

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 10/11/2022

Par : Cléa Laouar

Titre : Automatisation de la calibration des portefeuilles
répliquants par algorithme génétique

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

M. Nicolas Baradel *Baradel*

Entreprise : 

Nom : AXA France

Signature : *Zouag*

Membres présents du jury de l'Institut
des Actuaires

Jean-Marie Nessi

Tahmoor Qureshi

Directeur du mémoire en entreprise : *Zouag*

Nom : M. Zakaria Zouag

Signature : *Zouag*

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

Zouag

Secrétariat :

Signature du candidat

Bibliothèque :

Laouar

RÉSUMÉ

Mots clés : Portefeuilles répliquants, Algorithme génétique, Optimisation, Sélection

Les compagnies d'assurance sont fortement réglementées. Parmi les différentes contraintes auxquelles elles sont soumises, elles doivent respecter l'exigence en capital de solvabilité requis (SCR) issue du premier pilier de Solvabilité II. Le SCR de marché présente des difficultés calculatoires car contrairement à l'actif, le passif n'est pas facilement valorisable par formules fermées. Afin d'approximer le *Best Estimate Liabilities* (BEL), AXA France utilise la méthode des portefeuilles de réplication. Cette méthode consiste à déterminer un ensemble d'instruments financiers facilement valorisable capable de répliquer le BEL et possédant le même comportement face aux différents chocs de marché. La calibration des portefeuilles de réplication peut se scinder en deux sous-problématiques : le choix des instruments servant à répliquer le BEL d'une part, le calcul des nominaux qui leurs sont associés d'autre part.

La méthode actuelle de sélection d'instruments, à dire d'expert, est très chronophage. Ce mémoire a pour objectif de réduire le temps accordé à la sélection des instruments financiers. L'automatisation par algorithme génétique est envisagée. Ce choix se justifie par la quantité notable d'instruments financiers candidats formant la base de travail. L'algorithme génétique présente également l'avantage de proposer un ensemble de solutions potentielles et non une solution unique comme le feraient les algorithmes usuels. Plusieurs portefeuilles candidats peuvent ainsi être comparés. En un temps raisonnable d'exécution, l'algorithme parvient à produire des individus vérifiant les critères de performance mais le nombre d'instruments demeure au dessus du seuil de 50. Afin de réduire davantage la taille du portefeuille répliquant sans augmenter le temps d'exécution, deux algorithmes supplémentaires ont été implémentés : ils réduisent le nombre d'instruments en augmentant les indicateurs de performance clés.

ABSTRACT

Keywords : *Replicating portfolio, Genetic algorithm, Optimization, Selection*

Insurance companies are highly regulated. Among the various constraints they are subject to, they must respect the Solvency Capital Requirement (SCR) resulting from the first pillar of Solvency II. The market SCR is difficult to calculate because, unlike assets, liabilities are not easily valued by closed formulas. In order to approximate the Best Estimate Liabilities (BEL), AXA France uses the replicating portfolio method. This method consists in determining a set of financial instruments that can be easily valued, can replicate the BEL and have the same behavior in response to different market shocks. The calibration of replicating portfolios can be divided into two sub-problems : the choice of instruments used to replicate the BEL on the one hand, and the calculation of the associated nominal amounts on the other.

The current method of instrument selection, based on expert judgement, is very time consuming. The objective of this paper is to reduce the time spent on the selection of financial instruments. The automation using a genetic algorithm is considered. This choice is justified by the significant quantity of financial instruments candidates forming the database. The genetic algorithm also has the advantage of proposing a set of potential solutions and not a single solution as the most common algorithms would provide. Several candidate portfolios can thus be compared. In a reasonable execution time, the algorithm manages to produce individuals verifying the performance criteria but the number of instruments remains above the threshold of 50. In order to further reduce the size of the replicating portfolio without increasing the execution time, two additional algorithms have been implemented : they reduce the number of instruments by increasing the key performance indicators.

REMERCIEMENTS

Je tiens tout d'abord à remercier mon maître de stage M. Zakaria Zouaq, actuaire de l'équipe Modèles et Solvabilité II, pour son accueil, son suivi et le partage de son expertise tout au long de ce stage.

Je remercie ensuite l'ensemble de l'équipe Modèles et Solvabilité II : M. Alban Davand, M. Mohamed Amine Ouadeih, Mme Violaine Roville, Mme Abire Sabir et Mme Afnane Kotbi pour leurs conseils et l'excellent environnement de travail dans lequel j'ai eu la chance de réaliser ce stage.

J'adresse également mes remerciements à mon tuteur pédagogique, M. Nicolas Baradel, pour son suivi, ainsi qu'à mon professeur, Mme Caroline Hillairet, pour son aide précieuse au cours de cette année scolaire.

Enfin, je remercie mes proches pour leur soutien et la relecture de ce mémoire.

TABLE DES MATIÈRES

Résumé	2
Abstract	3
Introduction	8
1 L'assurance : un secteur d'activité aux contraintes réglementaires lourdes	9
1.1 Les trois piliers de Solvabilité II	9
1.1.1 Pilier 1 : Une exigence en capital	9
1.1.2 Pilier 2 : Des activités de contrôle et de supervision	10
1.1.3 Pilier 3 : Une discipline de marché	10
1.2 La contrainte SCR	11
1.2.1 La définition du SCR	11
1.2.2 L'évaluation économique du bilan pour le calcul du SCR	11
1.2.3 Le SCR en formule standard	13
1.2.4 Le SCR associé au risque de marché	16
1.3 Les différentes méthodes de calcul du SCR de marché en modèle interne	16
1.3.1 La méthode des Simulations dans les Simulations (SdS)	17
1.3.2 L'accélérateur SdS	18
1.3.3 Le Curve fitting	19
1.3.4 Le Least Square Monte Carlo	20
1.3.5 Le Portefeuille de Réplication	21
2 La méthode des portefeuilles répliquants	22
2.1 Les motivations d'AXA	22
2.2 Les données nécessaires à l'application de la méthode	23
2.2.1 Les zéros-coupons	23
2.2.2 Le zéro-coupon inflation	24
2.2.3 Les swaptions receveur et payeur	24
2.2.4 Les indices equity	25
2.2.5 Les call et put sur equity	26
2.2.6 Synthèse des instruments financiers utilisés	27
2.3 La description de la méthode	28
2.3.1 Les définitions et notations	28
2.3.2 Les caractéristiques d'un portefeuille répliquant pertinent	30
2.3.3 La formalisation du problème	30
2.3.4 La résolution pratique du problème d'optimisation	31

3	Les difficultés de la méthode des portefeuilles répliquants	32
3.1	Le calcul des nominaux d'un portefeuille répliquant	32
3.1.1	La formalisation du problème	32
3.1.2	La régression ridge	33
3.1.3	La factorisation LU	33
3.2	Les méthodes actuelles de sélection d'instruments	37
3.2.1	L'Orthogonal Matching Pursuit	37
3.2.2	La méthode "A dire d'expert"	39
3.2.3	Les inconvénients des méthodes actuelles	40
4	Le choix de l'algorithme génétique	42
4.1	L'origine de l'algorithme	42
4.2	Les cas d'utilisation et le principe de l'algorithme	43
4.3	La présentation de l'algorithme	44
4.3.1	La représentation des solutions	44
4.3.2	Les opérateurs d'évolution	47
4.3.3	Des paramètres additionnels	49
4.4	Un exemple introductif	50
4.4.1	La problématique	50
4.4.2	Les composantes de l'algorithme	50
4.4.3	Les opérateurs d'évolution	51
4.4.4	Les paramètres de l'algorithme	51
4.4.5	Les résultats	51
4.5	Les forces et faiblesses de l'algorithme	53
5	L'Application de l'algorithme génétique au cas des portefeuilles répliquants	55
5.1	Les données utilisées	55
5.1.1	Le passif à répliquer	55
5.1.2	Les instruments financiers facilement valorisables	55
5.1.3	Les sensibilités	57
5.2	La définition des paramètres du problème	57
5.2.1	Les gènes	57
5.2.2	L'individu	59
5.2.3	La génération	60
5.2.4	L'environnement	62
5.2.5	Le processus de croisement	62
5.2.6	Le processus de mutation	64
5.2.7	Le processus de sélection	66
5.2.8	Synthèse de l'ensemble des paramètres de l'algorithme génétique	67
5.3	Le choix de la fonction de coût	68
5.4	L'influence des paramètres	70
5.4.1	Les paramètres de croisement et de mutation	70
5.4.2	Le nombre de générations	70
5.4.3	Le nombre d'instruments initial	72
5.4.4	Le nombre d'individus par génération	74
5.5	Les résultats sur le <i>portefeuille A</i>	76
5.6	L'étude des topologies	83
5.7	La présentation de l'outil Python	86

6 Les Problématiques rencontrées et solutions préconisées	87
6.1 Le nombre d'instruments	87
6.2 Certains KPI non atteints	89
6.3 Le temps de calcul peut encore être réduit	90
6.4 Les topologies	91
Conclusion	92
Bibliographie	94
A La Value-at-Risk	95
B La formule du SCR	96
C Un exemple de réplication parfaite	97
D Le meilleur individu par génération de l'exemple introductif	99
E Le test de Kolmogorov-Smirnov	101
Note de synthèse	103
Executive summary	108

INTRODUCTION

La crise financière de 2008 a montré l'importance de la gestion des risques au sein des secteurs bancaire et assurantiel. Avec une structure en trois piliers, la réglementation Solvabilité II impose aux compagnies d'assurance de démontrer que la dimension "risques" est intégrée dans les décisions stratégiques. Parmi les différentes contraintes auxquelles ces compagnies sont soumises, elles doivent respecter l'exigence en capital de solvabilité requis (SCR) issue du premier pilier nécessaire au calcul du ratio de solvabilité II. Ce ratio met en évidence la capacité d'une compagnie d'assurance à respecter ses engagements vis-à-vis de ses assurés.

AXA est un leader mondial de l'assurance qui doit donc se soumettre à la réglementation Solvabilité II. Dans son communiqué de presse datant du 3 août 2022, la compagnie d'assurance présente un ratio de solvabilité II égal à 227% et en hausse de 10 points par rapport à la fin de l'année 2021 [1].

Au sein d'AXA France, l'équipe Modèles et Solvabilité II est en charge du calcul du ratio de solvabilité et du SCR. Afin d'avoir un modèle adapté au profil de risque de la compagnie, Axa France a fait le choix du modèle interne. Pour calculer le SCR de marché, la méthode des portefeuilles de réplication est utilisée. Cette méthode consiste à déterminer un ensemble d'instruments financiers facilement valorisable capable de répliquer le *Best Estimate Liabilities*. Cependant, la méthode actuelle permettant de sélectionner les instruments financiers utilisés pour la réplication est très chronophage. Par conséquent, l'objectif de ce mémoire est de proposer une solution moins coûteuse en temps capable d'automatiser la sélection d'instruments financiers.

Cette solution s'inscrit donc dans un contexte fortement réglementé dans lequel il convient de revenir dans ce mémoire (1). Toute compagnie d'assurance se doit aujourd'hui, entre autres, de respecter le SCR. Pour le calcul du SCR de marché, Axa France utilise la méthode des portefeuilles de réplication (2). Cette méthode présente deux difficultés : le calcul des nominaux d'une part et la sélection des instruments d'autre part (3). Si les résultats actuels associés aux méthodes de sélection d'instruments d'AXA sont probants, il n'en demeure pas moins que le temps de travail alloué à l'obtention des résultats est trop lourd. Aussi, il a été recherché une solution d'automatisation du procédé de sélection. Pour ce faire, l'algorithme génétique (4) a été retenu pour être appliqué aux portefeuilles répliquants (5). Les limites inhérentes à ce choix ont été contournées (6).

L'ASSURANCE : UN SECTEUR D'ACTIVITÉ AUX CONTRAINTES RÉGLEMENTAIRES LOURDES

1.1 Les trois piliers de Solvabilité II

Le milieu de l'assurance est **fortement réglementé**. Afin de garantir une protection appropriée des assurés, Solvabilité 2 a succédé à Solvabilité 1 en se basant sur des éléments économiques et non plus comptables. Partant d'une démarche similaire à Bale II, Solvabilité II est une réforme économique qui concède une **harmonisation du régime assurantiel au niveau européen**.

Le principal objectif de la réglementation Solvabilité II est d'**adapter les exigences en fonds propres** des compagnies d'assurance de façon proportionnelle en l'égard de la nature, de l'ampleur et de la complexité des risques auxquels elles font face. Ces exigences permettent d'une part de garantir les **engagement longs termes** des assureurs vis à vis de leurs assurés et, d'autre part, de **faire face à d'éventuels événements exceptionnels** capables d'affecter la solvabilité d'une compagnie d'assurance[16].

La réglementation Solvabilité II repose sur **trois piliers** : un pilier quantitatif, un pilier qualitatif et un pilier de transparence. A titre de rappel, ces trois piliers sont succinctement présentés dans les trois sous-sections qui suivent.

1.1.1 Pilier 1 : Une exigence en capital

Ce premier pilier est un pilier quantitatif d'exigence en capital. Les assureurs doivent en effet disposer de fonds propres suffisamment conséquent pour couvrir leurs expositions aux risques. Afin de quantifier le capital requis, deux indicateurs pour lesquels les normes de calcul sont définies dans ce pilier sont étudiés : le **Solvency Capital Requirement (SCR)** et le **Minimum Capital Requirement (MCR)**. Le premier définit le seuil de solvabilité en dessous duquel le régulateur peut intervenir et exiger la prise de mesures de redressement. Le second définit le seuil en dessous duquel la compagnie d'assurance court un risque irrecevable, déclenchant une intervention automatique du régulateur qui peut décider de suspendre les activités de l'entreprise.

L'enjeu de ce premier pilier est de **calculer un capital réglementaire qui reflète la réalité économique de l'assureur**.

1.1.2 Pilier 2 : Des activités de contrôle et de supervision

Ce second pilier fixe les normes qualitatives de suivi des risques en interne. Les principaux objectifs sont :

1. identifier les risques de l'entreprise
2. fixer les limites de tolérance au risque
3. intégrer les indicateurs de solvabilité dans les prises de décision
4. suivre régulièrement les indicateurs de risque afin d'anticiper toute dégradation de la solvabilité
5. définir et appliquer un plan d'action si la limite fixée n'est pas atteinte.

L'enjeu de ce second pilier est de **renforcer les systèmes de gouvernance et de gestion des risques**.

1.1.3 Pilier 3 : Une discipline de marché

Ce dernier pilier définit les obligations de publications des entreprises d'assurance. Ces obligations concernent les informations auxquelles le public (assurés et investisseurs) a accès. Cela est fait dans le but d'augmenter la transparence de la compagnie et d'améliorer la qualité des informations destinées aux autorités de marché afin de mieux évaluer la gestion des risques d'une entreprise d'assurance.

L'enjeu de ce troisième et dernier pilier est la **promotion de la transparence financière à des fins de communication et de contrôle**.

La figure 1.1 synthétise l'approche en trois piliers de Solvabilité II.

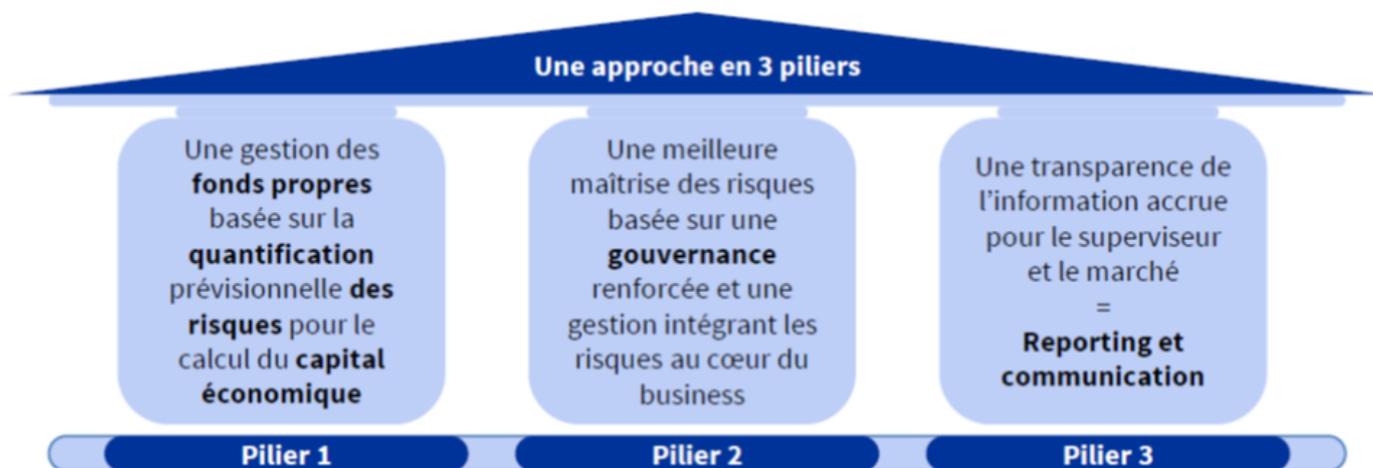


FIGURE 1.1 – Les trois piliers de Solvabilité II

L'objectif de l'étude présentée dans ce mémoire est de proposer une méthode permettant d'accélérer le calcul du SCR imposé par le pilier 1 de Solvabilité II en risque de marché.

1.2 La contrainte SCR

1.2.1 La définition du SCR

Le SCR est un indicateur du premier pilier de Solvabilité II. Il s'agit du **capital nécessaire pour absorber un événement bicentenaire dans l'année**. Plus précisément, le SCR correspond au capital économique minimal dont doit disposer une compagnie d'assurance pour faire face à une ruine économique à horizon 1 an avec une probabilité de 99.5%.

Trois notions fondamentales apparaissent :

- la **ruine économique** correspond à une situation dans laquelle les fonds propres d'une compagnie sont négatifs ou nuls : $[ruine\ économique] = [Fonds\ Propres \leq 0]$
- l'**horizon d'une année** impose de pouvoir obtenir la distribution des fonds propres économiques dans un an
- le **seuil de 99.5%** représente le seuil de confiance. La probabilité de l'événement *ruine économique* est dans ce cas inférieure à 0.5% : $\mathbb{P}(ruine\ économique) < 0.5\%$.

Le SCR est alors défini par [2], [7] :

$$SCR = \arg \min_{x \in \mathbb{R}} \mathbb{P}(Fonds\ Propres_{t=1} \leq 0 | Fonds\ Propres_{t=0} = x) < 0.5\% \quad (1.1)$$

$$= Fonds\ Propres_{t=0} - ZC(0, 1) \times VaR_{99.5\%}(Fonds\ Propres_{t=1}) \quad (1.2)$$

où

- $ZC(0, 1)$ est le prix en $t = 0$ d'un zéro coupon de maturité 1 an
- $VaR_{99.5\%}$ est la Value-at-Risk au seuil de 99.5%, soit le quantile 0.5%, dont la formule est rappelée en Annexe A [10].

La démonstration entre l'égalité des formules (1.1) et (1.2) est donnée en Annexe B.

Le SCR, indicateur du pilier 1, permet *in fine* de calculer le ratio de solvabilité (SII) qui montre la capacité d'une compagnie d'assurance à tenir ses engagements.

$$ratio_{SII} = \frac{Fonds\ Propres}{SCR}.$$

Ce ratio doit être supérieur à 1 et plus il est élevé, plus la compagnie peut faire face à des chocs extrêmes.

1.2.2 L'évaluation économique du bilan pour le calcul du SCR

Le SCR du premier pilier de Solvabilité II est basé sur l'évaluation économique du bilan de la compagnie d'assurance. Le calcul du SCR s'appuie en effet sur la **projection du bilan à horizon 1 an**. Afin de présenter les différentes méthodes de calcul du SCR, il est donc nécessaire de commencer par définir les éléments du bilan économique.

Chaque élément du bilan est calculé à partir de sa **valeur de marché**, c'est à dire au montant pour lequel cet élément (à l'actif ou au passif) peut être échangé entre deux parties informées et consentantes dans des conditions normales de concurrence.

Le bilan économique d'une compagnie d'assurance se présente tel que sur la figure 1.2 [16].

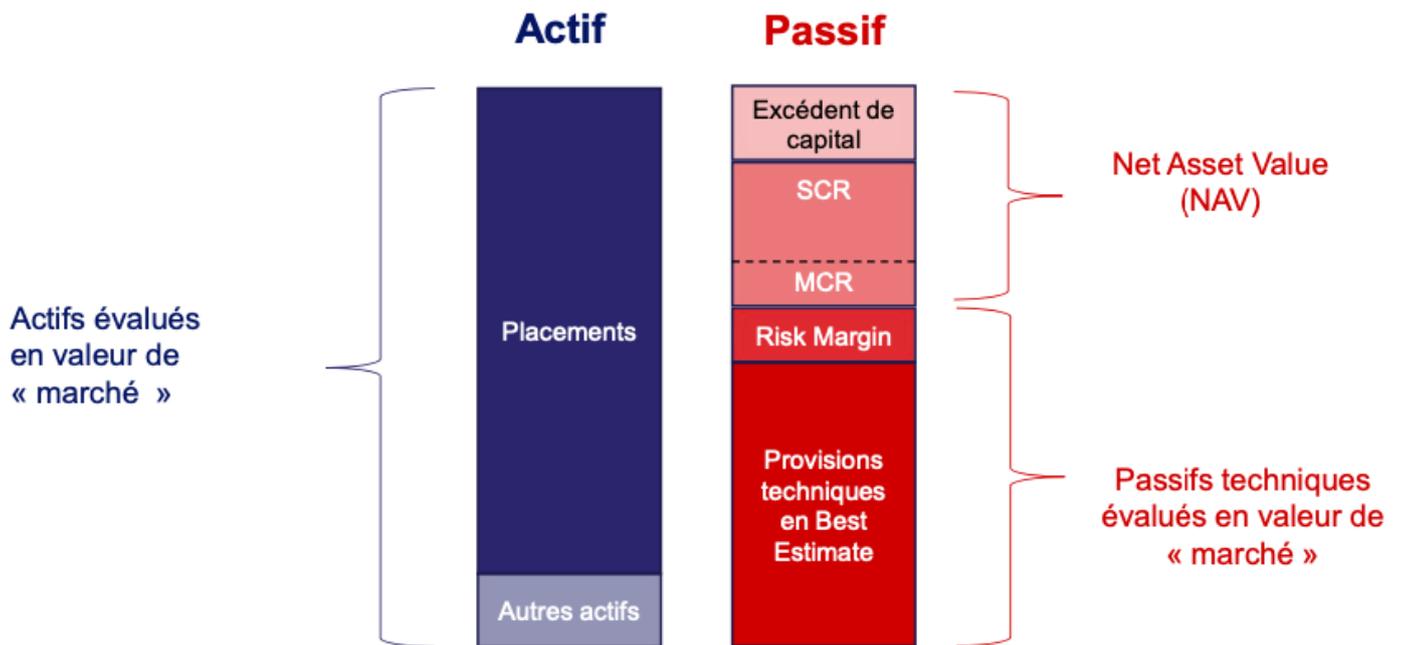


FIGURE 1.2 – Le bilan économique d’une compagnie d’assurance

L’**actif** à l’instant t , noté A_t , représente l’ensemble des placements de l’entreprise en valeur de marché.

Le **passif** se décompose en provisions techniques et en Net Asset Value (NAV_t). Les **provisions techniques** correspondent au prix qu’exigerait un tiers pour reprendre les engagements de l’assureur. Ils sont composés de la valeur économique du passif, noté BEL_t , et de la Risk Margin notée RM_t . La **valeur économique du passif**, aussi appelée Best Estimate Liabilities (BEL), est la meilleure vision économique des engagements de l’assureur. Elle est calculée comme la meilleure estimation des flux futurs des passifs d’assurance. La **Risk Margin** est le coût du capital que devrait lever une entité pour couvrir son exigence en capital jusqu’à l’extinction des passifs.

La **Net Asset Value**, autre composante du passif, correspond aux capitaux disponibles sous Solvabilité II. Il s’agit de la valeur économique de l’entreprise. Elle se calcule comme la différence entre le total d’actifs et le total de provisions techniques en valeur de marché. La Net Asset Value est formée des deux indicateurs de solvabilité du pilier 1 que sont le SCR et le MCR ainsi que d’un excédent de capital.

Afin de calculer le SCR qui est basé sur l’évaluation économique du bilan, Solvabilité II offre deux possibilités aux compagnies d’assurance.

La première possibilité est de recourir à une **formule standard** délivrée par le régulateur. La seconde, plus chronophage et mobilisant plus de moyens, est d’utiliser un **modèle spécifique développé en interne** et qui sera donc plus adapté au profil de risque de la compagnie. La figure 1.3 synthétise les inconvénients et les avantages liés à l’utilisation de ces deux possibilités.

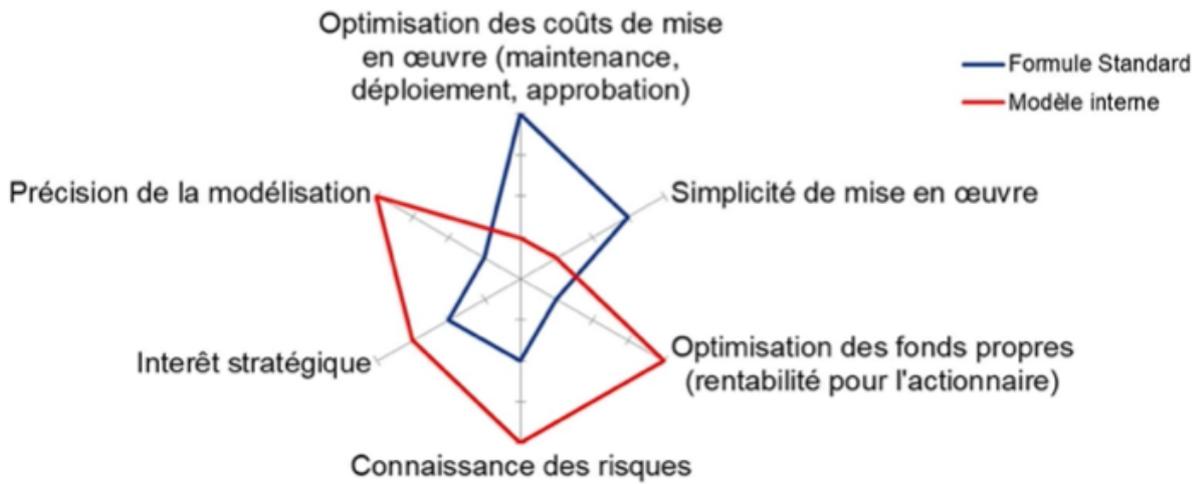


FIGURE 1.3 – Les avantages et inconvénients de la formule standard et du modèle interne

1.2.3 Le SCR en formule standard

La formule standard propose une méthode de calcul simplifiée du SCR. Cette méthode est donnée par le régulateur.

La formule du SCR est la suivante [2] :

$$SCR = BSCR + Ajustements + SCR_{Op} \quad (1.3)$$

où

- $BSCR$ est le capital requis de base
- $Ajustements$ représente les ajustements au titre de la capacité d'absorption des pertes
- SCR_{Op} est le chargement en capital au titre du risque opérationnel.

Comme illustré sur la figure 1.4, la formule standard se base sur une **approche modulaire** : un calcul de besoin en capital est effectué individuellement pour chaque module de risque. Le SCR global étant obtenu en agréant l'ensemble des SCR par risque selon la formule (1.3).



FIGURE 1.4 – La structure modulaire de la formule standard

Le BSCR

Chacun des SCR des 6 modules de risque présents sur la figure 1.4 est calibré pour que la probabilité de ruine à horizon 1 an soit égale à 0.5%.

Quatre de ces 6 modules sont scindés en sous-modules. Un **SCR sous-modulaire** est calculé pour chacun de ces sous-modules. Pour ce faire, ce SCR sous-modulaire est calculé comme l'impact d'un choc sur la NAV :

$$SCR_{\text{sous-module}} = \Delta NAV = NAV_{\text{scénario central}} - NAV_{\text{choc}}$$

où la NAV est obtenu tel qu'en chaque instant t , $NAV_t = A_t - BEL_t$, et le choc est donnée par le régulateur.

La figure 1.5 permet de comprendre ce calcul de SCR par la variation du bilan économique à la suite d'un choc. Pour cette figure, l'exemple du sous-module *action* appartenant au module *Marché* est présenté. L'occurrence du risque de marché *action* se traduit ici par la baisse des actions et donc la baisse de la valeur de l'actif. En assurance vie, le passif absorbe une partie de la perte subie à l'actif. Au total, la NAV de ce scénario choqué baisse par rapport au scénario central, ce qui permet d'en déduire le SCR du sous-module *action*.

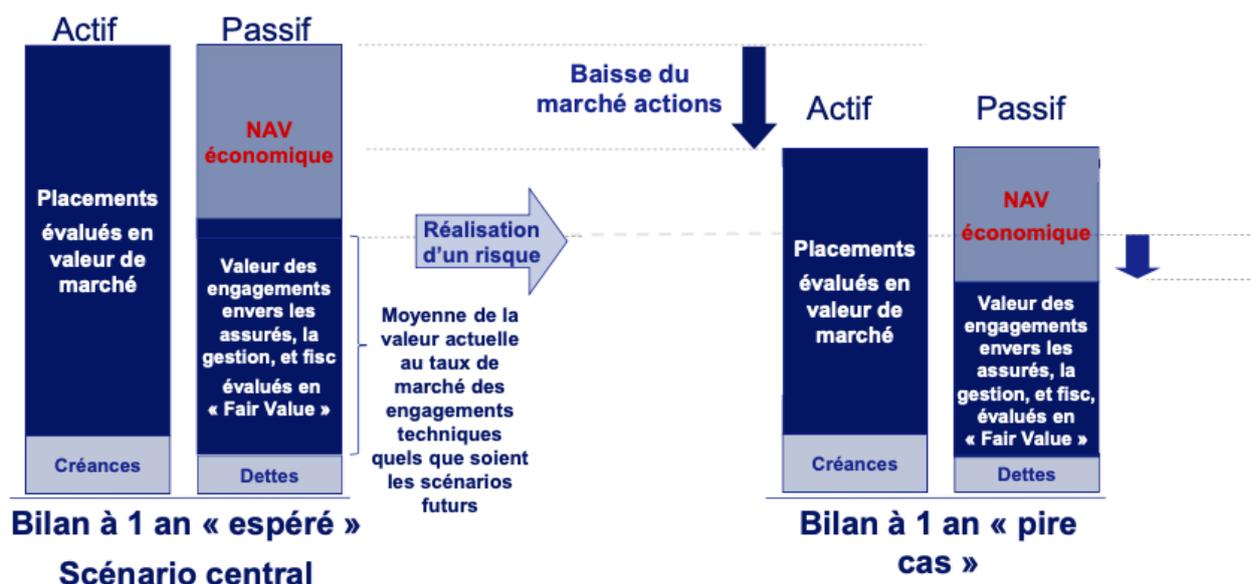


FIGURE 1.5 – La variation du bilan économique à la suite d’un choc action à la baisse

Ensuite, une fois les SCR sous-modulaire calculés, l’**effet de diversification intra-modulaire** est pris en compte par une agrégation des capitaux via une matrice de corrélation. Ainsi, pour les modules *Non Vie*, *Vie*, *Santé* et *Marché*, la formule du **SCR modulaire** est obtenue en tenant compte de la diversification :

$$SCR_{module} = \sqrt{\sum_{i,j} Corr_{i,j} \times SCR_{sous-module\ i} \times SCR_{sous-module\ j}} \quad (1.4)$$

où

- $SCR_{sous-module\ i}$ et $SCR_{sous-module\ j}$ sont les SCR des sous-modules d’un même module
- $Corr_{i,j}$ est le coefficient de la matrice de corrélation fournie par le régulateur.

Une fois l’ensemble des SCR sous-modulaires puis des 6 SCR modulaires obtenu, le **BSCR** peut alors être calculé. De nouveau, une matrice de corrélation est utilisée afin de tenir compte de l’effet de diversification inter-modulaire.

Pour tout $k, l \in \{Non\ Vie, Vie, Sante, Marche, Credit, Actifs\ Incorporiels\}$:

$$BSCR = \sqrt{\sum_{k,l} Corr_{k,l} \times SCR_k \times SCR_l}. \quad (1.5)$$

Les ajustements

Les **ajustements** permettent de tenir compte de la capacité de l’assureur à absorber les pertes futures via les mécanismes de participation aux bénéfices avec les assurés ou via les impôts différés.

Le SCR Opérationnel

Le **module opérationnel** a pour objectif de capter le risque de pertes résultant de procédures internes, de membres du personnel ou de systèmes inadéquats ou défectueux, ou encore d’événements extérieurs.

$$SCR_{Op} = \min(0.3 \times BSCR, Op) + 0.25 \times Exp_{ul} \quad (1.6)$$

où

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

Le calcul du SCR est réalisé au moins **une fois par an** mais son suivi doit être fait en continu.

1.2.4 Le SCR associé au risque de marché

Dans ce mémoire, le module étudié est le module du risque de *marché*. C'est un des principaux risques auquel est confrontée une compagnie d'assurance vie. En effet, du fait de l'inversion du cycle de production, les assureurs investissent les primes perçues dans des actifs financiers. Ces derniers doivent respecter des contraintes de rentabilité, de liquidité et de sûreté de placement. Les assureurs sont par conséquent exposés aux aléas des marchés financiers.

Le risque de marché est défini comme étant le risque d'une **perte potentielle de valeur du portefeuille sur un horizon donné**. Cette perte est engendrée par une détérioration de l'environnement financier.

Le risque de marché est composé des **6 facteurs de risque** suivants [9] :

- Le risque de **taux d'intérêt** : c'est le risque lié à la sensibilité aux changements affectant la courbe des taux d'intérêt ou la volatilité des taux d'intérêt
- le risque sur **actions** : c'est le risque lié à la sensibilité aux changements affectant le niveau ou la volatilité des actions
- le risque sur **actifs immobiliers** : c'est le risque lié à la sensibilité aux changements affectant le niveau ou la volatilité des actifs immobiliers
- le risque de **spread** : c'est le risque lié à la sensibilité aux changements affectant le niveau ou la volatilité des spreads de crédit par rapport à la courbe des taux d'intérêt sans risque
- le risque **monétaire** : c'est le risque lié à la sensibilité aux changements affectant le niveau ou la volatilité des taux de change
- le risque de **concentration** : c'est l'ensemble des risques liés à un manque de diversification du portefeuille d'actifs, ou à une exposition importante au risque de défaut d'un seul émetteur de valeurs mobilières ou d'un groupe d'émetteurs liés.

La difficulté du calcul du SCR de marché en assurance vie réside dans l'**interaction actif-passif** et nécessite donc l'utilisation d'un modèle ALM. Il faut pouvoir valoriser le passif dans un mouvement d'actif. Cette interaction est principalement due à la capacité de l'assureur à modifier les taux servis aux assurés, ce qui affecte le BEL.

1.3 Les différentes méthodes de calcul du SCR de marché en modèle interne

Afin de calculer le SCR, Axa a fait le choix du modèle interne. Un modèle interne est un modèle propre à l'entreprise soumis à l'approbation des autorités de contrôle. Il permet de réaliser une modélisation personnalisée du portefeuille. Le SCR calculé est alors fonction des risques réellement supportés par Axa et non plus basé sur l'approche forfaitaire de la formule standard.

Pour calculer le SCR de marché, la distribution de la NAV à horizon 1 an est nécessaire pour

obtenir le quantile à 0.5%. Contrairement à l'actif, le BEL n'est pas calculable par formules fermées. Cela rend complexe la détermination de la distribution de la NAV. Plusieurs méthodes permettent néanmoins d'approximer le BEL, celles-ci sont présentées dans cette section.

1.3.1 La méthode des Simulations dans les Simulations (SdS)

La méthode des simulations dans les simulations est l'approche classique concernant le calcul du SCR de marché. L'objectif est de calculer la distribution de la NAV après simulation d'un certain nombre de facteurs de risques financiers ou techniques. Une fois la distribution obtenue, il est alors facile d'obtenir le quantile à 0.5% [2], [7], [14], [26].

Cette méthode se scinde en deux parties :

1. La **prédiction** des variables économiques (taux, actions, ...) entre les dates $t = 0$ et $t = 1$
2. L'**évaluation** du bilan économique conditionnellement à la réalisation des variables économiques.

Il est important de rappeler qu'en assurance vie, il existe des interactions entre l'actif et le passif du bilan économique. Le passif doit être évalué dans un mouvement d'actif. L'utilisation d'un modèle ALM est alors nécessaire.

Evaluation monde réel

La prédiction des variables économiques entre les dates $t = 0$ et $t = 1$ doit être réalisée sous la probabilité dite *monde réel*. Sous cette probabilité, le comportement historique et statistique des grandeurs financières est représenté le plus fidèlement possible.

Une simulation *monde réel* correspond à un ensemble de prédictions de variables économique. Un univers économique réel est ainsi modélisé.

Axa utilise 11000 simulations *monde réel* qui représentent donc $P = 11000$ trajectoires d'ensemble de variables économiques qui reflètent l'ensemble des conditions de marché étudié.

Les simulations *monde réel* sont appelées **simulations primaires** dans la méthode des Simulations dans les Simulations. Elles servent à obtenir une distribution empirique.

Evaluation risque neutre

Chacune des $P = 11000$ simulations primaires engendre un bilan économique stressé. L'actif de ce bilan stressé s'obtient directement par une valorisation à l'aide des prix disponibles sur le marché (*mark-to-market*) ou de formules fermées (*mark-to-model*).

Cependant, concernant les provisions techniques, les interactions actif passif issues des mécanismes de participation aux bénéficiaires et des comportements de rachats dynamiques des assurés ne permettent pas une approche *mark-to-model* par formules fermées [7].

Ainsi, afin de valoriser le passif du bilan, pour chaque simulation primaire, des simulations dites **secondaires** sont nécessaires. Elles sont réalisées en univers *risque neutre* et utilisent le modèle ALM pour tenir compte des interactions actif passif.

La valorisation *risque neutre* consiste à simuler les flux futurs d'un titre de manière *market consistent* et à considérer la moyenne de ces flux actualisés au taux sans risque. En effet, en l'absence d'opportunité d'arbitrage, les processus de prix évoluent en moyenne au taux sans risque sous la probabilité *risque neutre*.

Le modèle interne d'Axa utilise $S = 4000$ simulations secondaires avec un horizon de $T = 60$ ans pour chacune d'entre elle.

La figure 1.6 [7] schématise le fonctionnement de la méthode des Simulations dans les Simulations.

