

Mémoire présenté devant l'ENSAE Paris
pour l'obtention du diplôme de la filière Actuariat
et l'admission à l'Institut des Actuaires
le 16/03/2022

Par : **Hugo Boccovi**

Titre : **Optimisation de l'allocation du risque entre
différentes Lignes d'activité au sein d'un portefeuille
d'assurance**

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

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

Membres présents du jury de la filière

Entreprise :

AXA GIE 


Nom :

Signature :

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

Directeur du mémoire en entreprise :

Nom : François TCHANGAI

Signature : 

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


Secrétariat :

Signature du responsable entreprise



Bibliothèque :

Signature du candidat



Remerciements

Avant toute chose, je souhaite remercier chaleureusement François Tchangai pour son accompagnement précieux tout au long du stage, et ses nombreuses suggestions et commentaires pertinents sans quoi cette étude n'aurait pas eu lieu.

Je souhaite également remercier particulièrement Christian Goanvic, ainsi que Xavier Serval et son équipe, qui ont su évaluer la pertinence de l'étude et offrir des perspectives sur les améliorations et développements possibles de l'analyse.

Résumé

Cette étude, réalisée sur un portefeuille d'assurance de AXA, a pour objectif d'élaborer un algorithme capable de déterminer la manière optimale de développer les lignes d'assurance d'une entreprise. La question est donc d'identifier de manière algorithmique les lignes d'assurance pour lesquelles il est optimal d'augmenter ou de diminuer l'exposition au risque. L'approche développée pour y parvenir consiste à utiliser les données du modèle interne afin de calculer des indicateurs de risque et de rendement du portefeuille d'assurance. De la sorte, il est possible d'adapter les travaux réalisés en analyse financière au sujet de l'optimisation de portefeuille, et donc de déterminer une frontière efficiente.

Dès lors, le corps de cette étude consiste tout d'abord à déterminer comment évoluent les indicateurs de risque et de rendement en fonction de l'évolution des lignes d'assurance. Une attention particulière a notamment été donnée à l'impact que la réassurance pouvait avoir sur ces indicateurs. Cela permet d'écrire de manière formelle le programme d'optimisation à résoudre.

Une fois ce problème écrit, sa résolution peut être effectuée à l'aide d'algorithmes d'optimisations. Dans le cadre de cette étude, c'est l'algorithme génétique NSGA-II qui a été retenu. De la sorte, la frontière efficiente, a pu être tracée. Enfin, une attention particulière a pu être donnée à l'application pratique des résultats de l'algorithme, ainsi qu'à son périmètre d'application. Pour ce faire, la significativité et la stabilité des résultats de l'algorithme génétique ont fait l'objet d'une étude approfondie.

Abstract

This study, performed on an AXA insurance portfolio, aims to formulate an algorithm capable of finding the optimal way of developing the lines of business of an insurance company. The goal is thus to identify the lines for which it can be said to be optimal to increase or decrease risk exposure.

To reach this goal, AXA's internal model was used to compute risk and return measures for the insurance portfolio. This makes it possible to use the analyses of quantitative finance on portfolio optimization to find the efficient frontier.

As such, this study opens by finding how the risk and return metrics evolve depending on the size of the relevant insurance lines. A particular point of interest at this stage was the impact that a reinsurance structure could have on these metrics. This made it possible to formulate the optimization problem.

Once a formal optimization problem has been written, it can be solved by using an optimization algorithm. Here, the genetic algorithm NSGA-II was chosen. As such, the efficient frontier could be found. Finally, the practical applications of the results of the algorithm were considered. To that end, the statistical significance and stability of the results were studied.

Table des matières

Table des figures	4
Introduction	6
1 Les enjeux stratégiques de cette étude	8
1.1 Les défis et enjeux de l'activité d'assurance	8
1.1.1 Des sources de revenus génératrices de risque	9
1.1.2 La notion de Capital risque	11
1.2 Problématique : la recherche d'une méthode pour optimiser le profil rendement/risque d'un portefeuille de lignes d'assurance	13
1.3 Les méthodes existantes mais insuffisantes pour répondre à cette attente . .	13
1.3.1 Les indicateurs concentrés uniquement sur la performance technique .	14
1.3.2 Les mesures de risque usuelles	15
1.3.3 Le Ratio Économique Combiné : Un premier indicateur qui tente de concilier risque et rendement	16
1.3.4 Le Ratio Économique combiné comme indicateur cependant statique et inapte à l'optimisation	17
1.4 Les caractéristiques de la base de données d'assurance utilisée pendant cette étude	18
1.4.1 Une base de données constituée de simulations de variables aléatoires	18
1.4.2 Une base de données agrégée par ligne d'assurance	20
2 L'écriture du programme d'optimisation brut de réassurance	22
2.1 La modélisation de l'optimisation d'un portefeuille d'actifs : la frontière efficiente de Markowitz	22
2.1.1 Le problème à la base des mathématiques financières : l'optimisation de l'arbitrage entre rendement et risque	23
2.1.2 Le résultat du programme d'optimisation : la frontière efficiente . . .	24
2.1.3 Passer du problème d'optimisation d'un portefeuille d'actifs financiers à un portefeuille de lignes d'assurance	26
2.2 Le rendement économique d'un portefeuille de lignes d'assurance	27
2.2.1 Définir le rendement d'une ligne d'assurance	28
2.2.2 Définir le rendement du portefeuille d'assurance	30
2.2.3 Les hypothèses liées à l'impact du facteur d'évolution sur le portefeuille d'assurance	31

2.2.4	Estimer le rendement après application d'une stratégie d'évolution . . .	34
2.3	Une mesure de risque du portefeuille d'assurance : le Coût en capital	35
2.3.1	Le constat à l'origine de la notion de coût en capital	36
2.3.2	La définition formelle du coût en capital	37
2.3.3	Les hypothèses nécessaires au calcul du coût en capital	38
2.3.4	Écrire le programme d'optimisation de l'assureur	40
3	L'impact d'une structure de réassurance donnée sur le programme d'optimisation	43
3.1	La structure de réassurance du portefeuille étudié	43
3.1.1	La réassurance comme composante essentielle du profil rendement / risque d'une entreprise d'assurance	44
3.1.2	Les différents traités de réassurance et leur impact sur les sinistres qu'ils couvrent	45
3.1.3	La répercussion des traités de réassurance sur les hypothèses d'évolution	48
3.2	Modéliser l'impact de la réassurance	51
3.2.1	Modéliser les différentes fonctions de Réassurance	51
3.2.2	Intégrer l'impact des résidus	56
3.3	Analyse de la pertinence de la modélisation choisie	59
3.3.1	Analyse graphique des hypothèses posées	59
3.3.2	Tester la significativité des fonctions de réassurance avec le test du R^2	62
3.3.3	Écriture du nouveau programme d'optimisation	64
4	Résolution du programme d'optimisation par algorithmes génétiques	66
4.1	Un aperçu des différents algorithmes d'optimisation	66
4.1.1	L'optimisation par la force brute	67
4.1.2	Les algorithmes du gradient	67
4.1.3	Recherche des racines du gradient	68
4.2	Résolution de problèmes multi-objectif avec un algorithme génétique	69
4.2.1	Présentation de l'algorithme génétique NSGA-II	70
4.2.2	L'application du NSGA-II au problème d'optimisation de l'assureur .	71
4.3	Vérifier la fiabilité de l'algorithme à travers plusieurs itérations	76
4.3.1	Générer une base de données des résultats en multipliant les itérations de l'algorithme NSGA-II	76
4.3.2	Mettre en place un test de significativité des résultats	78
4.3.3	Stabilisation des résultats obtenus	80

5	Application sur un portefeuille d'assurance d'AXA	83
5.1	Synthétiser la totalité de la frontière efficiente	83
5.1.1	Synthétiser simplement les informations pour chaque ligne d'assurance	84
5.1.2	L'analyse empirique des résultats	85
5.2	Segmenter la frontière efficiente selon les besoins	88
5.2.1	La division de la frontière efficiente en plusieurs sous-segments	88
5.2.2	La restriction de la frontière efficiente au segment maximisant le Ratio Sharpe	91
6	Limites et perspectives d'évolution du modèle	94
6.1	Analyse des hypothèses posées	95
6.1.1	Les imprécisions des hypothèses de linéarité	95
6.1.2	Un exemple de modélisation plus proche de la réalité	97
6.2	Optimisation par la réassurance plutôt que par la taille des lignes d'assurance	98
6.2.1	Programme d'optimisation en fonction de la réassurance	98
6.2.2	Les potentielles limites d'une telle approche	100
6.3	L'analyse de la robustesse du modèle aux différentes données utilisées	101
6.3.1	Tests de sensibilité traditionnels	102
6.3.2	Modification du test de sensibilité pour éviter le cumul des itérations	103
6.4	Autres considérations	104
6.4.1	Impact des délais d'ajustement et coûts de transaction	105
6.4.2	Limite du coût opérationnel	105
6.4.3	Prise en compte des synergies entre lignes d'assurance	106
	Conclusion	108
	Annexe	110
A	Récapitulatif des variables et notations utilisées	110
B	Méthode des régressions segmentées	111
C	Définition formelle de l'opérateur de croisement NSGA-II	112
D	Définition formelle de l'opérateur de mutation NSGA-II	114
	Bibliographie	115

Table des figures

1.1	Décomposition du portefeuille financier d'AXA (source : [5])	10
1.2	Illustration du bilan d'une société d'assurance	11
1.3	Bilan économique Solvabilité 2 d'une société d'assurance	12
1.4	Calcul du Ratio Économique Combiné	17
1.5	Extrait de la base de données utilisée (données anonymisées)	20
2.1	Représentation de la décision effectuée par l'investisseur financier	24
2.2	représentation de la Frontière efficiente (source : [18])	25
2.3	Représentation de la décision effectuée par l'assureur	27
3.1	Impact des traités de type <i>Excédent de sinistre</i> (source : [15])	48
3.2	Représentation de la relation entre $\epsilon_i(\delta_i)$ et $\epsilon_i(1)$	58
3.3	Sinistres ATT nets rapportés aux ATT bruts	60
3.4	Sinistres ATY nets rapportés aux ATY bruts	61
3.5	Sinistres CAT nets rapportés aux CAT bruts	62
3.6	Statistique de R^2 pour les échantillons d'entraînement et de test de chaque ligne d'assurance	64
4.1	Procédure de l'algorithme NSGA-II (source : [10])	71
4.2	Convergence de l'algorithme NSGA-II	73
4.3	Courbe du facteur d'évolution d'une plus petite ligne après 100 itérations	74
4.4	Ligne d'assurance pour laquelle la convergence est volatile	75
4.5	Illustration de la multiplication des itérations de frontières efficaces	78
4.6	La distribution uniforme des paramètres δ sans fonction objectif	79
4.7	Espérance et dispersion de $\delta_{i,q}^*$ au quantile 0.1 pour une ligne i donnée	82
5.1	Synthèse des Δ par ligne d'assurance sur la frontière efficiente totale	85
5.2	Profil rendement/risque de chaque ligne d'assurance en fonction de la recommandation de l'algorithme	87
5.3	Division de la frontière efficiente en trois segments	90
5.4	Synthèse des Δ par ligne d'assurance sur différents segments de frontière efficiente	90
5.5	Illustration des gains de performance des portefeuilles selon la position sur la frontière efficiente	92
5.6	Intervalles maximisant le ratio Sharpe sur les optima du NSGA-II	93

6.1	Illustration théorique des comparaisons de résultats avec et sans variation des données	104
-----	---	-----

Introduction

La diminution progressive des taux d'intérêts au cours de la dernière décennie s'est soldée par l'apparition d'une situation jusque-là inédite dans la sphère économique mondiale : l'apparition de taux d'intérêt à long-terme durablement proches de 0, voire négatifs. Un des secteurs d'activité affectés directement par ce phénomène est le secteur de l'assurance. La nature même de l'activité d'assurance impose en effet une certaine durée entre la signature d'un contrat et l'indemnisation des sinistres. De la sorte, en plus du revenu technique, qui découle directement de la différence entre les primes perçues et les sinistres payés, il est aussi possible de dégager du revenu financier en faisant fructifier les primes perçues sous forme d'actifs financiers. Or, le contexte des taux faibles diminue mécaniquement la rentabilité de ces actifs ; c'est pourquoi il devient de plus en plus essentiel d'optimiser le revenu technique.

Cette optimisation est cependant d'autant plus difficile que la notion de risque est une notion centrale aux problématiques actuarielles. En effet, l'assurance est une activité qui est soumise en permanence à des objectifs de gestion du risque, que ce soit en raison de stratégies de développement ou de contraintes réglementaires. Dès lors, il est naturel de s'interroger sur la possibilité de répartir l'exposition au risque d'une société d'assurance entre ses différentes lignes d'activité de manière optimale, selon des objectifs de risque et de rendement.

Toutefois, s'il existe à l'heure actuelle un certain nombre de métriques qui permettent d'évaluer la rentabilité ou le risque de l'activité d'assurance de manière individuelle, telles que le *Ratio de Pertes*, le *Ratio combiné* ou le *Capital de Solvabilité Requis*, peu permettent de prendre en compte ces deux aspects en même temps. De même, peu de méthodes permettant l'arbitrage entre ces métriques de risque et rendement ont été développées.

L'objectif de ce mémoire est donc de mettre au point un algorithme qui déterminerait l'ensemble des répartitions optimales d'exposition au risque entre les différentes lignes d'assurance, en prenant en compte les problématiques de risque et de rendement qui caractérisent le secteur de l'assurance. Pour ce faire, la méthode qui a été retenue adapte les travaux de finance quantitative sur l'optimisation du couple rendement/risque d'un portefeuille financier au secteur de l'assurance. *In fine*, l'objectif est donc de produire une frontière efficiente.

Les étapes de cette étude seront donc les suivantes :

Tout d'abord, une analyse préliminaire des différentes problématiques qui caractérisent le secteur de l'assurance sera effectuée. De la sorte, il sera possible de présenter les métriques de rendement et de risque souvent utilisées, ainsi que de déterminer précisément quels facteurs seront à prendre en compte lors du développement de l'algorithme cible.

Ensuite, le programme d'optimisation du couple rendement/risque d'un portefeuille d'assurance pourra être formalisé. Pour ce faire, il sera nécessaire d'identifier quels indicateurs

utiliser pour mesurer ces deux dimensions, et de déterminer comment ceux-ci évoluent en fonction de la répartition de l'exposition au risque entre les lignes du portefeuille.

Ici, une attention significative sera notamment portée sur la réassurance et l'impact qu'elle peut avoir sur ce programme d'optimisation.

Par la suite, l'enjeu sera de mettre au point une approche permettant de résoudre ce programme d'optimisation. Celui-ci, complexe, ne permet en effet pas nécessairement de calculer une solution explicite. C'est pourquoi un algorithme d'optimisation génétique sera utilisée pour approcher l'optimum théorique.

Cette étude se conclura donc par une évaluation de la pertinence du modèle. Pour ce faire, une première application à un portefeuille d'assurance d'AXA permettra de mesurer sa performance sur un exemple concret, avant de rappeler ses limites théoriques et potentiels axes d'évolution.

Chapitre 1

Les enjeux stratégiques de cette étude

Ce chapitre vise à introduire cette étude et explorer les enjeux stratégiques qui lui sont liés. La partie 1.1 introduira donc les problèmes soulevés à l'ensemble de l'activité de l'assurance, en développant notamment l'importance de la notion de risque.

Ces problèmes et défis relevés amèneront donc la problématique principale qui a motivé cette étude, énoncée en partie 1.2, c'est-à-dire la question de l'optimisation d'un portefeuille de lignes d'assurance en prenant en compte le risque.

La partie 1.3 pourra ensuite décrire les méthodes existantes pour tenter de répondre à cette problématique, tout en soulignant pourquoi elles sont insuffisantes. Enfin, une partie 1.4 explorera les données utilisées qui seront utilisées dans le reste de l'étude pour répondre à cette problématique.

1.1 Les défis et enjeux de l'activité d'assurance

Ce qui distingue le secteur de l'assurance d'autres secteurs d'activité est la notion définie comme *l'inversion du cycle de production* [13]. En effet, dans le cas des industries "traditionnelles", la vente d'un bien ou d'un service ne s'effectue qu'après que celui-ci a été produit, ou du moins que ses coûts soient connus. La caractéristique de l'assurance est donc le fait que ce processus soit inversé : le prix, c'est-à-dire le montant de prime perçue, est fixé au moment de la signature du contrat, tandis la prestation, c'est-à-dire l'indemnisation du sinistre, peut avoir lieu plusieurs années ensuite.

La conséquence de cette inversion est que le revenu généré par un contrat d'assurance est nécessairement la source d'un certain niveau de risque. Puisque le montant de l'indemnisation, et même sa survenance sont inconnues au moment où le contrat est conclu, il est difficile de garantir que la prime soit suffisante pour générer un revenu positif. La notion de risque est donc au cœur de l'activité d'assurance.

Afin de comprendre les enjeux principaux de ce secteur d'activité, la section 1.1.1 reviendra sur les différentes manières de générer du revenu en assurance, ainsi que de déterminer les types de risque qu'elles génèrent.

Cela permettra ensuite d'explorer en section 1.1.2 la notion de capital risque, nécessaire pour que les assureurs puissent garantir leur pérennité.

1.1.1 Des sources de revenus génératrices de risque

Le déroulement d'un contrat d'assurance non-vie peut de manière générale se résumer aux cinq étapes suivantes :

1. L'assureur fixe la prime
2. L'assuré et l'assureur signent le contrat
3. L'assuré verse la prime d'assurance à l'assureur
4. L'assuré subit ou non un sinistre
5. Dans le cas où il y a survenance d'un sinistre, l'assureur indemnise l'assuré

Ainsi, quand une prime est perçue, l'assureur prévoit d'en utiliser une certaine proportion pour indemniser des sinistres futurs. Cela se répercute sur le bilan d'une entreprise d'assurance sous la forme d'un montant important sur le poste de provisions, qui correspond à une estimation des charges futures de l'entreprise. Ces primes peuvent en outre être investies sur les marchés financiers afin de les fructifier jusqu'au moment où l'indemnisation doit avoir lieu (c.f. Figure 1.2).

Cela permet donc de dégager un revenu financier en plus du revenu technique.

Les revenus techniques

Les revenus techniques correspondent à la part du revenu qui découle directement de l'activité d'assurance. Ils peuvent donc se définir de la manière suivante :

$$\text{RevenuSouscription} = \text{Prime} - (\text{Frais} + \text{CoutSinistre})$$

La particularité de l'assurance est cependant le fait qu'au moment de la souscription, non seulement le coût, mais aussi la survenance même du sinistre est aléatoire. Le risque associé à un contrat d'assurance correspond donc à la possibilité que le revenu de souscription dévie par rapport à l'attendu, et se caractérise donc par la distribution de la loi de probabilité de ce montant de souscription.

Les revenus financiers

Comme souligné ci-dessus, l'indemnisation du sinistre ne survient qu'après la souscription du contrat et le paiement de la prime. Toutefois, selon le type de garantie, le paiement peut avoir lieu plus ou moins longtemps après le constat du sinistre, et être réparti sur une durée de temps plus ou moins longue. Dans le cas de la responsabilité civile par exemple, certains cas doivent passer devant un tribunal et l'identification du montant à indemniser peut prendre jusqu'à plusieurs décennies. Le revenu financier qu'il est possible de dégager dépend donc de la ligne d'assurance considérée, puisque celui-ci est d'autant plus important que les engagements sont longs.

Cependant, tout comme le revenu technique, ce revenu est générateur de risque. Les actifs financiers sont en effets considérés comme risqués par définition*. En pratique, les entreprises d'assurance tentent de limiter ce risque en concentrant les investissements vers les actifs financiers les moins risqués, comme les obligations, mais un risque, comme le risque

*. La notion théorique de taux sans-risque est toutefois souvent appliquée aux obligations émises par certains états, comme les États-Unis ou l'Allemagne

de contrepartie existe toujours (c.f. Figure 1.1).

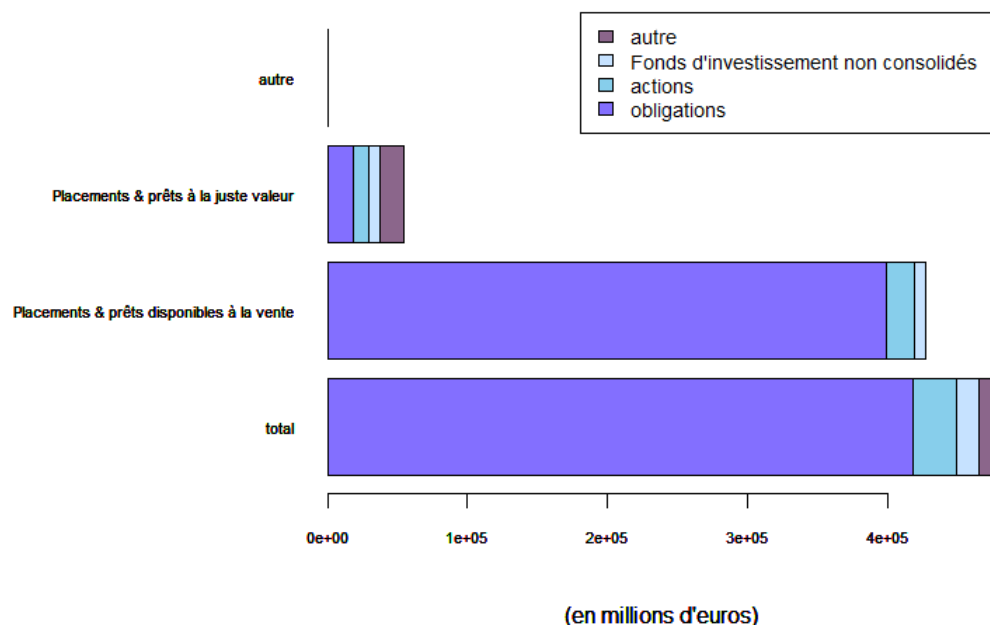


FIGURE 1.1 – Décomposition du portefeuille financier d'AXA (source : [5])

La décomposition des risques entre les revenus techniques et les revenus financiers

Les revenus générés par l'activité d'assurance impliquent donc un certain nombre de risques. Ceux-ci peuvent être décomposés en plusieurs facteurs de risque de la manière suivante :

- - Les risques de marché :
Risques de fluctuations des taux de rendement ou du cours des actions sur les marchés financiers
- - Risque de crédit :
Risques de défaut ou de taux de non-recouvrement des contreparties
- - Risques Opérationnels :
Risques liés à la conduite de l'activité d'assurance (risque de fraude, de pannes, de destruction des actifs immobilier, etc.)
- - Risques-vie :
Risques liés aux activités d'assurance vie (risque de longévité, risques de mortalité, etc.)
- -Risques non-vie :

Risques liés aux activités d'assurance non-vie. Ils peuvent se décomposer de la manière suivante :

1. risques de souscription liés aux catastrophes (CAT)
2. risques de souscription non liés aux catastrophes (non-CAT)
3. risques de réserve

Les risques de souscription non-CAT peuvent également se diviser entre sinistre attritionnels et sinistres atypiques.

La matérialisation de ces risques peut en outre être aggravée par l'existence de corrélations, que ce soit des corrélations entre lignes d'assurance ou entre types de risque.

Actifs	Passifs
Primes / autres créances nées d'opérations d'assurance	Capitaux propres
Créances nées d'opérations de réassurance	Passifs financiers (emprunts, dettes financières, etc.)
Placements / revenus financiers	Provisions techniques liées aux activités d'assurance
Actifs corporels / incorporels	Dettes / passifs liés à la réassurance
Autres actifs	Autres passifs

FIGURE 1.2 – Illustration du bilan d'une société d'assurance

Ainsi, si la maximisation du rendement est bien évidemment un des principaux objectifs des entreprises d'assurance, cette optimisation ne peut être atteinte sans la considération du risque associé.

1.1.2 La notion de Capital risque

Afin de se prévenir contre les risques qui découlent de ses activités, une entreprise d'assurance est donc tenue de détenir une quantité de capital suffisante pour couvrir ses engagements en cas de survenance de ces risques.

Pour les entreprises régulées dans l'Union Européenne, une limite inférieure à cette quantité de capital est notamment définie par la directive européenne de 2009 connue sous le nom

de **Solvabilité-II** [1].

Cette directive repose sur trois piliers : les exigences quantitatives, les activités de contrôle et la transparence de l'information. En particulier, le pilier d'exigences quantitatives définit le Capital de Solvabilité requis (aussi appelé (*Solvency Capital Requirement*), ou SCR), c'est-à-dire la quantité minimale de fonds propres nécessaire pour qu'une entreprise d'assurance puisse couvrir ses engagements 99.5% du temps. Autrement dit, l'objectif est de garantir qu'il est possible de faire face à un choc bicentenaire.

De manière plus formelle, en notant X la variable aléatoire correspondant la différence entre la valeur de l'actif et la valeur des engagements au passif, le SCR est défini de la manière suivante :

$$SCR = VaR_{0.005}(X)$$

avec :

$$\mathbb{P}(X \geq VaR_{0.005}(X)) = 0.005$$

Il sert donc à maîtriser la probabilité de ruine d'une entreprise d'assurance. En effet, ce montant de capital garantit qu'une entreprise d'assurance ne fera faillite en moyenne qu'au plus "une fois tous les 200 ans".

Actifs	Passifs	
Primes / autres créances nées d'opérations d'assurance	Capital excédentaire	Fonds propres
Créances nées d'opérations de réassurance	SCR	
Placements / revenus financiers	Marge de risque (risk margin)	Provisions techniques
Actifs corporels / incorporels	Valeur de marché des engagements	
Autres actifs		

FIGURE 1.3 – Bilan économique Solvabilité 2 d'une société d'assurance

Ce capital prudentiel n'est toutefois pas sans conséquences pour l'entreprise qui le détient. Deux sources de capital principales peuvent en effet être distinguées : le résultat des années précédentes et l'apport de capital des actionnaires. Dans la mesure où l'actionnariat exige un rendement sur le capital investi, y faire appel pour couvrir les besoins en capital peut être coûteux.

Ainsi, une ligne d'assurance risquée sera non seulement plus gourmande en capital, mais nécessitera aussi une rémunération des actionnaires plus grande ; c'est pourquoi il est question de **coût du capital**.

L'optimisation d'un portefeuille d'assurance peut donc se traduire par la recherche du rendement maximal sous contrainte de risque. Or, dans le contexte actuel d'un faible revenu financier engendré par des taux d'intérêt bas, il devient d'autant plus important de se concentrer sur le revenu technique et le risque qu'il génère.

1.2 Problématique : la recherche d'une méthode pour optimiser le profil rendement/risque d'un portefeuille de lignes d'assurance

Le point de départ de cette étude est donc le constat suivant : étant donné le contexte économique actuel, il est d'une grande importance de maximiser le revenu technique, tout en maîtrisant le risque généré. Pour ce faire, un des leviers d'action des entreprises d'assurance est la stratégie de souscription et appétence au risque associée. Le choix des lignes d'activités à développer est donc un enjeu majeur.

L'objectif de cette étude est donc d'aboutir à une méthode qui permette de déterminer la configuration optimale d'évolution de l'exposition au risque de chaque ligne d'activité. En somme, le but est d'être capable d'affirmer quelles lignes croître et quelles lignes réduire afin d'optimiser le profil de rendement et de risque de l'entreprise d'assurance.

La problématique est donc la suivante :

Comment faire évoluer le portefeuille d'assurance de manière à maximiser le rendement et minimiser le risque, et existe-t-il une combinaison d'évolution des lignes d'assurance qui soit strictement optimale ?

Étant donné la place centrale qu'occupe l'arbitrage entre le risque et le rendement au sein de l'activité d'assurance, un certain nombre de méthodes offrent déjà un élément de réponse à cette problématique. Toutefois, les indicateurs existants sont à l'heure actuelle insuffisants dès lors que l'objectif est de concilier le risque et le rendement conjointement de manière algorithmique.

1.3 Les méthodes existantes mais insuffisantes pour répondre à cette attente

Si l'approche développée lors de cette étude cherche à aborder cette problématique sous un nouvel angle, un certain nombre de méthodes sont déjà en place pour déterminer comment

arbitrer entre le risque et le rendement d'un portefeuille d'assurance. En particulier, plusieurs indicateurs permettent de mesurer un portefeuille selon l'une de ces deux dimensions.

Cependant, ces indicateurs sont en général insuffisants pour répondre à la problématique telle qu'elle est posée ici. En effet, les indicateurs de rendement, explorés en section 1.3.1, sont pour la plupart incapables d'incorporer une dimension de risque, tandis que les indicateurs de risque explorés en section 1.3.2 négligent la performance technique. Or, l'objectif est ici de mettre en place une méthode qui permettrait d'optimiser ces deux facteurs conjointement. Il existe certes des indicateurs qui tentent de concilier risque et rendement, tels le *Ratio Économique Combiné* défini en section 1.3.3. Toutefois, la section 1.3.4 montrera que ceux-ci restent inaptes à la résolution du problème d'optimisation.

1.3.1 Les indicateurs concentrés uniquement sur la performance technique

Le Ratio de Pertes

Dès qu'il est question d'évaluer la performance technique d'un portefeuille d'assurance non-vie, un des indicateurs les plus utilisés est le *Ratio de Pertes*. Cette mesure, relativement simple, se définit pour une ligne comme pour un portefeuille par le rapport entre la charge des sinistres et les primes perçues [22] :

$$RatioPerte = \frac{Sinistres}{Primes} \quad (1.1)$$

L'avantage principal de cet indicateur est sa capacité à faciliter les comparaisons entre différentes lignes d'assurance, malgré un potentiel effet volume. En effet, un ratio à valeur inférieure à 1 dénote une ligne ou un portefeuille dont la profitabilité actuarielle pure est positive, tandis qu'un ratio supérieur à 1 suppose que les primes perçues ne suffisent pas pour compenser les sinistres.

Toutefois, évaluer cet indicateur pour une décision en cours implique de ne considérer que les primes et les sinistres sous l'angle de la *meilleure estimation*. Autrement dit, ce ratio ignore totalement non seulement les écarts à la moyenne possible des sinistres, mais aussi ceux des primes. La dimension de risque est donc totalement ignorée.

En outre, cet indicateur ne permet pas de mesurer la performance technique réelle d'une ligne, puisqu'un certain nombre de facteurs qui ont un impact important sur le rendement sont ignorés.

Le Ratio Combiné

Le *Ratio combiné* tente de pallier certaines faiblesses du Ratio de Pertes en mesurant de manière plus précise la rentabilité réelle d'une ligne d'assurance [22]. Sa définition est donc la suivante :

1.3. Les méthodes existantes mais insuffisantes pour répondre à cette attente

$$RatioCombiné = \frac{Sinistres + Frais - RésultatRéassurance}{Primes} \quad (1.2)$$

Si cet indicateur donne une vision plus complète que le Ratio de Perte de la performance réelle de la ligne, il ne permet pas de s'approcher d'une réponse à la problématique posée ici. En effet, tout comme le Ratio de Perte, la dimension de risque est totalement ignorée ici.

1.3.2 Les mesures de risque usuelles

Tout comme pour le rendement, il est possible de déterminer des indicateurs capables de capter le risque d'un portefeuille ou d'une ligne d'assurance dès lors que la loi de probabilité du résultat est connue.

La valeur Exposée au risque

La Valeur à Risque (aussi appelée *Value-at-risk* ou *VaR*), cherche à déterminer le niveau de pertes à la probabilité α [14]. En notant X la variable aléatoire correspondant au revenu technique généré par une ligne ou un portefeuille d'assurance, la VaR est définie telle que :

$$\mathbb{P}(X \geq VaR_\alpha(X)) = \alpha$$

Cette mesure de risque a pour spécificité d'être relativement intuitive. Si le paramètre est par exemple fixé à $\alpha = 0.005$, cela signifie que la probabilité d'obtenir un résultat inférieur à $VaR_\alpha(X)$ est de 0.5 %. Autrement dit, sous l'hypothèse où l'entreprise conserverait le même profil au cours du temps, $VaR_\alpha(X)$ désigne le niveau de pertes auquel l'entreprise peut être exposée une fois toutes les 200 ans en moyenne.

Cet indicateur, souvent utilisé par les entreprises ou les instances réglementaires pour maîtriser le risque peut toutefois présenter des faiblesses. La première est le fait que, si cet indicateur donne un seuil de pertes qui ne devrait être dépassé qu'une fois tous les 200 ans, aucune information n'est donnée sur le montant de ce dépassement. Autrement dit, la *VaR* ne fait pas la différence entre les cas où ce seuil n'est atteint que faiblement et les cas où le dépassement est très significatif.

En outre, cet indicateur ne permet cependant pas de résoudre la problématique posée ici de manière convaincante. À l'inverse du *Ratio de Pertes* ou du *Ratio combiné*, la *Valeur exposée au risque* ne donne qu'une information sur la volatilité du rendement, et non sur sa valeur moyenne. Il n'est donc pas possible d'optimiser le portefeuille en raisonnant exclusivement sur cette mesure.

Le déficit Prévu

L'objectif du *déficit prévu* (*Expected Shortfall*) est de pallier la première faiblesse de la *Valeur exposée au risque* en calculant la moyenne des pertes dans le cas où la *VaR* est

atteinte [2]. En notant $ES_\alpha(X)$ le *déficit prévu* de niveau de risque α , sa définition est donc la suivante :

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_a(X) da \quad (1.3)$$

Outre le fait que cette mesure est plus difficile à évaluer, car elle nécessite une connaissance plus précise de la queue de distribution du revenu, le *déficit prévu* ne permet tout de même pas de répondre à la problématique posée.

En effet, à l'instar de la VaR , cet indicateur ne donne pas d'information sur le rendement moyen du portefeuille.

1.3.3 Le Ratio Économique Combiné : Un premier indicateur qui tente de concilier risque et rendement

Contrairement aux indicateurs précédents, le *Ratio Économique Combiné* permet d'intégrer à la fois une dimension de risque et une dimension de rendement. Son objectif est en effet de prendre en compte la totalité des facteurs qui ont un impact sur la performance d'une ligne ou d'un portefeuille. Pour ce faire, cette mesure part du *Ratio Économique* et ajoute de manière successive :

- Un ajustement pour prendre en compte la valeur-temps
- L'impact de l'imposition
- L'impact du coût du risque

Il est tout d'abord nécessaire de prendre en compte l'impact de la valeur-temps, car il existe parfois un délai significatif entre le moment où l'assureur perçoit la prime et le moment où le sinistre est indemnisé. Il est donc possible de fructifier les primes.

Or, ce délai, et donc le revenu financier, est d'autant plus important que la ligne d'activité considérée est longue. Il serait donc impossible d'omettre l'effet-temps du calcul sans biaiser les résultats vers les lignes courtes. L'impact de la valeur-temps pour la ligne d'activité i se calcule donc à travers la différence entre les résultats non actualisés et les résultats actualisés, qui prennent en compte la cadence de paiement :

$$VT_i = \frac{\text{Résultat Non Actualisé}_i - \text{Résultat Actualisé}_i}{\text{Primes}_i} \quad (1.4)$$

Afin de prendre en compte le risque, le *Ratio Économique Combiné* déprécie le *Ratio Combiné* par un facteur appelé *coût du risque*. En effet, les contraintes liées au capital risque sont d'autant plus lourdes que la ligne est risquée. Or, dès lors qu'il est nécessaire de faire appel à l'actionnariat pour subvenir aux besoins en capital de l'entreprise, la rémunération attendue des actionnaires fait peser des contraintes supplémentaires sur l'entreprise. C'est pourquoi le *coût du risque* qui s'ajoute au *Ratio Économique Combiné* est calculé de la manière suivante :

$$\text{CoutRisque}_i = \frac{SCR_i \times \text{objectif en capital} \times \text{prime de risque}}{\text{Primes}_i} \quad (1.5)$$

1.3. Les méthodes existantes mais insuffisantes pour répondre à cette attente

Ici, le SCR ajusté de l'objectif de l'entreprise en termes de ratio de couverture permet de rendre compte des besoins en capital. La prime de risque, elle, caractérise la rémunération du capital, c'est-à-dire la rémunération attendue par les actionnaires.

Ainsi, cet indicateur est capable de donner une vision plus précise et plus complète de la performance réelle d'une ligne d'assurance ou d'un portefeuille.

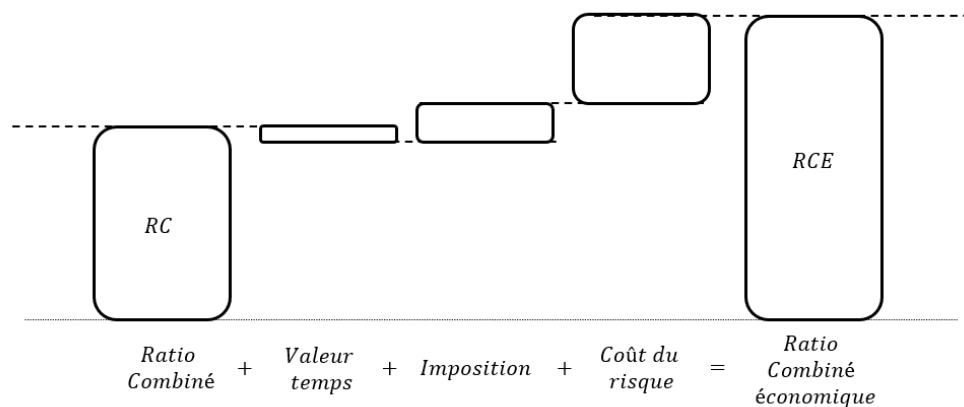


FIGURE 1.4 – Calcul du Ratio Économique Combiné

1.3.4 Le Ratio Économique combiné comme indicateur cependant statique et inapte à l'optimisation

Il apparaît donc que le *Ratio Économique Combiné* est un bon indicateur s'il est question d'évaluer la performance réelle de chaque ligne d'assurance en prenant en compte à la fois le risque et le rendement. C'est une mesure qui est notamment particulièrement adaptée dès qu'il est question de comparer différentes lignes entre elles. Toutefois, cet indicateur ne permet pas de répondre à la problématique posée dans cette étude.

L'objectif est en effet ici de déterminer quelles combinaisons d'évolution des lignes d'assurance qui composent un portefeuille peuvent être considérées comme optimales. Or, de par sa structure, cet indicateur est incapable de répondre aux questions suivantes :

- Comment le risque et le rendement du portefeuille évoluent-ils en fonction de la taille de la ligne d'assurance ?
- À quel point l'apport des différentes lignes d'assurances en termes de diversification est-il important ?
- Existe-t-il une évolution de la taille de la ligne d'assurance qui permette d'augmenter le rendement sans augmenter le risque, ou inversement ?

L'évaluation des différents indicateurs souvent utilisés pour évaluer le risque, le rendement ou la performance d'une ligne ou d'un portefeuille révèle donc que ces indicateurs ne peuvent en offrir une vision statique, et sont donc insuffisants dans le cadre de cette étude. Ainsi, il apparaît nécessaire de se tourner vers d'autres méthodes que les indicateurs usuels pour répondre à la problématique de cette étude.

1.4 Les caractéristiques de la base de données d'assurance utilisée pendant cette étude

Lors de cette étude, une approche permettant de répondre à la problématique posée a pu être appliquée à un portefeuille de lignes d'assurance de grands risques du groupe AXA. Par conséquent, il est nécessaire de rappeler les caractéristiques du portefeuille étudié et de la base de données associée afin de comprendre leur impact sur la modélisation.

Le fait que le portefeuille de lignes étudié ne contienne que des lignes de grands risques n'est par exemple non sans impact sur les analyses effectuées. En effet, de manière générale, le lien entre la prime perçue et la taille d'une ligne d'assurance pourrait être étudié à l'aide d'une étude d'élasticité. Cela se comprend par le fait que, si la demande est élastique, et que le marché est donc sensible à une diminution de tarifs, alors une augmentation de la taille d'une ligne pourrait se faire aisément aux prix d'une prime perçue plus faible. Toutefois, les données sur l'élasticité de la demande aux primes sont rarement disponibles et fiables dans le cas des grands risques. Cette analyse dut par conséquent être écartée lors de cette étude.

