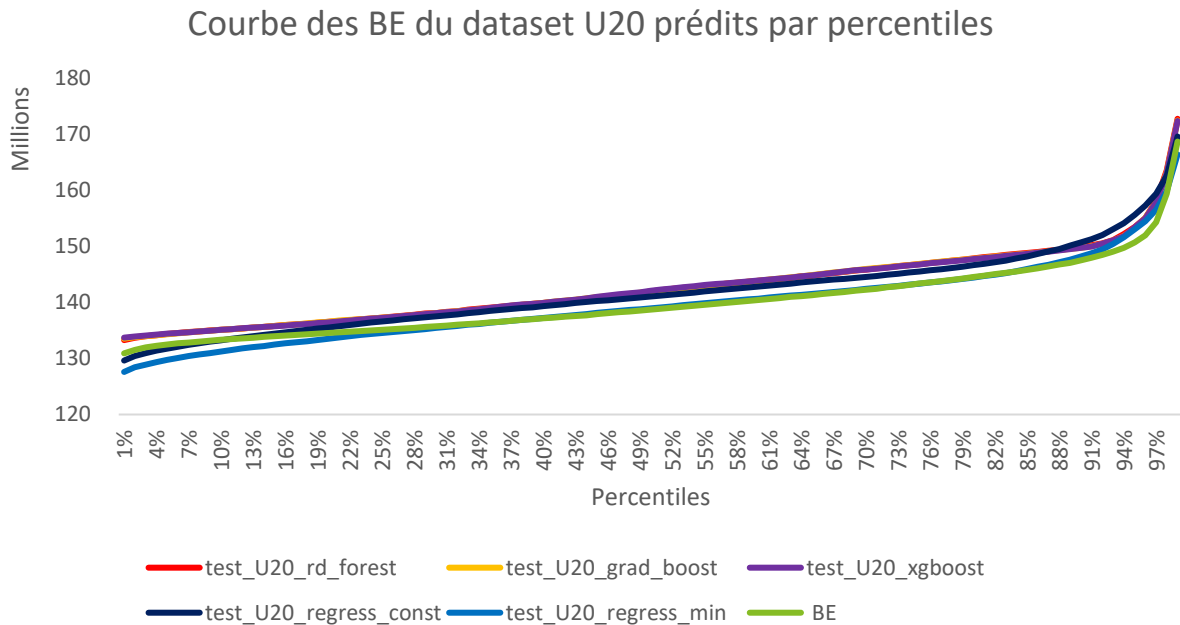
Courbe des BE du dataset D20 prédits par percentiles



6.19.5 Le dataset U20

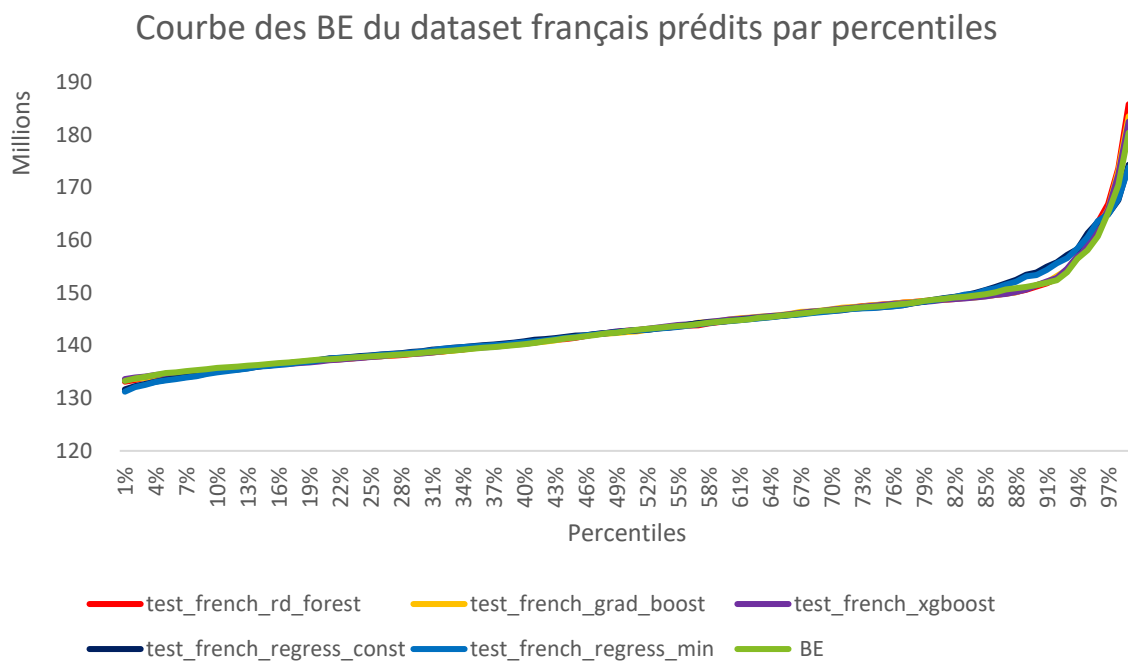
Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la hausse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles