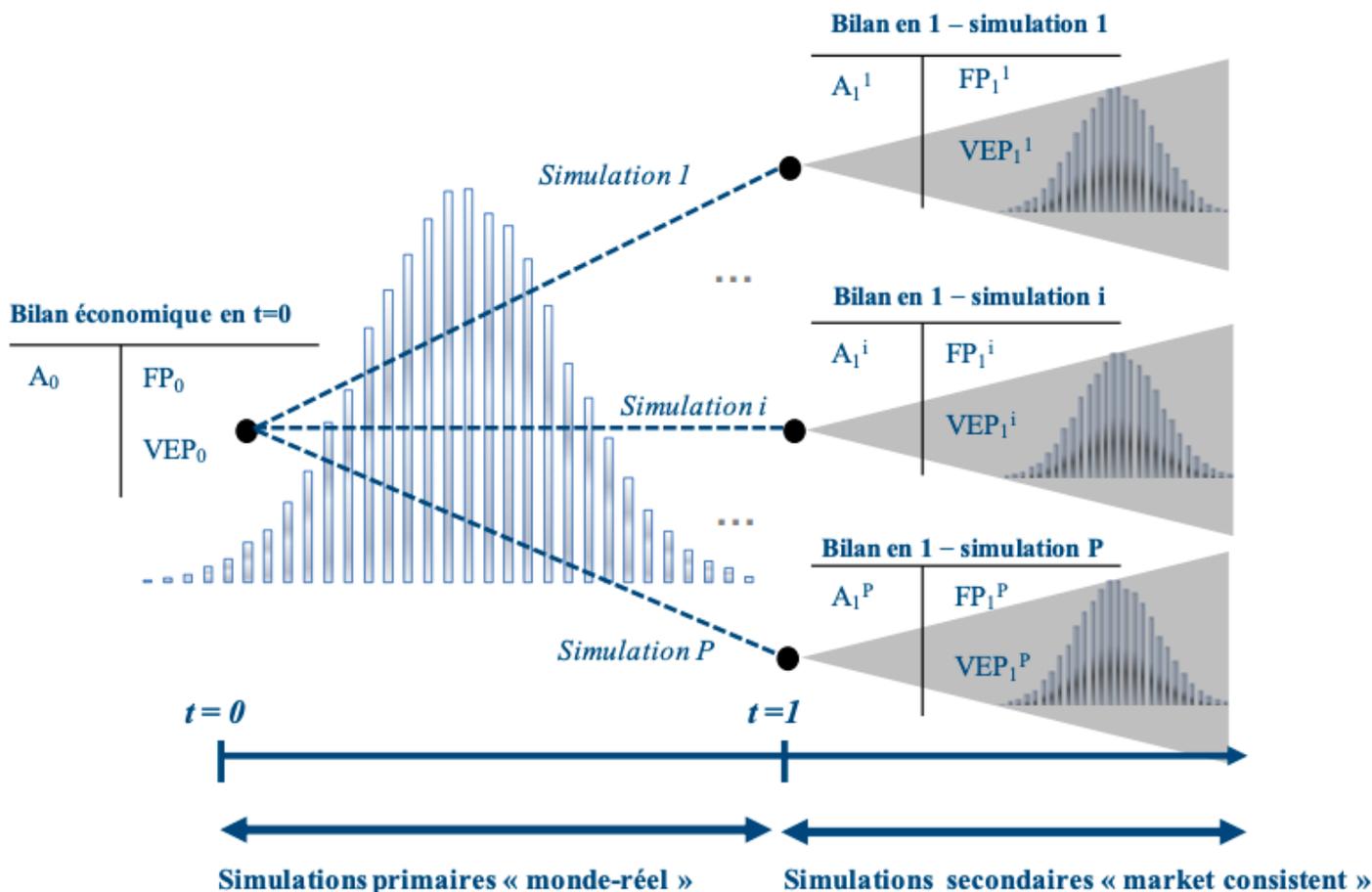


FIGURE 1.6 – La méthode des SdS

Bien que la méthode des simulations dans les simulations soit précise, sa **complexité** en $P \times S = 44$ millions de simulations dans le cas d'Axa mène à des temps de calcul irréalisables. En effet, pour une simulation primaire considérée, environ 4 heures sont nécessaires pour effectuer les $S = 4000$ simulations secondaires. Il faudrait alors $11000 \times 4 = 44000$ heures afin d'avoir la totalité des simulations de cette méthode. Cela représente **5 ans de calcul**, ce qui n'est pas envisageable en pratique.

Pour répondre à la problématique du temps de calcul très lourd imposé par la méthode des simulations dans les simulations, d'autres méthodes peuvent être employées.

1.3.2 L'accélérateur SdS

La méthode des Simulations dans les Simulations permet d'obtenir la distribution de la Net Asset Value afin d'en retirer le quantile à 0.5%. Néanmoins, seul ce quantile à 0.5% est indispensable au calcul du SCR de marché.

Partant de ce constat, il n'est pas nécessaire d'avoir la distribution complète de la NAV : seule la

queue de distribution est utile. Réduire le nombre de simulations primaires va donc réduire significativement le temps de calcul.

Dans le cas d'Axa, le quantile à 0.5% est la 55^{eme} pire valeur car le nombre de simulation primaire est $P = 11000$.

Devineau et Loisel [7] présentent les 4 étapes de l'accélérateur SdS :

1. La **construction de facteurs de risques élémentaires** (actions, taux, mortalité) permettant de traduire l'intensité du risque de chaque simulation primaire. La norme associée est également déterminée.
2. La définition d'une **région d'exécution** associée à un seuil fixé. C'est l'appartenance à cette région d'exécution qui va permettre de lancer une simulation primaire.
3. **Itérations sur le seuil** de la région de sorte à intégrer à chaque étape un nombre supplémentaire de scénarios à exécuter.
4. Le critère d'arrêt est le suivant : **l'algorithme s'achève lorsque les $N = 55$ (dans le cas d'AXA) pires valeurs de NAV sont stabilisées.**

La figure 1.6 [7] illustre la méthode.

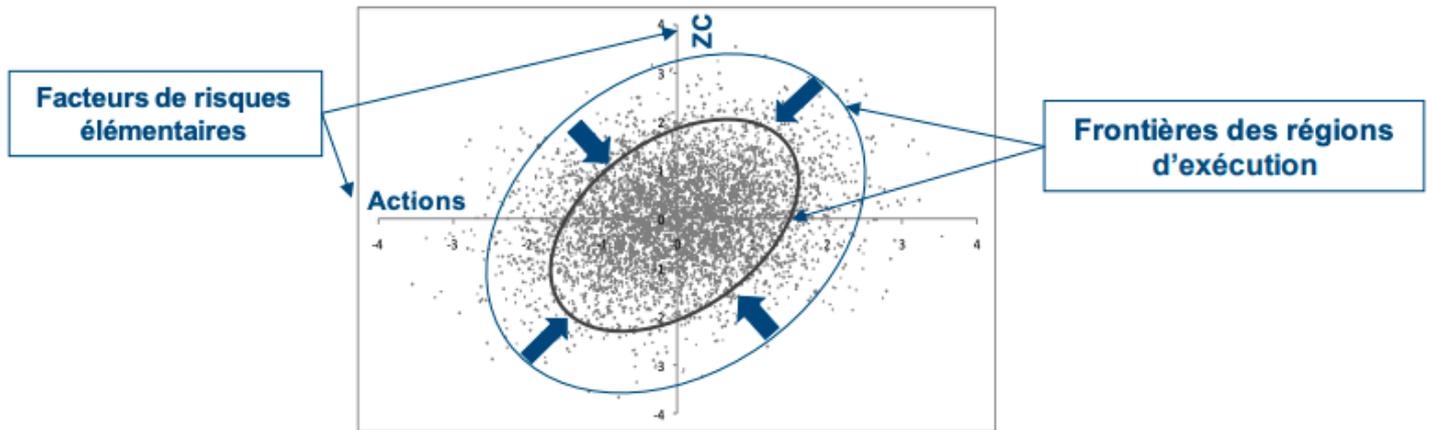


FIGURE 1.7 – La méthode accélérateur SdS

L'accélérateur SdS n'est pas la méthode favorisée par Axa. Cette méthode repose en effet sur des **hypothèses fortes de distribution** des facteurs de risque pour réduire le nombre de scénario primaires à exécuter. Ces hypothèses portent en général sur la normalité de la distribution.

De plus, par définition de la méthode, la distribution totale n'est pas à disposition.

1.3.3 Le Curve fitting

La méthode curve fitting est une autre méthode visant à réduire le nombre de simulations primaires. Le but de cette méthode est d'**utiliser un nombre limité de points issus des simulations primaires et d'interpoler la valeur du passif**. Cette dernière étant perçues comme une fonction complexe d'un nombre limité de facteurs de risques définis par les scénarios économique *monde réel*. Deux scénarios identiques conduisent alors à une même valorisation du BEL.

L'objectif de la méthode curve fitting est de ne pas réaliser les $S = 4000$ simulations secondaires pour plusieurs valeurs de BEL qui seraient proches.

La méthode curve fitting suit les grandes étapes suivantes [18] :

1. Le **choix des scénarios primaires retenus** doit être effectué dans l'optique de couvrir les risques pour lesquels la charge en capital doit être évaluée.
2. L'**interpolation** est alors réalisée sur ce nombre réduit de points $N \ll P = 11000$. Plusieurs méthodes sont possibles pour l'interpolation, par exemple polynomiale ou par morceaux.
3. L'utilisation de la **fonction d'interpolation** en étape 2. permet alors une **estimation rapide du BEL**.
4. Le **calcul de la distribution** de la Net Asset Value permet d'obtenir le **quantile à 0.5%**.

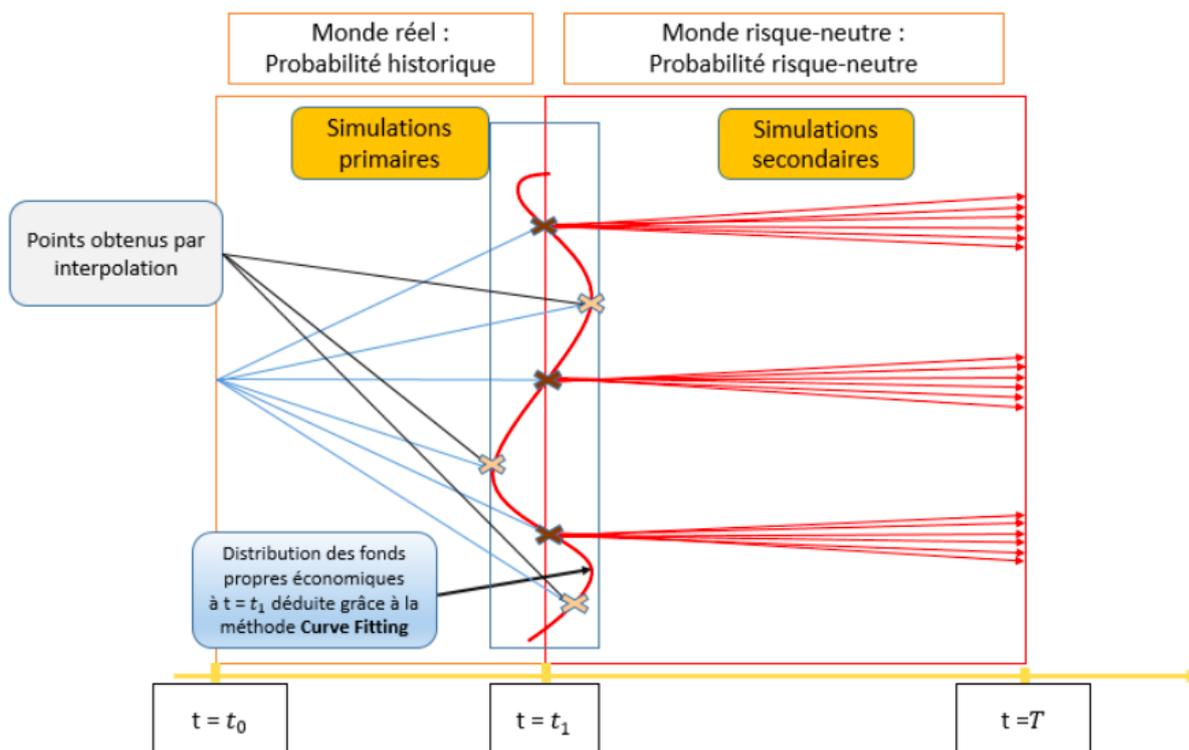


FIGURE 1.8 – La méthode curve fitting

Il s'agit d'une solution assez rapide, qui repose principalement sur la **régularité de la fonction d'interpolation**. Néanmoins, cette méthode n'a pas été retenue par Axa pour le calcul du SCR de marché car elle nécessite d'effectuer de **nombreuses calibrations** (par exemple une pour le risque de taux d'intérêt, une autre pour le risque actions ...).

En pratique, des calibrations sont nécessaires pour tout facteur de risque de marché pris en compte dans le calcul du SCR de marché.

Il est également très difficile dans la pratique d'effectuer une calibration lorsque de nombreux risques surviennent en même temps.

1.3.4 Le Least Square Monte Carlo

La méthode du Least Square Monte Carlo (LSMC) vise cette fois-ci à réduire le nombre de simulations secondaires.

$P = 11000$ simulations primaires sont réalisées. Cependant, au lieu d'exécuter $S = 4000$ simulations secondaires afin d'évaluer le passif du bilan, **seules quelques simulations secondaires sont utilisées** ($N \ll S = 4000$). Pour chacun des scénarios primaires, ces quelques simulations secondaires donnent une évaluation imprécise du passif. Cette inexactitude est toutefois

corrigée en effectuant une **régression à travers ces évaluations inexactes**. La courbe de régression ainsi obtenue est alors utilisée pour approximer le passif. La figure 1.9 [18] schématise cette approche.

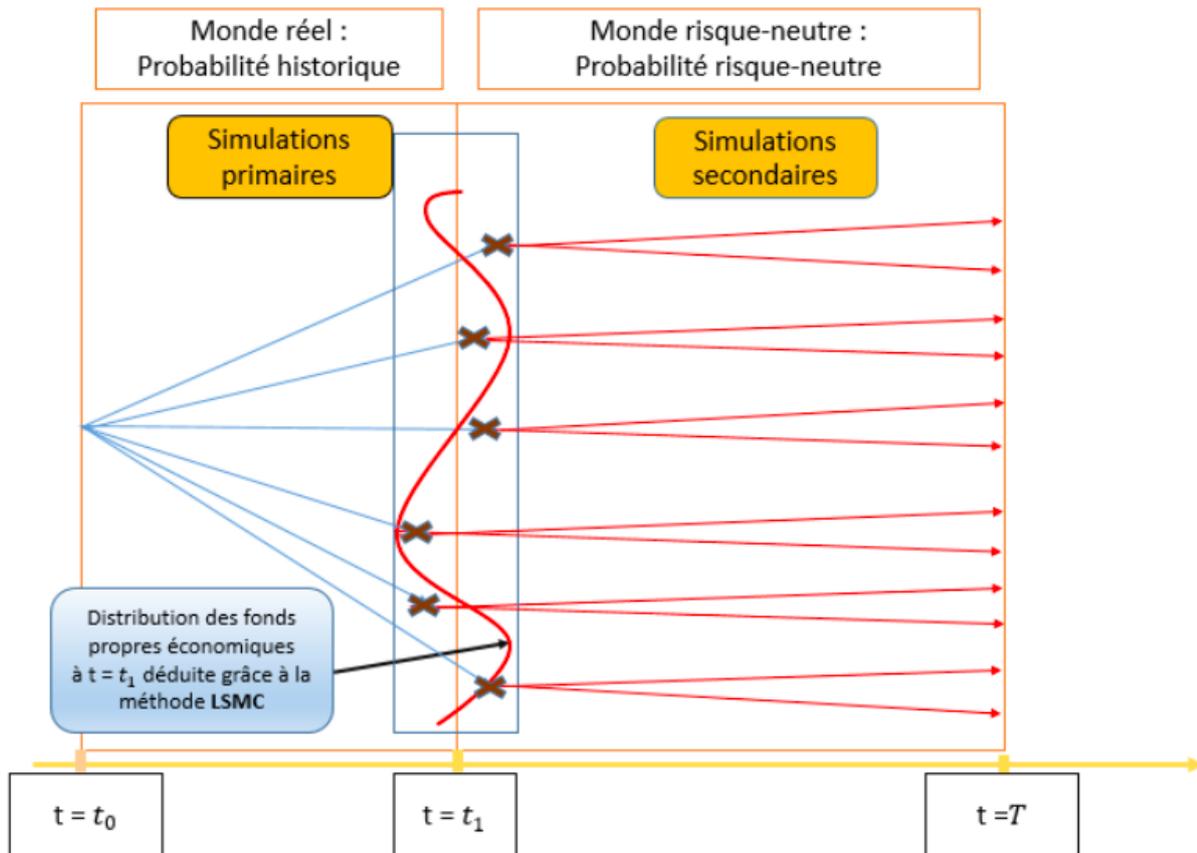


FIGURE 1.9 – La méthode LSMC

1.3.5 Le Portefeuille de Réplication

Tout comme la méthode LSMC, la méthode des portefeuilles répliquants permet de réduire le nombre de simulations secondaires.

C'est la méthode retenue par Axa afin de calculer le SCR de marché. Cette méthode utilise un **ensemble simple d'instruments financiers afin de répliquer le comportement du BEL**. Cela permet de ne pas avoir recours aux $S = 4000$ simulations secondaires, ce qui représente un gain de temps considérable.

Les raisons qui ont poussées Axa à retenir cette méthode, sa description ainsi que les éléments indispensables à sa mise en place sont présentés en détail dans le chapitre suivant.

LA MÉTHODE DES PORTEFEUILLES RÉPLIQUANTS

Afin de répondre à la contrainte du SCR imposé par la réglementation Solvabilité II, Axa a fait le choix d'utiliser la méthode des portefeuilles de réplication pour calculer le SCR de marché en modèle interne. Cette méthode est présentée en détail dans ce chapitre.

2.1 Les motivations d'AXA

La méthode des portefeuilles répliquants est la méthode privilégiée par Axa pour le calcul du SCR de marché en modèle interne. Cette méthode a pour objectif de **répliquer le Best Estimate Liabilities à l'aide d'instruments financiers simples à valoriser**. Le portefeuille déterminé doit posséder [6], [24] :

1. la même valeur de marché que le BEL
2. la même sensibilité que ce dernier face aux différents facteurs influençant le marché.

De ce fait, la réponse du portefeuille répliquant face à un mouvement de marché permet d'avoir une approximation du comportement et de la valeur du BEL.

Contrairement à la méthode des simulations dans les simulations, une fois le portefeuille répliquant déterminé, il n'est plus nécessaire de procéder aux simulations secondaires en monde risque neutre. C'est ainsi le calcul de $S = 4000$ simulations secondaires qui sont évitées pour chacune des $P = 11000$ simulations primaires. La durée de calcul du SCR de marché en modèle interne est donc significativement réduite. La première motivation de l'utilisation des portefeuilles répliquants est donc un **gain de temps considérable**.

Ensuite, la méthode des portefeuilles de réplication permet la **représentation exacte du passif** pour toutes les branches d'activité **P&C** ainsi que toutes les branches **Vie pour lesquelles il n'y a pas de rachats dynamiques par exemple**. Un exemple est donné en Annexe C.

Lorsque les branches d'activité Vie présentent des rachats dynamiques et des participations aux bénéficiaires, cette méthode est une **approximation du passif**. La qualité de la représentation peut et doit toutefois être testée.

Par l'utilisation d'instruments financiers simples, cette méthode fournit également une **explication économique de la structure du passif**. Cette explication économique permet alors :

- de donner un aperçu de l'activité et des risques financiers
- d'aider à concevoir des stratégies de couverture
- d'aider à concentrer la calibration du générateur de scénario économique (ESG) sur les instruments financiers les plus pertinents
- de remettre en question les résultats du système de projection.

On s'éloigne ainsi de l'aspect *boîte noire* que présentent les autres méthodes de calcul.

Enfin, les portefeuilles répliquants peuvent être utilisés pour recalculer le passif dans de **nouvelles conditions de marché** de manière **très rapide** (la fréquence pourrait être quotidienne si nécessaire).

Cela est très pratique pour évaluer l'impact des changements de modèle d'une part, et effectuer des stress tests ou des calculs rapides de roll-forward trimestriels d'autre part.

C'est pour l'ensemble de ces raisons que la méthode des portefeuilles répliquants a été retenue par Axa pour le calcul du SCR de marché en modèle interne. Il est cependant indispensable de rappeler que cette méthode est une approximation du BEL pour la branche d'activité Vie avec rachats dynamiques et participations aux bénéfices. Plusieurs mesures d'*accuracy* de la méthode doivent ainsi être mises en place.

2.2 Les données nécessaires à l'application de la méthode

La méthode des portefeuilles de réplication consiste à utiliser une combinaison d'instruments financiers simples afin d'approximer le passif d'assurance. Par conséquent, les caractéristiques du passif d'assurance seront naturellement **liées** à la nature ou aux caractéristiques des instruments financiers utiles à la réplication : la nature et les caractéristiques d'un instrument sont liées à des options de contrat comme des garanties ou des participations aux bénéfices par exemple.

Il est donc primordial de comprendre les caractéristiques du passif à répliquer afin de sélectionner les instruments pertinents lors d'une sélection dite *à dire d'expert*. Dans le cas d'une sélection automatisée, ce qui est l'objet de ce mémoire, cette compréhension est importante afin de pouvoir fournir une explication économique du passif en fonction des instruments financiers retenus et réciproquement, pour légitimer leur sélection.

Différents instruments peuvent être utilisées afin de répliquer le passif [23].

2.2.1 Les zéros-coupons

Des zéros-coupons de différentes maturités peuvent être intégrés dans un portefeuille de réplication.

Le détenteur d'un zéro-coupon de nominal 1 et de maturité t_1 recevra 1 unité de monnaie en date t_1 de façon certaine :

$$CF(t) = \begin{cases} 1 & \text{quand } t=t_1, \\ 0 & \text{sinon.} \end{cases}$$

Sur le graphique 2.1 est donné un exemple de rémunération d'un zéro-coupon de maturité $t_1 = 10$ ans.

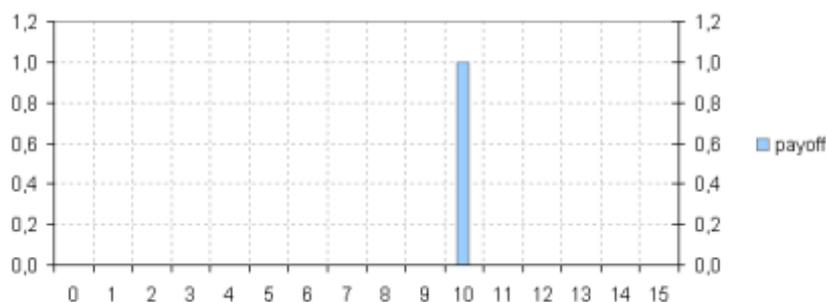


FIGURE 2.1 – La rémunération d’un zéro-coupon de maturité 10 ans

Les zéro-coupons sont utilisés pour **répliquer les passifs qui sont indépendants des actifs investis et du marché** : les flux de trésorerie sont certains. Ils représentent ainsi les primes d’assurance, les sinistres à payer et les garanties accordées aux assurés.

2.2.2 Le zéro-coupon inflation

Le détenteur d’un zéro-coupon inflation de nominal 1 et de maturité t_1 recevra une unité de monnaie multiplié par la valeur de l’inflation en date t_1 :

$$CF(t) = \begin{cases} \prod_{i=1}^{t_1} (1 + CPI(i)) & \text{quand } t=t_1, \\ 0 & \text{sinon} \end{cases}$$

où $CPI(i)$ est l’indice du prix à la consommation en date i .

Le zéro-coupon inflation usuellement utilisé dans la méthode des portefeuilles répliquants est le zéro-coupon inflation de maturité $t_1 = 10$ ans dont la rémunération est présentée figure 2.2.

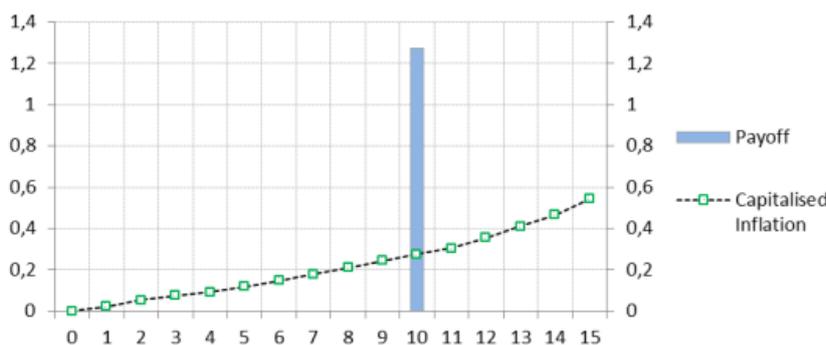


FIGURE 2.2 – La rémunération d’un zéro-coupon inflation de maturité 10 ans

L’utilisation du zéro-coupon inflation de maturité 10 ans permet de **réfléter la sensibilité du passif aux chocs sur l’inflation**.

2.2.3 Les swaptions receveur et payeur

Le détenteur d’un swaption receveur de maturité t_1 , de tenor t_2 et de strike K a le droit de conclure, en date t_1 , un swap receveur de maturité t_2 pour lequel il paierait le taux variable et recevrait le taux fixe.

Il n'exercera l'option que si les conditions garanties sont plus intéressantes que les conditions du moment :

$$CF(t) = \begin{cases} \max(K - S(t_1, t_2), 0) & \text{quand } t_1 + 1 \leq t \leq t_1 + t_2, \\ 0 & \text{sinon.} \end{cases}$$

La figure 2.3 illustre un swaption receveur de maturité $t_1 = 5$, de tenor $t_2 = 5$ et de strike $K = 3\%$.

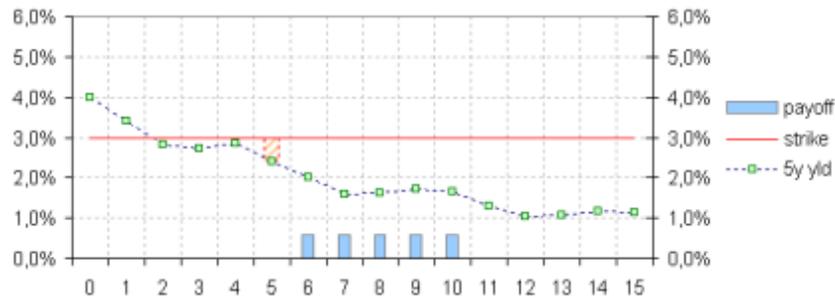


FIGURE 2.3 – La rémunération d'un swaption receveur

Le détenteur d'un swaption payeur de maturité t_1 , de tenor t_2 et de strike K a le droit de conclure, en date t_1 , un swap payeur de maturité t_2 pour lequel il recevrait le taux variable et paierait le taux fixe.

Il n'exercera l'option que si les conditions garanties sont plus intéressantes que les conditions du moment :

$$CF(t) = \begin{cases} \max(S(t_1, t_2) - K, 0) & \text{quand } t_1 + 1 \leq t \leq t_1 + t_2, \\ 0 & \text{sinon.} \end{cases}$$

La figure 2.4 illustre un swaption payeur de maturité $t_1 = 5$, de tenor $t_2 = 5$ et de strike $K = 3\%$.

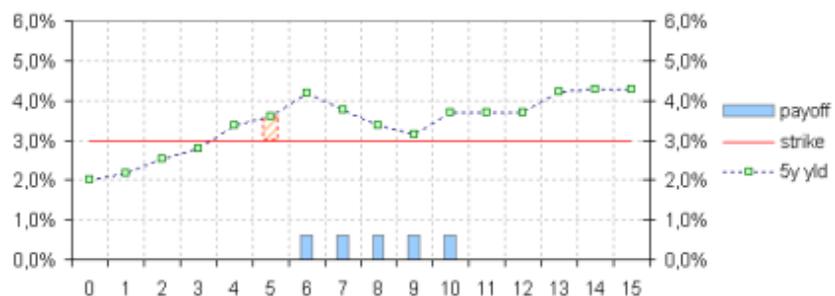


FIGURE 2.4 – La rémunération d'un swaption payeur

Les swaptions receveur et payeur permettent de **représenter les options offertes au titulaire d'une police d'assurance**. Ils servent également à **imiter les comportements de rachats dynamiques**.

2.2.4 Les indices equity

Le détenteur d'un indice equity de maturité t_1 et de tenor t_2 tient une position directe dans le base index sous-jacent qu'il vend progressivement entre $t_1 + 1$ et $t_1 + t_2$.

Il réalise alors des pertes et des profits. En notant $S(t)$ la *market value* de l'actif sous-jacent :

$$CF(t) = \begin{cases} \frac{S(t)}{t_2} & \text{quand } t_1 \leq t \leq t_1 + t_2 - 1, \\ 0 & \text{sinon.} \end{cases}$$

La figure 2.5 illustre un indice equity de maturité $t_1 = 5$ et de tenor $t_2 = 5$.

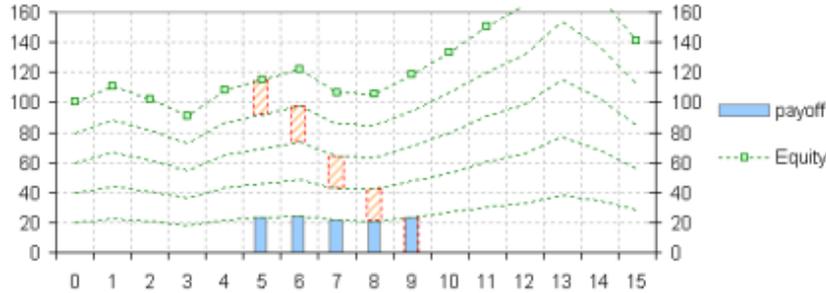


FIGURE 2.5 – La rémunération d'un indice equity

Les indices equity servent à **répliquer les expositions directes à différentes classes de risque**. Pour répliquer ces expositions, il est nécessaire de définir des indices adaptés. Ces indices equity permettent ainsi d'élargir le champ des facteurs de risque modélisés au sein du portefeuille de réplication.

Six indices equity sont utilisés par Axa [23] :

- l'indice *AllToRet_All* : composé à 100% d'equities, majoritairement des *listed equities* ou *private equities* selon les portefeuilles
- l'indice *CorpMort_All* : composé de *corporate spread* et de *mortgages*
- l'indice *QGovtGovt_All* : composé d'obligations souveraines et d'obligations détenues par des entreprises d'Etat
- l'indice *RE_All* : composé de 100% de Real Estate
- l'indice *Fake_All* : regroupe les compositions de *AllToRet_All*, *CorpMort_All* et de *RE_All* afin de capturer des effets croisés
- l'indice *AllSpreads_All* : regroupe les compositions de *CorpMort_All* et *QGovtGovt_All* afin de capturer des effets croisés.

Les indices equity sont utilisés pour élargir le champ des risques modélisés dans les portefeuilles répliquants et pour obtenir une meilleure adéquation avec l'allocation réelle des actifs.

2.2.5 Les call et put sur equity

Le détenteur d'une option call sur equity de maturité t_1 et de strike K a le droit d'acheter l'equity sous-jacent en date t_1 au prix K . En notant $S(t)$ la *market value* de l'actif sous-jacent :

$$CF(t) = \begin{cases} \max(S(t) - K, 0) & \text{quand } t=t_1, \\ 0 & \text{sinon.} \end{cases}$$

La figure 2.6 illustre un call sur equity de maturité $t_1 = 10$ et de strike $K = 120\%$.

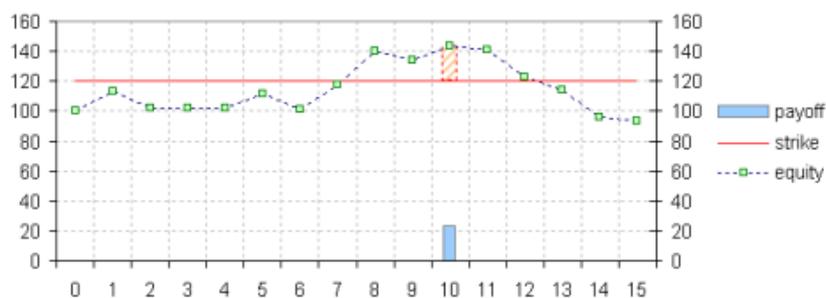


FIGURE 2.6 – La rémunération d'un call sur equity

Le détenteur d'une option put sur equity de maturité t_1 et de strike K a le droit de vendre l'equity sous-jacent en date t_1 au prix K . En notant $S(t)$ la *market value* de l'actif sous-jacent :

$$CF(t) = \begin{cases} \max(K - S(t), 0) & \text{quand } t=t_1, \\ 0 & \text{sinon.} \end{cases}$$

La figure 2.7 illustre un put sur equity de maturité $t_1 = 10$ et de strike $K = 80\%$.

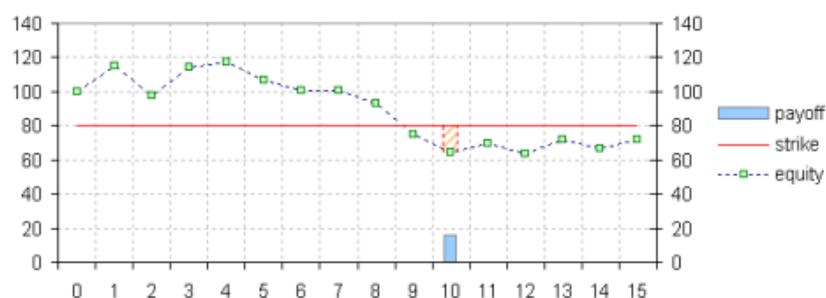


FIGURE 2.7 – La rémunération d'un put sur equity

Les put et call sur equity de maturités et strikes différents sont utilisés pour **capter les comportements non linéaires du passif**. Ils permettent en particulier de **représenter la participation aux bénéfices accordée aux assurés**, par laquelle ces derniers peuvent bénéficier de la hausse tout en étant protégés contre la baisse.

2.2.6 Synthèse des instruments financiers utilisés

L'ensemble des instruments financiers énumérés dans les sous-sections précédentes est synthétisé dans le tableau 2.1 suivant :

Type d'instrument	Effet capturé
Zero-coupon	Garanties
Zero-coupon inflation	Sensibilité à l'inflation
Swaption receveur	Convexité due aux rachats et réinvestissements
Swaption payeur	Convexité due aux rachats et réinvestissements
Indice equity	Exposition directe
Call sur equity	Participations aux bénéfices
Put sur equity	Participations aux bénéfices

TABLE 2.1 – Synthèse des instruments utilisés et les effets qu'ils capturent

Une **base de données de l'étude** est formée par un **ensemble de ces instruments financiers** aux diverses caractéristiques (maturité, strike, tenor). C'est à partir de cette base de données que la méthode des portefeuilles répliquants peut être employée.

2.3 La description de la méthode

La méthode des portefeuilles répliquants consiste à **déterminer un ensemble d'instruments financiers facilement valorisables afin de répliquer le passif d'assurance**. Afin de formaliser le problème d'optimisation associé à cette méthode, il est nécessaire de bien définir les éléments indispensables à l'obtention d'un portefeuille de réplication.

2.3.1 Les définitions et notations

Les différentes définitions et notations utilisées pour décrire formellement la méthode des portefeuilles de réplication sont présentées dans cette sous-section.

Définition : Scénario central (BC)

Le scénario central rassemble les conditions de marché à l'instant $t = 0$ de la valorisation du passif et dans lequel aucun stress n'est appliqué.

Soit $n = 4000$ le nombre de **trajectoires simulées en monde risque neutre**.

Soit $Y^{PV} = \begin{pmatrix} Y_1^{PV} \\ \vdots \\ Y_n^{PV} \end{pmatrix} \in \mathbf{R}^n$ les n **valeurs présentes du passif à répliquer**.

Chaque composante Y_i^{PV} pour $1 \leq i \leq n$, correspond à une trajectoire simulée en monde risque neutre dans le scénario central.

On a :

$$Y_i^{PV} = \sum_{t=1}^T \delta_{t,i}^{BC} C F_{t,i}^{BC,passif}$$

où

- $\delta_{t,i}^{BC}$ correspond au **déflateur** dans le scénario central à la date t pour une trajectoire risque neutre i ,

- $CF_{t,i}^{BC,passif}$ est le **cash flow du passif** dans le scénario central à la date t pour une trajectoire risque neutre i.

Pour chaque trajectoire risque neutre i, la valeur présente du passif correspond donc à la somme actualisée des flux du passif projetés sur T années ($T = 60$ ans en pratique pour Axa).

Le **BEL** est alors obtenu comme la meilleure estimation de la valeur économique du passif dans le scénario central : c'est la moyenne empirique du vecteur Y^{PV} :

$$BEL^{BC} = \overline{Y^{PV}} = \frac{1}{n} \sum_{i=1}^n Y_i^{PV} = \frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \delta_{t,i}^{BC} CF_{t,i}^{BC,passif}. \quad (2.1)$$

Définition : **Sensibilité (S)**

Une sensibilité est la variation de la valeur du BEL dans un scénario choqué par rapport au BEL du scénario central. Un scénario choqué est issu de la variation d'un ou plusieurs facteurs de risques. Ces chocs sont définis en amont par le comité de contrôle interne.

Soit $Y^S = \begin{pmatrix} Y_1^S \\ \vdots \\ Y_{n_S}^S \end{pmatrix} \in \mathbf{R}^{n_S}$ le vecteur composé des n_S **sensibilités financières du passif**.

En pratique, $n_S \in [[50, 60]]$ pour Axa. Le nombre de sensibilités financières varie en fonction de la nature et de la complexité du portefeuille de passif étudié.

Pour $1 \leq j \leq n_S$:

$$Y_j^S = \left[\frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T \delta_{t,i}^j CF_{t,i}^{j,passif} \right] - BEL^{BC} \quad (2.2)$$

où

- $\delta_{t,i}^j$ correspond au **déflateur** dans le scénario choqué j à la date t pour une trajectoire risque neutre i,
- $CF_{t,i}^{j,passif}$ est le **cash flow du passif** dans le scénario choqué j à la date t pour une trajectoire risque neutre i.

Soit p le **nombre d'instruments financiers** qui compose un portefeuille répliquant candidat. Pour rappel, il faut en pratique tester plusieurs combinaisons d'instruments financiers afin de trouver un portefeuille répliquant qui valide tous les critères de qualité du processus de validation. Deux combinaisons candidates n'auront pas forcément le même p . En pratique, $p \in [[30, 50]]$.

Soit p_{max} le **nombre total d'instruments financiers différents à disposition** pour la composition d'un portefeuille de réplication. C'est la taille de la base de données de l'ensemble des instruments financiers candidats. En pratique, p_{max} est de l'ordre de plusieurs milliers et dépend de la nature et de la complexité du portefeuille de passif à étudier.

Soit $X^{PV} = \begin{pmatrix} X_{1,1}^{PV} & \cdots & X_{1,p}^{PV} \\ \vdots & \ddots & \vdots \\ X_{n,1}^{PV} & \cdots & X_{n,p}^{PV} \end{pmatrix} \in \mathbf{R}^{n \times p}$ la matrice composée des n **valeurs présentes**

dans le scénario central de l'ensemble des instruments financiers candidats. Chaque colonne de la matrice est donc le vecteur des valeurs présentes d'un instrument financier considéré pour chacune des trajectoires risque neutre.

Soit $X^S = \begin{pmatrix} X_{1,1}^S & \cdots & X_{1,p}^S \\ \vdots & \ddots & \vdots \\ X_{n_S,1}^S & \cdots & X_{n_S,p}^S \end{pmatrix} \in \mathbf{R}^{n_S \times p}$ la matrice composée des n_S **sensibilités financières des instruments candidats**.

Comme précédemment, les sensibilités financières sont définies comme la variation entre la valeur présente dans le scénario stressé et la moyenne empirique des valeurs présentes dans le scénario central.

Soit $\beta \in \mathbf{R}^p$ l'ensemble des **nominaux associés aux p instruments candidats** d'un portefeuille répliquant.

Soit SCR_{marche}^{an-1} le **SCR du risque de marché de l'année passée** calculé par modèle interne pour le portefeuille de passif à répliquer.

L'ensemble de ces notations vont permettre de formaliser le problème d'optimisation résultant de la méthode des portefeuilles de réplication.

2.3.2 Les caractéristiques d'un portefeuille répliquant pertinent

Un portefeuille de réplication pertinent doit vérifier plusieurs critères de performances [24].

Tout d'abord, un portefeuille répliquant doit **correctement répliquer les valeurs présentes du passif** pour l'ensemble des trajectoires simulées en monde risque neutre. Cela se traduit par la proximité des vecteurs Y^{PV} et $X^{PV}\beta$.

Ensuite, la qualité d'un portefeuille de réplication est liée à la précision avec laquelle ce portefeuille candidat peut évaluer la valeur de marché du passif dans **différentes conditions de marché**. C'est à l'aide des n_S sensibilités étudiées que le pouvoir prédictif est évalué. Pour toute sensibilité $1 \leq j \leq n_S$, les valeurs Y_j^S et $(X^S\beta)_j$ doivent être comparables. Comme $X^S\beta \in \mathbf{R}^{n_S}$, la notation $(X^S\beta)_j$ correspond à la valeur de la j^{ieme} sensibilité du portefeuille répliquant.

Un portefeuille répliquant doit également éviter d'avoir recours à des instruments en **position acheteuse-vendeuse**. En effet, posséder un call sur equity de maturité t et de strike $K = 120\%$ ainsi qu'un put sur le même equity, de même maturité t et de strike $K = 80\%$ aura un effet compensatoire. Si les nominaux associés à ces positions acheteuses-vendeuses sont conséquents, cela pourrait alors mener à un comportement anormal du portefeuille répliquant lors de l'application de chocs financiers.