Cet exemple illustre donc la nécessité de se pencher sur les données disponibles pour ce portefeuille avant d'explorer les modèles développés.

La base de données utilisée est en effet d'autant plus particulière qu'il s'agit des sorties du modèle interne AXA. Ainsi, la section 1.4.1 montrera que les données utilisées sont en réalité les simulations d'un certain nombre de variables calibrées au préalable.

En outre, la section 1.4.2 explorera le fait les données utilisées ont pour caractéristiques d'être agrégées par lignes d'assurance de grands risques, ce qui réduit mécaniquement le niveau de granularité auquel il fut possible de travailler lors de cette étude.

1.4.1 Une base de données constituée de simulations de variables aléatoires

De manière générale, la base de données utilisée lors de cette étude est composé de diverses variables annuelles pour chacune des $n = 55$ lignes d'assurance qui constituent le portefeuille dont l'optimisation est recherchée. Pour chaque ligne, les lois de probabilité de variables telles les frais, les primes perçues, les sinistres, ou encore le résultat sont calibrées dans le cadre du modèle interne utilisé pour estimer le SCR.

C'est donc les simulations de ces différentes variables qui composent la base de données utilisée lors de cette étude. Ainsi, pour chaque ligne d'assurance considérée, les lois de probabilités calibrées au sein du modèle interne apparaissent dans la base de données sous la forme de $S = 50000$ réalisations pour chacune des variables.

Autrement dit, pour une ligne d'assurance $i \in [1, n]$, la base de données comprend les simulations suivantes :

- $Frais_i = (Frais_{i,1}, \dots, Frais_{i,S})$
- $Primes_i = (Primes_{i,1}, \dots, Primes_{i,S})$
- $Sinistres_i = (Sinistres_{i,1}, \dots, Sinistres_{i,S})$
- $Resultat_i = (Resultat_{i,1}, \dots, Resultat_{i,S})$

Ces variables sont en outre décomposées entre réalisations nettes de réassurance et réalisations brutes de réassurance. Pour toute réalisation $s \in [1, S]$, les variables suivantes sont donc observées :

- $Frais_{i,s} = (Frais_{i,s}^{brut}, Frais_{i,s}^{net})$
- $Primes_{i,s} = (Primes_{i,s}^{brut}, Primes_{i,s}^{net})$
- $Sinistres_{i,s} = (Sinistres_{i,s}^{brut}, Sinistres_{i,s}^{net})$
- $Resultat_{i,s} = (Resultat_{i,s}^{brut}, Resultat_{i,s}^{net})$

Une caractéristique supplémentaire de cette base de données est le fait que les sinistres sont soumis à une décomposition supplémentaire, selon le type de risque auxquels ils correspondent. Ainsi, une réalisation de la variable des sinistres annuels correspond à la somme de la réalisation des sinistres atypiques, des sinistres attritionnels, et des sinistres liés aux catastrophes :

$$\begin{aligned} Sinistres_{i,s}^{brut} &= Sinistres_{i,s}^{ATT,brut} + Sinistres_{i,s}^{ATY,brut} + Sinistres_{i,s}^{CAT,brut} \\ Sinistres_{i,s}^{net} &= Sinistres_{i,s}^{ATT,net} + Sinistres_{i,s}^{ATY,net} + Sinistres_{i,s}^{CAT,net} \end{aligned}$$

Enfin, il convient de noter que ces variables ne sont pas simulées de manière indépendante. Pour une simulation s , l'équation suivante est par exemple vérifiée :

$$Resultat_{i,s} = Primes_{i,s} - (Charges_{i,s} + Sinistres_{i,s}) \quad (1.6)$$

Cette non-indépendance trouvera surtout son importance lors des calculs de risque. Les variables de sinistres sont en effet simulées ou agrégées de sorte à respecter des matrices de corrélations données :

$$\begin{aligned} \Omega_{lignes} &= \begin{bmatrix} 1 & \rho_{1,2} & \dots \\ \vdots & \ddots & \\ \rho_{1,n} & & 1 \end{bmatrix} \\ \Omega_{risques} &= \begin{bmatrix} 1 & \rho_{ATT,ATY} & \rho_{ATT,CAT} \\ \rho_{ATY,ATT} & 1 & \rho_{ATY,CAT} \\ \rho_{CAT,ATT} & \rho_{CAT,ATY} & 1 \end{bmatrix} \end{aligned}$$

La base de données utilisée lors de cette étude ne correspond donc pas aux observations réelles de diverses variables, mais aux simulations de ces variables calibrées sous plusieurs hypothèses. Ainsi, une estimation des lois de probabilités de chaque variable est observée. L'objectif de cette étude n'est pas de remettre en question ces hypothèses, mais de les utiliser afin de mettre en place un modèle d'optimisation. La possibilité d'imprécisions au sein de ces estimations pourra être étudiée, mais le développement du modèle sera en général effectué sous l'hypothèse que les distributions de probabilité sont connues.

Numéro de simulation	Résultat brut actualisé ligne 1	Résultat net actualisé ligne 1	Primes brutes perçues actualisées ligne 1	Sinistres brut ATT & ATY actualisé ligne 1	Sinistres net ATT & ATY actualisé ligne 1	Sinistres brut CAT actualisé ligne 1	Sinistres net CAT actualisé ligne 1	Frais bruts actualisés	Frais nets actualisés
1	940098	1101411,58	3390324	715001	657800,92	948359	891457,46	786866	739654,04
2	2742536	2813611,99	4029797	121575	113064,75	428144	402455,36	737542	700664,9
3	3239231	3322844,55	4813032	397399	369581,07	663813	643898,61	512589	476707,77
4	6291964	6426384,88	8627861	399964	363967,24	995091	915483,72	940842	922025,16
5	2799370	2911275,7	4684281	450647	423608,18	916544	852385,92	517720	497011,2
6	1403542	1488265,78	2767130	409373	409373	838658	763178,78	115557	106312,44
7	5806344	5842176,72	7336406	439829	439829	493021	493021	597212	561379,28
8	6659840	6710904,17	8030363	918795	900419,1	114999	109249,05	336729	309790,68
9	6930272	6967951,73	9028009	376265	372502,35	835118	810064,46	886354	877490,46
10	1747711	1805030,59	3523534	989034	939582,3	359084	355493,16	427705	423427,95
11	561551	650937,79	2772922	499211	494218,89	835858	768989,36	876302	858775,96

FIGURE 1.5 – Extrait de la base de données utilisée (données anonymisées)

1.4.2 Une base de données agrégée par ligne d'assurance

Une caractéristique supplémentaire et importante de la base de données utilisée est en outre le fait que chaque observation est agrégée au niveau de la ligne d'assurance.

Pour déterminer la loi de probabilité de la sinistralité annuelle brute de la ligne $i \in [1, n]$ ($Sinistres_i^{brut}$) le modèle interne simule en réalité plusieurs sinistres. Leur nombre peut être simulé à partir d'hypothèses de fréquence, tandis que leur montant peut être simulé à partir d'hypothèses de sévérité.

Ainsi, pour chaque ligne $i \in [1, n]$ et pour chaque simulation $s \in [1, S]$, en notant $C_{i,s}$ la fréquence de sinistres, $K_{i,s}$ le nombre de contrats associés à cette ligne et $(Sinistres_{i,s}^{brut})_c$ le montant du c -ième sinistre simulé :

$$Sinistres_{i,s}^{brut} = \sum_{c=1}^{C_{i,s}} (Sinistres_{i,s}^{brut})_c$$

$$Primes_{i,s}^{brut} = \sum_{k=1}^{K_{i,s}} (Primes_{i,s}^{brut})_k$$

$$Frais_{i,s}^{brut} = \sum_{k=1}^{K_{i,s}} (Frais_{i,s}^{brut})_k$$

Ainsi, la réalisation agrégée $Sinistres_{i,s}^{brut}$ est observée, contrairement à la réalisation du sinistre individuel $(Sinistres_{i,c}^{brut})_s$ qui ne l'est pas.

Cette agrégation trouvera notamment son importance lorsqu'il sera question de prendre en compte l'impact de la réassurance. En effet, bien que les traités d'assurance soient souvent définis pour toute une ligne, l'application de la réassurance se fait sinistre par sinistre. Ainsi, en notant $f_{i,c}^{(s)}$ l'impact de l'ensemble des traités de réassurance qui influent sur le c -ième sinistre (avec $c \in [1, C_{i,s}]$), la variable des sinistres nets de la ligne i s'obtient de la façon suivante :

$$\begin{aligned}
 Sinistres_{i,s}^{net} &= \sum_{c=1}^{C_{i,s}} f_{i,c}^{(s)}((Sinistres_{i,s}^{brut})_c) \\
 &= f_{i,s}(\sum_{c=1}^{C_{i,s}} (Sinistres_{i,s}^{brut})_c) \\
 &= f_i(Sinistres_{i,s}^{brut}) + \epsilon_{i,s}
 \end{aligned}$$

Ainsi, puisque les structures précises des traités $f_{i,c}$ et des sinistres simulés $Sinistres_{i,c}^{brut}$ sont inobservées, les fonctions de réassurance exactes ne peuvent être déterminées.

La base de données utilisée pose donc des contraintes particulières lors du développement du modèle d'optimisation. En effet, si sa nature permet de déterminer de manière presque automatique les lois de probabilités des diverses variables importantes à la réalisation du modèle, le fait qu'elle soit agrégée signifie que certaines informations ne peuvent pas être connues précisément.

En outre, il convient de noter que les lois de probabilités ne sont connues que grâce à des simulations réalisées sous certaines hypothèses. Certes, une discussion autour de la pertinence de ces hypothèses dépasse largement le cadre de cette étude. Toutefois, il sera nécessaire de prendre en compte l'incertitude que cela engendre lors de l'étude des résultats du modèle qui sera développé dans les parties suivantes.

Ainsi, afin de répondre à la problématique posée, deux modèles similaires fondés sur les travaux de mathématiques financières d'optimisation du portefeuille ont été développés et testés sur le portefeuille d'assurance d'AXA.

L'objectif du premier, développé en partie 2, est d'optimiser le portefeuille brut de réassurance. Autrement dit, l'objectif stratégique est d'optimiser dans un premier temps le portefeuille de base, afin de décider de la structure de réassurance souhaitée dans un second temps.

L'objectif du second, développé en partie 3, est au contraire d'optimiser le portefeuille net de réassurance. Ainsi, la question est de déterminer, avec une structure de réassurance donnée, quelle combinaison d'évolution des lignes d'activité permettrait d'aboutir au portefeuille d'assurance optimal.

Chapitre 2

L'écriture du programme d'optimisation brut de réassurance

L'enjeu de cette étude est donc le développement du modèle qui soit capable de déterminer le portefeuille d'assurance optimal d'une entreprise d'assurance avant réassurance. Autrement dit, l'objectif est de déterminer l'allocation optimale de l'exposition au risque d'une entreprise d'assurance entre ses différentes lignes d'activité. Ce n'est donc qu'une fois cette optimum trouvé que l'entreprise cherchera à optimiser sa structure de réassurance, et c'est donc un aspect qui est volontairement écarté dans cette partie.

Or, un des problèmes usuels en mathématiques financières est le problème d'optimisation d'un portefeuille financier. La question est la suivante : si un investisseur souhaite investir sur un marché financier, existe-t-il une d'allocation de richesse entre les actifs disponibles sur le marché qui soit optimale ?

Dans le cadre de la finance quantitative, il est en général supposé qu'un investisseur est rationnel et averse au risque. Ainsi, son objectif est de maximiser le rendement espéré du portefeuille tout en en minimisant le risque. Le problème d'optimisation qui en découle est donc similaire à celui que cette étude cherche à résoudre, et c'est pourquoi les travaux d'analyse financière qui répondent à cette question ont été utilisés comme socle à partir duquel l'algorithme recherché a pu être développé.

Ainsi, cette partie, centrée sur le développement du modèle théorique développé lors de cette étude commencera par la présentation de problématiques de mathématiques financières. En particulier, la section 2.1 reviendra sur la frontière efficiente, notion que Markowitz utilise pour la première fois en 1952 pour résoudre un problème financier d'optimisation à double-objectif.

Il existe toutefois des différences significatives entre les secteurs d'assurance et de finance, ce qui empêche l'application immédiate de ces travaux à un portefeuille d'assurance. En particulier, il conviendra de revenir sur le calcul du rendement et de déterminer une mesure de risque adéquate, ce qui sera effectué dans les sections 2.2 et 2.3 respectivement.

2.1 La modélisation de l'optimisation d'un portefeuille d'actifs : la frontière efficiente de Markowitz

L'analyse que Markowitz offre dans sa publication de 1952 part d'un constat : l'idée selon laquelle un investisseur rationnel n'optimiserait son portefeuille d'actifs qu'en fonction

de son rendement espéré est erronée [18]. En effet, si le risque associé à un actif peut être rattaché au rendement par la notion de la prime de risque, il est cependant impossible de prendre en compte la diversification entre les actifs d'un portefeuille avec la seule notion de rendement espéré.

Dans le but de démontrer cette hypothèse et d'offrir une résolution théorique au problème d'optimisation de l'investisseur, les travaux de Markowitz commencent par une définition formelle du problème d'optimisation, explorée en section 2.1.1, avant de résoudre ce problème en utilisant la notion de Frontière Efficiente définie en section 2.1.2.

2.1.1 Le problème à la base des mathématiques financières : l'optimisation de l'arbitrage entre rendement et risque

Par souci de simplicité, le modèle présenté ici sera celui d'un modèle à deux périodes, notées t_0 et t_1 , et le taux d'actualisation sera fixé à 0. Le marché à partir duquel l'investisseur construit son portefeuille dispose donc des caractéristiques suivantes :

1. Il existe N actifs financiers disponibles sur le marché
2. La valeur de l'actif $i \in (1, \dots, N)$ est notée $V_{i,t \in (t_0, t_1)}$
 - - La valeur initiale V_{i,t_0} est supposée connue
 - - La valeur future V_{i,t_1} est aléatoire
3. Le rendement espéré de l'actif i sur la période est noté $r_i = \mathbf{E}\left[\frac{V_{i,t_1}}{V_{i,t_0}}\right]$
4. La volatilité du rendement de l'actif i est notée $\sigma_i = \sqrt{\text{Var}\left(\frac{V_{i,t_1}}{V_{i,t_0}}\right)}$
5. Soit $i, j \in (1, \dots, N)$ deux actifs financiers. On note alors $\rho_{i,j}$ leur coefficient de corrélation, et $\Omega = (\sigma_{i,j})_{1 \leq i, j \leq N}$ la matrice de covariance correspondante.

Une des hypothèses fondamentales de ce problème d'optimisation est le fait que l'investisseur est averse au risque. Autrement dit, l'objectif de l'investisseur est de maximiser le rendement de son portefeuille tout en minimisant le risque qu'il génère.

Afin d'aboutir à cet objectif, l'investisseur dispose d'une richesse initiale V_0 qu'il est libre de répartir entre chaque actif présent sur le marché. En notant $(\omega_1, \dots, \omega_N)$ la proportion de V_0 allouée à chaque actif, la décision de l'investisseur est donc soumise aux contraintes suivantes :

- (c1) : $\sum_{i=1}^N \omega_i = 1$
- (c2) : $\omega_i \geq 0$ pour tout $i \in (1, \dots, N)$

Il convient ici de noter que la contrainte (c2) peut être levée si les positions courtes sont autorisées, mais cela ne change pas les conclusions dans ce cadre.

En t_1 , la valeur du portefeuille associé à la décision $\omega = (\omega_1, \dots, \omega_N)$ est donc de $V_1 = V_0 \sum_{i=1}^N \omega_i \frac{V_{i,t_1}}{V_{i,t_0}}$.

Le rendement espéré du portefeuille prend ainsi la valeur suivante :

$$r^{(\omega)} = \sum_{i=1}^N \omega_i r_i \tag{2.1}$$

De même, le risque associé à cette décision, noté $\sigma^{(\omega)}$ est tel que :

$$\sigma^{(\omega)^2} = Var\left(\frac{V_0}{V_1}\right) = \sum_{i=1}^N \sum_{j=1}^N \omega_i \omega_j \sigma_{i,j}$$

Le programme de l'investisseur prend ainsi la forme suivante :

$$\begin{aligned} & \min_{\omega_{i(1 \leq i \leq N)}} (-r^{(p)}, \sigma^{(p)}) \\ & \text{s.c.} \sum_{i=1}^N \omega_i^{(p)} = 1 \\ & \forall i \in (1, \dots, N), \omega_i^{(p)} \geq 0 \end{aligned} \tag{2.2}$$

L'enjeu de ce problème d'optimisation est donc dorénavant de déterminer la forme de l'ensemble des solutions.

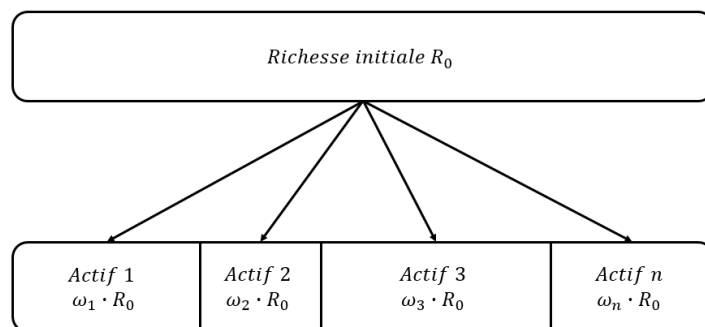


FIGURE 2.1 – Représentation de la décision effectuée par l'investisseur financier

2.1.2 Le résultat du programme d'optimisation : la frontière efficiente

Afin de résoudre le programme d'optimisation de l'investisseur, il convient d'introduire deux notions supplémentaires. Tout d'abord vient la notion d'efficience du marché, qui suppose que r_i , σ_i et Ω sont supposés connus par l'investisseur.

Ensuite vient la notion de portefeuilles admissibles. Les contraintes placées sur les poids ω_i donnent en effet lieu à un ensemble infini, mais délimité de stratégies d'allocations d'actifs. On appelle donc **Portefeuille admissible** tout portefeuille P dont les poids respectifs $\omega^{(p)}$ respectent les conditions $\sum_{i=1}^N \omega_i^{(p)} = 1$ et $\omega_i^{(p)} \geq 0$. L'ensemble de ces portefeuilles admissibles sera noté A^P , et l'ensemble des rendements et volatilités associés à ces portefeuilles respectivement $r^{(A^P)}$ et $\sigma^{(A^P)}$

Dans un tel cadre, s'il n'était question que d'optimiser le rendement, il suffirait de prendre une combinaison linéaire des actifs dont le rendement est maximal. En notant (a_1, \dots, a_A) l'ensemble des actifs tels que $\forall k \in (1, \dots, A), r_{a_k} = \max(r_1, \dots, r_N) = r_a$, un portefeuille à l'optimum serait donc tout portefeuille admissible qui soit de la forme suivante :

$$r^{(p_0)} = \sum_{k=1}^A \omega_{a_k} r_{a_k} = r_a \tag{2.3}$$

Or, un tel portefeuille est en réalité sous-optimal en général, car les notions de risque et de diversification sont dans ce cas ignorées.

Résoudre un tel programme d'optimisation multi-objectif revient donc finalement à déterminer l'ensemble des *optima* de Pareto, c'est-à-dire l'ensemble des allocations pour lesquelles il est impossible d'optimiser un critère davantage sans en pénaliser un autre. Cela signifie que la solution programme ci-dessus peut en réalité se réécrire comme les solutions de l'ensemble des problèmes d'optimisation à un objectif suivant :

$$\begin{aligned} & \max_{\omega_i (1 \leq i \leq N)} (r^{(p)}) \\ & s.c. \sum_{i=1}^N \omega_i^{(p)} = 1 \\ & \forall i \in (1, \dots, N), \omega_i^{(p)} \geq 0 \\ & \sigma^{(p)} \leq \sigma_0 \text{ avec } \sigma_0 \in [\min(\sigma^{(A_p)}), \max(\sigma^{(A_p)})] \end{aligned} \tag{2.4}$$

Ainsi, en notant $\sigma_0 \in [\min(\sigma^{(A_p)}), \max(\sigma^{(A_p)})]$ une volatilité atteignable et r_0 la solution au programme d'optimisation associé (2.4), l'ensemble des couples (r_0, σ_0) correspond aux solutions du programme d'optimisation de l'investisseur (2.2). Représenté graphiquement, cet ensemble solution est appelé **frontière efficiente**.

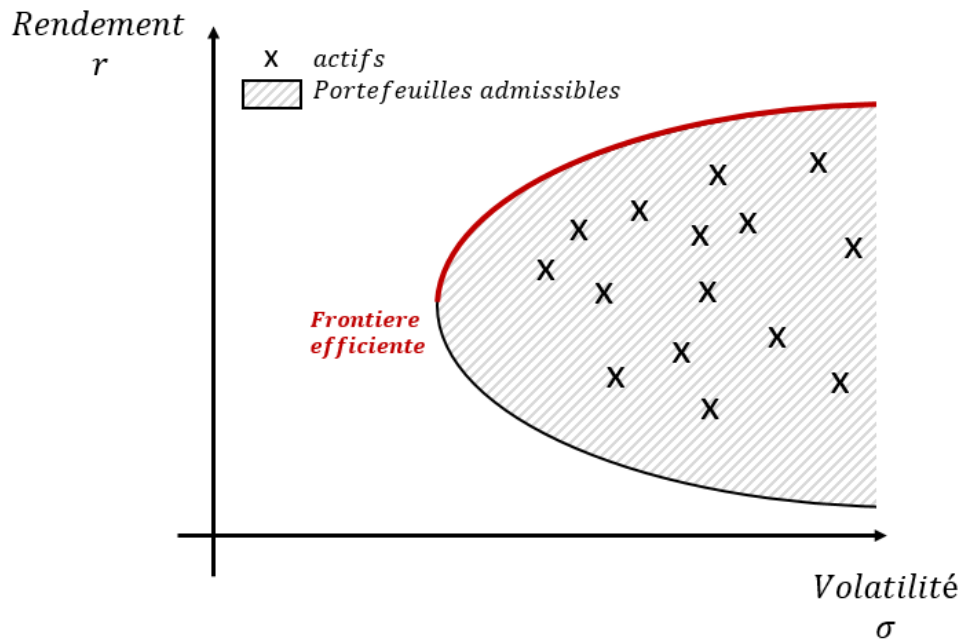


FIGURE 2.2 – représentation de la Frontière efficiente (source : [18])

Dès lors, l'objectif est non seulement d'adapter ce problème d'optimisation à un portefeuille de lignes d'assurance, mais aussi de déterminer si une frontière efficiente analogue

peut être déterminée.

2.1.3 Passer du problème d'optimisation d'un portefeuille d'actifs financiers à un portefeuille de lignes d'assurance

Comme mentionné précédemment, s'il est question dans les deux cas de maximiser le rendement tout en minimisant le risque, il existe des différences fondamentales entre l'optimisation d'un portefeuille d'assurance et celle d'un portefeuille d'actifs financiers. En particulier, si l'hypothèse simplificatrice d'absence de coûts de transaction peut être posée pour développer un modèle financier, en assurance, l'hypothèse selon laquelle il n'y a aucun coût de transaction, ou de friction lors de l'application d'une décision d'investissement est invraisemblable.

Admettons, par exemple, que l'actif i ne soit intéressant ni en termes de rendement, ni en termes de risque ou de diversification. Le poids optimal ω_i^* est donc 0. Dans ce cas, il serait aisé pour l'investisseur de ne pas acheter cet actif, ou bien de vendre la quantité d'actif dont il dispose.

À l'inverse, si une entreprise d'assurance constate qu'une ligne d'assurance est non pertinente du point de vue du profil rendement/risque, la possibilité de clore cette ligne immédiatement est pour le moins irréaliste. En effet, non seulement l'entreprise est-elle dans ce cas tenue de respecter ses engagements au long-terme déjà fixés, mais cela aurait aussi de nombreuses conséquences en terme d'emploi, d'image, ou de synergie avec le reste des lignes d'assurance.

Afin de pouvoir ultérieurement prendre cet aspect en compte sous formes de contraintes, le choix a été fait de ne pas raisonner en termes de proportion $(\omega_i)_{1 \leq i \leq n}$ de la valeur du portefeuille, mais d'optimiser l'évolution $(\delta_i)_{1 \leq i \leq n}$ de la taille de chaque ligne d'assurance. Ici, une évolution de la taille d'une ligne est définie comme la variation de l'exposition au risque de la ligne d'assurance.

En pratique, mesurer précisément l'exposition du risque n'est pas une tâche triviale. Une bonne mesure d'exposition doit en effet être capable de quantifier le risque assuré sur une période donnée de manière à fournir la meilleure estimation possible du prix du risque assuré. Toutes choses égales par ailleurs, la prime perçue doit donc être proportionnelle à la mesure de l'exposition au risque choisie.

Au cours de cette étude, le choix a été de mesurer l'évolution de l'exposition au risque par la variation du nombre de contrats associés à la ligne d'intérêt. Cette définition, relativement simple, se comprend aisément dans le cas de l'assurance des biens. Si une ligne d'assurance voit son nombre de contrats doubler, et que ces contrats sont relativement similaires à ceux déjà souscrits, il est de fait naturel de considérer que le risque a doublé.

Toutefois, si cette définition a l'avantage d'être intuitive, il convient de noter qu'il s'agit en réalité d'une simplification qui suppose que les risques soient homogènes. Or, cette hypothèse n'est notamment pas avérée dans le cas des grands risques. Elle sera toutefois maintenue afin de simplifier les calculs au cours de cette étude.

En notant $Rdmt(\delta_1, \dots, \delta_n)$ et $Risk(\delta_1, \dots, \delta_n)$ respectivement les valeurs de la mesure

de rendement et de la mesure du risque du portefeuille d'assurance après application des évolutions $(\delta_i)_{1 \leq i \leq n}$, le programme d'optimisation est donc le suivant :

$$\begin{aligned} \min_{\delta_i (1 \leq i \leq n)} & (-Rdmt(\delta_1, \dots, \delta_n), Risk(\delta_1, \dots, \delta_n)) \\ \text{s.c. } & \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n) \end{aligned} \tag{2.5}$$

Avec $\beta_i^{(min)}$ le taux d'évolution en deçà duquel réduire la ligne d'assurance i davantage serait irréaliste et $\beta_i^{(max)}$ son augmentation maximale. Si ces contraintes sont fixées, c'est parce qu'elles permettent d'intégrer divers phénomènes au programme d'optimisation :

- La contrainte d'évolution minimale $\beta_i^{(min)}$ reflète le fait qu'une réduction trop importante de l'activité d'une ligne n'est pas souhaitable, et que les engagements de l'assureur rendent un tel processus long à implémenter.
- La contrainte d'évolution maximale $\beta_i^{(max)}$ est fixée car, en pratique, il peut être très difficile de croître de manière significative sur une ligne d'assurance. Augmenter le nombre de contrats, par exemple, est une tâche ardue sans jouer sur le montant de primes demandée à l'assuré.

Il convient donc maintenant de définir des mesures de rendement et de risque cohérentes avec le problème posé. Ainsi, la valeur $Rdmt(\delta_1, \dots, \delta_n)$ sera définie en section 2.2, et la valeur $Risk(\delta_1, \dots, \delta_n)$ en section 2.3

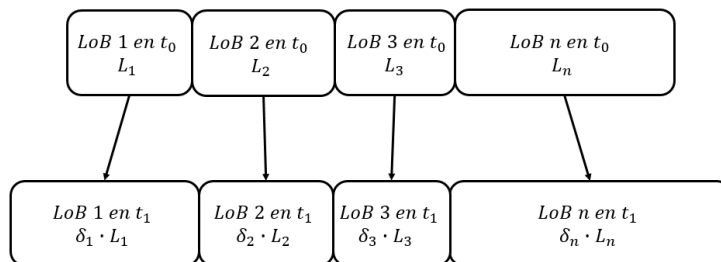


FIGURE 2.3 – Représentation de la décision effectuée par l'assureur

2.2 Le rendement économique d'un portefeuille de lignes d'assurance

Déjà défini en section 2.1.1, le rendement r_i d'un actif financier i n'est autre que l'évolution de sa valeur de marché sur une période donnée $V_i, t_1 / V_i, t_0$. Cependant, il serait peu pratique de définir le rendement d'une ligne d'assurance de manière analogue, car cela nécessiterait l'estimation de la valeur de cette ligne. Or, contrairement à un actif financier, une valeur de marché est rarement disponible, et il faudrait de toutes manières être capable de l'estimer après application de l'évolution δ_i .

L'objectif de cette sous-partie est donc tout d'abord de définir le rendement d'une ligne d'assurance i $Rdmt_i(1)$, ce qui sera effectué en section 2.2.1, puis celui du portefeuille d'assurance entier $Rdmt(1, \dots, 1)$ en section 2.2.2. Enfin, les hypothèses supplémentaires nécessaires à l'estimation du rendement après application de la stratégie de développement

$Rdmt(\delta_1, \dots, \delta_n)$ seront précisées en section 2.2.3.

2.2.1 Définir le rendement d'une ligne d'assurance

Le rendement espéré, de manière très générale, peut être défini comme la réponse à la question : "Quelle quantité de richesse cet investissement générera-t-il en moyenne ?" Dans le cas des actifs financiers, qui sont souvent des titres de créances échangeables, l'évolution de leur valeur de marché est un bon moyen d'obtenir la réponse à cette question. Sous l'hypothèse d'absence d'opportunité d'arbitrage, le prix d'un actif reflète en effet la valeur actualisée des flux futurs. Il y aurait ainsi équivalence entre détenir le titre à maturité ou le vendre sur les marchés financiers.

Or, une ligne d'assurance n'a en général pas vocation à être échangée, et c'est donc exclusivement la valeur actualisée des flux futurs qui intéresse une entreprise d'assurance. Dans cette étude, c'est donc le résultat actuariel de l'année $t + 1$ de la ligne qui sera retenu. Ainsi, avec les notations :

$$\left\{ \begin{array}{l} Result_i^{Disc} \text{ et } Result_i^{Undisc} = \text{Résultats techniques de } i \text{ actualisé et non actualisés} \\ Prime_i^{Disc} \text{ et } Prime_i^{Undisc} = \text{Primes de } i \text{ actualisées et non actualisées} \\ Frais_i^{Disc} \text{ et } Frais_i^{Undisc} = \text{Frais de } i \text{ actualisés et non actualisés} \\ Sinistre_i^{Disc} \text{ et } Sinistre_i^{Undisc} = \text{Sinistres de } i \text{ actualisées et non actualisées} \end{array} \right. \quad (2.6)$$

pour la ligne $i \in \{1, \dots, n\}$, avec $\delta_i = 1$, le rendement correspond à la variable aléatoire suivante :

$$Rdmt_i = Result_i^{Disc} \times (1 - TxImposition) \quad (2.7)$$

L'importance de la distinction entre les variables actualisées et les variables non actualisées apparaît dès lors que la dimension du temps est prise en compte. En effet, si les primes sont généralement acquises sur une période d'un an, le paiement des sinistres peut s'étaler sur une période bien plus longue. De la sorte, si le taux d'actualisation est positif, l'effet d'actualisation est plus fort sur les sinistres que sur les primes. Cela augmente donc le résultat actualisé par rapport au résultat non actualisé.

Afin de faire le lien entre le rendement et les indicateurs définis en partie 1.3.3, le rendement peut se réécrire de la manière suivante :

$$Rdmt_i = (Result_i^{Disc} + Result_i^{Undisc} - Result_i^{Undisc}) \times (1 - TxImposition)$$

Par définition la définition de $Result_i$ (1.6), l'égalité suivante est donc vérifiée :

$$\begin{aligned}
 Rdmt_i &= [(Prime_i^{Undisc} - Frais_i^{Undisc} - Sinistre_i^{Undisc}) + Result_i^{Disc} - Result_i^{Undisc}] \\
 &\quad \times (1 - TxImposition) \\
 &= Prime_i^{Undisc} \times \left(\frac{Prime_i^{Undisc} - Frais_i^{Undisc} - Sinistre_i^{Undisc}}{Prime_i^{Undisc}} \right. \\
 &\quad \left. - \frac{Result_i^{Disc} - Result_i^{Undisc}}{Prime_i^{Undisc}} \right) \times (1 - TxImposition) \\
 &= Prime_i^{Undisc} \times \left(1 - \frac{Frais_i^{Undisc} + Sinistre_i^{Undisc}}{Prime_i^{Undisc}} - \frac{Result_i^{Disc} - Result_i^{Undisc}}{Prime_i^{Undisc}} \right) \\
 &\quad \times (1 - TxImposition)
 \end{aligned}$$

Le rendement espéré avant application des facteurs d'évolutions, c'est-à-dire le rendement espéré lorsque $\delta_1 = \dots = \delta_n = 1$ est donc de :

$$\begin{aligned}
 Rdmt_i(1) &= \mathbb{E}[Rdmt_i] \\
 &= Prime_i^{Undisc} \times \left(1 - \mathbb{E} \left[\frac{Frais_i^{Undisc} + Sinistre_i^{Undisc}}{Prime_i^{Undisc}} \right] - \mathbb{E} \left[\frac{Result_i^{Disc} - Result_i^{Undisc}}{Prime_i^{Undisc}} \right] \right) \\
 &\quad \times (1 - TxImposition)
 \end{aligned}$$

Il apparaît donc que :

$Rdmt_i(1) = Prime_i^{Undisc} \times (1 - CR_i - ET_i) \times (1 - TxImposition)$ <p>avec :</p> $CR_i = \mathbb{E} \left[\frac{Frais_i^{Undisc} + Sinistre_i^{Undisc}}{Prime_i^{Undisc}} \right] \text{ le Ratio Combiné} \quad (2.8)$ $ET_i = \mathbb{E} \left[\frac{Result_i^{Undisc} - Result_i^{Disc}}{Prime_i^{Undisc}} \right] \text{ l'Effet-temps}$

Écrire le rendement de cette manière présente plusieurs intérêts. Tout d'abord, cette écriture permet de faire aisément le lien entre le rendement et le ratio combiné (1.3.3). En effet, plus le ratio combiné sera élevé, plus le rendement sera faible, et inversement.

De plus, cette écriture permet de mettre en lumière l'impact de l'effet-temps sur les différentes lignes d'assurance. En effet, certaines lignes d'assurance peuvent être considérées comme "longues", c'est-à-dire que les paiements de sinistres sont effectués après une durée de temps significative. Cela a donc pour impact d'augmenter la différence entre le résultat actualisé $Result^{Disc}$ et le résultat non actualisé $Result^{Undisc}$. Ainsi, en décomposant le rendement cette manière, cet effet peut être estimé de manière plus précise.

Enfin, cette écriture permet d'identifier les facteurs dont il faut mesurer la sensibilité à δ_i afin de calculer le rendement. Puisque le taux d'imposition ne dépend que du contexte économique et politique, seuls le montant de prime, le ratio combiné et l'effet-temps ne

peuvent dépendre de δ_i . Le calcul du rendement est donc automatique dès lors que les valeurs suivantes sont connues de manière explicite :

$$\begin{cases} Prime_i^{Undisc}(\delta_i) \\ CR_i^{Undisc}(\delta_i) \\ ET_i^{Undisc}(\delta_i) \end{cases}$$

Ainsi, l'objectif est désormais de déterminer comment calculer non plus le rendement d'une seule ligne d'assurance, mais le rendement de l'ensemble du portefeuille, ainsi que de déterminer comment évoluent le Ratio combiné, l'effet-temps et la prime en fonction des facteurs δ .

2.2.2 Définir le rendement du portefeuille d'assurance

Maintenant que le rendement espéré d'une unique ligne d'assurance $Rdmt_i(1)$ est connu, il est possible d'appliquer la même méthodologie au rendement espéré du portefeuille du portefeuille complet.

Ainsi, en notant $Rdmt_{(p)}(1)$ le rendement du portefeuille après application des évolutions, il apparaît que :

$$Rdmt_{(p)}(1) = Prime_{(p)}^{Undisc} \times (1 - CR_{(p)} - ET_{(p)}) \times (1 - TxImposition) \quad (2.9)$$

Or, la prime totale, les frais totaux, la sinistralité totale, et donc le résultat total du portefeuille correspond simplement à la somme des primes et des sinistres des lignes individuelles :

$$\begin{cases} Prime_{(p)}(1) & = \sum_{i=1}^n Prime_i(1) \\ Frais_{(p)}(1) & = \sum_{i=1}^n Frais_i(1) \\ Sinistres_{(p)}(1) & = \sum_{i=1}^n Sinistres_i(1) \\ Result_{(p)}(1) & = \sum_{i=1}^n Result_i(1) \end{cases} \quad (2.10)$$

Il est donc possible de démontrer que :

$$\begin{aligned} Rdmt_{(p)}(1) &= \sum_{i=1}^n Prime_i^{Undisc} \times (1 - CR_i - ET_i) \times (1 - TxImposition) \\ &= \sum_{i=1}^n Rdmt_i(1) \end{aligned} \quad (2.11)$$

Ainsi, pour calculer le rendement espéré du portefeuille total, il suffit de connaître le rendement des lignes individuelles.

L'enjeu est donc désormais d'être en mesure d'estimer le rendement des lignes individuelles après application de leurs stratégies d'évolutions respectives. Pour ce faire, un certain nombre d'hypothèses devront être posées sur l'impact des facteurs d'évolution δ sur les diverses variables d'une ligne d'assurance.

2.2.3 Les hypothèses liées à l'impact du facteur d'évolution sur le portefeuille d'assurance

Le rendement d'une ligne d'assurance et le rendement du portefeuille d'assurance étant maintenant définis dans le cas $(\delta_i)_{1 \leq i \leq 1} = 1$, il convient de déterminer comment évoluent ces rendements lorsque $(\delta_i)_{1 \leq i \leq 1} \neq 1$. Comme cela a été démontré en partie 2.2.1, cela revient donc à déterminer comment évoluent la prime, le Ratio Combiné et l'effet-temps en fonction de δ_i .

Pour ce faire, un certain nombre d'hypothèses devront être posées.

L'hypothèse de linéarité des primes

La première hypothèse posée sera l'hypothèse de linéarité des primes. Comme mentionné ci-dessus, le paramètre δ_i en fonction duquel l'optimisation est effectuée correspond à la croissance de la ligne d'assurance i , mesurée par l'évolution du nombre de contrats associés à cette ligne.

Ainsi, soit C_i le nombre de contrats de la ligne i , et soient $prime_{i,c}(1)$ et $prime_{i,c}(\delta_i)$ respectivement les primes perçues avec le contrat $c \in \{1, \dots, C_i\}$ avant et après évolution de la ligne. Les montants de prime totaux par ligne s'écrivent donc de la manière suivante :

$$\left\{ \begin{array}{l} Prime_i(1) = \sum_{c=1}^{C_i} prime_{i,c}(1) \text{ les primes totales de la ligne } i \text{ avant évolution} \\ Prime_i(\delta_i) = \sum_{c=1}^{\delta_i \times C_i} prime_{i,c}(\delta_i) \text{ les primes totales de la ligne } i \text{ après évolution} \end{array} \right.$$

Or, le montant de primes associé au contrat $c \in \{1, \dots, C_i\}$ ne relève que très peu de l'aléa. En supposant que les primes perçues par contrats sont indépendantes de l'évolution de l'exposition au risque, c'est-à-dire que $\forall c, prime_{i,c}(1) = prime_{i,c}(\delta_i)$, la prime totale de la ligne i après évolution se calcule donc la manière suivante :

$$\begin{aligned} Prime_i(\delta_i) &= \sum_{c=1}^{\delta_i \times C_i} prime_{i,c}(\delta_i) \\ &= \sum_{c=1}^{\delta_i \times C_i} prime_{i,c}(1) \\ &= \delta_i \times \sum_{c=1}^{C_i} prime_{i,c}(1) \end{aligned}$$

Dès lors, l'hypothèse de linéarité des primes fixées s'énonce telle que :

$$Prime_i(\delta_i) = \delta_i \times Prime_i(1)$$

L'hypothèse de linéarité des Frais

L'hypothèse de linéarité des frais consiste à considérer que ceux-ci évoluent de manière conjointe aux primes. Autrement dit, il est considéré que :

$$Frais_i(.) = a_i \times Prime_i(.)$$

Ce qui, par l'hypothèse de linéarité des primes, a la conséquence suivante sur l'évolution des frais :

$$Frais_i(\delta_i) = \delta_i \times Frais_i(1)$$

En pratique, cette hypothèse peut ne pas être avérée, car un certain volume de frais fixes indépendants du montant de primes seraient à prendre en compte. Toutefois, l'impact de cette hypothèse sur le programme d'optimisation n'est en réalité que marginal ; elle sera donc considérée afin de simplifier l'écriture des calculs.