6.19.6 Le dataset français

Graphique de l'évolution des BE par percentiles

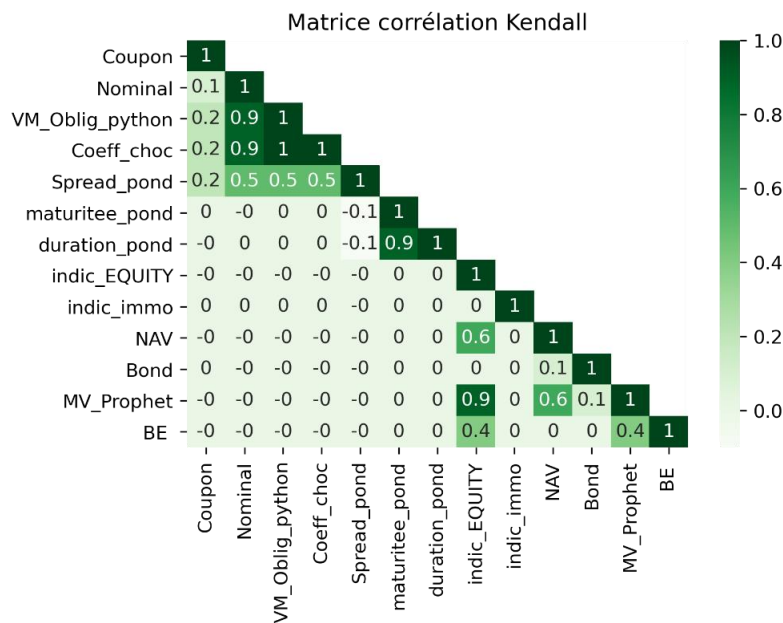
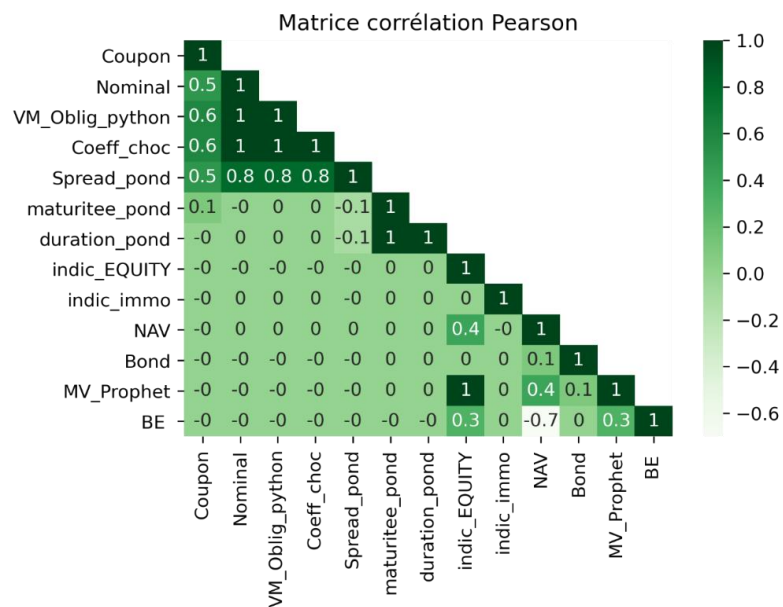