Il est d'autant plus important d'éviter ces positions acheteuses-vendeuses que le nombre p d'instruments candidats est limité : $p \in [[30, 50]]$. En effet, un nombre trop important d'instruments financiers au sein d'un portefeuille de réplication complexifie son **interprétation**.

2.3.3 La formalisation du problème

La recherche d'un portefeuille répliquant pertinent doit se faire en accord avec les caractéristiques de performances énoncés dans la sous-section précédente. Le problème d'optimisation sous contraintes permettant l'obtention de portefeuilles candidats est le suivant [12], [24] :

$$\arg \min_{\beta \in \mathbf{R}^p} \|Y^{PV} - X^{PV}\beta\|_2^2 \quad (2.3)$$

$$\text{sous les contraintes} \left\{ \begin{array}{l} R_{adj}^{2,PV} \geq 0.9, \\ \forall 1 \leq j \leq n_S, |Y_j^S - (X^S \beta)_j| \leq 0.02 \times SCR_{marche}^{an-1} \\ \frac{BEL^{BC}}{\sum_{k=1}^p |M_k| \times |\beta_k|} \geq 0.6 \\ F_\epsilon^{BC} \sim \Phi_{\mathcal{N}(0,1)} \\ p \ll p_{max} \end{array} \right.$$

où

- $R_{adj}^{2,PV}$ est le R^2 ajusté de la régression aux moindres carrés $\|Y^{PV} - X^{PV} \beta\|^2$
- F_ϵ^{BC} est la fonction de répartition des résidus de la régression
- $\Phi_{\mathcal{N}(0,1)}$ est la fonction de répartition de la loi normale centrée réduite
- $M \in \mathbb{R}^p$ est le vecteur des valeurs de marché des instruments financiers candidats.

La première contrainte est un **critère d'ajustement**. Elle permet de mesurer la qualité de l'ajustement du portefeuille répliquant sur le portefeuille de passif. Le seuil de 0.9 a fait l'objet d'un choix de la part du comité de contrôle interne.

La seconde contrainte porte sur les sensibilités financières. Elle mesure la **réponse du portefeuille candidat suite à des perturbations du marché**. Le portefeuille est satisfaisant s'il réagit de façon similaire au portefeuille de passif. Le critère retenu par le comité de contrôle interne est l'écart absolu ne dépassant pas 0.02 fois le SCR de marché de l'exercice antérieur calculé en modèle interne.

Le troisième critère porte sur les positions acheteuses-vendeuses. Il permet de **juger la structure globale de la composition du portefeuille de réplication**. Restreindre ces positions acheteuses-vendeuses revient à prévenir des comportements atypiques du portefeuille répliquant lors de l'application de chocs financiers.

La quantité $\frac{BEL^{BC}}{\sum_{k=1}^p |M_k| \times |\beta_k|}$ définit le *BEL ratio* ou *long short ratio*.

La quatrième contrainte est un critère de **normalité des résidus**. Elle est en pratique vérifiée par un test de Kolmogorov-Smirnov de niveau $\alpha = 0.05$. Ce test ainsi qu'une limite à son utilisation sont présentés en Annexe E.

Enfin, la dernière contrainte est une contrainte portant sur le **nombre d'instruments** que doit comporter un portefeuille de réplication candidat. Dans la pratique $p \in \llbracket 30, 50 \rrbracket$.

2.3.4 La résolution pratique du problème d'optimisation

Lors d'une régression, les **variables explicatives sont connues**. La difficulté dans le cas de la méthode des portefeuilles répliquants réside dans le fait que dans la formule (2.3), la matrice X^{PV} est un **paramètre du problème à déterminer**. Par conséquent, le problème d'optimisation tel que présenté dans la section précédente n'est pas résolu en pratique. En effet, Axa a fait le choix de **décomposer** ce problème d'optimisation en **deux sous-problèmes** :

1. la détermination d'un **ensemble d'instruments financiers de nominaux non nuls**
2. le **calcul des nominaux associés** à ces instruments.

Le chapitre suivant présente les méthodes utilisées pour répondre à ces deux problématiques.

LES DIFFICULTÉS DE LA MÉTHODE DES PORTEFEUILLES RÉPLIQUANTS

Le problème d'optimisation sous contraintes présenté dans le chapitre précédent n'est pas directement résolu. Ce problème est décomposé en deux sous problèmes : la détermination d'un ensemble d'instruments financiers de nominaux non nuls d'une part, et le calcul des nominaux qui leur sont associés d'autre part. Le travail effectué dans ce mémoire ne vise pas à modifier la seconde sous-problématique de calcul des nominaux. Seule la méthode de sélection d'instruments financiers, actuellement très chronophage, sera challengée. Ce chapitre présente les méthodes actuellement utilisées pour répondre à cette décomposition du problème initial.

3.1 Le calcul des nominaux d'un portefeuille répliquant

Le calcul des nominaux d'un portefeuille répliquant est la seconde partie du problème à traiter. Cette partie ne fera pas l'objet de modifications et sera **utilisée en l'état dans la suite de l'étude**.

3.1.1 La formalisation du problème

En supposant que l'ensemble des instruments financiers formant le portefeuille candidat soit **déterminé**, c'est à dire que X^{PV} et X^S sont connus, alors la fonction utilisée par Axa permettant de calculer les nominaux d'un ensemble préalablement fixé d'instruments est la suivante [25] :

$$C_{\theta,\lambda,k}(\beta) = (1 - \theta) \frac{\|Y^{PV} - X^{PV}\beta\|_2^2}{n} + \theta \sum_{j=1}^{n_S} k_j \left(Y_j^S - \sum_{k=1}^p X_{j,k}^S \beta_k \right)^2 + \lambda \|\beta\|_2^2 \quad (3.1)$$

où

— $\theta \in [0, 1]$ est un scalaire permettant d'**arbitrer** entre la précision de la répliation dans le scénario central (Y^{PV}) et celle des sensibilités financières (Y^S).

Plus θ se rapproche de 1, plus l'ajustement des sensibilités financières sera considéré.

— $k = \begin{pmatrix} k_1 \\ \vdots \\ k_{n_S} \end{pmatrix}$ tel que $\sum_{j=1}^{n_S} k_j = 1$ est un vecteur de **poinds** permettant d'attribuer une importance différente aux sensibilités financières.

- λ est un **coefficient de pénalisation** permettant de limiter l'amplitude des nominaux pour chacun des instruments du portefeuille répliquant candidat.

Les autres notations sont telles que définies dans le chapitre précédent.

3.1.2 La régression ridge

La fonction de coût de la section qui précède peut se réécrire sous la forme d'une régression ridge :

$$\begin{aligned} C_{\theta,\lambda,k}(\beta) &= (1 - \theta) \frac{\|Y^{PV} - X^{PV}\beta\|_2^2}{n} + \theta \sum_{j=1}^{n_S} k_j \left(Y_j^S - \sum_{k=1}^p X_{j,k}^S \beta_k \right)^2 + \lambda \|\beta\|_2^2 \\ &= (1 - \theta) \frac{\|Y^{PV} - X^{PV}\beta\|_2^2}{n} + \theta \sum_{j=1}^{n_S} \left(\sqrt{k_j} \left(Y_j^S - (X^S \beta)_j \right) \right)^2 + \lambda \|\beta\|_2^2 \\ &= (1 - \theta) \frac{\|Y^{PV} - X^{PV}\beta\|_2^2}{n} + \theta \|\tilde{Y}^S - \tilde{X}^S \beta\|_2^2 + \lambda \|\beta\|_2^2 \end{aligned}$$

avec $\tilde{Y}^S = \sqrt{k} Y^S$ et $\tilde{X}^S = \sqrt{k} X^S$.

Posons $Y = \begin{pmatrix} \sqrt{\frac{1-\theta}{n}} Y^{PV} \\ \sqrt{\theta} \tilde{Y}^S \end{pmatrix} \in \mathbf{R}^{n+n_S}$ et $X = \begin{pmatrix} \sqrt{\frac{1-\theta}{n}} X^{PV} \\ \sqrt{\theta} \tilde{X}^S \end{pmatrix} \in \mathbf{R}^{(n+n_S) \times p}$,

alors :

$$C_{\theta,\lambda,k}(\beta) = \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2.$$

Si $\lambda = 0$, il s'agit d'un problème de type Moindres Carrés Ordinaires.

Si $\lambda \neq 0$, il s'agit d'une régression ridge. Le vecteur des nominaux est alors estimé par

$$\hat{\beta} = \underbrace{(X^t X + \lambda I_p)}_{\Delta}^{-1} X^t Y.$$

La difficulté calculatoire porte sur l'inversion de la matrice $\Delta = X^t X + \lambda I_p$.

La méthode utilisée par Axa pour surmonter cette difficulté est la **factorisation LU** qui est présentée dans la section suivante.

3.1.3 La factorisation LU

La factorisation LU consiste à écrire une matrice carrée Δ en produit de matrice triangulaire inférieure (ne possédant que des 1 sur sa diagonale) L et triangulaire supérieure U :

$$\Delta = LU = \begin{pmatrix} 1 & & & \\ l_{2,1} & 1 & & (0) \\ \vdots & \ddots & \ddots & \\ l_{p,1} & \cdots & l_{p,p-1} & 1 \end{pmatrix} \times \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,p} \\ & u_{2,2} & & u_{2,p} \\ & (0) & \ddots & \vdots \\ & & & u_{p,p} \end{pmatrix}.$$

Si $\Delta = LU$, alors le système $\Delta \hat{\beta} = X^t Y \in \mathbf{R}^p$ peut s'écrire $L(U\hat{\beta}) = X^t Y$.

Si on pose $\hat{\alpha} = U\hat{\beta}$, alors il est possible de déterminer $\hat{\beta}$ en résolvant les deux équations suivantes :

En appliquant la matrice E_k , la k^{ieme} colonne est triangularisée. Cela signifie que par le produit matriciel à gauche, la matrice E_k effectue les opérations sur la colonne k pour les lignes allant de $k + 1$ à p afin d'annuler tous les coefficients qui se situent en dessous du k^{ieme} élément de la diagonale. Les colonnes précédentes, déjà triangularisées, ne sont pas affectées par la matrice E_k .

La matrice supérieure U est ainsi obtenue par la formule :

$$U = \prod_{i=1}^{p-1} E_{p-i} \times \Delta. \quad (3.4)$$

Illustrons par un exemple pour lequel $p = 3$ et $\Delta = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$.

Alors $E_1 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. En effet, afin de retirer le coefficient $\delta_{2,1} = -1$ situé sous le premier élément de la diagonale, l'opération suivante est effectuée :

$$\text{ligne}_2^\Delta \leftarrow \text{ligne}_2^\Delta - \left(-\frac{1}{2}\right) \times \text{ligne}_1^\Delta$$

où $-\frac{1}{2} = \gamma_{2,1}$.

Ainsi $E_1 \times \Delta = \begin{pmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & -1 & 2 \end{pmatrix}$, et $E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{2}{3} & 1 \end{pmatrix}$ car

$$\text{ligne}_3^{E_1 \times \Delta} \leftarrow \text{ligne}_3^{E_1 \times \Delta} - \left(-\frac{2}{3}\right) \times \text{ligne}_2^{E_1 \times \Delta}.$$

La matrice U est finalement obtenue par :

$$U = E_2 \times E_1 \times \Delta = \begin{pmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{pmatrix}.$$

La matrice triangulaire inférieure L

L'objectif est de déduire la matrice triangulaire inférieure L de la matrice triangulaire supérieure U .

La matrice E_k , utilisée pour l'obtention de la matrice U , peut se réécrire :

$$E_k = I_p - \gamma_k \times \iota'_k$$

où

— I_p est la matrice identité de taille p

— $\gamma_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \gamma_{k+1,k} \\ \vdots \\ \gamma_{p,k} \end{pmatrix}$ pour tout $1 \leq k \leq p$

— ι_k est la k^{ieme} colonne de la matrice identité : $\iota_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.

En posant :

$$E_k^{-1} = I_p + \gamma_k \times \iota'_k$$

on a bien :

$$E_k^{-1} \times E_k = E_k \times E_k^{-1} = I_p - \underbrace{(\gamma_k \times \iota'_k)^2}_{0_p} = I_p$$

avec E_k^{-1} une matrice triangulaire inférieure.

Comme

$$\underbrace{(E_{p-1} \dots E_1)^{-1} (E_{p-1} \dots E_1)}_{I_p} \Delta = (E_{p-1} \dots E_1)^{-1} \left(\underbrace{(E_{p-1} \dots E_1) \Delta}_U \right)$$

par associativité du produit matriciel, on a :

$$(E_{p-1} \dots E_1)^{-1} (E_{p-1} \dots E_1) \Delta = \Delta = (E_{p-1} \dots E_1)^{-1} \times U.$$

Ainsi, la matrice triangulaire inférieure L est obtenue par identification :

$$L = (E_{p-1} \dots E_1)^{-1} = E_1^{-1} \dots E_{p-1}^{-1}.$$

Par définition de la matrice E_k^{-1} , la matrice triangulaire inférieure L est égale à :

$$\begin{aligned} L &= E_1^{-1} \times E_2^{-1} \times \dots \times E_{p-1}^{-1} \\ &= (I_p + \gamma_1 \times \iota'_1) \times (I_p + \gamma_2 \times \iota'_2) \times \dots \times (I_p + \gamma_{p-1} \times \iota'_{p-1}) \\ &= I_p + \sum_{i=1}^{p-1} \gamma_i \times \iota'_i \end{aligned}$$

car pour tout $i \leq j$, $(\gamma_i \times \iota'_i) \times (\gamma_j \times \iota'_j) = 0_p$.

Pour effectuer la décomposition LU , il suffit donc de réaliser simultanément des opérations sur Δ et I_p . Pour chaque opération : $ligne_j \leftarrow ligne_j - ligne_k \times \gamma_{j,k}$ et, parallèlement, $\gamma_{j,k}$ est stocké dans l'élément j de la colonne k de la matrice identité I_p .

Avec l'exemple précédent pour lequel $p = 3$ et $\Delta = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$, la matrice L est donc

$$L = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{pmatrix}$$

car $\gamma_{2,1} = -\frac{1}{2}$ et $\gamma_{3,2} = -\frac{2}{3}$.

Remarquons qu'il est nécessaire que l'élément $\delta_{k,k}$ soit non nul. Dans le cas où $\delta_{k,k} = 0$, on peut se ramener à cette méthode en permutant les lignes de la matrice Δ de sorte à avoir $\delta_{k,k} \neq 0$. On aura alors la décomposition $P\Delta = LU$ avec P la matrice des permutations qui est orthogonale : $P^{-1} = P^t$. Cela permet de retrouver le vecteur $\hat{\beta}$ dans l'ordre attendu sans trop de complexité.

Le nombre d'opérations nécessaires à la résolution d'un système linéaire pour la décomposition LU est de l'ordre de $\frac{2}{3}p^3$.

Déterminer $\hat{\alpha}$

La résolution de $\hat{\alpha}$ dans l'équation (3.2) se fait par étapes successives dites de *descente* :

$$\text{Pour } i \text{ allant de } 1 \text{ à } p : \hat{\alpha}_i = (X^t Y)_i - \sum_{j=1}^{i-1} l_{i,j} \hat{\alpha}_j$$

où $L = (l_{i,j})_{1 \leq i, j \leq p}$ et $(X^t Y)_i$ est le i^{ieme} coefficient du vecteur $X^t Y$.

Déterminer $\hat{\beta}$

La résolution de $\hat{\beta}$ dans l'équation (3.3) se fait par étapes successives dite de *remontée* :

$$\text{Pour } i \text{ allant de } p \text{ à } 1 : \hat{\beta}_i = \frac{1}{u_{i,i}} \left(\hat{\alpha}_i - \sum_{j=i+1}^p u_{i,j} \hat{\beta}_j \right)$$

où $U = (u_{i,j})_{1 \leq i, j \leq p}$ et $\hat{\alpha}_i$ est le i^{ieme} coefficient du vecteur $\hat{\alpha}$.

Les nominaux associés aux instruments financiers retenus sont ainsi déterminés.

3.2 Les méthodes actuelles de sélection d'instruments

Une des difficultés liées à la méthode des portefeuilles de réplication réside dans la connaissance des instruments financiers à utiliser. En effet, dans les problèmes de régression usuels, l'ensemble X des variables explicatives est connu. Ce n'est cependant pas le cas pour la méthode des portefeuilles répliquants. Les instruments financiers doivent donc être déterminés en amont. Les deux méthodes actuellement employées sont présentées dans cette section. Ces deux méthodes se complètent car elles sont successivement utilisées par les actuaires d'Axa en charge du calcul du SCR de marché en modèle interne. L'Orthogonal Matching Pursuit est une méthode permettant de réduire le nombre d'instruments de travail à une centaine. C'est sur cette centaine d'instruments financiers qu'une équipe d'actuaires travaillent à *dire d'expert* afin de retenir une combinaison satisfaisante.

3.2.1 L'Orthogonal Matching Pursuit

L'Orthogonal Matching Pursuit est une variante de la méthode Matching Pursuit utilisée dans le traitement du signal : approximation, débruitage, problèmes inverses, détection, séparation de sources... [8]

Elle a été mise en place dans le cadre des portefeuilles répliquants afin de **sélectionner** les instruments financiers les plus **appropriés** à utiliser au sein d'un panel très large d'instruments

candidats.

Le problème duquel découle l'utilisation de l'Orthogonal Matching Pursuit est le suivant : Soit un signal Y et $X = (X_i)_{0 \leq i \leq p}$ un dictionnaire de signaux, appelés *atomes*, de la même dimension que Y . On cherche à trouver une **combinaison linéaire** d'éléments de X permettant d'écrire, ou d'**approximer**, le signal Y .

L'orthogonal Matching Pursuit permet de déterminer **itérativement** les atomes X_i constituant cette combinaison.

L'approche de l'Orthogonal Matching Pursuit est la suivante :

- Etape 1 : **Recherche** de l'atome X_{i_0} le plus **corrélé** à Y .
- Etape 2 : **Evaluation** du résidu $\epsilon = Y - \beta_{i_0} X_{i_0}$ en projetant Y sur X_{i_0}
- Etape 3 : **Répéter** les étapes 1 et 2 en remplaçant Y par le résidu ϵ . Le résidu de l'étape $k - 1$ est projeté sur l'atome X_{i_k} évalué en étape k .

A chaque itération, le sous-espace des atomes sélectionnés s'**accroît** d'un élément.

La projection de Y se fait sur le sous-espace des atomes sélectionnés jusqu'alors. Cela implique qu'**à chaque étape** de l'Orthogonal Matching Pursuit, les coefficients associés aux différents atomes sélectionnés X_i sont **recalculés**.

Dans le cas des portefeuilles répliquants, Y est le passif à répliquer et $X = (X_i)_{0 \leq i \leq p}$ est la matrice dont les colonnes sont constituées des valorisations des instruments financiers candidats : chaque colonne est un X_i donc la valorisation d'un instrument financier dans les différents scénarii est un atome.

Afin d'évaluer l'atome (donc l'instrument financier) le plus corrélé à Y , on évalue la projection de Y sur l'ensemble des atomes. L'instrument financier retenu est celui possédant le coefficient le plus élevé.

Le calcul des coefficients provenant de la projection de Y sur les différents atomes se fait en utilisant la décomposition LU exposée dans la section précédente.

Pour présenter l'algorithme de l'Orthogonal Matching Pursuit, introduisons quelques notations :

- Ω l'ensemble des instruments candidats
- Γ^k la composition du portefeuille répliquant à l'étape k de l'algorithme
- $\Omega^k = \Omega \setminus \Gamma^k$ l'ensemble des instruments candidats non sélectionnés à l'étape k
- X^k la matrice associée à la composition Γ^k
- β^k le vecteur des nominaux associé à la matrice X^k
- $\epsilon^k = Y - X^k \beta^k$ le vecteur de résidus à l'étape k
- α_i^k le produit scalaire entre le vecteur X_i des valorisations de l'instrument $i \in \Omega$ et le vecteur des résidus
- p_{OMP} le nombre d'instruments en fin d'algorithme. C'est ici un critère d'arrêt. Ce nombre est fixé par l'utilisateur.

L'algorithme d'Orthogonal Matching Pursuit est alors :

1. Sélection par l'utilisateur de l'ensemble de travail Ω et choix de p_{OMP}
2. Initialisation des variables : $\epsilon^0 = Y$, $\Gamma^0 = \emptyset$

3. Pour k allant de 1 à p_{OMP} :

- (a) Produit scalaire entre les résidus et les instruments candidats : α_i^k
- (b) Recherche de l'instrument maximisant le produit scalaire associé aux résidus :
 $i^k = \arg \max_{i \in \Omega^{k-1}} |\alpha_i^k|$
- (c) Ajout de l'instrument au portefeuille : $\Gamma^k = \Gamma^{k-1} \sqcup \{i^k\}$
- (d) Calcul des coefficients du nouveau portefeuille : $\beta^k = \arg \min_{\beta} \|Y - X^k \beta\|_2^2 + \lambda \|\beta\|_2^2$
- (e) Calcul des résidus ϵ^k .

La base de l'ensemble de travail Ω est à **déterminer au préalable** par l'utilisateur. Cette étape est très importante car ce sont les éléments de cette base qui seront sélectionnés par l'algorithme. Il peut être tentant d'augmenter la granularité de la base. Cependant, cela augmente considérablement le temps de calcul. Cela introduit également une forte colinéarité entre deux instruments aux caractéristiques proches. En pratique, avoir une granularité trop grande mène à un portefeuille résultant constitué de groupes d'instruments aux propriétés similaires. Cela peut avoir les différentes **conséquences** suivantes :

- donner de mauvais résultats sur les sensibilités
- impossibilité de répondre à la contrainte du BEL ratio (2.3)
- présenter des situations acheteuses-vendeuses
- overfitting.

La valeur de p_{OMP} retenue par l'utilisateur est en général supérieure au nombre désiré d'instruments financiers composant le portefeuille de réplication. En effet, il est en pratique très difficile de répondre à toutes les contraintes du problème (2.3) si $p_{OMP} = p$. L'Orthogonal Matching Pursuit permet en réalité d'obtenir une **base de travail restreinte** avec $p_{OMP} \in [70, 120]$ sur laquelle les actuaires peuvent travailler à *dire d'expert*.

3.2.2 La méthode "A dire d'expert"

La méthode à *dire d'expert* consiste à sélectionner les instruments financiers utilisés pour la réplication à partir du **jugement de l'actuaire** réalisant l'étude. Cette méthode requiert un haut niveau de maîtrise des instruments étudiés, de leurs caractéristiques et de leur interprétation.

Les différentes étapes usuellement adoptées pour cette approche sont exposées ci-après.

Etape 1 : Choix de la base de l'ensemble des instruments candidats Ω .

Comme énoncé dans la sous-section précédente, le choix de la **granularité** est primordial et requiert une attention particulière. Ce choix dépend du portefeuille étudié. En pratique, Ω est de l'ordre de 5000 instruments.

Etape 2 : Application de l'Orthogonal Matching Pursuit.

La seconde étape consiste à choisir la taille de la base restreinte sur laquelle l'actuaire souhaite travailler. Il détermine le p_{OMP} qu'il désire et applique l'**Orthogonal Matching Pursuit**. Cette étape est coûteuse en temps : afin de trouver le meilleur instrument à ajouter à la base restreinte, chaque instrument de la base Ω (et non déjà retenu) est testé lors de chacune des p_{OMP} itérations.

A l'issue de l'étape 2, la base de travail restreinte est composée de p_{OMP} instruments financiers.

Etape 3 : Réduction du nombre d'instruments composant le portefeuille.

A partir du **jugement d'expert** de l'actuaire, des instruments du portefeuille sont **éliminés**. Ces éliminations successives doivent permettre de réduire la taille du portefeuille en augmentant, ou *a minima* conservant, les performances du portefeuille de réplication. Ces performances concernent les contraintes énoncées dans la problématique (2.3), à savoir :

- $R^2 \geq 0.9$
- vérifier que l'erreur de réplication sur chaque sensibilité soit inférieure à 2% du SCR de marché de l'exercice précédent calculé en modèle interne
- BEL ratio ≥ 0.6
- vérifier la normalité des résidus.

Ces 4 contraintes de performance sont centrales à la recherche d'un portefeuille de réplication pertinent. Pour la suite de l'étude, ces contraintes de performance seront mentionnées par l'acronyme *KPI* pour *Key Performance Indicator*.

Etape 4 : Modification des instruments composant le portefeuille.

Il peut arriver que les KPI ne soient pas augmentées ou conservées en réduisant le nombre d'instruments. Dans de tels cas, la performance des KPI est différée au profit de la réduction du nombre d'instruments : il est nécessaire de réduire les performances des KPI afin de réduire le nombre d'instruments du portefeuille.

Cependant, cette réduction des KPI ne doit être que **provisoire**. En effet, après avoir réussi à réduire le nombre d'instruments qui compose le portefeuille, il est essentiel de satisfaire de nouveau ces KPI. Pour cela, il est possible de modifier certaines caractéristiques de certains instruments présents au sein du portefeuille.

Illustrons ce cas de figure par un exemple. Supposons qu'on dispose d'un portefeuille de 40 instruments financiers pour lequel seule la contrainte du R^2 est inférieure au seuil critique de 0.9. Si ce portefeuille est composé d'un zéro-coupon de maturité 10 ans, et que modifier ce zéro-coupon par un autre zéro-coupon de maturité 8 ans permet de passer au dessus du seuil de 0.9 (et ainsi de satisfaire l'ensemble des KPI), alors la modification est conservée.

Etape 5 : Vérification du portefeuille répliquant

Valider les KPI ne suffit pas à fournir un portefeuille de réplication recevable. L'étape ultime de validation suggère de regarder plusieurs graphiques appelés *Topologies*. Ces topologies permettent de vérifier que le portefeuille répliquant candidat se **comporte bien comme le portefeuille de passif à répliquer** face aux différents chocs. Si les topologies sont validées, le portefeuille répliquant peut être retenu.

Remarque :

Les différentes étapes présentées dans cette sous-section ne sont données qu'à titre **indicatif**. Elles suivent l'ordre réalisé usuellement par les actuaires de l'équipe *Modèles et Solvabilité 2*. Néanmoins, par principe de la méthode *à dire d'expert*, si l'expert estime ne pas avoir besoin de suivre les étapes précédemment énoncées afin de déterminer les instruments financiers composant le portefeuille répliquant, son **expertise prime** sur ces dites étapes.

3.2.3 Les inconvénients des méthodes actuelles

Les méthodes utilisées actuellement possèdent plusieurs inconvénients.

Lors du calcul des nominaux, la fonction de coût à minimiser $C_{\theta,\lambda,k}(\beta)$ dépend des paramètres θ , λ et du vecteur de poids k . Ces **paramètres ne sont pas connus** et sont déterminés

à dire d'expert. Si par exemple les performances liées aux sensibilités ne sont pas suffisantes, il peut être pertinent d'augmenter θ , et donc d'accorder plus de poids à ces sensibilités. Mais dans quelle mesure augmenter θ ? La réponse repose sur l'appréciation de l'utilisateur.

La problématique de détermination des paramètres θ , λ et du vecteur de poids k concerne également les méthodes actuelles de sélection d'instruments financiers. Les poids associés aux instruments doivent être calculés pour chaque portefeuille candidat afin de pouvoir observer les KPI.

A ce problème s'ajoute l'aspect **chronophage** des calculs lors de la phase de sélection d'instruments. En effet, l'Orthogonal Matching Pursuit est par construction très lourde en temps. De plus, la méthode *à dire d'expert* nécessite de mobiliser un actuaire plusieurs jours afin de calibrer un portefeuille répliquant. Plusieurs jours sont nécessaires entre la composition et la validation d'un portefeuille de réplication. Réduire ce temps permettrait de mobiliser ces actuaires sur d'autres problématiques. L'automatisation du processus de calibration paraît donc essentielle.

LE CHOIX DE L'ALGORITHME GÉNÉTIQUE

Une des difficultés liées à la méthode des portefeuilles de réplication repose sur la connaissance des instruments financiers à utiliser. Le chapitre précédent expose deux méthodes actuellement utilisées au sein d'AXA France mais qui présentent différentes limites. Ce chapitre introduit une méthode de sélection d'instruments visant à challenger les méthodes actuelles : l'algorithme génétique.

4.1 L'origine de l'algorithme

Un algorithme génétique est un **algorithme d'intelligence artificielle évolutionniste**. Cette caractérisation d'évolutionniste provient du fait qu'un algorithme génétique se base sur la **théorie de l'évolution** de Charles Darwin [4]. Afin de présenter l'algorithme génétique, commençons par rappeler ce qu'est le phénomène biologique de sélection naturelle dans la théorie de l'évolution.

La théorie de la **sélection naturelle** a été introduite par Charles Darwin. Cette théorie stipule que les individus les plus **adaptés** à un environnement connaissent un **meilleur taux de survie** et de plus grandes chances de reproduction que les autres. Les caractères faisant la force d'adaptation d'une espèce étant transmissibles par **héritage génétique**, ils se propagent au sein de la population. De génération en génération, les caractères héréditaires favorables deviennent plus communs, contrairement aux caractères héréditaires défavorables. Au fil des générations, il en résulte une **sélection d'individus adaptés** à un environnement et possédant des caractères favorables à l'espèce.

La sélection naturelle proposée par Darwin comme moyen d'expliquer l'évolution biologique repose sur **trois hypothèses**. La première est que le caractère soumis à la sélection doit être **héréditaire**. La seconde repose sur le fait qu'il doit y avoir une **variabilité du caractère** parmi les individus formant la population. Enfin, la dernière hypothèse est que la variabilité du caractère doit mener à des **différences dans la survie** ou le succès de la reproduction, et donc de l'extension de ce caractère dans la population. Si ces trois hypothèses sont vérifiées, alors la variation du caractère conduit au phénomène d'évolution au cours du temps au sein d'une population.

C'est en s'inspirant de ce phénomène biologique que les algorithmes génétiques ont été développés par John Holland dans les années 1960. En effet, le phénomène de sélection naturelle peut être appréhendé comme un processus qui se répète dans le temps et dont le but est d'**optimiser**

un individu au sein d'une population par l'optimisation de son patrimoine génétique. Cet individu répond alors au mieux à son environnement qui peut être perçu comme la problématique donnée [13].

A la fin des années 1980, David Goldberg popularise l'utilisation des algorithmes génétiques avec la publication de son livre sur le sujet [11]. Par le parallèle avec un **phénomène biologique connu** du plus grand nombre, la **simplicité de mise en oeuvre** et l'**efficacité d'exécution**, le champ d'application des algorithmes génétiques est vaste. De l'économie à la finance en passant par la théorie des jeux, les algorithmes génétiques sont des algorithmes d'optimisation puissants utilisés pour répondre à diverses problématiques [27].

4.2 Les cas d'utilisation et le principe de l'algorithme

Un algorithme génétique est un **algorithme d'optimisation stochastique** basé sur la génétique et sur le phénomène de sélection naturelle.

Cet algorithme témoigne de performances notables par rapport aux algorithmes plus classiques d'intelligence artificielle lorsque l'**espace des solutions est relativement grand**. Si l'espace des solutions n'est pas conséquent, il est plus sûr de parcourir entièrement ce dernier afin de trouver la solution optimale. Utiliser un algorithme génétique dans une telle situation pourrait mener à une solution non optimale, par exemple à un extremum local, en un temps qui ne sera pas forcément inférieur. A l'inverse, si l'espace des solutions est grand et qu'un **arbitrage exactitude de la solution/temps de calcul** se présente, il est approprié d'appréhender le problème en utilisant un algorithme génétique. Il sera alors préféré d'avoir une **bonne solution rapidement** plutôt qu'avoir la **solution optimale en une durée indéfinie**.

De ce fait, les algorithmes génétiques sont utilisés dès lors qu'il n'existe pas d'algorithme déterministe adapté et raisonnable ou lorsqu'il y a nécessité d'obtenir une solution très réactive aux conditions environnantes.

Le **principe de l'algorithme génétique** est simple, il se présente comme suit :

- Une population de solutions potentielles arbitrairement choisies représente la **première génération G0**. Au sein de cette génération G0, l'ensemble des caractères (gènes) différent d'un individu à l'autre.
- La **performance** de chacun des individus formant cette première génération est évaluée.
- C'est sur la base de ces différentes performances qu'**une nouvelle génération G1 naît** à partir de trois opérateurs évolutionnaires : la sélection, le croisement et la mutation.
- Cette nouvelle génération G1 va à son tour **générer une nouvelle population** d'individus mieux adaptés, et ce cycle se renouvelle jusqu'à l'obtention d'une solution satisfaisante.

Le coeur de l'algorithme repose sur les trois opérateurs évolutionnaires que sont la sélection, le croisement et la mutation.

La sélection retient les individus les mieux adaptés à la problématique. Le croisement permet le mélange par la reproduction des gènes des individus choisis. Enfin, la mutation altère de manière aléatoire certains gènes d'un individu. C'est ce dernier opérateur de mutation qui va permettre de ne pas converger trop rapidement vers une des solutions du problème, celle-ci pourrait en effet représenter un extremum local et donc ne pas être la plus optimale.

La figure 4.1 ci-dessous schématise le fonctionnement de l'algorithme.

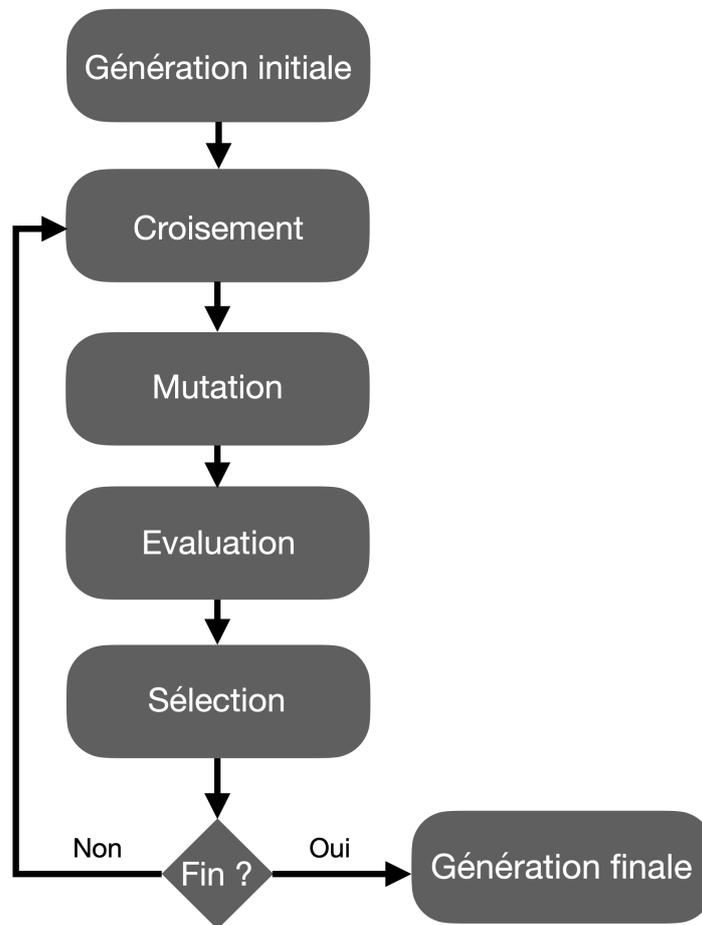


FIGURE 4.1 – Principe de l'algorithme génétique

4.3 La présentation de l'algorithme

Du fait du parallèle avec la théorie évolutionniste de Charles Darwin, l'algorithme génétique utilise un vocabulaire biologique qu'il convient de définir dans cette partie.

4.3.1 La représentation des solutions

Un algorithme génétique est défini par **cinq composantes** qui sont les suivantes [17] :

- Des gènes
- Des individus
- Des générations
- Un environnement
- Une fonction d'évaluation.

Les quatre premières composantes, qui définissent le problème à résoudre, peuvent être perçues comme des ensembles qui interagissent entre eux selon une **relation d'inclusion** telle que présentée sur la figure 4.2. La fonction d'évaluation, dernière composante d'un algorithme

génétique, sert quant à elle à **ordonner** les différents individus d'une génération selon la **qualité de leurs gènes**. C'est cet ordre qui va permettre de sélectionner les individus qui bénéficieront de la sélection naturelle et qui transmettront leurs gènes aux générations futures.

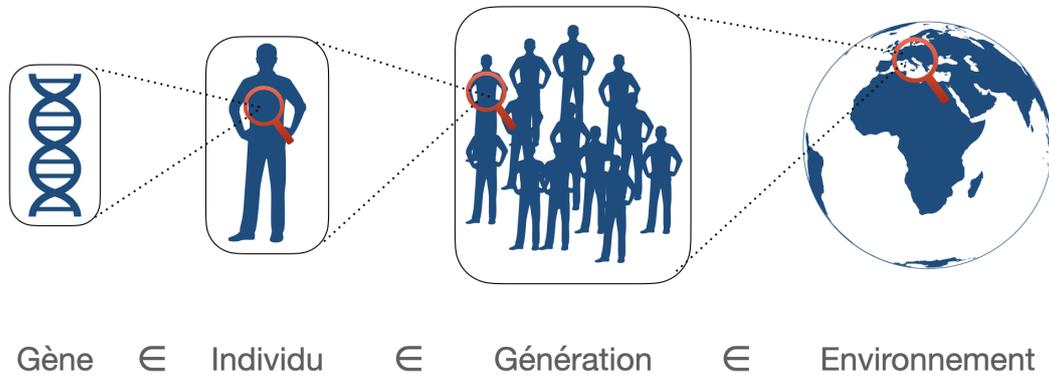


FIGURE 4.2 – Composantes d'un algorithme génétique

Afin de définir correctement les différentes composantes d'un algorithme génétique, prenons le cas d'un problème d'optimisation à P paramètres.

Les gènes

Un algorithme génétique sert à obtenir une solution approchée d'un problème d'optimisation d'un ou plusieurs paramètres.

Un gène représente **une des solutions potentielles d'un des paramètres** du problème d'optimisation. Dans le cas d'un problème d'optimisation à trois paramètres, trois gènes différents seront étudiés : le premier gène est une solution potentielle au premier paramètre, le second gène est une solution potentielle au second paramètre et le troisième et dernier gène, au dernier paramètre.

Soient p_1, \dots, p_P les P paramètres du problème d'optimisation.

Notons g_1, \dots, g_P P gènes. g_1, \dots, g_P sont des solutions potentielles à p_1, \dots, p_P . On a égalité des dimensions :

$$\dim(g_1) = \dim(p_1), \dots, \dim(g_P) = \dim(p_P).$$

Les individus

Un individu est formé par un ensemble de gènes représentant les solutions potentielles des paramètres du problème.

Notons I un individu, g_1, \dots, g_P P gènes du problème d'optimisation, on a alors

$$I = \{g_1, \dots, g_P\}.$$

Un individu est ainsi une **réponse potentielle au problème** d'optimisation.

Les générations

Une génération est un ensemble d'individus à un instant donné. Soient G une génération et I_1, \dots, I_d d individus, on a :

$$G = \{I_1, \dots, I_d\}.$$

Une génération représente ainsi un **ensemble de solutions potentielles** au problème d'optimisation.

L'environnement

L'environnement, noté \mathcal{E} , est l'**espace de définition des solutions potentielles du problème**. Dans le cas d'un problème d'optimisation à P paramètres p_1, \dots, p_P , la dimension de l'environnement, et donc de l'espace de recherche, est alors :

$$\dim(\mathcal{E}) = \dim(p_1) \times \dots \times \dim(p_P).$$

La fonction d'évaluation

La fonction d'évaluation a un rôle primordial car c'est elle qui permet de faire le **lien entre le problème d'optimisation et l'algorithme de résolution**. En ordonnant les individus d'une génération, elle sélectionne ceux qui auront de meilleures chances de survie et de reproduction, et permet ainsi de se rapprocher d'une des solutions.

Le choix de la fonction d'évaluation ainsi que sa complexité dépendent du problème. Dans certains cas, le processus de sélection d'une fonction d'évaluation est simple, mais lorsque des contraintes viennent s'ajouter au problème, le choix de la fonction peut rapidement se complexifier.