L'hypothèse de linéarité de l'actualisation

La seconde hypothèse posée sera ici l'hypothèse de linéarité de l'actualisation. Au cours de cette étude, il est donc considéré que la courbe des taux d'actualisation r est constante, ou tout du moins ne dépend pas de δ_i .

En effet, le taux d'actualisation découle du fait qu'il peut exister un temps de latence entre le moment où la prime est perçue et le moment où le sinistre est indemnisé, ce qui permet aux assureurs de fructifier les primes.

Cette hypothèse revient donc à supposer que :

- la décision d'évolution d'une entreprise d'assurance n'a pas d'impact significatif sur les prix sur les marchés financiers
- la décision d'évolution n'a pas d'impact sur la structure des placements effectués

En pratique, il est peu probable que la performance d'une entreprise d'assurance particulière ait un impact significatif sur les marchés financiers dans leur ensemble. Cette première hypothèse est donc raisonnable.

De même, l'actualisation est ici effectuée avec le taux sans risque. Ainsi, s'il est possible que la performance d'une ligne puisse influencer la structure réelle des placements, cela n'a pas d'impact sur le programme d'optimisation. C'est donc uniquement de la durée de la ligne que va dépendre l'effet-temps. Or, la durée ne reflète que la durée de temps moyenne sur laquelle les sinistres sont payés, et est donc indépendante du facteur δ .

Cette hypothèse est donc justifiée en pratique.

En supposant qu'un sinistre est payé en $t + 1$, l'impact de cette hypothèse est le suivant :

$$\begin{cases} Sinistre_i^{Disc}(1) = \frac{Sinistre_i^{Undisc}(1)}{(1+r)} \text{ par définition} \\ Sinistre_i^{Disc}(\delta_i) = \frac{Sinistre_i^{Undisc}(\delta_i)}{(1+r)} \text{ par définition} \end{cases} \Rightarrow \frac{Sinistre_i^{Undisc}(\delta_i)}{Sinistre_i^{Undisc}(1)} = \frac{Sinistre_i^{Disc}(\delta_i)}{Sinistre_i^{Disc}(1)} \quad (2.12)$$

Autrement dit :

$$\begin{aligned} \text{si } \text{Sinistre}_i^{\text{Undisc}}(\delta_i) &= \alpha_i \text{Sinistre}_i^{\text{Undisc}}(1) \\ \text{alors } \text{Sinistre}_i^{\text{Disc}}(\delta_i) &= \alpha_i \text{Sinistre}_i^{\text{Disc}}(1) \end{aligned} \quad (2.13)$$

L'hypothèse de linéarité des Sinistres :

La dernière hypothèse posée sera l'hypothèse de la linéarité des sinistres en fonction de l'exposition au risque. Pour l'énoncer, le nombre attendu de sinistres commencera par être étudié. Ainsi, en notant $b_{i,c}(1)$ et $b_{i,c}(\delta_i)$ les variables aléatoires correspondant respectivement au nombre de sinistres associés contrat $c \in \{1, \dots, C_i\}$ avant et après évolution de la ligne i :

$$\left\{ \begin{array}{l} B_i(1) = \sum_{c=1}^{C_i} b_{i,c}(1) \text{ Le nombre total de sinistres de la ligne } i \text{ avant évolution} \\ B_i(\delta_i) = \sum_{c=1}^{\delta_i \times C_i} b_{i,c}(\delta_i) \text{ Le nombre total de sinistres de la ligne } i \text{ après évolution} \end{array} \right.$$

Par un raisonnement analogue à celui effectué pour les primes, en supposant que $b_{i,c}(1)$ et $b_{i,c}(\delta_i)$ sont de même loi, cela implique que :

$$\mathbb{E}[B_i(\delta_i)] = \delta_i \times \mathbb{E}[B_i(1)]$$

Or, en notant $\text{sinistre}_{i,b}(1)$ et $\text{sinistre}_{i,b}(\delta_i)$ les variables aléatoires correspondant aux montants du b -ième sinistre respectivement avant et après évolution de la ligne, la sinistralité totale est de :

$$\left\{ \begin{array}{l} \text{Sinistre}_i(1) = \sum_{b=1}^{B_i(1)} \text{sinistre}_{i,b}(1) \text{ avant évolution} \\ \text{Sinistre}_i(\delta_i) = \sum_{b=1}^{B_i(\delta_i)} \text{sinistre}_{i,b}(\delta_i) \text{ après évolution} \end{array} \right.$$

Ainsi, en supposant que $\text{sinistre}_{i,b}(1)$ et $\text{sinistre}_{i,b}(\delta_i)$ sont de même loi et que les variables de coûts de sinistres et de nombre de sinistres $\text{sinistre}_{i,b}$ et B_i sont indépendantes, la propriété de l'espérance fréquence-sévérité indique que :

$$\begin{aligned} \mathbb{E}[\text{Sinistre}_i(\delta_i)] &= \mathbb{E}[B_i(\delta_i)] \times \mathbb{E}[\text{sinistre}_{i,b}(\delta_i)] \\ &= \delta_i \times \mathbb{E}[B_i(1)] \times \mathbb{E}[\text{sinistre}_{i,b}(1)] \\ &= \delta_i \times \mathbb{E}[\text{Sinistre}_i(1)] \end{aligned}$$

Ainsi, l'hypothèse de linéarité des sinistres s'appuie sur ce dernier résultat pour considérer que :

$$\text{Sinistre}_i(\delta_i) = \delta_i \text{Sinistre}_i(1)_i$$

Il convient cependant à ce stade de noter que l'hypothèse selon laquelle les montants de primes par contrat $prime_{i,c}$, le nombre de sinistres par contrats $b_{i,c}$ et le montant indemnisé par sinistre $sinistre_{i,b}$ sont indépendants du facteur d'évolution δ est une hypothèse assez forte.

En effet, pour la ligne $i \in 1, \dots, n$, un facteur δ_i supérieur à 1 revient par exemple à augmenter la souscription de contrats au sein de la ligne. Or, cela se peut se révéler très difficile sans mettre en place une politique tarifaire agressive.

Ainsi, le facteur δ_i ne peut en réalité dépasser un certain seuil qu'à condition de diminuer le taux de marge. Autrement dit, il faudrait dans cas soit accepter de percevoir une prime $prime_{i,c}$ plus faible, soit d'accepter des contrats plus risqués, et donc accepter de voir les montants ou fréquences de sinistres $b_{i,c}$ et $sinistre_{i,b}$ augmenter.

Si ces phénomènes sont intégrés au programme d'optimisation par la contrainte $\delta_i \leq \beta_i^{(max)}$, il serait aussi possible de les intégrer directement au sein de la fonction objectif. Cela pourrait par exemple se faire en ajoutant les paramètres $\alpha_1, \alpha_2 > 0$ à la modélisation de la manière suivante :

$$— \text{Sinistre}_i(\delta_i) = \delta_i \times \text{Sinistre}_i(1) \times (1 + \alpha_1)$$

$$— \text{Prime}_i(\delta_i) = \delta_i \times \text{Prime}_i(1) \times (1 - \alpha_2)$$

Cependant, la calibration de ces paramètres doit dans ce cas faire l'objet d'une étude approfondie et dépasse le cadre de l'analyse menée ici.

Enfin, il convient de rappeler que l'impact de la réassurance est ici négligé. En effet, l'objectif de cette partie est d'optimiser le portefeuille d'assurance avant réassurance, ce qui revient à déléguer le problème de la structure de réassurance à l'entreprise.

Cependant, lorsque l'optimisation sera faite sous l'hypothèse d'une structure de réassurance donnée dans la partie 3, il sera nécessaire d'ajuster ces hypothèses pour prendre en compte l'impact des traités non-proportionnels.

2.2.4 Estimer le rendement après application d'une stratégie d'évolution

Maintenant l'impact du facteur d'évolution δ sur l'actualisation et sur les sinistres et frais définis, il est possible de déterminer le ratio combiné et l'effet-temps après application de la stratégie d'évolution. Il est donc possible de déterminer la valeur du rendement espéré du portefeuille.

Ainsi, en utilisant l'hypothèse de linéarité des sinistres et des frais, il apparaît que l'impact de δ_i sur le ratio combiné de la ligne $i \in 1, \dots, n$ est le suivant :

$$\begin{aligned}
 RC_i(\delta_i) &= \mathbb{E} \left[\frac{Frais_i^{Undisc}(\delta_i) + Sinistre_i^{Undisc}(\delta_i)}{Prime_i^{Undisc}(\delta_i)} \right] \\
 &= \mathbb{E} \left[\frac{\delta_i (Frais_i^{Undisc}(1) + Sinistre_i^{Undisc}(1))}{\delta_i Prime_i^{Undisc}(1)} \right] \\
 &= RC_i(1)
 \end{aligned} \tag{2.14}$$

De même, l'impact de δ_i en utilisant les hypothèses de linéarité de l'actualisation et des sinistres et frais, l'impact sur l'effet-temps est le suivant :

$$\begin{aligned}
 ET_i(\delta_i) &= \mathbb{E} \left[\frac{Result_i^{Undisc}(\delta_i) - Result_i^{Disc}(\delta_i)}{Prime_i^{Undisc}(\delta_i)} \right] \\
 &= \mathbb{E} \left[\frac{\delta_i (Result_i^{Undisc}(1) - Result_i^{Disc}(1))}{\delta_i Prime_i^{Undisc}(1)} \right] \\
 &= ET_i
 \end{aligned} \tag{2.15}$$

Sous ces hypothèses, le ratio combiné et l'effet-temps sont donc tous deux stables en fonction du paramètre d'évolution δ . Ainsi, il est trivial de démontrer que le rendement après évolution du portefeuille d'assurance total peut s'écrire de la manière suivante :

$$\begin{aligned}
 Rdmt_i(\delta_i) &= \delta_i \times Rdmt_i(1) \\
 Rdmt(\delta_1, \dots, \delta_n) &= \sum_{i=1}^n \delta_i \times Rdmt_i(1)
 \end{aligned} \tag{2.16}$$

Le programme d'optimisation se réécrit donc de la manière suivante :

$$\begin{aligned}
 \min_{\delta_{i(1 \leq i \leq n)}} & \left(- \sum_{i=1}^n \delta_i \times Rdmt_i(1), Risk(\delta_1, \dots, \delta_n) \right) \\
 s.c. & \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n)
 \end{aligned} \tag{2.17}$$

Ainsi, une fois le vecteur $(Rdmt_1(1), \dots, Rdmt_n(1))$ déterminé, il n'est plus nécessaire de recalculer le rendement à chaque étape de l'algorithme d'optimisation, ce qui diminuera sa complexité de manière significative.

Il convient maintenant de définir le risque associé à un portefeuille d'assurance $Risk(\delta_1, \dots, \delta_n)$ afin de résoudre le programme d'optimisation souhaité.

2.3 Une mesure de risque du portefeuille d'assurance : le Coût en capital

Maintenant la mesure de rendement $Rdmt(\delta_1, \dots, \delta_n)$ adéquatement définie, il convient de définir la mesure de risque $Risk(\delta_1, \dots, \delta_n)$. S'il est possible de définir le risque d'une comme la volatilité associée au rendement, en notant donc $Risk(\delta_1, \dots, \delta_n) = \sqrt{Var(Rdmt(\delta_1, \dots, \delta_n))}$,

un autre choix est effectué ici.

En effet, si l'écart-type est souvent utilisé en finance quantitative, le coût en capital a été préféré lors de cette étude dans un objectif d'allocation stratégique du capital.

Ainsi, cette partie commencera par donner l'intuition derrière la notion de coût en capital en section 2.3.1, et expliquer pourquoi cette notion répond aux problématiques des assureurs sur le risque. Ensuite, la définition formelle de cette mesure pourra être donnée en section 2.3.2, et son mode de calcul sera exploré en section 2.3.4.

2.3.1 Le constat à l'origine de la notion de coût en capital

La notion de coût en capital vient du constat suivant : pour une entreprise d'assurance, le risque a un coût.

Afin d'assurer sa pérennité et sa capacité à satisfaire ses engagements à indemniser ses clients, une entreprise d'assurance se doit en effet de détenir une certaine quantité de fonds propres. C'est justement cette quantité de fonds propres qui est encadrée par la législation *Solvabilité II*, qui doit notamment être suffisante pour limiter la probabilité de ruine de l'entreprise d'assurance à au plus 0.5%.

Il convient toutefois de noter qu'en pratique, les entreprises d'assurances limitent rarement la quantité de capital au capital de solvabilité requis. En réalité, elles voudront plutôt maintenir leur niveau de capital au-dessus de leur cible de ratio de couverture, c'est-à-dire le rapport entre le capital détenu et le capital réglementaire.

Dès lors, la question de la provenance de ce capital se pose. Deux sources principales sont à distinguer :

1. Les capitaux apportés par les actionnaires
2. Les capitaux acquis par l'activité économique

Ainsi, dans le cas d'une insuffisance de fonds propres, l'entreprise d'assurance sera amenée à demander plus de fonds propres aux actionnaires, en contrepartie d'une rémunération du capital supplémentaire.

Le montant exact de cette rémunération peut cependant être difficile à estimer, en particulier à long-terme, car elle dépend des taux d'intérêts et des taux d'actualisation. C'est pourquoi la notion de coût en capital développée lors de cette étude cherche à répondre à la question suivante :

"Si l'exposition au risque de la ligne d'assurance i évolue de δ_i , quelle sera la nouvelle quantité de capital requise par ces engagements ?"

Autrement dit, la question est de déterminer la quantité de fonds propres nécessaires suite à cette évolution afin de s'assurer que la probabilité de ruine ne dépasse pas 0.5%. Il convient maintenant d'établir une définition formelle de ce coût en capital.

2.3.2 La définition formelle du coût en capital

Comme mentionné précédemment, les fonds propres proviennent soit de l'activité d'assurance, soit des actionnaires. Autrement dit, c'est une fonction de la distribution de la perte, c'est-à-dire de l'opposé du résultat du contrat :

$$\begin{aligned} Capital_{\text{contrat } c} &= f_{\text{contrat } c}(-Resultat_{\text{contrat } c}) \\ &= f_{\text{contrat } c}((Sinistre_{\text{contrat } c} + Frais_{\text{contrat } c}) - Prime_{\text{contrat } c}) \end{aligned}$$

En notant $Resultat_{t_0}^{(p)}$ le résultat du portefeuille enregistré pour l'année considérée, une estimation naïve de coût en capital du portefeuille serait donc la suivante :

$$CeC^{(p)} = -(VaR_{0.005}[Resultat_{t_0}^{(p)}] - \mathbb{E}[Resultat_t^{(p)}]) \quad (2.18)$$

Cependant, cette estimation est incomplète. En effet, le montant de sinistre est une quantité aléatoire connue seulement au moment du paiement de l'ensemble des sinistres survenus lors de l'année de référence. Ainsi, pour $t = 1$ l'année où le sinistre est survenu, seule une estimation de ce résultat peut être enregistrée. Cela entraîne l'existence d'un risque de déviation entre le sinistre réellement payé et le sinistre originellement enregistré, appelé *risque de réserve*.

Ainsi, en notant \mathcal{F}_1 l'ensemble des informations connues en $t = 1$, la perte enregistrée pour le contrat c sera la suivante :

$$-Resultat_{\text{contrat } c}^1 = \mathbb{E}[Sinistre_{\text{contrat } c} + Frais_{\text{contrat } c} - Prime_{\text{contrat } c} | \mathcal{F}_1]$$

C'est pourquoi cette estimation a vocation à être ré-estimée à chaque année $t \in [2, \dots, T]$ avec T la date à laquelle il est estimé que le sinistre sera totalement payé. Cela signifie que le résultat enregistré à chaque année t est la différence entre cette nouvelle estimation et l'estimation de l'année précédente :

pour $t \in [2, \dots, T]$:

$$Resultat_{\text{contrat } c}^t = \mathbb{E}[Sinistre_{\text{contrat } c} + Frais_{\text{contrat } c} - Pime_{\text{contrat } c} | \mathcal{F}_t] - Resultat_{\text{contrat } c}^{t-1}$$

Or, plus une ligne est longue, et plus la durée entre la survenance du sinistre et sa clôture est élevée. Cela signifie qu'en réalité, il peut exister des écarts significatifs entre la perte initialement enregistrée et la perte effective : c'est là qu'intervient le risque de réserve.

Dès lors la question est la suivante : Comment tenir compte de ce risque de réserve lors du calcul du coût en capital ?

Si l'objectif est de calculer une mesure telle le *ratio de solvabilité Requis*, le risque de réserve doit être calculé en tenant compte des réserves liées contrats souscrits les années passées.

Cependant, ici, l'objectif n'est pas de déterminer la quantité de capital à détenir à des fins réglementaires ou prudentielles, mais de déterminer l'impact futur de la décision d'évolutions $\delta = (\delta_1, \dots, \delta_n)$. Ainsi, c'est ici l'ensemble des risques de réserves futurs engendrés par cette décision qui doit être évaluée.

C'est pourquoi, en notant $T = \sup_{\text{contrat}}(t_{\text{contrat}})$ la durée à partir de laquelle tous les sinistres seront payés, avec $\delta_1 = \dots = \delta_n = 1$, le coût en capital est défini de la manière suivante :

$$C_e C^{(p)}(1) = \sum_{t=1}^T \frac{-(VaR_{0.005}[Resultat_t^{(p)}] - \mathbb{E}[Resultat_t^{(p)}])}{(1+r)^t} \quad (2.19)$$

Cette écriture permet donc de rendre compte non seulement du montant de capital requis pour une année donnée avec le terme en $t = 1$, mais aussi du risque de réserve engendré par la souscription avec les termes $t \in \{2, \dots, T\}$.

Dès lors, la question est maintenant de déterminer comment calculer le coût en capital après l'application des évolutions $(\delta_i)_{i \in [1, \dots, n]}$

2.3.3 Les hypothèses nécessaires au calcul du coût en capital

L'enjeu principal du calcul du coût en capital est donc le calcul des quantités suivantes :

- $\forall t \in [1, T], VaR_{0.005}[Resultat_t^p]$
- $\forall t \in [1, T], \mathbb{E}[Resultat_t^p]$

Pour ce faire, il est possible d'utiliser la base de données correspondant aux sorties du modèle interne AXA. Celle-ci comprend en effet les distributions des différents types de sinistres qui correspondent au risque de souscription, c'est-à-dire les sinistres attritionnels, les sinistres atypiques et les sinistres liés aux catastrophes naturelles. Autrement dit, pour chaque ligne $i \in 1, \dots, n$, les distributions des lois de probabilités sont données sous la forme de $S = 50\,000$ réalisations pour les variables suivantes :

- $\forall t \in [1, T], Sinistres_{i,t}^{ATT}$
- $\forall t \in [1, T], Sinistres_{i,t}^{ATY}$
- $\forall t \in [1, T], Sinistres_{i,t}^{CAT}$

Les lois des variables $Resultat_{i,t}, t \geq 2$ ont notamment pu être estimés au sein du modèle interne en projetant l'écoulement des réserves générées par la souscription, tout en considérant le risque de déviation de ces réserves. Elles sont donc considérées ici comme données.

L'enjeu est donc désormais d'agrèger les sinistres des différents types de risques et des différentes lignes d'assurance de sorte à déterminer la distribution de la loi de probabilité du résultat de l'ensemble du portefeuille, ou du moins l'espérance et le quantile.

2.3. Une mesure de risque du portefeuille d'assurance : le Coût en capital 39

Déterminer l'espérance de la somme des sinistres

Dans le cas de l'espérance, l'agrégation est triviale. En effet, par la linéarité de l'espérance, en notant n le nombre de lignes d'assurance, il apparaît immédiatement que :

$$\begin{cases} \mathbb{E} \left[Resultat_t^{(p)} \right] = \sum_{i=1}^n \mathbb{E} [Resultat_{t,i}] \\ \mathbb{E} \left[Sinistre_t^{(p)} \right] = \sum_{i=1}^n \mathbb{E} [Sinistre_{i,t}^{ATT}] + \mathbb{E} [Sinistre_{i,t}^{ATY}] + \mathbb{E} [Sinistre_{i,t}^{CAT}] \end{cases}$$

Déterminer le quantile de la somme des sinistres

Contrairement à l'espérance, il n'existe pas de formule explicite permettant de connaître le quantile d'une somme de variables aléatoires en général. Toutefois, le quantile du résultat ou de la sinistralité sont des quantités que toute entreprise d'assurance est tenue de pouvoir calculer, notamment dans le cadre du SCR.

C'est pourquoi cette étude fut effectuée en reprenant les hypothèses du modèle interne AXA. Cela signifie que les matrices de corrélations suivantes sont supposées connues :

$$\Omega^{risques} = \begin{pmatrix} 1 & \rho_{ATT,ATY} & \rho_{ATT,CAT} \\ \rho_{ATY,ATT} & 1 & \rho_{ATY,CAT} \\ \rho_{CAT,ATT} & \rho_{CAT,ATY} & 1 \end{pmatrix}$$

$$\Omega^{lignes} = \begin{pmatrix} \rho_{1,1} & \rho_{1,2} & \cdots & \rho_{1,n} \\ \rho_{2,1} & \rho_{2,2} & \cdots & \rho_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n,1} & \rho_{n,2} & \cdots & \rho_{n,n} \end{pmatrix}$$

De même, il est supposé que ces matrices permettent d'agréger les résultats en suivant les formules suivantes :

Agrégation inter-lignes des risques non-CAT

$$VaR \left(Resultat_{lignes\ 1, \dots, n}^{(ATT\&ATY)} \right) = \sqrt{\left(VaR_1^{(ATT\&ATY)}, \dots, VaR_n^{(ATT\&ATY)} \right) \times \Omega^{lignes} \times \left(VaR_1^{(ATT\&ATY)}, \dots, VaR_n^{(ATT\&ATY)} \right)^T}$$

Agrégation inter-risques

$$VaR^{(p)} = \sqrt{VaR^{(CAT)^2} + VaR^{(ATT\&ATY)^2} + 2 \times \rho_{CAT,ATY\&ATT} VaR^{(CAT)} VaR^{(ATT\&ATY)}}$$

Ces hypothèses, fausses en général, sont en réalité vérifiées dans le cas où les variables Résultat sont Gaussiennes. Cela se démontre aisément en utilisant les propriétés de la loi normale. En effet, si $Z = (Z_1, \dots, Z_m)$ est un vecteur gaussien de moyenne μ et de matrice de covariance $\Omega = (\sigma_{i,j})_{1 \leq i,j \leq m}$, les propriétés suivantes sont vérifiées :

$$VaR_\alpha(Z_i) = \mu_i + \sigma_i \Phi^{-1}(\alpha)$$

avec :

$$\sigma_{\Sigma Z} = \sqrt{(\sigma_1, \dots, \sigma_n) \omega (\sigma_1, \dots, \sigma_n)^T}$$

Φ la fonction de répartition de la loi normale

La seule exception est le cas des Résultats CAT, puisque les sinistres associés à ce type de risque touchent souvent d'emblée plusieurs lignes en même temps. Les simulations effectuées au sein du modèle interne prennent donc déjà en compte les corrélations existantes entre les lignes. Autrement dit, pour toute simulation au sein du modèle interne $s \in 1, \dots, S$, il est possible d'agrèger les sinistres CAT de la manière suivante :

Agrégation inter-lignes des risques CAT

$$VaR \left(\text{Sinistre}_{s, \text{lignes } 1, \dots, n}^{(CAT)} \right) = VaR \left(\sum_{i=1}^n \text{Sinistres}_{i,s}^{CAT} \right)$$

Ainsi, s'il est certes nécessaire d'émettre des hypothèses sur les méthodes d'agrégation afin de calculer le quantile de la somme des sinistres, aucune hypothèse n'a dû être ajoutée à cette étape. Cela est dû au fait que les hypothèses ont déjà été posées dans le cadre de la calibration du modèle interne.

Dès lors, il suffit d'utiliser ces hypothèses avec les hypothèses sur l'impact du facteur d'évolution δ fixées en partie 2.2.3 pour déterminer le coût en capital après stratégie d'évolution, et donc écrire le programme d'optimisation.

2.3.4 Écrire le programme d'optimisation de l'assureur

Maintenant que les hypothèses sur les méthodes d'agrégation et sur l'impact des facteurs d'évolution sont posées, il devient possible d'estimer le coût en capital de l'ensemble du portefeuille après application de la stratégie δ .

Pour ce faire, les hypothèses fixées en partie 2.2.3 sur l'évolution des sinistres seront notamment reprises. Ainsi, pour une ligne $i \in \{1, \dots, n\}$:

2.3. Une mesure de risque du portefeuille d'assurance : le Coût en capital 41

$$\begin{cases} Resultat_i^{(ATT)}(\delta_i) = \delta_i \times Resultat_i^{(ATT)} \\ Resultat_i^{(ATY)}(\delta_i) = \delta_i \times Resultat_i^{(ATY)} \\ Resultat_i^{(CAT)}(\delta_i) = \delta_i \times Resultat_i^{(CAT)} \\ Resultat_i(\delta_i) = \delta_i \times Resultat_i \end{cases}$$

Ces hypothèses permettent d'obtenir l'espérance suivante :

$$\mathbb{E} \left[Resultat_t^{(p)}(\delta_1, \dots, \delta_n) \right] = \sum_{i=1}^n \delta_i \times \mathbb{E} [Resultat_{t,i}]$$

Ainsi que le quantile suivant :

$$VaR_{0.005} \left(Resultat_t^{(p)}(\delta) \right) = \sqrt{VaR_{0.005}^{(CAT)}(\delta)^2 + VaR_{0.005}^{(ATT\&ATY)}(\delta)^2 + 2 \times \rho_{CAT,ATY\&ATT} VaR_{0.005}^{(CAT)}(\delta) VaR_{0.005}^{(ATT\&ATY)}(\delta)}$$

Cela signifie donc qu'il est possible de calculer le coût en capital total après application de la stratégie δ :

$$CeC^{(p)}(\delta_1, \dots, \delta_n) = \sum_{t=1}^T \frac{-(VaR_{0.005}[Resultat_t^{(p)}(\delta_1, \dots, \delta_n)] - \mathbb{E}[Resultat_t^{(p)}(\delta_1, \dots, \delta_n)])}{(1+r)^t} \quad (2.20)$$

Maintenant que les mesures de risque et de rendement sont définies et que leurs méthodes de calcul respectives ont été déterminées, le programme d'optimisation de l'investisseur est donc bien défini :

$$\begin{array}{l} \min_{\delta_i (1 \leq i \leq n)} \left(- \sum_{i=1}^n \delta_i R_{dmt_i}(1), CeC^{(p)}(\delta_1, \dots, \delta_n) \right) \\ \text{s.c.} \quad \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n) \\ \sum_{i=1}^n \delta_i \times Primes_i > \alpha_0 \sum_{i=1}^n Primes_i \end{array}$$

où $\alpha_0 \in [0, 1]$ permet de fixer le seuil en dessous duquel la prime totale ne peut pas descendre, afin d'empêcher une diminution trop importante de l'activité.

L'objectif de cette partie était de déterminer un modèle théorique qui permettrait de répondre à la problématique d'optimisation, et donc de déterminer l'ensemble des combinaisons d'évolution des lignes d'assurance qui soient strictement optimales.

L'enjeu principal était donc d'adapter les travaux de mathématiques financières effectués sur les frontières efficientes aux problématiques actuarielles. Ce processus a pu être effectué en plusieurs étapes :

- Tout d'abord, des mesures de risque et de rendement cohérentes avec le secteur de l'assurance ont dû être définies.
- Ensuite, la question était de déterminer comment ces indicateurs évoluent en fonction des facteurs d'évolution δ , ce qui a été effectué en posant certaines hypothèses.
- Enfin, les méthodes de calcul concrètes de ces indicateurs ont pu être mises en place en se fondant notamment sur les hypothèses posées dans le cadre de la calibration du modèle interne.

C'est donc au terme de ce processus que le programme d'optimisation ci-dessus a été défini.

Toutefois, si poser un certain nombre d'hypothèses est certes nécessaire lors de la résolution de ce problème, celles posées jusque-là négligent l'impact de la réassurance, ce qui peut conduire à biaiser fortement les résultats. En outre, si le problème d'optimisation a certes été posé, son ensemble de solution n'est pour le moment pas défini de manière explicite.

Dès lors, les parties suivantes de cette étude se consacreront d'abord à l'intégration d'une structure de réassurance au programme d'optimisation, puis aux méthodes utilisées pour déterminer l'ensemble solution de ce programme.

Chapitre 3

L'impact d'une structure de réassurance donnée sur le programme d'optimisation

Le programme d'optimisation exploré jusqu'ici, défini au cours de la partie 2, n'a considéré que les variables brutes de réassurance. Autrement dit, toutes les hypothèses posées lors de la formulation du problème d'optimisation ont été formulées en ignorant l'impact que la réassurance pouvait avoir sur les différentes variables.

Résoudre ce programme peut trouver son utilité en pratique. En effet, il peut être efficace pour un assureur de tenter d'optimiser d'abord le rendement et le risque d'un portefeuille brut avant d'essayer de lui appliquer la structure de réassurance optimale. Toutefois, une autre approche serait de considérer le programme d'optimisation net de réassurance, de manière à connaître immédiatement l'effet final des stratégies de souscription sur le portefeuille. Ces deux approches peuvent conduire à des résultats très différents, notamment dans le cas où le portefeuille d'assurance conjugue des lignes pour lesquelles l'impact de la réassurance est limité et d'autres qui sont protégées par la réassurance de manière significative.

C'est pourquoi cette partie se consacrera à l'intégration de l'impact de la réassurance dans le programme d'optimisation. Pour ce faire, il sera notamment possible de considérer la structure de réassurance du portefeuille étudié comme figée, et d'observer son impact sur les mesures de risque et de rendement déjà identifiées (parties 2.2.1 et 2.3.2).

Tout d'abord, la section 3.1 de cette partie décrira donc la structure de réassurance d'AXA afin de déterminer son impact sur les hypothèses. Ensuite, l'objectif de la section 3.2 sera de développer une modélisation cohérente avec les phénomènes identifiés. En particulier, la méthode des *régressions segmentées* sera étudiée. Cette partie se conclura donc avec la section 3.3 sur une étude de la pertinence des hypothèses posées et de la modélisation proposée.

3.1 La structure de réassurance du portefeuille étudié

Dès lors que l'objectif de cette partie est d'intégrer l'impact de la réassurance à la modélisation proposée, il est nécessaire de comprendre comment une structure de réassurance peut modifier le profil de rendement et de risque du portefeuille étudié. L'objectif de cette section est donc de décrire de manière qualitative le type de phénomènes qu'il conviendra de modéliser.

C'est donc dans cette optique que la section 3.1.1 rappellera de manière générale les

raisons qui guideraient un assureur vers la réassurance, ainsi que les conséquences générales de la réassurance sur le portefeuille d'assurance.

Les différents types de traités de réassurance pourront ensuite être décrits en section 3.1.2, de manière à identifier ceux qui auront un impact lors de l'écriture du programme d'optimisation.

3.1.1 La réassurance comme composante essentielle du profil rendement / risque d'une entreprise d'assurance

Si la réassurance occupe une place si fondamentale au sein de la stratégie de l'assureur, c'est parce que c'est un outil important de la gestion des risques.

Le principe d'un traité de réassurance est en effet le suivant : le réassureur s'engage à prendre en charge certains risques encourus par le portefeuille d'assurance de l'assureur, tandis que l'assureur s'engage à verser une certaine quantité de primes au réassureur en contrepartie.

Pour ce faire, il est par exemple possible de céder les risques sur un portefeuille complet, d'identifier les risques couverts au préalable, de partager l'ensemble des primes et des risques, ou bien de ne céder que les risques dépassant un certain montant [7].

Ainsi, en cédant une partie bien définie de son risque en contrepartie de primes versées au réassureur, l'assureur est capable de modifier directement le profil de risque et de rendement de son portefeuille, et donc de le rendre plus adapté à son appétit au risque.

Les rôles de la réassurance au sein d'un portefeuille d'assurance sont donc les suivants [6] :

1. Maîtriser la probabilité de ruine et limiter l'impact d'un risque hors normes
Les sinistres particulièrement importants peuvent demander une capacité anormalement élevée, ce que la réassurance permet de compenser.

2. Lisser le résultat d'un portefeuille
Une sinistralité plus fréquente ou plus grave qu'à l'accoutumée peut entraîner des dérives du résultat de l'assureur pour une année donnée. La réassurance permet de mitiger cet effet, par exemple en prenant en charge les grands risques.

3. Diversifier les risques à une échelle mondiale
Les risques de type catastrophes dont la survenance peut immédiatement engendrer une grande quantité de sinistres peuvent être dangereux pour un assureur, car ils sont difficiles à diversifier au sein d'un portefeuille d'assurance. C'est pourquoi il peut être intéressant de les céder au réassureur, pour qui il est plus simple de les diversifier sur le plan géographique.

La structure de réassurance est donc centrale au sein de la gestion des risques d'un assureur, dans la mesure où elle agit à plusieurs niveaux et remplit plusieurs rôles.

Cela signifie notamment que les besoins en réassurance de différentes entreprises d'assurance ou différents portefeuilles peuvent varier de manière significative. C'est pourquoi la réassurance prend en réalité plusieurs formes, et peut avoir un impact complexe sur le profil de risque d'un portefeuille.

3.1.2 Les différents traités de réassurance et leur impact sur les sinistres qu'ils couvrent

Selon leur taille, leurs types de risque encourus, ou encore leur appétit au risque, les besoins en réassurance peuvent donc être très différents d'un portefeuille d'assurance à un autre. C'est pourquoi la réassurance n'existe pas sous la forme d'une formule unique, mais prend plutôt la forme de différents types de traités.

En particulier, la distinction est souvent faite entre la réassurance obligatoire et la réassurance facultative, ainsi qu'entre la réassurance proportionnelle et la réassurance non proportionnelle [7].

La distinction entre la réassurance facultative et la réassurance obligatoire est le type de risque cédé. Ainsi, dans le cas de la réassurance obligatoire, le traité s'applique automatiquement sur l'ensemble d'un portefeuille, et la cession des risques est obligatoire dès lors que les différentes modalités sont respectées. À l'inverse, la réassurance facultative ne porte que sur une police ou un bien identifiable, et les assureurs et réassureurs sont libres d'accepter ou refuser le traité.

Dans le cas de la réassurance proportionnelle ou non-proportionnelle, c'est la manière dont la cession est effectuée que la distinction est faite.

Dans le cas de la réassurance proportionnelle, la cession est basée sur le risque. Le partage du risque et des primes est notamment effectué selon un taux de cession, qui peut être fixe ou variable.

À l'inverse, dans le cas de la réassurance non-proportionnelle, la cession est fondée sur le sinistre. Là, la couverture offerte par le réassureur dépend des sinistres qui ont lieu, et est donc dissociée des primes cédées.

Les formes de réassurance principales qui seront explorées lors de cette étude sont donc les traités de type *Quote-part* et *Excédent de plein*, qui correspondent à de la réassurance proportionnelle, et ceux de type *Excédent de sinistre par risque*, *Excédent de sinistre par évènement* et *Stop loss*, qui constituent de la réassurance non-proportionnelle.

Les traités de type *Quote-Part*

Dans le cas d'un traité *Quote-Part*, le taux de cession qui détermine le montant de primes et de sinistres cédés est fixe. Soit $\alpha_i^{(1)}$ le taux de cession d'un traité de *Quote-Part* appliqué aux risques d'une ligne d'assurance $i \in 1, \dots, n$. En notant $(Sinistre_{i,k}^{brut}, Sinistre_{i,k}^{net})$ les montants bruts et nets d'un sinistre individuel $k \in 1, \dots, n_i$, et $(Prime_{i,k}^{brut}, Prime_{i,k}^{net})$ les montants de prime associé, l'égalité suivante est donc vérifiée :

$$\begin{aligned} Prime_{i,k}^{net} &= (1 - \alpha_i) \times Prime_{i,k}^{brut} \\ Sinistre_{i,k}^{net} &= (1 - \alpha_i) \times Sinistre_{i,k}^{brut} \end{aligned}$$

Les traités de type *Excédent de plein*

Comme mentionné précédemment, si le traité de type *Excédent de plein* correspond aussi

à de la réassurance proportionnelle, le taux de cession $\alpha_i^{(2)}$ n'est pas défini directement. Ici, c'est le montant de risque cédé qui est d'abord défini à l'aide des variables suivantes :

- La rétention R_i donne le montant d'exposition à partir duquel le risque est cédé
- La capacité K_i donne le risque maximal cédé à l'assureur

En notant C_k l'exposition au risque d'une police k , le taux de cession est donc de :

$$\alpha_{i,k}^{(2)} = \frac{\min(K_i, \max(0, C_k - R_i))}{C_k}$$

Ainsi, les primes et sinistres nets prennent les valeurs suivantes :

$$\begin{aligned} Prime_{i,k}^{net} &= (1 - \alpha_{i,k}^{(2)}) \times Prime_{i,k}^{brut} \\ Sinistre_{i,k}^{net} &= (1 - \alpha_{i,k}^{(2)}) \times Sinistre_{i,k}^{brut} \end{aligned}$$

Les traités de type *Excédent de sinistre par risque*

Les traités de type *Excédent de sinistre* correspondent à de la réassurance non-proportionnelle. Ce type de traité peut cependant être appliqué par risque ou par évènement. Lorsqu'un tel traité est appliqué par risque, les sinistres concernés sont évalués de manière individuelle afin de déterminer le montant de sinistre cédé ; à l'inverse des traités par évènement où tous les sinistres liés à un seul évènement sont considérés de manière conjointe.

Dans ce type de traités, le montant de la prime cédée $\beta_{i,k}^{(1)} \times Prime_{i,k}^{brut}$ peut être fixe ou calculé comme une fonction de la prime brute. Toutefois, il n'existe pas de relation de proportionnalité entre cette prime cédée et le montant de sinistre cédé, qui est défini à l'aide des variables suivantes :

- La Priorité P_i donne le montant de sinistre à partir duquel le risque est cédé
- La Couverture C_i donne le risque maximal cédé à l'assureur

Dans le cas de l'excédent de sinistre par risque, la priorité et la couverture sont appliquées aux sinistres concernés de manière individuelle. Pour le sinistre $k \in 1, \dots, n_i$, le montant de sinistre cédé est donc le suivant :

$$SinistreCede_{i,k} = \min(C_i, \max(Sinistre_{i,k}^{brut} - P_i, 0))$$

Ainsi, les primes et sinistres nets ne sont plus proportionnels et prennent les valeurs suivantes :

$$\begin{aligned} Prime_{i,k}^{net} &= (1 - \beta_{i,k}^{(1)}) \times Prime_{i,k}^{brut} \\ Sinistre_{i,k}^{net} &= Sinistre_{i,k}^{brut} - \min(C_i, \max(Sinistre_{i,k}^{brut} - P_i, 0)) \end{aligned}$$

Les traités de type *Excédent de sinistre par évènement*

Les traités de type *Excédent de sinistre par évènement* diffèrent donc des traités *Excédent de sinistre par risque* en la manière dont les différentes variables sont appliquées. L'application ne se fait en effet pas par sinistres individuels, mais en agrégeant différents les différents sinistres qui ont pour origine un seul et même évènement identifié. Leur objectif est donc souvent de se prévenir contre le risque de catastrophes.