6.20 Annexe Base de données GSE variant, chocs obligataires et réallocation dynamique avec trois pools de TMG

6.20.1 Tableau temps d'apprentissage

Algorithme	Temps d'apprentissage
Regression_linéaire_const	0,21
Regression_linéaire_min	0,31
grad_boost	2,89
XGboost	12,51
rd_forest	36,38

6.20.2 Les matrices de corrélations



6.20.3 Le dataset test

Graphique des BE par percentile

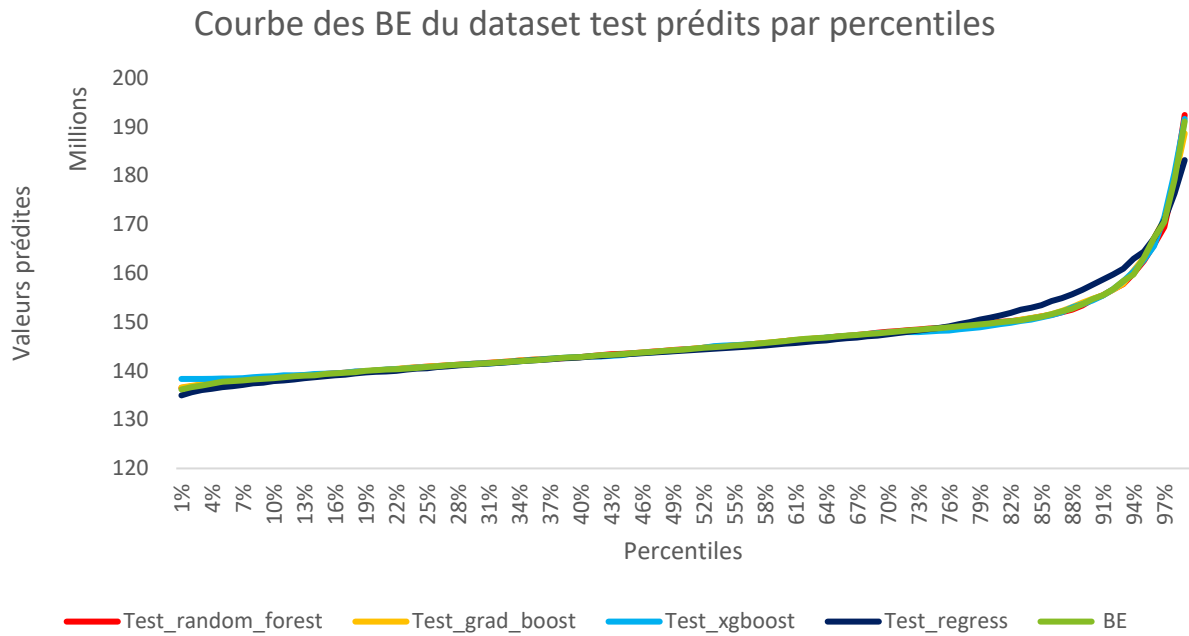
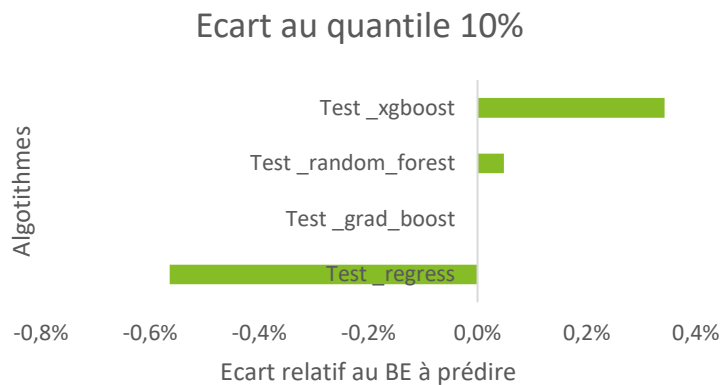