Afin de montrer ce processus de sélection, prenons l'exemple du problème de minimisation de la fonction de Rosenbrock provenant de l'étude de De Jong (1975) [5] :

$$F : [-2.048, 2.048]^2 \longrightarrow \mathbb{R} \\ (x_1, x_2) \longmapsto 100(x_1^2 - x_2)^2 + (1 - x_1^2).$$

Chaque individu, qui représente un couple de 2 gènes $\{g_1, g_2\}$, peut facilement être évalué. En effet, si $I_1 = \{0.231, -1.892\}$, alors $F(I_1) = 379.034$.

Il est important de noter que dans ce cas, chacun des individus appartenant à l'environnement (ou espace de recherche) $\mathcal{E} = [-2.048, 2.048]^2$ est une solution réalisable. Cependant, il est parfois nécessaire de faire la distinction entre espace de recherche et espace de recherche réalisable. En ajoutant une contrainte au problème précédent, par exemple

$$x_1^2 < x_1 + \frac{1}{x_2},$$

alors l'individu I_1 ne fait plus partie des solutions réalisables car la contrainte n'est plus vérifiée :

$$0.053 = 0.231^2 > 0.231 + \frac{1}{-1.892} = -0.297.$$

La question de la **gestion des individus irréalisables** apparaît alors et c'est par la fonction d'évaluation que cette question doit être traitée.

En général, un **espace de recherche** \mathcal{E} est constitué de deux sous-ensembles disjoints de **sous-espaces réalisable** \mathcal{F} et **irréalisable** \mathcal{U} .

$$\mathcal{E} = \mathcal{F} \sqcup \mathcal{U}.$$

Aucune hypothèse n'est faite sur ces sous-espaces, en particulier, aucune hypothèse de convexité n'est nécessaire.

En résolvant un problème d'optimisation, un optimum réalisable est désiré. Cependant, au cours du processus de recherche d'un algorithme génétique, des individus présentant des solutions réalisables et non réalisables peuvent être générés, par exemple :

$$G_1 = \{I_1, I_2, I_3, I_4\} \text{ avec } I_1, I_3 \in \mathcal{F} \text{ et } I_2, I_4 \in \mathcal{U}.$$

Le problème du traitement des individus irréalisables I_2, I_4 se pose donc. Ce problème, loin d'être trivial, soulève plusieurs questions qui doivent être abordées lors du choix de la fonction d'évaluation. Dans son étude, Z. Michalewicz (1995) [19] pose plusieurs questions qui méritent d'être relevées lors de la détermination de la fonction d'évaluation :

- Comment comparer deux solutions, et donc deux individus réalisables $I_1, I_3 \in \mathcal{F}$?
Cette question est généralement la plus simple à appréhender car la solution est souvent donnée dans le problème d'optimisation (c'est la fonction de Rosenbrock F dans notre exemple)
- Comment comparer deux individus irréalisables $I_2, I_4 \in \mathcal{U}$?
- Tout individu réalisable est-il préférable à tout individu irréalisable ?
Soit $x \mapsto eval(x)$ une fonction d'évaluation d'un problème de minimisation,
 $\forall (I, J) \in \mathcal{F} \times \mathcal{U}, eval(I) < eval(J)$?
Rappel : dans un problème de minimisation, l'opérateur "inférieur à" témoigne de meilleures performances.
- Les individus irréalisables doivent-ils être retirés de l'environnement ? C'est-à-dire avoir $\mathcal{E} = \mathcal{F}$?
Ou doivent-ils être considérés afin d'apporter plus de diversité et donc éviter une convergence prématurée vers un minimum local ? $\mathcal{E} = \mathcal{F} \sqcup \mathcal{U}$.

Toutes ces questions reviennent à poser celle de l'**utilisation d'individus irréalisables** pour leurs **potentiels d'information** concernant la recherche de l'optimum : un individu irréalisable pouvant être plus proche de la solution optimale que de nombreux autres individus réalisables.

En résumé, dans des **cas d'optimisation sous contraintes** pour lesquels la fonction d'évaluation n'est **pas facilement identifiable**, il n'existe pas de règle permettant de définir à coup sûr une bonne fonction d'évaluation. La **compréhension** du problème et de ses contraintes est essentielle afin de déterminer une fonction capable de mener à des solutions satisfaisantes et ainsi répondre au problème d'optimisation.

4.3.2 Les opérateurs d'évolution

Il existe trois opérateurs d'évolution qui ont un rôle très important dans la réussite d'un algorithme génétique [15], [27]. Le parallèle avec la théorie de l'évolution rend facilement compréhensible le principe de chacun de ces opérateurs qui agissent selon des critères qui leurs sont propres.

La sélection

Le processus de sélection consiste à retenir certains individus afin d'avoir un **ensemble de solutions capable de converger** vers l'optimum du problème posé. Cet opérateur est l'application directe du phénomène d'**adaptation** à un environnement dans la théorie évolutionniste de Charles Darwin. Les individus retenus sont en effet les mieux adaptés à leur environnement, sous réserve d'un bon choix de la fonction d'évaluation.

Différentes techniques de sélection peuvent être envisagées.

La plus intuitive est la **sélection par rang**. Cette technique consiste à retenir systématiquement les individus ayant les meilleures performances en termes de fonction d'évaluation.

La méthode de la **probabilité de sélection proportionnelle à l'adaptation** est une autre possibilité. Avec cette méthode, pour chaque individu, la probabilité d'être sélectionné est proportionnelle à l'adaptation au problème. Les individus ayant de bonnes performances ont donc plus de chance d'être sélectionnés.

La méthode du **tournoi** est également répandue : deux individus sont tirés aléatoirement au sein d'une population et le meilleur des deux est retenu dans la nouvelle population. Cette procédure est répétée jusqu'à ce que la nouvelle population soit complète.

Il est également possible de sélectionner aléatoirement, uniformément et sans intervention de la fonction d'évaluation certains individus d'une génération. C'est la **sélection uniforme**.

Il est aussi envisageable de mélanger plusieurs des méthodes énumérées ci-dessus afin de sélectionner des individus : par exemple procéder à une sélection par rang pour la moitié des individus sélectionnés et utiliser la sélection uniforme pour l'autre moitié. Cependant, ne permettant pas de générer de nouveaux individus, l'opérateur de sélection doit être complété par d'autres opérateurs.

Le croisement

Le croisement est le processus qui simule le phénomène de **reproduction** au cours de l'évolution. C'est avec cet opérateur que de nouveaux individus sont créés et que le mélange des gènes, et donc le **mélange d'information**, se fait.

Le principe est le suivant : parmi les individus retenus après l'opérateur de sélection, deux individus parents sont choisis aléatoirement et forment un couple. Ce couple génère des enfants qui héritent des gènes de leurs parents.

Cet héritage peut se faire selon différents procédés.

Il est par exemple possible de localiser un site pivot d'échange de gènes et, avec une probabilité p_C , l'échange a lieu entre les gènes des deux parents comme illustré sur la figure 4.3.(a). Ce procédé s'appelle le **croisement simple**.

Le **croisement multiple**, ou croisement par blocs, repose sur le même principe sauf que le nombre de points pivot est multiple (figure 4.3.(b) [15] avec 2 points pivot). Chacun des nouveaux individus créés hérite ainsi des gènes de ses deux individus parents.

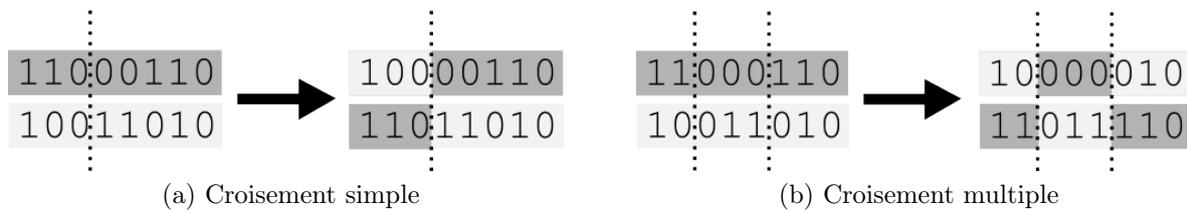


FIGURE 4.3 – Deux méthodes de croisement de gènes de deux individus parents

L'opérateur de croisement ne permet cependant pas de générer un gène qui ne serait pas déjà présent parmi les individus de la génération reproductrice. Ainsi, si ce gène manquant fait en réalité parti de la solution du problème d'optimisation et qu'il n'est pas déjà présent dans la population, les seuls opérateurs de sélection et de croisement ne permettront pas de converger vers la solution optimale. C'est pour pallier ce problème qu'intervient l'opérateur de mutation.

La mutation

La mutation est le procédé qui va apporter de la **diversité** au sein d'une population. C'est l'application du phénomène de **variation** dans la théorie de l'évolution. Un ou plusieurs gènes d'un individu vont être altérés avec une probabilité p_M . La mutation peut donc être **simple** ou **multiple**.

Grâce à cet opérateur, il sera possible de **s'éloigner d'un extremum local** et d'**éviter une convergence prématurée** vers une solution non optimale. L'opérateur de mutation permet ainsi l'exploration de tous les points de l'espace de recherche, ce qui permet d'obtenir la propriété d'ergodicité. La figure 4.4 [15] schématise le fonctionnement de la mutation avec l'exemple d'une mutation simple.

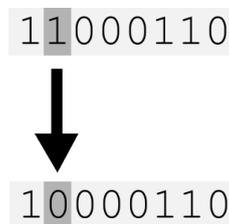


FIGURE 4.4 – Mutation d'un gène d'un individu

Avec les trois opérateurs d'évolution que sont la sélection, le croisement et la mutation, les algorithmes génétiques peuvent être appliqués.

4.3.3 Des paramètres additionnels

Les opérateurs d'évolution présentés précédemment sont caractérisés par un certain nombre de paramètres qui sont présentés dans cette sous-section. La valeur de ces paramètres influence la réussite d'un algorithme génétique.

Le premier paramètre déterminant est la **taille des générations** N . Les différentes générations issues d'un l'algorithme génétique sont souvent considérées constantes en nombre. Lorsque ce nombre N est grand, le temps de calcul peut s'avérer conséquent, mais lorsqu'il est trop faible, le risque de convergence prématurée augmente. Le choix de ce paramètre est important, il faut faire un compromis entre temps de calcul et convergence prématurée.

Les deux autres paramètres essentiels et déjà énoncés précédemment sont les **probabilités de croisement** et **de mutation** p_C et p_M .

Plus la probabilité de croisement p_C est grande, plus la population subit de changements importants. Sa valeur est généralement comprise dans l'intervalle $[0.5, 0.9]$.

La probabilité de mutation doit quant à elle rester relativement faible. En effet, une probabilité p_M élevée pourrait rendre difficile la convergence du fait de la haute variabilité des individus. Néanmoins, afin de ne pas imposer une valeur trop faible de p_M , une **stratégie dite élitiste** peut être employée. Cette stratégie consiste en la conservation d'un pourcentage de meilleurs individus d'une génération à l'autre, par exemple 5% des meilleurs individus. Les enfants issus des différents croisements et mutations viennent alors compléter la nouvelle génération. De ce fait, les individus aux meilleures performances ne sont pas altérés et sont conservés dans la génération suivante.

Les valeurs des paramètres N , p_C et p_M sont inconnues et sont fixes au cours des différentes itérations d'un algorithme génétique. Elles sont déterminées par expérience et ne peuvent donc pas être améliorées une fois l'algorithme lancé.

Des travaux ont cependant été menés afin de proposer un moyen d'introduire l'amélioration de ces paramètres au cours des différentes itérations de l'algorithme : il s'agit d'un **méta algorithme génétique** [21]. Deux algorithmes travaillent simultanément : l'un recherche l'individu optimal, comme le ferait l'algorithme classique, tandis que l'autre recherche la valeur optimale des paramètres. Cependant, l'utilisation d'un tel meta algorithme génétique serait extrêmement chronophage et n'est donc envisageable que dans des cas où l'environnement \mathcal{E} n'est pas trop grand.

4.4 Un exemple introductif

Afin de présenter l'application de l'algorithme génétique au cas plus complexe des portefeuilles répliquants, il semble pertinent de commencer par exposer dans cette section la puissance de cet algorithme sur un cas d'école.

4.4.1 La problématique

Dans cet exemple simple, on cherche à retrouver les mots "*Portefeuille répliquant*" dans cet ordre. Cet exemple est un problème de **maximisation** sans contrainte dont le résultat attendu est connu.

4.4.2 Les composantes de l'algorithme

Tout d'abord, il faut définir correctement les composantes de l'algorithme.

Dans cet exemple :

- un **gène** g correspond à un **caractère** parmi l'ensemble des lettres de l'alphabet latin (majuscules et minuscules) ainsi que les huit caractères : *espace* . [\ \] ^ _ ' soit 60 possibilités au total.
- un **individu** I est un ensemble formé de P gènes (ici $P = 23$, c'est la longueur de la phrase cible) : $I = \{g_1, \dots, g_P\}$
- une **génération** G est un ensemble de N individus : $G = \{I_1, \dots, I_N\}$. Pour cet exemple, une génération est composée de $N = 2000$ individus.

- l'**environnement** \mathcal{E} est l'ensemble des combinaisons possibles de $P = 23$ paramètres du problème. Sachant qu'il y a 60 possibilités pour chaque paramètre, $\dim(\mathcal{E}) = 60^{23} \simeq 7.9 \times 10^{40}$.
- la **fonction d'évaluation** f à **maximiser** dans ce cas précis est la proportion de gènes, et donc de caractères, correcte :

$$\forall i \in [[1, N]], f(I_i) = \sum_{j=1}^P \frac{\mathbb{1}_{g_j=p_j}}{P}.$$

Pour rappel, cette fonction de coût est possible dans ce cas précis car la phrase cible $\{p_1, \dots, p_{23}\}$ est connue.

4.4.3 Les opérateurs d'évolution

Une fois les composantes de l'algorithme déterminées, il faut préciser les méthodes retenues pour chacun des opérateurs d'évolution.

Opérateur de sélection

Pour cet exemple simple, l'opérateur de sélection retenu est la sélection par rang, c'est-à-dire que les individus sélectionnés sont ceux présentant les meilleures performances.

Opérateur de croisement

Dans cet exemple, deux parents engendrent deux enfants. L'opérateur de croisement utilisé est le croisement simple à un point de pivot choisi aléatoirement dans l'intervalle $[[1, P]]$. Avant ce point de pivot et avec probabilité de croisement p_C , le premier enfant hérite des gènes du premier parent et après ce point pivot, il hérite des gènes du second parent. Le même procédé est appliqué au deuxième enfant mais il hérite d'abord des gènes du parent 2, puis du parent 1 après le point pivot.

Opérateur de mutation

Ici, chacun des gènes d'un enfant peut subir une mutation avec probabilité p_M . Il s'agit donc d'une mutation multiple.

4.4.4 Les paramètres de l'algorithme

Les paramètres de l'algorithme sont les suivants :

- Le nombre d'individus par génération : $N = 2000$
- La probabilité de croisement : $p_C = 1$, il s'agit d'un cas de croisement certain
- La probabilité de mutation : $p_M = 0.1$.

4.4.5 Les résultats

L'algorithme génétique repose sur de l'aléatoire. Ainsi, les résultats obtenus varieront d'une exécution à l'autre.

Avec les composantes, opérateurs et paramètres définis dans les sous-sections précédentes, la phrase cible est obtenue en moins de 100 générations.

Le résultat est obtenu en **74 générations**. Le détail des meilleurs individus par génération est présenté en Annexe D. Le tableau 4.1 synthétise le tableau en Annexe D en ne présentant que les meilleurs individus de 10 générations avec, en rouge, les caractères correctement placés. On constate que dès la trentième génération, le meilleur individu a plus de la moitié de ses gènes correctement placés.

Génération	Meilleur individu	Fonction d'évaluation
1	\wedge Tr.wSvnr_ib_■_nVelpN Cle	$\frac{3}{23}$
5	oYutek'uxjiL_■_RfgNyXaF\ t	$\frac{6}{23}$
10	P^lzCAuuYlCZ^GepoiQEaDt	$\frac{8}{23}$
20	LRrueieicibDIImhepli quant	$\frac{12}{23}$
30	Purttfkuillh_■_hepliquanI	$\frac{17}{23}$
40	PortefeuilhQkepliTuagt	$\frac{18}{23}$
50	PRrtefeuille_■_GepGiquant	$\frac{3}{20}$
60	Purtefeuille_■_cepliquant	$\frac{22}{23}$
70	Portefeuille_■_RKpliquant	$\frac{22}{23}$
74	Portefeuille_■_Repliquant	$\frac{23}{23}$

TABLE 4.1 – Meilleurs individus de 10 générations

La moyenne de la fonction d'évaluation à maximiser par génération apparaît figure 4.5. On observe que cette moyenne ne fait qu'augmenter jusqu'à ce que la solution désirée soit atteinte pour un individu de la génération 74. L'algorithme s'achève alors car la solution est obtenue en moins de 100 générations : le critère d'arrêt de l'algorithme est vérifié.

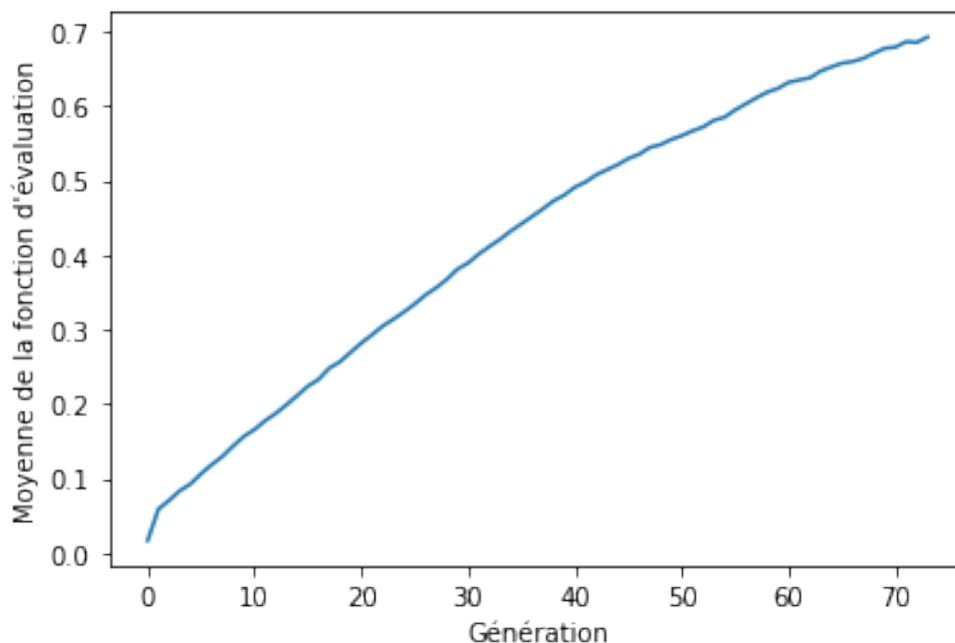


FIGURE 4.5 – Moyenne de la fonction d'évaluation par génération

Sur cet exemple simple, l'algorithme témoigne de bonnes performances d'optimisation. En effet, le résultat ciblé est obtenu en un temps limité (8 secondes). Afin de voir les performances

de l'algorithme génétique, comparons le désormais à l'aide de cet exemple face à une recherche aléatoire. Le critère de performance observé sera le temps de calcul.

L'environnement \mathcal{E} est l'ensemble des combinaisons possibles de $P = 23$ paramètres du problème. Sachant qu'il y a 60 possibilités pour chaque paramètre, $\dim(\mathcal{E}) = 60^{23} \simeq 7.9 \times 10^{40}$. Il existe donc près de 7.9×10^{40} combinaisons de 23 caractères avec les 60 caractères possibles.

Soit $(X_i)_{i \in \{1, \dots, P\}}$ une suite de variables aléatoires indépendantes et identiquement distribuées. Pour tout $i \in [1, P]$, $X_i \sim \mathcal{U}[1, 60]$ est le caractère situé en i^{eme} position. Alors $X = \{X_1, \dots, X_{23}\} \in \mathcal{E}$ est la variable aléatoire qui représente une séquence de $P = 23$ caractères. On a :

$$\begin{aligned} \mathbb{P}(X = \textit{Portfeuille Repliquant}) &= \mathbb{P}(X_1 = P, X_2 = o, X_3 = r, X_4 = t \dots, X_{23} = t) \\ &= \mathbb{P}(X_1 = P) \times \mathbb{P}(X_2 = o) \times \mathbb{P}(X_3 = r) \times \dots \times \mathbb{P}(X_{23} = t) \\ &= \frac{1}{60^{23}} \\ &= 1.27 \times 10^{-41}. \end{aligned}$$

Ainsi, la probabilité d'obtenir la combinaison de 23 caractères *Portfeuille Repliquant* dans cet ordre est infinitésimal.

Soit Y la variable aléatoire qui vaut 1 si $X = \textit{Portfeuille Repliquant}$ et 0 sinon.

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{P}(X = \textit{Portfeuille Repliquant}) \\ &= \frac{1}{60^{23}}. \end{aligned}$$

En moyenne, 60^{23} ensembles composés de 23 caractères sont nécessaires pour espérer avoir la combinaison *Portfeuille Repliquant*.

Sachant que sur Python, 24673 ensembles de 23 caractères peuvent être générés en 1 seconde, il faudrait en moyenne 10^{28} années pour obtenir la combinaison de 23 caractères désirée. Cet exemple témoigne de la pertinence de l'algorithme génétique lorsque l'environnement \mathcal{E} est très grand.

4.5 Les forces et faiblesses de l'algorithme

Bien que l'algorithme génétique soit un algorithme de machine learning élaboré il y a plus d'une soixantaine d'années, son utilisation reste apprécié pour diverses problématiques. Ceci témoigne de certains avantages liés à cette méthode.

Dans leurs travaux, Lerman et Ngouenet [17] présentent **quatre points** qui distinguent les algorithmes génétiques des autres algorithmes de machine learning.

Tout d'abord, les algorithmes génétiques offrent la possibilité d'utiliser un **codage de paramètre** et non le paramètre en lui-même. Ce premier point n'est ni une force, ni une faiblesse mais est une différence notable avec la plupart des autres algorithmes communément étudiés.

Ensuite, les algorithmes génétiques fonctionnent par génération, et donc par **ensemble de solutions potentielles**. Ainsi, dans certains problèmes d'optimisation pour lesquels la solution

exacte n'est pas recherchée, il est possible de considérer différents individus de la génération finale. La recherche se fait par ensemble de points et non pas par point unique comme pour les algorithmes plus classiques. Cela a pour avantage de proposer plusieurs solutions pour un algorithme génétique étudié.

Une autre force de l'algorithme génétique réside dans l'utilisation de la fonction étudiée. En effet, contrairement à bon nombre d'autres algorithmes, il n'est **pas nécessaire d'avoir davantage d'information sur la fonction étudiée** (dérivé, convexité, ...). Seules les valeurs de celle-ci sont nécessaires.

Enfin, l'algorithme génétique utilise des **mécanismes simples** issus de l'observation de la nature, il est **aisément mis en application** et son utilisation dans de nombreux domaines scientifiques témoigne de son **efficacité**.

L'algorithme génétique possède toutefois des défauts. Le principal défaut est la **convergence vers des extremum locaux**. Cette convergence est due à l'exploration d'une même région de l'environnement \mathcal{E} , en particulier lorsque des individus I très similaires apparaissent en nombre au sein des différentes générations. Les nombreux gènes communs entre les bons individus augmentent ce risque. L'opérateur évolutionniste de **mutation** doit alors être exploité correctement afin de s'éloigner d'un possible extrema local.

En conclusion, les algorithmes génétiques présentent plusieurs avantages. Ils sont simples d'utilisation et sont très appréciés lorsque l'espace de recherche \mathcal{E} est très grand. Son application au problème de sélection d'instruments dans le cadre des portefeuilles de réplcation semble donc appropriée.

L'APPLICATION DE L'ALGORITHME GÉNÉTIQUE AU CAS DES PORTEFEUILLES RÉPLIQUANTS

L'algorithme génétique est un algorithme d'optimisation. Ce chapitre présente l'application de cet algorithme à la sélection d'instruments financiers dans le problème d'optimisation des portefeuilles de réplication.

5.1 Les données utilisées

Afin de pouvoir appliquer la méthode de réplication, trois types de données sont nécessaires :

- le passif à répliquer Y
- les instruments financiers facilement valorisables
- les sensibilités financières.

5.1.1 Le passif à répliquer

Le premier type de données correspond au passif Y à répliquer. Axa France est composé de plusieurs fonds. Par souci de confidentialité, ces fonds ne seront pas nommés. Dans ce chapitre, les résultats présentés porteront sur la réplication du passif d'un fonds noté *portefeuille A*. Y est un vecteur de taille 4000 qui correspond aux 4000 simulations en monde risque neutre.

5.1.2 Les instruments financiers facilement valorisables

Un portefeuille répliquant consiste à utiliser une combinaison d'instruments financiers simples à valoriser afin d'approximer le passif.

Les différents instruments à disposition pour cette étude, leurs caractéristiques et les effets qu'ils capturent ont été présentés en détail au chapitre 2, section 2.2.

Le tableau 5.1 ci-dessous est un rappel des types d'instruments disponibles.

Type d'instrument	Effet capturé	Paramètres
Zero-coupon	Garanties	Maturité
Zero-coupon inflation	Sensibilité à l'inflation	Maturité
Swaption receveur	Convexité due aux rachats et réinvestissements	Maturité, tenor, strike
Swaption payeur	Convexité due aux rachats et réinvestissements	Maturité, tenor, strike
Indice equity	Exposition directe	Maturité, tenor
Call sur equity	Participations aux bénéfiques	Maturité, strike
Put sur equity	Participations aux bénéfiques	Maturité, strike

TABLE 5.1 – Synthèse des instruments utilisables, les effets qu'ils capturent et leurs paramètres

Déterminer la base de travail de l'étude consiste à déterminer l'ensemble des instruments financiers potentiellement intégrables à un portefeuille de réplication candidat. Il faut alors déterminer, pour chaque type d'instrument présent dans le tableau 5.1 si :

1. ce type d'instrument est retenu pour le fonds étudié
2. l'intervalle de maturité possible et le pas associé
3. l'intervalle de tenor possible et le pas associé si ce type d'instrument possède un tenor
4. l'intervalle de strike possible et le pas associé si ce type d'instrument possède un strike.

Les résultats présentés dans ce chapitre, pour approximer le passif du portefeuille A, sont obtenus à partir de la base de données formée des instruments de la table 5.2. L'intervalle et le pas des tenor ne sont volontairement pas renseignés ici car ils sont déduits de la maturité et de l'horizon de travail de 60 années.

Type d'instrument	Paramètre	Intervalle	Pas
Zero-coupon	Maturité	[1,60]	1
Zero-coupon inflation	Maturité	[10,10]	
Swaption receveur	Maturité	[7,50]	5
	Tenor	-	-
	Strike	[1%,5%]	1%
Swaption payeur	Maturité	[7,50]	5
	Tenor	-	-
	Strike	[1%,5%]	1%
Indice equity (×6)	Maturité	[1,60]	5
	Tenor	-	-
Call sur equity (×6)	Maturité	[1,60]	5
	Strike	[80%,120%]	5%
Put sur equity (×6)	Maturité	[1,60]	5
	Strike	[80%,120%]	5%

TABLE 5.2 – Description de la base de données composée des instruments financiers

Cela représente un total de $p_{max} = 4178$ instruments financiers.

5.1.3 Les sensibilités

Les sensibilités financières utilisées pour ce chapitre sont au nombre de 58. Elles représentent des chocs financiers et servent à vérifier que le portefeuille répliquant se comporte bien comme le passif dans des situations de stress de marché.

Certaines de ces sensibilités doivent obligatoirement être vérifiées afin que le contrôle interne valide l'utilisation du portefeuille répliquant déterminé. D'autres sensibilités sont facultatives, mais leurs validations restent préférables.

5.2 La définition des paramètres du problème

Les fonds étudiés présentent des caractéristiques de participation aux bénéfices et de rachats qui écartent une possible réplification parfaite. Dans l'optique d'une approximation du passif, l'algorithme génétique permet alors d'obtenir un ensemble de solutions potentielles. Afin de pouvoir prendre en compte les différentes spécificités liées à la méthode des portefeuilles de réplification, l'algorithme génétique n'a pas été utilisé à l'aide d'un package mais a été entièrement implémenté. Cela permet une plus grande liberté dans le choix de la représentation des paramètres du problème.

5.2.1 Les gènes

Comme présenté dans le chapitre précédent, les gènes correspondent aux paramètres à déterminer du problème d'optimisation. Afin de pouvoir correctement définir les gènes de l'algorithme génétique, il convient de commencer par définir les objectifs recherchés.

Tout d'abord, le premier élément recherché est un ensemble d'instruments financiers capable de répliquer le passif d'assurance. Cet ensemble d'instruments financiers sera par conséquent un gène du problème d'optimisation.

Ensuite, dans le chapitre *Les difficultés de la méthode des portefeuilles répliquants*, une limite a été énoncée concernant la méthode de calcul actuelle des nominaux. Cette limite concerne la détermination des paramètres θ , λ et du vecteur de poids k associé aux sensibilités dans la fonction de coût $C_{\theta,\lambda,k}(\beta)$. En effet, il n'existe pas de méthode permettant de déterminer ces paramètres qui sont ainsi évalués à *dire d'expert*. Afin de surmonter cette limite, θ , λ et le vecteur k seront intégrés aux paramètres du problème d'optimisation de l'algorithme génétique.

Les gènes de l'algorithme génétique seront alors formés d'un ensemble d'instruments financier et des valeurs de θ , λ et du vecteur k .

Maintenant que les gènes ont été définis, il est nécessaire de déterminer la manière dont ils seront représentés. Cette représentation est effectivement importante car c'est sur celle-ci que s'appuient les opérateurs d'évolution *croisement* et *mutation*. Une représentation adéquate réduit la complexité de ces deux opérateurs d'évolution.

La représentation de l'ensemble des instruments

Pour représenter les instruments retenus, le choix effectué a été de créer un vecteur g_1 de taille p_{max} , taille de la base de données composée de tous les instruments possibles (section 5.1.2) :

$$g_1 = \begin{pmatrix} g_{1,1} \\ \dots \\ g_{p_{max},1} \end{pmatrix} \in \mathbb{R}^{p_{max}}.$$

Chaque coefficient de ce vecteur ne prendra que 2 valeurs : $\forall 1 \leq i \leq p_{max}$,

$$g_{i,1} = \begin{cases} 1 & \text{si l'instrument est retenu pour cet ensemble candidat,} \\ 0 & \text{si il n'est pas retenu.} \end{cases}$$

Il s'agit d'une représentation binaire d'un ensemble d'instruments financiers qui est illustré en tableau 5.3.

0	0	1	0	0	...	0	0	0	1	0	1
---	---	---	---	---	-----	---	---	---	---	---	---

TABLE 5.3 – Exemple de représentation d'un gène *instruments financiers*

Cette représentation possède plusieurs avantages qui sont en particulier liés à l'utilisation des opérateurs de croisement et de mutation.

Le principal avantage est que cette représentation permet de ne pas avoir un nombre fixe d'instruments financiers. En effet, le vecteur g_1 de taille constante p_{max} , a comme nombre d'instruments financiers la somme de ses éléments :

$$p = \sum_{i=1}^{p_{max}} g_{i,1}.$$

Il est ainsi aisé de faire évoluer le nombre d'instruments avec les opérateurs de croisement et de mutation sans avoir à modifier la structure du gène g_1 . Le mécanisme de modification du nombre d'instruments ainsi que les autres avantages liés à cette représentation seront présentés en détail dans les sous-sections croisement et mutation.

Enfin, il est important de souligner que chaque coefficient de ce premier gène correspond à un instrument en particulier. L'ordre des instruments a son importance et cette notion d'ordre sera introduite dans la sous-section concernant l'opérateur de croisement. De plus, il est possible de noter que le nombre d'instruments sélectionnés est très petit devant la taille du vecteur : $p \ll p_{max}$. Il y a donc une sous-représentation des 1 par rapport à la quantité de 0 au sein d'un tel vecteur.

La représentation de θ

Le paramètre θ appartient à l'intervalle $[0, 1]$. Il permet de faire un arbitrage entre l'ajustement des sensibilités financières et celle du scénario central. Plus θ se rapproche de 1, plus les sensibilités sont favorisées.

La représentation du gène θ est un nombre réel compris dans l'intervalle $[a_\theta, b_\theta]$ où $(a_\theta, b_\theta) \in [0, 1]^2$ sont les bornes de l'intervalle laissées en paramètres de l'algorithme génétique. L'utilisateur aura ainsi le contrôle sur l'espace des valeurs possibles de θ .

La représentation de λ

Le paramètre λ correspond au paramètre de pénalisation dans la régression ridge. C'est un nombre réel qui est en pratique très petit dans le cadre de la méthode des portefeuilles répliquants. L'intervalle de valeurs de λ est par expérience le suivant : $\lambda \in [5.10^{-7}, 8.10^{-6}]$.

Comme pour le gène θ , le gène λ est un nombre réel compris dans l'intervalle $[a_\lambda, b_\lambda]$ où $(a_\lambda, b_\lambda) \in [5.10^{-7}, 8.10^{-6}]^2$ sont les bornes de l'intervalle laissées en paramètre de l'algorithme génétique. L'utilisateur aura de nouveau le contrôle sur l'espace des valeurs possibles de λ .

La représentation du vecteur de poids k

Le vecteur de poids $k = \begin{pmatrix} k_1 \\ \vdots \\ k_{n_S} \end{pmatrix}$ tel que $\sum_{j=1}^{n_S} k_j = 1$ permet d'attribuer une importance différente entre les $n_S = 58$ sensibilités financières.

Ce vecteur de poids k sera représenté par un vecteur g_4 de taille n_S , nombre de sensibilités financières :

$$g_4 = \begin{pmatrix} g_{1,4} \\ \vdots \\ g_{n_S,4} \end{pmatrix} \in \mathbb{R}^{n_S}.$$

Chaque coefficient de ce vecteur pourra prendre les valeurs suivantes : $\forall 1 \leq i \leq n_S$,

$$g_{i,4} = \frac{k_i}{\sum_{j=1}^{n_S} k_j} \text{ avec } k_i \in \mathbb{N}.$$

Une représentation possible du gène g_4 , où chaque sensibilité a le même poids, est exposée en tableau 5.4 ci-dessous.

$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	\dots	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$
----------------	----------------	----------------	----------------	----------------	---------	----------------	----------------	----------------	----------------	----------------	----------------

TABLE 5.4 – Exemple de représentation d'un gène *vecteur de poids k*

5.2.2 L'individu

Par définition, un individu est un ensemble de gènes. Par conséquent, tout individu de l'algorithme génétique sera composé de :

- g_1 : le gène représentant un ensemble d'instruments financiers
- g_2 : le gène représentant le paramètre θ
- g_3 : le gène représentant le paramètre λ
- g_4 : le gène représentant le vecteur de poids k .

Afin d'avoir un bon suivi de l'algorithme génétique, il a été décidé de compléter la représentation d'un individu par un identifiant unique.

Cet identifiant a plusieurs intérêts :

- identifier les individus de façon unique
- suivre les résultats en cours d'algorithme

- vérifier la cohérence de l'évolution en début d'algorithme : on s'attend à ce que les nouveaux individus en haut de classement surpassent les anciens en début d'algorithme *a minima*
- définir un critère d'arrêt basé sur l'apparition sur plusieurs générations successives des meilleurs individus en haut de classement.

L'identification commence à 1 et est incrémentée d'une unité à chaque nouvel individu.

5.2.3 La génération

Une génération est un ensemble d'individus à un instant donné. Les individus étant eux-mêmes un ensemble formé des 4 gènes précédemment présentés dans cette section.

La manière dont est définie la première génération est importante car c'est à partir d'elle que découle l'ensemble des solutions potentielles que représentent les générations futures. Dans les problématiques usuelles nécessitant l'utilisation d'un algorithme génétique, la coutume est de définir la génération initiale de façon aléatoire. Cependant, le cas des portefeuilles répliquants étant très spécifique, il est nécessaire de définir la génération initiale en imposant quelques contraintes.

La génération initiale

La génération initiale G_0 est une première solution possible au problème. Il est donc indispensable que les exigences potentielles concernant les solutions soient valides dès la première génération. La façon dont a été définie cette génération G_0 est présentée ici.

Le premier paramètre à définir est le nombre d'individus n_I qui composent une génération. Ce nombre est fixe de génération en génération. Ensuite, chaque individu I_0 composant la génération initiale est généré par le même procédé. Etant défini par leurs gènes, définissons les individus de G_0 en définissant chacun de leurs gènes.

Les instruments financiers

Le gène g_1 associé aux instruments financiers est un vecteur de taille p_{max} . Dans ce vecteur, la position unique d'un coefficient correspond à la position unique d'un instrument financier en particulier. La position d'un instrument n'est pas déterminée à partir de la base de travail de départ. En effet, dans cette base de travail, les instruments sont regroupés par classes. Les zéro-coupons sont par exemple regroupés et sont situés au début de la base. Le choix retenu a été d'attribuer de façon aléatoire une position à tous les instruments financiers candidats. Les positions sont mémorisées pour tout individu de toute génération de l'algorithme. Cela a pour avantage d'éviter les dépendances entre les instruments lors de la phase de croisement.

Afin de capturer l'effet de l'inflation, il est obligatoire que chaque portefeuille répliquant candidat possède le zéro-coupon inflation 10 ans. Pour satisfaire ce besoin, la première étape est de déterminer la position i_{zc_infl} associée au zéro-coupon inflation 10 ans. Chaque individu de chaque population aura son gène g_1 tel que le coefficient i_{zc_infl} ait la valeur 1.

Des flux de trésorerie sont certains. Il est donc indispensable d'avoir un minimum de zéro-coupons au sein d'un portefeuille de réplification candidat en début d'algorithme *a minima*. Effectuer une sélection aléatoire des instruments pourrait mener à une sous-représentativité des zéro-coupons dans la population initiale G_0 . En effet, la proportion d'instruments diffère

d'une classe d'instruments à l'autre. Ainsi, par défaut, l'algorithme propose 5 zéro-coupons choisis aléatoirement parmi les zéro-coupons de la base de travail pour chaque individu de la première génération. Ce nombre est laissé en paramètre de l'algorithme génétique ce qui offre une liberté à l'utilisateur. Il est important de noter que cette contrainte est imposée au niveau de la génération initiale. Elle peut ne plus être respectée lors des générations futures.

Tous les individus de la première génération G_0 ont le même nombre d'instruments initial p_0 . Ce nombre va pouvoir évoluer à la hausse ou à la baisse grâce aux opérateurs de croisement et de mutation et pourra ainsi varier entre les individus d'une même génération $G_{i,i \geq 1}$. Néanmoins, pour la première génération G_0 , le nombre d'instruments p_0 est égal pour tout individu $I_0 \in G_0$ et est un paramètre défini par l'utilisateur de l'algorithme.

Parmi ces p_0 instruments, 6 sont déjà connus. Il s'agit du zéro-coupon inflations de maturité 10 ans, commun à tous les individus, et de 5 zéro-coupons dont les maturités peuvent ne pas être égales pour chaque individu. Les $p_0 - 6$ autres instruments sont sélectionnés de manière aléatoire parmi les instruments non retenus jusqu'alors. Il s'agit d'une sélection aléatoire sans remise. Un individu de la génération initiale peut par conséquent détenir plus de 5 zéro-coupons.

Le paramètre θ

Pour chaque individu $I_0 \in G_0$, le paramètre θ est choisi de façon aléatoire dans l'intervalle $[a_\theta, b_\theta]$ où $(a_\theta, b_\theta) \in [0, 1]^2$ sont les bornes de l'intervalle laissées en paramètres de l'algorithme génétique.

Le paramètre λ

Pour chaque individu $I_0 \in G_0$, le paramètre λ est choisi de façon aléatoire dans l'intervalle $[a_\lambda, b_\lambda]$ où $(a_\lambda, b_\lambda) \in [0, 1]^2$ sont les bornes de l'intervalle laissées en paramètres de l'algorithme génétique.

Le vecteur de poids k

Le gène g_4 associé aux poids des sensibilités financières est un vecteur de taille $n_S = 58$. Dans ce vecteur, chaque coefficient correspond au poids d'une des 58 sensibilités financières étudiées. L'ordre des sensibilités financières n'aura cette fois-ci aucune importance sur les opérateurs de croisement et de mutation définis.