En notant cette fois-ci A_j l'ensemble des lignes d'assurance couvertes par le même traité j , le montant de sinistres cédés est donc le suivant :

$$\text{SinistreCede}_{A_j} = \min \left(C_j, \max \left(\text{Sinistre}_{A_j}^{\text{brut}} - P_j, 0 \right) \right)$$

Les primes et sinistres nets ne sont donc aussi plus proportionnels et prennent les valeurs suivantes :

$$\begin{aligned} \text{Prime}_{A_j}^{\text{net}} &= (1 - \beta_{A_j}^{(1)}) \times \text{Prime}_{A_j}^{\text{brut}} \\ \text{Sinistre}_{A_j}^{\text{net}} &= \text{Sinistre}_{A_j}^{\text{brut}} - \min \left(C_j, \max \left(\text{Sinistre}_{A_j}^{\text{brut}} - P_j, 0 \right) \right) \end{aligned}$$

Les primes et sinistres nets liés à l'évènement j ne sont donc plus proportionnels, et prennent les valeurs suivantes :

$$\begin{aligned} \text{Prime}_{A_j}^{\text{net}} &= (1 - \beta_{A_j}^{(1)}) \times \text{Prime}_{A_j}^{\text{brut}} \\ \text{Sinistre}_{A_j}^{\text{net}} &= \text{Sinistre}_{A_j}^{\text{brut}} - \min \left(C_j, \max \left(\text{Sinistre}_{A_j}^{\text{brut}} - P_j, 0 \right) \right) \end{aligned}$$

Les traités de type *Stop loss*

Puisque les traités de types *Stop loss* correspondent aussi à des formes de réassurance non proportionnelle, leur fonctionnement est proche des traités de type *Excédent de sinistre*. Ces traités sont notamment définis par une priorité et une couverture.

Cependant, ce qui distingue le traité *Stop loss* est le fait que ce traité s'applique sur la totalité des sinistres encourus sur une période, et que le montant de cession est défini comme un taux en fonction de la prime. Son objectif est donc de limiter la perte totale encourue sur une période. La priorité et la couverture du traité appliqué sur la ligne $i \in \{1, \dots, n\}$ sont donc définies comme des taux de la prime totale perçue $\sum_{k=1}^{n_i} \text{Prime}_{i,k}^{\text{brut}}$.

Les taux de couverture et de priorité sont respectivement notés γ_i^C et γ_i^P . Dans ce cas, le traité d'applique sur le montant de la sinistralité totale $\sum_{k=1}^{n_i} \text{Sinistre}_{i,k}^{\text{brut}}$ compris entre $\gamma_i^P \times \sum_{k=1}^{n_i} \text{Prime}_{i,k}^{\text{brut}}$ et $(\gamma_i^P + \gamma_i^C) \times \sum_{k=1}^{n_i} \text{Prime}_{i,k}^{\text{brut}}$

Autrement dit, le montant de sinistres cédés est le suivant :

$$\text{SinistreCede}_i = \min \left(\gamma_i^C \times \sum_{k=1}^{n_i} \text{Prime}_{i,k}^{\text{brut}}, \max \left(\sum_{k=1}^{n_i} \text{Sinistre}_{i,k}^{\text{brut}} - \gamma_i^P \times \sum_{k=1}^{n_i} \text{Prime}_{i,k}^{\text{brut}}, 0 \right) \right)$$

Les primes et sinistres nets liés de la ligne i sont donc aussi décorrélés, et prennent les valeurs suivantes :

$$Prime_i^{net} = \sum_{k=1}^{n_i} (1 - \beta_{i,k}^{(3)}) \times Prime_{i,k}^{brut}$$

$$Sinistre_{i,(j)}^{net} = \sum_{k=1}^{n_i} Sinistre_{i,k}^{brut} - \min \left(\gamma_i^C \times \sum_{k=1}^{n_i} Prime_{i,k}^{brut}, \max \left(\sum_{k=1}^{n_i} Sinistre_{i,k}^{brut} - \gamma_i^P \times \sum_{k=1}^{n_i} Prime_{i,k}^{brut}, 0 \right) \right)$$

Il convient de noter qu'en pratique, les traités de réassurance comprennent souvent des clauses additionnelles telles la clause de franchise annuelle ou bien de reconstitution. Si ces clauses ont certes un impact sur les montants de primes et de sinistres cédés, ils ne modifient pas le principe de base du traité de réassurance.

Par simplicité, ces clauses seront donc négligées lors de la modélisation de l'impact de la réassurance sur les indicateurs de risque et de rendement du portefeuille.

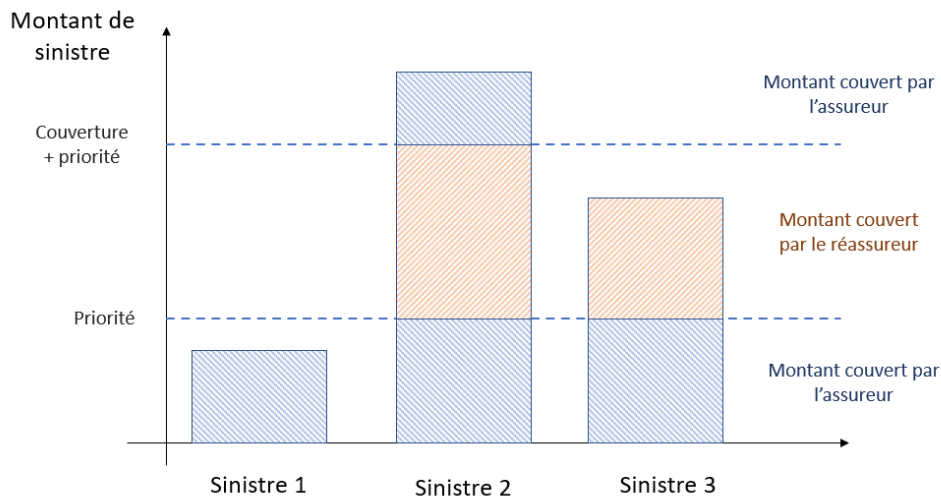


FIGURE 3.1 – Impact des traités de type *Excédent de sinistre* (source : [15])

Les traités de réassurance peuvent donc avoir des effets divers sur la sinistralité nette d'un portefeuille. Dès lors, il est nécessaire de réévaluer les hypothèses fixées en parties 2.2.3 et 2.3.4 afin de déterminer si elles restent valables pour un portefeuille net de réassurance.

3.1.3 La répercussion des traités de réassurance sur les hypothèses d'évolution

Lors du modèle développé en partie 2, il a été supposé que les sinistres et frais bruts évoluaient de manière linéaire en fonction de l'exposition au risque d'une ligne d'assurance. De même, il a été proposé de mesurer l'évolution de cette exposition au risque avec la

variation de la prime brute. Autrement dit, pour une ligne d'assurance $i \in 1, \dots, n$ les hypothèses fixées étaient les suivantes :

$$\left\{ \begin{array}{l} \delta_i = \frac{(Prime_i^{brut})_{t_1}}{(Prime_i^{brut})_{t_0}} \\ Prime_i^{brut}(\delta_i) = \delta_i \times Prime_i^{brut}(1) \\ Frais_i^{brut}(\delta_i) = \delta_i \times Frais_i^{brut}(1) \\ Sinistre_i^{brut}(\delta_i) = \delta_i \times Sinistre_i^{brut}(1) \end{array} \right. \quad (3.1)$$

Une méthode naïve qui permettrait d'obtenir de telles formules d'évolution nettes de réassurance serait de supposer l'existence d'un facteur de réassurance par ligne α_i qui soit fixe et proportionnel :

$$\left\{ \begin{array}{l} Prime_i^{net}(\delta_i) = \delta_i \times \alpha_i \times Prime_i^{brut}(1) \\ Frais_i^{net}(\delta_i) = \delta_i \times \alpha_i \times Frais_i^{brut}(1) \\ Sinistre_i^{net}(\delta_i) = \delta_i \times \alpha_i \times Sinistre_i^{brut}(1) \end{array} \right. \quad (3.2)$$

Ce qui permettrait de raisonner de manière analogue à la modélisation brute et de poser les hypothèses suivantes :

$$\left\{ \begin{array}{l} Prime_i^{net}(\delta_i) = \delta_i \times Prime_i^{net}(1) \\ Frais_i^{net}(\delta_i) = \delta_i \times Frais_i^{net}(1) \\ Sinistre_i^{net}(\delta_i) = \delta_i \times Sinistre_i^{net}(1) \end{array} \right. \quad (3.3)$$

Toutefois, une telle modélisation suppose en réalité que l'hypothèse suivante soit vérifiée :

$$\frac{Prime_i^{net}}{Prime_i^{brut}} = \frac{Frais_i^{net}}{Frais_i^{brut}} = \frac{Sinistre_i^{net}}{Sinistre_i^{brut}}$$

Or, l'étude des différents traités de réassurance effectuée ci-dessus (3.1.2) montre que ces relations de linéarité ne sont pas vraies en général. En particulier, il n'existe pas nécessairement de relation de proportionnalité entre les primes et les sinistres, notamment dans le cas de la réassurance non-proportionnelle.

Ainsi, il est nécessaire de passer par une étude du passage des observations brutes de réassurance aux observations nettes de réassurance afin de modéliser le programme d'optimisation net.

De manière générale, l'impact de la réassurance sur le portefeuille avant application des facteurs d'évolution δ peut être modélisé de la façon suivante :

Soit $i \in 1, \dots, n$ une ligne d'assurance du portefeuille, $s \in 1, \dots, S$ une simulation, et n_i^{Prime} , n_i^{Frais} et $n_i^{Sinistre}$ respectivement le nombre de versements de primes, de paiement de frais et de sinistres associés à cette ligne*. Alors, en notant f_i^{Prime} , f_i^{Frais} et $f_i^{Sinistre}$ l'impact de la structure de réassurance sur les primes, frais et sinistres individuels :

*. En pratique, il peut être difficile de faire correspondre les frais aux lignes en raison de l'existence de frais fixes. Ceux-ci sont ignorés ici, car ils n'ont que peu d'impact sur le portefeuille.

$$\begin{aligned} (Prime_i^{net})_s &= \left(\sum_{k=1}^{n_i^{Prime}} f_i^{Prime}(Prime_{i,k}^{brut}) \right)_s \\ (Frais_i^{net})_s &= \left(\sum_{k=1}^{n_i^{Frais}} f_i^{Frais}(Frais_{i,k}^{brut}) \right)_s \\ (Sinistre_i^{net})_s &= \left(\sum_{k=1}^{n_i^{Sinistre}} f_i^{sinistres}(Sinistre_{i,k}^{brut}) \right)_s \end{aligned}$$

Ainsi, sous les hypothèses d'évolution des variables brutes, chaque simulation des variables nettes de réassurance après application de la stratégie δ correspondrait aux équations suivantes :

$$\begin{aligned} (Prime_i^{net}(\delta_i))_s &= \left(\sum_{k=1}^{n_i^{Prime}} f_i^{Prime}(\delta_i \times Prime_{i,k}^{brut}) \right)_s \\ (Frais_i^{net}(\delta_i))_s &= \left(\sum_{k=1}^{n_i^{Frais}} f_i^{Frais}(\delta_i \times Frais_{i,k}^{brut}) \right)_s \\ (Sinistre_i^{net}(\delta_i))_s &= \left(\sum_{k=1}^{n_i^{Sinistre}} f_i^{sinistres}(\delta_i \times Sinistre_{i,k}^{brut}) \right)_s \end{aligned}$$

Toutefois, comme décrit précédemment en partie 1.4.2, la base de données utilisée lors de cette étude ne comprend que des valeurs agrégées au niveau de la ligne d'assurance, ce qui signifie que les valeurs précises de ces quantités ne peuvent pas être calculées avec les données disponibles au sein de la base de données utilisée lors de cette étude :

- les traités de réassurance appliqués aux sinistres individuels ne sont pas compris dans la base de données utilisée.
Les fonctions f_i^{Prime} , f_i^{Frais} et $f_i^{Sinistre}$ ne sont donc pas connues.
- Le nombre de versements de primes, le nombre de paiements de frais ou survenance des sinistres n_i^{Prime} , n_i^{Frais} et $n_i^{Sinistre}$ ne sont pas connus.
- Les simulations qui composent de la base de données ne sont retenues qu'après l'agrégation par ligne d'assurance.
Les simulations individuelles $Prime_{i,k}^{brut}$, $Frais_{i,k}^{brut}$ et $Sinistre_{i,k}^{brut}$ ne sont donc pas observées.

L'impact de la réassurance ne peut donc qu'être quantifié au niveau de la ligne d'assurance par la modélisation suivante :

$$\begin{aligned}
(Prime_i^{net}(\delta_i))_s &= \hat{f}_i^{Prime} (\delta_i \times Prime_i^{brut}(1))_s + (\epsilon_i^{Prime}(\delta_i))_s \\
(Frais_i^{net}(\delta_i))_s &= \hat{f}_i^{Frais} (\delta_i \times Frais_i^{brut}(1))_s + (\epsilon_i^{Frais}(\delta_i))_s \\
(Sinistre_i^{net}(\delta_i))_s &= \hat{f}_i^{Sinistre} (\delta_i \times Sinistre_i^{brut}(1))_s + (\epsilon_i^{Sinistre}(\delta_i))_s
\end{aligned}$$

Dès lors, l'objectif de cette partie est de développer une méthode qui permette de fournir des estimateurs cohérents des fonctions de réassurance \hat{f}_i^{Prime} , \hat{f}_i^{Frais} et $\hat{f}_i^{Sinistre}$.

3.2 Modéliser l'impact de la réassurance

L'étude menée jusqu'ici au cours de cette partie a non seulement démontré la nécessité de modéliser le programme d'optimisation de l'assureur net de réassurance, mais a aussi mis en avant les difficultés liées à cette modélisation. Les objectifs de cette section sont donc de déterminer une manière d'approcher les fonctions de réassurance de manière à prendre en compte les différents types de traités.

3.2.1 Modéliser les différentes fonctions de Réassurance

Afin d'intégrer l'impact de la réassurance au programme d'optimisation, la première étape est donc d'identifier une modélisation cohérente des fonctions de réassurance identifiées en partie 3.1.3. Autrement dit, il est question de déterminer la formes des fonctions \hat{f}_i^{Prime} , \hat{f}_i^{Frais} et $\hat{f}_i^{Sinistre}$ au sein des équations suivantes :

$$\begin{aligned}
(Prime_i^{net}(\delta_i))_s &= \hat{f}_i^{Prime} (\delta_i \times Prime_i^{brut}(1))_s + (\epsilon_i^{Prime}(\delta_i))_s \\
(Frais_i^{net}(\delta_i))_s &= \hat{f}_i^{Frais} (\delta_i \times Frais_i^{brut}(1))_s + (\epsilon_i^{Frais}(\delta_i))_s \\
(Sinistre_i^{net}(\delta_i))_s &= \hat{f}_i^{Sinistre} (\delta_i \times Sinistre_i^{brut}(1))_s + (\epsilon_i^{Sinistre}(\delta_i))_s
\end{aligned}$$

Or, l'analyse de l'application des différents traités de réassurance a permis de constater le fait que l'impact de la réassurance sur la prime était différent de son impact sur les sinistres. Il conviendra donc de procéder par étapes, et d'évaluer une à une les formules correspondant à chacune de ces fonctions.

Par souci de simplicité, seuls les traités de type *Quote-Share* et *Excédents de sinistres* seront considérés lors de cette analyse. Si cette restriction réduit certes l'exactitude de la modélisation présentée ici, l'analyse statistique de la méthode utilisée permet tout de même de démontrer la pertinence de cette approche (voir partie 3.3)

La fonction de réassurance des Frais

En réalité, contrairement aux primes ou aux sinistres, les frais de l'assureur ne sont pas directement cédés au réassureur. Toutefois, pour compenser les coûts encourus par l'assureur dès lors qu'il est nécessaire de souscrire de nouveaux contrats ou de faire appel à un courtier,

le réassureur verse une commission à la cédante. Ici, la notion de frais nets de réassurance désignera donc la différence entre les frais de l'assureur et cette commission.

Celle-ci est définie de manière générale à l'aide d'un taux appliqué sur la prime brute. En pratique, ce taux de commission, noté α_i^F , est le fruit de négociations entre le réassureur et la cédante, et dépend surtout du profil de risque de la ligne d'assurance. Il est donc considéré comme indépendant du facteur d'évolution δ , ce qui a la conséquence suivante sur les frais nets :

$$\begin{aligned} Frais_i^{net}(\delta_i) &= \sum_{k=1}^{n_i^{Frais}} (1 - \alpha_i^F) \times \delta_i \times Primes_{i,k}^{brut}(1) \\ &= \delta_i \times \sum_{k=1}^{n_i^{Frais}} (1 - \alpha_i^F) \times Primes_{i,k}^{brut}(1) \\ &= \delta_i Frais_i^{net}(1) \end{aligned}$$

Ainsi, tout comme pour les frais bruts, l'hypothèse de linéarité en fonction du facteur δ sera conservée pour les frais nets.

Vient ensuite la question des fonctions de réassurance des sinistres. Dans la mesure où c'est ici que différents traités de réassurance auront un impact différent selon le type de sinistre, il est pertinent de distinguer les sinistres attritionnels, les sinistres atypiques et les sinistres CAT.

La fonction de réassurance des Sinistres Attritionnels

Par définition, les sinistres attritionnels correspondent aux sinistres dont le montant est relativement faible de manière individuelle, mais dont la fréquence peut être élevée. Sur le plan de la réassurance, la conséquence de cette définition est que les sinistres attritionnels sont exclusivement touchés par les traités de type *Quote-Share* ; ils sont en effet trop faibles pour activer les traités en excédents de sinistres. En notant α_i le taux de cession du traité et n_i^{ATT} le nombre de sinistres attritionnels associés à la ligne $i \in 1, \dots, n$, le montant de sinistres cédés est donc le suivant :

$$SinistreCede^{ATT}(\delta_i) = \sum_{k=1}^{n_i^{ATT}} \alpha_i \times Sinistre_{i,k}^{ATT,brut}(1)$$

La modélisation de la sinistralité nette attritionnelle se fait donc de la manière suivante :

$$\begin{aligned}
Sinistre_i^{ATT,net}(\delta_i) &= \delta_i \times \sum_{k=1}^{n_i^{ATT}} (1 - \alpha_i) \times Sinistre_{i,k}^{ATT,brut}(1) \\
&= \delta_i \times \sum_{k=1}^{n_i^{ATT}} (1 - \alpha_i) \times Sinistre_{i,k}^{ATT,brut}(1) \\
&= \delta_i \times Sinistre_i^{ATT,net}(1)
\end{aligned}$$

L'hypothèse de linéarité en fonction du facteur d'évolution δ peut donc être conservée pour les sinistres attritionnels.

La fonction de réassurance des Sinistres Atypiques

Le cas des sinistres atypiques correspond aux sinistres qui sont relativement peu fréquents, mais dont le montant dépasse un certain seuil et dont la cause ne rentre pas dans la liste des évènements identifiés comme des catastrophes.

Cela signifie que les types de traités qui ont un impact sont ici les suivants :

- *Quote-part*
- *Excédent de pertes par sinistre*

Si l'impact du traité de type *Quote-part* est identique à la situation des sinistres attritionnels, la conséquence des traités de type *Excédent de pertes par sinistre* est plus complexe. En effet, en notant $k \in 1, \dots, n_i^{ATY}$ un des sinistres atypiques de la ligne i , P_i la priorité du traité est C_i sa couverture, l'impact de ce traité diffère selon le montant de sinistre atteint :

$$\begin{aligned}
SinistreCede_{i,k}^{ATY,ExcSns} &= \left(Sinistre_{i,k}^{ATY,brut} - P_i \right) \mathbf{1}_{P_i \leq Sinistre_{i,k}^{ATY,brut} \leq (P_i + C_i)} + \\
&\quad C_i \mathbf{1}_{Sinistre_{i,k}^{ATY,brut} \geq (P_i + C_i)}
\end{aligned}$$

Une simulation de la sinistralité atypique nette au sein du portefeuille correspond donc à une observation de la variable aléatoire suivante :

$$\begin{aligned}
Sinistre_i^{ATY,Net} &= (1 - \alpha_i) \sum_{k=1}^{n_i^{ATY}} \left(Sinistre_{i,k}^{ATY,brut} - P_i \right) \mathbf{1}_{P_i \leq Sinistre_{i,k}^{ATY,brut} \leq (P_i + C_i)} + \\
&\quad C_i \mathbf{1}_{Sinistre_{i,k}^{ATY,brut} \geq (P_i + C_i)}
\end{aligned}$$

Cela permet notamment de confirmer comme fausse l'hypothèse selon laquelle il existerait un réel $a \in \mathbb{R}$ tel que :

$$\begin{aligned}
Sinistre_i^{ATY,Net}(\delta_i) &= a \times \delta_i \times Sinistre_i^{ATY,Brut}(1) \text{ ou bien} \\
Sinistre_i^{ATY,Net}(\delta_i) &= a \times \delta_i \times Sinistre_i^{ATY,Net}(1)
\end{aligned}$$

Il est donc *a priori* impossible de modéliser le lien entre les sinistres atypiques nets et les sinistres atypiques bruts avec une relation linéaire simple de manière exacte. Dès lors, l'enjeu

est de proposer une modélisation qui permette d'estimer les phénomènes de seuil identifiés et de minimiser la quantité ϵ_i^{ATY} .

C'est à pourquoi la méthode proposée considère une relation linéaire segmentée pour passer de la sinistralité brute à la sinistralité nette.

L'objectif de cette modélisation est de faire la distinction entre différents groupes de simulations $s \in \{1, \dots, S\}$, et notamment entre :

- les simulations s "faibles" de $Sinistres_i^{ATY}$, pour lesquelles l'hypothèse selon laquelle la plupart des sinistres n'ont pas atteint la priorité est raisonnable
- les simulations s "intermédiaires" de $Sinistres_i^{ATY}$, où un grand nombre de sinistres ont probablement dépassé la priorité
- les simulations s "élevées" de $Sinistres_i^{ATY}$, où la couverture a été atteinte pour un certain nombre de sinistres

Cette distinction prend la forme d'une fonction de réassurance atypique linéaire par morceaux. Autrement dit, avec x_i une valeur de sinistre atypique potentielle, la fonction f_i^{ATY} est de la forme :

$$f_i^{ATY}(x_i) = \beta_i^{ATY,(0)} + \beta_i^{ATY,(1)} x_i + \beta_i^{ATY,(2)} x_i \mathbf{1}_{x_i \geq x_i^{ATY,(2)}} + \beta_i^{ATY,(3)} x_i \mathbf{1}_{x_i \geq x_i^{ATY,(3)}}$$

La s -ième observation de la sinistralité atypique nette de la ligne i s'écrit donc :

$$\left(Sinistre_i^{net}(\delta_i) \right)_s = f_i^{ATY} \left(\delta_i \times Sinistre_i^{brut}(1) \right)_s + \left(\epsilon_i^{Sinistre}(\delta_i) \right)_s$$

La fonction de réassurance des Sinistres CAT

Sur le plan de la réassurance, ce qui distingue les sinistres atypiques des sinistres CAT est le fait que les sinistres CAT sont engendrés par une catastrophe naturelle ou causée par l'homme. Cette distinction est d'une grande importance, car cela signifie que les sinistres CAT sont soumis à un traité de réassurance de type *Excédent de sinistre par évènement* en plus des traités de type *Quote-part* et *Excédent de sinistre par risque*.

En outre, là où les traités d'excédent par risque permettent une analyse ligne d'assurance par ligne d'assurance, les évènements CAT peuvent justement toucher plusieurs lignes d'assurance à la fois. C'est pourquoi certains traités *Excédent de sinistre par évènement* peuvent conjointement plusieurs lignes d'assurance à la fois.

Afin de prendre cette nouvelle dimension, un procédé similaire à celui effectué pour les sinistres atypiques est proposé. Seulement, plutôt que de n'effectuer la régression segmentée que sur les sinistres de la ligne correspondante, la régression sera aussi effectuée sur la

somme des sinistres des lignes d'assurance couvertes par le même traité de réassurance.

Ainsi, en notant $T(i)$ l'ensemble des lignes d'assurance qui appartiennent au même traité de réassurance que la ligne i , la modélisation suivante est obtenue :

soit $x_{T(i)} = \sum_{j \in T(i)} x_j$ une somme de sinistres fictifs. Alors :

$$f_i^{CAT}(x_i, x_{T(i)}) = \beta_i^{CAT,(0)} + \beta_i^{CAT,(1)} x_i + \beta_i^{CAT,(2)} x_i \mathbf{1}_{x_i \geq x_i^{CAT,(2)}} + \beta_i^{CAT,(3)} x_i \mathbf{1}_{x_i \geq x_i^{CAT,(3)}} + \gamma_i^{CAT,(1)} x_{T(i)} + \gamma_i^{CAT,(2)} x_{T(i)} \mathbf{1}_{x_{T(i)} \geq x_{T(i)}^{CAT,(2)}} + \gamma_i^{CAT,(3)} x_{T(i)} \mathbf{1}_{x_{T(i)} \geq x_{T(i)}^{CAT,(3)}}$$

Cette double régression segmentée permet donc de prendre en compte non seulement l'impact du traité de type *Excédent de sinistre par évènement*, mais aussi ceux de type *Quote-part* et *Excédent de sinistre par risque*.

La fonction de réassurance de primes

Contrairement à la situation brute de réassurance, l'hypothèse selon laquelle les primes nettes évolueraient de façon linéaire avec l'exposition au risque est erronée. Cela est dû à l'existence de traités non-proportionnels.

Le montant de prime cédée est en effet le fruit de négociations entre l'assureur et le réassureur. Or, dans le cas d'une augmentation de l'exposition au risque sur la ligne i , le nombre de sinistres atteignant la priorité d'un traité de type *excédent de sinistre* sera plus élevée. Si $\delta_i > 1$, L'impact cumulé des traités proportionnels et non-proportionnels sur les sinistres de la ligne i est donc le suivant :

$$\frac{SinistreCede_i^{net}(\delta_i)}{SinistreCede_i^{net}(1)} > \delta_i$$

Dès lors, une augmentation de la prime cédée serait nécessaire afin de maintenir le traité de réassurance. C'est pourquoi il est considéré que la prime nette n'évolue pas de façon linéaire, mais s'ajuste en fonction de la sinistralité cédée :

$$Prime_i^{net}(\delta_i) = \frac{\mathbb{E}[SinistreCede_i^{net}(\delta_i)]}{\mathbb{E}[SinistreCede_i^{net}(1)]} \times Prime_i^{net}(1)$$

L'estimation des fonctions de réassurance

Ainsi, les fonctions de réassurance théoriques des primes, des frais, et de tous les types de sinistres correspondant aux risques de souscription ont été définies. Dès lors, l'enjeu principal de la modélisation de la réassurance est l'estimation de ces fonctions. L'effet de la réassurance sur les sinistres atypiques et CAT en particulier a été théorisé comme étant linéaire par morceau ; il convient donc de déterminer une méthode qui permette d'estimer les points de changements, c'est-à-dire les points où la pente de la fonction de réassurance change.

C'est pourquoi l'estimation de ces fonctions a été effectuée par la méthode de la régression linéaire segmentée, théorisée par Muggeo en 2003 [21].

Soit Y la variable d'intérêt et X une variable explicative. Cette méthode permet en effet de déterminer en une fois non seulement la localisation des points de changement, mais aussi les paramètres linéaires de la régression de Y sur X (c.f. annexe B).

Dans le cas des sinistres atypiques, avec $i \in 1, \dots, n$ une ligne d'assurance, il suffit donc de considérer :

- $Y = \text{Sinistre}_i^{ATY,net}$
- $X = \text{Sinistre}_i^{ATY,brut}$

Tandis que dans le cas des sinistres CAT, en notant aussi $\text{Sinistre}_{T(i)}^{CAT,brut}$ la somme des sinistres CAT qui relèvent du même traité de type *Excédent de Sinistre par évènement* que la ligne i , il suffit de considérer :

- $Y = \text{Sinistre}_i^{CAT,net}$
- $X = \left(\text{Sinistre}_i^{CAT,brut}, \text{Sinistre}_{T(i)}^{CAT,brut} \right)$

Cela permet donc d'obtenir les fonctions f_i^{ATY} et f_i^{CAT} lors de la modélisation suivante :

$$\begin{aligned} \text{Sinistre}_i^{ATY,net}(\delta_i) &= f_i^{ATY} \left(\delta_i \times \text{Sinistre}_i^{ATY,brut}(1) \right) + \epsilon_i^{ATY, \text{Sinistre}}(\delta_i) \\ \text{Sinistre}_i^{CAT,net}(\delta_i) &= f_i^{CAT} \left(\delta_i \times \text{Sinistre}_i^{CAT,brut}(1), \sum_{k \in T(i), k \neq i} \delta_k \text{Sinistre}_k^{CAT,brut}(1) \right) \\ &\quad + \epsilon_i^{CAT, \text{Sinistre}}(\delta_i) \end{aligned}$$

La dernière étape de la modélisation théorique consiste donc à étudier la question des résidus $\epsilon_i^{ATY, \text{Sinistre}}(\delta_i)$ et $\epsilon_i^{CAT, \text{Sinistre}}(\delta_i)$.

3.2.2 Intégrer l'impact des résidus

L'objectif final de cette partie est de proposer une modélisation qui permette d'intégrer l'impact de la réassurance dans le programme d'optimisation de l'assureur défini en partie 2. En particulier, il est nécessaire de reconstruire un vecteur de S observations des variables de primes, de frais et de sinistres nettes de réassurance après application des facteurs δ .

Dans le cas des sinistres atypiques et CAT, une fois les fonctions de réassurance estimées, les composantes du vecteur de sinistralité nette connues sont les suivantes :

- f_i^{ATY} et f_i^{CAT} les fonctions de réassurance des sinistres atypiques et CAT
- $\text{Sinistre}_i^{brut}(\delta_i)$ le vecteur d'observations de sinistralité brute après application des facteurs d'évolution δ

Il ne reste donc qu'à estimer les résidus après application des facteurs δ , c'est-à-dire les variables $\epsilon_i^{CAT}(\delta_i)$ et $\epsilon_i^{ATY}(\delta_i)$.

Or, si les résidus après application de δ sont inconnus, il est trivial de déterminer le vecteur des résidus lorsque $\delta_i = 1$. En effet, pour une simulation $s \in 1, \dots, S$, les variables de résidus sont définies de la manière suivante :

$$\begin{aligned}\epsilon_i^{ATY}(1)_s &= \left(\text{Sinistre}_i^{ATY,net}(1) \right)_s - f_i^{ATY} \left(\text{Sinistre}_i^{ATY,brut}(1)_s \right) \\ \epsilon_i^{CAT}(1)_s &= \text{Sinistre}_i^{CAT,net}(1)_s - f_i^{CAT} \left(\text{Sinistre}_i^{CAT,brut}(1)_s, \sum_{k \in T(i)} \text{Sinistre}_k^{CAT,brut}(1)_s \right)\end{aligned}$$

La méthode retenue consiste à poser une condition supplémentaire sur la relation entre les résidus après application de la stratégie δ . Ainsi, la relation suivante est supposée :

$$\begin{aligned}\epsilon_i^{ATY}(\delta_i)_s &= \epsilon_i^{ATY}(1)_s + \nu^{ATY}(\delta_i)_s \\ \epsilon_i^{CAT}(\delta_i)_s &= \epsilon_i^{CAT}(1)_s + \nu^{CAT}(\delta_i)_s \\ &\text{avec :} \\ \mathbb{E}[\nu^{ATY}(\delta_i)] &= 0 \\ \mathbb{E}[\nu^{CAT}(\delta_i)] &= 0\end{aligned}$$

Dès lors, il suffit de considérer les estimateurs de résidus suivants :

$$\begin{aligned}\epsilon_i^{AT\hat{Y}}(\delta_i) &= \epsilon_i^{ATY}(1) \\ \epsilon_i^{CA\hat{T}}(\delta_i) &= \epsilon_i^{CAT}(1)\end{aligned}$$

Et il est donc possible de fournir une estimation de la totalité du vecteur de sinistralité nette en considérant l'estimateur suivant pour chaque simulation $s \in 1, \dots, S$:

$$\begin{aligned}\text{Sinistre}_i^{ATY,net}(\delta_i)_s &= f_i^{ATY} \left(\delta_i \times \text{Sinistre}_i^{ATY,brut}(1)_s \right) + \epsilon_i^{ATY}(1)_s \\ \text{Sinistre}_i^{CAT,net}(\delta_i)_s &= f_i^{CAT} \left(\delta_i \times \text{Sinistre}_i^{CAT,brut}(1)_s, \sum_{k \in T(i)} \delta_k \text{Sinistre}_k^{CAT,brut}(1)_s \right) + \epsilon_i^{CAT}(1)_s\end{aligned}$$

Intuitivement, cette approche revient en réalité à obtenir le nouveau vecteur en traduisant chaque point le long de la tendance identifiée à l'aide des fonctions de réassurance (c.f. Figure 3.2).

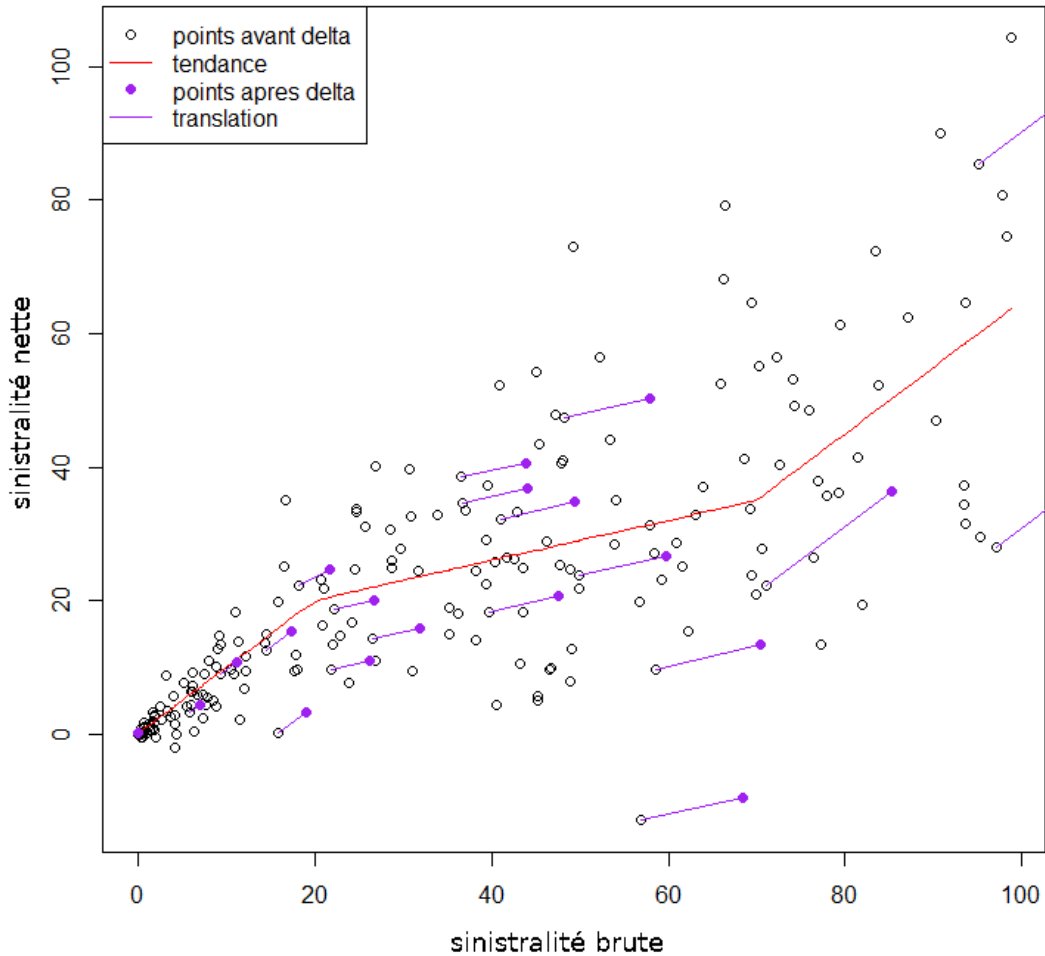


FIGURE 3.2 – Représentation de la relation entre $\epsilon_i(\delta_i)$ et $\epsilon_i(1)$

Une limite de cette modélisation est toutefois le fait que la variance associée aux termes $\nu_i^{ATY}(\delta_i)$ et $\nu_i^{CAT}(\delta_i)$ est négligée. Il serait possible de prendre cette variance en compte, par exemple en ajoutant un facteur au résidu $\epsilon_i(1)$ en fonction de la valeur de δ_i :

$$\begin{aligned}\epsilon_i^{AT\hat{Y}}(\delta_i) &= \epsilon_i^{ATY}(1) \times a \times \delta_i \\ \epsilon_i^{C\hat{A}T}(\delta_i) &= \epsilon_i^{CAT}(1) \times a \times \delta_i\end{aligned}$$

Il apparaît en effet par observation graphique que la dispersion autour de la tendance est d'autant plus forte que la valeur du sinistre brute est élevée. La propriété d'homoscédasticité n'est donc pas respectée. Un test aurait notamment permis de s'en assurer, et une estimation du paramètre a aurait pu être effectuée.

Ainsi, s'il est relativement simple de prendre en compte l'impact de la réassurance pour les primes, les frais et les sinistres attritionnels, il est nécessaire de complexifier le modèle dès lors qu'il existe des sinistres atypiques ou CAT.

Afin de s'assurer de la pertinence de cette approche, il est donc nécessaire de l'appliquer sur

la base de donnée étudiée, et de déterminer si les résultats sont cohérents avec les hypothèses posées.

3.3 Analyse de la pertinence de la modélisation choisie

L'analyse menée au cours de cette partie a donc permis de déterminer une modélisation qui permet de tenir compte de l'impact de la réassurance sur le programme d'optimisation en fonction des facteurs δ définis en partie 2.

Cette modélisation a toutefois requis l'ajout de certaines hypothèses. Dès lors, il convient d'appliquer la méthode proposée à la base de données étudiée afin de déterminer l'adéquation des hypothèses posées. Cette vérification se fera en deux temps :

Tout d'abord, les hypothèses posées seront vérifiées par analyse graphique en section 3.3.1. Ensuite, la pertinence des fonctions de réassurance sera examinée à l'aide d'un test d'adéquation en section 3.3.2.

3.3.1 Analyse graphique des hypothèses posées

Lors de cette modélisation de l'impact de la réassurance, les hypothèses les plus importantes relèvent des formes des fonctions de réassurance choisies. C'est pourquoi ce processus de vérification commencera par une comparaison des vecteurs d'observations bruts avec les vecteurs d'observation nets.

En particulier, les variables d'intérêt sont ici les différents types de sinistres d'un échantillon de lignes d'assurance, choisies de sorte à refléter l'ensemble des phénomènes observés au sein de la base de données.

Sinistres Attritionnels

Dans le cas des sinistres attritionnels, la sinistralité nette est supposée comme prenant la forme suivante :

$$\begin{aligned} \text{Sinistres}_i^{ATT,net}(\delta_i) &= \sum_{k=1}^{n_i^{ATT}} (1 - \alpha_i) \times \delta_i \times \text{Sinistres}_{i,k}^{ATT,brut}(1) \\ &= (1 - \alpha_i) \times \text{Sinistres}_i^{ATT,brut}(\delta_i) \end{aligned}$$

La fonction de réassurance attritionnelle f_i^{ATT} est donc la suivante :

$$\begin{aligned} f_i^{ATT} : \mathbb{R} &\longrightarrow \mathbb{R} \\ x &\longmapsto (1 - \alpha_i) \times x \end{aligned}$$

Cela est dû au fait que, parmi les types de traités pris en compte, seule les traités de type *Quote-part* s'appliquent à ce type de sinistre.

De fait, l'analyse graphique des sinistres nets rapportés aux sinistres bruts réalisée en figure 3.3 montre la pertinence de cette hypothèse. La relation de linéarité notamment est

parfaitement vérifiée dans le cas de la *ligne 1*, la *ligne 2* et la *ligne 4*.

Toutefois, si le cas de la *ligne 3* semble s'approcher d'une relation linéaire, il est toute de même possible d'observer un certain nombre de résidus non-nuls. C'est donc en cette disparité qu'il est possible de constater l'impact de la décision de négliger l'impact des traités de type *Stop-loss*, qui ont un impact non proportionnel, mais couvrent les sinistres attritionnels.