Tableau des écarts relatifs au BE

Ecart relatif	Test _random_forest	Test _grad_boost	Test _XGboost	Test _regress
9%	0,05%	0,00%	0,34%	-0,56%
19%	-0,05%	-0,05%	-0,06%	-0,23%
29%	0,01%	0,00%	0,03%	-0,08%
39%	-0,04%	-0,03%	-0,02%	-0,09%
49%	0,05%	0,01%	-0,06%	-0,11%
59%	-0,07%	-0,02%	-0,10%	-0,39%
69%	0,11%	-0,04%	0,06%	-0,28%
79%	-0,01%	-0,05%	-0,38%	0,64%
89%	-0,14%	0,21%	0,00%	1,92%
99%	0,69%	-1,28%	0,27%	-4,15%

Graphique des écarts relatifs au quantile 10%



Graphique des écarts relatifs au quantile 90%

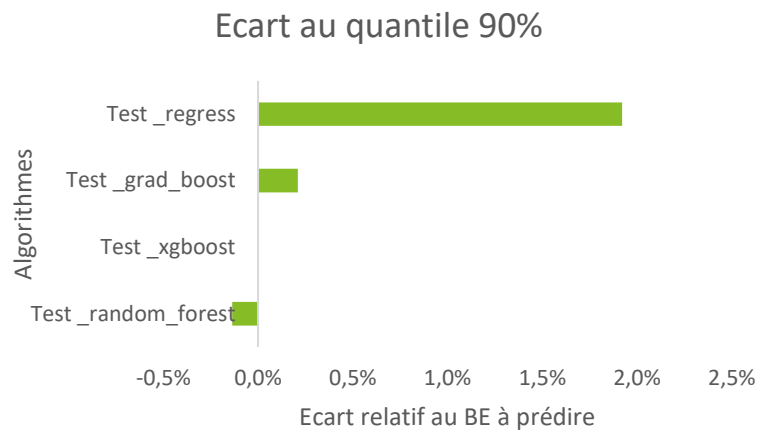
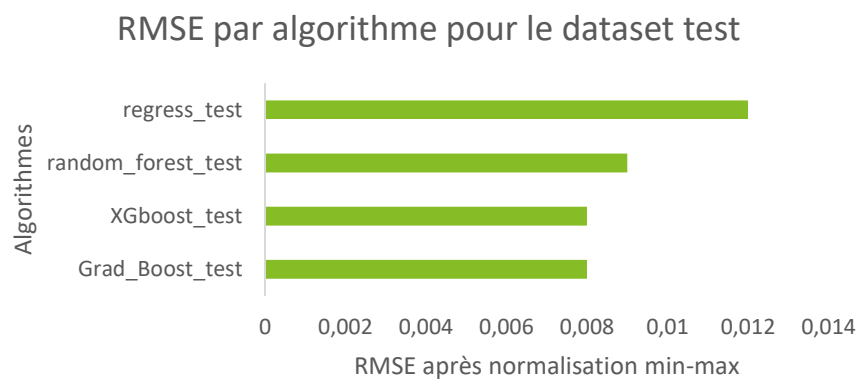


Tableau RMSE



6.20.4 Le dataset D20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la baisse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