Le choix retenu pour définir les gènes g_4 de tous les individus de la génération initiale est d'accorder le même poids aux sensibilités financières : $\forall 1 \leq i \leq n_S, g_{i,4} = \frac{1}{n_S}$.

$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	\dots	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$	$\frac{1}{58}$
----------------	----------------	----------------	----------------	----------------	---------	----------------	----------------	----------------	----------------	----------------	----------------

Ce choix a été effectué car en pratique, une majorité de sensibilités doivent avoir le même poids. Les quelques modifications sur la répartition des poids sont laissées à l'opérateur de mutation au fil des générations.

L'identification des individus

Le premier individu de la génération initiale est identifié par le matricule 1. Pour chaque nouvel individu, son matricule est incrémenté d'une unité par rapport au matricule du dernier

individu généré.

Le Critère d'arrêt

La méthode des portefeuilles de réplication vise ici à obtenir une approximation du passif. Dans cette optique, il est nécessaire de définir un bon critère d'arrêt afin que l'algorithme ne tourne pas indéfiniment mais qu'il apporte de bon résultat. Une solution simple et facile à mettre en place est de définir :

1. un nombre de générations maximum n_G
2. un nombre n_0 d'occurrences de n_{best} meilleurs individus au cours des n_0 dernières générations

Prenons l'exemple de $n_G = 10$, $n_0 = 5$ et $n_{best} = 3$. Alors l'algorithme s'arrêtera si les $n_{best} = 3$ meilleurs individus apparaissent en tête de classement lors de $n_0 = 5$ générations consécutives. Si ce cas de figure ne se présente pas, l'algorithme s'arrêtera après avoir généré les individus de la 10^{ème} génération.

Ce critère d'arrêt permet à l'utilisateur d'avoir un contrôle sur le temps de calcul qu'il est prêt à accorder à l'algorithme afin de trouver un portefeuille candidat.

5.2.4 L'environnement

L'environnement \mathcal{E} est l'espace de définition des solutions potentielles du problème. Ces solutions peuvent ne pas répondre aux contraintes imposées par la problématique des portefeuilles de réplication :

$$\mathcal{E} = \mathcal{F} \sqcup \mathcal{U}$$

où \mathcal{F} est le sous-espace des solutions vérifiant toutes les contraintes et \mathcal{U} est le sous-espace ne vérifiant pas au moins une des contraintes. \mathcal{E} représente l'ensemble des combinaisons de p instruments ($p \in \{1, \dots, p_{max}\}$), de θ , λ , et du vecteur de poids k . Cet espace est par définition très grand.

5.2.5 Le processus de croisement

Le croisement est le processus qui simule le phénomène de reproduction au cours de l'évolution. C'est avec cet opérateur que de nouveaux individus sont créés et que le mélange des gènes, et donc le mélange d'information, se fait.

Les individus de chaque génération sont séparés de façon aléatoire en deux listes. Les individus de la première liste forment les $parent_1$ et ceux de la seconde liste forment les $parent_2$. Chaque couple de 2 individus parents va engendrer 4 enfants. Si la génération G_0 possède $n_I = 100$ individus, alors 50 couples sont formés aléatoirement et engendreront $4 \times 50 = 200$ nouveaux individus.

Le procédé sélectionné pour effectuer les croisements dépend du gène considéré.

Les instruments financiers

Le gène des instruments financiers est de grande taille (p_{max}) et présente une sous-représentativité des 1 par rapport aux 0. Le processus retenu pour le croisement des gènes des instruments fi-

nanciers a donc été le croisement multiple par blocs.

Le procédé se décompose selon les étapes suivantes :

1. Choix du nombre de blocs n_{blocs} en paramètre de l'algorithme génétique
2. Choix de la probabilité de croisement $proba_C^{g_1}$ en paramètre de l'algorithme génétique
3. Générer $n_{blocs} - 1$ nombres entiers aléatoires compris entre 0 et p_{max} qui détermineront les sites de croisement par blocs.
Ils permettent de compartimenter le gène en n_{blocs} .
4. Pour chacun des n_{blocs} blocs :
 - (a) Générer un nombre compris entre 0 et 1
 - (b) Si ce nombre est inférieur à la probabilité de croisement $proba_C^{g_1}$, alors le bloc de $g_1^{enfant_1}$ reçoit le bloc du $g_1^{parent_2}$ et le bloc de $g_1^{enfant_2}$ reçoit le bloc du $g_1^{parent_1}$
 - (c) Sinon, le bloc de $g_1^{enfant_1}$ reçoit le bloc du $g_1^{parent_1}$ et le bloc de $g_1^{enfant_2}$ reçoit le bloc du $g_1^{parent_2}$
5. Recommencer 1 fois les étapes 1 à 4 pour générer 2 autres gènes g_1 de 2 autres enfants.

La figure 5.1 synthétise le procédé de croisement pour le gène des instruments financiers. Ce processus permet de faire évoluer le nombre d'instruments financiers grâce à l'échange par blocs.

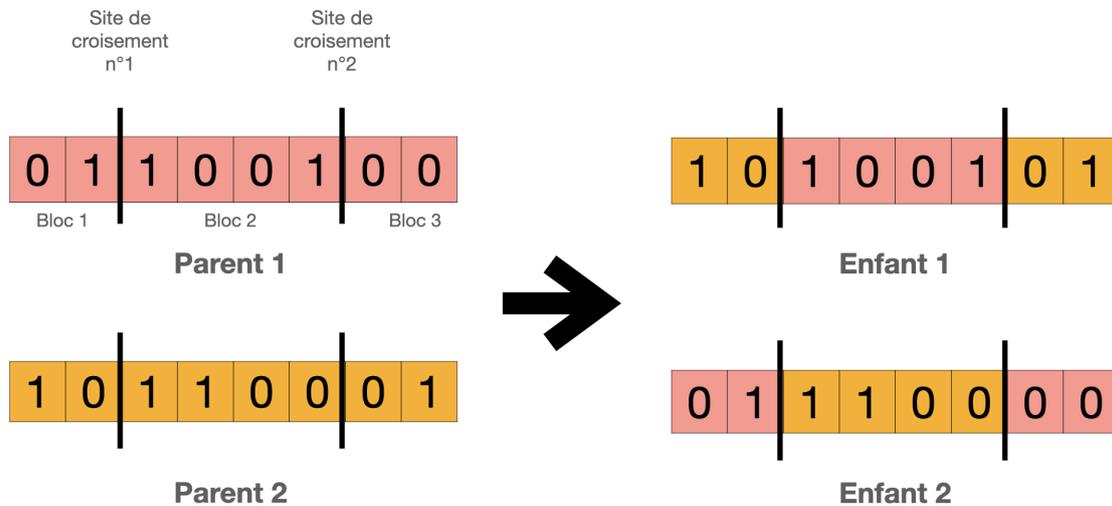


FIGURE 5.1 – Exemple d'un croisement par blocs pour $n_{blocs} = 3$

Les paramètres θ et λ

Le couple de 2 parents va engendrer 4 enfants. Chacun de ces enfants héritera aléatoirement du paramètre θ ou λ de l'un des deux parents.

Le vecteur de poids k

Le gène associé au vecteur de poids est de taille $n_S = 58$. Comme pour le gène des instruments financiers, le procédé de croisement retenu est le croisement multiple mais cette fois-ci

en chacune des positions associées aux sensibilités. C'est plus précisément les numérateurs qui subiront un échange d'information. Au préalable, les coefficients des poids des sensibilités financières auront été mis sous forme de fraction de même dénominateur.

Le procédé se décompose selon les étapes suivantes :

1. Choix de la probabilité de croisement $proba_C^{g_4}$ en paramètre de l'algorithme génétique
2. Pour chacune des 58 positions associées aux 58 sensibilités financières :
 - (a) Générer un nombre entre 0 et 1
 - (b) Si ce nombre est inférieur à la probabilité de croisement $proba_C^{g_4}$, alors la position de $g_4^{enfant_1}$ reçoit le numérateur de $g_4^{parent_2}$ et la position de $g_4^{enfant_2}$ reçoit le numérateur de $g_4^{parent_1}$
 - (c) Sinon la position de $g_4^{enfant_1}$ reçoit le numérateur de $g_4^{parent_1}$ et la position de $g_4^{enfant_2}$ reçoit le numérateur de $g_4^{parent_2}$
3. Ajuster les dénominateurs de tous les coefficients de sorte que la somme des poids soit égale à 1.
4. Recommencer 1 fois les étapes 1 à 3 pour générer 2 autres gènes g_4 de 2 autres enfants.

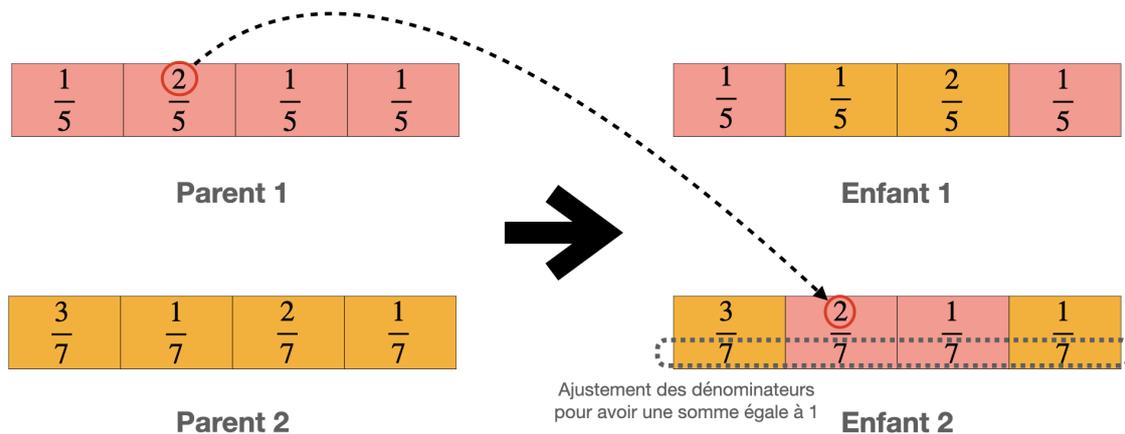


FIGURE 5.2 – Exemple d'un croisement de poids pour $n_S = 4$ sensibilités

5.2.6 Le processus de mutation

La mutation est le procédé qui va apporter de la **diversité** au sein d'une population. C'est l'application du phénomène de **variation** dans la théorie de l'évolution. Grâce à cet opérateur, il sera possible de **s'éloigner d'un extremum local** et d'**éviter une convergence prématurée** vers une solution non optimale.

Ce processus intervient après le processus de croisement dans la chronologie de l'algorithme.

Le procédé retenu dépend également du gène considéré.

Les instruments financiers

Un couple de 2 parents génère 4 enfants. La mutation s'opère sur les enfants et suit le procédé suivant :

1. Choisir la probabilité de mutation $proba_M^{g_1}$ en paramètre de l'algorithme
2. Générer un nombre aléatoire entre 0 et 1
3. Si ce nombre est inférieur à $proba_M^{g_1}$:
 - (a) Générer un nombre entier parmi $\{1,2,3\}$ tel que $\mathbb{P}(1) = \frac{7}{10}$, $\mathbb{P}(2) = \frac{2}{10}$ et $\mathbb{P}(3) = \frac{1}{10}$. Ce nombre entier correspond au nombre de mutations à effectuer.
 - (b) Pour le nombre de mutations à effectuer :
 - i. Générer aléatoirement un nombre entier compris entre 0 et $p_{max} - 1$. Ce nombre correspond à un instrument via sa position et ne peut pas correspondre à la position du zéro-coupon inflation 10 ans dont la présence est obligatoire.
 - ii. Si le coefficient associé à cette position est de 1, le mettre à 0 : on retire l'instrument
 - iii. Sinon, le coefficient associé à cette position est de 0, le mettre à 1 : on rajoute l'instrument.

La figure 5.3 illustre ce processus.

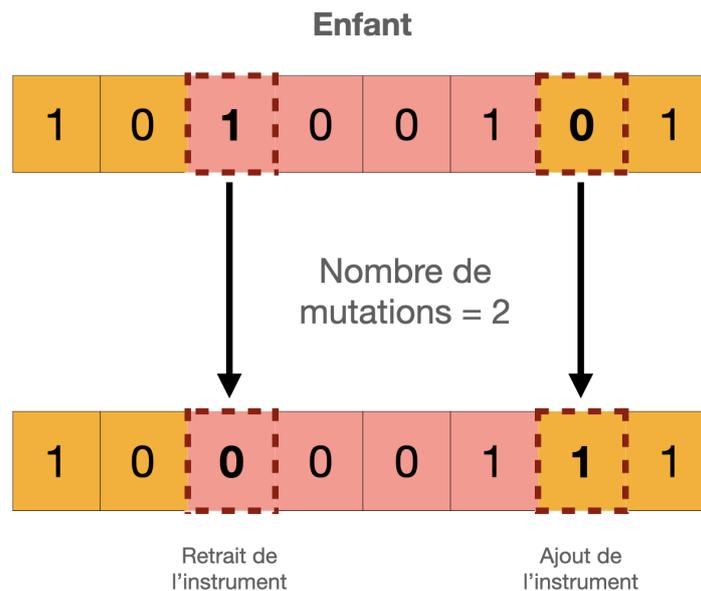


FIGURE 5.3 – Exemple de mutations d'instruments financiers

Les paramètres θ et λ

La mutation s'opère également sur les enfants. Cette fois-ci, la mutation simple est utilisée car les gènes associés aux paramètres θ et λ n'ont qu'une seule composante.

Une probabilité $proba_M^{g_2, g_3}$ est choisie par l'utilisateur en paramètre de l'algorithme génétique. Ensuite, pour chaque enfant, deux nombres compris entre 0 et 1 sont générés, un pour θ et un pour λ .

Si le premier nombre est inférieur à $proba_M^{g_2, g_3}$, alors un nouveau $\theta \in [a_\theta, b_\theta]$ est attribué à l'enfant considéré.

Si le second nombre est inférieur à $proba_M^{g_2, g_3}$, alors un nouveau $\lambda \in [a_\lambda, b_\lambda]$ est attribué à l'enfant considéré.

Le vecteur de poids k

La mutation s'opère de la manière suivante sur chaque enfant :

1. Choix de la probabilité de mutation $proba_M^{g^4}$ en paramètre de l'algorithme
2. Pour chaque sensibilité :
 - (a) Générer un nombre aléatoire entre 0 et 1.
 - (b) Si ce nombre est inférieur à $proba_M^{g^4}$:
 - i. Générer un nombre entier parmi $\{1,2,3,4,5\}$ tel que $\mathbb{P}(1) = 80\%$, $\mathbb{P}(2) = 10\%$, $\mathbb{P}(3) = 5\%$, $\mathbb{P}(4) = 3\%$ et $\mathbb{P}(5) = 2\%$.
 - ii. Attribuer ce nombre au numérateur.
 - iii. Ajuster les dénominateurs de tous les coefficients de sorte que la somme des poids soit égale à 1.

La figure 5.4 illustre la mutation des poids des sensibilités financières.

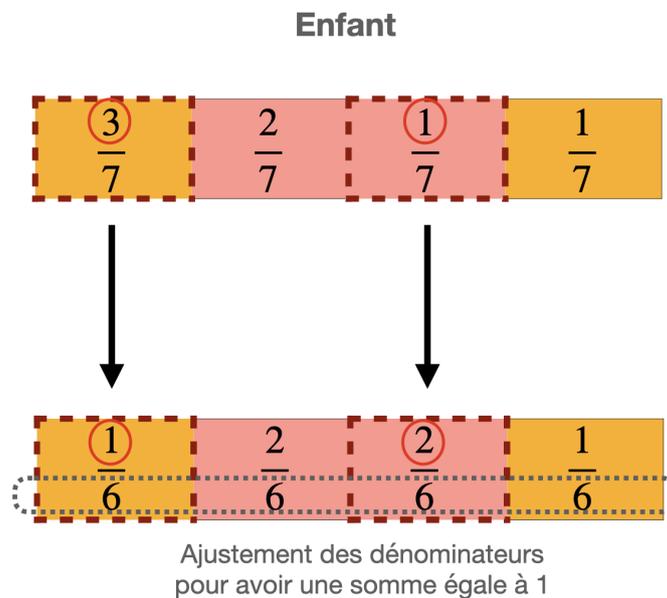


FIGURE 5.4 – Exemple de mutations de poids pour $n_S = 4$ sensibilités

5.2.7 Le processus de sélection

Le processus de sélection consiste à retenir certains individus afin d'avoir un ensemble de solutions capable de converger vers l'optimum du problème posé. C'est l'application directe du phénomène d'adaptation à un environnement dans la théorie évolutionniste de Charles Darwin. Sous réserve d'un bon choix de la fonction d'évaluation, les individus retenus sont les mieux adaptés à leur environnement.

Le processus de sélection exposé ici repose sur le principe de stabilité du nombre d'individus par génération : n_I est constant tout au long de l'algorithme.

Comme chaque couple de parents engendre 4 enfants, après le processus de croisement, il y a $n_I + 4 \times \frac{n_I}{2}$ individus au total.

Avec $n_I = 100$, il y a par exemple 300 individus après les processus de croisement et de mutation : 100 parents et 200 enfants.

Afin de conserver les meilleurs individus de la génération précédente, le choix a été fait de retenir l'ensemble de ces $3 \times n_I$ individus en amont du processus de sélection : un parent peut ainsi survivre et appartenir à la génération suivante. Cela présente l'avantage de conserver les individus très performants qui auraient pu engendrer de mauvais descendants par croisement avec un parent moins performant.

La technique de sélection retenue est une combinaison des méthodes de sélection par rang et de sélection uniforme. Le principe est le suivant :

- Choix du nombre d'individus par génération n_I en paramètre de l'algorithme génétique
- Sélection des $\lfloor 0.8 \times n_I \rfloor$ meilleurs individus parmi les parents et enfants pour former la génération suivante
- Des individus sont choisis aléatoirement parmi la population restante. Ils permettent de compléter la génération suivante jusqu'à obtenir n_I individus et favorise la diversité.

La figure 5.5 schématise ce procédé.

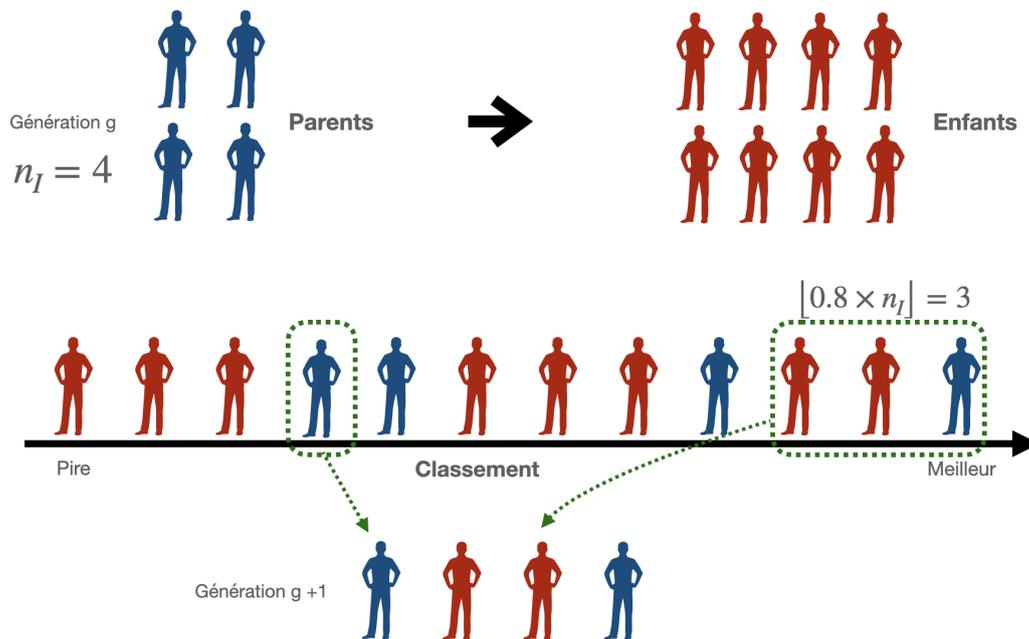


FIGURE 5.5 – Exemple du processus de sélection pour $n_I = 4$ individus par génération

5.2.8 Synthèse de l'ensemble des paramètres de l'algorithme génétique

Beaucoup de paramètres de l'algorithme génétique ont été énoncés dans cette section. Dans un souci de clarté, le tableau 5.5 permet de les synthétiser.

Nom	Description	Utilité
$[a_\theta, b_\theta]$	Bornes des valeurs possibles de θ	Initialisation
$[a_\lambda, b_\lambda]$	Bornes des valeurs possibles de λ	Initialisation
n_I	Nombre d'individus que contient chaque génération	Initialisation
$ZC_{initial}$	Nombre de zéro-coupons initial, par défaut égal à 5	Initialisation
p_0	Nombre d'instruments initial pour tous les individus de la génération initiale G_0	Initialisation
n_G	Nombre de générations maximum	Critère d'arrêt
n_{best}	Nombre de meilleurs individus devant apparaître en tête de classement pour valider le critère d'arrêt	Critère d'arrêt
n_0	Nombre d'occurrences de générations permettant de valider le critère d'arrêt	Critère d'arrêt
n_{blocs}	Nombre de blocs pour compartimenter le gène instruments financiers	Croisement
$proba_C^{g1}$	Probabilité de croisement pour le gène instruments financiers	Croisement
$proba_C^{g4}$	Probabilité de croisement pour le gène poids des sensibilités financières	Croisement
$proba_M^{g1}$	Probabilité de mutation pour le gène instruments financiers	Mutation
$proba_M^{g2,g3}$	Probabilité de mutation pour les gènes θ et λ	Mutation
$proba_M^{g4}$	Probabilité de mutation pour le gène poids des sensibilités financières	Mutation

TABLE 5.5 – Ensemble des paramètres de l'algorithme génétique

A ces paramètres s'ajoutent deux paramètres qui sont définis en amont de l'algorithme : la taille de la base de travail p_{max} et le nombre de sensibilités financières n_S .

5.3 Le choix de la fonction de coût

La fonction de coût a un rôle primordial car c'est elle qui permet de faire le lien entre le problème d'optimisation et l'algorithme de résolution. En ordonnant les individus d'une génération, elle sélectionne ceux qui auront de meilleures chances de survie et de reproduction, et permet ainsi de se rapprocher d'une des solutions du problème et donc d'un portefeuille de répliation candidat pertinent.

Rappelons qu'un portefeuille de répliation doit obligatoirement vérifier les indicateurs de performances clés (KPI) que sont :

- $R^2 \geq 0.9$
- vérifier que l'erreur de répliation sur chaque sensibilité soit inférieure à 2% du SCR de marché de l'exercice précédent calculé en modèle interne
- $BEL\ ratio \geq 0.6$
- vérifier la normalité des résidus.

La première piste envisagée a été de réutiliser la fonction permettant de calculer les nominaux $C_{\theta,\lambda,k}(\beta)$. Cependant, cette fonction de coût ne permettait pas de pénaliser un individu qui ne respectait pas le critère de normalité des résidus. De plus, un portefeuille de répliation

doit également vérifier la contrainte de posséder un nombre d'instruments raisonnable $p \leq 50$. Cette contrainte n'est de nouveau pas pénalisante avec cette fonction de coût.

Il a donc été décidé de définir une fonction de coût qui dépend directement des KPI et du nombre d'instruments d'un individu.

Une fonction de coût pertinente est alors une fonction qui pénalise lorsqu'un critère n'est pas respecté, et qui récompense si celui-ci est validé.

En prenant l'exemple du KPI R^2 , si le R^2 de l'individu considéré est supérieur à 0.9, il ne doit pas être pénalisé. A l'inverse, si le R^2 est inférieur à ce seuil de 0.9, un coût doit être attribué à cet individu. Plus le R^2 s'éloigne de 0.9 à la baisse, plus ce coût doit être élevé. Afin de traduire cette exigence sur la fonction de coût, la fonction partie positive $()_+$ sera exploitée. Concernant le R^2 , un individu devra donc minimiser la valeur $(1 - \frac{R^2}{0.9})_+$.

Une démarche similaire est adoptée à propos du BEL ratio, du nombre d'instruments et des sensibilités financières. Afin de satisfaire le BEL ratio, un individu devra minimiser la valeur $(1 - \frac{BEL\ ratio}{0.6})_+$. Pour le nombre d'instruments financiers, une valeur de référence définie à 35 instruments a été décidée : un individu devra minimiser la valeur $(\frac{p}{35} - 1)_+$ pour avoir un nombre d'instrument satisfaisant. Enfin, pour chaque sensibilité financière j , un individu devra minimiser la valeur : $(\frac{|Y_j^S - (X^S\beta)_j|}{0.02 \times SCR_{marche}^{an-1}} - 1)_+$.

Le coût associé à la normalité des résidus est appréhendé par la p -value du test de Kolmogorov-Smirnov au niveau $\alpha = 0.05$ présenté en Annexe E. Un individu devra minimiser la valeur $\mathbb{1}_{p_value \leq 0.05}$.

La fonction de coût finale retenue est alors la suivante :

$$C = (1 - \frac{R^2}{0.9})_+ + (1 - \frac{BEL\ ratio}{0.6})_+ + (\frac{p}{35} - 1)_+ + \mathbb{1}_{p_value \leq 0.05} + \sum_{j=1}^{ns} (\frac{|Y_j^S - (X^S\beta)_j|}{0.02 \times SCR_{marche}^{an-1}} - 1)_+. \quad (5.1)$$

C'est une fonction à **minimiser**.

Deux remarques sont à effectuer avec la fonction de coût ainsi définie.

La première est qu'elle permet de considérer un individu qui n'appartient pas à l'espace des solutions possibles : on permet à un individu de ne pas respecter l'ensemble des KPI. Cette tolérance réside dans la complexité à obtenir un portefeuille qui satisfierait toutes les contraintes. En effet, il n'est pas certain d'avoir un tel individu dans un temps raisonnable (en particulier à cause du critère du nombre d'instruments). C'est notamment la raison pour laquelle le critère *Satisfaire l'ensemble des KPI* n'a pas été défini en tant que critère d'arrêt de l'algorithme. Bien que l'algorithme soit capable de fournir un individu répondant à l'ensemble des critères pour plusieurs fonds, il n'est pas nécessaire de se restreindre à de tels individus tant que l'utilisateur de l'algorithme est un actuair. Ce dernier aura en effet les connaissances et l'expérience nécessaires afin de valider ces contraintes à condition que celles-ci ne soient pas trop éloignées des seuils définis.

La seconde remarque concerne le poids accordé aux sensibilités financières dans la fonction de coût. Telle que définie actuellement, en début d'algorithme, la fonction de coût détermine un ordre de recherche. Le R^2 , le BEL_ratio , le nombre d'instruments et la p_value du test

de Kolmogorov-Smirnov sont négligeables devant le coût associé au non-respect du critère des sensibilités financières. Ceci est délibéré car en pratique, il est difficile de satisfaire le respect du critère de sensibilités. Il faut d'abord travailler sur une combinaison d'instruments permettant de capter la majorité de ces sensibilités avant de chercher à obtenir les autres KPI. Cet ordre de recherche est bien traduit dans la fonction (5.1).

5.4 L'influence des paramètres

L'algorithme génétique adapté au cas des portefeuilles répliquants possède plusieurs paramètres. Ces derniers peuvent difficilement être étudiés par une approche de type *Grid Search* en pratique car le temps de calcul associé à cette recherche serait trop longue. L'approche challengeant la méthode à *dire d'expert* perdrait alors de son intérêt car l'algorithme génétique vise à réduire le temps de recherche d'un portefeuille répliquant candidat. De plus, certains paramètres tels que les bornes du paramètre θ et le nombre de zéro-coupons initial doivent être choisis par l'expertise de l'actuaire en fonction des attentes du fonds étudié. Par conséquent, une étude de certains paramètres de l'algorithme génétique sera présentée dans cette section. Le comportement de l'algorithme et l'influence de ces paramètres sur les performances du portefeuille A seront observés. Les résultats présentés portent sur une exécution d'algorithme. L'algorithme génétique étant basé sur de l'aléa, les résultats peuvent légèrement varier d'une exécution à l'autre.

5.4.1 Les paramètres de croisement et de mutation

Les paramètres de croisement pour les gènes instruments financiers g_1 et sensibilités financières g_4 correspondent à l'hérédité du gène du $parent_1$ ou du $parent_2$. Dans le cas où $proba_C < 0.5$, alors le $parent_1$ a plus de chances de transmettre ses gènes à l'enfant. Si à l'inverse $proba_C > 0.5$, c'est le $parent_2$ qui a plus de chances de transmettre ses gènes à l'enfant. Dans la mesure où pour chaque individu d'une génération, son attribution en tant que $parent_1$ ou $parent_2$ est aléatoire, $proba_C^{g_1} = proba_C^{g_4} = 0.5$ a été retenu pour les gènes g_1 et g_2 dans ce chapitre.

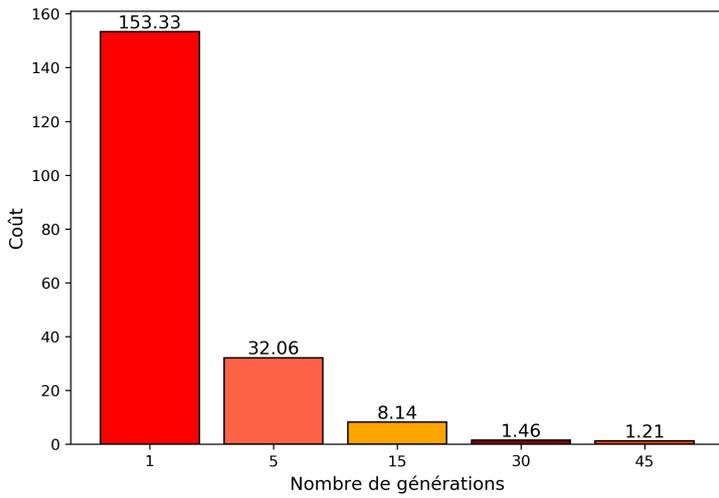
Pour la probabilité de mutation, elle diffère selon les gènes. En effet, l'objectif de mutation de chacun des gènes est différent. Pour le gène des instruments financiers g_1 , un nombre de mutations a été limité à 3 mutations maximum par individu à chaque itération. Ainsi, avoir une probabilité $proba_M^{g_1}$ conséquente pour ce gène n'impactera pas énormément la convergence. Cependant, concernant les gènes θ et λ , cette problématique de convergence intervient. Il faut alors avoir une probabilité suffisamment faible pour éviter une divergence et une probabilité suffisamment élevée pour apporter de la diversité.

Pour le gène du vecteur de poids k , il est souhaitable de conserver une majorité de numérateur égaux à 1 pour avoir la plupart des sensibilités avec le même poids. Par conséquent, tel que défini dans le processus de mutation, avoir une probabilité $proba_M^{g_4}$ élevé n'est pas dérangeant.

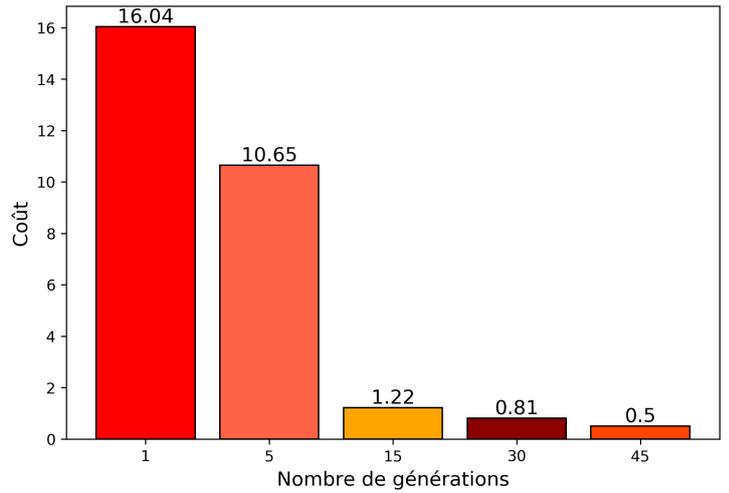
Dans ce chapitre, il sera retenu $proba_M^{g_1} = proba_M^{g_4} = 0.5$ et $proba_M^{g_2, g_3} = 0.2$.

5.4.2 Le nombre de générations

En considérant le nombre de générations défini par l'utilisateur comme seul critère d'arrêt, observons l'influence de ce nombre sur la fonction de coût et les KPI pour $n_G \in \{1, 5, 15, 30, 45\}$. Les autres paramètres de l'algorithme sont fixés.



(a) Coût moyen des individus de la génération finale



(b) Coût du meilleur individu de la génération finale

FIGURE 5.6 – Coût moyen et meilleur coût de la génération finale

Comme attendu, la figure 5.6 montre que les coûts moyens des individus de la dernière génération et du meilleur individu diminuent rapidement quand le nombre de générations augmente. Afin de constater que la fonction de coût commence par capter les sensibilités financières, observons le nombre de sensibilités non validées et les autres KPI du meilleur individu en fonction du nombre de générations.

Nombre de générations	Sensibilités non validées	R^2	BEL ratio	Normalité des résidus	Nombre d'instruments	Temps (mn)
1	15	85.02%	40.28%	Non	60	01
5	9	85.96%	51.45%	Non	65	09
15	2	81.55%	44.94%	Oui	64	27
30	0	90.73%	55.85%	Oui	61	55
45	0	88.64%	64.09%	Oui	52	84

TABLE 5.6 – KPI du meilleur individu des générations 1, 5, 15, 30 et 45

Le tableau 5.6 permet bien de constater que la fonction de coût commence par essayer de capturer un grand nombre de sensibilités financières. Avec les meilleurs individus des générations 1, 5, 15, 30 et 45, on remarque que le nombre de sensibilités financières non validées passe rapidement de 15 à 2. C'est lorsque ce nombre est faible que les autres KPI commencent à être vérifiés.

De plus, il est possible de voir que les résultats associés aux générations 30 et 45 ne présentent pas de différence conséquente en termes de KPI. La différence se fait sur le nombre de génération qui a diminué pour 45 générations. Cependant, le temps de calcul est plus important avec 45 générations. Par conséquent, un actuair e peut décider de gagner en temps de calcul en réduisant son exigence sur la validation de l'ensemble des KPI tant que ceux-ci sont proches des seuils définis et en autorisant un nombre d'instruments plus élevé. Sa connaissance peut en effet lui permettre de modifier certains instruments afin de valider le ou les KPI restant(s) et de réduire le nombre d'instruments. Un arbitrage temps de calcul/capacité à valider les KPI

restants et réduction du nombre d'instruments peut s'opérer.

Enfin, ce tableau permet de noter que la complexité associée à ce paramètre est en $O(n_G)$ et est donc linéaire selon le nombre de générations. Ainsi, l'utilisateur de l'algorithme peut directement agir sur le paramètre n_G tout en ayant une estimation du temps de calcul qu'il souhaite accorder à l'algorithme afin de trouver un ensemble de solutions potentielles.

5.4.3 Le nombre d'instruments initial

Soit le nombre de générations n_G fixé à 30 et les autres paramètres de l'algorithme définis comme lors de la sous-section précédente, observons l'influence du nombre p_0 d'instruments initial.

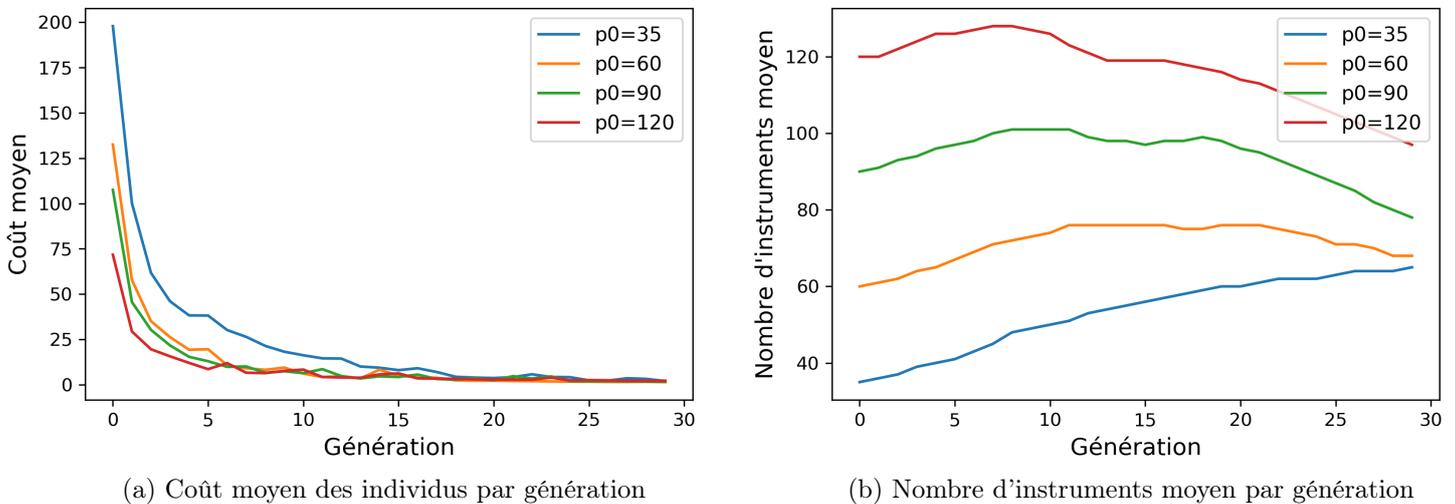


FIGURE 5.7 – Coût moyen et nombre d'instruments moyen par génération fonction du nombre initial d'instruments p_0

La figure 5.7 présente le coût moyen des individus par génération (5.7a) et le nombre d'instruments moyen par génération (5.7b) en fonction de p_0 .

On constate que le coût moyen par génération est plus grand sur les premières générations lorsque p_0 est faible. Cela s'explique par le fait qu'un grand nombre d'instruments capte plus facilement les sensibilités financières. Ces dernières ayant plus de poids en début d'algorithme, la fonction de coût sera plus faible lorsque le nombre d'instruments est important. C'est ce que l'on peut observer sur la figure 5.7b : afin de capter les sensibilités financières, la courbe bleue ($p_0 = 35$) ne cesse d'augmenter son nombre d'instruments moyen au fil des générations. A l'inverse, un changement de monotonie est observable sur les autres courbes. Cela indique que les individus ont réussi à obtenir les instruments nécessaires à remplir les conditions de l'optimisation. Les autres instruments non utiles ont pu commencer à être retirés, ce qui réduit le nombre d'instruments moyen. Néanmoins, on constate avec la courbe rouge ($p_0 = 120$) que lorsque le nombre d'instruments initial est trop important, le nombre d'instruments moyen met plus de temps à diminuer.

Avec un nombre de générations commun de $n_G = 30$, on remarque que le nombre d'instruments est encore loin de la convergence attendue vers 35 instruments. Cependant, atteindre ce point

de convergence pourrait être très long en pratique.

Le tableau 5.7 expose les KPI du meilleur individu pour chaque exécution associée aux différents p_0 . On constate bien que lorsque p_0 est trop important, la réduction d'instruments met plus de temps. En effet, avec $p_0 = 120$ (courbe rouge dans les graphiques 5.7) le meilleur individu possède 83 instruments contre moins de 70 pour les autres p_0 .

Nombre d'instruments initial	Sensibilités non validées	R^2	BEL ratio	Normalité des résidus	Nombre d'instruments	Temps (mn)
35	1	87.97%	39.99%	Oui	61	53
60	0	90.39%	62.04%	Oui	61	56
90	0	89.54%	56.56%	Oui	68	57
120	0	90.86%	56.45%	Oui	83	60

TABLE 5.7 – KPI du meilleur individu pour des nombres d'instruments initiaux de 35, 60, 90 et 120

Le temps de calcul augmente peu avec le nombre d'instruments initial p_0 . Cela est dû à la représentation des instruments dans l'algorithme génétique. Comme présentés en début de chapitre, les instruments financiers sont représentés par un vecteur de taille p_{max} composé de 0 et de 1. Dans ce vecteur, les 1 sont au nombre de p , le nombre d'instruments d'un individu. Ainsi, p_0 n'a aucune influence sur la taille du vecteur mais sur le nombre de 1 pour chaque individu de la génération initiale. Cela n'influence donc pas le temps de calcul. La légère différence de temps s'explique par le calcul des nominaux car le vecteur β dans la fonction de coût dépend bien du nombre d'instruments.