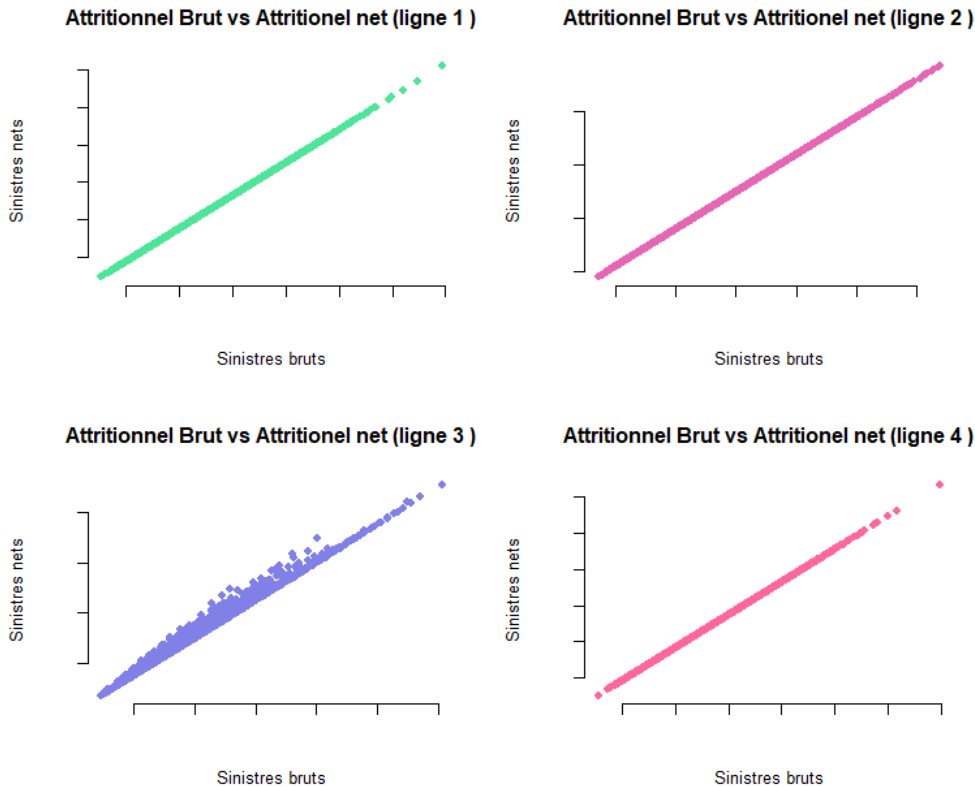


FIGURE 3.3 – Sinistres ATT nets rapportés aux ATT bruts

Ainsi, une modélisation plus précise serait la suivante :

$$\text{Sinistre}_i^{ATT,net}(\delta_i)_s = (1 - \alpha_i) \times \text{Sinistre}_i^{ATT,brut}(\delta_i)_s + \epsilon_i^{ATT}(\delta_i)_s$$

Il serait donc possible de prendre en compte l'impact de ce résidu non-nul en utilisant un procédé analogue à celui des sinistres atypiques et CAT (c.f. section 3.2.2). En pratique, l'impact de cette imprécision sur les résultats de cette modélisation est cependant minime.

Sinistres Atypiques

Dans le cas des sinistres atypiques, l'analyse graphique réalisée en [figure 3.4](#) révèle que l'hypothèse de linéarité serait vérifiée dans le cas de certaines lignes. Celles-ci correspondent en réalité aux lignes non couvertes par des traités de type *Excédent de sinistre*.

Toutefois, les résidus sont souvent bien plus significatifs ici. C'est le phénomène que montre notamment la ligne *ligne 2*.

En outre, la *ligne 3* est un exemple de ligne d'assurance pour laquelle l'hypothèse de linéarité est manifestement erronée. À l'inverse, la relation changeante entre les sinistres bruts et nets est manifestement bien prise en compte par la régression segmentée effectuée. Il convient en outre de constater la dispersion du nuage de point. Cela montre en effet que négliger les résidus au sein de l'analyse risquerait de sous-estimer la variance présente au sein des différents vecteurs d'observation.

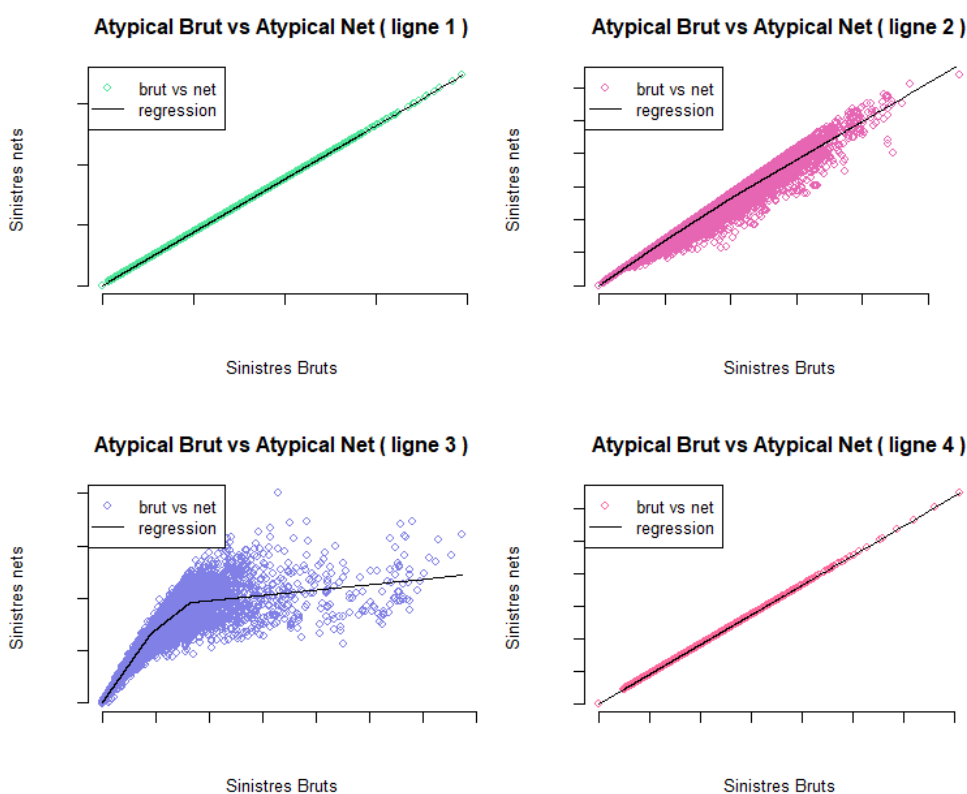


FIGURE 3.4 – Sinistres ATY nets rapportés aux ATY bruts

Sinistres CAT

Enfin, la figure 3.5 révèle que les sinistres CAT ont la distinction d'être les plus volatiles des trois types de sinistres. Cela s'explique notamment par l'impact que peuvent avoir les sinistres des autres lignes d'assurances.

Ici, la pertinence des régressions segmentées est d'autant plus apparente, notamment en regardant la tendance identifiée dans le cas de la *ligne 1* et la *ligne 2*. Cependant, l'estimation peut être imprécise pour les extrêmes lorsque très peu de simulations sont largement supérieures aux autres.

Il convient aussi de remarquer l'impact de l'intégration des sinistres de l'ensemble des lignes du traité au sein de la régression. C'est en effet cette partie de la modélisation qui est

mise en évidence par la volatilité de la courbe de régression au sein de la *ligne 1* et la *ligne 2*.

C'est pourquoi la *ligne 3* et la *ligne 4* ne présentent pas de phénomène similaire : aucune autre ligne d'assurance n'était couverte par le même traité de type *Excédent de sinistre par évènement*.

Cependant, l'étude de la *ligne 3* révèle qu'une dispersion trop importante du nuage de point diminue la précision de la régression segmentée. Là, différentes tendances n'ont pas pu être clairement identifiées.

C'est pourquoi la prise en compte des résidus est tout aussi cruciale pour les sinistres CAT que les sinistres atypiques.

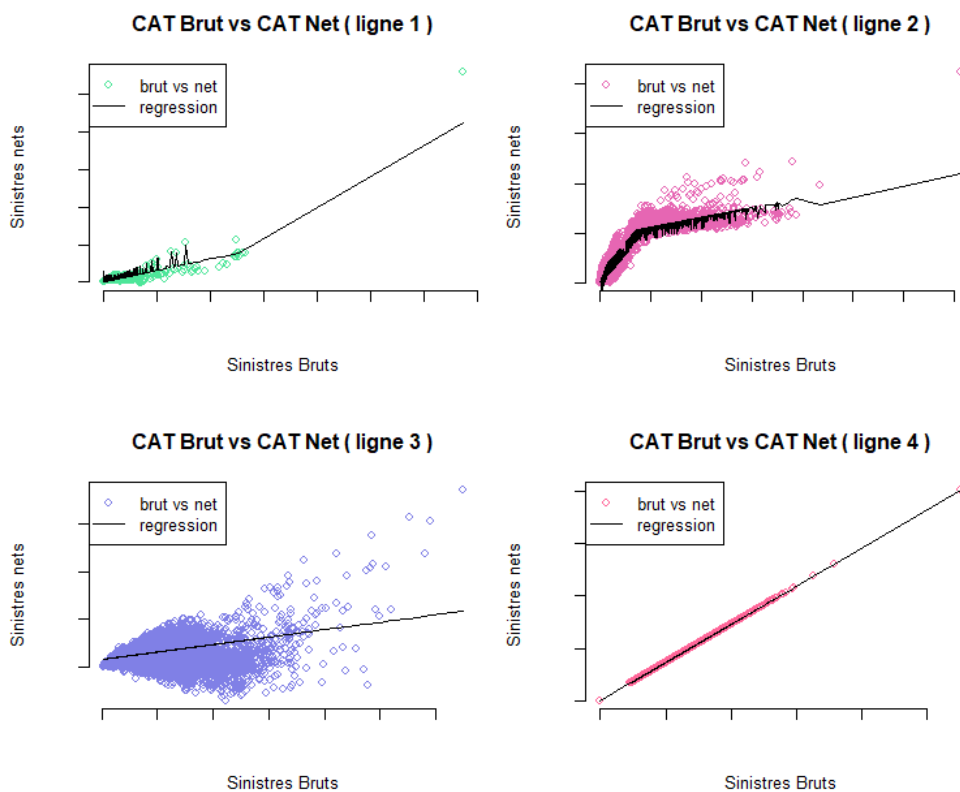


FIGURE 3.5 – Sinistres CAT nets rapportés aux CAT bruts

Ainsi, la modélisation proposée dans cette partie se révèle majoritairement efficace dans la prise en compte de l'impact de la réassurance sur la sinistralité.

Toutefois, si cette première analyse graphique s'avère encourageante, une analyse quantitative et systématique reste nécessaire afin de confirmer cette interprétation.

3.3.2 Tester la significativité des fonctions de réassurance avec le test du R^2

Afin de confirmer l'intuition fournie par cette première analyse graphique, il se révèle nécessaire de proposer une méthode qui permette de déterminer la pertinence de cette mé-

thode de manière plus systématique. C'est à cette fin qu'une approche reposant sur le test du R^2 a été développée.

Dans une régression, le R^2 , aussi appelé *coefficient de détermination*, est une mesure qui permet de déterminer à quel point la variance de la variable d'intérêt peut s'expliquer par la variable explicative [25].

En notant $(X_s)_{1 \leq s \leq S}$ le vecteur des observations des variables explicatives, $(Y_s)_{1 \leq s \leq S}$ les observations des variables d'intérêt et f la régression effectuée, la définition du R^2 est donc la suivante :

$$R^2 = \frac{\sum_{s=1}^S (Y_s - f(X_s))^2}{\sum_{s=1}^S (Y_s - \bar{Y})^2}$$

avec :

$$\bar{Y} = \frac{1}{S} \sum_{s=1}^S Y_s \quad \text{la moyenne empirique de la variable d'intérêt}$$

Cet indicateur permet donc les interprétations suivantes :

- Un modèle qui prédit systématiquement la moyenne aura un R^2 de 0
- Un modèle dont la capacité de prédiction est parfaite aura R^2 de 1
- Un modèle dont les prédictions sont moins bonnes que la moyenne peut avoir un coefficient de R^2 négatif

L'objectif de cet indicateur était en réalité double. Son intérêt premier était donc de vérifier la capacité prédictive des fonctions de réassurance choisies. Cependant, un deuxième objectif était aussi de s'assurer qu'il n'y avait pas de "sur-apprentissage". Autrement dit, l'enjeu était d'éviter une situation dans laquelle les fonctions de réassurance colleraient "trop bien" aux données, et seraient donc inadaptées dès lors que de nouvelles observations sont introduites [12].

Pour ce faire, pour chaque ligne d'assurance et chaque type de sinistre testé, un échantillon aléatoire d'observations appelé "échantillon test" et noté S^{test} a été retiré des S simulations. L'estimation des fonctions de réassurance a donc été effectuée sur les $S^{entraînement}$ observations restantes.

Ainsi, en comparant les statistiques de R^2 sur ces deux échantillons, il est possible de déterminer si du sur-apprentissage a eu lieu.

Les résultats de cette analyse, qui apparaissent sur la [figure 3.6](#), permettent tout d'abord de s'assurer de la pertinence de cette méthode pour les sinistres atypiques.

En effet, à l'exception des lignes d'assurance pour lesquelles les sinistres atypiques étaient nuls (lignes 11-12, 14-15, 42-44), le R^2 est systématiquement supérieur à 0.8, et presque identique pour les deux échantillons.

Dans le cas des sinistres CAT, le résultat est à l'image de l'analyse graphique menée en partie 3.3.1. Ainsi, si pour la majorité des lignes, le niveau de R^2 est comparable à celui des sinistres atypiques, il peut descendre jusqu'à 0.4, comme c'est le cas pour la ligne 5.

Ces valeurs correspondent aux lignes pour lesquelles la relation entre les sinistres bruts et les sinistres nets sont moins directes (c.f ligne 3 de la figure 3.5).

C'est aussi pour ces lignes qu'il peut exister des différences significatives entre le R^2 de l'échantillon d'entraînement et celui de l'échantillon test.

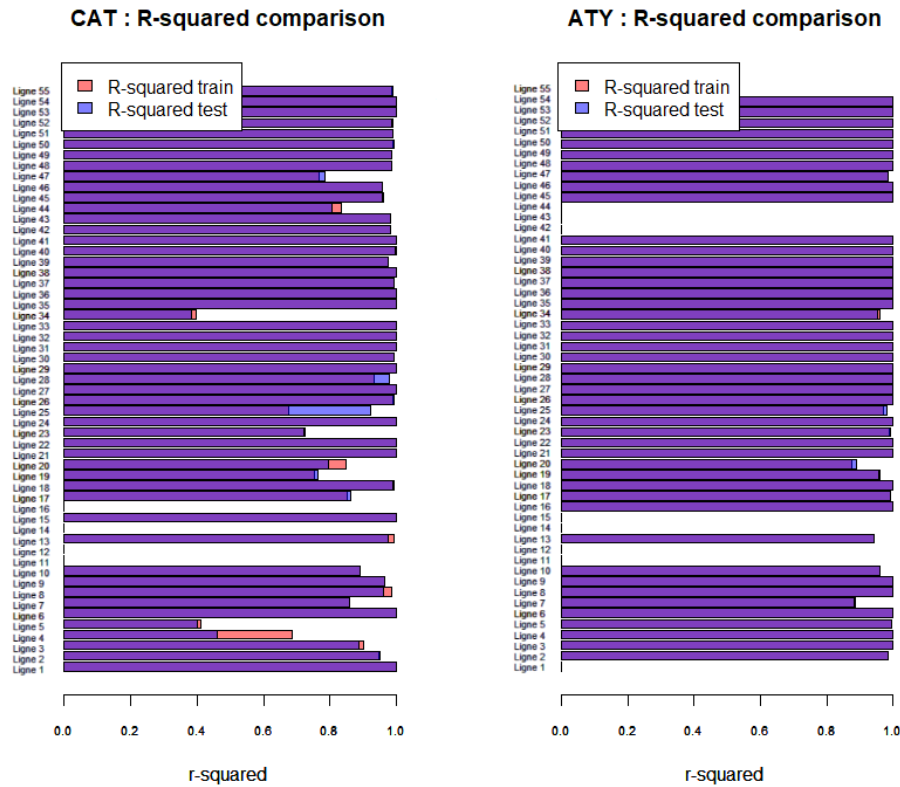


FIGURE 3.6 – Statistique de R^2 pour les échantillons d'entraînement et de test de chaque ligne d'assurance

3.3.3 Écriture du nouveau programme d'optimisation

Cette analyse, d'abord graphique puis quantitative, montre donc la pertinence de la méthode utilisée pour modéliser la totalité des fonctions de réassurance atypiques et la majorité des fonctions de réassurance CAT.

Toutefois, il existe certaines lignes pour lesquelles la modélisation par régression linéaire segmentée ne permet pas de prédire la sinistralité nette de manière précise.

Il convient cependant de noter que dans ces cas individuels, les nuages de points comparant la sinistralité nette à la sinistralité brute montrent une dispersion des observations élevée. Il serait donc difficile de mettre au point une méthode qui permette d'obtenir des fonctions de réassurance plus précises.

En outre, le traitement des résidus exploré en partie 3.2.2 permet en réalité de s'assurer que le vecteur d'observation après δ est toujours relativement proche du vecteur d'observations initial.

Le programme d'optimisation de l'assureur peut donc effectivement être écrit avec des valeurs nettes de réassurance :

$$\begin{array}{l}
 \min_{\delta_i (1 \leq i \leq n)} \left(- \sum_{i=1}^n \delta_i Rdm t_i^{net}(1), CeC_{(p)}^{net}(\delta_1, \dots, \delta_n) \right) \\
 \text{s.c.} \quad \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n) \\
 \sum_{i=1}^n \delta_i \times Primes_i^{net} > \alpha_0 \sum_{i=1}^n Primes_i^{net}
 \end{array}$$

Cependant, si ce programme est effectivement défini, sa résolution n'est manifestement pas triviale. Dès lors, l'enjeu principal de cette étude est la mise au point d'une approche qui permette de résoudre ce programme d'optimisation, ainsi que l'interprétation de ses résultats.

Chapitre 4

Résolution du programme d'optimisation par algorithmes génétiques

Lors de la formulation initiale du problème de frontière efficiente posée par Markowitz, les hypothèses posées permettaient de déterminer la solution du problème de manière explicite [18]. La courbe de la frontière efficiente était donc connue avec certitude. En pratique, la plupart des problèmes de ce type ne permettent cependant pas de connaître cet optimum théorique. Ces situations requièrent par conséquent un moyen de l'approcher et de trouver l'ensemble solution qui s'en rapproche le plus. C'est à cette fin que divers algorithmes de résolution de problèmes d'optimisation ont été mis au point.

Pour rappel, le problème d'optimisation qui caractérise cette étude, présenté en parties 2 et 3 est le suivant :

$$\begin{aligned} \min_{\delta_i (1 \leq i \leq n)} & \left(- \sum_{i=1}^n \delta_i Rdm_{t_i}(1), CeC_{(p)}(\delta_1, \dots, \delta_n) \right) \\ \text{s.c.} & \quad \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n) \\ & \quad \sum_{i=1}^n \delta_i \times Primes_i > \alpha_0 \sum_{i=1}^n Primes_i \end{aligned}$$

Déterminer l'ensemble solution explicite de ce programme serait une tâche ardue ; c'est pourquoi l'enjeu de cette partie est de déterminer quel algorithme d'optimisation appliquer à ce problème.

Ainsi, cette partie commencera par donner un aperçu de différents algorithmes d'optimisation possibles en section 4.1. Il sera ensuite possible d'explorer en détail les algorithmes génétiques, finalement retenus dans le cadre de cette étude en section 4.2. Une faiblesse de l'algorithme génétique est cependant le fait que ses propriétés de convergence peuvent être difficiles à déterminer. C'est pourquoi la section 4.3 analysera la convergence de l'algorithme une fois appliqué au problème posé lors des précédentes parties.

4.1 Un aperçu des différents algorithmes d'optimisation

Afin d'être en mesure de résoudre le problème d'optimisation de la manière la plus précise possible, il est nécessaire d'explorer au préalable les différents types d'algorithmes d'optimisation possibles. Le but ici n'est pas de fournir une liste exhaustive de la totalité des algorithmes d'optimisation ; seulement de rappeler leurs propriétés principales afin de

comprendre le choix de l'algorithme retenu.

4.1.1 L'optimisation par la force brute

La nécessité d'employer un algorithme d'optimisation se comprend aisément après avoir examiné le problème d'optimisation de cette étude sous l'angle de la *force brute*.

L'optimisation par *force brute* correspond à la méthode de résolution "naïve" d'un problème d'optimisation. En effet, cette méthode consiste tout simplement à calculer la fonction objectif pour toutes les valeurs de paramètres possibles, et de ne conserver que celles dont la performance est sur l'optimum.

Lorsque les paramètres évoluent au sein d'un espace continu, comme c'est le cas dans pour les paramètres $\beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)}$, il est possible de discrétiser l'espace en considérant un nombre P de valeurs uniformément réparties sur l'espace. Ainsi, soit A_i l'intervalle $[\beta_i^{(min)}, \beta_i^{(max)}]$ discrétisé tel que :

$$A_i = \left\{ \beta_i^{(min)} + \frac{\beta_i^{(max)} - \beta_i^{(min)}}{P} \times 1, \dots, \beta_i^{(min)} + \frac{\beta_i^{(max)} - \beta_i^{(min)}}{P} \times P \right\}$$

l'optimisation se ferait en calculant la totalité des valeurs des espaces suivants :

$$\begin{aligned} RdmtAdmissible &= \{ Rdmt_{(p)}(\delta_1, \dots, \delta_n), \delta_i \in A_i \} \\ CeCAdmissible &= \{ CeC_{(p)}(\delta_1, \dots, \delta_n), \delta_i \in A_i \} \end{aligned}$$

Cette approche permet notamment de s'assurer que la distance maximale obtenue entre une solution optimale estimée $\hat{\delta}^*$ et la solution optimale théorique δ^* est au plus le pas. Dans cet exemple, la distance maximale serait donc de $\beta_i^{(min)} + \frac{\beta_i^{(max)} - \beta_i^{(min)}}{P}$.

Toutefois, toute mise en pratique de cette méthode est très limitée par le nombre d'évaluations à effectuer. Tester P valeurs pour chaque paramètre $\delta_1, \dots, \delta_n$ revient en effet à évaluer P^n fonctions objectifs.

Dans le cadre de cette étude, tester 10 valeurs pour chaque paramètre δ_i reviendrait donc à calculer le rendement et le coût en capital un total de 10^{55} fois. Le nombre d'opérations à effectuer dans le cadre de l'optimisation par la force brute est donc bien trop grand pour un problème de cette échelle. Il est par conséquent nécessaire d'utiliser un algorithme d'optimisation plus efficace.

4.1.2 Les algorithmes du gradient

Les algorithmes du gradient, méthode développée par Cauchy, cherchent à contourner le problème du nombre d'évaluations de la méthode par la force brute en utilisant le gradient de la fonction objectif [8]. Son principe est le suivant :

Soit F une fonction objectif différentiable définie sur C , et ∇F son gradient. Cette méthode repose sur la propriété suivante, qui découle de la définition du gradient :

$$\begin{aligned} &\text{si } a > 0 \text{ et } x \in C, \text{ , alors :} \\ &F(x - a\nabla F(x)) < F(x) \end{aligned}$$

Dès lors, il est possible de s'approcher de la valeur du minimum local avec la suite suivante :

$$\begin{cases} x_0 \in C \\ x_{n+1} = x_n - a_n \nabla F(x_n) \end{cases}$$

Ainsi, en répétant le processus de manière itérative jusqu'à ce que $|F(x_{n+1}) - F(x_n)| < \epsilon$, où $\epsilon > 0$ fixé, l'algorithme est garanti de converger vers un minimum local.

Deux faiblesses de cet algorithme sont cependant à constater. Tout d'abord, cette méthode repose sur le fait que la fonction F est différentiable, et qu'il est possible de déterminer son gradient. Ensuite, si la convergence vers un minimum local est garanti, il n'est pas possible de s'assurer que le minimum obtenu est un optimum local.

Or, la forme de la fonction objectif du programme d'optimisation de l'assureur rend difficile le processus d'obtention de la fonction gradient, ou même la démonstration que la fonction objectif est différentiable.

En outre, l'objectif de cette étude n'est pas de trouver un optimum local, mais bel et bien de déterminer les optimaux globaux, c'est-à-dire l'ensemble des solutions pour lesquelles il est impossible d'améliorer le rendement sans détériorer le risque, ou inversement.

C'est pourquoi cette méthode est peu adaptée au problème posé ici.

4.1.3 Recherche des racines du gradient

Une autre approche à la question de l'optimisation est la suivante : si tous les optima locaux intérieurs d'une fonction continue sont les points pour lesquels le gradient est nul, il suffit d'annuler le gradient pour déterminer les optima d'une fonction. C'est là la méthode de Newton [17].

Le point de départ de cet algorithme est le développement de Taylor de degré un, dont la définition est la suivante :

Si f définie sur C est une fonction différentiable, alors pour $x_0 \in C$, l'approximation suivante est vérifiée :

$$\begin{aligned} &\forall x \in C : \\ &f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0) \end{aligned}$$

Il est donc possible de s'approcher de la racine de la fonction f par l'équation suivante :

$$0 = f(x_0) + f'(x_0)(x - x_0)$$

$$\Rightarrow x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Dès lors, il est possible de s'approcher fortement de la valeur de la racine de la fonction f , et donc d'un optimum local de sa primitive F à chaque itération de la suite suivante :

$$\begin{cases} x_0 \in C \\ x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \end{cases}$$

Cet algorithme n'est cependant lui non plus pas sans faiblesse. La principale est le fait que, pour déterminer un optimum de F , il est nécessaire d'estimer non seulement sa dérivée première f , mais aussi sa dérivée seconde f' . Or, il est difficile de garantir que la fonction objectif F soit deux fois dérivable, ou d'estimer la matrice Hessienne de cette fonction.

En outre, il est possible que la racine obtenue ne soit en réalité pas un optimum local, mais un point selle.

Enfin, tout comme pour l'algorithme de descente du gradient, cet algorithme ne permet pas de déterminer l'optimum global, mais seulement un optimum local.

Ainsi, s'il existe une grande diversité des algorithmes d'optimisation, tous ne sont pas adaptés à la résolution du problème de l'assureur tel qu'il est posé dans cette étude.

Cependant, si les algorithmes présentés ci-dessus ont l'avantage de présenter des propriétés de convergence claires, il existe aussi un certain nombre d'algorithmes d'optimisation pour lesquels la convergence est plus difficile à étudier. Ces algorithmes appelés *heuristiques* nécessitent souvent moins d'hypothèses, et s'appliquent donc plus facilement à une variété de problèmes d'optimisation.

C'est notamment dans cette catégorie que se situent les algorithmes génétiques, qui correspondent à la méthode finalement retenue pour cette étude.

4.2 Résolution de problèmes multi-objectif avec un algorithme génétique

Les algorithmes génétiques, classés comme algorithmes heuristiques, correspondent à un type de méthode d'optimisation reposant sur le concept *d'évolution* au sens darwinien du terme [19].

Ils fonctionnent donc en considérant une population d'individus, c'est-à-dire de valeurs possibles pour chacun des paramètres, puis en sélectionnant les "meilleurs" d'entre eux, et en leur appliquant un processus de mutation et de croisement.

Le programme d'optimisation de l'assureur développé en parties 2 et 3 sera donc résolu à

l'aide d'un algorithme génétique qui a déjà fait ses preuves dans d'autres problèmes d'optimisation de portefeuille [20] : l'algorithme NSGA-II.

Il convient donc dans un premier temps de définir cet algorithme (4.2.1). Il sera ensuite possible de l'appliquer directement au problème, et donc d'obtenir des premiers résultats (4.2.2).

4.2.1 Présentation de l'algorithme génétique NSGA-II

L'algorithme NSGA-II, développé par Deb, Pratap, Agarwal, et Meyarivan [10] a été choisi pour mener cette étude en raison de sa capacité à résoudre un problème d'optimisation multi-objectif sous contraintes, ainsi que sa vitesse de convergence vers les optima de Pareto.

L'objectif de cet algorithme est en effet de déterminer une population de N individus qui parcourent l'ensemble des optima de Pareto du problème. En tant qu'algorithme génétique, son principe est de générer une population au hasard sur l'ensemble des solutions, puis d'utiliser les meilleurs individus de cette population pour générer une population à l'aide d'opérateurs de croisement et de mutation. Cette procédure est ensuite répétée de manière itérative pour converger vers une population qui parcourt l'ensemble des optima du problème.

Cet algorithme repose donc sur trois opérateurs particuliers :

1. L'opérateur de croisement :
l'opérateur de croisement promu est l'opérateur *de croisement binaire simulé*, dont l'objectif est d'adapter l'opérateur *au niveau du bit* à l'ensemble des réels (c.f. Annexe C)
2. L'opérateur de mutation :
Dans le cas de l'opérateur de mutation, c'est l'opérateur *polynomial* qui est promu (c.f. Annexe D)
3. L'opérateur de *fitness* :
Pour comparer le *fitness* des individus, ceux-ci sont triés par non-domination. Ainsi, les individus qui ne sont pareto-dominés par aucun individu de la population sont classés dans l'ensemble \mathcal{F}_1 . Puis, ceux qui ne sont dominés que par les individus de \mathcal{F}_i sont classés dans l'ensemble \mathcal{F}_{i+1} , et cette procédure est répétée de manière itérative jusqu'à $\mathcal{F}_{q+1} = \emptyset$.

L'algorithme consiste donc à 6 étapes :

1. Initialisation de la population P_0 de N individus, et utilisation de processus de croisements et de mutations au sein de P_0 afin de créer la génération Q_0 de N individus
2. Création de la population $R_g = P_g \cup Q_g$

3. Classement de la population R_g au sens \mathcal{F}
4. Sélection des N meilleurs individus de R_g pour former la population P_{g+1} .
Dans le cas où il faudrait départager les individus de \mathcal{F}_i , ceux-ci sont classés selon leur *crowding distance*, qui mesure la densité des populations dans le voisinage, de manière à choisir les individus les mieux répartis sur l'ensemble des solutions.
5. Création de la génération Q_{g+1} de N individus à l'aide de croisements et mutations au sein de la population P_{g+1}
6. Répétition des étapes 2 à 5 jusqu'à I itérations

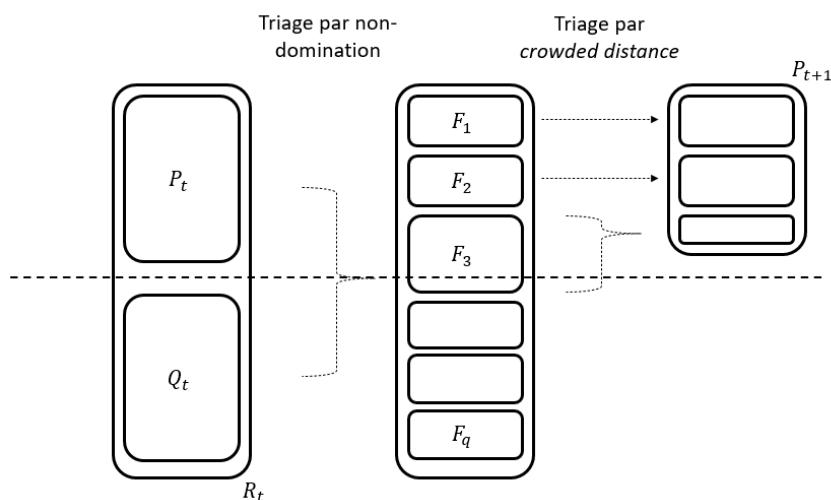


FIGURE 4.1 – Procédure de l'algorithme NSGA-II (source : [10])

En appliquant cet algorithme au problème d'optimisation posé (4.1), il est donc possible de converger par itérations vers l'ensemble des optima de Pareto, malgré le fait que la définition explicite de ces optima ne soit pas connue.

Cet algorithme a donc comme avantage majeur le fait que très peu d'hypothèses doivent être posées afin de l'appliquer à un problème d'optimisation. Cependant, ce manque d'hypothèses se traduit aussi par une difficulté à déterminer ses propriétés de convergence théoriques.

C'est pourquoi l'enjeu est maintenant d'étudier la convergence lorsque le NSGA-II est appliqué au programme d'optimisation de l'assureur.

4.2.2 L'application du NSGA-II au problème d'optimisation de l'assureur

Maintenant l'algorithme NSGA-II défini, il peut être utilisé dans le cadre de cette étude pour déterminer l'ensemble des optima du problème d'optimisation. En particulier, c'est au programme d'optimisation suivant qu'il est appliqué :

$$\begin{aligned}
 \min_{\delta_i (1 \leq i \leq n)} & \left(- \sum_{i=1}^n \delta_i Rdm_{t_i}(1), C_e C^{(p)}(\delta_1, \dots, \delta_n) \right) \\
 \text{s.c.} & 0.5 \leq \delta_i \leq 1.2 \quad \forall i \in (1, \dots, n) \\
 & 0.9 \sum_{i=1}^n Prime_i \leq \sum_{i=1}^n \delta_i Prime_i
 \end{aligned} \tag{4.1}$$

Autrement dit, pour tout $i \in [1, \dots, n]$, il a été fixé $\beta_{min} = 0.5$ et $\beta_{max} = 1.2$, pour une diminution maximale par ligne de 50 % et une augmentation maximale de 20 %. En outre, une contrainte supplémentaire empêchant la prime totale de diminuer de plus de 10 % a été fixée.

Puisque la convergence théorique de l'algorithme n'est pas assurée, il est impératif d'observer ses résultats afin de déterminer s'ils sont satisfaisants. Pour ce faire, deux types de sorties de l'algorithme peuvent être étudiés : l'apparence de la fonction objectif, et l'apparence des paramètres obtenus.

Étudier la convergence à l'aide de la fonction objectif

Étudier la fonction objectif revient à observer les valeurs que prennent les fonctions de rendement et de risque après un certain nombre d'itérations du NSGA-II. Cela signifie donc que l'apparence de la frontière efficiente doit être étudiée.

C'est à cette fin que la [figure 4.2](#) montre les valeurs que parcourent la frontière efficiente selon le nombre d'itérations de l'algorithme.

Puisque l'objectif est ici de minimiser le coût en capital tout en maximisant le rendement, une solution est d'autant plus optimale qu'elle se situe en haut à gauche sur la figure. C'est donc une représentation analogue à la frontière efficiente usuelle en mathématiques financières (voir partie 2.1.2).

L'analyse de cette figure révèle donc que l'algorithme permet bien de s'approcher de la frontière efficiente théorique. En effet, l'augmentation du nombre d'itérations de l'algorithme permet de translater la totalité de la frontière efficiente obtenue vers le quadrant supérieur gauche de la figure.

Il apparaît par ailleurs qu'il est bien possible d'optimiser davantage le portefeuille d'assurance étudié. Cela s'observe par le fait que le portefeuille actuel, c'est-à-dire le portefeuille pour lequel $\delta_1 = \dots = \delta_n = 1$, a un rendement plus faible et un risque plus élevé qu'un grand nombre de valeurs sur la frontière efficiente.

Cela ne permet cependant pas de garantir le fait que l'optimum théorique est atteint. C'est pourquoi une étude supplémentaire des paramètres qui permettent d'obtenir ces combinaisons de rendement et de coût en capital est nécessaire.

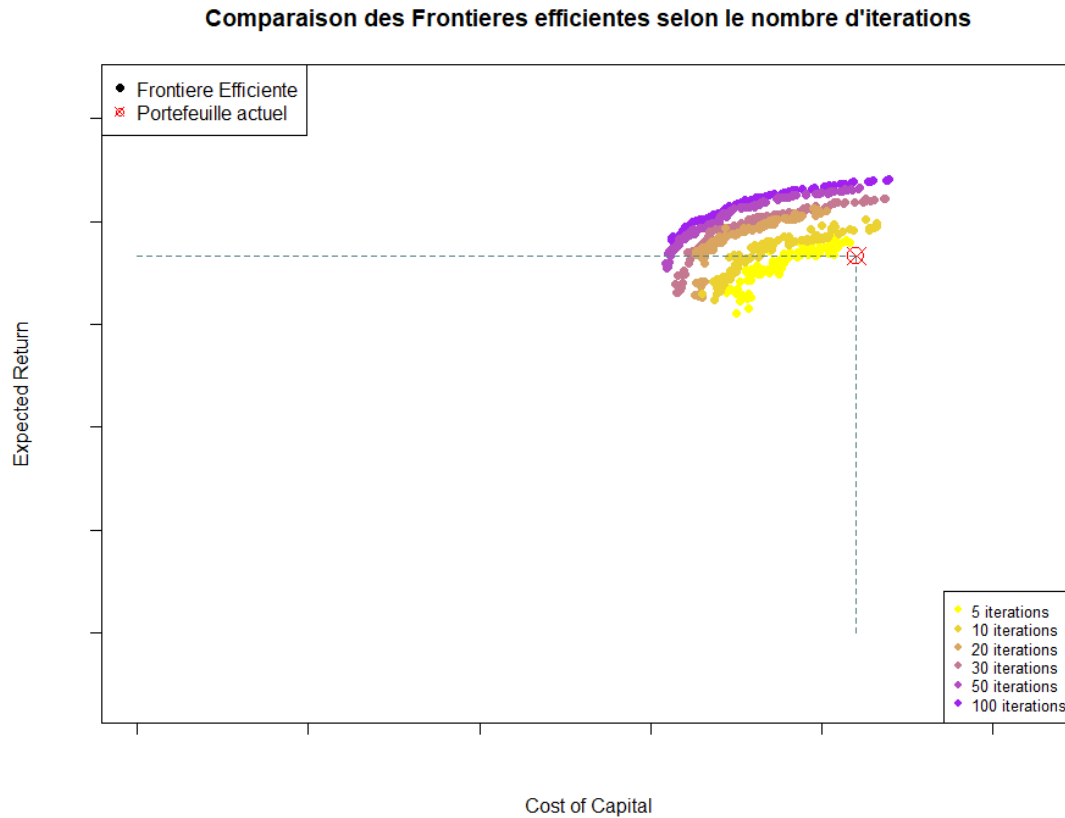


FIGURE 4.2 – Convergence de l’algorithme NSGA-II

Étudier la convergence en observant les paramètres

Tandis que l’analyse de la frontière efficace permet de déterminer l’optimum trouvé par l’algorithme d’optimisation, l’analyse des paramètres qui permettent d’atteindre cet optimum révèle aussi des informations sur ses propriétés de convergence. C’est à cette fin que la *courbe du facteur d’évolution* est étudiée ici.

Définition : fonction du facteur d’évolution d’une ligne d’assurance

Soient G le nombre de générations utilisé lors de l’algorithme NSGA-II et Q le nombre de points considérés sur la frontière efficace

Soit $(CeC_{(1)g}^*, \dots, CeC_{(Q)g}^*)$ les valeurs ordonnées des valeurs de CeC obtenues par l’algorithme NSGA-II à la g -ième génération

Alors, la fonction d’évolution de la ligne $i \in \{1, \dots, n\}$ notée $evol_{i,g}$ est la fonction qui prend en argument une position sur la frontière efficace, et qui renvoie la valeur de δ_i qui permet de l’atteindre :

$$\begin{aligned}
 evol_{i,g} : \{1, \dots, Q\} &\longrightarrow [\beta_i^{(min)}, \beta_i^{(max)}] \\
 n &\mapsto \delta_i^*, \text{ tq } CeC(\delta_i^*) = CeC_{(n)g}^*
 \end{aligned}
 \tag{4.2}$$

La courbe de la fonction $evol_{i,g}$ représente donc les facteurs d'évolution de la ligne i , notés $(\delta_{i(1)}, \dots, \delta_{i(Q)})$ qui permettent respectivement d'obtenir chaque couple $(CeC_{(q)g}^*, Rdm_{(q)g}^*)_{1 \leq q \leq Q}$ sur la frontière efficiente.

La Figure 4.3 montre donc l'apparence de cette courbe pour une ligne d'assurance avec un excellent profil rendement/risque selon le nombre de générations NSGA-II. Cette figure permet d'observer que, sur une telle ligne, l'algorithme a bien l'effet escompté :

- Lorsque le nombre de générations g est faible, le facteur d'évolution δ_i^* associé à cette ligne varie beaucoup selon la position sur la frontière efficiente
- Lorsque le nombre de générations g est élevé, le facteur d'évolution δ_i^* associé à une ligne très performante converge vers la contrainte supérieure quelle que soit la position sur la frontière efficiente.