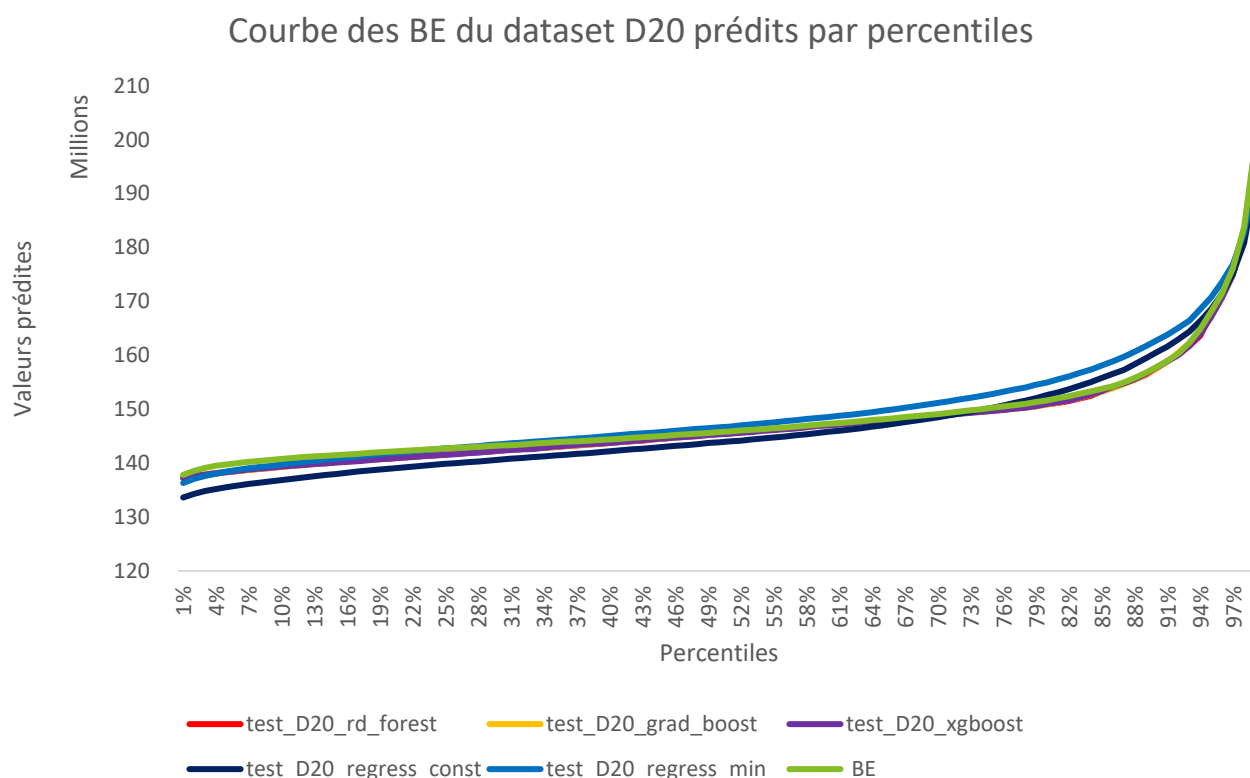


Tableau des écarts relatif des BE prédits

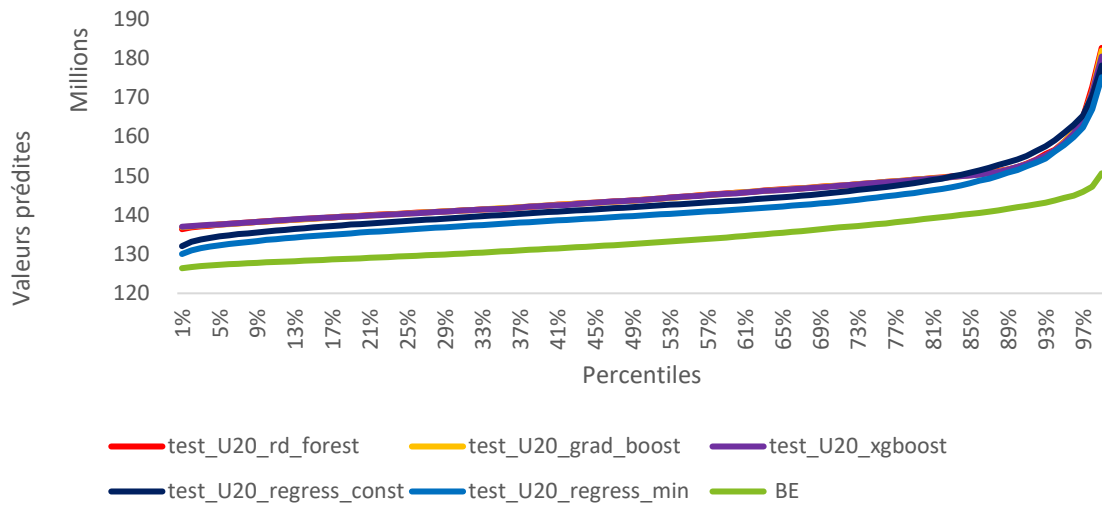
Ecart relatif	test_D20_rd_forest	test_D20_grad_boost	test_D20_XGboost	test_D20_regress_const	test_D20_regress_min
9%	-1,05%	-1,05%	-1,01%	-2,82%	-0,79%
19%	-0,85%	-0,89%	-0,91%	-2,25%	-0,25%
29%	-0,68%	-0,61%	-0,68%	-1,83%	0,16%
39%	-0,52%	-0,46%	-0,46%	-1,57%	0,41%
49%	-0,25%	-0,26%	-0,27%	-1,30%	0,59%
59%	-0,12%	-0,13%	-0,17%	-1,03%	0,87%
69%	-0,18%	-0,16%	-0,18%	-0,47%	1,30%
79%	-0,46%	-0,53%	-0,52%	0,47%	2,10%
89%	-0,27%	-0,21%	-0,14%	1,71%	3,13%
99%	-1,17%	-0,86%	-0,20%	-3,49%	-2,70%