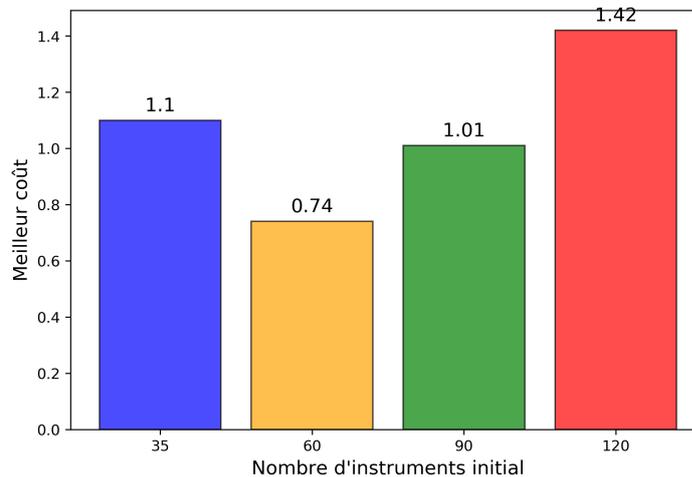


FIGURE 5.8 – Coût du meilleur individu en fonction du nombre initial d'instruments p_0

Avec ces différentes exécutions pour le *Portefeuille A*, on constate avec le tableau 5.7 et la figure 5.8 que le paramètre p_0 apportant les meilleures performances est $p_0 = 60$. Le meilleur individu satisfait l'ensemble des KPI et a un coût $C = 0.74$ qui correspond à la pénalisation de son nombre d'instruments $p = 61$.

5.4.4 Le nombre d'individus par génération

Soient le nombre de générations n_G fixé à 30 et le nombre d'instruments initial p_0 fixé à 60, observons l'influence du nombre n_I d'individus par génération. Les autres paramètres de l'algorithme sont définis comme précédemment.

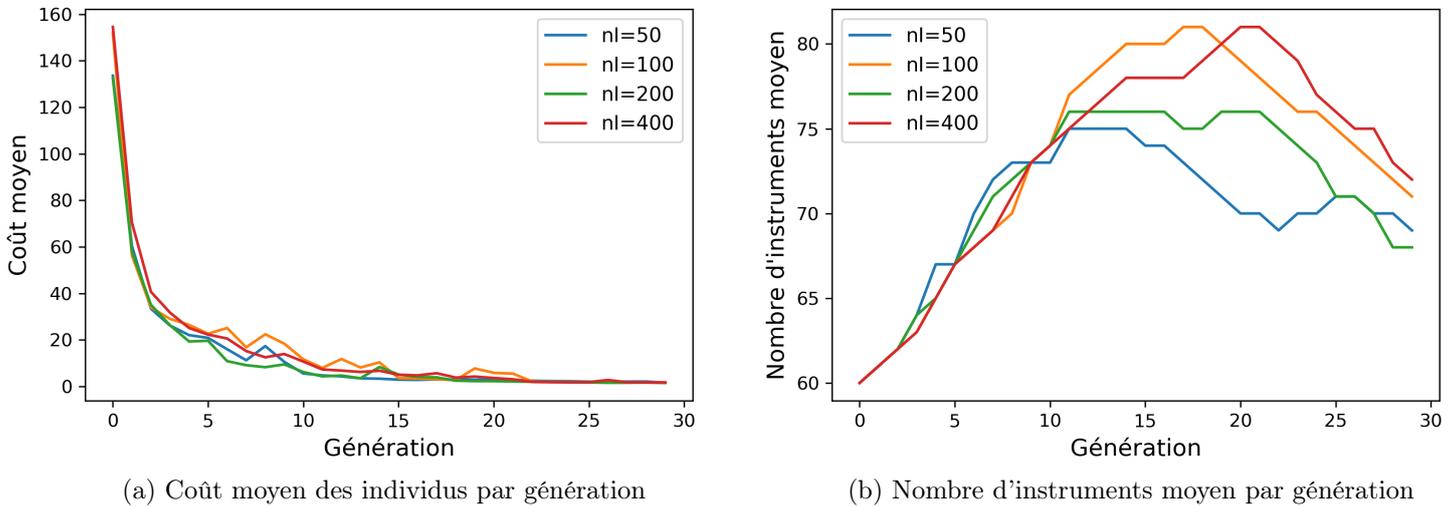


FIGURE 5.9 – Coût moyen et nombre d'instruments moyen par génération fonction du nombre d'individus par génération

La figure 5.9 présente le coût moyen des individus par génération (5.9a) et le nombre d'instruments moyen par génération (5.9b) en fonction de n_I .

La vitesse de convergence du coût moyen des individus par génération est similaire pour les 4 n_I étudiés. Le comportement du nombre d'instruments moyen est également semblable pour les 4 n_I étudiés : le nombre moyen d'instruments commence par augmenter puis décroît sur la figure 5.9(b). En revanche, le maximum diffère en fonction des courbes. On constate que les courbes rouge et orange (respectivement associées à $n_I = 400$ et $n_I = 100$) ont un nombre moyen d'instruments maximum de 81. Les courbes verte et bleue atteignent leur maximum en respectivement 76 et 75 instruments.

Nombre d'individus par génération	Sensibilités non validées	R^2	BEL ratio	Normalité des résidus	Nombre d'instruments	Temps (mn)
50	3	81.29%	44.03	Oui	61	13
100	0	86.23%	51.50%	Oui	70	28
200	0	90.39%	62.04%	Oui	61	56
400	1	88.80%	63.95%	Oui	62	112

TABLE 5.8 – KPI du meilleur individu pour des nombres d'individus par génération de 50, 100, 200 et 400

Le tableau 5.8 affiche les KPI du meilleur individu pour chaque exécution associée aux différents n_I . On remarque qu'avec un nombre d'individus trop faible par génération ($n_I=50$), presque tous les KPI du meilleur individu ne sont pas vérifiés. Cela provient d'un manque de diversité. En effet, en réduisant le nombre d'individus, on réduit le nombre de solutions

potentielles et donc le mélange de gènes est moins varié qu’avec un nombre d’individus plus important. Cela réduit ainsi l’exploration de l’environnement \mathcal{E} .

Ce tableau permet également de constater qu’en doublant le nombre d’individus par génération, on double le temps de calcul. La complexité algorithmique associée au nombre d’individus par génération est donc linéaire $O(n_I)$. Cela s’explique par le fait qu’en doublant le nombre d’individus, le nombre d’opérations nécessaire à l’attribution des gènes est doublé, tout comme le nombre de couples engendrant de nouveaux individus.

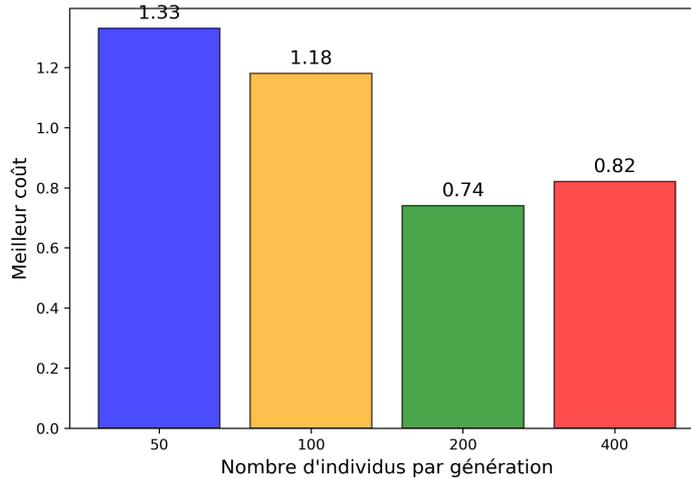


FIGURE 5.10 – Coût du meilleur individu en fonction du nombre d’individus par génération

Avec ces différentes exécutions pour le *Portefeuille A*, on observe avec le tableau 5.8 et la figure 5.10 que le paramètre n_I apportant les meilleures performances est $n_I = 200$. Le meilleur individu satisfait l’ensemble des KPI et a un coût $C = 0.74$ qui correspond à la pénalisation de son nombre d’instruments $p = 61$.

Comme annoncé en début de section, l’algorithme génétique repose sur de l’aléatoire. Pour de mêmes paramètres, les résultats peuvent légèrement varier d’une exécution d’algorithme à l’autre. Pour l’étude de l’influence du nombre d’instruments initial p_0 et du nombre d’individus par génération n_I , les performances exposées pour $p_0 = 60$ et $n_I = 200$ proviennent de la même exécution. Pour l’étude du nombre de générations n_G , bien que les paramètres soient les mêmes, il s’agit d’une autre exécution d’algorithme. Le tableau 5.9 qui rassemble les résultats de ces deux exécutions de même paramètres permet de souligner ce fait. On remarque bien que les KPI diffèrent. D’une exécution à l’autre les résultats varient légèrement du fait du caractère aléatoire sur lequel repose l’algorithme génétique.

Exécution	Sensibilités non validées	R^2	BEL ratio	Normalité des résidus	Nombre d’instruments	Temps (mn)
N°1	0	90.73%	55.85%	Oui	61	55
N°2	0	90.39%	62.04%	Oui	61	56

TABLE 5.9 – KPI de deux exécutions d’un algorithme de mêmes paramètres

5.5 Les résultats sur le *portefeuille A*

Le *Portefeuille A* est un portefeuille de retraite présentant différentes optionnalités. Cette section présente les résultats d'une exécution d'algorithme génétique qui vise à répliquer ce portefeuille. Choisissons d'observer les résultats obtenus par l'algorithme génétique pour un temps d'exécution maximal d'environ une heure. Pour cela, un nombre de générations $n_G = 30$ combiné à un nombre d'individus $n_I = 200$ sont retenus. L'ensemble des paramètres choisis pour exécuter l'algorithme sont réunis dans le tableau 5.10.

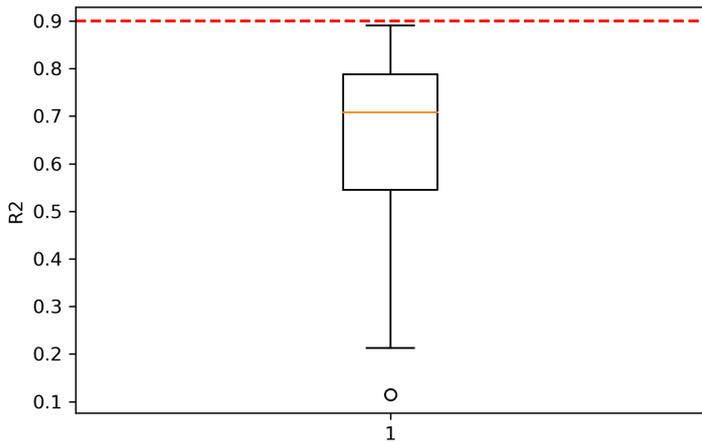
Paramètres	$[a_\theta, b_\theta]$	$[a_\lambda, b_\lambda]$	n_I	$ZC_{initial}$	p_0	n_G	n_{best}	n_0
Valeurs	$[0.5, 0.9]$	$[5.10^{-7}, 6.10^{-6}]$	200	5	60	30	5	5

Paramètres	n_{blocs}	$proba_C^{g1}$	$proba_C^{g4}$	$proba_M^{g1}$	$proba_M^{g2:g3}$	$proba_M^{g4}$
Valeurs	5	0.5	0.5	0.5	0.2	0.5

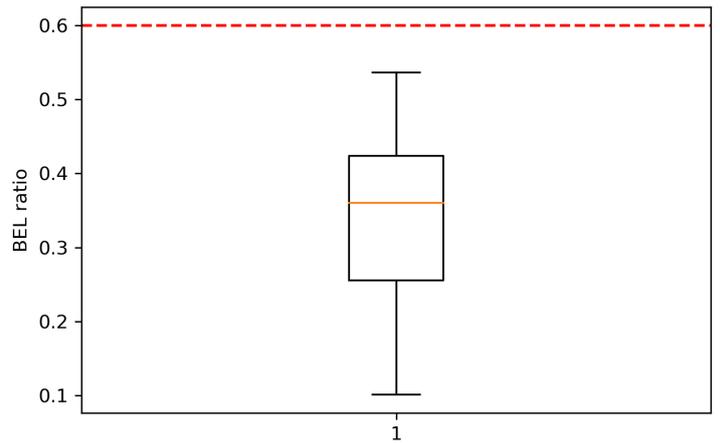
TABLE 5.10 – Paramètres utilisés pour l'exécution de l'algorithme génétique appliqué au *Portefeuille A*

Afin d'observer que l'algorithme optimise bien au fil des générations, commençons par regarder les résultats d'une exécution de génération initiale uniquement. Ces résultats serviront de comparaison avec ceux de la génération finale pour montrer que les individus évoluent. Pour rappel, bien que les individus de la première génération doivent respecter certaines contraintes associées aux paramètres de l'algorithme (posséder au moins 5 zéro-coupons, posséder le zéro-coupon inflation 10 ans, posséder $n_I = 200$ individus, avoir $p_0 = 60$ instruments par individu par exemple), ces individus reposent fortement sur de l'aléatoire. La figure 5.11 présente respectivement les boxplots du R^2 (5.11a), du *BEL ratio* (5.11b), le nombre de sensibilités non validées (5.11c) et la normalités des résidus (5.11d) des individus de la génération initiale G_0 .

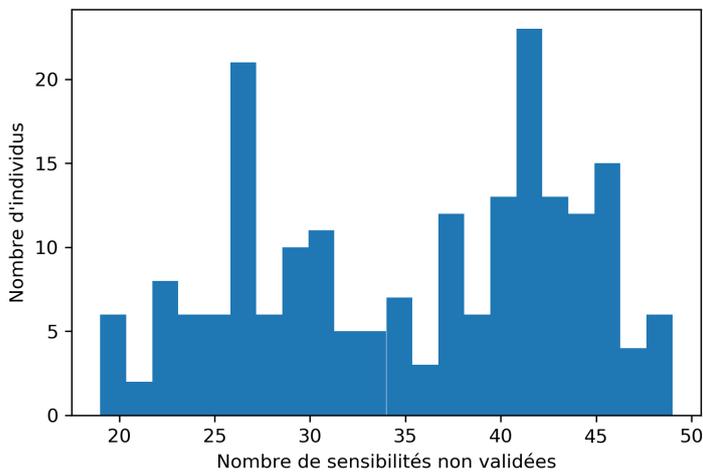
On constate que les critères de R^2 , de *BEL ratio* et de sensibilités ne sont respectés par aucun individu. Seuls 6 individus valident le critère de normalité des résidus (5.11d) mais ces individus ne valident pas au moins 19 sensibilités financières (5.11c). Enfin, le *BEL ratio* maximal au sein de cette première génération est de 53.63% et la médiane est à 36.72% (5.11b). Le R^2 est comparativement plus proche du seuil de 90% avec un maximum de 89.04% et une médiane de 70.75% (5.11b).



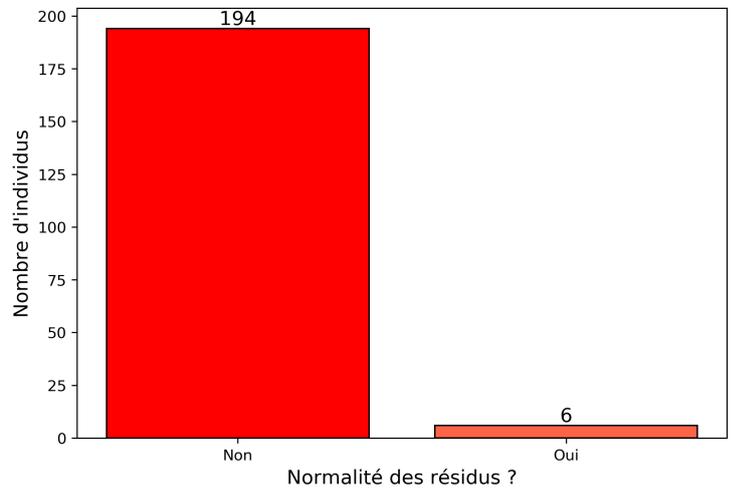
(a) Boxplot du R^2



(b) Boxplot du BEL ratio



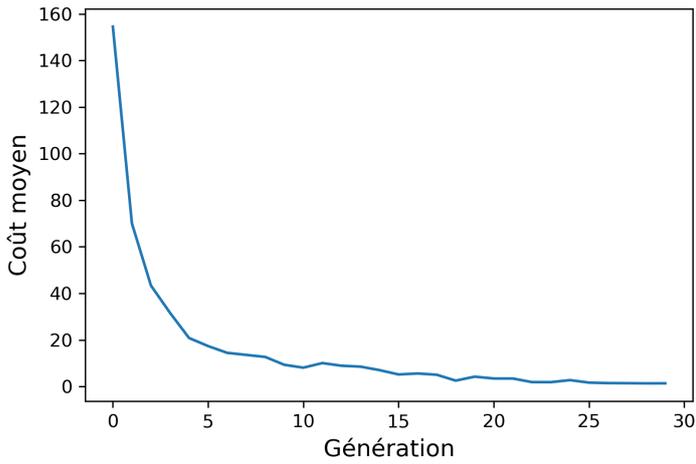
(c) Histogramme du nombre de sensibilités non validées



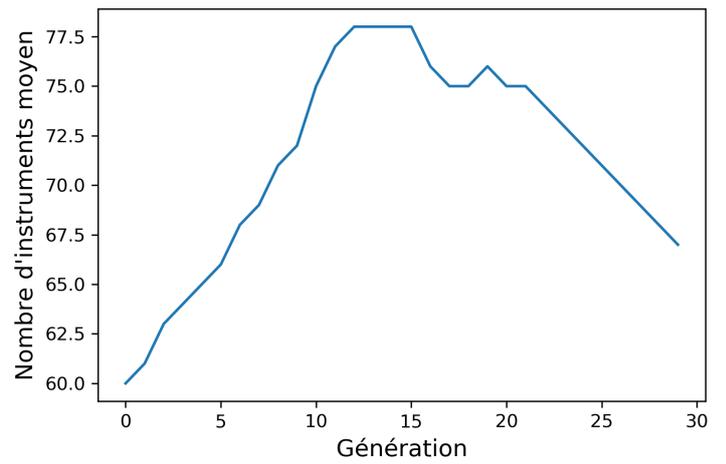
(d) Normalité des résidus

FIGURE 5.11 – R^2 , BEL ratio, sensibilités et normalité des individus de la génération initiale

Exécutons désormais l'algorithme génétique avec les paramètres du tableau 5.10. Les figures 5.12a et 5.12b montrent respectivement l'évolution du coût moyen et du nombre moyen d'instruments par génération. L'algorithme s'est arrêté après avoir engendré $n_G = 30$ générations en 58 minutes. Cela signifie qu'une même combinaison de $n_{best} = 5$ individus n'est pas apparue en tête de classement pendant $n_o = 5$ générations consécutives. De plus, il est possible de noter avec la figure 5.12b que le nombre moyen d'instruments est encore éloigné de l'objectif des 35 instruments. Le boxplot 5.13 permet de voir en détail qu'aucun individu n'a pu suffisamment réduire son nombre d'instruments en 30 générations. Ceci est une limite de l'utilisation de l'algorithme génétique qui réduit difficilement ce nombre dans un temps raisonnable.



(a) Coût moyen des individus par génération



(b) Nombre d'instruments moyen par génération

FIGURE 5.12 – Coût moyen et nombre d'instruments moyen par génération

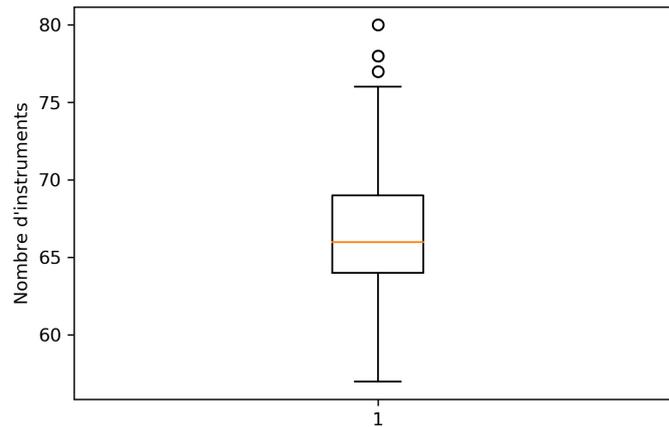
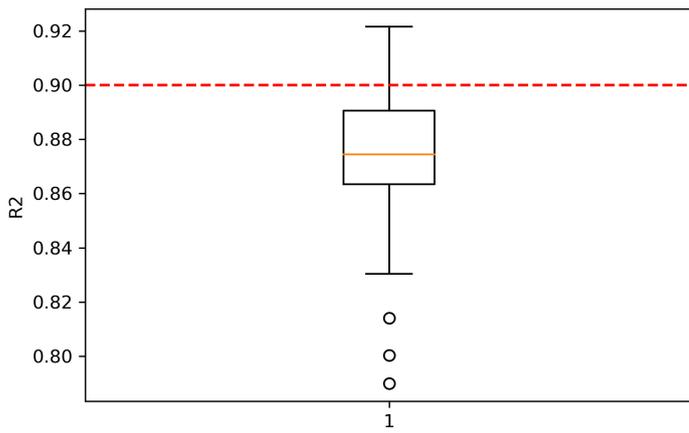


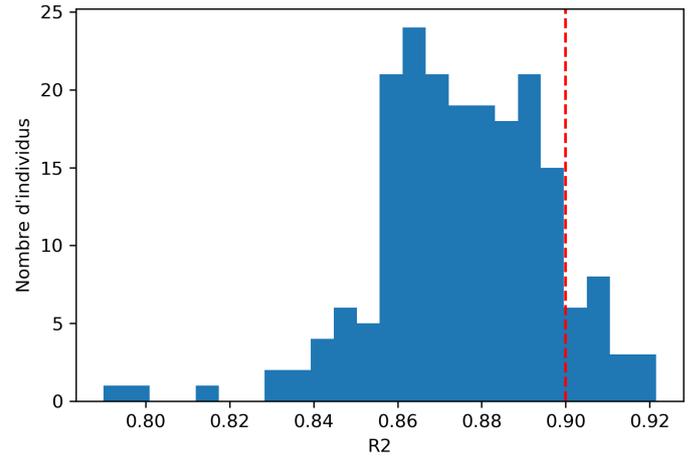
FIGURE 5.13 – Boxplot du nombre d'instruments de la génération finale

Avec les figures 5.14, 5.15 et 5.16, on constate que les individus se sont adaptés à leur environnement. En effet, les critères de R^2 , $BEL\ ratio$ et sensibilités qui n'étaient respectés par aucun individu le sont désormais par quelques-uns. Globalement, les individus possèdent des KPI proches des seuils. On remarque néanmoins que certains individus ont au moins le critère de R^2 ou de $BEL\ ratio$ qui ne soit pas à au moins 10 points de pourcentage des seuils définis. Il s'agit d'un individu pour le critère de R^2 , et de 39 individus pour le $BEL\ ratio$. Ces individus proviennent pour la plupart du processus de sélection qui retient 20% d'individus non optimaux à chaque génération (section 5.2.6). Leur présence sert à l'apport de diversité.

L'algorithme génétique a permis de passer d'une médiane de R^2 sur la population initiale de 70.75% (5.11a) à 87.44% pour la génération finale (5.14a). Le plus petit R^2 de la génération finale est égal à $\min_{G_f}^{R^2} = 79.00\%$ et est supérieur au troisième quartile de la génération initiale $Q3_{G_0}^{R^2} = 78.75\%$. Enfin, 19 individus valident le critère de R^2 supérieur à 90% (5.14b).



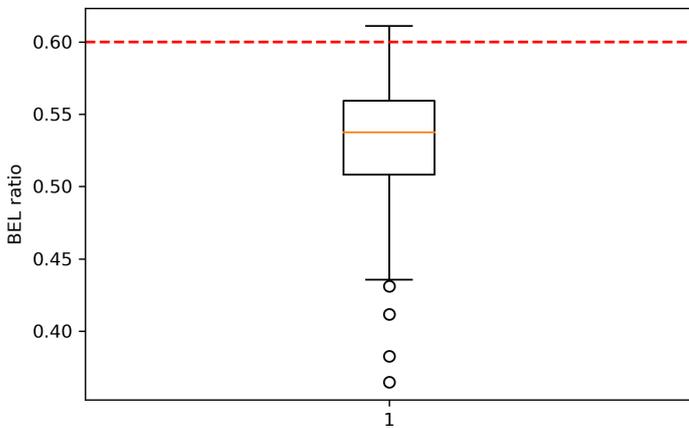
(a) Boxplot du R^2 des individus de la génération finale



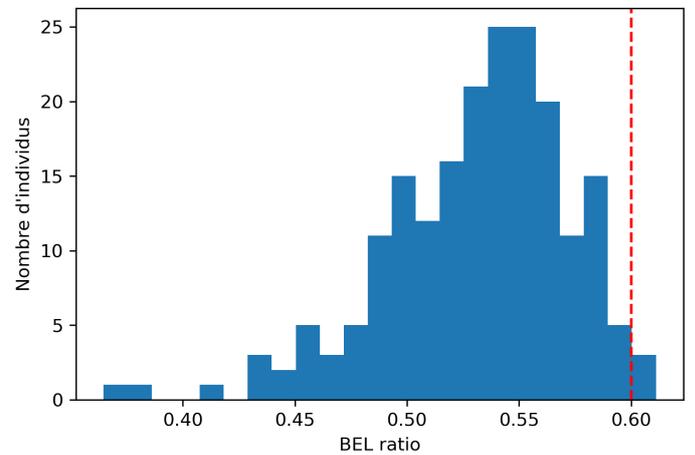
(b) Histogramme du R^2 des individus de la génération finale

FIGURE 5.14 – Boxplot et histogramme du R^2 des individus de la génération finale

L'algorithme génétique a également permis de passer d'une médiane de *BEL ratio* sur la population initiale de 36.72% (5.11b) à 53.74% pour la génération finale (5.15a). Le premier quartile de la génération finale étant égal à $Q1_{G_f}^{BEL\ ratio} = 50.81\%$, il est supérieur au troisième quartile de la génération initiale. Enfin, quatre individus de la génération finale valident le critère de *BEL ratio* supérieur à 60% (5.15b).



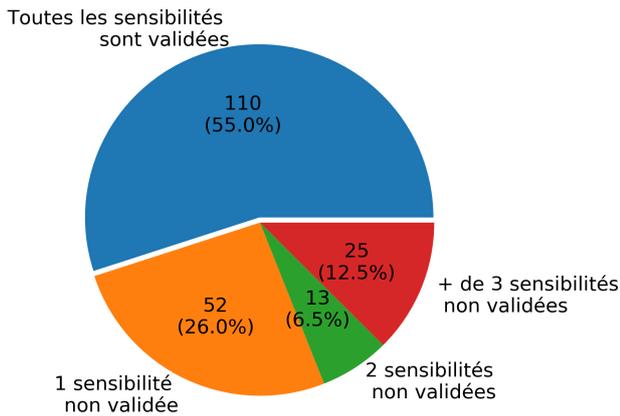
(a) Boxplot du *BEL ratio* des individus de la génération finale



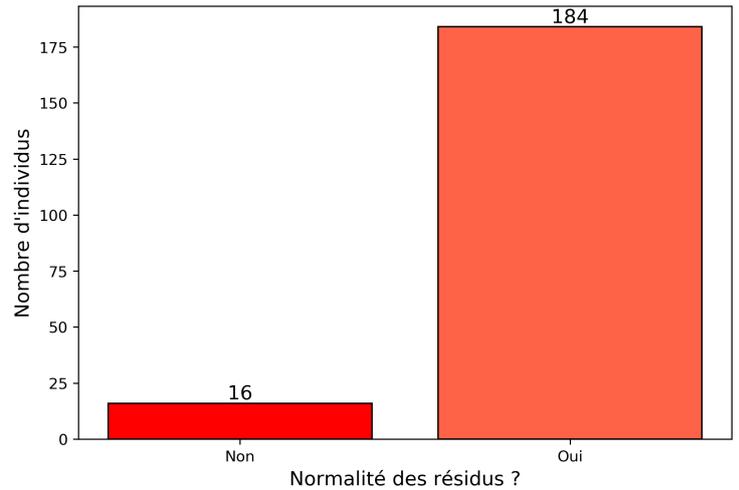
(b) Histogramme du *BEL ratio* des individus de la génération finale

FIGURE 5.15 – Boxplot et histogramme du *BEL ratio* des individus de la génération finale

Au sein de la génération finale, plus de la moitié des individus vérifient la totalité des sensibilités financières et plus de 75% des individus valident au moins 56 des 58 sensibilités financières (5.16a). Lors de la génération initiale, les 6 individus vérifiant le plus de sensibilités n'en vérifiait que 39 (5.11c). La figure 5.16b montre que la tendance concernant la normalité des résidus s'est inversée comparativement à la génération initiale : désormais 184 individus valident le critère de normalité contre seulement 6 pour la génération initiale (5.11d).



(a) Validation des sensibilités de la génération finale



(b) Normalité des résidus de la génération finale

FIGURE 5.16 – Normalité des résidus et validation des sensibilités de la génération finale

La figure 5.17 permet de voir que 3 individus valident les critères de R^2 , de $BEL\ ratio$ et de normalité des résidus. 54 individus sont à moins de 5 points de pourcentage de les satisfaire. Cela signifie que l'un ou plusieurs de ces critères sont à moins de 5 points de pourcentage du seuil défini, les autres étant validés. Le fait que ces individus ne soient pas majoritaires n'est pas dérangeant car c'est un unique portefeuille répliquant qui sera retenu au final. En pratique, à l'issue de l'algorithme génétique, 2 ou 3 individus sont retenus et étudiés afin de sélectionner l'individu permettant de fournir le portefeuille de réplcation le plus adapté.

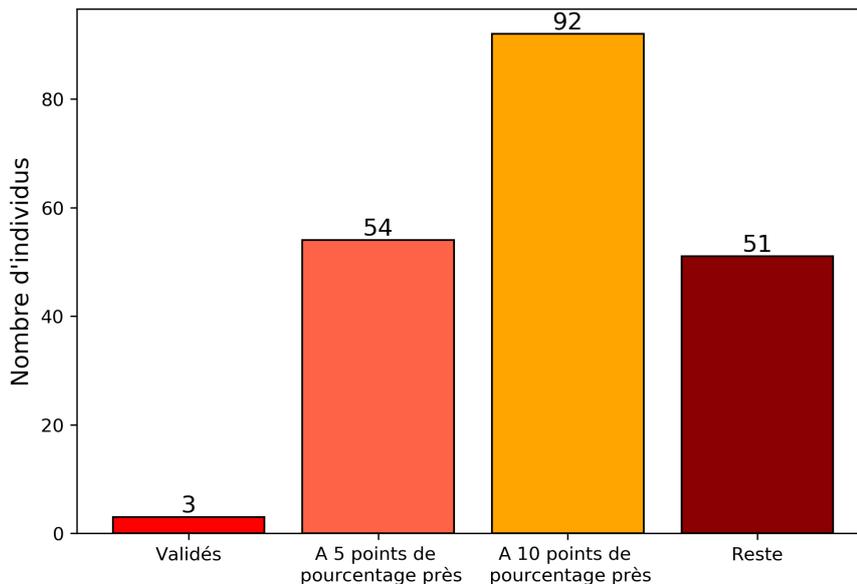


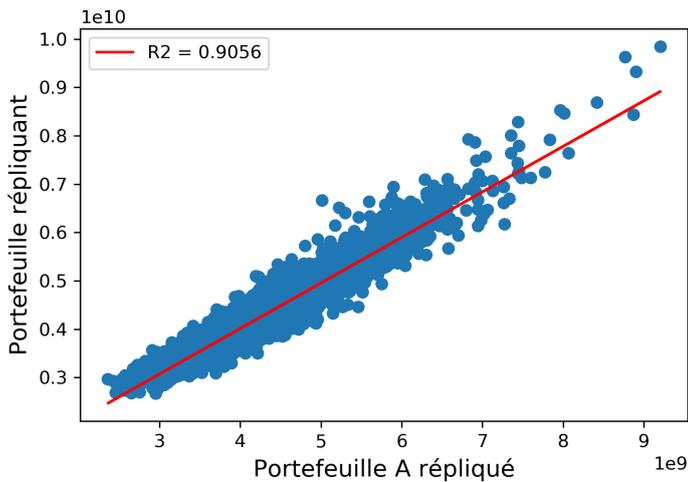
FIGURE 5.17 – Individus vérifiant les critères de R^2 , de $BEL\ ratio$ et de normalité des résidus

Regardons plus en détail les performances de 2 individus : par exemple le meilleur individu de la génération finale et le premier individu ne satisfaisant pas au moins l'un des KPI. L'identifiant

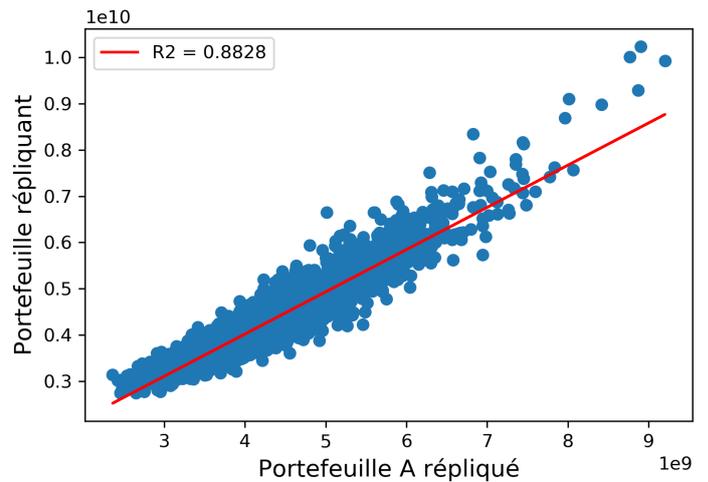
ainsi que les performances associés à chacun de ces deux individus sont exposés dans le tableau 5.11.

Classement au sein de la génération finale	Identifiant	Sensibilités non validées	R^2	BEL ratio	Normalité des résidus	Nombre d'instruments
1 ^{er}	9363	0	90.56%	61.11%	Oui	57
4 ^{eme}	10703	0	88.28%	60.02%	Oui	62

TABLE 5.11 – KPI de deux individus



(a) Individu n°9363

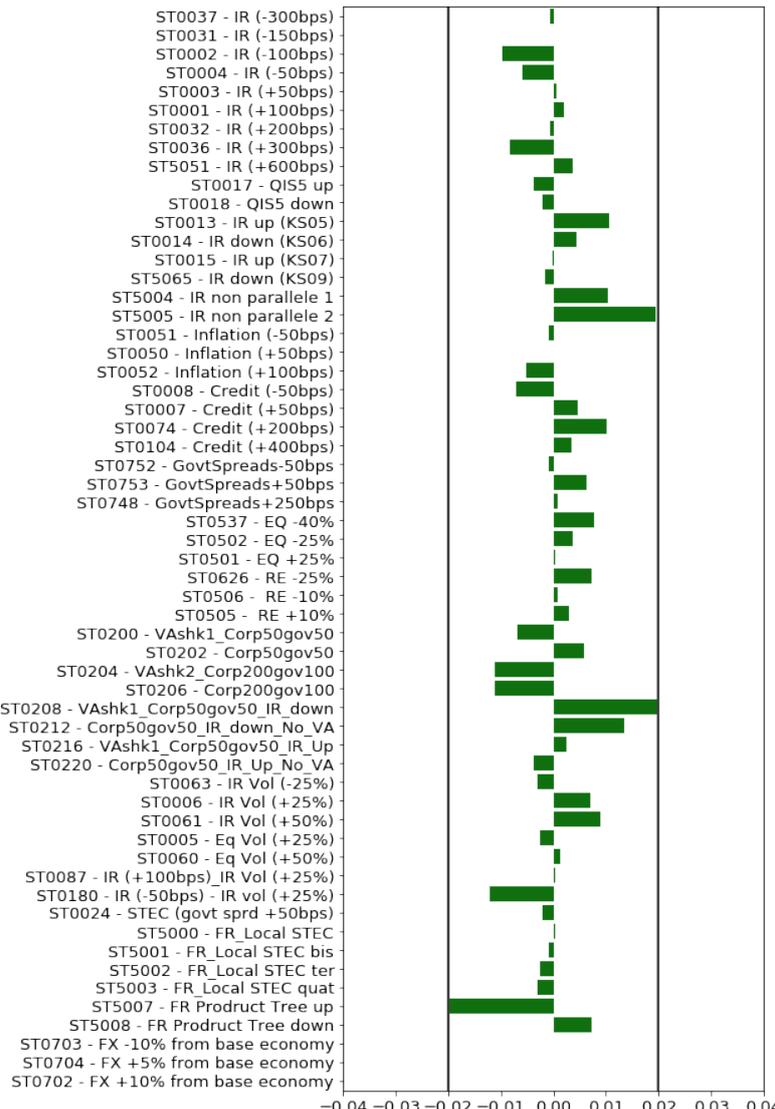


(b) Individu n°10703

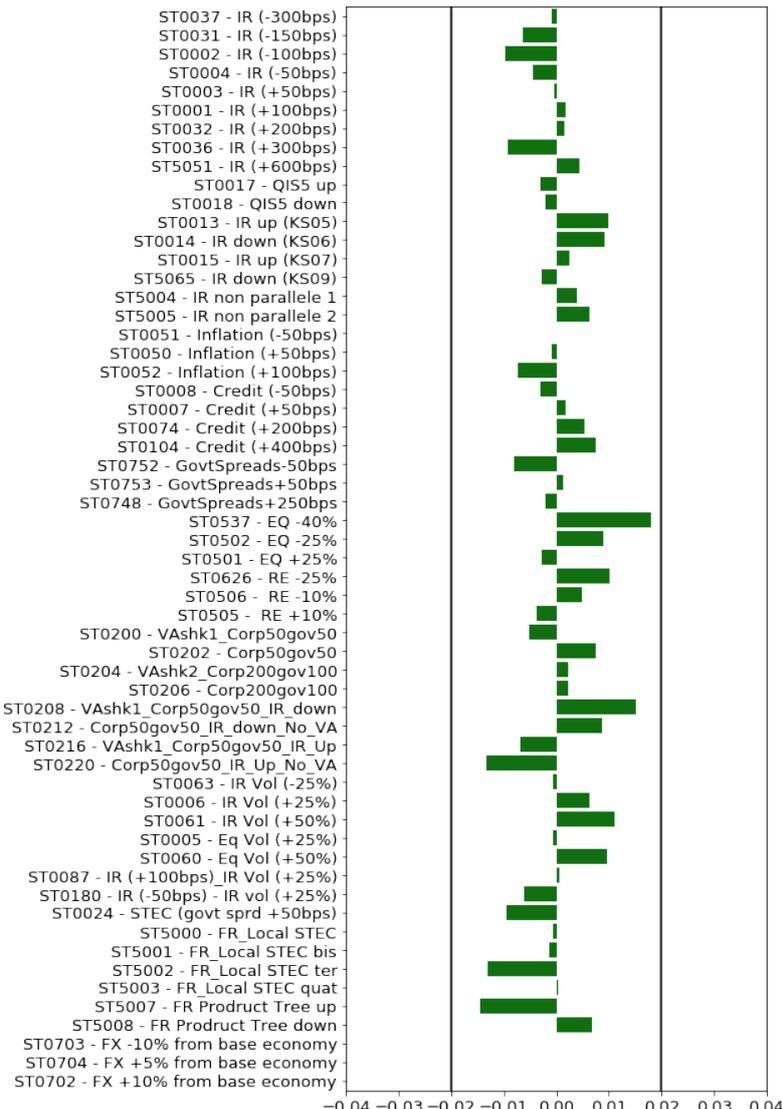
FIGURE 5.18 – Comparaison des 4000 scénarios entre le portefeuille répliquant et le portefeuille répliqué

Le meilleur individu dont l'identifiant est $ID9363$ valide tous les KPI. L'individu $ID10703$ ne valide pas le critère du R^2 mais est proche du seuil requis de 90% (tableau 5.11 et figure 5.18b). Ainsi, l'individu $ID9363$ réplique mieux le passif dans les 4000 scénarios secondaires que l'individu $ID10703$.