Cette approche semble par conséquent adéquate lorsque le profil rendement / risque d'une ligne d'assurance est très positif ou très négatif, puisque l'évolution optimale est alors d'aller coller à la contrainte.

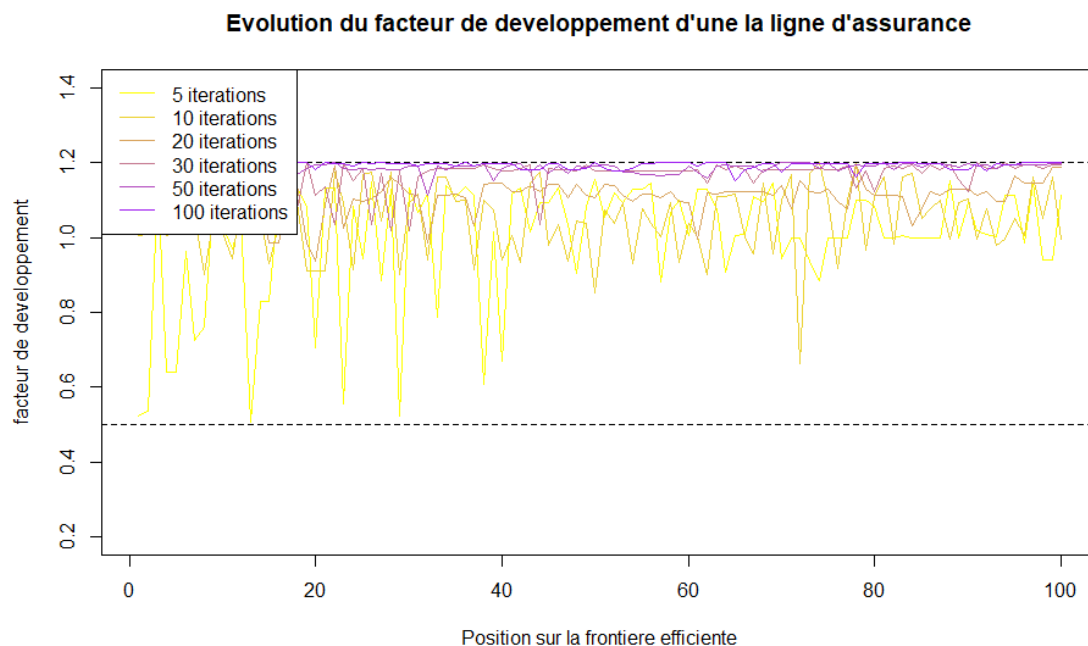


FIGURE 4.3 – Courbe du facteur d'évolution d'une plus petite ligne après 100 itérations

Cependant, lorsque le profil de la ligne d'assurance donne lieu à des conclusions moins évidentes, ou que la ligne a un impact plus ténue sur le portefeuille en raison de sa taille, la convergence n'est pas aussi précise. C'est ce que montre la Figure 4.4, sur laquelle est représentée la courbe d'évolution après 100 itérations d'une ligne peu risquée, mais à faible rendement. Là, trois phénomènes sont à constater :

- Puisque la ligne est peu risquée, le facteur δ_i^* sur la partie "à gauche" de la frontière efficiente est élevé
- La ligne étant peu rentable, le facteur d'évolution δ_i^* associé est faible sur la partie

"droite" de la frontière efficiente

- Après 100 itérations, aucune convergence n'est vraiment observée sur la partie intermédiaire de la frontière efficiente

L'étude de la courbe d'évolution révèle donc ici que le facteur d'évolution de l'algorithme NSGA-II δ_i^* peut éprouver des difficultés à converger sur certains segments de la frontière efficiente. En particulier, lorsque la ligne i est relativement petite, ou a un impact mitigé entre le risque et le rendement, la taille de ce segment de non-convergence a tendance à augmenter.

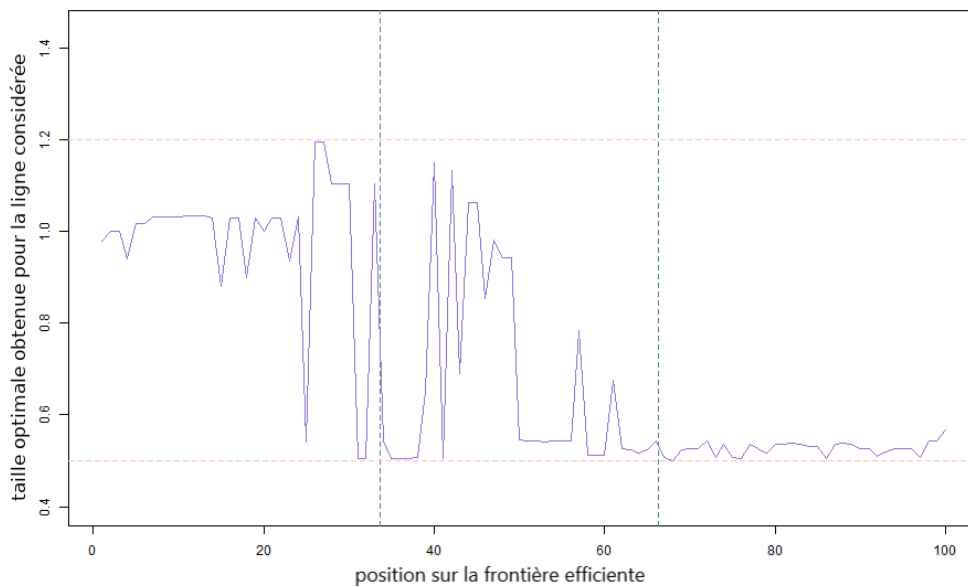


FIGURE 4.4 – Ligne d'assurance pour laquelle la convergence est volatile

La conséquence de cette volatilité est qu'il est très difficile de fournir une interprétation économique ou stratégique sur le segment de la frontière efficiente concernée. L'objectif de cette étude est en effet de déterminer de manière algorithmique comment développer le portefeuille d'assurance d'un assureur. Or, dès lors qu'une oscillation aussi significative est constatée, il devient impossible d'établir une recommandation d'évolution claire pour la ligne d'assurance qui permette d'atteindre le couple rendement / risque recherché.

L'étude des résultats d'une itération de l'algorithme NSGA-II permet donc d'établir que l'approche développée ici est très performante dès lors que l'objectif est de déterminer les lignes d'assurance dont l'impact sur le portefeuille est le plus important. Cependant, lorsque l'impact d'une ligne est plus disparate, la convergence vers un résultat stable se fait plus difficilement.

En pratique, lorsque le facteur d'évolution de la ligne i δ_i oscille entre deux valeurs extrêmes sur un segment de la frontière efficiente, aucune interprétation économique ne peut être faite.

C'est pourquoi il est nécessaire d'améliorer le modèle développé lors de cette étude davantage en mettant au point une méthode capable de déterminer la stabilité des facteurs d'évolution.

4.3 Vérifier la fiabilité de l'algorithme à travers plusieurs itérations

L'analyse des résultats de l'algorithme NSGA-II au problème met en réalité évidence la faiblesse principale de cet algorithme : sa stochasticité. Autrement dit, le fait que les sorties de cet algorithme soient aléatoires par définition signifie qu'il est difficile de garantir l'exactitude des résultats obtenus.

Il est donc nécessaire d'établir la nature de la volatilité des vecteurs de facteurs d'évolution $(\delta_{i,q}^*)_{1 \leq q \leq Q}$ obtenus. La question est notamment de déterminer si les oscillations constatées sont effectivement statistiquement significatives, si elles proviennent d'un manquement de l'algorithme NSGA-II, ou si les lignes d'assurance concernées ont un impact trop faible sur le portefeuille pour être modélisé.

C'est pourquoi cette partie se termine par une étude statistique des résultats obtenus.

Pour ce faire, la méthode proposée consiste à répéter la totalité de l'algorithme NSGA-II plusieurs fois afin d'obtenir un échantillon d'observations des facteurs d'évolution $(\delta_i^*)_{1 \leq i \leq n}$ (4.3.1).

Cet échantillon permettra ensuite de déterminer si les résultats obtenus sont statistiquement significatifs (4.3.2).

Enfin, il sera possible de mettre au point une approche qui permette de stabiliser les résultats, ainsi de comparer la volatilité des différentes lignes d'assurance (4.3.3).

4.3.1 Générer une base de données des résultats en multipliant les itérations de l'algorithme NSGA-II

L'incertitude liée aux résultats de l'algorithme relevée lors de la section précédente (4.2.2) porte principalement sur les paramètres δ^* utilisés pour tracer la frontière efficiente. En effet, puisque le NSGA-II est un algorithme stochastique, l'optimum δ^* obtenu est une variable aléatoire par définition, et il convient donc d'analyser ses propriétés.

Ainsi, avec $i \in \{1, \dots, n\}$ une ligne d'assurance et $q \in \{1, \dots, Q\}$ une position sur la frontière efficiente, l'objectif est de générer un vecteur de T observations indépendantes et identiquement distribuées du facteur d'évolution optimal $\delta_{i,q}^*$ identifié par l'algorithme.

Pour ce faire, il suffit de ré-exécuter le NSGA-II T fois, en gardant fixe le nombre de générations G et de points sur la frontière efficiente Q .

Cela permet donc d'avoir accès aux observations suivantes :

- $(Rdmt_{1t}^*, \dots, Rdmt_{Qt}^*)_{1 \leq t \leq T}$ les T réalisations des vecteurs des Q rendement optimaux obtenus par le NSGA-II
- $(CeC_{1t}^*, \dots, CeC_{Qt}^*)_{1 \leq t \leq T}$ les T réalisations des vecteurs des Q Coûts en Capital op-

- timaux correspondants
- $(\delta_{i,q_1}^*, \dots, \delta_{i,q_T}^*)$ les T réalisations du facteur d'évolution de la ligne i permettant d'obtenir le q -ième couple $(Rdmt_{q_t}^*, CeC_{q_t}^*)_{1 \leq t \leq T}$ correspondant

Il est important de noter que les résultats obtenus par le NSGA-II correspondent toujours à l'optimum trouvé par l'algorithme. Ainsi, Tous les couples $(Rdmt_{q_t}^*, CeC_{q_t}^*)_{1 \leq q \leq Q}$ obtenus sont ceux pour lesquels aucune combinaison de paramètres δ ne permet d'améliorer le rendement sans détériorer le risque ou inversement.

Ainsi, à chaque itération $t \in \{1, \dots, T\}$, il existe une suite $(u_q^{(t)})_{1 \leq q \leq Q}$ à valeurs dans $\{1, \dots, Q\}$ qui permette d'ordonner les vecteurs de rendement et de coût en capital conjointement, tels que :

$$Rdmt_{u_1^{(t)} t}^* \leq Rdmt_{u_2^{(t)} t}^* \leq \dots \leq Rdmt_{u_Q^{(t)} t}^*$$

$$CeC_{u_1^{(t)} t}^* \geq CeC_{u_2^{(t)} t}^* \geq \dots \geq CeC_{u_Q^{(t)} t}^*$$

Les notations sont donc les suivantes :

- $Rdmt_{(1)t}^* \leq Rdmt_{(2)t}^* \leq \dots \leq Rdmt_{(Q)t}^*$ les Q réalisations du rendement ordonnées selon la suite $(u_q^{(t)})_{1 \leq q \leq Q}$
- $CeC_{(1)t}^* \leq CeC_{(2)t}^* \leq \dots \leq CeC_{(Q)t}^*$ les Q réalisations du Coût en capital ordonnées selon la même suite $(u_q^{(t)})_{1 \leq q \leq Q}$
- $(\delta_{i,(1)t}^*, \dots, \delta_{i,(Q)t}^*)$ les facteurs d'évolution de la ligne i correspondants

En pratique, on constate peu de volatilité entre les différents couples ordonnés $(Rdmt_{(q)t}^*, CeC_{(q)t}^*)$ à chaque itération de la frontière efficiente (c.f [Figure 4.5](#)).

Les réalisations "ordonnées" de $\delta_{i,(q)}^*$ permettent donc d'obtenir un rendement et un coût en capital équivalents.

En outre, en gardant fixe le nombre de générations et de points sur la frontière efficiente à chaque itération de l'algorithme, il est possible de s'assurer que chaque itération du paramètre d'évolution de la ligne i en q -ième position $\delta_{i,(q)}^*$ est indépendante et identiquement distribuée.

Il est donc possible d'effectuer des tests statistiques sur les différentes réalisations de $\delta_{i,(q)}^*$ afin d'en déterminer les propriétés.

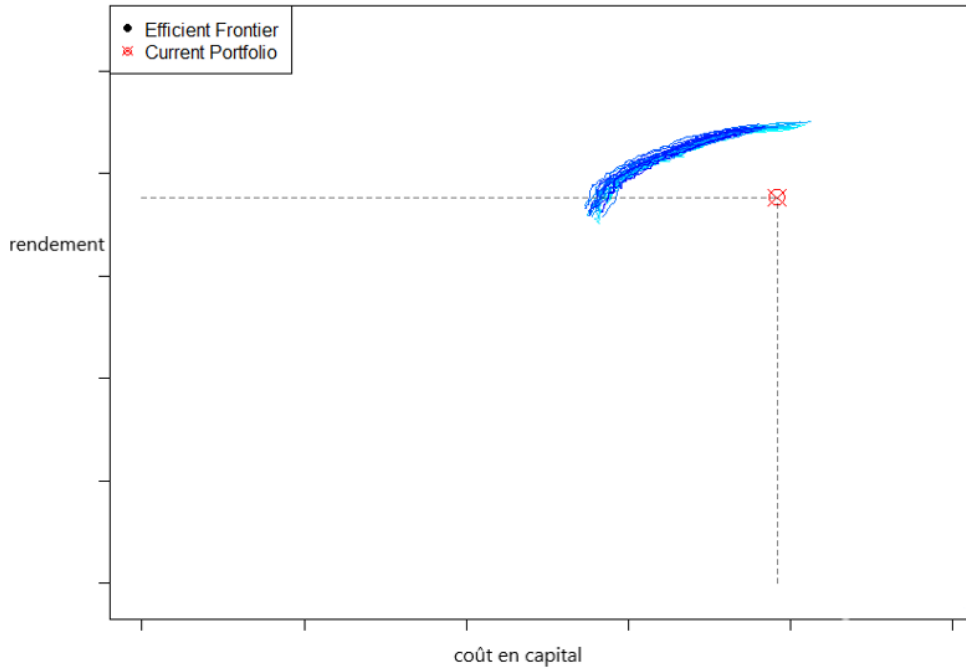


FIGURE 4.5 – Illustration de la multiplication des itérations de frontières efficaces

4.3.2 Mettre en place un test de significativité des résultats

La méthode mise au point ici est née du besoin de vérifier si les résultats de l'algorithme NSGA-II étaient statistiquement significatifs. En effet, lorsque des oscillations sont constatées sur le vecteur du facteur d'évolution de la ligne i en fonction de la position sur la frontière efficiente, plusieurs explications sont à distinguer (cf. Figure 4.4) :

- Ces oscillations sont réellement significatives, et correspondent aux valeurs optimales du vecteur $\left(\delta_{i,(q)}^*\right)_{1 \leq q \leq Q}$
- Le NSGA-II converge bien vers le paramètre optimal, mais des oscillations autour de l'optimum sont constatées en raison de la stochasticité de l'algorithme
- Le NSGA-II ne converge pas pour la ligne et le segment de la frontière efficiente concernée, et les valeurs obtenues doivent être ignorées.

Il est donc nécessaire de mettre tout d'abord au point un test statistique qui permette d'identifier les résultats qui relèvent de cette troisième catégorie. Autrement dit, l'objectif est de s'assurer que les résultats obtenus permettent effectivement d'améliorer le portefeuille sur les axes de rendement et de risque.

Or, sous l'hypothèse selon laquelle la fonction objectif du NSGA-II est totalement indépendante des paramètres δ , aucune convergence ne peut avoir lieu. Par conséquent, la distribution des paramètres δ devient celle d'une loi uniforme entre les contraintes. (c.f Figure 4.6)

La méthode proposée consiste donc à observer les T réalisations du facteur d'observation δ permettant d'obtenir le q -ième couple rendement / risque $\left(\delta_{i,(q)t}^*\right)_{1 \leq t \leq T}$, et à comparer la

distribution des observations avec celles d'une loi uniforme $\mathcal{U}[\beta_i^{min}, \beta_i^{max}]$.

Pour ce faire, un test de *Kolmogorov-Smirnov* est appliqué à chaque vecteur d'observations $(\delta_{i,(a)t}^*)_{1 \leq t \leq T}$.

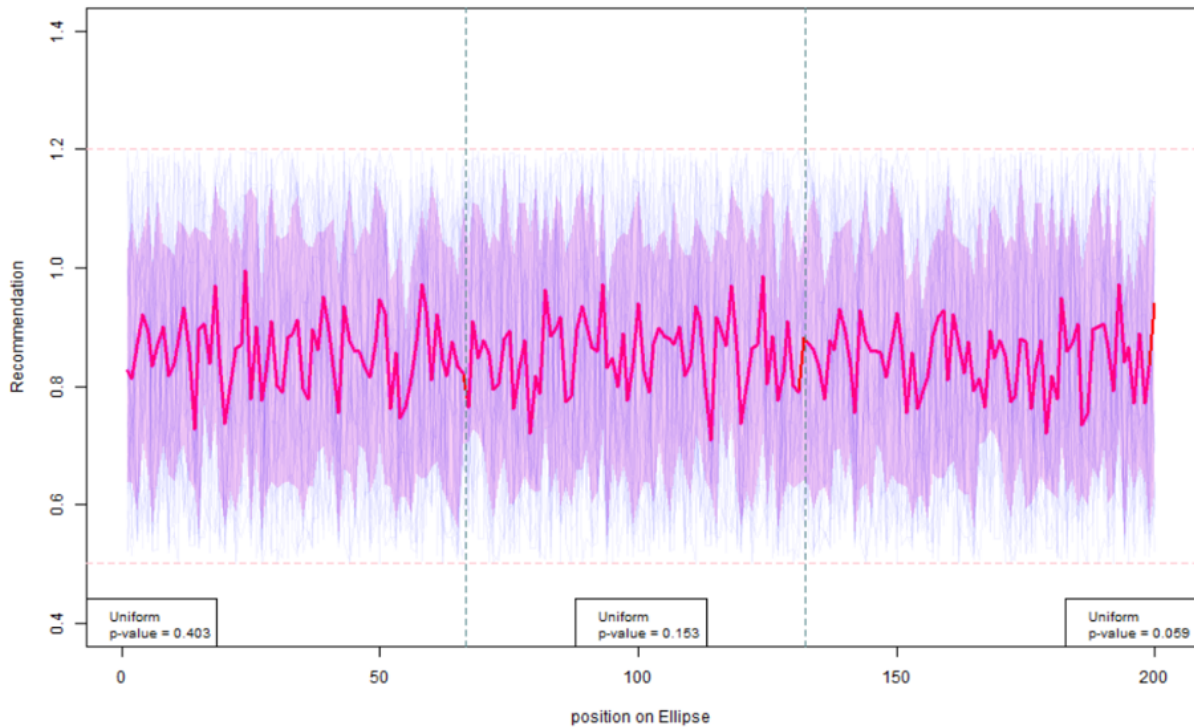


FIGURE 4.6 – La distribution uniforme des paramètres δ sans fonction objectif

Le test de Kolmogorov-Smirnov est une statistique de test qui vise à comparer la fonction de répartition empirique des observations d'une variable aléatoire à une fonction de répartition théorique [16].

En notant (X_1, \dots, X_N) les observations de la variable aléatoire X , la statistique de Kolmogorov-Smirnov se définit de la manière suivante :

Soit F_n la fonction de répartition empirique de X , définie par :

$$F_N(x) = \frac{1}{N} \sum_{k=1}^N \mathbb{1}(X_k \leq x)$$

Soit F une fonction de répartition théorique.

Alors, la statistique de test de Kolmogorov-Smirnov, notée KS_N , est définie par :

$$KS_N = \sup_x |F_N(x) - F(x)|$$

Intuitivement, la statistique KS_N donne donc l'écart maximal obtenu entre la fonction de répartition théorique et la fonction de répartition testée.

En calculant la probabilité asymptotique de $\mathbb{P}(KS_N < x)$, il est donc possible de déterminer un test permettant de rejeter ou non l'hypothèse nulle selon laquelle X suivrait la loi de F

[26].

En appliquant ce test au contexte de cette étude, pour une ligne d'assurance i , il est donc possible de déterminer pour chaque $q \in \{1, \dots, Q\}$ si le facteur d'évolution $\delta_{i,(q)}^*$ est significatif ou non.

- si l'hypothèse nulle est rejetée, les observations de $\left(\delta_{i,(q)t}^*\right)_{1 \leq t \leq T}$ ne suivent pas une loi uniforme entre les contraintes. Les résultats pour la ligne i sur le point q de la frontière efficiente sont donc significatifs.
- si l'hypothèse nulle n'est pas rejetée, il n'est pas possible de considérer que les observations de $\left(\delta_{i,(q)t}^*\right)_{1 \leq t \leq T}$ ne suivent pas une loi uniforme entre les contraintes. Les résultats pour la ligne i sur le point q de la frontière efficiente doivent donc être ignorés.

Cette méthode permet donc de déterminer si les résultats obtenus sont significatifs. Cependant, quand bien même ils le seraient, les oscillations possibles observées sur certains segments de la frontière efficiente rendent difficile le fait d'appliquer une stratégie concrète. C'est pourquoi l'enjeu est maintenant d'utiliser les multiples itérations du NSGA-II pour stabiliser les facteurs d'évolution δ de l'algorithme.

4.3.3 Stabilisation des résultats obtenus

L'objectif du test de significativité effectué jusqu'à présent était de déterminer si le NSGA-II permettait de converger vers l'optimum ou non. Toutefois, il est possible que les oscillations constatées soient dues, non pas au fait que l'algorithme ne converge pas vers l'optimum, mais à la volatilité inhérente aux algorithmes génétiques. C'est pourquoi il est maintenant question de stabiliser les résultats obtenus, et ainsi obtenir un vecteur de facteurs d'évolution $\left(\delta_{i,(q)}^*\right)_{1 \leq q \leq Q}$ pour chaque ligne i plus proche de l'optimum réel.

Pour ce faire, la méthode retenue consiste à calculer l'espérance de ce facteur $\delta_{i,(q)}^*$.

Déterminer l'espérance de $\delta_{i,q}^*$

Si la stochasticité inhérente au NSGA-II signifie que sa sortie pour le paramètre $\delta_{i,(q)}^*$ est une variable aléatoire, les multiples itérations réalisées en 4.3.1 permettent d'en obtenir une liste de réalisations indépendantes et identiquement distribuées. Il est donc possible d'appliquer la méthode suivante :

Soit $i \in \{1, \dots, n\}$ et $q \in \{1, \dots, n\}$.

Déterminer l'espérance du facteur d'évolution de la ligne i à la q -ième position sur la frontière efficiente $\delta_{i,q}^*$ se fait aisément en utilisant la *Loi Forte des Grands Nombres*. Celle-ci stipule en effet que, avec $T \in \mathbb{N}$ observations indépendantes et identiquement distribuées de la variable $\delta_{i,q}^*$ [11] :

$$\lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=1}^T \delta_{i,q_t}^* = \mathbb{E} [\delta_{i,q}^*]$$

Ainsi, pour stabiliser paramètres δ , il suffit de calculer la moyenne empirique. De la sorte, il est possible de soustraire la volatilité inhérente aux vecteurs $\delta_{i,(q)}^*$ obtenus par une itération de l'algorithme.

Quantifier la dispersion de $\delta_{i,q}^*$

Si le calcul de l'espérance du paramètre $\delta_{i,q}^*$ permet certes d'obtenir des résultats plus stables, il est tout de même intéressant d'analyser leur volatilité. En effet, si une ligne voit ses facteurs d'évolution δ_i soumis à de plus fortes oscillations, c'est que l'impact de cette ligne sur les fonctions objectifs est moins important. De la sorte, l'assureur disposera de plus de marge de manœuvre pour rester vers l'optimum.

Il serait alors intuitif d'utiliser des intervalles de confiance pour quantifier cette dispersion. Toutefois, l'objectif n'est pas ici de déterminer le seuil à partir duquel un résultat ne serait plus significatif, mais simplement de mettre au point un indicateur qui distingue les lignes pour lesquelles le facteur δ est volatile de celles où il ne l'est pas.

C'est pourquoi la méthode finalement retenue consiste à déterminer les quantiles empiriques.

Ainsi, soit α le niveau de confiance souhaité. Soit $i \in \{1, \dots, n\}$ et $q \in \{1, \dots, n\}$. La dispersion du facteur d'évolution sur le q -ième point de la frontière efficiente $\delta_{i,q}^*$ s'observe à l'aide des quantités suivantes :

$$\left\{ \begin{array}{l} Quantile^{sup} = \inf_x \left\{ x \mid \left(\sum_{t=1}^T \mathbb{1}(\delta_{i,q_t}^* \leq x) \right) > (1 - \alpha)T \right\} \\ Quantile^{inf} = \inf_x \left\{ x \mid \left(\sum_{t=1}^T \mathbb{1}(\delta_{i,q_t}^* \leq x) \right) > \alpha T \right\} \end{array} \right.$$

La dispersion correspond donc à l'écart entre la (αT) -ième plus petite observation de $\delta_{i,q}^*$ et la (αT) -ième plus grande observation.

Il convient toutefois de noter que cette méthode pose la question de la détermination du choix du nombre d'itérations T . Une valeur de T trop faible diminuerait la précision de l'estimateur de l'espérance, et ne permettrait donc pas de stabiliser les δ de manière adéquate. À l'inverse, une valeur de T trop élevée poserait des difficultés supplémentaires, en particulier au niveau du temps d'exécution de l'algorithme.

Le choix du paramètre T est donc le résultat de l'arbitrage entre précision et temps d'exécution.

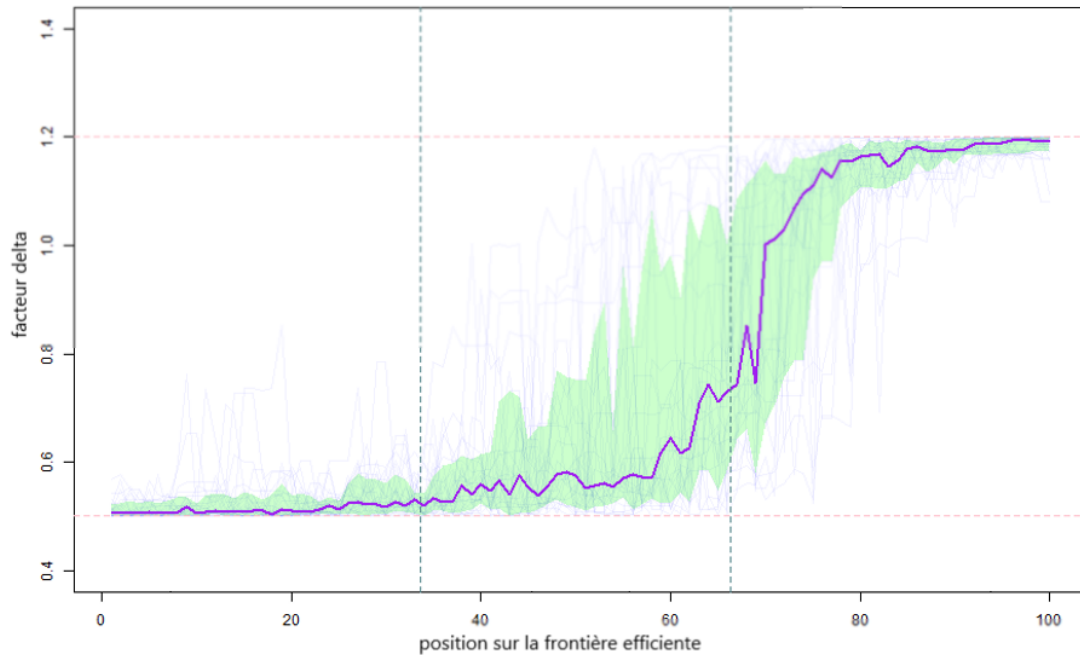


FIGURE 4.7 – Espérance et dispersion de $\delta_{i,q}^*$ au quantile 0.1 pour une ligne i donnée

En somme, parmi les différents algorithmes d'optimisation existants, l'algorithme génétique NSGA-II semblait être le mieux adapté pour résoudre le problème posé. De fait, aucune difficulté particulière n'a été constatée lors de l'identification de couples $(Rdmt, CeC)$ plus performants que le portefeuille de base.

Cependant, s'il est possible d'obtenir des résultats très pertinents pour les lignes dont l'impact est important, le facteur δ de lignes au profil plus mitigé peut être bien plus variable. Cette difficulté peut cependant être surmontée en multipliant les itérations du NSGA-II. De la sorte, il est possible d'effectuer des tests de significativité, ainsi que de stabiliser les résultats.

Dès lors, l'enjeu est de déterminer comment interpréter les résultats. En effet, au terme de l'étude réalisée, le modèle développé renvoie encore Q scénarios différents, correspondant à Q points différents sur la frontière efficiente obtenue. En outre, pour chacun de ces scénarios, en plus d'un facteur d'évolution $\delta_{i,q}$ moyen pour chaque ligne i , la dispersion est aussi à prendre en compte.

C'est pourquoi il est nécessaire de proposer une approche qui synthétise les résultats du modèle développé, et qui permette d'appliquer ses conclusions de manière concrète.

Chapitre 5

Application sur un portefeuille d'assurance d'AXA

L'objectif de cette étude, *in fine*, était de créer un outil capable de trouver de manière algorithmique quelle est l'évolution optimale de chaque ligne d'assurance qui comprend le portefeuille. Pour ce faire, les parties 2 et 3 se sont concentrées sur l'écriture du programme d'optimisation en lui-même, tandis que la partie 4 s'est concentrée sur ses méthodes de résolution.

De la sorte, le modèle développé jusqu'ici a permis d'obtenir les résultats suivants :

- Les couples de Rendement et de coût en capital atteignables en faisant varier l'exposition au risque des lignes d'assurance
- Les facteurs d'évolution δ à appliquer à chaque ligne d'assurance afin d'atteindre ces couples rendement / coût en capital
- Les résultats des tests de significativité permettant de déterminer si ces facteurs d'évolution sont statistiquement significatifs
- Des intervalles donnant la marge de manœuvre pour chaque ligne d'assurance si l'objectif est de rester proche de l'optimum

L'étape suivante est donc naturellement le passage de ces résultats théoriques à une décision pratique. Cependant, la grande quantité d'informations données rend ce processus non trivial, et ce d'autant plus que Q points de frontière efficiente sont donnés.

L'objectif de cette partie est donc de donner un aperçu des interprétations possibles des résultats ci-dessus. Cela se fera notamment à travers l'application du modèle au portefeuille d'assurance AXA, pour lequel des exemples de processus de décision seront mis en place.

Un premier processus consistant à synthétiser la totalité de la frontière efficiente de la manière la plus concise possible sera donc exploré dans une section 5.1.

L'idée de concentrer l'analyse sur le segment de la frontière efficiente désiré pourra ensuite être explorée en section 5.2.

5.1 Synthétiser la totalité de la frontière efficiente

L'objectif de cette première règle de décision mise au point est de simplifier le plus possible la lecture des résultats du modèle. C'est pourquoi le processus de décision mis au point ici vise à synthétiser simplement la totalité de la frontière efficiente.

Ici, cette synthèse passe tout d'abord par la réduction des Q points sur la frontière

efficente à un unique facteur par ligne d'assurance, effectuée en section 5.1.1. Ensuite, une fois ces uniques facteurs obtenus, la section 5.1.2 les analysera afin de déterminer notamment leur cohérence avec les résultats attendus.

5.1.1 Synthétiser simplement les informations pour chaque ligne d'assurance

La première difficulté lors de l'interprétation directe des sorties du modèle est la présence de Q points sur la frontière efficiente obtenue, ainsi que de n facteurs d'évolution pour chacun de ces points.

Ainsi, pour obtenir un facteur unique pour chaque ligne d'assurance $i \in \{1, \dots, n\}$, il est naturel de penser à calculer la moyenne empirique de son facteur d'évolution sur les Q observations. Autrement dit, le facteur retenu proposé ici est le suivant :

$$\Delta_i = \frac{1}{Q} \sum_{q=1}^Q \mathbf{E} [\delta_{i,q}^*]$$

Sous réserve de significativité, interpréter ce facteur agrégé Δ_i est assez intuitif :

- Plus Δ_i est supérieur à 1, plus il est recommandé, en moyenne, de faire croître la ligne d'assurance i afin de se rapprocher de la frontière efficiente
- Plus Δ_i est inférieur à 1, plus il est recommandé, en moyenne, de faire diminuer la ligne d'assurance i afin de se rapprocher de la frontière efficiente
- Plus Δ_i est proche de 1, plus il est recommandé, en moyenne, de maintenir la ligne d'assurance i à son niveau actuel afin de se rapprocher de la frontière efficiente

Une simplification supplémentaire peut en outre être effectuée; il suffit de classer les lignes d'assurance en trois catégories selon la valeur de leur facteur agrégé Δ_i .

Ainsi, en fixant les seuils $\Delta_i^{(croitre)}$ et $\Delta_i^{(decroitre)}$, la classification suivante peut être effectuée :

- Si le test de significativité ne permet pas de rejeter l'hypothèse nulle :
la ligne i est classée dans la catégorie "non significatif"
- Si $\Delta_i > \Delta_i^{(croitre)}$ et que l'hypothèse nulle est rejetée :
la ligne i est classée dans la catégorie "croître"
- Si $\Delta_i < \Delta_i^{(decroitre)}$ et que l'hypothèse nulle est rejetée :
la ligne i est classée dans la catégorie "croître"
- Si $\Delta_i \in [\Delta_i^{(decroitre)}, \Delta_i^{(croitre)}]$ et que l'hypothèse nulle est rejetée :
la ligne i est classée dans la catégorie "maintenir"

Intuitivement, cette approche permet notamment de faire passer dans la catégorie "maintenir" les lignes pour lesquelles le facteur $\delta_{i,q}$ change beaucoup selon la position sur la frontière efficiente q . Ainsi, il est possible d'obtenir une vision synthétique et simple de la frontière efficiente, qui peut se résumer à une seule valeur par ligne d'assurance (c.f. [Figure 5.4](#)).

Il convient cependant de noter que cette approche ne prend pas en compte l'intervalle

de dispersion des facteurs δ . Une manière d'intégrer cet aspect des résultats sans augmenter le nombre d'indicateurs serait de changer la méthode de classification.

Un exemple d'une telle méthode serait de considérer que la ligne est à croître si le facteur agrégé $\Delta_i > 1$ et à décroître si $\Delta_i < 1$, mais de fixer un seuil de dépassement maximal d :

- Si $\Delta_i > 1$, mais que $\sum_{q=1}^Q \mathbf{1}(\text{Quantile}_{i,q}^{inf} < 1) > d$:
La ligne i passe dans la catégorie "maintenir"
- Si $\Delta_i < 1$, mais que $\sum_{q=1}^Q \mathbf{1}(\text{Quantile}_{i,q}^{sup} > 1) > d$:
La ligne i passe dans la catégorie "maintenir"

Intuitivement, cette approche permet aussi de faire passer dans la catégorie "maintenir" les lignes pour lesquelles la dispersion de $\delta_{i,q}$ selon les itérations $t \in \{1, \dots, T\}$ est telle qu'il est difficile de fournir une recommandation unique.

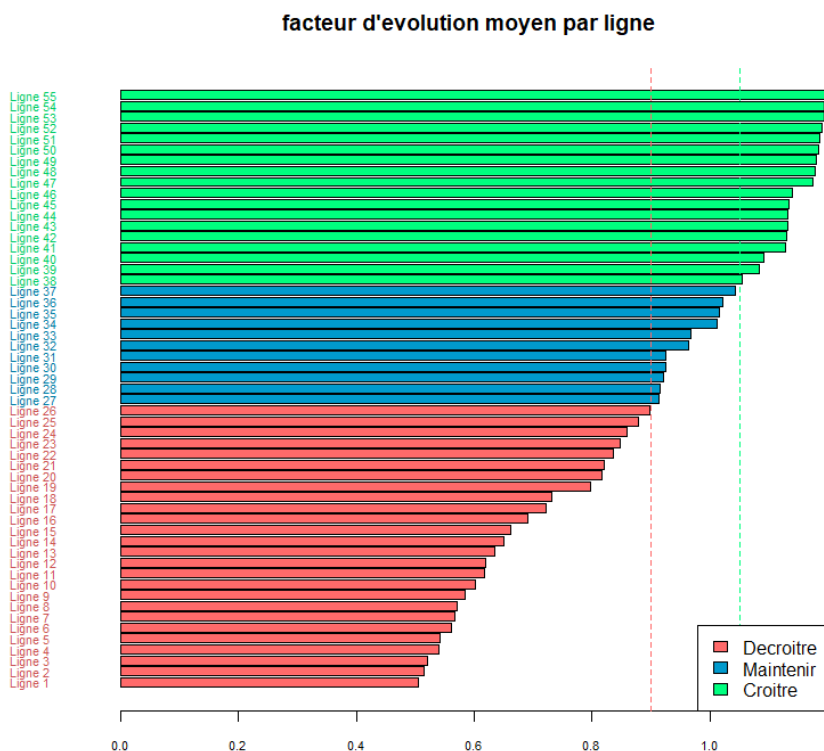


FIGURE 5.1 – Synthèse des Δ par ligne d'assurance sur la frontière efficiente totale

5.1.2 L'analyse empirique des résultats

Maintenant que les résultats de l'algorithme pour chaque ligne d'assurance ont été synthétisés en une classification simple, il convient d'étudier ces recommandations afin de déterminer leur pertinence.

Pour ce faire, la [Figure 5.2](#) représente la contribution de chaque ligne au coût en capital et au rendement du portefeuille initial en fonction de leur recommandation. Le rendement étant linéaire, la contribution au rendement peut être calculée directement. Ce n'est cependant pas le cas de coût en capital, et c'est pourquoi sa contribution est calculée en réallouant

le coût en capital total entre les différentes lignes. Pour ce faire, il est possible d'utiliser la méthode d'allocation de la covariance, qui repose sur la propriété suivante :

Si $(X_i)_{i \geq 1}$ est une suite de variables aléatoires, alors :

$$Variance \left(\sum_{i=1}^n X_i \right) = \sum_{i=1}^n Cov \left(X_i, \sum_{j=1}^n X_j \right)$$

Il est donc possible d'allouer la variance de chaque variable aléatoire telle que :

$$Variance \left(\sum_{i=1}^n X_i \right) = Variance \left(\sum_{i=1}^n X_i \right) \times \sum_{i=1}^n \frac{Cov \left(X_i, \sum_{j=1}^n X_j \right)}{Variance \left(\sum_{i=1}^n X_i \right)}$$

Appliquer cette méthode au portefeuille d'intérêt permet de calculer la contribution de chaque ligne au coût en capital de la manière suivante :

$$\begin{cases} ContribRdmt_i = \frac{Rdmt_i(1)}{Rdmt^{(p)}(1)} \\ ContribCeC_i = \frac{Cov(CeC_i(1), CeC^{(p)}(1))}{CeC^{(p)}(1)} \end{cases}$$

L'analyse de cette figure révèle que les lignes de catégorie croître sont concentrées dans le quadrant supérieur gauche, tandis que les lignes de catégorie "décroître" apparaissent dans le quadrant inférieur droit. Les exceptions principales à cette règle sont les lignes situées en dessous de l'axe des abscisses, pour lesquelles la recommandation est de décroître de manière quasi-systématique.

Ces résultats correspondent donc exactement aux attentes. En effet, l'objectif est de générer du rendement tout en maîtrisant le risque. Or, plus une ligne est située haut sur cette figure, plus son rendement est élevé, tandis qu'une ligne située vers la droite signifie que sa contribution au coût en capital de l'ensemble du portefeuille est important.