6.20.5 Le dataset U20

Afin de challenger les algorithmes, a été implémenté un nouveau jeu de 10000 scénarios stochastiques générés à l'aide de chocs de taux de 20bps à la hausse sur la courbe de taux EIOPA.

Graphique de l'évolution des BE par percentiles

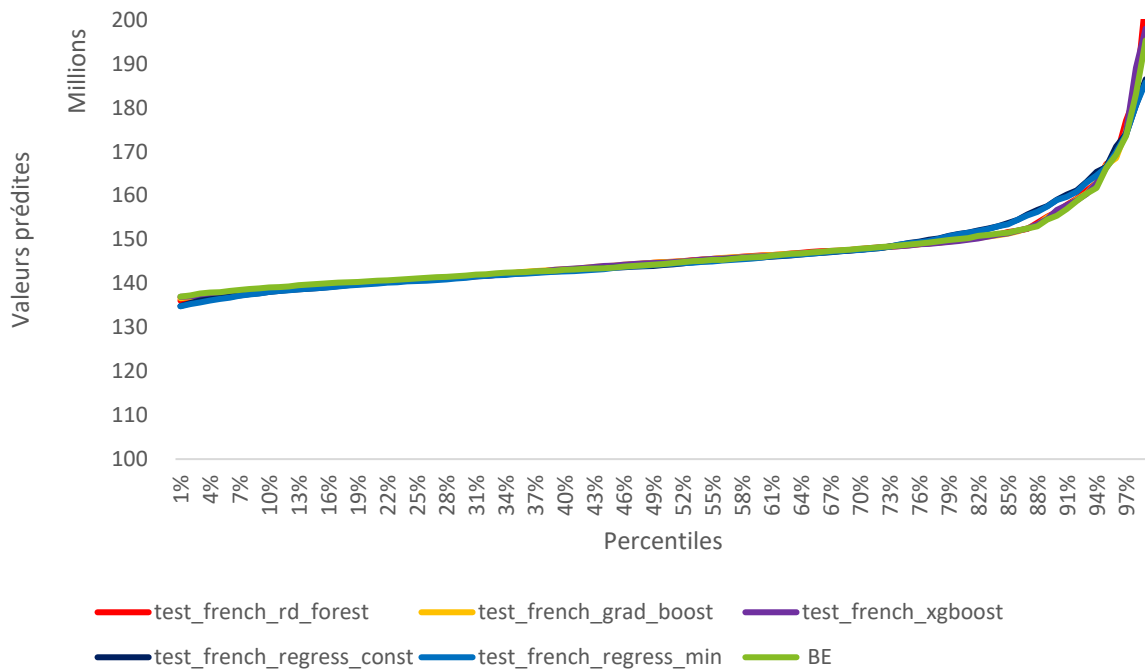
Courbe des BE du dataset U20 prédits par percentiles



6.20.6 Le dataset français

Graphique de l'évolution des BE par percentiles

Courbe des BE du dataset français prédits par percentiles



7 Bibliographie¹⁴⁹

Abdel MESSOUSSI (2017). *Application d'algorithmes de Machine Learning pour l'estimation du ratio de couverture d'un assureur-vie détenteur d'un produit épargne*. Mémoire d'actuariat, ENSAE

Alaa DAHER (2019). *Optimisation des réseaux cellulaires pour les communications professionnelles critiques*. Thèse, Télécom Paris

Amine DRISSI BOUTAYBI (2014). *Technique d'agrégation des risques*. Mémoire d'actuariat, EURIA

Arnaud BUZZI (2017). *Approximation du bilan économique sous Solvabilité II via des méthodes d'apprentissage automatique et application à l'ORSA*. Mémoire d'actuariat, Centrale-Dauphine.