Les deux individus vérifient le critère de toutes leurs sensibilités. Le détail des sensibilités est donné en figure 5.19a pour le meilleur individu $ID9363$ et en figure 5.19b pour l'individu $ID10703$.



(a) Individu *ID9363*



(b) Individu *ID10703*

FIGURE 5.19 – Détail des sensibilités des individus *ID9363* et *ID10703*

Bien que ces deux individus témoignent de bonnes performances, leur nombre d'instruments financiers demeure élevé. Le détail par type d'instruments est donné dans le tableau 5.12. Le meilleur individu *ID9363* possède moins de zéro-coupons et de swaption payeur que l'autre individu.

Type d'instruments	ID9363	ID10703
Zero-coupon	5	7
Zero-coupon inflation	1	1
Swaption receveur	11	15
Swaption payeur	14	15
Indice equity (×6)	5	4
Call sur equity (×6)	12	12
Put sur equity (×6)	9	8
Total	57	62

TABLE 5.12 – Détail des types d'instruments pour les deux individus *ID9363* et *ID10703*

L'individu *ID9363* montre de bonnes performances en satisfaisant l'ensemble des KPI. Cet individu possède de plus le nombre minimum d'instruments (57) parmi les individus de la génération finale. Une ultime étape doit être réalisée afin de valider la sélection du portefeuille répliquant, il s'agit d'une étude appelée étude des topologies.

5.6 L'étude des topologies

L'étude des topologies est une étude graphique qui permet d'analyser le comportement du portefeuille candidat dans le cadre de scénarios définis manuellement. Cette étude intervient une fois qu'un portefeuille candidat est déterminé. Rappelons que le portefeuille répliquant sera utilisé par les $P = 11000$ scénarios monde réel. Il est donc indispensable que ce portefeuille répliquant atteste d'un comportement cohérent face au mouvement des différents facteurs de risque de marché.

Le comportement du portefeuille répliquant est d'abord comparé aux sensibilités originales provenant directement du modèle ALM et représentés par des points sur le graphique 5.20. C'est une différence de BEL approximée par portefeuille répliquant qui est observée : on regarde $\Delta BEL^{RP} = BEL_{choquee}^{RP} - BEL_{scenario\ central}^{RP}$ pour différents mouvements de marché. Au-delà du fait que le portefeuille répliquant doit engendrer une courbe passant par ces différents points, il est attendu une certaine continuité logique de la part de celui-ci. Cela se traduit par une courbe lisse et non une courbe sinusoïdale qui marquerait un phénomène d'*overfitting*.

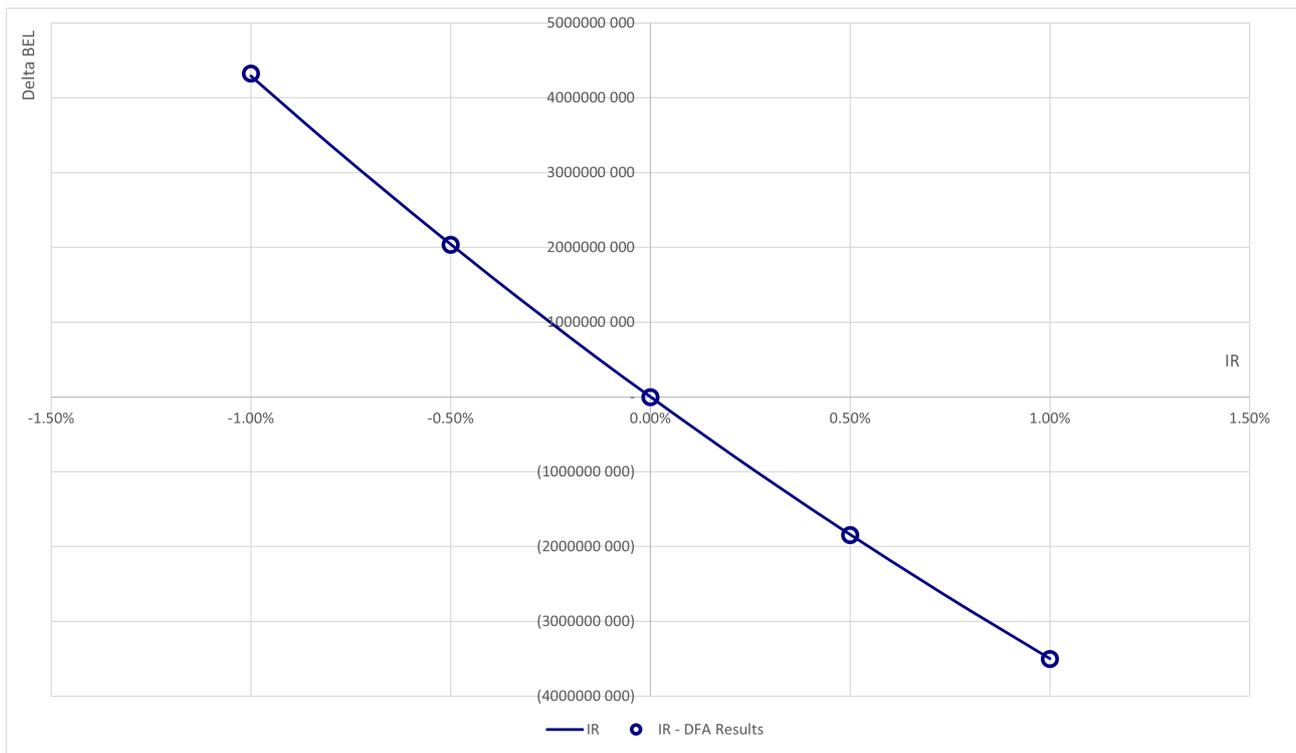


FIGURE 5.20 – Topologie présentant un comportement adéquat d’un portefeuille répliquant

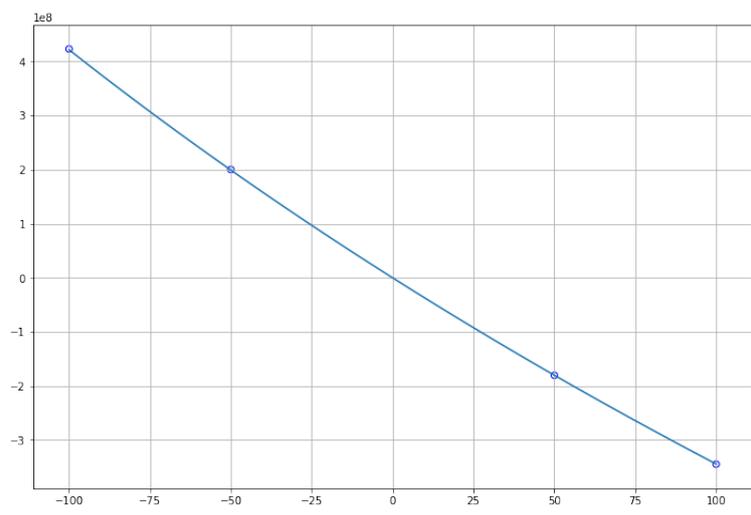
Par exemple, les topologies permettent d’observer le comportement face à l’évolution des taux. Il est en effet attendu que le BEL diminue lorsque les taux augmentent et inversement. C’est pourquoi l’allure de la courbe permet de vérifier la cohérence du portefeuille répliquant, et occasionnellement de vérifier qu’aucun instrument pertinent n’a été omis. Dans le cas où une forme anormale est observée dans la topologie, l’actuaire expérimenté peut essayer de la corriger en identifiant les instruments responsables. Si l’actuaire parvient à identifier le ou les instrument(s) responsable(s) de la forme anormale, les possibilités suivantes peuvent être envisagées :

- modifier la valeur du strike des instruments potentiellement problématiques
- modifier la maturité des instruments potentiellement problématiques
- ajout de cet instrument si la forme anormale provient de l’omission de celui-ci
- suppression d’un instrument si ce dernier est la cause de la forme anormale.

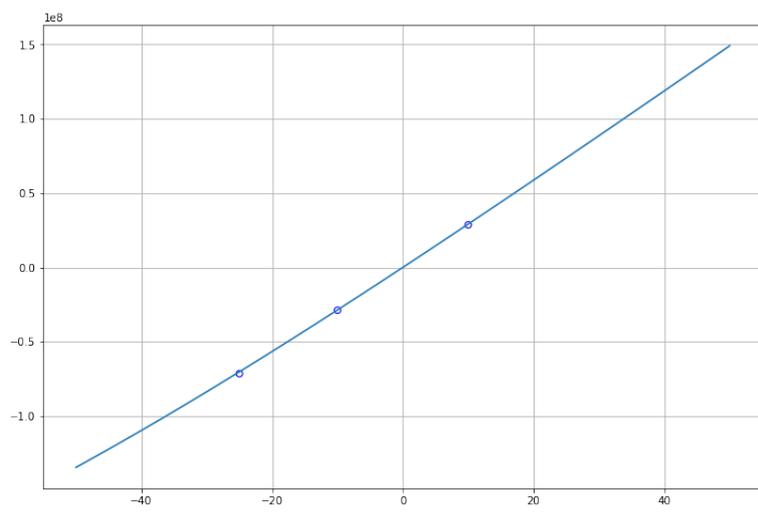
L’échange d’un instrument par un autre de même nature aux caractéristiques proches, l’ajout ou la suppression d’instrument(s) peuvent donc permettre d’améliorer la topologie.

La figure 5.21 présente les 5 topologies du meilleur individu $ID9363$. Son comportement face aux différents mouvements de marché est cohérent : les courbes passent bien par les points du modèle ALM et sont lisses. Il y a un bien décroissance du BEL^{RP} lors d’une hausse des taux (5.21a) et d’une hausse de spread (5.21c et 5.21d). Enfin, le BEL^{RP} augmente bien avec une hausse de l’immobilier (5.21b) et une hausse sur actions (5.21e).

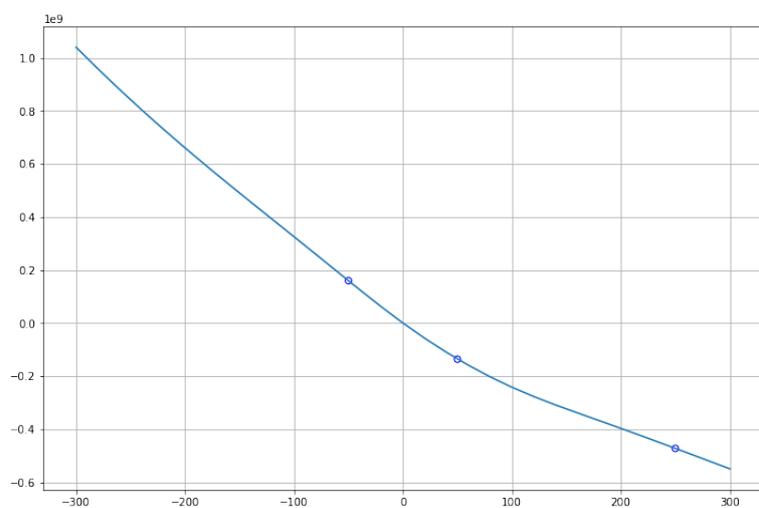
En moins d’une heure, l’algorithme génétique a permis d’obtenir l’individu $ID9363$ qui, avec ses 57 instruments, satisfait tous les KPI et vérifie les topologies.



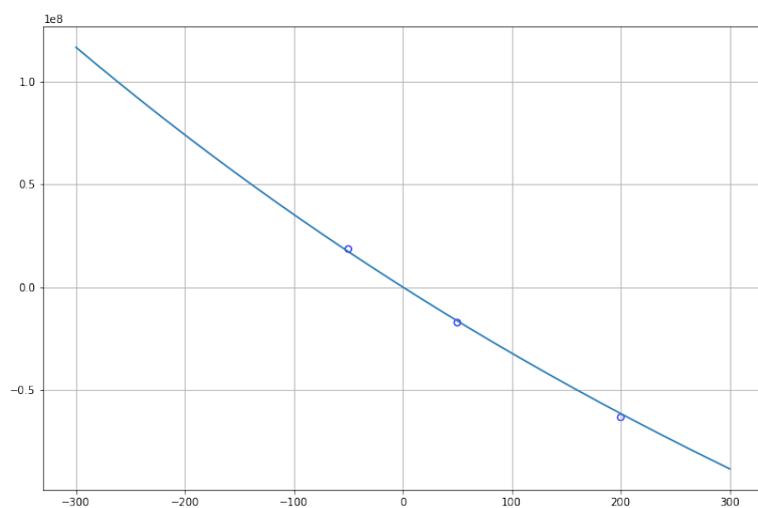
(a) Choc sur les taux d'intérêt



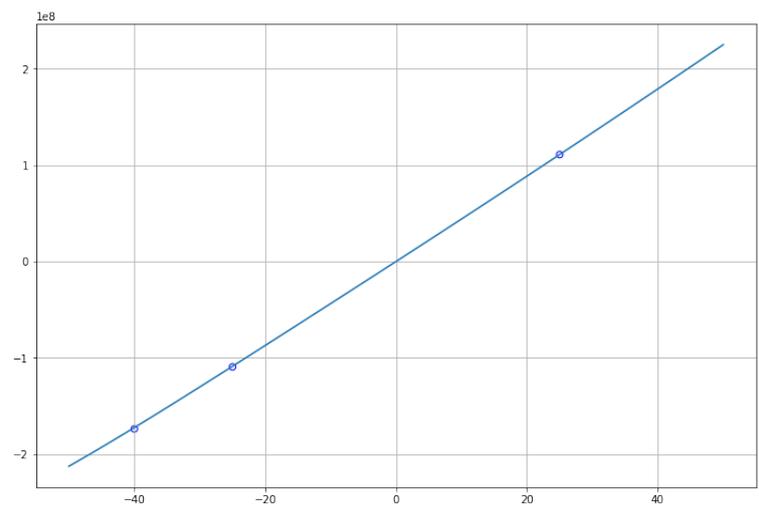
(b) Choc sur l'immobilier



(c) Choc sur le spread Govies



(d) Choc sur le spread de crédit



(e) Choc sur actions

FIGURE 5.21 – Topologies du meilleur individu *ID9363* sur les principaux facteurs de risque

5.7 La présentation de l’outil Python

L’algorithme génétique a été implémenté en langage Python sous la forme d’un outil. Les différentes fonctionnalités de l’outil sont exposées dans le diagramme de cas d’utilisation 5.22. Dans ce diagramme, on constate que deux algorithmes supplémentaires ont été introduits. Il s’agit de deux algorithmes visant à surpasser les limites rencontrées lors de l’utilisation de l’algorithme génétique. Ces deux algorithmes ainsi que les limites de l’utilisation de l’algorithme génétique sont présentés en détail dans le chapitre suivant.

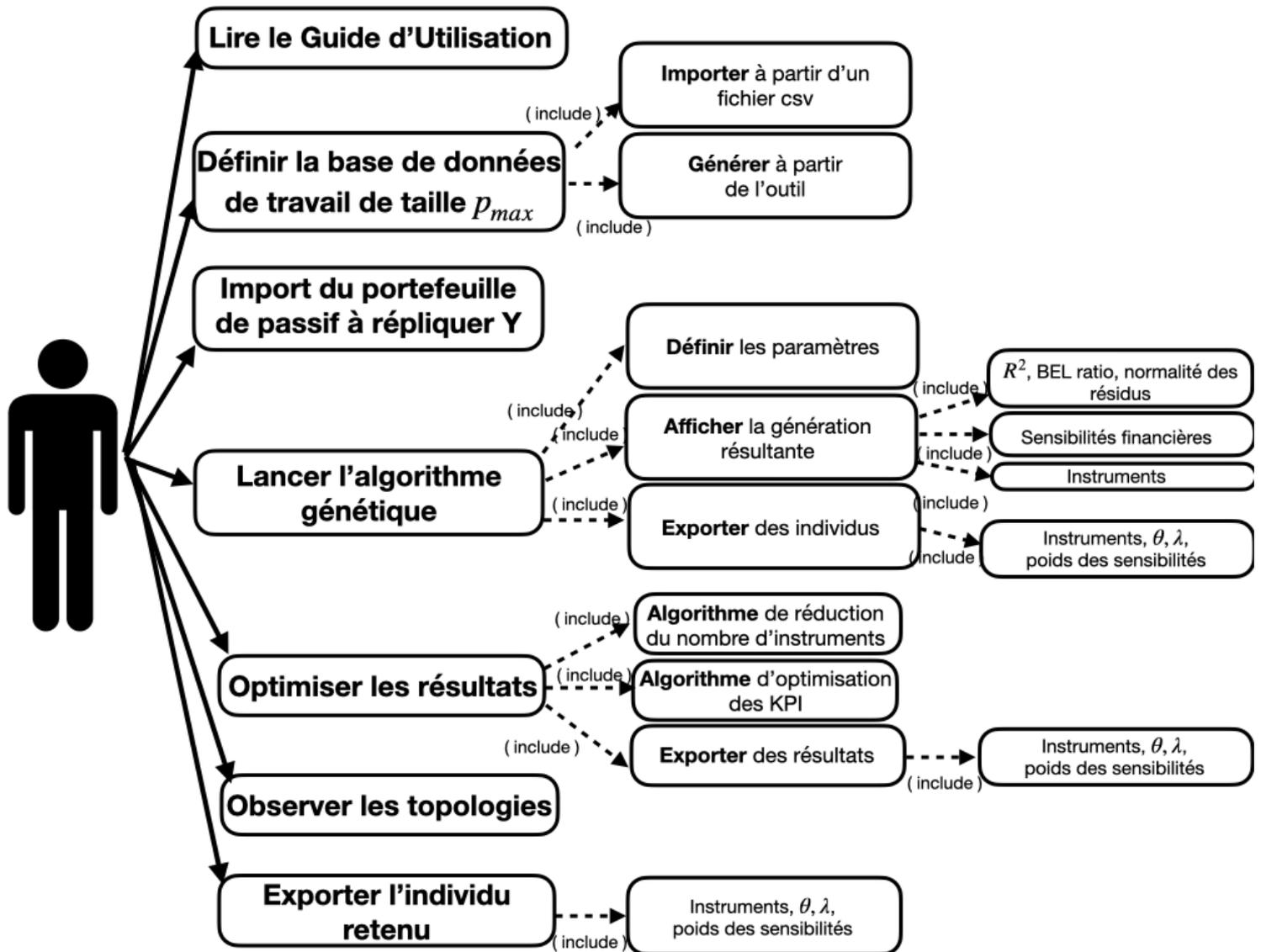


FIGURE 5.22 – Diagramme de cas d’utilisation de l’outil Python

LES PROBLÉMATIQUES RENCONTRÉES ET SOLUTIONS PRÉCONISÉES

L'utilisation de l'algorithme génétique appliqué au cas des portefeuilles de répliation possède quelques limites. Ce chapitre les énumère et propose des solutions afin d'y répondre.

6.1 Le nombre d'instruments

L'algorithme génétique avec la fonction de coût telle que définie actuellement réduit avec difficulté le nombre d'instruments d'un individu en un temps raisonnable. Bien qu'en augmentant le temps de calcul, l'algorithme génétique parvient à réduire ce nombre, cette réduction ne s'opère pas toujours en un temps relativement faible. L'objectif de l'algorithme génétique étant de proposer une solution peu coûteuse en temps, il a été décidé d'utiliser un algorithme supplémentaire afin d'effectuer ce travail de réduction.

Afin de dépasser cette limite, un algorithme imitant le travail manuel des actuaires a été implémenté. Il s'agit d'un algorithme qui réduit le nombre d'instruments financiers d'un portefeuille candidat en augmentant les KPI qui ne seraient pas validés jusqu'aux seuils définis.

Le fonctionnement de cet algorithme est le suivant :

1. Définir des seuils pour le R^2 , le *BEL ratio*, la p_{value} et les sensibilités financières. Par défaut, ces seuils sont respectivement de 0.9, 0.6, 0.05 et 0.02.
2. Lister les KPI qui ne respectent pas ces seuils
3. Si le nombre de KPI est supérieur à 1 :
 - Laisser le choix à l'utilisateur de l'ordre des KPI à optimiser
4. Pour chaque KPI non vérifié :
 - (a) Sélectionner le premier instrument du portefeuille de répliation candidat
 - (b) Tant que le KPI non vérifié n'atteint pas le seuil défini et qu'il reste des instruments à évaluer :
 - i. Retirer l'instrument SI son retrait améliore le KPI considéré SANS dégrader les KPI validant déjà les seuils définis.
 - ii. Sélectionner l'instrument suivant.

Cet algorithme permet de réduire le nombre d'instruments tout en optimisant les KPI qui n'atteindraient pas les seuils requis par le contrôle interne.

Si tous les KPI atteignent ces seuils, l'algorithme peut être utilisé en définissant des seuils légèrement supérieurs afin de procéder à la réduction. Dans un tel cas de figure, il est conseillé de ne pas choisir des seuils trop éloignés des seuils par défaut.

Lorsque les KPI n'atteignent pas les seuils par défaut, la troisième étape qui permet à l'utilisateur de choisir l'ordre dans lequel les KPI sont utilisés pour réduire le nombre d'instruments a son importance. En effet, un KPI non validé et étant proche du seuil défini peut être dégradé par le retrait d'un instrument si celui-ci permet d'augmenter le KPI actuellement optimisé. Prenons le cas du R^2 et du $BEL\ ratio$ devant respecter les seuils choisis de 0.9 et 0.6, les autres KPI étant vérifiés. Si on a initialement $R^2 = 0.87$ et $BEL\ ratio = 0.599$, alors il est plus pertinent de commencer par optimiser le $BEL\ ratio$ qui dépassera probablement plus facilement le seuil de 0.6 par suppression d'instruments.

Les résultats de l'application de cet algorithme à l'individu $ID10703$ sont présentés dans le tableau 6.1.

Application de l'algorithme de réduction d'instruments	Sensibilités non validées	R^2	$BEL\ ratio$	Normalité des résidus	Nombre d'instruments
Avant réduction	0	88.28%	60.02%	Oui	62
Après réduction	0	90.18%	60.87%	Oui	43

TABLE 6.1 – Comparaison des KPI avant et après utilisation de l'algorithme de réduction d'instruments sur l'individu $ID10703$

A l'issue de cet algorithme additionnel, l'individu $ID10703$ valide l'ensemble des KPI et est donc à présent un portefeuille de réplication pertinent. Son nombre d'instruments est passé de 62 à 43 et est donc inclus dans l'intervalle requis. Le tableau 6.2 permet d'observer sur quels types d'instruments s'est effectué l'élimination.

Type d'instruments	Avant réduction	Après réduction
Zero-coupon	7	5
Zero-coupon inflation	1	1
Swaption receveur	15	10
Swaption payeur	15	8
Indice equity (×6)	4	4
Call sur equity (×6)	12	9
Put sur equity (×6)	8	6
Total	62	43

TABLE 6.2 – Détail des types d'instruments avant et après algorithme de réduction d'instruments pour l'individu $ID10703$

6.2 Certains KPI non atteints

Lors de la génération finale, certains individus peuvent avoir des KPI proches des seuils requis par le contrôle interne sans les dépasser. L'individu *ID10703* en est un exemple. Lorsque ces individus ont un nombre d'instruments satisfaisant, l'algorithme de réduction d'instruments est à éviter car il pourrait engendrer une forme anormale dans les topologies par le retrait d'un instrument. Afin d'augmenter les KPI sans avoir à modifier le nombre d'instruments, un second algorithme a été implémenté. Son objectif est de modifier certaines caractéristiques des instruments du portefeuille répliquant candidat dans le but d'améliorer les KPI. La nature de l'instrument n'est pas modifiée, seuls sa maturité, son tenor et son strike pourront être échangés par une maturité, un tenor et un strike voisin.

Le principe de ce second algorithme est le suivant :

1. Définir un pas pour la maturité, le tenor et le strike. Ce pas servira à définir la notion de voisinage.
2. Générer une base de données des instruments voisins des instruments du portefeuille répliquant candidat
3. Définir des seuils pour le R^2 , le *BEL ratio*, la p_{value} et les sensibilités financières. Par défaut, ces seuils sont respectivement de 0.9, 0.6, 0.05 et 0.02.
4. Lister les KPI qui ne respectent pas ces seuils
5. Si le nombre de KPI est supérieur à 1 :
 - Laisser le choix à l'utilisateur de l'ordre des KPI à optimiser
6. Pour chaque KPI non vérifié :
 - Pour chaque instrument du portefeuille répliquant candidat :
 - (a) Pour chaque instrument voisin
 - i. Observer les KPI provenant de l'échange de l'instrument du portefeuille répliquant par l'instrument voisin
 - ii. Conserver cet instrument voisin dans une liste L SI il augmente le KPI considéré SANS dégrader les KPI validant déjà les seuils définis
 - (b) Si la liste L est non vide :
 - Echanger l'instrument par l'instrument voisin maximisant le KPI considéré dans le portefeuille répliquant.

Cet algorithme a une complexité plus importante que l'algorithme précédent. Par conséquent, il ne doit être utilisé que lorsque le nombre d'instruments du portefeuille répliquant candidat est raisonnable. Une exécution de l'algorithme de suppression d'instruments en amont de cet algorithme peut être pertinent dans certains cas, comme par exemple pour l'individu *ID10703*.

Enfin, remarquons que la principale faiblesse de l'algorithme génétique peut être dépassée à l'aide de ces deux algorithmes. En effet, converger vers un extremum local ne sera pas déroutant à condition que les KPI des individus de la génération finale soient suffisamment proches des seuils définis par le contrôle interne. Les algorithmes de réduction du nombre d'instruments et d'échange d'instruments pourront permettre de dépasser ces seuils et fournir un portefeuille répliquant répondant aux exigences attendues.

6.3 Le temps de calcul peut encore être réduit

Dans le chapitre précédent, section *Influence des paramètres*, il a été vu que certains paramètres influencent le temps de calcul. Cela provient notamment de différentes boucles itératives utilisées dans le code. Pour certaines de ces boucles, l'ordre d'exécution a son importance car il existe une dépendance entre les éléments de la boucle. Par exemple, il y a dépendance directe entre les individus des différentes générations : les individus de la génération $g + 1$ dépendent directement du profil d'individus de la génération g . Cependant, d'autres boucles itératives ne présentent pas de dépendances entre les éléments qu'elles génèrent. Par exemple, au sein du processus de croisement, chaque ensemble de 4 enfants dépend directement d'un couple de 2 parents et uniquement de ce couple. Un parent ne pouvant former un couple qu'avec un unique autre parent au sein d'une même génération, la création de 4 enfants est indépendante entre chaque couple de parents. Cette indépendance est retrouvée lors de la génération initiale : chaque individu de la génération initiale est généré aléatoirement selon le procédé énoncé dans le chapitre précédent et indépendamment des autres individus. De ce fait, dans le but de réduire davantage le temps de calcul de l'algorithme génétique, il pourrait être intéressant d'envisager de paralléliser certains calculs.

La parallélisation est un ensemble de techniques logicielles et matérielles permettant l'exécution simultanée de séquences d'instructions indépendantes sur des processeurs différents [20]. Afin de paralléliser l'algorithme génétique, certaines boucles séquentielles pourraient être transformées en boucles parallèles comme sur la figure 6.1. Ces boucles ne pourront être parallélisées qu'à condition que les éléments qui la composent soient indépendants entre eux [22].

Utiliser la parallélisation des calculs peut être une piste intéressante afin de réduire davantage le temps de calcul de l'algorithme génétique appliqué au cas des portefeuilles de réplique. Cette parallélisation pourrait par exemple être appliquée :

- à la génération initiale pour la création des individus initiaux
- aux croisements issus de chaque couple de 2 parents qui sont indépendants entre les couples
- aux mutations effectuées sur chaque enfant.

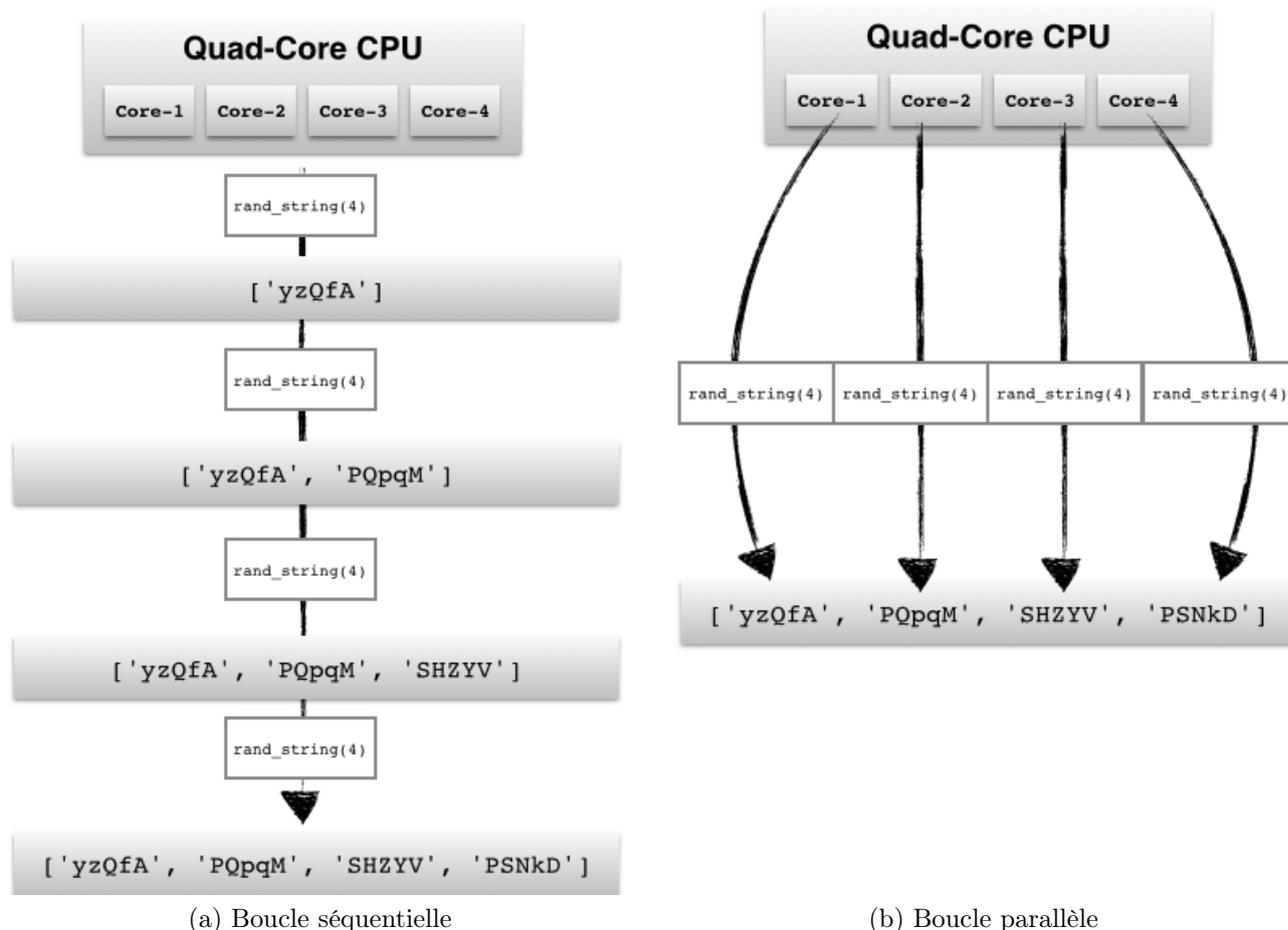


FIGURE 6.1 – Parallélisation d'une boucle

6.4 Les topologies

Enfin, une autre amélioration envisageable concerne l'étude des topologies. Actuellement, que ce soit avec la méthode *à dire d'expert* ou la méthode de l'algorithme génétique, l'étude des topologies n'est réalisée qu'une fois qu'un portefeuille candidat satisfait les KPI. Sachant que le comportement du portefeuille candidat face aux mouvements des facteurs de risque est connu, il peut être intéressant de procéder à cette vérification en amont. En pratique, cela pourrait se traduire par l'intégration du comportement attendu dans la fonction de coût de l'algorithme génétique. Avec l'exemple des taux d'intérêt, on souhaite que le BEL^{RP} diminue lorsque les taux augmentent. Ainsi, si on calcule l'évolution du BEL^{RP} en fonction des taux d'intérêt, on pourrait pénaliser lorsqu'il n'y a pas décroissance de la fonction et donc que la fonction dérivée n'est pas négative, et récompenser sinon. Il faut cependant être prudent avec l'ajout de telles contraintes dans la fonction de coût dans la mesure où cela pourrait davantage augmenter le temps de calcul.

CONCLUSION

La méthode des portefeuilles de réplication est une méthode qui peut se scinder en deux sous-problématiques : la sélection des instruments utilisés pour la réplication et le calcul des nominaux associés à ces instruments. Si l'approche permettant de calculer les nominaux est performante, la méthode actuelle permettant de sélectionner les instruments, très chronophage, mérite d'être challengée.

Ce mémoire propose l'apport d'un outil permettant d'automatiser le processus de sélection d'instruments à l'aide d'un algorithme génétique entièrement implémenté afin de satisfaire aux spécificités de la méthode des portefeuilles répliquants. L'objectif est de réduire le temps de calcul. Ce choix de l'algorithme génétique repose sur la quantité notable d'instruments financiers candidats formant la base de travail. De plus, cet algorithme a également pour avantage de proposer un ensemble de solutions potentielles formé par les individus de la génération finale. Cela permet de pouvoir comparer plusieurs portefeuilles candidats.

Au cours des premières générations et avec la définition de la fonction de coût, l'algorithme commence par générer des individus validant le critère associé aux n_S sensibilités financières de l'étude. Ce sont ensuite les indicateurs de performance du R^2 , du BEL ratio et de la normalité des résidus qui sont optimisés. Afin de satisfaire au critère portant sur le nombre d'instruments, un nombre plus conséquent de générations est nécessaire, ce qui augmente cependant le temps de calcul. Dans l'optique de conserver un temps raisonnable d'exécution d'algorithme, une tolérance sur un nombre d'instruments plus élevé a été accordée. Le nombre requis d'instruments peut alors s'obtenir, pour un individu, par l'utilisation d'un algorithme additionnel permettant de réduire le nombre d'instruments en augmentant les indicateurs de performance. De manière à diminuer davantage le temps de calcul de l'algorithme génétique, la parallélisation de certaines boucles pourrait être envisagée. Cette parallélisation n'est possible que si les éléments constituant la boucle sont indépendants les uns des autres. Cela est par exemple le cas pour les individus de la génération initiale, le processus de croisement pour chacun des couples ainsi que le processus de mutation effectué sur les enfants.

Par sa simplicité d'utilisation et de compréhension, l'algorithme génétique est un algorithme qu'il peut être intéressant d'appliquer à d'autres problématiques d'optimisation en assurance. Le secteur de l'assurance étant en perpétuelle évolution, l'automatisation de certains process représente un gain de temps, réduit le risque opérationnel et apparaît alors comme une nécessité. La réglementation des assureurs français, qui était autrefois à une échelle nationale, prend une ampleur européenne et pourrait se mondialiser. De nouveaux besoins d'automatisation peuvent alors apparaître et il pourrait être intéressant d'envisager l'algorithme génétique comme moyen de réponse si le problème d'optimisation présente un espace des solutions de grande taille.

BIBLIOGRAPHIE

- [1] T. Buberl. Résultats semestriels 2022. 2022.
- [2] M. Chauvigny. Cours de techniques de valorisation de portefeuilles et de calcul de capital économique en assurance-vie. *Ecole Nationale de la Statistique et de l'Administration Economique*, 2021-2022.
- [3] A. Dalalyan. Cours d'introduction à la statistique et à l'économétrie. *Ecole Nationale de la Statistique et de l'Administration Economique*, 2017-2018.
- [4] C. Darwin. *L'origine des espèces au moyen de la sélection naturelle ou la lutte pour l'existence dans la nature*. C. Reinwald, 1887.
- [5] K. A. De Jong. *An analysis of the behavior of a class of genetic adaptive systems*. University of Michigan, 1975.
- [6] L. Devineau and M. Chauvigny. Replicating portfolios : techniques de calibrage pour le calcul du capital économique solvabilité ii. 2010.
- [7] L. Devineau and S. Loisel. Construction d'un algorithme d'accélération de la méthode des «simulations dans les simulations» pour le calcul du capital économique solvabilité ii. *Bulletin Français d'Actuariat*, 10(17) :188–221, 2009.
- [8] C. Dossal. Matching pursuit et orthogonal matching pursuit. 2008.
- [9] J. D. Fermanian. Cours de foundations of risk management. *Ecole Nationale de la Statistique et de l'Administration Economique*, 2021-2022.
- [10] C. Francq. Cours de risk measures. *Ecole Nationale de la Statistique et de l'Administration Economique*, 2021-2022.
- [11] D. E. Golberg. Genetic algorithms in search, optimization, and machine learning. *Addion wesley*, 1989(102) :36, 1989.
- [12] F. Guillemot. Calibration de portefeuilles répliquants. 2021.
- [13] J. H. Holland. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.
- [14] S. Le Mer. Calcul du capital économique en assurance vie. 2009-2010.
- [15] T. Le Roux. *Algorithme génétique*. https://igm.univ-mlv.fr/~dr/XPOSE2013/tleroux_genetic_algorithm/index.html. Page consultée en juillet 2022.
- [16] F. Le Vallois. Cours de réglementation prudentielle en assurance. *Ecole Nationale de la Statistique et de l'Administration Economique*, 2021-2022.
- [17] I.-C. Lerman and R. F. Ngouenet. *Algorithmes génétiques séquentiels et parallèles pour une représentation affine des proximités*. PhD thesis, INRIA, 1995.
- [18] M. A. Mechergui. Evaluation du capital économique sous solvabilité 2 : Mise en place de l'approche curve fitting. 2018.

- [19] Z. Michalewicz et al. A survey of constraint handling techniques in evolutionary computation methods. *Evolutionary programming*, 4 :135–155, 1995.
- [20] G. Moebis. Calcul parallèle avec mpi. 2010.
- [21] S. Novkovic and D. Šverko. Genetic waste and the role of diversity in genetic algorithm simulations. In *Proceedings of the Second Workshop on Economics with Heterogeneous Interacting Agents*, pages 30–31, 1997.
- [22] S. Raschka. An introduction to parallel programming using python’s multiprocessing module. https://sebastianraschka.com/Articles/2014_multiprocessing.html. Page consultée en septembre 2022.
- [23] AXA Group Risk Management. Market risk - replicating portfolios - guidelines and process. *Document interne*, 2022.
- [24] AXA Group Risk Management. Market risk - replicating portfolios - theory and application. *Document interne*, 2022.
- [25] AXA France Solvabilité II. Processus de calibration des portefeuilles répliquants. *Document interne*, 2022.
- [26] A. Toukourou. Portefeuilles répliquant et calcul du capital économique en assurance vie : nouvelles approches de calibration. 2013.
- [27] T. Vallée and M. Yildizoğlu. Présentation des algorithmes génétiques et de leurs applications en économie. *Revue d’économie politique*, pages 711–745, 2004.

ANNEXE A

LA VALUE-AT-RISK

Soit L une variable aléatoire représentant une perte.

La Value-at-Risk, notée VaR, de niveau $\alpha \in (0, 1)$ et à horizon $h \in \mathbb{N}^*$ est le quantile α de la distribution de L :

$$VaR_{h,\alpha}(L) = \inf\{x : \mathbb{P}(L_h \leq x) \geq 1 - \alpha\}.$$

Lorsque la fonction de répartition de L est continue et strictement croissante, la Value-at-Risk de niveau α est égale au quantile $(1 - \alpha)$:

$$VaR(\alpha) := F^{-1}(1 - \alpha).$$

La Value-at-Risk est invariante par translation, positivement homogène et monotone. Ce n'est cependant pas une mesure de risque dite cohérente car elle n'est pas sous-additive [10].

ANNEXE B

LA FORMULE DU SCR

L'égalité entre les termes ci-dessous est démontrée par Devineau et Loisel (2009) [7] :

$$\begin{aligned} SCR &= \arg \min_{x \in \mathbb{R}} \mathbb{P}(Fonds Propres_{t=1} \leq 0 | Fonds Propres_{t=0} = x) < 0.5\% \\ &= Fonds Propres_{t=0} - ZC(0, 1) \times q_{0.5\%}(Fonds Propres_{t=1}). \end{aligned}$$

où

- $ZC(0, 1)$ est le prix en $t = 0$ d'un zéro coupon de maturité 1 an
- $q_{0.5\%}$ est le quantile 0.5%.