Cette analyse permet donc de vérifier la pertinence de l'approche développée ici de manière empirique, en confrontant les résultats de l'algorithme aux résultats attendus.

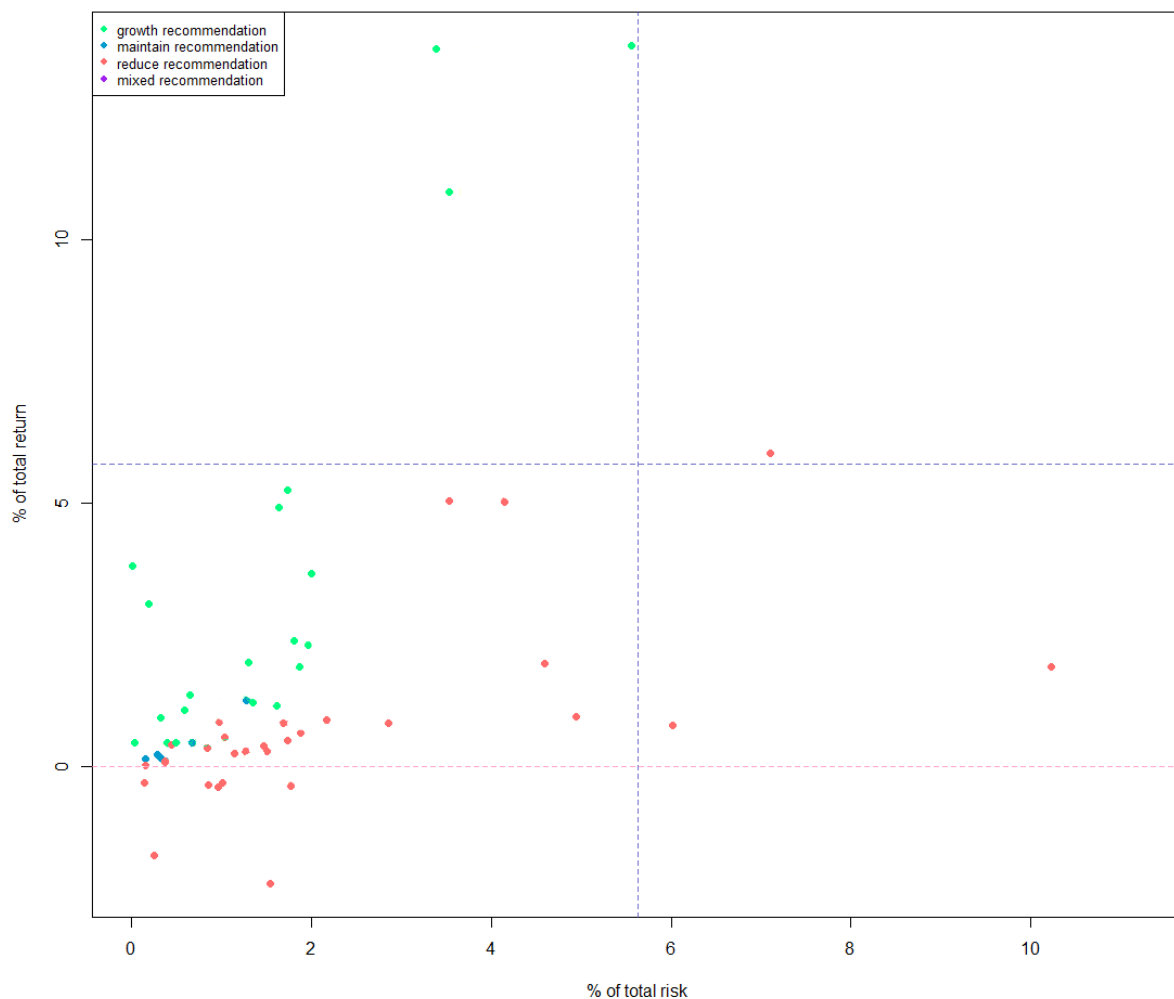


FIGURE 5.2 – Profil rendement/risque de chaque ligne d'assurance en fonction de la recommandation de l'algorithme

L'attrait principal de l'agrégation des résultats de la totalité de la frontière efficiente est donc la simplicité. Obtenir un résultat unique par ligne d'assurance permet notamment de faire ressortir de manière immédiate les lignes d'assurance les plus performantes, ainsi que celles qui le sont le moins.

Cependant, cette simplicité ne peut se faire qu'au prix d'une grande quantité d'informations. Si une ligne $i \in \{1, \dots, n\}$ donnée est classée dans la catégorie "maintenir", cette approche ne permet par exemple pas de faire la différence entre les situations suivantes :

1. Les cas où le facteur d'évolution $\mathbb{E}[\delta_{i,q}^*]$ est proche de 1 pour toutes les positions sur la frontière efficiente $q \in \{1, \dots, Q\}$
2. Les cas où, selon la position q , le facteur d'évolution $\mathbb{E}[\delta_{i,q}^*]$ passe de largement supérieur à 1 à largement inférieur à 1

Dans le cas d'une ligne d'assurance dans la deuxième catégorie, cela signifie que la décision "maintenir" peut en réalité n'être optimale que sur un segment très réduit de la frontière efficiente. De la sorte, appliquer cette règle de décision de manière uniforme à

toutes les lignes d'assurance revient finalement à choisir une combinaison de δ à appliquer qui peut être éloignée de tout point q sur la frontière efficiente. Cette combinaison serait donc sous-optimale.

Afin de pallier cette difficulté, un second type d'approche doit notamment être abordé.

5.2 Segmenter la frontière efficiente selon les besoins

En tenant de synthétiser la totalité de la frontière efficiente en un seul indicateur par ligne, un certain nombre de nuances au sein des résultats sont nécessairement mises à l'écart. C'est par exemple le cas dès lors qu'une ligne i voit son facteur d'évolution $\mathbf{E}[\delta_{i,q}^*]$ changer de façon significative selon la position sur la frontière efficiente q .

Pour pallier cette difficulté, deux types d'approches peuvent être envisagées. Une première approche revient à augmenter le nombre d'indicateurs retenus par ligne d'assurance, au prix d'un niveau de complexité plus important. Une deuxième approche serait de renoncer à synthétiser la totalité des informations présentes sur la frontière efficiente, et de ne se concentrer que sur la partie de la frontière efficiente la plus intéressante.

Dans le cas de cette première approche, présentée en section 5.2.1, il est possible de procéder en divisant la frontière efficiente en un certain nombre de sous-segments à analyser de manière successive.

Dans le cas de cette deuxième approche présentée en section 5.2.2, un indicateur peut être mis au point de manière à déterminer immédiatement quel segment est le plus intéressant. Ici, l'indicateur retenu sera construit de manière analogue au *Ratio Sharpe*, souvent utilisé en analyse financière.

5.2.1 La division de la frontière efficiente en plusieurs sous-segments

L'objectif de la séparation de la frontière efficiente en plusieurs sous-segments et de pallier les difficultés mises en avant lors de la précédente règle de décision, tout en conservant une méthodologie simple à appliquer en pratique. En effet, la totalité de ce nouveau processus de décision peut se résumer en deux étapes :

L'identification des sous-segments de la frontière efficiente

Afin d'identifier différents intervalles d'analyse sur la frontière efficiente, il suffit de la segmenter selon les quantiles empiriques. Ainsi, en notant (A_1, \dots, A_μ) les $\mu \in \mathbb{N}$ segments souhaités, leur définition est la suivante :

$$\begin{aligned}
A_1 &= \left\{ q \in [1, Q] \mid Rdm t_{(q)}^* \leq Rdm t_{(\lfloor Q \frac{1}{\mu} \rfloor)}^* \right\} \\
A_2 &= \left\{ q \in [1, Q] \mid Rdm t_{(\lfloor Q \frac{1}{\mu} \rfloor)}^* < Rdm t_{(q)}^* \leq Rdm t_{(\lfloor Q \frac{2}{\mu} \rfloor)}^* \right\} \\
&\vdots \\
A_\mu &= \left\{ q \in [1, Q] \mid Rdm t_{(\lfloor Q \frac{\mu-1}{\mu} \rfloor)}^* < Rdm t_{(q)}^* \leq Rdm t_{(Q)}^* \right\}
\end{aligned}$$

Ainsi, l'intervalle A_1 correspond à l'intervalle pour lequel les valeurs de Rendement et de Coût en Capital sont les plus faibles, tandis que A_μ correspond à celui pour lequel ces valeurs sont plus élevées.

Ainsi, avec par exemple $\mu = 3$, cela revient à identifier un intervalle concentré sur la réduction du risque, un intervalle concentré sur l'augmentation du rendement, et un intervalle médian (c.f. [Figure 5.3](#)).

L'analyse de chacun de ces segments

Maintenant que les segments d'intérêt (A_1, \dots, A_μ) ont été définis, il reste donc à synthétiser leurs informations. Pour ce faire, il suffit de reprendre l'analyse par la moyenne proposée en section 5.1.

De la sorte, μ ensembles de recommandations sont obtenus, ordonnés selon le niveau de rendement et de coût en capital qu'ils permettent d'atteindre.

En appliquant cette procédure au portefeuille AXA avec $\mu = 3$, il apparaît que plus le segment de frontière efficiente analysé est concentré sur le rendement, plus le nombre de lignes classées dans la catégorie "maintenir" ou "croître" est élevé.

Ce résultat correspond donc au résultat attendu d'une telle analyse. En effet, sous l'hypothèse qu'une ligne i soit rentable en moyenne, l'inclure dans le portefeuille d'assurance augmentera mécaniquement non seulement le rendement, mais aussi le risque. Ainsi, un facteur d'évolution δ supérieur à 1 pour une telle ligne aura le même effet.

En outre, si une ligne i change de recommandation entre deux segments, il est possible de postuler la conjoncture suivante :

- Si la ligne i passe de la catégorie "croître" à "décroître", alors cette ligne génère très peu de risque pour le portefeuille, mais a un rendement très faible, voir négatif.
- Si la ligne i passe de la catégorie "décroître" à "croître", alors c'est une ligne très risquée ou qui se diversifie mal, mais qui est particulièrement rentable.

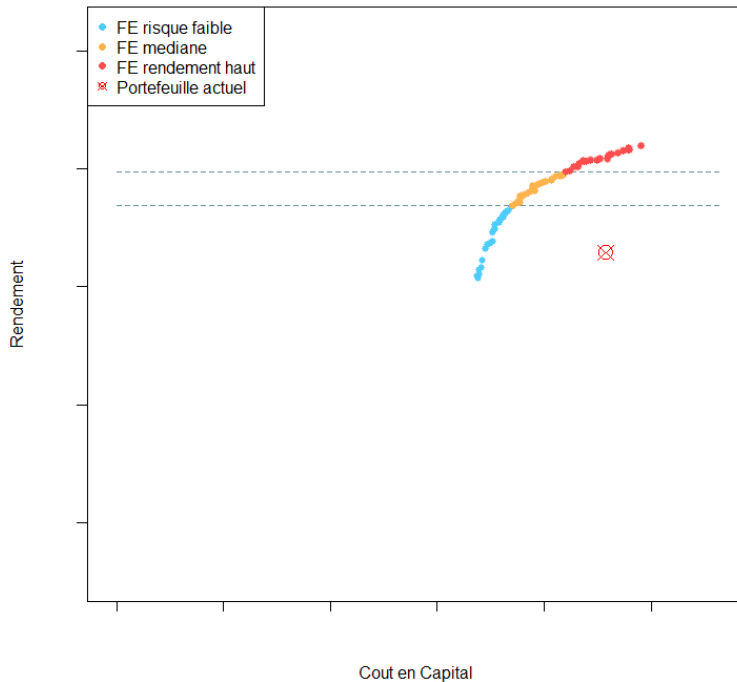


FIGURE 5.3 – Division de la frontière efficace en trois segments

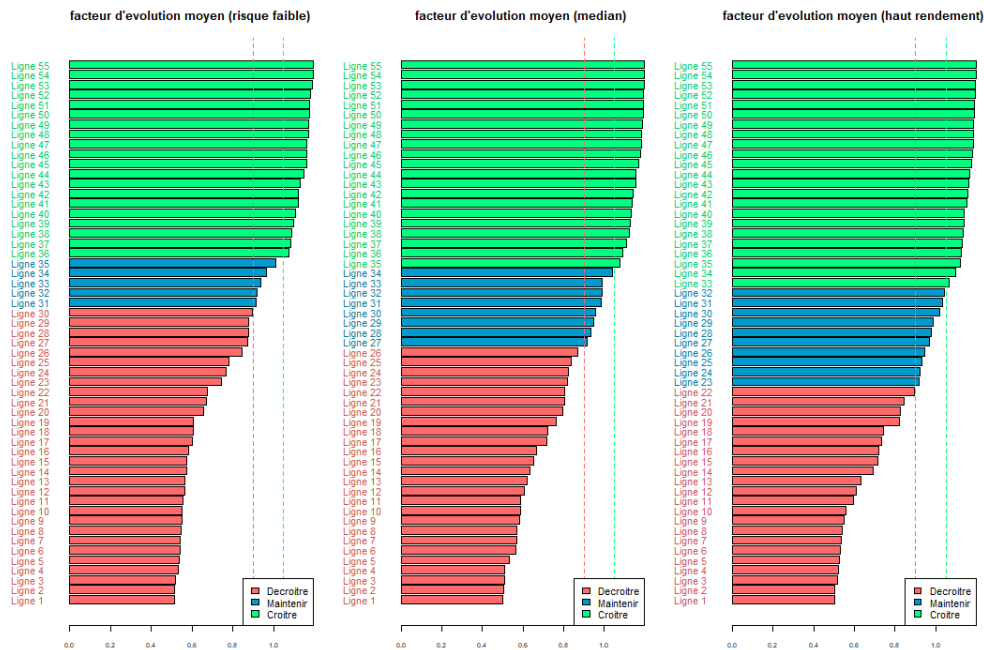


FIGURE 5.4 – Synthèse des Δ par ligne d'assurance sur différents segments de frontière efficace

Ainsi, le processus proposé consistant à diviser la frontière efficace en plusieurs segments d'intérêts présente de nombreux avantages sur le plan de l'interprétation des résultats ou de la prise de décision.

Tout d'abord, c'est une méthode qui reste relativement simple à appliquer, puisqu'elle

ne revient qu'à ajouter une étape au processus décrit en section 5.1.

En outre, des l'analyse des différences entre ces segments permet de déterminer aisément quel type d'impact une ligne d'assurance peut avoir sur le portefeuille.

Enfin, cette approche permet à l'entreprise d'assurance de déterminer quelle priorité poursuivre lors de l'optimisation du portefeuille. Si l'objectif est avant tout de réduire le risque, la liste de recommandation à suivre se rapprochera par exemple du segment peu risqué.

Si l'intérêt de la décomposition de la frontière efficiente en plusieurs segments est maintenant démontrée, cette méthode présente toutefois l'inconvénient d'augmenter la complexité de la règle de décision. En particulier, là où l'approche initiale permettait de synthétiser les informations en un indicateur par ligne d'assurance, ici, μ indicateurs sont donnés.

Dès lors, l'enjeu est de proposer une autre manière d'interpréter les résultats, qui permet notamment d'arbitrer entre les différents segments de frontière efficiente possibles.

5.2.2 La restriction de la frontière efficiente au segment maximisant le Ratio Sharpe

Jusqu'à présent, l'analyse des résultats s'est effectuée sous l'hypothèse qu'il était impossible d'arbitrer *a priori* entre les différents points sur la frontière efficiente. En effet, par définition, dès lors qu'un portefeuille d'assurance est situé sur cette frontière, il est impossible d'en augmenter le rendement en jouant sur les facteurs d'évolution δ sans en détériorer le risque, ou inversement.

Cependant, si un arbitrage théorique est impossible, un arbitrage pratique apparaît dès lors que la variation de ces quantités le long de la frontière efficiente est considérée. C'est ce constat qui a motivé l'identification d'un indicateur capable d'arbitrer entre ces différentes possibilités.

Le raisonnement est le suivant :

Soit p_0 un portefeuille situé sur la frontière efficiente. Plusieurs cas peuvent alors être distingués :

- Si $\frac{Rdmt(p_0)}{CeC(p_0)} < \max_p \left(\frac{Rdmt(p)}{CeC(p)} \right)$, un déplacement vers la droite de la courbe permet d'augmenter le rendement plus vite que le coût en capital
- Si $\frac{Rdmt(p_0)}{CeC(p_0)} > \max_p \left(\frac{Rdmt(p)}{CeC(p)} \right)$, un déplacement vers la droite de la gauche permet de diminuer le rendement moins vite que le coût en capital (cf. [Figure 5.5](#))

Dès lors, le point sur la frontière efficiente qui permet de maximiser les gains de rendement, tout en maîtrisant le gain de risque est le point q_0 , défini tel que :

$$\frac{Rdmt_{q_0}^*}{CeC_{q_0}^*} = \max_q \left(\frac{Rdmt_q^*}{CeC_q^*} \right)$$

Cet indicateur est notamment analogue au ratio Sharpe en analyse financière, qui permet de classer plusieurs portefeuilles entre eux en maximisant le rapport entre leur rendement et leur risque [24].

Une difficulté supplémentaire se doit toutefois d'être relevée. Ici, afin de stabiliser les paramètres δ , l'algorithme NSGA-II a été répété de manière à obtenir T frontières efficaces. Or, il est difficile de garantir que ce point soit proche pour toutes les frontières.

Afin de résoudre cette complexité, plutôt que de ne considérer que le point qui maximise le ratio Sharpe dans l'analyse, c'est ici l'intervalle des p points qui l'entoure qui sera retenu (cf. Figure 5.6).

Ainsi, en notant $q_0^{(t)}$ le point qui maximise le ratio rendement / coût en capital de l'itération t , les intervalles retenus pour chaque itération de l'algorithme NSGA-II $(B^{(t)})_{1 \leq t \leq T}$ peuvent être définis par :

$$B^{(t)} = \left\{ q \in [1, Q] \mid Rdmt_{(q_0^{(t)} - (p/2))}^{(t)} \leq Rdmt_{(q)}^{(t)} \leq Rdmt_{(q_0^{(t)} + (p/2))}^{(t)} \right\}$$

Une fois les intervalles d'intérêt ainsi définis pour chaque itération $t \in \{1, \dots, T\}$, l'analyse peut donc se faire en reprenant la méthode par la moyenne proposée en section 5.1. De la sorte, il est possible de conjuguer les avantages des deux précédentes méthodes. Tout d'abord, c'est ici un vecteur unique de recommandations pour chaque ligne qui est identifié. Ensuite, l'analyse est restreinte à un segment de frontière efficace suffisamment restreint pour diminuer le risque de variations importantes des paramètres δ sur le segment.

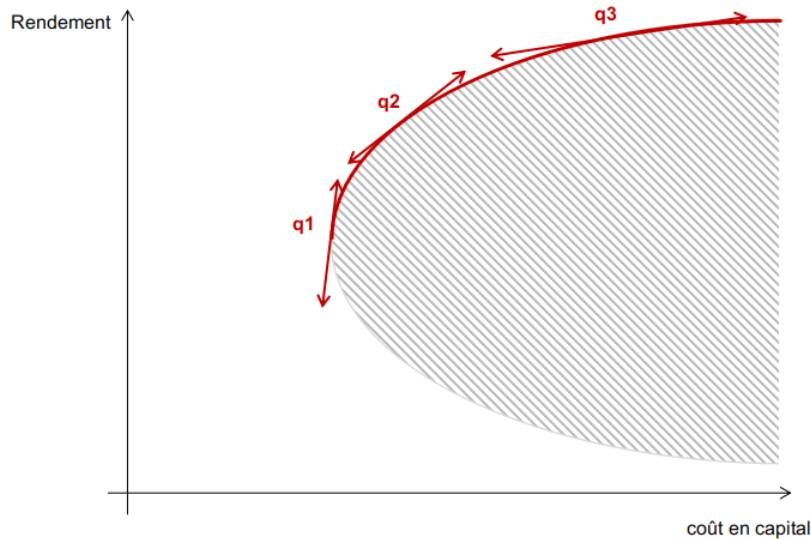


FIGURE 5.5 – Illustration des gains de performance des portefeuilles selon la position sur la frontière efficace

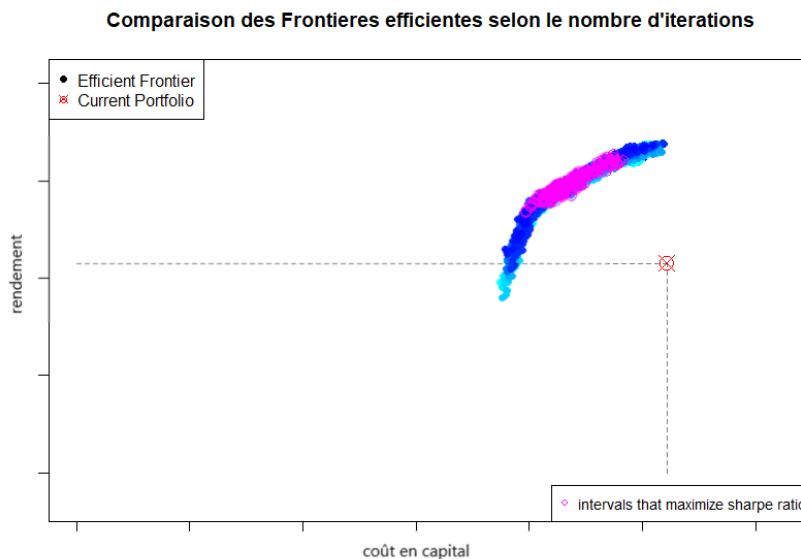


FIGURE 5.6 – Intervalles maximisant le ratio Sharpe sur les optima du NSGA-II

En maximisant le ratio Rendement / Coût en capital, il est donc possible d'obtenir une recommandation unique pour chaque ligne d'assurance à partir des résultats de l'algorithme développé lors de cette étude.

Cependant, contrairement aux deux méthodes précédentes, une partie des combinaisons de la frontière efficiente sont ignorées. Cette approche ne peut par conséquent pas être adaptée selon les besoins précis de l'assureur.

In fine, l'objectif de cette partie n'était pas de fournir une liste exhaustive des manières d'interpréter les informations présentées par les résultats de l'algorithme, mais simplement de donner des exemples d'approches qui permettraient de passer de l'algorithme théorique à une décision pratique.

De même, si les exemples ci-dessus permettent certes d'aider au processus de prises de décision d'une entreprise d'assurance, ils ne le remplacent pas. L'impact précis d'une évolution de la taille d'une ligne d'assurance sur l'activité totale d'un assureur est en effet de nature complexe, et un certain nombre de facteurs ont donc été écartés lors de cette étude.

Le périmètre précis d'application de ce modèle doit donc être établi. De la sorte, ses limites théoriques peuvent être exposées, et il sera possible d'identifier des pistes d'évolution ou d'amélioration.

Chapitre 6

Limites et perspectives d'évolution du modèle

Lors de cette étude, l'objectif était de développer une approche qui permette d'optimiser un portefeuille d'assurance de manière algorithmique selon une dimension de risque et une dimension de rendement. Pour ce faire, un modèle inspiré des travaux d'analyse financière a pu être construit, qui permet notamment de déterminer les tailles optimales théoriques des lignes d'assurance qui composent le portefeuille.

Cependant, en pratique, certaines limites inhérentes à la modélisation compliquent l'application immédiate de ses résultats par l'assureur. Par exemple, lors de l'écriture du programme d'optimisation, un certain nombre d'hypothèses explicites ont dû être posées afin de pouvoir déterminer comment évoluaient le rendement et le coût en capital en fonction de la taille d'une ligne d'assurance.

Une hypothèse implicite de stationnarité fut aussi posée. En effet, l'analyse réalisée jusqu'à présent écarte la possibilité d'un changement de l'environnement économique qui se répercuterait sur le portefeuille. De même, si un écart entre la sinistralité observée et la sinistralité modélisée rendait nécessaire la re-calibration du modèle interne utilisé pour calculer les indicateurs de risque et de rendement, cela serait susceptible de modifier les conclusions du modèle.

En outre, s'il est souvent relativement aisé de changer la composition d'un portefeuille financier, augmenter ou réduire la taille d'une ligne d'assurance peut avoir de nombreuses conséquences, qui n'ont jusqu'à présent pas été approfondies.

C'est pourquoi cette étude s'achèvera sur une analyse des limites du modèle d'optimisation tel qu'il a été développé au cours des précédentes parties. De la sorte, il sera non seulement possible de rappeler le périmètre précis d'application du modèle, mais aussi d'identifier des axes d'évolution et d'éventuels approfondissements.

Pour ce faire, la section reviendra 6.1 sur les hypothèses explicites posées lors du calcul des indicateurs de rendement et de coût en capital.

Afin de pallier la difficulté à se passer de ces hypothèses, une première approche sera de s'intéresser à une modélisation alternative à celle proposée lors de cette étude. En particulier, la question de l'optimisation de la réassurance plutôt que de la taille des lignes d'assurance sera explorée en section 6.2.

Ensuite, une approche permettant de mieux prendre en compte la sensibilité du modèle aux évolutions du risque inhérents aux lignes d'assurance ou du contexte économique pourra être exposée en section 6.3.

Enfin, la section 6.4 présentera d'autres considérations permettant d'affiner l'analyse effectuée pourront être présentées.

6.1 Analyse des hypothèses posées

Afin de résoudre un programme d'optimisation, il est nécessaire de définir non seulement la fonction objectif à minimiser, mais aussi les paramètres dont elle dépend. Ici, les paramètres correspondent aux évolutions possibles de l'exposition au risque des lignes d'assurance, notés $(\delta_1, \dots, \delta_n)$, tandis que la fonction objectif comprend la dimension du rendement et la dimension du coût en capital.

Ainsi, lors de l'écriture du programme d'optimisation réalisée en partie 2, un des enjeux fondamentaux était de déterminer comment les fonctions $Rdmt(\cdot)$ et $CeC(\cdot)$ évoluaient en fonction des paramètres δ .

Pour ce faire, certaines hypothèses ont cependant dû être posées. En particulier, il a été postulé que la sinistralité, les primes et les frais bruts de réassurance évoluaient de manière linéaire en fonction de l'exposition au risque. En pratique, il s'agit d'hypothèses très fortes, et il est nécessaire d'examiner leur impact sur les conclusions de l'étude.

6.1.1 Les imprécisions des hypothèses de linéarité

Si les imprécisions liées aux hypothèses de linéarité ont été brièvement mentionnées lorsque celles-ci ont été posées en partie 2.1.3, il est possible de les approfondir davantage. Les hypothèses de linéarité reviennent à supposer que les primes et les sinistres évoluent de manière similaire dès lors qu'un assureur tente d'accroître ou de décroître la taille d'une ligne d'assurance dans son portefeuille. Elles peuvent donc se résumer de la manière suivante :

$$\frac{Prime_i(\delta_i)}{Prime_i(1)} = \frac{Sinistres_i(\delta_i)}{Sinistres_i(1)} = \delta_i$$

Intuitivement, cela revient à supposer qu'il est possible de multiplier le nombre de contrats d'homogènes de la ligne i par δ_i sans modifier les termes du contrat. En effet, en notant C_i le nombre de contrats associés à cette ligne, les équations suivantes sont vérifiées :

$$\left\{ \begin{array}{l} Primes_i^{brut} = \sum_{k=1}^{C_i \times \delta_i} Primes_{i,k}^{brut} \\ Sinistres_i^{brut} = \sum_{k=1}^{C_i \times \delta_i} Sinistres_{i,k}^{brut} \end{array} \right.$$

Ainsi, si les vecteurs aléatoires $(Primes_{i,k}^{brut})_k$ et $(Sinistres_{i,k}^{brut})_k$ sont des vecteurs de variables indépendantes et identiquement distribuées, et que leur loi ne dépend pas de δ :

$$\left\{ \begin{array}{l} \frac{\mathbb{E}[Prime_i(\delta_i)]}{\mathbb{E}[Prime_i(1)]} = \frac{\sqrt{Variance[Prime_i(\delta_i)]}}{\sqrt{Variance[Prime_i(1)]}} = \delta_i \\ \frac{\mathbb{E}[Sinistres_i(\delta_i)]}{\mathbb{E}[Sinistres_i(1)]} = \frac{\sqrt{Variance[Sinistres_i(\delta_i)]}}{\sqrt{Variance[Sinistres_i(1)]}} = \delta_i \end{array} \right.$$

Cependant, en pratique, il est irréaliste de considérer que les primes et sinistres individuels par contrat $(Primes_{i,k}^{brut})_k$ et $(Sinistres_{i,k}^{brut})_k$ ne dépendent pas de δ . Cela est notamment dû au fait qu'il peut être extrêmement difficile de gagner des parts de marché sans avoir recours à une politique tarifaire agressive. Autrement dit, toute augmentation significative du nombre de contrats ne peut se faire qu'à condition d'accepter de réduire la prime.

Ainsi, dans le cas où le nouveau nombre de contrats C'_i est supérieur à C_i ($\delta_i > 0$), le montant des primes des nouveaux contrats est inférieur à celui des contrats précédents. En notant $Primes_{i,k}^{brut'}$ le nouveau montant de primes par contrat, l'inéquation suivante est donc vérifiée :

$$\begin{aligned} \forall k > C_i, Primes_{i,k}^{brut'} &< Primes_{i,k}^{brut} \\ \implies \sum_{k=C_i}^{C'_i} Primes_{i,k}^{brut'} &< \sum_{k=C_i}^{C'_i} Primes_{i,k}^{brut} \\ \implies \sum_{k=C_i}^1 Primes_{i,k}^{brut'} &< \sum_{k=C_i}^1 Primes_{i,k}^{brut} \\ \implies Primes_i^{brut}(\delta_i) &< \times \delta_i Primes_i^{brut}(1) \end{aligned}$$

Un raisonnement analogue peut être effectué dans le cas des sinistres. En effet, si un assureur souhaite acquérir un nombre supplémentaire de contrats en limitant la diminution de la prime, alors il est nécessaire d'accepter de souscrire des contrats plus risqués pour un niveau de primes équivalent.

Ainsi, en notant $Sinistre_{i,k}^{brut'}$ le nouveau montant de sinistres par contrat :

$$\begin{aligned} \forall k > C_i, Sinistre_{i,k}^{brut'} &> Sinistre_{i,k}^{brut} \\ \sum_{k=C_i}^{C'_i} Sinistre_{i,k}^{brut'} &> \sum_{k=C_i}^{C'_i} Sinistre_{i,k}^{brut} \\ \sum_{k=1}^{C'_i} Sinistre_{i,k}^{brut'} &> \sum_{k=1}^{C'_i} Sinistre_{i,k}^{brut} \\ \implies Sinistres_i^{brut}(\delta_i) &> \times \delta_i Sinistres_i^{brut}(1) \end{aligned}$$

La conséquence de cette non-linéarité est donc que, dans le cas où $\delta_i > 1$:

$$\frac{Prime_i(\delta_i)}{Prime_i(1)} < \frac{Sinistres_i(\delta_i)}{Sinistres_i(1)}$$

Le cas où $\delta_i < 1$ est toutefois plus nuancé. Il serait certes intuitif d'effectuer le raisonnement inverse à celui effectué lors d'une augmentation du nombre de contrats, en considérant que :

- Si l'objectif est de réduire le nombre de contrats, il est possible d'augmenter le taux de primes jusqu'à ce que seuls le nombre de contrats désirés ne soit pas résilié
- De même, limiter la souscription aux contrats moins risqués pour un niveau de primes équivalent permet de diminuer le montant moyen des sinistres des contrats restants

L'examen des hypothèses de linéarité révèle donc que hypothèse revient donc à surestimer la capacité de l'assureur à dégager du rendement et à diminuer le coût en capital en jouant sur les facteurs d'évolution δ . Il est donc probable qu'une modélisation plus complète du problème d'optimisation produise une courbe de frontière efficiente qui soit plus proche du portefeuille initial que celle obtenue lors de cette étude.

6.1.2 Un exemple de modélisation plus proche de la réalité

La réévaluation des hypothèses de linéarité révèle donc que ces hypothèses sont sans doute trop optimistes quant à l'effet d'une augmentation du nombre de contrats sur le profil rendement/coût en capital du portefeuille. Dès lors, un axe d'amélioration des travaux présentés lors de cette étude serait de modéliser l'impact non-linéaire du facteur δ sur les primes et les sinistres.

L'analyse réalisée ci-dessus révèle donc qu'une telle modélisation doit respecter les conditions suivantes :

- si $\delta_i > 1$, $Sinistres_i^{brut}(\delta_i) > \delta_i \times Sinistres_i^{brut}(1)$
- si $\delta_i > 1$, $Primes_i^{brut}(\delta_i) < \delta_i \times Primes_i^{brut}(1)$

Pour ce faire, une modélisation possible serait ici de prendre en compte un paramètre d'ajustement des primes a_i^{primes} et un paramètre d'ajustement des sinistres $a_i^{sinistres}$, tels que :

$$\begin{aligned} Sinistres_i^{brut}(\delta_i) &= \delta_i \times (1 + a_i^{sinistres}(\delta_i)) \times Sinistres_i^{brut}(1) \\ Primes_i^{brut}(\delta_i) &= \delta_i \times (1 + a_i^{primes}(\delta_i)) \times Primes_i^{brut}(1) \end{aligned}$$

Dès lors, la question est de savoir comment évoluent a_i^{primes} et $a_i^{sinistres}$ en fonction de (δ_i) . Il est ici possible de distinguer plusieurs cas :

- si a_i^{primes} (ou $a_i^{sinistres}$) se rapproche de 0 lorsque δ_i augmente, cela revient à considérer que plus l'exposition au risque augmente, moins il est nécessaire de d'appliquer une réduction supplémentaire de la prime afin de souscrire plus de contrats
- si a_i^{primes} (ou $a_i^{sinistres}$) s'éloigne de 0 lorsque δ_i augmente, cela revient à considérer que plus le nombre de contrats a augmenté, plus il est difficile d'augmenter encore davantage la souscription
- si a_i^{primes} (ou $a_i^{sinistres}$) est constant en fonction de δ_i , une diminution du taux de primes aura toujours un impact équivalent sur l'acquisition de contrats

La première intuition serait donc de considérer que a_i^{primes} et $a_i^{sinistres}$ s'éloignent de 0 lorsque δ_i augmente. Toutefois, c'est le fait de tenter de quantifier cette évolution qui révèle la difficulté principale liée à cette approche : la calibration des paramètres.

Cette calibration est en effet complexifiée par le fait que toute régression ne peut ici pas se faire sans précaution particulière. Par exemple, effectuer une régression du montant des primes sur le nombre de contrats en utilisant plusieurs années comme observations reviendrait à négliger la tendance du marché de l'assurance, comme le cycle de marché. À l'inverse, si les observations considérées sont les montants et volumes de contrats de plusieurs lignes d'assurance, ce sont les caractéristiques respectives des différentes lignes qui sont ignorées.

Ainsi, la réévaluation des hypothèses posées lors de cette étude révèle que ces hypothèses sont en réalité peu conservatrices, et qu'une modélisation alternative pourrait représenter plus fidèlement la réalité. Cependant, une telle modélisation ne peut avoir lieu sans une analyse supplémentaire approfondie de l'impact du volume de contrats sur la sinistralité et les primes individuelles. Cela nécessiterait toutefois un volume de données bien plus important et d'un spectre bien plus large. Ce n'est en effet que sous ces conditions qu'il serait possible de distinguer la tendance générale du marché de l'impact réel de la variation du nombre de contrats souscrits.

Dès lors, si les travaux réalisés répondent à la problématique de l'optimisation en fonction de la taille des lignes d'assurance, il est possible de s'interroger sur la possibilité de résoudre ce problème en fonction de paramètres alternatifs.

6.2 Optimisation par la réassurance plutôt que par la taille des lignes d'assurance

Si les hypothèses posées lors de l'écriture du programme d'optimisation sont difficiles à soutenir en réalité, c'est parce qu'elles supposent que l'assureur est en mesure de contrôler librement la taille des lignes d'assurance de son portefeuille. Or, si des variations marginales sont certes envisageables, des variations plus significatives sont difficiles à atteindre sans effectuer de changements structurels.

Dès lors, il convient de s'interroger sur la possibilité d'optimiser le profil rendement/risque du portefeuille d'assurance en jouant sur un paramètre dont l'assureur a un plus grand contrôle. Un tel levier d'action serait donc la réassurance, dont le principe même permet à l'assureur de maîtriser le risque. Si l'analyse approfondie d'une telle méthode dépasse le cadre des travaux réalisés ici, un exemple de mise en œuvre d'un programme d'optimisation en fonction de la réassurance peut être donné afin d'explorer un axe d'évolution possible de cette étude.

6.2.1 Programme d'optimisation en fonction de la réassurance

L'écriture du programme d'optimisation en fonction de la réassurance peut en réalité se faire en utilisant un raisonnement analogue à celui effectué lors des travaux présentés jusqu'ici. Certes, le paramètre en fonction duquel l'optimisation est réalisée diffère, mais tant que l'objectif reste l'augmentation du rendement sous condition de maîtrise du risque, les fonctions objectif du programme d'optimisation peuvent être conservées.

Ainsi, en notant γ un paramètre permettant de mesurer la réassurance, le programme d'optimisation à résoudre est le suivant :

$$\min_{\gamma} (Rdmt^{(p)}(\gamma), CeC^{(p)}(\gamma))$$

Dès lors, la question est de déterminer comment mesurer la réassurance γ , et quel impact ce paramètre a sur les fonctions objectif.

Définition du paramètre de réassurance γ

La difficulté principale lors de l'écriture de ce programme est la complexité que la structure de réassurance peut prendre. Un exemple de modélisation consisterait donc à simplifier le programme en ne considérant que des traités de type quote-part. Ainsi, pour la ligne d'assurance $i \in \{1, \dots, n\}$, l'impact de la réassurance serait déterminé par un taux de cession unique γ_i .

L'optimisation se ferait donc uniquement en fonction du vecteur de taux de cession $\{\gamma_1, \dots, \gamma_n\}$, qui capterait à lui seule la totalité de la réassurance étudiée.

L'impact de γ sur les fonctions objectif

Dès lors, l'enjeu serait de déterminer l'impact du paramètre γ sur les fonctions $Rdmt$ et Cec . Puisque les traités de quote-part ont un impact linéaire sur les primes et les sinistres, les équations suivantes seraient vérifiées :

$$\begin{aligned} Primes_i^{net}(\gamma_i) &= (1 - \gamma_i) \times Primes_i^{brut} \\ Sinistres_i^{net}(\gamma_i) &= (1 - \gamma_i) \times Sinistres_i^{brut} \end{aligned}$$

Le cas des frais peut être traité de manière analogue à la partie 3. Si les frais ne sont pas directement cédés au réassureur, la différence entre les frais bruts et la commission versée à la cédante peut s'interpréter comme des frais nets de réassurance.

Le taux de cette commission n'est *a priori* pas de γ , mais il est possible de poser l'hypothèse suivante sans impact significatif sur le résultat :

$$Frais_i^{net}(\gamma_i) = (1 - \gamma_i) \times Frais_i^{brut}$$

En reprenant les méthodes d'agrégation de la partie 2, les fonctions de rendement et de coût en capital après réassurance γ prennent donc les valeurs suivantes :

$$\begin{cases} Rdmt^{(p)}(\gamma) = \sum_{i=1}^n (1 - \gamma_i) Resultat_i \\ Cec^{(p)}(1) = \sum_{t=1}^T \frac{-(VaR_{0.005}[\sum_{i=1}^n (1 - \gamma_i) Resultat_{i,t}] - \mathbb{E}[\sum_{i=1}^n (1 - \gamma_i) Resultat_{i,t}])}{(1 + r)^t} \end{cases}$$

Il serait ensuite possible de résoudre le programme d'optimisation par une approche similaire reposant sur les algorithmes génétiques.

Ainsi, un problème d'optimisation alternatif en fonction de la réassurance serait une évolution possible de l'étude réalisée. Une telle approche permettrait donc d'optimiser le profil

rendement / risque d'un portefeuille d'assurance en jouant sur des facteurs plus aisément accessibles pour l'assureur, et d'éventuelles conclusions pourraient être plus faciles à appliquer.