Aurélien SURAY (2020). *Développement d'un outil de prédiction des résultats basé sur les techniques d'intelligence artificielle*. Mémoire de fin d'étude pour l'obtention du titre d'ingénieur, ESILV

Banque de France (2009). *DIRECTIVE 2009/138/CE DU PARLEMENT EUROPÉEN ET DU CONSEIL du 25 novembre 2009 sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II)*

ECB (2022). *Statistical Data Warehouse*

EIOPA (2015). *RÈGLEMENT DÉLÉGUÉ (UE) 2015/35 DE LA COMMISSION du 10 octobre 2014 complétant la directive 2009/138/CE du Parlement européen et du Conseil sur l'accès aux activités de l'assurance et de la réassurance et leur exercice (solvabilité II)*

EIOPA (2019). *ARTIFICIAL INTELLIGENCE GOVERNANCE PRINCIPLES: TOWARDS ETHICAL AND TRUSTWORTHY ARTIFICIAL INTELLIGENCE IN THE EUROPEAN INSURANCE SECTOR*

Eve ZURFLUH (2019). *Utilisation du Machine Learning dans l'estimation du ratio de solvabilité d'un assureur vie et application aux Reverse Stress Tests*. Mémoire d'actuariat, ENSAE

Fédération Française des Société d'Assurance (2009). *Réavis de la norme Solvabilité II*

France assurance (2021). *L'assurance Française, Données clefs 2020*

Hélène LASSARTESES (2013). *Intérêt de l'investissement en actions protégées dans un contrat d'épargne*. Mémoire d'actuariat, ESSEC

INSEE (2022). *Indice des prix à la consommation - résultats provisoires (IPC) - mai 2022*

¹⁴⁹ La bibliographie a été enrichie avec d'autres mémoires portant sur l'application du Machine Learning pour la prédiction de résultats ALM

Institut des ACTUAIRES (2018) *Document d'orientation, les générateurs de scénarios économiques, éléments d'analyse et exemple de pratiques actuarielles applicables au marché français*

Jeremy LESNE (2012). *Calculation of the required capital for interest rates and equity risks for variable annuity products*. Mémoire d'actuariat, EURIA

Le Monde (2022). *Wall Street chute, inquiète de l'inflation et de la perspective d'une hausse des taux*

Rémi GAUVILLE (2017). *Projection du ratio de solvabilité : des méthodes de Machine Learning pour contourner les contraintes opérationnelles de la méthode des SdS*. Mémoire d'actuariat, EURIA

Ruocong ZHANG (2014). *Apprentissage statistique en gestion de portefeuille*. Thèse, Télécom Paris

Silvia BUCCI (2021). *Etude et implémentation de techniques d'analyse de sensibilité dans les modèles de tarification Non-Vie*. Mémoire d'actuariat, ENSAE

Styven PONNUSAMY (2020). *Méthodes de sélection d'instruments financiers pour le calibrage de portefeuilles répliquants*. Mémoire d'actuariat ISUP

Valentin SAVIDAN (2015). *Comptabilisation de la valeur temps des options et garanties intrinsèques d'un contrat d'épargne Euro selon la norme IFRS4 - phase II*. Mémoire d'actuariat, DUAS

Valentin CERISIER (2021). *Application de méthodes de Machine Learning dans le calcul de la solvabilité infra-annuelle*. Mémoire d'actuariat, Dauphine

Vanda ROUSSEAU (2011). *Le Best Estimate appliqué aux contrats d'épargne et de retraite selon les règles de Solvabilité II*. Mémoire d'actuariat, ISFA

Zong Xiangyu (2021). *Machine Learning in stock indices trading and pair trading*. Phd thesis, University of Glasgow