Le SCR tel que défini par la première égalité et Solvabilité II est le niveau minimal de fonds propres en date initiale permettant de satisfaire la contrainte $\mathbb{P}(FP_1 < 0) \leq 0.5\%$.

La quantité $-ZC(0, 1) \cdot q_{0.5\%}(FP_1)$ apparaît comme un surplus algébrique qu'il faut ajouter aux fonds propres en date initiale afin de garantir la condition $\mathbb{P}(FP_1 < 0) = 0.5\%$.

Hypothèse : $S = -ZC(0, 1) \cdot q_{0.5\%}(FP_1)$ est investi (ou désinvesti) en cash et évolue la première année au taux zéro-coupon 1 an.

Soient FP_1^{adjust} , A_1^{adjust} et VEP_1^{adjust} les variables économiques après ajout du montant S . On a alors :

$$\begin{aligned} \mathbb{P}(FP_1 < 0) &= \mathbb{P}(A_1^{adjust} - VEP_1^{adjust} < 0) \\ &\simeq \mathbb{P}(A_1 + \frac{S}{ZC(0, 1)} - VEP_1 < 0) \\ &= \mathbb{P}(FP_1 + \frac{S}{ZC(0, 1)} < 0) \\ &= \mathbb{P}(FP_1 < \frac{S}{ZC(0, 1)}) \\ &= 0.5\%. \end{aligned}$$

Cela conclut la démonstration.

ANNEXE C

UN EXEMPLE DE RÉPLICATION PARFAITE

Considérons un portefeuille d'épargne en euros avec un Taux Minimum Garanti (TMG) et une participation aux bénéfiques (PB) [12]. L'assuré confie son épargne à l'assureur en date $t = 0$ et reçoit, à maturité ($t = T$), son épargne capitalisée au TMG et sa PB, tous deux fixés à la signature du contrat. Les caractéristiques de ce contrat sont :

- Maturité $T = 10$ ans
- Prime $P = 1000$
- TMG $r_G = 2\%$
- Taux de PB : $\tau = 85\%$
- Allocation d'actifs : $\delta_A = 20\%$ d'actions et $\delta_O = 80\%$ d'obligations au taux sans risque $r = 3\%$
- L'assuré n'a pas la possibilité de racheter son contrat.

Du point de vue de l'assuré, le payoff du contrat est :

$$\text{Payoff} = G + PB.$$

Avec G le montant garanti tel que :

$$G = P(1 + r_G)^T \simeq 1219.$$

La participation aux bénéfiques PB est telle que :

$$PB = \tau \left(P \left[\delta_A \frac{S_T}{S_0} + \delta_O (1 + r)^T \right] - G \right)_+$$

où

- S_t désigne la valeur des actions en date t
- $(\cdot)_+$ désigne la fonction partie positive : $(x)_+ = \max(x, 0)$.

Par définition de G , on a :

$$\begin{aligned} PB &= \tau \left(P \left[\delta_A \frac{S_T}{S_0} + \delta_O (1 + r)^T \right] - P(1 + r_G)^T \right)_+ \\ &= \tau P \delta_A \left(\frac{S_T}{S_0} - \frac{1}{\delta_A} \left[(1 + r_G)^T - \delta_O (1 + r)^T \right] \right)_+. \end{aligned}$$

En posant $\omega_C = \tau P \delta_A = 170$, il est possible d'exprimer PB comme le payoff de ω_C calls sur

action de maturité T et de strike $K = \frac{1}{\delta_A} \left[(1 + r_G)^T - \delta_0(1 + r)^T \right] = 71.9\%$.

La partie garantie s'interprète alors comme le flux de trésorerie de $\omega_{ZC} = P(1 + r_G)^T = 1219$ zéro-coupons de maturité $T = 10$ ans.

Cet exemple de répliation parfaite est un cas idéal. En réalité, les contrats doivent intégrer de nombreux autres paramètres, comme une allocation dynamique d'actifs, des rachats de contrats, des frais...

Dans ce contexte, l'objectif de la répliation n'est pas de répliquer parfaitement la valeur du passif mais d'en obtenir une bonne approximation.

ANNEXE D

LE MEILLEUR INDIVIDU PAR GÉNÉRATION DE L'EXEMPLE INTRODUCTIF

Génération	Meilleur individu
1	^Tr.wSvnr_ib nVelpN Cle
2	E[rwqcwuZlQyfRWNtzG]u.t
3	V[rwqHquZlQE KsBfij_nxC
4	V[rwqHquZlQE KsBfiqTPOz
5	oYutek'uxjiL RfgNyXaF\t
6	YFfthiecZilsQqeGRiGSRet
7	EqrtufIJllqZCYehlhq^^Wt
8	goytek'uxjeL RfgNij_nx\
9	P^lzCAuuYICZ^GepoiQEang
10	P^lzCAuuYICZ^GepoiQEaDt
11	sYrt gaRMllfWSeql^ixatt
12	PuiaeAuuYICZ^GepoiQEaDt
13	PC_IRfeuTpl^.sBfijuQVt
14	bKrheSehal_JpRehxiNWanI
15	Pdrpbfv.cmlY .PplY.WanI
16	gorhZceRxll aseHliquQVC
17	PYrtnOh Yllh nIplGslanI
18	m.rMe euisDhhR^pliquuh\
19	P[rthfouNlahxzepI Gqm_pt
20	LRrueieicibDImhepliquatt
21	PSrtqHquGll Res.^quRet
22	PoltuFeDibDImhepliquatt
23	PSrtqcquGll Res.^quant
24	Sowt geuYlleV[vpliquanI
25	ortekeuClleZReGRiGuGet
26	gowt geuYlleV[vpliquant
27	Pdrpbfkuillh hepliqua[A
28	Pdrpbfkuillh hepliquanI
29	PC]tRfeuTllh neHli^uant

Génération	Meilleur individu
30	Purffkuillh hepliquanI
31	POrthfeuTllh Rbpliquatt
32	PoltRfeuTll Resliqatt
33	PdrJefkuillI Rbpliquang
34	cTrtequeullh RepliqjanU
35	Poetcfueiilhchepliquant
36	PowtefeuiilhQkepliTuagt
37	ZortufeuiOly RbpNiquant
38	Portufeaille Rqpliq amO
39	P^rtgfsuillg RxpGiquant
40	PortefeuiilhQkepliTuagt
41	ZortufeuiOly Rbpliquant
42	Eortefe\ll Vepliquant
43	PortefhuiCdq xepliquant
44	zorthfeuiCle hepliqkant
45	cTrtefeuille Reslissant
46	PovteWeuille RepliOuaVt
47	Portefhuillx Repoiquant
48	Portefeuille cePliquaNt
49	PRrtefeuille GepGiquant
50	PRrtefeuille GepGiquant
51	PortefeuTlld RepliTuanS
52	Pqrtefeuiilh RIpliquant
53	Portefhuill_ RKpliquant
54	Porteffuille hFpliquant
55	UorteWeuille Rypliquant
56	[ortefeuille nepliquaFt
57	D]rtefeuille Repliquaet
58	PortRpeuille RepAiquant
59	PortefLuille Repliquynt
60	Purtefeuille cepliquant
61	Portefeuille cepliquant
62	PortefeuiAle Repliquunt
63	PortefeuiAle Repliquant
64	Portefeuille ReplG\uant
65	PortefeuiCle RepViquant
66	Portefeuille nepliquant
67	PortefeKille cepliquant
68	Portefeuille Repliqkant
69	Portefeuille Repliqkant
70	Portefeuille RKpliquant
71	Portefeuille ReTliquant
72	Portefeuille ReTliquant
73	Portefeuille Renliquant
74	Portefeuille Repliquant

LE TEST DE KOLMOGOROV-SMIRNOV

Le test de Kolmogorov-Smirnov est un test d'ajustement à une loi continue. Soit un échantillon (X_1, \dots, X_n) un échantillon d'une loi inconnue P . L'hypothèse nulle est

$$\mathcal{H}_0 : \text{la loi } P \text{ a pour fonction de répartition } F_0$$

où F_0 est la fonction de répartition d'une loi continue donnée.

Si l'hypothèse \mathcal{H}_0 est correcte, alors la fonction de répartition empirique \hat{F} de l'échantillon doit être proche de F_0 . La fonction de répartition empirique est la fonction de \mathbb{R} dans $[0,1]$, qui vaut :

$$\hat{F}(x) = \begin{cases} 0 & \text{pour } x < X_{(1)} \\ \frac{i}{n} & \text{pour } X_{(i)} \leq x \leq X_{(i+1)}, \forall 1 \leq i \leq n \\ 1 & \text{pour } x \geq X_{(n)} \end{cases}$$

où les $X_{(i)}$ sont les statistiques d'ordre de l'échantillon.

On mesure l'adéquation de la fonction de répartition empirique à la fonction F_0 par la distance de Kolmogorov-Smirnov, qui est la distance de la norme uniforme entre fonctions de répartition. Afin de la calculer, il suffit d'évaluer la différence entre \hat{F} et F_0 aux points $X_{(i)}$.

$$D_{KS}(F_0, \hat{F}) = \max_{1 \leq i \leq n} \left\{ \left| F_0(X_{(i)}) - \frac{i}{n} \right|, \left| F_0(X_{(i)}) - \frac{i-1}{n} \right| \right\}.$$

Sous l'hypothèse \mathcal{H}_0 , la loi de la statistique $D_{KS}(F_0, \hat{F})$ ne dépend pas de F_0 car les image de X_i par F_0 sont des variables aléatoire de la loi $\mathcal{U}(0, 1)$.

Sous l'hypothèse \mathcal{H}_0 , on a pour tout $t \geq 0$:

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\mathcal{H}_0}(\sqrt{n}D_{KS}(F_0, \hat{F}) \leq t) = 1 - 2 \sum_{k=1}^{+\infty} (-1)^{k+1} \exp(-2k^2 t^2).$$

La série converge très rapidement.

Si \mathcal{H}_0 est fausse, $\sqrt{n}D_{KS}(F_0, \hat{F})$ tend vers $+\infty$ avec n . Le test est donc nécessairement unilatéral à droite : on rejète des valeurs trop grandes [3].

Remarque :

Dans cette étude, la normalité des résidus est testée. C'est donc l'appartenance à la famille de lois $F(0, \hat{\sigma}^2)$ et non à une loi précise qui est testée.

Dans ce cas, il peut être noté l'existence d'un biais qui subsiste asymptotiquement, ce qui réduit la puissance du test.

Un test alternatif permettant d'étudier la normalité des résidus pourrait être le test de Jarque Bera dont l'hypothèse nulle est :

$$\mathcal{H}_0 : \text{les données suivent une loi normale}$$

et dont la statistique de test fait intervenir les moments d'ordre 3 et 4.

NOTE DE SYNTHÈSE

Dans son communiqué de presse datant du 03 août 2022, AXA présente un ratio de solvabilité II égal à 227% et en hausse de 10 points par rapport à la fin de l'année 2021 [1]. Ce communiqué de presse fait suite à une des exigences issue de la réglementation Solvabilité II qui requiert le calcul ainsi que la publication du ratio de solvabilité des entreprises d'assurance.

Une des composantes nécessaire au calcul de ce ratio est le capital de solvabilité requis (SCR) qui correspond au capital économique minimal dont doit disposer une compagnie d'assurance pour faire face à une ruine économique à horizon 1 an avec une probabilité de 99.5%. Il est obtenu par agrégation de SCR évalués par module de risque.

Au sein d'AXA France, le SCR de marché est calculé en modèle interne par la méthode des portefeuilles de réplication par l'équipe Modèles et Solvabilité II. Cette méthode consiste à déterminer un ensemble d'instruments financiers facilement valorisable capable de répliquer le *Best Estimate Liabilities*. Deux sous-problématiques apparaissent : la sélection des instruments utilisés pour la réplication et le calcul des nominaux associés à ces instruments.

L'approche permettant d'obtenir les nominaux est basée sur la régression ridge et la factorisation LU, elle ne fait pas l'objet de modifications. Seule la méthode actuelle de sélection d'instruments, qui est très chronophage, est challengée dans ce mémoire.

Une combinaison d'instruments appropriée doit répondre à certaines exigences. En effet, le comité de contrôle interne d'Axa France impose plusieurs contraintes permettant d'évaluer la qualité d'un portefeuille de réplication. Un portefeuille de réplication pertinent doit satisfaire les exigences suivantes :

- $R^2 \geq 0.9$
- $BEL\ ratio \geq 0.6$
- vérifier que l'erreur de réplication sur chaque sensibilité financière soit inférieure à 2% du SCR de marché de l'exercice précédent calculé en modèle interne
- vérifier la normalité des résidus.

R^2 , $BEL\ ratio$, sensibilités financières et normalité des résidus forment les indicateurs de performance d'un portefeuille répliquant.

En plus de ces indicateurs de performance, une contrainte supplémentaire porte sur le nombre d'instruments composant un portefeuille répliquant.

L'ensemble des instruments financiers candidats permettant de répliquer le BEL ainsi que les effets qu'ils capturent sont résumés dans le tableau ci-dessous.

Une base de données formée de ces instruments pour des ensembles de maturité, tenor et strike différents est de l'ordre de plusieurs milliers d'instruments. Cependant, en pratique, le nombre p d'instruments financiers composant un portefeuille répliquant doit être inférieur à 50.

Type d'instrument	Effet capturé
Zero-coupon	Garanties
Zero-coupon inflation	Sensibilité à l'inflation
Swaption receveur	Convexité due aux rachats et réinvestissements
Swaption payeur	Convexité due aux rachats et réinvestissements
Indice equity	Exposition directe
Call sur equity	Participations aux bénéfiques
Put sur equity	Participations aux bénéfiques

TABLE E.1 – Synthèse des instruments utilisés et les effets qu'ils capturent

L'espace des combinaisons possibles des instruments constituant un portefeuille répliquant candidat est donc très grand. C'est une des raisons pour laquelle l'utilisation d'un algorithme génétique a été envisagée pour challenger la méthode *à dire d'expert*.

Un algorithme génétique est un algorithme d'intelligence artificielle évolutionniste basé sur la théorie de l'évolution de Charles Darwin. Croisements, mutations et sélections permettent d'optimiser un individu (solution potentielle) au sein d'une population (ensemble de solutions potentielles) par l'optimisation de son patrimoine génétique (paramètres). Cette optimisation peut s'opérer grâce à une fonction d'évaluation (ou fonction de coût) qui permet d'ordonner les différents individus selon la qualité de leurs gènes. Cela en fait une composante essentielle du processus de sélection.

L'ensemble d'individus de la génération finale peut être perçu comme l'ensemble des solutions répondant au mieux au problème d'optimisation donné. Ainsi, dans le cadre des portefeuilles répliquants, cet algorithme présente également l'avantage de proposer un ensemble de solutions potentielles formé par les individus de la génération finale. Cela permet de pouvoir comparer plusieurs portefeuilles candidats pour une unique exécution d'algorithme.

Afin de pouvoir répondre aux spécificités de la méthode des portefeuilles de réplication, l'automatisation de la sélection d'instruments par algorithme génétique a été entièrement implémentée. La fonction de coût retenue, basée sur les indicateurs de performance d'un portefeuille répliquant, est la suivante :

$$C = \left(1 - \frac{R^2}{0.9}\right)_+ + \left(1 - \frac{BEL\ ratio}{0.6}\right)_+ + \left(\frac{p}{35} - 1\right)_+ + \mathbb{1}_{p_value \leq 0.05} + \sum_{j=1}^{n_S} \left(\frac{|Y_j^S - (X^S \beta)_j|}{0.02 \times SCR_{marche}^{an-1}} - 1\right)_+$$

où

- p est le nombre d'instruments du portefeuille répliquant
- p_value est la p-valeur du test de Kolmogorov-Smirnov
- Y_j^S est la sensibilité j du passif à répliquer
- X^S est le vecteur des sensibilités du portefeuille répliquant
- β est l'ensemble des nominaux associés aux p instruments financiers
- $(\cdot)_+$ est la fonction partie positive.

Il s'agit d'une fonction à minimiser : plus C s'approche de zéro, meilleur sera l'individu considéré.

La fonction de coût ainsi définie imite la recherche à *dire d'expert*. En effet, le critère des sensibilités est d'abord optimisé car l'ensemble des sensibilités a plus de poids que les autres indicateurs de performance en début d'algorithme.

Dans l'optique de présenter les performances de l'algorithme génétique pour la sélection d'instruments financiers, une exécution d'algorithme d'une durée de 58 minutes a permis d'obtenir les résultats qui suivent. Cette exécution est appliquée à un portefeuille de retraite à répliquer, noté *Portefeuille A*.

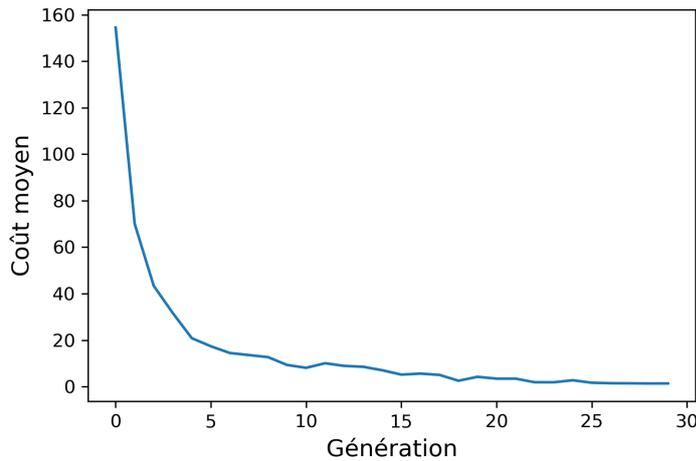
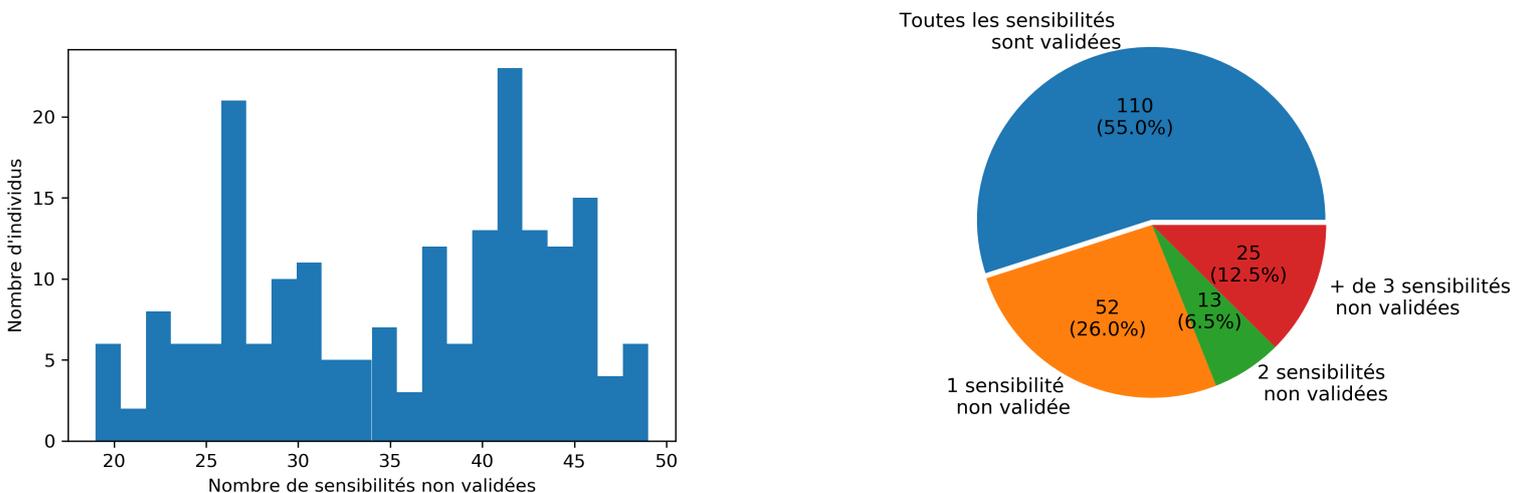


FIGURE E.1 – Coût moyen des individus par génération

Sur la figure ci-dessus, la décroissance exponentielle de la fonction de coût moyenne au fil des générations montre bien que l'algorithme génétique optimise les individus au cours des générations.



(a) Nombre de sensibilités non validées des individus de la génération initiale

(b) Validation des sensibilités de la génération finale

FIGURE E.2 – Comparaison du critère de sensibilités financières pour les générations initiale et finale

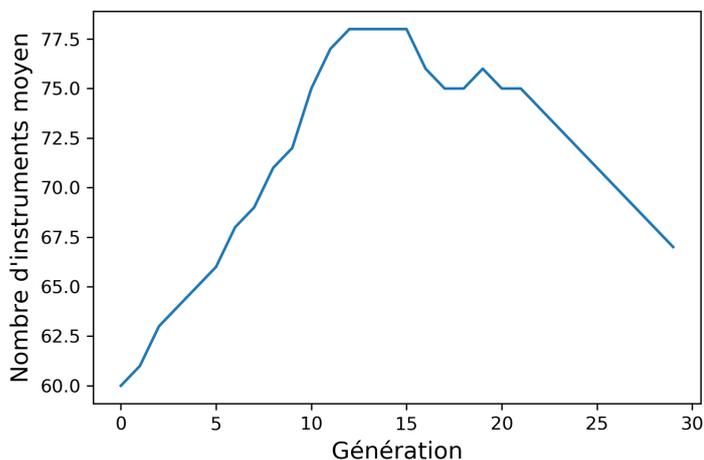
L'histogramme et le diagramme en secteurs permettent de comparer les performances des générations initiale et finale sur l'indicateur de validation des sensibilités financières. Sur ce seul critère de sensibilité financière, lorsque l'individu de la génération initiale avec la meilleure performance ne valide pas 19 des 58 sensibilités, plus de 75% des individus de la génération finale en valident au moins 57.

Les indicateurs de performance du R^2 , du $BEL\ ratio$ et de la normalité des résidus sont également optimisés par l'algorithme génétique. Le tableau ci-dessous en témoigne en détaillant les performances de 2 individus : le meilleur individu de la génération finale et le premier individu ne satisfaisant pas au moins l'un des indicateurs de performance.

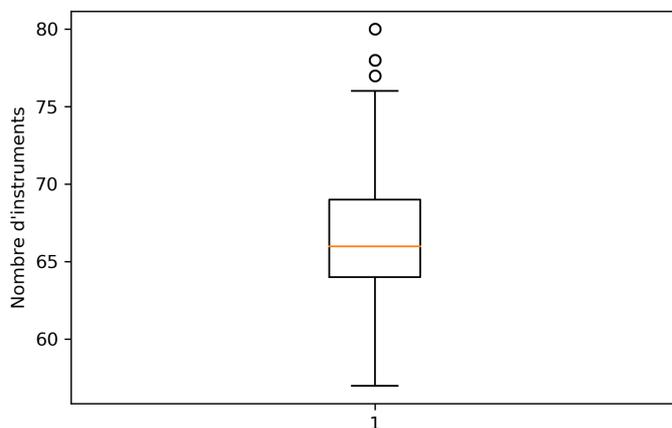
Classement au sein de la génération finale	Identifiant	Sensibilités non validées	R^2	$BEL\ ratio$	Normalité des résidus	Nombre d'instruments
1 ^{er}	9363	0	90.56%	61.11%	Oui	57
4 ^{eme}	10703	0	88.28%	60.02%	Oui	62

TABLE E.2 – Performance de deux individus de la génération finale

Globalement, l'algorithme génétique parvient à produire des individus vérifiant les indicateurs de performance ou, *a minima*, s'approchant des seuils définis. Cependant, l'évolution du nombre moyen d'instruments par génération permet de noter que ce nombre est encore éloigné de l'objectif des 35 instruments défini dans la fonction de coût. Le boxplot précise ce constat car aucun individu n'a pu suffisamment réduire son nombre d'instruments en 30 générations. Ceci est une limite de l'utilisation de l'algorithme génétique qui réduit difficilement ce nombre en un temps raisonnable.



(a) Nombre d'instruments moyen par génération



(b) Boxplot du nombre d'instruments de la génération finale

FIGURE E.3 – Nombre d'instruments moyen par génération et nombre d'instruments de la génération finale

Afin de satisfaire au critère portant sur le nombre d'instruments, un nombre plus conséquent de générations est nécessaire, ce qui augmente toutefois le temps de calcul. Dans le but de conserver un temps raisonnable d'exécution d'algorithme, une tolérance sur un nombre d'instruments plus élevé a été accordée. Le nombre requis d'instruments peut alors s'obtenir, pour un individu, par l'utilisation d'un algorithme additionnel permettant de réduire le nombre d'instruments en augmentant les indicateurs de performance. Un instrument sera retiré s'il améliore

l'indicateur non validé sans dégrader les autres. Le tableau ci-dessous permet d'apprécier les performances de cet algorithme sur un individu ne satisfaisant pas le critère de R^2 (mais étant proche du seuil de 0.9), et possédant un nombre d'instruments supérieur à 50. A l'issue de cet algorithme additionnel et sur l'individu considéré, le nombre d'instruments passe de 62 à 43 et tous les indicateurs de performance sont vérifiés.

Application de l'algorithme de réduction d'instruments	Sensibilités non validées	R^2	<i>BEL ratio</i>	Normalité des résidus	Nombre d'instruments
Avant	0	88.28%	60.02%	Oui	62
Après	0	90.18%	60.87%	Oui	43

TABLE E.3 – Comparaison des indicateurs de performance avant et après utilisation de l'algorithme de réduction d'instruments sur un individu

Enfin, de manière à diminuer davantage le temps de calcul de l'algorithme génétique, la parallélisation de certaines boucles pourrait être envisagée. Cette parallélisation n'est possible que si les éléments constituant la boucle sont indépendants les uns des autres. Cela est par exemple le cas pour les individus de la génération initiale, le processus de croisement pour chacun des couples ainsi que le processus de mutation effectué sur les enfants.

Par sa simplicité d'utilisation et de compréhension, l'algorithme génétique est un algorithme qu'il peut être intéressant d'appliquer à d'autres problématiques d'optimisation en assurance. Le secteur de l'assurance étant en perpétuelle évolution, l'automatisation de certains process présente un gain de temps, réduit le risque opérationnel et apparaît alors comme une nécessité. La réglementation des assureurs français, qui était autrefois à une échelle nationale, prend une ampleur européenne et pourrait se mondialiser. De nouveaux besoins d'automatisation peuvent alors apparaître et il pourrait être intéressant d'envisager l'algorithme génétique comme moyen de réponse si le problème d'optimisation présente un espace des solutions de grande taille.

EXECUTIVE SUMMARY

In a press release published on August 3, 2022, AXA presents a solvency II ratio equal to 227% and up by 10 points compared to the end of 2021 [1]. This press release follows one of the requirements resulting from the Solvency II regulation which requires the calculation as well as the publication of the solvency ratio of insurance companies.

One of the components necessary for the calculation of this ratio is the Solvency Capital Requirement (SCR) which corresponds to the minimum economic capital that an insurance company must have to face an economic ruin at a horizon of 1 year with a probability of 99.5%. It is obtained by aggregating SCRs evaluated per risk module.

Within AXA France, the market SCR is calculated using the internal model with the replicating portfolio method by the model and Solvency II team. This method consists of determining a set of financial instruments that can be easily valued and that are capable of replicating the Best Estimate Liabilities. Two sub-problems appear : the selection of the instruments used for replication and the calculation of the nominal amounts associated with these instruments. The approach to obtain the nominals is based on ridge regression and LU factorization, and is not subject to change. Only the current method of instrument selection, which is very time consuming, is challenged in this paper.

An appropriate combination of instruments must meet certain requirements. Indeed, Axa France's internal control committee imposes several constraints to assess the quality of a replicating portfolio. A relevant replicating portfolio must meet the following requirements :

- $R^2 \geq 0.9$
- $BEL\ ratio \geq 0.6$
- ensure that the replicating error of each financial sensitivity is below 2% of the previous year's market SCR calculated using the internal model
- ensure the normality of the residuals.

R^2 , $BEL\ ratio$, financial sensitivities and normality of residuals constitute the key performance indicators of a replicating portfolio.

In addition to these key performance indicators, there is an other requirement regarding the number of instruments composing a replicating portfolio. The table below summarizes all the financial instruments that are candidates for replicating the BEL and the effects they capture. A database of these instruments for different maturity, tenor and strike sets is estimated to contain several thousand instruments. However, in practice, the number p of financial instruments constituting a replicating portfolio should be less than 50.

Type of instrument	Captured effect
Zero-coupon	Guarantees
Zero-coupon inflation	Sensitivity to inflation
Receiver swaption	Convexity due to lapse and reinvestment
Payer swaption	Convexity due to lapse and reinvestment
Equity index	Direct exposure
Equity call	Profit sharing
Equity put	Profit sharing

TABLE E.4 – Summary of the instruments used and the effects they capture

The universe of possible combinations of instruments composing a potential replicating portfolio is therefore very large. This is one of the reasons why the use of a genetic algorithm has been considered to challenge the current method.

A genetic algorithm is an evolutionary artificial intelligence algorithm based on Charles Darwin's evolution theory. Crossovers, mutations and selections allow to optimize an individual (potential solution) within a population (set of potential solutions) by optimizing its genetic patrimony (parameters). This optimization can be done thanks to an evaluation function (or cost function) which allows to order the different individuals according to the quality of their genes. It is thus an essential component of the selection process.

The set of individuals of the final generation can be seen as the set of solutions that best fit the given optimization problem. Thus, in the context of replicating portfolios, this algorithm also has the advantage of proposing a set of potential solutions formed by the individuals of the final generation. This enables the comparison of several candidate portfolios for a single execution of the algorithm.

In order to meet the specificities of the replicating portfolio method, the automation of the instrument selection by genetic algorithm has been fully implemented. The chosen cost function, based on the key performance indicators of a replicating portfolio, is :

$$C = (1 - \frac{R^2}{0.9})_+ + (1 - \frac{BEL\ ratio}{0.6})_+ + (\frac{p}{35} - 1)_+ + \mathbb{1}_{p_value \leq 0.05} + \sum_{j=1}^{n_S} (\frac{|Y_j^S - (X^S \beta)_j|}{0.02 \times SCR_{marche}^{an-1}} - 1)_+$$

where

- p is the number of instruments in the replicating portfolio
- p_value is the p-value of the Kolmogorov-Smirnov test
- Y_j^S is the sensitivity j of the BEL to be replicated
- X^S is the vector of sensitivities of the replicating portfolio
- β is the set of nominals associated with the p financial instruments
- $(\cdot)_+$ is the positive part function.

It is a function to minimize : the closer C is to zero, the better the individual will be.

The cost function defined in this way reproduces the expert opinion research. Indeed, the criterion of sensitivities is first optimized because the set of sensitivities has more weight than the other key performance indicators at the beginning of the algorithm.

In order to present the performance of the genetic algorithm for the selection of financial instruments, an algorithm execution of 58 minutes duration has allowed to obtain the following results. This execution is applied to a pension portfolio to be replicated, noted *Portfolio A*.

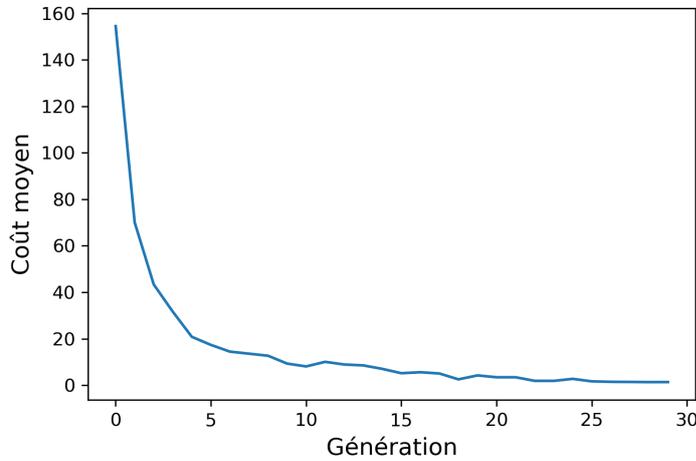


FIGURE E.4 – Average cost of individuals per generation

In the figure above, the exponential decrease of the average cost function over the generations shows that the genetic algorithm optimizes individuals over the generations.

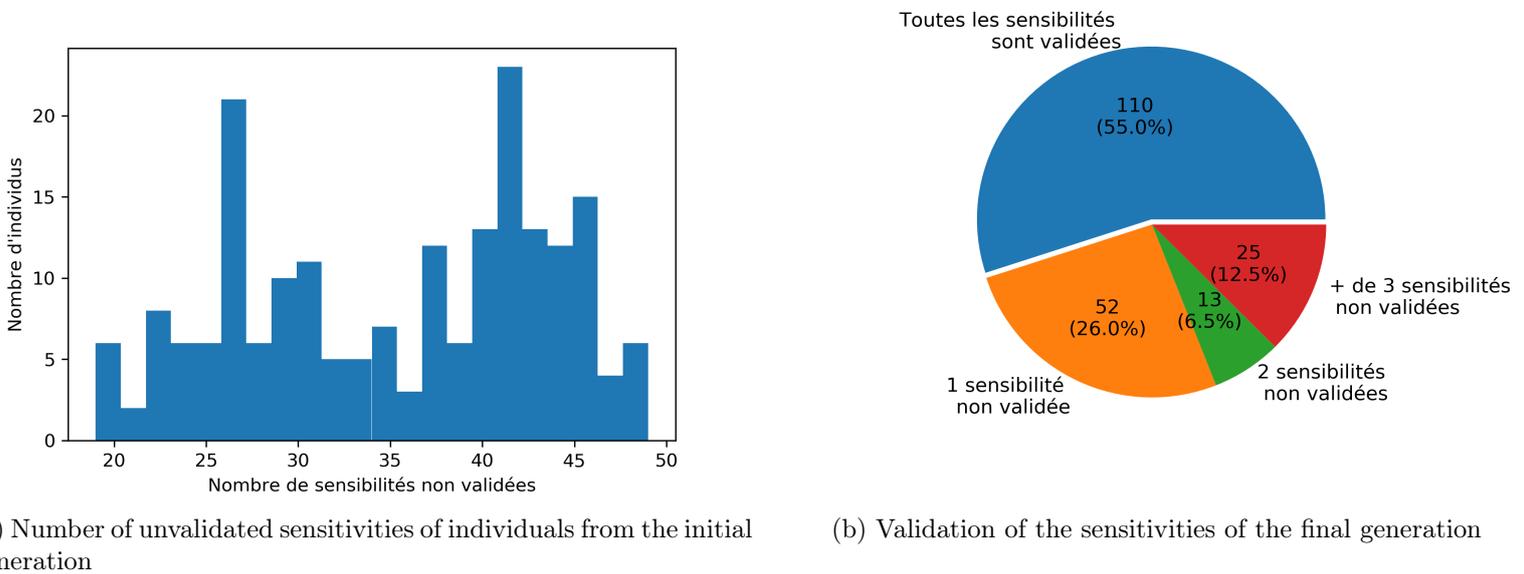


FIGURE E.5 – Comparison of financial sensitivities for the initial and final generations

The histogram and pie chart enable a comparison of the performance of the initial and final generations on the financial sensitivity validation indicator. On this single criterion of financial sensitivity, the best individual from the initial generation does not validate 19 of the 58 sensitivities while more than 75% of the individuals from the final generation validate at least 57 sensitivities.

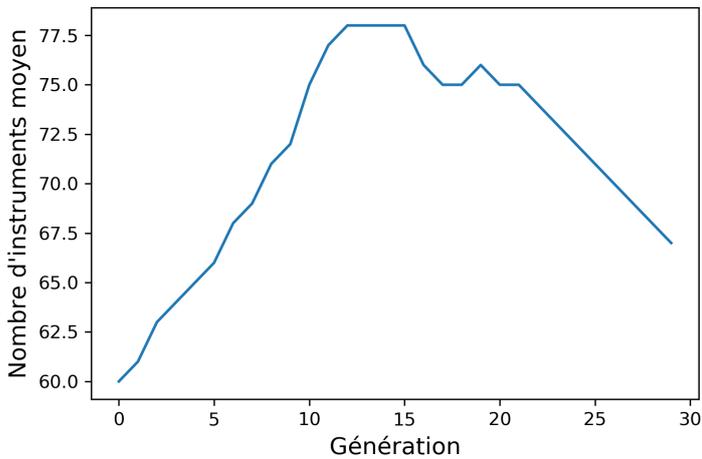
The key performance indicators of R^2 , *BEL ratio* and residuals normality are also optimized

by the genetic algorithm. The table below illustrates this by detailing the performances of 2 individuals : the best individual of the final generation and the first individual not satisfying at least one of the indicators.

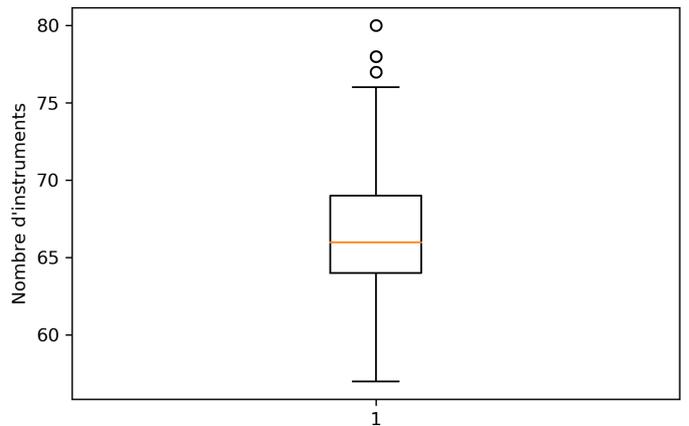
Ranking within the final generation	ID	Non-validated sensitivities	R^2	BEL ratio	Normality of the residuals	uombre of instruments
1 st	9363	0	90.56%	61.11%	Yes	57
4 th	10703	0	88.28%	60.02%	Yes	62

TABLE E.5 – Performance of two individuals of the final generation

Overall, the genetic algorithm manages to produce individuals that verify the key performance indicators or, at least, approach the defined thresholds. However, the evolution of the average number of instruments per generation allows to note that this number is still far from the objective of 35 instruments defined in the cost function. The boxplot specifies this observation because no individual has been able to sufficiently reduce its number of instruments in 30 generations. This is a limitation of the use of the genetic algorithm which hardly reduces this number in a reasonable time.



(a) Average number of instruments per generation



(b) Boxplot of the number of instruments of the final generation

FIGURE E.6 – Average number of instruments per generation and number of instruments in the final generation

To satisfy the criterion of the number of instruments, a larger number of generations is required, which however increases the computation time. To maintain a reasonable algorithm execution time, a tolerance on a higher number of instruments has been granted. The required number of instruments can then be obtained, for an individual, by using an additional algorithm to reduce the number of instruments by increasing the key performance indicators. An instrument will be removed if it improves the non-validated indicator without degrading the others. The table below shows the performance of this algorithm for an individual who does not meet the R^2 criterion (but is close to the 0.9 threshold), and who has a number of instruments greater than 50. At the end of this additional algorithm and for the given individual, the number of instruments is reduced from 62 to 43 and all the performance indicators are verified.

Use of the instrument reduction algorithm	Non-validated sensitivities	R^2	<i>BEL ratio</i>	Normality of the residuals	Number of instruments
Before	0	88.28%	60.02%	Yes	62
After	0	90.18%	60.87%	Yes	43

TABLE E.6 – Comparison of the key performance indicators before and after using the instrument reduction algorithm on a given individual

Finally, to further decrease the computation time of the genetic algorithm, the parallelization of some loops could be considered. This parallelization is only possible if the elements constituting the loop are independent from each other. This is for example the case for the individuals of the initial generation, the crossing process for each of the couples as well as the mutation process carried out on the children.

Because of its simplicity of use and understanding, the genetic algorithm is an algorithm that could be interesting to apply to other optimization problems in insurance. The insurance industry is constantly evolving and the automation of certain processes saves time and reduces operational risk. The regulation of French insurers, which used to be on a national scale, is now taking on a European dimension and could even become a matter of globalization. New automation needs may then appear and it could be interesting to consider the genetic algorithm as a solution if the optimization problem has a large solution space.