6.2.2 Les potentielles limites d'une telle approche

Il convient toutefois de rappeler que l'approche donnée en exemple ici ne reste qu'une simplification de l'impact réel de la réassurance. En réalité, il serait nécessaire de considérer plusieurs problématiques essentielles afin de pouvoir obtenir une bonne interprétation de ce modèle.

L'impact de la réassurance non-proportionnelle

La problématique la plus évidente mise à l'écart dans l'approche définie ci-dessus est l'impact de la réassurance non-proportionnelle. En effet, si l'impact des traités de type *quote-part* est simple à modéliser, en pratique, optimiser la structure de réassurance passerait également par la mise en place de traités de type *Excédents de pertes* par sinistres ou par évènement.

Or, modéliser de tels traités n'est pas trivial. Tout d'abord, le calcul de l'impact de ces traités sur les sinistres se heurte au même problème que celui identifié en partie 3. Sans connaître la distribution des sinistres individuels qui caractérise une ligne d'assurance, l'impact d'un nouveau traité non-proportionnel peut être très difficile à estimer.

Ensuite, la définition même des paramètres d'optimisation est ici complexe. Un traité de type *Excédents de pertes* par sinistre est défini à l'aide de deux paramètres : la priorité et la limite. Si cette réassurance est prise en compte en plus des quote-parts, cela signifie que trois paramètres par ligne d'assurance sont nécessaires ; plus encore si des traités *Excédents de pertes* par évènements sont pris en compte.

La conséquence d'une multiplication des paramètres d'optimisation est donc non seulement la convergence plus lente et plus incertaine de l'algorithme génétique, mais aussi une interprétation plus difficile de l'impact de chaque paramètre.

La prise en compte de la structure de réassurance existante

Plutôt que d'écrire le programme d'optimisation de la réassurance en considérant que les nouveaux traités venaient s'appliquer à un portefeuille brut de réassurance, il est aussi possible de supposer qu'une structure de réassurance est déjà mise en place. Ce cas est en effet bien plus probable en pratique. Dès lors, la question est de déterminer quel impact cela aurait sur le programme d'optimisation de réassurance.

Deux possibilités peuvent ici être considérées. La première consiste simplement à considérer que les traités en fonction desquels l'optimisation a lieu viennent s'ajouter à la structure de réassurance actuelle. Autrement dit, il serait question d'ajouter davantage de réassurance à chaque ligne d'activité.

La seconde option serait à l'inverse de tenter d'optimiser la structure de réassurance existante. En effet, puisque les primes, sinistres et frais bruts et nets de réassurance sont connus, il est aisé de déterminer les primes, sinistres et frais cédés. Dès lors, l'objectif serait par exemple d'appliquer un facteur à la réassurance existante afin de déterminer l'impact d'une augmentation ou d'une diminution de la réassurance existante.

Une telle approche permettrait de pallier les problèmes identifiés avec l'écriture de ce programme. Non seulement cela permettrait-il d'éviter la multiplication des paramètres par ligne d'assurance, mais il serait en outre possible d'augmenter le rendement net moyen, par exemple en diminuant la réassurance existante sur des lignes rentables. Toutefois, ici encore, la question de l'assurance non-proportionnelle se pose. En pratique, un "doublement" de la réassurance serait simple à modéliser dans le cas d'un traité de *quote-part*, mais imprécis dans le cas d'un traité d'*excédent de pertes*.

Ainsi, une optimisation en fonction de la réassurance serait une modélisation alternative à celle proposée lors de cette étude. En particulier, une telle approche permettrait d'éviter de poser les hypothèses de linéarité qui peuvent conduire à une surestimation de la marge d'optimisation dans le cas des travaux initiaux.

Toutefois, la mise en place pratique d'un tel programme d'optimisation conduit à des difficultés supplémentaires. L'existence des traités non-proportionnels signifie en effet qu'il est difficile de procéder à une optimisation parfaite, du moins sans connaître la distribution des sinistres individuels.

Finalement, ces deux approches ne sont en réalité pas opposées, mais bien complémentaires. Plusieurs manières de conjuguer ces approches peuvent être considérées :

- Optimiser d'abord le rendement et coût en capital bruts, puis optimiser la structure de réassurance en appliquant de nouveaux traités
- Effectuer une première optimisation de la structure de réassurance, puis optimiser la taille des lignes d'assurance en considérant la nouvelle structure

6.3 L'analyse de la robustesse du modèle aux différentes données utilisées

En plus des hypothèses explicites posées lors de l'écriture du programme d'optimisation, la modélisation effectuée lors de cette étude repose en réalité aussi sur une hypothèse implicite : celle selon laquelle les données utilisées sont correctes. Cette supposition, qui peut sembler évidente, se doit pourtant d'être évaluée avec attention.

La base de données au cœur de cette étude est la sortie du modèle interne AXA, qui permet de déterminer la distribution de probabilités de diverses variables, telles le montant de primes ou la sinistralité. Les données utilisées sont donc en réalité des simulations, et

sont donc sujettes à deux types de volatilité :

- La volatilité liée aux simulations, c'est-à-dire le risque que l'échantillon des simulations tirées ne soit pas représentatif de la distribution de probabilité réelle
- La volatilité liée à la calibration, c'est-à-dire le risque que les lois de probabilité utilisées lors des simulations ne soient pas fidèles à la réalité

En plus de la base de données principale, certains paramètres économiques ont aussi été utilisés au cours de ce modèle. C'est par exemple le cas de la courbe des taux "sans-risques", utilisée afin d'actualiser le coût en capital. Si le risque que ce paramètre soit mal évalué à un instant donné est presque nul, puisqu'il s'agit d'une donnée de marché, sa valeur au cours du temps peut elle être amenée à évoluer.

Si ces données sont sources d'incertitudes pour la modélisation, prendre cet aspect en compte dans la modélisation est une tâche ardue. En effet, si la volatilité liée aux simulations peut être calculée si la loi de probabilité sous-jacente est connue, le risque de calibration ou la volatilité des données de marché peuvent être très difficile à quantifier.

L'enjeu est donc de déterminer comment prendre en compte les risques que ces incertitudes font peser sur la modélisation.

6.3.1 Tests de sensibilité traditionnels

Le concept des tests de sensibilité des résultats aux données est simple : si une incertitude existe sur un type de données particulier, il suffit de faire varier la valeur des données en question et de multiplier les itérations du modèle. De la sorte, il devient possible de déterminer un échantillon de valeurs possibles pour chaque résultat du modèle, et donc de déterminer la sensibilité du modèle aux données [23].

Afin d'illustrer cette méthode, l'exemple du risque de calibration sera considéré. Ici, R correspondra au nombre de réévaluations du modèle après avoir fait varier les données. Ainsi, soit $r \in \{1, \dots, R\}$ un indice de réévaluation, $i \in \{1, \dots, n\}$ une ligne d'assurance, et $s \in \{1, \dots, S\}$ un indice de simulation des variables de primes, de frais et de sinistres. Dans le cadre de cette étude, le risque de calibration peut par exemple être étudié en appliquant la transformation suivante aux données :

$$\begin{cases} Prime_{i,s}^r = Prime_{i,s} \times (1 + \epsilon_i) + \nu_{i,s} \\ Frais_{i,s}^r = Frais_{i,s} \times (1 + \epsilon_i) + \nu_{i,s} \\ Sinistres_{i,s}^r = Sinistres_{i,s} \times (1 + \epsilon_i) + \nu_{i,s} \end{cases}$$

Intuitivement, c'est ici le paramètre ϵ_i qui permet de simuler l'impact d'une loi de probabilité d'espérance différente, tandis que le paramètre $\nu_{i,s}$ permet de prendre en compte les variations marginales.

Appliquer cette méthodologie permettrait donc d'obtenir R vecteurs de résultats diffé-

6.3. L'analyse de la robustesse du modèle aux différentes données utilisées 103

rents, et donc de déterminer à quel point ceux-ci sont sensibles à une éventuelle erreur de calibration.

6.3.2 Modification du test de sensibilité pour éviter le cumul des itérations

Appliquer cette méthodologie dans le cadre de cette étude révèle cependant une difficulté particulière : le temps d'exécution. En effet, T itérations ont déjà été employées afin de stabiliser les résultats de l'algorithme génétique au cours de cette étude (c.f partie 4.3.1). Ainsi, si le paramètre d'actualisation r venait à prendre successivement les R valeurs $\{r_1, \dots, r_R\}$, le nombre d'évaluations de la fonction objectif serait alors de :

$$Q \text{ points sur la frontière} \times G \text{ générations} \times T \text{ itérations} \times R \text{ valeurs testées}$$

Il serait donc préférable d'éviter un tel cumul des itérations.

Pour ce faire, une méthode possible serait d'exécuter T premières itérations de l'algorithme NSGA-II, de manière identique à l'approche préconisée en partie 4.3.1. Puis, une fois une première distribution des paramètres δ obtenue, l'objectif serait de répéter ces T itérations, mais en appliquant la transformation ci-dessus à chaque itération. ainsi, le nombre d'itérations de l'algorithme génétique serait limité à $2 \times T$ plutôt qu'à $R \times T$.

Pour la ligne $i \in \{1; \dots, n\}$, deux vecteurs de résultats seraient donc obtenus :

$$\left\{ \begin{array}{l} \delta_i^{1*} = \left(\delta_{i,1t}^1, \dots, \delta_{i,Q_t}^1 \right)_{1 \leq t \leq T} \\ \text{pour lequel les données sont les données initiales} \\ \delta_i^{2*} = \left(\delta_{i,1t}^2, \dots, \delta_{i,Q_t}^2 \right)_{1 \leq t \leq T} \\ \text{pour lequel les données sont transformées à chaque itération } t \end{array} \right.$$

Ainsi, en comparant les distributions δ_i^{1*} et δ_i^{2*} , il est possible de déterminer dans quelle mesure d'éventuelles imprécisions lors de la calibration des données du modèle interne ont un impact important sur les résultats du modèle.

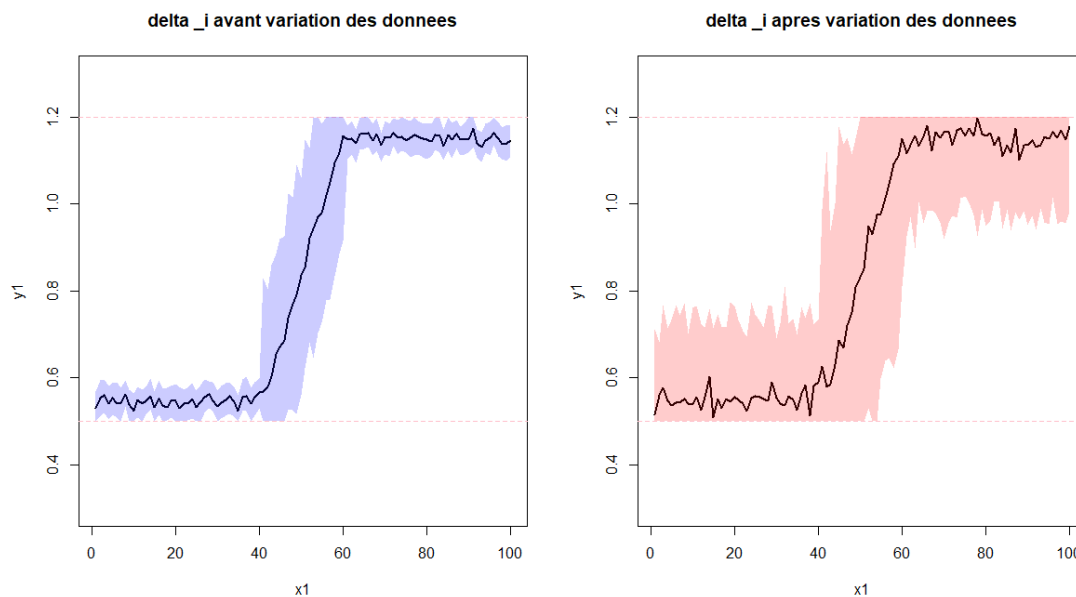


FIGURE 6.1 – Illustration théorique des comparaisons de résultats avec et sans variation des données

En outre, cette approche n'est pas limitée à la base de données sortie du modèle interne, mais pourrait être utilisée pour tester la sensibilité des résultats à plusieurs facteurs. Dans le cadre de cette étude, les facteurs d'intérêts seraient donc non seulement les données de primes, sinistres et frais, mais aussi la courbe des taux utilisée pour l'actualisation, et même les paramètres d'ajustement des hypothèses a_i^{primes} et $a_i^{sinistres}$ identifiés en partie 6.1.

La difficulté principale derrière la prise en compte de l'impact non-linéaire des facteurs d'évolution δ était en effet le fait que les paramètres précis pourraient être très difficiles à estimer. Toutefois, en laissant a_i^{primes} et $a_i^{sinistres}$ varier au sein d'intervalles donnés, il devient possible de quantifier leur impact possible sur le programme d'optimisation. En particulier, il est possible de déterminer si, quelles que soient les valeurs de ces paramètres, l'algorithme continue à identifier des solutions pour lesquelles le rendement et le coût en capital sont tous deux meilleurs que le portefeuille actuel.

Un axe d'évolution des travaux présentés ici serait donc les tests de sensibilité des résultats. En appliquant une telle approche, il serait non seulement possible de prendre en compte l'incertitude inhérente aux données, mais aussi de quantifier l'impact des hypothèses posées.

6.4 Autres considérations

Au cours de cette partie, la limite principale de cette étude a été identifiée comme étant les hypothèses de linéarité, posées au moment de l'écriture du programme d'optimisation. Suite à cela, une analyse supplémentaire a pu révéler que ces hypothèses, jugées nécessaires au moment de la réalisation des travaux, pouvaient tout de même être allégées de diverses

manières. Par exemple, considérer un programme d'optimisation en fonction d'autres paramètres que les facteurs d'évolution δ , ou réaliser des tests de sensibilité permettrait de rendre compte de leur impact sur le résultat.

Toutefois, la question de l'optimisation d'un portefeuille d'assurance est complexe, et l'impact réel ou la faisabilité d'un résultat se doivent donc d'être nuancés par plusieurs facteurs. C'est pourquoi cette partie s'achèvera sur une liste de différentes considérations jusque-là peu abordées.

6.4.1 Impact des délais d'ajustement et coûts de transaction

Contrairement au cadre d'un portefeuille financier, l'ajustement d'un portefeuille d'assurance est un processus de long-terme. Là où les actifs financiers généralement considérés sont des titres négociables, qui peuvent être échangés à tout moment sur les marchés financiers, les contrats d'assurance engagent en général l'assureur pour une durée déterminée. Dès lors, si l'objectif est d'optimiser un portefeuille d'assurance en fonction de la taille des lignes qui le composent, toute évolution $\delta \neq 1$ est en réalité soumise à de longs délais d'ajustement. Certains coûts de transaction existent certes aussi dans lors de l'optimisation d'un portefeuille financier, mais leur impact est donc bien moindre. Cet effet est en outre d'autant plus important que l'évolution requise est importante.

Cet aspect est donc difficile à intégrer directement au sein du modèle. C'est donc justement pour capter ce genre d'effets que les contraintes sur les facteurs d'évolution δ_i ont été fixées lors de l'écriture du programme d'optimisation :

$$\begin{aligned} \min_{\delta_{i(1 \leq i \leq n)}} & \left(-Rdmt^{(p)}(\delta_1, \dots, \delta_n), CeC^{(p)}(\delta_1, \dots, \delta_n) \right) \\ \text{s.c.} & \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n) \end{aligned}$$

Une seconde approche pour incorporer cet effet, bien que plus difficile à mettre en place de manière efficace, serait de pénaliser la fonction objectif dès lors que $\delta \neq 1$:

$$\begin{aligned} \min_{\delta_{i(1 \leq i \leq n)}} & \left(-(Rdmt^{(p)}(\delta_1, \dots, \delta_n) - f(\delta_1, \dots, \delta_n)), CeC^{(p)}(\delta_1, \dots, \delta_n) \right) \\ \text{s.c.} & \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n) \end{aligned}$$

La complexité d'une telle approche serait cependant la calibration de la fonction f , de sorte que la pénalisation soit fidèle à la réalité.

6.4.2 Limite du coût opérationnel

Ici, les coûts opérationnels sont distingués des coûts de transaction afin de mettre en lumière deux phénomènes différents. En effet, il peut être coûteux de mettre en place une telle politique d'évolution de la taille des lignes d'assurance à l'échelle de la totalité d'un

portefeuille. Il serait en pratique plus simple de ne s'intéresser qu'aux lignes d'assurance dont l'impact est le plus important.

Dès lors, une dimension supplémentaire pourrait être prise en compte au sein du programme d'optimisation : la notion de minimisation du nombre de lignes à optimiser.

Pour ce faire, plusieurs approches pourraient être envisagées. La première consiste à considérer directement le nombre de lignes optimisées comme une dimension du problème d'optimisation, qui s'écrirait donc de la manière suivante :

$$\min_{\delta_i (1 \leq i \leq n)} \left(-Rdmt^{(p)}(\delta_1, \dots, \delta_n), CeC^{(p)}(\delta_1, \dots, \delta_n), \sum_{i=1}^n \mathbf{1}(\delta_i \notin [1 - \epsilon, 1 + \epsilon]) \right)$$

$$s.c. \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n)$$

La deuxième serait de contraindre le programme d'optimisation, de sorte que l'algorithme génétique ne conserve que les meilleures solutions pour lesquelles le nombre de lignes optimisées est en dessous d'un certain seuil :

$$\min_{\delta_i (1 \leq i \leq n)} \left(-Rdmt^{(p)}(\delta_1, \dots, \delta_n), CeC^{(p)}(\delta_1, \dots, \delta_n) \right)$$

$$s.c. \beta_i^{(min)} \leq \delta_i \leq \beta_i^{(max)} \quad \forall i \in (1, \dots, n)$$

$$\sum_{i=1}^n \mathbf{1}(\delta_i \notin [1 - \epsilon, 1 + \epsilon]) \leq n_0$$

Toutefois, ces méthodes ont l'inconvénient de rajouter soit une contrainte, soit une dimension au problème d'optimisation. Cela peut donc avoir un impact sur la vitesse convergence de l'algorithme génétique. Une dernière méthode possible serait donc d'éviter ce problème en enlevant successivement les n_0 lignes d'assurance dont l'impact est le moins évident de la liste des lignes à optimiser. Ainsi, en notant r une répétition de l'algorithme, les lignes considérées pour l'itération $r + 1$ seraient limitées aux suivantes :

$$lignes_{r+1} = \left\{ i \in lignes_r \mid \sum_{j \in lignes_r, j \neq i} \mathbf{1}(|1 - \Delta_i| > |1 - \Delta_j|) > n_0 \right\}$$

6.4.3 Prise en compte des synergies entre lignes d'assurance

Un facteur pour l'instant négligé au cours de cette étude est l'existence de synergies entre certaines lignes d'assurance. En effet, en pratique, des types contrats d'assurance peuvent être fixés de manière conjointe, ce qui signifie que l'exposition au risque de leurs lignes d'assurance respectives évoluent de manière similaire.

Afin d'incorporer cet effet dans le programme d'optimisation, l'approche la plus simple est de considérer que leur évolution possible est identique. Ainsi, si les lignes i_1 et i_2 sont

identifiées comme ayant des synergies entre elles, cela serait modélisé de la manière suivante :

$$\delta_{i_1} = \delta_{i_2}$$

Cela reviendrait donc à enlever une dimension aux paramètres d'optimisation chaque fois qu'une synergie est identifiée.

Dans le cas où l'évolution de deux lignes est corrélée, mais pas identique, une seconde approche serait de considérer que ces lignes évoluent systématiquement dans le même sens, mais pas nécessairement du même taux. Ainsi, si les lignes i_3 et i_4 ont ce type de synergie, cela peut être pris en compte en optimisant en fonction d'un paramètre δ correspondant à la ligne "principale", ainsi que d'un paramètre α qui rend compte de la différence d'évolution entre ces deux lignes :

$$\delta_{i_4} = 1 + (\delta_{i_3} - 1) \times \alpha_{i_4}$$

De la sorte, l'optimisation se ferait avec le même nombre de paramètres que la modélisation initiale, mais les lignes identifiées évolueraient toujours dans le même sens par définition.

L'enjeu principal de cette dernière partie était donc d'évaluer les travaux effectués au cours de cette étude, de sorte à identifier si des éléments importants n'avaient pas été pris en compte, ou s'il était possible d'identifier des axes d'évolution pour une éventuelle étude ultérieure.

Il convient toutefois de noter que, si l'objectif était d'identifier les éléments les plus importants, la liste des facteurs identifiés au cours de cette partie n'est pas exhaustive. C'est par exemple le cas de l'impact qu'une politique d'augmentation ou de diminution rapide du nombre de contrats sur une ligne d'assurance peut avoir sur la réputation de l'assureur. De même, les coûts de publicité ou l'impact social d'une diminution de l'activité sur une ligne importante n'ont pas été considérés ici.

In fine, l'optimisation d'un portefeuille d'assurance est un problème qui, à l'image de l'activité même de l'assurance, peut être extrêmement complexe et nécessiter la prise en compte d'une grande variété de facteurs. De la sorte, si les travaux réalisés ici correspondent à une première étude sur le sujet, ceux-ci peuvent toujours être affermis par des recherches supplémentaires.

Conclusion et perspectives

La problématique au coeur de cette étude consistait à déterminer s'il était possible de faire évoluer l'exposition au risque des lignes d'assurance d'un portefeuille de manière à maximiser le revenu technique tout en minimisant le risque. Afin d'y répondre, les travaux effectués se sont concentrés sur un portefeuille d'AXA pour lequel les lois de probabilité des diverses variables, telles la sinistralité ou les primes perçue par ligne, étaient données sous la forme de 50 000 simulations.

Les travaux effectués se sont donc organisés en plusieurs parties :

- Tout d'abord, le rendement et le coût en capital, mesures respectivement du revenu technique et du risque, ont été définies. De la sorte, il a suffi de déterminer comment ces quantités évoluaient en fonction des variations d'exposition au risque des lignes d'assurance afin d'écrire le programme d'optimisation.
- Si le programme d'optimisation était considéré brut de réassurance, alors il aurait été relativement aisé de mesurer l'impact de la taille des lignes d'assurance à condition de poser certaines hypothèses. Toutefois, cela aurait couru le risque de biaiser les résultats en faveur des lignes qui reposent peu sur la réassurance. C'est pourquoi une attention particulière a été donnée à la modélisation de l'impact non-linéaire de la réassurance sur les différentes variables.
- Une fois le programme net de réassurance écrit, le processus d'optimisation put avoir lieu. Cependant, des solutions explicites du programme d'optimisation n'ont pas pu être identifiées, et c'est donc un algorithme génétique qui fut utilisé pour déterminer l'ensemble des optima. Ainsi, une frontière efficiente donnant l'ensemble des meilleures évolutions possibles de la taille des lignes d'assurance fut obtenue.
- L'étude s'est donc poursuivie avec une étude des résultats obtenus, notamment afin de s'assurer de la pertinence des conclusions produites. Plusieurs manières d'interpréter les sorties de l'algorithme développé ont notamment été données, afin d'illustrer le passage de la modélisation théorique à une règle de décision pratique.

Si le modèle développé lors de ces travaux répond donc à la problématique au coeur de l'étude, certaines limites sont tout de même à signaler. En particulier, les hypothèses de linéarité posées afin de quantifier l'impact de la taille des lignes d'assurance sur les variables de primes et de sinistralité peuvent ne pas refléter la réalité. De même, d'autres facteurs tels l'impact des synergies entre lignes d'assurance, les temps et coûts de l'application des résultats, ou les coûts opérationnels ont été ignorés dans le corps des travaux. Des analyses complémentaires permettraient toutefois de mieux tenir compte des imprécisions liées aux hypothèses. C'est notamment le cas des tests de sensibilité qui permettraient de quantifier l'impact des hypothèses ou des divers paramètres sur les résultats du modèle.

Cependant, construire un modèle capable de prendre en compte la totalité des effets de l'évolution de la taille d'une ligne d'assurance est en réalité une tâche pour le moins ardue, si ce n'est impossible. La prétention de l'algorithme développé lors de cette étude n'est donc pas de fournir une recommandation parfaite, mais d'informer sur la performance strictement quantitative des lignes d'assurance du portefeuille d'intérêt. C'est pourquoi les recommandations mises en avant par l'algorithme développé ne peuvent pas être appliquées de manière unilatérale.

In Fine, le but de l'approche développée ici n'est donc que d'aider au processus de décision, pas de le remplacer. Certes, des travaux supplémentaires peuvent être effectués afin d'améliorer le modèle développé ici. Cependant, c'est dans cette optique d'assistance que des potentielles évolutions de ces travaux doivent être considérées.

Annexes

A Récapitulatif des variables et notations utilisées

SCR	<i>Ratio de solvabilité requis</i>
VaR_α	Valeur exposée au risque de probabilité α
ES_α	Déficit prévu de probabilité α
δ_i	facteur d'évolution à tester sur la ligne i
$VT_i(\delta_i)$	Ratio de valeur-temps de la ligne i après application du facteur d'évolution δ_i
$RC_i(\delta_i)$	Ratio combiné de la ligne i après application du facteur d'évolution δ_i
$Primes_{i,s}$	s-ième simulation du modèle interne de la variable de primes pour la ligne i
$Frais_{i,s}$	s-ième simulation du modèle interne de la variable de frais pour la ligne i
$Sinistres_{i,s}$	s-ième simulation du modèle interne de la variable de sinistres pour la ligne i
ATT	Sinistres Attritionnels
ATY	Sinistres Atypiques
CAT	Sinistres liés aux catastrophes naturelles
r	Taux d'actualisation
$Rdmt_i$	fonction de rendement de la ligne d'assurance i
$Rdmt^{(p)}$	fonction de rendement du portefeuille total
CeC_i	fonction de coût en capital de la ligne d'assurance i
$CeC^{(p)}$	fonction de coût en capital du portefeuille total
$(\delta_{i,q})_t$	Valeur du facteur d'évolution δ_i sur le q-ième point de la frontière efficiente, à la t-ième itération de l'algorithme
$\Delta_{i,T}$	facteur d'évolution moyen de la ligne i après les T itérations de l'algorithme

B Méthode des régressions segmentées

La méthode de la régression linéaire segmentée, théorisée par Muggeo en 2003 [21], permet de déterminer en une fois non seulement la localisation des points de changement, mais aussi les paramètres linéaires de la régression.

Pour ce faire, en notant Ψ le point de changement, Y la variable d'intérêt et Z , une variable explicative, Muggeo propose la modélisation suivante :

$$Y = \alpha Z + \beta(Z - \Psi)_+$$

La pente de la régression est donc de α avant le point de changement, et $\alpha + \beta$ après. Or, en notant $(-1)\mathbb{1}(Z > \Psi^{(s)})$ la valeur dérivée de $(Z - \Psi)_+$ au point $\Psi^{(s)}$, le développement de Taylor démontre que :

$$(Z - \Psi)_+ \approx (Z - \Psi^{(s)})_+ + (\Psi - \Psi^{(s)})(-1)\mathbb{1}(Z > \Psi^{(s)})$$

Avec une égalité stricte dans le cas où $\Psi = \Psi^{(s)}$.

Dès lors, il est possible de tester le modèle suivant :

$$Y = \alpha Z + \beta(Z - \Psi^{(s)})_+ - \gamma\mathbb{1}(Z > \Psi^{(s)})$$

dans lequel α est la pente avant tout point de changement, $\alpha + \beta$ la pente après le point de changement possible $\Psi^{(s)}$, et γ le "saut" entre la pente avant $\Psi^{(s)}$ et la pente après $\Psi^{(s)}$.

γ est donc nul si et seulement si $\Psi^{(s)} = \Psi$, et il est donc possible de se rapprocher par itérations du point de changement Ψ en considérant $\Psi^{(s+1)} = \Psi^{(s)} + \frac{\gamma}{\beta}$.

En répétant ce processus par itérations pour $\Psi^{(s+1)}$ jusqu'à obtenir une valeur de γ non-différente de 0 de manière significative, il est donc possible d'estimer à la fois les coefficients linéaires α et β et le point de changement Ψ .

C Définition formelle de l'opérateur de croisement NSGA-II

L'opérateur de croisement suggéré par Deb et Agrawal lors de la définition de l'algorithme NSGA-II est l'opérateur *de croisement binaire simulé*, dont l'objectif est d'adapter l'opérateur *au niveau du bit* à l'ensemble des réels [3].

Fonctionnement de l'opérateur de croisement au niveau du bit

L'opérateur de croisement *au niveau du bit*, défini uniquement lorsque les individus parents sont des entiers, est le suivant :

Soit $\delta_i^{(1)}$ et $\delta_i^{(2)}$ deux individus de la génération "parent". Alors, ceux-ci seront croisés avec une probabilité $p^{(crois)}$, et engendreront dans ce cas deux "enfants" $\delta_i'^{(1)}$ et $\delta_i'^{(2)}$ définis par le processus suivant :

Tout d'abord, $\delta_i^{(1)}$ et $\delta_i^{(2)}$ sont convertis en écriture binaire. De la sorte, ils peuvent être entièrement caractérisés par les l bits $(a_0^{(1)}, \dots, a_{l-1}^{(1)}) \in \{0, 1\}$ et $(a_0^{(2)}, \dots, a_{l-1}^{(2)}) \in \{0, 1\}$ tels que :

$$\begin{cases} \delta_i^{(1)} = \sum_{k=0}^{l-1} a_k^{(1)} 2^k \\ \delta_i^{(2)} = \sum_{k=0}^{l-1} a_k^{(2)} 2^k \end{cases}$$

Ensuite, un point de croisement $l_0 \in [0, l - 1]$ est défini. Les individus enfants sont obtenus en échangeant les *bits* situés après le point l_0 entre les parents :

$$\begin{cases} \delta_i'^{(1)} = \sum_{k=0}^{l_0-1} a_k^{(1)} 2^k + \sum_{k=l_0}^{l-1} a_k^{(2)} 2^k \\ \delta_i'^{(2)} = \sum_{k=0}^{l_0-1} a_k^{(2)} 2^k + \sum_{k=l_0}^{l-1} a_k^{(1)} 2^k \end{cases}$$

Si cet opérateur est utilisé dans le cadre de paramètres discrets, c'est parce que ses propriétés sont intéressantes du point de vue de l'algorithme génétique. En particulier, il est possible de démontrer que la moyenne des individus "enfants" est systématiquement égale à celles des individus "parents". De même, la dispersion des enfants est d'autant plus importante que l'écart entre les parents est large :

$$\left\{ \begin{array}{l} \frac{\delta_i^{(1)} + \delta_i^{(2)}}{2} = \frac{\delta_i'^{(1)} + \delta_i'^{(2)}}{2} \\ \text{Les valeurs de } \left| \frac{\delta_i^{(1)} - \delta_i^{(2)}}{\delta_i'^{(1)} - \delta_i'^{(2)}} \right| \text{ sont concentrées vers } 1 \end{array} \right.$$

Fonctionnement de l'opérateur *de croisement binaire simulé*

L'objectif de l'opérateur *de croisement binaire simulé* est donc de créer un opérateur dont les propriétés sont similaires à l'opérateur *au niveau du bit*, mais qui soit défini pour les réels. Pour ce faire, cet opérateur prend en argument un paramètre de dispersion η . Puis, en reprenant les mêmes notations que pour l'opérateur binaire, les individus "enfants" sont définis de la manière suivante :

$$\begin{cases} \delta_i'^{(1)} = \frac{1}{2} (1 + \beta_i) \delta_i^{(1)} + (1 - \beta_i) \delta_i^{(2)} \\ \delta_i'^{(2)} = \frac{1}{2} (1 - \beta_i) \delta_i^{(1)} + (1 + \beta_i) \delta_i^{(2)} \end{cases}$$

où β_i est défini par :

$$\beta_i = \begin{cases} (2\mu)^{\frac{1}{1+\eta}} & \text{si } u \leq \frac{1}{2} \\ \left(\frac{1}{2(1-\eta)}\right)^{\frac{1}{1+\eta}} & \text{si } u > \frac{1}{2} \end{cases} \quad \text{avec } u \sim \mathbf{U}[0, 1]$$

En définissant l'opérateur *de croisement binaire simulé* de sorte, il est donc possible de construire un opérateur aux propriétés proches de l'opérateur binaire, mais qui soit défini pour les réels. En particulier, les propriétés de stabilité de la moyenne et d'héritage de la dispersion sont maintenues.

D Définition formelle de l'opérateur de mutation NSGA-II

Lorsque Deb et Agrawal ont défini l'algorithme d'optimisation NSGA-II en 1999, deux opérateurs de mutation ont été défini : un opérateur de mutation *au niveau du bit*, et un opérateur de mutation *polynomial* [10].

Tout comme dans le cas de l'opérateur de croisement, l'opérateur de mutation *au niveau du bit* n'est pas défini pour les paramètres non discrets. Son principe était de convertir les individus de la population à faire muter en écriture binaire, puis de faire passer chaque *bit* de chaque individu de 0 à 1 ou de 1 à 0 avec une probabilité donnée.

Toutefois, contrairement au cas du croisement, l'opérateur privilégié dans le cas des paramètres à valeurs réelles n'est pas une simulation de l'opérateur binaire, mais repose directement sur une définition polynomiale. C'est pourquoi c'est principalement sur l'opérateur *polynomial* qu'il convient de s'intéresser.

L'opérateur de mutation polynomial, appliqué à un individu d'une population à tester, est défini à l'aide de deux arguments : le paramètre d'encombrement $\eta \in \mathbb{R}$, et la probabilité de mutation $p^{(mut)} \in [0, 1]$. Ainsi, soit δ_i l'individu parent avant mutation.

À la prochaine génération, un individu muté δ'_i est alors généré avec probabilité $p^{(mut)}$, et est défini par [9] :

$$\delta'_i = \begin{cases} \delta_i + \mu_1 (\delta_i - \beta_i^{(min)}) & \text{si } u \leq \frac{1}{2} \\ \delta_i + \mu_2 (\beta_i^{(max)} - \delta_i) & \text{si } u > \frac{1}{2} \end{cases} \quad \text{avec } u \sim \mathbf{U}[0, 1]$$

où μ_1 et μ_2 sont définis par :

$$\begin{cases} \mu_1 = (2u)^{\frac{1}{1+\eta}} - 1 \\ \mu_2 = 1 - (2(1-u))^{\frac{1}{1+\eta}} \end{cases}$$

Intuitivement, μ_1 et μ_2 sont définis de sorte que $\mu_1, \mu_2 \in [0, 1]$, et que leur valeur soit d'autant plus proche de 0 que η soit élevé. De la sorte, δ'_i reste systématiquement dans l'intervalle $[\beta_{min}, \beta_{max}]$, et la mutation est d'autant plus importante que η est faible.

En pratique, la littérature indique comme valeur standard de $p^{(mut)}$ la valeur $\frac{1}{n}$, de sorte qu'une variable environ mute par individu, tandis que η peut être dans l'intervalle $[20, 100]$ sans que des différences significatives de convergence soient observées [4].

Bibliographie

- [1] Solvabilité ii. *Directive 2009/138/CE du Parlement européen et du Conseil* (2009).
- [2] ACERBI, C., AND TASCHE, D. Expected shortfall : a natural coherent alternative to value at risk. *Economic Notes*. 31 (2) (2002).
- [3] AGRAWAL, R. B., AND DEB, K. Simulated binary crossover for continuous search space. *Complex Systems* (1995).
- [4] AWAD, M., AND KHANNA, R. Multiobjective optimization. *Efficient Learning Machines* (2013).
- [5] AXA. *Document d'Enregistrement Universel 2020*. Rapport Financier Annuel, 2020.
- [6] BRAHIN, P., CHATAGNY, J., HABERSTICH, U., LECHNER, R., AND SCHRAFT, A. The essential guide to reinsurance. *Swiss Re* (2013).
- [7] BUGMANN, C. Proportional and non-proportional reinsurance. *Swiss Re* (1997).
- [8] CAUCHY, A. Méthode générale pour la résolution des systèmes d'équations simulatées. *C. R. Acad. Sci. Paris* (1847).
- [9] DEB, K., AND DEB, D. Analysing mutation schemes for real-parameter genetic algorithms. *Artificial Intelligence and Soft Computing* (2014).
- [10] DEB, K., PRATAP, A., AGARWAL, S., AND MEYARIVAN, T. A fast and elitist multiobjective genetic algorithm : Nsga-ii. *IEEE Transactions on Evolutionary Computation* (2002).
- [11] DEKKING, F. M., KRAAIKAMP, C., LOPUHAÄ, H. P., AND MEESTER, L. E. *A Modern Introduction to Probability and Statistics*. 2005.
- [12] EVERITT, B., AND SKRONDAL, A. Cambridge dictionary of statistics. *Cambridge University Press* (2010).
- [13] G. PLANTIN, AND ROCHET, J. 4. inversion of the production cycle and capital structure of insurance companies. *When Insurers Go Bust : An Economic Analysis of the Role and Design of Prudential Regulation* (2009).
- [14] JORION, P. *Value at Risk : The New Benchmark for Managing Financial Risk (3rd ed.)*. McGraw-Hill. 2006.
- [15] KHISAMOVA, G. *Optimization of profit and minimization of probability of ruin for insurer by choosing reinsurance treaty*. 2017.
- [16] KOLMOGOROV, A. *Sulla determinazione empirica di una legge di distribuzione*. 1933.
- [17] LAGRANGE, J. *Traité de la Résolution des Équations Numériques de Tous les Degrés*. 1826.
- [18] MARKOWITZ, H. Portfolio selection. *The Journal of Finance* (1952).
- [19] MITCHELL, M. An introduction to genetic algorithms. *The MIT Press* (1996).
- [20] MONIN, C. Optimisation multiobjectif de l'allocation stratégique par un algorithme génétique. *Institut de Science Financière et d'Assurances* (2014).
- [21] MUGGEO, V. Estimating regression models with unknown break-points. *Statistics in Medicine* 22 (2003).

- [22] RUBIN, H. Dictionary of insurance terms. *4th Ed. Baron's Educational Series* (2000).
- [23] SALTELLI, A. Sensitivity analysis for importance assessment. *Risk Analysis* (2002).
- [24] SHARPE, W. F. Mutual fund performance. *Journal of Business*. 39 (S1) : 119–138 (1966).
- [25] SLINKER, B., AND GLANTZ, S. Primer of applied regression and analysis of variance. *McGraw-Hill* (1990).
- [26] SMIRNOV, N. Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics* (1948).

Résumé —

Cette étude, réalisée sur un portefeuille d'assurance de AXA, a pour objectif d'élaborer un algorithme capable de déterminer la manière optimale de développer les lignes d'assurance d'une entreprise. La question est donc d'identifier de manière algorithmique les lignes d'assurance pour lesquelles il est optimal d'augmenter ou de diminuer l'exposition au risque. L'approche développée pour y parvenir consiste à utiliser les données du modèle interne afin de calculer des indicateurs de risque et de rendement du portefeuille d'assurance. De la sorte, il est possible d'adapter les travaux réalisés en analyse financière au sujet de l'optimisation de portefeuille, et donc de déterminer une frontière efficiente.

Dès lors, le corps de cette étude consiste tout d'abord à déterminer comment évoluent les indicateurs de risque et de rendement en fonction de l'évolution des lignes d'assurance. Une attention particulière a notamment été donnée à l'impact que la réassurance pouvait avoir sur ces indicateurs. Cela permet d'écrire de manière formelle le programme d'optimisation à résoudre.

Une fois ce problème écrit, sa résolution peut être effectuée à l'aide d'algorithmes d'optimisations. Dans le cadre de cette étude, c'est l'algorithme génétique NSGA-II qui a été retenu. De la sorte, la frontière efficiente, a pu être tracée.

Enfin, une attention particulière a pu être donnée à l'application pratique des résultats de l'algorithme, ainsi qu'à son périmètre d'application. Pour ce faire, la significativité et la stabilité des résultats de l'algorithme génétique ont fait l'objet d'une étude approfondie.

Mots clés : Optimisation / Portefeuille / assurance / NSGA-II / Lignes d'assurance / AXA / régression segmentée

ENSAE
5 Avenue Le Chatelier,
Campus Paris-Saclay
91120 Palaiseau